Generating Function Analysis of Wireless Networks and ARQ Systems

by

Shihyu Chang

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Electrical Engineering: Systems)
in The University of Michigan
2006

Doctoral Committee:

Professor Wayne E. Stark, Co-Chair
Associate Professor Achilleas Anastasopoulos, Co-Chair
Professor Arthur G. Wasserman
Assistant Professor Mingyan Liu

ACKNOWLEDGEMENTS

I am most grateful to my Ph.D. advisors Professor Wayne E. Stark and Professor Achilleas Anastasopoulos, for their invaluable guidance and continued support during the course of this research work at the University of Michigan. I am also extremely grateful to Professor Stark for his patient rectification of my English pronunciation and academic writing skills.

I would like to thank the other members of my committee, Professor Mingyan Liu and Professor Arthur G. Wasserman, for their time and effort in reading this work and providing worthwhile comments and suggestions.

I would also like to present my gratitude to my colleagues, Kar-Peo Yar, Chih-Wei Wang, Jinho Kim and Changhun Bae, at the Wireless Communication Lab at the University of Michigan. I have enjoyed supportive and resourceful discussion with them during our group meetings. In addition, I would like to acknowledge my financial support from the Office of Naval Research under Grant N00014-03-1-0232.

Finally, my sincere thanks goes to my parents and siblings. Without their support and love, it is impossible for me to finish my Ph.D study at the University of Michigan. They have always been there to encourage me when I needed them.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

LIST OF APPENDICES

Appendix

CHAPTER I

Introduction

Future wireless systems may change the way people communicate, shop and work significantly by establishing ubiquitous communication among people and devices. For example, globalization of business will be realized by trading among companies located at different countries via internet; distance eduction enables students and teachers to communicate through information technology without the necessity for the students and teachers to be physically in same location and time; mobile access to work environment will allow people to work anywhere in the world. These goals will be fulfilled by interconnecting any devices at anywhere and anytime through wired and wireless communication.

A lot of challenges have not been solved in implementing the future wireless systems. Some of them are:

- Channel and network characteristics are random and time varying.
- Because most devices are battery powered and the energy of a battery is a limiting resource, energy efficient network techniques are crucial to design of a network system to meet some performance criteria.
- In order to have seamless communication between all existing different wired and wireless communication protocols, vertical handoff algorithms need to be

developed.

- Advanced coding schemes and multiple antenna systems are indispensable and an intelligent controller for radio resources is still missing.

In this thesis, we will concentrate on discussing the issue of energy efficiency. To combat severe channel conditions of wireless networks compared to wired networks, we need more energy in packet transmission to decrease packet error probability. On the other hand, if less energy is used, the loss of data packets could increase the delay to transmit a packet successfully. For instance, decreasing the transmission delay by increasing the transmission rate often results in less energy efficiency. Therefore, the requirements for optimizing performance are often contradictory and there is a fundamental tradeoff between energy and delay. These two factors are crucial in designing an energy efficient wireless network. In the rest of this thesis, we will determine the tradeoff between energy and delay analytically.

1.1 Wireless Network Architectures and Protocol Layers

We will describe briefly network architectures and the wireless protocol stack.

1.1.1 Wireless Network Architectures

There are two different network architectures: infrastructure and *ad hoc* networks. A wireless network in which stations communicate with each other by first going through an access point (AP) is called an infrastructure network. In an infrastructure network, wireless stations can communicate with each other or can communicate through a wired network with other stations not in radio range. A set of wireless stations which are connected to an AP is referred to as a basic service set (BSS). Most corporate wireless local area networks (LANs) operate in infrastructure mode

because they require access to the wired LAN in order to use services such as file servers or printers. Fig. 1.1 shows an infrastructure network. The big circle around each AP represents the communication range of each AP.

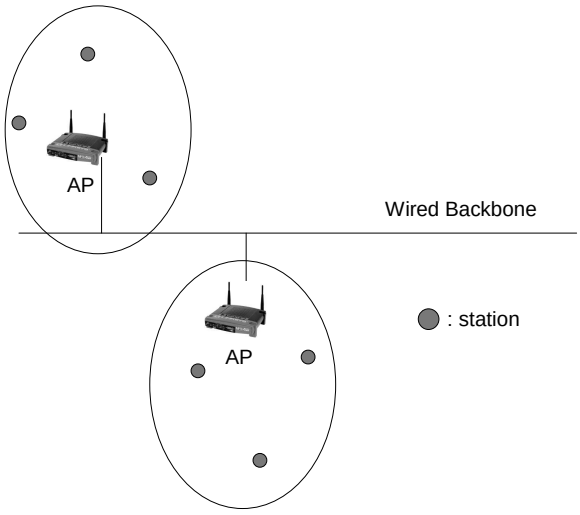

Figure 1.1: An infrastructure network.

A wireless network in which stations communicate directly with each other, without the use of an AP is called a network with *ad hoc* mode. An *ad hoc* network architecture is also referred to as peer-to-peer architecture or an independent basic service set (IBSS). *Ad hoc* networks are useful in cases that temporary network connectivity is required, and are often used for battlefields or disaster scenes. Fig. 1.2 shows an *ad hoc* network.

