ADDRESSING THE HIDDEN TERMINAL PROBLEM IN MU-MIMO WLANS WITH RELAXED ZERO-FORCING APPROACH

by

Sanjeeb Shrestha

Dissertation submitted in fulfilment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

School of Computing and Communications
University of Technology Sydney,
Australia

3rd of July 2017

CERTIFICATE OF ORIGINAL AUTHORSHIP

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of the requirements for a degree to any other university or institution other than University of Technology Sydney, Australia.

I also certify that the thesis is an original piece of research and it has been written by me.

In addition, I certify that all information sources and literature used are indicated in the thesis.

Last but not the least, this research is supported by an Australian Government Research Training Program Scholarship.

. . . . . .

Sanjeeb Shrestha

ACKNOWLEDGMENTS

I would like to express my sincerest gratitude to my principal supervisor, Dr Gengfa Fang, who has constantly encouraged me to strive for excellence and given me freedom and trust to find my own way in this project.

I am profoundly grateful to my co-supervisor, Professor Xiaojing Huang, for his remarkable insights and invaluable guidance without which this project would not have been possible. I would also like to acknowledge my alternate-supervisor Professor Eryk Dutkiewicz for his guidance and support from time to time. A special thanks to Dr Keith Imrie for his thorough grammatical editing and help with my English expression for my papers/thesis. Also, I would like to extend my warm regards to all my colleagues in the Department for their help and companionship.

I am sincerely thankful to the Australian Government, for an Australian Postgraduate Award, to Macquarie University for a generous PGRF award and Intel top-up scholarship, and University of Technology Sydney for Faculty of Engineering and Information Technology Scholarship, which provided me a conducive environment for my research and helped me explore new ideas.

Last but not least, I am forever indebted to my parents, my wife Sulochana, my son Yash and my daughter Yashasvi for their endless love, support and encouragement during this time and throughout my life. I cannot thank you enough; without your presence I would not have travelled so far. I am always extremely grateful to the Almighty God and His countless blessings.

To my parents, my wife Sulochana, my son Yash, and my daughter Yashasvi

ABSTRACT

An ever-increasing data rate demand, mainly due to the proliferation of numerous smart devices, enterprises' mission critical networks, and industry automation, has mounted tremendous pressure on today's Wireless Local Area Networks (WLANs). Several avenues such as bandwidth, constellation density, the Multiple Input Multiple Output (MIMO) technique, etc., have been explored, e.g., IEEE802.11n/ac standards, to keep up with the demand. Future WLAN standard, e.g., IEEE802.11ax, with potential technologies such as uplink Multi-User (MU)-MIMO, full duplex transmission, etc., is anticipated by 2019.

Having said that, there has been a strong emphasis on solving the technical issues with WLANs along with the addition of new frontiers in order to cope with the data rate demanded. One such appending decade-long issue is the inevitable Hidden Terminal (HT) problem in a distributive, decentralised and densely deployed WLANs, which fundamentally arises because of the transmission time overlaps between different transmitters operating at a particular frequency. The consequence is that it causes collisions of signals, which sharply reduces the system throughput.

In the context of MU-MIMO based WLANs, several designs for a general network scenario, without the consideration of the HT problem, have been proposed, bringing efficiency by avoiding the collision of signals. However, a dedicated design, which could effectively address the HT problem in MU-MIMO WLANs

and also become interoperable (with legacy standards) and feasible with existing hardware, is lacking to the best of our knowledge.

In this thesis, we propose a solution for the HT problem which has three fundamental attributes.

First, a) at the Physical (PHY) layer, the Zero-forcing (ZF) transmission strategy with fairness and throughput aware precoding is proposed, b) a hybrid scheduling scheme, combining the packet position-based First In First Out (FIFO) and channel quality-based scheme, namely the Best of the Two Choices, is designed, c) at the Medium Access Control (MAC) layer, Degrees-of-Freedom (DoF) based Transmission Opportunity (TXOP) for Access Points (APs) is developed which is backed by an extended Point Coordination Function (PCF), d) an explicit channel acquisition framework is proposed for ZF which has a reduced signaling time overhead of 98.6740 μs compared to IEEE802.11ac. e) performance evaluation methodologies are: i) hardware testbed results of the PHY strategy, which shows a received SNR gain of about 6 dB on average, and about 10 dB in comparison to the HT scenario, ii) simulation results of the MAC design, which shows a constant throughput gain of $4-5$ times w.r.t. the popular Request to Send/Clear to Send (RTS/CTS) solution.

Second, to address the interoperability issue, we purposefully use the standard frame format except for some required logical changes. Notably, the transition mechanism of our design, and for any MAC that uses standard frame formats, is investigated meticulously. The transition condition, transition steps and transition frame formats are detailed.

Third, to address a practical constraint of an imperfect Channel State Information (CSI) at APs, a) we incorporate the Finite Rate Feedback (FRF) model in our solution. The effects on system parameters such as quantisation error bounds,

throughput loss w.r.t. perfect CSI, etc., are discussed with closed-form analytical expressions, b) instead of an ideal ZF technique, a Relaxed ZF (RZF) framework is considered, in which the interference and power constraints of the optimisation problem are relaxed to the interference upper bound and to the maximum transmit power respectively. Our results lead to a distributive algorithm for calculating the optimal ZF precoding vector which suits the distributive, decentralised and uncoordinated nature of MU-MIMO WLANs.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

We live in a world with ever-increasing mobility. The pace of life in our cities is growing at an exponential rate[1]. To keep up with the momentum, access to information on the go has become an integral part of our lifestyle. Unlike wired Ethernet, the promise of wireless technology, to provide seamless connectivity and an increasing data rate, has made wireless communication systems a principal candidate for the task. Both consumers and enterprises have been relying heavily on wireless communication systems for connecting people with the information cloud, things (i.e., forming Internet of Things (IoT)) and one another regardless of the location.

We have seen overwhelming technical advances in the field of wireless communication systems in the last decade. Over the years, Wireless Local Area Network (WLAN) has evolved and emerged as one of the successful wireless communication technologies among consumers, enterprises and industry automation. Moreover, in recent years specifically,

[1]Professor Richard Wiseman, in collaboration with the British Council, discovered that the pace of life is now 10 % faster than in the early 1990s [1].

in order to meet the ever-increasing throughput demand, up to 1000 fold, WLANs based on Multi-User Multiple Input Multiple Output (MU-MIMO) have gathered tremendous attention and have been at the epicentre of research in academia and industry.

WLANs have gained much more attention than other wireless communication systems, principally due to their inherent advantages such as: a promise of gigabit throughput, inexpensive hardware, ease of deployment, decentralised control, flexibility, scalability, etc. Meanwhile, the IEEE 802.11 standardisation committee made a significant policy departure by adopting a Single User-Multiple Input Multiple Output (SU-MIMO) in the IEEE 802.11n standard in 2009. Later in 2013, the Multiple Input Multiple Output (MIMO) technique was continued in the following IEEE 802.11ac standard as MU-MIMO. Now, the much anticipated IEEE802.11ax standard (also refered to as High Efficiency WLANs (HEW))[2], with the provision of uplink MU-MIMO, is expected to be finalised by the end of 2019. The fundamental principle to adopting MIMO in these standards is to reap the advantages of the MIMO technique and provide a gigabit wireless throughput experience to users[2].

Having said that, since MIMO becomes an integral part of these standards, there is an intransigent requirement to incorporate the MIMO technique specifically at the physical (PHY) layer. The consequence is that the legacy Medium Access Control (MAC) protocol, which was basically designed focusing on the Single Input Single Output (SISO) technique, needs to be revisited and overhauled. Nevertheless, the new MAC has to be compatible with the legacy standards.

The other major technological challenge to adopting the MIMO technique in WLANs comes from within its characteristics: scalability and decentralised control. Due to the inherent advantages and scalability, WLANs are ubiquitous. However, given the limited

[2]Additionally, in the IEEE 802.11ac standard, channel bandwidth has been significantly increased up to 160 MHz and the modulation scheme to 256 Quadrature Amplitude Modulation (QAM).

bandwidth of the Industrial Scientific and Medical (ISM) bands, many adjacent networks have to operate at the same frequency. In such a context, the lack of centralised control due to the distributed nature of MU-MIMO WLANs would undoubtedly exacerbate the notorious Hidden Terminal (HT) problem. Specifically, the HT problem is defined as the scenario in WLANs where transmitters/Access Points (APs) do not sense each others' transmission but interfere with each other at the intended receiver, causing signal collision and decoding failure, leading to a huge throughput degradation. The HT problem has been a prominent research interest since the advent of WLANs in the late 1990s.

1.2 The Context

The consequence of the HT problem would be more catastrophic in MU-MIMO based WLANs because: a) an HT terminal reduces throughput by more than 40% (referenced to legacy standards) [3], and this would be a huge setback when additional resources (such as an additional pair of antennas) are already required for the MIMO technique to increase the throughput of WLANs; b) the MIMO technique itself is susceptible to interference, in addition to the HT problem, which would cause further interference. This may result in a very complex scenario for WLANs.

The IEEE standards supporting the MIMO technique, e.g., IEEE802.11n/ac have remained completely open on what MIMO technique would be used to manage interference in the network. Specifically downlink and uplink MIMO techniques are not defined. For instance, the basic guideline for the IEEE802.11ac at the downlink is to null-steer beams from APs to clients. However, depending on the design objectives, the transmission strategy can be varied. In such a context, there has been a tremendous attention in the research community for designing a MU-MIMO WLAN networks which can reap the advantage of the MIMO technique and provide the throughput demanded.

In recent years, several designs have been proposed for MU-MIMO WLANs considering different downlink, uplink, scheduling and channel access strategies. However the inevitable HT problem, which is highly likely to become prevalent in future MU-MIMO WLANs given WLANs' ubiquitous existence and decentralised control, has not be addressed sufficiently. The HT problem, in most designs, has been either overlooked, by the high-level statement that the design was also capable of solving the HT problem without explicit results, or, avoided as an optional issue to be dealt with in the main design.

In this thesis, the notorious HT problem in MU-MIMO WLANs is considered. A dedicated design, which takes the popular Zero-forcing (ZF) technique as a transmission strategy at the PHY layer and a protocol at MAC based on the PHY layer scheme, is presented. The interoperability issue of the design is checked and ensured to make the design interoperable with any design that uses standard frame formats. The performance evaluation tools are testbed and simulation results.

1.3 Motivation of the Research

The popularity of WLANs is growing and this trend will continue in the near future. Their inherent advantages have made them almost ubiquitous. However, due to data-hungry multi-media services such as video on demand (VoD), High Definition Television (HDTV), live streaming, etc., the WLAN technology has been under constant pressure to meet the ever-increasing data-rate demands. Several avenues have been explored to keep up with the traffic demand, e.g., bandwidth, constellation density, the MIMO technique, etc. Many IEEE WLAN standards are evolving, e.g., IEEE 802.11n/ac, and IEEE 802.11ax (to be finalised by 2019). However, there has been a strong emphasis in the research community on solving the technical issues along with the addition of new features. One such appending decade-long issue is the HT problem, which specifically arises from the

transmission time overlaps between different APs. The consequence is that it causes collisions of signals and reduces the network throughput by as much as 40%.

Conventionally, solutions of the HT problem are developed from the perspective of a single antenna (Single In Single Out) SISO system. Related studies on the HT problem on legacy standards have suggested that the HT problem cannot be addressed by a channel-sensing mechanism such as Carrier Sense Multiple Access/Collosion Avoidance (CSMA/CA) in WLANs. Popular solutions such as the Busy-tone scheme, Request to Send/Clear to Send (RTS/CTS) scheme and their various derivatives do exist. However, due to the advent of the MIMO technique in WLANs, there has been a paradigm shift from single-antenna systems to multiple-antenna systems. The consequence is that the legacy MAC designs and solutions need to be revisited and overhauled.

Over time, various MAC designs for MU-MIMO WLANs have been proposed considering different access mechanisms, downlink, uplink, channel acquisition and scheduling schemes. However, a dedicated design that: a) addresses the inevitable HT problem in MU-MIMO WLANs, and b) becomes compatible with any design that uses standard frame formats has not been considered in the prior proposed designs, to the best of our knowledge.

The need to address the HT problem in MU-MIMO WLANs is paramount. First, due to the ubiquitous presence of WLANs, many adjacent MU-MIMO WLANs will be operating at the same frequency. In such a context, the distributive, decentralised and densely deployed MU-MIMO WLANs will further aggravate the HT problem. Second, the throughput loss due to the HT problem in MU-MIMO WLANs would be severe when additional resources, e.g., multiple antennas, interference mitigation approaches, etc., have already been placed in MU-MIMO WLANs.

Thus we clearly see the need to address the HT problem in MU-MIMO WLANs.

1.4 Research Objectives

The prime objective of this research is to develop a solution to the notorious HT problem in MU-MIMO WLANs. The envisioned solution is expected to have three fundamental attributes: a) simplicity, b) effectiveness, and c) interoperability with legacy standards. The solution will need to take into account various practical constraints. Bearing in mind that the solution needs to be functional, the specific research objectives are as follows:

- To develop an effective PHY layer scheme to address the HT problem in MU-MIMO WLANs with a simple, pragmatic ZF transmission strategy. In the context of MU-MIMO WLANs, the ZF precoding need to be fairness and throughput aware.
- To design a suitable MAC protocol based on the PHY layer scheme. Specifically, concurrent transmissions (as a result of the PHY layer scheme) need to be appropriately accommodated. The downlink and uplink transmissions are to be detailed.
- To provide a transition condition, transition steps and transition frame formats to make the design at the PHY and MAC layers interoperable with any designs that use standard frame formats.
- Considering feedback constraints, to develop a design suitable for MU-MIMO WLANs. Specifically, for the Finite Rate Feedback (FRF) model, quantisation error bound, throughput loss w.r.t. perfect CSI, etc., are to be investigated in the context of MU-MIMO WLANs.

1.5 Scope of the Thesis

This thesis deals with a design to address the inevitable HT problem in MU-MIMO WLANs. Fundamentally, the design considers the sub-optimal yet simple and pragmatic

ZF technique as a downlink transmission strategy at the PHY layer. The explicit channel-acquisition method, by which APs compute CSI by estimating a training sequence sent by APs to clients and then clients estimate the CSI and send it back to APs, is adopted to obtain CSI at APs for the ZF technique. However, owing to the practical constraints, partial CSI with a finite number of quantisation bits is used. The primary objective in taking the practical constraints into account is to make the design robust and feasible in practice. A hybrid scheduling scheme, combining the packet-based FIFO and channel quality-based scheme, namely the Best of the Two Choices, is implemented to select clients in downlink and uplink transmissions.

The design at the MAC layer adopts and expands the Point Coordinated Function (PCF) as a medium-access mechanism, in which transmitters/APs play a pivotal role in managing downlink and uplink transmissions. The main justification for adopting PCF as a medium-access mechanism is that PCF is a part of the IEEE 802.11 standard, and can support concurrent transmissions as desired for MU-MIMO WLANs, with its Contention Free Period (CFP). Additionally, in most envisioned MU-MIMO WLAN applications, APs can have a calculated amount of sophistication and can play a coordinating role, whereas receivers are thought to have a simple architecture.

The interoperability issue is taken into account while designing the MAC protocol. Specifically, we characterise the design as a dedicated one for addressing the HT problem in MU-MIMO WLAN settings, which can work seamlessly for any other design that uses standard frame formats. In order to ensure the interoperability issue: a) we purposefully use the standard frame format except for some logical changes; b) the transition mechanism of our MAC for any standard Distributive Coordinated Function (DCF) is described meticulously. The transition condition, transition steps and transition frame formats are detailed.

The handshaking overheads and the performance gain of our MAC in addressing the

HT problem w.r.t. the popular RTS/CTS scheme is presented. However the theoretical analytic performance evaluation of our MAC design, like that for the DCF mechanism, is beyond the scope of this thesis. Our evaluation tools include results from testbed work and simulations.

1.6 Organisation of the Thesis

This thesis has nine chapters, divided into several sections and subsections for a detailed presentation of the specific topics. A brief outline of the contents is given as follows.

Chapter 2 is the background and related-work section, where the context of the research and the current state of the art on the solution of the HT problem in WLANs are comprehensively detailed. The pioneering solution of the HT problem, e.g., the Busy-tone solution, the RTS/CTS solution and their various derivatives, and the latest MU-MIMO based MAC designs, are studied and their pros and cons are detailed. A summary is presented in Section 2.1.

Chapter 3 studies the HT-based system model and investigates design opportunities. The fundamental assumption for our system model, i.e., that our system model is in the weak interference regime, is discussed. In terms of design opportunities, we first compare and contrast our system model with the conventional Zeroforming Beamforming (ZFB). Second, the application-specific requirements of the design in relation to our system model is discussed. In order to address the HT problem, the need to eliminate the collision of signals and to maintain the concurrent transmission among the desired clients with fairness is highlighted. Similarly, the requirement for the development of a medium-access mechanism which can work independently without centralised control is emphasised. The implication of the design is discussed and a summary is presented in Section 3.3.

Chapter 4 describes fairness and network throughput-aware precoding. This chapter

develops a solution at PHY layer based on the application-specific requirements presented in Chapter 3. We present channel acquisition via channel sounding and a Transmission Opportunity (TXOP) decision based on Degrees-of-Freedom (DoF) at APs. A concurrent transmission algorithm to find clients among desired clients and a detailed illustration for calculating the precoding vector are illustrated. We discuss the performance of the concurrent transmission algorithm and present a chapter summary in Section 4.3.

Chapter 5 is a description of the experimental prototype setup. The implementation of the Universal Software Radio Peripheral2 (USRP2) platform is described meticulously. Practical issues of the implementation process are discussed. We present the results based on the analysis of the raw received signal, the impact on the received Signal to Noise Ratio (SNR) and the analysis of Effective SNR (ESNR). A chapter summary is given in Section 5.7.

Chapter 6 describes the wireless medium-access design in which protocols at the MAC layer are designed based on PHY layer solutions. The sequence of steps used in the development of the MAC protocol is presented. We first start with the proposed procedure for acquiring CSI from clients and evaluate its performance in terms of available air-time utilisation and capacity gain. Second, we discuss the transmission opportunity, fairness mechanism and firmness index of the design for APs and heterogeneous clients. Third, a detailed medium-access (both downlink and uplink) procedure, incorporating a CFP of the extended PCF, is presented. Some points worthy of note are presented, followed by a chapter summary in Section 6.5.

Chapter 7 adds a practical dimension to the proposed design for solving the HT problem in MU-MIMO WLANs. The practical constraints for our system model, such as channel estimation error, limited bandwidth of the feedback path, delay in feedback path, etc, are discussed. The need for the addition of the FRF model in our design, due to the imperfect CSI availability at APs is illustrated. We then develop our system model with

the ZF technique incorporating the FRF in MU-MIMO WLANs. Numerical analysis for expected error-bound, rate with FRF and average rate reduction for ZF due to FRF are presented. A chapter summary is presented in Section 7.6.

Chapter 8 presents the relaxed ZF framework for FRF-based MU-MIMO WLANs. First, the determination of the number of feedback bits B for heterogeneous antenna clients in MU-MIMO WLANs is discussed. Second, the optimisation problem, in which the ZF technique is relaxed with the number of quantisation bits B, is formulated. Later, a distributive algorithm for an optimal ZF precoding vector based on the subgradient solution of the dual problem of the optimisation problem is developed. The specific attribute of the solution is that the optimal ZF precoding vector can be tuned with the number of quantisation bits B. Simulation results are given, followed by a chapter summary in Section 8.5.

Chapter 9 deals with conclusions and future work. The conclusion of the thesis is presented in this section. Some interesting and challenging future work suggested by the research is outlined in Section 9.1.

1.7 Thesis Contributions

The major contributions of this thesis are as follows.

- We propose a dedicated solution to address the HT problem in MU-MIMO WLANs which possesses reduced signalling time overheads, as low as 98.6 μs, compared to IEEE802.11ac, and can also achieve a constant network throughput gain of 4-5 times compared to the popular RTS/CTS solution. Additionally, a feasibility study of the proposed ZF transmission strategy at the PHY layer is conducted in a USRP2/GNURadio testbed prototype, which shows a notable received SNR gain of about 10 dB w.r.t. the HT scenario.

- Fairness and network throughput-aware ZF precoding: at the PHY layer, the conventional ZF transmission strategy is further developed to suit the fairness and network throughput requirements of MU-MIMO WLANs. A hybrid scheduling scheme combining a packet position-based FIFO (for the fairness issue) and the channel quality-based scheme, namely the Best of the Two Choices (for throughput), is proposed.

- Degrees-of-Freedom (DoF) based medium-access: a mechanism to access the wireless medium in the HT scenario is designed based on the available DoF at APs. Specifically, we propose that the TXOPs for APs in the HT scenario are to be decided based on the DoF available at APs in a MU-MIMO WLAN setting. A simple counter-based fairness algorithm for managing TXOPs for APs is designed. The Jain fairness index for the algorithm is around 90%. Additionally, the traditional PCF is further extended to suit the concurrent transmission both in the downlink and uplink.

- The solution is interoperable with legacy standards: a seamless transition between our solution and other legacy standards is ensured by the use of the standard frame formats except for some logical changes required in them. A threshold based transition condition is devised which is backed by series of transition steps. The necessary frame formats are meticulously detailed.

- The solution is robust and meets practical constraints: in order to address the practical constraint of imperfect CSI at APs, the FRF based system is considered. In the context of MU-MIMO WLANs, the effect of the quantisation of CSI on system parameters such as quantisation error bounds, throughput loss w.r.t. perfect CSI, choice of an appropriate number of feedback bits, etc., are discussed with closed-form analytical expressions. Furthermore, we consider a Relaxed ZF (RZF) framework

to calculate the optimal ZF precoding, in which the interference and power constraints of the optimisation problem are relaxed to the interference upper bound and to the maximum transmit power constraint respectively. Lastly, a distributed algorithm is presented for calculating the optimal ZF precoding vector which suits the distributive, decentralised and uncoordinated nature of MU-MIMO WLANs.

Chapter 2

Background and Related Work

2.1 Background

The exponential increase in the data rate demand, mainly due to the proliferation of numerous smart devices and data-hungry multi-media applications, has been pushing the capacity of the existing Wireless Communication technology into its limits. In order to cope with the demand, different frontiers such as: channel bandwidth, constellation density, MU-MIMO technology, carrier aggregation, advanced medium-access techniques, traffic offloading, etc., have been exploited and new wireless standards have been evolving.

Given the ever increasing popularity and the ubiquitous presence of WLANs, several IEEE 802.11 standards and amendments exist today, e.g., IEEE 802.11a/b/g/n/ac. In recent years, prominent new features from the PHY layer perspective, SU-MIMO and MU-MIMO technologies, have now become an integral part of the released standard, IEEE801.11n/ac, and will continue to be in the future standard, IEEE 802.11ax. The fundamental objective for adopting the MIMO technique in these IEEE standards is to reap the advantage of the well-known capacity promise suggested by the MIMO technique in the literature [4, 5, 6, 7, 8].

Having said that, the peak PHY layer data rate cannot be directly realised at the MAC layer in terms of network throughput, principally due to a) control overheads of management frames and b) poor spectrum utilisation, and frame loss and retransmission because of the collision of signals.

Control overheads arise from the necessary management frames at the MAC layer and can be lowered with an effective design. However, poor spectrum utilisation and frame collision are MAC design constraints which are intertwined. A large volume of literature suggesting various collision avoidance mechanism does exist, however, the root cause for frame collisions and retransmissions, the transmission time overlap, given the distributive, decentralised and uncoordinated nature of WLANs, has not been satisfactorily addressed. Related work suggests a holistic approach involving both the PHY and the MAC layers which would be beneficial.

In this thesis, we deal with the notorious frame collision problem in WLANs that arises from the transmission time overlap among different transmitters, i.e., the HT problem.

Specifically, the HT problem is defined as the scenario in WLANs where transmitters do not sense each others transmission but interfere with each-other at the intended receiver, causing frame collision and decoding failure, leading to a huge throughput degradation. The HT problem has been a prominent research interest since the advent of WLANs in the late 1990s.

For instance, consider a HT scenario as shown in Fig. 2.1, where for simplicity we have taken three networks i, j and k respectively out of K networks. The transmission range, APs, clients and desired transmission for the networks are shown in colours: green, blue and violet respectively. All the ith, jth and kth network APs, AP1, AP2 and AP3, are out of carrier sensing range of each other, so they cannot hear each other's transmissions. However, they can transmit to their clients at the same time, causing collision of signals and decoding failure. The red dotted arrows show the interference from each of the three

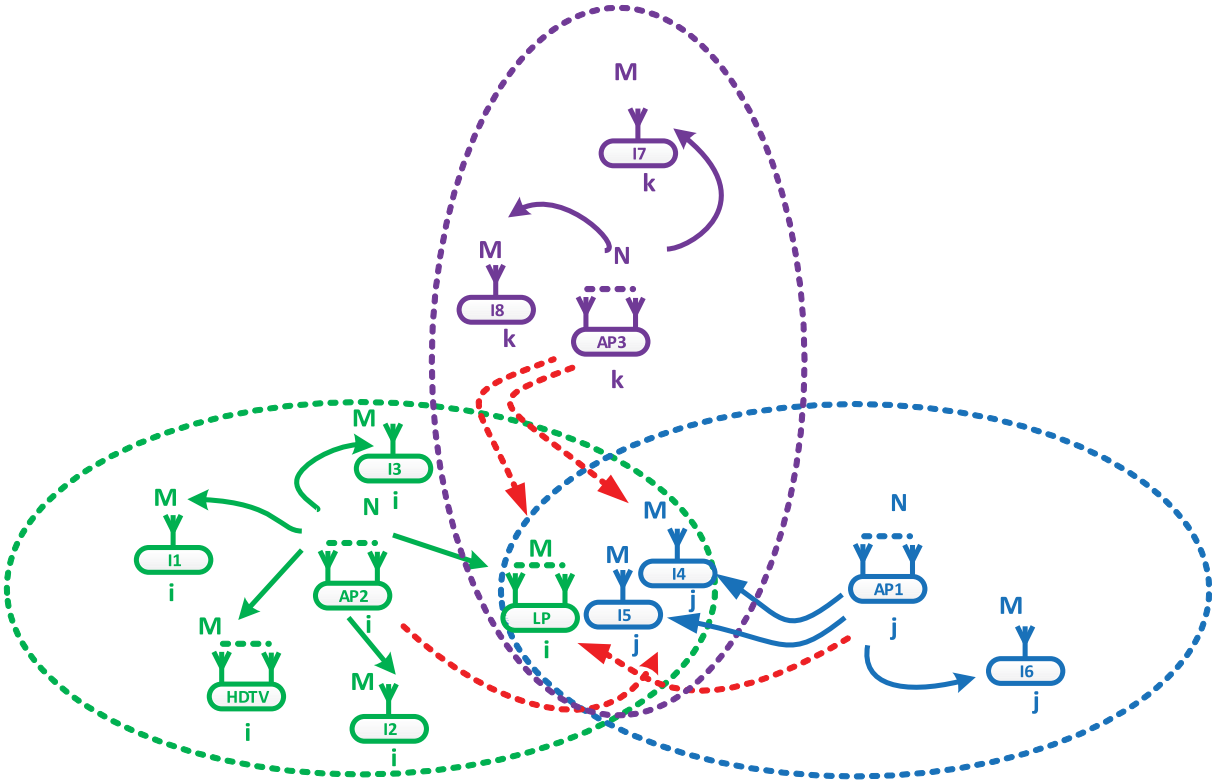

Figure 2.1: The Hidden-Terminal problem with a specific three-network scenario, i.e., K=3.

APs, considering downlink transmission, which are responsible for frame collision.

The ever increasing throughput demand on the one hand and the decades-long HT problem on the other hand may cast our higher bit-rate aspiration into oblivion. Thus the effect of the HT problem cannot be overlooked, especially at the present time, given the ubiquitous presence of WLANs[1] which extends from personal multimedia services (e.g., heterogeneous hardware such as iPhones, Laptops, HDTV, Tablets, etc.) to the growing dependence of enterprises on WLANs for mission-critical networks and for industry automation.

Moreover, the HT problem is further aggravated by WLANs' dynamic topologies (e.g.

[1]The penetration of WLANs is recorded as almost more than 70% in household and business enterprises and about 30% in public sectors such as transport, leisure, events, streets, etc. [9].

plug and play, bring your own device trend, BYOD), complex and irregular traffic patterns (e.g. voice, video and data traffics) and operation in the unlicensed band with inexpensive hardware without a centralised control (one of the beauties of WLANs). Measurements from production LANs [10] show that collisions of signals in WLANs are bound to happen, where transmission loss due to interference among 50% of sender-receiver pairs has a 2.5% probability. Similarly, the study in [3] reveals that HTs lead to about 40-42 % of collision loss, which results in a detrimental effect on system throughput. Thus, in the context of WLANs, the impact of the HT problem on network throughput cannot be overlooked.

WLANs have undergone tremendous changes since the first IEEE802.11 standard in 1997; the context of study and approaches to solution regarding the HT have varied. One of the prominent changes affecting the approach to the HT problem has been with the introduction of the MIMO technique in the IEEE802.11n standard in 2009. With the adoption of multiple antennas both at APs and clients, a whole new spatial dimension is in effect, in contrast to the single-antenna based SISO WLANs; this has changed our approach to deal with the problem arising from the transmission time overlap problem (i.e., HT problem). Multiple simultaneous transmission exploiting the spatial domain have now become a fundamental approach to the design constraints (of the PHY and the MAC layers) since the adoption of the MIMO technique in the WLAN standard.

2.2 Related Work

We shall first start with recounting the pioneering work regarding the HT scenario and then present the latest trends. The IEEE 802.11 standard proposed the CSMA/CA with RTS/CTS mechanism as a part of the DCF for medium access for WLANs in 1997. Since then IEEE802.11 CSMA/CA with RTS/CTS has been a de-facto mechanism to avoid collisions. However, there exist inherent limitations as to how it treats interference at

the receiver, related to the carrier sensing at the transmitter. The fact is that successful transmissions mostly depend on an interference-free condition at the receiver. Theoretical and experimental work on CSMA/CA [11, 12] showed that the CSMA/CA mechanism degrades performance due to poor spatial reuse. Besides, it also fails to address the HT problem [13] and capture-effect [14] issues.

In early years, a receiver-initiated busy-tone scheme was proposed to solve the HT problem for Packet Radio Networks (PRN), and was found to be effective in eliminating collisions caused by HTs [13]. Similarly, the Busy-tone Multiple Access (BTMA) scheme [15] was proposed, assuming that all stations are within the transmission range of APs and a busy tone using a separate channel can be sent to clients informing them of the status of the data channel, for solving the HT problem in WLANs. However, these schemes required a dedicated channel for the busy tone, which is not desirable in wireless networks.

Karn proposed the RTS/CTS mechanism as a part of Multiple Access with Collision Avoidance (MACA) [16] to address the HT problem. Later an amendment to MACA, Multiple Access with Collision Avoidance for Wireless (MACAW) [17], was proposed for single-channel WLANs at the Palo Alto Research Center, with a provision for an extra data-sending (DS) control frame in between CTS and the data frame for avoiding the collision of ACKs and other frames, a RTS or a data frame from other stations. Although the transmission time was cautiously checked, both methods did not sense a channel before transmission, running the risk of collision of signals anyway. Floor Acquisition Multiple Access (FAMA) [18] and Multiple Access via Collision Avoidance with Enhanced Parallelism (MACA-P) [19] were also proposed as a solution for HTs. However, experimental results showed that RTS/CTS significantly reduces the overall throughput [20].

A recent study proposed a lightweight wireless handshake [21] where the header of the

payload and Acknowledgement (ACK) are separated and designed to act like RTS/CTS. Nonetheless, packet decoding in dynamic channels is a fundamental limitation for that approach. In addressing the HT in WLANs, adopting the solution used in Code Division Multiple Access (CDMA) is not viable as it requires tight power control and special codes [22] and at high SNRs the performance is degraded. An alternative technique like Zigzag decoding [23] analysed collisions of packets with strategically selected collision patterns, which showed a significant packet loss reduction from 72.06 % to about 0.7% in a testbed of 14 USPR nodes. Moreover, it needs to have a collision-free chunk to bootstrap decoding in an irregular traffic pattern such as in WLAN scenarios.

As the capacity and spectral efficiency for a multi-antenna system were known [4, 5, 6, 7, 8], the IEEE802.11 standard started adopting the MIMO technique as an integral part of the standard. Since then, the MAC protocol for WLANs, ad hoc and mobile networks, which were originally designed from the perspective of a single omni-directional antenna, has been undergoing tremendous changes. The development of a new MAC is basically focused on spatial reuse accredited by the use of multiple antennas at both transmitters and receivers.

We shall now briefly discuss the evolution of MAC protocols since the advent of the the MIMO technique in WLANs. The aim is to present the context for the development of our MAC scheme in this thesis to address the HT problem in WLAN.

Firstly, we differentiate MAC schemes based on directional antenna usage and beamforming. Secondly, we focus on the downlink, uplink, channel acquisition, scheduling, etc., schemes based on which MAC schemes for WLANs are designed so far.

Based on the MIMO technique, there are two approaches: directional antenna [24, 25, 26] and beamforming [27, 28, 29]. Since directional antennas are only suitable for environments with sparse scattering and low angular spread, the beamforming technique has been the obvious choice at the PHY layer [30]. The reason is that the beamforming

technique is applicable for both sparse and rich scattering environments. In [29], a beamforming technique for null steering similar to [31] was used. However, it did not limit the channel bandwidth by splitting the channel into control and data sub-channels as in [31]. The flip side of the scheme was the inefficient use of the spectrum because, after every successful communication session, nodes were supposed to keep silent for a certain period of time specified in the field-silent period in RTS and CTS so that they would not disturb neighbors' communications. Likewise, a beamforming technique enabling spatial reuse was presented in reference to IEEE802.11s in [32]. However, the scheme relies on optional Mesh Coordinated Channel Access (MCCA) for channel access, which is challenging from synchronisation aspects in such a decentralised mesh network.

Specifically, depending on the downlink, uplink, channel acquisition and scheduling strategies, various MAC protocols for MIMO-based WLANs have been designed. The downlink MIMO channel is known as the MIMO broadcast channel where APs transmit to their clients. For this process to succeed, there are two important issues to be addressed. First, APs select a group of clients under a certain criterion which can be referred to as a scheduling problem. Second, APs precode the transmission to remove interference to undesired clients during transmission, which is a multi-user interference cancellation problem.

Similarly, the uplink channel of MIMO is known as a MIMO multiple-access channel, where multiple clients transmit to APs. As multiple clients transmit simultaneously to APs, APs need to decode the transmission from clients which is the multiuser detection problem. Depending on the ways to address downlink and uplink problems for MU-MIMO based WLANs, MAC designs vary from each other.

Related work shows that the ZF technique has been the most used downlink transmission strategy based on which MAC protocols are designed [33, 34, 35, 36, 37]. The channel feedback mechanism for these designs use both implicit feedback, meaning APs

compute CSI by estimating a training sequence sent by clients, in the uplink channel, and explicit feedback, meaning APs retrieve CSI by estimating a training sequence sent by APs to clients and then clients estimate the CSI and send it back to APs [38, 39]. For instance, [33] uses explicit feedback whereas [37] uses implicit feedback. The scheduling schemes followed are greedy [33], highest SNR [40] and priority queue [41]. Most of the performance evaluation methods were simulation based, while testbed analysis was used in [33].

Other transmission strategies such as Minimum Mean Square Error (MMSE) are used in [42] whereas beamforming and leakage coding are used in [40] and [41] respectively for the MAC protocol design based on a downlink transmission strategy..

Likewise, for the MAC protocol based on MIMO uplink transmission, various transmission strategies such as CDMA [43], MMSE [44, 45], blind detection [46] are used. Almost all the MAC protocols use implicit feedback. The scheduling schemes followed are best CSI [44], subcarrier [47], etc. Similar to the MAC based on a downlink, performance evaluation for most of the MACs based on uplink also used a simulation tool, and testbed evaluation is used in [47].

Some MAC proposals are integrated, meaning that they consider both the uplink and downlink in a single MAC protocol [48, 49].

In [48], group RTS is used in downlink and group CTS is used in uplink. A common control channel is used to exchange information on packet status. In [49], multipacket transmission in the downlink and multiple control-packet reception in the uplink are used.

The studies in [50, 51] use PCF as a medium access-mechanism where both the downlink and uplink transmissions are initiated and coordinated by APs. In [50], frequency signatures are used to differentiate simultaneously received control frames and greedy scheduling is adopted with fairness consideration. It is shown that the MAC design outperforms IEEE 802.11n SU-MIMO and MU-MIMO [41]. Similarly [51] uses the synergy

between interference alignment and interference cancellation known as IAC and presents the idea to decode more packets at APs than the number of antennas at APs. The MAC design emphasises the role of a leader AP which manages the transmissions both in downlink and uplink. A concurrency algorithm is designed to select clients in downlink and uplink. Both experimental and simulation results are presented.

The approach where a leader AP coordinates and schedules transmissions is further supported by a survey paper on MU-MIMO MAC [52, p.174] which anticipates that the forthcoming IEEE 802.11ax standard (to be finalised by 2019), would use APs to control and coordinate uplink MU-MIMO in the standard. The central role of APs for channel acquisition in IEEE 802.11ac is taken into account to make such predictions.

Related studies explore different downlink, uplink, channel acquisition and scheduling strategies on MAC design for MIMO based WLANs. Several interesting features are discussed. However, we believe that the inevitable HT problem on these MAC designs has not been sufficiently addressed. The HT problem is either overlooked mostly by the high-level statement that the proposed MAC design was also capable of addressing the specific HT issue without explicit evaluation results,[47, p.1418], or avoided by taking HT as an optional issue to be dealt with in the main design [53, p.153].

The prime objective of the design in this thesis is to address the notorious HT problem in a MU-MIMO based WLAN in an effective way [54, 55, 56, 57, 58, 59]. The proposed design to address the HT problem has following principal attributes: a) The medium-access mechanism is PCF, b) downlink and uplink transmission are controlled and coordinated by APs, c) downlink transmission strategy is the ZF technique d) channel acquisition is done explicitly e) a hybrid scheduling scheme with packet position-based FIFO and channel quality-based scheme, namely the Best of the Two Choices is adopted, f) performance evaluation tools are testbed and simulation results.

Furthermore, we add the another important dimension in our design from a practical

viewpoint. Since we use the ZF technique as a transmission strategy at the PHY layer for our MAC protocol design to address the HT problem in MU-MIMO-based WALN, there are few technical PHY layer challenges to be considered. First, the knowledge of CSI at APs/transmitters is the prerequisite for the ZF technique. However, owing to constraints such as: channel estimation errors at the receivers/clients, limited bandwidth of the feedback path, delay in feedback path, cost of feedback overheads, etc., perfect CSI cannot be obtained at APs. Thus the way out may be to provide a partial CSI in a quantised form to the APs and thoroughly analyse the impact of CSI quantisation in the system. Second, unlike cellular systems, MU-MIMO WLANs possess heterogeneous-antenna clients. Thus the choice of the number of quantisation bits, B, for all the clients is critical from the quantisation error perspective. Third, due to the quantisation error of channels, there will always be residual interference, in addition to thermal noise, at undesired clients, which cannot be completely removed[2]. This incurs throughput loss and affects the calculation of the optimal precoding vector.

