Shanghai Jiao Tong University
University of Michigan- Shanghai Jiao Tong University Joint Institute

Full Duplex Communication with Half Duplex Clients: MIMO Case

by

Orhan Okan Mutgan

A thesis submitted in partial satisfaction of the
requirements for the degree of Master in
Electrical and Computer Engineering at Shanghai Jiao Tong University

Committee in charge:
Professor Xudong Wang (Chair)
Assistant Professor Weikang Qian
Assistant Professor Chong Han

Shanghai
March, 2017

上海交通大学

交大密西根学院

基于MIMO的AP与半双工客户端的全双工通信

Orhan Okan Mutgan

上海交通大学密西根学院硕士学位论文

通信与网络专业

委员会成员：

王旭东（主席）

钱炜慷

韩充

上海

2017年3月

Abstract

Full duplex (FD) communication is an attractive topic in recent years since it allows a radio to receive and transmit simultaneously in the same frequency band. Subsequently, this results in doubled capacity and reduced bandwidth consumption. Recent research shows the hardware feasibility of enabling full duplex communication in commercial networks. However, there exist many users that only operate in half duplex (HD) mode, therefore they do not support FD communication. On the other hand, FD functionality can be inserted into access point (AP) more conveniently. Within this context, communication network in this study concerns coexistence of full duplex and half duplex communication, where AP supports FD communication and users serve as HD nodes. Conversely, inter-node interference (INI) between users occurs in such scenario, specifically, uplink transmission causes interference at downlink transmission. In order to mitigate the effect of INI, this work exploits IA and MIMO technology.

Current studies realize FD communication for single antenna users. Some researchers also consider FD communication for multiple antenna users. However, simultaneous communication of different number of antenna users needs to be investigated. In this sense, this study focuses on designing a MAC protocol for users equipped with a different number of antennas. Under the topology of one full duplex AP and numerous half duplex users, the proposed MAC protocol aims at selecting clients to maximize throughput and ensure fairness, i.e. achieving the best trade-off between throughput and fairness. In this regard, this work suggest four selection schemes for AP to evaluate the throughput and fairness, and consequently obtains the highest trade-off. Simulation results show that proportional fairness selection scheme achieves the most desirable performance among the proposed four selection schemes. Overhead and complexity are also examined and compared for the proposed schemes, and analysis is given to demonstrate their improvement.

摘要

全双工通信技术允许一个通信设备在同一频带上同时收发，进而能够减少带宽消耗，使通信容量翻倍，因此近年来吸引了学术界的广泛研究。最近的研究验证了在商用网络中使用全双工通信技术的可行性，但是大多数客户端设备依然是半双工设备。考虑到全双工通信技术能便利地应用于接入点（AP），因此有必要研究接入点工作在全双工模式，客户端工作在半双工模式的应用场景。在这种场景下，客户端节点间存在相互干扰，具体来讲，上行通信会对下行通信造成干扰。本文利用干扰对齐（IA）和多输入多输出（MIMO）技术克服节点间干扰，实现可靠的全双工通信。

目前的全双工通信研究大多仅考虑了具有单个天线的客户端，部分研究也考虑了多天线客户端，但是不同天线数量的用户之间的全双工通信还未被纳入研究。因此，本文针对不同天线数量的用户之间的全双工通信提出了一种新的 MAC 协议。本文所研究的通信网络拓扑包含一个全双工接入点和若干个半双工客户端，本文研究的目的是从这些客户端中选出最优的可通信客户端，实现最大的网络流通量，并保证公平性，也就是说，在网络流通量和公平性之间取得最佳的折中。本文比较了四种不同的客户端选取方案，评估了在这四种方案下的网络流通量和公平性，以获得最佳的折中方案。本文的仿真结果显示，正比公平（proportional fairness）选取方案在所比较的四种方案中能实现最佳的网络性能。另外，本文还对所提出的 MAC 方案作了网络开销和复杂度的分析。

Contents

Chapter 1

Introduction

It has been a long-held assumption that it is generally not possible for radios to receive and transmit on the same frequency band at the same time because of the interference that results (Goldsmith, 2005). This statement indicates that radios cannot transmit and receive simultaneously, i.e. they operate in half duplex mode. Recent works have shown that it is nowadays possible for a radio to transmit and receive concurrently, i.e. they operate in full duplex mode, hence conceivably doubling the link capacity.

Several research has shown the implementation of full duplex radios (Choi et al., 2010; Jain et al., 2011; Duarte et al., 2012; Hong et al., 2013), and some even with the use of multiple antennas (Aryafar et al., 2012). The key challenge in realizing such a FD device is the self-interference (SI) generated by the Tx antenna at the Rx antenna. To achieve FD, a radio has to cancel the SI, which can be several orders of magnitude stronger than the received signal. Various researchers addressed the SI problem in the analog domain through antenna (Choi et al., 2010; Aryafar et al., 2012; Everett et al., 2011) and RF cancellation (Radunovic et al., 2010; Jain et al., 2011; Duarte et al., 2012; Bhadaria et al., 2013), while others addressed it through digital cancellation (Duarte et al., 2012; Bhadaria et al., 2013).

Considering the fact that SI is addressed efficiently in the previous studies, current works about full duplex networks mainly focus on point to point scenario. In such scenario, both nodes operate in FD fashion and transmit(receive) to(from) each other. However, it is significant to consider FD communication where HD clients also exist in the network, since it is easier to implement FD functionality in AP but it is not practical to adopt FD capabilities to the legacy clients. Therefore, a FD AP with HD clients network seems more realistic in today's commercial networks.

Enabling FD AP with half duplex clients introduces a new form of interference. Since AP performs FD, it receives packets from clients and transmits packets to other clients simultaneously. While an HD client is transmitting to AP, another HD client is receiving from AP. Because uplink (UL) and downlink (DL) transmissions take place at the same time on the same frequency, a different type of interference, called *inter-node interference (INI)*, arises between uplink and downlink clients as shown in Fig. 1.1. One simple approach to address INI is the separation of the clients. UL and DL clients can be placed far away from each other to mitigate the effect of INI. If the separation is distant enough, the transmissions can be realized.

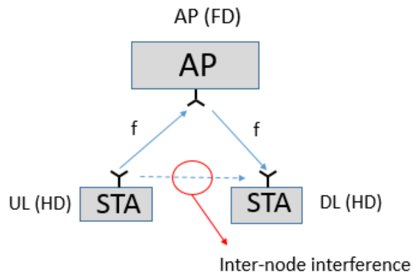


Figure 1.1: Inter-node Interference

However, this approach reduces the gain of FD session and does not seem practical in real environment. Another simple but more feasible approach is using the side channel. Devices of clients most probably have both 3G/4G and WiFi interfaces. As shown in (Bai and Sabharwal, 2012), UL clients can send a copy of the interference to the DL client using one channel (e.g WiFi interface). On the other channel (e.g 3G/4G), the FD communication can be realized. After receiving the coded version of interference on the side channel, DL client can remove the interference on the main channel, hence FD session can be performed. Nevertheless, this approach requires additional bandwidth, as well as utilizing two interfaces, which leads to high power consumption. A more reliable approach is shown in (Sundaresan et al., 2014), where interference alignment and MIMO are exploited to overcome the inter-node interference problem. The idea here is to use multiple antennas in the client devices. Adding multiple antennas has the advantage of utilizing interference alignment. Interference alignment is a linear precoding technique that attempts to align interfering signals at the receiver. Precoding is realized at the transmitters by the use of multiple antennas. Transmitters multiply their data streams by a precoding vector, so that the interference streams at the receiver lie on same subspace, allowing the desired signal to be decoded in the presence of interference.

Apart from PHY layer attributes, there have been several proposals for MAC protocols to enable FD communication (Sahai et al., 2011; Singh et al., 2011; Kim et al., 2013; Tang and Wang, 2015). The network structure of (Kim et al., 2013) and (Tang and Wang, 2015) are similar to the scenario in this study. In both work, AP serves as a FD node, while HD clients establish a link (transmission and reception) with AP. The performance in both cases depends on the interference between clients. (Kim et al., 2013) regards the interference as a source that reduces the channel capacity, but still steady communications are achievable by alleviating the transmission rate. (Tang and Wang, 2015) exploits capture effect, in which the receiver decodes the stronger signal correctly, when it receives two colliding packets. Nevertheless, none of the MAC work mentioned above takes into account of clients equipped with multiple antenna. Even though (Sundaresan et al., 2014) mainly focuses on exploiting MIMO and interference alignment, it also proposes a basic MAC protocol. However, the MAC protocol that is proposed needs to be improved to fully take advantage of full duplex communication. Therefore, in this paper, we study a novel MAC protocol by utilizing exclusive physical layer features such as MIMO and IA in a network where AP supports full duplex communication while all clients operate in half duplex fashion.

This thesis is divided into five chapters. Chapter 1 describes previous work and gives motivation for the work performed in this thesis. Chapter 2 presents the approach taken and the background theory required for the analysis. Chapter 3 provides an insight for the need of the MAC design and the detailed methodology used in the MAC protocol. Chapter 4 represents the performance evaluation and discusses the significance of the results. Finally, Chapter 7 summarizes the main conclusions of the thesis and presents an outlook for future work.

Chapter 2

Background theory

In this section, we present a brief introduction to MIMO and Interference Alignment. Multiple input multiple output (MIMO) and Interference Alignment (IA) are two key concepts in our design, thus we better provide a background for the reader. After completing this section, we hope that the reader grasps the idea of MIMO and IA, consequently, the usage of them in our design.

2.1 MIMO Premier

To proceed our work and to understand the interference alignment, it is important to know some basic properties of MIMO. MIMO utilizes multiple antennas both at transmitter and receiver as shown in Fig. 2.1. In the figure, channel coefficients are represented by h_{ij} from transmitting antenna i to receiving antenna j, and transmitted signals are represented by x_1 and x_2. When transmitter transmits a signal on his first antenna, the receiver receives h_{11} times x_1 on its first antenna and h_{12} times x_1 on its second antenna. Since transmitter has two antennas, it can concurrently transmit another signal on the second antenna. The receiver receives $h_{21}x_2$ on its first antenna and $h_{22}x_2$ on its second antenna. Thus, the second signal also creates a vector in the space defined by the two antennas. Consequently, receiver receives a linear combination of the two transmitted signals, on each antenna. So, the two-antenna receiver receives the following signals:

$$y_1 = h_{11}x_1 + h_{21}x_2 \tag{2.1}$$

$$y_2 = h_{12}x_1 + h_{22}x_2 \tag{2.2}$$

Intuitively, the receiver receives a vector whose direction is determined by the channel as shown in Fig. 2.1.

Note that it is convenient to use two-dimensional vectors to represent the system (Tse and Viswanath, 2005).

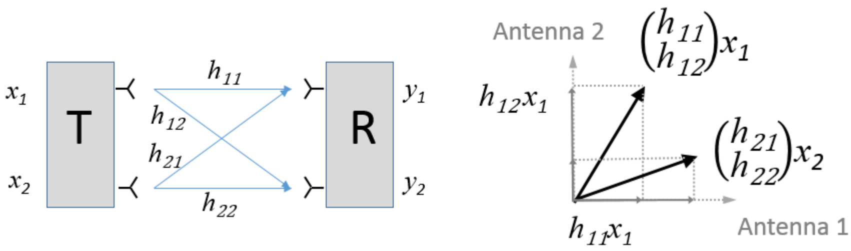


Figure 2.1: Received signals at the receiver

After understanding the transmission process, decoding is also significant and worth recalling. To understand the decoding process, it is important to know that an n-antenna node receives signals in n-dimensional space. For example, a one-antenna client receives a signal at only one antenna; so it receives signals in one-dimension. Correspondingly, a client with two antennas receives a signal on both of its antennas. Hence, the received signal is a vector in a two-dimensional space. Similarly, a node with three antennas receives a signal in a three-dimensional space as shown in Fig. 2.2. Equivalently, an n-antenna node transmits signals in n-dimensional space as well. For example, a two-antenna transmitter transmits a three-dimensional vector.

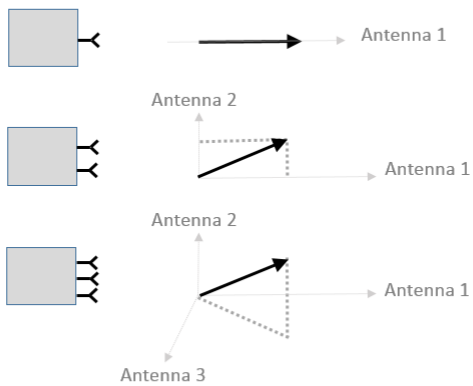


Figure 2.2: Antenna–dimension relationship

Let us look back at the example in Fig. 2.1 to comprehend decoding process. Receiver receives the sum of these vectors as shown in Fig. 2.3. Since the two signals x_1 and x_2 are transmitted concurrently, they interfere. In order to decode a signal, the receiver projects on a direction orthogonal to the interference from the other signal. For example, to decode x_1, the receiver has to cancel interference from x_2, which it can do by projecting on a direction orthogonal to x_2. The signal after projection is a scaled version of the original signal. So, x_1a is the decoded signal of the original signal x_1. Equivalently, to decode x_2, the receiver has to eliminate interference from $x_1^{'}$, which can be achieved by projecting on the direction orthogonal to x_1. The decoded signal is $x_2^{'}$. Thus, a receiver with two-antennas can decode two concurrent packets. The whole process is shown in Fig. 2.3.

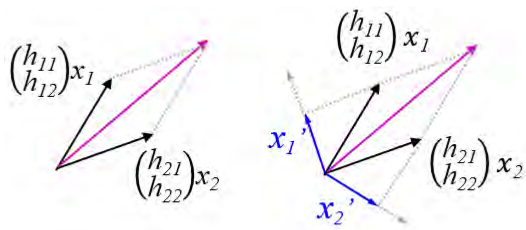


Figure 2.3: Decoding process

In decoding process, it is important to recall that the angle (θ) between the received signals affects the performance. Depending on the angle θ, the projected signal might have a large or small amplitude. A larger amplitude corresponds to a higher signal-to-noise ratio (SNR), thus it gives a higher throughput. Consequently, smaller amplitude results in lower SNR and hence yields lower throughput. For example, consider two cases where the receiver receives the same signals x_1 and x_2 but with different angles as shown in Fig 2.4. This difference in angle might be caused by different channel conditions. On the left side of the Fig. 2.4, since the angle between received signals is small, corresponding decoded signals $x_1^{'}$ and $x_2^{'}$ yield a small amplitude, thus the loss in decoding will be very high, consequently, transmission rate will be low. However, on the right side of Fig. 2.4, since the angle between received signals is large, decoded signals $x_1^{'}$ and $x_2^{'}$ achieve a larger amplitude and hence grant a higher transmission rate. One can directly realize that the optimum case is where the angle is 90 degrees. In that case, projection does not induce a loss, hence exactly the same amplitude will be obtained from the original signal. (Shen et al., 2012) presents a good understanding of the relationship between angle, SNR and the performance in relation.

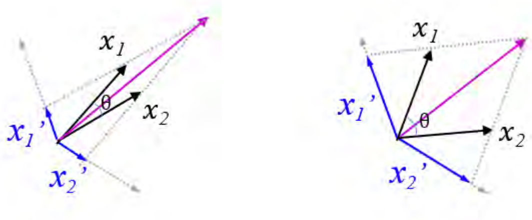
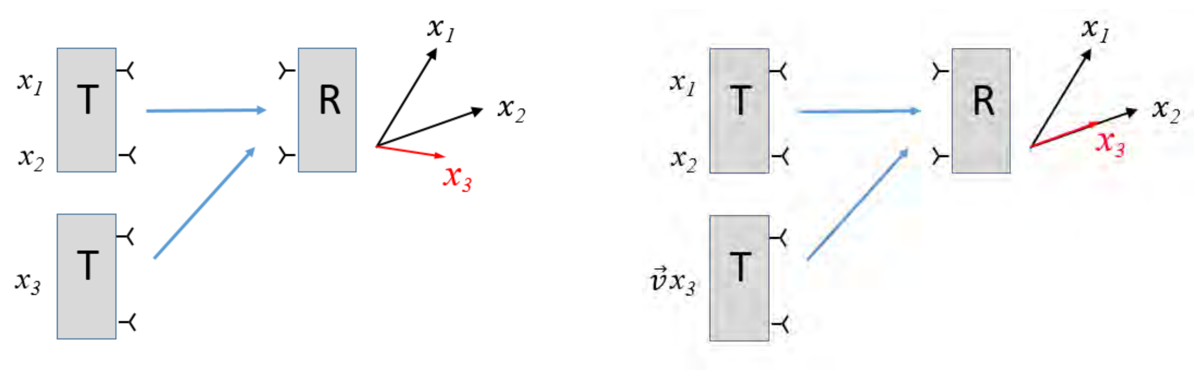


Figure 2.4: Dependency on angle θ

What if three concurrent signals are transmitted to two-antenna receiver? Let us look at Fig. 2.5a. In addition to the two signals, x_1 and x_2, the receiver now receives a third signal x_3. Now the receiver has the



(a) Without IA (b) With IA

Figure 2.5: Multiple streams

three packets along three different directions, which it cannot decode. This is because to decode say x_1, it has to find a direction orthogonal to interference from both x_2 and x_3. However, since the receiver has only two antennas, the space is two-dimensional and the receiver cannot find a direction that is orthogonal to both x_2 and x_3. Thus the receiver cannot decode any of these packets. In general, n-antenna receivers can decode up to n concurrent signals (Tse and Viswanath, 2005).

One of the biggest advantage of MIMO is that transmitter can control the vectors along which its signal is received. A transmitter can use precoding vector to change how its signal is received at a specific receiver. To do so, it multiplies the transmitted signal by a pre-coding matrix $\vec{v}$ as shown in Fig 2.5b. The process of aligning the signals on the same space is called interference alignment, and more details will be given in the next section.

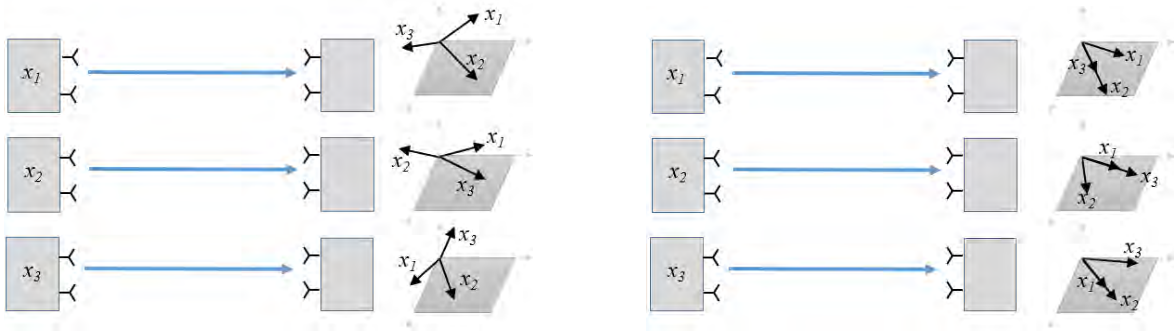
2.2 Interference Alignment

Interference alignment (IA) is a transmission technique that mitigates the effect of interference. The idea of interference alignment is to attempt to align (more accurately by coding) interfering signals over multiple dimensions such as time, frequency, or space (antenna), where aforementioned concepts of interference alignment are introduced in (Cadambe and Jafar, 2008). By coding over multiple dimensions, transmitted signals are constructed to compound (which is also expressed as aligning) the interfered signals received at each receiver into a lesser dimensional subspace, thereby allowing to maximize the number of desired signals that can be simultaneously communicated over the communication channel. Afterwards, simple interference cancellation algorithms at the receiver side can be facilitated. Our focus in this work is on spatial interference alignment. In this thesis, interference alignment (IA) refers to spatial interference alignment.