1.1.2 Protocol Layers

The concepts of protocol layering provides a basis for knowing how a complicated set of protocols cooperate together with the hardware to provide a complete wireless network system. Although new protocol stacks such as the infrared data association (IRDa) protocol stack for point-to-point wireless infrared communication and

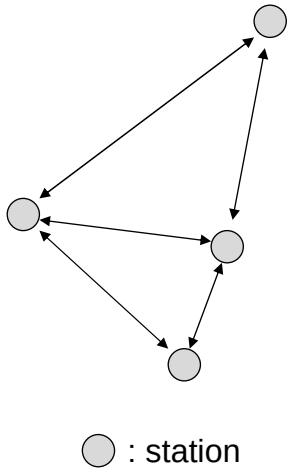

Figure 1.2: An *ad hoc* network. Each station communicates mutually without the help of AP.

Table 1.1: The OSI network's seven-layer model.

Application and Services layer
Presentation layer
Session layer
Transport layer
Network layer
Date link layer A. Logical control sublayer B. Media access control sublayer
Physical layer

the wireless application protocol (WAP) Forum protocol stack for building more advanced services have been proposed for wireless networks [2, 1], we will concentrate our discuss on the traditional OSI protocol stack. There are seven layers regulated by OSI, shown in Table 1.1.

Physical Layer:

The physical layer (PHY) is made up by radio frequency (RF) circuits, modulation, and channel coding systems. The major functions and services performed by the physical layer are :

- establishment and termination of a connection to a communications medium;

- conversion between the representation of digital data in user equipment and the corresponding signals transmitted over a communications channel.

Data Link Layer:

The function of data link layer is to establish a reliable logical link over the unreliable wireless channel. The data link layer is responsible for security, link error control, transferring network layer packets into frames and packet retransmission. The media access control (MAC), a sublayer of data link layer is responsible for the task of sharing the wireless channel used by stations in the network.

Network Layer:

The task of network layer is to rout packets, establish the network service type (connectionless vs. connection-oriented), and transfer packets between the transport and link layers. In the scenario of mobility of stations, this layer is also responsible for mobility management.

Transport Layer:

The transport layer provides transparent transfer of data between hosts. It is usually responsible for end-to-end error recovery and flow control, and ensuring complete data transfer. In the Internet protocol suite this function is achieved by the connection oriented transmission control protocol (TCP). The purpose of the transport layer is to provide transparent transfer of data between end users, thus relieving the upper layers from any concern with providing reliable and cost-effective data transfer.

Session Layer:

The session layer sets up, coordinates, and terminates conversations, exchanges, and dialogs between the applications at each end. It deals with session and connection coordination.

Presentation Layer:

The presentation layer converts incoming and outgoing data from one presentation format to another.

Application and Services Layer:

Source coding, digital signal processing (DSP), and context adaption are implemented in this layer. Services provided at this layer depend on the various users requirements.

Layered architectures have been used in most data networks such as Internet. Since all layers of the protocol stack affect the energy consumption and delay for the end-to-end transmission of each bit, an efficient system requires a joint design across all these layers. In this thesis, we focus on the lowest two layers of OSI seven layer architecture mentioned in Sec. 1.1.2. In the rest of section, we are going to determine the tradeoff between energy and delay in wireless networks taking into account the data link and physical layers.

1.2 Motivation

Time varying channels (e.g., multipath fading, shadowing) are often encountered in wireless and mobile systems. Given the energy used per information bit, a wireless communication system adopting forward-error-correction (FEC) can provide higher protection with lower code rate. However, a system with lower code rate needs longer delay to transmit the same information packet due to the fixed bandwidth. If we want to decrease the delay to transmit the same information packet, the system requires higher energy per information bit to achieve the same reliability as a system with lower code rate. Another example is shown in Fig. 1.3. The horizontal axis represents the system time and the vertical axis indicates the signal-to-noise ratio (SNR) of the channel. We plot a channel SNR realization with respect to the system

time in Fig 1.3. In Fig. 1.3(a), the transmitter waits until the channel SNR becomes high before transmissions. In this case, the system will spend large delay to finish transmission. In Fig. 1.3(b), the transmitter transmits immediately even when the channel SNR is low. However, the transmitter needs to spend more energy in transmission in order to keep the same reliability as in the previous case. Therefore, there is a fundamental tradeoff between energy and delay of a wireless communication system.