Having said that, in order to make our design feasible for practical consideration, we take into account the practical constraints and consider a Limited Feedback/FRF-based MU-MIMO WLANs. Particularly at the PHY layer, we focus on: a) the upper bound of the quantisation error and the throughput loss due to quantisation bits, B, b) the calculation of the optimal ZF precoding vector considering the residual interference at undesired clients (which cannot be completely removed) due to the channel quantisation. For this, we first analyse the upper bound of the quantisation error incurred by the quantised CSI and calculate the average throughput loss. For the number of quantisation bits B, an analytical expression relating B, M and θ is derived, where θ represents the

[2]This is because the beamformer is based on the quantised channels from both the desired and undesired clients; however, the channels in the downlink or uplink are real (not quantised). This will be discussed further.

slope of the $\Delta R(P)$ vs B curve as shown in Fig. 8.1.

Second, we quantify the upper bound of the residual interference and relax the interference constraint with the upper bound in order to calculate the optimal ZF precoding vector. The interesting result is that the optimal precoding vector can be tuned with B, used for channel quantisation.

Third, owing to the decentralized control and distributive nature of MU-MIMO WLANs, we present a distributive algorithm for calculating the optimal ZF precoding vector for MU-MIMO WLANs. Specifically, our distributive algorithm takes into account the residual interference and tunes the optimal precoding vector according to B.

The fundamental concept of FRF is that it uses a finite B bits to feedback the instantaneous CSI from the receiver to the transmitter. Basically, each client quantises its channel to B bits and feedbacks the quantisation index instantaneously to the transmitters. The transmitter then uses these quantised CSI for beamforming. FRF was studied in [60, 61, 62] and the effects of CSI quantisation have been primarily studied in [63, 64] considering different scenarios. However, most work in FRF is modeled for a MISO system and is related to cellular systems. In [60], it is shown that ideal beamforming is impossible because as the quality of the side channel degrades, there is an inherently lower achievable performance. Similarly, [65] showed the effectiveness of a small number of bits per antenna in an FRF system.

In [66], a Multiple Input Single Output (MISO) downlink where each base station is equipped with the number of antennas greater than that in the mobile stations was considered in addition to FRF. It is shown that, for target performance, diversity of users reduces the number of CSI feedback bits. The key result of [63] is that Zero-forcing-Dirty Paper Coding (ZF-DPC) with combined quantised feedback achieves most of the capacity achieved by the perfect CSI feedback. In [64] a downlink MISO system with FRF and Zero-forcing beamforming (ZFB) is studied, showing that the throughput of the system,

for the number of antennas equal at both the mobile stations and the base stations, is very sensitive to CSI accuracy, and the number of bits must be increased with SNR to ensure that the throughput grows with SNR. Few have considered MIMO settings and estimated error bound and capacity. Related studies in [67, 68] calculate the asymptotic capacity for beamforming and have shown that the Random Vector Quantization (RVQ) scheme is asymptotically optimal. The capacity bound for transmitting antennas, receiving antennas and a number of feedback bits B approaching to infinity has been derived there. Our work is partly motivated by the aforementioned work; however, as the scenario we have considered is different, the error bound derived earlier becomes invalid.

Since we consider MU-MIMO WLANs, clients are heterogeneous and the change in the number of antennas at clients, M, should be reflected in a quantisation error, throughput loss and in the choice of the number of quantisation bits, B. Furthermore, as mentioned earlier, there will always be residual interference, in addition to thermal noise, at undesired client which cannot be removed although a ZF transmission strategy is used. It is because in FRF system, the beamformer is based on the quantised channels from the desired and the undesired clients, however, the downlink and the uplink channels of desired and undesired clients are real (not in quantised form). Thus the beamformer cannot completely eliminate interference to the undesired clients as there is no orthogonality between quantised beamformer and the unquantised channels of undesired clients.

For instance, say $\hat{\mathbf{v}}_j$, is the quantised beamformer calculated for the jth AP taking into account the quantised desired channel, say $\hat{\mathbf{H}}_{jj}$, and an undesired quantised channel, say $\hat{\mathbf{H}}_{ji}$, for a FRF-based MU-MIMO WLANs. However, during the downlink and the uplink, the desired and the undesired channels are real, i.e., not quantised channels, say $\tilde{\mathbf{H}}_{jj}$ and $\tilde{\mathbf{H}}_{ji}$ respectively. Thus, $\hat{\mathbf{v}}_j$ cannot remove interference to the undesired clients as $\tilde{\mathbf{H}}_{ji}^H \hat{\mathbf{v}}_j \neq \mathbf{0}$, $\forall i \neq j$ because $\hat{\mathbf{v}}_j$ projects $\hat{\mathbf{H}}_{jj}$ on to the complementary orthogonal column space of $\hat{\mathbf{H}}_{ji}$ and not $\tilde{\mathbf{H}}_{ji}$. Thus there is always residual interference given the

constraint $\tilde{\mathbf{H}}_{ji}^{H}\hat{\mathbf{v}}_j \neq \mathbf{0}$, $\forall i \neq j$. However, the optimal ZF precoding vector is vital for the ZF transmission strategy. Thus we consider the optimal ZF precoding vector with relaxed residual interference to undesired clients up to a predetermined level.

To this end, we first quantify the residual interference, $\tilde{\mathbf{H}}_{ji}^{H}\hat{\mathbf{v}}_j$, and relax the interference and power constraints for the calculation of the optimal ZF precoding vector. The interesting result is that the optimal ZF precoding vector can be tuned with B, used for channel quantisation.

There has been some significant prior work regarding the RZF framework, specifically for MISO systems [69, 70, 71]. In [69] beamforming vectors for the rate maximization are considered with the given set of interference leakage levels for the MISO system. However, the leakage levels are not quantified and the solution has been suggested by the matrix inversion lemma forming sequential orthogonal projection combining (SOPC). The necessary condition for a Pareto Optimal rate tuple for MISO-IC, depending on a particular set of interference temperature (IT) constraints, is derived in [70], where the channel capacity subject to IT to other clients is considered. The key finding of [71] is that the original family of non-convex optimization problems, pertaining to the interference channel with single user detection, can be converted to the convex optimization problem, and beamforming is able to achieve all boundary points in the achievable rate region.

Although the research framework is RZF, similar to what we consider, the scenarios considered are different, leading to different constraints and solutions. Also, the approaches to solutions are different depending on the complexity, practicality and objective of the solutions. In this work, we principally focus on the influence of the quantisation error on the optimal ZF precoding vector in FRF-based MU-MIMO WLANs.

For the purpose of exposition, we consider a MU-MIMO-based WLANs scenario, where there are K APs operating in a particular frequency band in a particular area. There are heterogeneous clients associated with each AP which may lie in the transmission

overlapping region of APs, giving rise to the HT problem. Our considered case may resemble a residential area or various enterprises in an urban setting, where MU-MIMO WLANs are densely deployed, distributive and uncoordinated. Owing to the feedback constraints, as described earlier, each heterogeneous clients feeds back their channels to APs, in the form of a quantisation index after quantising with (quantisation bits) B, i.e., the system follows the FRF model. APs calculate the optimal ZF precoding vector with the quantised channels and use the suboptimal but pragmatic ZF transmission strategy for serving client requests.

In such a context, although the ZF technique is used, interference to undesired clients cannot be removed completely as the beamformer is based on quantised channels. Thus we relax the interference leakage to a predetermined level. Furthermore, since the transmissions of APs overlap, we model the scenario as an MU-MIMO based interference channel (IC). We then focus on calculating the optimal precoding vector at APs considering the interference leakage to the undesired clients. For this, we first quantify the upper bound of interference leakage in terms of the quantisation error and calculate the optimal ZF precoding vector. As the quantisation error fundamentally depends on the quantisation bits B, we find that the optimal ZF precoding vector can be tuned with B. The obtained precoding vector permits interference to the undesired clients up to a certain predetermined level at the time of transmission. Finally, we present a distributive algorithm for APs to calculate the optimal ZF precoding vector suited for the decentralized and distributive nature of MU-MIMO WLANs.

2.3 Summary

This chapter presents a comprehensive literature review on the solutions of the HT problem in WLANs. First, we start with the pioneering work that discovered the HT problem

and suggested the initial solutions. The Busy-tone scheme, the RTS/CTS scheme and their various derivatives are discussed. Their limitations are stated. Alternative solutions such as light-weight RTS/CTS and the Zigzag solution for HT are presented. Second, the change in the solution approach to HT problem due to the advent of MIMO technique is outlined.

In order to present the context of our proposed solution to address the HT problem in MU-MIMO WLANs, different MAC designs based on the MIMO technique, evolved over time, considering medium access mechanism, downlink, uplink, scheduling, channel acquisition are discussed. Our brief MAC design synopsis: the ZF technique as transmission strategy, PCF as medium access mechanism, a hybrid scheduling scheme with, packet position-based FIFO and the channel quality-based scheme, namely the Best the Two Choices and explicit channel acquisition, is presented. Third, given the practical channel feedback constraints, the necessity of an important practical dimension, the partial CSI feedback, for our design is discussed. Since the ZF technique is an integral part of our MAC design, the effect of partial CSI on the system parameters are discussed. Prior work relating to the partial CSI context in the cellular system is surveyed. The difference in the context and the need of new analysis is presented. Finally, our solution approach at the PHY layer is concluded with a brief discussion on the optimal ZF precoding vector and the distributive algorithm for the APs in MU-MIMO WLANs.

Chapter 3

Hidden-Terminal Based System Model in WLANs and Design Opportunities

This chapter presents the system model we have considered for the HT problem in MU-MIMO WLANs and the design challenges and opportunities based on the system model. Initially, we discuss our system model considering only two APs in a MU-MIMO WLAN, for simplicity. Later, we develop a generic model, with K APs and clients associated with them, suited for a general case in MU-MIMO WLAN settings. In addition, design challenges and opportunities are throughly analysed based on the HT model. Finally, two specific requirements for addressing the HT problem in MU-MIMO WLANs are illustrated which led to the design and the development of our proposed solution.

3.1 System Model

Consider a HT scenario as shown in Fig. 3.1 where the ith network AP2 and the jth network AP1 are out of carrier sensing range of each other, so they cannot hear each other's transmissions. However, AP1 and AP2 can transmit to their clients, 'I4', 'I5' and 'I6', and 'I1', 'LP', 'I2', 'HDTV', 'I3', respectively, at the same time, causing collision of signals and decoding failure.

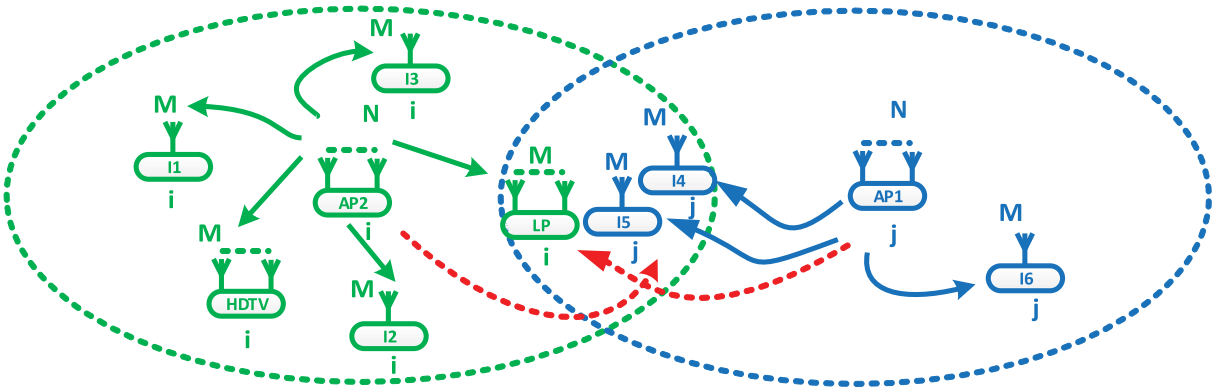

Figure 3.1: The HT problem with a specific two-network scenario, i.e., K=2.

For simplicity and ease of discussion, we assume clients 'I1'(1), 'I2'(1), 'I3'(1), 'I4'(1), 'I5'(1) and 'I6'(1) have 1 antenna whereas clients 'LP'(2) and 'HDTV'(2) have 2 antennas. The corresponding antenna/s associated with the clients is/are shown in parentheses whenever necessary.

Suppose there are K APs who are out of carrier sensing range of each other and are transmitting simultaneously[1]. Clearly, at the jth reference client 'I4', we have collision of signals that are coming from $K-1$ HT nodes. For simplicity and ease of discussion, we take the jth network AP1 and client 'I4' as a reference as shown in Fig. 3.1. We consider N transmitting antennas (different for each AP) at the APs and M receiving antennas

[1]It is to note that in the K APs scenario, 'I4', 'I5' and 'LP' in Fig. 3.1 are considered in the common coverage of K APs.

(different for each client) at the clients.

Our considered case may resemble a residential area or various enterprises in an urban setting, where MU-MIMO WLANs are densely deployed, distributive and uncoordinated. The received signal at the jth network client 'I4' is now given by

$$\mathbf{y} = \sum\nolimits_{i=1}^{K} \mathbf{H}_{ij}^{H} \mathbf{x}_i + \mathbf{w} \tag{3.1}$$

where the received signal is $\mathbf{y} \in \mathbb{C}^{M\times 1}$, $\mathbf{H}_{ij}$ is the channel associated with the ith AP to the jth client, $\mathbf{H}_{ij} \in \mathbb{C}^{N\times M}$ and transmitted signal $\mathbf{x}_i \in \mathbb{C}^{N\times 1}$. The noise term is represented by $\mathbf{w} \in \mathbb{C}^{M\times 1}$ and is modelled as circularly symmetric additive white Gaussian noise with zero mean and σ^2 variance. All the APs satisfy the transmit power constraint P, i.e., $\mathbb{E}\{\|\mathbf{x}_i^2\|\} < P$, for $i \in \{1, .., j-1, j, ...K-1, K\}$. The concatenation of channels at the jth network client 'I4' is given by

$$\mathbf{H}_{total} = \left[\mathbf{h}_{1j}^{H}, \mathbf{h}_{2j}^{H}, ...\mathbf{h}_{jj}^{H}, ...\mathbf{h}_{K-1j}^{H}, \mathbf{h}_{Kj}^{H}\right] \tag{3.2}$$

where $\mathbf{H}_{total}$ is a $[M \times KN]$ matrix with the jth row equal to the channel of the K APs to the jth antenna element of the jth client with M antennas.

The fundamental assumption regarding our system model is as follows. We assume that the interference level at each receiver/client in our system model is in the weak interference regime, i.e., the interference is at $0 < a = b \leqslant a^*$ where a and b are interference coefficients (if a two-interferer based interference channel (IC) is considered) and a^* is the critical value where there is a sudden change in the slope between the achievable rates R_1 and R_2 [72]. By this we mean that, first, the receivers make no attempt to decode and subtract the received interference from the $K-1$ APs to get their desired signal. Second, the only way to get rid of the collision of signals at the receivers in our system model is by preventing the interference reaching the receivers. The main justification for this assumption is that, in most envisioned MU-MIMO WLAN applications, clients would use receivers with a simple architecture. Additionally, in a dynamic environment

of WLANs, interference subtraction is difficult as the receivers do not know the coding and modulation schemes used by the interfering transmitters.

Our assumption is directly in contrast with the presence of very strong interference, where Carleial [73] in 1978 showed the striking fact that very strong interference is as innocuous as no interference. The interfering signals are so strong in this case that the receivers may decode them reliably and subtract them from the received signal to obtain the desired signal.

3.2 Design Opportunities

It is evident from the system model present in Section 3.1 that, in the HT scenario, collision of signals, causing decoding failure at receivers, is due to the presence of weak interference from the undesired transmitters. Thus the natural choice in order to get rid of collision of signals at receivers is to remove interference at the receivers. We use the ZF technique [74, 75] for the mitigation of interference at receivers. Our choice, the ZF technique, to remove interference at receivers is principally motivated by the following perspective: today's WLANs use multiple antennas both at transmitter and receiver to form MIMO systems. In such a context, ZF is a very promising candidate as it is found to be both effective and pragmatic in mitigating interference at receivers.

Studies [76, 77, 78] discuss the sub-optimal performance of the ZF technique with respect to its optimal counterpart, Dirty Paper Coding (DPC) [79]. However, owing to the complexities of the optimal solution [80, 81, 82], leading to practically unrealisable hardware design, ZF is simple and realisable in hardware platforms. Additionally, it possesses an acceptable capacity, which basically characterises the capacity versus SNR curve as in DPC [83, 78], even though it has an absolute rate/power offset.

Having said that, the principal technique ZF in its conventional form cannot provide

an apt solution in MU-MIMO WLAN settings. This is due to the application-specific requirements of WLANs (which are to be discussed shortly) which fundamentally give rise to two basic prerequisites for the ZF precoding vector in MU-MIMO WLANs. First, to cancel interference to undesired clients and remove collision of signals. Second, to maintain concurrent transmission to desired clients with optimal fairness and network throughput. Thus we propose a fairness and a throughput-aware ZF precoding at PHY to address the HT problem in MU-MIMO WLANs.

3.2.1 ZF Precoding for MU-MIMO WLANs in Our Design

To place the proposed ZF precoding algorithm of our design in context, we begin with a brief introduction to conventional Zero-forcing Beamforming (ZFB) with reference to a Multi User-MIMO-interference channel (MU-MIMO-IC). Subsequently, we discuss the application-specific requirements of MU-MIMO WLANs. The aim is to highlight the need to adopt and enhance the classical ZFB to effectively address the HT problem in MU-MIMO WLANs. Finally, we present an algorithm to precode APs' transmissions in a HT scenario which not only cancels interference to undesired clients and removes collision of signals but also maintains concurrent transmission to desired clients with optimal fairness and network throughput.

3.2.2 Primer-Zero-forcing Beamforming in Multiuser Interference Channel

ZFB is a transmission strategy where the SNR of each desired stream is maximised subject to the constraint that interference to other clients is completely eliminated [74, 75].

Consider a MU-MIMO-IC with K transmitters having N antennas and their corresponding receivers, each with a single antenna. The interference to the undesired receivers

is completely eliminated by ZFB. Thus, the best ZF beamforming vector for any ith transmitter is given by solving the following optimisation problem for $i \in \{1, ...K\}$

$$\begin{aligned} \max_{\mathbf{v}_i} \log\left(1 + \frac{\left|\mathbf{h}_{ii}^H \mathbf{v}_i\right|^2}{\sigma_i^2}\right) & \\ s.t \left|\mathbf{h}_{ij}^H \mathbf{v}_i\right|^2 &= 0 \, \forall j \neq i \\ \left\|\mathbf{v}_i\right\|^2 &\leqslant P_i \end{aligned} \tag{3.3}$$

where $\left|\mathbf{h}_{ij}^H \mathbf{v}_i\right|^2 = 0$ is the ZF leakage constraint of the ith transmitter to the jth receiver. The optimisation problem has a non-trivial solution given by $\mathbf{v}_i^{ZF} = c \prod^{\perp}_{[h_{i1},.....h_{ii-1},h_{ii+1}.....h_{iK}]} \mathbf{h}_{ii}$, where c is the scalar satisfying the transmit power constraint. The necessary condition for the non-trivial solution is that the number of antennas at transmitter N is greater than or equal to the number of antennas at receiver M, i.e., $N \geqslant M$.

The Signal to Interference plus Noise Ratio (SINR) for the jth client in a network is given by

$$SINR_j = \frac{\left|\mathbf{h}_{jj}^H \mathbf{v}_j\right|^2}{\sigma_j^2 + \sum\limits_{i \neq j}^{K-1} \left|\mathbf{h}_{ij}^H \mathbf{v}_i\right|^2}. \tag{3.4}$$

3.2.3 Application-specific Requirements of MU-MIMO WLANs

The HT scenario in MU-MIMO WLANs fundamentally corresponds to a multiuser IC as discussed in Subsection 3.1.1. In such cases, one may wonder whether the solution to the optimisation problem in (3.3) can be applied to remove interference to the undesired clients and manage collision of signals arising from the HT problem.

However, the solution is not so straightforward. We particularly see two specific requirements of MU-MIMO WLAN settings while addressing the HT problem that exists there.

First, unlike cellular systems, MU-MIMO WLANs have heterogeneous antenna clients

(it may be at home networks or office network settings or elsewhere) such as iPhones, laptops, HDTVs, etc. The numbers of antennas these clients possess are different, perhaps due to space and cost constraints. For instance, iPhones have a single antenna whereas laptops and HDTVs can have two or more antennas. As there are heterogeneous antenna clients, fairness in clients' access with network throughput is equally important with interference mitigation to the clients.

We see that fairness and network throughput in the context of heterogeneous antenna clients are important because it is highly probable that clients with increasingly greater numbers of antennas with high channel gains may become a more lucrative incentive for APs than clients with a smaller numbers of antennas with low channel gains in MU-MIMO WLANs. This may lead to APs choosing only clients with a larger number of antennas, resulting in a deep unfairness in client access with a smaller number of antennas.

Thus the design challenge as a whole is, not only to mitigate interference to undesired clients, but also maintain the balance and serve heterogeneous antenna clients, with two contrasting objectives: a) fairness in client access among (heterogeneous) desired clients, b) maximise the network throughput. These specific objectives put a stringent requirement on the precoding vector while we address the HT problem in MU-MIMO WLANs. Shortly, the implications of these specific objectives and our approach to the solution will be discussed. Without loss of generality, we focus on the downlink.

For instance, consider the previous MU-MIMO WLAN setting in Fig. 3.1. Specifically, whenever say the 6-antenna AP2, decides to transmit, it has to satisfy two conditions. First, remove interference to the undesired clients ('I4' and 'I5') to get rid of the collision when AP1 transmits to them. Second, maintain concurrent transmission among the desired clients ('I1', 'LP', 'I2', 'HDTV' and 'I3') with fairness. It is worthwhile to note that in our design both conditions have to be satisfied at the same time by the APs (focusing on the downlink) with an appropriate precoding vector based on ZFB.

For calculation of the precoding vector, APs have to consider the channel realisation associated with the undesired clients, since interference cancellation is an important step in our design (The detailed process for acquiring CSI is given in Section 6.2). Having said that, it also becomes equally important for AP2 to consider who among the desired clients 'I1', 'LP','I2', 'HDTV', and 'I3' are to be served concurrently.

We highlight that the choice of the desired clients is vital in two respects. First, it is related to the concurrent transmission among the desired clients, which in fact is associated with the AP's network objectives such as maximising the network throughput or maintaining fairness among the served clients, and or both. We may consider this scenario as a popular load-balancing problem [84] among desired clients. Second, the choice of the desired clients forms an integral part in the precoding vector (a detailed description is given in Section 4.1) which AP2 needs to calculate before transmission.

Thus the design challenge for the HT problem with heterogeneous desired and undesired clients in MU-MIMO WLANs is to calculate a suitable precoding vector at the APs so that it can achieve the following simultaneously: a) remove interference to undesired clients and avoid collision of signals. b) maintain fairness and network throughput among the desired clients with concurrent transmissions.

Second, from a general WLAN network perspective, meaning that there are numbers of APs operating at the same frequency in a certain area, there can be more than two APs involved in causing the HT problem. In such a context, merely an AP-based fairness and interference mitigation precoding approach will not be adequate to completely address the HT problem. The implications may be as follows: a) It is with high probability in a dynamic MU-MIMO WLAN setting that there can be more than one AP involved in creating a HT scenario.

b) No sooner had the interferences from one AP been removed than the same receiver might start experiencing interference from other APs.

In such a context, it seems that a centralised coordination and control mechanism for MU-MIMO WLAN networks would properly address the HT problem. However, this would not be a feasible approach, as MU-MIMO WLAN networks are decentralised and distributive in nature. Hence another specific requirement is to develop a medium-access mechanism which not only solves the HT problem in the MU-MIMO WLANs but also works independently at each AP without any centralised coordination. Specifically, there is a need for an appropriate decision-making scheme at the medium access level which can address the TXOP issue among APs.

We shall now discuss fairness and network throughput in concurrent transmissions and TXOP among APs in more detail.

3.2.4 Implications of the Design

In order to address the application-specific requirements of MU-MIMO WLANs, the research issues are: a) fairness in client access and throughput of the network. b) transmission opportunity for APs from a WLAN network perspective. We shall discuss each of them as follows:

a) Fairness and network throughput: The ZF solution allows APs to transmit to their desired clients simultaneously/concurrently, while at the same time cancelling interference to undesired clients in their transmission range in order to avoid collision of signals. However, we recall the fact that interference cancellation by APs to undesired clients with ZF requires precisely as many DoFs of APs as the number of antennas at the undesired clients. This means that APs with the remaining DoF have to operate concurrent transmissions among the desired clients.

For instance, consider the previous network at Fig. 3.1. Whenever AP2 is supposed to transmit, addressing the HT problem existing there by using our design costs AP2 exactly 2 DoFs to remove interference to the undesired clients (i.e., 'I4'(1) and 'I5'(1)). Recall

that the number in parentheses indicates the number of antenna/s each client possesses. Thus AP2 can have only 4 DoFs, by which it can serve at most four antennas concurrently while cancelling interference to the undesired clients at the same time.

However, there are 5 desired clients; namely 'I1', 'LP', 'I2', 'HDTV', and 'I3' in the ith network to be served. Now the critical question is how AP2 with TXOP chooses the clients from the desired clients so that the remaining DoF that AP2 has (i.e., 4) can be best utilised in favor of network objectives such as maximising the network throughput or maintaining fairness among the served clients, or both. We may view this scenario as a popular load-balancing problem. Among many possible solutions, we focus our discussion on the fairness issue among the desired clients and the throughput of the network. The detailed discussion for the fairness and the network throughput-aware ZF precoding is given in Section 4.1.

b) transmission opportunity for APs: It is worthwhile to note that, instead of using the total available DoF of the APs, we are interested in whether or not APs which are willing to transmit satisfy the required DoFs of the network. It is because in the HT scenario, whenever APs transmits, they have to cancel interference to the number of antennas at the undesired clients, also known as required DoFs. Hence the remaining DoFs=total DoFs - required DoFs. In such a context, in order to ensure interference-free transmission of APs, and hence collision-free reception, it becomes mandatory for APs that, whatever time instant APs decide to transmit, they satisfy the required DoF for transmission.

The aim of such provision in the design is to avoid the situation when transmitters win contention but do not possess the required DoF, resulting in a collision of signals. Thus we clearly see the need to design a medium-access mechanism based on both time and the remaining DoF of transmitters, unlike the traditional time-division based (RTS/CTS) medium-access mechanism in WLANs.

3.3 Summary

Two important aspects of our research are presented. First, a system model for the HT problem in MU-MIMO WLANs is illustrated. Second, design challenges and opportunities based on the system model is explored. Initially, a specific case with only two APs and the clients associated with them are considered for simplicity and ease of discussion. Subsequently, a more generic model for the HT problem in MU-MIMO WLANs is discussed. Drawing on the design challenges and opportunity, we principally identified two specific requirements of MU-MIMO WLANs while addressing the HT problem: First, fairness and network throughput from a single-AP perspective. Second, TXOPs for APs from a multiple-AP perspective. Finally, we found out that there is a clear need to come up with a new design for solving the HT problem in MU-MIMO WLANs.

Chapter 4

Fairness and Network Throughput-Aware Precoding

4.1 Introduction

In the previous chapter, we discussed the application-specific requirements of MU-MIMO WLANs and identified the need for fairness and throughput-aware precoding vectors based on ZF to address the HT problem in MU-MIMO WLANs. We highlighted that the precoding vectors should possess two key attributes: a) to cancel interference to the undesired clients and b) to maintain both fairness in client access and network throughput while addressing the HT problem in MU-MIMO WLANs. In this chapter, we focus on how such fairness and throughput-aware precoding vectors can be obtained in MU-MIMO WLAN settings. The content of this chapter constitutes a published work by the author in [54].

4.2 Steps to Obtain the Precoding Vector

For illustrative purposes, we proceed through a sequence of steps in order to obtain the precoding vector. Every step is discussed in the following subsections which ultimately yields a fairness and throughput-aware precoding vectors based on ZF.

4.2.1 Channel Acquisition via Channel Sounding

As CSI is the prime prerequisite for the ZF technique, transmitters/APs in our design first use a channel sounding process and measure CSI. Since APs need to know the CSI of both the desired and the undesired clients in order to effectively address the HT problem [1], channel sounding is performed for both the desired and undesired clients within their transmission range. Additionally, APs need to distinguish between the channels from the desired and the undesired clients. To do this APs use Association IDs (AID) (which are assigned by the APs to their desired clients during the association process. A more rigorous description of our channel sounding process including the distinction among the desired and the undesired clients is given in Section 6.2.). We here emphasise that, upon the completion of channel sounding, APs in our design will have CSI from both the desired, i.e., $\mathbf{H}_{ii_Q} = \left[\mathbf{h}_{ii_1}, \mathbf{h}_{ii_2}, ..., \mathbf{h}_{ii_Q}\right]$, and undesired, i.e., $\mathbf{H}_{ii_P} = [\mathbf{h}_{ii_1}, \mathbf{h}_{ii_2}, ..., \mathbf{h}_{ii_P}]$ clients, and ultimately form a total CSI, $\mathbf{H} = \left[\mathbf{h}_{ii_1}, \mathbf{h}_{ii_2}, ..., \mathbf{h}_{ii_Q}, \mathbf{h}_{ij_1}, \mathbf{h}_{ij_2}, ...\mathbf{h}_{ij_P}\right]$, where the numbers of desired clients and undesired clients are Q and P respectively.

4.2.2 TXOP Decision

The decision on whether or not APs are allowed to transmit at a given time instant, known as TXOP, fundamentally leads APs to calculate the precoding vector. APs with

[1] This is required as per the objective of our design, for maximising the throughput among desired clients and removing interference to undesired clients.

no TXOP do not calculate the precoding vector at that instant as they are not allowed to transmit at that time. In this section, we briefly discuss how TXOPs for APs are decided in our design, which further clarifies the context for which precoding vectors are calculated.

APs know the number of antennas they possess, N in our system model, meaning that they have N DoFs, and the number of antennas at undesired clients (i.e., PM in our system model) due to channel sounding as per Subsection 4.2.1. Then, APs simply check whether $N > PM$ for TXOP. Note that our design checks $N > PM$ rather than $N \geqslant PM$ at APs. The aim is to ensure that APs at least have one DoF after TXOP.

If the condition is satisfied, APs win TXOP and remain in an 'Active mode', otherwise they lose TXOPs at that instant and remain in a 'Silent Mode'.

Put another way, 'Active mode' and 'Silent mode' APs are decided on whether or not the available DoFs each AP possesses, are sufficient for transmission at that instant. This is an important decision because those APs who have available DoF have the ability to remove interference to the undesired clients. So they will remain in the 'Active mode' and become ready for concurrent transmission in the next Step. Otherwise, APs will remain in the 'Silent mode', because if they transmit with the lack of available DoF they end up interfering with some undesired clients within their transmission range, causing collision of signals. More details about the modes of APs are discussed in Subsection 6.2.2. We here highlight that 'Active Mode' APs are supposed to calculate fairness and throughput-aware precoding vectors.

4.2.3 Concurrent Transmission Algorithm to Find Clients Among Desired Clients

Before the 'Active mode' APs attempt concurrent transmissions, they should know how many concurrent transmissions have to be made. APs calculate this by checking how

many remaining DoFs they have at that instant, i.e., $if(N - PM) > 0$. The obvious reason is that we can at most have as many concurrent transmissions as the remaining DoFs that APs have in the 'Active mode'.

For instance, let $(N - PM) = D$, where D is the number of remaining DoFs of APs. Now, APs run the concurrency/scheduling algorithm and select D out of Q channel realisations, keeping in mind network throughput and fairness among the desired clients. For instance, $\mathbf{H}_{ii_D} = [\mathbf{h}_{ii_1}, \mathbf{h}_{ii_2}, ..., \mathbf{h}_{ii_D}]$ are the channel realisations of the selected clients by the concurrency algorithm, where $\mathbf{H}_{ii_D}$ is a subset of $\mathbf{H}_{ii_Q}$, i.e., $\mathbf{H}_{ii_D} \subset \mathbf{H}_{ii_Q}$, and the number of channel realisations is less than (if there are insufficient client requests on uplink) or equal to the remaining DoFs, i.e., $\mathbf{H}_{ii_D} \leq (N - PM)$.

The mechanism on how $\mathbf{H}_{ii_D} = [\mathbf{h}_{ii_1}, \mathbf{h}_{ii_2}, ..., \mathbf{h}_{ii_D}]$ are selected is part of the concurrency/scheduling algorithm and detailed discussions are given in Section 4.2. We here underline that D clients are selected among Q desired clients by concurrency algorithm.

4.2.4 Calculation of Precoding Vector

After $\mathbf{H}_{ii_P}$ and $\mathbf{H}_{ii_D}$ are determined as per Subsection 4.2.1 and Subsection 4.2.3, the precoding vector is calculated. For example, the precoding vector $\mathbf{v}_i \in \mathbb{C}^{N\times 1}$ for AP2 is given by

$$\mathbf{v}_i = \frac{\prod^{\perp}_{\mathbf{H}_{ij_P}} \mathbf{H}_{ii_D}}{\left\| \prod^{\perp}_{\mathbf{H}_{ij_P}} \mathbf{H}_{ii_D} \right\|} \mathbf{U} \tag{4.1}$$

where $\prod^{\perp}_{\mathbf{H}_{ij_P}} = \mathbf{I_N} - \mathbf{H}_{ij_P}(\mathbf{H}_{ij_P}{}^H \mathbf{H}_{ij_P})^{-1} \mathbf{H}_{ij_P}{}^H$ denotes the projection onto the orthogonal complement of the column space of $\mathbf{H}_{ij_P}$. $\mathbf{I}_N$ represents the identity matrix of size N. $\mathbf{U} \in \mathbb{C}^{DM\times 1}$ is a unit vector acting as a demultiplexer where $\mathbf{U}^H\mathbf{U} = 1$. The condition $N \geq PM$ has to be satisfied in order to take the left inverse.[2]

[2]The right inverse can still be taken if $N < PM$, however this will not increase the concurrent transmissions as the number of DoFs of the network is min{N,PM}.

The received signal at the jth reference client 'I4' of the jth network AP1 is given by

$$\mathbf{y} = \mathbf{H}_{jj_D}^H \frac{\prod_{\mathbf{H}_{ji_P}}^{\perp} \mathbf{H}_{jj_D}}{\left\| \prod_{\mathbf{H}_{ji_P}}^{\perp} \mathbf{H}_{jj_D} \right\|} \mathbf{U}\, s_j + \sum_{i \neq j}^{K-1} \mathbf{H}_{ij_P}^H \frac{\prod_{\mathbf{H}_{ij_P}}^{\perp} \mathbf{H}_{ii_D}}{\left\| \prod_{\mathbf{H}_{ij_P}}^{\perp} \mathbf{H}_{ii_D} \right\|} \mathbf{U}\, s_i + \mathbf{w}. \tag{4.2}$$

4.3 Concurrent Transmission Algorithm

The concurrent transmission algorithm is an integral part of our design which selects D clients out of Q desired clients. It helps the 'Active mode' APs decide which of the desired clients within the network are to be served concurrently. This is one of the fundamental steps in our scheme, as it determines fairness and throughput-aware precoding vectors, as described in Subsection 4.2.3. In this section, we illustrate how the concurrent transmission algorithm is designed and later present the performance of our concurrent transmission algorithm.

4.3.1 Concurrent Transmission Algorithm Design

For instance, let us consider the previous network with AP2 with 4 remaining DoF and there are 5 desired clients namely, 'I1', 'LP', 'I2', 'HDTV', 'I3', in the network to be served. Since AP2 in the network has only 4 remaining DoFs, it can at most serve 4 antennas concurrently. Now the critical question is how AP2 with TXOP chooses the clients among 'I1', 'LP', 'I2', 'HDTV' and 'I3' so that the objectives such as maximising the network throughput or maintaining fairness among the served clients or/and both is/are met. We may view this scenario as a popular load-balancing problem among the desired clients.

For equal fairness among the clients within a network, desired clients can be served based on FIFO packet queues. As for example, suppose that the packets are queued thus: I1(1), LP(2), I2(1), HDTV(2) and I3(1). Recall that the number in parentheses indicates the number of antenna/s each client possesses. In the FIFO algorithm, the first 4 packets

in the queue are taken (because 6-antenna AP2 has only 4 remaining DoFs while it has to null signals to 2 antennas of the undesired clients) and will be served first and the remainder will be served accordingly. However, AP2 will be unable to use 1 DoF in some FIFO queue patterns, for example, a) I1(1), I2(1), I3(1), b) I1(1), LP(2) and c) HDTV(2), I1(1), where two or three consecutive FIFO clients already occupy 3 DoFs (out of 4 DoFs) of AP2 and is followed by a 2-antenna client request, either LP(2) or HDTV(2). This is because the selection of 2-antenna clients would exceed the remaining DoFs of APs at that instant.

Additionally, clients are chosen according to the packet queue request, irrespective of the channel quality. Thus this method is highly oblivious to the network throughput, though it can ensure fairness in terms of FIFO request queues.

For maximising the network throughput, one can choose the Brute-Force approach, where all the combinations among the desired clients, subject to the number of antennas being equal to the remaining DoFs of AP (i.e.,4 in the example) are checked and the best among the combinations that maximises the throughput are chosen.

As in the case considered with clients queue: 'I1'(1), 'LP'(2), 'I2'(1), 'HDTV'(2) and 'I3'(1), we simply use a combination formula [85, p.17], i.e., $\begin{pmatrix} n \\ k \end{pmatrix} = \frac{n!}{k!(n-k)!}$, whenever $k \leqslant n$, for $C(n,k)$, where n and k are 5 and 3 respectively, resulting in 10 combinations[3] of desired clients. Out of the 10 possible combinations, four combinations are unable to satisfy the constraint that the total number of antennas should be equal to the remaining DoFs of AP, are discarded. Thus there are 6 combinations among the 5 desired clients: I1(1), LP(2), I2(1) HDTV(2), and I3(1).

We choose the best combination among the 6 possible combinations for concurrent transmission that maximise the throughput at that instant. Thus the Brute-Force method

[3]We take a Singular Value Decomposition for each combination of channels and the position of the channel realisation does not matter in the throughput calculation.

will ensure a maximum throughput of the network. Nonetheless, it would undoubtedly end up with unfairness among the clients who cannot maximise throughput because they are not chosen for transmission. Also, this method can be very cumbersome for a large number of users, as the combinations grow because all the possible combinations of clients have to be checked.

In order to balance both the fairness and the throughput of the network, one possible solution would be the combination of the FIFO and the Best of the Two Choices. For fairness, we always select the first client in the queue for transmission. For the rest of the clients, we use the Best of the Two Choices. The Best of the two Choices is one of the standard approaches for reducing the complexity of combinatorial problems [84].

For general illustrative purposes, we use the popular balls and bin model to show the superior performance of load balancing by the Best of the Two Choices [84]. Suppose that n balls are placed into n bins, with each ball being placed into a bin chosen independently and uniformly at random. Let the load of a bin be the number of balls in that bin after all balls have been placed. It is well known that, with high probability, the maximum load upon completion will be approximately $\frac{\ln n}{\ln(\ln(n))}$. Thus, it is evident that the load is not balanced in the system.

Suppose that the balls are placed sequentially, so that for each ball we choose two bins independently and uniformly at random and place the ball into the less-full bin (breaking ties arbitrarily). In this case, the maximum load drops to $\frac{\ln(\ln(n))}{\ln 2} + O(1)$ with high probability [84].

Thus, we pick the first client in the queue for transmission to ensure fairness and the rest of the clients are chosen in accordance with a randomized design that exploits the Best of the Two Choices. In our considered scenarios of desired clients, 'I1', 'LP', 'I2', 'HDTV' and 'I3', this method results in 4 combinations out of all 10 possible combinations. Thus, we take the best combinations among the 4 which ensures the maximum network

throughput. Hence, the FIFO combined with the Best of the Two Choices balances both fairness and network throughput.