There have been some studies discussing that pre-processing transmitted signals at the transmitters to align interference at the receivers increases the total capacity (Cadambe and Jafar, 2008; Ali et al., 2008; Changho and Tse, 2008; Jafar and Shamai, 2008). Trying to obtain the maximum number of signals that can be achieved through interference alignment has also been researched in different network structures (Cadambe and Jafar, 2008; Ayach et al., 2013). Various studies look at the interference alignment from the theoretical perspective, however, some researchers have presented a system design and implementation to show that IA works in practise (Gollakota et al., 2009; Lin et al., 2011; Kumar et al., 2013; Adib et al., 2013). This work takes the same approach, and builds the system on the theory of interference alignment.

To illustrate the IA concept, consider an example shown in Fig 2.6. Three transmitters with two antennas each, send a single data stream to the receivers with two antennas each. Since the nodes have two antennas, the transmitted and received signals lie in two dimensional space. Each transmitter has a data stream for one receiver. If no coding at the transmitters is processed, i.e. if transmitters send their signals without processing them, after the signals pass the communication channel, receivers observe the received signals as shown in Fig. 2.6a. As discussed in the previous section, if a receiver has two antennas, it can only decode two signals. In other words, the receiver has two degrees of freedom (DoF). However, as shown in the figure, a receiver receives three independent streams lying in three dimensions, thus a receiver cannot decode these signals, it can only decode maximum two streams. Nonetheless, assuming the channel is known at the transmitters, the transmitter can pre-process (or perform coding) the signals before the transmission. In this manner, the interfering signals will be aligned at each receiver as shown in Fig. 2.6b. As one can figure out, the receiver now observes one desired signal and two interfering signals on the same space (as if they are just one signal now), giving two dimension to the receiver. Hence, the receiver can decode its desired signal within its two DoF. The pre-process is realized through a procedure known as pre-coding. In this procedure, the signal at the transmitter is multiplied by a precoding vector $\vec{v}$ before being sent to the communication channel. In other words, by changing the pre-coding vector $\vec{v}$, transmitter can control the vector along which the receiver receives the signal.

Let's define H_{ij} as the channel matrix between transmitter i and receiver j. In fact, when a transmitter desires to send a signal, it first multiples its signal by the precoding vector, then transmits the signal through the antennas onto the channel. Then, signal passes the channel and arrives at the receiver. Mathematically speaking, the signal is first multiplied by the precoding vector, then by the channel matrix. Hence, receivers observe the signals as shown in Fig 2.7.



(a) Without IA (b) With IA

Figure 2.6: Received streams

For the alignment, we need to ensure:

$$H_{21}\vec{v_2} = H_{31}\vec{v_3} \tag{2.3}$$

$$H_{12}\vec{v_1} = H_{32}\vec{v_3} \tag{2.4}$$

$$H_{13}\vec{v_1} = H_{23}\vec{v_2} \tag{2.5}$$

The three equations above align the packets at each receiver to make sure that two interfering signals lie along the same vector (space). These are three linear equations with three unknown variables (assuming channel matrices are known) and they can be solved easily. If we choose the pre-coders at the different transmitters based on the above constraint, we can realize the desired alignment at the receiver side.

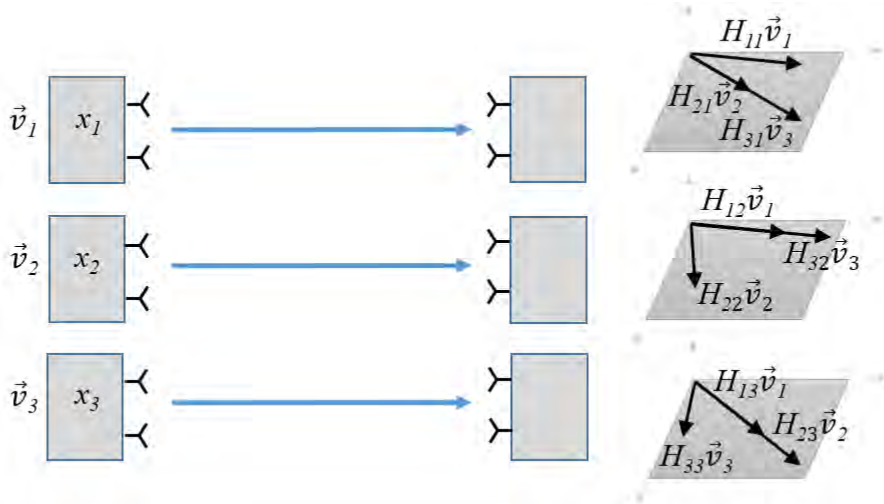


Figure 2.7: Interference Alignment

In decoding, as also described in section 2.1, projection is required to process. If the first receiver wants to decode its desired signal (which is x_1), it projects it on a vector orthogonal to the aligned interference, i.e. a vector orthogonal to $H_{21}\vec{v_2}$ and $H_{31}\vec{v_3}$. As it can also be seen in the figure, since these two vectors are aligned, there is a vector that is perpendicular to both of them, and hence the receiver can perform the decoding. If transmitters do not align the signals, the receiver can not decode the signals since $H_{21}\vec{v_2}$ and $H_{31}\vec{v_3}$ would have different directions, and there is not a single vector that is orthogonal to both. The vector that is orthogonal to interfering signals is referred to as decoding vector in this thesis. Moreover, one might ask here if it is possible for the first receiver to decode the signals x_1 and x_2. In fact, these two signals are the interfering signals for the first receiver. Thus, first receiver has no intention to decode them, it just regards them as the interfering signals. Furthermore, from the perspective of the receiver, since these two signals are aligned on the same direction and seen as if there is only one signal, receiver could not separate them to decode. So, if the first receiver desires to decode these two packets, it could not perform decoding for them.

Chapter 3

Our Mac Protocol

It is well known that wireless network can be two types, centralized network and distributed network. Centralized network is realized by a controller (e.g. Access Point (AP) or Base Station). Controller needs to manage the entire transmission both in uplink and downlink. Users should listen to the controller and act according to it. Controller is responsible for controlling the actions, maintaining the data transmission and acknowledging users. On the other hand, distributed network does not need a controller, users have to manage the transmission between themselves by communicating with each other. Our work is based on centralized network, where AP manages the communication in entire network. Furthermore, in order to be the controller in a centralized network, AP needs to have a scheduler. The goal of a AP scheduler in a wireless network is to schedule the downlink and uplink transmission of data packets to the numerous users in an effective and fair way. There exist various scheduling algorithms in wireless networks. Each scheduling algorithm has different purpose. Mostly, scheduling algorithms concern two parameters: throughput and fairness. In this sense, we demonstrate four client selection schemes to evaluate the trade-off between throughput and fairness, and then conclude the performance among them. These four selection schemes are Maximum Throughput Oriented Scheme (MT), Round Robin Scheme (RR), Remainder Choice Scheme (RC) and Proportional Fairness Scheme (PF). Note that scheduling is referred to as a selection scheme in our work. Whenever reader sees scheduling and selection scheme in this work, he immediately needs to pay attention that they have the same meaning.

In addition to four selections schemes, we also show how to reduce overhead and decrease the complexity in our system for a robust implementation. Briefly, choosing DL clients first, then choosing UL clients afterwards will result in low overhead and less complexity than choosing DL clients and UL clients jointly. Note that this approach (choosing first DL clients then UL clients) will be utilized no matter which selection scheme is used in our network. In other words, one can consider that our MAC algorithm prioritizes low overhead and less complexity, then client selection schemes will contribute to the trade-off between throughput and fairness. In the upcoming sections, we present our motivation for this work and the design of our MAC protocol.

3.1 Motivation

Since full duplex (FD) networks are studied for the past years, it is worth mentioning to distinguish its behaviour from half duplex networks. As we provide the definition of full duplex in the previous sections, we are now familiar with the concept of how a full duplex node operates, that is, a FD node can transmit and receive at the same time on the same frequency. One of the main challenges in this process is the self-interference (SI) generated by the Tx antenna at the Rx antenna. As discussed earlier, there have been some researches done to suppress this interference. Hence, in our work we assume the perfect SI cancellation whenever we utilize a FD node.

Furthermore, recent studies mainly focus on communication between two FD nodes. Considering the fact that FD is a new developing technology and there are still many HD devices being used in today's systems, it is plausible to consider a network where both FD and HD nodes exist at the same time. Since it is a burden to implement FD features to a legacy HD device, we consider our network with a FD AP and many HD clients, as FD features can be implemented easier to the access point (AP). In such topology, a HD client is transmitting to FD AP, while FD AP is transmitting to a HD client(s). Note that both uplink and downlink transmission use same frequency, since a FD AP can support such a technology. However, such a scenario introduces a new form of interference, called inter-node interference (INI), arises between uplink (UL) and downlink (DL) users as illustrated in Fig. 1.1. This type of interference significantly affects the performance of the system. (Sundaresan et al., 2014) shows that in a network with one FD AP, one HD client transmitting to AP, and one HD clients receiving from AP, INI can reduce SINR more than 10 dB even for relatively large distances between the UL and DL clients. Thus, the INI between the pairs should be eliminated for a solid communication.

To address the INI problem, this work takes advantage of interference alignment and MIMO. Specifically, if HD clients are equipped with multiple antennas (see Fig. 3.1), HD UL clients can pre-code their transmitted signals (which are interfering signals for HD DL clients) to align them at the HD DL clients. To do so, transmitters need to know the channel state information (CSI) between UL and DL clients. If CSI is available both at UL and DL clients, UL clients can design their precoders, while DL clients construct their decoding vectors for such alignment.

In such a network where FD AP and HD clients coexist, as discussed earlier, pre-coding is needed for

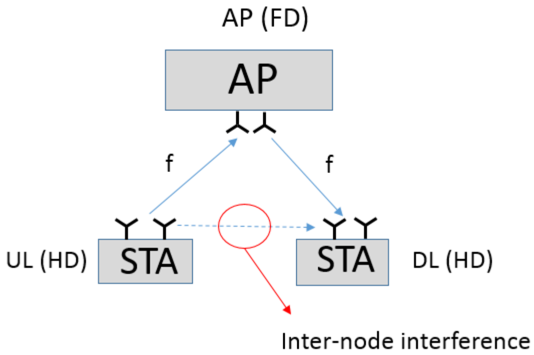
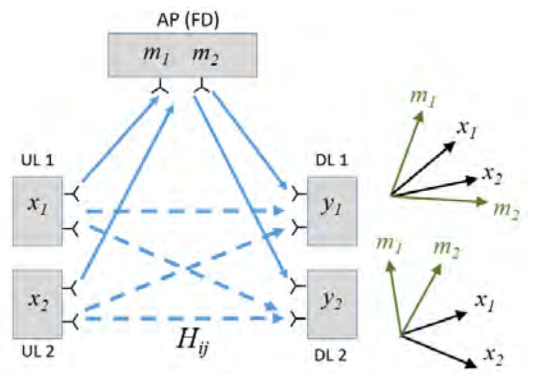
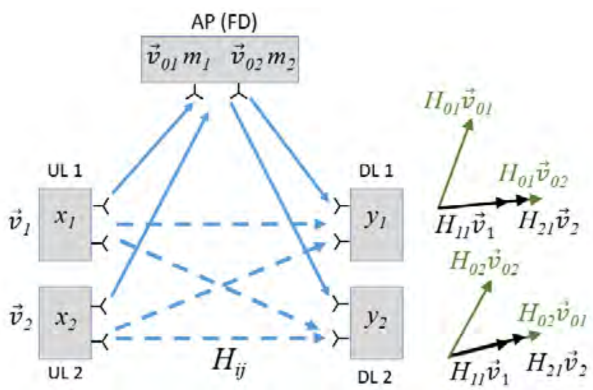


Figure 3.1: INI can be addressed through MIMO and IA



(a) Without IA



(b) With IA

Figure 3.2: Received streams in FD session

interference alignment. See Fig. 3.2a what receivers observe without interference alignment. DL clients receive x_1 and x_2 (black in the figure) as interfering signals. However, DL clients have their desired signals (m_1 for first DL client, m_2 for second DL client –green in the figure–) transmitted from AP. Without alignment, a DL client will receive 4 signals. For example, for the first DL client, the desired signal is only m_1, while other signals (x_1, x_2 and m_2) are interference signals for the DL client. Since a DL client has two antennas (DoF $=$ 2), it can only decode two (dimensional) signals. Thus, it is impossible for the DL client to decode its desired signal in this scenario. Now, let's look at the case with interference alignment. Since UL clients align their transmitted signals (by pre-coding) at the DL clients as if they are one signal, DL clients observe interference only in one dimension as shown in Fig. 3.2b.

Mathematically, alignment at the UL clients is as follows:

$$H_{11}\vec{v_1} = H_{21}\vec{v_2} \tag{3.1}$$

$$H_{12}\vec{v_1} = H_{22}\vec{v_2}. \tag{3.2}$$

The solution of these two equations are:

$$\vec{v_1} = eig(H_{11}^{-1}H_{21}H_{22}^{-1}H_{12}) \tag{3.3}$$

$$\vec{v_2} = H_{22}^{-1}H_{12}\vec{v_1}. \tag{3.4}$$

Note that, since channel matrix is 2×2, $\vec{v_1}$, it is composed of two eigenvectors, and we are free to choose one arbitrarily. After determining the $\vec{v_1}$, $\vec{v_2}$ will be determined according to the constraint above.

Nonetheless, pre-coding only at the UL clients is not enough for DL clients to decode their signals. Because, even DL clients receive the interfering signals along one direction, they also receive signals from AP. A DL client only has one desired signal from AP, but it also receives signals transmitted from AP to other DL clients. Signals for other DL clients also form interference for a DL client. Thus, apart from pre-coding at UL clients, AP also has to pre-code its signals for the DL clients as shown in Fig. 3.2b. Alignment process is easy:

$$H_{01}\vec{v_{02}} = H_{11}\vec{v_1} \tag{3.5}$$

$$H_{02}\vec{v_{01}} = H_{12}\vec{v_1}, \tag{3.6}$$

where H_{0j} is the channel matrix between AP (AP is represented by 0) and DL client j, $\vec{v_{01}}$ and $\vec{v_{02}}$ are the precoding vectors for AP, and y_j is the received signal for the j-th DL client. Hence, corresponding pre-coding vectors yield:

$$\vec{v_{01}} = H_{02}^{-1}H_{12}\vec{v_1} \tag{3.7}$$

$$\vec{v_{02}} = H_{01}^{-1}H_{11}\vec{v_1}. \tag{3.8}$$

Let us now consider the scenario where there are three UL clients and three DL clients as shown in Fig. 3.3. As similar to the two-antenna case, both pre-coding at UL clients and at AP are necessary.

In this case, three UL clients intend to transmit one stream to AP, while AP has one stream for each DL client. Streams transmitted from UL clients to AP will cause interference at the DL clients. UL clients can find

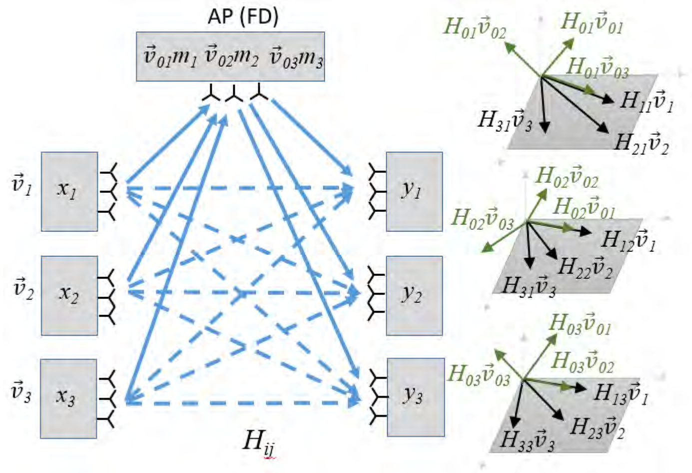


Figure 3.3: IA in three-antenna case

their pre-coding vectors to align their transmitted signals at the receivers as follows:

$$\vec{v_1} = eig(H_{11}^{-1}H_{21}H_{22}^{-1}H_{12}) \tag{3.9}$$

$$\vec{v_2} = H_{22}^{-1}H_{12}\vec{v_1} \tag{3.10}$$

$$\vec{v_3} = H_{33}^{-1}H_{13}\vec{v_1}. \tag{3.11}$$

Moreover, AP also needs to pre-code its signals for the DL clients as indicated below:

$$\vec{v_{01}} = H_{02}^{-1}H_{12}\vec{v_1} \tag{3.12}$$

$$\vec{v_{02}} = H_{03}^{-1}H_{13}\vec{v_1} \tag{3.13}$$

$$\vec{v_{03}} = H_{01}^{-1}H_{11}\vec{v_1}. \tag{3.14}$$

By doing so, it is assured that interference will lie on the same space.

One might ask here, what if we choose more clients? No matter how appealing one desires to transmit as many streams as possible, in our scenario we have some constrains. For example, consider the topology shown in Fig. 3.2. We have two two-antenna clients in UL and two two-antenna clients in DL. In this scenario, we have already shown that interference alignment is possible. Let us say, if we add a third client for UL, then we will have too many constraints, hence could not realize IA. In other words, we cannot find pre-coding vectors for three UL clients that are able to align all the interfering streams at the DL clients. In fact, two examples shown in Fig. 3.2 and Fig. 3.3 are studied in (Sundaresan et al., 2014), and a lemma is given:

“For N antenna HD clients, only four clients are necessary for IA, if N is even. If N is odd, six clients are necessary.”

In our work, we intend to build our MAC protocol on this statement. Our network consists of two types of HD clients; HD clients with two antennas and HD clients with three antennas, which is a common case in commercial networks (Halperin et al., 2010). For two-antenna HD clients, we can choose maximum two clients for UL, and two clients for DL transmission. Choosing more clients than this aforementioned lemma does not not yield interference alignment. Note that we do not have to choose two UL and two DL clients exactly, the interference alignment scheme will work even there is one two-antenna UL client and two two-antenna DL clients. Similarly, for three-antenna case, we are free to choose maximum three three-antenna clients for UL, and three three-antenna clients for DL. We provide our lemma:

Lemma 1:

- If the number of 2-antenna UL CLTs ≥ 2 and the number of 2-antenna DL CLTs ≥ 2, choose 2 CLTs for UL, choose 2 CLTs for DL.
- If the number of 3-antenna UL CLTs ≥ 3 and the number of 3-antenna DL CLTs ≥ 3, choose 3 CLTs for UL, choose 3 CLTs for DL.

