Improving Throughput and Efficiency for WLAN: Sounding, Grouping, Scheduling

Xiaofu Ma

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
In
Electrical Engineering

Jeffrey H. Reed, Chair
Vuk Marojevic
Carl Dietrich
Jung-Min Park
James Mahaney

Sep. 22nd, 2016
Blacksburg, Virginia

Keywords: Wireless local area networks, Graph theory, Optimization

Improving Throughput and Efficiency for WLAN: Sounding, Grouping, Scheduling

Xiaofu Ma

(ABSTRACT)

Wireless local area networks (WLANs) have experienced tremendous growth with the proliferation of IEEE 802.11 devices in past two decades. Wireless operators are embracing WLAN for cellular offloading in every smartphone nowadays [1]. The traffic over WLAN requires significant improvement of both WLAN throughput and efficiency.

To increase throughput, multiple-input and multiple-output (MU-MIMO) has been adopted in the new generation of WLAN, such as IEEE 802.11ac. MU-MIMO systems exploit the spatial separation of users to increase the network throughput. In an MU-MIMO system, efficient channel sounding is essential for achieving optimal throughput. We propose a dynamic sounding approach for MU-MIMO systems in WLANs. We analyse and show that the optimal sounding intervals exist for single user transmit beamforming (SU-TxBF) and MU-MIMO for given channel conditions. We design a low-complexity dynamic sounding approach that adjusts the sounding interval adaptively in real-time. Through our collected over-the-air channel measurements, we demonstrate significant throughput improvements using our proposed dynamic sounding algorithm while being compliant with IEEE 802.11ac standard.

Subsequently, we investigate the user grouping problem of downlink WLANs with MU-MIMO. Particularly, we focus on the problem of whether SU-TxBF or MU-MIMO should be utilized, and how many and which users should be in a multi-user (MU) group. We formulate this problem for maximizing the system throughput subject to the multi-user air time fairness (MU-ATF) criterion. We show that hypergraphs provide a suitable mathematical model and effective tool for finding the optimal or close to optimal solution. We show that the optimal grouping problem can be solved efficiently for the case where only SU-TxBF and 2-user MU groups are allowed in the system. For the general case, where any number of users can be assigned to groups of different sizes, we develop an efficient graph matching algorithm (GMA) based on

graph theory principles with near-optimal performance. We evaluate the proposed algorithm in terms of system throughput using an 802.11ac emulator using collected channel measurements from an indoor environment and simulated channel samples for outdoor scenarios. We show that the approximate solution, GMA, achieves at least 93% of the optimal system throughput in all considered test cases.

A complementary technique for MU-MIMO is orthogonal frequency-division multiple access (OFDMA), which will be the key enabler to maximize spectrum utilization in the next generation of WLAN, IEEE 802.11ax. An unsolved problem for 802.11ax is maximizing the number of satisfied users in the OFDMA system while accommodating the different Quality of Service (QoS) levels. We evaluate the possibility of regulating QoS through prioritizing the users in OFDMA-based WLAN. We define a User Priority Scheduling (UPS) criterion that strictly guarantees service requests of the spectrum and time resources for the users with higher priorities before guaranteeing resources to those of lower priority. We develop an optimization framework to maximize the overall number of satisfied users under this strict priority constraint. A mathematical expression for user satisfaction under prioritization constraints (scheduler) is formulated first and then linearized as a mixed integer linear program that can be efficiently solved using known optimization routine. We also propose a low-complexity scheduler having comparable performance to the optimal solution in most scenarios. Simulation results show that the proposed resource allocation strategy guarantees efficient resource allocation with the user priority constraints in a dense wireless environment. In particular, we show by system simulation that the proposed low-complexity scheduler is an efficient solution in terms of (1) total throughput and network satisfaction rate (less than 10% from the upper bound), and (2) algorithm complexity (within the same magnitude order of conventional scheduling strategy.

Improving Throughput and Efficiency for WLAN: Sounding, Grouping, Scheduling

Xiaofu Ma

(GENERAL AUDIENCE ABSTRACT)

We are now living in a world with seamlessly wireless connections. Our smart phones, tablets, personal computers, etc., enable us to hear, see and communicate with family members, friends and colleagues who could be on the other side of the earth that is thousands of miles away from us. Sharing travel photos, learning breaking news and syncing work documents, etc., can be done immediately by simply touching the screen of our mobile devices. Today's dedicated wireless infrastructures extensively broaden our view of imagining what the future world for our daily life would be.

Among all the wireless infrastructure, Wi-Fi continues to become the one that carries the most amount of digital data transmission. Because of the technology advances and Wi-Fi standardization, we have seen a dramatic reduction of the Wi-Fi device cost and increased deployment of Wi-Fi technology to cover almost every home and office as well as public areas, such as hotel, airport, and hospital. The increased diversity of devices such as smartphones, tablet, laptops, set-top boxes, media centers, televisions, and wireless display adapters requires significant improvement of both throughput and efficiency for new Wi-Fi systems.

In this dissertation, we aim to investigate the theoretical foundations and practical algorithms for the advanced wireless technologies, and take Wi-Fi as an example to demonstrate the data rate and user experience improvement. The theoretical study is critical as it can be used as a guideline for system design, and the wireless algorithms are valuable for the deployment by the commercial wireless network system.

Dedication

To my parents.

Acknowledgements

I am heartily thankful to my advisor Dr. Jeffrey H. Reed for the inspiration, guidance and support throughout my Ph.D. study at Virginia Tech. Dr. Reed encouraged me and led me through both my research and my personal goals, and afforded me the opportunities to participate in exciting projects which have greatly expanded my knowledge base and professional expertise.

I would like to thank the rest of my committee: Dr. Vuk Marojevic for the insightful discussions and invaluable feedback on my academic research throughout; Dr. Carl Dietrich, my supervisor on the FRA project, for his guidance, support, and general understanding while I worked on my dissertation; Dr. Jung-Min Park and Dr. James Mahaney for providing me with useful suggestions and alternative points of view.

I extend great gratitude to Dr. Haris Volos, Dr. Taeyoung Yang and Dr. Qinghai Gao's guidance and help on both my research works and career planning during the important stages in my Ph.D study.

I sincerely thank all my colleagues in Wireless@Virginia Tech Group, and all my friends at Blacksburg with whom I have experienced such a wonderful journey of life for the last five years.

Last but not least, no word can express my deep gratitude for my parents and family for believing in me, and for their love.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

Introduction

Various wireless devices, infrastructure and applications has been enabled thanks to the unlicensed spectrum utilization. Starting from 2.4 GHz industrial, scientific, and medical (ISM) band, then the 5 GHz unlicensed national information infrastructure (U-NII) band, to the millimeter-wave band in 60 GHz, the Federal Communications Commission (FCC) has approved multiple radio bands for unlicensed commercial use. Most recently in 2014, the FCC agreed to release more spectrum in the 5 GHz band for the increasing wireless device users and applications.

The unlicensed band utilization in ISM band stimulates the boost of wireless innovations with various wireless devices such as Wi-Fi, ZigBee [2], Bluetooth [3] and cordless phones, etc, while the 5 GHz is mainly used by Wi-Fi systems with less cross-technology congestion. The success of the unlicensed band also attract the participation of the cellular systems to improve cellular network efficiency and capacity by developing new techniques, such as LTE-unlicensed (LTE-U) technology [4] and MulteFire technology [5]. Among all the technologies in the unlicensed band, the Wi-Fi system is the dominated one, and it is predicted that by 2019, more than 60% of the data traffic in the U.S. will be carried over WLAN [6].

1.1 Overview of Wireless Local Area Networks

Since 1985 when the FCC allowed the use of experimental ISM bands for the commercial application, the wireless local area network (WLAN) market has achieved tremendous growth in the last two decades with the proliferation of IEEE 802.11 devices. IEEE 802.11 working group approves many amendments, particularly for higher speed physical layer transmission, namely IEEE 802.11a, 11b, 11g, 11n, 11ac, etc., for enhancement of higher level service support like 11e and 11i, for wireless access in vehicular environments, 11p, and for mesh networking, 11s. The list of most popular past and present 802.11 amendments [7] is shown in Table 1. The other recent and on-going IEEE 802.11 [7] amendments are listed in Table 2.

Table 1 Major IEEE 802.11 amendments.

802.11 Amendments	**Major Features**
802.11a (1999)	Up to 54 Mb/s in the 5 GHz unlicensed band. Orthogonal frequency division multiplexing (OFDM) in physical layer.
802.11b (1999)	Up to 11 Mb/s to in the 2.4 GHz unlicensed band High rate direct sequence spread spectrum (HR/DSSS) in physical player.
802.11d (2001)	Assist the development of WLAN devices to meet the different regulations enforced in various countries.
802.11c (2001)	Bridge standard to incorporate bridging in access points or wireless bridges.
802.11f (2003)	Describe an Inter Access Point Protocol (IAPP) that enables user roaming in different access points among multivendor systems
802.11g (2003)	Up to 54 Mb/s in the 2.4 GHz unlicensed band.

802.11h (2003)	Solve the coexistence problem of satellites and radar that both use the same 5 GHz frequency band. Provide dynamic frequency selection (DFS) and transmit power control (TPC).
802.11i (2004)	Specify WLAN security mechanisms to replace the previous Wired Equivalent Privacy (WEP).
802.11j (2004)	Designed for Japanese market in the 4.9 to 5 GHz band to conform to the Japanese regulations.
802.11k (2008)	Enable radio resource management that improves the way traffic is distributed in the network.
802.11e (2005)	Provide a set of Quality of Service (QoS) enhancements for WLAN applications through the MAC layer modifications.
802.11r (2008)	Define the fast and secure handoffs from one base station to another in order to permit continuous connectivity.
802.11n (2009)	Up to 100 Mb/s to in the 2.4 GHz or 5 GHz unlicensed bands. MIMO technology in physical layer.
802.11p (2010)	Define wireless access in vehicular environments (WAVE) for applications such as toll collection, vehicle safety services, and commerce transactions via cars.
802.11s (2011)	Enable both broadcast/multicast and unicast delivery in the multi-hop topologies. Defining the wireless mesh network.
802.11u (2011)	Add features that improve interworking with external networks.
802.11v (2011)	Define wireless network management that enables devices to exchange information about the network RF environment and network topology.
802.11y (2008)	Enables high powered data transfer equipment to on a co-primary basis in the 3650 to 3700 MHz band in the U.S.
802.11z (2010)	Extensions to direct link setup.

Table 2 Recent and on-going IEEE 802.11 amendments.

Amendment	Band	Major Features
802.11aa (2012)	2.4, 5 GHz	Robust streaming of audio/video transport streams.
802.11ac (2014)	5 GHz	Very high throughput WLAN. MU-MIMO in physical layer.
802.11af (2014)	470-790 MHz (EU)	WLAN in the TV White Space.
802.11ah (2016)	54–72, 76–88, 174–216, 470–698, 698–806, 902–928 MHz (US)	WLAN for the sub 1 GHz license exempt operation.
802.11ax (2019)	863–868MHz (EU), 755–787 (China) 916.5–927.5MHz (JP), 2.4, 5GHz	High efficiency WLANs
802.11ae (2012)	2.4, 5 GHz	Efficient frame management
802.11ad (2012)	57.05–64 GHz (US), 57–66 GHz (EU) 59–62.90 GHz (China), 57–66 GHz (JP)	Very high-throughput WLAN in the 60 GHz band
802.11ai (2016)	-	Fast initial link setup
802.11aj (2016)	45, 59–64 GHz	WLAN in the Milli-Meter Wave in China
802.11aq (2016)	-	Pre-association discovery
802.11ak (2017)	-	Enhancements for transit links within bridged networks
802.11ay (2017)	-	Enhancements for Ultra High Throughput in or around the 60 GHz Band
802.11az(TBD)	-	Next Generation Positioning

Because of the technology advances and IEEE 802.11 standardization, we have seen a dramatic reduction of the WLAN device cost and increased deployment of WLAN technology to cover almost every home and office as well as public areas, such as hotel, airport, and hospital. The increasing high data rate provided by the WLAN enables new applications to emerge. Audio and video streaming through wireless connection are becoming more and more popular with consumers. Tablet, laptop and phone manufactures desire their end user devices to interact wirelessly to local services or peripherals such as video cameras. Service providers want to support multiple high definition (HD) streams in homes, offices, campuses or in hospitals. The increased diversity of devices such as smartphones, tablet, laptops, set-top boxes, media centers, televisions, and wireless display adapters requires significant improvement of both throughput and efficiency for new WLAN systems.

We classify the major WLAN-related research efforts in literature into different areas: (1) high-throughput WLANs; (2) LTE-U and WLAN coexistence; (3) Cellular/WLAN interworking; (4) M2M communications; (5) WLANs operating in TV white spaces; (6) programmable WLANs. Table 3 lists the example references.

Table 3 Major research areas in WLANs.

Category	Reference	Major contribution
High-throughput WLANs	[8]	The authors provide a testbed-based performance evaluation of IEEE 802.11ac with 91% throughput improvement vs. IEEE 802.11n.
	[9]	The authors propose an analytical model based on Markov chains to evaluate the performance of an IEEE 802.11ac access point enabling the TXOP sharing mechanism.
	[10], [11]	IEEE 802.11aa-based algorithms are proposed for supporting voice/video traffic.
WLAN and LTE-U coexistence	[12], [13]	The impact of LTE-U on WLAN performance is investigated based on the probabilistic and numerical analyses
	[14]	The coexistence mechanisms are proposed, including static muting, listen-before-talk (LBT), and other sensing-based schemes that make a use of the existing WLAN channel reservation protocol.
	[15]	The authors investigate the fairness issue for WLAN and LTE-U coexistence, and analyse the performance of proportional fair allocation which can be implemented without explicit coordination among the different networks.

Cellular/WLAN interworking	[16], [17]	Interworking solutions, such as roaming, have already been investigated between cellular and Wi-Fi.
	[18]	Stochastic geometry has been adopted to predict the performance bounds of the heterogeneous networks which cellular and WLAN systems operate parallelly.
	[19]	The offloading performance of the femtocells and Wi-Fi are evaluated in an example indoor scenario.
M2M communications	[20]	A feasibility study of the system coverage and throughput for an IEEE802.11ah WLAN is provided based on extensive simulation.
	[21]	The throughput boundary under different transmission power levels and data rates are presented for M2M WLAN systems.
	[22]	The trade-offs between power consumption and system delay are investigated with analysis of the impact of the network size.
WLANs operating in TV white spaces	[23], [24]	Those works focus on the mutual impacts of different technologies operating in the same bands for WLAN systems.
	[25]	Radio Environment Maps (REM) or White Space Maps (WSM) are proposed to predict the interference between devices.
	[26]	The basic IEEE802.11af communication schemes of using spectrum are presented for more accurate sharing of the radio resources.
Programmable Wireless LANs	[27]	The authors design a software defined platform, OpenFlow, which allows flexible control for heterogeneous switches and routers.
	[28]	An SDN demonstration, ODIN, was introduced as a programmable enterprise WLAN framework, which can support a wide range of reconfigurable functionalities, such as authentication, authorization, policy, mobility control, and load balancing.

In order to achieve the very high throughput (VHT) WLAN towards the gigabit era and beyond, multiple antenna techniques have become more and more crucial and indispensable. In the next subsection, we will present the overview about the multiple antenna techniques, the WLAN standard efforts, and the state-of-art in literature on multiple antenna WLANs. This discussion helps set the context for the innovations presented in this dissertation.

1.2 Multiple Antenna WLAN

1.2.1 Why Multiple Antenna Techniques in WLAN?

It is widely known that multiple antenna communication systems outperform single antenna systems with respect to channel capacity [29], [30]. In many cases, however, the number of antennas in a multiple antenna system is limited by the size and cost constraints of the end user devices. Although the cost of implementing multiple antenna systems has dropped thanks to

technological advances, energy-efficiency and form factor requirements still leave size as a major challenge for multiple antenna system deployments.

A typical multiple antenna system in a WLAN includes one access point (AP) and several mobile stations. APs bridge traffic between mobile stations and other devices in the network. The APs have been MIMO capable since the release of IEEE 802.11n at 2009 [31]. The new standard IEEE 802.11ac [32] further adds multi-user MIMO (MU-MIMO) as the key technology to improve system throughput, and these systems are starting to be widely deployed. In a beamforming system, transmit precoding techniques [33], [34] enable a multiple antenna transmission system to steer one single or multiple spatial streams to one or multiple end-users simultaneously sharing the same frequency band. In the mode of single user transmit beamforming (SU-TxBF), the transmitter focuses the energy towards one certain direction. Compared with the omnidirectional transmission, SU-TxBF results in a higher signal-to-interference-plus-noise ratio (SINR) at the target receiver, enabling improved transmission data rate. When it comes to multi-user MIMO (MU-MIMO) enabled beamforming system, multiple spatial streams are transmitted simultaneously and directed to multiple, spatially separated receivers. And it has been proved theoretically that the network capacity of the MU-MIMO system increases linearly with the number of transmit antennas or the number of receive antennas, whichever is lower [35], [36]. The MU-MIMO system throughput improvement has also been demonstrated empirically in [37], [38], [39]. MU-MIMO system relaxes the requirements on the end user devices compared to traditional MIMO system because the end user devices do not necessarily need multiple antennas. In other words, the benefits of a multiple antenna system are achievable through MU-MIMO even in a network with single-antenna end user devices, because the energy can be focused at the AP.

Therefore, MIMO and MU-MIMO Techniques, adopted by the IEEE 802.11 standard, is one of the most crucial techniques that lead WLANs towards the very high throughput era. Industrial and academic efforts have made to explore the benefit of applying multiple antenna techniques to WLAN. The standardization efforts at PHY layer advance the MIMO capacity in the WLAN systems, whereas the research works at MAC layer focus on the effective radio resource management in the multi-user systems.

1.2.2 Standardizations for Multiple Antenna WLAN

MIMO comes to WLAN starting from IEEE 802.11n which achieves the peak throughput above megabit. The first generation of products typically operate in the 2.4 GHz band only, with up to two spatial streams and 40 MHz channel width. For robustness, devices began employing STBC. Due to the number of sub-options with TxBF, it was difficult to select a common set of sub-options to certify. This greatly limited the deployment of TxBF in IEEE 802.11n devices. The more practical and more flexible multiple antenna techniques are adopted by IEEE 802.11ac and will be combined with other PHY layer techniques, such as OFDMA in IEEE 802.11ax.

(1) VHT WLAN -- IEEE 802.11ac

As an evolution of IEEE 802.11n, IEEE 802.11ac (so called Very High Throughout WLAN) adds 80 MHz, 160 MHz, and non-contiguous 160 MHz (80 + 80 MHz) channel bandwidths for transmission [40], respectively, to the standard with the goal of improving throughput. The channel allocation for IEEE 802.11ac is shown in Fig. 1.

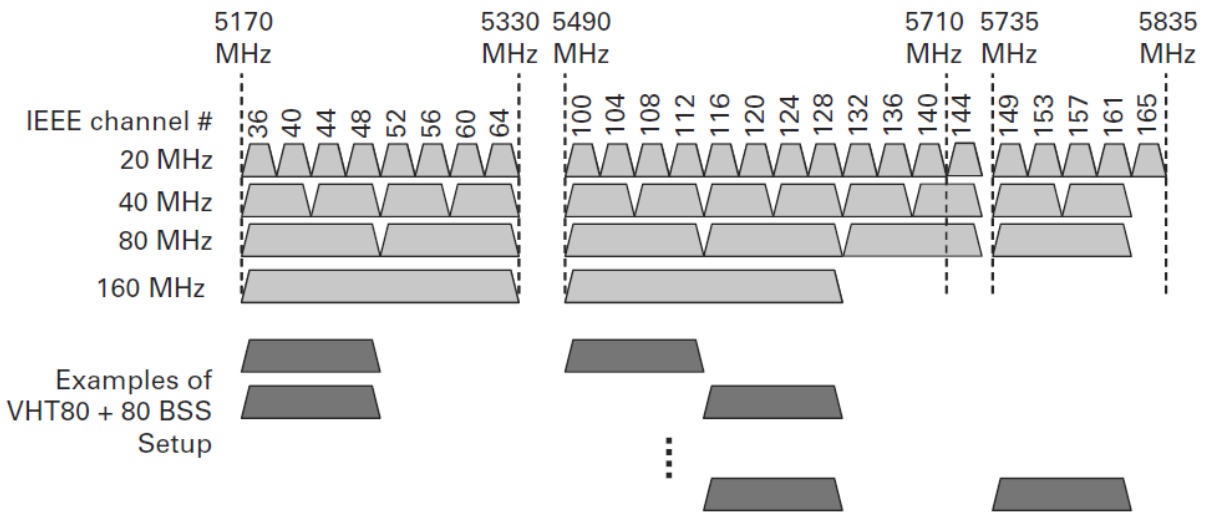


Fig. 1 Channel allocation for IEEE 802.11ac [40].

Another major throughput enhancement feature is the adaptation of the downlink MU-MIMO. In addition, the modulation constellation size increased from 64 QAM to 256 QAM with the evolution of IEEE 802.11n to IEEE 802.11ac. To support higher data rates, the packet aggregation size limit increased in IEEE 802.11ac. Previously in IEEE 802.11n, transmit beamforming had many options which made interoperability across manufacturers a very difficult task. In IEEE 802.11ac, implicit feedback is not supported for channel sounding, and transmit beamforming is limited to only the explicit feedback mechanism. Channel sounding for transmit beamforming is limited to Null Data Packet (NDP) [40].

(2) High Efficiency WLAN (HEW) -- IEEE 802.11ax

IEEE 802.11ax (so called High Efficiency WLAN) standard goes beyond throughput enhancements of 802.11ac with new techniques that enable more efficient utilization of the existing spectrum. This standard is still under development with submission of a draft ballot expected in July 2017. It is anticipated that actual commercial release will take place in late 2019 at the earliest [41]. A goal of IEEE 802.11ax is to support at least one operation mode that supports

at least four times improvement in the average throughput per station in a dense deployment scenario over IEEE 802.11ac [41]. Backward compatibility and coexistence with legacy IEEE 802.11 devices will be guaranteed. IEEE 802.11ax focuses on improving metrics that better reflect user experience instead of simply improving aggregate throughput. Improvements will be made to support environments such as wireless corporate office, outdoor hotspot, dense residential apartments, and stadiums [42]. The basic technology parameters and features of IEEE 802.11ac and IEEE 802.11ax are listed in Table 4.

Table 4 Basic Technology Parameters and Features for IEEE 802.11ac and 802.11ax.

	IEEE 802.11ac	IEEE 802.11ax
Frequency Bands	5 GHz	2.4 GHz and 5 GHz
Bandwidth (MHz)	20, 40, 80, 80+80, 160	Unlikely to change from 802.11ac
Modulation and coding schemes	1/2 BPSK, 1/2 QPSK, 3/4 QPSK, 1/2 16-QAM, 3/4 16-QAM, 2/3 64-QAM, 3/4 64-QAM, 5/6 64-QAM, 3/4 256-QAM, 5/6 256-QAM	Unlikely to change from 802.11ac
Number of spatial streams	1, 2, 3, 4, 5, 6, 7, 8	Unlikely to change from 802.11ac
Channel sounding	Explicit feedback	Unlikely to change from 802.11ac
Advanced technologies	Frame aggregation, Transmit beamforming Downlink MU-MIMO	Frame aggregation, Transmit beamforming Downlink and uplink MU-MIMO Downlink and uplink OFDMA, Full Duplex

1.2.3 Research Efforts for Multiple Antenna WLAN

Considerable research efforts have been made to explore the multiple antenna benefit across different layers of WLANs. Studies on PHY layer have mainly focused on the design and evaluation of signal processing, such as the transmit beamforming [43], [44], decoding approaches

[45], [46], [47], and MU-MIMO-OFDM peak-to-average power ratio (PAPR) reduction [48], [49]. In the multiple user access scenario, the drop in efficiency with increasing data rate usually comes from the frame overhead and inefficient resource management with poor scaling of throughput above the MAC with PHY data rate. Recent research efforts thus mainly focus on seeking for significant system performance gain in throughput.

System level evaluations and demonstrations have been made for MU-MIMO WLAN. In [50], the authors proposed a simulation framework for analysing the benefit of IEEE 802.11ac by using MU-MIMO OFDM. In [51], the authors evaluate the main features of IEEE 802.11ac in a fully connected wireless mesh network using an analytic model and simulations. The benefit of distributed antenna systems was evaluated in [52] to advance IEEE 802.11ac Networks. Particularly for precoding approaches, the performance comparison has been conducted in [53] for IEEE 802.11ac systems. An experimental evaluation of multi-user beamforming was presented in [37] by implementing a customized multiple antenna testbed. A real-time capture and analysis tool for MU-MIMO channel state information was demonstrated in [54] by using an FPGA-based platform.

Packet aggregation and fragmentation have been investigated as one way to improve system throughput. In [55], the authors evaluated the performance of frame aggregation in the IEEE 802.11ac system with fixed aggregation size and limited buffer length. The authors of [56] investigated the effects of space-time block coding (STBC) and MU-MIMO with a packet aggregation, and proposed the packet aggregation for users with multiple addresses. Packet fragmentation was demonstrated in [57] to eliminate the waste of medium resources in terms of meaningless packet pads in IEEE 802.11ac systems.

The antenna design is also crucial for MU-MIMO system performance. In [58], the authors investigated the performance MU-MIMO system as the function of the spacing and the number of access point's uniform circular array (UCA) antenna elements in an LOS environment. Different optimal antenna spacing was derived for different number of antenna elements. The mutual coupling issue was discussed in [59], and the authors proposed an efficient coupling reduction scheme for WLAN MIMO applications at 5.8 GHz. A dual-band dual-polarized compact bowtie dipole antenna array was designed in [60] for anti-interference as well as flexible beam switching in MIMO WLAN systems. While beamforming manipulates the signal timing, BeamFlex [61] was developed to alter the signal's phase and control the antenna pattern that transmits the beamformed signal. Particularly, the access point can automatically try different combinations of antenna elements to create optimized antenna tuned yield the highest possible system data rates.

Varies of CSI acquisition and compression methods have been investigated as a key for reducing inefficient transmission overhead. The limitation of limited CSI is described and evaluated in [62], [63]. Although only explicit sounding (with explicit CSI feedback) is allowed for exchanging the CSI information in WLAN, the work of [64], [40] compared the feedback methods in WLAN to demonstrate the pros and cons of the explicit and implicit channel sounding (without CSI feedback). The effect of CSI compression on the feedback overhead is investigated in [65], and the adaptive compression was proposed for improved throughput in [66]. The authors of [67] proposed a novel pilot and unitary matrix design to reduce the training frame length. In [68], the authors designed the codebook methods in which multiple codebooks were prepared between the transmitter and receiver such that only the index number was required instead of the CSI feedback.

In order to achieve the full MU-MIMO potentials for the end user, resource control in MAC layer plays an essential role. With the adoption of the MU-MIMO transmission in the 802.11ac amendment, the original MAC layer function has been modified accordingly to enable very high throughput through multiple data streams. The TXOP sharing mechanisms [69], [70] which aim at achieving better Quality of Service (QoS) for WLANs, becomes the most significant change of the basic MAC function of the current 802.11 standard [71]. It has been implemented in all manufactured Wi-Fi products nowadays as the fundamental mechanisms for VHT WLAN. The TXOP sharing mechanism extends the original Distributed Coordination Function (DCF) and allows primary channel access such that there is no limitation of the amount of packets that can be transmitted within the bounded TXOP limit period. Many resource management schemes considering the TXOP sharing mechanism have been proposed in literature. The authors of [72] investigated the MU-MIMO transmission in WLAN by extending the MAC protocol with training functionalities to support efficient multi-user transmission. In [73], an efficient and heuristic MU-MIMO transmission method was proposed and compared with the beam-forming (BF) based approach. A simple scheme for switching MU-MIMO and single user transmission was presented in [74]. Cluster-based resource control schemes [75] enabled efficient multiuser MIMO transmissions in WLANs with the evaluations of impact of different clustering methods on throughput. The authors of [76] jointly consider scheduling and frame aggregation for IEEE 802.11ac system by designing an ant colony optimization heuristic to solve it. Fairness issues were studied in [77], and a multiuser proportional fair scheduling scheme was proposed to schedule multi-user transmissions while providing a high degree of fairness.

Large scale Wi-Fi deployment is another focus for practical scenarios. For example, the enterprises such as corporate offices, hotels, airport, university campuses, hospital, and municipal

downtowns, typically deploys hundreds to thousands of APs inside buildings and across campuses. As a practical and efficient approach to solve the complex radio resource management problem, almost all major WLAN vendors such as Cisco Inc, Aruba Networks, Ruckus Wireless, Meru Networks, Symbol Technologies, and Juniper Inc. adopted the centralized approach by using the WLAN controller for WLAN management. This practice of centralized management has fuelled the dramatic proliferation of WLANs in recent year [78]. The WLAN controller enables the easier frequency band assignment and interference management [79], cheaper high-density WLANs deployment [80] and lower system energy consumption [81].

This subsection serves as a comprehensive literature review for the state-of-art for the multiple antenna techniques in WLAN. And the more specific description of the related work for the research problems in this dissertation are summarized in the next section together with the problem statements and lists of major contributions. Table 5 provides the summary of the research works mentioned above.

Table 5 Multiple Antenna Techniques in WLAN.

Category	Reference	Major Features
System level evaluations and demonstrations	[50]	A simulation framework for analysing the benefit of IEEE 802.11ac by using MU-MIMO OFDM.
	[51]	Main features evaluation of IEEE 802.11ac in a fully connected wireless mesh network.
	[52]	Distributed antenna system evaluation in IEEE 802.11ac Networks.
	[53]	Precoding approach evaluation and comparison in IEEE 802.11ac systems.
	[37]	An experimental evaluation of multi-user beamforming by implementing a customized multiple antenna testbed.