Out of the three concurrent transmission algorithms discussed above, based on fairness and network throughput, we use the FIFO combined with the Best of The Two Choices as a part of the concurrency algorithm in our design. This gives us D selected desired clients leading to $\mathbf{H}_{ii_D} = [\mathbf{h}_{ii_1}, \mathbf{h}_{ii_2}, ..., \mathbf{h}_{ii_D}]$ as described in Subsection 4.2.3. The basic pseudo-code of the FIFO combined with the Best of the Two Choices is given as **Algorithm 1**.

For instance, given AP2, $N = 6$, with the desired clients' queue 'I1'(1), 'LP'(2), 'I2'(1), 'HDTV'(2), 'I3'(1) and the undesired clients 'I4'(1) and 'I5'(1), $D = N - P = 6 - 2 = 4$. Applying FIFO + the Best of the Two Choices, APs can have 4 possible combinations $\mathbf{H}_{i_D i_D} \in \mathbb{C}^{D_N \times D_M}$= 'I1'(1),'LP'(2),'I2'(1) or 'I1'(1),'LP'(2),'I3'(1) or 'I1'(1),'HDTV'(2),'I2'(1) or 'I1'(1),'HDTV'(2),'I3'(1) (Note that, for Fairness, the first client in the queue, 'I1'(1), is always chosen irrespective of the channel quality). APs check each 4 combinations with the standard MIMO capacity, i.e., $C = \sum_{i=1}^{D} \log_2 \left(1 + \frac{SNR}{N}.\lambda_i\right)$. The combination that can maximise throughput is taken for concurrent transmission, hence maximum throughput is ensured. It is imperative to note that the collection of the desired clients, i.e., $\mathbf{H}_{ii_D} = [\mathbf{h}_{ii_1}, \mathbf{h}_{ii_2}, ..., \mathbf{h}_{ii_D}]$, is an integral component of the precoding vector for a Fairness and Network throughput-aware ZF precoding vector at APs as shown in the analytical expression in (4.1).

In a nutshell, a simple diagrammatic illustration for obtaining fairness and throughput-aware precoding vectors based on ZF is given in Fig. 4.1. The fairness and network throughput-aware ZF Precoding vector at APs satisfies one of the application-specific requirements described in Chapter 3.

Algorithm 1 Concurrent Transmission Algorithm Overview

procedure FIFO COMBINED WITH THE BEST OF THE TWO CHOICES

$D \leftarrow i - j_P$; APs calculate remaining DoF, where i and j_P are number of antennas at APs and the undesired clients at an instant.

APs then calculate $\mathbf{H}_{i_D i_D} \in \mathbb{C}^{D_N \times D_M}$out of $\mathbf{H}_{ii_Q} \in \mathbb{C}^{N \times QM}$ where $\mathbf{H}_{ii_Q} = \left[\mathbf{h}_{ii_1}, \mathbf{h}_{ii_2}, ..., \mathbf{h}_{ii_Q}\right]$ by Concurrent Transmission Algorithm (i.e., $FIFO$ + $Best\ of\ the\ Two\ Choices$). Since P antennas of APs are used for interference cancellation, we are actually interested in $\mathbf{H}_{i_D i_D} \in \mathbb{C}^{D_N \times D_M}$.

We use the following loop:

    for *iter* 1 to number of possible combinations **do**

        $offset$=*iter*-1; for selecting different clients

        for N 1 to D **do** ;N represents number of antennas in APs

            for M_d 1 to D **do**;M_d represents total number of antennas at desired clients

                if M_d==1 **then**

                    $\mathbf{H}_{i_D i_D}(N,M_d)$=$\mathbf{H}_{ii_Q}(N,M_d)$; always takes the first client in the queue.

                else

                    $\mathbf{H}_{i_D i_D}(N,M_d)$=$\mathbf{H}_{ii_Q}(N,M_d + offset)$; for selecting the different clients.

                end if

            end for

        end for

    end for

end procedure

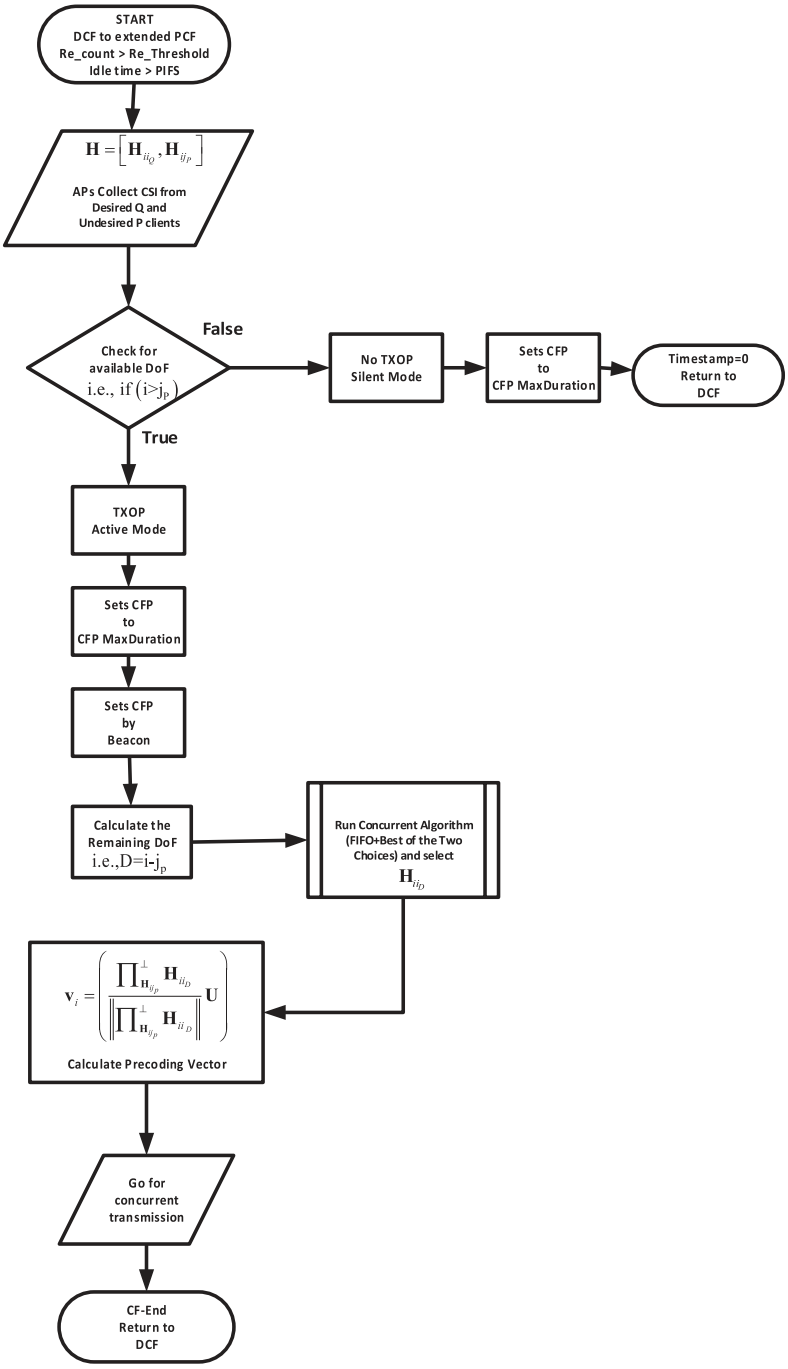

Figure 4.1: Diagrammatic illustration of the design in the PHY layer. The figure basically illustrates our way to calculate a fairness and throughput-aware ZF precoding vector at AP.

4.3.2 Performance of the Concurrent Transmission Algorithm

Subsection 4.1.3 discusses the need for the concurrent transmission algorithm and Section 4.1 presents the three possible options for concurrent transmission in detail. With the simulation studies, we checked the performance of the three concurrent transmission algorithms considering the previous MU-MIMO WLAN where there is a 6-antenna AP with 5 desired and 2 undesired clients. Since there are $N - I = 6 - 2 = 4$ concurrent

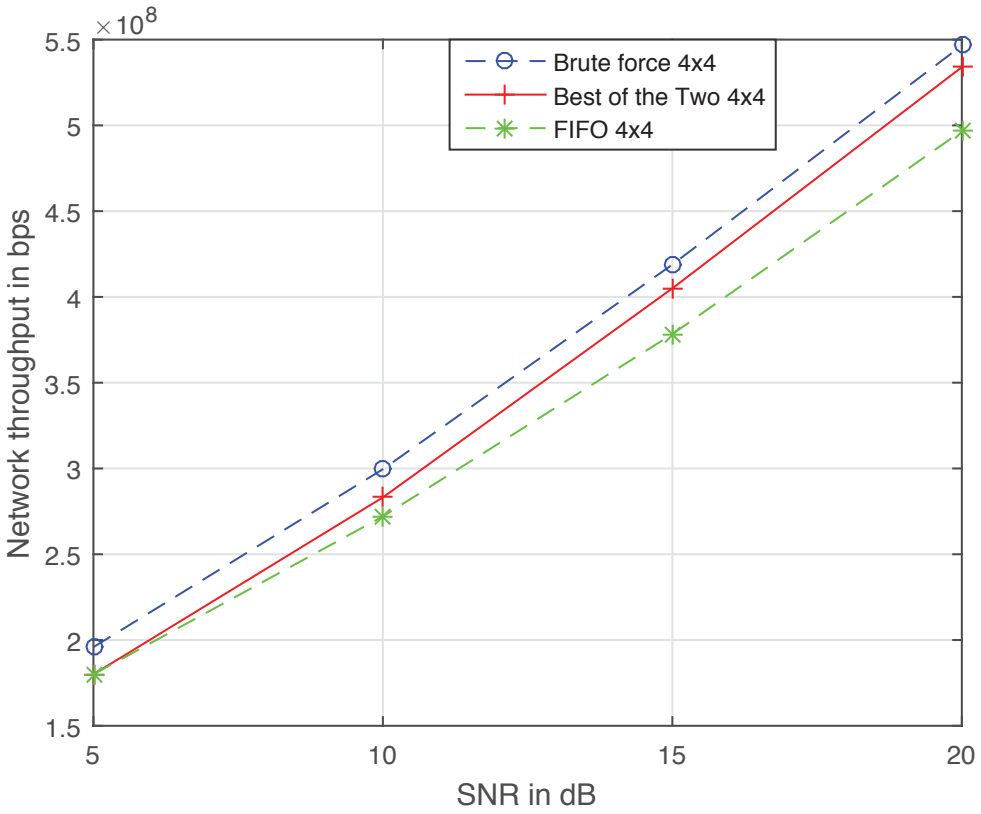

Figure 4.2: Throughput comparison among FIFO, Brute-Force and FIFO combined with Best of the Two Choices. We discuss three concurrency algorithms for APs to transmit with their desired clients within the network after winning TXOP. The throughput comparison is made among FIFO, Brute-Force and the FIFO combined with the Best of the Two Choices.

transmissions to maintain, each concurrent trasmission algorithm, FIFO, Brute-Force and

the FIFO combined with the Best of the Two Choices, was used to calculate the network throughput with increasing SNR values. The simulation result at Fig. 4.2 shows that, of the three, Brute-Force has a higher throughput, followed by FIFO combined with Best of the Two Choices, and FIFO. The reason is that Brute-Force always maximises the network throughput by selecting the best combination of served clients. FIFO combined with Best of the Two Choices takes care of the fairness issue while maximising the throughput, so can have a lower throughput. FIFO is oblivious to the network throughput as it cares most about the fairness issue.

A closer study of Fig. 4.2 shows that the maximum network throughput of the considered network can never exceed the throughput of the Brute-Force algorithm. However, in the scenarios when the best channel clients queue subsequently, the FIFO algorithm can attain the throughput performance of the Brute-Force algorithm. This case applies to the FIFO combined with the Best of the Two Choices algorithm as well, given that the first client in the queue happens to be among one of the best channels available.

4.4 Summary

In this chapter, one of the major aspects of our solution for the HT problem in MU-MIMO WLANs, fairness and throughput-aware precoding based on ZF, is thoroughly discussed. In order to explain how such precoding vectors are obtained, we adopted a step-by-step approach for illustrative purposes and explained the procedure of calculating the precoding vectors. Starting with the process of CSI collection for the both desired and undesired clients, we discussed TXOP decision for APs. Subsequently, the need for a concurrent transmission algorithm and its rigorous description is presented. Additionally, a simulation study of the concurrent transmission algorithm, FIFO combined with the Best of the Two Choices, comparing it with the FIFO and Brute-Force approaches is shown

in Fig. 4.2. Finally, a fairness and throughput-aware precoding vector is calculated as per analytical expression (4.1). Additionally, an illustrative flow diagram for the whole process is shown in Fig. 4.1.

Chapter 5

Experimental Study

5.1 Introduction

We use an experimental prototype in a laboratory setting to demonstrate that a precoding vector based on ZF is feasible and effective to solve the HT problem in MU-MIMO WLANs. The experimental prototype is set up in Wireless Research Centre, Macquarie University, and principally consists of the URSP2 [86] and other accessories such as RFX2400 daughterboards, Jackson Labs equipment along with an external Global Positioning System (GPS) antenna, Personal Computers (PCs), etc. The content of this chapter constitutes a published work by the author in [54, 58].

5.2 Experimental Prototype Setup

A simple illustrative block diagram of our experimental prototype setup is given in Fig. 5.1. Specifically, we create an HT scenario with the help of USRP2s by operating them at the same frequency, i.e., 2.45 GHz, and placing them within the transmission range of each other. A more detailed description of our experimental setup is presented in Subsections 5.2.1, 5.2.2 and 5.2.3 respectively. The performance of our testbed is basically measured

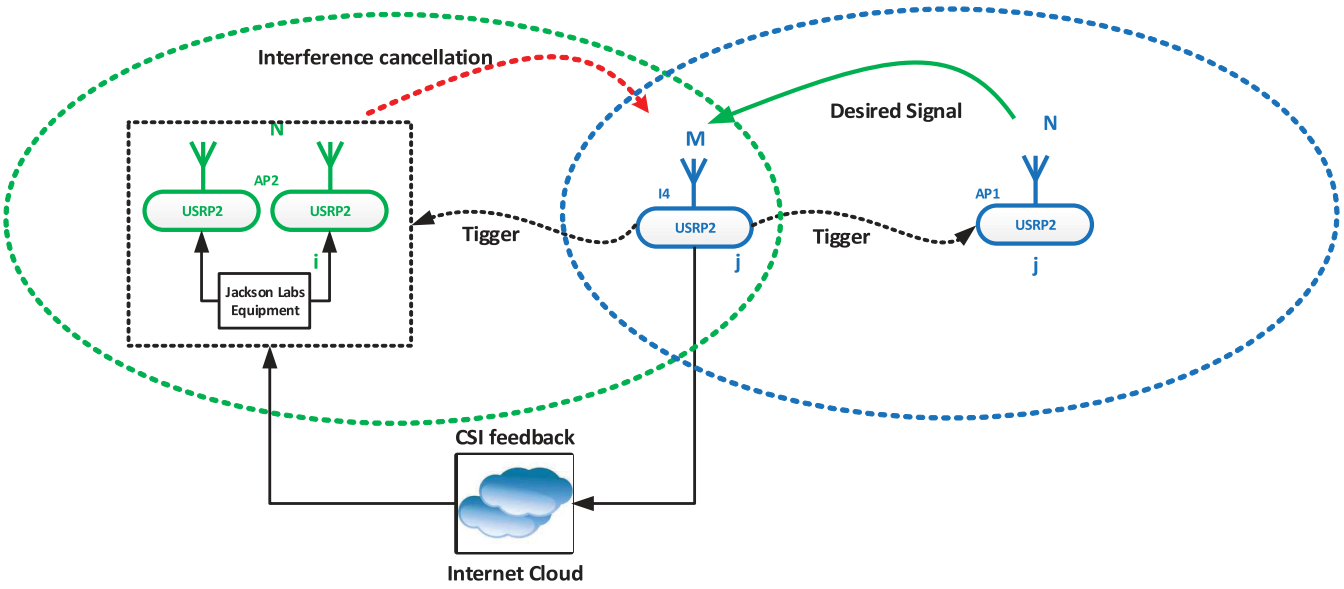

Figure 5.1: Basic block diagram of our hardware prototype implementation.

by studying the received SNR[1] at the jth client,'I4', under three settings: First, in the presence of the HT scenario. Second, while applying our design in the presence of the HT scenario and third, in the absence of the HT scenario.

We then compare and analyse the raw received signals, the SNR and the Effective SNR (ESNR)[2]. The fundamental objective of these studies is to quantify the received SNR at client 'I4' and establish the basis for performance comparison. Clearly, the higher is the received SNR at client 'I4', while applying our design to address the HT scenario, the better is the design's performance. The reason is that with the effective interference

[1]In the networking community we often compare throughput results. However, they do not bring much insight for radios that do not have proper rate adaptation. Since current GNU Radios do not yet support rate adaptation and they do not have a stable implementation of the set of modulation and coding scheme that allows rate adaptation, it is unfair to compare the throughput because the system can have the same throughput yet can have significantly higher SNR. A higher SNR could have allowed better modulation and coding schemes to achieve higher throughput.

[2]This is defined as the SNR that would give the same error performance on a narrowband channel [87].

management measure, precoding based on ZF, the SNR at client 'I4' is increased.

5.2.1 The USRP2 Platform

We primarily use USRP2, a hardware platform for a software-defined radio (SDR), to set up the HT prototype in laboratory settings. The RFX2400 daughterboards (2.3 - 2.9 GHz) are used for the Radio Frequency (RF) transmitters and receivers and Jackson Labs equipment for the clock to maintain synchronisation. The standard GNU Radio libraries [88] are used in the Ubuntu 11.04 environment. The experiment is carried out in the indoor environment with stationary nodes and with an operating frequency of 2.45 GHz. Table 5.1 shows the PHY layer parameters.

Table 5.1: PHYSICAL LAYER PARAMETERS

Parameters	**Value**
Carrier Frequency	2.45 GHz
Number of Subcarriers	64
Occupied Subcarriers	48
Bandwidth	1.5625 MHz
Fast Fourier Transform (FFT) length	64
Orthogonal Frequency Division Multiplexing (OFDM) Symbol+CP	64+64
Symbol time per Subcarrier	81.92 μs
Subcarrier Spacing	24.414 kHz
Transmission rate	585 kbps
Modulation	BPSK

5.2.2 Implementations

We construct the testbed with the help of four USRP2s equipped with RFX2400 daughter-boards. We configure two USPR2s consisting of a single antenna each, to work as a single node. In order to synchronise the two USPR2s, an external clock is provided by Jackson Labs equipment, which uses an external GPS antenna to fine-tune the clock, for the construction of a two-antenna single node. These two antennas represent the ith network AP2 as in Fig. 3.1. The remaining USRP2s, consisting of a single antenna each, are configured as the jth network AP1 and the jth network client 'I4'.

We basically focus our prototype implementation to measure the effectiveness of ZF precoding vectors to address the HT problem. The fundamental objective is to verify that our approach is effective and feasible at the hardware level. However, because of the USRP2/GNU Radio's inability to support numerous instantaneous channels between APs and clients, we take a small sample, two APs, AP1 and AP2, having a single and double antenna(s) respectively and the client 'I4' having a single antenna, from Fig. 3.1, for our case. Although a small sample is taken, the fundamental approach, ZF precoding vectors, in general can be used[3] by other APs and clients in the network to address the HT problem.

We create the HT scenario by placing the jth network AP1 and the ith network AP2 equidistant from the jth network client 'I4'. We use the same RF antenna (50 mW) to ensure equal transmission powers at both APs. This condition is vital because the transmission would otherwise demonstrate the capture effect [14], where one transmission would capture the link and start transmission while the other terminal is affected by or may starve for transmission. AP1 and AP2 are synchronised by a trigger signal from client 'I4' in order to make sure that they transmit at the same time.

[3]We may require sophisticated hardware which can handle numerous channels at the same time.

5.2.3 Channel Feedback

The implementation of the precoding vectors based on ZF to address the HT problem in our prototype setting is subject to two prerequisites: First, CSI measurement at clients. Second, timely CSI feedback to APs from clients.

In order to satisfy the first requirement, we use Time Division Duplexing (TDD) and achieve the CSI at APs with the help of the preambles[4]. The packet preamble is augmented with each OFDM data symbol[5], which consists of 7 identical OFDM symbols each of length 64. Between the OFDM symbols there are guard intervals of length 64. These form the cyclic prefix (CP) of our design[6].

Specifically, first the APs (AP1 and AP2) send the packet preambles. Second, the clients within the transmission range receive the preambles and update them to the host PC. The reception of the preamble is an important step as it contains crucial information such as OFDM symbol timing, carrier frequency offset (CFO) and channel estimation. Third, the host PC calculates the channel frequency response as shown in Fig. 5.2 and feeds it back to AP2 (In our prototype setting we use the University's Dynamic Host Control Protocol (DHCP) server for feedback purposes). After getting CSI at AP2, AP2 calculates the precoding vector for transmission. The precoding vector is then multiplied by the transmitting symbols of AP2. As a result, interference is removed at the undesired client 'I4'(1). Hence, whenever the jth network AP1 and the jth network AP2 transmit

[4]It is to be emphasised that we use a packet preamble for packet detection, OFDM symbol timing determination, carrier frequency offset (CFO) estimation and channel estimation. No separate pilot signals of any patterns or any pilot signals embedded as a part of the desired signal are used for symbol timing determination, CFO estimation and channel estimation.

[5]Each data OFDM symbol has 64 subcarriers with 48 data symbols transmitted over 48 subcarriers and 16 null subcarriers.

[6]The OFDM FFT size and the CP size are kept equal to deal with additional delays due to channel propagation and hardware turn-around time in our prototype settings. However usually, in a real implementation, the CP length is kept smaller than the OFDM FFT size.

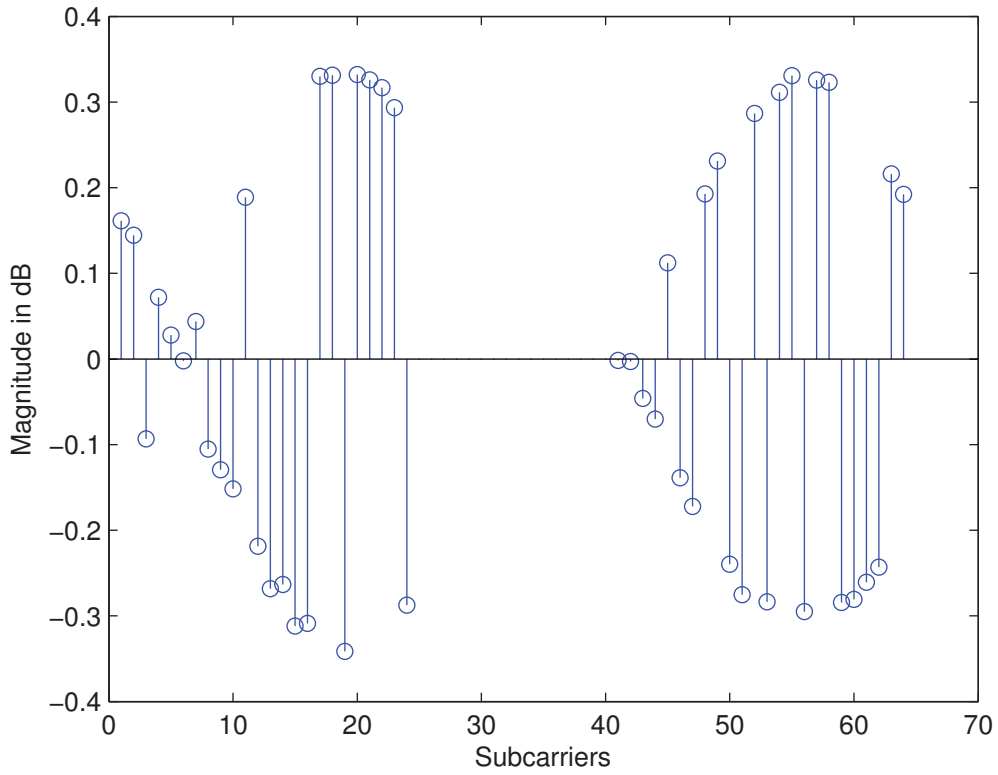

Figure 5.2: The frequency response of the channel with 64 subcarriers out of which 48 are occupied subcarriers and the remainder are unoccupied subcarriers. The response is obtained by estimating the channel with the preambles sent by APs.

at the same time, there will not be a collision of signals at the jth desired client 'I4'(1). Thus the jth network client 'I4'(1) is now able to detect the desired signal from the jth network AP1 and logs the results to the host PC. The host PC extracts the raw received signals with offline decoding using MATLAB® .

In order to satisfy the second requirement, we measure the feedback delay time (T_f) of our testbed environment (which is found to be 4.871 ms). Then, we compare it with the standard Coherence Time (CT), of 21.2 ms, measured by MacLeod et al. in [89] for ISM wireless indoor environments. This study is vital, as stale CSI degrades the performance in terms of interference management.

Comparison shows that T_f is about a fifth of the standard CT. This ensures that the channel changes more infrequently than the feedback delay time T_f of our testbed environment, meaning that the precoding vector is up to date with respect to changes in the channel conditions.

The comparison of T_f and the standard CT is validly made, as the measurement environment of the standard CT and our experimental environment are similar. Both are for the ISM band and are inside a laboratory with electronic equipment, desks, tiled floors, two walls at two sides and two glass walls.

5.3 Practical Issues

A few practical issues have to be taken care of to achieve the desired signal reception at the receiver. They are frequency offset, time synchronisation and packet detection, multipath and complexity.

5.3.1 Frequency Offset and Phase Tracking

Due to engineering constraints there are always small variations in the carrier frequency between the transmitter and the receiver which cause a phase rotation in the received signal. Before these variations accumulate with time and cause a decoding failure, CFO must be precisely estimated and compensated for at the receiver. The CFO of our testbed is estimated with the help of the packet preambles as described above. The host PC of the clients receives the 'cfo.dat' file which is analysed and compensated for[7] by MATLAB® during offline decoding.

[7]The frequency offset can cause a linear displacement in the phase of the received signal. Thus the frequency offset is compensated for by multiplying digital signal samples by $e^{j2\pi n\delta f_c}$, where δf_c is the frequency offset.

5.3.2 Time Synchronisation and Packet Detection

In order to prevent inter-symbol interference between the transmitted signals all the transmitters are synchronised with the CP of an OFDM symbol. We use an OFDM FFT size and CP size equal to 64 to deal with additional delays due to channel propagation and hardware turnaround time.

For packet detection, we set a predefined power threshold suitable for our platform and compare it with the computed total power of the preamble of the detected packet. We capture only packets with the preamble which have a total power larger than the preset power threshold. This is done to avoid the noise signals which may slip into the receiver during packet reception at clients. The noise level of the platform varies differently with the environment; we however, set the power threshold to be 0.3 dB.

5.3.3 Multipath

Signals travel through different paths before they are received at the receiver. We used an indoor environment in our testbed. Since OFDM is used, the effect of fading is largely minimised.

5.4 Performance Evaluation from the Testbed

We measure the ability of our testbed to address the HT problem as follows: we consecutively perform three things: First, measure the received SNR at the jth network client 'I4' under the HT scenario which is constructed by synchronising AP1 and AP2 to transmit at the same time. Second, apply our design, where both AP1 and AP2 are transmitting but the ith AP2 uses the precoding vector to cancel interference at the jth network client 'I4', and measure the received SNR at 'I4'. Third, measure the received SNR at 'I4', when the HT problem is eliminated by removing AP2 from the transmission range of

AP1. The received SNR at 'I4' in this setting gives the optimal measure to compare our design performance.

Finally, we analyse the raw received signals, compare the received SNR and the Effective SNR (ESNR)[8] with the aforementioned settings.

5.4.1 Analysis from the Raw Received Signal

After the HT condition is satisfied, both AP1 and AP2 start transmitting. We observe

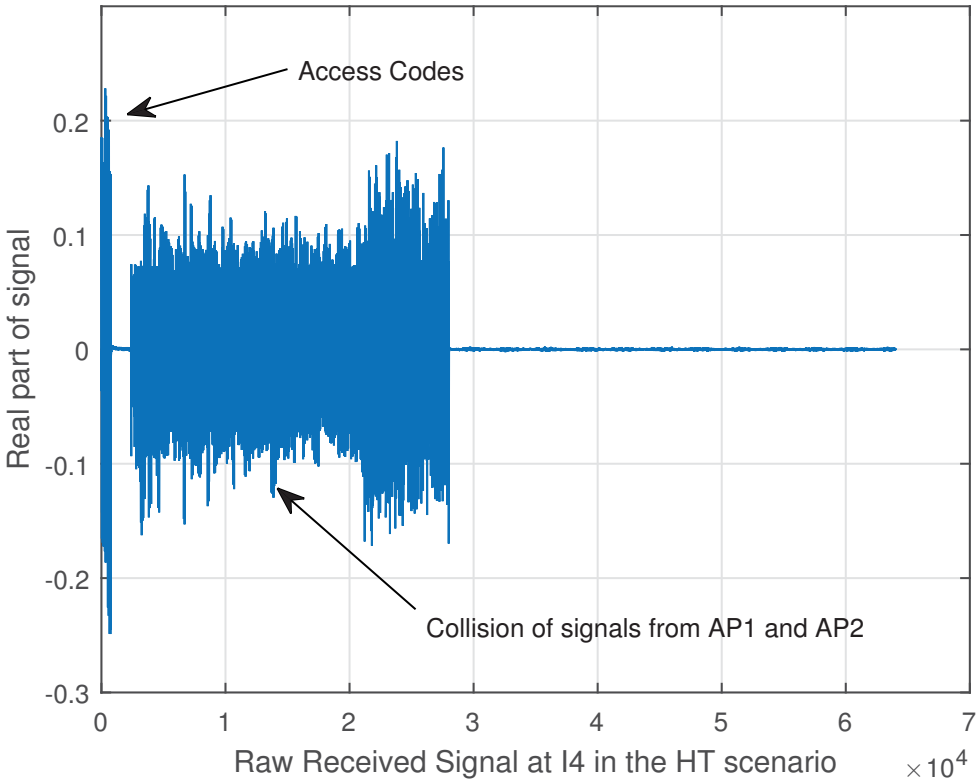

Figure 5.3: The decoded raw samples in a collided form at client 'I4' from AP1 and AP2 in the HT Scenario.

that the jth network client 'I4' is totally flooded with signals from both AP1 and AP2

[8]This is defined as the SNR that would give the same error performance on a narrowband channel [87].

(i.e., the signals interfere with each other) as shown in Fig. 5.3. The observed y-axis shows the real part of the signal. The signals lie within the range of -0.2 to +0.2 dB, however with some irregularities and spikes[9].

We calculate and use the precoding vector at the ith network AP2 to cancel the interference at the jth network undesired client 'I4'. This results in a raw received signal at the jth network client 'I4' from just its jth network AP1 as shown in Fig. 5.4. Also, we view the signal received from AP2. This is intentionally done to show what the AP2 signal would look like if it were not managed. It is the part of AP2's signal where we deliberately did not apply interference cancellation.

5.4.2 The Impact on the Received SNR

We compare the received SNRs at the jth network client 'I4': a) under the HT scenario, b) with our proposed solution and c) with the collision-free transmission. The SNR values for the 48 occupied subcarriers of the OFDM signals are plotted in Fig.5.5. We see a significant gain, in the received SNR (blue dots) of the jth network client 'I4', after applying our design, relative to the SNR under the HT scenario (green dots) . This improvement in SNR comes from the successive jth network AP1 transmission to its jth network client 'I4'. The gain in SNR is significant because, in the HT scenario, the signal transmission is marred by interference from the ith network AP2. However, after implementing our design the interference is mitigated and a significant SNR gain is obtained.

In addition to this, there is an average of about 4-5 dB difference in SNR between the received SNR of the collision-free transmission (red dots) and the received SNR with our

[9]The principle of superposition among the collided signals causes signals to add up constructively or destructively (noise is present as well) which may have caused the received signal to be irregular in pattern and with spikes.

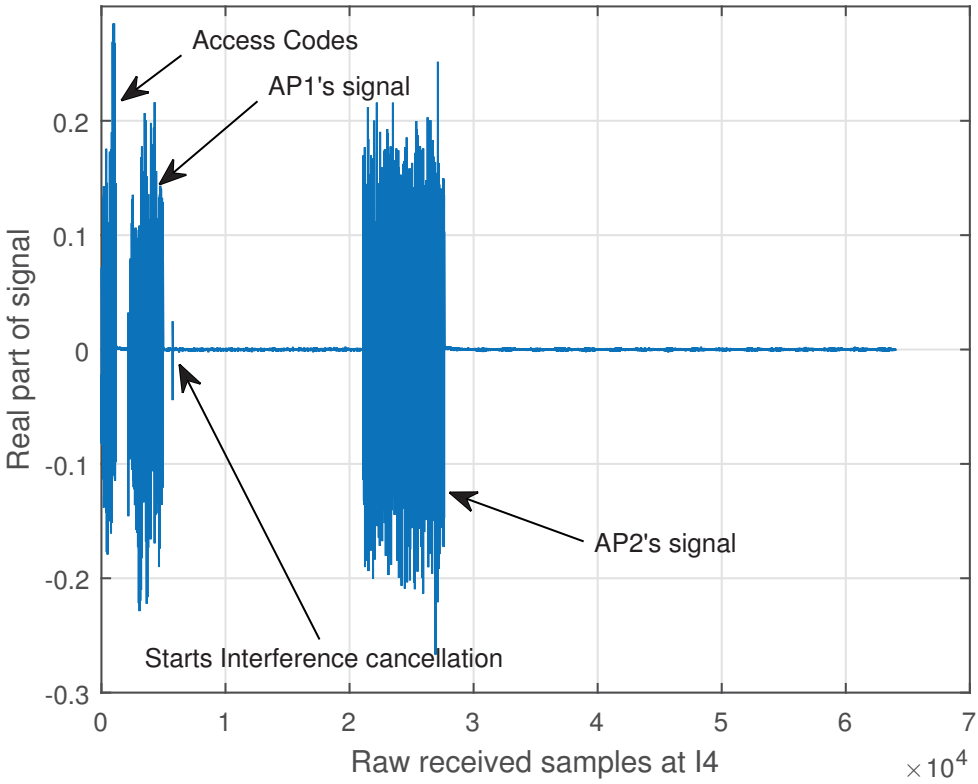

Figure 5.4: The raw received signals after applying our scheme in the HT scenario where AP2 cancels interference at I4, and as a result the AP1 signal is only seen at I4. However, some part of the AP2 signal is also present. This is deliberately done to show what the AP2 signal would look like when interference cancellation is not applied.

design (blue dots) in Fig. 5.5. The SNR gain in a collision-free transmission is at the upper bound that our design is supposed to achieve.

Despite imperfections in interference mitigation caused by hardware offsets and other implementation limitations, our design possesses an acceptable received SNR gain of about 6 dB on average, with the 48 occupied subcarriers. This gain is about 10 dB in comparison to transmission in the HT scenario.

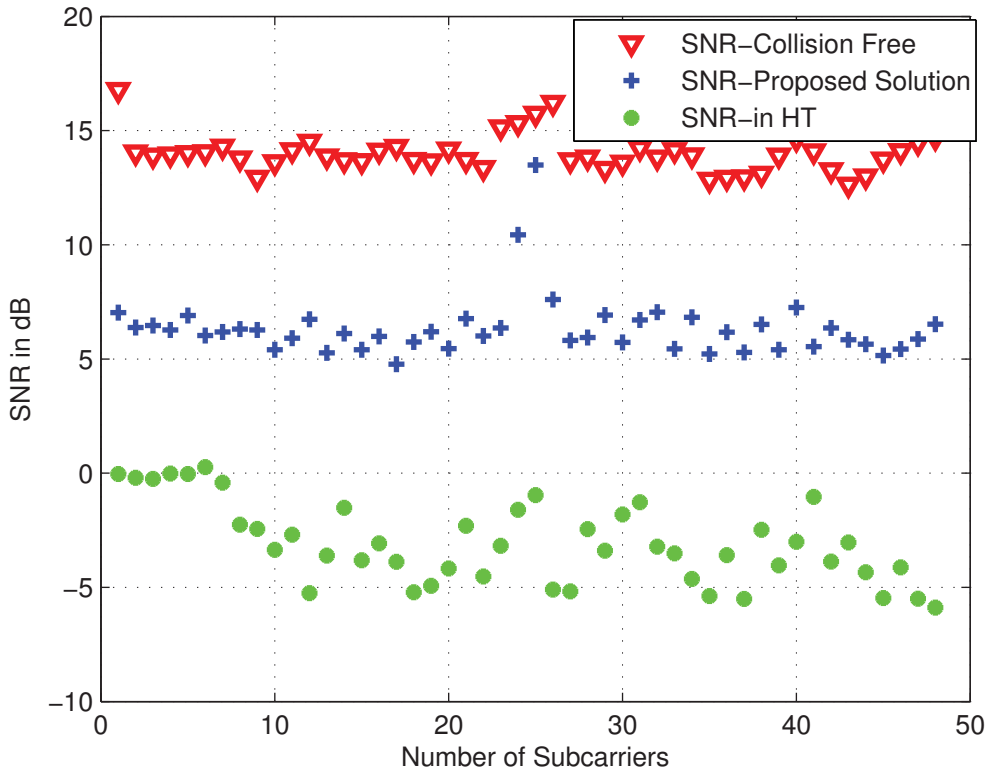

Figure 5.5: Comparison of the received SNR per subcarrier with the collision-free jth network AP1 and the jth network client I4 transmission, the proposed solution and the HT scenario. The received SNRs in collision-free transmissions, the proposed solution and the HT scenario are plotted separately, and present them in one combined figure. The gain in SNRs is associated with 48 occupied subcarriers.

5.4.3 Analysis of ESNR

For multicarrier systems like OFDM, subcarriers may undergo different levels of fading, and these channel qualities cannot simply be represented by the overall Received Signal Strength Indication (RSSI)[10]due to frequency selectivity [90]. In such a context, the ESNR can be used as an important metric for performance evaluation. We use the availability

[10]For a narrowband channel which shows an almost flat fading response to all the subcarriers, RSSI would be a better choice for measuring the success and failure of packet delivery.