Lemma 2:

- If the number of 2-antenna UL CLTs < 2 or the number of 2-antenna DL CLTs < 2, choose max 2 CLTs for UL, choose max 2 CLTs for DL.
- If the number of 3-antenna UL CLTs < 3 or the number of 3-antenna DL CLTs < 3, choose max 3 CLTs for UL, choose max 3 CLTs for DL.

These two lemmas basically state that 1) if there exist more than 4 two-antenna clients (minimum 2 clients for DL and minimum 2 clients for UL transmission) and 6 three-antenna clients (minimum 3 clients for DL and minimum 3 clients for UL transmission), choose maximum 4 two-antenna clients (2 clients for DL transmission and 2 clients for UL transmission) and 6 three-antenna clients (3 clients for DL transmission and 3 clients for UL transmission), 2) If there exist less than 4 two-antenna clients and 6 three-antenna clients, choose maximum number of existing clients. Note that minimum 5 antennas at the AP are sufficient to perform our MAC protocol.

Our MAC protocol aims to achieve two purposes, first, choosing best clients to achieve best performance in terms of throughput and fairness, second, reducing the overhead and decreasing the complexity of the system. It is known that there is a trade-off between throughput and fairness, i.e. if we want to maximize the throughput, we will have fairness issue in our network, which means, we only let clients with best channel conditions to transmit, hence clients with poor channel conditions will suffer to get a chance to transmit. Similarly, if we only focus on fairness, the throughput of the system will decrease, thus we cannot achieve good performance. Note that different fairness schemes are proposed in research of communications for many years, such as fair queueing (Demers et al., 1989; Greenberg and Madras, 1992), round robin (Miao et al., 2016; Nagle, 2003), weighted round robin (Greenberg and Madras, 2006), deficit round robin (Shreedhar and Varghese, 1996), proportional fairness (Kelly et al., 1998; B-Porat et al., 2011), and GPS-scheduling (Wang, 2005). Each one of these fairness schemes accomplishes different performances. In our network, we focus on four schemes (MT, RR, RC and PF) to recognize the performance of our topology, and we believe that these four schemes provide a proper insight to observe the behaviour of the system.

Moreover, there are some points needed to be emphasized. Our network is based on a centralized scheme, where AP acts as a controller. As presented previously, interference between clients needs to be aligned by pre-coding transmitted signals at the uplink users. Besides, AP also has to precode its signals for the downlink transmission. In order to obtain pre-coding vectors, channel parameters need to be known. Channel parameters are significant for a reliable communication in such scenario, thereupon, it is a feasible strategy to have AP as a controller. Putting AP in charge of being a controller makes sure that channel parameters for interference alignment will be known for the desired transmitters. Consider a scenario where AP does not act as a controller, contrariwise, clients try to capture the channel in a random way like CSMA/CA. In this scenario, not only some nodes suffer from not capturing the channel and not having a change to transmit, but also it becomes complicated for AP to track channel parameters. Furthermore, because of random access' nature, more delay might occur in the network if collision takes place between users. Thus, AP controlling the entire operation helps all clients for a reliable transmission. AP will demand from clients which packets to send, choose which clients should transmit to which client, and which clients should be served next. Here, AP is responsible for making decisions, clients do not have to burden with complex computations or complicated algorithms. Clients that are announced by AP only need to send packets that AP asked for. More specifically, UL clients need to send packets for channel estimation, and computing pre-coding vectors for transmission. Meanwhile, DL clients should estimate the channel parameters, feedback them to the AP for user selection, and prepare their receiver's parameters. This operation will be explained with more details in the next sections. For now, we present four selection schemes. Afterwards, we tend to present overhead-complexity analysis and our MAC protocol in deeper sense.

3.2 Selection Schemes

This section introduces four selection schemes for our MAC protocol, namely, Maximum Throughput Oriented Scheme (MT), Round Robin Scheme (RR), Remainder Choice Scheme (RC) and Proportional Fairness Scheme (PF). Our objective is to demonstrate these four selection schemes in details to lead the reader to comprehend the idea of each algorithm and their functionality.

3.2.1 Maximum Throughput Oriented Scheme (MT)

As mentioned earlier, we first look at how to maximize the throughput in our network. In this manner, MT scheme operates to achieve maximum throughput in the topology. So far, we have shown that UL clients and AP pre-code their signals for interference alignment at the DL clients. However, interference alignment does not directly contribute to improve the throughput. The reason is that pre-coding vectors are chosen according to the interference alignment, but, actual data transmission takes place between UL clients and AP, as well as between AP and DL clients. Otherwise stated, actual data transmission will be affected by pre-coding vectors, since pre-coding vectors are determined for interference alignment. Recall that decoding process at a receiver depends on the angle between the received signals (as long as DoF is satisfied) as shown in Fig. 2.4. For example, for a two-antenna receiver, an angle of 80 degrees between two received signals yields a better performance than an angle of 30 degrees. Note that an angle of 90 degrees would deliver the best result with no loss in the signals.

In the previous section, we introduced the selection process of the pre-coding vectors for both UL clients and AP. It is important to remember that pre-coding vectors of AP depend on the pre-coding vectors of UL clients. It also means that DL transmission explicitly depends on the UL transmission (more specifically, depends on the UL pre-coding vectors). Our design intends to choose the signals with high angle both at UL and DL transmission. In such a design, after AP collects the necessary information (we will give more details about collecting information in section 3.4) about the clients, it first makes calculations for the clients that provide higher angle in their transmission, and then inform those that could achieve the best result to start transmissions. However, these calculations are made between all possible UL-DL pairs in the network, which increases the complexity of the system and adds more overhead. Hence, our design adopts an approach to not looking at all DL clients at the same time, it just chooses two DL clients for two-antenna case, and three DL clients for three-antenna case. By doing so, AP will only calculate the corresponding angle (which will give the maximum throughput) between all UL clients and chosen DL clients. This approach results in decreasing the complexity and reducing the overhead in the system (overhead and complexity issue will be discussed more in sections 3.5 and 3.6). Note that AP is the coordinator in our system, and responsible for requesting information from clients, computing the calculations, and informing which clients to start transmission.

Our MAC protocol works as follows: 1) First choose DL clients: We previously showed that if there are many clients in our network, AP chooses two UL clients and two DL clients for the two-antenna case, and three UL clients and three DL clients for three-antenna case. Our design chooses two DL clients (for two-antenna case) or three DL clients (for three-antenna case). 2) Then choose UL clients: After determining DL clients, AP calculates the possible transmission rates between all UL and chosen DL clients. After the calculation, UL clients that yield the maximum throughput will be selected for the transmission. 3) Start transmission: After the selection of UL and DL clients, AP will inform which clients to transmit, and which clients it will transmit to. 4) Acknowledgement: After completing data transmission, AP informs UL clients about successful data transmission, hereupon, DL clients also send their ACK packets to AP for acknowledgement.

An Example Scenario

To illustrate our idea, let us look at an example of two-antenna case. In this setup, there are three two-antenna UL clients (namely, A, B and C), and three two-antenna DL clients (namely, K, L and M) as shown in Fig. 3.4. Recall that since it is a two-antenna case, AP chooses two UL and two DL clients.

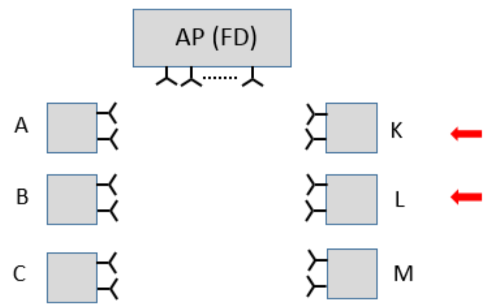


Figure 3.4: Three UL clients and three DL clients example

After AP knows the existing clients in the network, it first tends to choose DL clients. In our example, AP decides to choose K an M as DL clients. Then it calculates the possible angle values (transmission rates) between the chosen DL clients and uplink clients, i.e. A&B –K&L, A&C –K&L and B&C –K&L. The calculation process is shown in Fig. 3.5.

In Fig. 3.5, first case tells us what if AP chooses A and B as UL clients. In that scenario, AP receives the streams x_A and x_B from the UL client A and B, respectively, with an angle of 62 degrees. Furthermore, the

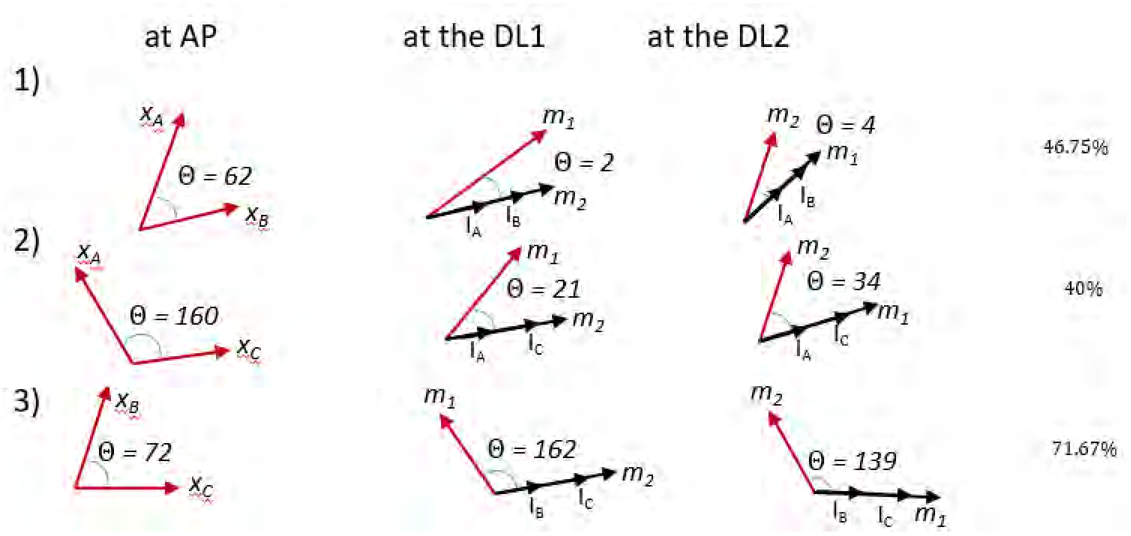


Figure 3.5: Throughput gain

receiver of K observes that the angle between its desired signal m1 and interference signals (I_A comes from UL clients A, I_bA comes from UL client B, and m_2 comes from AP) is 2 degrees. Similarly, DL client L receives its desired signal with an angle of 4 degrees. After decoding at each node, the total gain of this selection will give 46.75%. In other words, we will only obtain 46.75% of the transmitted signals. Note that in the decoding process, nodes will decode only their intended signals (those are shown red in the figure). Equivalently, second case will yield a 40% gain, and third case will yield 71.67% gain. Therefore, AP recognizes that third case maximizes the throughput, hence it will choose B and C as UL clients. At the end, B and C become UL clients, while K and M are chosen as DL clients.

3.2.2 Round Robin Scheme (RR)

While Maximizing Throughput Oriented Scheme (MT) targets at maximizing the throughput, it might be considered as unequal; since MT chooses the clients that enhance the transmission rate, some nodes in the network might suffer from not getting a chance to transmit. In order to be fair, AP has to let other clients to join the transmission. In this sense, round robin will allow each client to access the channel for transmission.

Round robin scheme is a scheduling scheme to choose all elements in a list equally in circular order, usually from the top to the bottom of a list and then starting again at the top of the list and so on. This procedure continues until all the tasks of the elements in the list are completed. Round robin algorithm can be applied to various operations such as computing, networking and so on. Generally speaking, each element in the list takes an equal share of a resource in turn. This resource is mostly a time, however, one can use round robin algorithm with other resources for instance frequency, packet size etc. For example, consider a communication network where AP can only assign a certain bandwidth to its users. In such scenario, the resource is bandwidth (frequency) and this resource should be assigned to clients equally, i.e. same amount of bandwidth should be given to each client. Therefore, each client will obtain same amount of bandwidth and the resource allocation is accomplished fairly. For time-sharing round robin algorithm, RR scheduler gives each task a time slot and interrupts the task if it is not finished by then. The task is proceeded in the next time slot that is appointed to that task. Note that time-sharing round robin scheduling is also called time slicing. The value of time slot is dependable on some parameters of the system. For example, if the processor is fast enough to process many computations, time slot can be a short time interval, since processor can complete the task in a short time. When processor needs more time for computation, time interval may be chosen longer, or task will be continued to be processed in the next time slot. Reader might consider (time-sharing) round robin scheme like ATM, where some clients line up to withdraw money from. Each client has a certain amount of time to complete his transaction, let's say, each client is supposed to have maximum two minutes to complete the transaction. If a client completes his transaction within two minutes, ATM will serve to the next client in the queue. If the client could not complete his transaction within two minutes, ATM will terminate his transaction process and start the next client's transaction. First client then has to move to the end of the queue, and wait for other clients' transaction completion to proceed his money withdrawal.

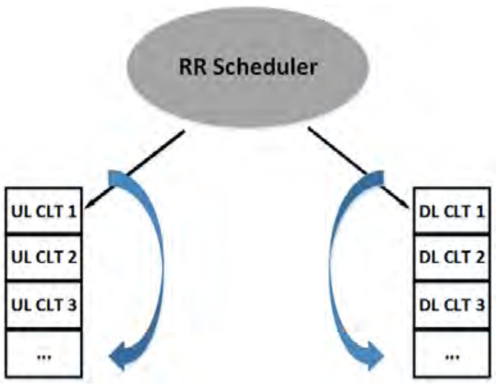


Figure 3.6: Round Robin Scheme

In our communication network, time is the resource for round robin scheduling, which means, when AP runs RR scheduling, it should devote certain time slots to certain user in turn. Thus, each user will have a chance to join the transmission no matter if its channel is good or bad, i.e. users can achieve high data rate or low data rate, depending on the channel state. This will result in equality among the users by giving them same amount of time. Here, one might claim that RR scheduling scheme has the potential to lower the data rate in the network, since allocation is realized based on equal time sharing to the users rather than data rate.

Round robin scheme in our network is shown in Fig. 3.6. Recall that for two-antenna clients, AP should select four clients. RR scheduler will give each client same amount of time to join transmission. However, as mentioned, AP should choose two clients for UL and two clients for DL. Thus, AP first chooses two DL clients (remember that our MAC algorithm always starts with choosing DL clients first then choosing UL clients), namely, DL client 1 and DL client 2. Afterwards, AP will let UL client 1 and UL client 2 transmit for UL communication. Reader can think this approach as a "group" round robin scheme, that is to say, instead of determining one client to transmit, AP groups the clients for the UL and DL transmission. In the next round, DL client 3 and DL client 4 will be chosen for DL transmission, correspondingly, UL client 3 and UL client 4 will be the users for UL transmission. This procedure will continue until all the clients will be served and given same amount of time to transmit. When all clients complete their transmissions turn by turn, AP will start again from the beginning, i.e. AP starts picking DL client 1 and DL client 2 for DL transmission, and UL client 1 and UL client 2 for UL transmission. For three-antenna case, the idea is same. In three-antenna case, AP should choose three DL clients and three UL clients. First, scheduler constitutes DL1, DL2, and DL3 for DL transmission, then UL1, UL2 and UL3 for UL transmission. After one round of transmission, AP will schedule DL4, DL5 and DL6 for downlink transmission, and UL4, UL5 and UL6 for uplink transmission, and so on.

3.2.3 Remainder Choice Scheme (RC)

Another selection scheme that is applied in our topology is remainder choice scheme. So far, we have seen maximum throughput oriented scheme, which aims at choosing clients to maximize the network throughput (MT), and round robin scheme (RR), which gives equal amount of time to the clients. In MT, some nodes do not have a high opportunity to receive/send since AP does not select them for transmission. Hence, although MT contributes higher data rates to the network, it does not provide equality between users. On the other hand, RR is purely an equal scheme, because no matter what condition clients have, they will have a chance to join the transmission. RR scheme might increase the data rate, but, because its focus on equal time sharing rather than network throughput, it will most probably bring low network throughput.

Remainder choice scheme tries to combine the maximum throughput algorithm from MT and equal time sharing idea from RR. The purpose of RC is to be equal to all clients and achieve higher throughput in the network at the same time. The selection process goes as follow: in the beginning, as usual, AP chooses DL clients first. AP will store which DL clients are served. Then UL clients will be chosen according to the maximum throughput algorithm discussed above. Again, AP will store which UL clients are served. In the next round, AP will choose other DL clients who did not participate in the previous round. UL clients with best channel who did not take part in the previous round will join the transmission. This process will continue until all the clients are picked at least once in the network. Afterwards, the algorithm will start from the beginning by choosing the DL clients first, and then choosing UL clients that will achieve higher data rate. Reader might discover that RC is a mixture of MT and RR. In fact, what RC actually does is the combination of both. Choosing clients that did not participate in the previous round is a similar idea as round robin, and selecting the clients that will provide high throughput is the idea behind MT. Actually, remainder choice name comes from choosing among remainders. One short notice is that RC scheme is more close to MT than to RR. One can easily claim that RC is also concentrates on higher throughput, while giving each client an opportunity for transmission. An example of RC is shown in Fig. 3.7. In this example, there are 5 two-antenna DL clients and 5 two-antenna UL clients. Scheduler first picks DL1 and DL2 as downlink clients, then it computes the possible UL-DL pairs (DL1-DL2 & UL1-UL2, DL1-DL2 & UL1-UL3 and so on). After estimating the throughput for each pair, AP decides on UL1 and UL2 for UL transmission. In the second round, AP cannot choose DL1 and DL2 for downlink transmission since they were already engaged in the first round, same for UL1 and UL2. In the second round, AP first selects DL3 and DL4 as downlink clients, then computes the possible throughputs for DL3-DL4&UL clients, and finds out that UL3 and UL5 will result in higher capacity. Same logic applies in the coming rounds. Note that idea is the same for three antenna clients, with a difference of choosing 3 clients for both UL and DL transmission.

	UL1	UL2	UL3	UL4	UL5	DL1	DL2	DL3	DL4	DL5
R1	•	•				•	•			
R2	x	x	•		•	x	x	•	•	
....										