	[54]	A real-time capture and analysis tool for MU-MIMO channel information by using an FPGA-based platform.
Packet aggregation and fragmentation	[55]	Frame aggregation in the IEEE 802.11ac system with fixed aggregation size and limited buffer length.
	[56]	The effects of space-time block coding (STBC) and MU-MIMO with a packet aggregation.
	[57]	Packet fragmentation to eliminate the waste of meaningless packet pads in IEEE 802.11ac systems.
Antenna design	[58]	The spacing and the number of access point's uniform circular array (UCA) antenna elements for MU-MIMO system in an LOS environment.
	[59]	Efficient coupling reduction scheme for WLAN MIMO applications at 5.8 GHz.
	[60]	Dual-band dual-polarized compact bowtie dipole antenna array for anti-interference and flexible beam switching in MIMO WLAN systems.
	[61]	BeamFlex that was developed to alter the signal's phase and control the antenna pattern that transmits the beamformed signal.
CSI acquisition and compression methods	[62], [63]	The limitation of limited CSI.
	[64], [40]	Pros and cons demonstration of the explicit and implicit channel sounding.
	[65]	The effect of CSI compression on the feedback overhead.
	[66]	Adaptive compression for improved throughput.
	[67]	A novel pilot and unitary matrix design to reduce the training frame length.
	[68]	The codebook methods that only the index number was required instead of the CSI feedback.
	[69], [70]	The TXOP sharing mechanism which aims at achieving better Quality of Service (QoS) for WLANs

Resource control methods in MAC layer	[72]	An MU-MIMO training scheme in WLAN by extending the MAC protocol for efficient multi-user transmission.
	[73]	An efficient and heuristic MU-MIMO transmission method.
	[74]	A simple scheme for switching MU-MIMO and single user transmission.
	[75]	Cluster-based resource control schemes for efficient multiuser MIMO transmissions in WLANs.
	[76]	Joint consideration of scheduling and frame aggregation for IEEE 802.11ac system.
	[77]	A multiuser proportional fair scheduling scheme for scheduling multi-user transmissions while providing a high degree of fairness.
Large scale Wi-Fi deployment	[78]	A centralized management for the dramatic proliferation of WLANs.
	[79]	The WLAN controller that enables the easier frequency band assignment and interference management,
	[80]	A WLAN controller for cheaper high-density WLANs deployment
	[81]	A WLAN controller for lower system energy consumption.

1.2.4 Scope of Dissertation

In this work, we focus on improving two major features of the new generation of WLAN, MU-MIMO and OFDMA for the cross layer design. While there are extensive works on MU-MIMO WLAN, the current literature lacks system-level analyses that considers sounding interval as another degree of freedom to be managed and optimized. In addition, the available techniques do not provide the optimal framework to decide whether SU-TxBF or MU-MIMO should be used, and how many and which users should be assigned to an MU group. We investigate the channel

sounding and user grouping for the downlink transmission for MU-MIMO to increase the system throughput. We design a dynamic sounding algorithm and efficient user grouping algorithms that are all rigorously compliant with IEEE 802.11ac standard and can be easily integrated into the commercial WLAN systems. As for OFDMA, QoS assurance and prioritization must be considered as part of the system design for next generation of WLAN, which bring challenges for resource allocation design. Thus, we address the downlink spectrum utilization in prioritized OFDMA system and explore regulating QoS through prioritizing the users in the OFDMA WLAN for IEEE 802.11ax and other future systems.

Considering the WLAN controller is deployed widely in practice for efficient radio resource management by enabling easier frequency band assignment across neighbouring WLAN access points for avoiding interference. The research in this dissertation focus on the system of one access point with its associated users.

1.2.5 Dissertation Organization

The main body of this dissertation is organized into seven chapters discussing throughput and efficiency improvement in WLANs. Chapter 2 summarizes overviews three research problems addressed in this dissertation. For each research problem, a literature review is presented first in order to appreciate context of the research problem, followed by the problem statement. A summary of our contributions is presented for each research problem. Detailed technical contributions are presented starting from Chapter 3.

In Chapter 3, we present the technical work for the dynamic sounding for MU-MIMO in WLAN. We evaluate the trade-off between sounding overhead and throughput. A practical dynamic sounding approach is proposed for dynamic environment. At the end of the chapter, we

evaluate the proposed approach through using an 802.11ac emulator that uses our measured channel information.

In Chapter 4, we present the technical work for the user grouping for MU-MIMO in WLAN. We first presents the problem statement with the system description and derives the complexity of the exhaustive search solution. Then, we model the grouping problem using a hypergraph, and show that the maximum hypergraph matching is the optimal solution. A more efficient algorithm, which is based on graph matching, is proposed to provide the near-optimal but more computationally efficient solution. The experimental evaluation of the proposed algorithm is demonstrated at the end of this chapter.

In Chapter 5, we present the technical work for the spectrum utilization in prioritized OFDMA system in WLAN. The system description is introduced at the beginning of the chapter, and then we present an optimization framework for prioritized OFDMA system and a low-complexity scheduler for prioritized WLAN system. The evaluation results are shown to validate the theoretical work.

The last chapter concludes the dissertation and looks forward to some future research.

Chapter 2

Problem Statement and Related Work

In this chapter, we overview three research problems addressed in this dissertation. For each research problem, a literature review is presented first in order to appreciate context of the research problem, followed by the problem statement. A summary of our contributions is presented for each research problem. At the end of this section, we present the basics of the VHT Frame format, MAC header, PHY preamble, frame aggregation, and block acknowledgement mechanism, which are important concepts throughout the dissertation for understanding our major technical designs evaluations and contributions.

2.1 Research Problem # 1: Dynamic Sounding for MU-MIMO in WLANs

2.1.1 Context

The basics of MU-MIMO have been well studied by the wireless research community. With M antennas at a transmitter and N antennas at a receiver, MIMO capacity increases linearly with the min(M, N) [35]. Similarly, the system capacity of an MU-MIMO system, consisting of one M-antenna transmitter and K single-antenna users, increases linearly with minimum of M and K [82]. The capacity of MU-MIMO Gaussian broadcast channels is investigated in [83], [84], [36]. The

symmetry of uplink and downlink transmissions in MU-MIMO is studied in [85], [86]. The diversity and multiplexing gains are explored in [87], [82] for MU-MIMO systems.

Given the scenario when the number of users is larger than the number of antennas equipped at the transmitter, users need to be grouped for multi-user transmission. To achieve increased system throughput in MU-MIMO systems with limited transmit antennas, many research efforts have focused on the strategy of user grouping. The previous studies assume perfect CSI information is known by the transmitter. The scheme presented in [88] chooses users based on the estimated capacity to greedily group users to improve system throughput. Channel orthogonality can also become an effective metric to group users. For example, a heuristic approach is proposed in [89] to gather the most orthogonal users in one group and disperse highly-correlated users in different time slots. User group membership and identifiers for managing MU-MIMO downlink transmissions in WLAN are presented in [90].

Media access control (MAC) protocol design is another research area in MU-MIMO systems. An analytical model is presented in [91] to characterize the maxmum throughput of a CSMA/CA-based MAC protocol for MU-MIMO WLAN. To improve the throughput, a modified CSMA/CA protocol for MU-MIMO systems is proposed in [73], focusing on ACK-replying mechanisms with CSI assumed to be known for the MMSE precoding. A MAC protocol for downlink MU-MIMO transmissions is presented in [92] to schedule the frames in the order of arrival, assuming the CSI is obtained through the training sequence in the preamble of the control packet. A proportional fair allocation mechanism of MU-MIMO spatial streams and station transmission opportunities is adopted in [93] that assumes the AP has full knowledge of the channel. This past work has paved the road towards practical implementation of the MU-MIMO in WLAN systems; however, it does

not consider channel sounding overhead and imperfection of CSI can significantly impact MU-MIMO throughput, especially for rapidly changing channels.

The impact of CSI has been analysed for MU-MIMO in commercial wireless systems based on numerical simulation. Training and scheduling aspects of MU-MIMO schemes that rely on the use of out dated CSI are investigated in [94], showing that reduced but useful MU-MIMO throughput can be achieved even with fully outdated CSI. The channel feedback mechanisms is studied in [95] for the downlink of MU-MIMO-based FDD systems, considering user diversity and the channel correlations in both time and frequency. Recent literature also provides experimental results for MU-MIMO throughput. In [96], the system data rate of the MU-MIMO system using channel feedback is demonstrated to have much better performance than without channel feedback. The effect of CSI compression on the feedback overhead is investigated in [66]. To reduce the sounding overhead, implicit sounding (with no CSI feedback) may outperform explicit sounding (with explicit CSI feedback) with lower overhead [40]. However, implicit sounding requires more computation for both channel calibration and transceiver RF chain calibration, such as random phase and amplitude differences in RF hardware, to maintain full channel reciprocity. In addition, imperfect CSI at the transmitter can significantly degrade the system throughput of the MU-MIMO system more than that of a basic MIMO system. The standard form of channel sounding in IEEE 802.11ac is only explicit and requires the use of channel measurement frames [97]. The non-negligible overhead produced by explicit channel sounding motivates our study in finding the optimal sounding interval. The relationship of sounding interval and MU-MIMO system throughput needs both theoretical and experimental investigations. The current literature lacks system-level analyses that considers sounding interval as another degree of freedom to be managed and optimized.

2.1.2 Problem Statement

MU-MIMO is a promising technology to increase the system throughput of WLAN systems. Channel sounding strategy requires balancing channel sounding overhead with CSI inaccuracy. Less frequent sounding causes more cross-user interference because of the out-dated CSI; whereas too frequent sounding reduces the effective transmission time. The problem thus consists in finding a suitable operational point that trades the CSI accuracy for effective transmission time for the current radio environment. Practical issues for IEEE 802.11ac and future WLAN systems include:

(1) What is the optimal sounding interval in terms of the system throughput?

(2) How to dynamically adapt the sounding interval for rapidly changing radio environments?

2.1.3 Contributions

This work makes a concrete step towards advancing in channel sounding for MU-MIMO systems by providing a theoretical analysis and experimental results based on extensive over-the-air channel measurements in the indoor WLAN environment. The main contributions are summarized as follows:

(1) We characterize and evaluate the selection of the channel sounding interval for single user transmit beamforming (SU-TxBF) and MU-MIMO in WLAN. We conduct extensive channel measurements in both the static and dynamic indoor scenarios through test nodes with Qualcomm 802.11ac Chipset QCA9980. We demonstrate the existence of an optimal sounding interval for given channel conditions.

(2) We propose a low-complexity dynamic sounding approach to improve the effective throughput in the real indoor dynamic environment. We evaluate the proposed approach through using an 802.11ac emulator that uses our measured channel information. We show

that our proposed approach can achieve up to 31.8 % throughput improvement over the static sounding approach in the real indoor environment.

2.2 Research Problem # 2: User Grouping of MU-MIMO in WLANs

2.2.1 Context

In the past few years, the research efforts on MU-MIMO user selection for downlink transmission can be divided into user pairing when $N_u = 2$ and user grouping when $N_u \geq 3$. For user pairing, a low complexity solution is proposed in [98] for improved system throughput based on SINR threshold, but it cannot achieve maximal system throughput because it is a greedy approach. The maximal throughput and polynomial time solution is presented in [99], where the authors analyze the user pairing for downlink MU-MIMO with zero forcing precoding and show that the optimal solution can be found using a combinatorial optimization [100].

For user grouping, two commonly used metrics for grouping users are estimated capacity and user orthogonality, which are both calculated based on the channel states. The orthogonality-based (or Frobenius norm-based) grouping approaches, such as [89], greedily groups the most orthogonal users in one group with respect to channel-norm-related parameters and disperse highly-correlated users over different time slots. Capacity-based grouping approaches, such as [88], adopt estimated capacity as the metric to greedily group users for improving the system throughput. These user grouping strategies are designed primarily for cases where multiuser diversity offers abundant channel directions, i.e., the transmitter can find user groups with good spatial separation among users. However, the available techniques do not provide the optimal framework to decide whether SU-TxBF or MU-MIMO should be used, and how many and which users should be assigned to an MU group. These are important concerns in real environments with unpredictable channel correlations among users. Particularly for scenarios where the channel correlation is high, flexible

group sizes can be beneficial and SU-TxBF can outperform MU-MIMO. This has been demonstrated in different radio environments, especially in outdoor scenarios [101]. Since WLAN deployments in the coming years will also include both indoor and outdoor areas [1], a need arises for the WLAN APs to select beamforming modes according to the radio environment. This work, therefore, provides a framework for answering the question of how to group users for commercial WLAN scenarios. The classification of the approaches are summarized in Table 6.

Table 6 MU-MIMO User Pairing and Grouping Approaches.

Reference	Evaluation Tools	Group Category	MUD	Grouping Criterion	Optimality
Huang et al. [98]	Simulation	User pairing	MMSE	User paired based on SINR threshold	Heuristics
Hottinen et al. [99]	Simulation	User pairing	Zero forcing	User paired based on capacity	Optimum
Dimic et al. [88]	Simulation	User grouping	Zero forcing	User grouped based on capacity	Heuristics
Yoo et al. [89]	Simulation	User grouping	Zero forcing	User grouped based on channel orthogonality	Heuristics
Thapa et al. [74]	Simulation	User grouping	MMSE	User selected by instantaneous data rate	Heuristics

2.2.2 Problem Statement

In a typical WLAN scenario, it is the access point (AP) that directs simultaneous multi-stream transmissions to multiple stations. The challenge is that the number of mobile stations needing to be served can be much larger than the number of transmit antennas, N_t, at the AP. In such loaded WLAN systems, the maximum allowed number of users per multi-user (MU) group, N_u, that the AP can serve simultaneously is less than the number of mobile stations in the network. Thus, selecting sets of user(s) using SU-TxBF or MU-MIMO transmission along with scheduling all these groups over successive time slots is essential for achieving high system throughput while guaranteeing user fairness. The problem consists of deriving an optimal method for the selection of SU-TxBF or MU-MIMO transmission that maximizes the downlink system throughput with

MU-MIMO user fairness. Our objective is to develop an efficient AP transmission selection scheme and MU group assignment under the MU-MIMO air time fairness criterion. This requires dynamically assigning the stations to groups for time slotted transmission.

2.2.3 Contributions

The main contributions are summarized as follows.

(1) We model the user grouping problem using a hypergraph and show that the maximum hypergraph matching provides the optimal solution for choosing between SU-TxBF and MU-MIMO. We demonstrate an approach to determine how many users and which users to assign to the MU groups to maximize system throughput subject to MU-MIMO air time fairness.

(2) We develop an efficient and scalable algorithm based on graph matching for solving the above problem and evaluate its performance using an 802.11ac emulator that uses collected channel measurements in an office environment and simulated outdoor channles. The results show that the proposed algorithm achieves at least 93% of the optimal system throughput, which outperforms today's state-of-art algorithms.

2.3 Research Problem # 3: Spectrum Utilization in Prioritized OFDMA-Based WLANs

2.3.1 Context

IEEE 802.11ac has already achieved peak data rates at the gigabit level, whereas 802.11ax, the next generation WLAN, is investigating advanced technologies that enable more efficient spectrum utilization using metrics that better reflect the user experiences than simple throughput [102]. The continuously increasing market for WLAN and the additional competition for unlicensed spectrum operation with other systems, such as LTE-Unlicensed, requires major changes to IEEE 802.11 to continue improving the throughput and efficiency. OFDMA technology

is considered the key enabler for next generation WLANs [103]. With OFDMA, the entire channel is divided into several sub-channels where different users use different sub-channels simultaneously.

Many resource allocation algorithms for OFDMA-based systems have been proposed for 4G cellular technologies and will likely be proposed for 5G systems. Schemes have focused on trying to maximize the overall system throughput [104] or maximize the total date rates with each user having individual target rate [105], [106] as the system constraints. The maximization of the weighted sum rate and minimization of the weighted sum power consumption are investigated in [107] where the authors propose a Lagrange dual decomposition method. Resource allocation approaches focusing on the bit error rate constraint are studied in [108], [109]. The objectives of these efforts are to improve overall system performance; however, this may lead to unfair resource allocation among users, because the resource may be allocated to only a few of the most resource-efficient users in the extreme case.

Quality of Service (QoS) prioritization brings challenges for WLAN. In the hospital WLAN scenario, for instance, one major challenge of device management comes from the different devices' QoS requirements and operational priorities [110], [111]. Although the efficiency improvement of the medical systems has been demonstrated by [112], [113] through deploying the WLAN infrastructure and dedicated WLAN medical devices, the QoS assurance and prioritization are not considered in these works which only focus on the basic functionality of using mobile devices through WLAN. Particularly, life-critical biomedical transmissions should be given higher priority compared to the less important and non-time critical applications [110]. In addition, the priorities for individual patients may vary based on their different emergency levels. Another important scenario that requires QoS prioritization is the Industry 4.0 based smart factory [114].

Although WLAN technique based systems [115], [116] have been created for monitoring and controlling the manufacturing process, QoS prioritization is still a challenging task. For instance, in a manufacturing environment, the WLAN service may break down, simply because it may reach the capacity limits due to the frequent and random WLAN service requests from non-essential personal devices [117].

Thus, QoS assurance and prioritization must be considered as part of the system design for next generation of WLAN [42], which bring challenges for resource allocation design. In the dense wireless environments of the future, it is important and necessary to guarantee that radio resources are allocated to the high priority devices before addressing low priority devices. The application and user priority requirements increase the design complexity of the resource allocation algorithms, and will be growing problem as systems increase in scale.

There has been huge interest in the research community on QoS provisioning that classifies users into real-time users and non-real time users, which can be referred for WLAN protocol design. The authors in [118] divide the secondary users in the cognitive radio network into real-time users and non-real time users. Each real-time user is assigned a fixed number of sub-channels, and the supported numbers of real-time and non-real time users are determined accordingly. However, this approach is not applicable to OFDMA systems and does not manage sub-channels with more than two QoS levels. A scheduling algorithm for OFDMA system that considers user fairness is proposed in [119]. The paper proposes to assign the resources that remain unused by real-time user to non-real-time traffic, but does not address different levels of user priorities.

By setting aside a fixed amount of resources for real-time users, those users may occupy more resource than necessary, which would lead to the wasted channel allocation and unnecessary penalization of non-real time users. Since fixed availability of resources cannot satisfy all user

requirements at all times, a dynamic resource allocation algorithm capable of addressing the user requirements and priorities is needed. Furthermore, the number of priority levels needs to be more than two for practical scenarios. Therefore, we propose a framework based on a user priority scheduling criterion to prioritize the users and applications. We propose an efficient strategy to allocate resources under this user priority scheduling criterion. The goal of the prioritized WLAN system is to satisfy as many user requirements as possible insuring the user priority levels. The related works are summarized in Table 7.

Table 7 OFDMA Scheduling Approaches.

Example Reference	Evaluation Tools	Priority Considered?	Major system Considerations
Jiho et al. [104]	Simulation	No	Maximize the total data rate
Jiao et al. [105]	Simulation	No	Guarantee individual target rate
Kibeom et al. [107]	Simulation	No	Minimize the power consumption
Huiling et al. [108]	Simulation	No	Minimize the transmission bit error rate
Fraimis et al [119]	Simulation	Yes	Separate users into two priority levels

2.3.2 Problem Statement

To increase the system throughput and spectrum efficiency, orthogonal frequency-division multiple access (OFDMA) will play an important role in the next generation of WLAN systems. However, maximizing the number of the satisfied users in the OFDMA system while insuring different Quality of Service (QoS) levels is very challenging. Specifically, the problem is how to maximize the number overall number of satisfied users under strict priority constraints (guarantees for service requests of the spectrum and time resources for selected users having higher priority before accommodating those of lower priority).

2.3.3 Contributions

The major contributions are summarized as follows:

(1) We formulate a framework for resource allocation constrained by User Priority Scheduling (UPS) criterion that maximizes the number of the satisfied users while maintaining user priorities.

(2) To analyse the solution boundary, we linearize the scheduler algorithm so that it can be addressed using a mixed integer linear programing, which has well known solutions.

(3) For practical real-time implementation, we develop a low-complexity scheduler that achieves near optimal performance in most scenarios.

2.4 Background on VHT WLAN Frame Format and Aggregation

Two key standard amendments in MAC layer for performance improvements in high throughput WLAN are the frame aggregation and block acknowledgement, which are modified to make better use of PHY capability. In this section, we present the basics of the VHT Frame format, MAC header, PHY preamble, frame aggregation, and block acknowledgement mechanism. These are important concepts throughout the dissertation for understanding our major technical designs evaluations and contributions.

2.4.1 Frame aggregation

Internet data moves down from the application layer, through transport layer and network layer to the link layer based on the Open Systems Interconnection (OSI) model. The data at the Logical Link Control (LLC) sublayer comprises the MAC Service Data Unit (MSDU), which contains the data payload that includes IP packets with LLC data.

The MSDU is encapsulated with a frame header including three segments, the destination address (DA), source address (SA), and the byte number of the MSDU. The encapsulated frame is

then concatenated with up to 24 bits to align on a 4-byte word boundary. The aggregate MSDU (A-MSDU) may contain multiple such subframes as the data payload. Fig. 2 shows the A-MSDU payload.

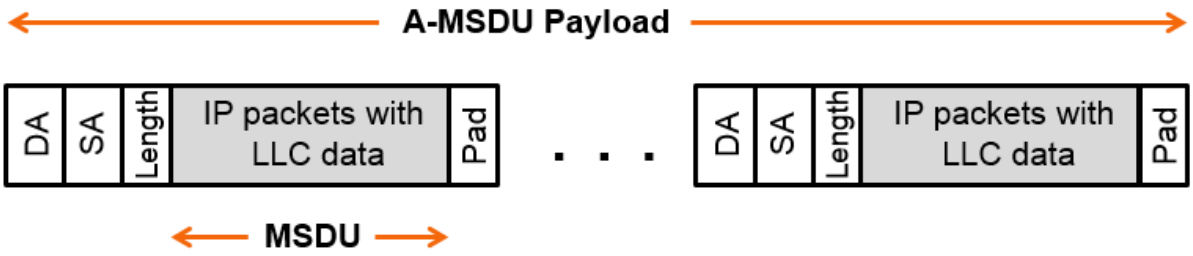


Fig. 2 A-MSDU payload.

The A-MSDU is prefixed with the MPDU header or MAC header and is postfixed with the 4-byte Frame Check Sequence (FCS). This encapsulated frame is called MAC protocol data unit (MPDU). The structure of MPDU is shown in Fig. 3, and MPDU header format is illustrated in Fig. 4.

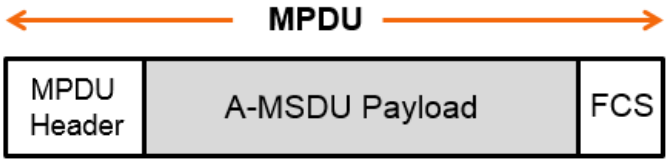


Fig. 3 MPDU structure.

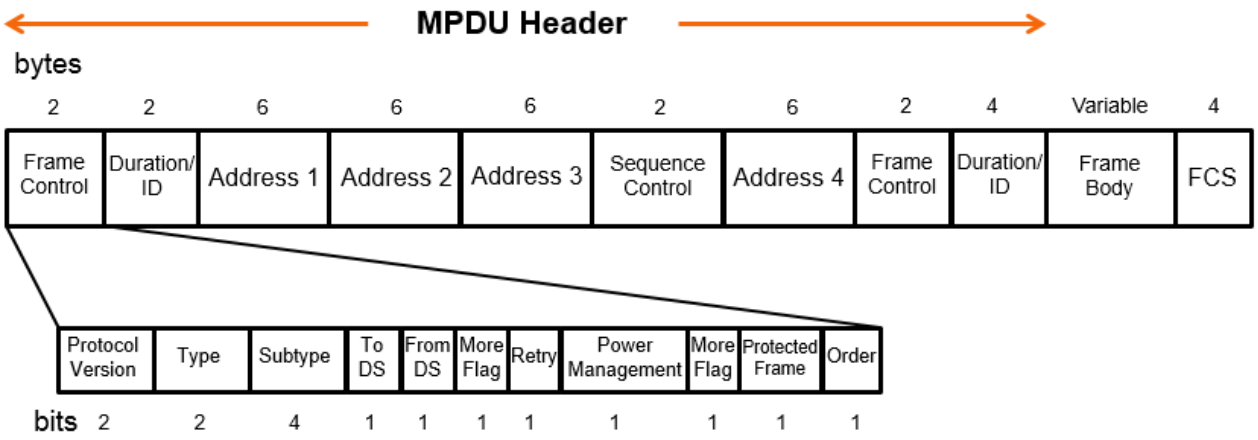


Fig. 4 MPDU header.

Multiple MPDUs are concatenated together to form an aggregate MPDU (A-MPDU), which is the frame to deliver to the physical layer as the Physical Layer Service Data Unit (PSDU). Such A-MPDU's structure is illustrated in Fig. 5. Between every two successive MPDUs, a 32-bits long delimiter is inserted with four fields for End of File (EOF), MPDU length, Cyclic Redundancy Check (CRC), and signature, respectively. A 0–3 bytes pad is postfixed to the MPDU (except for the last one) to round the subsequent delimiter to a 32-bit word boundary.

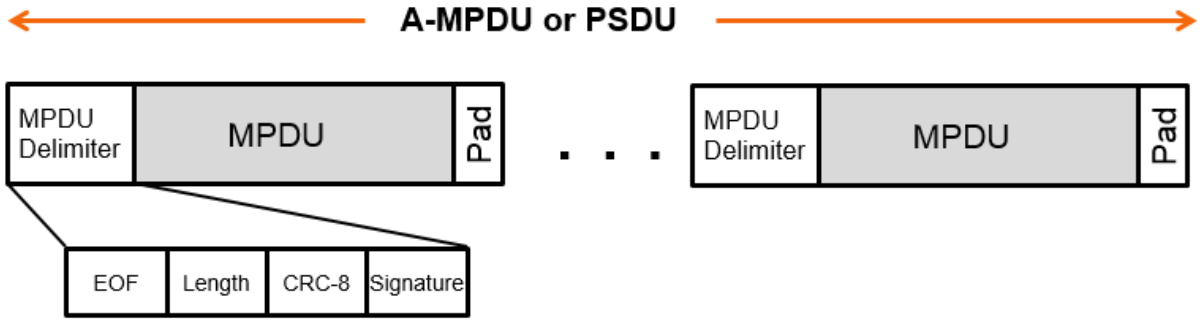


Fig. 5 A-MPDU structure.

The physical layer preamble, which includes the legacy portion and a VHT portion, is encapsulated with the A-MPDU to be a PPDU. Fig. 6 shows a complete VHT packet structure for an example MU-3 transmission, and the function for each filed is shown in Fig. 6.

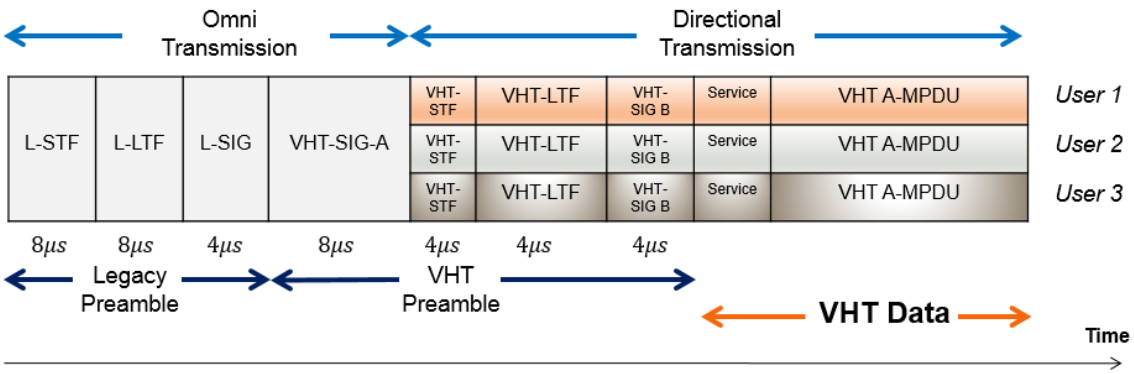


Fig. 6 VHT Packet structure for an example MU-3 transmission.

Table 8 Fields in VHT packet preamble

Field	Full Name	Major Function
L-STF	Legacy short training field	▪ Start-of-packet detection. ▪ Automatic gain control (AGC) setting. ▪ Initial frequency offset estimation. ▪ Initial time synchronization.
L-LTF	Legacy long training field	▪ More accurate time estimation. ▪ Channel estimation.
L-SIG	Legacy signal field	▪ Modulation and coding configuration. ▪ Data length setting. ▪ Parity checking.
VHT-SIG-A	VHT signal field A	▪ Bandwidth setting. ▪ Number of spatial stream configuration. ▪ Group ID control. ▪ Guard interval configuration.
VHT-STF		▪ Gain control estimation for MIMO operation.
VHT-LTF		▪ Long training sequences for MIMO channel estimation.
VHT-SIG-B	VHT signal field B	▪ Data length and the MCS configuration for single-user or multi-user modes.
Service	-	▪ Scrambler initialization.

2.4.2 Block Acknowledgement

The block acknowledgement adopts a single Block Acknowledgement (BA) frame to acknowledge the block data frame transfer. In this way, the transmission efficiency improves

because the acknowledgement is much shorter using by BA compared with the scheme that replies one ACK frame for every individual data frame.

Fig. 7 illustrates the example block acknowledgement in a session where block data transfer takes place. The access point initiates MU-MIMO transmission with A-MPDU towards multiple stations, and retrieve the BAs from stations by sending out Block Acknowledgement Request (BAR) frames to solicit an immediate BA from a specified station each time. BAR frame is not needed for the first station which is specified in the data frame, but starting from the second and subsequent stations, BAR frame for each station is required. By checking the BA, the access point prepares the MPDUs that are not expired and not acknowledged for retransmission, whereas the acknowledged MPDUs are released.

