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

---

Impact of Physical Layer Techniques on Wireless Networks

by

Shanshan Wu

A thesis submitted in partial fulfillment of the
requirements for the degree of Master of Science in
Electronics Science and Technology at Shanghai Jiao Tong University

---

Committee in charge:
Associate Professor Xudong Wang, Chair
Associate Professor Jun Zhang
Assistant Professor Xinen Zhu

Shanghai
January, 2014

Abstract

This thesis provides frameworks for evaluating the impact of different physical layer techniques on the network performance. The emphasis of the framework is on how to analytically characterize various performance metrics as functions of different parameters. By analyzing those functions, insights for design and optimization over different parameters can be obtained. Two physical layer techniques are considered separately in this thesis, because of their potential for improving network throughput. They are two-way relaying and multi-user MIMO (MU-MIMO). Each of them corresponds to a research topic.

In the first research topic, we consider routing path selection in a two-way relay network. Information theoretical analysis is carried out to derive bandwidth efficiency and energy efficiency of a linear multi-hop network with transmission through two-way relay channels. Such analysis provides a framework of routing path selection by considering bandwidth efficiency, energy efficiency, and latency subject to physical layer constraints such as the transmission rate, processing power, path length, and the number of relays. This framework provides insightful guidelines for future routing protocol design in a two-way relay network.

In the second research topic, we develop an analytical model to investigate the saturation throughput and mean access delay performance of a recently proposed CSMA/CA-based MU-MIMO wireless LAN. We propose and model a distributed opportunistic transmission scheme, where clients are able to contend for the concurrent transmission opportunities only when their concurrent rates exceed a threshold. Comparisons with simulation results show that our analytical model provides a close estimation of the network performance. By means of the proposed model, we evaluate the throughput and delay performance with respect to different network parameters, including the backoff window sizes, the number of antennas at the AP, the network size, and the threshold of the opportunistic transmission scheme. The analytical model and performance analysis provide optimization over various parameters and an in-depth understanding of the random access MU-MIMO transmission scheme.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In the last decade, wireless communication networks have experienced an explosive growth in users' demand for high data rates to meet the need of new services like mobile micro-blogging, online gaming, and high-definition video streaming. This demand, as forecasted by Cisco (2013), will continue to escalate over the next five years. The key to meeting this rapidly increasing demand involves how to effectively incorporate powerful new physical layer techniques into wireless systems. However, many proposed physical layer techniques are built upon a simple network model with certain assumptions. Although those techniques are able to offer significant performance gains under their assumed network models, it is still unknown how they will perform when they are employed in a different network scenario. To address this problem, this thesis proposes mathematical frameworks for evaluating the impact of certain physical layer techniques on the network performance under a network scenario different from their assumed network models.

In this thesis two physical layer techniques are considered because of their potential for improving network throughput. They are two-way relaying and multi-user MIMO (MU-MIMO). Each of them corresponds to a research topic. For each topic, a framework is developed to analytically characterize network performance as functions of different parameters when the

corresponding physical layer technique is applied. By investigating those functions, insights for design and optimization over key parameters can then be obtained. Detailed introduction of the two physical layer techniques and the research topics are presented in the following two sections.

1.1 Two-Way Relaying

A typical model for a two-way relay channel (TWRC) contains three nodes (Popovski and Yomo, 2006a,b): two source nodes $\mathbb{A}$ and $\mathbb{B}$ exchange data via a relay $\mathbb{R}_1$ (see Fig. 1.1). Assuming half-duplex transmission, then an amplify-and-forward (AF) TWRC (Popovski and Yomo, 2006b) works as follows. In the first time slot, $\mathbb{A}$ and $\mathbb{B}$ transmit simultaneously to $\mathbb{R}_1$, where the superimposed signals get amplified and transmitted in the second time slot. After receiving it, $\mathbb{A}$ and $\mathbb{B}$ subtract their own signal to extract the desired data. By exploiting bi-directional interference, TWRC enables two source nodes to exchange data every two time slots, which is half the time needed in a traditional routing scheme. In spite of the advantage of TWRC in a three-node scenario, how to efficiently utilize it to acquire performance gains in a large wireless network remains an unsolved problem.

The objective of this research topic is to provide insightful guidelines towards routing path selection when transmitting data through TWRCs (Fig. 1.1), by choosing energy efficiency (EE), bandwidth efficiency (BE), and latency as performance metrics. To achieve this goal, this research topic is divided into three tasks. In the first task, we develop a framework to evaluate performance of a linear multi-hop network. In the second task, the EE-BE performance of a linear multi-hop network is numerically investigated with respect to different variables, such as the transmission rate and the total number of relays. In the third task, we apply the proposed framework to routing path selection in a two-way relay network. Specifically, a general objective function is defined to integrate the three performance metrics. Based on this function, an

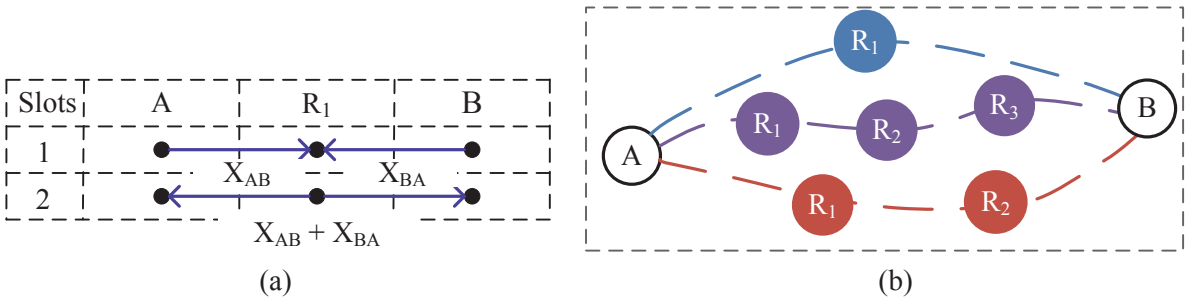


Figure 1.1: (a) Illustration of an AF TWRC; (b) $\mathbb{A}$ and $\mathbb{B}$ want to exchange data through AF TWRCs, which one of the three routes performs the best?

optimal route to provide the best tradeoff between EE, BE, and latency can be determined. Note that although there are different two-way relaying techniques (e.g., decode-and-forward), we focus on AF TWRC because of its simplicity in implementation and comparable performance with other techniques (Popovski and Yomo, 2006a,b).

1.2 Multi-User MIMO

A multiuser MIMO (MU-MIMO) network contains a multi-antenna access point (AP) and multiple clients. Those clients usually have small physical sizes and limited power. Hence, each client is normally equipped with a single transmit antenna. Multiple clients can communicate concurrently with the AP in both the uplink (many-to-one) and downlink (one-to-many) (Tse and Viswanath, 2005). With spatial multiplexing and antenna diversity, an MU-MIMO system offers a high network throughput that increases with the number of antennas at the AP. However, the distributed coordination function (DCF) in current wireless LANs (WLANs) only allows one client to transmit at a time, and hence underutilizes the MU-MIMO capability in the uplink. Moreover, a random access-based MAC protocol is highly preferred in a WLAN because it allows users to access the medium in a simple manner. Therefore, how to enable multiple clients to transmit concurrently while keeping the random access property becomes a hot topic recently (Tan et al., 2009; Lin et al., 2011; Shen et al., 2012). Tan et al. (2009) have developed a CSMA/CA-based MAC protocol in an MU-MIMO WLAN, which enables multiple

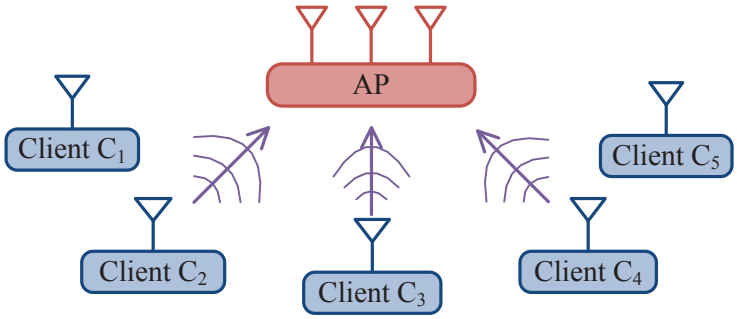


Figure 1.2: An example of MU-MIMO network with a three-antenna AP and five single-antenna clients.

clients to communicate with the AP concurrently. This transmission scheme is then improved by Shen et al. (2012), where the optimal bit rate adaptation scheme in a CSMA/CA-based MU-MIMO WLAN is proposed and the optimal bit rate is picked for each client by considering the interference from ongoing transmissions. In (Lin et al., 2011), a similar CSMA/CA-based MAC protocol is developed in an MIMO network, in which nodes with more antennas can join the ongoing transmissions without interfering them.

The objective of this research topic is to provide a in-depth understanding of the transmission mechanism in a CSMA/CA-based MU-MIMO WLAN (Tan et al., 2009; Shen et al., 2012). To achieve this goal, this research topic is divided into four tasks. In the first task, a theoretical model is developed to characterize the network performance of the uplink channel in a CSMA/CA-based MU-MIMO WLAN under the saturation condition. Our derivation is based on Bianchi's Markov chain model Bianchi (2000), but is different from Bianchi's model in three aspects: the derivation of conditional collision probability, the formulation of saturation throughput and mean access delay. These three aspects essentially capture the difference between a conventional 802.11 MAC protocol and a CSMA/CA-based MAC in an MU-MIMO WLAN. The concurrent transmission rates are formulated by assuming that clients experience i.i.d. time-varying Rayleigh fading. In the second task, the CSMA/CA-based MU-MIMO transmission scheme is revised by incorporating the opportunistic communication mechanism. An analytical model is also developed to characterize the network performance of this opportunis-

tic transmission scheme. In the third task, comparisons between the analytical and simulation results are carried out to verify our model. Particularly, we are interested in two questions: 1) whether our analytical model can closely predict the saturation throughput and mean access delay; 2) how the accuracy of the analytical model varies with respect to different network parameters and what are the possible reasons for this accuracy fluctuation. In the fourth task, the throughput and delay performance are evaluated by means of the developed model. By investigating the influence of different parameters, such as the number of antennas at the AP and the threshold defined in the opportunistic scheme, we are able to gain insights into the random access MU-MIMO transmission mechanism.

Chapter 2

Theoretic Study on Routing Path Selection in Two-Way Relay Networks

This chapter presents the concrete steps we take to complete the first research topic. In Section 2.1, the system model and the applied scheduling scheme are described. An information theoretical framework is developed in Section 2.2 to derive the EE, BE, and latency of a linear multi-hop network, assuming that Hop-by-Hop scheduling scheme is used to enable transmission through TWRCs. Besides, an optimal power allocation scheme is derived, which allows a multi-hop network to consume the smallest energy with a given transmission rate. Numerical evaluation of the network performance is carried out in Section 2.3. This chapter is summarized in Section 2.4.

2.1 System Model

We consider a linear multi-hop network with two source nodes $\mathbb{A}$ and $\mathbb{B}$, and k relays $\mathbb{R}_1$,...,$\mathbb{R}_k$ in between. $\mathbb{A}$ and $\mathbb{B}$ want to exchange packets with each other, with the help of k relays over AF TWRCs. Every node is assumed to have single antenna and one-hop transmission range, and operate in half-duplex mode. Nodes within two hops cannot transmit simultaneously, except

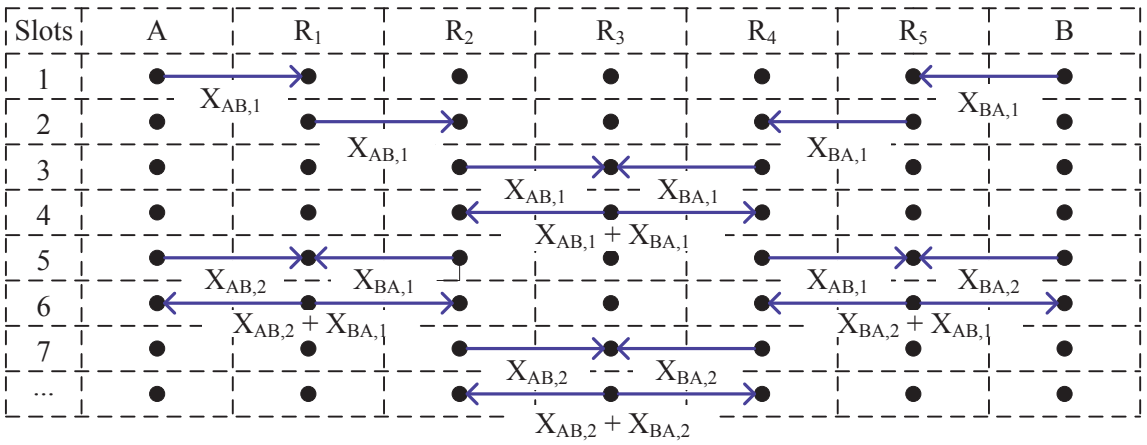


Figure 2.1: Illustration of the Hop-by-Hop scheme.

in TWRCs.

To enable efficient transmissions over AF TWRCs, a simple scheduling scheme is proposed by (You et al., 2011), which allows $\mathbb{A}$ and $\mathbb{B}$ to exchange one packet every four time slots. Note that a different multi-hop scheme exists Popovski and Yomo (2006b), which allows $\mathbb{A}$ and $\mathbb{B}$ to exchange one packet every two time slots. However, it is not applicable in real systems, because noise will accumulate rapidly as packets traverse the network, resulting in a fast decreasing SNR for new transmitted packets. Therefore, we consider the scheduling scheme in (You et al., 2011). To see how this scheme works, let us take $k = 5$ as an example (see Fig. 2.1). Let $X_{\mathbb{AB},n}$ ($X_{\mathbb{BA},n}$) be the n-th packets that $\mathbb{A}$ ($\mathbb{B}$) wants to send to $\mathbb{B}$ ($\mathbb{A}$). During the first two time slots, $X_{\mathbb{AB},1}$ and $X_{\mathbb{BA},1}$ are inserted into the network and are forwarded to relays $\mathbb{R}_2$ and $\mathbb{R}_4$, respectively. An AF TWRC is then formed among the three relays in the center, allowing $\mathbb{R}_2$ and $\mathbb{R}_4$ to exchange their packets. In the fifth and sixth time slots, two TWRCs are formed symmetrically on both sides of the network, where $\mathbb{R}_2$ ($\mathbb{R}_4$) swaps its packet with $\mathbb{A}$ ($\mathbb{B}$), so that $X_{\mathbb{BA},1}$ and $X_{\mathbb{AB},1}$ reach their destinations, and two new packets $X_{\mathbb{AB},2}$ and $X_{\mathbb{BA},2}$ enter the system. By then, packets are transmitted in a stable and recursive pattern: during every four time slots, the source nodes will insert one new pair of packets into the system, and receive one pair of packets from the other side; all the relays in between help forward data over AF TWRCs.

The scheduling scheme proposed in (You et al., 2011) only considers networks with *odd*

number of relays. In the next section, we generalize this scheme to take into account all values of k, and name it a Hop-by-Hop scheme because TWRCs are formed every two hops. As illustrated in Fig. 2.3–2.7, the recursive pattern varies for different values of k: when k is odd, all the nodes are involved in TWRCs; when k is even, one of the source nodes will perform traditional transmission.

2.2 Performance Measures

In this section we investigate the performance of linear multi-hop networks with a Hop-by-Hop scheduling scheme. The performance metrics of interest are: EE (bits/J), BE (bits/s/Hz), and latency (time slots/bit). EE and BE are measures of the efficiency that a network utilizes energy and spectrum to transmit data (Bae and Stark, 2009) (Verdú, 2002), respectively, while latency refers to the time elapse for each bit to reach its destination after being sent out. By deriving those metrics, we are then motivated to determine the best performance that a network can achieve. Specifically, we are going to find the highest EE when BE is given. Since EE and BE depends on how power is provisioned at each node, an optimal power allocation scheme associated with the highest EE value will also be determined.

Since Shannon's capacity formula is applied to derive the optimal power allocation, all sources of interference are taken into account in the SINR part. Therefore, the complexity of derivation increases as the number of relays (i.e., k) grows. In this chapter, we only consider small-scale multi-hop networks with $0 \leq k \leq 6$. However, our framework can be easily extended to the case when $k > 7$. Moreover, as discussed in Section 2.3, analyzing 7 cases (i.e., $0 \leq k \leq 6$) is enough to reveal the performance trend of linear multi-hop networks.

For each value of k and each node i, let x_i and y_i denote the transmitted and received symbols, respectively. z_i represents a zero-mean AWGN with power spectral density N_0. Let P_i be the transmission energy carried by each transmitted symbol x_i. Then P_i is the transmission

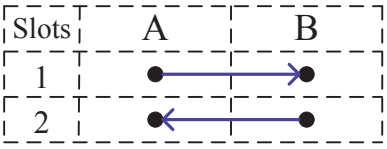


Figure 2.2: Recursive pattern of the Hop-by-Hop scheme when $k = 0$.

power (unit: J/s) evaluated in a time scale of one symbol, so it has a unit of J/symbol or J/channel use. Let P_{proc} denote the processing energy, which can be calculated as (Wang et al., 2006):

$$P_{proc} = (1/\eta - 1)P_{tr} + P_0, \tag{2.1}$$

where η is a constant representing the drain efficiency of the power amplifier at the transmitter. The first term of Eq. (2.1) denotes processing energy in power amplifier, which is linearly proportional to the transmission energy P_{tr}. The second term P_0 denotes energy consumed in radio electronics (at both transmitter and receiver sides) other than power amplifier, and is assumed to be a constant value. Since P_{proc} depends on the network scenario, in the following derivation, we will use $P_{proc,k}$ and $P_{0,k}$ to indicate networks with k relays.

For simplicity, we assume that all the nodes are equally spaced and transmit with the same rate R bits/channel use. Here each channel use occupies 1 second $\times$ hertz. The channel is assumed to be reciprocal. Let h denote the channel between two consecutive nodes. The relationship between h and distance d is $|h|^2 = d^{-\alpha}$, where α is pass-loss exponent and normally $2 \leq \alpha \leq 4$. This is a common assumption on the relationship between channel gain and distance, e.g., see (Bae and Stark, 2009).

2.2.1 Direct Transmission

As shown in Fig. 2.2, in the case of $k = 0$, $\mathbb{A}$ communicates with $\mathbb{B}$ directly.

BE and EE

Define $U = P_{\mathbb{A}} + P_{\mathbb{B}}$, then it denotes the transmission energy during two channel uses. According to Eq. (2.1), the corresponding processing energy in the two channel uses is: $P_{proc,0} = (1/\eta - 1)U + P_{0,0}$. Then the total energy is: $U + P_{proc,0} = U/\eta + P_{0,0}$. Hence,

$$\text{BE} = R, \;\; \text{EE} = 2R/(U/\eta + P_{0,0}). \tag{2.2}$$

Optimal Power Allocation

Given BE, i.e., given R, EE is maximized when U reaches its minimum, which is achieved when we use capacity-achieving coding, i.e.,

$$R = \log_2(1 + |h|^2 P_i/N_0), \; i = \mathbb{A}, \; \mathbb{B}. \tag{2.3}$$

Accordingly, the minimum U and the optimal powers are

$$U_{\min} = 2N_0|h|^{-2}(2^R - 1), \;\; P_{\mathbb{A}} = P_{\mathbb{B}} = U_{\min}/2. \tag{2.4}$$

2.2.2 Two-Way Relay Channel

This case is depicted in Fig. 1.1, where $\mathbb{A}$, $\mathbb{R}_1$, and $\mathbb{B}$ form a TWRC. Packets are exchanged every 2 time slots.