Figure 3.7: Remainder Choice Scheme

Let's briefly look at the simplicity of the three selection schemes we have considered so far, namely, maximum throughput oriented scheme, round robin scheme and remainder choice scheme. Round robin scheme is fairly simple and easy to implement. In RR, AP does not have to calculate or compute any metric such as throughput estimation (which is the case in MT and RC), on the contrary, AP just equally allocates certain time slots to each user in turn. Here, what AP has to know is just which clients would like to be receiving from AP, and which clients would like to transmit to AP. Selection process does not cover anything else, AP needs to record the clients that are chosen, hence, in the next round, AP will be aware of which client should be served next. MT includes many computations to determine the clients who will reach the maximum throughput. After choosing DL clients, scheduler at the AP will examine all the possible uplink- (chosen) downlink pairs, then calculate the angles for each pair which will give the throughput, and AP decides to pick the ones that have higher throughput. As one can see that this selection process consists of large number of calculations. This is the consequence if we would like to achieve higher data rate in the system. Nevertheless, as we will see in the complexity analysis in section 3.6, the number of computations might be even larger if AP looks at all DL clients (instead of determining DL clients first), and pairs them with all the UL clients, then makes the calculations. Thus, choosing DL clients first is a good strategy to decrease the complexity. Finally, RC's simplicity is more than RR but less than MT. As stated previously in MT, after determining DL clients, AP has to calculate all possible DL-UL pairs to detect the clients for maximum throughput. RC also calculates the pair's throughput, and make the decision based on the calculation. However, in RC, next rounds will consist of fewer calculations, since AP will ignore clients that participated in the previous rounds. This will result in less computation at the AP side, hence one can regards RC as a simpler scheme than MT.

3.2.4 Proportional Fairness Scheme (PF)

We now present our last selection scheme, i.e. proportional fairness scheme. PF is a scheduling algorithm that tries to manage equilibrium between two objectives: Attempting to maximize total data rate and ensuring all users to have at least essential level of service simultaneously. In other words, PF intends to enhance throughput performance by taking advantage of channel alteration while preserving fairness among the clients. This is enabled by appointing each user a scheduling metric that is proportional to its anticipated data rate and average data rate that the user consumed until then. Note that compared to MT, RR and RC, PF is the first scheme that tries to maintain the throughput-fairness balance between the users.

In our network, AP operates on a time-sharing manner, i.e. TDMA based scheme. All selection schemes we have discussed so far rely on TDMA based operation, where AP is the controller and assigns certain time period to its users both in uplink and downlink transmission. This time period is divided into time slots and the AP dedicates each time slot to send data to a specific user. Time slot in our communication network depends on the coherent time of the channel for the intended user. Our research assumes that assigned time slots cover coherence time of the channel for the specific user. Stated in other words, when actual data transmission occurs for a certain user, the channel that the user utilizes does not change until the data transmission is completed. Scheduler determines which of the clients will be served. In TDMA based schemes, scheduler makes the decision by using priority coefficient to each user for every time slot. The user with the highest coefficient value will be selected for transmission. Mathematically, as i denotes the user index,

$$Q_i(t) = \frac{R_i(t)}{A_i(t)} \tag{3.15}$$

Where $R_i(t)$ is the estimated data rate in time slot t. $R_i(t)$ depends on the SNR value after decoding procedure at the receiver. As emphasized earlier, this SNR value also relies on the angle between the received signals arrived at the decoder. Recall that the higher the angle is, the less will be lost from the SNR. $A_i(t)$ is the throughput average until time slot t (not including), and $Q_i(t)$ is the corresponding priority coefficient. User with the highest $Q_i(t)$ value will be selected to to join transmission in time slot t. The average throughput is updated after every time slot by,

$$A_i(t+1) = \frac{(t-1)}{t} A_i(t) + \frac{1}{t} R_i(t) S_i \tag{3.16}$$

As one can immediately see that updated average throughput depends on average throughput and data rate. In equation 3.16, S_i is added to data rate. If user i gets the channel for transmission in time slot t, S_i will be assigned 1, otherwise, if user i does not get the channel for transmission in time slot t, S_i will be assigned 0. This means that in the next time slot, it is harder for a user to acquire the highest priority coefficient. This is the consequence of a user that joined the transmission in time slot t. If a user has a chance to join transmission, his updated average throughput will be higher, and this will decrease his potential priority coefficient in the upcoming time slots. Furthermore, let us look at equation 3.15. The name of proportional fairness comes from the proportion of data rate and average throughput. Specifically, let us consider a user whose average throughput is 2 Mbps due to poor channel conditions until now, while assuming network's average throughput is 5 Mbps. It is obvious that this user cannot contribute too much to the network throughput, however, this user also needs to have a chance to engage in the transmission for the purpose of a fair network. If this user has a potential to achieve 10 Mbps for the current time slot, because of his high priority coefficient, AP will schedule this user for the transmission. In this sense, even though this user could not obtain high throughput previously, it is now his chance to achieve high throughput. This idea will lead network to be more fair to the "poor" users and accomplish bigger throughput values in network. Furthermore, when looking at the equation 3.16, one can realize that, after a large amount of time, updated average throughput will mostly depend on average throughput, since for large t values, $\frac{1}{t}$ will become less. Hence, updated average throughput will not be affected by the user whether he utilized that time slot for the transmission or not. One point to mention is that if S_i is assigned by 1, scheduler will run maximum throughput oriented scheme.

As mentioned, our MAC protocol first chooses downlink users, then uplink users will be determined. AP will first announce which downlink clients are elected, then it will collect necessary information (channel parameters) to adjudicate which uplink clients should be picked for that time slot. Remember that using equation in 3.16 will help AP to make a decision who to choose. However, choosing downlink users before AP collects information to make decision means that downlink clients should be chosen before the intended time slot. Chronology take places in the following order: 1) AP first decides DL clients, 2) AP will ask for channel parameters for the intended time slot, 3) UL clients will send frames for channel estimation, 4) DL clients feedback channel parameters to AP, 5) AP makes decision which client to transmit/receive. Here, step 2 to 5 takes places for the intended time slot, while step 1 happens before the intended time slot. In this manner, AP cannot simply benefit eq. 3.15 to choose downlink clients. So, AP requires to employ modified equation as given in 3.17. Only modification in this equation is that AP calculates estimated throughput from the previous time slot instead of the current time slot. Note that average throughput remains the same since average throughput is calculated until the intended time slot, therefore no change is needed for the denominator of the equation.

$$Q_i(t) = \frac{R_i(t_{last})}{A_i(t)} \tag{3.17}$$

As a last point, simplicity of the proportional fairness selection scheme should also be discussed briefly. Since PF scheduler needs to calculate the priority coefficient in eq. 3.15 and eq. 3.17, this scheme is a bit more complicated than other schemes as discussed earlier. Because RR does not necessitate to compute any metric or equation, it is considered as the simplest scheme among our selection schemes, while MT and RC are required to calculate the estimated throughput (through angle between received signals). PF takes a small step forward and needs the calculate the priority coefficient. This coefficient relies on, firstly, estimated throughput in the intended time slot for transmission, and secondly, average throughput until that intended time slot. In short words, while MT and RC are only responsible for computing estimated throughput, PF needs to compute both estimated throughput and average throughput. One might question the necessity of using PF scheduler due to its more complex structure, however, as we will see in the section 4, PF ensures desirable throughput-fairness trade-off.

All in all, each selection scheme obtains different performance in terms of fairness and throughput. We proposed four selection schemes. One can also design different selection schemes for the MAC protocol. Besides, varied combinations of the four proposed selection scheme also can be considered. In this work, four selection schemes help reader to observe different scenarios. For example, MT scheme achieves the highest throughput, hence it is a reference point to see the highest throughput in our network. After other three selection schemes achieve their throughput performance, they can be compared to MT scheme to examine the throughput difference between them and MT scheme. Similarly, as shown in the next sections, PF accomplishes best fairness in the network. Comparing other selection schemes with PF assists reader to see the distinctness with respect to fairness. Combination or other selection schemes designs are left for further studies. We believe that proposed selection schemes in this work provide an insight to see distinct performance from the viewpoint of fairness and throughput.

3.3 Fairness Index

There are some crucial concepts in communication technologies that are addressed by researches in academia and industry, and fairness plays an important role among these concepts. Fairness is a multidisciplinary subject which is usually related to resource allocation. Resource in this definition extends to various topics. For instance, let's consider a company with many employees that are simply doing the same job under same circumstances. Company naturally has to pay salary to its employees. Considering that all employees are equal and doing the same job, they expect to earn same amount of money. Hence, company should provide same salary to its employees, which is regarded as the fairest way. If an employee earns less than the other, it will be unfair for that employee. Here, resource allocation is fulfilled based on money. Similarly, in communication networks, different resources are expected to be shared fairly amongst all operations. All users wish to acquire the bandwidth and also the quality of service (QoS) fairly.

To illustrate the idea of fairness in wireless communication, let's look at an example, where AP communicates with four users namely, x_1, x_2, x_3 and x_4 as shown in Fig. 3.8. Users x_1, x_2, x_3 and x_4 communicate with AP. AP assigns communication channel to the users in wireless access manner. In this simple example, some fairness related points can be examined. For example, nodes are supposed to obtain fair opportunity to access the channel, bandwidth should be shared fairly, nodes should get fair Quality of Service and so on. Briefly, various performance issues have to be plausible and fair. In order for a system to be fair, there is no single approach to be fair at all times. Fairness should be defined according to the network needs and implementation of the system. For instance, in our example, let's suppose that user x_1 and x_2 pay more money to the service provider to be served more frequent than x_3 and x_4 in order to get more bandwidth. AP logically should give more chance to the users x_1 and x_2. In this sense, x_3 and x_4 cannot claim to demand more opportunity to access channel since priority is users x_1 and x_2. However, in our network, we assume that all clients are equal and they are supposed to be served in a fair way.

A question might arise in reader's mind, how to measure fairness? To determine whether a system is fair or not, fairness measures are used. Fairness measures or metrics are used in computer/network architecture to dictate whether users or applications are receiving a fair share of system resources. Even though there are several ways to measure fairness, we benefit *Jain's Index* in this work. Reader might refer to fairness survey of in (Shi et al., 2013) to learn about more about fairness measures. Jain's index was first proposed in (Jain et al.,

Figure 3.8: A simple illustration of a wireless network with one access point and four users

1984), and it follows four features:

- Population size independence: The index should be applied to any number of users.
- Scale and metric independence: The index should not depend on any scale or metric, i.e. the measurement can be applicable to any unit.
- Boundedness: The index should be a value between 0 and 1.
- Continuity: The index function should be continuous.

Jain's index is defined as,

$$F = \frac{(\sum_{i=1}^{n} x_i)^2}{n(\sum_{i=1}^{n} x_i^2)}, \tag{3.18}$$

where $0 \leq F \leq 1$. F is the fairness index in the equation, and it measures the fairness of any system. For the purpose of our design, n represents the number of clients in the network, i is the user index, and x_i is the average throughput of the user i. We run our design repeatedly, at the end each user will achieve an average throughput. This average throughput will imply how much throughput a user gets from the total throughput, and fairness index will tell us how fair the throughput is divided among users. To understand fairness index better, let's look at the table 3.1. This table actually shows a demonstration of the example illustrated in Fig. 3.8. Recall that, there exist four users x_1, x_2, x_3 and x_4. In case 1, x_1 gets most of the bandwidth (70%) in the network, while x_3 never had a chance to access the channel for transmission. In this case, as table also displays, bandwidth is not allocated equally among users, therefore, fairness index value becomes 0.4505. On the other hand, if bandwidth of the system is divided equally among users (case 4), fairness index reaches 1. Note that allocation is fairer when fairness index is closer to 1, and less fair when it is closer to 0.

	Case 1	Case 2	Case 3	Case 4
x_1	70%	40%	35%	25%
x_2	5%	0%	30%	25%
x_3	0%	50%	20%	25%
x_4	25%	10%	15%	25%
F	0.4505	0.5952	0.9091	1

Table 3.1: Fairness Index

There is one last point needed to be mentioned about fairness index. When we presented round robin selection scheme, we also referred to equality/fairness among users. However, fairness demonstrated in round robin selection scheme and fairness illustrated in Jain's index have a slight difference. In RR, idea to be fair is about giving equal chance to the users to access channel for transmission. Hence, in RR, each user has an equal opportunity for transmission. But, fairness index specifies fair resource allocation between users in terms of bandwidth. In this work, if not otherwise stated, whenever we imply fairness, we refer to the concept of fairness index that is described in this section.

3.4 Timeline

In this section, we present the operation timeline of our MAC design. Our MAC design is a TDMA based scheme, where AP is the controller and gives certain timeslots to the users to transmit/receive. To illustrate our idea, let us consider the example scenario discussed in section 3.2.1. Operation starts with AP sending "P" packet. It is a broadcast packet to inform all users, therefore, all UL and DL users receive this packet. This packet includes which DL clients are chosen, so each user now knows selected DL clients. Furthermore, this packet also notifies UL and DL users to poll them in a specific order. After UL and DL users receive this packet, they all know which UL and DL clients will send packets to AP in which order. Upon receiving P packet, each UL client sends an "R" packet in polling order that was announced in P packet. DL clients and AP will receive R packet; AP will use this packet to determine the channel state information of the uplink transmissions, i.e. channel coefficients between UL users and AP, while DL clients use this packet to estimate the channel between UL and DL clients (which can be referred to as interference channel). R packet is crucial in our network, since both AP and DL users should examine this packet for their purposes. AP resolves this packet to identify channel parameters between UL and AP, and DL clients determine interference channel parameters (channel between UL and DL clients, i.e. INI). At this point, DL clients now know the channel parameters between UL and DL clients, which will be used for interference alignment by UL clients. Both of these channel parameters will be utilized by AP to calculate necessary pre-coding vectors. However, DL clients determining INI channel parameters is not enough in our network, hence, DL clients need to give this information to AP. After all UL clients send their "R" packets, chosen DL client sends "C" packet to the AP in a specific order that was declared in P packet. Note that only chosen DL clients will feedback packet C (in Fig. 3.4, K and L are chosen DL clients, so M will not participate in operation), thereby, reducing the overhead in the system. Packet C includes the CSI between UL and DL clients. Furthermore, AP will also use packet C to estimate CSI between AP and DL clients. AP in this moment has all essential information, that is, CSI between UL and DL clients, CSI between UL clients and AP, and CSI between (selected) DL clients and AP. Here, since AP knows all the channel state information, it can make the calculations as discussed in section 3.2. Remember that AP makes calculations according to the operating scheduler, which can be either MT, RR, RC or PF. After the calculations, AP will send a "start" packet to inform which UL clients are chosen. Additionally, because AP has all the information for calculations, it can compute pre-coding vectors as well, hence, it computes the pre-coding vector $\vec{v_1}$ and sends it as well. Chosen UL clients use this information for themselves, i.e. client that is assigned for $\vec{v_1}$ will use this pre-coding vector for its transmission, while client assigned as a second client will compute its pre-coding vector $\vec{v_2}$ based on $\vec{v_1}$. The reason behind pre-coding vector $\vec{v_1}$'s announcement by AP for the first UL client is to avoid UL clients to compute its pre-coding vector by itself. Because AP can calculate this pre-coding vector, it helps uplink client to avoid extra computation. However, even though AP can also calculate second client's pre-coding vector, it does not announce this pre-coding vector in order to keep overhead low in the network. Remember that pre-coding vectors are calculated with equations presented in section 3.1. Moreover, AP will also compute its pre-coding vectors for the DL transmission meanwhile. In Fig. 3.9, selected clients are marked with red arrow. After the selection process, chosen clients will start transmission at the same time. When transmissions are completed, AP will send a block ACK to inform UL clients, just after that, each DL client will send its ACK. This whole procedure is just one round, in the following rounds, AP will start again with sending "P" packet to inform DL clients that are chosen, and so on. Note that there are two moments when AP makes selection. First time is when AP broadcasts P packet, which is for DL client selection and polling. Second time is in start packet for UL selection. Both decisions are made upon the selection schemes discussed in section 3.2. However, when comparing computations in P packet and start packet, start packet requires more computation. This is because, in P packet, AP only determines DL clients, on the other hand, start packet consists of UL clients selection, and pre-coding vectors computation for both UL and DL clients. Whole operation is shown in Fig. 3.9. As stated formerly, our network concerns clients with different number of antennas. Example in Fig 3.9 only shows the scenario with two antennas. After understanding the example in Fig 3.9, it is relatively easy to comprehend the scenario where two-antenna and three-antenna clients exist in the network at the same time.

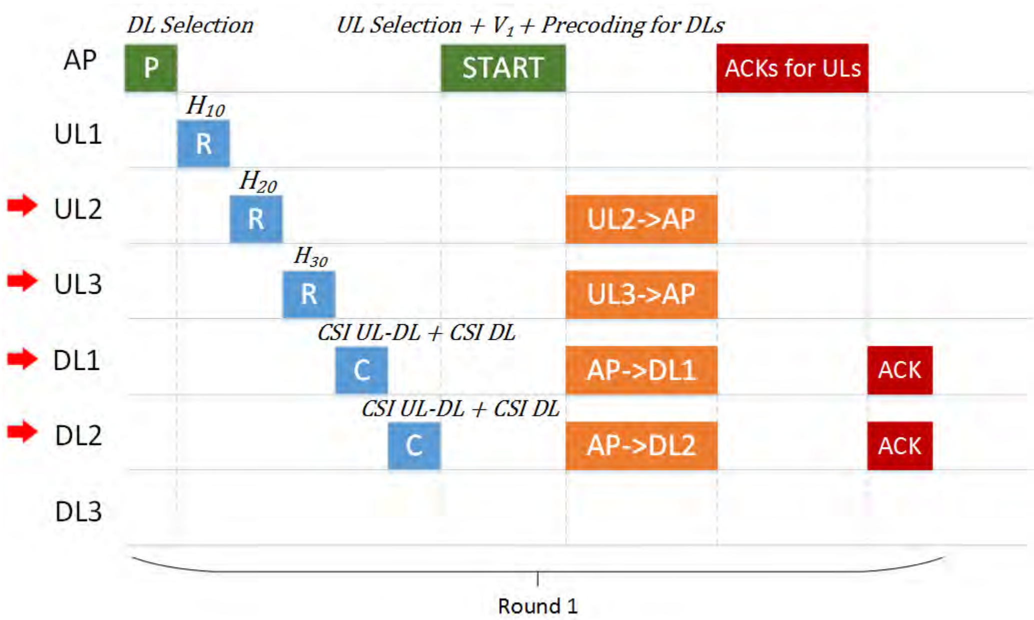


Figure 3.9: MAC protocol timeline for two-antenna users

The idea is identical to the two-antenna case. Reader may think that two-antenna and three-antenna cases are independent from each other. In other words, AP runs the same MAC algorithm for both two-antenna case and three-antenna case. Two-antenna and three-antenna cases are divided in TDMA based way. The difference here is that AP is supposed to consider selection scheme for both cases (two-antenna and three antenna cases), correspondingly, it should make more calculations to determine pre-coding vectors and to select clients. Note that polling of P packet first calls for two-antenna clients, then three-antenna clients. Similarly, P packet should also specify both two-antenna and three-antenna DL clients. This procedure is shown in Fig 3.10. Pay attention that superscript indicates antenna number.