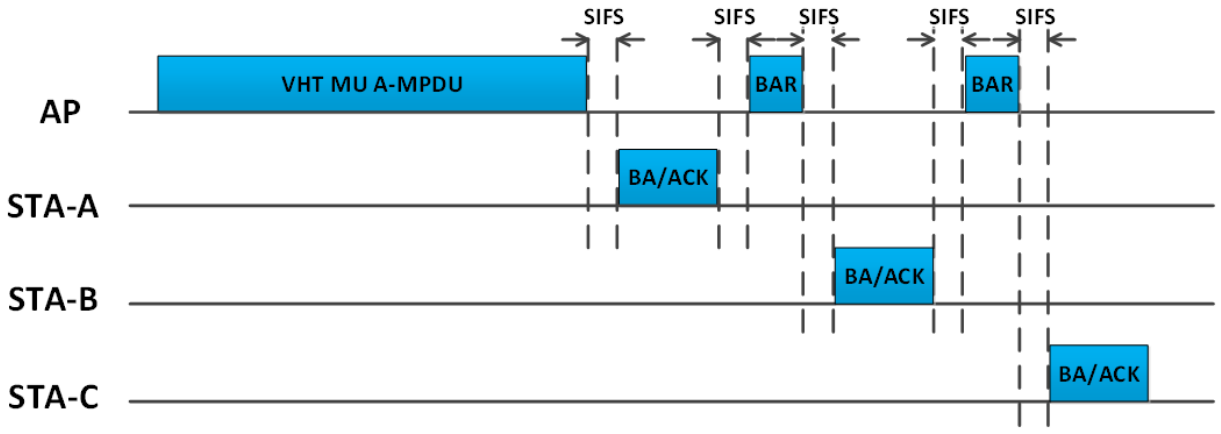


Fig. 7 Block acknowledgement mechanism.

Chapter 3

Dynamic Sounding For MU-MIMO in WLAN

In this chapter, we present the technical work for the dynamic sounding for MU-MIMO in WLAN. Our contributions are: (1) We characterize and evaluate the selection of the channel sounding interval for single user transmit beamforming (SU-TxBF) and MU-MIMO in WLAN. We conduct extensive channel measurements in both the static and dynamic indoor scenarios through test nodes with Qualcomm 802.11ac Chipset QCA9980. We demonstrate the existence of an optimal sounding interval for given channel conditions. (2) We propose a low-complexity dynamic sounding approach to improve the effective throughput in the real indoor dynamic environment. We evaluate the proposed approach through using an 802.11ac emulator that uses our measured channel information. We show that our proposed approach can achieve up to 31.8% throughput improvement in the real indoor environment.

The outline of this chapter is as follows. We evaluate the trade-off between sounding overhead and throughput in Section 3.1. Practical dynamic sounding approaches are proposed and evaluated in Section 3.2 and Section 3.3, and Section 3.4 concludes this work.

3.1 Optimal Sounding Interval

The aim of the sounding interval strategy is to maximize the system effective spectral efficiency which is defined as the total actual data rate delivered to all stations over transmission bandwidth. This effective spectral efficiency further depends on two major factors: the effective transmission percentage and the instantaneous data rate as a function of time. We emphasize the effective spectral efficiency in SU-TxBF and MU-MIMO scenarios, both of which exist in the new generation WLAN systems.

We evaluate the trade-off between the accuracy of CSI and the effective transmission time. Our goals in this section are to (1) show the existence of the optimal sounding interval which is the one among all the constant sounding intervals that leads to the highest effective spectral efficiency, and (2) analyse the dependency of this optimal sounding interval on the different system parameters.

We present the channel model in Section 3.1.1. The sounding interval optimization problem is formulated in Section 3.1.2 for SU-TxBF and MU-MIMO scenarios. Section 3.1.3 presents the evaluation of the sounding interval with an IEEE 802.11ac emulator with real channel measurements in different indoor scenarios.

Table 9 Symbols Used for Analysis.

Symbol	Description
$h(t,\tau)$	Channel impulse response in the time domain with t as the time variable and τ as the delay variable.
T_Δ	The time gap between starting time of two successive sounding operations.
T_S	The time consumed for one explicit channel sounding operation.
T_D	The time consumed for each data transmission.
Δ_t	The block length of the block fading.

$\gamma_i(T_\Delta)$	The SINR for stream i at time T_Δ after the latest sounding operation.
N_{ss}	The total number of spatial streams.
$\boldsymbol{h_{s(i)}}$	The frequency-domain channel response vector of spatial stream i in SU-TxBF case.
$\boldsymbol{w_{s(i)}}$	The streering vector for spatial stream i in SU-TxBF case.
$\boldsymbol{h_{n,i}}$	The frequency-domain channel response vector of spatial stream i for user n in MU-MIMO case.
$\boldsymbol{w_{n,i}}$	The streering vector for spatial stream i at user n in MU-MIMO case.
N_u	The number of users in a multi-user transmission group.
T_{NDPA}	Time duration for Null Data Packet Announcement (NDPA).
T_{SIFS}	Time duration for Short Interframe Space (SIFS).
T_{NDP}	Time duration for Null Data Packet (NDP).
T_{CBF}	Time duration for Compressed Beamforming (CBF) Report.
T_{BRP}	Time duration for Beamforming Report Poll (BRP).

3.1.1 Channel Model

In this work, we denote by $h(t,\tau)$ as the impulse response of the channel at time t, where τ is the delay variable. Due to the multipath fading effect, $h(t,\tau)$ can be expressed as

$$h(t,\tau) = \sum_{k=1}^{N} a_k(t)\delta(\tau - \tau_k)\ , \tag{1}$$

where N represents the number of taps, a_k represents the weight of the tap k and τ_k represents the delay on the kth tab. The impulse responses are based on the cluster model [40], and the power of the delayed responses decays linearly on a log-scale. To compute the power for each tap at each delay, the powers of the taps in overlapping clusters are summed at each delay. In the MIMO operation analysis, we mainly adopt the frequency domain representation. The system analysis is presented for only one single subcarrier for notation simplicity, but the analysis applies to any subcarrier in the frequency band. Table 9 summarizes major notations that we use in our analysis.

3.1.2 Trade-off Between Accurate CSI and Sounding Overhead

We denote by T_Δ as the time consumed for each data transmission period, which includes two parts: channel sounding and data transmission. T_Δ can also be viewed as the time gap between two successive sounding operations. We denote by T_s the operation time for a sounding operation. The value of T_s differs for different number of user(s) served by the SU-TxBF or MU-MIMO transmission. Therefore, the effective transmission time percentage is $\left(1-\frac{T_s}{T_\Delta}\right)$. We assume the channel undergoes block fading and each block lasts for a very small time Δ_t. Thus, the effective spectral efficiency $C_e(T_\Delta)$ during T_Δ can be characterized as:

$$C_e(T_\Delta) = \left(1-\frac{T_s}{T_\Delta}\right)\cdot\frac{1}{M}\cdot\sum_{m=1}^{M} C(T_s + m\cdot\Delta_t)\ . \tag{2}$$

Here, $M = \lfloor T_\Delta/\Delta_t \rfloor$, and $C(t)$ is the spectral efficiency. $C(t)$ is measured as the expected value over channel realization and transmission of $\sum_{i=1}^{N_{ss}} log_2(1+\gamma_i(t))$, where N_{ss} is the number of transmission stream and $\gamma_i(T_\Delta)$ represents the SNR of the ith spatial stream. It is obvious from (2) that $\left(1-\frac{T_s}{T_\Delta}\right)$, the first part of $C_e(T_\Delta)$, increases with respect to T_Δ. The value of the second part of the equation, $\frac{1}{M}\cdot\sum_{m=1}^{M} C(T_s + m\cdot\Delta_t)$, also changes with T_Δ.

The problem formulation for the optimal sounding interval can therefore be written as

$$max\ \ C_e(T_\Delta) \qquad s.t.\ T_\Delta \geq T_S\ . \tag{3}$$

To optimize the effective spectral efficiency with respect to the channel sounding interval T_Δ, we need to analyze the impact of T_Δ on $\gamma_i(T_\Delta)$.

We denote the number of transmit antenna as N_t. In the case of SU-TxBF using zero-forcing precoding, the SINR of the spatial stream i after t_Δ from the channel estimation time t_0 can be expressed as (without losing generality, we set t_0 to be 0)

$$\gamma_i(t_\Delta) = \frac{P_i|\boldsymbol{h}^*_{s(i)}(t_\Delta)\cdot\boldsymbol{w}_{s(i)}(t_0)|^2}{N_0+\sum_{j\neq i}P_j|\boldsymbol{h}^*_{\boldsymbol{s(i)}}(t_\Delta)\cdot\boldsymbol{w}_{\boldsymbol{s(i)}}(t_0)|^2}\ , \tag{4}$$

where P_i is the power allocated for spatial stream i, $\boldsymbol{h}_{s(i)}(t_\Delta)$ is the 1-by-N_t frequency-domain channel response vector of spatial stream i at time t_Δ, $\boldsymbol{w}_{s(i)}(t_0)$ is the N_t-by-1 streering vector for spatial stream i based on the channel estimation at time t_0, and N_0 is the channel noise power. The expression $\sum_{j\neq i}P_j\,|\boldsymbol{h}^*_{\boldsymbol{s(i)}}(t_\Delta)\cdot\boldsymbol{w}_{\boldsymbol{s(i)}}(t_0)|^2$ is the term of cross-stream interference for stream i at the receiver.

In the MU-MIMO case, when zero-forcing precoding is used for multi-stream steering, the SINR of the spatial stream i for user n after t_Δ from the channel estimation time t_0 is

$$\gamma_{n,i}(t_\Delta) = \frac{P_{n,i}|\boldsymbol{h}^*_{\boldsymbol{n,i}}(t_\Delta)\cdot\boldsymbol{w}_{\boldsymbol{n,i}}(t_0)|^2}{N_0+\sum_{j\neq i}P_{n,j}|\boldsymbol{h}^*_{\boldsymbol{n,i}}(t_\Delta)\cdot\boldsymbol{w}_{\boldsymbol{n,j}}(t_0)|^2+\sum_{m\neq n}P_{n,j}|\boldsymbol{h}^*_{\boldsymbol{n,i}}(t_\Delta)\cdot\boldsymbol{w}_{\boldsymbol{m,j}}(t_0)|^2}\ , \tag{5}$$

where $P_{n,i}$ is the power allocated for stream i for user n, $\boldsymbol{h}_{n,i}(t_\Delta)$ is the 1-by-N_t frequency-domain channel response vector of spatial stream i for user n at time t_Δ, and $\boldsymbol{w}_{n,i}(t_\Delta)$ is the N_t-by-1 streering vector for spatial stream i at user n based on the channel estimation at time t_0. The expression $\sum_{j\neq i}P_{n,j}\,|\boldsymbol{h}^*_{\boldsymbol{n,i}}(t_\Delta)\cdot\boldsymbol{w}_{\boldsymbol{n,j}}(t_0)|^2$ in (5) is the cross-stream interference at user n, and $\sum_{n\neq m}P_{n,j}\,|\boldsymbol{h}^*_{\boldsymbol{n,i}}(t_\Delta)\cdot\boldsymbol{w}_{\boldsymbol{m,j}}(t_0)|^2$ is the cross-user interference.

In the next subsection, we evaluate the transmission trade-off using our real channel measurement in different indoor environment.

3.1.3 Evaluation

3.1.3.1 Experimental System Setup

In this section, we evaluate the dependency of the optimal sounding interval on the different system parameters in a WLAN based MU-MIMO environment. We consider a typical WLAN topology with one AP and 12 stations. The AP is equipped with four antennas and each station is

equipped with one antenna. The system can serve up to three stations with downlink multi-user transmission simultaneously, which is typical for a practical WLAN scenario.

We measure the actual channel information through the four test nodes. Each test node is equipped with four antennas and a 4×4 four-stream 802.11ac Qualcomm transceiver chipset QCA9980. The over-the-air transmission were conducted over a 40 MHz channel in an office environment. One test node operates as the transmitter AP, and the other three work as the receiver stations. The 12 receive antennas are spaced apart from each other and distributed randomly in the office. Those 12 sets of channel samples are used to mimic the channel from one four-antenna AP to 12 single-antenna stations.

We conducted the channel measurements under two scenarios namely **low Doppler** and **high Doppler** scenarios. The low Doppler scenario refers to a low-mobility radio environment with no human activity or device movement during the channel measurement collection. Under the high Doppler scenario, we collect the channel samples when significant human movement exists in the measurement environment. Specifically, random rotating and shaking was conducted for each terminal. Notice that there are still random variations in either the low or high Doppler scenario.

With the measured channel information, trials with different sounding intervals can be conducted over the identical channels for comparison. We implement the measurement-driven MU-MIMO-OFDM emulator rigorously according to the IEEE 802.11ac specification. This emulator together with the over-the-air channel information are used to test the performance of the media access control (MAC) and physical (PHY) layer algorithms. Singular value decomposition (SVD) is employed as the precoding scheme, and the MMSE receiver is implement to mitigate the cross-user interference.

Table 10 Evaluation Parameters.

Parameters	Values
Maximum AMPDU Duration	2 ms
MPDU Length	1556 bytes
MSDU Length	1508 bytes
SIFS Duration	0.016 ms
Bandwidth	40 MHz
Number of OFDM Subcarriers for Data	108
Guard Interval for OFDM Symbol	400 ns

Since the focus of this work is the selection of the sounding interval, we are carefully about eliminating other possible effects on the system performance. In particular, we assume that the AP adapts the optimal modulation and coding scheme (MCS) during each simulation trial. The evaluation parameters of the IEEE 802.11ac system are summarized in Table 10.

In a single user transmission scenario, the sounding overhead of 802.11ac comes from the explicit feedback mechanism, i.e., the information exchange between the AP and the single mobile station. The procedure is described in Fig. 8. This information exchange during each channel sounding process goes like this: the AP first broadcasts a Null Data Packet Announcement (NDPA) within the system. Then, after a Short Interframe Space (SIFS) time interval, the AP sends out a Null Data Packet (NDP) to sound the transmission channel. From the station's perspective, after receiving the NDP packet from AP, it waits for a SIFS interval and responds to the AP with a compressed beamforming (CBF) report. Therefore, the time required for each channel sounding operation under a single user transmit beamforming can be expressed as:

$$T_{S\_SU} = T_{NDPA} + T_{SIFS} + T_{NDP} + T_{SIFS} + T_{CBF} + T_{SIFS} \quad , \tag{6}$$

where T_{NDPA}, T_{SIFS}, T_{NDP} and T_{CBF} are the time durations for NDPA transmission, SIFS, NDP transmission and CBF transmission, respectively.

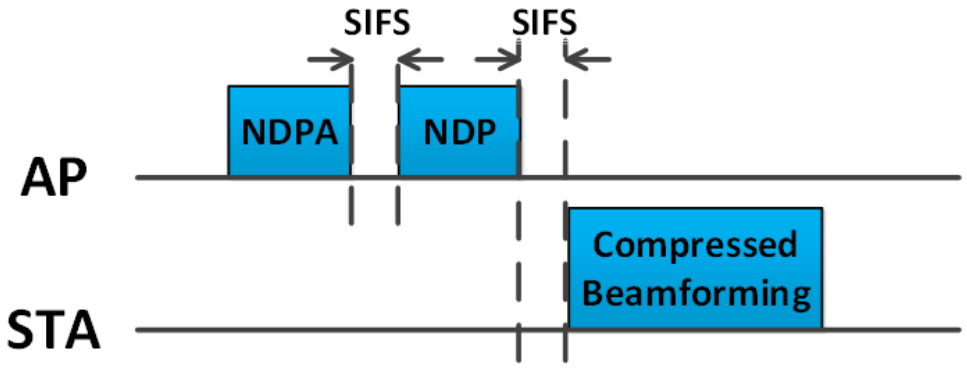


Fig. 8 Channel sounding for single-user transmit beamforming.

The sounding duration for MU-MIMO in WLAN differs from that of the single user case in that the sounding overhead includes the beamforming report poll(s) for all stations as well as a sequence of stations' CBF reports. The procedure is described in Fig. 9. The sounding procedure for MU-MIMO initiates exactly the same as the SU-TxBF sounding procedure. The AP first broadcasts an NDPA. After a SIFS time interval, the AP sends out an NDP to sound the transmission channel. However, to retrieve the channel information from each STA, the MU sounding procedure needs Beamforming Report Poll (BRP) frame(s) to collect responses from all stations one by one. Poll frame is not needed for the first station which is specified in the NDPA frame, but starting from the second and subsequent stations, BRP frame for each station is required. The AP will integrate all the received CBF reports together into a steering matrix. Therefore, the time required for one channel sounding operation under an MU-MIMO transmit beamforming can be extended from (6) and expressed as

$$T_{S\_MU} = T_{NDPA} + T_{SIFS} + T_{NDP} + T_{SIFS} + N_u \cdot (T_{SIFS} + T_{CBF}) + (N_u - 1) \cdot (T_{SIFS} + T_{BRP}) \,, \tag{7}$$

where N_u represents the number of users in the multi-user transmission group, and T_{BRP} are time duration for a BRP frame transmission.

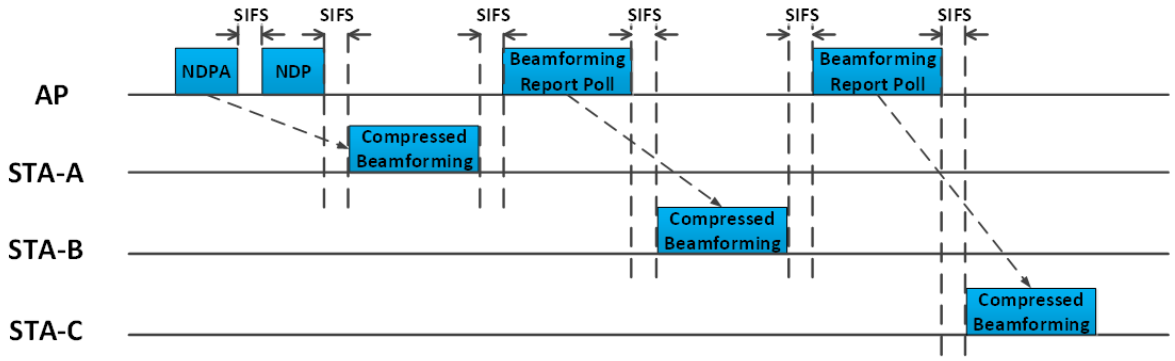


Fig. 9 Channel sounding for MU-MIMO transmission.

3.1.3.2 Evaluation Results

3.1.3.2.1 Sounding Interval vs. SINR.

Fig. 10 shows how the output SINR changes as a function of the sounding interval in the low Doppler scenario. We denote SU, MU2, and MU3 as SU-TXBF transmission, MU-MIMO transmission with 2 users, and MU-MIMO transmission with 3 users, respectively. We observe that the SINR value decreases when the sounding gap increases. This is because the CSI accuracy for beam steering degrades with the increasing time gap between the channel estimation and the data transmission. We also observe that the SINR per spatial stream decreases with the increasing number of users in the MU transmission in the system. More precisely, after 200 ms, the SINR for SU-TxBF keeps nearly the same value, while MU2 drops from 31 dB to 27 dB, and MU3 drops more significantly from 27 dB to 22 dB. MU-MIMO is more sensitive to inaccurate CSI because the spatial stream steering is done in such a way that the inter-user interference is minimized for MU transmission as opposed to maximizing the user's SINR for SU-TxBF.

In the high Doppler scenario, shown in Fig. 11, we observe that the SINR values decreases with the increasing of sounding gap and with the number of users allowed to simultaneously transmit. We find the degradation of the SINR is much more severe for the high Doppler scenario

than the low Doppler scenario. For example, at 200 ms, the SINR for MU2 and MU3 under high Doppler scenario are 14 dB and 8 dB, which are much lower than those in the low Doppler scenario. This illustrates that the impact of the fast changing channel on SINR for MU transmission is much more significant than that for SU transmission. And as the number of users in the MU transmission increases from 2 to 3 users, the SINR degradation is even more.

We evaluate the performance of SU-TxBF with a four-antenna enabled AP and a four-antenna enabled station. We adopt the IEEE Channel D [40], which is the typical channel model for the indoor environment with delay spread to be 50ms as shown in Table 11. The number of spatial streams ranges from 1 to 4.

Table 11 Evaluation Parameters.

Model	RMS Delay Spread (ns)	Environment
A	0	N/A
B	15	Residential
C	30	Residential/small office
D	50	Typical office
E	100	Large office
F	150	Large space (indoor/outdoors)

Fig. 12 plots the performance of the SU SINR per spatial stream over the sounding gap. We observe that under the same sounding gap, the SINR decreases with respect to the number of spatial streams. This indicates that double or triple the spatial streams for such a SU transmission does not necessarily double or triple the data rate, as the SINR per stream cannot maintain the same. As opposed to the MU-MIMO transmission that requires the interference cancellation to mitigate the cross-user interference which degrade the MU performance greatly, the SINR degradation for SU, which has no cross-user interference, keeps a similar gradient as the number of antenna equipped in each station increases.

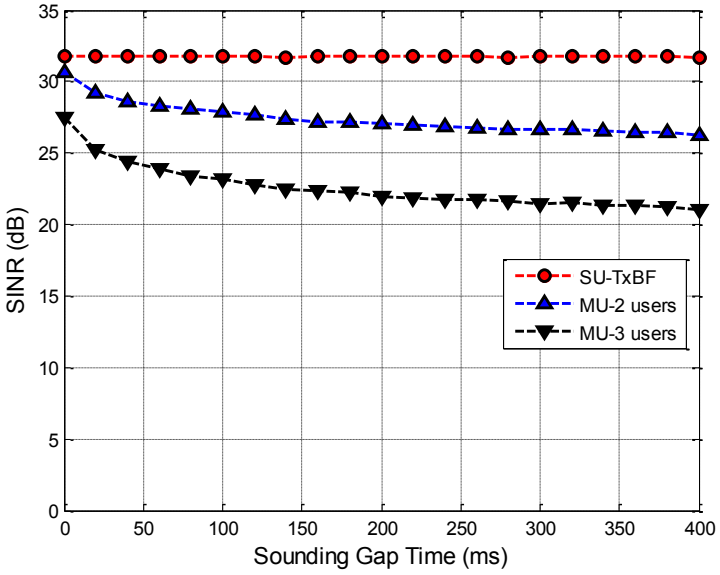


Fig. 10. SINR as a function of sounding interval in the low Doppler scenario.

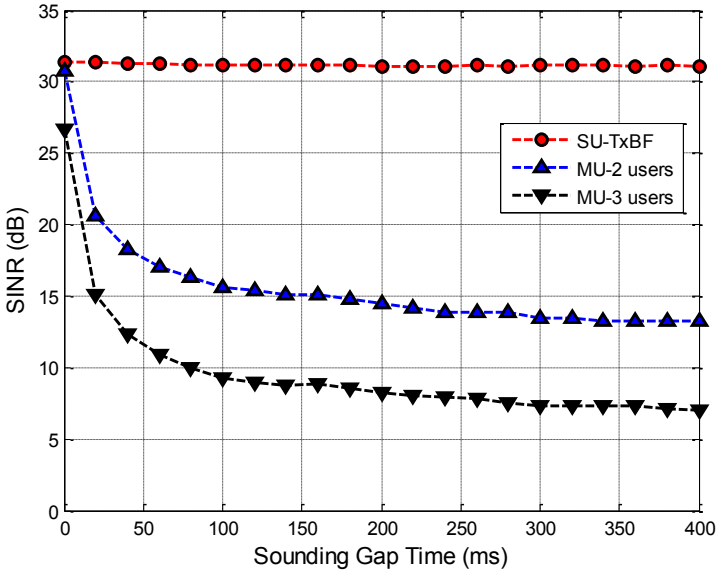


Fig. 11. SINR as a function of sounding interval in the high Doppler scenario.

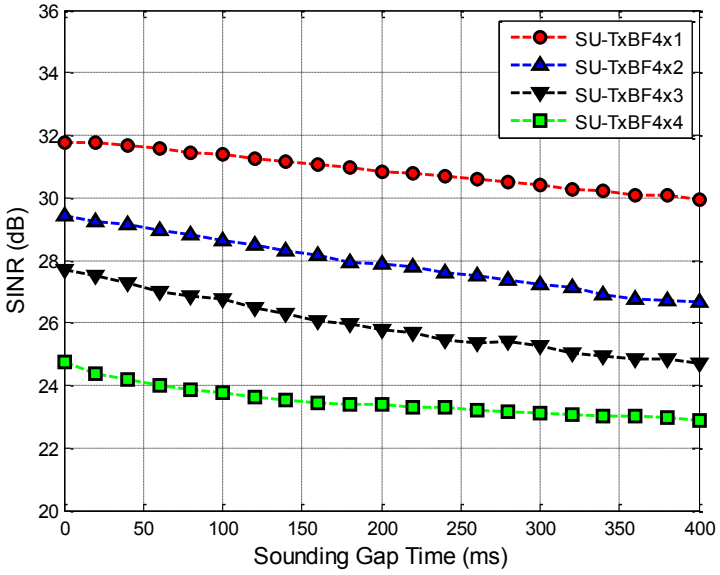


Fig. 12. SINR per stream as a function of sounding interval for SU-TxBF using IEEE channel D.

3.1.3.2.2 Sounding Interval vs. Throughput

In this section, we evaluate the impact of sounding interval on the 802.11ac system throughput in the low and high Doppler scenarios.

Fig. 13 plots the total PHY layer throughput with respect to the sounding gap for the low Doppler scenario. In particular, we compare the system throughputs under SU, MU2, and MU3 cases with sounding intervals ranging from 2 ms to 400 ms. It is observed in Fig. 4 that the total PHY layer throughput decreases monotonically under MU2 and MU3 scenarios but stays relatively the same under the SU scenario. This illustrates that the cross-user interference due to the outdated CSI significantly degrades the SINRs of the MU transmission whereas the impact of outdated CSI on the SINR under the SU scenario is not that significant. We also observe from Fig. 4 that the drop of the system throughput under the same sounding gap is enlarged by the number of users within MU transmission. For example, when the sounding gap increases from 2 to 50 ms, the PHY

throughput of MU3 drops by 18% compared to only 5.6% drop for MU2. This is because the increased number of users within an MU transmission worsens the cross-user interference.

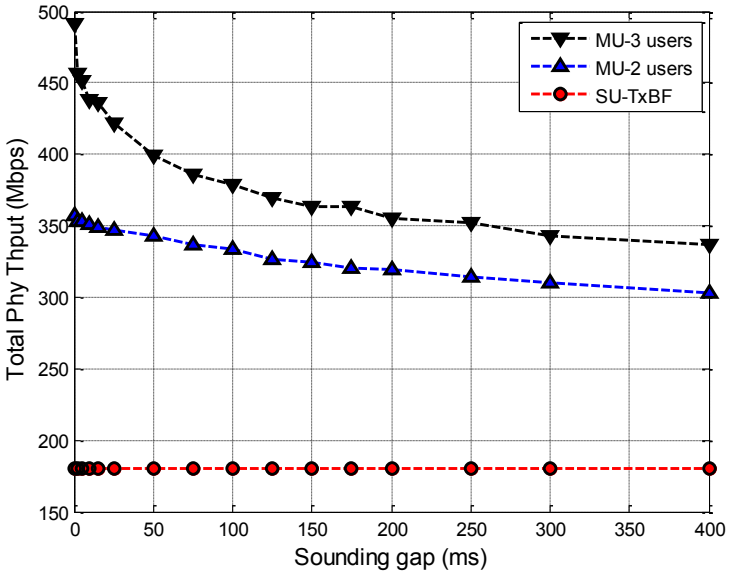


Fig. 13. Total PHY throughput as a function of sounding interval in the low Doppler scenario.

When it comes to the high Doppler scenario, the throughput drops dramatically compared to the low Doppler scenario as the sounding interval increases, as shown in Fig. 14. When the sounding gap is less than 10 ms, the PHY layer throughput of MU3 outperforms that of MU2, and both outperform SU-TxBF. When the sounding gap approaches 50 ms, the PHY layer throughputs of MU2 and MU3 drop to the same level as SU-TxBF. Beyond 50 ms, MU2 has a higher total PHY layer throughput than MU3 and both are below SU-TxBF throughput.

We observe from Fig. 13 and Fig. 14 that SU-TxBF's PHY layer throughput remains almost the same for low and high Doppler scenarios. This confirms to the SINR results shown in Fig. 10 and Fig. 11. With a sufficient high SINR for both low and high Doppler scenario, SU-TxBF can support the highest MCS, and hence the same high throughput. In contrast, the PHY layer throughputs of MU2 and MU3 clearly drop, and MU3's throughput drops faster than that of MU2

when we increase the amount of sounding gap. This can also be explained by the MU2 and MU3's SINR reduction that causes the reduced throughput. We also observe that for the same length of sounding gap, a throughput difference exists between the low and high Doppler scenarios, and this difference is more significant when the number of users within MU transmission is larger. For example, when the sounding gap is 50 ms, the throughput difference for MU2 between high and low Doppler scenarios is 165 Mbps, whereas the difference becomes 220 Mbps for MU3. When the sounding gap increases to 100 ms, the difference becomes 180 Mbps and 230 Mbps, respectively.

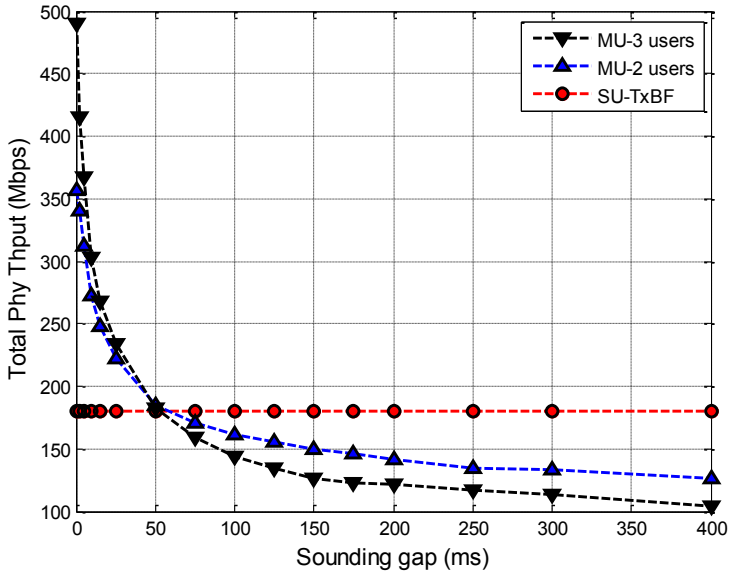
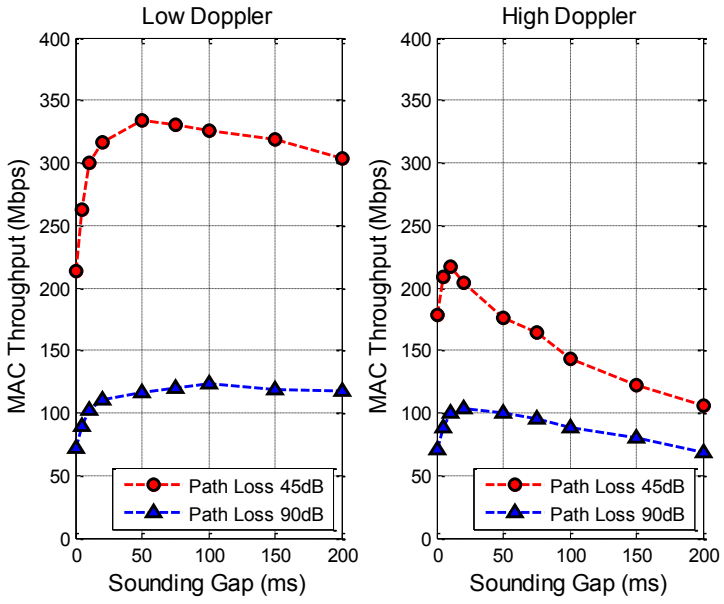


Fig. 14. Total PHY throughput as a function of sounding interval in the high Doppler scenario.