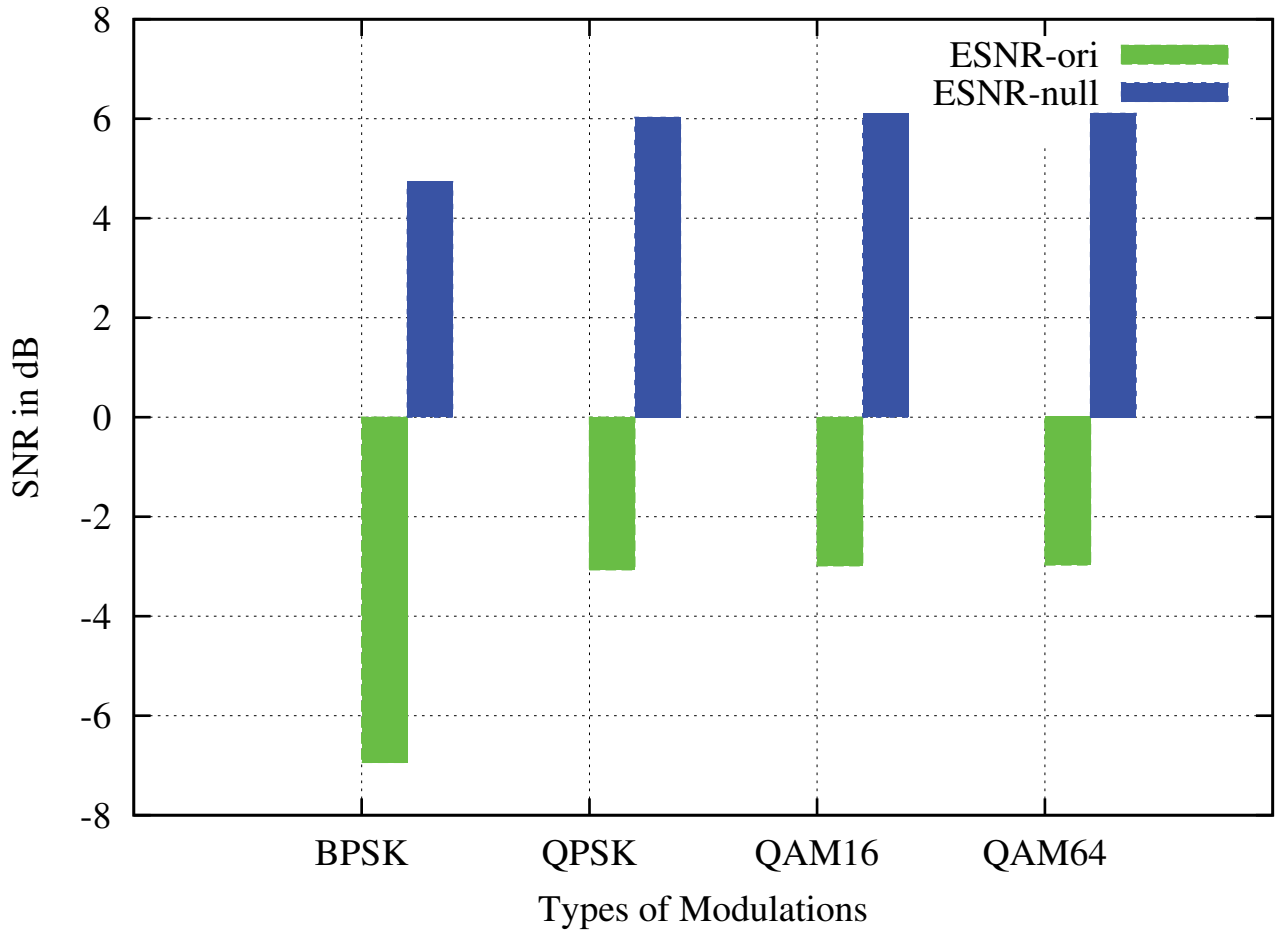

Figure 5.6: ESNR comparison of different modulation schemes in the HT Scenario. The figure shows an Effective SNR comparison: a) when the HT scenario is present (ESNR-ori), and b) ESNR after applying our scheme (ESNR-null). Different modulation schemes Binary Phase Shift Keying (BPSK), Quadrature Phase Shift Keying (QPSK), 16QAM and 64QAM are used for comparison.

of CSI at the subcarrier levels, as shown in Fig. 5.2, to measure the ESNR at the jth network client 'I4', by averaging the subcarriers' BER and then inverse mapping the Bit Error Rate (BER) to the SNR. From Fig. 5.6, we observe a rise in the ESNR value by about 10 dB for each modulation scheme after applying our design to address the HT problem.

5.5 Summary

In this chapter, we present two important facet of our research. First, a detailed description of our experimental setup to deal with the HT problem in laboratory settings. Second, obtained results from the platform. Additionally, the two fundamental prerequisites, CSI measurement at clients and timely CSI feedback to APs from clients, are highlighted and illustrated. Our USRP2 platform, in compliance with the prerequisites, is described along with the accessories such as RFX2400 daughterboards, Jackson Labs equipment, RF antennas, etc. The practical issues for the platform are presented. Finally, the results from the platform in terms of the received SNR are discussed, being obtained from three different settings in the laboratory: a) in the HT scenario, b) in the HT scenario while applying our design and c) without the HT scenario.

Chapter 6

Degrees-of-Freedom-Based Medium Access

6.1 Introduction

In Chapter 4, a PHY layer scheme, fairness and throughput-aware precoding based on the ZF precoding, is proposed to satisfy the first application-specific requirement of MU-MIMO WLANs while addressing the HT problem. The objective of PHY scheme is to allow multiple APs in MU-MIMO WLANs to transmit concurrently to the desired clients without interference to undesired clients in the HT scenario. Also, the PHY layer scheme should maintain fairness in clients' access and network throughput.

In this chapter, a MAC protocol based on the PHY scheme is developed. Fundamentally, our MAC protocol uses the fact that APs willing to transmit at any time instant should have required DoF for transmission in order to avoid collision of signals in the HT scenario[1]. Our DoF based MAC protocol is discussed in two sections. First, we present

[1]This may also be referred to as the second application-specific requirement of our design as described in Chapter 3.

our Wireless Medium Access procedure where channel acquisition method and DoF based TXOP for APs is meticulously detailed. The performance of channel acquisition method and the fairness algorithm for TXOP is presented. Second, we discuss the Medium Access Function where we adopt and expand the conventional PCF and use the CFP of PCF for the downlink and uplink data traffic. The content of this chapter constitutes a published work by the author in [54, 56, 57].

6.2 Wireless Medium Access Procedure

Unlike the traditional time-division-based medium-access mechanisms such as RTS/CTS, we clearly see the need for a new mechanism for the Wireless Medium Access (WMA) which would ensure that APs will have the required DoFs at the time when they transmit. In this section, we rigorously discuss a new mechanism to access the wireless medium which would not only check time but also required DoFs before making APs eligible for transmission.

Basically the challenge for such WMA mechanism is two-fold. First, in order to get a suitable precoding vector as described in Chapter 4, the CSI associated with both the desired and the undesired clients within the transmission range of the APs/transmitters have to be obtained, but the signalling/handshaking time overhead associated with this process must be as low as possible. The detailed mechanism for a modified seamless channel sounding process is given in Subsection 6.2.1.

Second, the mechanism for medium access for APs and clients is vital. Unlike traditional time-division-based decisions, APs/transmitters in our design decide TXOPs based on the available DoFs of the APs. Thus we need to design a mechanism that checks DoFs at APs and decides TXOPs accordingly. The provision for DoF-based TXOPs for APs is presented in Subsection 4.1.2, whereas the decision procedure is presented in Subsection

6.2.3.

Other important aspects are to move the complexity of signal processing to the APs and dynamically change the combination of concurrent clients in bursty traffic demand, while respecting fairness. We shall discuss these aspects further.

6.2.1 Acquiring CSI Associated with Transmitter and Clients

While IEEE802.11ac incorporates MU-MIMO and provides the basic framework for channel estimation with the null data packet in the standard (called channel sounding), it has noticeable signaling time overheads. We adopt and modify the channel sounding of IEEE802.11ac with a view to reducing the signalling time overheads and better utilising the precious air time of the APs. Keeping in mind that the channel sounding process should be as standard-compliant as possible, for being amenable to commercial adaptation, we reuse some data frames from the standard in our design. Additionally, IEEE 802.11ac leaves the determination of the MU-MIMO precoding scheme open and implementation-dependent. We take ZF beamforming as the principal transmission strategy and design the sounding process accordingly.

The channel sounding is initiated by those APs who have packets in the queue for transmission. Since APs need to find the channels associated with desired and undesired clients for Fairness and Throughput-aware precoding, they first transmit a broadcast frame called 'B_frame' so that the clients within the APs' transmission range can report their channels to the APs. The frame formats used in channel sounding are shown in Fig. 6.1.

Similar to the Null Data Packet (NDP) announcement frame of 802.11ac, APs broadcast a control frame, B_frame, of 25 bytes. This frame contains a 6-byte field for the transmitter address (i.e., AP) and separate address fields for a set of multiple station information records used to request multi-user feedback. Most importantly, APs assign Association

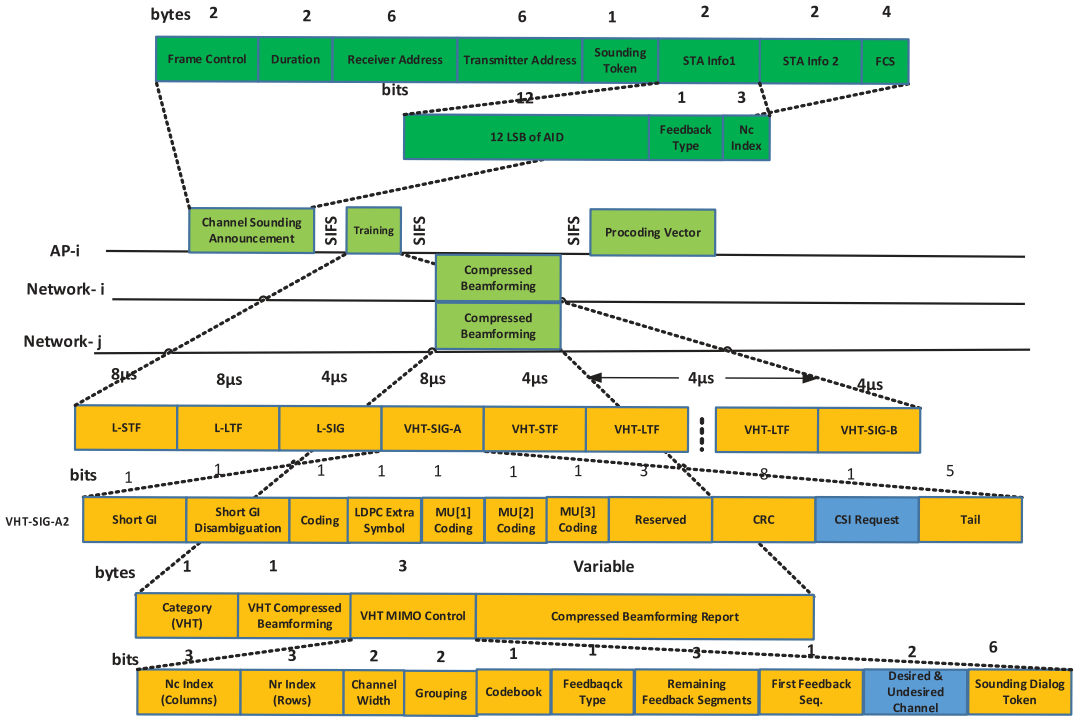

Figure 6.1: Frame formats and basic Channel Sounding process. The Channel Sounding process and the frame formats used for channel sounding. We present the Channel Sounding process considering the ith network AP2 and clients from both the ith and the jth networks.

Identifications (AIDs) to the clients upon association which are included inside the 12 Least Significant Bits (LSB) of Station (STA) information.

The second step is to send the training symbols by APs for channel measurement. This is done by the training frame, T_frame, without a data field. Unlike NDP frames in 802.11ac, the T_frame in our design consists of 1-bit CSI request field at Very Hight Throughput-Signalling-A2 (VHT-SIG-A2) in the VHT-SIG-A field. The tail consists of only 5 bits compared to 6 bits in 802.11ac. This change in tail bits is to make 24 bits correspond to one OFDM symbol.

The function of the 1-bit CSI request field is to ask both desired and undesired clients their CSI within the transmission range of APs. Unlike IEEE 802.11ac, it is worthwhile

to note that there are no separate polling frames to acquire CSI from the undesired client. All the clients (desired and undesired) respond with their associated CSI after the CSI request field is received as a part of the training frame.

The removal of the separate polling frame in our design in fact decreases the signalling time overheads in the channel sounding process. Specifically, for the two-network scenario we have considered in previous examples, it removes 2 Short Inter-frame Spacing (SIFS) times and the time for 1 Beamforming Poll frame which equals 98.67 μs. The detailed calculation for signalling time overhead is presented in Subsection 6.2.2.

Each client analyses the training symbols in the Physical Layer Convergence Protocol (PLCP) header (of the T_frame) and measures the channel between the APs and themselves. It is obvious that the clients within the overlapping region would hear multiple B_frame broadcasts. Each client responds to the channel request on a FIFO basis in the uplink.

After the reception of the T_frame, the third step is to feedback the measured channels. Owing to the limited feedback channel, the channels (in the form of the matrices) are compressed and sent in the form of Very Hight Throughput (VHT) Compressed Beamforming frames[2].

Since the APs need to differentiate channels associated with the desired and the undesired clients, our design uses 2 bits from the reserved bits of the Compressed Beamforming Action Frame of 802.11ac. The 2 bits can have at most 4 logical combinations, which are used to distinguish between the desired and the undesired clients. The reserved field bits are set to 00 or 11 in the Compressed Beamforming Action Frame by the clients who have AID (i.e., desired clients). Otherwise the bits are set to 01 or 10 by the clients who do not have AID (i.e., undesired clients). Hence, upon the reception of the Compressed Beam-

[2]A more rigorous discussion on the limited feedback channel, leading to the adoption of the FRF model and its implications on our proposed design, is presented in Chapters 7 and 8 respectively.

forming Action Frame, APs check the reserved field and distinguish the CSI feedback between the desired and the undesired clients, $\mathbf{H}_{ii_Q}$ and $\mathbf{H}_{ii_P}$.

In terms of power, regulation can limit the transmit power based on the number of antennas used at the transmitter so that transmit beamforming does not increase the maximum distance range.

Fig. 6.1 represents a basic diagram of the frame formats and the channel measurement process for the APs in a typical ith network, AP2 and two clients, each from the ith and jth networks.

6.2.2 Performance Evaluation

We compare the signalling time overhead of our WMA with the traditional RTS/CTS mechanism in the context of the HT problem in WLANs. Also, we compare the signalling time overhead of our WMA and the IEEE802.11ac standard. We further compare the total network throughput gain of our WMA and the traditional RTS/CTS. The throughput of our WMA and IEEE802.11ac is also compared. As shown in Fig. 6.1, the payload is not transmitted until we access all channels and calculate the precoding vectors for the APs. This period is defined as the signalling period.

In a typical HT scenario for two networks, we calculate the signalling time overheads associated with different frames as shown in Fig. 6.1. Since a PLCP preamble and a PLCP header are added to an MAC Protocol Data Unit (MPDU) to create a PLCP Protocol Data unit (PPDU), the transmission duration of the broadcast B_frame is given by $T_{B\_frame} = PLCP\ frame + \left[\frac{25}{BpS(m)}\right] \times tSymbol = 40 + \left[\frac{25}{3}\right] \times 4\ \mu s = 73.33\ \mu s$. We use the same PLCP frame as IEEE802.11ac, because it is designed to be compatible with legacy standards. $tSymbol = 4\ \mu s$ is the OFDM symbol interval. The type of modulation used is BPSK with data rate 6 Mbps and code rate $\frac{1}{2}$. The training frame has the same format as the VHT PPDU except for the data field, so the transmission

duration is $T_{T\_frame} = (8 \times 5) = 40\ \mu s$.

Unlike IEEE802.11ac, there is no provision for the Beamforming Report Poll in our design, thus there is no signalling time overhead associated with it. In the IEEE802.11ac standard, the Beamforming Report Poll frame constitutes a signalling time overhead of $T_{BR_{Poll}} = 40 + \left[\frac{20}{3}\right] \times 4\ \mu s = 66.67\ \mu s$, considering BPSK modulation with data rate 6 Mbps and code rate $\frac{1}{2}$.

We further calculate the time duration of VHT Compressed beamformimg assuming payload length $l = 200$ octets, $T_{CB_{\text{Report}}=}40 + \left[\frac{5+l}{3}\right] \times 4\ \mu s = 313.33\ \mu s$.

Thus, for a typical two network, $K = 2$, the signalling time overhead is given by $T_{OH} = T_{B_{frame}} + T_{T\_frame} + 2 \times T_{CB_{\text{Report}}} + 3 \times SIFS = 787.99\ \mu s$, where $SIFS = 16\ \mu s$. The traditional signaling time overhead for the RTS/CTS scheme is given by $T_{RTS/CTS} = T_{DIFS} + T_{RTS} + T_{CTS} + 2 \times SIFS = (34 + 50.33 + 42.33 + 2 \times 16) = 158.7\ \mu s$ whereas the IEEE802.11ac signalling time overhead is given by $T_{OH_{ac}} = T_{NDP_{ann}} + T_{NDP} + 2 \times T_{CB_{\text{Report}}} + T_{BR_{Poll}} + 5 \times SIFS = 886.66\ \mu s$.

Based on the calculations of the signalling time overheads, we study their impact on the network throughput. The simulations are carried out for a 6-antenna AP2 consisting of 5 desired clients, 'I1', 'LP','I2','HDTV', and 'I3', and 2 undesired clients, 'I4' and 'I5', in the transmission range. There are 4 clients to be served concurrently. Typically, $N = 6$ and $I = 2$ (from the same jth network), so the number of concurrent transmissions after TXOP is $N - I = 4$.

We first take an arbitrary air time $t = 20\ ms$ and compare the throughput gain for our WMA with the traditional RTS/CTS and MAC of IEEE802.11ac at SNR = 5, 15 and 25 dB respectively.

Our simulation results in Fig. 6.2 reveal that the RTS/CTS scheme has an early gain in capacity at around 157.8 μs and IEEE802.11ac MAC has a gain at around 886.66 μs, whereas our WMA protocol gain in capacity is around 787.99 μs. This is an expected

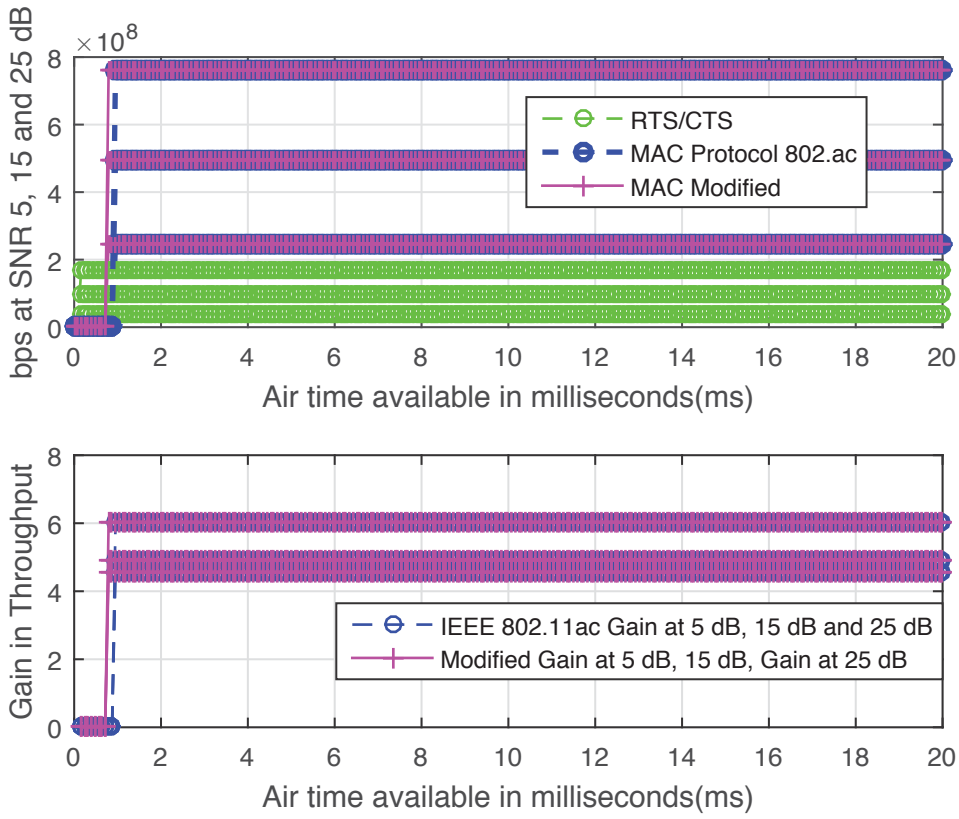

Figure 6.2: Capacity comparison between RTS/CTS, our WMA and IEEE802.11ac MAC: air-time 20 ms.

behaviour, as our WMA has a signaling time overhead that is higher than RTS/CTS and lower than IEEE802.11ac MAC. It is interesting to observe that our WMA protocol and IEEE802.11ac have about 4-5 times capacity gain compared to the RTS/CTS scheme, however our WMA has the advantage of obtaining the gain as early as 98.67 μs compared to IEEE802.11ac.

Normally, a lower slignaling time overhead is desired so that the available transmission time can be better utilized for packet transmissions. Owing to the design constraints, we cannot lower the signalling time overhead below the traditional RTS/CTS, however, we can reduce the signalling time overhead by 98.67 μs compared to IEEE802.11ac. Additionally, a constant network capacity gain of 4-5 times can be achieved compared to

RTS/CTS.

The gain in capacity comes from the concurrent transmissions (made possible by the precoding vector) that take place once the handshaking process is completed. Thus the slightly longer signalling time in our WMA protocol is compensated for by the throughput gain contributed by concurrent transmissions. In summary, the large signalling overhead of IEEE802.11ac is reduced by 98.67 μs and a constant capacity gain of 4-5 times, compared to RTS/CTS, is achieved by our WMA protocol.

Furthermore, we have a closer look with a short available air time $t = 2\ ms$ with all other factors remaining the same. Fig. 6.3 shows that we have the same throughput gain

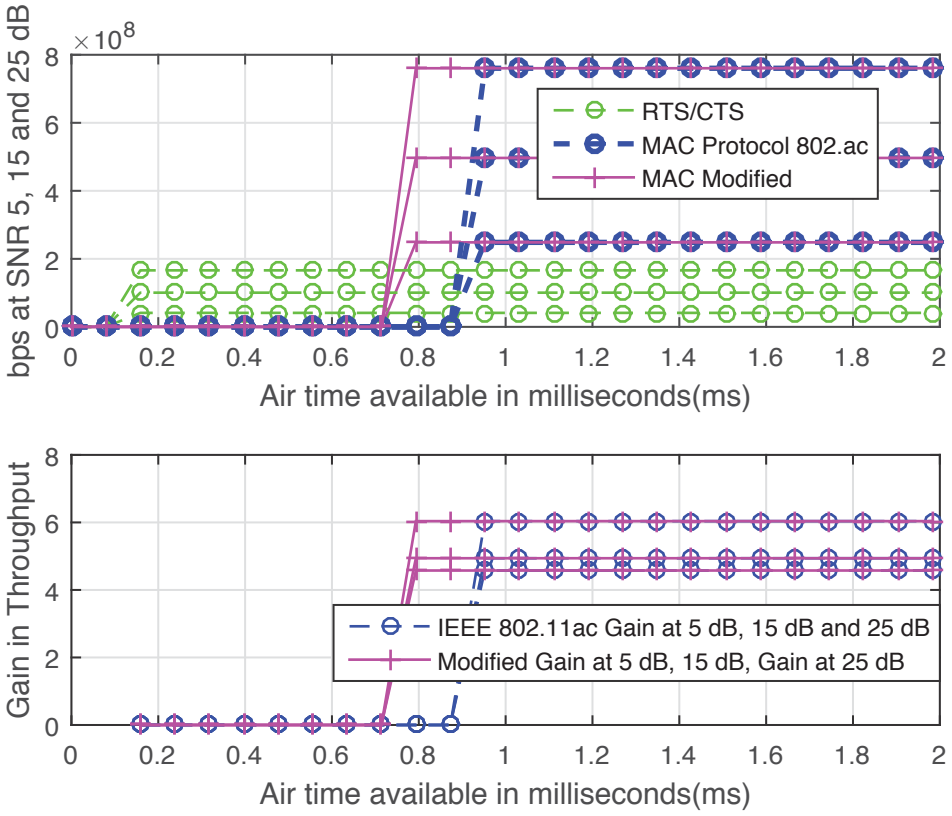

Figure 6.3: Capacity comparison between RTS/CTS and our WMA and IEEE802.11ac MAC, air-time 2 ms.

of around 4-5 times compared to RTS/CTS, and this gain is as early as 98.67 μs before

IEEE802.11ac, as in $t = 20\ ms$. Thus, from Fig. 6.2 and Fig. 6.3 it is clear that the gain holds for the available air time satisfying $t > 787.99\ \mu s$. Thus, as long as the available air time is $t > 787.99\ \mu s$, a constant gain in the range of 4-5 times with respect to the traditional RTS/CTS and a signalling overhead improvement of 98.67 μs compared to IEEE802.11ac is achieved by our WMA at SNR = 5, 15 and 25 dB.

6.2.3 Transmission Opportunity for APs with Heterogeneous Antennas

We present a simple DoF-based approach to decide the TXOPs among APs who compete for the medium at any instant. Thus APs calculate the precoding vector only when they have sufficient DoF for transmission[3], otherwise, the APs end up causing interference to other clients.

Before competing for TXOP, APs perform Channel Sounding and acquire CSI from Q desired (i.e., $\mathbf{h}_{ii_d}$) and P undesired (i.e., $\mathbf{h}_{ij_u}$) clients in their transmission range, where $\underset{i_d \neq j_u}{i_d} \in \{1.....QM\}$ and $\underset{j_u \neq i_d}{j_u} \in \{1.....PM\}$ as described in Subsection 4.1.1. Second, APs make a decision for TXOP (i.e., choose to remain in the 'Active' or the 'Silent' Mode), if $(i > j_u)$, for example for AP2, where i is the number of antennas at AP2, $i \in \{1.....N\}$, and j_u is the number of antennas at undesired clients within range. Note that our design checks $i > j_u$ rather than $i \geqslant j_u$ for TXOP. The aim is to ensure that every AP has at least one DoF after winning TXOP, because interference cancellation costs exactly as many DoFs as the number of antenna(s) [8].

Suppose that we have 3 APs in a network having N_1, N_2 and N_3 antennas respectively. Each of these APs has I_1, I_2, I_3 antennas for undesired clients in the overlapping region.

[3]For instance, in a network of 3 ongoing concurrent transmissions, a two-antenna AP does not qualify for TXOP, however APs with a greater number of antennas than 3 can obtain the precoding vector and can transmit concurrently.

The 3 APs satisfying $N_1 > I_2 + I_3$, $N_2 > I_1 + I_3$ and $N_3 > I_1 + I_2$ will have TXOP and will remain in the 'Active' mode, otherwise the APs will not have TXOPs and will remain in the 'Silent' mode at that instant. Hence the TXOP among the APs are decided on the available DoF.

6.2.4 Fairness among APs with Heterogeneous Antennas

Since TXOP for APs is based on the DoF of APs, it is possible for APs with a large number of antennas (i.e., consisting of higher DoFs) would win TXOP all the time. This may result in a deep unfairness for APs having fewer antennas. In order to prevent APs with fewer antennas starving for TXOP, our design executes a simple Fairness Algorithm for TXOPs, which runs at APs. Since MU-MIMO WLANs are distributed in nature, the Fairness Algorithm for TXOPs is designed to ensure fairness among APs without requiring coordination with each other.

Specifically, we assign all APs two types of credit counters ('S_Counter' and 'F_Counter', initialised to 0) and a credit threshold ('C_threshold', set to a constant value). Each time APs get the TXOP, the 'S_counter' will be incremented, otherwise the 'F_Counter' will be increased. If the 'F_Counter' crosses the 'C_threshold', the corresponding AP directly qualifies for TXOP. The basic pseudocode is given in **Algorithm 2**.

6.2.5 Fairness Index of the Algorithm for TXOPs

We evaluate the Fairness Index of our **Algorithm 2**, Fairness Algorithm for TXOPs, according to the Jain Fairness Index [91], which fundamentally gives a quantitative measure to any resource sharing or allocation problem. We consider 3 APs with 2, 3 and 4 antennas, for which the fairness in throughput is studied. We assume that each AP, after winning TXOP, has at least one stream for transmission and has at most a number of transmission streams equal to the number of antennas at the AP. For instance, a

Algorithm 2 Fairness Algorithm for TXOPs

procedure FAIRNESS

initialization $S\_counter = 0$; $F\_counter = 0$; $C\_threshold$

if $i > j_u$ **then**

$S\_counter \leftarrow S\_counter + 1$.

if $S\_counter == 2 * C\_threshold$ **then**

Reset $S\_counter$ to 0

end if

else

$F\_counter \leftarrow F\_counter + 1$.

if $F\_counter == 2 * C\_threshold$ **then**

Reset $F\_counter$ to 0

end if

end if

if $S\_counter \leqslant C\_threshold$ **then**

Wins TXOP 'Active' Mode

else

'Silent' Mode

end if

if $F\_counter \leqslant C\_threshold$ **then**

'Silent' Mode

else

'Active' Mode

end if

end procedure

3-antenna AP has at least 1 transmission stream (worst case) and at most can have 3 transmission streams (best case) after TXOP. Let R_i be the ith AP that has TXOP and n be the number of such APs; then the Jain Fairness Index is given by

$$f(R) = \frac{\left|\sum_{i=1}^{n} R_i\right|^2}{\sum_{i=1}^{n} R_i^2}, \; R_i \geq 0. \tag{6.1}$$

R_i is calculated as the ratio of the maximum throughput and the least expected throughput so that for the AP with 3 antennas,

$$R_i = \frac{Throughput\ when\ three\ streams\ are\ used}{Throughput\ when\ single\ stream\ is\ used}. \tag{6.2}$$

Similarly, R_i can be calculated for the remaining APs. The Fairness Index for $n = 3$ (in our case), is calculated and is discussed further. We discuss the Fairness Index of **Algorithm 2**, Fairness Algorithm for TXOPs, considering three APs, i.e., $n = 3$, with the number of antennas 2, 3 and 4 respectively.

Fig. 6.4 shows the system throughput with the credit counters where 'C_threshold' is arbitrarily taken to be 6. Also, we show the maximum and minimum stream-based throughput of APs when the credit counter gradually increases to 'C_threshold'. We also calculate the Jain Fairness Index according to (6.1) for three APs and present the Jain Fairness Index with the credit counters. It is shown that our fairness algorithm has a Fairness Index greater than 0.9. Thus the use of our fairness algorithm provides a more than 90% fair share among three APs in terms of throughput.

6.3 Wireless Medium-access Function

Since our design permits concurrent transmissions between the APs and the clients after the APs win the TXOP, we need to develop a mechanism to manage these concurrent transmissions. Notably, concurrent transmissions take place after the determination of

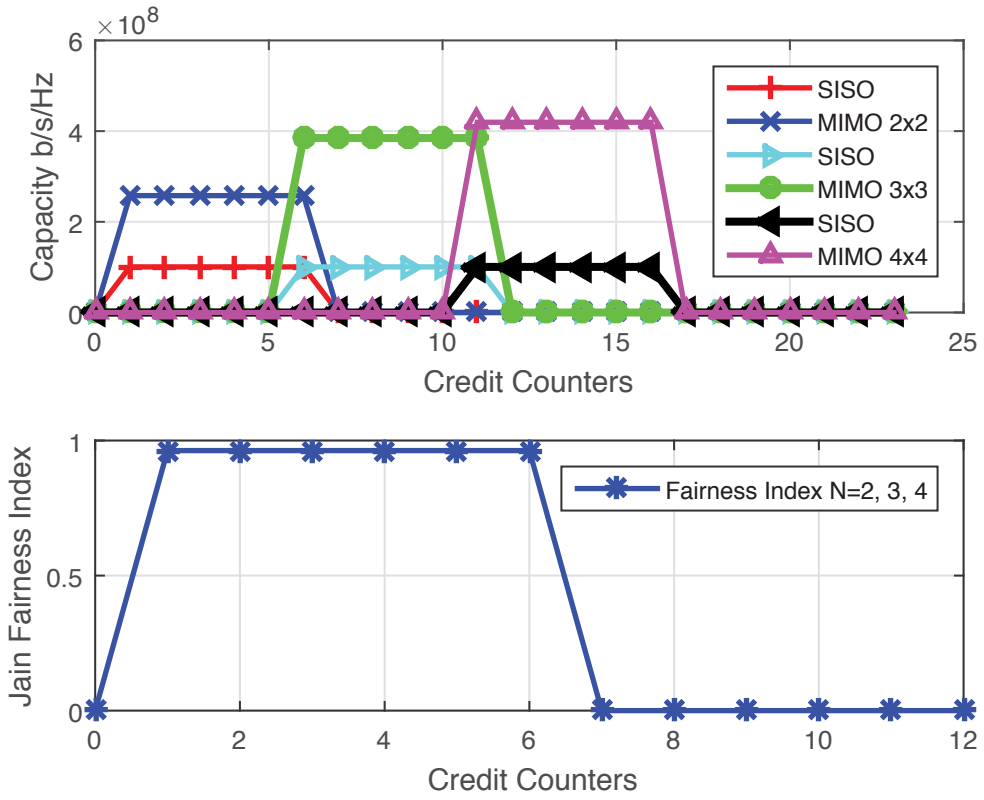

Figure 6.4: Throughput of three APs with the Fairness Index. Basically we present a Jain Fairness Index considering 3 APs and clients, i.e., 4×4, 3×3 and 2×2 systems.

TXOP and there is no contention among the APs during concurrent transmissions. We exploit this fundamental attribute of the concurrent transmission and adopt and expand the IEEE802.11 PCF mode to address the concurrent transmissions. The reason is that concurrent transmissions are contention free and they can be governed by the CFP of PCF (the detailed process is discussed below). Besides, the PCF is the part of the IEEE802.11 standard and its expansion to suit our PHY layer solution would be a more straight forward way to satisfy the interoperability issues.

6.3.1 Contention Free Period and Contention Period

Similar to the PCF in IEEE802.11, we designate each AP in the network as a point coordinator. We divide time into CFP and Contention Period (CP) as shown in Fig. 6.5.

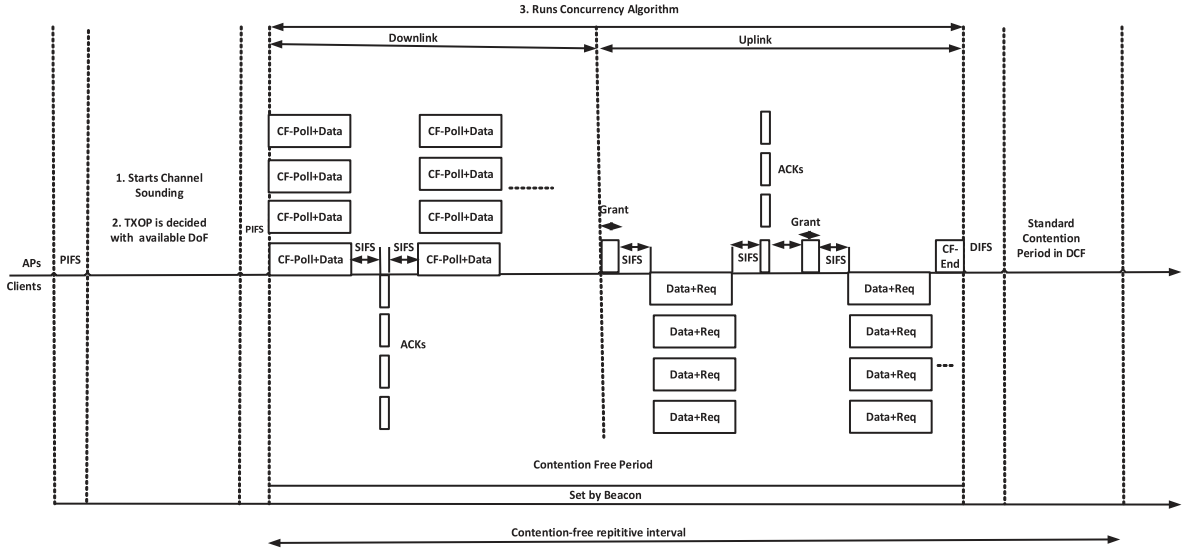

Figure 6.5: We extend the traditional PCF for concurrent transmissions. Since concurrent transmissions are contention free, the CFP of PCF can be used for this purpose. The figure shows 4 concurrent transmissions as an example.

The beginning of the CFP is marked by the transmission of a beacon by APs. This sets the duration of the current CFP. During the CFP, APs run the concurrency algorithm as described in Section 4.2 and select the clients for transmissions both at downlink and uplink. The CFP is followed by the CP, during which any clients can contend for the medium using IEEE802.11 and point-to-point MIMO. The objective of the CFP is to manage the concurrent transmissions as much as possible so that the network throughput can be increased. The duration of the CFP depends on the traffic congestion and may increase or decrease with the rise and fall of the traffic congestion. However, at the CFP, the APs serve at least one packet (on downlink and uplink) to all clients that have pending traffic. In contrast, the duration of the CP is constant. Similarly to IEEE802.11,

we set the minimum length of CP to be equivalent to the time required to transmit and acknowledge one maximum-size frame. However, sometimes it is possible for the contention-based service to run past the expected beginning of the CFP, which we call 'foreshortening' of CFP.

6.3.2 Medium Access to Downlink

Fig. 6.5 presents a series of events that takes place in this process. APs run the concurrency algorithm and select the group of clients for transmission at that instant. After that, APs transmit their downlink packets and poll the clients for uplink data with the help of the CF-Poll+Data frame. This process is similar to the existing PCF except that the current PCF polls the individual clients one by one whereas, with our MAC, APs poll for uplink traffic a group of clients selected by the concurrency algorithm.

The CF-Poll+Data frame contains two parts. The first part of the frame is shown in Fig. 6.6, and contains IDs of the clients that are selected as the result of the concurrency algorithm. The IDs are given to the clients upon association to assist with the control and management function. It also contains the frame id, Fid, the address of APs and the checksum of their broadcast. The APs and the clients use this checksum to test whether or not the received data is correct. The second part of the frame is the list of the concurrent downlink data of the APs to the clients.

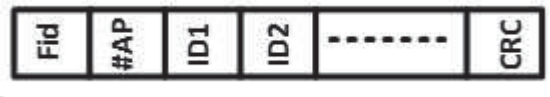

Figure 6.6: Metadata structure.

For instance, a downlink transmission of an AP with 4 antennas is given in Fig. 6.5. Upon reception of data at clients in downlink, the clients send Acknowledgments (ACKs) to AP. The order in which they send these ACKs is the same order as the IDs in the

Data+Poll frame. Basically, clients send ACKs using the traditional MIMO. Thus each received data at clients is acknowledged.

6.3.3 Medium Access to Uplink

The uplink data transmission at the CFP is initiated by the Grant frame[4] broadcast by AP. The Grant frame consists of the IDs of the clients determined by the concurrency algorithm that runs at APs. The clients in the uplink transmit the Data+Req frame. This frame contains the uplink data and, if there are more data in the clients to send, it also contains a request frame. The request frame is for transmission of the new frame in the uplink. However, APs do not entertain the new requests immediately. Instead, APs record the IDs of the Clients who have sent requests along with their uplink data, and then polls them in the next cycle. Upon reception of the poll request, Clients with the new data can transmit to APs. APs decode the Data+Req frame sent by Clients by using the standard MIMO decoding method.

Upon reception of the Data+Req frame, APs confirm the received data by the broadcast of the ACK frames within their transmission range. The clients in the transmission range receive the ACK frames. However, the clients who are not selected for the uplink transmission discard the ACKs. Thus clients are acknowledged for successful reception of data in the uplink transmission.

The end of the CFP is marked by the CF-End frame broadcast by APs. This frame indicates to the clients the end of the CFP. This frame prepares APs and clients to go back to the standard contention mode in DCF.

[4]The Grant frame is similar to the traditional CF-Poll, with frame type value '10' and subtype value '0110'. The difference between the traditional CF-Poll and the Grant frame is that APs send a CF-Poll Frame on a Client-by-Client basis whereas a Grant frame in our scheme is sent to a group of Clients selected by the concurrency algorithm.

6.3.4 When to Initiate the Extended PCF?

Traditionally the HTs are managed by the RTS/CTS mechanism. Basically, the RTS function determines whether the APs should use the CSMA/CA or the RTS/CTS mechanism to send the payload. Usually, the RTS threshold value is set in the range of 0-2347, and any data bytes beyond the threshold exchange data by the RTS/CTS mechanism.