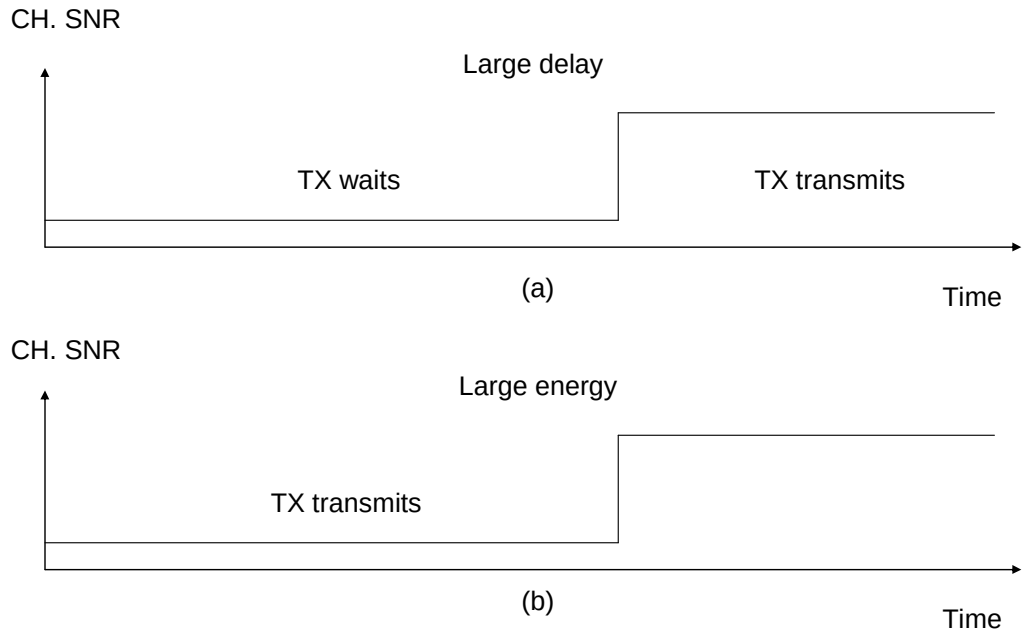

Figure 1.3: The illustration of fundamental tradeoff between energy and delay.

Hidden stations in a wireless network refer to stations which are out of the communication range of other stations. Take a physical star topology with an AP with many stations surrounding it in a circular fashion; each station is within communication range of the AP, however, not each station can communicate with each other. For example, it is likely that the station at the far edge of the circle can access the AP since the distance between the station and AP is less than the communication

range, say **r**, but it is unlikely that the same station can detect a station on the opposite end of the circle because the distance between these two stations exceeds the communication range **r**. These two stations are known as hidden stations with respect to each other. Fig. 1.4 illustrates the hidden stations problem of a wireless network. Hidden stations problem leads to difficulties in media access control.

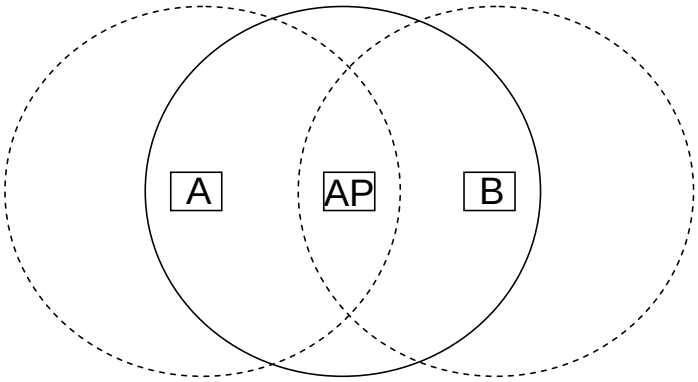

Figure 1.4: The hidden stations problem. Station A and B are hidden station of each other.

In order to solve the problem of hidden stations [33], which occurs when some stations in the network are unable to detect each other, the 802.11 protocol uses a mechanism known as request-to-send/clear-to-send (RTS/CTS). Before transmitting the data packet, the source station sends a request-to-send (RTS) packet. If the RTS packet is received correctly at the destination station (there is no collision with other RTS packets sent by other competing stations and the receiver correctly decodes the packet) the destination station broadcasts a clear-to-send (CTS) packet. If the CTS packet is successfully received by the source station, the channel reservation is successful and the source station will begin to send data and wait for the acknowledgement (ACK) packet. The source station detects an unsuccessful channel reservation by the lack of a correct CTS packet. The MAC protocol investigated in

this thesis is RTS/CTS protocol adopted in IEEE 802.11 standard.

It is a well known fact that a packet with more redundant bits have smaller error probability. This means that longer packets can be transmitted successfully with lower energy demand. However, longer packets require larger delay to transmit. In order to determine the tradeoff between energy and delay of a wireless system, we need to have a relation between error probability, energy and delay for each type of packet in the system. In this thesis, we will use the reliability function bounds for a channel to determine the packet error probability. Let K be the number of information bits in a packet and N be the number of coded symbols for this packet. Then there exists an encoder and decoder for which the packet error probability $P_{e,K,N}$ is bounded by

$$P_{e,K,N} \leq 2^{K-NR_0}, \tag{1.1}$$

where R_0 is the cutoff rate determined by channel characteristics. For an additive white Gaussian noise (AWGN) channel using binary input the cutoff rate is $R_0 = 1 - \log_2(1 + e^{-E_c/N_0})$, where E_c/N_0 is the received signal-to-noise ratio per coded symbol.

1.3 Literature Review for IEEE 802.11 DCF Analysis

We will begin with a brief introduction of IEEE 802.11 protocol followed by reviewing literatures about its analysis. The IEEE 802.11 protocol for wireless LANs is a multiple access technique based on carrier sense multiple access/collision avoidance (CSMA/CA). The basic operation of this protocol is described as follows. A station with a packet ready to transmit listens the activity of transmission channel. If the channel is sensed idle, the station captures the channel and transmits data packets. Otherwise, the station defers transmission and keeps in the backoff state. There are

two basic techniques to access the medium. The first one is called distributed coordination function (DCF). There is no centralized coordinator in the system to assign the medium to the users in the network. This is a random access scheme, based on the CSMA/CA protocol. Whenever the packets collide or have errors, the protocol adopts a random exponential backoff scheme before retransmitting the packets. Our work is based on using DCF to access the medium. Another medium access method is called point coordinator function (PCF) to implement medium access control. There is a centralized coordinator to provide collision free and time bounded services.