The actual throughput that can be achieved and delivered to the end users also depends on the sounding overhead. We call this effective throughput or MAC throughput. Fig. 15 shows the effective throughput as a function of sounding gap for three user MU-MIMO. As can be seen for the low Doppler scenario (Fig. 15a), when the sounding gap is small, the sounding overhead dominates the air time, which causes the low effective throughput. When the sounding frequency

decreases, the sounding overhead correspondingly drops and the throughput increases. Since the environment is very static, i.e., the channel environment is relatively stable, outdated channel information does not significant affect the throughput in the low Doppler scenario even when the sounding interval becomes as large as 100 ms. In the high Doppler scenario (Fig. 15b), similar to the low Doppler situation, the sounding overhead leads to throughput degradation when the sounding interval is set very small. In this high Doppler scenario, the best sounding interval is around 10 to 20 ms, which leads to the highest throughput. If the sounding interval keeps increasing, the total effective throughput degrades fast due to the inter-user interference caused by the rapidly changing environment and inaccurate CSI.



(a) (b)

Fig. 15. Effective throughput (MAC throughput) changes with sounding intervals under different path losses for (a) the low Doppler and (b) the high Doppler scenarios.

We observe from Fig. 15 that the optimal sounding interval varies not only under different channel variation conditions, but also under different instantaneous path losses. In particular, the effective throughout degrades as the pass loss increases in either low or high Doppler scenario.

This is reasonable as larger pass loss leads to lower SINR. Fig. 15 also shows the dependency of the optimal sounding interval on the path loss in high and low Doppler scenarios. In the low Doppler scenario, the optimal sounding interval is 50 ms if the path loss is 45 dB, whereas the optimal value becomes around 100 ms if the path loss is 90 dB. A similar phenomenon can be observed in the high Doppler scenario, where the optimal sounding interval under 45 dB pass loss is around 10 ms, which is lower than that under 90 dB pass loss (around 20ms).

In this section, we showed the existence of the optimal fixed sounding interval and its dependency on the channel variation conditions and path losses. The channel variation indicates the device or environment motion, and the path loss reflects the positions of the AP and the station. Since the motion in the environment and the positions of the stations are dynamic and unpredictable, it is necessary to design an algorithm that dynamically determines a suitable operational point that trades the CSI accuracy for the effective transmission time as a function of the real-time radio environment.

3.2 Sounding in Dynamic Environment

So far, we have investigated the determination of the optimal sounding interval in a radio environment characterized by low or high Doppler. In dynamic indoor environments, the channel conditions change unpredictably because of the random movement of human, reflectors, and wireless devices, etc. The sounding interval thus requires dynamic adaptation based on the instantaneous channel conditions. We propose a dynamic sounding approach with the capability of (1) being aware of the instantaneous channel condition, (2) inferring the suitable sounding interval from the estimated statistics of the current channel condition, and (3) adjusting the sounding interval based on internal feedback. We first presents the proposed dynamic sounding approach, followed by the evaluation results based on the collected channel measurements.

3.2.1 Proposed Dynamic Sounding Approach

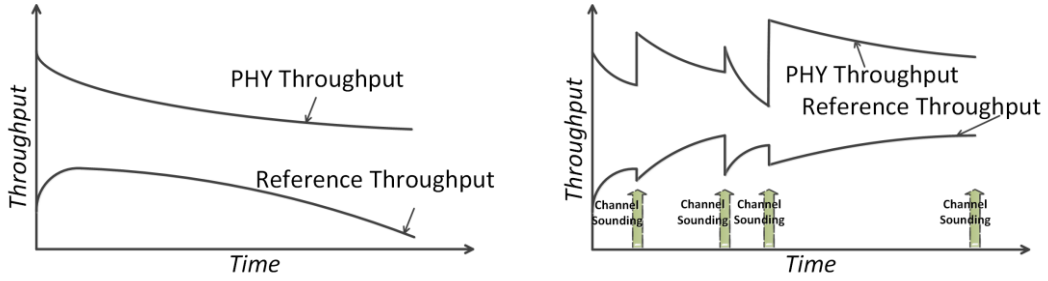
To trace the time variation of the radio environment, a Doppler spectrum profile can be maintained through the feedback CSI, and the optimal sounding interval can be estimated accordingly. However, both the Doppler spectrum profile creation and the theoretical calculation of the optimal sounding interval require a lot of computing power and time, which is very challenging for the real-time system implementations. To simplify the system design, we propose a sounding adjustment that depends on the accumulated data rate, which is easy to collect from the AP.

We consider the sounding overhead and the instantaneous transmission condition to estimate the time correlation of the channel variation and to calculate the effective throughput. We denote the accumulated effective throughput as the reference throughput, and will use this throughput to adjust the sounding interval. All 802.11ac data frames are sent in an aggregated MAC protocol data unit (AMPDU). We calculate the reference throughput after the nth AMPDU transmission since the latest sounding operation using

$$R_{TH}(n) = \frac{\sum_{m=1}^{n}\sum_{j=1}^{N_u} D(m,j)}{(T_s+\sum_{m=1}^{n} T_{AMPDU}(m))} \quad , \tag{8}$$

which is the ratio between the estimated accumulated successfully-transmitted AMPDUs during the downlink MU transmission and the time period which includes the overhead of the latest sounding operation and the time for accumulated downlink AMPDU transmissions.$D(m,j)$ stands for the successfully transmitted data amount for the mth AMPDU of user j since the latest sounding operation, and $T_{AMPDU}(m)$ stands for the time duration of the mth downlink AMPDU transmission.

In our proposed dynamic sounding approach, the AP regularly estimates the reference throughput as a function of time after the latest sounding operation. After a sounding operation, the reference throughput initially increases in general as the sounding overhead dominates the effective throughput. After a while, the reference throughput drops as the channel estimation precision decay becomes the dominant factor for reducing the effective throughput. Unless there is another sounding operation, this reduction in throughput becomes worse. This impact of timely channel estimation on the reference throughput over time is illustrated in Fig. 16 (a). Since the reference throughput indicates the accumulated MAC layer throughput since the latest sounding operation, its value is always below the PHY layer throughput. Without sounding, the PHY layer throughput cannot increase over time because of the degrading accuracy of the CSI. The basic idea of our dynamic sounding approach is to trigger the sounding operation when the reference throughput starts degrading, which is illustrated in Fig. 16 (b). After each trigger, the PHY layer throughput is expected to be pulled up to a higher value. Notice that the reference throughput is not the instantaneous channel statistic, but a metric with the value accumulated over time, so the impact of the instantaneous fluctuation of the noise on the trigger of the sounding operation is limited. The proposed approach works while there are uplink transmissions, since the reference throughput is a metric to that actually reflects the throughput drop trend of the downlink transmission over time. Table 12 provides the details of the proposed dynamic sounding approach. Since the major calculation is dominated by (8), the time complexity of the algorithm is very low and can be easily implemented in the MU-MIMO systems.



(a) **(b)**

Fig. 16. Dynamic sounding approach: (a) reference throughput over time after the latest sounding operation, (b) reference throughput over time under dynamic sounding approach.

Table 12 Algorithm # 1.

Dynamic Sounding Approach

Stage 1: Initialization

Flag for whether sounding is needed: $s = 1$.

Flag for whether the current AMPDU packet is the first one after sounding: $f = 0$.

The transmitted AMPDU index: $n = 0$.

Stage 2: MU Downlink Operation

while MU Downlink Transmission is needed

  if $s = 1$

    Operate channel sounding by sending NDP, NDPA and collect CBF(s) from STA(s).

    $f = 1$.

  end if

  Transmit AMPDU using rate adaptation algorithm.

  $n = n + 1$.

  Update $R_{TH}(n)$ using (8)

  if $(f = 1)\ or\ (R_{TH}(n-1) < R_{TH}(n))$

    Set $s = 0$ and prepare for sending the next AMPDU.

    $f = 0$.

  else

    Set $s = 1$ and prepare for sounding.

  end if

end while

3.2.2 Evaluation Results

We consider a WLAN with one AP and 12 stations. We use the 802.11ac emulator with our over-the-air channel measurements as the evaluation platform which is discussed in Section 3.1. As the system can support up to three stations for MU-MIMO transmission, we consider any three stations out of the 12 are grouped.

We evaluate the performance of the proposed approach under the collected over-the-air channel measurements, including the low and high Doppler scenarios. In addition, we mimic the dynamic scenario by alternating between samples from the low and high Doppler scenarios. Specifically, we construct the channel environment as an alternation between high and low Doppler conditions, and each segment period for low or high Doppler channel samples is set to be 50ms.

We compare the proposed dynamic sounding approach with two baseline fixed sounding interval schemes, representing the optimal fixed intervals for low and high Doppler scenarios. The exact sounding interval values for them are the same as it is in Section 3.1. We call those two schemes low Doppler approach (LDA) and high Doppler approach (HDA). Fig. 17 and Fig. 18 show that dynamic sounding approach even outperforms the approaches with the optimal fixed sounding interval solutions.

Fig. 17 indicates the time stamps for channel soundings as operated by the AP using the three schemes in an example 400 ms period of time under the alternation high-low Doppler condition. In Fig. 17, each black vertical bar stands for an occurrence of a sounding operation. The grey and white areas represent the time periods under high Doppler and low Doppler channel conditions. As expected, under both conditions, the sounding interval under HDA is always around 11ms and the average sounding interval under LDA is around 43 ms. In contrast, the sounding interval of

our dynamic sounding approach varies according to the change of the radio environment, which illustrates its channel awareness functionality. Particularly, the average sounding interval of our proposed approach in the high Doppler scenario is around 11ms, and it increases to around 41 ms per sounding operation under the slowly changing environment in the low Doppler scenario. As can be observed from Fig. 17, the operation activity of the dynamic sounding approach stays almost the same as that of LDA in low Doppler scenario and that of HDA in the high Doppler scenario. This illustrates the accuracy of the channel awareness process of the proposed dynamic sounding approach.

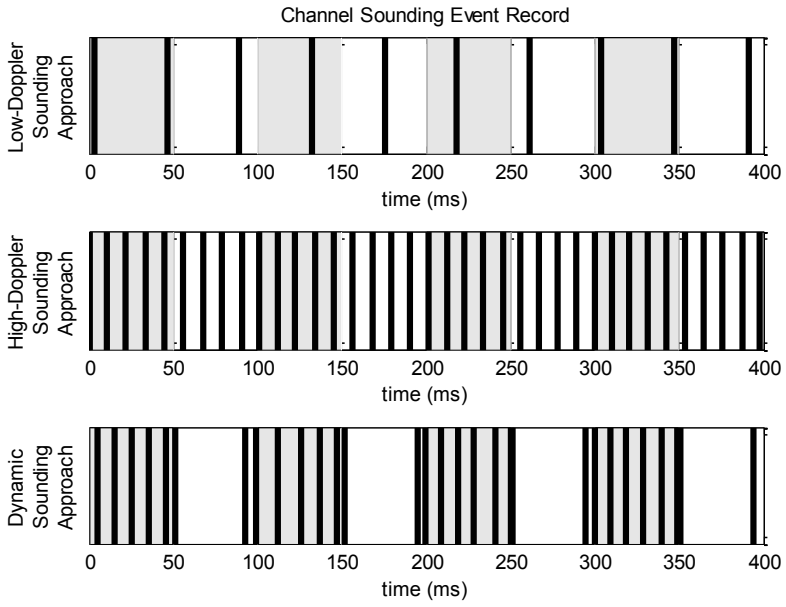


Fig. 17. Example channel sounding event record under three schemes – low Doppler approach (LDA), high Doppler approach (HDA), and dynamic sounding approach.

We define the throughput improvement of approach X over approach Y to be $(R_X - R_Y)/R_Y$, where R_X and R_Y are the throughputs achieved by approach X and Y. Fig. 18 shows the throughput improvement of the proposed dynamic sounding approach over LDA (green bars) and over HDA (yellow bars) in high Doppler, low Doppler, and alternating conditions. We observe that in the

high Doppler scenario, the throughput improvement of dynamic sounding approach over LDA is as high as 31.8%. This illustrates the benefit of operating the sounding fast in the high-variation radio environment. In addition, the throughput improvement of dynamic sounding approach over HDA is 8.6%, which means that dynamic sounding approach outperforms the scheme that uses the best fixed-sounding interval. This is because in the high Doppler scenario the channel variation levels change over time, and dynamic sounding approach can capture this change timely and then results in a better performance. Similarly, we observe a 14.3% improvement better of using dynamic sounding approach over HDA in the low Doppler scenario. Dynamic sounding and LDA almost achieve the same performance, where the difference is 3.1%. In the alternation conditions, however, none of the fixed approaches achieve the same level of performance as the dynamic sounding approach. In particular, our proposed approach is 26.2% better than LDA, and 10.9% better than HDA. Our approach explores the change of the channel environment and trades the CSI accuracy for effective transmission time as the function of the real-time channel variation.

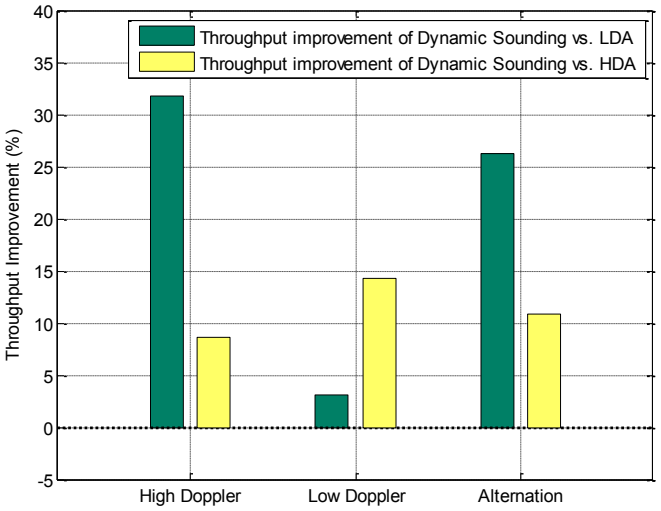


Fig. 18. Throughput improvement of dynamic sounding over LDA and HDA (using the optimal rate adaptation).

3.3 The Impact of Rate Adaptation

Rate adaptation exploits multiple transmission rates under different environments for better achievable throughput. The MCS index for IEEE 802.11ac is listed in Table 13.

Table 13 IEEE 802.11ac MCS index [40].

MCS Index	Modulation	Coding
0	BPSK	1/2
1	QPSK	1/2
2	QPSK	3/4
3	16-QAM	1/2
4	16-QAM	3/4
5	64-QAM	2/3
6	64-QAM	3/4
7	64-QAM	5/6
8	256-QAM	5/6
9	256-QAM	5/6

In this section, we investigate the assumption of using optimal MCS. Our goal is twofold. First, we design a practical and efficient rate adaptation algorithm as attached to the dynamic sounding approach. Second, we evaluate the loss of performance (against the dynamic sounding approach with optimal MCS) when the proposed rate adaptation approach is used.

3.3.1 Proposed Rate Adaptation for Dynamic Sounding

The proposed rate adaptation procedure is illustrated in Fig. 19. When the transmission starts with a channel sounding, the MCS is selected for the first A-MPDU based on the instantaneous SINR calculated from the fresh channel estimation. Starting from the second A-MPUD, the CSI accuracy will degrade. Thus, we adapts the packet error rate (PER) as the indicator of the real-time transmission status. In this way, the MU transmission rate is tracked within a sounding interval.

The MCS will perform rate drops: if A-MPDU PER is greater than a threshold δ_1, the MCS index will drop by 1; if AMPDU PER is greater than a threshold δ_2, the MCS will drop by 2. Notice that within a sounding interval, the MCS index does not go up.

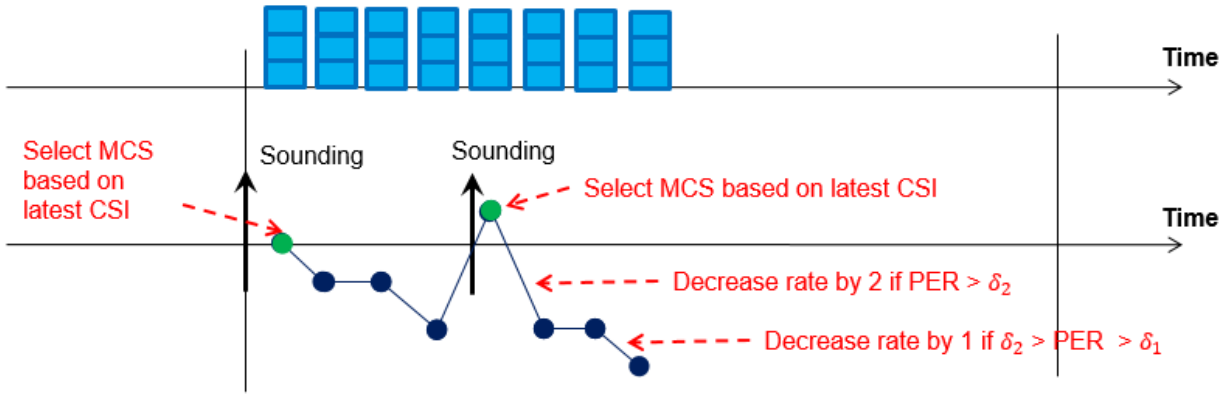


Fig. 19. Rate adaptation scheme in the dynamic sounding approach.

3.3.2 Performance Impact of Rate Adaptation

We continue to use the 802.11ac emulator with our over-the-air channel measurements as the evaluation platform which is discussed in Section 3.1. The only difference is that the rate adaptation is not assumed to be the one which selects the optimal MCS. Based on empirical offline simulation results, we set δ_1 and δ_2 with the values of 0.1 and 0.4, respectively.

Table 14 shows the throughput gap in three scenarios between (1) dynamic sounding with the proposed rate adaptation and (2) the previous dynamic sounding with optimal rate adaptation. Since the optimal MCS cannot be selected all the time, the proposed rate adaptation scheme lead to up to 9.2% throughput loss among all the test scenarios. The highest loss happens in the dynamic scenario which alternates low and high Doppler samples. This is because in such an environment, the channel condition changes between low Doppler and high Doppler much more significantly than the other two scenarios. The required MCS change thus should be faster than either the low

or high Doppler scenario. Therefore, due to the heuristic nature of selecting MCS, the throughput loss is the most obvious in the alternation scenario.

Table 14 Throughput Gap in Different Scenarios.

Scenario	Throughput Gap (%)
High Doppler	7.3%
Low Doppler	2.6%
Alternation	9.2%

We compare the throughput improvement of dynamic sounding over the fixed sounding interval approaches when the proposed rate adaptation algorithm is integrated. To make fair comparison, we configure that the proposed rate adaptation is adopted by both the dynamic sounding approach and the fixed sounding interval approaches (LDA and HDA).

Fig. 20 shows the throughput improvement of the updated dynamic sounding approach over updated LDA (green bars) and over updated HDA (yellow bars) in high Doppler, low Doppler, and alternating conditions. As can be seen, the throughput improvement becomes larger when the proposed rate adaptation is considered. Especially in the alternation scenario, the improvements of dynamic sounding vs. LDA and HDA are 27.8% and 14.5%. These values are much higher than the results (19.8% and 10.9% as shown in Fig. 18) which selects the best MCS. Even though both dynamic sounding approach and the approach with fixed sounding intervals are impacted by adopting the proposed heuristic rate adaptation, the dynamic sounding works better with it. This means that the proposed rate adaption fits well in the dynamic sounding approach. In both the high and low Doppler scenario, the results are only slightly improved. This illustrate that the rate adaptation's impact is larger in the alternation scenario (where MCS change should be faster) than in either pure low Doppler scenario or high Doppler scenario.

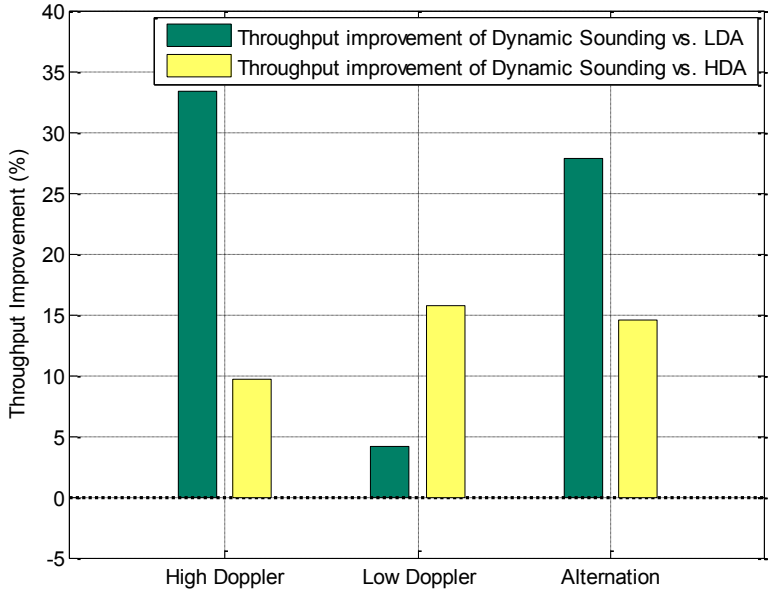


Fig. 20. Throughput improvement of dynamic sounding over LDA and HDA (using the proposed rate adaptation).

3.4 Conclusion and Future Works

In this work, we present a framework to evaluate the channel sounding with application to MU-MIMO in WLAN. Using our collected channel measurements, we discuss the trade-off between the accuracy of the CSI and the effective transmission time. The optimal sounding interval is analysed for the given channel environment. For practical scenarios where the indoor radio environment can change unpredictably, a low-complexity dynamic sounding approach is proposed that updates the sounding interval to improve the effective throughput. We use reference throughput as the metric to determine the sounding interval. Under our collected channel measurements, we show significant throughput improvements (up to 31.8%) for 802.11ac systems by using our proposed dynamic sounding approach, especially in the highly dynamic environments.

Chapter 4

User Grouping For MU-MIMO in WLAN

In this chapter, we present the technical work for the user grouping for MU-MIMO in WLAN. Selecting the improper MIMO mode would lead to system performance deficiency, whereas grouping the users optimally could improve the throughput significantly. Our contributions are: (1) We model the user grouping problem using a hypergraph and show that maximum hypergraph matching provides the optimal solution for choosing between SU-TxBF and MU-MIMO. We demonstrate an approach to determine how many users and which users to assign to the MU groups to maximize system throughput subject to MU-MIMO air time fairness. (2) We develop an efficient and scalable algorithm based on graph matching for solving the above problem and evaluate its performance using an 802.11ac emulator that uses collected channel measurements in an office environment and simulated outdoor channels. The results show that the proposed algorithm achieves at least 93% of the optimal system throughput, and outperforms today's state-of-art algorithms.

The outline of this chapter is as follows. In Section 4.1, we provide the background materials on graph matching. Section 4.2 presents the problem statement with the system description and derives the complexity of the exhaustive search solution. In Section 4.3, we model the grouping

problem using a hypergraph, and show that the maximum hypergraph matching is the optimal solution. A more efficient algorithm, which is based on graph matching, is proposed to provide the near-optimal but more computationally efficient solution. The experimental evaluation of the proposed algorithm is presented in Section 4.4 to validate the theoretical work, and Section 4.5 concludes this work.

4.1 Graph Matching

In mathematics, graph theory is the study of vertices and edges, and one of the prime research areas in discrete mathematics. Graphs are mathematical structures for describing pairwise relations between objects. Many practical problems with different many types of relations and processes can be modelled and analysed using graph theory.

This section presents a concise introduction to the concepts and tools in graph theory that are used throughout our MU-MIMO user grouping design and analysis. We begin with the basics of graph and then describe the set of graph matching problems and their properties. Specifically for the matchings in hypergraph, the details are introduced at the end of this section.

4.1.1 Graph

A **graph** $G = (V, E)$ comprises a set V of **vertices** together with a set E of **edges** as shown in Fig. 21. Each edge is associated with two vertices. $v_i \subset V$ represents a vertex in the graph with the index i, and $e_j \subset E$ represents an edge with index j.

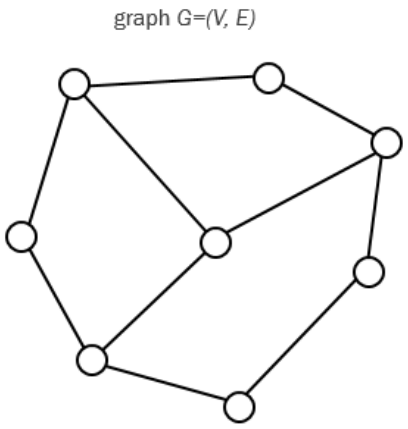


Fig. 21 A graph with vertices and edges.

If the graph is **weighted**, each edge is given a numerical value as its weight. W is the set of the **weights**, and w_i is denoted as the weight for vertex v_i. If the edges of a graph are directed from on vertex to another, the graph is called **directed**. If there is no distinction between the two vertices associated with each edge, the graph is called **undirected**.

The number of vertices of a graph G is its **order**, represented by $|G|$. According to the orders, graphs can be **finite** or **infinite**. In our analysis, we consider the graphs to be weighted, undirected and finite unless otherwise stated.

4.1.2 Graph Matching

A **matching** S of a graph G is a subset of edges such that no two edges share a common vertex. That is to say, a matching S satisfies (1) $S \subset E$ and (2) $\forall\, e_i, e_j \subset S$, $e_i \cap e_j = \phi$. The size of a matching $|S|$ is the edge number of that matching.

A vertex is called **matched** if it belongs to one of the edges in the matching. Otherwise, it is an **unmatched** vertex. An edge is called matched if it is in the matching. Otherwise, it is an unmatched edge.

For a weighted graph, a matching is **maximum** if it has the largest possible sum weight value across all the vertices in that matching. Note that there may be several maximum matchings in a given graph.

A matching is **complete** if it contains all the vertices in $|G|$. In some works, the term **perfect** matching is also used. For example, in Fig. 22, the bold edges stands for complete matching. If the sum of the weights are maximal, this matching is a maximum matching.

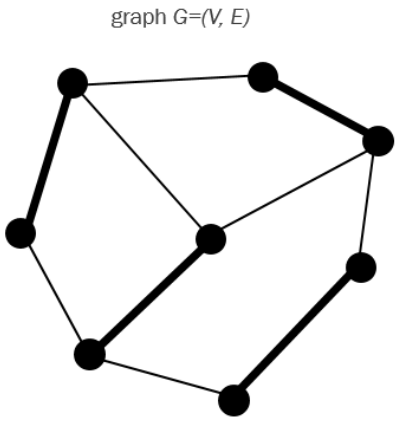


Fig. 22 A matching in a graph.

Particularly of the WLAN user grouping, we adopt tools in graph matching to model and design efficient algorithms. Here we demonstrate one simple example. In a graph as shown in Fig. 23, each vertex in a graph stands for a station. The weight of the edge between two stations stands for the data rate when we put those two STAs in one group for MU2 transmission. For instance, if station A and B are grouped together, their MU2 data rate is 580Mbps, whereas the MU2 data rate becomes 300Mbps if station B and C form a 2-user group. Thus, the goal of achieving the highest network throughput drive us to find a set of edges that the sum of weights has a maximal value, and each vertex happens only once for user fairness. In this example, after we transverse all the

possible combinations, the best solution is shown in Fig. 23, which means A and B are in one group, C and F are in one group, D and E are in one group, and H and G are in one group.

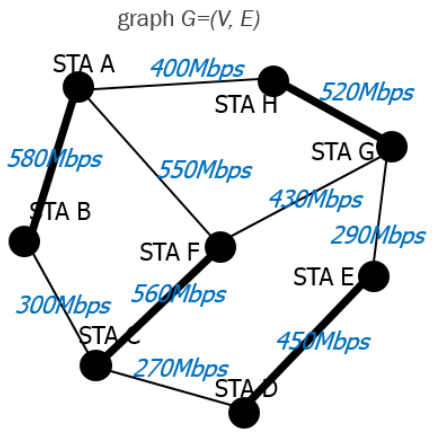


Fig. 23 An example of using graph matching for user grouping.

Notice that this is exactly the maximum matching problem in group theory. However, this model cannot work for the MU2 + SU case, where MU2 and SU transmissions are all allowed, because every edge in the ordinary graph has to connect with exact two vertices. Thus, bipartite graph and hypergraph are investigated for the more general WLAN user grouping cases. Here, we introduce the basics of those graph theory tools. The user grouping modelling, algorithm design and system evaluation are presented later in Section, 4.2, 4.3, and 4.4.

4.1.3 Matchings in Weighted Bipartite Graph

A **weighted bipartite graph** is a weighted graph whose vertices can be divided into two disjoint sets such that no edge connects two vertices of the same set as shown in Fig. 24. A **balanced** bipartite graph is one that has an equal number vertices in the two disjoint sets. In a weighted bipartite graph, a maximum weighted bipartite matching is a matching where the sum of the weights of the edges in the matching has a maximal value.

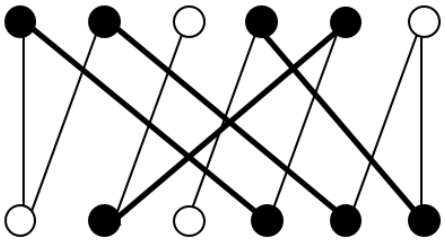


Fig. 24 A weighted bipartite graph.