BE and EE

Define the transmission energy during the two channel uses as $T = P_{\mathbb{A}} + P_{\mathbb{R}_1} + P_{\mathbb{B}}$, then following the same procedure as in previous case, we get the total energy: $T/\eta + P_{0,1}$. Hence,

$$\text{BE} = R, \;\; \text{EE} = 2R/(T/\eta + P_{0,1}). \tag{2.5}$$

Optimal Power Allocation

Given BE, the maximum EE is achieved when T is minimum. As shown in Fig. 1.1, $\mathbb{A}$, $\mathbb{B}$, and $\mathbb{R}_1$ forms a TWRC. Assuming that the signal received at $\mathbb{R}_1$ is amplified by β, i.e.,

$$\beta = \sqrt{P_{\mathbb{R}_1}/(|h|^2 P_{\mathbb{A}} + |h|^2 P_{\mathbb{B}} + N_0)}, \tag{2.6}$$

after self-interference cancellation, signals at $\mathbb{A}$ and $\mathbb{B}$ become

$$\begin{aligned} y_{\mathbb{A}} - h^2\beta x_{\mathbb{A}} &= h^2\beta x_{\mathbb{B}} + h\beta z_{\mathbb{R}_1} + z_{\mathbb{A}}, \\ y_{\mathbb{B}} - h^2\beta x_{\mathbb{B}} &= h^2\beta x_{\mathbb{A}} + h\beta z_{\mathbb{R}_1} + z_{\mathbb{B}}. \end{aligned} \tag{2.7}$$

Energy is used with maximum efficiency when transmission is done by capacity-achieving coding, i.e.,

$$R = \log_2(1 + \frac{|h|^4\beta^2 P_{\mathbb{B}}}{(|h|^2\beta^2 + 1)N_0}) = \log_2(1 + \frac{|h|^4\beta^2 P_{\mathbb{A}}}{(|h|^2\beta^2 + 1)N_0}). \tag{2.8}$$

Here β is assumed to be a real number. If it is a complex number, then we need to use $|\beta|$ instead of β during derivation. Rearrange Eq. (2.8), substitute it into T, and let the derivative of T with respect to β be equal to zero gives

$$\beta^2 = \sqrt{(2^{R+1} - 2)|h|^{-4}/(2^{R+1} - 1)}. \tag{2.9}$$

Checking the second derivation of T shows that when β satisfies Eq. (2.9), T reaches its mimimum

$$T_{\min} = 2N_0|h|^{-2}(\sqrt{(2^{R+1} - 1)(2^{R+1} - 2)} + 2^{R+1} - 2), \tag{2.10}$$

which gives a maximum EE. The corresponding optimal power allocation at $\mathbb{A}$, $\mathbb{B}$, and $\mathbb{R}_1$ can be found by substituting Eq. (2.9) into Eqs. (2.8) and (2.6). Their expressions can be found in our report (Wu and Wang, 2012) and are omitted here for saving space.

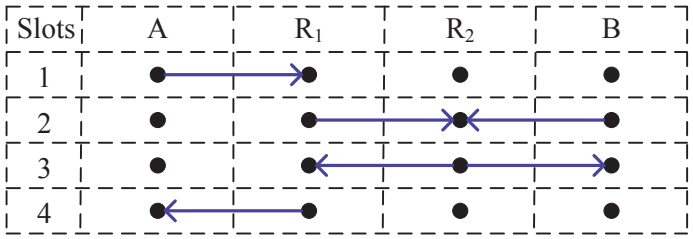


Figure 2.3: Recursive pattern of the Hop-by-Hop scheme when $k = 2$.

2.2.3 Two Relays

The recursive pattern for $k = 2$ is shown in Fig. 2.3. During the first and fourth time slots, $\mathbb{A}$ sends and receives one packet from $\mathbb{R}_1$ through unicast transmission. During the second and third time slots, $\mathbb{R}_1$, $\mathbb{R}_2$, and $\mathbb{B}$ form a TWRC. Packets between $\mathbb{A}$ and $\mathbb{B}$ are exchanged every 4 time slots.

BE and EE

Let $U = P_{\mathbb{A}} + P_{\mathbb{R}_1}$ and $T = P_{\mathbb{R}_1} + P_{\mathbb{R}_2} + P_{\mathbb{B}}$ be the transmission energy dissipated in unicast channel and TWRC, respectively, following the same procedure as in the case of $k = 0$ gives the total energy: $(U + T)/\eta + P_{0,2}$. Hence,

$$\text{BE} = R/2, \quad \text{EE} = 2R/((U + T)/\eta + P_{0,2}). \tag{2.11}$$

Note that U and T have already been defined in the previous cases, but they are reused here to denote energy consumption in the same transmission scenario, i.e., U for unicast channel and T for TWRC. For the same reason, symbols like S and F are reused in the following derivation.

Optimal Power Allocation

Given BE, EE is maximized when $U + T$ is minimized. Since energy in U and T are consumed in different time slots, $U + T$ achieves minimum when both of U and T are minimized. On the other hand, the minimum of U and T, and the associated optimal power allocation scheme

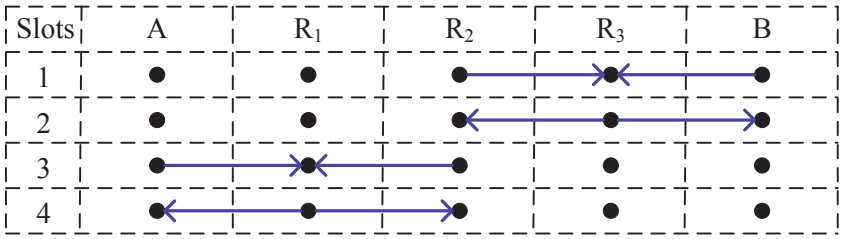


Figure 2.4: Recursive pattern of the Hop-by-Hop scheme when $k = 3$.

have been found in the cases of $k = 0$ and $k = 1$, so previous results, e.g., Eqs. (2.4) and (2.10), can be directly applied here.

2.2.4 Three Relays

The case of $k = 3$ is shown in Fig. 2.4, where the left half nodes and the right half nodes form two TWRCs, respectively. $\mathbb{A}$ and $\mathbb{B}$ exchange one packet every 4 time slots.

BE and EE

Let $T_1 = P_{\mathbb{A}} + P_{\mathbb{R}_1} + P_{\mathbb{R}_2}$ and $T_2 = P_{\mathbb{R}_2} + P_{\mathbb{R}_3} + P_{\mathbb{B}}$ represent the transmission energy dissipated in the two TWRCs, following the same procedure as in the case of $k = 0$ gives us the total energy: $(T_1 + T_2)/\eta + P_{0,3}$. Hence,

$$\text{BE} = R/2, \quad \text{EE} = 2R/((T_1 + T_2)/\eta + P_{0,3}). \tag{2.12}$$

Optimal Power Allocation

Given BE, EE is maximized when $T_1 + T_2$ is minimized. For the same reason as in the case of $k = 2$, previous results, e.g., Eq. (2.10), can be directly applied here to get the minimum of $T_1 + T_2$ and the associated optimal power allocation scheme.

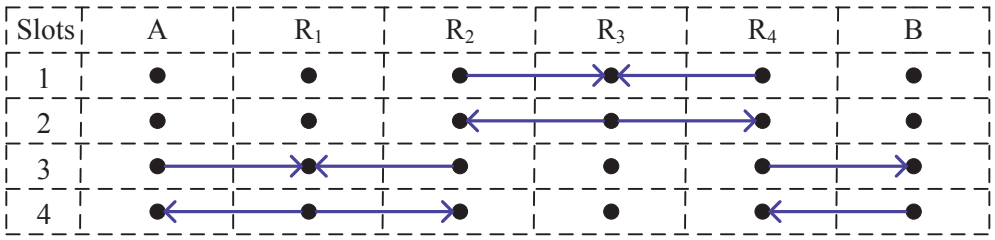


Figure 2.5: Recursive pattern of the Hop-by-Hop scheme when $k = 4$.

2.2.5 Four Relays

This case is shown in Fig. 2.5. During the first two time slots, data is exchanged between $\mathbb{R}_2$ and $\mathbb{R}_4$. During the last two time slots, each of $\mathbb{A}$ and $\mathbb{B}$ sends and receives a new packet. Accordingly, a recursive pattern of 4 time slots is formed.

BE and EE

Let $T = P_{\mathbb{R}_2} + P_{\mathbb{R}_3} + P_{\mathbb{R}_4}$, $F = P_{\mathbb{A}} + P_{\mathbb{R}_1} + P_{\mathbb{R}_2} + P_{\mathbb{R}_4} + P_{\mathbb{B}}$ denote the transmission energy consumed during the first two and last two time slots, respectively. Following the same procedure as in the case of $k = 0$, we get the total energy consumption: $(T + F)/\eta + P_{0,4}$. Hence,

$$\text{BE} = R/2, \quad \text{EE} = 2R/((T + F)/\eta + P_{0,4}). \tag{2.13}$$

Optimal Power Allocation

Given BE, EE is maximized when $T + F$ is minimized. Since T and F represents energy consumed in different time slots, $T + F$ is minimized when both of T and F are minimized. $T_{\min}$ have been derived in the case of $k = 1$, i.e., Eq. (2.10), so we only need to find $F_{\min}$.

During the third and fourth time slots, $\mathbb{A}$ communicates with $\mathbb{R}_2$ through a TWRC. At the same time, $\mathbb{B}$ and $\mathbb{R}_4$ perform unicast transmission. Therefore, we have

$$y_{\mathbb{R}_1} = hx_{\mathbb{A}} + hx_{\mathbb{R}_2} + z_{\mathbb{R}_1} + \sqrt{3^{-\alpha}} hx_{\mathbb{R}_4}, \tag{2.14}$$

$$x_{\mathbb{R}_1} = \beta y_{\mathbb{R}_1}, \tag{2.15}$$

$$y_{\mathbb{A}} = h x_{\mathbb{R}_1} + z_{\mathbb{A}} + \sqrt{5^{-\alpha}} h x_{\mathbb{B}}, \tag{2.16}$$

$$y_{\mathbb{R}_2} = h x_{\mathbb{R}_1} + z_{\mathbb{R}_2} + \sqrt{3^{-\alpha}} h x_{\mathbb{B}}, \tag{2.17}$$

$$y_{\mathbb{B}} = h x_{\mathbb{R}_4} + z_{\mathbb{B}} + \sqrt{3^{-\alpha}} h x_{\mathbb{R}_2} + \sqrt{5^{-\alpha}} h x_{\mathbb{A}}, \tag{2.18}$$

$$y_{\mathbb{R}_4} = h x_{\mathbb{B}} + z_{\mathbb{R}_4} + \sqrt{3^{-\alpha}} h x_{\mathbb{R}_1}, \tag{2.19}$$

where $\sqrt{3^{-\alpha}}$ and $\sqrt{5^{-\alpha}}$ come from two assumptions: 1) relays are equally spaced along a line, and 2) relationship between the channel h and distance d is $|h|^2 = d^{-\alpha}$.

EE achieves maximum when capacity-achieving coding is applied, therefore, the following SINR values at receivers $\mathbb{A}$, $\mathbb{R}_2$, $\mathbb{B}$, and $\mathbb{R}_4$, respectively, should all be equal to $2^R - 1$:

$$\frac{|h|^4 \beta^2 P_{\mathbb{R}_2}}{3^{-\alpha}|h|^4 \beta^2 P_{\mathbb{R}_4} + (|h|^2\beta^2 + 1)N_0 + 5^{-\alpha}|h|^2 P_{\mathbb{B}}}, \tag{2.20}$$

$$\frac{|h|^4 \beta^2 P_{\mathbb{A}}}{3^{-\alpha}|h|^4 \beta^2 P_{\mathbb{R}_4} + (|h|^2\beta^2 + 1)N_0 + 3^{-\alpha}|h|^2 P_{\mathbb{B}}}, \tag{2.21}$$

$$\frac{|h|^2 P_{\mathbb{R}_4}}{3^{-\alpha}|h|^2 P_{\mathbb{R}_2} + 5^{-\alpha}|h|^2 P_{\mathbb{A}} + N_0}, \tag{2.22}$$

$$\frac{|h|^2 P_{\mathbb{B}}}{3^{-\alpha}|h|^2 P_{\mathbb{R}_1} + N_0}. \tag{2.23}$$

Additionally, Eqs. (2.15) and (2.14) give

$$P_{\mathbb{R}_1} = \beta^2(|h|^2 P_{\mathbb{A}} + |h|^2 P_{\mathbb{R}_2} + 3^{-\alpha}|h|^2 P_{\mathbb{R}_4} + N_0). \tag{2.24}$$

There are five equations, i.e., (2.20)–(2.24), and five variables, i.e., $P_{\mathbb{A}}$, $P_{\mathbb{R}_1}$, $P_{\mathbb{R}_2}$, $P_{\mathbb{R}_4}$, $P_{\mathbb{B}}$.

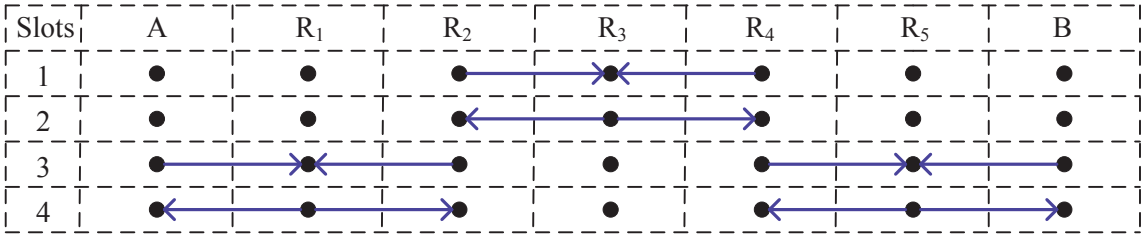


Figure 2.6: Recursive pattern of the Hop-by-Hop scheme when $k = 5$.

Solving these equations and substituting the results into the formula of F gives

$$F = (c_1\beta^2 + \frac{c_2}{|h|^4\beta^2} + c_3|h|^{-2})N_0 \geq (2\sqrt{c_1c_2} + c_3)N_0|h|^{-2}, \tag{2.25}$$

where $\{c_i\}_{i=1}^3$ are complicated functions of R and α. Accordingly, the minimum value of F is $F_{\min} = (2\sqrt{c_1c_2} + c_3)N_0|h|^{-2}$, achieved when $\beta^4 = c_2|h|^{-4}/c_1$. Here we do not prove that $c_1c_2 > 0$, but $F_{\min}$ are achievable as demonstrated in Section 2.3. The corresponding optimal power allocation can be obtained by solving Eqs. (2.20)–(2.24) and substituting the optimal β.

The complete expressions of $F_{\min}$ and the optimal power, as well as the detailed derivation are available in Appendix B, which can also be found in (Wu and Wang, 2012).

2.2.6 Five Relays

The recursive pattern for $k = 5$ is shown in Fig. 2.6, which has been analyzed in Section 2.1.

BE and EE

Define $T = P_{\mathbb{R}_2} + P_{\mathbb{R}_3} + P_{\mathbb{R}_4}$ and $S = P_A + P_{\mathbb{R}_1} + P_{\mathbb{R}_2} + P_{\mathbb{R}_4} + P_{\mathbb{R}_5} + P_B$ as the transmission energy consumed during the first and last two time slots, respectively. Following the same procedure as in the case of $k = 0$ gives the total energy consumption: $(T + S)/\eta + P_{0,5}$. Hence,

$$\text{BE} = R/2, \quad \text{EE} = 2R/((T + S)/\eta + P_{0,5}). \tag{2.26}$$

Optimal Power Allocation

Similar to the previous case, given BE, EE is maximized when both T and S are minimized. Since $T_{\min}$ is given in Eq. (2.10), we only need to find $S_{\min}$ and the corresponding power.

During the third and fourth time slots, two TWRCs are formed. Due to symmetry, we only need to consider one of them, and the total power is minimized when $P_{\mathbb{A}} = P_{\mathbb{B}}$, $P_{\mathbb{R}_2} = P_{\mathbb{R}_4}$, $P_{\mathbb{R}_1} = P_{\mathbb{R}_5}$. Accordingly, $S = 2(P_{\mathbb{A}} + P_{\mathbb{R}_1} + P_{\mathbb{R}_2})$. Considering the left TWRC formed by $\mathbb{A}$, $\mathbb{R}_1$, and $\mathbb{R}_2$ gives

$$y_{\mathbb{R}_1} = hx_{\mathbb{A}} + hx_{\mathbb{R}_2} + \sqrt{3^{-\alpha}}hx_{\mathbb{R}_4} + \sqrt{5^{-\alpha}}hx_{\mathbb{B}} + z_{\mathbb{R}_1}, \tag{2.27}$$

$$x_{\mathbb{R}_1} = \beta y_{\mathbb{R}_1}, \tag{2.28}$$

$$x_{\mathbb{A}} = hx_{\mathbb{R}_1} + \sqrt{5^{-\alpha}}hx_{\mathbb{R}_5} + z_{\mathbb{A}}, \tag{2.29}$$

$$x_{\mathbb{R}_2} = hx_{\mathbb{R}_1} + \sqrt{3^{-\alpha}}hx_{\mathbb{R}_5} + z_{\mathbb{R}_2}, \tag{2.30}$$

where $\sqrt{3^{-\alpha}}$ and $\sqrt{5^{-\alpha}}$ follow the same reason as in the previous case.

Energy is minimized when capacity-achieving coding is used, so the following SINR values at receivers $\mathbb{A}$ and $\mathbb{R}_2$ should be equal to $2^R - 1$:

$$\frac{|h|^4\beta^2 P_{\mathbb{R}_2}}{|h|^2(P_{\mathbb{R}_1} - \beta^2|h|^2(P_{\mathbb{A}} + P_{\mathbb{R}_2})) + 5^{-\alpha}|h|^2 P_{\mathbb{R}_5} + N_0}, \tag{2.31}$$

$$\frac{|h|^4\beta^2 P_{\mathbb{A}}}{|h|^2(P_{\mathbb{R}_1} - \beta^2|h|^2(P_{\mathbb{A}} + P_{\mathbb{R}_2})) + 3^{-\alpha}|h|^2 P_{\mathbb{R}_5} + N_0}. \tag{2.32}$$

From Eqs. (2.27) and (2.28), we have

$$P_{\mathbb{R}_1} = \beta^2|h|^2(P_{\mathbb{A}} + P_{\mathbb{R}_2} + 3^{-\alpha}P_{\mathbb{R}_4} + 5^{-\alpha}P_{\mathbb{B}} + N_0|h|^{-2}). \tag{2.33}$$

Similar to the case of $k = 4$, we can solve the 3 equations, i.e., (2.31)–(2.33) to get $P_{\mathbb{A}}$, $P_{\mathbb{R}_1}$,

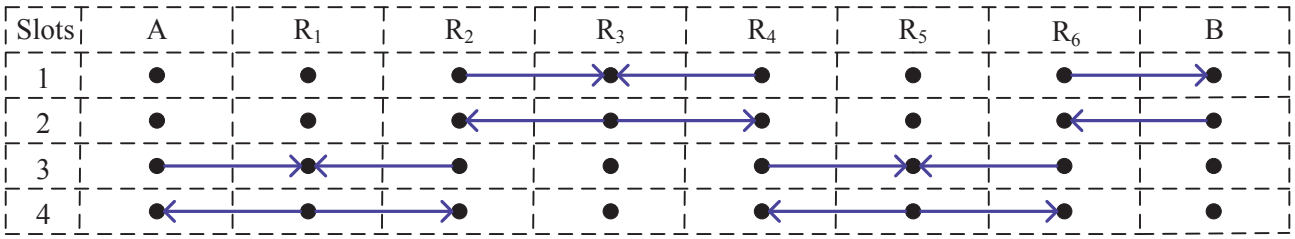


Figure 2.7: Recursive pattern of the Hop-by-Hop scheme when $k = 6$.

and $P_{\mathbb{R}_2}$, and then substitute them into the formula of S to get

$$S = (c_4\beta^2 + \frac{c_5}{|h|^4\beta^2} + c_6|h|^{-2})N_0 \geq (2\sqrt{c_4c_5} + c_6)N_0|h|^{-2}, \tag{2.34}$$

where $\{c_i\}_{i=4}^6$ are complicated functions of R and α. Accordingly, $S_{\min} = (2\sqrt{c_4c_5} + c_6)N_0|h|^{-2}$, achieved when $\beta^4 = c_5|h|^{-4}/c_4$. Here we do not prove that $c_4c_5 > 0$, but $S_{\min}$ are achievable as demonstrated in Section 2.3. The corresponding optimal power can be obtained by substituting the optimal β in Eqs. (2.31)–(2.33).

Similar to $F_{\min}$, the complete expressions of $S_{\min}$ and the optimal power, as well as the detailed derivation are available in Appendix B, which can also be found in (Wu and Wang, 2012).

2.2.7 Six Relays

As shown in Fig. 2.7, the system works like the case of $k = 4$ during the first two time slots, and the case of $k = 5$ during the last two time slots.