Because the analysis of network performance in this thesis is based on Bianchi's distributed coordination function (DCF) analysis, we will give a literature review of papers which use Bianchi's DCF analysis in this section for comparison. From these papers, we classify them into three categories. The first category is to use Bianchi's model to perform the cross-layer analysis (PHY and MAC). The second category is to use Bianchi's model to perform the priority and scheduling analysis. We put other applications in the third category due to their variety. Some use Bianchi's model to improve channel utilization and some modifies Bianchi's model to consider non-saturation traffic scenario. We will begin with the first category.

1.3.1 PHY and MAC Cross-layer Analysis

In [24], a theoretical cross-layer saturation goodput model for the IEEE 802.11a PHY layer and DCF basic access scheme MAC protocol is developed. The proposed analytical approach relates the system performance to channel load, contention window (CW) resolution algorithms, distinct modulation schemes (BPSK and QPSK), FEC schemes (convolutional code), receiver structures (maximum ratio combining) and channel models (uncorrelated Nakagami-m fading channel). In [32], the authors study the impact of frequency-nonselective slowly time-variant Rician fading chan-

nels on the performance of single-hop *ad hoc* networks using the IEEE 802.11 DCF in saturation. They study the throughput performance of the four-way handshake mechanism under direct sequence spread spectrum (DSSS) with differential binary phase shift keying (DBPSK) modulation.

In an *ad hoc* network, it is important that all stations are synchronized to a common clock. Synchronization is necessary for frequency hopping spread spectrum (FHSS) to ensure that all stations hop at the same time; it is also necessary for both FHSS and direct sequence spread spectrum (DSSS) to perform power management. In [25],the authors evaluates the synchronization mechanism, which is a distributed algorithm, specified in the IEEE 802.11 standards. By both analysis and simulation, it is shown that when the number of stations in an IBSS is not very small, there is a non-negligible probability that stations may get out of synchronization. The more stations, the higher probability of asynchronism. Thus, the current IEEE 802.11s synchronization mechanism does not scale; it cannot support a large-scale *ad hoc* network. To alleviate the asynchronism problem, this paper also proposes a simple modification to the current synchronization algorithm. The modified algorithm is shown to work well for large *ad hoc* networks.

In [46], a cross-layer analytical approach from both the PHY layer and the MAC layer to evaluate the performance of the IEEE 802.11 is developed. From the PHY layer, this analytical approach incorporates the effects of both capture and directional antenna, while from the MAC layer, the model takes account of the CSMA/CA protocol. Through a cross-layer modelling technique, this analytical framework can provide valuable insights of the PHY layer impact on the throughput performance of the CSMA/CA MAC protocol. These insights can be helpful in developing a MAC protocol to fully take advantage of directional antennas for enhancing the

performance of the WLAN. In [36], the proposed model computes its saturation throughput by relating to the positions of the other concurrent stations. Further, this model provides the total saturation throughput of the medium. They solve the model numerically and show that the saturation throughput per station is strongly dependent not only on the stations position but also on the positions of the other stations.

1.3.2 Priority and Scheduling Analysis

In [28, 29], the authors develop two mechanisms for QoS communication in multi-hop wireless networks. First, they propose distributed priority scheduling, a technique that piggybacks the priority tag of a stations head-of-line packet onto handshake and data packets; e.g., RTS/DATA packets in IEEE 802.11. By monitoring transmitted packets, each station maintains a scheduling table which is used to assess the station's priority level relative to other stations. They then incorporate this scheduling table into existing IEEE 802.11 priority back-off schemes to approximate the idealized schedule. Second, they observe that congestion, link errors, and the random nature of medium access prohibit an exact realization of the ideal schedule. Consequently, they provide a scheduling scheme named multi-hop coordination so that downstream stations can increase a packets relative priority to compensate for excessive delays incurred upstream. From these aspects, the authors develop an analytical model to quantitatively explore these two mechanisms according to the model of Bianchi. In the former mechanism, they study the impact of the probability of overhearing another packets priority index. In the latter mechanism, the proposed analytical model is provided for multi-hop coordination and it is used to compare the probability of meeting an end-to-end delay bound over a multi-hop path with and without coordination.

In [31], the authors develop a model-based frame scheduling scheme, called MFS, to enhance the capacity of IEEE 802.11-operated wireless LANs. In MFS each station estimates the current network status by keeping track of the number of collisions it encounters between its two consecutive successful frame transmissions, and, based on the the estimated information, computes the current network utilization. The result is then used to determine a scheduling delay that is introduced (with the objective of avoiding collision) before a station attempts for transmission of its pending frame. In order to accurately calculate the current utilization in WLANs, they develop an analytical model that characterizes data transmission activities in IEEE 802.11-operated WLANs with/without the RTS/CTS mechanism, and validate the model with ns-2 simulation.

In [7], a number of service differentiation mechanisms have been proposed for, in general, CSMA/CA systems, and, in particular, the 802.11 enhanced DCF. An effective way to provide prioritized service support is to use different inter frame spaces (IFS) for stations belonging to different priority classes. This paper proposes an analytical approach to evaluate throughput and delay performance of IFS based priority mechanisms for different priority classes. This work extends previous work of Bianchi by adding a further state to model different IFS for different priority classes. However, the model does not rely on traditional multi-dimensional Markov chains because the crucial assumption of a constant probability to access the channel in a given time slot is not always correct. For an example, the model fails when the difference between the IFS of two classes is greater than the minimum contention window.