The alternating path and augmenting path are the important concepts of the maximum matching for the either ordinary graphs or weighted bipartite graphs. Given a matching S, an **alternating path** is the one that begins with an unmatched vertex and alternates between matched and unmatched edges. An **augmenting path** is an alternating path that starts from and ends on unmatched vertices as shown in Fig. 25.

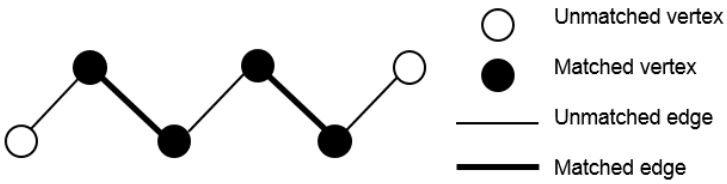


Fig. 25 An augmenting path.

It can be proved that a maximum matching is achieved if and only if the graph does not have any augmenting path. Finding the augmenting paths efficiently is the key for maximizing the matching in graphs with varies of topologies.

Blossom algorithm provides the polynomial time solutions by searching the shortest path in the augmenting path algorithm. The matching of a graph is constructed by iteratively improving an initial empty matching along augmenting paths. The key mechanism is that an odd-length cycle in the graph (blossom) is contracted to a single vertex, with the search continuing iteratively in the

contracted graph. Other than Blossom algorithm, Hungarian algorithm is another polynomial time solution for maximum matching and combinatorial optimization problems.

4.1.4 Hungarian Algorithm

(a) Algorithm Basics

Hungarian Algorithm is based on the Kuhn-Munkres Theorem to convert the maximum weight matching problem into finding the right labeling function with perfect matching on the corresponding equality subgraph.

A labeling for graph $G = (V, E)$ is a function $l: V \to \mathbb{R}$, such that: $\forall (u, v) \in E: l(u) + l(v) \geq w((u, v))$, where $w(\cdot)$ stands for the weight of an edge in the graph. An equality subgraph is a subgraph $G_l = (V, E_l) \subseteq G$, with a labeling l, such that $E_l = \{(u, v) \in E: l(u) + l(v) = w((u, v))$.

Kuhn-Munkres Theorem: Given labeling l, if S is a perfect matching on G_l, then S is maximal-weight matching of G.

The key procedure of Hungarian algorithm is to maintain both a matching S and equality graph G_l. The initial starting condition is $S = \emptyset$, a valid l and a input bipartite weighted graph with disjoint vertex sets V_1 and V_2. In each iteration, it either augments S or improves the labeling $l \to l^*$. The algorithm continues until S becomes a perfect matching on G_l.

The searching begins with an initial matching S, and a valid labeling $l := \forall x \in X, y \in Y: l(y) = 0, l(x) = max_{y^* \in Y}(w(x, y^*))$. The following two steps are executed until S becomes a perfect matching: (i) Augment the matching. (ii) If augmenting path does not exist, improve the labeling $l \to l^*$ and go to step (i).

The matching augments as follows. If a path alternates between $E_l - S$ and S, it is augmenting for S on G_l, and its starting and ending vertices are unmatched in S. Track and record a candidate

augmenting path starting at an unmatched vertex u, $u \in V, u \notin S$. If an unmatched vertex v can be found, then an augmenting path P from u to v is recorded. We then flip the matching by replacing the edges in the current matching S with the edges in the augmenting path that are in $E_l - S$. Notice that this step increases the size of the matching, i.e., $|S^*| > |S|$, because the starting and ending vertices are both unmatched.

The labeling improves as follows. We use $P \subseteq V_1$ and $Q \subseteq V_2$ represent the current candidate augmenting alternating path between the matching S and the other outside edges in $E_l - S$. The neighbours to the node in S along E_l is denoted by $N_l(P) = \{v | \forall u \in P: (u, v) \in E_l\}$. If $N_l(P) = Q$, which means the alternating path cannot increase and augment, labeling then need to improve. A calculation for δ_l is required: $\delta_l = min_{u \in P, v \notin Q}\{l(u) + l(v) - w((u, v))\}$. Then, labeling $l \rightarrow l^*$ improves: if $r \in P$, $l^*(r) = l(r) - \delta_l$; if $r \in Q$, $l^*(r) = l(r) + \delta_l$; otherwise, $l^*(r) = l(r)$.

(b) Algorithm Complexity

The size of the matching, $|S|$, increases by one edge for each iteration through either step (i) or (ii). Thus, the total rounds can be $O(|V| = n) = O(n)$ total rounds. For augmenting the matching, if the right vertex exists, it requires $O(|V|)$ to find it out, and $O(|V|)$ to flip the matching. For improving the label, finding δ_l and updating the labeling take O($|V|$) time. In addition, in the case when no augmenting path can be found, improving the labeling can occur $O(|V|)$ times. Therefore, one single iteration takes total $O(|V|^2)$ work. Considering the overall iteration number is $O(|V|)$, the total running time of Hungarian algorithm is $O(|V|^3) = O(n^3)$.

4.1.5 Blossom Algorithm

(a) Algorithm Basics

A blossom in a graph is defined as a cycle of odd length ($2k + 1$ edges) with k matching edges. The major reduction of the augmenting path searching mechanism for blossom algorithm comes from the Cycle Shrinking Lemma.

Cycle Shrinking Lemma: Assume that a blossom B in the graph G is vertex-disjointed from with the rest of the matching S. Reconstruct the graph G into G^* by contracting B to a single vertex. In such conditions, the matching S^* in G^* induced by S is maximum in G^* if and only if S is maximum in G.

In other words, an equivalent way of find the maximum matching in the original graph is to find the maximum matching in the shrunk graph which is constructed based on blossom detection and contraction. Specifically, the blossom algorithm contract the graph for fast searching of augmenting path, and it lifts the blossom in to generate the complete results in the original graph. The contraction and lifting of a blossom in a graph is illustrated in Fig. 26.

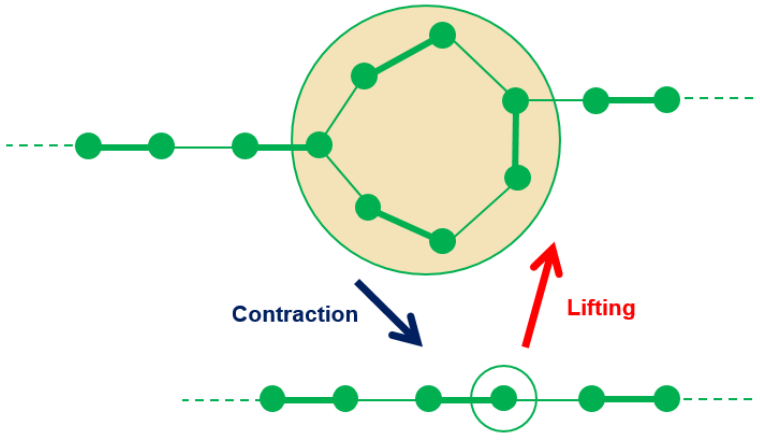


Fig. 26 Blossom contraction and lifting.

The first step of blossom algorithm is to search for an augmenting path using the data structure of a forest F which is an undirected graph consisting of a disjoint union of trees. All the vertices in F will be assigned a label, either “odd” or “even”.

The set of unmatched vertices in G is denoted by $U \subseteq V$ for a given matching S. All the vertices in U are labeled as "even" in the beginning. Now, U is regarded as the initial forest F. This forest grows for every vertex in U in the following way: (1) We find an edge that connects an "even" vertex within F with a matched vertex v outside F. Find the neighbour vertex of v in S, and denote it to be v^*. (2) Expand the forest $F = F \cup (u, v) \cup (v, v^*)$ with v to be "odd" and v^* to be "even".

After the forest construction is finished, the following steps are executed in each iteration of the algorithm: (1) Find an augmenting path, (2) Detect a blossom and contract the graph. (3) Locate an augmenting path P^* in the contracted graph and lifts it to the original graph. (4) Decide whether there are no augmenting paths.

(b) Algorithm Complexity

There are $O(n)$ iterations at the stage of the forest growing, because there can be $O(n)$ augmenting paths. In each iteration, $O(n)$ blossoms can be found for one augmenting path to shrink. Identifying a blossom or an augmenting path requires $O(m)$ at each level. Therefore, it takes $O(mn^2)$ for reducing the graph and finding the maximum matching. In addition, the time complexity for expanding is $O(mn^2)$, because it requires $O(m)$ time to expanding one single blossom. Thus, considering contraction, searching and expansion, the total time complexity of blossom algorithm is $O(mn^2)$.

4.1.6 Matchings in Hypergraph

A **hypergraph** is a generalization of a graph. It consists of vertices and hyperedges, where each **hyperedge** connects any non-zero number of vertices from V. A hypergraph is equivalent to an ordinary graph, if every hyperedge is associated with exact 2 vertices. In a weighted hypergraph,

a single numerical value is associated with each hyperedge. Different from the ordinary graphs, hypergraphs are not straight-forwardly to visualize due to the arbitrary number of vertices for each hyperedge.

Similarly as the ordinary graph, in a hypergraph $G = (V, E)$, any subset $S \subset E$ is called a matching if any two hyperedges in the subset do not have a common vertex, i.e., $\forall\, e_i, e_j \subset S$, $e_i \cap e_j = \phi$. If every vertex in the hypergraph is incident to exactly one hyperedge of the matching, it is called a complete matching. For the weighted hypergraph, maximizing the matching is to find a matching that maximizes the weighted sum of the matching.

Although there is polynomial time solutions for the ordinary graph matching problem, hypergraph matching is NP-hard and no good approximation algorithm exists. Existing solutions are almost all based on local search. Local search can be considered as an extension of the augmenting path algorithm. The nature of the local search method is to use heuristics for finding an approximate solution by applying local moves.

4.2 Problem Context and Formulation

In this section, we present the problem context including the description of the downlink MU-MIMO transmission (Section 4.2.1) and the MU-MIMO air time fairness (Section 4.2.2). The formulation of the problem with the solution space analysis is given in Section 4.2.3.

4.2.1 Downlink MU-MIMO transmission

We consider a downlink MU-MIMO scenario in WLAN which consists of one AP with N_t antennas and M mobile stations, each equipped with 1-receive antenna. The AP serves all mobile stations. We consider two transmission modes: SU-TxBF and MU-MIMO. We refer to them as SU and MU_X, where X stands for the number of users in the MU group. The AP supports both SU

and MU_X transmission modes and can flexibly switch between them and between MU_X and MU_Y ($X \neq Y$). We denote N_u to be the maximum number of users in an MU group for the considered system. The AP acquires the channel state information (CSI) of all active users before performing downlink MU transmission in a transmit opportunity (TXOP) period, similar to the MAC protocol used in [120], [92], [121].

4.2.2 Fairness Criterion

Throughput maximization and fairness consideration are two important criteria for WLAN design and deployment. Throughput maximization aims to either maximize the individual throughput of each station or the overall throughput of the network. However, such strategies can result in unfair quality of service (QoS) delivery among the mobile stations.

Fairness in a non-MU-MIMO WLAN system, such as IEEE 802.11a/b/g, can be classified into two categories: throughput fairness and air time fairness (ATF). Systems employing throughput fairness provide equal throughput for each individual station in the network. However, as observed in [122], when one or more mobile stations use a lower bit rate than the others, the aggregate throughput of all stations is drastically reduced. This is because lower-rate stations capture the channel for more time. This penalizes the stations capable of a higher transmission rate. The extreme case is where one station's rate is so low that it occupies most of the air time. To overcome this deficiency of throughput fairness, ATF is introduced in [123] and [124]. With ATF, each competing node receives an equal share of the wireless channel time. It is demonstrated that this notion of fairness can lead to significant improvements in aggregate throughput while still guaranteeing that no node receives worse channel access than it would if all stations were using the same transmission rate. In addition, it is shown that the long-term individual throughput of a station is proportional to that station's data rate capability. Thus, ATF helps to control

transmissions in such a way that it gives WLAN users a better and more predictable wireless experience. Under the concept of ATF, user fairness is achieved with SISO transmissions when each user gets equal time to transmit in the network.

We extend ATF to MU-MIMO ATF for the downlink MU-MIMO WLAN scenario. Consider that the AP serves the stations in turns in a Round-Robin manner [125]. If time duration T_{SU} is allocated for a single-user transmission and time duration T_{MU} is allocated for a multi-user transmission, then $T_{MU} = T_{SU} \cdot N_{mu}$ defines MU-MIMO ATF, where N_{mu} is the number of users in the multi-user transmission group. In other words, MU-ATF entails that the transmission time allocated to an $MU_{N_{mu}}$ group is N_{mu} times that of an SU.

There are two explanations why MU-MIMO ATF is more applicable than ATF across groups. First, encouraging higher order MU transmission increases the system throughput. This is because with a larger percentage of total transmission time allocated to MU transmission there is higher spectral efficiency and thus higher throughput. Second, if the ATF is equal across groups, the actual data rate for each user in an MU transmission may be lower than that for single user transmission. This could discourage MU-MIMO transmission. To illustrate this, consider a six-station scenario where there is one SU transmission (Station-A), one MU group with two users (Station-B, Station-C), and one MU group with three users (Station-D, Station-E, Station-F). We denote the data rate for SU, MU_2, and MU_3 as r_{su}, r_{mu2}, and r_{mu3}, where typically $r_{mu2} \leq 2 \cdot r_{su}$ and $r_{mu3} \leq 3 \cdot r_{su}$ under the same channel conditions. If the air time fairness were equal across groups, then the throughput of devices operating in MU transmission (B, C, D, E, and F, in this case) would be potentially lower than the legacy devices or MU-disabled 802.11ac devices. This discourages MU transmission and MU devices. Therefore, we consider the following multi-user air time fairness (MU-ATF) criterion: the users in a group take turns being the primary user, and

ATF is then the fairness among all of the primary users in an MU group and the SU receivers. For the previous 6-station example, this air time fairness criterion would then lead to transmission period **A**, (**B**C), (**C**B), (**D**EF), (**E**FD), and (**F**DE).

4.2.3 Problem Formulation and Solution Space

Consider M single-antenna mobile stations, Station-1 to Station-M, are divided into K groups. We denote the user index set of the kth group as $G_k = \{\pi_k(1), \dots, \pi_k(|G_k|)\}$, where $\pi_k(i)$ is the index of the ith station in group k. Notice that $G_k \subset \{1, \dots, M\}$, and $|G_k| \leq N_u$ for any k. In other words, each group contains at most N_u stations being a subset of the M stations in the system, where N_u is the system-specific maximum group size. The estimated capacity expression is

$$R(G_k) = B \cdot \sum_{m \in G_k} log_2\left(1 + \frac{P_m \cdot |\boldsymbol{h}_m \boldsymbol{w}_m|^2}{N_0 + \sum_{m \in G_k, i \neq m} P_i \cdot |\boldsymbol{h}_m \boldsymbol{w}_i|^2}\right), \quad (9)$$

where P_m is the power allocated for user m, $\boldsymbol{h}_m$ is the frequency-domain channel response vector for user m, and $\boldsymbol{w}_m$ is the steering vector for user m. $R(G_k)$ is the estimated capacity of SU-TxBF where $|G_k| = 1$; is the estimated capacity for the MU-MIMO transmission of group G_k where $|G_k| \geq 2$. Thus, the user grouping problem for throughput maximization subject to MU-ATF can be formulated as follows:

$$\begin{aligned} \text{Maximize} \quad & \sum_{k=1}^{K} |G_k| \cdot R(G_k), \\ \text{subject to} \quad & \bigcup_{k=1}^{K} G_k = \{1, \dots, M\}, \\ & G_p \cap G_q = \emptyset \quad (p \neq q, 1 \leq p \leq K, 1 \leq q \leq K), \end{aligned}$$

where K and G_k (for $k = 1 \dots K$) are the decision variables. The MU-ATF is implicitly guaranteed in the maximization expression through the $|G_k|$ term. And the constraint ensures that each user appears either as a single user transmission or in one MU group.

A straightforward way to guarantee the maximum throughput is to conduct an exhaustive search over all possible grouping candidates. There are many ways to group the users even in small networks. For example, in a network of only six stations where up to three stations can be grouped for MU transmission as shown in Fig. 27, (1) SU-TxBF can be used for each station, (2) stations can be grouped using three MU_2 groups with $C(6,2) \cdot C(4,2) \cdot C(2,2) = 90$[1] combinations; (3) stations can be grouped using two MU_3 groups with 20 combinations; and (4) the six stations can be grouped as one *SU*, one MU_2 and one MU_3 with 60 combinations, and so forth. Table 15 lists the number of combination for all possible divisions of 6 stations. Note that the total number of combination is the sum over all rows in Table 15.

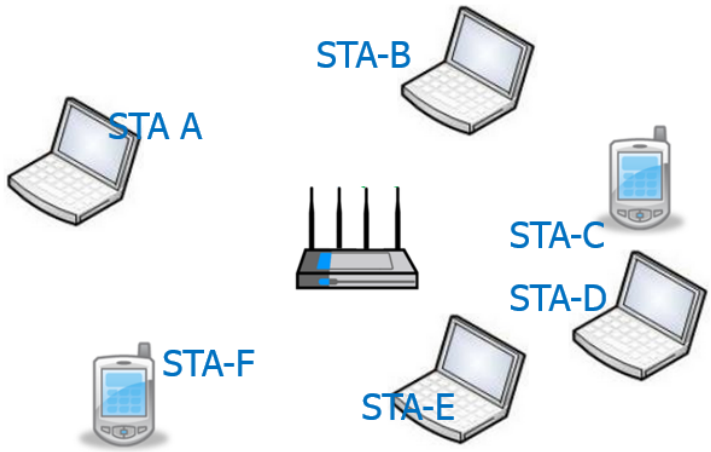


Fig. 27. Example WLAN with one access point and 6 stations.

[1] $C(n,k) = \frac{n!}{k!(n-k)!}$

Table 15 Grouping Division for the 6-Station Case with Group Size of 1-3 Users.

Grouping Division	***Number of Combinations***
Two MU_3 Groups	$C(6,3) = 20$
One MU_3 Group, one MU_2 Group and one SU Group	$C(6,3) \cdot C(3,2) = 60$
One MU_3 Group and three SU Groups	$C(6,3) = 20$
Three MU_2 Groups	$C(6,2) \cdot C(4,2) = 90$
Two MU_2 Groups and Two SU Groups	$C(6,2) \cdot C(4,2) = 90$
One MU_2 Group and four SU Groups	$C(6,2) = 15$
Six SU Groups	$C(6,6) = 1$

The number of possible grouping combinations increases exponentially with the number of stations. It is of order 10^{n-5} for $6 < n < 20$ stations and $N_u = 3$, and larger for larger MU groups. Table 16 shows a few examples. Thus, the exhaustive search is not scalable and infeasible for practical applications.

Table 16 Number of Combinations for Different Station Number.

Number of Stations	***Number of Combinations***
6	296
9	55,897
12	$\sim 2.8 \times 10^7$
15	$\sim 2.7 \times 10^{10}$
18	$\sim 4.9 \times 10^{13}$

4.3 Hypergraph Modeling and Algorithm Design

In this section, we use graph theory tools to model and solve the user grouping problem. We show that the maximum hypergraph matching provides the optimal solution for maximizing system throughput subject to MU-MIMO air time fairness when we can choose between SU-TxBF and MU-MIMO. We develop algorithms for the special case where the system supports MU_2 and SU only and for the general case where any number of users can be grouped in an MU transmission. We show that an optimal solution of polynomial time complexity exists for the special case and propose a scalable heuristic algorithm for the general case.

4.3.1 Hypergraph Modeling

We use the hypergraph to model the grouping problem. A graph $G = (V, E)$ comprises a set V of vertices together with a set E of edges. Each edge is associated with two vertices. If the graph is weighted, each edge is given a numerical value as its weight. A hypergraph [126] is a generalization of a graph. It consists of vertices and hyperedges, where each hyperedge connects any non-zero number of vertices from V. In a weighted hypergraph, a single numerical value is associated with each hyperedge. An example hypergraph with twelve vertices and ten hyperedges is shown in Fig. 28. In this example, e_1, e_2 and e_9 are hyperedges that connect two vertices, hyperedges e_3, e_5, e_7 and e_8 connect three vertices, whereas e_4, e_6 and e_{10} are each associated with a single vertex. We denote $|e_i|$ the number of vertices that e_i connects.

In a hypergraph, any subset $S \subset E$ is called a matching if any two hyperedges in the subset do not have a common vertex, i.e., $e_i, e_j \subset S$, $e_i \cap e_j = \phi$. Subsets $\{e_1, e_3, e_4\}$, $\{e_7, e_9\}$, and $\{e_1, e_3, e_4, e_7, e_9\}$ are three example matchings for the hypergraph of Fig. 28. If every vertex in the hypergraph is incident to exactly one hyperedge of the matching, it is called a complete matching. The matching $\{e_1, e_3, e_4, e_6, e_7, e_9\}$ is a complete matching of the hypergraph example.

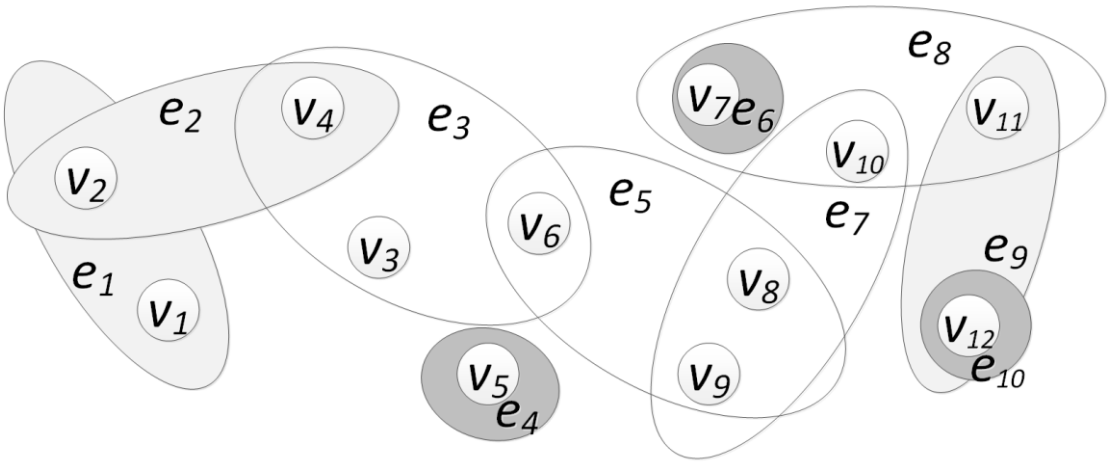


Fig. 28. Example of a hypergraph with twelve vertices ($\boldsymbol{v_1}$ to $\boldsymbol{v_{12}}$) and ten hyperedges ($\boldsymbol{e_1}$ to $\boldsymbol{e_{10}}$).

In our grouping problem, vertex v_i in the hypergraph represents station i in the network and hyperedge e_j stands for an SU-TxBF or MU group. For example, e_4 represents a single user group with station 5, whereas e_3 represents a MU_3 group with stations 3, 4 and 6. Depending on the number of vertices that e_j connects, its edge weight, w_j, corresponds to the estimated capacity $R(G_k)$ according to (9) for either SU-TxBF or MU-MIMO transmission of the given group (1 user, 2 users or 3 … etc.). Therefore, the problem of maximizing the throughput under the MU-ATF criterion can be reformulated as finding a complete matching $S \subset E$ of hyperedges such that

(1) for any two distinct hyperdeges $e_i, e_j \subset S$, $e_i \cap e_j = \phi$, and

(2) $\sum_{e_i \in S}(w_i \cdot |e_i|)$ is maximized.

The first statement ensures that any station appears in only one single group, whereas the second maximizes the sum throughput based on the MU-ATF criterion. Our system design problem can, thus, be equivalently transformed to the maximum matching problem [127] in a weighted hypergraph known from graph theory. The objective is to find a matching that maximizes the sum of the weights in the matching, which is called maximum matching in graph theory.

Finding the maximum matching is a fundamental problem in combinatorial optimization and has various applications.

In a hypergraph where each hyperedge can connect any non-zero number of vertices, the maximum matching problem becomes very complex and was proven to be NP-hard [127]. All the well-known approximation algorithms for the hypergraph matching problem are based on local search methods [128]. A promising algorithm comes from Berman [129] who developed a $\left(\frac{k+1+\epsilon}{2}\right)$-approximation algorithm. There are two issues with adopting these algorithms for our application. Since all user grouping combinations are allowed, the calculation of all the corresponding hypergraph weights is proportional to $|V|^{N_u}$, which requires significant computation for reasonable problem sizes. In addition, all the local-search-based algorithms are heuristic in nature and thus do not guarantee finding the optimal solution.

Here we propose a new design approach for optimizing the sum-throughput under MU-ATF criterion while reducing the computational complexity. We start with the case where $N_u = 2$, i.e., only MU_2 and SU are supported by the system. We then extend this to the general case where MU_X (X being any positive number) and SU are supported.

4.3.2 Algorithm Design

Solutions to the throughput maximization problem under the MU-MIMO ATF criterion requires new algorithms for effectively selecting the transmission mode. We first consider the $N_u = 2$ case and show that the optimal grouping problem can be efficiently solved for systems that allow only SU-TxBF and 2-user MU groups (Section 4.3.2.1). We then develop an efficient graph matching algorithm based on graph theory principles for the general case where any number of users can be assigned to groups of different sizes (Section 4.3.2.2).

4.3.2.1 $MU_2 + SU$

The $MU_2 + SU$ grouping problem is inherently more difficult to solve than MU_2 grouping. MU_2 user pairing [99] can be optimally solved for maximum throughput by directly adopting the combinatorial optimization algorithm [130]. However, this method cannot be used for solving our $MU_2 + SU$ case subject to the MU-ATF criterion. Here, we show that the grouping problem in this case can be modeled as a symmetric maximum matching problem in a weighted bipartite graph. This problem can be optimally and efficiently solved by the Blossom algorithm [131], which is a polynomial time solution.

The weighted bipartite graph is a graph whose vertices are divided into two disjoint sets V_1 and V_2. The matching then happens between vertices in V_1 and vertices in V_2. Sets V_1 and V_2 in our problem each contain all vertices in the system, where a vertex stands for a mobile station in the network. The weight of the edge connecting a station in V_1 with a station in V_2 represents the data rate of the MU_2 group of stations if the two connected stations are not the same; otherwise, the weight indicates for the SU-TxBF data rate for that station. Thus, the candidate matchings of the bipartite graph provide the candidate user grouping solutions, and the matching with the maximum sum weight is the optimal grouping for $MU_2 + SU$ case.

For example, in the six-station scenario, all six stations (A - F) are assigned to both sets, V_1 in the first row and V_2 in the second of Fig. 29, which illustrates the matching (**A**), (**B**C), (**C**B), (**D**E), (**E**D), (**F**). The weight of the edges between a station in V_1 and a station in V_2 corresponds to the data rate of the MU_2 group if the connected stations are different. The edge connecting the same station corresponds to SU-TxBF. For the matching of Fig. 29, A and F are SU-TxBF, and BC and DE form two MU2 groups.

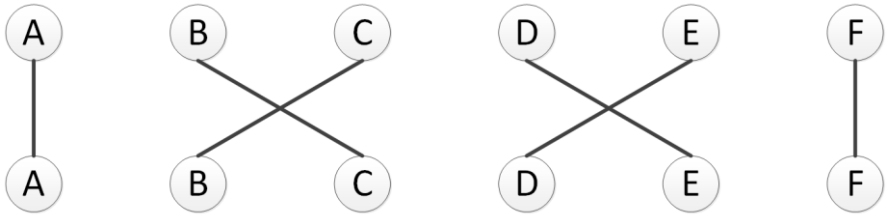


Fig. 29. Symmetric matching in a bipartite graph.

Notice that the matching in Fig. 29 is a symmetric matching: if B in V_1 connects to C in V_2, then C in V_1 connects to B in V_2. This is a characteristic of our $MU_2 + SU$ problem. The symmetric matching of the weighted bipartite graph can be optimally solved using the Blossom algorithm, which has polynomial time complexity [131].

If there were no requirement on matching symmetry, B in V_2 could, for instance, connect to D in V_1 even if B in V_1 does not connect to D in V_2. In such a case where the matching is not required to be symmetric, a polynomial time solution, Hungarian algorithm [130], can be used. The Hungarian algorithm is not suitable for our MU_2+SU case, but we will use it as part of the solution for the MU_X+SU case.

4.3.2.2 $MU_X + SU$

We propose an efficient graph matching algorithm (GMA) for solving the MU_X+ SU problem, which directly reduces to the optimal solution for MU_2+SU case when $X = 2$. The basic design idea is to repeatedly apply symmetric weighted bipartite graph matching to form higher order MU groups and split the low data rate groups. The following steps describe the operations for obtaining the final grouping results, illustrated in Fig. 30 for a twelve-station scenario. In Step 1, the Blossom algorithm partitions the users into MU_2 or SU. These groups are then sorted in descending order of data rates in Step 2. The low throughput groups are decomposed into single user groups for the asymmetric weighted bipartite graph matching, which is solved by the Hungarian algorithm in

Step 3. Notice that the purpose of the decomposition is to split the low throughput user groups as necessary for the input for Hungarian algorithm; therefore, the decomposition is not a function of any system parameters, such as network scale or user orthogonality. In Step 4, for each new group, if the group provides a higher total data rate, the grouping result is kept; otherwise, the previous group assignment remains. This procedure is repeated recursively for $N_u > 3$. More precisely, a single run is needed for MU_3, two runs are needed for MU_4, and so forth.

This procedure is illustrated in Fig. 30 for twelve stations. The Blossom algorithm is applied to the weighted bipartite graph, grouping the station into MU_2 or SU groups. The MU_2 groups here are (AG), (EH), (BJ), (DK), and (FI), whereas C and L remain as SU groups. The resulting groups are sorted according to their throughput and are divided into two sets, one containing the higher throughput groups and the other the lower throughput SU-TxBF users. Asymmetric weighted bipartite matching is then applied to the two disjoint sets using the Hungarian algorithm and leads to (EHF), (DKC), (AGI) and (BJL). If a newly formed group, such as (EFH), leads to a higher throughput than before, the grouping result is accepted; otherwise, the previous group assignment—(EH) and (F)—is kept. Here, the throughput of (EFH) is higher than the sum throughput of (EH) and (F). On the other hand, (DK) and (C) is a better option than (CDK). If the maximum number of users in a group is $N_u = 3$, the final grouping result is obtained as (EFH), (AGI), (BJL), (DK) and (C). If more users per MU group were permitted, the procedure would resume at Step 2. Compared with local search methods, the weight computation cost is reduced from $O(|V|^{N_u})$ to $O(N_u|V|^2)$.