In contrast to this traditional scheme, we set the retransmission threshold at the APs as *Re_Threshold*. Generally, when the transmitted data are not acknowledged at the APs, either due to data or ACK loss, APs retransmit the data and retry up to a certain count, *Re_count*. We compare *Re_count* and *Re_Threshold* and, if ($Re\_count > Re\_Threshold$), the extended PCF is initiated. The provision of *Re_Threshold* is similar to the traditional RTS threshold value. However, we choose a retransmission threshold, *Re_Threshold*, for initiation of the extended PCF. This is because the presence of the HT scenario causes collisions of signals and there is a high possibility of occurrence of retransmissions. In such a context, our proposed extended PCF comes into play to address the HT scenario.

6.4 Some Worthwhile Points

In this section, we discuss some worthwhile points in relation to our proposed MAC protocol design in the following subsections.

6.4.1 Are Active and Silent Modes Fixed?

The active and silent modes for APs are not fixed. Since the number of antennas that APs possess is always fixed, so N is fixed. However PM is not fixed; it is variable and depends on two components: First, the number of Undesired Clients, P, present in any AP's transmission range. Second, the number of heterogeneous antennas, M, that each Undesired Client possesses at the given instant where there exists the HT scenario.

Since PM vary from Case to Case, this will further vary two components in our scheme. First the remaining DoFs at APs are variable, because DoF=N-PM (this determines the number of concurrent transmissions in the scheme). Second, the TXOP of APs varies because TXOP = $(N > PM)$. Thus with variable PM, at one instant APs can satisfy $N > PM$ and will have TXOP, while at the another APs may not satisfy $N > PM$ and lose TXOP. APs with TXOP will remain in an Active mode, otherwise they will remain in a Silent mode. Additionally, to ensure fairness for the APs who may remain in the Silent mode for a long time and starve for TXOP, we have proposed Algorithm 2, a Fairness Algorithm for TXOPs.

6.4.2 How does the Concurrent Transmission Algorithm go out of Active and Silent Modes?

The CPF MaxDuration and Timestamp field of the Beacon frame are used to manage the transition of the Active and the Silent Mode APs. The algorithm presented with the flow diagram runs on top of the standard DCF, which is invoked when we have an HT scenario in MU-MIMO WLANs. Thus, both the Active and Silent Mode APs at the end of the algorithm return to contention-based DCF. The process works as follows: After TXOP is decided, both the Active and Silent mode APs set CFP to the maximum duration. Similar to traditional PCF, this is done by setting bit 15 to 1 out of 2 bytes in the CFP MaxDuration field in the CF parameter set [92]. The Active mode APs inform this decision to their Clients by sending a Beacon frame. Upon reception of the Beacon frame, Clients set their Network Allocation Vector (NAV) to 32768. On the other hand, Silent mode APs do not transmit any Beacon frame, however they constantly monitor the Timestamp and wake up at the elapse of CFP. Since the duration set for both the Active and the Silent mode APs is the same, the times when the Silent mode APs wake up, and the transmission of CF-End, marking the end of CFP by the Active mode APs, coincide.

6.4.3 If There is No Contention Among the APs, How will They Synchronise?

We see the need of synchronisation in our scheme in two stages. Stage 1 synchronisation is needed to invoke an extended PCF when all APs in the vicinity start experieo ncing the HT problem. However the challenge comes from the decentralised and distributed nature of MU-MIMO WLANs, where different APs do not communicate with each other. In such a context, we need a synchronisation mechanism which can synchronise APs independently before we initiate an extended PCF. We describe this synchronisation mechanism in Stage 1 which is achieved as follows: We use *Re_Threshold* and the Point Coordinated Function Infer-frame Spacing (PIFS) interval to synchronise APs before we initiate an extended PCF. Since there exists the HT scenario among APs, there will be a collision of data packets and ACKs are not received at APs. This makes APs retry for transmission of data packets. Our design exploits this well-known process of APs in WLANs and sets the same retransmission threshold, *Re_Threshold*, for all APs. Thus, whenever there is an HT problem in the network, APs in the vicinity reach their *Re_Threshold* more or less at the same time. Now this is the time for APs to initiate an extended PCF.

Additionally, to better synchronise the initiation of the extended PCF, we make APs wait longer, for a specified PIFS interval. This process will give extra time to any APs who lag to reach *Re_Threshold*. Thus all APs will be synchronised and will initiate an extended PCF at the same time.

Stage 2 Synchronisation is needed during the CFP operation. The Synchronisation is obtained via different types of frames used in CFP, such as Management Frames, Control Frames and Data Frames.

For example Ack in our scheme are synched as follows: Acks are sent after an SIFS interval whenever APs or Clients receive data packets in the uplink or downlink. Since

the concurrent transmission algorithm is used to select a group of Clients who are to be served at an instant, the transmission of Ack after packet reception can be done by using the traditional MIMO technique, both in uplink and downlink.

6.5 Summary

In this chapter, we basically focus on developing the design for satisfying the second application-specific requirement while addressing the HT problem in MU-MIMO WLANs. We fundamentally highlight that both time and DoFs are vital in our design for medium access in order to remove interference to undesired clients. The design challenges of WMA mechanism, low signalling time overhead, DoF-based TXOPs and the concurrent transmission-based access mechanism, are discussed and analysed thoroughly. As a result, we finally come up with: a) a new channel sounding process with reduced signalling time overhead, b) DoF-based TXOPs, c) a fairness algorithm with a fairness index greater than 0.9, and d) an extended PCF, giving a detailed description of uplink and downlink on an event-by-event basis.

Chapter 7

The ZF Technique with Finite Rate Feedback in MU-MIMO WLANs

7.1 Introduction

In the previous chapters, we discuss and illustrate the effectiveness of the ZF technique to address the HT problem in MU-MIMO WLANs. Principally, we addressed the two specific design requirements: a) fairness and network throughput, and b) TXOP decision of APs, while efficiently addressing the HT problem in MU-MIMO WLANs. In other words, we proposed the fairness and network throughput-aware ZF precoding at the PHY layer and designed a MAC protocol based on the PHY layer. In this chapter, we consider a practical scenario in a MU-MIMO WLAN settings and identify the limitations and the implications, when we attempt to satisfy the fundamental prerequisite of the ZF technique. It is imperative to note that: a) the ZF technique requires CSI at APs from clients, and b) the effectiveness of the technique to address the HT problem basically depends on the quality of the channels at the APs. However, owing to the constraints in MU-MIMO WLAN settings, such as channel estimation errors at receivers/clients, limited

bandwidth of the feedback path, delay in feedback path[1], cost of the feedback overheads, etc., it is not practically possible to obtain the perfect CSI at APs. This condition may pose a serious question on our design approach, using the ZF technique to address the HT problem in MU-MIMO WLANs. We see that one of the possible ways out may be to work with a partial CSI at APs in terms of quantised bits B, acknowledging that this does not represent the actual channel between APs and clients, and measuring the effect of imperfect CSI/quantised CSI on the system parameters.

It is worthwhile to note that our proposed design in the previous chapters inarguably supports the notion of partial CSI at APs in terms of quantised bits B. Specicially, during the channel souding process at MAC, the desired and undesired quantised channels (due to the consideration of the practical diemension in this chapter) can be fedback to APs by the VHT MIMO control field inside the compressed beamforming frame as shown in Fig. 6.1. The concurrent transmission algorithm then selects the quantisation indices of the desired channels before the calculation of the optimal ZF precoding vector.

Related work on cellular systems supports the study of imperfect CSI as being a more practical approach to the constraints in hand. We have been partly motivated by the work on imperfect CSIs in cellular systems.

Having said that, the HT problem which we consider in MU-MIMO WLAN settings is in a totally different context, and the system parameters such as error bound derived earlier become invalid. Specifically, when MU-MIMO WLANs are considered, the total feedback bandwidth T_{fb} and the number of feedback bits B are finite, whereas the numbers of antennas at the clients are heterogeneous.

In such a context, we quantify the quantisation error bound and study its impact on the system parameters. We derive a closed-form expression for the upper bound of the

[1]We consider zero delay feedback path, however studying the proposed model considering delay in the feedback path would be an interesting avenue for future research.

channel quantisation error and the average rate reduction due to the quantisation error with respect to the perfect CSI at APs in a MU-MIMO WLAN settings. The content of this chapter constitutes a published work by the author in [55].

7.2 System Model

Similarly to Chapter 3, we consider a MU-MIMO WLAN setting, where AP1 and AP2 are hidden terminals to each other. Each AP has heterogeneous clients, iPhones, laptops, HDTVs, etc., with a variable number of antennas. Since our proposed design uses the ZF technique to address the HT problem, we need to further develop a system model on top of the system model in the previous chapter, taking into account the practical constraints of MU-MIMO WLANs. Thus we add: a) the CSI feedback path of a finite bandwidth T_{fb} for the clients to send CSI to APs, and b) channel quantisation bits B for the representation of a partial CSI at APs, in the system model.

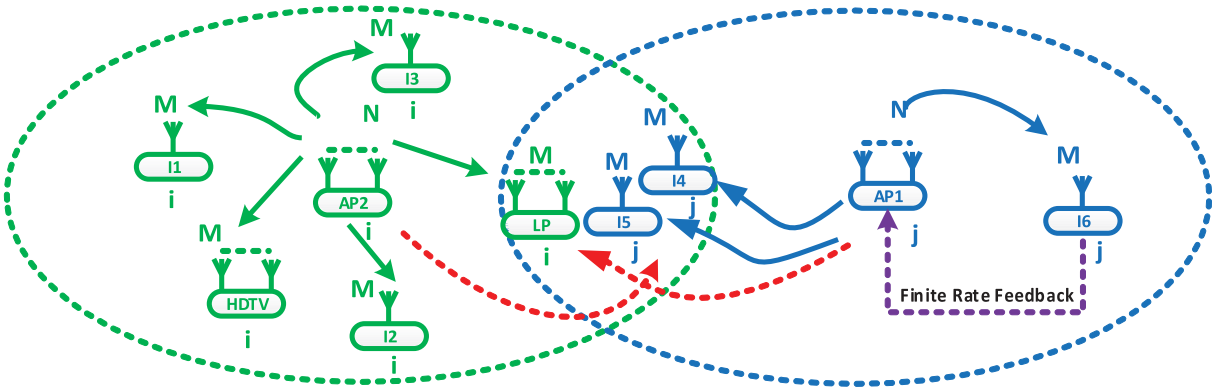

Figure 7.1: A diagrammatic representation of System Model with FRF, where a dotted violet line between the jth AP and jth client 'I6' represents a wireless FRF. However, the system model assumes the similar wireless FRFs that exist between all APs and clients in order to feedback CSI in quantised form.

We make the following assumptions pertaining to the partial CSI. First, the feedback

path is error free, low rate and with zero delay. Second, the channel varies according to the block fading model, where channels change independently from block to block. Third, the clients have a perfect CSI which allows us to neglect the channel estimation error.

Each client at the beginning of the block fading quantises their channels to B bits and feeds back the quantisation index instantaneously to the APs. The quantisation is performed using a quantisation codebook C based on Random Vector Quantisation (RVQ) [93]. The codebook is known to both the APs and the clients. We normalise the CSI of each client. For instance, for a jth client having M antennas, the channels with the jth AP can be written as $\mathbf{H}_{jj} = [\mathbf{h}_{j1}, \mathbf{h}_{j2}, ...\mathbf{h}_{jj}, ...\mathbf{h}_{jM}]$. After normalisation, we have $\tilde{\mathbf{H}}_{jj} \triangleq \frac{\mathbf{H}_{jj}}{\|\mathbf{H}_{jj}\|^2}$. Now, for channel quantisation, each of the columns of the normalised channel matrix $\tilde{\mathbf{H}}_{jj} = \left[\tilde{\mathbf{h}}_{j1}, \tilde{\mathbf{h}}_{j2}, ...\tilde{\mathbf{h}}_{jj}, ...\tilde{\mathbf{h}}_{jM}\right]$, where $\tilde{\mathbf{h}}_{jj} \in \mathbb{C}^{N\times 1}$ for the jth receiving antenna, first individually quantises with a codebook consisting of quantisation vectors $C \triangleq \{\mathbf{w}_1, \mathbf{w}_2, ...\mathbf{w}_{2^B}\}$ that form a minimum angle[2] to it [61, 62]. Thus

$$\hat{\mathbf{h}}_{jj} = \arg \min_{\mathbf{w}\in C} \sin^2(\angle(\tilde{\mathbf{h}}_{jj}, \mathbf{w})) \tag{7.1}$$

and $\hat{\mathbf{H}}_{jj} = \left[\hat{\mathbf{h}}_{j1}, \hat{\mathbf{h}}_{j2}, ...\hat{\mathbf{h}}_{jj}, ...\hat{\mathbf{h}}_{jM}\right]$. The quantisation error is given by

$$Z \triangleq \sin^2(\angle(\tilde{\mathbf{h}}_{jj}, \hat{\mathbf{h}}_{jj})). \tag{7.2}$$

It is to be noted that the increase in the number of antennas at the clients, M, effectively increases the quantisation codebook size from 2^B to $M.2^B$.

Let I_U^F be the set of indices of clients. Then $I_j^F = \{1 \leq j \leq U\}$ is the feedback index of the jth client out of U clients who feedback to the jth AP, i.e., $I_j^F \in I_U^F$. The Channel Quality Indicator (CQI) is not included in T_{fb}, and is perfectly known to the APs.

[2]Angle between two vectors, $\bar{\mathbf{u}}$ and $\bar{\mathbf{v}}$ refers to the standard convention $cos\theta = \frac{\{\bar{\mathbf{u}}.\bar{\mathbf{v}}\}}{\{\|\bar{\mathbf{u}}\|.\|\bar{\mathbf{v}}\|\}}$.

7.3 Expected Quantisation Error for MU-MIMO WLANs

In MU-MIMO WLAN setting, the number of antennas at APs, N, is fixed whereas the number of antennas at clients, M, is variable. Taking into account the FRF model in MU-MIMO WLANs, we have the following conditions:

Condition 1- The feedback parameter B is an offline design parameter, meaning that per-user feedback bits are decided beforehand for a given T_{fb}, SNR, N and M. Thus B does not vary dynamically from one coherence block to another. The main reason for keeping B fixed is because of wireless standards that are currently being finalised and the lack of flexibility to vary per-user feedback.

Condition 2- The feedback bandwidth T_{fb}, feedback bits B and the number of antennas at the APs N are fixed, whereas the number of antennas at the clients M are heterogeneous and are dynamically varying with the incoming and outgoing clients in a MU-MIMO WLANs. The implication is described as follows.

For instance, for $\mathbf{H}_{jj}$, when $M = 1$ then all B bits are used to quantise the channel vector of size $N \times 1$. However, when $M > 1$ the rank of the channel matrix $\mathbf{H}_{jj}$ is increased with increasing i.i.d. columns. Hence, assuming an isotropic distribution of B, the number of feeback bits per column is given by $\alpha = \frac{B}{M}$.

The complementary cumulative distribution of the quantisation error Z [94] is given by

$$Pr(\sin^2(\angle(\tilde{\mathbf{h}}_{jj}, \hat{\mathbf{h}}_{jj})) \geqslant z) = (1 - z^{N-1})^{2^{(\alpha)}}. \tag{7.3}$$

Thus we can write

$$\mathbb{E}[\sin^2(\angle(\tilde{\mathbf{h}}_{jj}, \hat{\mathbf{h}}_{jj}))] = \int_0^1 \left(1 - z^{N-1}\right)^{2^{(\alpha)}} dz. \tag{7.4}$$

The Integral representation of the Beta function [95, p.5] gives

$$\beta\left(c, \tfrac{a}{b}\right) = b \int_0^1 z^{a-1}\left(1 - z^b\right)^{c-1} dz, \quad a > 0,\, b > 0,\, c > 0. \tag{7.5}$$

Substituting $a = 1$, $b = N - 1$ and $c = (2^{(\alpha)} + 1)$ we get

$$\beta\left(2^{(\alpha)} + 1, \frac{1}{N-1}\right) = N - 1 \int_0^1 z^{1-1}\left(1 - z^{N-1}\right)^{2^{(\alpha)}} dz. \tag{7.6}$$

From (7.4) and (7.6) we get

$$\begin{aligned} &\mathbb{E}\left[\sin^2(\angle(\tilde{\mathbf{h}}_{jj}, \hat{\mathbf{h}}_{jj}))\right] \\ &= \tfrac{1}{N-1}\beta\left(2^{(\alpha)} + 1, \tfrac{1}{N-1}\right) \\ &= \frac{2^{(\alpha)}\Gamma\left(2^{(\alpha)}\right)\Gamma\left(1+\frac{1}{N-1}\right)}{\Gamma\left(2^{(\alpha)}+1+\frac{1}{N-1}\right)}. \end{aligned} \tag{7.7}$$

For the analytical expression (7.7) we have used the property $\Gamma(x+1) = x\Gamma(x)$. Now, calculating for $N > 2$ we get

$$\begin{aligned} &\mathbb{E}\left[\sin^2(\angle(\tilde{\mathbf{h}}_{jj}, \hat{\mathbf{h}}_{jj}))\right] \\ &\leq \frac{2^{(\alpha)}\Gamma\left(2^{(\alpha)}\right)}{\Gamma\left(2^{(\alpha)}+1+\frac{1}{N-1}\right)} \\ &= \frac{\Gamma\left(2^{(\alpha)}+1\right)}{\Gamma\left(2^{(\alpha)}+1+\frac{1}{N-1}\right)}. \end{aligned} \tag{7.8}$$

The inequality follows from the convexity of the gamma function [96] $\Gamma(x) \leq 1$ for $1 \leq x \leq 2$ and the fact that $\Gamma(1) = \Gamma(2) = 1$. Now by applying Kershaw's inequality for the gamma function [97] we have

$$\frac{\Gamma(x+s)}{\Gamma(x+1)} < \left(x + \frac{s}{2}\right)^{s-1}, \ \forall x > 0,\ 0 < s < 1. \tag{7.9}$$

Substituting $x = (2^{(\alpha)} + \frac{1}{N-1})$ and $s = (1 - \frac{1}{N-1})$ in (7.9) we can write (7.8) as

$$\begin{aligned} &\mathbb{E}\left[\sin^2(\angle(\tilde{\mathbf{h}}_{jj}, \hat{\mathbf{h}}_{jj}))\right] \\ &< \left(2^{(\alpha)} + \tfrac{1}{N-1} + \tfrac{1}{2}\left(1 - \tfrac{1}{N-1}\right)\right)^{-\frac{1}{N-1}} \\ &< \left(2^{(\alpha)} + \tfrac{N}{2(N-1)}\right)^{-\frac{1}{N-1}}. \end{aligned} \tag{7.10}$$

In (7.10), for $N = 2$, the term $(\frac{N}{2(N-1)}) = 1$ and for $N > 2$, $(\frac{N}{2(N-1)}) < 1$, thus we can approximately rewrite (7.10) as

$$\mathbb{E}\left[\sin^2(\angle(\tilde{\mathbf{h}}_{jj}, \hat{\mathbf{h}}_{jj}))\right] < 2^{-\left(\frac{B}{M(N-1)}\right)}. \tag{7.11}$$

Note that (7.11) follows for M=1, since $\tilde{\mathbf{h}}_j$ and $\hat{\mathbf{h}}_j$ are vectors of size $N \times 1$. When $M > 1$, the expression can be rewritten as

$$\mathbb{E}\left[\sin^2(\angle(\tilde{\mathbf{H}}_{jj}, \hat{\mathbf{H}}_{jj}))\right] < 2^{-\left(\frac{B}{M(N-1)}\right)}, \forall M \geq 1. \tag{7.12}$$

Note in (7.12) that we are concerned with the angle between the normalized and the estimated channel vector components and not with the angle between two matrices. Thus the expectation of the quantization error is upper bounded by $2^{-\left(\frac{B}{M(N-1)}\right)}$.

7.4 Rate Reduction for Zero-forcing Beamforming with FRF in MU-MIMO WLANs

Since our design uses the ZF technique to remove interference to the undesired clients in order to address the HT problem, it becomes imperative to study the implication(s) it brings to the system. Furthermore, we consider a FRF system in our design, owing to the practical constraints for system deployments. So, it becomes essential to measure the rate reduction for the ZF technique with FRF in MU-MIMO WLANs. Thus we compare the rate, when using the ZF technique with FRF, with respect to the rate when using the ZF technique with the perfect CSI, in MU-MIMO WLANs.

The ZF precoding vector when CSI is perfectly known to the jth AP is given by

$$\mathbf{v}_j^{zf} = \frac{\prod_{\mathbf{H}_{ji}}^{\perp} \mathbf{H}_{jj}}{\left\| \prod_{\mathbf{H}_{ji}}^{\perp} \mathbf{H}_{jj} \right\|} \mathbf{U} \tag{7.13}$$

where $\prod_{\mathbf{H}_{ji}}^{\perp} = \mathbf{I}_N - \mathbf{H}_{ji}(\mathbf{H}_{ji}{}^H \mathbf{H}_{ji})^{-1} \mathbf{H}_{ji}{}^H$ denotes the projection onto the orthogonal complement of the column space of $\mathbf{H}_{ji}$. $\mathbf{I}_N$ represents the identity matrix of size N. $\mathbf{U} \in \mathbb{C}^{M \times 1}$ is a unit vector acting as a demultiplexer where $\mathbf{U}^H \mathbf{U} = 1$. The rate associated

with this transmission is given by

$$R_j^{ZF} = N\mathbb{E}\left[\log_2\left(1+\frac{\gamma\|\mathbf{H}_{jj}\|^2\left\|\tilde{\mathbf{H}}_{jj}^H\mathbf{v}_j^{zf}\right\|^2}{\sigma^2}\right)\right]. \tag{7.14}$$

However, with a FRF, the precoding vector is calculated in the same way as in (7.13) but with the quantised channel information $\hat{\mathbf{H}}_{jj}$ and $\hat{\mathbf{H}}_{ji}$. The rate for ZFB with FRF is given by

$$R_j^{FRF} = N\mathbb{E}\left[\log_2\left(1+\frac{\gamma\|\mathbf{H}_{jj}\|^2\left\|\tilde{\mathbf{H}}_{jj}^H\hat{\mathbf{v}}_j\right\|^2}{\sigma^2+\sum_{i\neq j}^{K}\gamma\|\mathbf{H}_{ij}\|^2\left\|\tilde{\mathbf{H}}_{ij}^H\hat{\mathbf{v}}_i\right\|^2}\right)\right]. \tag{7.15}$$

Although ZF beamforming is used, there is residual interference because the beamformer is based on the quantised CSI. Thus we analyse the interference term i.e., $\left\|\tilde{\mathbf{H}}_{ij}^H\hat{\mathbf{v}}_i\right\|^2$ in (7.15) further.

Considering $M = 1$, since RVQ is used, the quantisation error Z is isotropically distributed in $\mathbb{C}^N$. Thus, with the magnitude of Z, the channel direction can be written as the sum of two vectors, one in the direction of the quantisation, and the other isotropically distributed in the null space of quantisation [64], as

$$\begin{aligned}\tilde{\mathbf{h}}_{ij} &= \sqrt{(1-Z)}\,\hat{\mathbf{h}}_{ij}+\sqrt{Z}\,\mathbf{g}\\ \left|\tilde{\mathbf{h}}_{ij}^H\hat{\mathbf{v}}_i'\right|^2 &= (1-Z)\left|\hat{\mathbf{h}}_{ij}^H\hat{\mathbf{v}}_i'\right|^2+Z\left|\mathbf{g}^H\hat{\mathbf{v}}_i'\right|^2\\ &= Z\left|\mathbf{g}^H\hat{\mathbf{v}}_i'\right|^2\end{aligned} \tag{7.16}$$

where $\mathbf{g}\in\mathbb{C}^{N\times 1}$ and $\hat{\mathbf{v}}_i'\in\mathbb{C}^{N\times 1}$ are vectors in the $(N-1)$-dimensional null space of $\hat{\mathbf{h}}_{ij}$. Thus the quantity $\left|\mathbf{g}^H\hat{\mathbf{v}}_i'\right|^2$ is beta $(1, N-2)$ distributed which is independent of Z. Since the beta random variable has support $[0,1]$, then

$$\left|\tilde{\mathbf{h}}_{ij}^H\hat{\mathbf{v}}_i'\right|^2 = Z\left|\mathbf{g}^H\hat{\mathbf{v}}_i'\right|^2 \leq \sin^2(\angle(\tilde{\mathbf{h}}_{jj},\hat{\mathbf{h}}_{jj})), \tag{7.17}$$

where $\mathbf{g}\in\mathbb{C}^{N\times 1}$ and $\hat{\mathbf{v}}_i'\in\mathbb{C}^{N\times 1}$ are vectors in the $(N-1)$-dimensional null space of $\hat{\mathbf{h}}_j$. The quantity $\left|\mathbf{g}^H\hat{\mathbf{v}}_i'\right|^2$ is beta $(1, N-2)$ distributed and is independent of Z.

For $M > 1$, the expectation of the interference term, i.e., $\mathbb{E}\left[\left\|\tilde{\mathbf{H}}_{ij}^H \hat{\mathbf{v}}_i\right\|^2\right]$, is the product of the exception of the quantisation error and the expectation of beta $(1,N-2)$ random variables, which is equal to $\frac{1}{N-1}$. Thus from (7.12) and (7.17)

$$\mathbb{E}\left\|\tilde{\mathbf{H}}_{ij}^H \hat{\mathbf{v}}_i\right\|^2 = \left(\frac{1}{(N-1)}\right) 2^{-\left(\frac{B}{M(N-1)}\right)}, \forall M > 1, \tag{7.18}$$

where $\hat{\mathbf{v}}_i$ is calculated in the same way as in (7.13) but with the quantised channel information $\hat{\mathbf{H}}_{jj}$ and $\hat{\mathbf{H}}_{ij}$.

Since FRF with B bits per client is used, this incurs a throughput loss relative to perfect CSIT

$$\begin{aligned}
&\Delta R(P) \triangleq \tfrac{1}{N}(R_j^{ZF} - R_j^{FRF}) \\
&= \mathbb{E}\left[\log_2\left(\sigma^2 + \gamma \left\|\mathbf{H}_{jj}\right\|^2 \left\|\tilde{\mathbf{H}}_{jj}^H \mathbf{v}_j^{zf}\right\|^2\right)\right] \\
&-\mathbb{E}\left[\log_2\left(1 + \frac{\gamma\|\mathbf{H}_{jj}\|^2\left\|\tilde{\mathbf{H}}_{jj}^H \hat{\mathbf{v}}_j\right\|^2}{\sigma^2 + \sum_{i\neq j}^K \gamma\|\mathbf{H}_{ij}\|^2\left\|\tilde{\mathbf{H}}_{ij}^H \hat{\mathbf{v}}_i\right\|^2}\right)\right] \\
&= \mathbb{E}\left[\log_2\left(\sigma^2 + \gamma \left\|\mathbf{H}_{jj}\right\|^2 \left\|\tilde{\mathbf{H}}_{jj}^H \mathbf{v}_j^{zf}\right\|^2\right)\right] \\
&-\mathbb{E}\left[\log_2\left(\begin{array}{l}\sigma^2 + \sum_{i\neq j}^K \gamma\|\mathbf{H}_{ij}\|^2\left\|\tilde{\mathbf{H}}_{ij}^H \hat{\mathbf{v}}_i\right\|^2 + \\ \gamma\|\mathbf{H}_{jj}\|^2\left\|\tilde{\mathbf{H}}_{jj}^H \hat{\mathbf{v}}_j\right\|^2\end{array}\right)\right] \\
&+\mathbb{E}\left[\log_2\left(\sigma^2 + \sum_{i\neq j}^K \gamma\|\mathbf{H}_{ij}\|^2\left\|\tilde{\mathbf{H}}_{ij}^H \hat{\mathbf{v}}_i\right\|^2\right)\right] \\
&\leq^a \mathbb{E}\left[\log_2\left(\sigma^2 + \gamma \left\|\mathbf{H}_{jj}\right\|^2 \left\|\tilde{\mathbf{H}}_{jj}^H \mathbf{v}_j^{zf}\right\|\right)^2\right] \\
&-\mathbb{E}\left[\log_2\left(\sigma^2 + \gamma\|\mathbf{H}_{jj}\|^2\left\|\tilde{\mathbf{H}}_{jj}^H \hat{\mathbf{v}}_j\right\|^2\right)\right] \\
&+\mathbb{E}\left[\log_2\left(\sigma^2 + \sum_{i\neq j}^K \gamma\|\mathbf{H}_{ij}\|^2\left\|\tilde{\mathbf{H}}_{ij}^H \hat{\mathbf{v}}_i\right\|^2\right)\right] \\
&=^b \mathbb{E}\left[\log_2\left(\sigma^2 + \sum_{i\neq j}^K \gamma\|\mathbf{H}_{ij}\|^2\left\|\tilde{\mathbf{H}}_{ij}^H \hat{\mathbf{v}}_i\right\|^2\right)\right].
\end{aligned} \tag{7.19}$$

The inequality (a) follows because $\sum_{i\neq j}^K \gamma\|\mathbf{H}_{ij}\|^2\left\|\tilde{\mathbf{H}}_{ij}^H \hat{\mathbf{v}}_i\right\|^2 \geq 0$ and $\log(.)$ is a monotonically increasing function whereas the equality (b) follows because $\mathbb{E}\left[\log_2\left(\sigma^2 + \gamma \left\|\mathbf{H}_{jj}\right\|^2 \left\|\tilde{\mathbf{H}}_{jj}^H \mathbf{v}_j^{zf}\right\|\right)\right.$ $\left.\mathbb{E}\left[\log_2\left(\sigma^2 + \gamma\|\mathbf{H}_{jj}\|^2\left\|\tilde{\mathbf{H}}_{jj}^H \hat{\mathbf{v}}_j\right\|^2\right)\right]\right.$. Note that $\mathbf{v}_j^{zf}$ and $\hat{\mathbf{v}}_j$ are isotropically distributed

unit vectors independent of $\tilde{\mathbf{H}}_{jj}^{H}$. Applying Jensen's inequality to the upper bound in b and exploiting the independence of the channel norm (which satisfies $\mathbb{E}\left[\left\|\tilde{\mathbf{H}}_{ij}^{H}\right\|^2\right] = N$) and the channel direction, we get

$$\begin{aligned}\Delta R(P) &\leq \log_2\left(\sigma^2 + \gamma\,(N-1)\,\mathbb{E}\left[\|\mathbf{H}_{ij}\|^2\right]\mathbb{E}\left[\left\|\tilde{\mathbf{H}}_{ij}^{H}\hat{\mathbf{v}}_i\right\|^2\right]\right)\\ &= \log_2\left(\sigma^2 + P\,(N-1)\,\mathbb{E}\left[\left\|\tilde{\mathbf{H}}_{ij}^{H}\hat{\mathbf{v}}_i\right\|^2\right]\right).\end{aligned} \tag{7.20}$$

Thus, from (7.18) and (7.20), we get

$$\Delta R(P) \leq \log_2\left(\sigma^2 + P\,2^{-\frac{B}{M(N-1)}}\right),\ \forall M > 1. \tag{7.21}$$

7.5 Numerical Analysis

In this section, we present a numerical analysis based on the analytical expressions derived in the previous sections.

7.5.1 Error Bound Analysis

Fig. 7.2a is plotted according to the analytical expression (7.12) which shows that, for an arbitrary fixed number of feedback bits $B = 24$ and with an increasing number of antennas at the clients, the error bound is increased. This is an expected result because B is constant and, whenever M increases, the limited B bits have to quantise the increased channel realisations with the dimensionality of M. This gives rise to the error bound. Also, with an increasing number of antennas at the APs, as for example $N = 4$, 6 and 8, we see that the error bound increases. Putting this another way, with an increment in M, the number of feedback bits per antenna, i.e., $\alpha = \frac{B}{M}$, is decreased. Fig. 7.2b shows that, with fixed B, the error bound increases with decreasing α. In both of these plots we consider $N \geq M$.

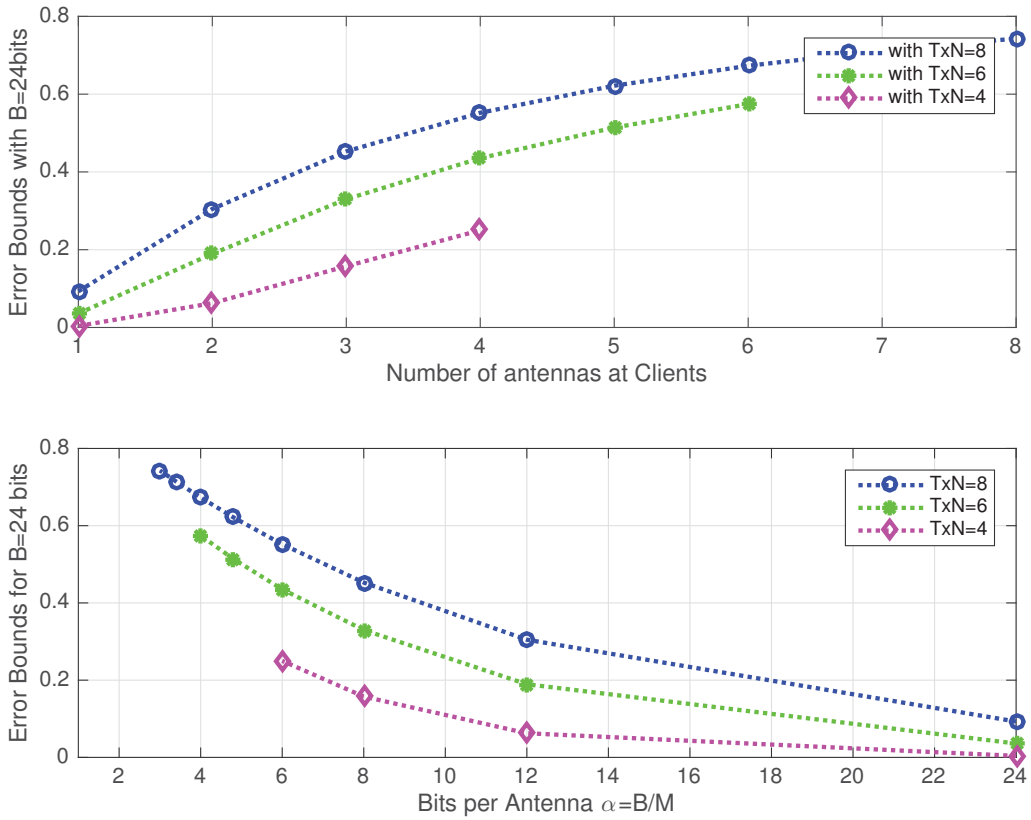

Figure 7.2: Error Bounds with: a. number of antennas at clients M and b. Feedback bits per antenna α.

7.5.2 Rate with Finite Rate Feedback

We analyse the average rate with FRF at the APs, perfect CSI at the transmitters/APs (CSIT) and without CSI, i.e., noCSIT at the transmitter/APs. Assuming $M = 1$, $\mathbf{h}_{jj}$, the ergodic capacity, is given by $C_{CSIT} = \mathbb{E}[\log_2(1 + P\|\mathbf{h}_{jj}\|^2)]$. With noCSIT the optimum transmission strategy is to transmit equal power independently to the N transmit antennas at the APs: $C_{noCSIT} = \mathbb{E}[\log_2(1 + \frac{P}{N}\|\mathbf{h}_{jj}\|^2)]$, where $\gamma = \frac{P}{N}$. With FRF the average rate achieved assuming RVQ is $C_{FRF} = \mathbb{E}[\log_2(1 + P\|\mathbf{h}_{jj}\|^2\cos^2(\angle\tilde{\mathbf{h}}_{jj}, \hat{\mathbf{h}}_{jj}))] \approx \mathbb{E}[\log_2(1 + P\|\mathbf{h}_{jj}\|^2(1 - 2^{-\frac{B}{M(N-1)}}))]$.

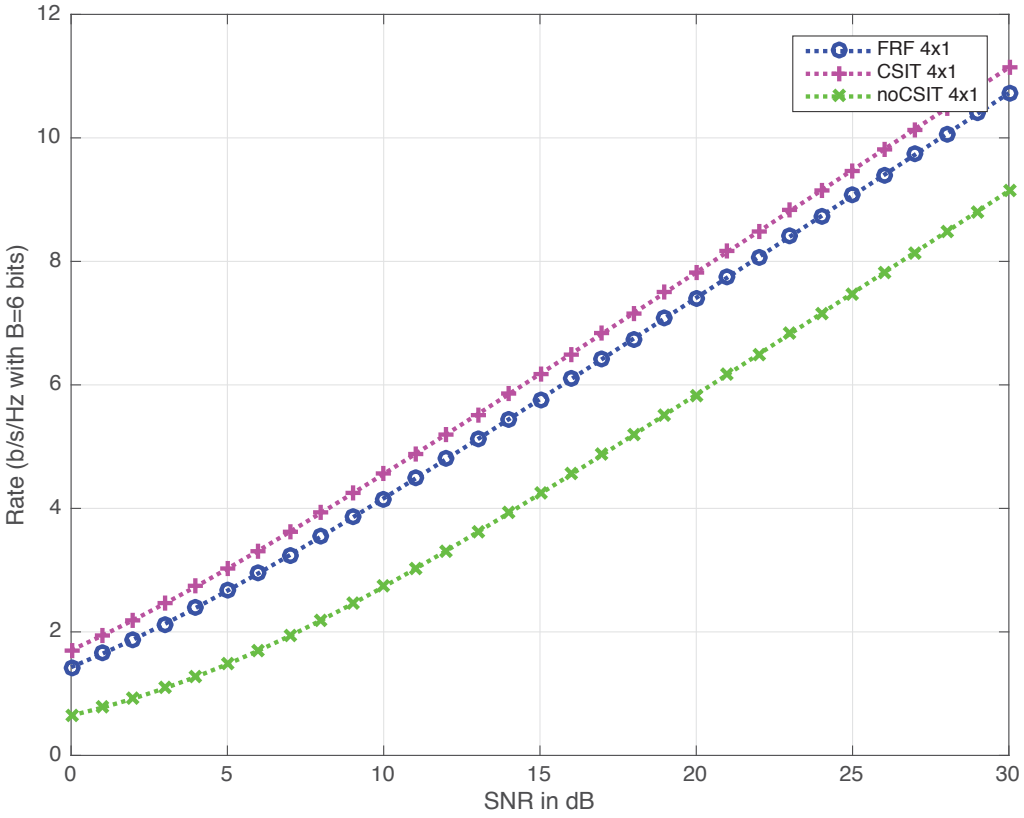

Figure 7.3: 4×1 MISO capacity with $B = 6$ bits.

Fig. 7.3 shows a capacity comparison among C_{CSIT}, C_{FRF} and C_{noCSIT} for $N = 4$, $M = 1$ and $B = 6$ bits. C_{FRF} lies between C_{CSIT} and C_{noCSIT} for a 4×1 MISO system with $B = 6$ bits. Note that C_{FRF} lies below C_{CSIT}, which is an expected result as the quantised forms of the channels are used for transmissions. The capacity reduction in comparison to C_{CSIT} is due to the SNR degradation in dB which is given by

$$\begin{aligned} &\Delta SNR_{dB} = 10\log_{10}\left(P\|\mathbf{h}_j\|^2\right) \\ &-10\log_{10}\left(P\|\mathbf{h}_j\|^2\left(1+2^{-\frac{B}{M(N-1)}}\right)\right) \\ &= -10\log_{10}\left(1+2^{-\frac{B}{M(N-1)}}\right). \end{aligned} \tag{7.22}$$

For $N = 4$, $M = 1$ and $B = 6$ bits, $\Delta SNR_{dB} = 1.25$ dB. Thus, the capacity in C_{FRF} is seen to be within about 1.25 dB of C_{CSIT} in Fig. 7.3. A similar calculation between

C_{CSIT} and C_{noCSIT} gives an SNR degradation of $10\log_{10}N$ dB, which is about 6 dB for a 4×1 MISO system with $B = 6$ bits. Thus we observe three capacity curves at gaps of 1.25 dB and 6 dB from C_{CSIT} respectively.