1.3.3 Other Research

In [43], the authors analyze the performance of channel utilization, based on data burst transmissions, supported by the emerging IEEE 802.11e. They develop an analytical framework to evaluate the impact of different access modes (i.e., 2-way/4-way handshaking) and acknowledgment policies (i.e., immediate/block ACK) on the overall system performance. Through the analytical modelling, they show that, given a data packet size and a retransmission limit, the access mode and the ACK policy have a great impact on the overall system throughput, and some optimizations are possible. For example, they show that the block ACK is generally not useful for low data rates and low value for retransmission limit, while it is very attractive for high data rate transmissions. Another interesting conclusion is that the optimal selection between immediate and block ACK does not depend on the number of contending stations. They quantify these comparisons by providing the efficiency thresholds needed to select the best possible mechanism. Finally, they discuss the role of the block ACK protection mechanisms, i.e., of the HOB (head of burst) immediate ACK.

Most of analytical models proposed so far for EEE 802.11 DCF focus on saturation performance. In [41, 30], the authors develop an analytic model for unsaturation throughput evaluation of 802.11 DCF, based on Bianchi's model. The model explicitly takes into account both the carrier sensing mechanism and an additional backoff interval after successful frame transmission, which can be ignored under saturation conditions. Expressions are also derived by means of the equilibrium point analysis in [41]. In [19], a model is proposed to predict the throughput, delay and frame dropping probabilities of the different traffic classes in the range from a lightly loaded, non-saturated channel to a heavily congested, saturated medium. Furthermore, the model describes differentiation based on different AIFS-values (Arbitration

Inter Frame Space), in addition to the other adjustable parameters (i.e. window-sizes, retransmission limits etc.) also encompassed by previous non-saturated models. AIFS differentiation is described by a simple equation that enables access points to determine at which traffic loads starvation of a traffic class will occur.

In [52], the authors propose a new contention algorithm called parallel contention algorithm that divides the subcarriers into multiple groups to reduce the contention time. They analyze the proposed scheme by extending the Markov chain model and verify the accuracy of the analysis through the simulations. The protocol performs well especially when the transmission speed and the number of users are getting higher, thereby achieving a better performance improvement ratio than the original IEEE 802.11a standard.

1.4 Thesis Contributions and Outline

The most important contribution of this thesis is to represent the operation of a system by a state diagram and use the generating function approach to derive its energy and delay consumption. The goal of this thesis is to investigate the energy and delay tradeoff of a wireless communication system. In the first part of this thesis, we will concentrate on communication networks, and in the second part of this thesis, we will study a single link wireless ARQ communication system.

Previous research on performance evaluation of 802.11 has been carried out by two methods. Crow [18], [6] and Weinmiller et al. [48] used computer simulations to evaluate the network throughput. In [23, 17, 11], the system performance was evaluated by an analytical model. Bianchi [5] used a simple but accurate model that characterizes the random exponential backoff protocol. These papers did not incorporate channel noise in the analysis, which is an important factor in wireless network.

Although Hadzi-Velkov and Spasenovski [22] considered the effects of packet errors in the analysis, they did not relate the packet error probability to the energy used and the number of redundant bits used for error control coding. We extend the results from Bianchi [5] and Hadzi-Velkov [22] by considering the effects packet errors in the analysis. In their original work, they did not relate the packet error probability to the energy used and the number of redundant bits used for error control coding. The motivation for this thesis is to understand the role energy and codeword length (number of redundant bits) at the PHY layer have on the total energy and delay of the network. We propose a state diagram representation the operation of the MAC layer and obtain the joint generating function of the energy and delay by incorporating the effects of the PHY layer. Next, we optimize numerically over the code rate for each type of packet to minimize the average transmission delay. We use the random coding bound to represent the packet error probability as a function of the delay and energy. By changing the signal-to-noise ratio, the energy-delay tradeoff curves for minimum delay are obtained. Finally, we propose an approximation method to express the energy-delay tradeoff curves analytically and show the proposed approximation is extremely accurate especially when the number of information bits per packet is large.

Another contribution of this thesis is to apply our proposed generating function method to the analysis and design of other wireless network protocols. The first proposed protocol is an energy adaptation scheme with original IEEE 802.11 protocol in which the transmitter will increase the energy level per coded symbol whenever it suffers an unsuccessful transmission. The numerical results show that the proposed protocol can improve system performance significantly when the channel condition is bad. By using Reed-Solomon codes we can optimize the system performance over

the code rate and energy per coded bit. Although the packet error probabilities are evaluated with Reed-Solomon codes over an additive white Gaussian noise (AWGN) channel, the framework of our analysis can be used for other coding and modulation schemes over various wireless channels. Finally, we also compare the system performance between Reed-Solomon codes and convolutional codes. The second proposed protocol is an ARQ mechanism for data packets transmission with 802.11 protocol. The motivation of this analysis is to demonstrate that the generating function approach can be applied to analyze more function layers jointly by including the analysis of logical link control (LLC) sublayer into the original protocol (MAC and PHY layers only). The numerical results show that the IEEE 802.11 original protocol and the proposed one have almost identical performances and are equally sensitive to the knowledge of the channel quality at the transmitter.