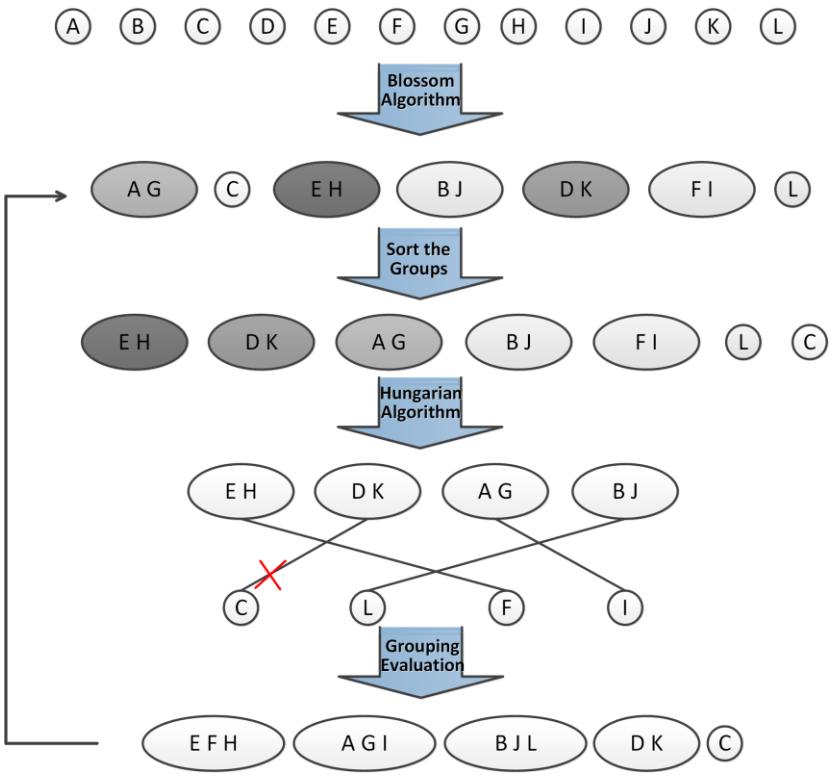


Fig. 30. Flowchart of the proposed graph matching algorithm (GMA).

Table 17 provides the algorithm description in the form of a pseudo code. Lines 1-9 represent the use of the Blossom algorithm to group users into $MU_2 + SU$ groups. Lines 12-27 describe the sorting and splitting of the grouping results. Lines 28-35 correspond to the use of the Hungarian algorithm for forming higher order MU groups. The process then loops back to line 11 for MU_4 or higher-order groups. It is easy to observe that this approach provides a solution in polynomial time because the main processes are the (1) Blossom algorithm, which has the complexity of $O(|V|^3)$, (2) the sorting operation, which has a complexity of $O(|V|log|V|)$, and (3) the Hungarian algorithm, which is of complexity of $O(|V|^3)$. Hence, the overall complexity order is $O(N_u|V|^3)$.

Table 17 Algorithm # 2.

Heuristic Algorithm for N_u-User Grouping

Calculate the achievable throughput for each single user as GMA's input;
Calculate the achievable MU_2 throughput for every two users as GMA's input;
Calculate the best MU_2 group G using Blossom algorithm;
for $i = |G|:-1:1$
    if SU transmission of the two users in $G\{i\}$ is better than MU_2
        Add two SU groups;
        Remove $G\{i\}$;
    end if
end for
$k = 2;$
while $k < N_u$
    $G_{result} = \emptyset$; $k = k + 1$;
    Put the groups into two sets S_1 and S_2;
    $S_1 = G$; $S_2 = \emptyset$;
    Sort the groups in $|S_1|$ according to their achievable throughput;
    while $|S_1| > |S_2|$
        Put the user(s) of the lowest group of $|S_1|$ into $|S_2|$ as SU group;
        Remove this lowest group from $|S_1|$;
    end while
    while $|S_1| \neq |S_2|$
        Put the last user of $|S_2|$ into $|S_1|$;
        Remove this user from $|S_2|$;
        if $|S_1| > |S_2|$
            Put this SU group into G_{result};
            Remove this user from $|S_1|$;
        end if
    end while
    Calculate the groups G_{MU} based on Hungarian algorithm;
    for $i = |G_{MU}|: -1: 1$
        if the new pair achieves better throughput
            Add the new group into G_{result};
        else
            Add the previous group into G_{result};
        end if
    end for
end while

4.4 Evaluation

4.4.1 Experimental System Setup

We obtained channel information measurements for our validation using four MIMO test nodes, each equipped with four-antennas and Qualcomm Chipset QCA9980 which is a four-spatial-stream IEEE 802.11ac transceiver chipset. We conducted the over-the-air transmission over the 40 MHz channel in an office environment. One test node operated as the transmitter AP, and the other three nodes were used to emulate twelve single-antenna receiver stations. We positioned the twelve receive antennas away from one another, at random locations in the office. The twelve sets of captured channel samples were then used to mimic the channel from one four-antenna AP to twelve single-antenna stations. We conducted the channel measurements while there was significant human movement in the measurement environment.

We implemented the measurement-driven MU-MIMO-OFDM emulator rigorously according to the IEEE 802.11ac specifications [40]. This emulator was seeded with the over-the-air channel information to test the performance of the media access control (MAC) and physical (PHY) layer algorithms. The settings of the IEEE 802.11ac system are summarized in Table 18.

Table 18 System Settings.

Parameters	***Values***
Maximum AMPDU Duration	2 ms
MPDU Length	1556 bytes
MSDU Length	1508 bytes
SIFS Duration	0.016 ms
Bandwidth	40 MHz
Number of OFDM Subcarriers for Data	108
Guard Interval for OFDM Symbol	400 ns

To evaluate the performance of our proposed group matching algorithms (GMAs), we compare them with three other state-of-the-art user grouping approaches (a) Zero Forcing with Selection (ZFS) [88], (b) Semi-orthogonal User Selection (SUS) [89], and (c) Random Selection (RS). ZFS selects the user with the highest channel capacity to be the first member in a group and then finds the next user based on the potential sum rate in each selection iteration. SUS picks an ungrouped user to be added to a group based on the qualified and highest effective channel norm value to the existing group. We select the best-performing SUS variant for our evaluation. Random selection randomly selects a user until the cardinality of the group reaches N_u. We also implement the full search (exhaustive search), which provides the optimal system throughput subject to MU-ATF. The algorithms are compared based on measured and simulated channels.

4.4.2 Evaluation based on the Measured Channels

Fig. 31 shows the system throughput as a function of the total number of users in the network for the system in which the largest MU transmission group size is two. We observe that the Blossom algorithm achieves the same system throughput as the full search no matter how many users exist in the network. This is because when only MU_2 and SU are allowed in the system, the Blossom algorithm provides the optimal solution as discussed in Section 4.3.2.1. The performance of random selection is the lowest among all approaches, which is as expected because it does not take any channel information into consideration when forming transmission groups. The performance of ZFS and SUS are better than random selection, but worse than the optimal solution because they are both heuristic in nature. We also observe that the system throughput increases with the number of users in the network for the Blossom algorithm, ZFS, and SUS. This is because the search space for each algorithm becomes larger so there is a higher chance to find a better grouping result.

Fig. 32 shows the system throughput results for the scenario where up to 3 stations can be grouped for MU transmission. We observe that our proposed GMA is closest to the optimal solution which is bounded by full search. The proposed algorithm outperforms ZFS and SUS. The reason for this can be explained by the group splitting and regrouping mechanisms discussed in Section 4.3.2.2. This is to say, effective reshuffling of the users to different groups leads to high system throughput, which is very close to the optimum. From Fig. 32, we observe that our proposed algorithm leads to a system throughput of 98% of the optimal throughput for the considered scenario.

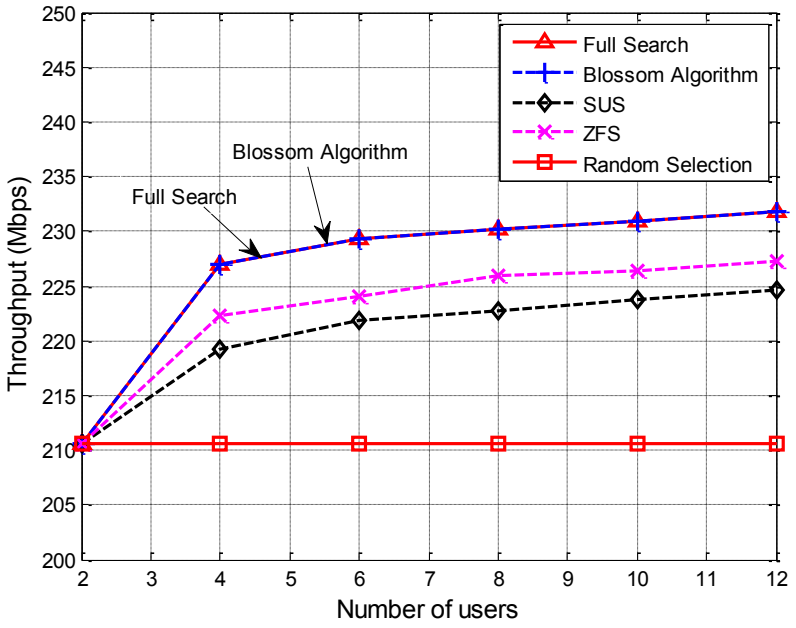


Fig. 31. System throughput as a function of the total number of the uses in the network when $\boldsymbol{N_u = 2}$.

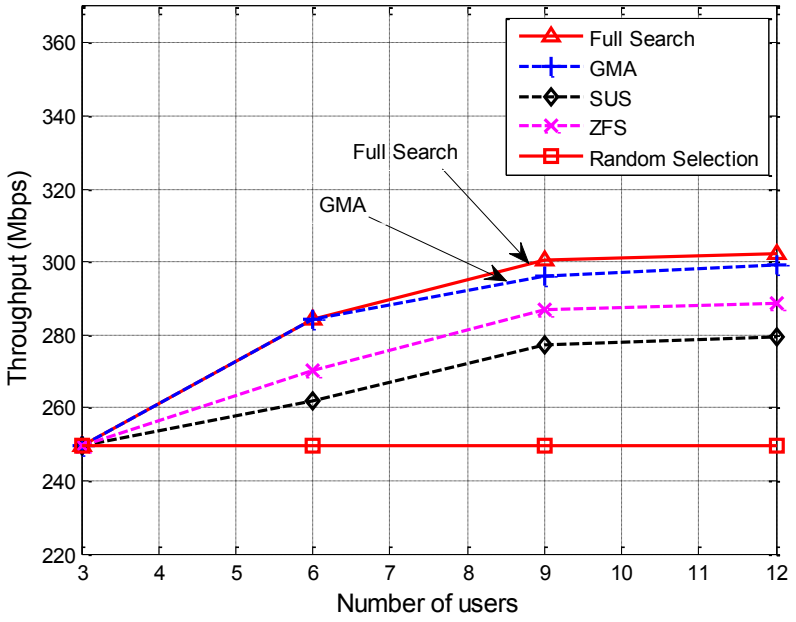


Fig. 32. System throughput as a function of the total number of the uses in the network when $\boldsymbol{N_u = 3}$.

When we examine how users are grouped, we find that MU transmission is preferred by all approaches. That is to say, the more users can be grouped for transmission, the higher the system throughput becomes. This means that the correlation among the channels perceived at the different antennas is very low, which can be explained by the random placement of the receive antennas in the office. To mimic an outdoor environment where there is typically much higher channel correlation and line-of-sight (LOS) effects, we evaluate the algorithms using simulated correlated channels in the next section.

4.4.3 Evaluation based on the Simulated Channels

We have observed that the all the methods attempt to group as many users as possible for transmission. The low channel correlation explains why even the random selection approach can lead to a reasonable system throughput. Now we consider the outdoor case where there is typically

a higher correlation between users' channels and line-of-sight (LOS) effects than indoor environments. The channel is simulated using a Rician fading channel where the k factor is set to 8 dB.

Fig. 33 shows the system throughput for the different algorithms as a function of the correlation coefficient in the scenario where there are 3 correlated users. We observe that the throughput performance of the full search, the proposed GMA, ZFS, and SUS do not depend on the value of the correlation coefficient ρ. This is because MU_3 is always used as long as none of the three correlated users are grouped together. On the other hand, the performance of random selection degrades with increasing value of ρ. This is reasonable because the correlated users can be grouped together and higher correlation among users in the MU transmission leads to significantly lower data rates.

An even more severe throughput drop for the random selection is observed in Fig. 34, which shows the results for the case of 6 correlated users in the network. For example, the system throughput is around 235 Mbps in the 3-correlated-user scenario if ρ equals 0.6 and drops to 160 Mbps when the number of correlated users increases to six. This is because with six as opposed to three correlated users, there is a higher chance that the random selection would group correlated users together for MU transmission. Fig. 34 clearly indicates the benefit of flexible group sizes: The throughput performances of all approaches degrade as the value of channel correlation index ρ increases. Random selection always chooses maximum group sizes, which leads to a considerably higher relative throughput loss than other approaches for $\rho \geq 0.2$. This is so because GMA, SUS, and ZFS avoid grouping the correlated user together and, thus, MU_2 or even SU transmission may be preferred over MU_3. The degradation from MU_3 to MU_2 and/or SU leads to the system throughput drops for GMA, SUS, and ZFS. Among all the analyzed algorithms, the

proposed GMA performs closest to optimum and achieves at least 93% of the optimal system throughput.

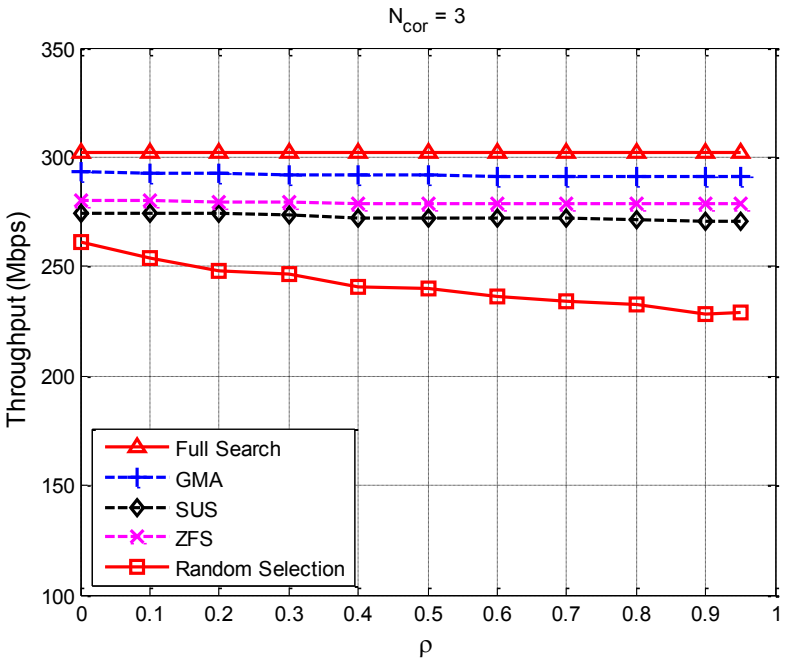


Fig. 33. Throughput as a function of ρ when there are 3 correlated users in the network.

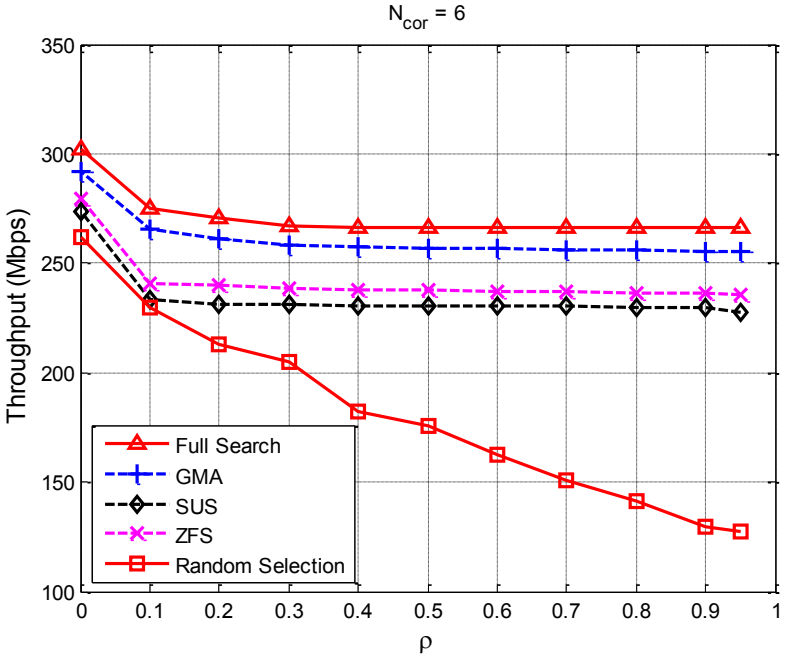


Fig. 34. Throughput as a function of ρ when there are 6 correlated users in the network.

We then measured the runtime of grouping functionality among all the approaches. To compare the complexity of the approaches, we set the random selection as the baseline. Fig. 35 plots the relative complexity with respect to random selection in dB.

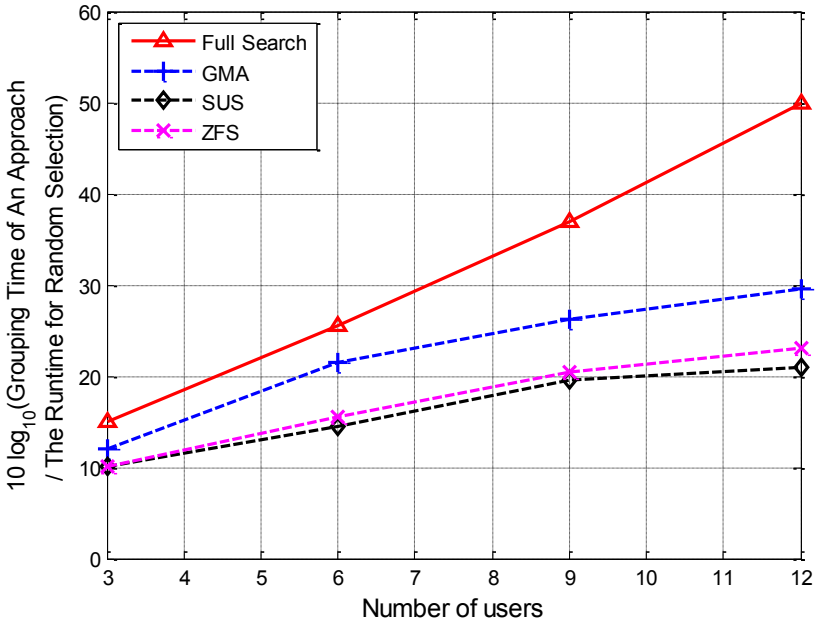


Fig. 35. Relative complexity with respect to random selection vs. network size.

As expected, full search needs considerably more time to complete and is not a feasible solution when the network size increases. Whereas GMA requires more time than SUS and ZFS, all the heuristic algorithms increase with the network size much slower than full search as they are all polynomial time solution. Notice that the Blossom and Hungarian algorithms can be parallelized as investigated in [132], [133]. Using Application-Specific Integrated Circuits (ASICs) for hardware acceleration, the runtime can be further reduced.

Notice that in this chapter, similar to [89], [88], [120], [92], [121], we set the AP to acquire the channel state information (CSI) of all active users before performing downlink MU transmission in a TXOP period. The focus of this work is to provide a low-complexity grouping algorithm to

effectively select user(s) for single user transmit beamforming or multi-user MIMO transmission. The proposed grouping strategy works with any sounding interval and independent of how the sounding interval is decided. Large-scale networks, which are beyond the scope of this work, can be divided into a few small scale network sets to be served in turn and apply the proposed grouping strategy.

4.5 Conclusion and Future Works

This work formulates the optimization problem for choosing among different MU-MIMO transmission modes and SU-TxBF for practical WLAN deployments for emerging indoor-outdoor scenarios. A hypergraph matching-based solution is proposed and demonstrated to be effective for solving the user grouping problem that maximizes throughput while providing MU-MIMO air time fairness. The optimal grouping result is first derived for the 2-user group case, followed by an extension to the more general case where any number of users can be assigned to groups of different sizes as a function of the radio environment. A computationally efficient approach is proposed for practical system implementation. Using our collected indoor channel measurements, we have evaluated the proposed algorithms under the specification framework of IEEE 802.11ac. The results demonstrate the suitability of the algorithms for low and high correlated channels, emulating indoor and outdoor practical WLAN deployments. Our proposed algorithm achieves at least 93% of the optimal system throughput in all test cases and is within 98% of the optimal solution for the conducted indoor experiments. The work can be extended for analyzing the system delay in scenarios where users have different traffic patterns and queuing lengths. There are further two directions to continue the work. One is to design and implement the parallelized version of the proposed algorithm for ASIC design. The second is to jointly analyze the user grouping and sounding problem as well as the implications of large-scale networks.

Chapter 5

Spectrum Utilization in Prioritized OFDMA System in WLANs

In this chapter, we develop a prioritization scheme for the spectrum utilization in prioritized OFDMA WLAN systems. Our contributions are: (1) Formulate a framework for resource allocation constrained by User Priority Scheduling (UPS) criterion that maximizes the number of the satisfied users while maintaining user priorities. (2) Analyse the solution boundary using a linearized scheduler algorithm so that it can be addressed using a mixed integer linear programing. (3) Develop a low-complexity scheduler for practical real-time implementation that achieves near optimal performance in most scenarios.

The outline of this chapter is as follows. In Section 5.1, the system description is introduced. In Section 5.2, we present an optimization framework for prioritized OFDMA system and a low-complexity scheduler for prioritized WLAN system. The simulation results are presented in Section 5.3, and Section 5.4 concludes this work.

5.1 System Model

5.1.1 Communication Protocol

The mandatory part of the current 802.11 MAC is called the distributed coordination function (DCF), which is based on Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA). The multi-user transmission feature at the MAC layer level requires the inclusion of Enhanced Distributed Channel Access (EDCA) Transmission Opportunity (TXOP) and the Aggregated MAC Protocol Data Unit (A-MPDU) with Compressed Block ACK mechanism. EDCA enables data transmission consecutively as long as the whole transmission time does not exceed the EDCF TXOP limit, which is determined and announced by the access point (AP). In the IEEE 802.11ac standard, the downlink multi-user MIMO transmission is orchestrated according to the TXOP sharing rules. For downlink OFDMA functionality proposed in 802.11ax, the TXOP mechanism is similar to [134], [135], [136] , where the AP attempts to reserve radio resources according to the traditional DCF on behalf of all the stations. After obtaining the resources, the AP collects the channel states information (CSI) of each station, and allocates subsets of these radio resources—essentially OFDM subcarriers and time slots—to different stations for simultaneous transmission during the TXOP.

5.1.2 Radio Resource Management Considerations

Insuring prioritization is the key consideration and hard requirement for formulating our radio resource management algorithm. Yet, it is also important to satisfy as many user requirements as possible. Thus, devices with higher priority should access the channels before those with lower priority when resources are not enough.

5.2 Scheduling Strategy Design

Consider a WLAN system with one AP, a set of N stations and M available subcarriers. Each TXOP is divided into K time slots. The requirement for each station's data rate depends on (1) its delay requirement and (2) the traffic amount in its buffer queue. Therefore, the requirement R_i for each station can be calculated by the AP when starting TXOP transmission.

In this section, we first formulate the prioritized OFDMA system as a non-linear optimization problem. To simplify the calculation, we then find the equivalent linear formulation by adding auxiliary variables. A low complexity approach is proposed at the end of this section. Table 19 summarizes all notations adopted in the model.

Table 19 Symbols Used for Problem Formulation.

Symbol	**Description**
$\mathcal{N}$	*Set of stations* $\{1, 2, \dots, N\}$
$\mathcal{M}$	*Set of subcarriers* $\{1, 2, \dots, M\}$
K	*Number of time slots in each TXOP*
$r_{i,j}$	*Real transmission rate towards station* i *on subcarrier* j
$a_{i,j,k}$	*Binary indicator to mark whether subcarrier* j *at slot* k *is assigned to station* i
r_i^*	*Station* i*'s real transmission rate in TXOP*
R_i	*Station* i*'s required transmission rate in TXOP*
p_i	*The priority ranking of station* i *(integer values)*
β_i	*Shortage below the target rate* R_i *for station* i
α_i	*Surplus above the target rate* R_i *for station* i

5.2.1 Non-prioritized OFDMA System

To maximize the system throughput of the OFDMA system in each TXOP, the objective function can be formulated as

$$\sum_{i=1}^{N}\sum_{j=1}^{M}\sum_{k=1}^{K} r_{i,j}a_{i,j,k} \ . \tag{10}$$

where, $r_{i,j}$ indicates the effective transmission rate towards station i on subcarrier j based on the subcarrier state delivered from the CSI feedback of each station, and $a_{i,j,k} \in \{0,1\}$ indicates whether subcarrier j at slot k is assigned to station i.

Since $M$is the total number of available subcarriers to be assigned to stations in each time slot, the frequency resource constraint becomes

$$\sum_{i=1}^{N}\sum_{j=1}^{M} a_{i,j,k} \leq M, \forall k = 1,2, \dots K. \tag{11}$$

There are K time slots in each TXOP, and then the time resource constraint is correspondingly,

$$\sum_{k=1}^{K} a_{i,j,k} \leq K, \forall i \in \mathcal{N}, \forall j \in \mathcal{M}. \tag{12}$$

We classify station i as satisfied if its actual data rate r_i^* meets its requirement R_i:

$$r_i^* = \sum_{j=1}^{M}\sum_{k=1}^{K} r_{i,j}. a_{i,j,k} \geq R_i \ , \forall i \in \mathcal{N}. \tag{13}$$

In summary, the problem in the OFDMA system is to maximize the objective function given in (10) subject to constraints (11), (12) and (13). Notice that this is without the consideration of the prioritized application and users.

This optimization problem does not always have a feasible solution, as the service requirements from the stations can be in excess of the limited radio resources. The current WLAN system typical provides best effort for all the users, and every user will be degraded. Therefore, an alternative priority based OFDMA optimization scheduling considering the prioritized users and applications is required.

5.2.2 Prioritized OFDMA System

To this end, the ***User Priority Scheduling (UPS)*** criterion is proposed as a set of rules:

Rule 1: Time-frequency resource is allowed to be assigned to station i if and only if all the requirements for stations with higher priority than station i are satisfied.

Rule 2: The number of stations being satisfied is as large as possible under the constraint of Rule 1.

Rule 1 guarantees the strict priority order among users, while Rule 2 balances the resource allocation in order to maximize the user satisfaction capacity, defined as the number of the satisfied users following Rule 1. This is essentially a hierarchical scheduling system, where users are treated equally at each level of priority and the scheduler works its way down the hierarchy. The scheduler that satisfies the UPS criterion is termed as the prioritized scheduler. The optimal scheduling decision in each TXOP guarantees (a) the prioritized services for each individual user, and (b) the maximum number of satisfied users under the priority order.

Particularly for our prioritized OFDMA system, $p_i \in \mathcal{N}$ denotes the priority ranking of station i. Without loss of generality, we assume that station s has a higher priority than station t if $p_s < p_t$. If r_s^* and r_t^* represent the net transmitted data amount during the TXOP for station s and station t,

$$(r_s^* - R_s) \cdot r_t^* \geq 0,\ r_t^* \geq 0, \tag{14}$$

provides the formal description of Rule 1 in the UPS criterion. The inequalities in (14) guarantee that r_t^* is zero if $r_s^* < R_s$, which means that resource for the lower priority station t cannot be allocated unless the required data rate for the higher priority station s is satisfied. If $r_s^* \geq R_t$, station t may be allocated resources. Therefore, (14) constrains the scheduler to allocate the resources in the order of priority.

The objective function (10) aims at maximizing the total data rate, which may lead to the unbalanced allocation of resources. In the extreme case, the scheduler may assign all resources to only one or two stations. If the highest priority user happens to have the highest data rate, the network throughput would be maximized. However, the scheduler is expected to satisfy as many stations' requirements as possible subject to the priority order. Thus, we formulate the problem with an alternative objective function as required by Rule 2 in the UPS criterion.

We denote $\beta_i \geq 0$ as the data-rate shortage below the target data rate R_i for station i and $\alpha_i \geq 0$ as the surplus above R_i. Then,

$$-\beta_i \leq (\sum_{j=1}^{M} \sum_{k=1}^{K} r_{i,j} a_{i,j,k}) - R_i \leq \alpha_i \ , \ \forall i \in \mathcal{N}, \tag{15}$$

where $\alpha_i \geq 0$, $\beta_i \geq 0$, $i \in \mathcal{N}$. In order to maximize the number of users being satisfied under the constraint of Rule 1, the overall data rate shortage should be minimized. This mitigates the resource imbalance while maximizing the user satisfaction capacity following Rule 1. The alternative problem formulation for guaranteeing the UPS criterion in the prioritized OFDMA system then becomes

$$min \ \sum_{i=1}^{N} \beta_i, \tag{16}$$

subject to (11), (12), (15) and priority constraint using (14). If the radio resources are sufficient and of high channel quality, the objective value of (16) can be zero, which means that all station requirements are satisfied. In such a case, an additional optimization stage can be easily integrated into the scheduler to optimize other desired system parameters, which is beyond the scope of this work.

5.2.3 *Linearized Problem Formulation of the Prioritized OFDMA System*

One issue of the prioritized OFDMA system scheduling as defined in the previous section is the nonlinearity of the problem, making it difficult to be solved in general. Specifically, expanding the first term in (14), we get

$$r_s^* r_t^* = \sum_{j_s=1}^{M} \sum_{k_s=1}^{K} \sum_{j_t=1}^{M} \sum_{k_t=1}^{K} r_{s,j_s} . r_{t,j_t} . (a_{s,j_s,k_s} . a_{t,j_t,k_t}), \quad (17)$$

which is non-linear due to the product term $a_{s,j_s,k_s} . a_{t,j_t,k_t}$. We transform the above problem into an equivalent linear problem by introducing auxiliary variables using Lemma 1.