BE and EE

Define $F = P_{\mathbb{R}_2} + P_{\mathbb{R}_3} + P_{\mathbb{R}_4} + P_{\mathbb{R}_6} + P_{\mathbb{B}}$ and $S = P_{\mathbb{A}} + P_{\mathbb{R}_1} + P_{\mathbb{R}_2} + P_{\mathbb{R}_4} + P_{\mathbb{R}_5} + P_{\mathbb{R}_6}$ as the transmission energy during the first and the last two time slots, respectively. Following the same procedure as in the case of $k = 0$ gives the total energy consumption: $(F + S)/\eta + P_{0,6}$.

Hence,

$$\text{BE} = R/2, \;\; \text{EE} = 2R/((F+S)/\eta + P_{0,6}). \tag{2.35}$$

Optimal Power Allocation

Similar to the previous case, given BE, EE is maximized when both F and S are minimized. Because of the same recursive patterns, $F_{\min}$ and $S_{\min}$ as well as the optimal power allocation have been derived in previous cases of $k = 4$ and $k = 5$, e.g., in Eqs. (2.25) and (2.34).

2.2.8 Latency

As shown in Fig. 2.3–2.7, packets are forwarded to the next hop per time slot, so the latency of a multi-hop network with k relays is $k + 1$ time slots/bit.

2.3 Numerical Analysis

In previous section, we have derived EE, BE, and latency of multi-hop networks with different number of relays. We also found the condition when EE is maximized for a given BE, i,e, when the network consumes the smallest energy for a given end-to-end rate. The goal of this section is to numerically investigate those results. Note that the EE discussed in this section always refers to the *maximum* EE for a given BE.

2.3.1 Evaluation of the EE-BE Performance

The path loss exponent α is 4 and noise power spectral density N_0 is -174 dBm/Hz. To capture the effect of processing energy, we consider two cases: an extreme case of zero processing energy, i.e., $\eta = 1$ and $P_{0,k} = 0$ for all k; and a specific case with* $\eta = 0.75$. For the second case, we assume that $P_{0,k}$ is proportional to the number of senders and receivers in a recursive pattern,

*This drain efficiency is achievable for high-class power amplifiers (Wang et al., 2006).

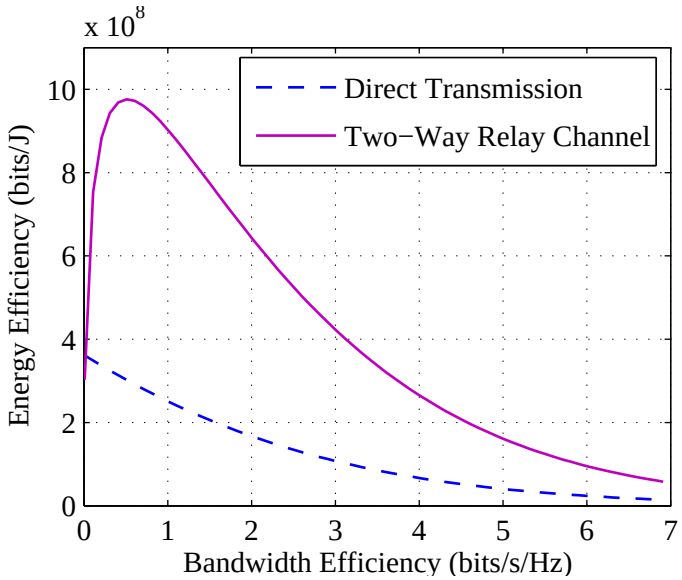


Figure 2.8: Comparison of EE-BE relation for a multi-hop network with end-to-end distance 1000 m: $k = 0$ versus $k = 1$ with zero processing energy.

i.e., $P_{0,k}/P_{0,0} = m/4$, where m is the number of active nodes in a recursive pattern for network with k relays. Let $P_{0,0} = 5 \times 10^{-7}$ mJ/channel use, then $P_{0,1} = 6/4P_{0,0}$, $P_{0,2} = 10/4P_{0,0}$, $P_{0,3} = 12/4P_{0,0}$, $P_{0,4} = 16/4P_{0,0}$, $P_{0,5} = 18/4P_{0,0}$, $P_{0,6} = 22/4P_{0,0}$.

In Fig. 2.8 we compare the EE-BE performance for $k = 0$ and $k = 1$, which corresponds to transmission directly or through a TWRC. The processing energy is ignored. Unlike direct transmission, where EE and BE are involved in a tradeoff relation, transmission through a TWRC presents a different EE-BE interaction: EE first increases and then decreases with BE, achieving maximum when BE (which in this case equals the transmission rate per hop) is around 0.6 bits/s/Hz. In other words, at a low transmission rate, we can decrease energy consumption and increase transmission rate at the same time. Besides, when transmission rate approaches zero, direct transmission tends to have higher EE than TWRC, and hence performs better.

In Fig. 2.9 we plot EE and BE of a linear multi-hop network with a fixed distance between $\mathbb{A}$ and $\mathbb{B}$, but different number of relays. The processing energy is ignored. Given the length of a multi-hop network and a certain BE, EE increases with increased number of relays. This is

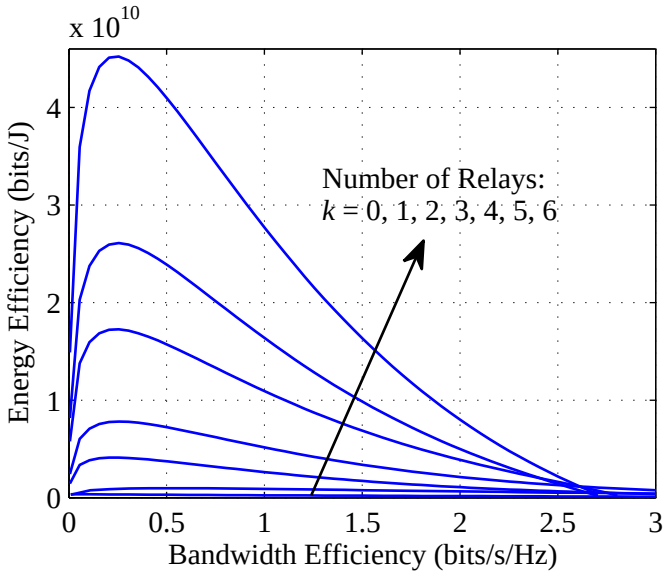


Figure 2.9: Comparison of EE-BE relation for a multi-hop network with end-to-end distance 1000 m: $0 \leq k \leq 6$ with zero processing energy.

because when the length of a multi-hop network is fixed, more relays means smaller distance in each hop, and hence smaller transmission energy consumed in each hop. Although more relays increases the number of devices consuming energy, the decrease due to a smaller distance per hop dominates the total energy.

In Fig. 2.10 we depict EE and BE for the previous case, but with processing energy considered. As shown in this figure, when transmission rate is low, routes with fewer relays have higher EE. This is because at low transmission rate, transmission energy is small, then processing energy dominates the total energy consumption. Since $P_{0,k}$ is assumed to be proportional to the number of active nodes in a recursive pattern, routes with fewer relays dissipate less processing energy and hence have larger EE. As rate increases, the EE-BE relation resembles that in Fig. 2.9, in the sense that a network with more relays tends to have larger EE. When BE approaches 2.8 bits/channel use, a sharp decrease happens for EE of networks with more than 3 relays. This is because an upper bound exists for the transmission rate, due to interference influence. Taking $k = 4$ as an example, the SINR value in Eq. (2.20) equals to $2^R - 1$, since $P_{\mathbb{R}_2}$ in the numerator, and $P_{\mathbb{R}_4}$, $P_{\mathbb{B}}$ in the denominator are of the same order, R is sure to have an

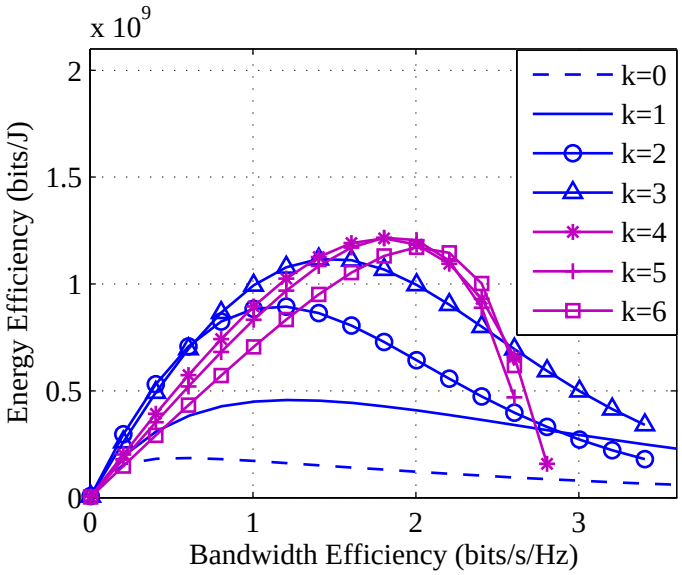


Figure 2.10: Comparison of EE-BE relation for a multi-hop network with end-to-end distance 1000 m: $0 \leq k \leq 6$ with processing energy considered.

upper bound. No interference influence (other than in the TWRC) exists in the transmission scheme when $k \leq 3$, so rate is bounded only for networks with $k \geq 4$.

2.3.2 Optimal Routing Path Selection

To apply our results to the selection of an optimal routing path in Fig. 1.1, we assume that: 1) relays along each route are nearly equidistant; 2) each route has a small curvature so that it can be treated as a straight line. The two assumptions are not always true practically, but are reasonable because routes satisfying the two assumptions perform better. A TWRC with relay in the middle has the highest sum rate and lowest outage probability (Yi et al., 2011). Besides, transmission through a path with large curvature means that a large amount of energy is spent on moving data around the source node rather than forwarding it to the destination, so it is energy inefficient. Previous analysis (see Fig. 2.9) tells us that a route with more relays consumes smaller transmission energy. However, more relays incurs larger latency. In order to find an optimal routing path that gives the best tradeoff between those performance metrics,

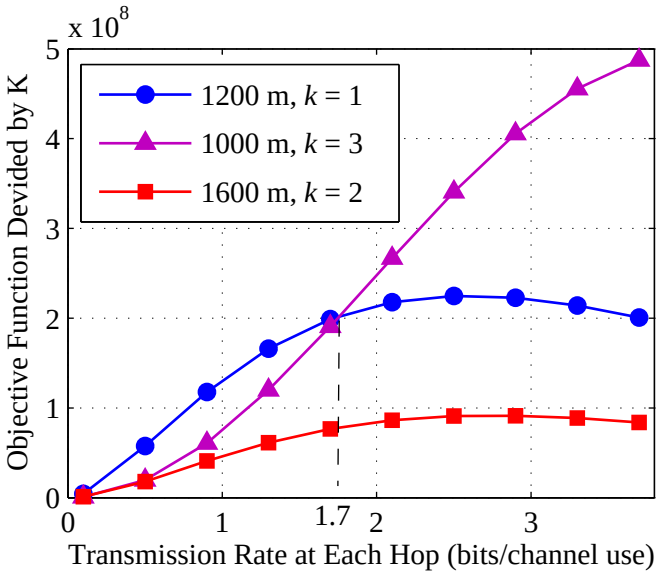


Figure 2.11: Comparison of the three routes in Fig. 1.1, with processing energy considered.

we can define a general performance metric as

$$M = \frac{\mathrm{EE}/\mathrm{EE}_{\max} \times \mathrm{BE}/\mathrm{BE}_{\max}}{\mathrm{latency}/\mathrm{latency}_{\max}} = \frac{\mathrm{EE} \times \mathrm{BE}}{\mathrm{latency}} \times K, \tag{2.36}$$

where $\mathrm{EE}_{\max}$, $\mathrm{BE}_{\max}$, and $\mathrm{latency}_{\max}$ are the maximum achievable EE, BE, and latency of a system. Therefore, K is a constant formed by the three maximum values. M is an objective function that is to be maximized. Note that the definition of this general performance metric is not unique. For example, if one cares more about EE when selecting the routing path, one can design an objective function that gives EE more "weight".

In Fig. 2.11 we depict M/K for the three routes in Fig. 1.1, assuming that their length and the number of relays are: 1200 m with $k = 1$, 1000 m with $k = 3$, and 1600 m with $k = 2$. When the transmission rate is smaller than 1.7 bits/channel use, the first route is optimal, mainly because it has the lowest latency and smallest processing energy consumption. When the transmission rate is large, the second route is optimal, which is consistent with our previous analysis: since the second route has both the shortest length and the largest number of relays among the three routes, it consumes the smallest transmission energy, and hence has the largest

M value. This example illustrates the necessity of jointly considering the transmission rate, processing energy, path length, and the number of relays, in order to select an optimal path that provides the best tradeoff between EE, BE, and latency.

2.4 Summary

In this chapter, an information theoretical framework has been developed to study routing path selection for amplify-and-forward two-way relay networks. Energy efficiency, bandwidth efficiency, and latency for linear multi-hop networks have been formulated under a Hop-by-Hop scheduling scheme. An optimal power allocation scheme has been derived. It allows minimum energy consumption for a given transmission rate. Numerical investigation has provided guidelines for future routing protocol design.

To select an optimal path that provides the best tradeoff between energy efficiency, bandwidth efficiency, and latency, a joint consideration of the transmission rate, processing energy, path length, and the number of relays is necessary. At a low transmission rate, where processing energy dominates the total energy consumption, route with fewer relays has higher energy efficiency and lower latency. As rate increases, transmission energy becomes dominant. Route with a shorter length and more relays consumes less energy at the cost of higher latency. Moreover, an upper bound exists for the transmission rate of routes with more than three relays, which is due to interference influence in the Hop-by-Hop scheduling scheme.

Chapter 3

Performance Analysis of Random Access Multi-User MIMO Wireless LANs

This chapter presents the concrete steps we take to complete the second research topic. In Section 3.1 we describe the main features of a CSMA/CA-based MU-MIMO WLAN. An analytical model is derived in Section 3.2 to compute the saturation throughput and mean access delay of the uplink channel. A simple distributed opportunistic scheme is also introduced and modeled. Comparisons between simulation results and analytical results are carried out in Section 3.3. Related work is presented in Section 3.5. Throughput performance evaluation based on the proposed model is presented in Section 3.4. This chapter is summarized in Section 3.6.

3.1 CSMA/CA-Based MU-MIMO WLAN

In this section we summarize the main characteristics of a CSMA/CA-based MU-MIMO WLAN (see (Tan et al., 2009) for detailed descriptions). We focus on the throughput performance of the uplink channel throughout this chapter, because the downlink channel has been analyzed before, e.g., by Tse and Viswanath (2005).

In the standard 802.11 DCF access scheme, as shown in Fig. 3.1, only one client is allowed

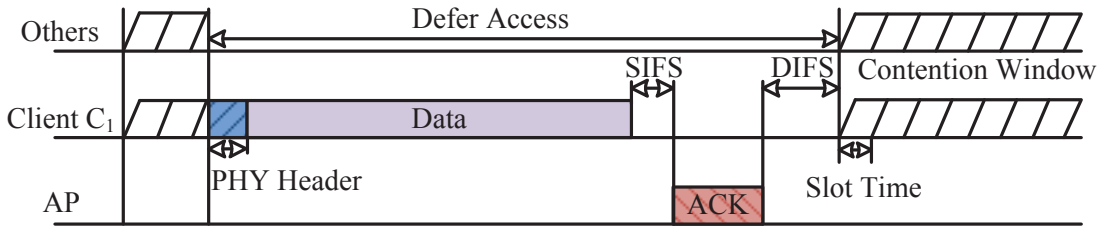


Figure 3.1: Illustration of the standard 802.11 DCF access scheme.

to transmit at a time (*IEEE Std. 802.11*, 2012). Clients who want to transmit data enter the contention period: their *backoff counter*s are reduced each time the channel is sensed idle for a *time slot*. Client that wins the contention, i.e., has zero backoff counter, transmits data, while other clients defer their access to the channel (and stop reducing their backoff counters) until they find the medium is idle for an interval of *distributed interframe space* (DIFS). After that, a new contention period starts.

Unlike the standard 802.11 DCF access scheme, a CSMA/CA-based MU-MIMO WLAN allows multiple clients to transmit concurrently to the AP. For ease of description, let us consider a simple network with one three-antenna AP and five single-antenna clients (Fig. 1.2). Since the AP has three antennas, the maximum number of concurrent clients cannot exceed three (assume that all the clients know this threshold through the AP's beacons). To achieve this goal, every client maintains a *transmission counter* that counts the current number of the concurrent streams by detecting their preambles. Preamble detection can be realized by correlating the received signals with the known preamble. If the transmission counter is smaller than the threshold (which in this case is three since the AP has three antennas), all the rest clients will continue to contend for concurrent transmission opportunities. For example, as shown in Fig. 3.2, when Client C_5 wins the contention and begins transmission, each of the rest clients monitors the channel and detects Client C_5's preamble*, and then increases the transmission

*Sometimes clients also need to decode MAC header of the first contention winner (Tan et al., 2009) before they start to compete for the concurrent transmission opportunities. However, it will not affect the following derivation process of our analytical model. Therefore, here we assume that the contention period starts when previous contention winner finishes transmitting its PHY header (i.e., PLCP preamble and PLCP header), as shown in Fig. 3.2.

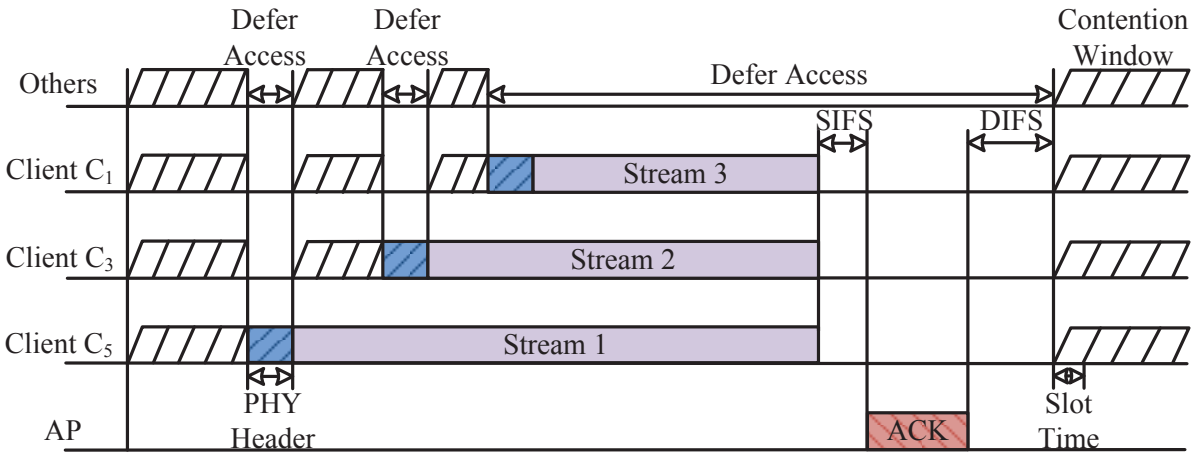


Figure 3.2: Illustration of the CSMA/CA-based MU-MIMO WLAN.

counter from zero to one. Since there is only one client transmitting, which is smaller than three, the rest four clients will continue to contend for the second transmission opportunity. When Client C_3 wins the second contention and transmits, the rest clients behave similarly to the previous case when Client C_5 wins the channel[†] except that the transmission counters are increased from one to two. When the transmission counter is updated to three, i.e., no concurrent transmission opportunity remains, the rest clients will defer their access to the channel until the channel is idle for an interval longer than DIFS. Assuming that the three concurrent transmissions end at the same time (which can be realized by packet fragmentation and aggregation (Lin et al., 2011) (Shen et al., 2012)), the AP will then send an ACK-to-All message to the three clients in the downlink channel through transmit beamforming (Tse and Viswanath, 2005).