3.5 Reducing Overhead

The role of MAC protocols is to coordinate the access to the channel to make it possible for several users to communicate within a multiple access network that incorporates a shared medium. In order to control the communication, network has to benefit "non-data" packets that are required to attain a particular goal, which is referred to as overhead in communication networks. A typical example of overhead is RTS/CTS/ACK packets in 802.11 CSMA/CA. These packets do not contain any "real" data for the destined users, conversely, these packets help the network to establish a communication link between nodes (RTS/CTS), or to acknowledge if the transmission succeeded (ACK). These packets are essential in a network for a reliable and robust communication,

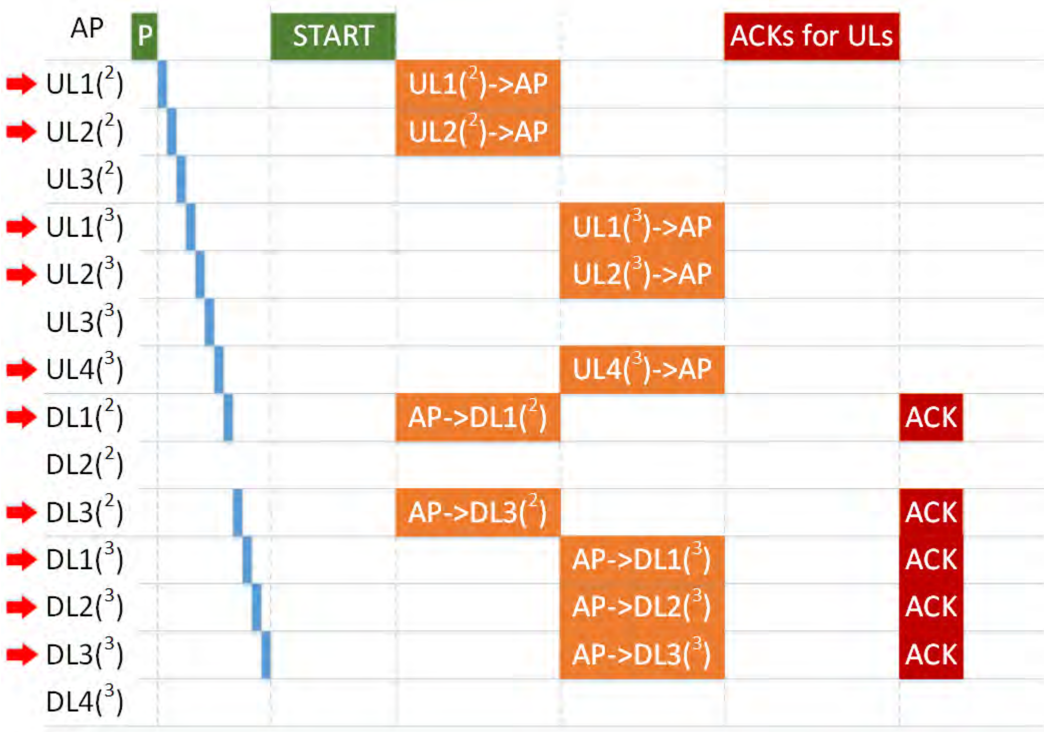


Figure 3.10: MAC protocol timeline for two- and three antenna users

however, they consume network's resources such as time, bandwidth, memory etc. Consequently, it is not desired to utilize large number of overhead in a communication network. Smart designs always try to keep overhead as less as possible.

Another type of overhead is brought by framing in OSI model (see Fig. 3.11). OSI model divides the communication process into seven layers and provided a framework for overall communications process. Each layer can be considered independently from each other, i.e. each layer has its own protocols, standards and services. A layer can act individually, even though two or more layers might cover some concepts together. Each layer adds header to the block of information it accepts from layer above. This type of header increases the network's overhead, nevertheless, communication network needs to sacrifice its resources to some header portion to apply certain protocols and standards.

Some studies pay close attention to the overhead problem in communication systems. For example, in (Lin et al., 2012), authors present a light-weight wireless handshake mechanism to overcome overhead issue in their MAC protocol. Furthermore, (Dunn, 2010) analyses overhead constraints in wireless networks, and it is shown that on average approximately 71% of an 802.11b packet transmission is used for communication of the end user's data, the remaining 29% of a packet is consumed by overhead in the packets. In our wireless network, packets P, R and C cause overhead to our system. Particularly, R and C packets potentially might increase

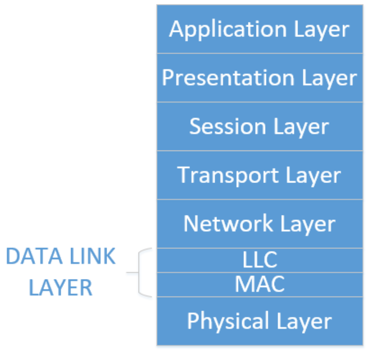
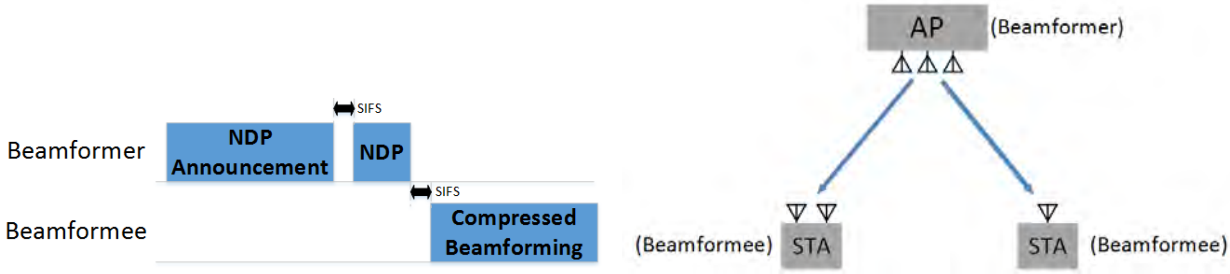


Figure 3.11: 7 Layer OSI Model

overhead in our network, since they depend on the number of uplink and downlink users in our system.

We reduce overhead in our network with two approaches. First approach is to restrict number of downlink users. In our MAC protocol, after AP broadcasts P packet, uplink and downlink users send R and C packets in polling order. These packets are vital in our design, because this packets will be used to determine channel coefficients, accordingly, pre-coding vectors for transmitters. P packet announces chosen downlink clients (2 DL clients for two-antenna case, and 3 DL clients for three-antenna case). If P packet does not decide DL clients first, all DL clients in the network are supposed to send C packets. Since AP limits the number of DL clients, our network has to deal with certain number of DL clients (maximum 5 DL clients in two- and three-antenna cases). This will result in significant overhead reduction. Second approach is to use PHY layer packets instead of using MAC layer packets. As explained earlier in this section, OSI layer is divided into seven layers. MAC layer belongs to layer 2 (Data Link Layer), while PHY layer is layer 1. Because of the framing nature of layers, a MAC packet consists of MAC packet and PHY packet. In other words, if we want to use MAC packet, we also have to add PHY packet to send this MAC packet in real environment. In this sense, using MAC packet introduces more overhead to the network. Hence, if PHY packet can do the same job as MAC packet, network can avoid using MAC packets, thus reducing the overhead to a greater extent. In fact, since AP limits DL clients to a certain number, accordingly, it also means that less C packets will be sent to AP. Notwithstanding, there is no limit for UL clients, so the network might experience large number of R packets, equivalently, high overhead. As most of the overhead is caused by R in our wireless network, we would like to design PHY packets for our R packet, rather than using MAC packets. Within this direction, our network benefits similar procedure that 802.11ac proposes.

802.11ac defines transmit beamforming technique. A device that forms its transmitted signals is called a beamformer, and a receiver of such signals is called a beamformee. A single device may behave both as a beamformer and a beamformee. In 802.11ac, transmit beamforming is enabled only at AP, therefore, in



(a) Sounding Procedure (b) Beamformer and Beamformee

Figure 3.12: Sounding procedure in 802.11ac

802.11ac, beamformer is AP, and beamformees are clients. Beamforming in 802.11ac is a method that AP transmits its signals in the direction of the intended users. Transmit beamforming needs the information of the channel parameters to calculate a steering matrix that is employed to the transmitted signal to optimize reception at one or more receivers. The steering matrix allows transmitter's antennas to shape their signals towards targeted users. In our system, steering matrix is equivalent to the uplink users' precoding vectors. Transmit beamforming is performed by a process called "sounding". Sounding procedure works as follows: the beamformer starts the process by sending Null Data Packet Announcement (NDP-A), which identifies beamformees. After beamformees receive this packet, they put off channel access until sounding sequence is over. NDP-A assigns Association ID (AID) to the clients, this identification helps beamformer and beamformees to communicate in PHY layer packet level, instead of using MAC layer packets. Beamformer then sends Null Data Packet (NDP). This packet is also referred to as null data packet since it does not contain any data. NDP packet is a PHY packet, it also includes AID, so clients do not rely on MAC packet to understand that the packet is for themselves. Using NDP packet leads to not using MAC packets to identify clients. Since communication network does not use MAC packet here, overhead is reduced. Beamformees, which are assigned in NDP-A packet to respond back to the beamformer, use this packet to analyse channel and estimate channel parameters, and hence the steering matrix. Beamformees examine NDP packet for channel estimation, and compute a feedback matrix. They then send feedback matrix back to the beamformer. The beamformer receives the feedback matrix and calculates the steering matrix to direct transmissions toward the beamformees. Reader might refer to (Perahia and Stacey, 2013; Gast, 2013; IEEE, 2013) for more details about sounding process in 802.11ac. Sounding procedure is shown in Fig 3.12a.

In sounding procedure, NDP-A and Compressed Beamforming are MAC layer packets, while NDP packet is PHY layer packet. Before sounding procedure begins, stations must be in the appropriate connection state with AP. In this manner, authentication and association are the first steps in network attachment for stations.

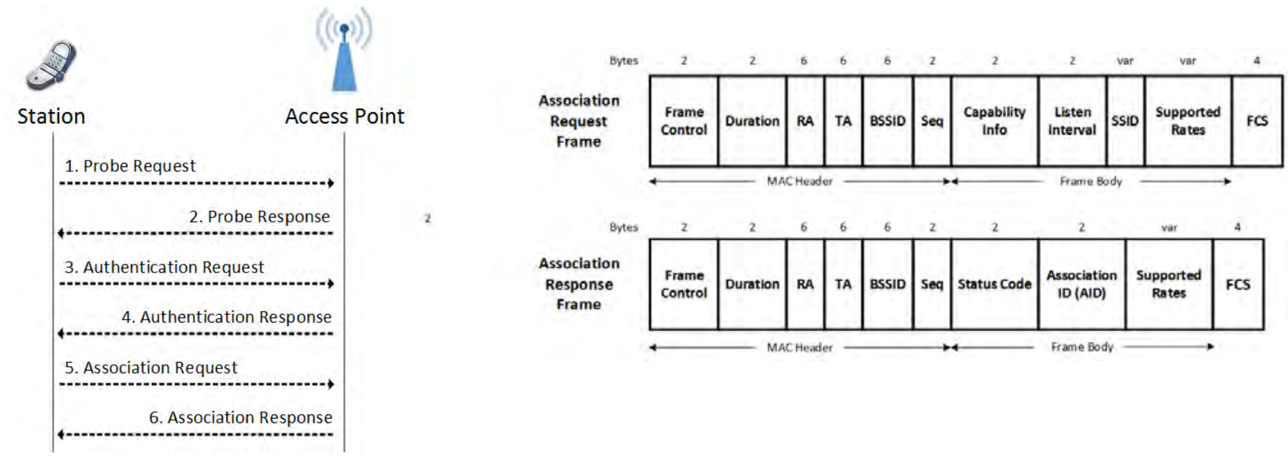
First, a station sends probe request to discover 802.11 networks within its proximity. Probe requests advertise the station's supported data rates and 802.11 capabilities. AP receiving the probe request check to see if the station has at least one common supported data rate. If they have compatible data rates, a probe response is sent advertising the SSID (wireless network name), supported data rates, encryption types if required, and other 802.11 capabilities of the AP. Afterwards, an authentication request is sent from the station to AP. Next, an authentication response with a success message is delivered from AP to the station. Once authentication is complete, stations can associate (register) with an AP to gain full access to the network. Association allows the AP to record each mobile device so that frames are properly delivered. If the elements in the association request match the capabilities of the AP, the AP will create a unique 16 bits Association ID (AID) for the station and respond with an association response with a success message granting network access to the mobile station. Fig. 3.13a shows the authentication and association steps. Pay attention that packets in steps 1-6 are MAC layer packets (more specifically, management frames) that contain station and AP's MAC addresses. In association process, AP assigns a unique AID to the station based on the MAC address. Thus, AP utilizes MAC address of the station to appoint an AID (an integer value) from the range 1-2,007 (Fig. 3.13b).

Algorithm 1: Assigning unique AID

```
initialization:
int i, j=0;
Create an integer list of AID: AID[]={1,2,..,2007};
if i-th client associates with AP then
    Assign AID[j] in the list to the i-th client;
    j++:
end
```

For example, suppose that there are two users (user1 and user2) in the network. Let's define their 48 bits MAC addresses as user 1 $\rightarrow$ AB:.. and user 2 $\rightarrow$ CD:.. . Upon association with AP, AP assigns them an integer value (16 bits unique AID) with respect to their MAC addresses: user 1 $\rightarrow$ AB:.. $\rightarrow$ 1 and user 2 $\rightarrow$ CD:.. $\rightarrow$ 2. 1 and 2 are unique AID given for the user 1 and user 2. In order for AP to guarantee unique AID assignment to the users, AP first creates a list of all possible AIDs. Next, AP assigns elements in the list one by one to each user. By doing so, it is assured that each client has a unique ID. Note that AID can used for a variety of purposes in the communication network. In the proposed MAC protocol, AID is adopted for R packets in sounding procedure, because AID can be inserted in PHY layer packet. Therefore, MAC protocol exploits PHY layer packet for R packets to identify the clients according to their AID. Reader might refer to (Gast, 2005) for more details about authentication and association procedure.

As explained, beamformer (AP) first assigns AID to the beamformees (clients), later AP announces a PHY packet (which includes AID's of clients), then clients can figure out by this PHY packet that the packet is for themselves, afterwards, clients feedback CSI to the AP. In our design, we modify the procedure in some context. In our MAC protocol, 'R' packet (packets from UL clients) are crucial because this packet will be used to estimate the channel coefficients between the UL client and AP, as well as, interference channel between UL



(a) Authentication and Association Steps (b) Association Frame Formats

Figure 3.13: Authentication and Association Procedure

clients and DL clients. Hence, it makes more sense in our design for UL clients to employ NDP packets. AP should announce an NDP-A packet (same as in 802.11ac), this will assign unique AIDs to each client, then UL clients should announce 'R' packets (which is similar to NDP packet in 802.11ac). Afterwards, DL clients can estimate the interference channel coefficients by examining 'R' packet, as well as, AP can estimate the channel coefficients between UL and AP by exploiting this 'R' packet. At the end, DL clients should feedback channel coefficients (as in Compressed Beamforming Matrix in 802.11ac) to the AP by using 'C' packet. The whole procedure is shown in Fig. 3.14.

Now, let us look at the frame formats of P, R and C packets. They are the modified version of NDP-A, NDP and Feedback Matrix packet in 802.11ac. More details about 802.11ac packet formats, please refer to (Perahia and Stacey, 2013; Gast, 2013; IEEE, 2013; Liao et al., 2014). P packet contains frame control, duration, receiver address (RA), transmitter address (TA), sounding token, station information (STA info) and frame check sequence (FCS) as shown in 3.15. Here, TA is the address of AP, and RA is the broadcast address. Upon association to the AP, users are assigned an association ID. STA info part contains intended user's association ID (AID). Note that, even though AID is 16 bits, STA info includes 12 LSB of that 16 bits.

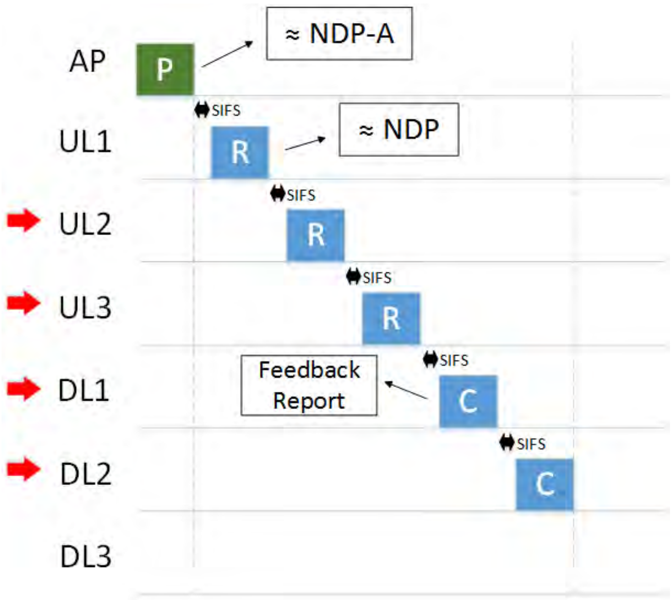


Figure 3.14: P, R and C packets

Using AID is crucial for the intended users, because after P packet is announced, users will act based on this AID in PHY layer packet, ensuing in not needing to use MAC layer packet, thereby reducing overhead significantly. If there are more than one user in the network, each user's STA info is added to the P packet. Note that our algorithm polls clients in a specific order. In this manner, the order of the STA info in P packet also polls clients in that order. Particularly, P packet first contains two-antenna UL users, then three-antenna UL users, thereafter two-antenna DL users, and afterwards three-antenna DL users. This means that users should send their packets in this order. After users receive P packet, they all know the polling order of the users. Hence, each user will listen to the other users, and start deliver their packets when it is their turn to send. First, UL users will reply with R packets in the order of P packet's arrangement. R packet consists of L-STF, L-LTF, L-SIG, VHT-SIG-A, VHT-STF, VHT-LTF and VHT-SIG-B as shown in Fig. 3.15. The signal portions starting with L- imply that this packet is recognizable by the legacy 802.11 users such as 802.11n/a/b/g etc., while VHT- targets at 802.11ac users. In this packet, VHT-SIG-A plays an important role. VHT-SIG-A contains 9 bits partial AID. The partial AID is a signature of the user's presence. Partial AID allows users and AP to immediately identify the user's existence. For example, if an uplink users transmits its R packet, downlink users and AP can recognize the uplink user by looking at its partial ID. In consequence, while AP can exploit that packet for uplink channel estimation, downlink users utilize that packet for interference channel evaluation. Besides, in an R packet there is a VHT-LTF for each spatial stream used in transmission. Depending on the number of transmitted streams, it can contain from 1 to 8 symbols. In our network, since only two-antenna and three-antenna users exist, two-antenna users should include two VHT-LTF, and three-antenna users should use three VHT-LTF symbols. After two-antenna and three antenna UL users send their R packets, now it is time for two-antenna and three-antenna DL users to send their C packet. Remember that each DL user transmits their C packets in the order of P packet. C packet accommodates Compressed Beamforming Report, which is the feedback report and will be used by AP to compute pre-coding vectors for both uplink and downlink communication links. C packet's frame structure is also shown in Fig 3.15. Note that, each packet (P, R and C) has to be sent after SIFS time interval.