For wireless ARQ systems, we extend the traditional Markov chain model for the channel state as well as the transmitter state [13] by using a state diagram that takes into account the states of transmitter, receiver and channel jointly. The states of the transmitter can be used to model the different packet lengths adopted by the transmitter and the states of the receiver can be utilized to model the receiver memory content [27]. From the system state diagram, we are able to characterize the joint energy and delay distribution of the system incorporating physical layer characteristics (packet error probability as a function of energy and delay) through generating function approach. The effect of transition probability which depends on the packet length is also investigated. As the numerical results demonstrate, the time-varying characteristics of the channel have a great influence on system performance especially at low channel SNR.

The outline of the rest of the thesis is as follows. In Chapter 2, we will use the

proposed generating function method to analyze the energy and delay consumption of wireless networks and discuss the tradeoff between energy and delay. The application of generating function method in designing and analyzing other wireless network protocols are presented in Chapter 3. In Chapter 4, we give the analysis of energy and delay expense for ARQ systems over time varying channels and derive the cutoff rate for different memory receiver structure. Finally in Chapter 5, we briefly summarize the conclusions from the thesis and suggest possible future research directions.

CHAPTER II

Energy-Delay Analysis of MAC Protocols in Wireless Networks

2.1 Introduction

Recently there has been considerable interest in the design and performance evaluation of wireless local area networks (WLANs). Some WLANs must operate solely on battery power. In such cases it is important to consider energy consumption in the system design and analysis. It is possible to reduce energy consumption by increasing delay incurred. Two critical components of a wireless network are the medium access control (MAC) protocol and the physical layer (PHY). The MAC protocol resolves conflicts between users attempting to access the channel. Generally users make reservations for transmissions in a decentralized way. Thus there is some amount of delay in accessing the channel and there is energy used in reserving the channel. An important component of the PHY layer is forward error control coding, which mitigates the effect of channel noise at the receiver. By transmitting redundant bits in addition to information bits, error control coding reduces the energy needed for transmission at the expense of increased delay.

There are many MAC protocols that have been developed for wireless voice and data communication networks. Typical examples include the time-division multiple access (TDMA), code-division multiple access (CDMA), and contention-based

protocols such as IEEE 802.11 [3], [4]. In this paper, we adopt the MAC protocol used in the 802.11 standard. There are two basic techniques to access the medium in the 802.11 standard. The first one called the distributed coordination function (DCF), is employed when there is no centralized coordinator in the system to assign the medium to users in the network. The DCF is a random access scheme, based on carrier sense multiple access with collision avoidance (CSMA/CA). When packets collide or have errors, the transmitter performs a random backoff before retransmitting the packets. Another MAC method in the 802.11 standard called point coordinator function (PCF), is used when there is a centralized coordinator to coordinate the access of the medium. In this paper we focus on the DCF protocol for accessing the medium.

In order to combat the problem of hidden terminals [33], which occurs when some stations in the network are unable to detect each other, the 802.11 protocol uses a mechanism known as request-to-send/clear-to-send (RTS/CTS). Before transmitting the data packet, the source station sends a request-to-send (RTS) packet. If the RTS packet is received correctly at the destination station (there is no collision with other RTS packets sent by other competing stations and the receiver correctly decodes the packet) the destination station broadcasts a clear-to-send (CTS) packet. If the CTS packet is successfully received by the source station, the channel reservation is successful and the source station will begin to send data and wait for the acknowledgement (ACK) packet. The source station detects an unsuccessful channel reservation by the lack of a correct CTS packet.

Many previous research on performance evaluation of 802.11 has been based on an analytical model proposed by Bianchi [5]. Bianchi used a simple but accurate model that characterizes the random exponential backoff protocol. In [22, 46, 32, 24], the

authors used Bianchi's model to perform the cross-layer analysis (PHY and MAC). For example, a theoretical cross-layer saturation goodput model for the IEEE 802.11a PHY and MAC layers was developed in [24]. The proposed analytical approach relates the system performance to channel load, contention window (CW) resolution algorithms, distinct modulation schemes (BPSK and QPSK), FEC schemes (convolutional code), receiver structures (maximum ratio combining) and channel models (uncorrelated Nakagami-m fading channel). In [28, 29, 7, 31], the authors adopted Bianchi's model to perform the priority and scheduling analysis. For example, in [7], a number of service differentiation mechanisms have been designed for CSMA/CA systems, and, in particular, the 802.11 enhanced DCF. The authors proposed an effective way to provide prioritized service support by using different inter frame spaces (IFS) for stations belonging to different priority classes. Although some of the above papers considered the effects of packet errors in the analysis, they did not relate the packet error probability to the energy used and the number of redundant bits used for error control coding. The motivation for this paper is to understand the role of energy and codeword length (number of redundant bits) at the PHY layer have on the total energy and delay of the network. The contributions in this chapter are as follows:

1. We propose a state diagram representing the operation of the MAC layer and obtain the joint generating function of the energy and delay by incorporating the effects of the PHY layer. This is a universal approach and could be applied to other MAC protocols.

2. By taking the partial derivative for the joint generating of the energy and delay, we determine the average energy and delay of a successful packet transmission by taking the packet error probability into consideration. The results obtained

from the generating function approach will be agree with the results derived from the renewal cycle method proposed in [22].

3. We optimize numerically over code rate to have minimum average transmission delay over different packet by introducing the random coding bound to represent the packet error probability. By changing the signal-to-noise ratio, the energy-delay tradeoff curves for minimum delay are obtained.