A collision happens when two clients win the contention at the same time slot. As an example, consider the beginning period in Fig. 3.6, where both Client C_4 and Client C_5 win the second concurrent transmission opportunity. Although the rest clients can detect the preamble, they do not know that the preamble is actually two overlapping preambles. As a result, Client C_3 wins the third contention and becomes the fourth concurrent transmitter. For successful

[†]Note that the rest clients will stop backoff when they detect a new preamble. Since preamble detection can be done within a slot time, they will not reduce their backoff counters once a client wins the channel and starts to transmit.

decoding, the AP needs to estimate each transmitter's channel parameters using their preambles[‡]. Since the two frames of Client C_4 and Client C_5 overlap together, it is hard for the AP to nullify the interference of the two overlapped frames to extract Client C_3's frame. Besides, because the AP fails to decode the two overlapped frames, it cannot perform successive interference cancellation to extract the first contention winner's (i.e., Client C_1's) frame. In sum, the AP encounters a decoding failure when a collision happens (Tan et al., 2009). In the case of decoding failure, no ACK message is sent to the concurrent transmitters, as shown in Fig. 3.6. Besides, each of the concurrent clients will select a random backoff time interval to prevent future collisions. Here we apply the binary exponential backoff mechanism which is also used in the conventional 802.11 WLANs. This backoff mechanism works as follows. Each client selects a backoff time interval from the uniform distribution over $[0, CW]$. CW means *contention window* and is set as $2^k - 1$, where k is a positive integer (e.g., $CW = 15$, $CW = 31$). At first, $k = k_{\min}$, $CW = CW_{\min}$, k is increased by one when a client is involved in a collision, until CW reaches $CW_{\max}$. CW is reset to $CW_{\min}$ when the client successfully transmits a packet.

3.2 Modeling the Uplink Channel of a CSMA/CA-based MU-MIMO WLAN

In this section we propose an analytical model to compute the *saturation throughput* and the *mean access delay* of the uplink channel in a CSMA/CA-based MU-MIMO WLAN, under the saturation condition that each client has data to send all the time. For simplicity, only single-antenna clients are considered. Generalization to multi-antenna clients is our future work. As mentioned in the previous section, we focus on the uplink channel with no downlink transmission from AP, except ACKs.

[‡]Readers who are interested in the detailed decoding process can refer to (Tan et al., 2009), (Shen et al., 2012), and Section 8.3 in (Tse and Viswanath, 2005).

Our analysis is based on Bianchi's Markov chain model Bianchi (2000). However, Bianchi's analysis is proposed for the standard 802.11 DCF scheme, so we need to tailor it to accommodate the CSMA/CA-based MAC protocol that allows concurrent transmissions. This section consists of six subsections. In the first subsection we apply the discrete-time Markov chain model to compute the transmission probability τ of each client, which is derived as a function of the conditional collision probability p. The variable p is assumed to be constant for all the clients, and is computed in the second subsection. In the third subsection we calculate the transmission rates of the concurrent streams. The saturation throughput and mean access delay are formulated as functions of τ in the fourth and fifth subsections. An opportunistic transmission scheme is considered and modeled in the last subsection.

3.2.1 Transmission Probability

We first focus on the backoff behavior of a single client, say, Client C_1. Let $b(t)$ be the value of its backoff counter at time t, then $b(t)$ follows a stochastic process. An example of this process is shown in Fig. 3.3. During the contention period, $b(t)$ is reduced by one every slot time. When some client wins the contention, $b(t)$ stays in its value for a certain time interval. The length of this interval depends on who the contention winner is and whether the concurrent transmission opportunity remains. For example, if Client C_2 wins the channel and there is one more chance for concurrent transmissions, then $b(t)$ remains unchanged when Client C_2 transmits its PHY header and continues to decrease after that. Each time $b(t)$ reduces to zero, Client C_1 wins the contention and starts to transmit. After finishing transmission, Client C_1 draws a value from the uniform distribution over $[0, CW]$ and assigns it to $b(t)$.

By regarding $(CW, b(t))$ as the states of a backoff counter, we can describe the change of $(CW, b(t))$ as a bidimensional discrete-time Markov chain. The state transition probabilities are shown in Fig. 3.4, where p is the conditional probability that Client C_1 encounters a failed transmission when it has won the channel. Although the MAC protocol in a CSMA/CA-based

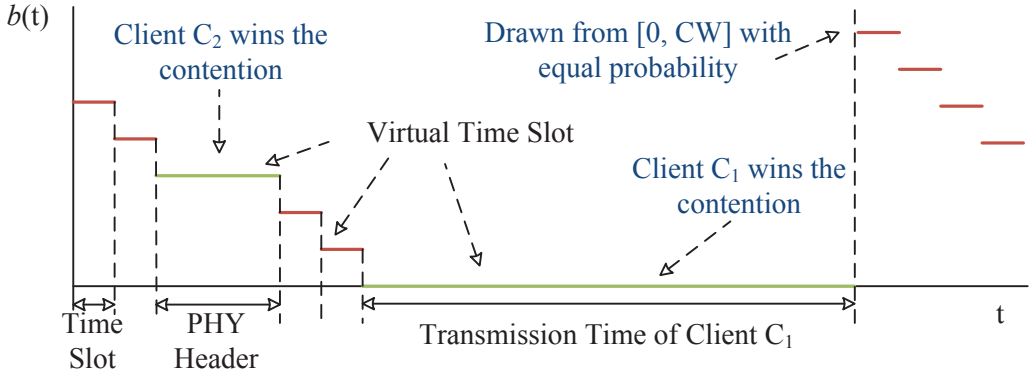


Figure 3.3: Example of the stochastic process of Client C_1's backoff counter.

MU-MIMO WLAN is different from that in a conventional 802.11 WLAN, the state transition behavior, i.e., the Markov chain, of a client's backoff counter in the two WLANs are identical. This is because the backoff mechanism used in the two WLANs are the same. In the Markov chain, let τ be the probability that $b(t) = 0$, i.e., Client C_1 is in transmission state, then Bianchi's result can be applied here (see (Bianchi, 2000) for details):

$$\tau = \frac{2(1-2p)}{(1-2p)(W+1) + pW(1-(2p)^m)}, \tag{3.1}$$

where $W = CW_{\min} + 1$ and $2^m W = CW_{\max} + 1$. Since clients are assumed to have packets to transmit at all times (i.e., the saturation condition), in the long term, all the clients share the same transmission probability τ and conditional collision probability p. Therefore, Eq. (3.1) holds true not only for Client C_1 but also for all the clients in the network.

Note that Eq. (3.1) is derived based on the transition process of the Markov chain without considering the real time elapse. Actually, how long a client stays in its current state before jumping to the next state is different for different states, as described at the beginning of this subsection. Here, we give it a general name *virtual time slot*, meaning the time interval between two consecutive states (see Fig. 3.3). Although the Markov models in a CSMA/CA-based MU-MIMO WLAN and a conventional 802.11 WLAN are identical, the length of virtual time slots in the two WLANs are different, which is a reason why Bianchi's method is not applicable to

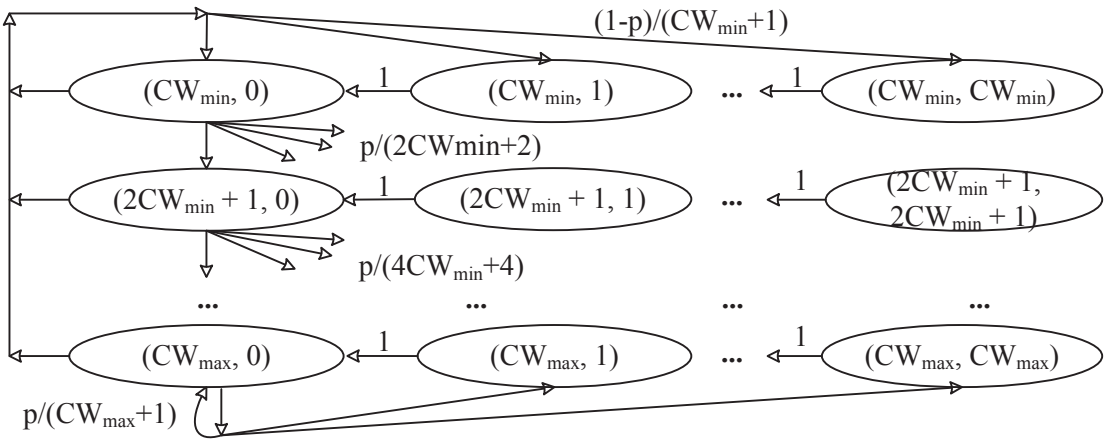


Figure 3.4: Discrete-time Markov chain model for the backoff counter proposed by Bianchi (2000).

the derivation of conditional collision probability and saturation throughput, as illustrated in the next two subsections.

3.2.2 Conditional Collision Probability

As defined in the previous subsection, the conditional collision probability p is the probability that a client encounters transmission failure given that it has won the channel. Let N be the total number of clients in a WLAN. In Bianchi's analysis (Bianchi, 2000), $p = 1 - (1 - \tau)^{N-1}$, corresponding to the probability that, in a virtual time slot, when a client (say, Client C_1) transmits, at least one of the remaining $N - 1$ clients transmits at the same time. However, in a CSMA/CA-based MU-MIMO WLAN, this result does not hold true, for the following two reasons. First, when Client C_1 transmits in a virtual time slot, the number of the remaining clients that can transmit at the same virtual time slot is unknown, since some of the clients (e.g., Client C_3 and Client C_5 in Fig. 3.2) may have already been involved in the ongoing transmissions. Second, when Client C_1 starts to transmit, collisions may happen not only at Client C_1 but also at clients that transmit concurrently with it. As shown in Fig. 3.6, although Client C_1 does not encounter collisions when it first wins the channel, it still fails to transmit its data because Client C_4 and Client C_5 collide with each other. Since Bianchi's result is not

applicable, we propose a new approach to compute p.

Define M as the maximum number of clients that can transmit concurrently and successfully in a CSMA/CA-based MU-MIMO WLAN with only N single-antenna clients, then

$$M = \min\{\text{the number of antennas at AP}, N\}. \tag{3.2}$$

We use the term *round* to denote the time interval spent by a transmission with M (or more than M if collision happens) concurrent clients. Note that it is possible that fewer than M clients transmit in a round, when no clients win the contention before the ongoing transmission ends. Here we ignore this probability, which consequently results in a limitation of the analytical model. This limitation will be discussed in Section 3.3. A transmission round can *succeed* (or *fail*), corresponding to whether the AP can perform successful decoding in that round (see Fig. 3.6). According to the definition of the conditional collision probability, p can be represented as

$$\begin{aligned} p &= P(\text{Client C}_i \text{ fails}|\text{Client C}_i \text{ transmits}) \\ &= P(r \text{ fails}|\text{Client C}_i \text{ transmits in round } r), \end{aligned} \tag{3.3}$$

where $i \in \{1, 2, \ldots, N\}$ denotes the client that we are interested in, and r is a randomly chosen round. Let $\mathcal{R}_s$ and $\mathcal{R}_f$ denote the sets of successful and failed rounds, respectively. Then $r \in \mathcal{R}_s$ (or $r \in \mathcal{R}_f$) means that r succeeds (or fails). Accordingly, $1 - p$ can be calculated as

$$\begin{aligned} 1 - p &= P(r \in \mathcal{R}_s|\text{Client C}_i \text{ transmits in round } r) \\ &= \frac{P(\text{Client C}_i \text{ transmits in } r \text{ and } r \in \mathcal{R}_s)}{P(\text{Client C}_i \text{ transmits in round } r)} \\ &= \frac{P(\text{Client C}_i \text{ transmits in } r|r \in \mathcal{R}_s)P(r \in \mathcal{R}_s)}{1 - P(\text{Client C}_i \text{ does not transmit in round } r)}. \end{aligned} \tag{3.4}$$

As discussed in the last paragraph of Section 3.1, the AP encounters a decoding failure as long as a collision happens. Note that many factors, e.g., fading, interference, collisions, can cause transmission failure, however, here we focus on investigating the effect of collisions. Therefore, $P(r \in \mathcal{R}_s)$ represents the probability that no clients transmit at the same time in a round. This probability certainly depends on the number of allowed concurrent transmissions in a round and the number of clients competing for those transmission opportunities, i.e., M and N. In the following derivation, we use $P_s(M, N)$ instead of $P(r \in \mathcal{R}_s)$ to highlight its dependence on M and N.

In a successful round, there are exactly M concurrent transmissions and hence M contention periods, so $P_s(M, N)$ can be computed as the probability that, at the end of each contention period, only one client wins the transmission opportunity. Let A_j ($j \in \{1, 2, \ldots, M\}$) be the event that exactly one client wins the j-th contention, i.e., no collision happens in the j-th concurrent transmission. $P_s(M, N)$ can then be represented as

$$P(A_1)P(A_2|A_1)\cdots P(A_M|A_1, A_2, \ldots, A_{M-1}). \tag{3.5}$$

Since τ is the probability that Client C_i transmits ($i \in \{1, 2, \ldots, N\}$) in a randomly chosen virtual time slot, and a virtual time slot is the same as a conventional time slot during the contention period, we can compute Eq. (3.5) as

$$\begin{aligned} P_s(M, N) =& \frac{N\tau(1-\tau)^{N-1}}{1-(1-\tau)^N} \frac{(N-1)\tau(1-\tau)^{N-2}}{1-(1-\tau)^{N-1}} \\ & \cdots \frac{(N-M+1)\tau(1-\tau)^{N-M}}{1-(1-\tau)^{N-M+1}}, \end{aligned} \tag{3.6}$$

where there are M terms multiplying together, and each term corresponds to a contention period. For the first term, the denominator $1-(1-\tau)^N$ denotes the probability that at least one of the N clients transmits in a time slot, while the numerator $N\tau(1-\tau)^{N-1}$ denotes the probability that exactly one of the N clients transmits in a time slot. Therefore, the first term

represents the probability that given a time slot where at least one client wins the contention[§], exactly one client transmits in that time slot. All the rest $M-1$ terms can be explained in the same way, except that they are computed under the collision-free condition of the previous contentions.

Recall that our aim is to compute the uplink throughput when every client has data to send all the time. Under this saturation condition, each client will have an equal probability to join a successful round. Consequently, the probability that Client C_i is among the M concurrent clients of a successful round is equal to the probability that i is among the M clients that are randomly picked from N clients, i.e.,

$$P(\text{Client } \mathrm{C}_i \text{ transmits in } r | r \in \mathcal{R}_s) = \frac{\binom{N-1}{M-1}}{\binom{N}{M}} = \frac{M}{N}. \tag{3.7}$$

To compute $P(\text{Client } \mathrm{C}_i \text{ does not transmit in round } r)$, note that if Client C_i is sure to be silent in a round, then the whole network will act as if Client C_i were not there, i.e., there were $N-1$ clients. Therefore, we have

$$P(r \in \mathcal{R}_s | \text{Client } \mathrm{C}_i \text{ does not transmit in } r) = P_s(M', N-1), \tag{3.8}$$

where M' is defined as the maximum number of allowed concurrent transmissions in a network with the same AP but with $N-1$ single-antenna clients, i.e.,

$$M' = \min\{M, N-1\}. \tag{3.9}$$

[§]This condition restricts the time slot to be at the end of a contention period.

Based on Bayes' Theorem, we get

$$\begin{aligned} & P(\text{Client C}_i \text{ does not transmit in round } r) \\ &= \frac{P(\text{Client C}_i \text{ does not transmit in } r | r \in \mathcal{R}_s) P(r \in \mathcal{R}_s)}{P(r \in \mathcal{R}_s | \text{Client C}_i \text{ does not transmit in round } r)} \\ &= \frac{(1 - \frac{M}{N}) P_s(M, N)}{P_s(M', N-1)}. \end{aligned} \tag{3.10}$$

Substituting Eqs. (3.7) and (3.10) into Eq. (3.4) gives

$$p = 1 - \frac{\frac{M}{N} P_s(M, N)}{1 - \frac{(1-\frac{M}{N}) P_s(M,N)}{P_s(M', N-1)}}, \tag{3.11}$$

where $P_s(M, N)$ is calculated by Eq. (3.6). Now we have two non-linear equations of p and τ, i.e., Eqs. (3.1) and (3.11). The value of τ can be determined by solving these equations.

3.2.3 Transmission Rates

The transmission rates of concurrently transmitting clients depend on their channels and how they interact in the decoding procedure. For the CSMA/CA-based MU-MIMO WLAN, the AP uses *zero-forcing with successive interference cancellation* (ZF-SIC)[¶] to decode the M independent data streams (Tse and Viswanath, 2005). The decoding procedure contains M stages, as shown in Fig. 3.5. In the k-th stage, the AP decorrelates and decodes the $(M+1-k)$-th stream, and then subtracts off the decoded stream from the received vector so that in the $(k+1)$-th stage, there are $M-k$ remaining streams. According to this decoding procedure, when the AP decodes the k-th stream, only the interfering streams that join before the k-th stream need to be considered, since streams that join after it have already been removed. This property allows that the k-th concurrent client transmits at a rate that is determined by the channels of itself and the previous $k-1$ contention winners.

[¶]The zero-forcing operation is also known as decorrelator or interference nulling.

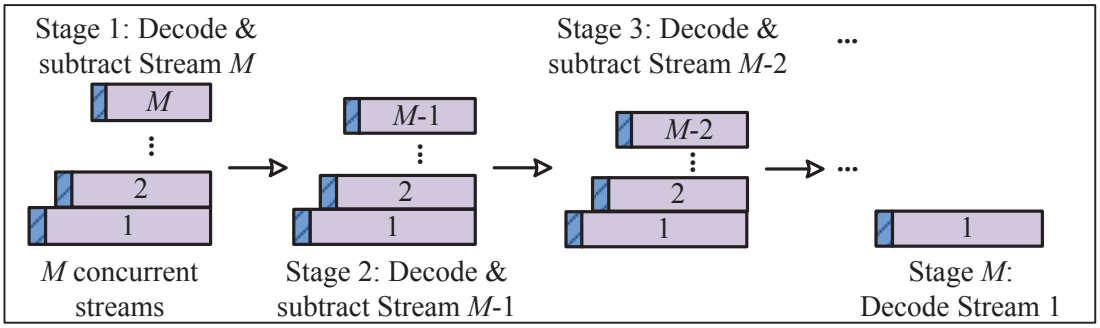


Figure 3.5: The decoding procedure of ZF-SIC.

To illustrate how ZF-SIC works, let us consider a network with n-antenna AP. After decoding and removing $M-k$ streams, the remaining received vector at the AP for a symbol time can be written as

$$\mathbf{y} = \sum_{1}^{k} \mathbf{h}_i x_i + \mathbf{w}, \tag{3.12}$$

where $\mathbf{y}$, $\mathbf{h}_i$, and $\mathbf{w}$ are $n \times 1$ vectors. The i-th contention winner transmits a data symbol x_i through a channel $\mathbf{h}_i$. The additive white noise vector is denoted by $\mathbf{w}$, which follows a circular symmetric distribution $\mathcal{CN}(0, N_0\mathbf{I}_n)$. We assume that the data streams, the channel vectors, and the noise vectors are all independent. To decorrelate x_k, the AP projects the received $\mathbf{y}$ onto the *null space* of the matrix $[\mathbf{h}_1\mathbf{h}_2 \ldots \mathbf{h}_{k-1}]^T$, where $[\cdot]^T$ is the transpose operator. Under the assumption of independent channel vectors, the dimension of this null space is

$$d_k = n - k + 1. \tag{3.13}$$

We can construct a $d_k \times n$ matrix $\mathbf{Q}_k$, with its rows representing an orthogonal basis of this null space. Then the projection operation can be expressed as multiplying $\mathbf{Q}_k$ and $\mathbf{y}$, which yields

$$\mathbf{Q}_k\mathbf{y} = \mathbf{Q}_k\mathbf{h}_k x_k + \mathbf{Q}_k\mathbf{w}. \tag{3.14}$$

Accordingly, the k-th stream can be decoded and then removed from Eq. 3.12. The AP will continue to decode the $(k-1)$-th stream following a similar procedure.