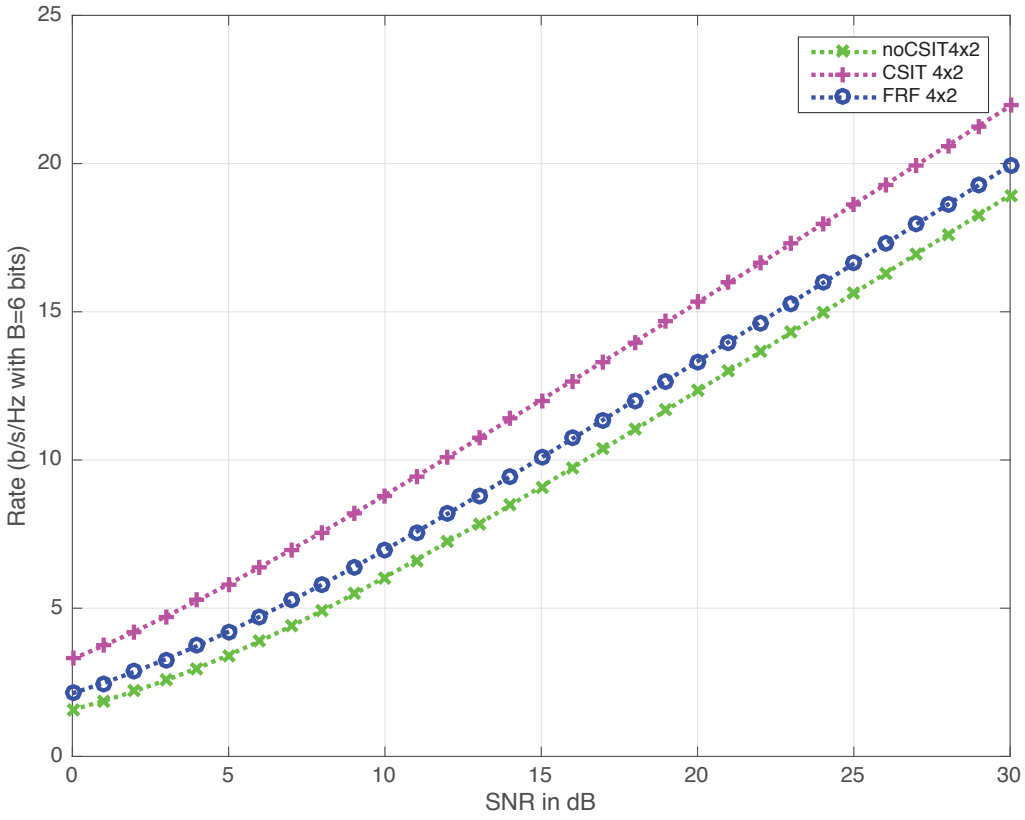

Figure 7.4: 4×2 MIMO capacity with $B = 6$ bits.

In Fig. 7.4, we compare the capacity among C_{CSIT}, C_{FRF} and C_{noCSIT} for $N = 4$, $M = 2$ and $B = 6$ bits. We observe that C_{FRF} lies below C_{CSIT}. The reason for this capacity degradation is the use of quantised CSI for transmission.

It is worthwhile to compare Fig. 7.3 and Fig. 7.4 and note that the C_{FRF} reduction with respect to C_{CSIT} is greater in a 4×2 MIMO system than in a 4×1 MISO system, as we can see a larger gap in the former. This is because, with increasing numbers of antennas at the clients M, the feedback bits per antenna $\alpha = \frac{B}{M}$ decreases. Also, we

know from the error bound analysis and Fig. 7.2 that, with decreasing α, the error bound increases. Now, the error bound versus the capacity in Fig. 7.5 shows that, with an increase in error bound, C_{FRF} decreases. Thus, C_{FRF} lies below C_{CSIT} for increasing M, as shown in Fig. 7.4. The amount of C_{FRF} reduction is given by the SNR degradation,

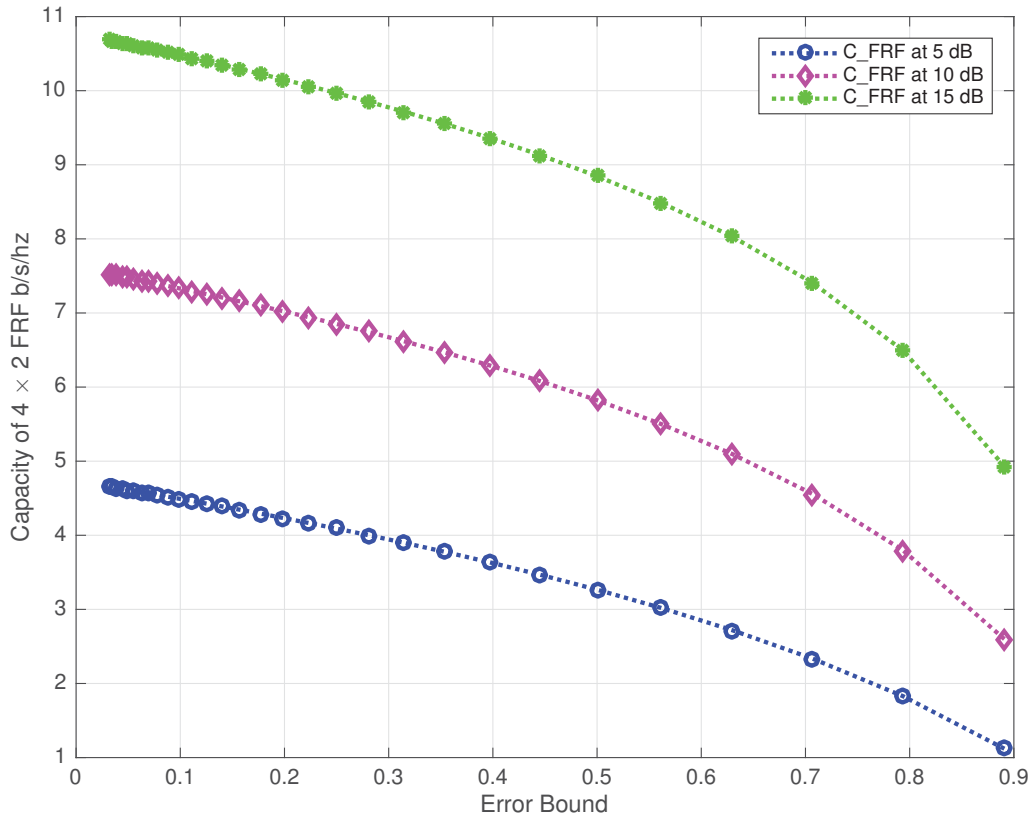

Figure 7.5: 4×2 MIMO capacity with error bound.

ΔSNR_{dB}. For $N = 4$, $M = 2$ and $B = 6$ bits, $\Delta SNR_{dB} = 3$ dB. Thus, C_{FRF} performs within about 3 dB of C_{CSIT} as shown in Fig. 7.4.

7.5.3 Average Rate Reduction for Zero-forcing Beamforming due to FRF

The analytical expression in (7.21) gives the upper bound of the throughput loss incurred by FRF with respect to CSIT, i.e., $\Delta R = C_{CSIT} - C_{FRF}$. In MU-MIMO with $B = 6$ bits, $N = 4$ and 3 clients with $M = 1$, 2 and 4 receiving antennas, the throughput loss for C_{FRF} in comparison to C_{CSIT} is plotted in Fig. 7.6. We observe that, with an increase in M, the throughput loss ΔR for C_{FRF} also increases. This is an expected result because, with increasing M, α decreases, and as a result the error bound increases. With increasing error bound C_{FRF} decreases. Thus the throughput loss, ΔR, increases with increasing M as shown in Fig. 7.6. Additionally, from (7.15) and (7.21), the expression for R_j^{FRF} has $\Delta R(P)$ in the denominator. This term reduces R_j^{FRF} with increasing SNR, which as a result maintains a constant gap with C_{CSIT} with increasing SNR.

7.6 Conclusion

This chapter adds a practical dimension to our proposed design, using the ZF technique to address the HT problem in MU-MIMO WLANs. Specifically, we consider some inevitable practical constraints, such as: channel estimation errors at receivers/clients, limited bandwidth of the feedback path, cost of the feedback overheads, etc., when implementing the ZF technique in MU-MIMO WLANs, and introduce an FRF approach in our design on top of the previously proposed design, where APs having the perfect CSI. We quantify both the quantisation error bound and the average rate reduction for FRF in a MU-MIMO WLAN setting. The feedback bandwidth T_{fb}, the number of feedback bits B, the number of antennas at the APs, N, are fixed but the number of antennas at clients M is variable. We derive a closed form expression for both the quantisation error bound and the average rate reduction with respect to perfect CSI. Our numerical studies show

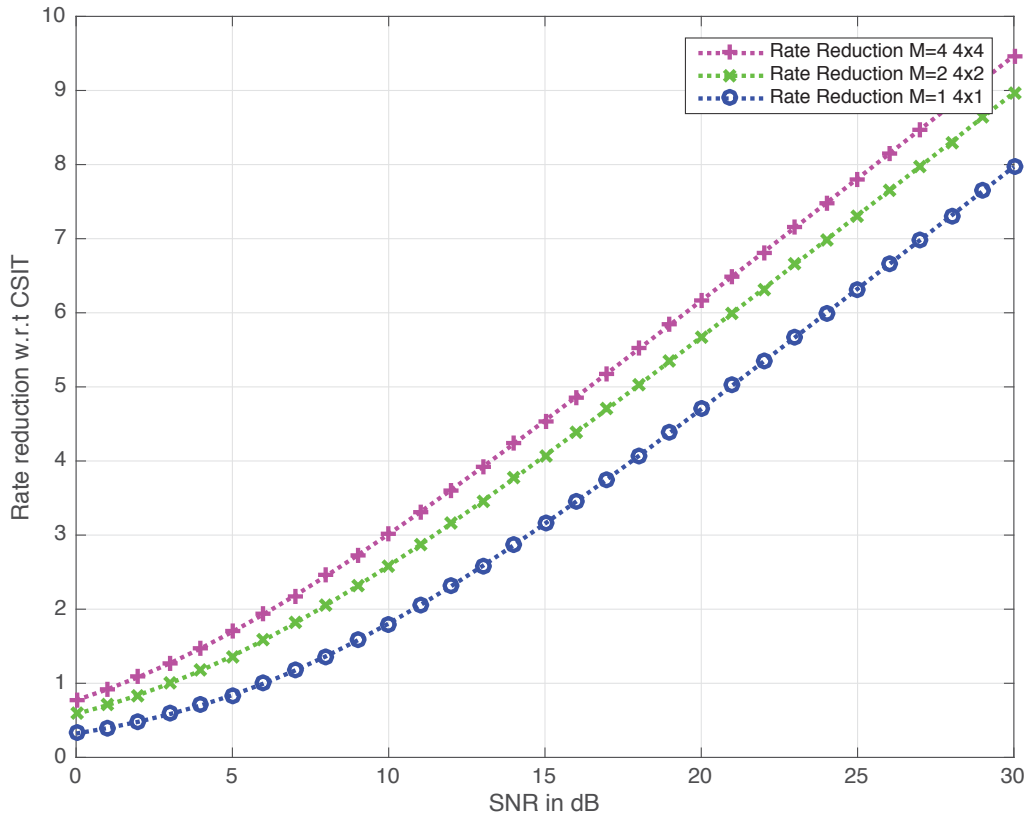

Figure 7.6: Average rate reduction due to Finite Rate Feedback when $B = 6$ bits.

that, with increasing M, both the quantisation error bound and the average rate reduction with respect to perfect CSI increase. Specifically, Chapter 7 studies the implications of adopting the practically motivated FRF model in MU-MIMO WLANs. However, the fundamental results, for instance, the increase of both the quantisation error bound and the average rate reduction with increasing M, do not justify the essence of the additional resource at clients, i.e., with $M > 1$. So, this is not a desired scenario for a MU-MIMO WLAN setting. Thus the relation of B with variable M is further analysed in Chapter 8.

Chapter 8

Relaxed Zero-forcing Precoding Based on Channel Quantisation for Finite Rate Feedback based MU-MIMO WLANs

8.1 Introduction

In the previous chapter, we consider a practical scenario for a MU-MIMO WLAN and discuss the FRF model owing to the inevitable constraints relating to channel feedback for ZF precoding. Basically, the quantisation bits relating to quantisation error and the average throughput loss of ZF w.r.t. perfect CSIT are meticulously presented. In this chapter, we further develop the topic pertaining to the practical dimension and do the followings.

First, the key observation in Chapter 7 is that with increasing M, the average throughput loss for ZF w.r.t. perfect CSIT also increases as shown in Fig. 7.6. Since it does

not justify the essence of the additional resource, this is not a desired scenario for a MU-MIMO WLAN settings. Thus the relation of B with variable M is further analysed in this chapter.

Specifically, the reason for throughput loss is due to the decrement in feedback bits per column, i.e., $\alpha = \frac{B}{M}$. As feedback bits B remains fixed and M increases with the heterogeneous antenna clients, it increases the upper bound of the quantisation error according to (7.12) and subsequently increases the throughput loss as shown in (7.21).

Thus there is a need to find the relation of B with M that checks the quantisation error with increasing M. We derive an analytical expression that relates B with M and limits the quantisation error bound to a constant level irrespective of the heterogeneous clients with a variable receiving antennas, M.

Second, as mentioned earlier, there will always be a residual interference, in addition to the thermal noise, at undesired client which cannot be removed although ZF transmission strategy is used.

It is because in FRF system, the beamformer is based on the quantised channels from the desired and the undesired clients, however, the downlink and the uplink channels of desired and undesired clients are real (not in quantised form). Thus the beamformer cannot completely eliminate interference to the undesired clients as there is no orthogonality between quantised beamformer and the unquantised channels of undesired clients.

For instance, say $\hat{\mathbf{v}}_j$, is the quantised beamformer calculated for the jth AP taking into account the quantised desired channel, say $\hat{\mathbf{H}}_{jj}$, and an undesired quantised channel, say $\hat{\mathbf{H}}_{ji}$, for an FRF-based MU-MIMO WLAN. However, during the downlink and the uplink, the desired and the undesired channels are real, i.e., not quantised channels, say $\tilde{\mathbf{H}}_{jj}$ and $\tilde{\mathbf{H}}_{ji}$ respectively. Thus, $\hat{\mathbf{v}}_j$ cannot remove interference to the undesired clients as $\tilde{\mathbf{H}}_{ji}^H \hat{\mathbf{v}}_j \neq \mathbf{0}$, $\forall i \neq j$ because $\hat{\mathbf{v}}_j$ projects $\hat{\mathbf{H}}_{jj}$ on to the complementary orthogonal column space of $\hat{\mathbf{H}}_{ji}$ and not $\tilde{\mathbf{H}}_{ji}$. Thus there is always residual interference given the constraint

$\tilde{\mathbf{H}}_{ji}^{H}\hat{\mathbf{v}}_j \neq \mathbf{0}$, $\forall i \neq j$.

Since our design considers a practical case and uses the ZF transmission strategy for addressing the HT problem, we need the optimal precoding vector based on quantised channel. The channel sounding process in our design supports the quantised channels and such desired and undesired quantised channels are fedback by the VHT MIMO control field inside the compressed beamforming frame during channel sounding as shown in Fig. 6.1. The concurrent transmission algorithm then selects the quantisation indices of the desired channels before the calculation of the optimal ZF precoding vector.

In order to calculate the optimal precoding vector, we first use the upper bound of the residual interference as quantified, $\mathbb{E}\left\|\tilde{\mathbf{H}}_{ij}^{H}\hat{\mathbf{v}}_i\right\|^2 = \left(\frac{1}{(N-1)}\right)2^{-\left(\frac{B}{M(N-1)}\right)}$, $\forall M > 1$, in (7.18) and relax the interference constraint and power constraint of the optimisation problem with the interference upper bound and the total power P_j respectively.

Although the research framework is the RZF, similar to those in [69, 70, 71], however, the scenarios considered are different, leading to different constraints and solutions. Also, approaches to solutions are different depending on complexity, practicality and objective of the solutions.

We principally focus on the influence of quantisation error on the optimal ZF precoding vector in FRF based MU-MIMO WLANs for addressing the HT problem. The interesting result is that the optimal precoding vector can be tuned with B, used for channel quantisation. The content of this chapter constitutes a journal submission made by the author in [59].

8.2 The Feedback Bits B

The choice of feedback bits B is fundamental for an FRF system. Since B bits are used for channel quantization, this forms a principal basis for the upper bound of the channel

quantization error as shown in (7.12). This in turn determines whether the channel at transmitters/APs is fine or coarse. Apparently, the average throughput loss depends on the channel quantization error as shown in the analytical expression (7.21).

In the previous chapter, in order to show the characteristic trend, we used B randomly, i.e., $B = 24$ bits and $B = 6$ bits in our numerical studies. However, in this section, we primarily concentrate on the choice of feedback bits B for FRF-based MU-MIMO WLANs.

8.2.1 Determining Feedback Bits B

Specifically, given the number of antennas at APs, N, heterogeneous antenna clients, M, and noise variation σ^2, we derive a closed-form expression for feedback bits B for FRF-based MU-MIMO WLANs.

We express B as a function of θ, where θ is the slope of the $\Delta R\,(P)$ vs B curve as shown in Fig. 8.1. Since we are interested in measuring the absolute unit change in the y-axis, i.e., $|\Delta R\,(P)\,|$ w.r.t. the absolute unit change in the x-axis, i.e., $|B|$, it is worthwhile to note that θ is given by $\theta = \tan^{-1}\left(\frac{|\Delta R(P)|}{|B|}\right)$. Since θ represents the slope, it is measured in degrees. Put another way, we define θ as an angle (measured in degrees) subtended by the tangent on the curve $\Delta R\,(P)$ vs B.

Fig. 8.1, shows the characteristic curve according to the analytical expression (7.21). It shows the variation of the average throughput loss (of FRF with CSIT), $\Delta R(P)$, w.r.t. quantisation bits B, for $N = 4$, SNR = 10 dB and $M = 1$, 2, 3 and 4. Since $\Delta R(P)$ is the measure of the closeness between the average rate of FRF and CSIT, we see that this information is vital for determining the feedback bits B in MU-MIMO WLANs.

With increasing B, $\Delta R(P)$ decreases initially but later becomes asymptotic with the x-axis with increasing B. We may regard this as a high-gain region, meaning that with a small increment in B, there is a rapid decrement of $\Delta R\,(P)$, and a low-gain region, meaning that with a large increment of B, there is little decrement of $\Delta R\,(P)$ in Fig. 8.1.

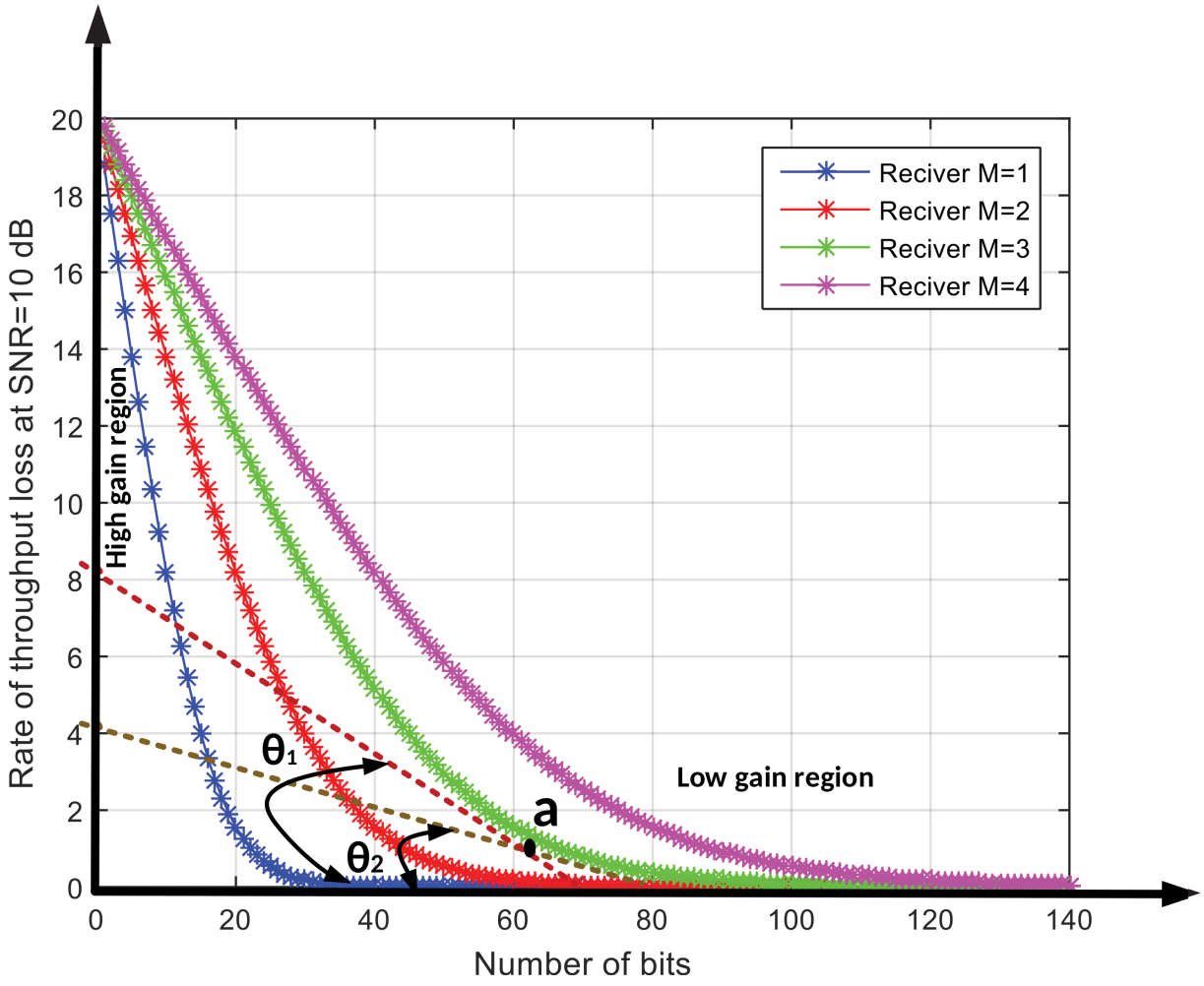

Figure 8.1: Average rate reduction due to Finite Rate Feedback vs feedback bits B.

This means that in a high-gain region, with little increment of B, the gap between C_{CSIT} and C_{FRF} decreases rapidly, whereas in a low-gain region the gap decreases slowly with increasing B.

Since the second derivative of $\Delta R\,(P)$ w.r.t. B, i.e., $\frac{d^2\Delta R(P)}{dB^2}$, is zero, there is no inflection point on Fig. 8.1. This means that there is no minima or maxima on the curve which can suggest a minimum and maximum B for $\Delta R\,(P)$. In fact, the appropriate choice of B becomes a system parameter and can vary from system to system. However, the chosen B for a system decides how well the system could perform relative to the perfect CSIT.

Thus we have $\frac{d\Delta R(P)}{dB} = \tan\theta$ which expresses B in terms of θ. We derive a closed-form expression for B with varying M and θ in Lemma 1.

Lemma 1: The number of feedback bits B in MU-MIMO WLANs is given by

$$B = -\frac{M(N-1)}{\ln(2)}\left(\left(\ln\left(\frac{P}{\tan\theta}\right) + \ln\left(\frac{1}{M(N-1)} + 1\right)\right) + \ln\left(\sigma^2\right)\right).$$

The negative form of the expression is due to the (negative) slope of the $\Delta R(P)$ vs B curve. We take an absolute value of feedback bits, i.e., $|B|$, since negative feedback bits B does not hold any physical meaning.

Proof: See Appendix C.1.

Given $SNR = 10$ dB, $N = 4$ and variable M, we plot B as a function of θ in Fig. 8.2a, according to (appendix C.2). With increasing slope θ, feedback bits B decreases. This is an expected behavior as per the nature of the average throughput loss. For instance, in a reference point a in Fig. 8.1, we see that with increasing slope of the $\Delta R(P)$ vs B curve, B decreases, i.e., $\theta_1 > \theta_2$, but $B_1 < B_2$.

Specifically in Fig. 8.2a, we observe that, given $\theta = 1$, the corresponding B is about 28 bits for $M = 1$. It is interesting to note that prior work on the optimisation of ZF beamforming [98, p.6], considering feedback bits B, number of transmitting antennas, $N = 4$, receiving antennas $M = 1$, $SNR = 10$ dB and total aggregate feedback load T_{fb}, where $\frac{T_{fb}}{B}$=number of users selected for feedback, suggested a similar finding for about $B = 25$ bits. We infer that our analytical expression (appendix C.2) complies with the results in [98] for $M = 1$.

Additionally, it can be deduced that, in the prior study, the maximum closeness be-

tween the rate with FRF and CSIT was observed to be in the range $1 \leq \theta < 2$ degrees, which resulted in $B = 25$ bits. However, our analytical expression (appendix C.2) benefits two-fold: a) it is able to find B, for even $\theta < 1$ degree, and can go closer to CSIT to the degree we please, b) it can find B for M=1 (cellular case) as well as for $M > 1$, which is very useful for heterogeneous antenna clients in MU-MIMO WLANs.

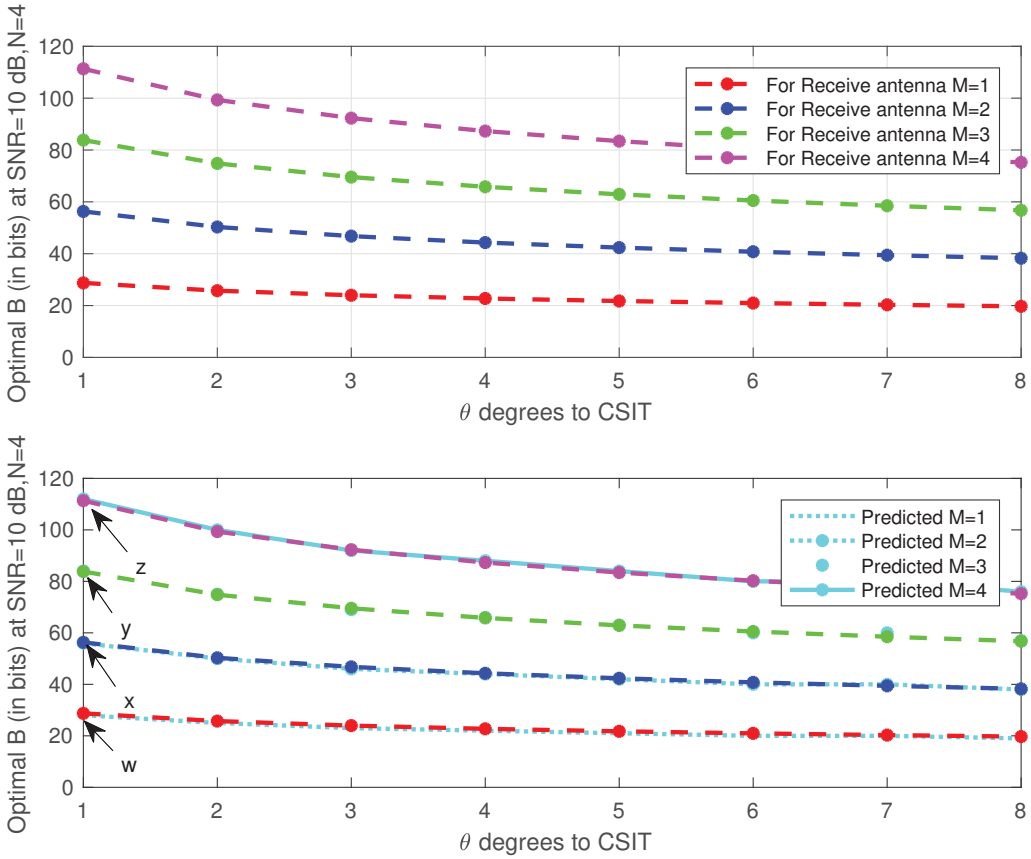

Figure 8.2: a) Feedback bits B as a function of θ, b) Predicted bits B.

8.2.2 Feedback Bits B for Heterogeneous Antenna Clients in MU-MIMO WLANs

From Fig. 8.2a and Fig. 8.2b, we see different values of B corresponding to different values of M for any given slope. For instance, at θ=1, the values of B for each M heterogeneous antenna clients are different as shown in the reference points, w, x, y and z respectively. For a MU-MIMO WLAN setting, it is crucial that, with additional antennas at clients, the quantisation error would not increase, resulting in a decrement of the average rate. Thus, we derive a closed-form expression, which enables us to predict the number of feedback bits B in order to limit the quantisation error to a certain level and maintain the average throughput loss, $\Delta R(P)$, irrespective of the increment in the number of receiving antennas M.

Lemma 2: The analytical expression to predict B with variable M and θ is given by

$$B = \frac{M(1-N)}{\log(2)} \left(\log_{10} \left(1 - 10^{\frac{10\log_{10}(\eta)}{10}} \right) \right), \text{ where}$$

$$\eta = \left(1 - 2^{-\frac{\left(\left(\ln\left(\frac{P}{\tan\theta}\right)\right)+\ln\left(\frac{1}{M(N-1)}+1\right)\right)+\ln\left(\sigma^2\right)}{\ln(2)}} \right).$$

Proof: See Appendix C.2.

We plot Fig. 8.2b according to (appendix C.2). We observe that there is a close match between the number of feedback bits B proposed by (appendix C.2) with increasing slope θ. Thus we conclude that, given slope θ, N, SNR and M, the analytical expression (appendix C.2) is able to predict the feedback bits B proposed by the analytical expression in (appendix C.2). This would bring a huge advantage in MU-MIMO WLAN settings in determining B. For example, in a MU-MIMO WLAN, already operating with a particular preset slope θ, there is a need for all joining heterogeneous antenna clients M in a network to have appropriate B in order to limit the quantisation error. In such a context, our

analytical expression (appendix C.2) can predict B for all M. The predicted B limits the quantisation error, i.e., $2^{-\left(\frac{B}{M(N-1)}\right)}$, for all M by maintaining $\alpha = \frac{B}{M}$ constant. This as a result ensures a constant average throughput loss w.r.t. CSIT irrespective of the number of antennas, M, at heterogeneous receivers.

8.3 Zero-forcing Relaxed with Quantisation Bits B and Distributive Algorithm

In this section, we first formulate an optimisation problem for finding the optimal ZF precoding vector considering the MU-MIMO WLAN settings. The constraint of the optimisation problem is relaxed with the channel quantisation bits B. Channel quantisation for MU-MIMO WLANs is discussed in the previous sections, and specifically the quantised bits B is meticulously studied in Section **??**. We then find the solution of the optimisation problem and calculate the optimal precoding vector. Second, we present a distributive algorithm for MU-MIMO WLANs which has two principal attributes: a) it takes into account interference leakage due to the channel quantisation bits B, b) it calculates the optimal precoding vector, which can be tuned with B.

8.3.1 Formulation

We recall the fact from (7.15) in the previous section that ZF beamforming will still have some residual interference because the beamformer is based on a quantised CSI which is not orthogonal to the physical channel (not the quantised channel). Mathematically, for any jth client, $\left\|\tilde{\mathbf{H}}_{ij}^{H}\hat{\mathbf{v}}_i\right\|^2 \neq 0$, $\forall j \neq i$. Similarly, for any ith client, $\left\|\tilde{\mathbf{H}}_{ji}^{H}\hat{\mathbf{v}}_j\right\|^2 \neq 0$, $\forall i \neq j$.

The ZF precoding vector at APs considering a FRF system is given by

$$\hat{\mathbf{v}}_j = \frac{\prod_{\hat{\mathbf{H}}_{ji}}^{\perp} \hat{\mathbf{H}}_{jj}}{\left\|\prod_{\hat{\mathbf{H}}_{ji}}^{\perp} \hat{\mathbf{H}}_{jj}\right\|}\mathbf{U}. \tag{8.1}$$

However, $\tilde{\mathbf{H}}_{ji}$ is the normalised real physical channel, which is not quantised. Although we consider block fading and assume that the channel does not change for the duration of the transmission, the orthogonality between $\tilde{\mathbf{H}}_{ji}$ and $\hat{\mathbf{v}}_j$ cannot be established, because $\tilde{\mathbf{H}}_{ji}$ is the normalised real channel but $\hat{\mathbf{v}}_j$ is calculated with quantised channels $\hat{\mathbf{H}}_{jj}$ and $\hat{\mathbf{H}}_{ji}$ and not with $\tilde{\mathbf{H}}_{jj}$ and $\tilde{\mathbf{H}}_{ji}$. Since $\prod^{\perp}_{\hat{\mathbf{H}}_{ji}}$ is at the orthogonal complement of the column space of $\hat{\mathbf{H}}_{ji}$, it is not orthogonal with $\tilde{\mathbf{H}}_{ji}$.

The quantisation error Z basically depends on B as per (7.2) and the expected value, i.e., $\mathbb{E}\left\|\tilde{\mathbf{H}}^H_{ij}\hat{\mathbf{v}}_i\right\|^2 = \left(\frac{1}{(N-1)}\right)2^{-\left(\frac{B}{M(N-1)}\right)}$, as per (7.18). Thus we form the interference leakage constraint of the optimisation problem as $\left\|\tilde{\mathbf{H}}^H_{ji}\hat{\mathbf{v}}_j\right\|^2 \neq \phi_{ji}$, where $\phi_{ji} = \left(\frac{1}{(N-1)}\right)2^{-\left(\frac{B}{M(N-1)}\right)}$ for all $i \neq j$ receivers.

The physical meaning of ϕ_{ji} is that it controls the level of interference from the jth AP to the ith client, where $i \neq j$. Also, we see that ϕ_{ji} is a function of B for given SNR, N and M as per (7.18). Thus, the amount of interference leakage can be controlled by B.

In this section we are interested in finding the optimal ZF precoding vector, which can be tuned with B. In order to find such an optimal ZF precoding vector, we relax the constraint of the optimisation problem to the interference leakage ϕ_{ji}, and formulate the optimisation problem as follows:

Problem 1: For each transmitter $j \in \{1, ..., K\}$,

$$\begin{aligned} \underset{\hat{\mathbf{v}}_j}{Max} \ & \log\left(1 + \frac{\gamma\|\mathbf{H}_{jj}\|^2\left\|\tilde{\mathbf{H}}^H_{jj}\hat{\mathbf{v}}_j\right\|^2}{\sigma_j^2 + \sum\limits_{i\neq j}^{K}\gamma\|\mathbf{H}_{ij}\|^2\left\|\tilde{\mathbf{H}}^H_{ij}\hat{\mathbf{v}}_i\right\|^2}\right) \\ s.t. \quad & \left\|\tilde{\mathbf{H}}^H_{ji}\hat{\mathbf{v}}_j\right\|^2 \leq \phi_{ji}\sigma_i^2 \quad \forall i \neq j \\ & \|\hat{\mathbf{v}}_j\|^2 \leq P_j \end{aligned} \tag{8.2}$$

where $\phi_{ji} = \mathbb{E}\left\|\tilde{\mathbf{H}}^H_{ji}\hat{\mathbf{v}}_j\right\|^2 = \left(\frac{1}{(N-1)}\right)2^{-\left(\frac{B}{M(N-1)}\right)}$ as per (7.18) is the interference leakage from the transmitter/AP j to the client i and each AP has a transmit power constraint, $\|\hat{\mathbf{v}}_j\|^2 \leq P_j$, $j = 1, ..., K$.

We assume that the residual interference due to the quantised CSI is greater than the power of the thermal noise at each client, i.e., for the jth client, $\gamma\|\mathbf{H}_{ij}\|^2 \left\|\tilde{\mathbf{H}}_{ij}^H \hat{\mathbf{v}}_i\right\| >> \sigma_j^2$.

As the logarithmic objective function complicates the solution, we use the monotonicity of the logarithm and convert Problem 1 into an equivalent problem, Problem 2, as follows:

Problem 2: For each transmitter $j \in \{1, ..., K\}$,

$$\begin{aligned} \underset{\hat{\mathbf{v}}_j}{Max} \quad & \left\|\tilde{\mathbf{H}}_{jj}^H \hat{\mathbf{v}}_j\right\|^2 \\ s.t. \quad & \left\|\tilde{\mathbf{H}}_{ji}^H \hat{\mathbf{v}}_j\right\|^2 \leq \phi_{ji}\sigma_i^2 \quad \forall i \neq j \\ & \|\hat{\mathbf{v}}_j\|^2 \leq P_j. \end{aligned} \tag{8.3}$$

In order to write an equivalent convex optimisation problem, Problem 3, for Problem 2, we make $\tilde{\mathbf{H}}_{jj}^H \hat{\mathbf{v}}_j$ real and non-negative [99] as follows:

Problem 3: For each transmitter $j \in \{1, ..., K\}$,

$$\begin{aligned} \underset{\hat{\mathbf{v}}_j}{Max} \quad & \tilde{\mathbf{H}}_{jj}^H \hat{\mathbf{v}}_j \\ s.t. \quad & \left\|\tilde{\mathbf{H}}_{ji}^H \hat{\mathbf{v}}_j\right\|^2 \leq \phi_{ji}\sigma_i^2 \quad \forall i \neq j \\ & \|\hat{\mathbf{v}}_j\|^2 \leq P_j. \\ & \tilde{\mathbf{H}}_{jj}^H \hat{\mathbf{v}}_j \geq \mathbf{0}. \end{aligned} \tag{8.4}$$

Since Problem 3 is equivalent to Problem 1 and Problem 2, we solve Problem 3 in the subsequent section to find the optimal ZF precoding vector which can be tuned with the given relaxed constraint.

8.3.2 The Optimal ZF Precoding Vector

We use Karush-Kuhun-Tucker (KKT) conditions and solve Problem 3 to find the optimal ZF precoding vector. The Lagrangian of Problem 3 can be written as

$$\begin{aligned} & L\left(\hat{\mathbf{v}}_j, \boldsymbol{\lambda}_j, \mu_{jj}\right) = \tilde{\mathbf{H}}_{jj}^H \hat{\mathbf{v}}_j - \textstyle\sum_{i=1, i\neq j}^{K} \lambda_{ji} \left(\left\|\tilde{\mathbf{H}}_{ji}^H \hat{\mathbf{v}}_j\right\|^2 - \phi_{ji}\sigma_i^2 \right) \\ & -\lambda_{jj}\left(\|\hat{\mathbf{v}}_j\|^2 - P_j\right) + \mu_{jj}\left(\tilde{\mathbf{H}}_{jj}^H \hat{\mathbf{v}}_j\right) \end{aligned} \tag{8.5}$$

where $\boldsymbol{\lambda}_j = [\lambda_{j1}, ..., \lambda_{jK}] \geq \mathbf{0}$ and $\mu_{jj} \geq 0$ are non-negative dual variables for Problem 3.

In order to calculate the optimal ZF precoding vector $\hat{\mathbf{v}}_j^{opt}$, we take the gradient of the Lagrangian and equate it to zero. However, the Lagrangian function in (8.5) is a complex-valued function composed of complex variables such as $\tilde{\mathbf{H}}_{ji}$ and $\hat{\mathbf{v}}_j$. Thus we need to take a complex differential of the complex-valued Lagrangian function.