4. We propose an approximation method to express the energy-delay tradeoff curves analytically. The comparison of the energy-delay tradeoff curves evaluated from this approximation method with the exact energy-delay tradeoff curves (from numerical optimization) indicates that this approximation method is extremely accurate especially when the number of information bits per packet is large.

The remainder of this chapter is organized as follows. In Section 2.2, we give a brief description for the protocol used in our analysis and introduce the system assumptions. In Section 2.3, we discuss our system state diagram and utilize it to derive the joint generating function of energy and delay. Then the average energy with outage delay constraint and average delay with outage energy constraint are analyzed with generating function. The proposed approximation method for energy-delay tradeoff curves is given in Section 2.4. We demonstrate that the energy and delay relationship with random coding under AWGN channel through numerical method in Section 2.5. Finally, Section 2.6 gives the conclusion and future research.

2.2 System Description

The wireless networks that we analyze here have the following network layer specifications. First, each station with a fixed position can hear (detect and decode) the

transmission of $n-1$[1] other stations in the network. Second, stations always have a packet ready to transmit. Third, each station uses the 802.11 MAC protocol. At the PHY layer, a packet of K information bits is encoded into a packet of N coded symbols. It is assumed that the receivers have no multiple-access capability (i.e., they can only receive one packet at a time) and they cannot transmit and receive simultaneously. The packet error probability depends on the parameters K, N, and E_c/N_0, where E_c is the received coded symbol energy and N_0 is the one-sided power spectral density level of the thermal noise at the receiver.

In the following we give a brief description of the most salient features of the IEEE 802.11 MAC protocol (more details can be found in [3] and [4]). When a station is ready to transmit a packet, it senses the channel for DIFS seconds. If the channel is sensed idle, the transmission station picks a random number j, uniformly distributed in $\{0, 1, \ldots, W_i - 1\}$, where $W_i = 2^i W$ is the contention window (CW) size, i is the contention stage (initially $i = 0$), and W is the minimum CW size. A backoff time counter begins to count down with an initial value j: it decreases by one for every idle slot of duration σ seconds (also referred to as the standard slot) as long as the channel is sensed idle, stops the count down when the channel is sensed busy, and reactivates when the channel is sensed idle again. The station transmits an RTS packet when the counter counts down to zero. After transmitting the RTS packet, the station will wait for a CTS packet from the receiving station. If there is a collision of the RTS packet with other competing stations or a transmission error occurs in the RTS or CTS packet, the transmitting station doubles the CW size (increases the contention stage i by one) and picks another random number j as before. If there are no collisions or errors in the RTS and CTS packets, the station

[1] $n-1$ is the number of stations in the communication range of the reference station.

begins to transmit the data packet and waits for an acknowledgment (ACK) packet. However, if the data or the ACK packet is not successfully received, the CW size will also be doubled (the contention stage i will increase by one) and the transmitting station will join the contention period again. The contention stage is reset (i is set to zero) when the transmitting station receives an ACK correctly. A time diagram indicating the sequence of these events is depicted in Fig. 2.1. It is also noted that there is a maximum CW size (or equivalently, a maximum contention stage, m); when the transmitter is in this maximum stage and needs to join the contention period again, it does not increase further the CW, but picks a random number in $\{0, 1, \ldots, W_m - 1\}$.

2.3 Energy-Delay Analysis

In this section, we analyze the energy and delay characteristics of the wireless networks described above. The delay T_d of each data packet is defined as the time duration from the moment the backoff procedure is initiated until DIFS seconds after the ACK packet is received correctly by the transmitting station, as shown in Fig. 2.1. Similarly, the energy E_t is defined as the energy consumed by both transmitting and receiving stations in the duration of T_d. Without loss of the generality, for notational simplicity we assume that the propagation loss between transmitter and receiver is one (0 dB). We also assume that the propagation time is negligible. In this chapter, we only consider the energy consumption for packet transmission and omit the energy required for signal processing and channel sensing. The system parameter SIFS is defined as the time between the end of a packet reception, say RTS and the beginning of a packet transmission, say CTS. This time includes the time required for decoding a packet and other processing functions at the receiver.

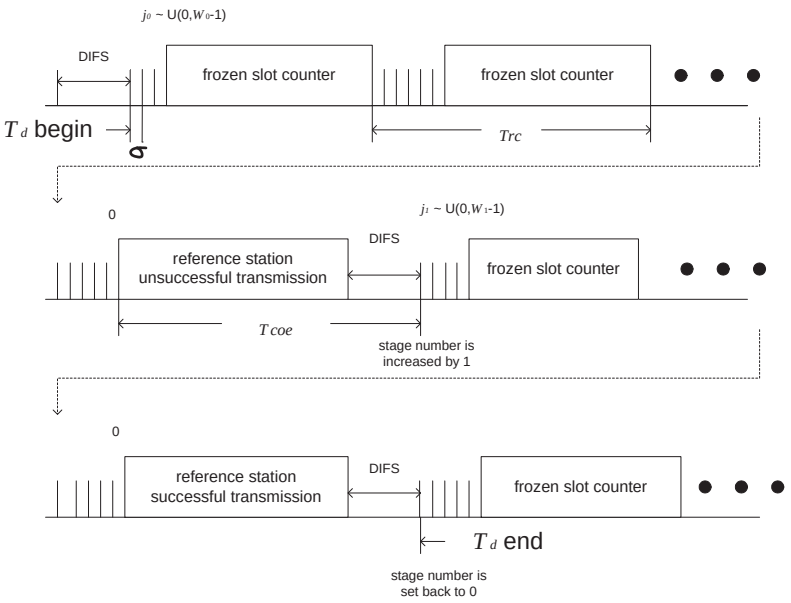

Figure 2.1: Timing diagram for the protocol under investigation. The numbers j_0 and j_1 represent random numbers at stage $i = 0$, and $i = 1$, respectively.