To characterize the resulting rate differentiation, note that in Eq. 3.14, $\mathbf{Q}_k\mathbf{w}$ is still white noise, distributed as $\mathcal{CN}(0, N_0\mathbf{I}_{d_k})$. Let $P = E[||x||^2]$ be the transmission power of each client, and B be the channel bandwidth, then the maximum data rate achieved by the k-th concurrent client is

$$R_k = B\log_2(1 + P||\mathbf{Q}_k\mathbf{h}_k||^2/N_0), k = 1, \ldots, M. \tag{3.15}$$

We consider a time-varying i.i.d. Rayleigh fading channel model, with coherence time as a transmission round, which means that each client's channel remains unchanged during a round, but is independent across successive transmission rounds. Let $\mathbf{h}_i \sim \mathcal{CN}(0, \mathbf{I}_n)$, we can now calculate the average data rates of concurrent date streams. Since $\mathbf{Q}_k\mathbf{Q}_k^* = \mathbf{I}_{d_k}$, the distribution of $\mathbf{Q}_k\mathbf{h}_k$ is $\mathcal{CN}(0, \mathbf{I}_{d_k})$. Thus $||\mathbf{Q}_k\mathbf{h}_k||^2$ is distributed as $\chi^2_{2d_k}$, i.e., it is Chi-squared distributed with $2d_k$ degrees of freedom. This result also holds true for the first contention winner, because when $k = 1$, $||\mathbf{h}_1||^2$ follows the distribution χ^2_{2n}. Accordingly, the average transmission rate of the k-th concurrent data stream can be computed as

$$E[R_k] = \int_0^{+\infty} B\log_2(1 + Px/N_0) f_{\chi^2_{2d_k}}(x)\mathrm{d}x, \tag{3.16}$$

where $f_{\chi^2_{2d_k}}(\cdot)$ denotes the PDF for $\chi^2_{2d_k}$ distribution with $k = 1, 2, \ldots, M$.

3.2.4 Saturation Throughput

Saturation throughput refers to the network throughput when clients always have data to transmit. To formulate it, we first introduce a concept called *virtual transmission time*, as defined in (Calì et al., 2000). It represents the time elapse between two consecutive successful rounds (see Fig. 3.6). Let N_{fail} be the number of failed transmission rounds during the virtual transmission time and N_{idle} be the number of idle time slots between two consecutive rounds. Let T_j denote the transmission time of the j-th concurrent client in a round, where $j \in \{1, 2, \ldots, M\}$.

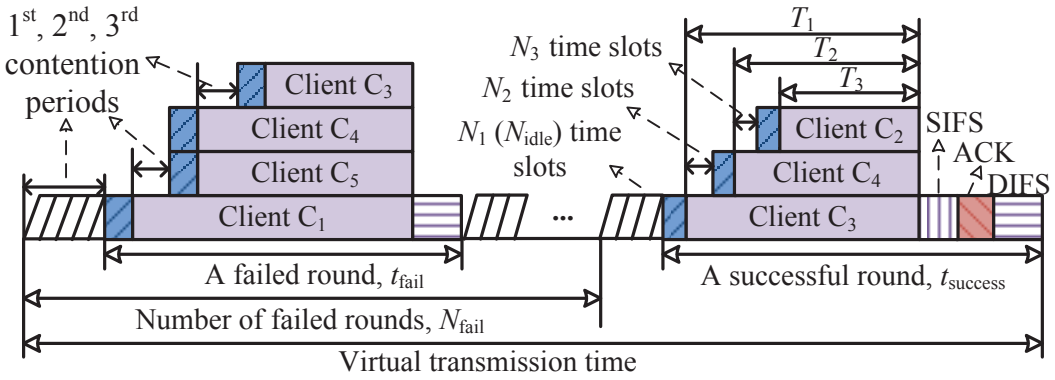


Figure 3.6: Illustration of the transmission structure in a CSMA/CA-based MU-MIMO WLAN.

Then N_{fail}, N_{idle}, and T_j are all random variables. Together with the transmission rates in Section 3.2.3, we can express the saturation throughput as

$$\rho = \frac{\sum_{j=1}^{M} E[R_j]E[T_j]}{E[N_{\text{fail}}]t_{\text{fail}} + t_{\text{success}} + (E[N_{\text{fail}}]+1)E[N_{\text{idle}}]t_{\text{slot}}}, \tag{3.17}$$

where t_{fail}, t_{success}, and t_{slot} are the time elapse of a failed round, a successful round and an idle time slot, respectively.

According to the definition, N_{fail} follows a geometric distribution with parameter $1 - P_s(M,N)$, which is the probability that a randomly chosen round fails, i.e.,

$$P(N_{\text{fail}} = k) = (1 - P_s(M,N))^k P_s(N,M), \tag{3.18}$$

where $k = 0, 1, 2, \ldots$, then the average number of failed rounds in a virtual transmission time is

$$E[N_{\text{fail}}] = \frac{1 - P_s(M,N)}{P_s(M,N)}. \tag{3.19}$$

In a successful transmission round, because the transmitting clients are forced to end simultaneously, the time they spent on data transmission can be calculated recursively as

$$T_{j+1} = T_j - t_{\text{PHY}} - N_{j+1}t_{\text{slot}}, \tag{3.20}$$

where t_{PHY} is the time needed to transmit a PHY header, and N_j ($j \in \{1, 2, \ldots, M\}$) is the number of idle time slots elapsing in the j-th contention period (see Fig. 3.6). Since N_{idle} is the number of idle time slots between two consecutive rounds, it actually equals to N_1. During the first contention period, there are N clients competing for the transmission opportunity, and each client transmits in a time slot with probability τ. Thus N_1 follows a geometric distribution with the parameter $(1-\tau)^N$, i.e.,

$$P(N_1 = k) = ((1-\tau)^N)^k(1-(1-\tau)^N). \tag{3.21}$$

Accordingly, we have

$$E[N_{\text{idle}}] = E[N_1] = \frac{(1-\tau)^N}{1-(1-\tau)^N}. \tag{3.22}$$

During the j-th contention period, where $j \geq 2$, there are $N-j+1$ clients competing for the jth concurrent transmission opportunity. However, all the contending clients must have nonzero backoff counters, for otherwise they would have been involved in the ongoing transmission and can no longer join the j-th contention. In other words, no client is able to win the j-th concurrent transmission opportunity in the first time slot of the contention period, i.e., $N_j \geq 1$ for $j \geq 2$. Therefore, the distribution of N_j is

$$P(N_j = k) = ((1-\tau)^{N-j+1})^{k-1}(1-(1-\tau)^{N-j+1}), \tag{3.23}$$

where $k \geq 1$ and $2 \leq j \leq M$. Then its expectation becomes

$$E[N_j] = \frac{1}{1-(1-\tau)^{N-j+1}}. \tag{3.24}$$

Assuming that $E[T_1]$ is known, we can then use Eqs. (3.20) and (3.24) to compute the expec-

tation of other T_js in a recursive manner, i.e.,

$$E[T_{j+1}] = E[T_j] - t_{\text{PHY}} - \frac{1}{1-(1-\tau)^{N-j}} t_{\text{slot}}. \tag{3.25}$$

Although each time different clients are engaged in a transmission round, how long a round lasts depends only on the data time of the first client (Fig. 3.6). Therefore,

$$t_{\text{success}} = t_{\text{PHY}} + E[T_1] + \text{SIFS} + \text{ACK} + \text{DIFS}, \tag{3.26}$$

$$t_{\text{fail}} = t_{\text{PHY}} + E[T_1] + \text{DIFS}. \tag{3.27}$$

The propagation delay is normally too small compared with the total frame transmission time, so it is omitted here. We also ignore ACK timeout in the theoretical analysis, and the resulting limitation is discussed in Section 3.3.3.

Substituting Eqs. (3.6), (3.16), (3.19), (3.22), and (3.25)–(3.27) into Eq. (3.17), with values of N, M, B, P/N_0, $E[T_1]$ as well as the value of τ calculated in the last subsection, we are now able to compute the saturation throughput ρ of the uplink channel in a CSMA/CA-based MU-MIMO WLAN.

3.2.5 Access Delay

In this subsection the mean access delay is determined for each client. The access delay is defined as the time experienced by a packet, from it first becoming the head of the queue to the time it is transmitted successfully. Under the saturation condition, all the clients have the same mean access delay. Let d denote the mean access delay of a given client, say, Client C_1. Then d refers to the average time between Client C_1's consecutively transmitted packets. According to Eq. (3.7), we know that Client C_1 needs to wait an average of $1/P(\text{Client C}_1 \text{ transmits in } r | r \in \mathcal{R}_s)$ successful rounds to get itself join a successful round. Based on the concept of virtual

transmission time, we can then calculate the mean access delay as

$$d = \frac{E[\text{Virtual transmission time}]}{P(\text{Client C}_1 \text{ transmits in } r | r \in \mathcal{R}_s)}. \tag{3.28}$$

According to Eqs. (3.7) and (3.17), we can get

$$d = \frac{E[N_{\text{fail}}]t_{\text{fail}} + t_{\text{success}} + (E[N_{\text{fail}}] + 1)E[N_{\text{idle}}]t_{\text{slot}}}{M/N}. \tag{3.29}$$

Note that Client C_1's packets are of varying sizes, which depend on the dimension Client C_1 occupies in each successful round. Therefore, d corresponds to the transmission of a packet with an average size $(\sum_{j=1}^{M} E[R_j]E[T_j])/M$.

3.2.6 Opportunistic Transmission

According to Section 3.2.3, the k-th stream experiences interference from the previous $k-1$ concurrent clients. After projecting $\mathbf{h}_k$ onto the subspace orthogonal to the one spanned by $\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_{k-1}$, the k-th stream achieves an SNR of $P||\mathbf{Q}_k\mathbf{h}_k||^2/N_0$. When the transmission power is given, the concurrent transmission rate is fully determined by $||\mathbf{Q}_k\mathbf{h}_k||$, which represents the effect of inter-stream interference. The value of $||\mathbf{Q}_k\mathbf{h}_k||$ depends on the interaction of the channels of the k concurrent streams, and is always less than or equal to $||\mathbf{h}_k||$, where equality is only achieved when $\mathbf{h}_k$ is orthogonal to the span of $\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_{k-1}$.

Considering a network where the clients fade independently, we are then interested in an opportunistic transmission scheme: during the j-th contention period ($j \geq 2$), if clients with large concurrent rates, i.e., large values of $||\mathbf{Q}_j\mathbf{h}_j||$, are given high probabilities of winning the contention, then the total network throughput can be improved. In this subsection we will model the saturation throughput and mean access delay of a simple distributed opportunistic scheme for a 2-antenna AP scenario. Our modeling method can be easily generalized to scenarios with more antennas at the AP, but considering a simple two antenna case is enough to reveal the

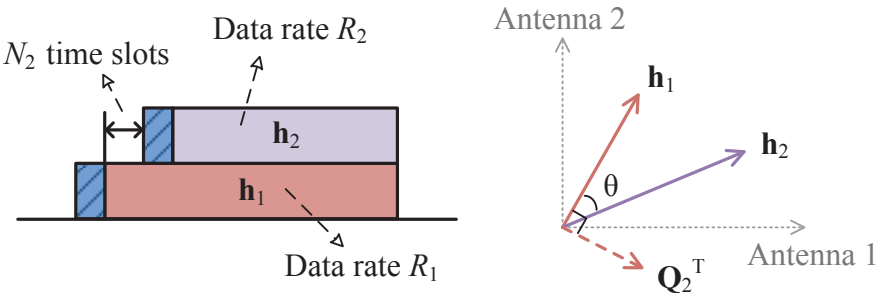


Figure 3.7: A geometric interpretation of $\mathbf{h}_1$, $\mathbf{h}_2$, and $\mathbf{Q}_2^T$.

influence of the threshold value on the network performance, as indicated in Section 3.4.4.

In Section 3.2.3, we have shown that the maximum data rates of two concurrent streams are

$$\begin{aligned} R_1 &= B\log_2(1+P||\mathbf{h}_1||^2/N_0), \\ R_2 &= B\log_2(1+P||\mathbf{Q}_2\mathbf{h}_2||^2/N_0), \end{aligned} \tag{3.30}$$

where $\mathbf{Q}_2$ is a 1×2 unit vector that is orthogonal to $\mathbf{h}_1$. A geometric interpretation of $\mathbf{h}_1$, $\mathbf{h}_2$, and $\mathbf{Q}_2^T$ is shown in Fig. 3.7. Similar to Section 3.2.3, we consider a network where the clients experience i.i.d. Rayleigh fading. The channel of each client remains unchanged during a round time, but is independently variable between successive rounds. Assuming that $\mathbf{h}_i \sim \mathcal{CN}(0,\mathbf{I}_2)$ for $i=1,2$, then $||\mathbf{Q}_2\mathbf{h}_2||^2$ follows a Chi-squared distribution with 2 degrees of freedom.

The opportunistic MAC protocol that we will model works as follows. In a transmission round, when a client wins the first contention, each of the rest clients would then calculate the value of $||\mathbf{Q}_2\mathbf{h}_2||^2$ by assuming that itself is the second contention winner. To do so, the rest clients must know their own channels as well as that of the first contention winner, which is possible according to (Shen et al., 2012). A client could learn its own channel through the reverse channel. Besides, the first winner can put its own channel information in its PLCP header so that other clients would be able to know it. After obtaining the value of $||\mathbf{Q}_2\mathbf{h}_2||^2$,

only the clients satisfying

$$||\mathbf{Q}_2\mathbf{h}_2||^2 \geq \mathrm{T} \tag{3.31}$$

is allowed to join the second contention, while others will defer their access until the next round. Here T acts as a threshold.

To formulate the opportunistic MAC protocol, we can follow the same procedure as developed in the previous subsections. Note that Eq. (3.1) still holds true here because the opportunistic transmission scheme does not change the backoff mechanism of the CSMA/CA-based MU-MIMO WLANs.

To derive the conditional collision probability p, we need to calculate $P_s(M, N)$ first (with $M = 2$ and $N > 2$), which represents the probability that, in a network with N clients, a randomly chosen round is successful. Let N_{join} denote the number of clients that can contend the second transmission opportunity in a successful round. Then N_{join} is a random variable because of the randomness of clients' channels. Denoted by p_{join} the probability that, given the first contention winner, i.e., given $\mathbf{Q}_2$, a randomly chosen client satisfies Eq. (3.31), i.e.,

$$p_{\text{join}}(\mathrm{T}) = P(||\mathbf{Q}_2\mathbf{h_2}||^2 \geq \mathrm{T}|\mathbf{Q}_2). \tag{3.32}$$

Then N_{join} follows a binomial distribution as

$$P(N_{\text{join}} = k) = \binom{N-1}{k} p_{\text{join}}^k (1 - p_{\text{join}})^{N-1-k}, \tag{3.33}$$

where $k = 0, 1, 2..., N-1$. To compute p_{join}, let $\theta \in [0, \pi]$ be the angle between $\mathbf{h}_1$ and $\mathbf{h}_2$ in the antenna space, as shown in Fig. 3.7. According to Shen et al. (2012), we have

$$||\mathbf{Q}_2\mathbf{h_2}||^2 = ||\mathbf{h_2}||^2 \sin^2(\theta). \tag{3.34}$$

We assume that $||\mathbf{h_2}||^2$ and θ are independent random variables. We already know that $||\mathbf{h_2}||^2 \sim$

χ^2_4. Assuming that θ is uniformly distributed in the interval $[0, \pi]$, Eq. (3.32) can then be computed as

$$\begin{aligned} p_{\text{join}}(\mathrm{T}) &= 1 - P(||\mathbf{Q}_2\mathbf{h_2}||^2 < \mathrm{T}|\mathbf{Q}_2) \\ &= 1 - P(||\mathbf{h_2}||^2 \sin^2(\theta) < \mathrm{T}) \\ &= 1 - \int_0^\infty P(||\mathbf{h_2}||^2 = x) P(\sin^2(\theta) < \frac{\mathrm{T}}{x}) \mathrm{d}x \\ &= 1 - \int_0^\infty f_{\chi^2_4}(x) \frac{\arcsin(\sqrt{\mathrm{T}/x})}{\pi/2} \mathrm{d}x, \end{aligned} \tag{3.35}$$

where $f_{\chi^2_4}(\cdot)$ denotes the PDF of a Chi-squared distribution with 4 degrees of freedom. Based on Eqs. (3.35) and (3.33), we can now calculate the probability of a successful round as the following expectation

$$P_s(2, N) = E\left[\frac{N\tau(1-\tau)^{N-1}}{1-(1-\tau)^N} \frac{N_{\text{join}}\tau(1-\tau)^{N_{\text{join}}-1}}{1-(1-\tau)^{N_{\text{join}}}}\right], \tag{3.36}$$

where $N > 2$ and $N_{\text{join}} = 0, 1, 2..., N-1$. In the case of $N_{\text{join}} = 0$, when a successful round contains only one stream, $P_s(M, N)$ reduces to the probability that only one client wins the first contention. Note that in Eq. (3.33), when p_{join} approaches 1, i.e., when less restriction is given on who can join the second contention, the value of N_{join} would be approaching $N-1$, and the above equation becomes the same as the one defined in Eq. (3.6).

Eq. (3.8) still hold true here because it is formulated without considering the detailed MAC protocol. Eq. (3.7) calculates the probability that Client C_i shows up in a randomly chosen successful round. Under the opportunistic transmission scheme, in each round the number of clients that can join the second contention is limited. However, since the clients' channels are assumed to be independent for different rounds, in the long run, each client has an equal probability to join a successful round and is able to transmit concurrently with any other clients.

Considering the special case of $N_{\text{join}} = 0$, we denote p_0 the probability that in a randomly chosen successful round, no client contends for the second concurrent transmission opportunity, i.e.,

$$p_0 = P(N_{\text{join}} = 0 | r \in \mathcal{R}_s). \tag{3.37}$$

Then Eq. (3.7) is changed to

$$P(\text{Client C}_i \text{ transmits in } r | r \in \mathcal{R}_s) = \frac{2}{N}(1 - p_0) + \frac{1}{N}p_0. \tag{3.38}$$

Using Eqs. (3.33) and (3.36), p_0 can be calculated as

$$p_0 = \frac{P(r \in \mathcal{R}_s \text{ and } N_{\text{join}} = 0)}{P(r \in \mathcal{R}_s)} = \frac{\frac{N\tau(1-\tau)^{N-1}}{1-(1-\tau)^N}(1 - p_{\text{join}})^{N-1}}{P_s(2, N)}. \tag{3.39}$$

The average transmission rate of the first contention winner is the same as that of the original MAC protocol, i.e.,

$$E[R_1] = \int_0^{+\infty} B \log_2(1 + Px/N_0) f_{\chi_4^2}(x) \mathrm{d}x, \tag{3.40}$$

where $f_{\chi_4^2}(\cdot)$ denotes the PDF of a Chi-squared distribution with 4 degrees of freedom. To compute the the second stream's average rate, note that according to the opportunistic transmission scheme, the value of $||\mathbf{Q}_2\mathbf{h}_2||^2$ falls in the range of $[\mathrm{T}, \infty)$. Since $||\mathbf{Q}_2\mathbf{h}_2||^2 \sim \chi_2^2$, we have

$$E[R_2] = \frac{\int_{\mathrm{T}}^{\infty} B \log_2(1 + Px/N_0) f_{\chi_2^2}(x) \mathrm{d}x}{\int_{\mathrm{T}}^{\infty} f_{\chi_2^2}(x) \mathrm{d}x}, \tag{3.41}$$

where $f_{\chi_2^2}(\cdot)$ denotes the PDF of a Chi-squared distribution with 2 degrees of freedom.

To compute the saturation throughput, note that since the number of clients competing for the second transmission opportunity is reduced to N_{join}, which is a random variable with

distribution given by Eq. (3.33), we can then change Eq. (3.24) to

$$E[N_2] = E\left[\frac{1}{1-(1-\tau)^{N_{\text{join}}}}\right]. \tag{3.42}$$

Accordingly, the average transmission time of the second contention winner, i.e., Eq. (3.25), becomes

$$E[T_2] = E[T_1] - t_{\text{PHY}} - E\left[\frac{1}{1-(1-\tau)^{N_{\text{join}}}}\right] t_{\text{slot}}. \tag{3.43}$$

Using the above results and following the same procedure in the previous subsections of Section 3.2, we are now able to calculate the saturation throughput and mean access delay of the opportunistic transmission scheme.

3.3 Model Validation

In this section comparisons between the analytical and simulation results are presented to validate our previous analysis. It includes three subsections. In the first and second subsections examples are shown that our analytical model can closely predict the uplink throughput and mean access delay of a wireless LAN. Two MAC protocols are simulated: a CSMA/CA-based MU-MIMO MAC and its opportunistic variation. In the third subsection we discuss about the main limitations of our analytical model.

3.3.1 CSMA/CA-based MU-MIMO WLANs

We use MATLAB to simulate the uplink channel of a CSMA/CA-based MU-MIMO WLAN. Our event-driven simulation program contains all major components of the MAC protocol, e.g., contention, PHY header, ACK, ACK timeout and the interframe spaces. The program also simulates the PHY layer as described in Section 3.2.3. Every client is assigned an $n \times 1$ channel vector, where n is the number of antennas at the AP. Each component of the channel vector

Table 3.1: Network parameters used to obtain numerical values.