Lemma 3: Given the Lagrangian function in (8.5), the function is not complex differential and its derivative cannot be taken.

Proof: See Appendix C.3.

Thus, in order to calculate the optimal ZF precoding vector $\hat{\mathbf{v}}_j^{opt}$, we take the complex/Wirtinger gradient of the Lagrangian and equate it to zero. In this case, the dual variables $\boldsymbol{\lambda}_j$, μ_{jj} are optimal, i.e., $\boldsymbol{\lambda}_j^*, \mu_{jj}^*$. Thus we have

$$\begin{aligned} \nabla_{\mathbf{v}_j^*} L\left(\hat{\mathbf{v}}_j, \boldsymbol{\lambda}_j^*, \mu_{jj}^*\right) \Big|_{\hat{\mathbf{v}}_j = \mathbf{v}_j^{opt}} &= 0 \\ &= \tilde{\mathbf{H}}_{jj} - \textstyle\sum_{j=1, j\neq i}^{K} \lambda_{ji}^* \tilde{\mathbf{H}}_{ji} \tilde{\mathbf{H}}_{ji}^H \mathbf{v}_j^{opt} \\ &\quad -\lambda_{jj}^* \mathbf{v}_j^{opt} + \mu_{jj}^* \mathbf{H}_{jj} \end{aligned} \tag{8.6}$$

where $\nabla_{\hat{\mathbf{v}}_j^*}$ is a Wirtinger gradient[1]. The complementary slackness theorem for Problem 3 shows that $\lambda_{ji}\left(\left\|\tilde{\mathbf{H}}_{ji}^H \hat{\mathbf{v}}_j\right\|^2 - \phi_{ji}\sigma_i^2\right) = 0$, $\forall i \neq j$, $\lambda_{jj}\left(\|\hat{\mathbf{v}}_j\|^2 - P_j\right) = 0$ and $\mu_{jj}\left(\tilde{\mathbf{H}}_{jj}^H \hat{\mathbf{v}}_j\right) = 0$.

For the optimality condition, $\boldsymbol{\lambda}_j^* > \mathbf{0}$ and $\mu_{jj}^* > 0$ should be true. Thus, when $\lambda_{ji}^* > 0$,

[1]The Wirtinger gradient is an alternative formulation to calculate the gradients of real value functions defined in complex domains. These non-holomorphic functions do not hold to the Cauchy Riemann (CR) conditions. However, if the function f is differentiable in the real sense, a form of Taylor's series expansion can still be found [100] [101, p.6-12]. Note that whenever the CR conditions are satisfied for f, the Wirtinger derivative degenerates to the standard complex derivative.

$\left(\left\|\tilde{\mathbf{H}}_{ji}^{H}\hat{v}_j\right\|^2\right) = \phi_{ji}\sigma_i^2$. This means that the interference leakage from the jth AP to the ith client is equivalent to $\left(\frac{1}{(N-1)}\right)2^{-\left(\frac{B}{M(N-1)}\right)}$. Similarly, when $\lambda_{jj}^* > 0$, $\left(\|\hat{\mathbf{v}}_j\|^2\right) = P_j$. This means that the jth AP uses full power P_j. Likewise, when $\mu_{jj}^* > 0$, $\left(\tilde{\mathbf{H}}_{jj}^{H}\hat{\mathbf{v}}_j\right) = \mathbf{0}$. However, the physical meaning of $\left(\tilde{\mathbf{H}}_{jj}^{H}\hat{\mathbf{v}}_j\right) = \mathbf{0}$ is that jth client has no data rate. Since this is not the case in MU-MIMO WLANs, we have $\mu_{jj}^* = 0$ and $\tilde{\mathbf{H}}_{jj}^{H}\hat{\mathbf{v}}_j > \mathbf{0}$. Thus we finally get

$$\hat{\mathbf{v}}_j^{opt} = \left(\sum_{i=1,\, i\neq j}^{K} \lambda_{ji}^*\tilde{\mathbf{H}}_{ji}\tilde{\mathbf{H}}_{ji}^{H} + \lambda_{jj}^*\mathbf{I}\right)^{-1}\tilde{\mathbf{H}}_{jj}. \tag{8.7}$$

We define $\mathbf{S} = \left(\sum_{i=1,\, i\neq j}^{K} \lambda_{ji}^*\tilde{\mathbf{H}}_{ji}\tilde{\mathbf{H}}_{ji}^{H} + \lambda_{jj}^*\mathbf{I}\right)$ which should be non-singular to calculate $\hat{\mathbf{v}}_j^{opt}$, i.e., $\mathbf{S}$ should have a full rank. In a rich scattering and distributive MU-MIMO WLAN environment, channels are highly uncorrelated and a full rank requirement for $\mathbf{S}$ is almost surely satisfied. However, $\mathbf{S}$ varies with the optimal dual variable λ_{jj} in two respects:

a) when $\lambda_{jj}^* > 0$, $\mathbf{S} = \left(\sum_{i=1,\, i\neq j}^{K} \lambda_{ji}^*\tilde{\mathbf{H}}_{ji}\tilde{\mathbf{H}}_{ji}^{H} + \lambda_{jj}^*\mathbf{I}\right)$ and $\hat{\mathbf{v}}_j^{opt}$ can be calculated as in (8.7).

b) when $\lambda_{jj}^* = 0$ then $\left\|\mathbf{v}_j^{opt}\right\|^2 < P_j$, meaning that full power P_j is not used at the jth AP and we have $\mathbf{S} = \left(\sum_{i=1,\, i\neq j}^{K} \lambda_{ji}^*\tilde{\mathbf{H}}_{ji}\tilde{\mathbf{H}}_{ji}^{H}\right)$. In this case, in order to find the inverse of $\mathbf{S}$, $N < K$. This is a condition when $\lambda_{ji} > 0$, $\forall i \neq j$ and $\phi_{ji} = \left(\frac{1}{(N-1)}\right)2^{-\left(\frac{B}{M(N-1)}\right)}$ i.e., the interference constraint is tight regardless of whether the transmit constraint for the jth AP is tight in Problem 3. This case can be true only when $N < K$.

The optimal ZF precoding vector $\hat{\mathbf{v}}_j^{opt}$ in (8.7) depends on the calculation of the optimal dual variable $\boldsymbol{\lambda}_j$, where $\boldsymbol{\lambda}_j = [\lambda_{j1}, \lambda_{j2}, ...\lambda_{jj-1}, \lambda_{jj}, \lambda_{jj+1}, ..., \lambda_{jK}] > \mathbf{0}$. However for $\boldsymbol{\lambda}_j > \mathbf{0}$, the maximum value can go up to $+\infty$. In order to find the greatest lower bound of $\boldsymbol{\lambda}_j$ for finding the least upper bound of $\hat{\mathbf{v}}_j^{opt}$, we need to form a dual problem to Problem 3.

The dual function of Problem 3 is given by

$$g\left(\boldsymbol{\lambda}_j\right) = \max_{\mathbf{v}_j} L\left(\hat{\mathbf{v}}_j, \boldsymbol{\lambda}_j, \mu_{jj}\right) \tag{8.8}$$

Thus the dual problem of Problem 3 is given by

$$\min_{\boldsymbol{\lambda}_j \geq 0} g\left(\boldsymbol{\lambda}_j\right) \tag{8.9}$$

where $\boldsymbol{\lambda}_j \geq \mathbf{0}$ is component-wise non-negative.

Since Problem 3 is convex and have strictly feasible points, Slater's condition for strong duality holds [102]. Thus, the duality gap between the optimal value of the primal and dual problem, is zero. This means that the primal Problem 3 can be solved by solving its dual problem in (8.9). In other words, just as we can find the optimal dual variable $\boldsymbol{\lambda}_j$ by solving (8.9), we can find the optimal ZF precoding vector $\hat{\mathbf{v}}_j^{opt}$ in (8.7).

8.3.3 Sub-gradient Based Solution for Dual Problem

From (8.9), the objective function of the dual problem is discrete and piecewise linear. This may not be differential at certain points of $\boldsymbol{\lambda}$ space. We use the sub-gradient method to solve the dual problem, which finds the optimal solution of the dual problem by an iterative method. In this method the Lagrange multipliers are updated according to the degree of constraint violation. We start with $\boldsymbol{\lambda}_j(0)$ and the sequence of $\boldsymbol{\lambda}_j(k)$ eventually converges to the optimal solution according to

$$\boldsymbol{\lambda}_j(k+1) = \boldsymbol{\lambda}_j(k) + \alpha(k)\boldsymbol{g}(k) \tag{8.10}$$

where k is the iteration index. $\alpha(k)$ is the step size satisfying $\sum_{i=0}^{\infty} \alpha^2\left(k\right) < \infty, \sum_{i=0}^{\infty} \alpha\left(k\right) < \infty$ and $\boldsymbol{g}(k)$ is the sub-gradient of Problem 3 at $\boldsymbol{\lambda}(k)$ given by $\boldsymbol{g}\left(k\right) = \left\|\tilde{\mathbf{H}}_{ji}\hat{\mathbf{v}}_j\right\|^2 - \phi_{ji}$. The update in the value of $\boldsymbol{\lambda}$ stops when the change of $\boldsymbol{\lambda}$ value is very small, i.e., $\boldsymbol{\lambda} < \xi$. ξ is a preset system parameter.

Since the optimal dual variable, $\boldsymbol{\lambda}^*$, can be obtained from the sub-gradient method as shown in (8.10), the optimal ZF precoding vector $\hat{\mathbf{v}}_j^{opt}$ can be calculated from (8.7).

8.3.4 Distributive Algorithm for Optimal ZF Precoding Vector Tuned with Quantised Bits B

Our distributive algorithm for calculating the optimal ZF precoding vector tuned with quantised bits B stands on two fundamentals: a) the relation of the quantisation error, Z, with quantised bits B as shown in (7.12), and b) the fact that $\left\|\tilde{\mathbf{H}}_{ji}^H \hat{\mathbf{v}}_j\right\| \neq 0$, $\forall i \neq j$, irrespective of any channel assumptions we make, as discussed in Subsection 8.3.1. Also, we have taken into account the role of Z in the throughput loss w.r.t. perfect CSIT for ZF as shown in (7.21).

The channel quantisation error produces an error offset of $2^{\frac{B}{M(N-1)}}$ in MU-MIMO WLANs, which basically depends on B, given N and variable M. The error offset reduces the average throughput loss (of the ZF transmission strategy) w.r.t perfect CSIT as shown in (7.21). Specifically, the less the quantisation error the closer the throughput gain w.r.t. perfect CSI.

Moreover, orthogonality is the basis of ZF, however, we see that $\left\|\tilde{\mathbf{H}}_{ji}^H \hat{\mathbf{v}}_j\right\| \neq 0$, $\forall i \neq j$ for any channel conditions in (7.15). This is because $\hat{\mathbf{v}}_j$ projects $\hat{\mathbf{H}}_{jj}$ into the complementary orthogonal column space of $\hat{\mathbf{H}}_{ji}$ and not that of $\tilde{\mathbf{H}}_{ji}$. Thus, there is no orthogonality between $\tilde{\mathbf{H}}_{ji}$ and $\hat{\mathbf{v}}_j$.

However, the expected value, i.e., $\mathbb{E}\left\|\tilde{\mathbf{H}}_{ij}^H \hat{\mathbf{v}}_i\right\|^2$ is equate to $\left(\frac{1}{(N-1)}\right) 2^{-\left(\frac{B}{MN-1}\right)}$, $\forall M > 1$, as shown in (7.18). Thus we relaxed the constraints of the optimisation problem, Problem 3, as $\left\|\tilde{\mathbf{H}}_{ji}^H \hat{\mathbf{v}}_j\right\|^2 \leq \phi_{ji}$, where, $\phi_{ji} = \left(\frac{1}{N-1}\right) 2^{-\left(\frac{B}{M(N-1)}\right)}$. The transmit power constraint is $\|\mathbf{v}_j\|^2 \leq P_j$, where P_j is the total power at APs/transmitter. We then solve the optimisation problem in order to find the optimal precoding vector $\hat{\mathbf{v}}_j^{opt}$. Since the constraint of Problem 3 is dependent on B, we can use B to determine the error offset

and the average throughput loss, and find the corresponding $\hat{\mathbf{v}}_j^{opt}$ for MU-MIMO WLANs. The distributive algorithm is presented in **Algorithm 3**.

Algorithm 3 Distributive Algorithm for APs in a FRF based MU-MIMO WLAN.

1: **procedure** SEARCH FOR $\boldsymbol{\lambda}_j$ BY SUB-GRADIENT METHOD

2: Initialize: $k = 1$, $\boldsymbol{\lambda}(k) > \mathbf{0}$,
$$\hat{\mathbf{v}}_j^k = \left(\sum_{i=1,\, i\neq j}^{K} \lambda_{ji}^k \tilde{\mathbf{H}}_{ji} \tilde{\mathbf{H}}_{ji}^H + \lambda_{jj}^k \mathbf{I} \right)^{-1} \tilde{\mathbf{H}}_{jj}, \text{ and } \xi > 0.$$

3: Determine: Feedback bits B, where $B = f(\theta)$ and $\theta = \tan^{-1}\left(\frac{d\Delta R(P)}{dB}\right)$. $\Delta R(P)$ is defined by B.

4: Estimate: $\mathbb{E}\left\|\tilde{\mathbf{H}}_{ji}^H \hat{\mathbf{v}}_j\right\|^2 = \phi_{ji}$, for B, set in Step 3, where $\phi_{ji} = \mathbb{E}\left\|\tilde{\mathbf{H}}_{ji}^H \hat{\mathbf{v}}_j\right\|^2 = \left(\frac{1}{N-1}\right) 2^{-\left(\frac{B}{M(N-1)}\right)}$, $\forall i \neq j$ and $M \geq 1$.

5: Update: $\boldsymbol{\lambda}(k+1) = \boldsymbol{\lambda}(k) + \alpha(k)\, \boldsymbol{g}(k)$, where $\boldsymbol{g}(k) = \left\|\tilde{\mathbf{H}}_{ji}^H \hat{\mathbf{v}}_j\right\|^2 - \phi_{ji}$ and $\alpha(k)$ is any step size satisfying: $\sum_{i=0}^{\infty} \alpha^2(k) < \infty, \sum_{i=0}^{\infty} \alpha(k) = \infty$.

6: **if** $(\boldsymbol{\lambda}(k+1) - \boldsymbol{\lambda}(k)) < \xi$ **then**, go to Step 12.

7: **else**

8: $\hat{\mathbf{v}}_j^{k+1} = \left(\sum_{i=1,\, i\neq j}^{K} \lambda_{ji}^{k+1} \tilde{\mathbf{H}}_{ji} \tilde{\mathbf{H}}_{ji}^H + \lambda_{jj}^{k+1} \mathbf{I} \right)^{-1} \tilde{\mathbf{H}}_{jj}$.

9: Val_optimised=$\tilde{\mathbf{H}}_{jj}^H \hat{\mathbf{v}}_j^{k+1}$.

10: **end if**

11: go to step 5.

12: Terminates with $\hat{\mathbf{v}}_j^k$ for the kth iteration.

13: **end procedure**

The distributive algorithm for MU-MIMO WLANs works as follows: First, the number of quantisation bits B is set for the network. This determines the error offset and the average throughput loss w.r.t. CSIT. This means that B determines how low from the perfect CSIT the average throughput of the heterogeneous clients in MU-MIMO WLANs are performing. In other words B decides the gap, $\Delta R(P) = C_{CSIT} - C_{FRF}$.

Note that the gap in the performance is irrespective of the number of antennas at heterogeneous clients in MU-MIMO WLANs, because appropriate B for variable M can be predicted by (appendix C.2), which keeps the quantisation error constant for all clients.

For instance, consider $B = 28$ bits for a MU-MIMO WLAN. Then a single-antenna client, $M = 1$, uses all $B = 28$ bits for channel quantisation, whereas a double-antenna client, $M = 2$, uses (appendix C.2) and predicts an appropriate B for quantisation as shown in Fig. 8.2b. This keeps the quantisation error constant for both $M = 1$ and $M = 2$.

Second, APs/transmitters in the network need the Lagrange multipliers $\boldsymbol{\lambda}_j$ (as per Step 8 according to the distributive algorithm) to calculate the optimal ZF precoding vector for serving clients' requests in the network. We use an iterative method to find $\boldsymbol{\lambda}_j$ as in Steps 5-11. The optimal precoding vector $\hat{\mathbf{v}}_j^{opt}$ and hence the optimised values can be calculated in Step 8 and Step 9 respectively.

Third, any changes in the system parameters are properly addressed. For instance, any change in M, due to any multi-antenna heterogeneous clients, is addressed with (appendix C.2) by predicting an appropriate B. This helps the heterogeneous clients to operate with the same ΔR (w.r.t. perfect CSIT) as a client with $M = 1$.

When the set system parameter B is changed, resulting in a change of the offset value w.r.t perfect CSIT, Step 4 in the distributive algorithm changes the offset. For that particular offset, the system works as usual. The choice of B and its implications are discussed in Section 8.2.

8.4 Simulation Results

In this section, we present the simulation results of our proposed distributive algorithm. We set up a MU-MIMO WLAN system with $N = 6$ antennas at AP, with a number

of clients $K = 4$ having $M = 1$ receiving antennas. Since a single client is considered as a desired client, we have $K - 1$ inequalities pertaining to interference leakage. The numbers of feedback bits are $B = 5$, 10, 15, 20, 25, 30. The number of iterations is 50 and the small positive incremental value is $\xi = 0.003$. The initial values of the Lagrange multipliers λ_{j1}, λ_{j2} and λ_{j3} are all 0.5. For λ_{jj}, we take two approaches: a) λ_{jj} as the ratio of the maximum power P_j and $\|\mathbf{v}_j\|^2$ i.e., $\lambda_{jj} = \frac{P_j}{\|\mathbf{v}_j\|^2} = 5$, b) similar to the interference leakage constraint, we take the power gradient, i.e., $\|\hat{\mathbf{v}}_j\|^2 - P_j$ and update λ_{jj} with the iterations. In this case, we set a maximum power of AP, $P_j = 27$ dBm, and the initial value of $\lambda_{jj} = 0.5$.

For $\lambda_{jj} = \frac{P_j}{\|\mathbf{v}_j\|^2} = 5$, we observe that in Fig. 8.3a and 8.3b, with an increasing number of iterations k, the value of the objective function $\mathbf{H}_{jj}^H \hat{\mathbf{v}}_j$, (regarded as optimal values given the optimisation problem) for different feedback bits B, (arbitrary values of B, taken as 5, 10, 15, 20, 25, 30, bits respectively), become asymptotic with the optimised values corresponding to $B = 30$ bits.

This is an expected behavior. Since our iterative distributive algorithm updates λ_{ji} in each iteration according to the pre-defined step length, the precoding vector $\hat{\mathbf{v}}_j$ approaches the optimal precoding vector, $\hat{\mathbf{v}}_j^{opt}$, gradually.

This continuously updates the optimised values $\tilde{\mathbf{H}}_{jj}^H \hat{\mathbf{v}}_j$. As λ_{ji} approaches λ_{ji}^*, $\hat{\mathbf{v}}_j$ approaches $\hat{\mathbf{v}}_j^{opt}$, meaning that the value of the objective function $\mathbf{H}_{jj}^H \hat{\mathbf{v}}_j$ is optimised, given $\lambda_{jj} = \frac{P_j}{\|\mathbf{v}_j\|^2} = 5$. However, since the numbers of feedback bits B considered are different, i.e., B=5, 10, 15, 20, 25 and 30, the values of the inequality constraint, $\phi_{ji} = \left(\frac{1}{N-1}\right) 2^{-\left(\frac{B}{M(N-1)}\right)}$, associated with each B are different, given fixed N and M. The variation in ϕ_{ji} for each B results in different $\hat{\mathbf{v}}_j$. This makes the optimal values asymptotic with with each other and with B=30 as iteration proceeds.

More precisely, for instance, consider $\lambda_{jj} = \frac{P_j}{\|\mathbf{v}_j\|^2} = 5$ and $B = 5$ bits. With increasing iterations, k, λ_{ji} is updated from its initial assumed arbitrary value of 0.5, as per Step

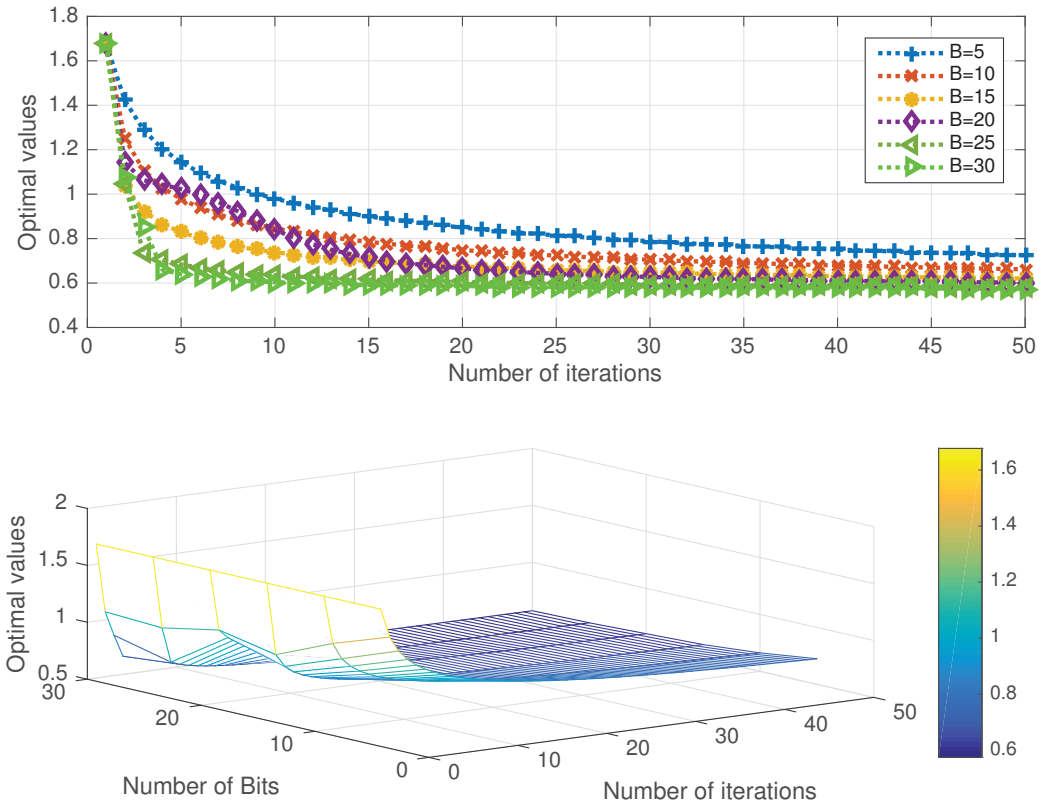

Figure 8.3: a) Optimal values vs Number of iterations with varying feedback bits B, when $\lambda_{jj} = \frac{P_j}{\|\mathbf{v}_j\|^2} = 5$; b) 3D view showing Optimal values, Number of iterations and Number of feedback bits B, when $\lambda_{jj} = \frac{P_j}{\|\mathbf{v}_j\|^2} = 5$.

5 in the algorithm. Accordingly, the precoding vector $\hat{\mathbf{v}}_j$ is calculated with the updated value λ_{ji}, and with $\lambda_{jj} = \frac{P_j}{\|\mathbf{v}_j\|^2} = 5$ as per Step 8.

As the process advances to a preset number of iterations, say $k = 50$ in our simulation, λ_{ji} is continuously updated according to Step 5.

Apparently, λ_{ji} converts the first inequality constraint, $\left\|\mathbf{H}_{ji}^H \hat{\mathbf{v}}_j\right\|^2 \leq \phi_{ji}$, of the optimisation problem, Problem 3, to the equality constraint, i.e., $\left\|\mathbf{H}_{ji}^H \hat{\mathbf{v}}_j\right\|^2 = \phi_{ji}$, with $\lambda_j{=}\lambda_j^*$, where $\phi_{ji} = \left(\frac{1}{N-1}\right) 2^{-\left(\frac{B}{M(N-1)}\right)}$. The corresponding precoding vector at λ_j^* is the optimal

precoding vector, $\hat{\mathbf{v}}_j^{opt}$, for $B = 5$ bits and for $\lambda_{jj} = \frac{P_j}{\|\mathbf{v}_j\|^2} = 5$. The value of the objective function is optimised. A similar process follows for the other numbers of feedback bits B, and their corresponding optimised values are calculated accordingly. However since the values of the inequality constraints, ϕ_{ji}, are different for different B for given N and M, the corresponding optimised values become asymptotic with each other and with B=30 bits, as iteration proceeds.

Notably, the values of the objective function for different B (other than $B = 30$ bits) do not approach the values with $B = 30$ bits, but become asymptotic with them. The asymptotic nature is because: $B = 30$ is the maximum feedback bits considered in the simulation and the average throughput associated with it is relatively closer to that of the perfect CSIT [2]. Thus this forms the boundary limits for the values of the objective function, irrespective of the optimised values calculated for different B with our distributive algorithm.

Put another way, as B increases $\phi_{ji} = \left(\frac{1}{N-1}\right) 2^{-\left(\frac{B}{M(N-1)}\right)}$ decreases for given N and M. When $\phi_{ji} = \mathbf{0}$, we have the perfect CSIT. Hence the throughput associated with $B = 30$, is relatively closer to the perfect CSIT and forms the boundary limits for the optimised values of the objective function lower than that of $B = 30$ bits. These boundary values are irrespective of the optimised values calculated for different B with our distributive algorithm.

The updated values of the Lagrange multipliers, λ_{j1}, λ_{j2} and λ_{j3} for arbitrary feedback bits $B = 5$, 10, 15 bits respectively, are shown in Fig. 8.4.

Unlike fixed $\lambda_{jj} = \frac{P_j}{\|\mathbf{v}_j\|^2} = 5$ in Fig. 8.3a and Fig. 8.3b, Fig. 8.5a and Fig. 8.5b show the optimised value for different B with increasing iterations. The power constraint,

[2]This is due to the quantisation error, i.e., $2^{-\left(\frac{B}{M(N-1)}\right)}$ in (7.21), that makes $\Delta R(P)$ smaller and narrows the gap with CSIT. Fig. 8.1 and the discussion therein (i.e., if $B_2 > B_1$, then $\theta_2 < \theta_1$) thus follows.

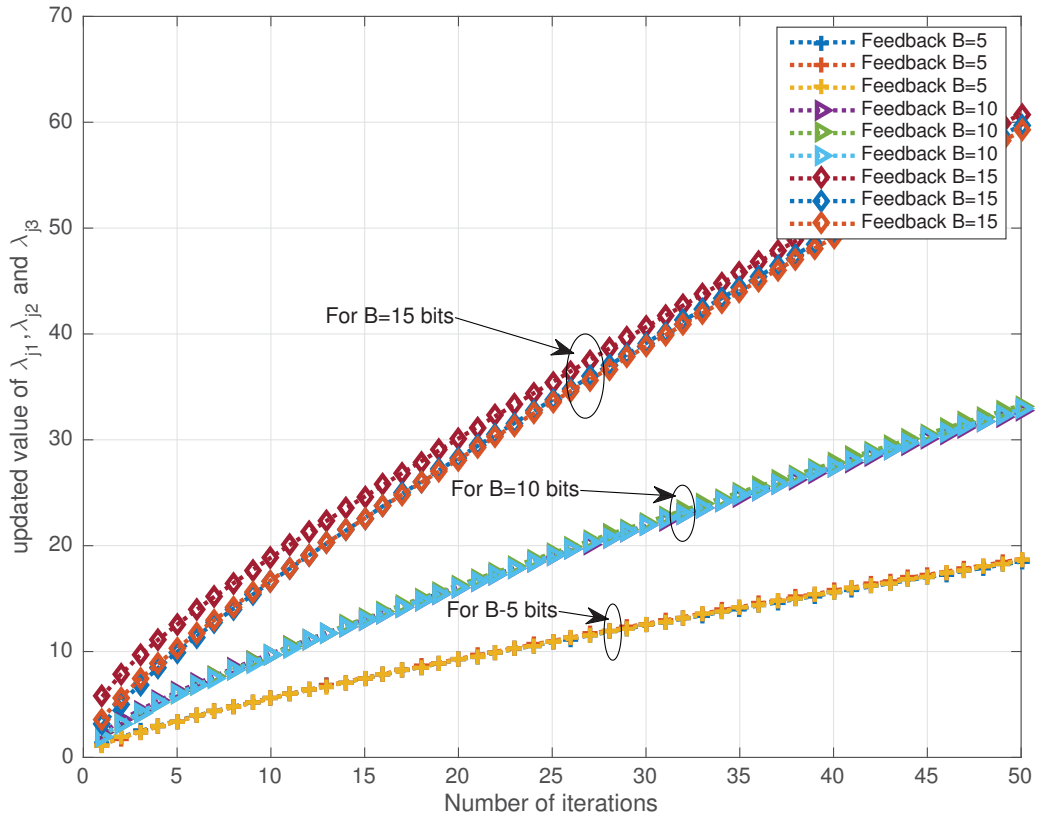

Figure 8.4: Values of λ_{j1}, λ_{j2} and λ_{j3} for feedback bits $B = 5$, 10, 15 bits respectively, when $\lambda_{jj} = \frac{P_j}{\|\mathbf{v}_j\|^2} = 5$.

$\|\mathbf{v}_j\|^2 \leq P_j$, is addressed by updating λ_{jj}, similar to with the interference constraints as done by λ_{ji}, for the optimisation problem, Problem 3. The total power considered is 27 dBm and the initial value for λ_{jj} is 0.5.

Specifically, we study the effect of both the interference constraint ϕ_{ji} and the total power constraint P_j, for the optimisation problem, Problem 3, in calculating the optimal precoding vector $\mathbf{v}_j^{opt}$ and hence the optimal values in Fig. 8.5a and Fig. 8.5b. Thus we update λ_{ji} continuously in each iteration, as shown in Fig. 8.6a and λ_{jj} as shown in Fig.

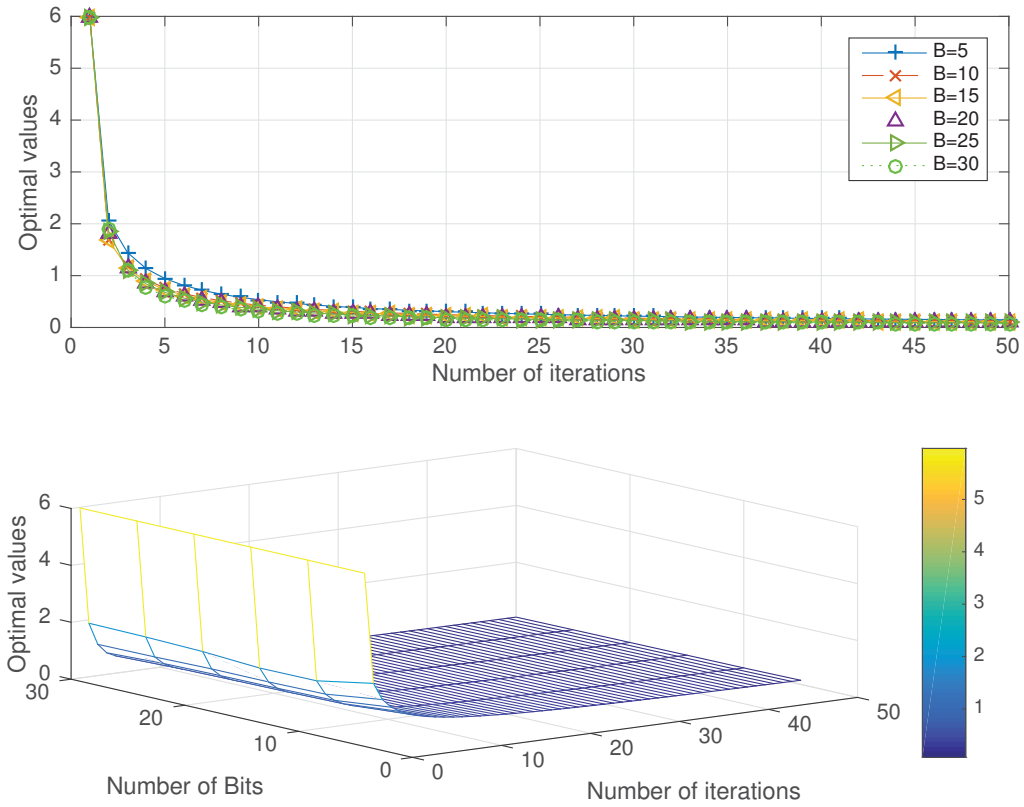

Figure 8.5: a) Optimal values vs Number of iterations with varying feedback bits B, when λ_{jj} uses power gradient i.e., $\|\hat{\mathbf{v}}_j\|^2 - P_j$, b) 3D view showing Optimal values, Number of iterations and Number of feedback bits B, when λ_{jj} uses power gradient i.e., $\|\hat{\mathbf{v}}_j\|^2 - P_j$.

8.6b, for the calculation of the optimal precoding vector[3] .

The asymptotic characteristic for the optimised values for different B in Fig. 8.5a and Fig. 8.5b is as expected. Since the total power constraint $P_j = 27$ dBm, is addressed by updating λ_{jj}, the asymptotic characteristic of the optimised value in Fig. 8.5a and Fig.

[3]In this context, we can thus consider the results in Fig. 8.3a and Fig. 8.3b as one of the optimised objective function cases when $\lambda_{jj} = 5$.

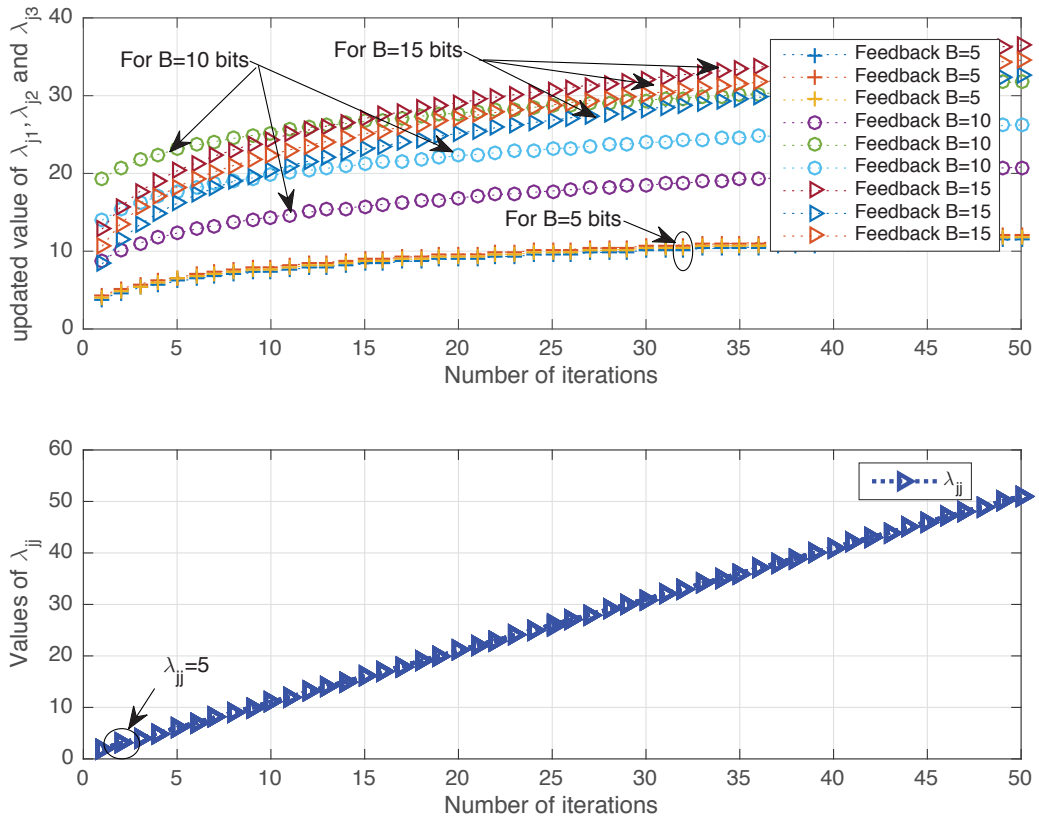

Figure 8.6: a) Values of λ_{j1}, λ_{j2} and λ_{j3} for feedback bits $B = 5$, 10, 15 bits respectively, when λ_{jj} uses the power gradient, i.e., $\|\hat{\mathbf{v}}_j\|^2 - P_j$, b) Values of λ_{jj} when λ_{jj} uses the power gradient, i.e., $\|\hat{\mathbf{v}}_j\|^2 - P_j$.

8.5b is explained by the different values of B that are used in channel quantisation.

Due to the range of the values of B (5, 10, 15, 20, 25 and 30 bits), the interference constraint ϕ_{ji} for each B is different. With the difference in the interference constraints ϕ_{ji}, the precoding vector $\hat{\mathbf{v}}_j$ is different and hence the optimised values. With increasing iterations, they become asymptotic with each other as seen in Fig. 8.5a and Fig. 8.5b.

In summary, the asymptotic characteristic of the optimal values is expected, as it explains the a varying value of the optimal precoding vector, with varying value of ϕ_{ji}, according to varying B. Thus our distributive algorithm is able to calculate the optimal precoding vector and hence the optimal values tuned with B.

8.5 Summary

In this chapter, we consider a FRF-based MU-MIMO WLAN with ZF downlink transmission strategy. We propose a new method to calculate the optimal ZF precoding vector, which relaxes the interference constraint to a pre-determined level, namely, the upper bound of the residual interference, $\frac{1}{(N-1)}2^{\left(\frac{B}{M(N-1)}\right)}$, $\forall M \geq 1$. We quantify the upper bound of both the quantisation error and the residual interference and find the optimal precoding vector that can be tuned with quantisation bits B, given, N, M and SNR for MU-MIMO WLANs. Additionally, we derive a closed-form expression which can predict B for heterogeneous antenna clients, i.e., for variable M, but produces a constant quantisation error in a MU-MIMO WLAN setting. Our results lead to a distributive algorithm for APs which runs independently to obtain the ZF precoding vector for downlink transmission.

The number of quantisation bits B is the key for representing channels in a quantised form for an FRF system. Although B is discussed in terms of θ and the corresponding B can be found given θ, where θ can go to any extent closer to the perfect CSIT, the

appropriate choice of B that can maximize the rate of the system to achieve the capacity of the MIMO downlink channel is still unknown. In other words, the choice of appropriate B that produces a suitable residual constraint, ϕ_{ji}, which leads to a point in the pareto optimal boundary, is a challenging unsolved information-theoretic problem.

Chapter 9

Conclusions and Future Work

9.1 Conclusions

This thesis studies the decades long HT problem in the context of MU-MIMO based WLANs and proposes a solution that has following key attributes: a) the downlink transmission strategy used is the ZF technique, which is further developed as a fairness and network throughput-aware ZF precoding, b) the medium access mechanism used is based on DoF of APs, which is backed by the extended PCF, c) channel acquisition is done explicitly by an enhanced channel sounding framework, d) a hybrid scheduling scheme with packet position-based FIFO and channel quality-based scheme, namely the Best of the Two Choices is used, f) performance evaluation tools are testbed and simulation results.

The proposed solution is feasible in existing hardware as our feasibility study of the PHY layer solution in the laboratory setting shows a significant received SNR gain of about 10 dB compared to the HT scenario. Furthermore, our simulation study of the MAC layer shows: a) a reduced signalling time overhead of 98.6 μs compared to IEEE802.11ac, b) a constant network throughput gain of 4-5 times w.r.t. the popular RTS/CTS solution. The proposed solution is interoperable with legacy standards as standard frame formats

are used except for some logical changes that are required. This chapter summarises the research work, discusses the open issues and presents avenues for future research.