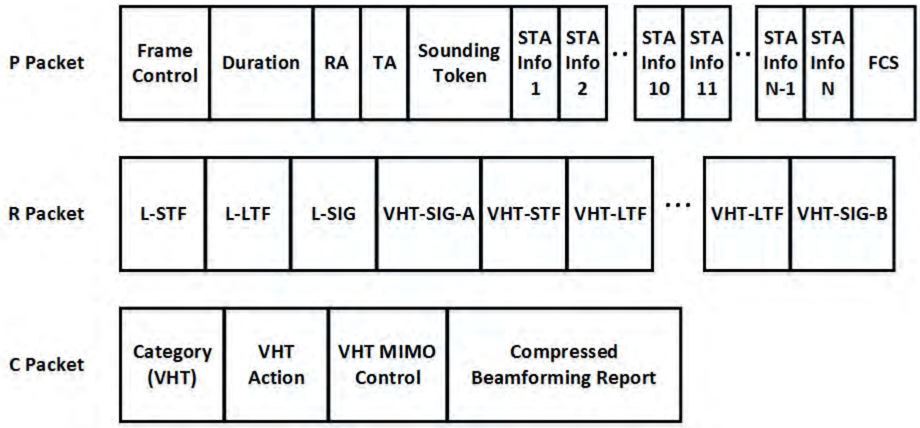


Figure 3.15: Frame Structures

In our MAC protocol, sounding procedure in 802.11ac is benefited. However, our approach uses some modifications of the standard packets in 802.11ac. To recall, P and C packets are MAC layer control packets, and R packets are PHY layer packets. P packet differs from standard NDP-A packet in two senses: 1- P packet contains each user's STA info, 2- P packet polls users in a specific order. To achieve these two purposes, we expand the type of control frames in 802.11ac. The format of the frame control is shown in Fig 3.16. In frame control field, 01 in type field indicates that the frame is a control frame. In subtype segment, different frames (RTS, CTS and ACK etc) are defined with particular values (for example, subtype of RTS frame is 1011). 802.11ac utilizes 0101 subtype value for NDP-A packet, and 0100 for Compressed Beamforming Report. There are still some reserved values (from 000 to 0011) in 802.11ac. We can explore one of these reserved values and define our P packet. Table 3.2 shows the details and the description of the P packet. Note that our algorithm only modifies NDP-A packet into P packet. Standard Beamforming Report packet is still used as C packet. Since P packet is defined for the purpose of sounding mechanism in our protocol, upon announcement of P packet, users are aware of the order of polling, and know to send the required packets (R and C packets).

As mentioned earlier in this chapter, our algorithm reduces overhead with two approaches, namely, limiting number of DL users, and using PHY layer packets (R packets) instead of using MAC layer packets. At this

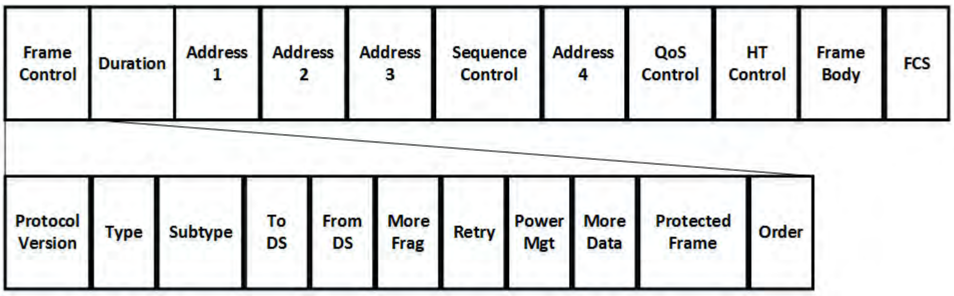


Figure 3.16: Frame Control Field in 802.11

Type	Type Description	Subtype	Subtype Description	Note
01	Control	0011	P Packet	Contains the information of each user Polls users in specific order
01	Control	0100	Beamforming Report	Standard Beamforming Report packet
01	Control	0101	NDP Announcement	Standard NDP-A packet

Table 3.2: Control Frames

moment, let us look at the overhead reduction standpoint. Our MAC protocol first chooses DL clients. In the presence of the both two-antenna and three-antenna clients, our algorithm only chooses maximum 5 downlink clients (2 from two-antenna and 3 from three-antenna clients). What if AP polls all downlink clients in the network? In that case, each downlink client has to send C packet to the AP. C packet contains channel parameters feedback report. It will also be used by AP to estimate downlink channel parameters for AP-DL clients' transmission. For one C packet, network consumes one 16μs SIFS + 40 μs PHY packet + 5 bytes + report size. Our MAC algorithm restricts DL users to maximum 5 clients, thus, the network will consume 5 x (16μs SIFS + 40 μs PHY packet + 5 bytes + report size) = 80 μs + 40 μs PHY packet + 5 bytes + report size (Let's define this value as unit k). This is the maximum overhead caused by DL clients in our design. If AP polls all DL clients in the network, overhead increases linearly as shown in Fig. 3.17. Hence, our algorithm notably reduces the overhead in the network.

3.6 Decreasing Complexity

One of the main insights of our MAC protocol is decreasing the complexity of the system. Remember that our proposed MAC protocol first determines downlink users. This engenders less complexity in the network. To recall, AP selects 4 two-antenna clients (2 UL and 2 DL) and 6 three-antenna clients (3 UL and 3 DL). Consider for a moment that AP polls all downlink and uplink clients. Instead of determining downlink clients first, AP asks all downlink clients to engage in the selection process. In this case, AP needs to make computations for each uplink-downlink group. Let's name this procedure *joint decision*. On the other side, let's define our MAC

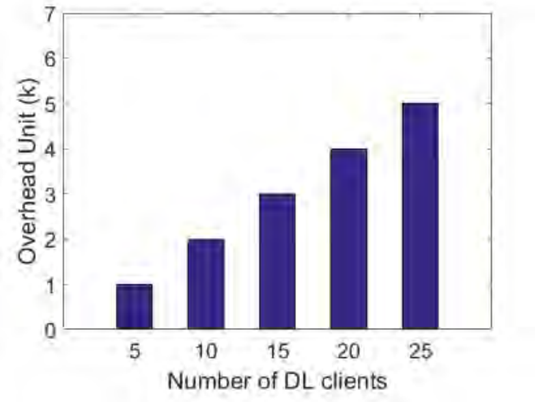


Figure 3.17: Overhead Increment

Protocol's idea of choosing downlink clients first as *first DL decision* for convenience in this chapter. In order for AP to make decision for selection, AP goes over possible uplink downlink groups. For simplicity, let k, l, p and r represent number of two-antenna downlink clients, number of two-antenna uplink clients, number of three-antenna downlink clients and number of three-antenna uplink clients, respectively. Generally speaking, choosing 2 clients among k two-antenna downlink users means:

$$C_k^2 = \frac{k!}{(k-2)!2!} = \frac{k(k-1)}{2} \tag{3.19}$$

Here, C_k^2 simply indicates the possible number of two-antenna downlink groups. For example, if $k = 5$, C_k^2 gives the value of 10, interpreting that 10 possible two-antenna downlink clients can be grouped by AP. For instance, if A, B, C, D and E represent the two-antenna downlink clients, (as known that) 2 of them will be chosen as DL clients. In this sense, there are 10 possible groups (A&B, A&C,A&D, A&E, B&C, B&D, B&E, C&D, C&E and D&E). Similarly, for 5 two-antenna uplink clients, there are $C_l^2 = C_5^2 = 10$ possible groups. Since joint decision resolves downlink and uplink clients together, it yields:

$$C_k^2 \times C_l^2, \{\text{for two-antenna clients}\} \tag{3.20}$$

$$C_p^3 \times C_r^3, \{\text{for three-antenna clients}\}. \tag{3.21}$$

Note that *joint decision* needs to consider all downlink and uplink clients jointly, hence, downlink and uplink computations depend on each other. Thereby, C_k^2 and C_l^2 (as well as C_p^3 and C_r^3) should be multiplied. We hope that it is so far clear to the reader that two-antenna and three-antenna cases are independent from each other, i.e. AP independently selects two-antenna clients and three-antenna clients. Even though communication network consists of two-antenna and three-antenna clients altogether, AP sees both cases as different processes. When AP collects information from all users in the network, it stores those information for two-antenna clients and three-antenna clients separately. Thereupon, selection process for two-antenna and three-antenna cases is conducted individually. Therefore, the number of groups AP needs to look at is the sum of the two cases (two-antenna and three-antenna cases):

$$(C_k^2 \times C_l^2) + (C_p^3 \times C_r^3) \tag{3.22}$$

In*first DL decision*, AP limits the number of DL groups. This will reduce the group number as follows:

$$1 \times C_l^2, \{\text{for two-antenna clients}\} \tag{3.23}$$

$$1 \times C_r^3, \{\text{for three-antenna clients}\}. \tag{3.24}$$

One notice here is that in 3.23 and 3.24, limiting downlink clients means assigning 1 to them. It does not indicate that AP chooses 1 downlink client. It indicates that AP chooses 1 DL client group (for example A&B).

Three of the four selection schemes (MT, RC and PF) have to make a decision for client selection with respect to a metric. Let's characterize this metric as *computation metric (CM).* For example, maximum throughput oriented scheme has to calculate possible data rates (R) to make the decision, so CM here signifies data rate R. Similarly, RC's decision also depends on data rate R, since it selects clients with regard to high data rates. Proportional fairness scheme has to determine priority coefficient (Q), thus its CM is Q. On the other hand, RR does not decide its clients based on any metric. It chooses clients in turn, i.e. clients are selected according to the RR's list one by one. Therefore, RR does not need any metric for client selection. Moreover, MT, PF and RC need different number of operations to calculate CM. As we will show in section 4.1, data rate is estimated by $R = log_2(1 + \|\vec{v}^T H\vec{w}\|^2)$, where H is the channel matrix, $\vec{v}$ and $\vec{w}$ are the corresponding encoding and decoding vectors (Tse and Viswanath, 2005). Remember that priority coefficient Q relies upon two parameters, expressly, data rate (throughput) R and average data rate (throughput) A. As seen, to calculate Q, one should calculate R as well. Within this context, in order to calculate the metric, AP needs addition, multiplication and logarithm operators. Note that subtraction can be regarded as addition, and division can be regarded as multiplication. What is more, different antenna users necessitate different number of operations for the calculation, because the size of pre-coding vector, channel matrix and decoding vector vary depending on the antenna number of the client (i.e. transmission of a two-antenna client requires 2×2 channel matrix, 2×1 pre-coding vector and 2×1 decoding vector, while three-antenna clients need 3×3 channel matrix, 3×1 pre-coding vector and 3×1 decoding vector). Required number of operations is indicated as:

- To calculate R, AP needs:
  - For 2-antenna clients → 7 multiplication + 4 addition + 1 log operations (z_1)

- For 3-antenna clients $\rightarrow$ 13 multiplication + 9 addition + 1 log operations (z_2)

- To calculate $Q = \frac{R}{A}$, AP needs:
  - For 2-antenna clients $\rightarrow$ 12 multiplication + 6 addition + 1 log operations (z_3)
  - For 3-antenna clients $\rightarrow$ 18 multiplication + 11 addition + 1 log operations (z_4),

where z_1, z_2, z_3 and z_4 are assigned to the total number of operations for simplicity. Pay attention that required number of operations is subject to the mathematical base. As an example, for two-antenna clients, MT should calculate $R = log_2(1 + \|\vec{v}^T H \vec{w}\|^2)$, where $\vec{v}$ is 2×1 vector, H is 2×2 matrix and $\vec{w}$ is 2×1 vector. Mark that vectors are actually equivalent to matrices. Hereof, these three matrices behove 6 multiplication and 3 addition. Consequently, matrix multiplication yields a single value. Afterwards, norm turns this value into a positive number (if necessary), then this value is squared (+1 multiplication) and added by one (+1 addition). At the end, AP takes the algorithm of this value, resulting in 7 multiplication + 4 addition + 1 log operations. Similar calculation method is applied to PF scheme. Having said that, a small notice is still worth mentioning for PF scheme. To recall, PF priority coefficient is $Q_i(t) = \frac{R_i(t)}{A_i(t)}$, where average throughput is updated at the end of the time slot by $A_i(t+1) = \frac{(t-1)}{t} A_i(t) + \frac{1}{t} R_i(t) S_i$. While AP makes the decision based on $Q_i(t) = \frac{R_i(t)}{A_i(t)}$ for the intended time slot t, it needs same number of operation as shown above (z_1 for two-antenna clients and z_2 for three-antenna clients. However, AP should also calculate $A_i(t)$ to determine $Q_i(t)$. $A_i(t)$ represents the average throughput until time slot t, which is updated at the end of time slot $t-1$. For example, at the end of 5^{th} time slot, the average throughput for 6^{th} time slot is calculated by $A_i(6) = \frac{4}{5} A_i(5) + \frac{1}{5} R_i(5)$ (assuming user i transmitted in 5^{th} time slot, hence $S_i = 1$). On the calculation process of $A_i(6)$, AP already knows the values $A_i(5)$ and $R_i(5)$, because these values are obtained after the transmission (at the end of 5^{th} time slot). Therefore, in the beginning of 6^{th} time slot when PF scheduler at the AP calculates $Q_i(6) = \frac{R_i(6)}{A_i(6)}$ for decision making, the values of $A_i(5)$ and $R_i(5)$ of the denominator of $A_i(6) = \frac{4}{5} A_i(5) + \frac{1}{5} R_i(5)$ are just values at this moment. Thus, in general, denominator $A_i(t+1) = \frac{(t-1)}{t} A_i(t) + \frac{1}{t} R_i(t) S_i$ requires 4 multiplication + 2 addition, while numerator $R_i(t)$ needs 7 multiplication + 4 addition + 1 log (for 2-antenna clients) or 13 multiplication + 9 addition + 1 log. Additionally, to get $Q_i(t)$, another division (multiplication) of $R_i(t)$ and $A_i(t)$ is mandatory, resulting in 12 multiplication + 6 addition + 1 log operations (for two-antenna clients) and 18 multiplication + 11 addition + 1 log operations (for two-antenna clients).

In subject to complexity, this work defines complexity based on the required number of operations. On that note, more operation implies more complexity. As emphasized already, *first DL decision* claims to decrease complexity. We now provide the idea behind it. Before that, in the beginning of this section, it has been already shown the possible group combination of the clients. For two-antenna clients, a group of 4 users will be selected, while a group of 6 users gets chosen among three-antenna clients. Considering the fact that 4-users group contains 4 streams, for a two-antenna group, 4 computation metric (CM) must be calculated. Similarly, for a three-antenna group, 6 metrics should be calculated. We define a mathematical formula for the total required number of operations as follows:

- In MT scheme:
  - Joint Decision
    - $[(C_k^2 \times C_l^2) \times 4 \times z_1] + [(C_p^3 \times C_r^3) \times 6 \times z_2] \leftrightarrow O(k^2l^2 + p^3r^3)$
  - First DL selection
    - $[(1 \times C_l^2) \times 4 \times z_1] + [(1 \times C_r^3) \times 6 \times z_2] \leftrightarrow O(l^2 + r^3)$
- In PF scheme:
  - Joint Decision
    - $[(C_k^2 \times C_l^2) \times 4 \times z_3] + [(C_p^3 \times C_r^3) \times 6 \times z_4] \leftrightarrow O(k^2l^2 + p^3r^3)$
  - First DL selection
    - $[(1 \times C_l^2) \times 4 \times z_3] + [(1 \times C_r^3) \times 6 \times z_4] \leftrightarrow O(l^2 + r^3)$

As seen in the formulas, Joint Decision suggests less number of operations than First DL decision, consequently, reduces complexity. In big O notation, this corresponds to the reduction from $O(k^2l^2 + p^3r^3)$ to $O(l^2 + r^3)$. Note that each selection scheme has different complexity. Even though they are different, we compare them according to the number of operations (how many additions, multiplications and logarithms are required for a specific scheme). For instance, it is obvious that PF ends up with higher complexity than MT scheme since PF requires more operations to calculate its metric. However, the main idea on decreasing complexity arises from First DL selection. All in all, choosing DL clients first will decrease the number of operations greatly.

# of 2-ant. CLTs		# of 3-ant CLT.		Total # of	Joint Decision			First DL Decision		
DL	UL	DL	UL	Clients	+	*	*Log*	+	*	*Log*
5	5	5	5	20	7 000	10 600	1 000	700	1 060	100
7	7	7	7	28	73 206	107 898	9 114	2 226	3 318	294
10	10	10	10	40	810 000	1 179 900	94 500	7 200	10 620	900

Table 3.3: Required number of operations for MT

# of 2-ant. CLTs		# of 3-ant CLT.		Total # of	Joint Decision			First DL Decision		
DL	UL	DL	UL	Clients	+	$*$	*Log*	+	$*$	*Log*
5	5	5	5	20	9 000	15 600	1 000	900	1 560	100
7	7	7	7	28	91 434	153 468	9 114	2 814	4 786	294
10	10	10	10	40	999 000	1 652 400	94 500	9 000	15 120	900

Table 3.4: Required number of operations for PF

One last point to mention is that we evaluate complexity analysis for MT and PF scheme. While RR does not calculate any metric to make a decision, RC needs to calculate its metrics based on the data rate R. As a matter of fact, defining the mathematical formula for the Rc is scheme is a bit harder than MT and PF, since RC scheme ignores some users after each round of transmission. In this sense, RC's mathematical formula should add a variable depending on the number of user. However, First DL selection also reduces the complexity of RC scheme as well. Finally, Table 3.3 and Table 3.4 shows required number of operations for MT and PF depending on the number of users and antenna in the network, where $+$, $*$ and *Log* represent the operators addition, multiplication and logarithm, respectively.