Parameter	Value
t_{slot}	9 μs
t_{PHY}	20 μs
SIFS	16 μs
DIFS	34 μs
ACK	39 μs
ACK timeout	70 μs
$E[T_1]$	2000 μs
B	20 MHz
P/N_0	10 dB
$CW_{\min}$	127
$CW_{\max}$	1023

is an i.i.d. $\mathcal{CN}(0,1)$ random variable. The channel vectors are generated at the beginning of every transmission round and remains unchanged during a round time. The channel vectors of a client for different rounds are independent.

The network parameters used to obtain numerical values for both analytical model and simulation program are outlined in Table 3.1, where the upper half values are defined by 802.11 standards for OFDM PHY layer with 20 MHz channel spacing, and the lower half values are either calculated or selected according to the standards (*IEEE Std. 802.11*, 2012). Note that our analytical model is derived using the mean frame transmission time of the first transmitter in a round, i.e., $E[T_1]$, with no restrictions on the probability distribution of T_1. For simplicity, we only consider the case of constant T_1 in the simulation program. Besides, we set ACK timeout to be 70 μs in the simulation program, a value that is long enough to cover an SIFS and ACK transmission. The ACK timeout values can be set to other values. As shown in Table 3.2, different ACK timeout values cause relatively small difference in performance. Unless otherwise specified, the numerical values obtained in the rest of this chapter are all based on the network parameters listed in Table 3.1.

In Fig. 3.8 we plot the saturation throughput and mean access delay using both simulation (symbols) and analytical (lines) results. Different network configurations are considered: the

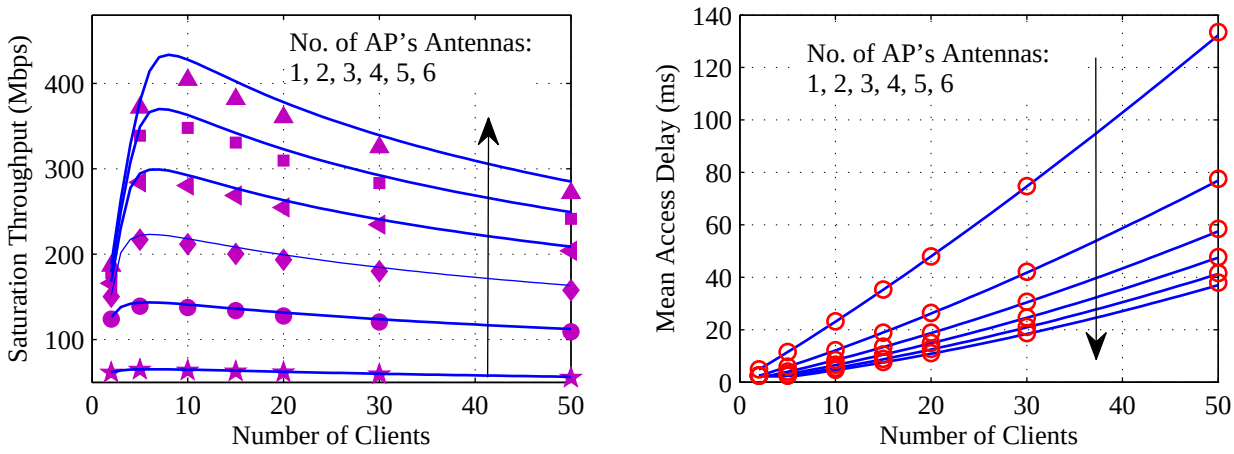


Figure 3.8: Saturation throughput (left) and mean access delay (right) for different network configurations: simulation (symbols) versus analysis (lines).

number of antennas at the AP varies from 1 to 6 and the number of clients ranges from 2 to 50. The figure indicates two thighs: 1) Our analytical model provides a close approximation of the saturation throughput and mean access delay in a CSMA/CA-based MU-MIMO WLAN; 2) The accuracy of the saturation throughput model degrades as the number of antennas at the AP increases, the reason for which is discussed in Section 3.3.3.

3.3.2 Opportunistic Transmission Scheme

We modify the simulation program of the previous subsection to characterize the opportunistic transmission scheme of Section 3.2.6. A network with a 2-antenna AP is considered. Every client has a 2×1 channel vector, with each component being an i.i.d. $\mathcal{CN}(0,1)$ random variable. The channel vectors remain unchanged for a round time and are independent between successive rounds. When a client wins the first contention, every other client calculates its concurrent transmission rate and check whether its concurrent rate is less than the threshold $B\log_2(1+P\mathrm{T}/N_0)$. If so, the corresponding client stops decreasing its backoff counter, i.e., defers its access to the channel, until the current round ends.

The saturation throughput and mean access delay are plotted in Fig. 3.9. Simulation (sym-

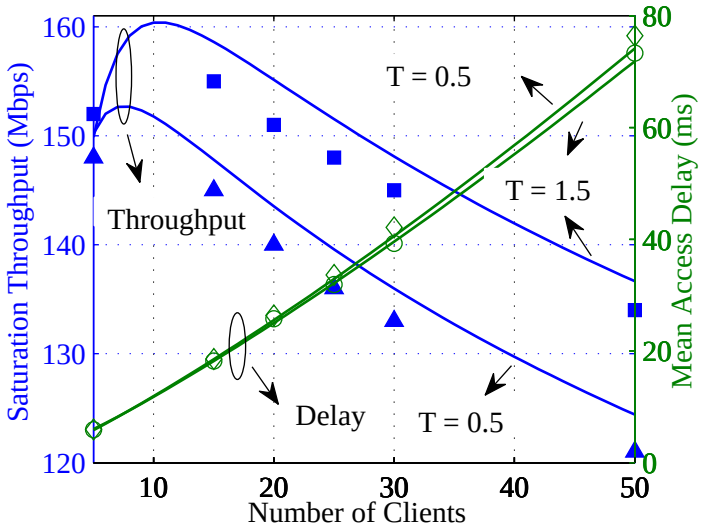


Figure 3.9: Saturation throughput and mean access delay of the opportunistic transmission scheme: simulation (symbols) versus analysis (lines).

bols) and analytical (lines) results are compared for networks with sizes varying from 5 to 50. The threshold T is set as 0.5 and 1.5. Note that the y-axis for saturation throughput ranges from 120 to 160 Mbps. Although differences exist between the simulation and analytical saturation throughput, the error percentage is less than 4%. Fig. 3.9 indicates that our analytical model can closely estimate the network performance of the opportunistic transmission scheme.

3.3.3 Limitations and Discussions

In this subsection we would discuss the limitations of our analytical model. Possible reasons are provided to explain why the accuracy of our model varies with different parameters.

Rounds with Less than M Concurrent Transmissions

As shown in Fig. 3.10, the accuracy of our saturation throughput model varies with respect to three parameters: frame transmission time, the number of antennas at the AP, and the number of total clients. The error percentage between simulation and analytical result is calculated as $|\rho_{\text{simulation}} - \rho_{\text{analytical}}|/\rho_{\text{simulation}}$. Comparing scenarios (a) and (b), scenarios (a) and (c) would

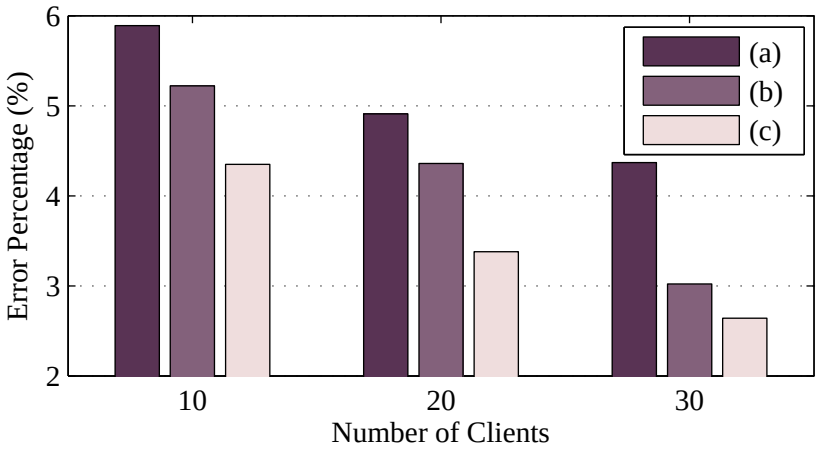


Figure 3.10: Error that occurs when the analytical model is used to estimate the saturation throughput in different scenarios: (a) $M = 6$ and $E[T_1] = 2000$ μs; (b) $M = 6$ and $E[T_1] = 4000$ μs; (c) $M = 4$ and $E[T_1] = 2000$ μs.

reveal that our model is more accurate in the case of long frame transmission time, and small AP's antennas. Besides, the error percentage decreases as the number of clients grows. This fluctuation of accuracy can be explained as follows. Our analytical model does not consider the situation that there may be less than M concurrent transmissions in a round, which happens when none of contending clients win the channel before the ongoing transmission ends. During the derivation of the analytical model (e.g., Eq. (3.6)), we simply assume that in every round, there are M (or more than M if collision happens) concurrent transmissions. This assumption holds true with a high probability when the frame transmission time is long, the number of antennas at the AP is small, and the number of total clients is large, which explains why the accuracy of our model is high under these situations.

To make the above explanation more convincing, we change t_{slot} from 9 μs to 1 μs and simulate the CSMA/CA-based MU-MIMO WLANs[‖]. The contention window sizes are set as $CW_{\min} = 511$ and $CW_{\max} = 1023$. Note that in Section 3.3.1 we set $CW_{\min} = 127$ and $CW_{\max} = 1023$. A larger $CW_{\min}$ is used here because we want to reduce the influence of backoff freezing operation, as explained in Section 3.3.3. Except t_{slot} and $CW_{\min}$, all the network

[‖]Please note that this change is only for illustrative purpose; it does not imply the existence of such an implementation in standard wireless networks.

parameters are the same as those in Table 3.1. Comparisons between the simulation (symbols) and analytical (lines) results are presented in Fig. 3.11, with M ranges from 7 to 20. This figure indicates that our model is extremely accurate even when the AP has a large number of antennas. This is because in this case a successful round is ensured to have M concurrent clients. If there is a successful round with only $M-1$ concurrent clients, then all the other clients must have a backoff time longer than $E[T_1]-(M-2)t_{\text{PHY}}$. For $E[T_1]=2000$ μs, $t_{\text{PHY}}=20$ μs, and $M=20$, this value is equal to 1640 μs. However, a client's backoff time is always less than $CW_{\max}t_{\text{slot}}=1023$ μs. Since 1023 μs is smaller than 1640 μs, it is impossible for a round to have less than M concurrent transmissions.

As demonstrated by Fig. 3.11, our model can accurately characterize a CSMA/CA-based MU-MIMO WLAN when there are always M concurrent clients in a successful round. This corresponds to the situation when all the dimensions at the AP are utilized by the concurrent streams. When the number of concurrent streams are less than the maximum number allowed, i.e., when the AP's antennas are underutilized, the achieved saturation throughput would be less than the maximum saturation throughput that can be achieved by the same AP. In other words, given the number of antennas at the AP, the maximum network throughput is achieved when all the AP's degrees of freedom are exploited by the concurrent streams, which corresponds to a situation that is characterized by our analytical model.

Backoff Freezing Operation

In (Bianchi and Tinnirello, 2005), Bianchi improves his analytical model by taking into account the *backoff freezing operation* in the 802.11 DCF standard. This operation happens after each successful transmission round. To simplify the description, let set A denote the clients who are involved in a randomly chosen successful round and set $\bar{A}$ be the rest clients. Note that the backoff counters of the clients in $\bar{A}$ are nonzero, for otherwise, they would have been involved in the transmission round. After the successful transmission, a contention period starts, and all

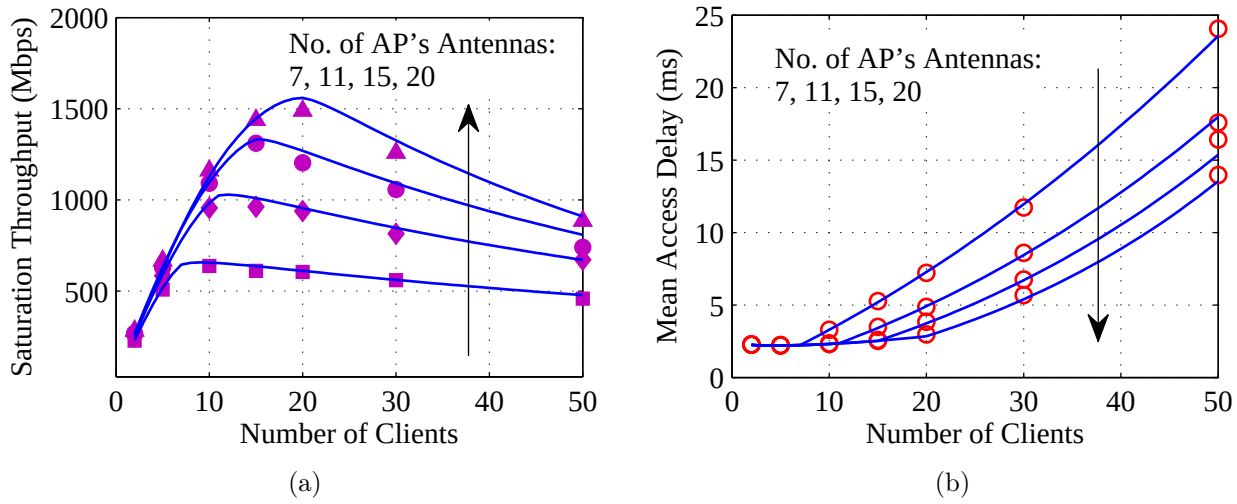


Figure 3.11: Comparisons between the simulation (symbols) and analytical (lines) results when $t_{\text{slot}} = 1$ μs and $CW_{\min} = 511$. Saturation throughput is shown in the left and mean access delay is shown in the right.

the clients compete for the transmission opportunity in the new round. Since all the clients in set A have zero backoff time, they will randomly choose a new backoff time from the interval [0, $CW_{\min}$] before the start of the contention period. If one client selects 0 as the new backoff time, it will win the channel automatically. However, no clients in $\bar{A}$ are able to win the channel at the first time slot of the contention period, because of their nonzero backoff counters. Therefore, at the beginning of the contention period after a successful transmission, the system does not behave as what is assumed when deriving the analytical model, i.e., every client has an equal probability τ of winning the transmitting opportunity.

It is too complex to consider the backoff freezing operation in the analysis of a CSMA/CA-based MU-MIMO WLAN. Fortunately, since the probability of selecting 0 as the new backoff time is $1/(CW_{\min}+1)$, the impact of the backoff freezing operation will reduce if $CW_{\min}$ is large. In Fig. 3.12 we plot the estimation error for different $CW_{\min}$s under two conditions: fixed m and fixed $CW_{\max}$. Other parameters are set as: $M = 5, N = 10, E[Data_1] = 2000$ μs. In both cases, the error between simulation and analytical results decreases when $CW_{\min}$ increases, as expected.

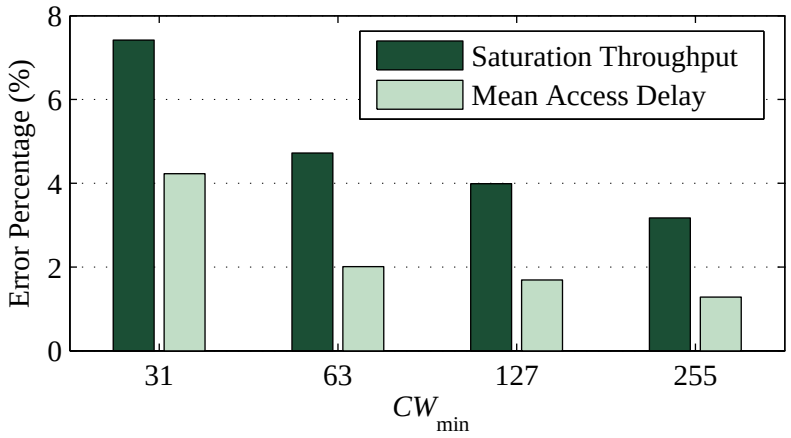


Figure 3.12: The error percentage between the analytical and simulation results for different $CW_{\min}$s, with $M = 5$, $N = 10$, and $E[T_1] = 2000$ μs.

Similar situation happens after a failed transmission round. Let A and $\bar{A}$ denote the set of transmitting clients and silent clients in a failed round, respectively, then clients in the two sets will not join the next round's contention at the same time. This is because clients in A cannot access the channel until the end of ACK timeout, while clients in $\bar{A}$ will start to decrement their backoff counters when they find the medium is idle for DIFS time, which is usually shorter than ACK timeout. In other words, during the beginning time slots of the contention, only clients in $\bar{A}$ participate, which obviously violates the assumptions in the analytical model that every client joins the contention and wins the channel with equal probability. Moreover, depends on the actual value of ACK timeout, the time slots for clients in A and $\bar{A}$ may become asynchronous. This asynchronous behavior makes the whole protocol even more complex and is certainly not considered by our analytical model.

In Table 3.2 we list the simulation results of the saturation throughput and mean access delay for different ACK timeout values, with $M = 5$, $N = 10$, and $E[T_1] = 2000$ μs. Note that the ACK timeout values are separated in two categories. The upper half values all equal a DIFS time period (i.e., 34 μs) plus a time period of multiple time slots (i.e., multiples of 9 μs). They follow this pattern to ensure that the time slots of all clients (i.e., clients in both A and $\bar{A}$) are synchronized. By contrast, the lower half values are selected so that the clients in the two

Table 3.2: The influence of different ACK timeout values on the saturation throughput with $M = 5$, $N = 10$, and $E[T_1] = 2000$ μs.

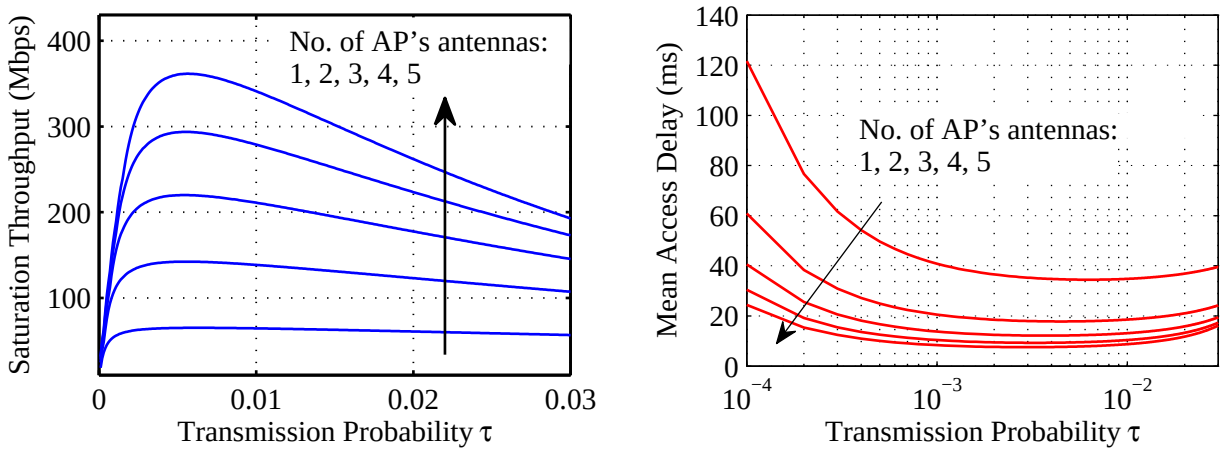
ACK timeout	Saturation Throughput (Mbps)	Mean Access Delay (ms)
70	346.55	5.44
97	346.56	5.46
124	347.33	5.43
60	362.42	5.24
80	365.12	5.21
100	361.64	5.26

sets experience asynchronous time slots. From this table, we find that the differences between the simulation results in each category are less than 1%, while the differences between the two categories are around 4%. This indicates that the main impact of ACK timeout on saturation throughput comes from the impact of asynchronous time slots. Our analytical model cannot characterize this impact of asynchronous time slots, which is certainly a limitation. However, since the differences between the saturation throughput of different ACK timeout values are relatively small, and considering the complex situation of asynchronous time slots, ignoring ACK timeout during analytical model derivation remains a reasonable simplification.