Lemma 1: Let a, b, w to be binary integers, $w = a \cdot b$ if and only if $w \leq a$, $w \leq b$ and $w \geq a + b - 1$.

Proof: Necessity: $0 \leq a \leq 1, 0 \leq b \leq 1,$ so $a \cdot b \leq a$. Since $w = a \cdot b$, we have $w \leq a$. Similarly, $w \leq b$. $0 \leq a \leq 1, 0 \leq b \leq 1$, so $(a-1)(b-1) \geq 0$. Thus, we have $a \cdot b - a - b + 1 \geq 0$ and $a \cdot b \geq a + b - 1$. We know $w = a \cdot b$, so $w \geq a + b - 1$. **Sufficiency:** Suppose the proposition is false. Thus, $w > a \cdot b$ or $w < a \cdot b$. When $> a \cdot b$, if $a = 1, b = 1$, then $c \leq 1$ and $c > 1$; thus, there's no solution for c. When $< a \cdot b$, if $a = 1, b = 1$, then $c \geq 1$ and $c < 1$. Thus, there's no solution for c, either. This leads to a contradiction.

Thus, we can substitute the term $a_{s,j_s,k_s} . a_{t,j_t,k_t}$ in (17) by w_{s,j_s,k_s,t,j_t,k_t} to obtain

$$r_s^* r_t^* = \sum_{j_s=1}^{M} \sum_{k_s=1}^{K} \sum_{j_t=1}^{M} \sum_{k_t=1}^{K} r_{s,j_s} . r_{t,j_t} . w_{s,j_s,k_s,t,j_t,k_t} , \quad (18)$$

with the additional constraints:

$$w_{s,j_s,k_s,t,j_t,k_t} \leq a_{s,j_s,k_s} \quad , \quad (19)$$

$$w_{s,j_s,k_s,t,j_t,k_t} \leq a_{t,j_t,k_t} \quad , \quad (20)$$

$$w_{s,j_s,k_s,t,j_t,k_t} \geq a_{s,j_s,k_s} + a_{t,j_t,k_t} - 1 . \quad (21)$$

The inequality in (14) then becomes

$$\sum_{j_s=1}^{M} \sum_{k_s=1}^{K} \sum_{j_t=1}^{M} \sum_{k_t=1}^{K} r_{s,j_s} r_{t,j_t} w_{s,j_s,k_s,t,j_t,k_t} - R_s r_t^* \geq 0 . \quad (22)$$

The actual transmission rate during the TXOP for any station should be nonnegative, i.e.,

$$r_i^* \geq 0, \forall i \in \mathcal{N} \tag{23}$$

Thus, the optimization problem is transformed to maximize objective function (8) subject to constraints (11), (12), (15), and (19) - (23). Thus, the problem formulation becomes the well-known mixed integer linear program (MILP) [137]. For solving the MILP, there exist highly efficient off-the-shelf solvers, such as CPLEX [138]. We use this CPLEX solver to find the optimal solution and upper bound for the prioritized scheduling problem.

5.2.4 A Low Complexity Solution

Notice that the auxiliary variables generated from (19) - (21) increases the number of the decision variables of the optimization problem. To reduce the complexity for practical implementation, we propose an efficient algorithm to find a solution. The basic idea of the heuristic approach is to divide the users into groups according to their priority levels, and the resource allocation is performed from the highest priority group until the lowest priority group. Specifically, the users are grouped into different sets G_u, where $p_i < p_j$ if station $i \in G_u$, $j \in G_v$, and $u < v$.

Starting from the highest priority group, the allocation is preformed to satisfy the requirement of the users in the current group while minimizing the number of the resource blocks used. We denote the already-allocated resource block by $b_{j,k} \in B$ if subcarrier j at slot k has been allocated for higher priority users. Thus, the problem of the resource allocation for each group can be formulated as

$$\text{min} \qquad \sum_{g \in G_u}^{|G_u|} \sum_{j=1}^{M} \sum_{k=1}^{K} a_{g,j,k} \,, \tag{24}$$

$$\text{s.t.} \quad r_g^* = \sum_{j=1}^{M} \sum_{k=1}^{K} r_{i,j} . a_{g,j,k} \geq R_g \,, \forall g \in G_u. \tag{25}$$

$$a_{i,j,k} = 0, \text{ if } b_{j,k} \in B, \forall i \in \mathcal{N} \tag{26}$$

Table 20 Algorithm # 3.

Heuristic Algorithm for Prioritized Scheduling

1 Calculate the data rate requirement for each user;
2 $u = 1$;
3 $G_u = \emptyset$;
4 Sort stations in set S according to stations priority.
5 $p_t = 1$;
6 **for** $q = 1{:}\,1{:}\,|S|$
7 **if** $p_q = p_t$
8 $G_u = G_u \cup \{q\}$;
9 else
10 $u = u + 1$;
11 $G_u = \{q\}$;
12 **end if**
13 **end for**
14 **while** $d = 1{:}\,1{:}\,u$
15 Solve the problem of min Eq. (24) s.t. (25) and (26);
16 **if** the problem has a feasible solution
17 Update the already-allocated block set B;
18 **else**
19 **for** $g = 1{:}\,1{:}\,|G_d|$
20 Solve the problem of min Eq. (24) s.t. (26) and $r_g^* = \sum_{j=1}^{M} \sum_{k=1}^{K} r_{i,j}.a_{g,j,k} \geq R_g$;
21 **end for**
22 break;
23 **end if**
24 **end while**

If the optimization problem does not have a feasible solution because of the insufficient radio resources, the available resource blocks will be allocated for the users one by one in the current group on a random basis. The detailed algorithm is provided in Table 20. Lines 1-13 represent the division of the users into groups according to their priority levels. Lines 14-24 describe the resource allocation operation from the highest priority group until the lowest priority group.

5.3 System Evaluation

In this section, we compare the performance of the proposed optimal scheduler method, the proposed low-complexity scheduler method, and the conventional method as discussed in section

5.2, which only try to maximize the system throughput, but it doesn't consider the user priority. The method with cognitive radio network's concept [118], which divides the users into two priority levels of users, is shown as a special case from the simulation results below. The evaluation goals in this section are twofold. First, we show the optimal scheduler's capability of satisfying the data rate requirements for the prioritized stations and its performance gain over the conventional method. Second, we provide insights on the impact of different environmental parameters on the performance of the proposed scheduling strategy.

5.3.1 System Settings

Consider a WLAN system consisting of N labeled stations with decreasing priorities (i.e., Station i's priority > Station j's priority, for $i < j$). We assume the received signal on each channel of each user undergoes Rayleigh fading. The average SINR for any station among the N stations is between 1 dB and 12 dB. It is assumed that all the channels are uncorrelated and the SINR for each station is independent and constant within one TXOP. The simulation results are averaged over 500 channel realizations. The simulation parameters are listed in Table 21.

Table 21 Simulation Parameters.

Parameters	**Values**
TXOP duration	3ms
MPDU length	1556 bytes
MSDU length	1508 bytes
Bandwidth	40 MHz
Number of sub-channels	8
Number of OFDM subcarriers	104
Number of subcarriers per sub-channel	13 (12 subcarrier for data and 1 for pilot)

We consider a hospital scenario first with Station 1 to 5 being the devices of life-critical device for patients used by hospital staff, non-life-critical devices used by hospital staff, device used by hospital staff for business administration or emails with large attachment, device of patient, and device of guest, respectively. Their ranking and data rate requirements are listed in Table 22 (adapted from [139]). Then, we evaluate the impact of the network size on the performance of the designed scheduling strategy. To eliminate the effects of other parameters, the data rate requirement of stations are set to be the same (20 Mbps).

Table 22 Examples Device Category with Different Transmission Priority.

Ranking	**Device Description**	**Data Rate**
1	Life-critical devices for patients used by hospital staff.	20 Mbps
2	Non-life-critical devices used by hospital staff.	20 Mbps
3	Devices used by hospital staff for business administration, emails with large attachment, etc.	20 Mbps
4	Devices of patient	20 Mbps
5	Devices of guests	20 Mbps

5.3.2 Results

The Quality of service (QoS) rate for a station is defined as the ratio of the assigned data rate over its required data rate. Fig. 36 shows one case of the scheduling result of QoS rates for the stations under different SINR conditions in hospital scenario with five prioritized devices. It can be observed that the optimal approach (OA) and sub-optimal approach (SA) strictly guarantee the priority order of the stations, while the conventional one does not have the priority guarantee across the network. For OA and SA, no resources are allocated for the station with lower priority if the QoS rate of the higher priority stations does not reach 100%. If SINR of the environment is 12 dB, all stations' requirements can be satisfied for OA, SA and the conventional scheme. For the radio environment with 9 dB SINR, OA still satisfies 4 stations' requirements, while the conventional

one does not cover any station's requirement fully, even though its total data rate is slight higher. SA in this case covers 4 stations fully with nearly similar performance as OA. If SINR decays further to 6 dB, OA can provide full rate guarantee for the first 3 stations with 46% gap for the station 4, while SA can also cover three stations with a very slight difference from OA. If SINR becomes 3 dB, the SA can provide enough date rate for the first station with 14% gap from the second station, while the first two stations can be fully satisfied by OA.

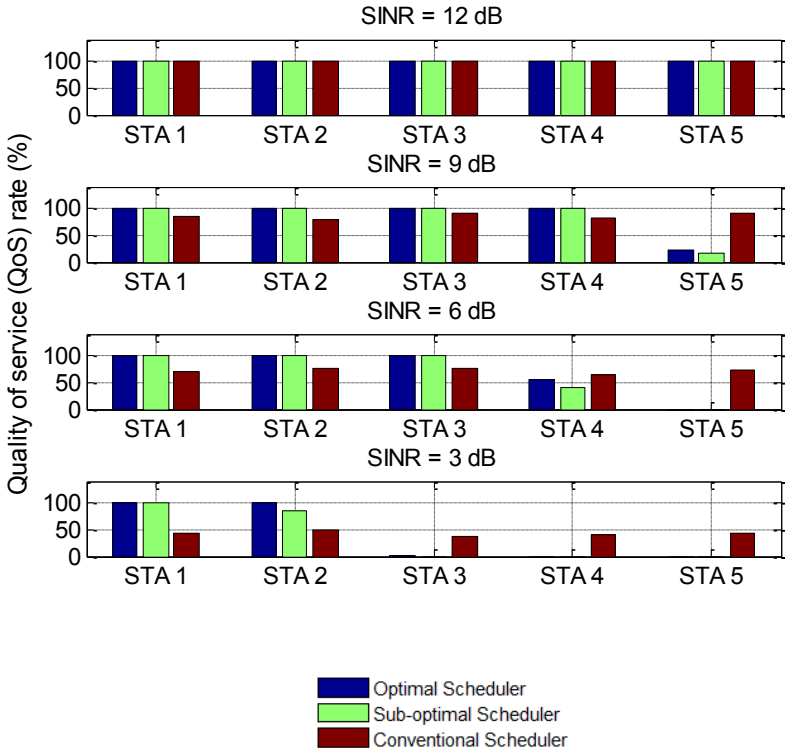


Fig. 36. Example case of QoS rates for the stations.

The average performance gap ratio is defined as the network QoS rate (addition of the individual station's QoS) gap of OA over that of SA. Fig. 37 shows how the average performance gap ratio varies with SINR conditions. It is observed that the difference between OA and SA changes with different SINR conditions. If SINR ranges from 2 dB to 8 dB, the gap becomes large. For example, the average gap is around 9% when the SINR is 4 or 5 dB. For the high or low SINR

scenarios, both the schedulers work well, and SA can achieve similar performance as OA. For example, if the SINR is very high (i.e., greater than 8 dB in this scenario), all five stations' data rate requirements are satisfied by OA; thus there's no data rate increase when SINR increases. In contrast, the total data rate by SA keeps rising. This is the reason why the total data rates of the two approaches gradually come closer to each other. If the SINR is low (i.e., less than 3 dB in this scenario), the difference between the two approaches decreases. This is because the amount of effective resource becomes less for the schedulers to assign.

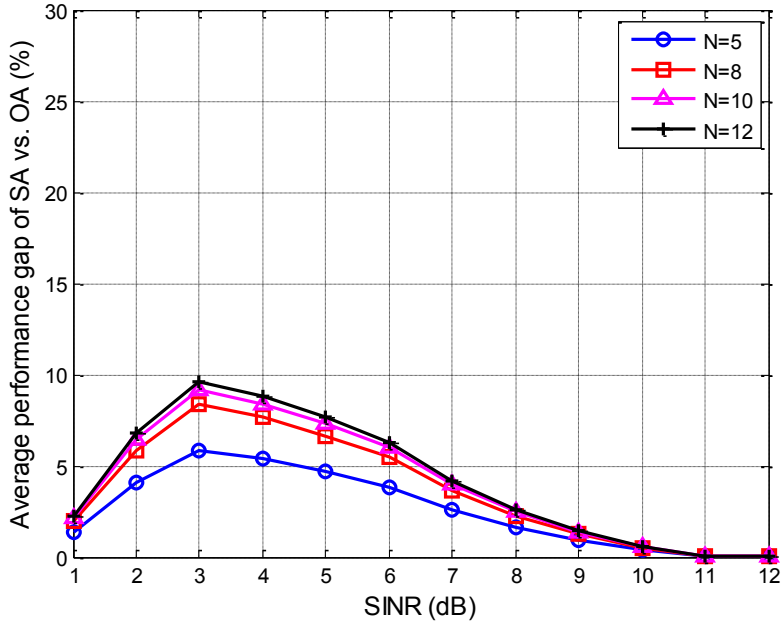


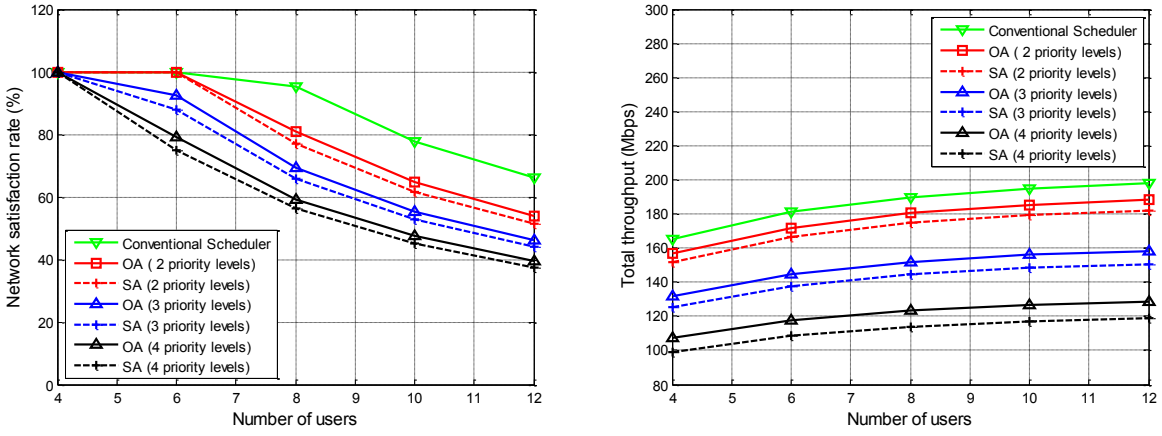
Fig. 37. Average gap ratio as a function of SINR.

It is observed in Fig. 37 that the gap rate increase with the number of users in the network. This can be interpreted as follows: when increasing the number of users in the network, the diversity of the resource blocks increases which directly affects the scheduling decisions and improves the over system performance. It is also noted that the gap is bounded by the larger number of users in

the network because of the limited radio resource which is shared by the increasing number of users.

Fig. 38 shows the system performance as a function of the number of users. Particularly, as shown in Fig. 38(a), the network satisfaction rate (total assigned data rate over the total required data rate) increases with the number the users under each approach. The decrease comes from the increasing users sharing the fixed amount of the radio resources, although the total network throughput increases with the user number as shown in Fig. 38(b) due to the user diversity of the OFDMA system.

It is observed that, without user priority consideration, both the highest network satisfaction rate and total network throughput can be achieved by all the approaches given the number of users in the network. The performance decrease when increasing the number of the priority levels. This can be explained from the formulation of the prioritized scheduling that adding more constraints to the problem decreases the optimal objective value. SA results in slightly lower throughput compared with OA, however, the gap is always within 10% for all our test cases.



(a) Network satisfaction rate (b) Total network throughput

Fig. 38. System performance as a function of the user number.

To compare the complexity of the approaches, we set the conventional method as the baseline. Fig. 39 plots the relative complexity with respect to the conventional method in dB. For multiple priority levels, as expected, OA's computation increases exponentially and is not a scalable approach. The difference between SA and the conventional method is quite small, and SA for multiple priority levels can be even slightly better than the conventional method. This is because even though SA requires multiple rounds of optimization solving, the size of the optimization solving in each round is smaller than the conventional method. Notice that the complexity of SA for 1 priority level is at the same level as the conventional approach.

In summary, SA is efficient solution in terms of (1) total throughput and network satisfaction rate (less than 10% from OA), and (2) algorithm complexity (within the same magnitude order of the conventional method).

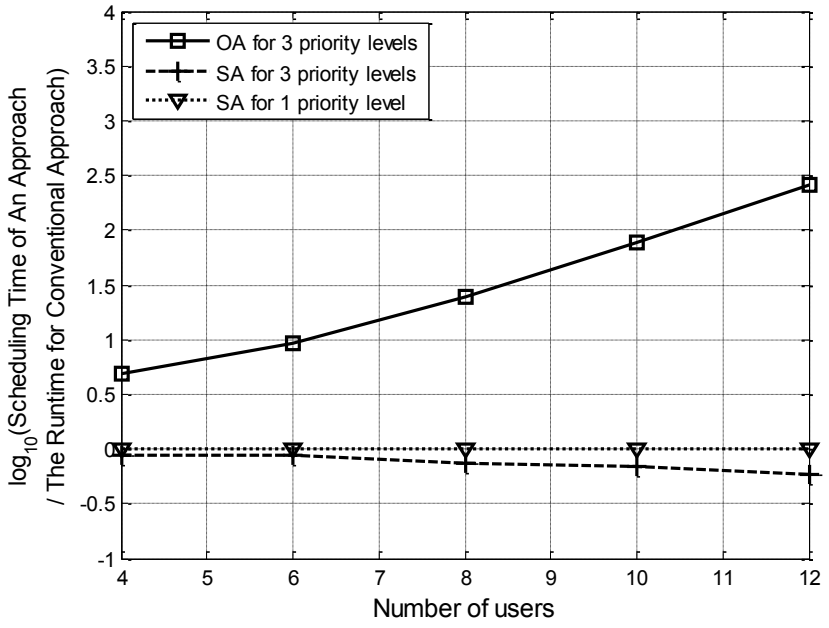


Fig. 39. Relative complexity with respect to the conventional method vs. network size.

5.4 Conclusion and Future Works

We have investigated the spectrum utilization for OFDMA systems in WLAN and characterized the behavior and constraints of prioritized wireless networks. We have proposed a User Priority Scheduling (UPS) criterion that aims at maximizing the number of satisfied users under the strict priority order. Using this criterion, an optimization framework for the prioritization of wireless networks has been proposed. The simulations results have indicated consistency with our analytical findings and illustrate the advantages of the proposed prioritized scheduling method. Our future work will focus on generalizing this framework to other wireless systems, such as LTE-A networks, to provide prioritized service for various applications, including public safety, healthcare, premium and general subscribers, and cognitive or policy based radios.

Chapter 6

Summary and Future Works

6.1 Summary for Channel Sounding

In this work, we focus on the two major features of the new generation of WLAN, MU-MIMO and OFDMA, to improve system throughput and efficiency.

MU-MIMO is a promising technology to increase the system throughput of WLAN systems. Channel sounding strategy requires balancing channel sounding overhead with CSI inaccuracy. Less frequent sounding causes more cross-user interference because of the out-dated CSI; whereas too frequent sounding reduces the effective transmission time. The problem thus consists in finding a suitable operational point that trades the CSI accuracy for effective transmission time for the current radio environment. We investigated channel sounding issues and answer the practical questions for IEEE 802.11ac and future WLAN systems including (1) what is the optimal sounding interval in terms of the system throughput? And (2) How to dynamically adapt the sounding interval for rapidly changing radio environments?

Our major contributions of the dynamic channel sounding work are:

(1) We characterize and evaluate the selection of the channel sounding interval for single user transmit beamforming (SU-TxBF) and MU-MIMO in WLAN. We conduct extensive channel

measurements in both the static and dynamic indoor scenarios through test nodes with Qualcomm 802.11ac Chipset QCA9980. We demonstrate the existence of an optimal sounding interval for given channel conditions.

(2) We propose a low-complexity dynamic sounding approach to improve the effective throughput in the real indoor dynamic environment. We evaluate the proposed approach through using an 802.11ac emulator that uses our measured channel information. We show that our proposed approach can achieve up to 31.8 % throughput improvement in the real indoor environment.

6.2 Summary for User Grouping

In a typical WLAN scenario, it is the access point (AP) that directs simultaneous multi-stream transmissions to multiple stations. The challenge is that the number of mobile stations needing to be served can be much larger than the number of transmit antennas, N_t, at the AP. In such loaded WLAN systems, the maximum allowed number of users per multi-user (MU) group, N_u, that the AP can serve simultaneously is less than the number of mobile stations in the network. Thus, selecting sets of user(s) using SU-TxBF or MU-MIMO transmission along with scheduling all these groups over successive time slots is essential for achieving high system throughput while guaranteeing user fairness. The problem consists of deriving an optimal method for the selection of SU-TxBF or MU-MIMO transmission that maximizes the downlink system throughput with MU-MIMO user fairness. Our objective is to develop an efficient AP transmission selection scheme and MU group assignment under the MU-MIMO air time fairness criterion. This requires dynamically assigning the stations to groups for time slotted transmission.

Our major contributions of the user grouping work are:

(1) We model the user grouping problem using a hypergraph and show that the maximum hypergraph matching provides the optimal solution for choosing between SU-TxBF and MU-MIMO. We demonstrate an approach to determine how many users and which users to assign to the MU groups to maximize system throughput subject to MU-MIMO air time fairness.

(2) We develop an efficient and scalable algorithm based on graph matching for solving the above problem and evaluate its performance using an 802.11ac emulator that uses collected channel measurements in an office environment and simulated outdoor channles. The results show that the proposed algorithm achieves at least 93% of the optimal system throughput, which outperforms today's state-of-art algorithms.

6.3 Summary for Priority Scheduling

To increase the system throughput and spectrum efficiency, orthogonal frequency-division multiple access (OFDMA) will play an important role in the next generation of WLAN systems. However, maximizing the number of the satisfied users in the OFDMA system while insuring different Quality of Service (QoS) levels is very challenging. Specifically, the problem is how to maximize the number overall number of satisfied users under strict priority constraints (guarantees for service requests of the spectrum and time resources for selected users having higher priority before accommodating those of lower priority).

Our major contributions of the user scheduling work are:

(1) We formulate a framework for resource allocation constrained by User Priority Scheduling (UPS) criterion that maximizes the number of the satisfied users while maintaining user priorities.

(2) We analyse the solution boundary by linearizing the scheduler algorithm so that it can be addressed using efficient mixed integer linear programing.

(3) We develop a low-complexity scheduler for practical real-time implementation that achieves near optimal performance in most scenarios.

6.4 Future Works

MU-MIMO and OFDMA in further generations of WLANs or/and other commercial systems provide very rich topics for future research.

Particularly for channel sounding, further works includes, but not limited to: (1) Analyze the impact of channel sounding (sounding interval/ quantization) on joint OFDMA and MU-MIMO systems; (2) Analyze the impact of channel sounding on the performance of the incoming spectrum sharing systems such as LTE-A and LTE-U; (3) Design efficient and dynamic algorithms for such systems, etc. As for user grouping, further works includes, but not limited to: (1) Extend the MU-MIMO grouping work to analyze the system delay theoretically in scenarios where users have different traffic patterns and queuing lengths; (2) Apply our design principle to other commercial MU-MIMO / massive MIMO systems for future networks for sharing spectrum; (3) Design and implement the parallelized version of the proposed algorithm to further reduce algorithm time complexity; (4) Jointly analyze the user grouping and sounding problem as well as the implications of large-scale networks, etc. Example scheduling design includes but not limited to: (1) joint OFDMA and MU-MIMO systems scheduling with guaranteed QoS; (2) Joint downlink/uplink OFDMA scheduling design, etc. In addition, new features in WLAN, such as

6.5 List of All Publications

6.5.1 Journal Manuscripts Under Review

1. Xiaofu Ma, Ji Wang, Qinghai Gao, Vuk Marojevic, and Jeffrey H. Reed, “Dynamic Sounding for Multi-User MIMO in Wireless LANs”, under review of *IEEE Transactions on Wireless Communications*.

2. Xiaofu Ma, Miao Yao, Munawwar Sohul, Qinghai Gao, Vuk Marojevic, and Jeffrey H. Reed, "Efficient Spectrum Utilization in Prioritized OFDMA-Based Wireless LANs", under review of *IEEE Transactions on Vehicular Technology*.

6.5.2 Published Journal Papers

1. Xiaofu Ma, Qinghai Gao, Vuk Marojevic, and Jeffrey H. Reed, "Hypergraph Matching for MU-MIMO User Grouping in Wireless LANs" [140], *Elsevier Journal of Ad Hoc Networks, Vol. 48, pp. 29-37, 2016*.
2. Munawwar M. Sohul, Miao Yao, Xiaofu Ma, Eyosias Y. Imana, Vuk Marojevic, Jeffrey H. Reed, "Next Generation Public Safety Networks: A Spectrum Sharing Approach" [141], *IEEE Communication Magazine, vol. 54, no. 3, pp. 30-36, 2016.*
3. Stephen M. Dudley, William C. Headley, Marc Lichtman, Eyosias Imana, Xiaofu Ma, Mahi Abdelbar, Aditya Padaki, Abid Ullah, Munawwar Sohul,Tae-young Yang, and Jeffrey H. Reed, "Practical Issues for Spectrum Management with Cognitive Radios" [142], *IEEE Proceedings, vol. 102, no. 3, pp. 242-264, 2014.*

6.5.3 Published Conference Papers

1. Xiaofu Ma, Sayantan Guha, Junsung Choi, Christopher R.Anderson, Randall Nealy, Jared Withers, Jeffrey H.Reed, Carl Dietrich, "Analysis of Directional Antenna for Railroad Crossing Safety Applications" [143], *IEEE CCNC,* 2017.
2. Xiaofu Ma, Thaddeus Czauski, Taeyoung Yang, and Jeffrey H. Reed, "Wirelessly Sensing Medication Administration" [144], *ACM Wireless Health, pp. 1-8, 2014.*
3. Xiaofu Ma, Yujun Liu, Dr. Karen A. Roberto, Dr. Jeffrey H. Reed, "Using Wireless Sensing Technology for Medication Management by Older Adults" [145], Abstract accepted by *Gerontological Society of America's 68th Annual Scientific Meeting*.
4. Munawwar M. Sohul, Xiaofu Ma, Taeyoung Yang, Jeffrey H. Reed, "Quality of Service Assurance for Shared Spectrum Systems" [146], *IEEE MILCOM, pp. 1471-1476, 2014.*

5. Xiaofu Ma, Haris I. Volos, Xiangwei Zheng, Jeffrey H. Reed, and Tamal Bose, “A Variation-Aware Approach for Task Allocation in Wireless Distributed Computing Systems” [147], *IEEE GLOBECOM, pp. 5006-5011, 2013.*

Bibliography

[1] V. Jones and H. Sampath, "Emerging technologies for WLAN," *IEEE Communications Magazine,* vol. 53, pp. 141-149, 2015.

[2] P. Kinney, "Zigbee technology: Wireless control that simply works," in *Communications design conference*, 2003, pp. 1-7.

[3] C. Bisdikian, "An overview of the Bluetooth wireless technology," *IEEE Communication Magazine,* vol. 39, pp. 86-94, 2001.

[4] A. Al-Dulaimi, S. Al-Rubaye, Q. Ni, and E. Sousa, "5G Communications Race: Pursuit of More Capacity Triggers LTE in Unlicensed Band," *IEEE Vehicular Technology Magazine,* vol. 10, pp. 43-51, 2015.

[5] *Introducing MulteFire: LTE-like performance with Wi-Fi-like simplicity*. Available: https://www.qualcomm.com/news/onq/2015/06/11/introducing-multefire-lte-performance-wi-fi-simplicity

[6] *Juniper Research*. Available: http://www.juniperresearch.com/press/press-releases/wifi-to-carry-60pc-of-mobile-data-traffic-by-2019

[7] G. R. Hiertz, D. Denteneer, L. Stibor, Y. Zang, X. P. Costa, and B. Walke, "The IEEE 802.11 universe," *IEEE Communications Magazine,* vol. 48, pp. 62-70, 2010.

[8] Y. Zeng, P. H. Pathak, and P. Mohapatra, "A first look at 802.11ac in action: Energy efficiency and interference characterization," in *2014 IFIP Networking Conference*, 2014, pp. 1-9.

[9] M. Yazid, A. Ksentini, L. Bouallouche-Medjkoune, A. D, x00Ef, and ssani, "Performance Analysis of the TXOP Sharing Mechanism in the VHT IEEE 802.11ac WLANs," *IEEE Communications Letters,* vol. 18, pp. 1599-1602, 2014.

[10] K. Maraslis, P. Chatzimisios, and A. Boucouvalas, "IEEE 802.11aa: Improvements on video transmission over wireless LANs," in *2012 IEEE International Conference on Communications (ICC)*, 2012, pp. 115-119.

[11] A. Banchs, A. d. l. Oliva, L. Eznarriaga, D. R. Kowalski, and P. Serrano, "Performance Analysis and Algorithm Selection for Reliable Multicast in IEEE 802.11aa Wireless LAN," *IEEE Transactions on Vehicular Technology,* vol. 63, pp. 3875-3891, 2014.

[12] A. Babaei, J. Andreoli-Fang, Y. Pang, and B. Hamzeh, "On the Impact of LTE-U on Wi-Fi Performance," *International Journal of Wireless Information Networks,* vol. 22, pp. 336-344, 2015.