Chapter 4

Performance Evaluation

4.1 Implementation Setup

Simulation is conducted in Matlab to evaluate the performance of our MAC protocol. In our topology, one AP supports full-duplex function, and various clients with different number of antennas operate on half duplex mode. Since AP runs full-duplex, we assume perfect self-interference cancellation in the device, and radio is able to communicate in a full duplex manner without any PHY constraints. Through this work, we only focus on two- and three-antenna users, consequently, in our network, communication only takes places for those users. To see the effects of our MAC protocol, we evaluate two sets of experiments: 1- Fixed number of users, where certain number of users (with antennas) exist in each round 2- Random number of users, where random number of users (with random number of antennas) appear in each round. As mentioned, AP acts as a controller, so none of the users try to perform random access in the network. Besides, there are no hidden nodes in the network, i.e. all devices as well as AP can hear each other's transmission. Our topology requires each user to be active for transmission at all times, to put in different way, uplink users always have some packets to send to AP, and downlink users are constantly ready to receive packets from AP. Data transmission begins after P, R and C packets' transmission, and it depends on the information examined from those packets. In this context, we build our channel with large coherence time, that is, channel parameters do not vary rapidly. Our channel parameters are composed of complex Gaussian random variable, i.e. $h = a + bj$, where a and b are normal (Gaussian) distributed random variables with zero mean and 1/2 variance $\sim N(0, 1/2)$. The reason behind choosing these values in channel model is that these values represent Rayleigh fading that is often experienced in an indoor environment where there is a large number of reflections (multipath) present. Choosing 1/2 suggests that the channel behaves constructively with the probability of 63%, and destructively with the probability of 63% (see Appendix A for more in-depth discussion for $N(0, 1/2)$). This is our simulation setup for a typical indoor environment. In our topology, only two-antenna and three-antenna users exist, hence channel matrix H can be either 2x2 or 3x3, depending on the user's number of antennas. Note that, for a 2x2 MIMO channel, H consists of 4 elements, each element is represented by $h = a + bj$ as shown above. For 3x3 MIMO channel, channel matrix H has 6 elements that have the form of $h = a + bj$. In our network, three channels play a crucial role for our MAC algorithm, namely, uplink channel (from uplink users to AP), downlink channel (from AP to downlink users) and interference channel (from uplink users to downlink users). All these three channels are independent from each other, thereby, in our simulation, we randomly create these three channel matrices. Furthermore, each selection scheme of our MAC protocol requires to calculate the possible metric. For example, MT scheme makes the decision based on estimated throughput, hence MT scheme only focuses on throughput (so, MT's metric is throughput). Similarly, RC scheme also selects clients based on user's estimated throughput, thus it's metric is calculated from throughput as well. PF utilizes priority coefficient, and this coefficient depends on two values; estimated throughput and average throughput. As seen in the priority coefficient numerator, this metric needs estimated throughput for calculation. In this sense, AP should calculate it's metric (which is priority coefficient for PF) as well. Note that average throughput actually depends on the previous throughput values. This previous throughput values should be stored at the AP's memory, so they can be used to calculate the priority coefficient. To calculate the (estimated) throughput, we use rate metric as follows:

$$R = log_2(1 + \|\vec{v}^T H \vec{w}\|^2), \tag{4.1}$$

where H is the channel matrix, $\vec{v}$ and $\vec{w}$ are the corresponding encoding and decoding vectors (Tse and Viswanath, 2005). Encoding and decoding vectors can be calculated as discussed in section 3.1. Both uplink (UL-AP) and downlink (AP-DL) transmission data rates are estimated by using this rate metric. Note that before uplink and downlink data transmissions start, all users and AP know the selected clients, correspondingly, encoding and decoding vectors. Thus, selected uplink clients and AP will compute their encoding vectors, while downlink users and AP determine their decoding vectors before the actual data transmission starts.

We evaluate our MAC design for different selection schemes, namely, maximum-throughput oriented scheme (MT), round robin scheme (RR), remainder choice scheme (RC) and proportional fairness scheme (PF). Scheduler at AP should be able to perform all client selections schemes. In fact, AP does not have to advertise which scheme it is performing, because our protocol steps do not depend on the selection scheme. No matter which selection scheme is adopted, our MAC protocol utilizes same steps. Thereby, selection schemes only affect the results, not the procedure. In other words, that timeline of the MAC protocol does not change with the selection scheme. For each selection scheme, timeline goes as usual: AP announcing DL clients first, then UL clients sending R packets followed by DL client's feedback to AP (C packet), afterwards AP makes the decision for

UL clients, and actual data transmission starts. At the end, acknowledgement packets are shared to terminate the transmission. Besides, we also evaluate our selection schemes based on different number of antennas. Considering the fact that a network might embrace different number of clients, the impact of number of clients is also considered in our simulation. Different number of clients actually change the performance of each selection scheme as shown in section 4.2. As stated earlier, our MAC protocol aims at the trade-off between fairness and throughput. In this context, we examine fairness index for each selection schemes for the trade-off. Simulation results also show that each selection scheme as well as number of users have different impact on throughput and fairness in the network. Next section provides more details about throughput and fairness.

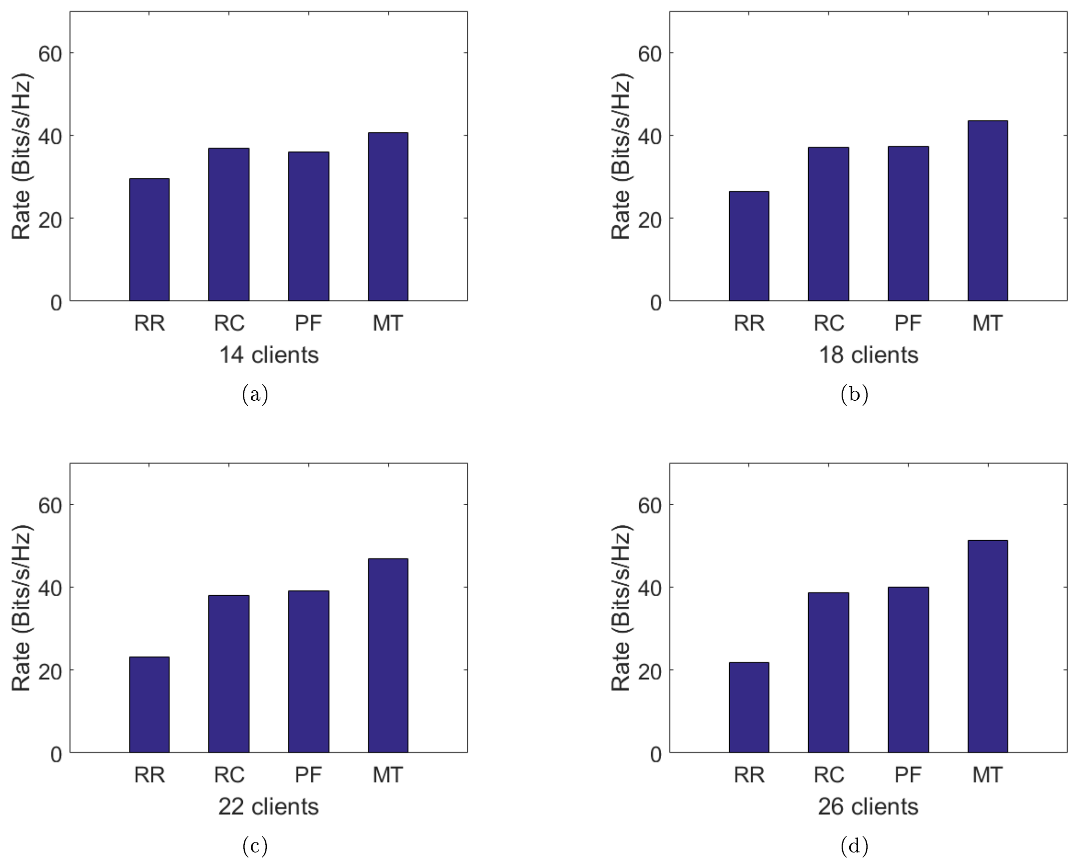
4.2 Simulation Results

This section presents proposed MAC protocol's simulation results on a deeper focus. Each scenario is explained in details and compared with other schemes to show the impact on the performance.

4.2.1 Throughput Analysis

Fixed Number Of Users

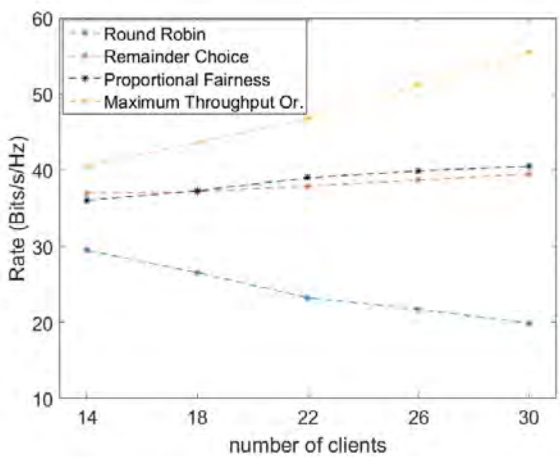
It is usual in the networking community to compare the throughput of different designs. For that matter, in this section we investigate the date rate of our four selection schemes under different number of clients existing in the network. In this set of experiment, fixed number of users exist in each round, i.e. in each round, networks consists of fixed number of users. Fig. 4.1 shows data rate results for different selection schemes. In the first experiment (Fig. 4.1a), we simulate our MAC design for 14 clients. Among these 14 clients, there are 6 two-antenna clients (3 for uplink and 3 for downlink), and 8 three-antenna clients (4 for uplink and 4 for downlink). Remember that our MAC protocol selects 4 two-antenna clients and 6 three-antenna clients, thus, rate values in the figure cover 4 streams for two-antenna clients and 6 streams for three-antenna clients, indicating 10 streams in total. As expected, MT achieves the highest rate, because MT only focuses on maximizing the total rate. MT only targets at clients whose channel is in best state. Clients with best channel conditions will result in gaining the maximum data rate in the network. RC also tries to maximize the rate while providing equal channel access to the users. In the beginning of RC, AP has a freedom to choose any clients in the network, therefore, it chooses clients that attain maximum rate. It also means that RC acts similar to MT at the beginning. However, in the next rounds, RC scheduler is forced to choose clients that did not participate in the previous round(s) (for equal channel access), as a result, RC scheduler ought to choose clients who will drop total rate of the network. After RC serves to all users in the network, it will again have a freedom to choose clients for maximum rate. To put differently, operation principle of RC is similar to MT in the start, afterwards, by reason of providing equal channel access to the users, RC cannot continue acting like MT, hence it does not reach as high rate as MT does. On the other hand, RR only concerns about giving equal transmission opportunity to users. It picks clients one by one, so, RR does not take client's channel condition into account. Although RR might choose clients with good channel conditions (by chance), it mostly would end up with selecting clients that do not have good channel states. In fact, RR results in poorest performance among four selection schemes in terms throughput and fairness. The reason of using RR is to see the effect of a simple selection scheme. It is uncomplicated and basic selection scheme that only allocates equal time share to the users. For a reliable implementation, it is not highly recommended to use RR due to it's relatively limited performance. Conversely, PF achieves a favourable data rate. To remind, PF determines clients with respect to the priority coefficient given in eq. 3.15. PF can achieve as high data rate as RC which is a throughput oriented scheme and focuses on large-scale number of data rate. In fact, in the next section, we will see that advantage of PF arises from fairness point of view.



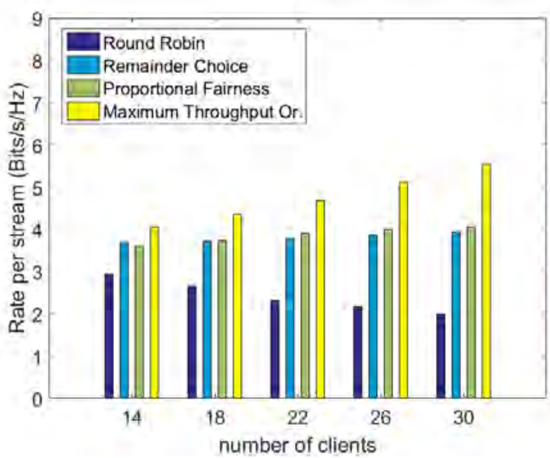
(a) (b) (c) (d)

Figure 4.1: Data rate for different scenarios (fixed number of users)

In the second experiment (Fig. 4.1b), we increase the number of clients in the network. We add 2 two-antenna clients (1 for uplink and 1 for downlink) and 2 three-antenna clients (1 for uplink and 1 for downlink). In this scenario, AP deals with more clients, thus, more computation is needed. We conclude that data rate of MT increases dramatically. The reason behind it is that there are more clients in the network, and MT is more flexible to select clients. If there exist more clients in the network, there will most probably be clients whose channel is in better condition. As a result of MT's maximum throughput selection algorithm, clients with better channel state will be chosen to achieve higher data rate. On the contrary, RR's data rate drops strikingly. Due to larger number of clients in the network, RR mostly picks clients with bad channel conditions. However, simulation results show that larger number of users does not affect RC's and PF's data rate significantly. PF balances the data rate with its priority coefficient. RC benefits large number of users similar as MT. As stated earlier, RC acts like MT in the beginning of early rounds of transmission, allowing it to choose clients more freely. As number of rounds take place in communication, RC is obliged to select among remaining users. This will surely reduce the data rate, nevertheless, overall performance is not influenced greatly in general. Third and fourth experiments' results in Fig. 4.1c and Fig. 4.1d are conducted under larger number of users in the system (10 two antenna clients -5 uplink and 5 downlink clients- in Fig. 4.1c, and 10 two antenna clients -5 uplink and 5 downlink clients- in fig. 4.1d). In summary, simulation results demonstrate that with large number of users, MT achieves higher data rate as a result of more freedom to choose, RR's data rate decreases greatly due to its ignorance of channel conditions, RC's rate performance stays stable by compensating itself after each round, PF obtains consistent data rate by applying priority coefficient. Fig. 4.2a displays overall look on different scenarios in our simulation. Besides, Fig. 4.2b illustrates data rate per user under various cases.



(a) Overall look on data rate

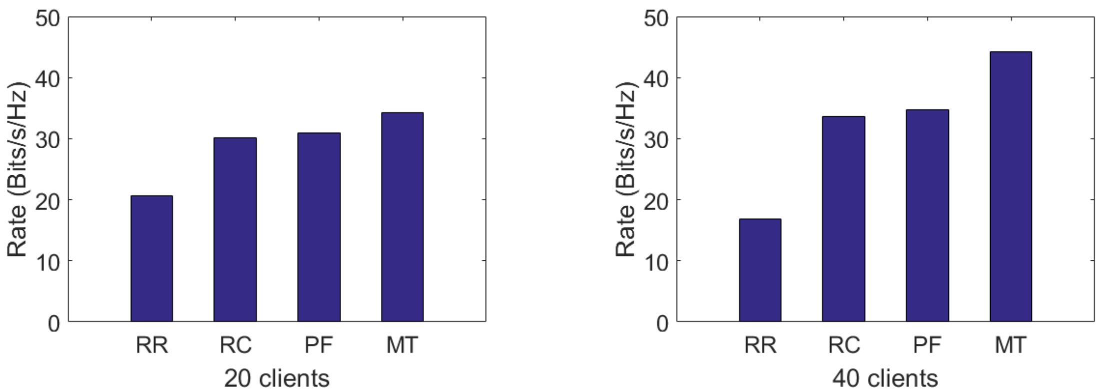


(b) Rate per stream

Figure 4.2

Random Number Of Users

In this set of experiments, we investigate random number of users (with random number of antennas) existing in the network. In each round, users appear in a random manner. For example, 3 two-antenna and 4 three-antenna clients might appear in the first round, while there might be 1 two-antenna and 10 three-antenna users in the second round. In the first simulation, we limit the maximum number of users to 20 clients. Second simulation is conducted under the network where maximum 40 clients can show up. In comparison with fixed number of users case, random number of users case achieves lower throughput. Due to the randomness of the users, in some rounds, small number of users participate in the transmission, causing degradation in throughput. When there are small number of users in the network, the effect of the selection schemes is underachieving. For example, when there is only one user in the network, AP should choose that user under any circumstances. No matter which scheme AP is running, that client has to be selected even his channel condition is really poor. However, when there is large amount of users, network is affected less by the randomness of the users. Simulation results for 20 and 40 random users with random number of antennas are shown in Fig. 4.3.

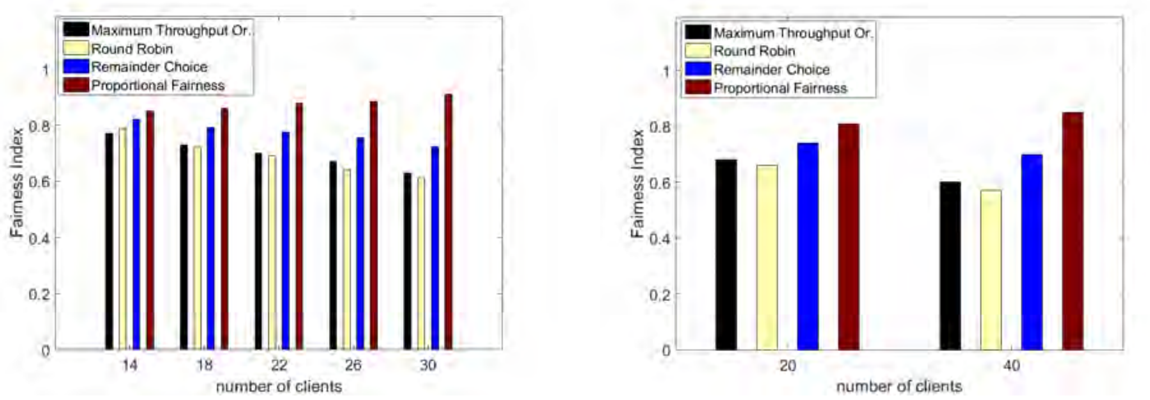


(a) (b)

Figure 4.3: Data rate for random number of users

4.2.2 Fairness Analysis

In this section, we carry out fairness analysis for our MAC protocol. To recall, fairness means fair sharing of the resources. As described in section 3.3, our network defines resource as bandwidth. In this context, after running our algorithm repeatedly, each user will obtain an average data rate through all the rounds that they participated in. For example, at the end of the day, user A acquires 3 Mbps, while user B attains 4 Mbps average data rate. No matter how many times they accessed the channel or which algorithm is applied to them, they will have gotten 3 and 4 Mbps average date rate. In this manner, we can conclude that user B consumes more bandwidth than user A in the network. In simple words, network resource is not shared equally between user A and user B. The fairest scenario would be if both users achieve same average data rate. To measure fairness in our network, we use Jain's index as explained in section 3.3. This index highlights how fair is the selection scheme. Remember that Jain's fairness index always lies between 0 and 1. A fairness index close to 1 indicates that the network is fairer, while small index implies that resource of the network is not allocated fairly.



(a) Fixed number of users

(b) Random number of users

Figure 4.4: Fairness index for different schemes

We examine fairness in our four selection schemes (MT, RR, RC, PF). In the first experiment shown in Fig. 4.1a, where there are 14 (fixed number of) users in the network, PF achieves the best fairness index among four selection schemes. As a matter of fact, when number of clients is not large (as in scenario with 14 clients), fairness index of each selection scheme is analogous to each other. The reason is that the small number of users acts similarly. For instance, even in RR scheme, because the number of users is small, they will access the channel very frequently. Users sometimes experience bad channel conditions, but once in a while, they join the transmission with good channel conditions. On account of the small number of users, one user is often able to access channel as if it constantly transmits its data. Conversely, this will not be the case when network contains larger number of users. If we look at large number of users, we see the difference of fairness index between different selection schemes. In random number of users case fairness index obtains similar performance as well. Simulation results show that only PF maintains steady fairness index. Other selection schemes such as MT, RR and RC eventuate in smaller fairness index. We can conclude that only PF achieves desirable fairness index, while other selection schemes focus either on maximizing data rate or allocating equal access time, resulting in worse fairness index. Fairness index for different selection schemes with fixed number of users and random number of users is shown in Fig. 4.4a and Fig 4.4b.