The HT problem is inevitable in a distributive, decentralised and densely-deployed WLAN, which causes collisions of signals and degrades the network throughput. This thesis extensively investigates various solutions to the HT problem in different settings. The pioneering Busy-tone solution, the popular RTS/CTS solution, their various derivatives and MIMO based MAC solutions in recent years are studied with their pros and cons.

It is observed that the HT problem, which fundamentally arises due to the transmission time overlaps between different transmitters, resulting in collisions of signals, can be dealt with by an effective utilisation of the time and spatial domains, provided by the use of additional antennas in MU-MIMO WLANs. Thus this thesis proposes that the ZF technique at PHY layer backed with an appropriate MAC layer design can be used to solve the HT problem in MU-MIMO WLANs. Having said that, the application-specific requirements of MU-MIMO WLANs cannot be overlooked. Thus Chapter 3 illustrates the specific requirements of MU-MIMO WLANs in the context of using the ZF technique and discusses the system model, design opportunities and implications for design. The major focus of Chapter 3 is on the illustration of the need for a ZF precoding vector which can: a) cancel interference to undesired clients; b) maintain both fairness and network throughput while addressing the HT problem in MU-MIMO WLANs.

One of the fundamental aspects of the design in the PHY layer is presented. Fairness and network throughput aware precoding is proposed at the PHY layer in the context of the specific requirements highlighted in Chapter 3. The steps for such a precoding vector are meticulously illustrated. The vital steps, such as channel acquisitions, TXOP decisions for APs, steps for precoding vector calculation, the concurrent transmission algorithm and its performance, are rigorously detailed. Since the major focus is on the

fairness in client access and network throughput, emphasis is given to the design of a concurrent transmission algorithm, which possesses attributes to balance both fairness and throughput of network. A hybrid approach, for the design of the concurrent transmission algorithm, is taken, and adopts the packet-based FIFO for ensuring fairness in client access and the channel quality based scheme, the Best of the Two Choices, for maximising system throughput.

Simulation studies show that the performance of the concurrent transmission algorithm lies between the Brute-force and the FIFO approaches. The concurrent transmission algorithm is an important component for the calculation of the ZF precoding vector. A fundamental PHY layer design is presented with a block diagram in Chapter 4.

An experimental prototype in a laboratory setting is presented to demonstrate that the ZF technique is feasible and effective in addressing the HT problem in MU-MIMO WLANs. The experimental prototype is set up at the Wireless Research Center at Macquaire University, and principally consists of the USRP2, RFX2400 daughter boards, Jackson Labs equipment along with GPS antennas and PCs. The operating frequency is 2.45 GHz, and other PHY layer parameters including prototype settings are presented in detail in Chapter 5. An HT scenario is created in the laboratory setting and the performance of the ZF technique is investigated. The experimental results show that there is about 10 dB SNR gain w.r.t. the HT scenario when the ZF technique is used to address the HT problem.

A mechanism that governs the TXOP decisions for APs/transmitters is critical in order to reap the advantages of the PHY layer design presented in Chapters 4 and 5. Thus a DoF based TXOP at the MAC layer is proposed which fundamentally allows APs to transmit only after: a) satisfying a required DoF to cancel interference to undesired clients; b) ensuring at least one concurrent transmission. In order to prevent APs with less DoF from starving for TXOPs, a simple counter-based fairness algorithm is presented

which runs independently at APs and possesses a Jain fairness index of as much as 90%.

Furthermore, channel acquisition with reduced signalling overheads by 98.6 μs compared to IEEE802.11ac is developed. The performance comparison of our MAC layer protocol to address the HT problem w.r.t. the traditional RTS/CTS shows that our design can achieve a constant network capacity gain of 4-5 times than RTS/CTS. Additionally, the transition mechanism of the design, for any standard MAC design that uses standard frame formats, is presented. The transition condition, transition steps and transition frame formats are detailed. Overall, a design from the MAC layer perspective is presented based on the PHY layer design.

Practical dimensions are considered on top of the solution proposed at PHY and MAC layers in Chapters 4, 5 and 6 respectively. Since the ZF technique is used at the PHY layer, knowledge of CSI at transmitters/APs is a prerequisite. However, owing to practical constraints such as channel estimation errors at receivers/clients, limited bandwidth of the feedback path, delay in feedback path, cost of the feedback overheads, etc., it is not practically possible to obtain the perfect CSI at APs. This may be a critical question from the practical perspective when the design is fundamentally based on the ZF technique. This bottleneck of the design is addressed with the partial CSI/FRF model, a researched topic in the field of cellular communication systems, in which a finite number of quantised bits represent CSI, and the effects on quantisation in the system parameters are analysed.

However, the context of a MU-MIMO WLAN is different, for example, involving heterogeneous antenna clients, fairness and throughput requirements, etc., which result in constraints different to those of cellular systems. Thus, the FRF model of CSI feedback is further developed to make it appropriate for MU-MIMO WLAN settings. The system parameters such as quantisation error bounds, throughput loss w.r.t. perfect CSI and appropriate numbers of feedback bits are analysed with closed-form mathematical expressions.

Since the channel acquisition mechanism at the MAC layer supports compressed channel feedback as shown in Fig. 6.1, the practically motivated quantised CSI of the design is easily adapted in the proposed design in Chapter 6.

Furthermore, a relaxed ZF framework is considered rather than an ideal ZF for the calculation of the optimal ZF precoding vector. The fact that there will always be residual interference, because the beamformer is based on the quantised CSI, is exploited to calculate the optimal ZF precoding vector. The proposed concurrent transmission algorithm is unchanged except that the quantisation indices of the channels are selected from the desired and undesired clients instead of the assumed perfect CSI. The optimal ZF precoding vector for the design is obtained by relaxing the interference and power constraints of the optimisation problem to the interference upper bound and to the total transmitter power respectively. Finally, the results of the consideration of the practical constraints lead to a distributive algorithm for calculating the optimal ZF precoding vector, which can be tuned with the quantisation bits, suitable for distributive, decentralised and uncoordinated MU-MIMO WLANs.

9.2 Future Work

The research work presented in this thesis provides a number of potential avenues for future work. Some interesting and challenging open questions are discussed.

9.2.1 Live Performance Measurements

This thesis presents an effective design to address the notorious HT problem in MU-MIMO WLANs with pilot study results in the laboratory prototype settings. A number of distributive algorithms feasible to run at AP is proposed. However, due to the constraints of the USRP2 (i.e., the lack of functionalities to support real-time processing), a real-time

performance measure at the hardware level is not possible. Thus it would be interesting to run a live network and test the performance of the design in the HT scenario. A complete practical implementation would be an interesting line of future work.

9.2.2 Frequency of CSI Measurements

For a design which uses the ZF technique, the state of the knowledge of CSI at a transmitter is critical for performance measurement in terms of interference mitigation. Although an efficient channel acquisition mechanism with reduced signalling time overhead can be designed, the frequency of CSI measurement is vital. A frequent channel acquisition may ensure the up-to-date CSI required for the ZF transmission, but adds additional overhead to the system. Infrequent channel acquisition, although it cuts overheads, may come at the cost of a stale CSI in terms of interference management. Thus the study on the frequency of channel acquisition, considering parameters such as CSI feedback latency, interference leakage, channel coherence time, etc., is still an open issue. Some work can be found in [103, 104], however this critical issue needs more attention.

9.2.3 Sufficiency Condition/s for the Pareto-boundary

The optimal ZF precoding vector for our design is calculated considering the RZF framework. The interference and power constraints of the optimisation problem are relaxed to the interference upper bound and to the total transmit power. The necessary conditions for the maximum rate, with the given finite numbers of quantisation bits B, are derived. However, the choice of the interference leakage, ϕ_{ji}, becomes arbitrary with the choice of B. Thus it is not guaranteed that the maximum rate achieved by clients in our design is also Pareto-optimal. The sufficient condition(/s) that determine(/s) the interference leakage constraint, which in turn ensure(/s) a point in the Pareto-optimal boundary, is a challenging, unsolved information-theoretic problem.

Appendix A

List of Publications

The research work during my PhD has been published/ currently under review in a number of research papers. They are listed below:

Journal Papers

1. **Shrestha, Sanjeeb**, Gengfa Fang, Eryk Dutkiewicz, and Xiaojing Huang. "Solving hidden terminal problem in MU-MIMO WLANs with fairness and throughput-aware precoding and a degrees-of-freedom-based MAC design." EURASIP Journal on Wireless Communications and Networking 2016, no. 1 (2016): 1-25.

2. **Shrestha, Sanjeeb**, Gengfa Fang, Eryk Dutkiewicz, and Xiaojing Huang. "Relaxed Zero-forcing for FRF based-MU-MIMO WLANs." (Submitted for review in the IEEE Transaction on Vehicular Technology)

3. **Shrestha, Sanjeeb**, Gengfa Fang, Eryk Dutkiewicz, and Xiaojing Huang. "Multivariable Analysis of the Relaxed Zero-forcing Precoding Vector in FRF based-MU-MIMO WLANs." (In preparations, intended to submit to IEEE Wireless Communications Letters)

Conference Papers

1. **Shrestha, Sanjeeb**, Gengfa Fang, Eryk Dutkiewicz, and Xiaojing Huang. "Effect of CSI quantization on the average rate in MU-MIMO WLANs." In 2016 13th IEEE Annual Consumer Communications & Networking Conference (CCNC), pp. 824-828. IEEE, 2016.

2. **Shrestha, Sanjeeb**, Gengfa Fang, Eryk Dutkiewicz, and Xiaojing Huang. "Medium Access Control Protocol to Address Hidden Terminals in MU-MIMO WLANs." In Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing (CIT/IUCC/DASC/PICOM), 2015 IEEE International Conference on, pp. 1638-1645. IEEE, 2015.

3. **Shrestha, Sanjeeb**, Gengfa Fang, Eryk Dutkiewicz, and Xiaojing Huang. "Zero-forcing precoding based MAC design to address hidden terminals in MU-MIMO WLANs." In Telecommunications (ICT), 2015 22nd International Conference On, pp. 283-288. IEEE, 2015.

4. **Shrestha, Sanjeeb**, Gengfa Fang, Eryk Dutkiewicz, and Xiaojing Huang. "Addressing hidden terminals in wlans with Zero-forcing coordinated beamforming." In Communications and Information Technologies (ISCIT), 2014 14th International Symposium on, pp. 249-253. IEEE, 2014.

Appendix B

Abbreviations

ACKs	Acknowledgments
AID	Association Identification
APs	Access Points
BER	Bit Error Rate
BPSK	Binary Phase Shift Keying
BTMA	Busy Tone Multiple Access
CDMA	Code Division Multiple Access
CFO	Carrier Frequency Offset
CFP	Contention Free Period
CP	Cyclic Prefix
CP	Contention Period
CQI	Channel Quality Indicator
CR	Cauchy Riemann
CSI	Channel State Information
CSIT	Channel State Information at Transmitter
CSMA/CA	Carrier Sense Multiple Access/Collision Avoidance

CT	Coherence Time
CTS	Clear to Send
DCF	Distributive Coordinated Function
DHCP	Dynamic Host Control Protocol
DoF	Degrees-of-Freedom
DPC	Dirty Paper Coding
DS	Data Sending
ESNR	Effective Signal to Noise Ratio
FAMA	Floor Acquisition Multiple Access
FIFO	First In First Out
FRF	Finite Rate Feedback
GPS	Global Positioning System
HDTV	High Definition Television
HEW	High Efficiency Wireless Local Area Network
HT	Hidden Terminal
IAC	Interference Alignment and Cancellation
IC	Interference Channel
ISM	Industrial Scientific and Medical
IT	Interference Temperature
IoT	Internet of Things
KKT	Karush-Kuhun-Tucker
LSB	Least Significant Bits
MAC	Medium Access Control
MACA	Multiple Access with Collision Avoidance
MACA-P	Medium Access via Collision Avoidance with Enhanced Parallelism
MCCA	Mesh Coordinated Channel Access

MACAW	Multiple Access with Collision Avoidance for Wireless
MIMO	Multiple Input Multiple Output
MISO	Multiple Input Single Output
MMSE	Minimum Mean Square
MPDU	Medium Access Control Packet Data Unit
MU-MIMO	Multi-User Multiple Input Multiple Outpur
MU-MIMO-IC	Multi-User Multiple Input Multiple Output-Interference Channel
NAV	Network Allocation Vector
NDP	Null Data Packet
OFDM	Orthogonal Frequency Division Multiplexing
PC	Personal Computer
PCF	Point Coordination Function
PHY	Physical
PIFS	Point Coordinated Function Inter frame Spacing
PLCP	Physical Layer Convergence Procedure
PPDU	Physical Layer Convergence Procedure Protocol Data Unit
PRN	Packet Radio Network
QAM	Quadrature Amplitude Modulation
QPSK	Quadrature Phase Shift Keying
RF	Radio Frequency
RSSI	Receiver Signal Strength Identification
RTS	Ready To Send
RVQ	Random Vector Quantisation
RZF	Relax Zeroforcing
SDR	Software Define Radio
SIFS	Short Inter Frame Spacing

SINR	Signal to Interference plus Noise Ratio
SISO	Single Input Single Output
SNR	Signal to Noise Ratio
SOPC	Sequential Orthogonal Projection Combining
STA	Station
SU-MIMO	Single User-Multiple Input Multiple Output
TDD	Time Division Duplexing
TXOP	Transmission Opportunity
USRP	Universal Software Radio Peripheral
VHT	Very High Throughput
VHT-SIG	Very High Throughput Signalling
VoD	Video On Demand
WLANs	Wireless Local Area Networks
WMA	Wireless Medium Access
ZF	Zeroforcing
ZFB	Zeroforcing beamforming

Appendix C

Lemma Proofs

C.1 Proof of Lemma 1

$$\begin{aligned}
\frac{d\Delta R(P)}{dB} &= \frac{d}{dB}\left(\log_2\left(\sigma^2 + P.2^{-\frac{B}{M(N-1)}}\right)\right) \\
&= \frac{d\left(\log_2\left(\sigma^2+P.2^{-\frac{B}{M(N-1)}}\right)\right)}{d\left(\sigma^2+P.2^{-\frac{B}{M(N-1)}}\right)}.\frac{d\left(\sigma^2+P.2^{-\frac{B}{M(N-1)}}\right)}{dB} \\
&= -\frac{1}{\ln(2)\left(\sigma^2+P.2^{-\frac{B}{M(N-1)}}\right)}.P\left(\frac{\ln(2).2^{-\frac{B}{M(N-1)}}}{M(N-1)}\right) \\
&= -\frac{P}{M(N-1)}.\frac{2^{-\frac{B}{M(N-1)}}}{\left(\sigma^2+P.2^{-\frac{B}{M(N-1)}}\right)}.
\end{aligned} \tag{C.1}$$

Furthermore,

$$\begin{aligned}
&\frac{d\Delta R(P)}{dB} = \tan\theta \\
&\tan\theta = -\frac{P}{M(N-1)}.\frac{2^{-\frac{B}{M(N-1)}}}{\left(\sigma^2+P.2^{-\frac{B}{M(N-1)}}\right)} \\
&-\frac{P}{\tan\theta}\frac{1}{M(N-1)} = \frac{\left(\sigma^2+P.2^{-\frac{B}{M(N-1)}}\right)}{\left(P.2^{-\frac{B}{M(N-1)}}\right)} \\
&-\frac{P}{\tan\theta}\frac{1}{M(N-1)} = \frac{\sigma^2}{\left(2^{-\frac{B}{M(N-1)}}\right)} + 1 \\
&-\frac{P}{\tan\theta}\left(\frac{1}{M(N-1)} - 1\right) = \frac{\sigma^2}{\left(2^{-\frac{B}{M(N-1)}}\right)}
\end{aligned}$$

and, taking the ln of both sides,

$$
\begin{aligned}
&-\ln\left(\tfrac{P}{\tan\theta}\left(\tfrac{1}{M(N-1)}+1\right)\right)=\ln\left(\sigma^2\right)-\ln\left(2^{-\frac{B}{M(N-1)}}\right)\\
&\left(-\ln\left(\tfrac{P}{\tan\theta}\left(\tfrac{1}{M(N-1)}+1\right)\right)-\ln\left(\sigma^2\right)\right)=\tfrac{B}{M(N-1)}.\ln\left(2\right)\\
&B=-\tfrac{M(N-1)}{\ln(2)}\left(\ln\left(\tfrac{P}{\tan\theta}\left(\tfrac{1}{M(N-1)}+1\right)\right)+\ln\left(\sigma^2\right)\right)\\
&B=-\tfrac{M(N-1)}{\ln(2)}\left(\left(\ln\left(\tfrac{P}{\tan\theta}\right)+\ln\left(\tfrac{1}{M(N-1)}+1\right)\right)+\ln\left(\sigma^2\right)\right).
\end{aligned}
$$

C.2 Proof of Lemma 2

We know from (7.22) that $\Delta SNR_{dB} = -10\log_{10}\left(1-2^{-\frac{B}{M(N-1)}}\right)$. The negative sign represents the decrement in SNR in dB. Thus we take the absolute value of ΔSNR_{dB}.

$$\tfrac{\Delta SNR_{dB}}{10}=\log_{10}\left(1-2^{-\frac{B}{M(N-1)}}\right)$$

taking the antilog of both sides,

$$
\begin{aligned}
10^{\frac{\Delta SNR_{dB}}{10}}&=1-2^{-\frac{B}{M(N-1)}}\\
10^{\frac{\Delta SNR_{dB}}{10}}-1&=-2^{-\frac{B}{M(N-1)}}
\end{aligned}
$$

taking the log of both sides,

$$
\begin{aligned}
&\log\left(1-10^{\frac{\Delta SNR_{dB}}{10}}\right)=-\tfrac{B}{M(N-1)}.\log\left(2\right)\\
&B=M\left(\frac{(1-N)\left(\log\left(1-10^{\frac{\Delta SNR_{dB}}{10}}\right)\right)}{\log(2)}\right)
\end{aligned}
$$

Substituting $|\Delta SNR_{dB}| = 10\log_{10}\left(1-2^{-\frac{B}{M(N-1)}}\right)$ and $|B| = \frac{M(N-1)}{\ln(2)}\left(\left(\ln\left(\frac{P}{\tan\theta}\right)+\ln\left(\frac{1}{M(N-1)}+1\right)\right)+\ln\left(\sigma^2\right)\right)$, we get

$$B=\frac{M(1-N)}{\log(2)}\left(\log_{10}\left(1-10^{\frac{10\log_{10}(\eta)}{10}}\right)\right), \tag{C.2}$$

where $\eta=\left(1-2^{-\frac{\left(\left(\ln\left(\frac{P}{\tan\theta}\right)\right)+\ln\left(\frac{1}{M(N-1)}+1\right)\right)+\ln\left(\sigma^2\right)}{\ln(2)}}\right)$.

C.3 Proof of Lemma 3

According to the definition of a complex-valued function, if $z = x + iy$, then a function $f(z)$ is simply a function $F(x, y) = u(x, y) + iv(x, y)$. When $f(z) = |z|$, then $F(x, y) = (u = x, v = 0)$, and it only takes a value on the real axis. Thus $f(z) = |z|$ does not satisfy the Cauchy Riemann conditions $\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}$ or $\frac{\partial v}{\partial x} = -\frac{\partial u}{\partial y}$. Hence, the complex-valued function $f(z) = |z|$ is not holomorphic.

Similar arguments follow for the complex-valued Lagrangian function in (8.6), as we see that complex-valued variables such as $\tilde{\mathbf{H}}_{ji}$ and $\hat{\mathbf{v}}_j$ are either absolute or only the real parts are taken, Hence, the complex-valued Lagrangian function is not holomorphic and its complex derivative cannot be taken.

Bibliography

[1] R. Wiseman, *Quirkology: how we discover the big truths in small things.* Basic Books, 2007.

[2] [Online]. Available: www.ieee802.org/11/Reports/tgax_update.htm

[3] A. Sheth, C. Doerr, D. Grunwald, R. Han, and D. Sicker, "MOJO: A distributed physical layer anomaly detection system for 802.11 WLANs," *ACM*, pp. 191–204, 2006.

[4] J. Winters, "On the capacity of radio communication systems with diversity in a rayleigh fading environment," *IEEE Journal on Selected Areas in Communications*, vol. 5, no. 5, pp. 871–878, 1987.

[5] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Technical Journal*, vol. 1, no. 2, pp. 41–59, 1996.

[6] E. Telatar, "Capacity of multi-antenna gaussian channels," *European transactions on telecommunications*, vol. 10, no. 6, pp. 585–595, 1999.

[7] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 5, pp. 684–702, 2003.

[8] D. Tse and P. Viswanath, *Fundamentals of Wireless Communications*. Cambridge University Press, 2005.

[9] [Online]. Available: http://www.infonetics.com/pr/2013/2Q13-Wireless-LAN-Market-Highlights.asp

[10] Y.-C. Cheng, J. Bellardo, P. Benkö, A. C. Snoeren, G. M. Voelker, and S. Savage, "Jigsaw: Solving the puzzle of enterprise 802.11 analysis," in *Proceedings of the 2006 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications*, ser. SIGCOMM '06. New York, NY, USA: ACM, 2006, pp. 39–50.

[11] M. Bertocco, G. Gamba, A. Sona, and S. Vitturi, "Performance measurements of CSMA/CA-based wireless sensor networks for industrial applications," in *Instrumentation and Measurement Technology Conference (IMTC 2007)*, May 2007, pp. 1–6.

[12] X. Yang and N. Vaidya, "On physical carrier sensing in wireless ad hoc networks," in *24th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 2005)*, vol. 4, March 2005, pp. 2525–2535 vol. 4.

[13] F. A. Tobagi and L. Kleinrock, "Packet switching in radio channels: part ii–the hidden terminal problem in carrier sense multiple-access and the busy-tone solution," *Communications, IEEE Transactions on*, vol. 23, no. 12, pp. 1417–1433, 1975.

[14] C. Ware, J. Judge, J. Chicharo, and E. Dutkiewicz, "Unfairness and capture behaviour in 802.11 adhoc networks," in *IEEE International Conference on Communications ICC 2000*, vol. 1. IEEE, 2000, pp. 159–163.

[15] C. L. Fullmer and J. Garcia-Luna-Aceves, "Solutions to hidden terminal problems in wireless networks," in *ACM SIGCOMM Computer Communication Review*, vol. 27, no. 4. ACM, 1997, pp. 39–49.

[16] P. Karn, "MACA-a new channel access method for packet radio," in *ARRL/CRRL Amateur radio 9th computer networking conference*, vol. 140, 1990, pp. 134–140.

[17] V. Bharghavan, A. Demers, S. Shenker, and L. Zhang, "MACAW: a media access protocol for wireless LANs," *ACM SIGCOMM Computer Communication Review*, vol. 24, no. 4, pp. 212–225, 1994.

[18] C. L. Fullmer and J. Garcia-Luna-Aceves, "Floor acquisition multiple access (FAMA) for packet-radio networks," in *ACM SIGCOMM computer communication review*, vol. 25, no. 4. ACM, 1995, pp. 262–273.

[19] A. Acharya, A. Misra, and S. Bansal, "MACA-P: a MAC for concurrent transmissions in multi-hop wireless networks," in *Proc. First IEEE International Conference on Pervasive Computing and Communications,(PerCom 2003)*. IEEE, 2003, pp. 505–508.

[20] P. C. Ng, S. C. Liew, K. C. Sha, and W. T. To, "Experimental study of hidden node problem in IEEE 802.11 wireless networks," *Sigcomm Poster*, p. 26, 2005.

[21] K. Lin, Y.-J. Chuang, and D. Katabi, "A light-weight wireless handshake," *ACM SIGCOMM Computer Communication Review*, vol. 42, no. 2, pp. 28–34, 2012.

[22] J. G. Andrews, "Interference cancellation for cellular systems: a contemporary overview," *Wireless Communications, IEEE*, vol. 12, no. 2, pp. 19–29, 2005.

[23] S. Gollakota and D. Katabi, "Zigzag decoding: combating hidden terminals in wireless networks," *ACM*, vol. 38, no. 4, 2008.

[24] R. Ramanathan, J. Redi, C. Santivanez, D. Wiggins, and S. Polit, "Ad hoc networking with directional antennas: a complete system solution," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 3, pp. 496–506, 2005.

[25] K. Sundaresan and R. Sivakumar, "A unified MAC layer framework for ad-hoc networks with smart antennas," in *Proceedings of the 5th ACM international symposium on Mobile ad hoc networking and computing*. ACM, 2004, pp. 244–255.

[26] H. Singh and S. Singh, "Smart-aloha for multi-hop wireless networks," *Mobile Networks and Applications*, vol. 10, no. 5, pp. 651–662, 2005.

[27] B. Hamdaoui and K. G. Shin, "Characterization and analysis of multi-hop wireless MIMO network throughput," in *Proceedings of the 8th ACM international symposium on Mobile ad hoc networking and computing*. ACM, 2007, pp. 120–129.

[28] J. Mundarath, P. Ramanathan, and B. D. Van Veen, "A cross layer scheme for adaptive antenna array based wireless ad hoc networks in multipath environments," *Wireless Networks*, vol. 13, no. 5, pp. 597–615, 2007.

[29] J.-S. Park, A. Nandan, M. Gerla, and H. Lee, "SPACE-MAC: Enabling spatial reuse using MIMO channel-aware MAC," in *IEEE International Conference on Communications (ICC 2005)*, vol. 5. IEEE, 2005, pp. 3642–3646.

[30] K. Huang, J. G. Andrews, D. Guo, R. W. Heath Jr, and R. A. Berry, "Spatial interference cancellation for multiantenna mobile ad hoc networks," *IEEE Transactions on Information Theory*, vol. 58, no. 3, pp. 1660–1676, 2012.

[31] J. Mundarath, P. Ramanathan, and B. D. Van Veen, "NULLHOC: a MAC protocol for adaptive antenna array based wireless ad hoc networks in multipath environments," in *Global Telecommunications Conference (GLOBECOM'04)*, vol. 5. IEEE, 2004, pp. 2765–2769.

[32] Y. Lebrun, K. Zhao, S. Pollin, A. Bourdoux, F. Horlin, S. Du, and R. Lauwereins, "Beamforming techniques for enabling spatial-reuse in MCCA 802.11 s networks," *EURASIP Journal on wireless communications and networking*, vol. 2011, no. 1, pp. 1–13, 2011.

[33] Z. Zhang, S. Bronson, J. Xie, and W. Hu, "Employing the one-sender-multiple-receiver technique in wireless LANs," *IEEE/ACM Transactions on Networking (TON)*, vol. 21, no. 4, pp. 1243–1255, 2013.

[34] B. Bellalta, J. Barcelo, D. Staehle, A. Vinel, and M. Oliver, "On the performance of packet aggregation in IEEE 802.11 ac MU-MIMO WLANs," *IEEE Communications Letters*, vol. 16, no. 10, pp. 1588–1591, 2012.

[35] G. Redieteab, L. Cariou, P. Christin, and J.-F. Hélard, "SU/MU-MIMO in ieee 802.11 ac: PHY+ MAC performance comparison for single antenna stations," in *Wireless Telecommunications Symposium (WTS), 2012*. IEEE, 2012, pp. 1–5.

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

[37] 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*. ACM, 2012, pp. 41–52.

[38] H. Lou, M. Ghosh, P. Xia, and R. Olesen, "A comparison of implicit and explicit channel feedback methods for MU-MIMO WLAN systems," in *24th Annual In-*

ternational Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC). IEEE, 2013, pp. 419–424.

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

[40] E. Kartsakli, N. Zorba, L. Alonso, and C. Verikoukis, "Multiuser MAC protocols for 802.11 n wireless networks," in *IEEE International Conference on Communications*. IEEE, 2009, pp. 1–5.

[41] L. X. Cai, H. Shan, W. Zhuang, X. Shen, J. W. Mark, and Z. Wang, "A distributed multi-user MIMO MAC protocol for wireless local area networks," in *Global Telecommunications Conference (GLOBECOM 2008)*, Nov 2008, pp. 1–5.

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

[43] W. L. Huang, K. B. Letaief, and Y. J. Zhang, "Joint channel state based random access and adaptive modulation in wireless LANs with multi-packet reception," *IEEE transactions on wireless communications*, vol. 7, no. 11, pp. 4185–4197, 2008.

[44] T. Tandai, H. Mori, K. Toshimitsu, and T. Kobayashi, "An efficient uplink multiuser MIMO protocol in IEEE 802.11 WLANs," in *IEEE 20th International Symposium on Personal, Indoor and Mobile Radio Communications*. IEEE, 2009, pp. 1153–1157.

[45] Z. Sheng and N. Zhisheng, "Distributed medium access control with SDMA support for WLANs," *IEICE transactions on communications*, vol. 93, no. 4, pp. 961–970, 2010.

[46] P. X. Zheng, Y. J. Zhang, and S. C. Liew, "Multipacket reception in wireless local area networks," in *IEEE International Conference on Communications*, vol. 8. IEEE, 2006, pp. 3670–3675.

[47] H. Li, K. Wu, Q. Zhang, and L. M. Ni, "CUTS: improving channel utilization in both time and spatial domain in WLANs," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 6, pp. 1413–1423, 2014.

[48] M. Zhao, M. Ma, and Y. Yang, "Applying opportunistic medium access and multiuser MIMO techniques in multi-channel multi-radio WLANs," *Mobile Networks and Applications*, vol. 14, no. 4, pp. 486–507, 2009.

[49] H. Li, A. Attar, and V. C. Leung, "Multi-user medium access control in wireless local area network," in *IEEE Wireless Communication and Networking Conference*. IEEE, 2010, pp. 1–6.

[50] H. Shen, S. Lv, Y. Sun, X. Dong, X. Wang, and X. Zhou, "Concurrent access control using subcarrier signature in heterogeneous MIMO-based WLAN," in *International Workshop on Multiple Access Communications*. Springer, 2012, pp. 109–121.

[51] S. Gollakota, S. D. Perli, and D. Katabi, "Interference alignment and cancellation," in *ACM SIGCOMM Computer Communication Review*, vol. 39, no. 4. ACM, 2009, pp. 159–170.

[52] R. Liao, B. Bellalta, M. Oliver, and Z. Niu, "MU-MIMO MAC protocols for wireless local area networks: A survey," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 1, pp. 162–183, 2014.

[53] K. C.-J. Lin, S. Gollakota, and D. Katabi, "Random access heterogeneous MIMO networks," in *ACM SIGCOMM Computer Communication Review*, vol. 41, no. 4. ACM, 2011, pp. 146–157.

[54] S. Shrestha, G. Fang, E. Dutkiewicz, and X. Huang, "Solving hidden terminal problem in mu-mimo wlans with fairness and throughput-aware precoding and a degrees-of-freedom-based MAC design," *EURASIP Journal on Wireless Communications and Networking*, vol. 2016, no. 1, pp. 1–25, 2016.

[55] ——, "Effect of CSI quantization on the average rate in MU-MIMO WLANs," in *13th IEEE Annual Consumer Communications & Networking Conference (CCNC)*. IEEE, 2016, pp. 824–828.

[56] ——, "Medium access control protocol to address hidden terminals in mu-mimo wlans," in *IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing (CIT/IUCC/DASC/PICOM)*. IEEE, 2015, pp. 1638–1645.

[57] ——, "Zero-forcing precoding based MAC design to address hidden terminals in MU-MIMO WLANs," in *22nd International Conference On Telecommunications (ICT)*. IEEE, 2015, pp. 283–288.

[58] ——, "Addressing hidden terminals in wlans with zeroforcing coordinated beamforming," in *14th International Symposium on Communications and Information Technologies (ISCIT)*. IEEE, 2014, pp. 249–253.

[59] ——, "Relaxed zeroforcing for finite rate feedback based MU-MIMO WLANs," *IEEE Transection on Vehicular Technology (Submitted for review)*.

[60] A. Narula, M. J. Lopez, M. D. Trott, and G. W. Wornell, "Efficient use of side information in multiple-antenna data transmission over fading channels," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, pp. 1423–1436, 1998.

[61] D. J. Love, R. W. Heath, and T. Strohmer, "Grassmannian beamforming for multiple-input multiple-output wireless systems," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2735–2747, 2003.

[62] K. K. Mukkavilli, A. Sabharwal, E. Erkip, and B. Aazhang, "On beamforming with finite rate feedback in multiple-antenna systems," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2562–2579, 2003.

[63] P. Ding, D. J. Love, and M. D. Zoltowski, "On the sum rate of channel subspace feedback for multi-antenna broadcast channels," in *Global Telecommunications Conference, 2005. GLOBECOM'05. IEEE*, vol. 5. IEEE, 2005, pp. 5–pp.

[64] N. Jindal, "MIMO broadcast channels with finite-rate feedback," *IEEE Transactions on Information Theory,*, vol. 52, no. 11, pp. 5045–5060, 2006.

[65] D. J. Love, R. W. Heath, V. K. 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, no. 8, pp. 1341–1365, 2008.

[66] T. Yoo, N. Jindal, and A. Goldsmith, "Multi-antenna downlink channels with limited feedback and user selection," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 7, pp. 1478–1491, 2007.

[67] W. Santipach and M. L. Honig, "Asymptotic capacity of beamforming with limited feedback," in *IEEE International Symposium on Information Theory*, 2004, pp. 290–290.

[68] ——, "Capacity of beamforming with limited training and feedback," in *IEEE International Symposium on Information Theory*. IEEE, 2006, pp. 376–380.

[69] J. Park, G. Lee, Y. Sung, and M. Yukawa, "Coordinated beamforming with relaxed zero forcing: The sequential orthogonal projection combining method and rate control," *IEEE Transactions on Signal Processing*, vol. 61, no. 12, pp. 3100–3112, 2013.

[70] R. Zhang and S. Cui, "Cooperative interference management with MISO beamforming," *IEEE Transactions on Signal Processing*, vol. 58, no. 10, pp. 5450–5458, 2010.

[71] X. Shang, B. Chen, and H. V. Poor, "On the optimality of beamforming for multi-user MISO interference channels with single-user detection," *arXiv preprint arXiv:0908.3171*, 2009.

[72] H. Te Sun and K. Kobayashi, "A new achievable rate region for the interference channel," *IEEE transactions on information theory*, vol. 27, no. 1, pp. 49–60, 1981.

[73] A. Carleial, "Interference channels," *IEEE Transactions on Information Theory*, vol. 24, no. 1, pp. 60–70, 1978.

[74] A. Wiesel, Y. C. Eldar, and S. Shamai, "Zero-forcing precoding and generalized inverses," *IEEE Transactions on Signal Processing*, vol. 56, no. 9, pp. 4409–4418, 2008.

[75] F. Boccardi and H. Huang, "Zero-forcing precoding for the MIMO broadcast channel under per-antenna power constraints," in *IEEE 7th Workshop on Signal Processing Advances in Wireless Communications, SPAWC'06*. IEEE, 2006, pp. 1–5.

[76] G. Caire and S. Shamai, "On the achievable throughput of a multiantenna gaussian broadcast channel," *IEEE Transactions on Information Theory*, vol. 49, no. 7, pp. 1691–1706, 2003.

[77] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Transactions on Signal Processing*, vol. 52, no. 2, pp. 461–471, 2004.

[78] J. Lee and N. Jindal, "High SNR analysis for mimo broadcast channels: Dirty paper coding versus linear precoding," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4787–4792, 2007.

[79] M. H. Costa, "Writing on dirty paper (corresp.)," *IEEE Transactions on Information Theory*, vol. 29, no. 3, pp. 439–441, 1983.

[80] R. Zamir, S. Shamai, and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," *IEEE Transactions on Information Theory*, vol. 48, no. 6, pp. 1250–1276, 2002.

[81] U. Erez and S. ten Brink, "A close-to-capacity dirty paper coding scheme," *IEEE Transactions on Information Theory*, vol. 51, no. 10, pp. 3417–3432, 2005.

[82] T. Philosof, U. Erez, and R. Zamir, "Combined shaping and precoding for interference cancellation at low SNR," in *IEEE international symposium on information theory*, 2003, pp. 68–68.

[83] N. Jindal, "High SNR analysis of MIMO broadcast channels," in *Proceedings International Symposium on Information Theory.* IEEE, 2005, pp. 2310–2314.

[84] M. Mitzenmacher, "The power of two choices in randomized load balancing," *IEEE Transactions on Parallel and Distributed Systems*, vol. 12, no. 10, pp. 1094–1104, 2001.

[85] V. N. Sachkov, "Combinatorial methods in discrete mathematics," no. 55, 1996.

[86] [Online]. Available: http://www.ettus.com/

[87] S. Nanda and K. M. Rege, "Frame error rates for convolutional codes on fading channels and the concept of effective e b/n 0," *IEEE Transactions on Vehicular Technology*, vol. 47, no. 4, pp. 1245–1250, 1998.

[88] [Online]. Available: http://www.gnu.org/

[89] H. MacLeod, C. Loadman, and Z. Chen, "Experimental studies of the 2.4-ghz ISM wireless indoor channel," in *Proceedings of the 3rd Annual Communication Networks and Services Research Conference*. IEEE, 2005, pp. 63–68.

[90] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Predictable 802.11 packet delivery from wireless channel measurements," *ACM SIGCOMM Computer Communication Review*, vol. 41, no. 4, pp. 159–170, 2011.

[91] R. Jain, D.-M. Chiu, and W. Hawe, "A quantitative measure of fairness and discrimination for resource allocation in shared computer systems," 1998.

[92] M. Gast, "802.11 wireless networks: the definitive guide," 2005.

[93] W. Santipach and M. L. Honig, "Signature optimization for CDMA with limited feedback," *IEEE Transactions on Information Theory*, vol. 51, no. 10, pp. 3475–3492, 2005.

[94] C. K. Au-Yeung and D. J. Love, "On the performance of random vector quantization limited feedback beamforming in a MISO system," *IEEE Transactions on Wireless Communications*, vol. 6, no. 2, pp. 458–462, 2007.

[95] A. K. Gupta and S. Nadarajah, *Handbook of beta distribution and its applications*. CRC Press, 2004.

[96] P. J. Davis, "Leonhard euler's integral: A historical profile of the gamma function: In memoriam: Milton abramowitz," *American Mathematical Monthly*, pp. 849–869, 1959.

[97] D. Kershaw, "Some extensions of W. Gautschi's inequalities for the gamma function," *Mathematics of Computation*, pp. 607–611, 1983.

[98] N. Ravindran and N. Jindal, "Multi-user diversity vs. accurate channel state information in MIMO downlink channels," *IEEE Transactions on Wireless Communications*, vol. 11, no. 9, pp. 3037–3046, 2012.

[99] M. Bengtsson and B. Ottersten, "Optimal downlink beamforming using semidefinite optimization," in *37th Annual Allerton Conference on Communication, Control, and Computing*, 1999, pp. 987–996.

[100] K. Kreutz-Delgado, "The complex gradient operator and the cr-calculus," *arXiv preprint arXiv:0906.4835*, 2009.

[101] P. Bouboulis, "Wirtinger's calculus in general hilbert spaces," *arXiv preprint arXiv:1005.5170*, 2010.

[102] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge University Press, 2004.

[103] G. Redieteab, L. Cariou, P. Christin, and J.-F. Helard, "PHY+ MAC channel sounding interval analysis for ieee 802.11 ac MU-MIMO," in *2012 International Symposium on Wireless Communication Systems (ISWCS)*. IEEE, 2012, pp. 1054–1058.

[104] R. Liao, B. Bellalta, J. Barcelo, V. Valls, and M. Oliver, "Performance analysis of ieee 802.11 ac wireless backhaul networks in saturated conditions," *EURASIP*

Journal on Wireless Communications and Networking, vol. 2013, no. 1, pp. 1–14, 2013.