[13] S. Sagari, S. Baysting, D. Saha, I. Seskar, W. Trappe, and D. Raychaudhuri, "Coordinated dynamic spectrum management of LTE-U and Wi-Fi networks," in *2015 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN)*, 2015, pp. 209-220.

[14] J. Jeon, H. Niu, Q. C. Li, A. Papathanassiou, and G. Wu, "LTE in the unlicensed spectrum: Evaluating coexistence mechanisms," in *2014 IEEE Globecom Workshops*, 2014, pp. 740-745.

[15] C. Cano and D. J. Leith, "Coexistence of WiFi and LTE in unlicensed bands: A proportional fair allocation scheme," in *2015 IEEE International Conference on Communication Workshop (ICCW)*, 2015, pp. 2288-2293.

[16] K. Samdanis, T. Taleb, and S. Schmid, "Traffic offload enhancements for eUTRAN," *IEEE Communications Surveys and Tutorials,* vol. 14, pp. 884-896, 2012.

[17] D. Astely, E. Dahlman, G. Fodor, S. Parkvall, and J. Sachs, "LTE release 12 and beyond," *IEEE Communications Magazine,* vol. 51, pp. 154-160, 2013.

[18] H. ElSawy, E. Hossain, and M. Haenggi, "Stochastic Geometry for Modeling, Analysis, and Design of Multi-Tier and Cognitive Cellular Wireless Networks: A Survey," *IEEE Communications Surveys and Tutorials,* vol. 15, pp. 996-1019, 2013.

[19] Hu, Liang, et al. "Realistic indoor wi-fi and femto deployment study as the offloading solution to lte macro networks." *2012 Vehicular Technology Conference*, 2012.

[20] A. Hazmi, J. Rinne, and M. Valkama, "Feasibility study of IEEE 802.11ah radio technology for IoT and M2M use cases," in *2012 IEEE Globecom Workshops*, 2012, pp. 1687-1692.

[21] W. Sun, M. Choi, and S. Choi, "IEEE 802.11 ah: A long range 802.11 WLAN at sub 1 GHz," *Journal of ICT Standardization,* vol. 1, pp. 83-108, 2013.

[22] D. Jung, R. Kim, and H. Lim, "Power-saving strategy for balancing energy and delay performance in WLANs," *Computer Communications,* vol. 50, pp. 3-9, 2014.

[23] C. Ghosh, S. Roy, and D. Cavalcanti, "Coexistence challenges for heterogeneous cognitive wireless networks in TV white spaces," *IEEE Wireless Communications,* vol. 18, pp. 22-31, 2011.

[24] A. Eksim, S. Kulac, and M. H. Sazli, "Effective cooperative spectrum sensing in IEEE 802.22 standard with time diversity," in *International Conference on Advances in Computational Tools for Engineering Applications*, 2009, pp. 528-531.

[25] I. F. Akyildiz, B. F. Lo, and R. Balakrishnan, "Cooperative spectrum sensing in cognitive radio networks: A survey," *Physical Communication,* vol. 4, pp. 40-62, 2011.

[26] A. B. Flores, R. E. Guerra, E. W. Knightly, P. Ecclesine, and S. Pandey, "IEEE 802.11af: a standard for TV white space spectrum sharing," *IEEE Communications Magazine,* vol. 51, pp. 92-100, 2013.

[27] N. McKeown, T. Anderson, H. Balakrishnan, G. Parulkar, L. Peterson, J. Rexford*, et al.*, "OpenFlow: enabling innovation in campus networks," *SIGCOMM Computer Communication Review,* vol. 38, pp. 69-74, 2008.

[28] L. Suresh, J. Schulz-Zander, R. Merz, A. Feldmann, and T. Vazao, "Towards programmable enterprise WLANS with Odin," presented at the Proceedings of the first workshop on Hot topics in software defined networks, Helsinki, Finland, 2012.

[29] E. Telatar, "Capacity of multi-antenna gaussian channels," *European Transactions on Telecommunications,* vol. 10, pp. 585-595, 1999.

[30] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*: Cambridge university press, 2005.

[31] E. Perahia, "IEEE 802.11n development: history, process, and technology," *IEEE Communications Magazine,* vol. 46, pp. 48-55, 2008.

[32] R. Nee, "Breaking the gigabit-per-second barrier with 802.11ac," *IEEE Wireless Communications,* vol. 18, pp. 4-4, 2011.

[33] U. C. Lai and R. D. Murch, "A transmit preprocessing technique for multiuser MIMO systems using a decomposition approach," *IEEE Transactions on Wireless Communications,* vol. 3, pp. 20-24, 2004.

[34] C. Windpassinger, R. F. H. Fischer, T. Vencel, and J. B. Huber, "Precoding in multiantenna and multiuser communications," *IEEE Transactions on Wireless Communications,* vol. 3, pp. 1305-1316, 2004.

[35] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE Journal on Selected Areas in Communications,* vol. 21, pp. 684-702, 2003.

[36] G. Caire and S. Shamai, "On the achievable throughput of a multiantenna Gaussian broadcast channel," *IEEE Transactions on Information Theory,* vol. 49, pp. 1691-1706, 2003.

[37] E. Aryafar, N. Anand, T. Salonidis, and E. W. Knightly, "Design and experimental evaluation of multi-user beamforming in wireless LANs," in *Proceedings of the sixteenth annual international conference on Mobile computing and networking*, 2010, pp. 197-208.

[38] H. Yu, L. Zhong, A. Sabharwal, and D. Kao, "Beamforming on mobile devices: a first study," presented at the Proceedings of the 17th annual international conference on Mobile computing and networking, Las Vegas, Nevada, USA, 2011.

[39] W.-L. Shen, K. C.-J. Lin, S. Gollakota, and M.-S. Chen, "Rate adaptation for 802.11 multiuser MIMO networks," *IEEE Transactions on Mobile Computing,* vol. 13, pp. 35-47, 2014.

[40] E. Perahia and R. Stacey, *Next generation wireless LANs: 802.11n and 802.11ac*: Cambridge University Press, 2013.

[41] IEEE802.11-14/0165r0, "P802.11ax Task Group Press Release".

[42] IEEE802.11-14/0165r1, "P802.11ax Task Group Press Release".

[43] J. B. Andersen, "Array gain and capacity for known random channels with multiple element arrays at both ends," *IEEE Journal on Selected Areas in Communications,* vol. 18, pp. 2172-2178, 2000.

[44] L. Lanante, M. Kurosaki, and H. Ochi, "Non-linear precoding for next generation multi-user MIMO wireless LAN systems," in *2011 13th International Conference on Advanced Communication Technology (ICACT),* , 2011, pp. 1211-1216.

[45] A. Burg, M. Borgmann, M. Wenk, M. Zellweger, W. Fichtner, and H. Bolcskei, "VLSI implementation of MIMO detection using the sphere decoding algorithm," *IEEE Journal of Solid-State Circuits,* vol. 40, pp. 1566-1577, 2005.

[46] Z. Bai, B. Badic, S. Iwelski, T. Scholand, R. Balraj, G. Bruck*, et al.*, "On the Equivalence of MMSE and IRC Receiver in MU-MIMO Systems," *IEEE Communications Letters,* vol. 15, pp. 1288-1290, 2011.

[47] Y. Yokota and H. Ochi, "Complexity reduction for higher order MIMO decoder using block diagonalization," in *2013 International Symposium on Intelligent Signal Processing and Communications Systems (ISPACS)*, 2013, pp. 235-239.

[48] S. Umeda, S. Suyama, H. Suzuki, and K. Fukawa, "PAPR Reduction Method for Block Diagonalization in Multiuser MIMO-OFDM Systems," in *Vehicular Technology Conference (VTC 2010-Spring), 2010 IEEE 71st*, 2010, pp. 1-5.

[49] G. Li, S. Yang, X. Mu, and L. Qi, "A new method for peak-to-average power ratio reduction in MIMO-OFDM system," in *ISPACS 2007. International Symposium on Intelligent Signal Processing and Communication Systems, 2007.* , 2007, pp. 614-617.

[50] R. P. F. Hoefel, "Multi-user OFDM MIMO in IEEE 802.11ac WLAN: A simulation framework to analysis and synthesis," in *2013 IEEE Latin-America Conference on Communications*, 2013, pp. 1-6.

[51] R. Liao, B. Bellalta, J. Barcelo, V. Valls, and M. Oliver, "Performance analysis of IEEE 802.11ac wireless backhaul networks in saturated conditions," *EURASIP Journal on Wireless Communications and Networking,* vol. 2013, pp. 1-14, 2013.

[52] J. Xiong, K. Sundaresan, K. Jamieson, M. A. Khojastepour, and S. Rangarajan, "MIDAS: Empowering 802.11ac Networks with Multiple-Input Distributed Antenna Systems," in *Proceedings of the 10th ACM International on Conference on emerging Networking Experiments and Technologies*, Sydney, Australia, 2014.

[53] D. Nojima, L. Lanante, Y. Nagao, M. Kurosaki, and H. Ochi, "Performance evaluation for multi-user MIMO IEEE 802.11ac wireless LAN system," in *14th International Conference on Advanced Communication Technology (ICACT)*, 2012, pp. 804-808.

[54] C. Hunter, P. Murphy, and E. Welsh, "Demo: real-time MU-MIMO channel analysis with a custom 802.11 implementation," in *Proceedings of the 20th annual international conference on Mobile computing and networking*, Maui, Hawaii, USA, 2014.

[55] B. Bellalta, J. Barcelo, D. Staehle, A. Vinel, and M. Oliver, "On the Performance of Packet Aggregation in IEEE 802.11ac MU-MIMO WLANs," *IEEE Communications Letters,* vol. 16, pp. 1588-1591, 2012.

[56] J. Cha, H. Jin, B. C. Jung, and D. K. Sung, "Performance comparison of downlink user multiplexing schemes in IEEE 802.11ac: Multi-user MIMO vs. frame aggregation," in *2012 IEEE Wireless Communications and Networking Conference (WCNC)*, 2012, pp. 1514-1519.

[57] C. Chung, T. Chung, B. Kang, and J. Kim, "A-MPDU using fragmented MPDUs for IEEE 802.11 ac MU-MIMO WLANs," in *2013 IEEE International Conference of IEEE Region 10 (TENCON 2013)*.

[58] H. Suzuki, D. B. Hayman, J. Pathikulangara, I. B. Collings, Z. Chen, and R. Kendall, "Design Criteria of Uniform Circular Array for Multi-User MIMO in Rural Areas," in *2010 IEEE Wireless Communication and Networking Conference*, 2010, pp. 1-6.

[59] J. OuYang, F. Yang, and Z. M. Wang, "Reducing Mutual Coupling of Closely Spaced Microstrip MIMO Antennas for WLAN Application," *IEEE Antennas and Wireless Propagation Letters,* vol. 10, pp. 310-313, 2011.

[60] W. C. Zheng, L. Zhang, Q. X. Li, and Y. Leng, "Dual-Band Dual-Polarized Compact Bowtie Antenna Array for Anti-Interference MIMO WLAN," *IEEE Transactions on Antennas and Propagation,* vol. 62, pp. 237-246, 2014.

[61] Ruckus Wireless Inc. Available: https://www.ruckuswireless.com/rucktionary.

[62] D. J. Love, R. W. Heath, V. K. N. Lau, D. Gesbert, B. D. Rao, and M. Andrews, "An overview of limited feedback in wireless communication systems," *IEEE Journal on Selected Areas in Communications,* vol. 26, pp. 1341-1365, 2008.

[63] A. D. Dabbagh and D. J. Love, "Feedback rate-capacity loss tradeoff for limited feedback MIMO systems," *IEEE Transactions on Information Theory,* vol. 52, pp. 2190-2202, 2006.

[64] H. Lou, M. Ghosh, P. Xia, and R. Olesen, "A comparison of implicit and explicit channel feedback methods for MU-MIMO WLAN systems," in *2013 IEEE 24th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, 2013, pp. 419-424.

[65] Y. Hatakawa, T. Matsumoto, and S. Konishi, "Development and experiment of linear and non-linear precoding on a real-time multiuser-mimo testbed with limited CSI feedback," in *2012 IEEE 23rd International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, 2012, pp. 1606-1611.

[66] X. Xie, X. Zhang, and K. Sundaresan, "Adaptive feedback compression for MIMO networks," in *Proceedings of the 19th annual international conference on Mobile computing & networking*, Miami, Florida, USA, 2013, pp. 477-488.

[67] R. Kudo, S. M. D. Armour, J. P. McGeehan, and M. Mizoguchi, "A channel state information feedback method for massive MIMO-OFDM," *Journal of Communications and Networks,* vol. 15, pp. 352-361, 2013.

[68] D. J. Love and R. W. Heath, "Limited feedback unitary precoding for spatial multiplexing systems," *IEEE Transactions on Information Theory,* vol. 51, pp. 2967-2976, 2005.

[69] M. Yazid, A. Ksentini, L. Bouallouche-Medjkoune, and D. Aissani, "Enhancement of the TXOP sharing designed for DL-MU-MIMO IEEE 802.11ac WLANs," in *2015 IEEE Wireless Communications and Networking Conference (WCNC)*, 2015, pp. 908-913.

[70] C. Zhu, A. Bhatt, Y. Kim, O. Aboul-magd, and N. Chiu, "MAC enhancements for downlink multi-user MIMO transmission in next generation WLAN," in *2012 IEEE Consumer Communications and Networking Conference (CCNC)*, 2012, pp. 832-837.

[71] E. Charfi, L. Chaari, and L. Kamoun, "Upcoming WLANs MAC access mechanisms: An overview," in *8th International Symposium on Communication Systems, Networks & Digital Signal Processing (CSNDSP)*, 2012, pp. 1-6.

[72] M. X. Gong, E. Perahia, R. Want, and M. Shiwen, "Training protocols for multi-user MIMO wireless LANs," in *IEEE 21st International Symposium on Personal Indoor and Mobile Radio Communications*, 2010, pp. 1218-1223.

[73] M. X. Gong, E. Perahia, R. Stacey, R. Want, and S. Mao, "A CSMA/CA MAC protocol for multi-user MIMO wireless LANs," in *IEEE Global Telecommunications Conference*, 2010, pp. 1-6.

[74] A. Thapa and S. Shin, "A MAC protocol to select optimal transmission mode in very high throughput WLAN: MU-MIMO vs. multiple SU-MIMO," in *Third Asian Himalayas International Conference on Internet*, 2012, pp. 1-5.

[75] A. Ettefagh, M. Kuhn, C. Eşli, and A. Wittneben, "Performance analysis of distributed cluster-based MAC protocol for multiuser MIMO wireless networks," *EURASIP Journal on Wireless Communications and Networking,* vol. 2011, pp. 1-14, 2011.

[76] M. Aajami and J. B. Suk, "Optimal TXOP Sharing in IEEE 802.11ac," *IEEE Communications Letters,* vol. 19, pp. 1141-1144, 2015.

[77] M. Esslaoui, F. Riera-Palou, and G. Femenias, "A fair MU-MIMO scheme for IEEE 802.11ac," in *2012 International Symposium on Wireless Communication Systems (ISWCS)*, 2012, pp. 1049-1053.

[78] V. Shrivastava, N. Ahmed, S. Rayanchu, S. Banerjee, S. Keshav, K. Papagiannaki*, et al.*, "CENTAUR: realizing the full potential of centralized wlans through a hybrid data path," in *Proceedings of the 15th annual international conference on Mobile computing and networking*, Beijing, China, 2009.

[79] A. Mishra, V. Brik, S. Banerjee, A. Srinivasan, and W. Arbaugh, "A Client-Driven Approach for Channel Management in Wireless LANs," in *25th IEEE International Conference on Computer Communications*, 2006, pp. 1-12.

[80] A. P. Jardosh, K. Papagiannaki, E. M. Belding, K. C. Almeroth, G. Iannaccone, and B. Vinnakota, "Green WLANs: On-Demand WLAN Infrastructures," *Mobile Networks and Applications,* vol. 14, pp. 798-814, 2008.

[81] A. P. Jardosh, G. Iannaccone, K. Papagiannaki, and B. Vinnakota, "Towards an Energy-Star WLAN Infrastructure," in *Eighth IEEE Workshop on Mobile Computing Systems and Applications*, 2007, pp. 85-90.

[82] D. Tse, P. Viswanath, and L. Zheng, "Diversity-multiplexing tradeoff in multiple-access channels," *IEEE Transactions on Information Theory,* vol. 50, pp. 1859-1874, 2004.

[83] W. Yu and J. M. Cioffi, "Sum capacity of Gaussian vector broadcast channels," *IEEE Transactions on Information Theory,* vol. 50, pp. 1875-1892, 2004.

[84] Z. Shen, R. Chen, J. G. Andrews, R. W. Heath, and B. L. Evans, "Sum Capacity of Multiuser MIMO Broadcast Channels with Block Diagonalization," *IEEE Transactions on Wireless Communications,* vol. 6, pp. 2040-2045, 2007.

[85] P. Viswanath and D. Tse, "Sum capacity of the vector Gaussian broadcast channel and uplink-downlink duality," *IEEE Transactions on Information Theory,* vol. 49, pp. 1912-1921, 2006.

[86] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels," *IEEE Transactions on Information Theory,* vol. 49, pp. 2658-2668, 2003.

[87] Z. Lizhong and D. Tse, "Diversity and multiplexing: a fundamental tradeoff in multiple-antenna channels," *IEEE Transactions on Information Theory,* vol. 49, pp. 1073-1096, 2003.

[88] 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, pp. 3857-3868, 2005.

[89] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE Journal on Selected Areas in Communications,* vol. 24, pp. 528-541, 2006.

[90] O. Aboul-Magd, U. Kwon, Y. Kim, and C. Zhu, "Managing downlink multi-user MIMO transmission using group membership," in *IEEE Consumer Communications and Networking Conference* 2013, pp. 370-375.

[91] S. Wu, W. Mao, and X. Wang, "Performance study on a CSMA/CA-based MAC protocol for multi-user MIMO wireless LANs," *IEEE Transactions on Wireless Communications,* vol. 13, pp. 3153-3166, 2014.

[92] R. Liao, B. Bellalta, C. Cano, and M. Oliver, "DCF/DSDMA: Enhanced DCF with SDMA downlink transmissions for WLANs," in *Baltic Congress on Future Internet Communications*, 2011, pp. 96-102.

[93] V. Valls and D. J. Leith, "Proportional fair MU-MIMO in 802.11 WLANs," *IEEE Wireless Communications Letters,* vol. 3, pp. 221-224, 2014.

[94] A. Adhikary, H. C. Papadopoulos, S. A. Ramprashad, and G. Caire, "Multi-user MIMO with outdated CSI: training, feedback and scheduling," in *49th Annual Allerton Conference on Communication, Control, and Computing*, 2011, pp. 886-893.

[95] N. Jindal and S. Ramprashad, "Optimizing CSI feedback for MU-MIMO: tradeoffs in channel correlation, user diversity and MU-MIMO efficiency," in *IEEE 73rd Vehicular Technology Conference*, 2011, pp. 1-5.

[96] H. V. Balan, R. Rogalin, A. Michaloliakos, K. Psounis, and G. Caire, "Achieving high data rates in a distributed MIMO system," in *Proceedings of the 18th annual international conference on mobile computing and networking*, Istanbul, Turkey, 2012, pp. 41-52.

[97] M. Gast, *802.11 ac: A survival guide*: O'Reilly Media, Inc., 2013.

[98] B. Huang, X. Fang, Q. Chen, R. He, and W. Chen, "A Pairing Scheduling Scheme for Virtual MIMO Systems," in *6th International Conference on Wireless Communications Networking and Mobile Computing*, 2010, pp. 1-6.

[99] A. Hottinen and E. Viterbo, "Optimal user pairing in downlink MU-MIMO with transmit precoding," in *6th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks and Workshops*, 2008, pp. 97-99.

[100] C. H. Papadimitriou and K. Steiglitz, *Combinatorial optimization: algorithms and complexity*: Courier Corporation, 1998.

[101] F. Kaltenberger, D. Gesbert, R. Knopp, and M. Kountouris, "Correlation and capacity of measured multi-user MIMO channels," in *IEEE 19th International Symposium on Personal, Indoor and Mobile Radio Communications*, 2008, pp. 1-5.

[102] M. X. Gong, B. Hart, and S. Mao, "Advanced Wireless LAN Technologies: IEEE 802.11AC and Beyond," *GetMobile: mobile computing and communications,* vol. 18, pp. 48-52, 2015.

[103] D.-J. Deng, K.-C. Chen, and R.-S. Cheng, "IEEE 802.11ax: Next generation wireless local area networks," in *10th International Conference on Heterogeneous Networking for Quality, Reliability, Security and Robustness*, 2014, pp. 77-82.

[104] J. Jiho and L. Kwang Bok, "Transmit power adaptation for multiuser OFDM systems," *IEEE Journal on Selected Areas in Communications,* vol. 21, pp. 171-178, 2003.

[105] W. Jiao, L. Cai, and M. Tao, "Competitive scheduling for OFDMA systems with guaranteed transmission rate," *Computer Communications,* vol. 32, pp. 501-510, 2009.

[106] T. Meixia, L. Ying-Chang, and Z. Fan, "Resource Allocation for Delay Differentiated Traffic in Multiuser OFDM Systems," *IEEE Transactions on Wireless Communications,* vol. 7, pp. 2190-2201, 2008.

[107] S. Kibeom, M. Mohseni, and J. M. Cioffi, "Optimal resource allocation for OFDMA downlink systems," in *IEEE International Symposium on Information Theory*, 2006, pp. 1394-1398.

[108] Z. Huiling and W. Jiangzhou, "Chunk-based resource allocation in OFDMA systems - part I: chunk allocation," *IEEE Transactions on Communications,* vol. 57, pp. 2734-2744, 2009.

[109] Z. Huiling and W. Jiangzhou, "Chunk-based resource allocation in OFDMA systems - Part II: joint chunk, power and bit allocation," *IEEE Transactions on Communications,* vol. 60, pp. 499-509, 2012.

[110] S. D. Baker and D. H. Hoglund, "Medical-Grade, Mission-Critical Wireless Networks," *IEEE Engineering in Medicine and Biology Magazine,* vol. 27, pp. 86-95, 2008.

[111] P. Phunchongharn, E. Hossain, D. Niyato, and S. Camorlinga, "A cognitive radio system for e-health applications in a hospital environment," *Wireless Commun.,* vol. 17, pp. 20-28, 2010.

[112] A. Soomro and D. Cavalcanti, "Opportunities and challenges in using WPAN and WLAN technologies in medical environments," *IEEE Communications Magazine,* vol. 45, pp. 114-122, 2007.

[113] Y. Zhang, N. Ansari, and H. Tsunoda, "Wireless telemedicine services over integrated IEEE 802.11/WLAN and IEEE 802.16/WiMAX networks," *IEEE Wireless Communications,* vol. 17, pp. 30-36, 2010.

[114] J. Lee, B. Bagheri, and H.-A. Kao, "A Cyber-Physical Systems architecture for Industry 4.0-based manufacturing systems," *Manufacturing Letters,* vol. 3, pp. 18-23, 2015.

[115] T. K. Refaat, R. M. Daoud, H. H. Amer, M. Hassan, and O. M. Sultan, "Workcell concatenation using wifi-based Wireless Networked Control Systems," in *17th IEEE International Conference on Electronics, Circuits, and Systems*, 2010, pp. 138-141.

[116] D. M. Anand, J. R. Moyne, and D. M. Tilbury, "Performance evaluation of wireless networks for factory automation applications," in *IEEE International Conference on Automation Science and Engineering*, 2009, pp. 340-346.

[117] D. Zuehlke, "Smartfactory–from vision to reality in factory technologies," in *Proceedings of the 17th International Federation of Automatic Control (IFAC) World Congress*, 2008, pp. 82-89.

[118] A. Alshamrani, X. Shen, and L.-L. Xie, "QoS Provisioning for Heterogeneous Services in Cooperative Cognitive Radio Networks," *IEEE Journal on Selected Areas in Communications,* vol. 29, pp. 819-830, 2011.

[119] I. G. Fraimis and S. A. Kotsopoulos, "QoS-Based Proportional Fair Allocation Algorithm for OFDMA Wireless Cellular Systems," *IEEE Communications Letters,* vol. 15, pp. 1091-1093, 2011.

[120] E. Kartsakli, N. Zorba, L. Alonso, and C. Verikoukis, "Multiuser MAC Protocols for 802.11n Wireless Networks," in *IEEE International Conference on Communications*, 2009, pp. 1-5.

[121] Z. Zhang, S. Bronson, J. Xie, and H. Wei, "Employing the One-Sender-Multiple-Receiver Technique in Wireless LANs," in *IEEE INFOCOM*, 2010, pp. 1-9.

[122] M. Heusse, F. Rousseau, G. Berger-Sabbatel, and A. Duda, "Performance anomaly of 802.11b," in *IEEE INFOCOM*, 2003, pp. 836-843.

[123] G. Tan and J. Guttag, "Time-based fairness improves performance in multi-rate WLANs," presented at *Proceedings of the annual conference on USENIX Annual Technical Conference*, Boston, MA, 2004.

[124] L. B. Jiang and S. C. Liew, "Proportional fairness in wireless LANs and ad hoc networks," in *IEEE Wireless Communications and Networking Conference*, Vol. 3, pp. 1551-1556, 2005.

[125] R. V. Rasmussen and M. A. Trick, "Round robin scheduling – a survey," *European Journal of Operational Research,* vol. 188, pp. 617-636, 2008.

[126] C. Berge and E. Minieka, *Graphs and hypergraphs, vol. 7*: North-Holland publishing company Amsterdam, 1973.

[127] Y. H. Chan and L. C. Lau, "On linear and semidefinite programming relaxations for hypergraph matching," *Mathematical programming,* vol. 135, pp. 123-148, 2012.

[128] E. Angel, "A survey of approximation results for local search algorithms," in *Efficient Approximation and Online Algorithms*, Springer, 2006, pp. 30-73.

[129] P. Berman, "A d/2 approximation for maximum weight independent set in d-claw free graphs," in *Algorithm Theory-SWAT 2000*, ed: Springer, 2000, pp. 214-219.

[130] H. W. Kuhn, "The Hungarian method for the assignment problem," *Naval Research Logistics Quarterly,* vol. 2, pp. 83-97, 1955.

[131] V. Kolmogorov, "Blossom V: a new implementation of a minimum cost perfect matching algorithm," *Mathematical Programming Computation,* vol. 1, pp. 43-67, 2009.

[132] C. N. K. Osiakwan and S. G. Akl, "A perfect speedup parallel algorithm for the assignment problem on complete weighted bipartite graphs," in *International Conference on Databases, Parallel Architectures and Their Applications*, 1990, pp. 293-301.

[133] S.-S. Ahn, S. Park, M. Chertkov, and J. Shin, "Minimum Weight Perfect Matching via Blossom Belief Propagation," in *Advances in Neural Information Processing Systems*, 2015, pp. 1288-1296.

[134] N. Bao, J. Li, W. Xia, and L. Shen, "QoS-aware resource allocation algorithm for OFDMA-WLAN integrated system," in *IEEE Wireless Communications and Networking Conference*, 2013, pp. 807-812.

[135] S. Valentin, T. Freitag, and H. Karl, "Integrating multiuser dynamic OFDMA into IEEE 802.11 WLANs - LLC/MAC extensions and system performance," in *IEEE International Conference on Communications*, 2008, pp. 3328-3334.

[136] H. M. Alnuweiri, Y. P. Fallah, P. Nasiopoulos, and S. Khan, "OFDMA-based medium access control for next-generation WLANs," *EURASIP Journal on Wireless Communication Networking,* vol. 2009, pp. 1-9, 2009.

[137] G. L. Nemhauser and L. A. Wolsey, *Integer and combinatorial optimization*: Wiley New York, 1988.

[138] *IBM ILOG CPLEX Optimizer*. Available: www.ibm.com/software/integration/optimization/cplex-optimizer

[139] *Wireless Considerations in Healthcare Environments*. Available: http://www.cisco.com/c/dam/en_us/solutions/industries/docs/healthcare/wireless_hc_environ061208.pdf

[140] Xiaofu Ma, Qinghai Gao, Vuk Marojevic, and Jeffrey H. Reed, "Hypergraph Matching for MU-MIMO User Grouping in Wireless LANs", *Elsevier Journal of Ad Hoc Networks*, Vol. 48, pp. 29-37, 2016.

[141] Munawwar M. Sohul, Miao Yao, Xiaofu Ma, Eyosias Y. Imana, Vuk Marojevic, Jeffrey H. Reed, "Next Generation Public Safety Networks: A Spectrum Sharing Approach", *IEEE Communication Magazine*, vol. 54, no. 3, pp. 30-36, 2016.

[142] Stephen M. Dudley, William C. Headley, Marc Lichtman, Eyosias Imana, Xiaofu Ma, Mahi Abdelbar, Aditya Padaki, Abid Ullah, Munawwar Sohul,Tae-young Yang, and Jeffrey H. Reed, "Practical Issues for Spectrum Management with Cognitive Radios", *IEEE Proceedings*, vol. 102, no. 3, pp. 242-264, 2014.

[143] Xiaofu Ma, Sayantan Guha, Junsung Choi, Christopher R.Anderson, Randall Nealy, Jared Withers, Jeffrey H.Reed, Carl Dietrich, "Analysis of Directional Antenna for Railroad Crossing Safety Applications", in *IEEE CCNC*, 2017.

[144] Xiaofu Ma, Thaddeus Czauski, Taeyoung Yang, and Jeffrey H. Reed, "Wirelessly Sensing Medication Administration", in *ACM Wireless Health*, pp. 1-8, 2014.

[145] Xiaofu Ma, Yujun Liu, Dr. Karen A. Roberto, Dr. Jeffrey H. Reed, "Using Wireless Sensing Technology for Medication Management by Older Adults", Abstract accepted by *Gerontological Society of America's 68th Annual Scientific Meeting*.

[146] Munawwar M. Sohul, Xiaofu Ma, Taeyoung Yang, Jeffrey H. Reed, "Quality of Service Assurance for Shared Spectrum Systems", in *IEEE MILCOM*, pp. 1471-1476, 2014.

[147] Xiaofu Ma, Haris I. Volos, Xiangwei Zheng, Jeffrey H. Reed, and Tamal Bose, "A Variation-Aware Approach for Task Allocation in Wireless Distributed Computing Systems", in *IEEE GLOBECOM*, pp. 5006-5011, 2013.