3.4 Implications from the Analytical Model

In this section the developed model is used to analyze the network performance with respect to different parameters. The CSMA/CA-based MU-MIMO WLAN is investigated in the first three subsections while performance of the opportunistic transmission scheme is analyzed in the last subsection.

As discussed in Section 3.3.3, our analytical model can accurately characterize the network performance when the number of concurrent streams in a successful round equals the number of antennas at the AP, i.e., when all the AP's dimensions are occupied. Therefore, performance analysis using our model can reveal the full influence of varying AP's antennas. To maintain a



(a) (b)

Figure 3.13: Saturation throughput and mean access delay versus τ for different numbers of antennas at the AP, with $N = 15$, $E[T_1] = 2000$ μs.

high accuracy when using the proposed model, in this section we will focus on networks with no more than 6 antennas at the AP.

3.4.1 Transmission Probability

In Section 3.2.4 the analytical throughput ρ is derived as a function of the transmission probability τ, and so is the mean access delay d in Section 3.8(b). To highlight their dependence on τ, in this subsection we express ρ as $\rho(\tau)$ and d as $d(\tau)$.

In Fig. 3.13, we plot $\rho(\tau)$ and $d(\tau)$ for $M = 1, 2, 3, 4, 5$, $N = 15$, and τ from 0 to 0.03. As shown in the figure, $\rho(\tau)$ is maximized at a certain transmission probability $\tau_{\max}$. Similarly, $d(\tau)$ achieves its minimum when τ reaches a certain value $\tau_{\min}$. Note that τ is an indication of clients' willingness in transmitting during a slot time. When τ is small, few clients tend to transmit in a time slot, so a large amount of time is wasted by idle time slots. When τ is large, the probability that two or more clients transmit in the same time slot is high, so the collision probability is large. Both cases would lead to a small saturation throughput and a large mean access delay.

Since τ is the solution of two nonlinear equations (i.e., Eqs. (3.1) and (3.11)), when M, N

Table 3.3: The maximum saturation throughput and minimum mean access delay achieved in Fig. 3.13.

M	$\rho_{\max}$ (Mbps)	$W_{\max}$	$d_{\min}$ (ms)	$W_{\min}$
1	65.07	$[312, 327]$	34.46	$[302, 338]$
2	142.3	$[338, 384]$	17.82	$[407, 487]$
3	219.9	$[350, 384]$	12.16	540
4	293.7	$[356, 364]$	9.296	605
5	361.5	$[344, 363]$	7.552	$[666, 689]$

are given, τ is fully determined by the backoff parameters, i.e., $CW_{\min}$ and $CW_{\max}$. Therefore, based on $\tau_{\max}$ and $\tau_{\min}$, we can obtain the optimal backoff parameters, which corresponds to the maximum throughput and minimum access delay. For simplicity, consider a special backoff strategy that employs constant window size, i.e., $CW_{\min} = CW_{\max}$. Eq. (3.1) then becomes: $\tau = 2/(W + 1)$, where $W = CW_{\min} + 1$. Using this simple relation between τ and W, optimal contention window sizes can be found. As shown in Table 3.3, $W_{\max}$ and $W_{\min}$ represent the optimal backoff window sizes** corresponding to $\rho_{\max}$ and $d_{\min}$, respectively.

3.4.2 Number of Antennas at the AP

In this subsection the influence of AP's antennas is evaluated. For convenience, ρ and d are expressed as $\rho(M)$ and $d(M)$.

As shown in Fig. 3.8(b), $d(M)$ decreases as M increases from 1 to 6. This is mainly because with more concurrent transmission opportunities, a client would have a larger chance to access the channel. However, since a transmission round of M concurrent clients fails when any one of the M clients encounters a collision, the failure probability of a round increases as M grows. Therefore, $d(M)$ decreases slowly and would finally increase at a large M.

In Fig. 3.14(a), we plot $\rho(M)$ for $E[T_1] = 2000$ and 4000 μs with $N = 30$. Saturation throughput achieved with fixed backoff parameters ($CW_{\min} = 127$ and $CW_{\max} = 1023$) is

**Due to the precision limitation of MATLAB, in the table $W_{\max}$ ($W_{\min}$) would be represented as intervals instead of a single value.

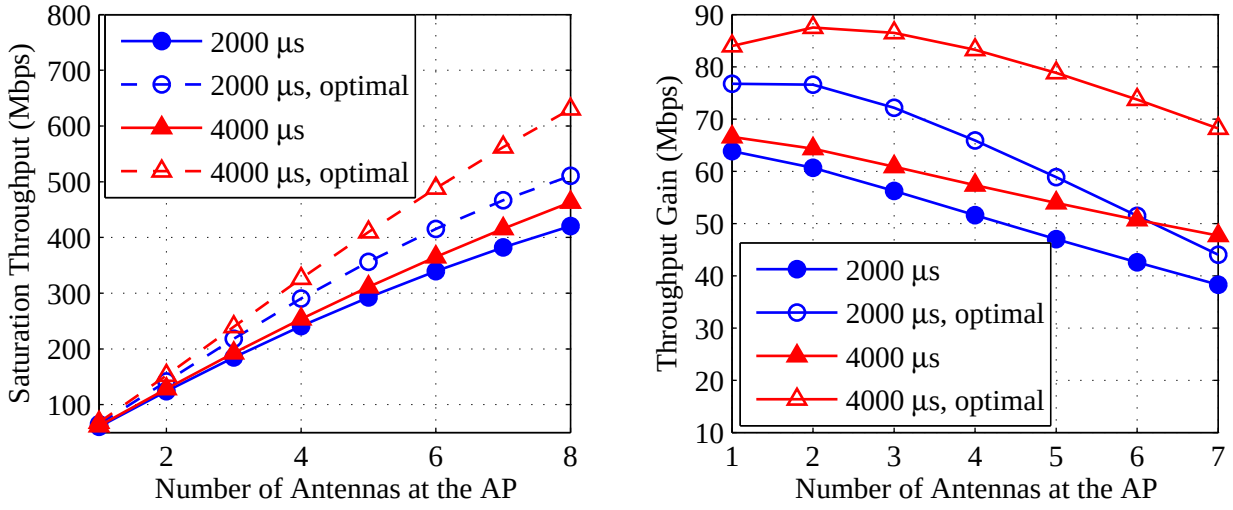


Figure 3.14: (a) Saturation throughput versus the number of antennas at the AP for different $E[T_1]$s, with $N = 30$. The solid lines denote the saturation throughput when $CW_{\min} = 127$ and $CW_{\max} = 1023$. The dashed lines correspond to the optimal saturation throughput evaluated at the optimal backoff parameters. (b) The throughput gain of adding one antenna to the current AP. The solid lines and dashed lines are calculated under the same scenarios as (a).

depicted in solid lines, while the maximum saturation throughput achieved at the optimal backoff parameters is shown in dashed lines. As shown in Fig. 3.14(a), $\rho(M)$ increases as M increases from 1 to 6. Let $\Delta\rho(M)$ denote the the amount of increased throughput when one more antenna is added to the M-antenna AP. In Fig. 3.14(b) we plot $\Delta\rho(M)$ under the same scenarios as in Fig. 3.14(a).

Fig. 3.14 indicates two things. First, the throughput is high with a large $E[T_1]$, mainly because the data transmitted during virtual transmission time is large when the frame transmission time is long. Second, the throughput gain of adding more antennas at the AP decreases as M grows large. The reasons are threefold.

- The frame transmission time of the M-th concurrent client (i.e., T_M) decreases as M increases. As shown in Fig. 3.2, transmission time is wasted by the contention periods in a round. The number of idle time slots during the contention can be reduced by choosing suitable backoff parameters. The resulting throughput increase is indicated by the dashed

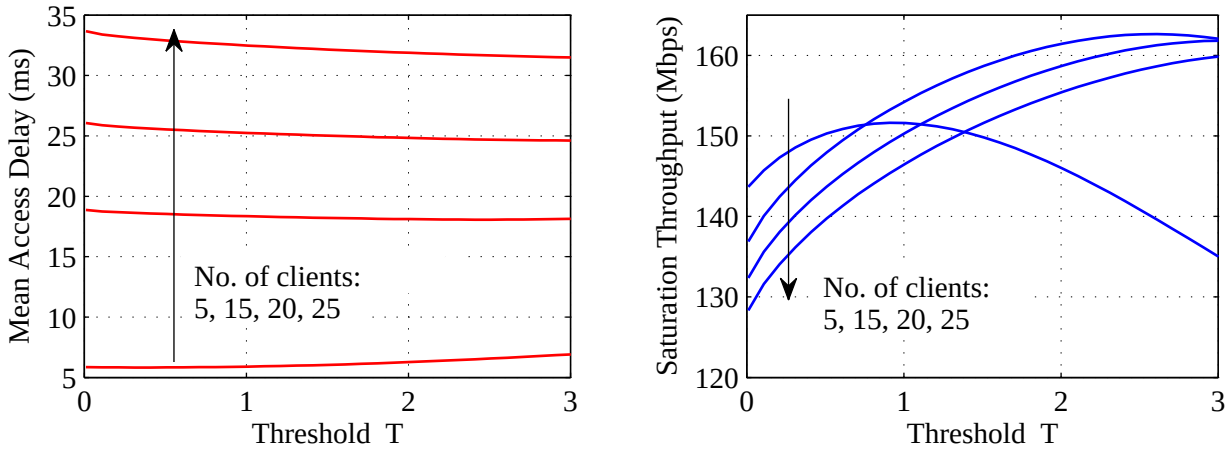
lines in Fig. 3.14.

- The date rate achieved by the M-th stream (i.e., R_M) decreases for a large M. According to the decoding procedure (Fig. 3.5), the M-th stream is decoded while the previous $M-1$ streams are treated as interference. The rate decrease due to interference can be avoided if the concurrent streams have orthogonal channel gains, which can hardly happen in a pure random access MAC protocol. However, we can consider an opportunistic MAC protocol, which gives clients with larger concurrent rate higher probabilities to join the ongoing transmission. A simple opportunistic transmission scheme is modeled in Section 3.2.6 and is analyzed in Section 3.4.4.

- A large M means a large chance for a round to fail, since any one of the M concurrent streams encounters a collision would result in transmission failure. The increased probability of a failed round is also a reason for the decrease of $\Delta\rho(M)$ as M grows.

3.4.3 Network Size

The network size refers to the total number of clients in the network, which is denoted previously as N. In this subsection we focus on $\rho(N)$ and $d(N)$. As shown in Fig. 3.8(a), $\rho(N)$ first increases and then decreases as N grows large. The reason is straightforward: when N is small, the number of concurrent clients in a successful round is limited by N (see Eq. (3.2)); when N is large, the number of clients that contend for each concurrent transmission opportunity is large, resulting in a large collision probability. How d varies with respect to N can be found in Fig. 3.8(b). Given the number of antennas at the AP, the mean access delay d increases with N due to increased collision probability.



(a) (b)

Figure 3.15: Mean access delay (above) and saturation throughput (below) versus threshold for the opportunistic transmission scheme in Section 3.2.6.

3.4.4 Threshold

In this subsection we will analyze how the threshold T affects the network performance in the opportunistic transmission scheme. In Section 3.2.6, a network with a 2-antenna AP is considered. In every transmission round only clients with concurrent rates larger than $B\log_2(1+PT/N_0)$ are allowed to contend for the concurrent transmission opportunity. When $T = 0$, the opportunistic scheme is just the original CSMA/CA-based MU-MIMO transmission scheme.

In Fig. 3.15(a) we plot $d(T)$ when the number of clients are 5, 15, 20, and 25. As shown in the figure, d increases slowly with T for $N = 5$ while decreases slowly for other cases. In Fig. 3.15(b) we plot $\rho(T)$ for different network sizes. The saturation throughput ρ tends to first increase with T and then decrease after T reaches a certain value. This inverted U-shaped curve can be seen for $N = 5$ when T goes from 0 to 3. The relationship of d and ρ versus T can be explained by the following reasons.

- As T increases, a network would obtain two benefits. The first benefit is an increased data rate of the second stream, as indicated in Eq. (3.41). The second benefit is a reduced collision probability, since a large T results in a smaller number of contending clients in

the second contention period. The two benefits would drive the saturation throughput to increase with T. Besides, the reduced collision probability leads to a reduced virtual transmission time (Fig. 3.6), which is the cause for the decreased mean access delay.

- A large T would degrade the network performance. When T is large, the average number of clients that can contend for the concurrent transmission (i.e., $E[N_{\text{join}}]$) is small. According to Eq. (3.43), the average transmission time of the second stream (i.e., $E[T_2]$) would then reduce. An extreme case would be that no client contends for the concurrent transmission, i.e., $N_{\text{join}} = 0$. In that case, T_2 is equal to 0, which corresponds to the worst case since one of AP's degrees of freedom is wasted. The probability of no contending clients in the second contention period is high when the network size is small, which explains why the network performance loss is prominent when $N = 5$.

3.5 Related Work

Many studies have been performed to design and analyze a wireless network that enables multiple concurrent transmissions in the uplink. In Zheng et al. (2006), Zheng *et al.* propose and analyze a RTS/CTS-based MAC protocol that supports multiple packet reception (MPR) in a WLAN. The proposed protocol is extended in Huang et al. (2008), where adaptive resource allocation and MPR are jointly considered through a cross-layer framework. In Jin et al. (2008) and Jin et al. (2011), Jin *et al.* compare the network performance of single-user MIMO and MU-MIMO schemes in the uplink WLAN, where MU-MIMO transmission is enabled when multiple clients win the contention at the same time. Throughput tradeoff between downlink and uplink in an MU-MIMO based WLAN is investigated in Jin et al. (2009). In Yoon et al. (2011), Yoon *et al.* develop, analyze, and implement a CSMA-based scheme that enables simultaneous concurrent transmissions from different clients in an ad hoc network.

Despite the many previous research efforts on the design and analysis of MU-MIMO schemes

in the uplink, a key difference exists between the MU-MIMO transmission schemes that are analyzed in previous research and the one we have analyzed in this paper. In the previous schemes, concurrent streams are transmitted at the same time by different clients, i.e., their transmissions start *synchronously*. However, in the CSMA/CA-based MU-MIMO WLAN of Section 3.1, clients are allowed to join the ongoing transmission one after another, resulting in *asynchronous* concurrent transmissions. This asynchronous characteristic causes two effects in the performance modeling and analysis procedure. The first effect is an increased complexity in performance modeling. Taking the conditional collision probability p as an example, if concurrent transmissions are synchronous, then p is simply calculated as the probability that the number of concurrent streams is larger than the maximum number allowed (see, e.g., Eq. (20) of Zheng et al. (2006), Eq. (13) of Jin et al. (2008), and Eq. (3) of Yoon et al. (2011)). However, in the case of asynchronous concurrent transmissions, the derivation of p is more complicated, as indicated in Section 3.2.2. The second effect is an increased complexity in performance analysis. As indicated in Section 3.4, when discussing about the influence of different parameters on the network performance, effects on the average transmission time of each concurrent stream, the collision probability, and the concurrent transmission rates have to be jointly considered. Modeling and analyzing the network performance by considering the two effects is the main contribution of this paper.

Over the past years, many efforts have been made to improve the throughput of MU-MIMO networks by selecting a subset of users to perform concurrent transmissions. Multi-user selection algorithms are proposed for both the downlink Shen et al. (2006) and uplink MU-MIMO systems Fan et al. (2009). Joint user/antenna selection algorithms are investigated in Chen et al. (2008). Most of the proposed algorithms are centralized, in which a scheduler is assumed to have the clients' channel information. In this paper we consider a simple distributed opportunistic scheduling scheme, where users contend for the concurrent transmission opportunities only when their concurrent rates are large enough. We model and analyze its throughput and

delay performance by considering both PHY and MAC layer influences.

3.6 Summary

In this chapter we modeled and investigated the saturation throughput and mean access delay of a CSMA/CA-based MU-MIMO WLAN, where clients are allowed to contend for concurrent transmission opportunities. Concurrent transmission rates were formulated for the ZF-SIC decoding mechanism, with clients experiencing i.i.d. time-varying Rayleigh fading. We also considered a simple distributed opportunistic transmission scheme, where clients are able to join the ongoing transmissions only when their concurrent rates exceed a threshold.

We derived analytical models to characterize the network performance of the CSMA/CA-based MU-MIMO WLAN and its opportunistic variation. Comparisons between simulation and analytical results were conducted to demonstrate the validity of our analytical model. Our analytical model was shown to be most accurate when the number of concurrent streams of a successful round was equal to the maximum number, a situation when all the degrees of freedom at the AP were occupied by the concurrent clients.

Based on the analytical model, the saturation throughput ρ and mean access delay d were investigated with respect to four parameters. The first one was the transmission probability τ. We showed that optimal backoff sizes could be derived for a network to achieve the maximum ρ and minimum d. The optimal backoff sizes balance the costs between collisions and idle time slots. The second parameter was the number of AP's antennas. The throughput gain of adding one antenna at the AP reduces as the total number of AP's antennas grows, due to a reduced concurrent frame time, a decreased concurrent rate, and an increased collision probability. The third parameter was the network size. An optimal network size was found, which provides a tradeoff between the cost of underutilized AP's antennas and the cost of increased collision probability. The fourth parameter was the threshold T of the opportunistic scheme. Our

model can be used to determine an optimal T that results from the joint effects of an increased concurrent rate, a reduced collision probability, and a decreased concurrent transmission time. In summary, our modeling and analysis provided an insightful understanding of the CSMA/CA-based MU-MIMO transmission scheme.

Chapter 4

Conclusion

Users' demands towards higher data rates and better QoS are growing rapidly. Meeting those demands in an efficient and effective manner becomes the central goal of next generation wireless networks. To achieve this goal, new physical layer technique has to be incorporated into the wireless systems, which brings about many challenging problems. A critical issue is to design upper layer protocols that can make the best use of those techniques employed in the physical layer. The aim of this thesis is to address those challenging issues by considering two promising physical layer techniques. The first one is two-way relaying. By developing an information theoretical framework to characterize the EE, BE, and latency of a linear multi-hop two-way relay network, we provided insightful guidelines towards optimal routing path selection and routing protocol design. The second physical layer technique that we considered is MU-MIMO. Specifically, our focus was on a recently proposed CSMA/CA-based MU-MIMO transmission scheme. By modeling and analyzing the saturation throughput and mean access delay of this random access MU-MIMO MAC protocol and its opportunistic variation, we optimized the network performance with respect to different parameters. Our modeling and analysis provided an in-depth understanding of the random access MU-MIMO transmission schemes. For both physical layer techniques considered in this thesis, the developed framework can be a helpful

tool for future study.

We have finished several high-quality papers by conducting the two research topics. Some research results have been published in IEEE GLOBECOM 2013 (Wu et al., 2013b) (Wu et al., 2013a) (Mao et al., 2013), while others are under review with IEEE Transactions on Wireless Communications. More information about our publications can be found in Appendix A.

4.1 Contributions

The contributions and important results of this thesis are summarized below.

- An information theoretical framework was developed to characterize the EE, BE, and latency of a linear multi-hop network, where Hop-by-Hop scheduling scheme was used to enable transmission through amplify-and-forward TWRCs.
- An optimal power allocation scheme was derived, based on which a multi-hop two-way relay network was able to consume the smallest energy at a given transmission rate.
- The EE-BE performance of a linear multi-hop two-way relay network was evaluated by means of the developed information theoretical framework. Specifically, we found two important relations: 1) When transmission rate is low, network with fewer relays presents higher EE and lower latency; 2) When transmission rate is high, network with more relays consumes less energy but have higher latency.
- An analytical model was developed to formulate the saturation throughput and mean access delay of the uplink channel in a CSMA/CA-based MU-MIMO WLAN. Comparisons between simulation and analytical results were conducted to verify our analytical model.
- A distributed opportunistic transmission scheme was proposed and modeled. In this scheme, clients are allowed to contend for the concurrent transmission opportunities only when their concurrent rates exceed a threshold.

- The saturation throughput and mean access delay were investigated using the developed analytical model. Specifically, we optimized the network performance over the back-off window size, the network size, and the threshold of the opportunistic transmission scheme. We found that the throughput gain from adding one antenna at the AP would reduce as the total number of antennas grows. Performance variation with respect to four parameters were analyzed thoroughly.