4.2.3 Fairness - Throughput Trade-off

This work claims to define a trade-off between throughput and fairness. As simulation results show, MT scheme acquires highest throughput, and PF achieves highest fairness. However, fairness in MT scheme is comparatively lower than that of PF and RC's. On the other hand, the performance of RR is not comparable with other selection schemes from the point of fairness and throughput. In order to see the difference between each selection scheme's fairness-throughput relationship, we created a figure shown in Fig. 4.5. According to the simulation results, each case is represented with a value in the figure. Note that throughput and fairness is normalized to see the real difference between each selection schemes under different number of users (fixed or random). Point (1,1) is the optimum point, i.e. it shows the best case in terms of throughput and fairness. The closer a scheme to the point (1,1), the better result it achieves with regard to fairness and throughput. If a scheme was on point (1,1), it means that the proposed scheme obtains highest throughput and highest fairness at the same time. However, it is not possible for a selection scheme to reach that point since each selections

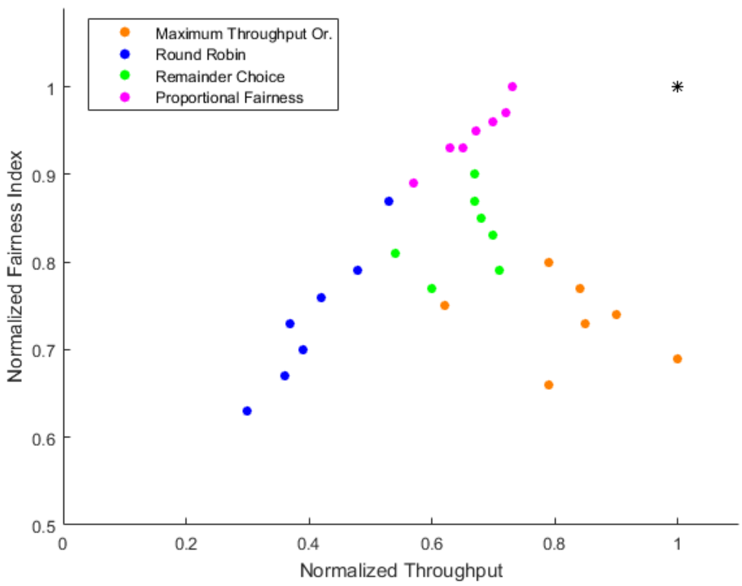


Figure 4.5: Fairness - Throughput Trade-off

scheme proposes different algorithm to attain a throughput and fairness value. As seen in the figure, MT values are relatively closer to higher throughput region, while PF values are comparatively lie on the fairer region. RR scheme reveals lower throughput and fairness, and RC scheme is situated in between. To dig deeper, we define an average distance of each selection scheme to the optimum point (1,1). Average distance is calculated based on the distance of each scheme's performance to the point (1,1). The shorter the average distance is, the closer is the scheme to the optimum point. PF scheme achieves the shortest average distance to the optimum point by 0.28. MT, RC and RR obtains 0.38, 0.43 and 0.64 respectively. Ultimately, PF is gains the shortest distance to the optimum point, meaning that it suggests the highest throughput-fairness trade-off. Nevertheless, other selection schemes also can be used in the network for the specific purposes.

Chapter 5

Conclusion

5.1 Contributions

This work introduces a MAC protocol for a full duplex network, where AP supports full duplex function, and clients are legacy half duplex users with multiple antennas. In this scenario, challenging problem of inter-node interference is addressed and overcome by interference alignment and MIMO technology. We built our MAC protocol to achieve: 1) Finding the best trade-off between throughput and fairness, 2) Reducing the overhead and decreasing the complexity. We designed and examined four different client selection schemes to determine the trade-off for each scheme, and simulation results show that proportional fairness selection scheme becomes the most favourable candidate among these four selection schemes in terms of high data rate achievement and fairness assurance. For the purpose of a reliable and practical implementation, proposed MAC protocol also targets at overhead and complexity issue. In the MAC protocol, the idea of determining downlink clients before uplink clients leads not only to reduced overhead but also to low complexity. While complexity analysis exhibits great improvement on the required number of operations at the AP, this study also demonstrates that our work restricts the linear overhead increase. We further designed PHY layer packets to keep the overhead at minimum level.

5.2 Future Work

The focus on this work was designing a MAC protocol for the scenario of an FD AP with HD clients. The perspective on the scenario given in this dissertation establishes the groundwork for many interesting extensions, some of which we review here.

Multi-cell networks: To adopt our design widely, it would be interesting to consider our scenario in multi-cell networks. In such research, INI has to be addressed more carefully, and new form of MAC protocol should be considered for both in-cell and inter-cell communication.

Heterogeneity: As PHY layer technologies arises rapidly, it is soon expected to see devices on the shelf that support full duplex capabilities. In this manner, a communication network consisting of both half duplex and full duplex users would be an appealing topic for future research.

Acknowledgments

First and foremost I would like to thank my advisor, Prof. Xudong Wang, for his support and guidance throughout my graduate studies. It has been a pleasure to work with him for almost three years. Under his supervision, I gained valuable knowledge in both academia and industry.

I also would like to thank my defense committee members, Assistant Prof. Weikang Qian and Assistant Prof. Chong Han for taking the time to serve on my committee. They helped me with their precious feedback and comments on my thesis study.

A special thanks also goes to WANG Lab members who I have enjoyed working and developing friendships with during my master studies in University of Michigan-Shanghai Jiao Tong University Joint Institute. It was a pleasure to get to know each of you.

Finally, I must express my very profound gratitude to my parents and to the people around me for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis. This accomplishment would not have been possible without any of these names mentioned above.

Appendix A

Channel Model with N(0,1/2)

In a wireless communication channel, the transmitted signal propagates from transmitter to receiver over multiple reflective paths. This leads to multipath fading which causes fluctuations in amplitude and phase. Communication channel can be modelled as a complex number. As α_i, τ_i and L represent attenuations, delays and the number of multipath channels, respectively, channel parameter is expressed as:

$$h = \sum_{i=0}^{L-1} \alpha_i e^{-j2\pi f_c \tau_i} = x + yj = ae^{j\theta}, \tag{A.1}$$

where

$$x = \sum_{i=0}^{L-1} \alpha_i cos(2\pi f_c \tau_i) \tag{A.2}$$

$$y = -\sum_{i=0}^{L-1} \alpha_i sin(2\pi f_c \tau_i). \tag{A.3}$$

Since α_i and τ_i are random in nature,x and y are the sum of a large number of components. By central limit theorem, x and y can be assumed to be Gaussian distributed random variables: $x \sim N(0, \sigma^2)$ and $y \sim N(0, \sigma^2)$. Probability density function of x and y are:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-x^2}{2\sigma^2}} \tag{A.4}$$

$$f_Y(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-y^2}{2\sigma^2}}. \tag{A.5}$$

Assuming that x and y are independent random variables, joint distribution of x and y is given as the product of the individual distributions:

$$f_{X,Y}(x,y) = f_X(x)f_Y(y) \tag{A.6}$$
$$= \frac{1}{\sqrt{2\pi\sigma^2}}e^{\frac{-x^2}{2\sigma^2}}\frac{1}{\sqrt{2\pi\sigma^2}}e^{\frac{-y^2}{2\sigma^2}} \tag{A.7}$$
$$= \frac{1}{2\pi\sigma^2}e^{\frac{-(x^2+y^2)}{2\sigma^2}}. \tag{A.8}$$

Equation in (A.6) shows the distribution of the fading channel coefficient in terms of the real and imaginary parts. This is one way to characterize the distribution of the fading channel coefficient. However, fading channel coefficient can also be characterized with regard to magnitude and phase. As seen in equation (A.1), magnitude and phase can be written as $a^2 = x^2 + y^2$ and $\theta = tan^{-1}\frac{y}{x}$, equivalently, $x = acos(\theta)$ and $y = asin(\theta)$. Given the joint distribution of x and y, joint distribution of a and θ can be derived as a product of joint distribution of x and y and the determinant of the Jacobian XY:

$$f_{A,\Theta}(a,\theta) = f_{X,Y}(x,y)\left|J_{XY}\right| \tag{A.9}$$
$$= \frac{1}{2\pi\sigma^2}e^{\frac{-(x^2+y^2)}{2\sigma^2}}a \tag{A.10}$$
$$= \frac{a}{2\pi\sigma^2}e^{\frac{-a^2}{2\sigma^2}}, \tag{A.11}$$

where determinant of $J_{XY} = a$. At this point, individual distributions of a and θ can also be derived from the joint distribution as follows:

$$f_A(a) = \int_{-\pi}^{\pi} f_{A,\Theta}(a,\theta)d\theta \tag{A.12}$$
$$= \frac{a}{\sigma^2}e^{\frac{-a^2}{2\sigma^2}} \tag{A.13}$$
$$f_\Theta(\theta) = \int_{0}^{\infty} f_{A,\Theta}(a,\theta)da \tag{A.14}$$
$$= \frac{1}{2\pi}. \tag{A.15}$$

(A.13) and (A.15) indicate that the amplitude of the channel is rayleigh distributed, and the phase of the channel is uniformly distributed.

Multipath signals can be constructive and destructive on the received signal. By using (A.13), the effect of the multipath signals can be observed. Taking $x \sim N(0, 1/2)$ and $y \sim N(0, 1/2)$ results in:

$$f_A(a) = 2ae^{-a^2}. \tag{A.16}$$

In order to find the probability of the channel that has destructive effect, integral of the probability density function (in A.13) should be taken. Integral bounds must be from 0 to 1, since the amplitude of the channel a attenuates the transmitted signal. Similarly, for constructive channel, a needs to be greater than 1, which corresponds to the integral bounds with 1 to ∞.

$$P(0 < a < 1) = \int_0^1 2ae^{-a^2} da \tag{A.17}$$

$$\approx 0.63, \tag{A.18}$$

and,

$$P(1 < a < \infty) = \int_1^\infty 2ae^{-a^2} da \tag{A.19}$$

$$\approx 0.37. \tag{A.20}$$

(A.18) and (A.20) show that the channel behaves constructively with the probability of 37%, and destructively with the probability of 63%. For larger σ^2 values, the probability of a constructive channel becomes higher. For example, when $\sigma^2 = 1$, the probability of a constructive channel increases to 61% (while destructive probability reduces to 39%). For $\sigma^2 = 2$, constructive and destructive probabilities yield 88% and 12%, respectively. Besides, since $h = x + yj$, choosing $x \sim N(0, 1/2)$ and $y \sim N(0, 1/2)$ suggests that h is a complex random variable with zero mean and unit variance of power. Since h has real and imaginary parts that are independent from each other, then it is logical to consider that half of the power is in the real part and half of the power is in the imaginary part.

Bibliography

Adib, F., S. Kumar, O. Aryan, S. Gollakota, and D. Katabi (2013). Interference alignment by motion. *Mobicom*, 279–290.

Ali, M. A., S. A. Motahari, and A. K. Khandani (2008). Communication over mimo x channels, interference alignment, decomposition and performance analysis. *IEEE Transactions on Information Theory 54*(8), 3457–3470.

Aryafar, E., A. Khojastepour, K. Sundaresan, S. Rangarajan, and M. Chiang (2012). Midu: Enabling mimo full duplex. *ACM Mobicom*.

Ayach, O. E., S. W. Peters, and R. W. Heath (2013). The practical challenges of interference alignment. *IEEE Wireless Communications Magazine 20*(1), 35–42.

B-Porat, U., A. Bremler-Barr, H. Levy, and B. Plattner (2011). On the vulnerability of the proportional fairness scheduler to retransmission attacks. *INFOCOM, 2011 Proceedings IEEE*.

Bai, J. and A. Sabharwal (2012). Distributed full-duplex via wireless side channels: Bounds and protocols. *Arxiv, abs/1212.5300*.

Bhadaria, D., E. McMilin, and S. Katti (2013). Full-duplex radios. *ACM Sigcomm*.

Cadambe, V. and S. A. Jafar (2008). Interference alignment and degrees of freedom of the k user interference channel. *IEE Transactions on Information Theory 14*(10), 3425–3441.

Changho, S. and D. Tse (2008). Interference alignment for cellular networks. *Allerton Conference*, 1037–1044.

Choi, J., M. Jain, K. Sirinavasan, P. Levis, and S.Katti (2010). Achieving single channel, full duplex wireless communication. *ACM Mobicom*.

Demers, D., S. Keshay, and S. Shenker (1989). Analysis and simulation of a fair queueing algorithm. *ACM SIGCOMM Computer Communication Review 19*, 1–12.

Duarte, M., C. Dick, and A. Sabharwal (2012). Experiment-driven characterization of full duplex wireless systems. *IEEE Transaction on Wireless Communications*.

Dunn, B. P. (2010, December). Overhead in communication systems as the cost of contraints. Master's thesis, University of Notre Dame.

Everett, E., M. Duarte, C. Dick, and A. Sabharwal (2011). Empowering full-duplex wireless communication by exploiting directional diversity. *Asilomar Conference*.

Gast, M. (2005). *802.11 Wireless Networks: The Definitive Guide* (2 ed.). USA: O'Reilly Media.

Gast, M. (2013). *802.11ac: A Survival Guide* (1 ed.). USA: O'Reilly Media.

Goldsmith, A. (2005). *Wireless Communications* (1 ed.). New York, USA: Cambridge University Press.

Gollakota, K., S. D. Perli, and D. Katabi (2009). Interference alignment and cancellation. *ACM Sigcomm*.

Greenberg, A. G. and N. Madras (1992). How fair is fair queuing. *Journal of the ACM (JACM) 39*, 568–598.

Greenberg, A. G. and N. Madras (2006). Weighted round-robin cell multiplexing in a general-purpose ATM switch chip. *IEEE Journal on Selected Areas in Communications 9*, 1265–1279.

Halperin, D., W. Hu, A. Sheth, and D. Wetherall (2010). 802.11 with multiple antennas for dummies. *ACM SIGCOMM Computer Communication Review 40*, 19–25.

Hong, S. S., J. Mehlman, and S. Katti (2013). Picasso: flexible rf and spectrum slicing. *ACM Sigcomm*.

IEEE (2013). IEEE standard for information technology- telecommunications and information exchange between systemslocal and metropolitan area networks– specific requirements–part 11: Wireless lan medium access control (mac) and physical layer (phy) specifications–amendment 4: Enhancements for very high throughput for operation in bands below 6 ghz. *IEEE 802.11ac*, 1–425.

Jafar, S. and S. Shamai (2008). Degrees of freedom of mimo x channel. *IEEE Transactions on Information Theory 54*, 151–170.

Jain, M., T. M. K. abd D. Bharadia, S. Seth, K. Srinavasan, P. Levis, S. Katti, and P. Sinha (2011). Practical, real-time, full duplex wireless. *ACM Mobicom*.

Jain, R. K., D. M. W. Chui, and W. R. Hawe (1984). Quantitative measure of fairness and discrimination for resource allocation in shared systems. *ACM Transaction on computer systems*.

Kelly, F. P., A. K. Maulloo, and D. K. H. Tan (1998). Rate control for communication networks: Shadow prices, proportional fairness and stability. *The Journal of the Operational Research Society 49*, 237–252.

Kim, J. Y., O. Mashayekhi, H. Qu, M. Kazandjieva, and P. Levis (2013). Janus: A novel mac protocol for full duplex radio. *Stanford University*.

Kumar, S., D. Cifuentes, S. Gollakota, and D. Katabi (2013). Bringing cross-layer mimo to today's wireless LANs. *ACM Sigcomm*, 387–398.

Liao, R., B. Bellalta, T. M. Cao, J. Barcelo, and M. Oliver (2014). Uni-MUMAC: a unified down/up-link MU-MIMO MAC protocol for IEEE 802.11ac WLANs. *Wireless Networks (WIREL NETW), springer*.

Lin, K., Y. J. Chuang, and D. Katabi (2012). A light-weight wireless handshake. *ACM SIGCOMM Computer Communication Review 42*, 28–34.

Lin, K. C., K. Gollakota, and D. Katabi (2011). Random access heterogeneous mimo networks. *ACM Sigcomm*.

Miao, G., J. Zander, K. W. Sung, and S. B. Slimane (2016). *Fundamentals of Mobile Data Networks* (1 ed.). Sweden: Cambridge University Press.

Nagle, J. (2003). On packet switches with infinite storage. *IEEE Transactions on Communications 35*, 435 – 438.

Perahia, E. and R. Stacey (2013). *Next Generation Wireless LANs: 802.11n and 802.11ac* (2 ed.). USA: Cambridge University Press.

Radunovic, B., D. Gunawardena, P. Key, A. Proutiere, N. Singh, V. Balan, and G. Dejean (2010). Rethinking indoor wireless mesh design: Low power low frequency, full duplex. *Wireless Mesh Networks*.

Sahai, A., G. Patel, and A. Sabharwal (2011). Pushing the limits of full-duplex: Design and real-time implementation. *arXiv:1107.0607*.

Shen, W. L., Y. C. Tung, K. C. Lee, K. C. Lin, , S. Gollakota, D. Katabi, and M. S. Chen (2012). Rate adaptation for 802.11 multiuser MIMO networks. *Mobicom*, 29–40.

Shi, H., R. V. Prasad, E. Onur, and I. G. M. M. Niemegeers (2013). Fairness in wireless networks - issues, measures and challenges. *IEEE Communications Surveys & Tutorials 16*, 5–24.

Shreedhar, M. and G. Varghese (1996). Efficient fair queuing using deficit round-robin. *IEEE/ACM Transactions on Networking 4*, 375–385.

Singh, N., D. Gunawardena, A. Proutiere, B. Radunovic, H. Balan, and P. Key (2011). Efficient and fair mac for wireless networks with self-interference cancellation. *In Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt) IEEE*, 94–101.

Sundaresan, K., M. Khojastepour, E. Chai, and S. Rangarajan (2014). Full-suplex without strings: Enabling fuplex with half-duplex clients. *Mobicom*, 55–66.

Tang, A. and X. Wang (2015). A-duplex: Medium access control for efficient coexistence between full duplex and half duplex communications. *IEEE Transactions on Wireless Communications 14*(10), 5871–5885.

Tse, D. and P. Viswanath (2005). *Fundamentals of Wommunication Communication* (1 ed.). USA: Cambridge University Press.

Wang, X. (2005). An FDD wideband CDMA MAC protocol with minimum-power allocation and GPS-scheduling for wireless wide area multimedia networks. *IEEE Transactions on Mobile Computing 4*(1), 16–28.