4.2 Future Work

In addition to the above contributions, we have also found many interesting areas for future research. They are outlined as follows.

For the first topic we study the performance of routing paths under the assumptions of equally spaced relays and a small curvature. Overcoming these limitations and considering more general cases (e.g., transmission with multiple frequencies and imperfect scheduling) would be an area of future research. Another important research area is to design an efficient MAC protocol that can exploit the advantages of two-way relaying technique in a wireless network.

For the second topic our investigation is restricted to the case of single-antenna clients. Considering the throughput and delay performance of a network with multi-antenna clients would be one direction for future research. Another area of research is to construct performance models under non-saturation conditions with traffic differentiation mechanisms such as EDCA. Besides, according to our modeling of an opportunistic transmission scheme, an interesting research direction would be to determine how a random access MU-MIMO WLAN performs with a different MAC protocol, and see whether a better transmission scheme can be designed.

Appendix A

Publications

- Wenguang Mao, Xudong Wang, and Shanshan Wu, "Distributed opportunistic scheduling with QoS constraints for wireless networks with hybrid links," submitted to *IEEE Transactions on Wireless Communications* (submission date: Oct. 5, 2013).

- Shanshan Wu, Wenguang Mao, and Xudong Wang, "Performance study on a CSMA/CA-based MAC protocol for multi-user MIMO wireless LANs," revised for *IEEE Transactions on Wireless Communications* (first round review result: major revision, Nov. 13, 2013).

- Shanshan Wu, Wenguang Mao, and Xudong Wang, "Performance analysis of random access multi-user MIMO wireless LANs," in *Proceedings of the IEEE GLOBECOM*, 2013.

- Shanshan Wu, Wenguang Mao, and Xudong Wang, "Information-theoretic study on routing path selection in two-way relay networks," in *Proceedings of the IEEE GLOBECOM*, 2013.

- Wenguang Mao, Shanshan Wu, and Xudong Wang, "QoS-oriented distributed opportunistic scheduling for wireless networks with hybrid links," in *Proceedings of the IEEE GLOBECOM*, 2013.

Appendix B

Detailed Derivation of the Optimal Power in Section 2.2

In this appendix we present the detailed procedure of deriving the optimal power allocation schemes in Section 2.2.5 and Section 2.2.6, which correspond to the linear multi-hop two-way relay networks with four relays and five relays, respectively.

B.1 Four Relays

As illustrated in Section 2.2.5, our objective is to find the minimum value of F under the constraints of Eqs. (2.20)-(2.24). The variable F is defined as follows

$$F = P_{\mathbb{A}} + P_{\mathbb{R}_1} + P_{\mathbb{R}_2} + P_{\mathbb{R}_4} + P_{\mathbb{B}}. \tag{B.1}$$

First of all, we rearrange Eqs. (2.20)-(2.24) to get the five variables, i.e., $P_{\mathbb{A}}$, $P_{\mathbb{R}_1}$, $P_{\mathbb{R}_2}$, $P_{\mathbb{R}_4}$, $P_{\mathbb{B}}$. From Eqs. (2.23) and (2.24), we can obtain

$$P_{\mathbb{R}_1} = \frac{3^{\alpha} P_{\mathbb{B}}}{2^R - 1} - 3^{\alpha}|h|^{-2} N_0. \tag{B.2}$$

Substituting Eqs. (B.2) and (2.22) into the expression of F gives

$$F = (1+(2^R-1)5^{-\alpha})P_{\mathbb{A}}+(1+(2^R-1)3^{-\alpha})P_{\mathbb{R}_2}+(\frac{3^\alpha}{2^R-1}+1)P_B+((2^R-1)-3^\alpha)|h|^{-2}N_0. \quad \text{(B.3)}$$

Eliminating $P_{\mathbb{R}_4}$ from Eqs. (2.20) and (2.23) gives

$$2^R h^4\beta^2 P_{\mathbb{R}_2} + |h|^4\beta^2(2^R-1)P_{\mathbb{A}} = (3^\alpha|h|^2 + 5^{-\alpha}|h|^2(2^R-1))P_{\mathbb{B}} + (2^R-1)(1-3^\alpha)N_0. \quad \text{(B.4)}$$

Similarly, eliminating $P_{\mathbb{R}_4}$ from Eqs. (2.21) and (2.23) gives

$$2^R h^4\beta^2 P_{\mathbb{A}} + |h|^4\beta^2(2^R-1)P_{\mathbb{R}_2} = (3^\alpha|h|^2 + 3^{-\alpha}|h|^2(2^R-1))P_{\mathbb{B}} + (2^R-1)(1-3^\alpha)N_0. \quad \text{(B.5)}$$

From Eqs. (B.4) and (B.5), we get

$$(2^{R+1}-1)|h|^4\beta^2 P_{\mathbb{A}} = a_1|h|^2 P_{\mathbb{B}} + (2^R-1)(1-3^\alpha)N_0, \quad \text{(B.6)}$$

$$(2^{R+1}-1)|h|^4\beta^2 P_{\mathbb{R}_2} = a_2|h|^2 P_{\mathbb{B}} + (2^R-1)(1-3^\alpha)N_0. \quad \text{(B.7)}$$

Substituting Eq. (2.22) into Eq. (2.23), together with Eqs. (B.6) and (B.7), we get

$$(1-a_5)|h|^2 P_{\mathbb{B}} = (a_3 + |h|^2\beta^2 a_4)N_0. \quad \text{(B.8)}$$

Substituting Eqs. (B.6), (B.7), and (B.8) into the expression of Eq. (B.3) gives

$$F = (a_8\beta^2 + \frac{a_7}{|h|^4\beta^2} + a_6|h|^{-2})N_0 \geq (2\sqrt{a_7 a_8} + a_6)N_0|h|^{-2}. \quad \text{(B.9)}$$

Therefore, the minimum of F is $F_{\min} = (2\sqrt{a_7 a_8} + a_6)N_0|h|^{-2}$, achieved when $\beta^4 = a_7|h|^{-4}/a_8$.

The corresponding optimal power allocation is

$$P_B = \frac{(a_3 + a_4|h|^2\beta^2)N_0}{(1-a_5)|h|^2}, \tag{B.10}$$

$$P_{\mathbb{A}} = \frac{a_1|h|^2 P_B + (2^R-1)(1-3^\alpha)N_0}{(2^{R+1}-1)|h|^4\beta^2}, \tag{B.11}$$

$$P_{\mathbb{R}_2} = \frac{a_2|h|^2 P_B + (2^R-1)(1-3^\alpha)N_0}{(2^{R+1}-1)|h|^4\beta^2}. \tag{B.12}$$

The optimal power allocated at $\mathbb{R}_1$ and $\mathbb{R}_4$ can then be easily computed from Eqs. (2.22) and (2.23). The variables $\{a_i\}_{i=1}^8$ are functions of R and α, which are shown below.

$$a_1 = -(2^R-1)^2 5^{-\alpha} + 2^R(2^R-1)3^{-\alpha} + 3^\alpha, \tag{B.13}$$

$$a_2 = 2^R(2^R-1)5^{-\alpha} - (2^R-1)^2 3^{-\alpha} + 3^\alpha, \tag{B.14}$$

$$a_3 = \frac{(2^R-1)^2}{2^{R+1}-1}(1-3^\alpha)(3^{-\alpha}2 + (45^{-\alpha}+27^{-\alpha})(2^R-1)) + (2^R-1), \tag{B.15}$$

$$a_4 = 9^{-\alpha}(2^R-1)^2 + 3^{-\alpha}(2^R-1), \tag{B.16}$$

$$a_5 = \frac{2^R-1}{2^{R+1}-1}[a_2(3^{-\alpha}+45^{-\alpha}(2^R-1)) + a_1(3^{-\alpha}+27^{-\alpha}(2^R-1))], \tag{B.17}$$

$$\begin{aligned} a_6 = {} & \frac{a_2a_4(1+(2^R-1)5^{-\alpha}) + a_1a_4(1+(2^R-1)3^{-\alpha})}{(2^{R+1}-1)(1-a_5)} \\ & + (2^R-1) - 3^\alpha + \frac{a_3}{1-a_5}(\frac{3^\alpha}{2^R-1}+1), \end{aligned} \tag{B.18}$$

$$\begin{aligned} a_7 = {} & \frac{(2^R-1)(1-3^\alpha)}{2^{R+1}-1}(2+(2^R-1)(3^{-\alpha}+5^{-\alpha})) \\ & + \frac{a_2a_3(1+(2^R-1)5^{-\alpha})}{(1-a_5)(2^{R+1}-1)} + \frac{a_1a_3(1+(2^R-1)3^{-\alpha})}{(1-a_5)(2^{R+1}-1)}, \end{aligned} \tag{B.19}$$

$$a_8 = \frac{a_4}{1-a_5}(\frac{3^\alpha}{2^R-1}+1). \tag{B.20}$$

Please note that for convenience, in Section 2.2.5 we use $\{c_i\}_{i=1}^3$ to express $F_{\min}$ instead of

$\{a_i\}_{i=1}^{8}$. Their relationships are $c_1 = a_8$, $c_2 = a_7$, and $c_3 = a_6$.

B.2 Five Relays

As illustrated in Section 2.2.6, our objective is to find the minimum value of S under the constraints of Eqs. (2.31)-(2.33). The variable S is defined as follows

$$S = 2(P_{\mathbb{A}} + P_{\mathbb{R}_1} + P_{\mathbb{R}_2}) \tag{B.21}$$

Similar to the previous section, we first rearrange Eqs. (2.31)-(2.33) to get the three variables, i.e., $P_{\mathbb{A}}$, $P_{\mathbb{R}_1}$ and $P_{\mathbb{R}_2}$. Substituting Eq. (2.33) into the definition of S gives

$$S = 2(1 + (1 + 5^{-\alpha})|h|^2\beta^2)P_{\mathbb{A}} + 2(1 + (1 + 3^{-\alpha})|h|^2\beta^2)P_{\mathbb{R}_2} + 2\beta^2 N_0. \tag{B.22}$$

We can subtract Eq. (2.31) from Eq. (2.32) and then obtain

$$|h|^4\beta^2(P_{\mathbb{A}} - P_{\mathbb{R}_2}) = (2^R - 1)(3^{-\alpha} - 5^{-\alpha})|h|^2 P_{\mathbb{R}_1}. \tag{B.23}$$

Substituting Eq. (2.33) into Eq. (B.23) gives

$$|h|^4\beta^2 P_{\mathbb{A}} = b_1|h|^4\beta^2 P_{\mathbb{R}_2} + b_2|h|^2\beta^2 N_0. \tag{B.24}$$

By substituting Eq. (B.24) into Eqs. (2.31) and (2.32), we can then get

$$|h|^4\beta^2 P_{\mathbb{R}_2} = (b_3|h|^2\beta^2 + b_4)N_0. \tag{B.25}$$

$$|h|^4\beta^2 P_{\mathbb{A}} = ((b_1 b_3 + b_2)|h|^2\beta^2 + b_1 b_4)N_0. \tag{B.26}$$

Substituting Eqs. (B.25) and (B.26) into Eq. (B.22) gives

$$S = (b_5\beta^2 + \frac{b_6}{|h|^4\beta^2} + b_7|h|^{-2})N_0 \geq (2\sqrt{b_5 b_6} + b_7)N_0|h|^{-2}. \tag{B.27}$$

Therefore, the minimum of S is $S_{\min} = (2\sqrt{b_5 b_6} + b_7)N_0|h|^{-2}$, achieved when $\beta^4 = b_6|h|^{-4}/b_5$. The corresponding optimal power allocation is

$$P_{\mathbb{A}} = \frac{((b_1 b_3 + b_2)|h|^2\beta^2 + b_1 b_4)N_0}{|h|^4\beta^2}, \tag{B.28}$$

$$P_{\mathbb{R}_2} = \frac{(b_3|h|^2\beta^2 + b_4)N_0}{|h|^4\beta^2}. \tag{B.29}$$

The optimal power allocated at $\mathbb{R}_1$ can be easily computed from Eq. (2.33). The variables $\{b_i\}_{i=1}^7$ are complicated functions of R and α, and their full expressions are

$$b_1 = \frac{1 + (2^R - 1)(3^{-\alpha} - 5^{-\alpha})(1 + 3^{-\alpha})}{1 - (2^R - 1)(3^{-\alpha} - 5^{-\alpha})(1 + 5^{-\alpha})}, \tag{B.30}$$

$$b_2 = \frac{(2^R - 1)(3^{-\alpha} - 5^{-\alpha})}{1 - (2^R - 1)(3^{-\alpha} - 5^{-\alpha})(1 + 5^{-\alpha})}, \tag{B.31}$$

$$b_3 = \frac{[(5^{-\alpha}2 + 25^{-\alpha})b_2 + 5^{-\alpha} + 1](2^R - 1)}{1 - (2^R - 1)[b_1 5^{-\alpha}(2 + 5^{-\alpha}) + 5^{-\alpha} + 3^{-\alpha} + 15^{-\alpha}]}, \tag{B.32}$$

$$b_4 = \frac{2^R - 1}{1 - (2^R - 1)[b_1 5^{-\alpha}(2 + 5^{-\alpha}) + 5^{-\alpha} + 3^{-\alpha} + 15^{-\alpha}]}, \tag{B.33}$$

$$b_5 = 2(b_1 b_3 + b_2)(1 + 5^{-\alpha}) + 2b_3(1 + 3^{-\alpha}) + 2 \tag{B.34}$$

$$b_6 = 2(b_1 + 1)b_4 \tag{B.35}$$

$$b_7 = 2(b_1 b_3 + b_2) + 2b_1 b_4(1 + 5^{-\alpha}) + 2b_3 + 2b_4(1 + 3^{-\alpha}). \tag{B.36}$$

Please note that for convenience, in Section 2.2.6 we use $\{c_i\}_{i=4}^6$ to express $F_{\min}$ instead of $\{b_i\}_{i=1}^8$. Their relationships are $c_4 = b_5$, $c_5 = b_6$, and $c_6 = b_7$.

Bibliography

Bae, C. and W. E. Stark (2009). End-to-end energy–bandwidth tradeoff in multihop wireless networks. *IEEE Transactions on Information Theory 55*(9), 4051–4066.

Bianchi, G. (2000). Performance analysis of the IEEE 802.11 distributed coordination function. *IEEE Journal on Selected Areas in Communications 18*(3), 535–547.

Bianchi, G. and I. Tinnirello (2005). Remarks on IEEE 802.11 DCF performance analysis. *IEEE Communications Letters 9*(8), 765–767.

Calì, F., M. Conti, and E. Gregori (2000). Dynamic tuning of the IEEE 802.11 protocol to achieve a theoretical throughput limit. *IEEE/ACM Transactions on Networking 8*(6), 785–799.

Chen, R., Z. Shen, J. G. Andrews, and R. W. Heath (2008). Multimode transmission for multiuser MIMO systems with block diagonalization. *IEEE Transactions on Signal Processing 56*(7), 3294–3302.

Cisco (2013). Cisco visual networking index: global mobile data traffic forecast update, 2012-2017. *Cisco White Paper*.

IEEE Computer Society LAN MAN Standards Committee and Others (2012), IEEE Std. 802.11-2012, *Wireless LAN medium access control (MAC) and physical layer (PHY) specifications*.

Fan, B., W. Wang, Y. Lin, L. Huang, and K. Zheng (2009). Spatial multi-user pairing for uplink virtual-MIMO systems with linear receiver. In *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE.

Huang, W. L., K. Letaief, and Y. J. Zhang (2008). Cross-layer multi-packet reception based medium access control and resource allocation for space-time coded MIMO/OFDM. *IEEE Transactions on Wireless Communications 7*(9), 3372–3384.

Jin, H., B. C. Jung, H. Y. Hwang, and D. K. Sung (2008). Performance comparison of uplink wlans with single-user and multi-user MIMO schemes. In *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1854–1859. IEEE.

Jin, H., B. C. Jung, H. Y. Hwang, and D. K. Sung (2009). A throughput balancing problem between uplink and downlink in multi-user MIMO-based WLAN systems. In *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE.

Jin, H., B. C. Jung, and D. K. Sung (2011). A tradeoff between single-user and multi-user MIMO schemes in multi-rate uplink wlans. *IEEE Transactions on Wireless Communications 10*(10), 3332–3342.

Lin, K. C.-J., S. Gollakota, and D. Katabi (2011). Random access heterogeneous MIMO networks. In *Proceedings of ACM SIGCOMM*, pp. 146–157. ACM.

Mao, W., S. Wu, and X. Wang (2013). Performance analysis of random access multi-user MIMO wireless lans. In *Proceedings of the IEEE Global Communications Conference (GLOBECOM)*. IEEE.

Popovski, P. and H. Yomo (2006a). The anti-packets can increase the achievable throughput of a wireless multi-hop network. In *Proceedings of the IEEE International Conference on Communications (ICC)*, pp. 3885–3890. IEEE.

Popovski, P. and H. Yomo (2006b). Bi-directional amplification of throughput in a wireless multi-hop network. In *Proceedings of the IEEE 63rd Vehicular Technology Conference (VTC)-Spring*, pp. 588–593. IEEE.

Shen, W.-L., Y.-C. Tung, K.-C. Lee, K. C.-J. Lin, S. Gollakota, D. Katabi, and M.-S. Chen (2012). Rate adaptation for 802.11 multiuser MIMO networks. In *Proceedings of the 18th Annual International Conference on Mobile Computing and Networking (MOBICOM)*, pp. 29–40. ACM.

Shokrollahi, A. (2006). Raptor codes. *IEEE Transactions on Information Theory 52*(6), 2551–2567.

Shen, Z., R. Chen, J. G. Andrews, R. W. Heath, and B. L. Evans (2006). Low complexity user selection algorithms for multiuser MIMO systems with block diagonalization. *IEEE Transactions on Signal Processing 54*(9), 3658–3663.

Tan, K., H. Liu, J. Fang, W. Wang, J. Zhang, M. Chen, and G. M. Voelker (2009). SAM: enabling practical spatial multiple access in wireless LAN. In *Proceedings of the 15th Annual International Conference on Mobile Computing and Networking (MOBICOM)*, pp. 49–60. ACM.

Tse, D. and P. Viswanath (2005). *Fundamentals of wireless communication*. Cambridge University Press.

Verdú, S. (2002). Spectral efficiency in the wideband regime. *IEEE Transactions on Information Theory 48*(6), 1319–1343.

Wang, Q., M. Hempstead, and W. Yang (2006). A realistic power consumption model for wireless sensor network devices. In *Proceedings of the 3rd Annual IEEE Communications Society on Sensor and Ad Hoc Communications and Networks (SECON), 2006*, Volume 1, pp. 286–295. IEEE.

Wu, S., W. Mao, and X. Wang (2013a). Information-theoretic study on routing path selection in two-way relay networks. In *Proceedings of the IEEE Global Communications Conference (GLOBECOM)*. IEEE.

Wu, S., W. Mao, and X. Wang (2013b). Performance analysis of random access multi-user MIMO wireless LANs. In *Proceedings of the IEEE Global Communications Conference (GLOBECOM)*. IEEE.

Wu, S. and X. Wang (2012, July). Information-theoretic study on routing path selection in two-way relay networks. http://arxiv.org/abs/1207.6650.

Yi, Z., M. Ju, and I.-M. Kim (2011). Outage probability and optimum power allocation for analog network coding. *IEEE Transactions on Wireless Communications 10*(2), 407–412.

Yoon, S., I. Rhee, B. C. Jung, B. Daneshrad, and J. H. Kim (2011). Contrabass: Concurrent transmissions without coordination for ad hoc networks. In *Proceedings of the IEEE Conference on Computer Communications (INFOCOM)*, pp. 1134–1142. IEEE.

You, Q., Z. Chen, Y. Li, and B. Vucetic (2011). Multi-hop bi-directional relay transmission schemes using amplify-and-forward and analog network coding. In *Proceedings of the IEEE International Conference on Communications (ICC)*. IEEE.

Zhang, S., S. C. Liew, and P. P. Lam (2006). Hot topic: physical-layer network coding. In *Proceedings of the 12th Annual International Conference on Mobile Computing and Networking (MOBICOM)*, pp. 358–365. ACM.

Zheng, P. X., Y. J. Zhang, and S. C. Liew (2006). Multipacket reception in wireless local area networks. In *Proceedings of the IEEE International Conference on Communications (ICC)*, pp. 3670–3675. IEEE.