2.3.1 Three Nonlinear System Equations

In order to analyze energy and delay relationships, we need to define two random processes to characterize the backoff counter state and the CW size. The first process $b(\tau)$ represents the backoff time counter for the reference station (this is the station for which we evaluate energy and delay). The second random process $s(\tau)$ is used to represent the CW stage $i \in \{0, 1, \ldots, m\}$ of the station at time τ. In order to analyze the energy and delay characteristics it is sufficient to only consider the time instances the backoff counter (and CW stage) changes value. To this end we further define the discrete-time random processes $b_t = b(\tau_t)$ and $s_t = s(\tau_t)$, where τ_t is the time instance of the t-th change in value of $b(\tau)$. A realization of these random processes is shown in Fig. 2.2. In this realization, $b_3 = 4$ is the value of the counter just before being frozen and $b_4 = 3$ is the value of the counter after the channel has been sensed idle for DIFS seconds, and the counter becomes active again.

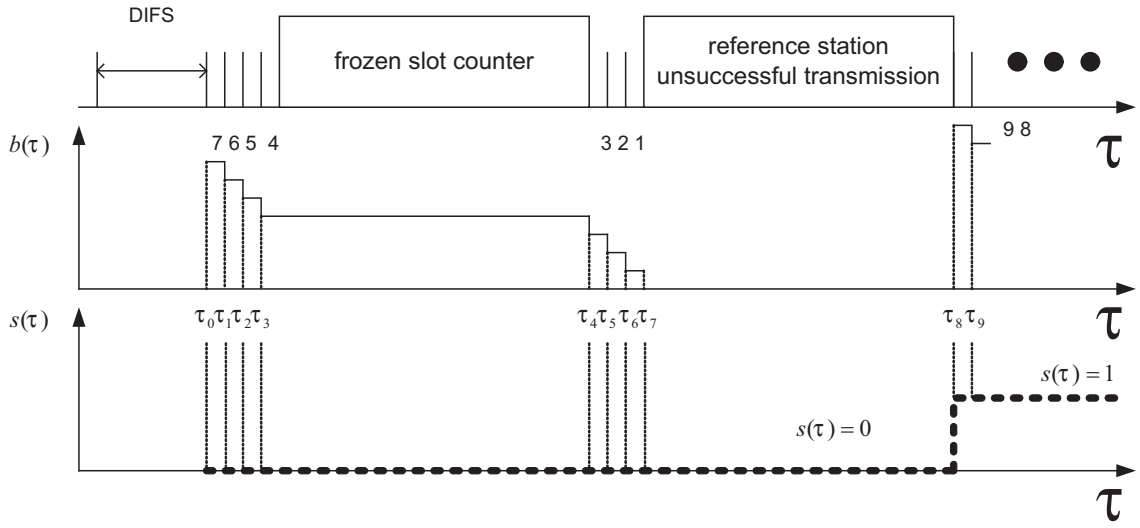

Figure 2.2: Illustration of the random processes, $b(\tau)$, $s(\tau)$ and their discrete-time counterparts b_t, s_t.

The first assumption made in this analysis is that the event of packet collision is independent of past collisions and thus independent of the contention stage. The second assumption is that the packet collision is identical for all states (values of b_t and s_t) of a user. As verified in [5], these two assumptions are extremely accurate when the number of stations in the network is large (say greater then 10). As a result of the above assumptions, the two dimensional process (s_t, b_t) is a discrete-time Markov chain. With the assumption that errors in transmission can occur only due to collisions, the one-step transition probabilities developed in [5] are given by $P\{i,k \mid j,l\} = P\{s_{t+1} = i, b_{t+1} = k \mid s_t = j, b_t = l\}$ with

$$P\{i,k \mid i,k+1\} = \quad 1, \quad 0 \le k \le W_i - 2, 0 \le i \le m \tag{2.1a}$$

$$P\{0,k \mid i,0\} = \quad \tfrac{1-p_c}{W_0}, \quad 0 \le k \le W_0 - 1, 0 \le i \le m \tag{2.1b}$$

$$P\{i,k \mid i-1,0\} = \quad \tfrac{p_c}{W_i}, \quad 0 \le k \le W_i - 1, 1 \le i \le m \tag{2.1c}$$

$$P\{m,k \mid m,0\} = \quad \tfrac{p_c}{W_m}, \quad 0 \le k \le W_m - 1, \tag{2.1d}$$

where p_c is the conditional collision probability, i.e., the probability of a collision given a packet transmitted on the channel. The first equation in (3.1) corresponds to the decrement of the backoff counter at the beginning of each time slot. The second equation accounts for the fact that a new packet following a successful packet transmission starts at contention stage $i = 0$, and thus the backoff counter is initially uniformly chosen in the range of $\{0, 1, \ldots, W_0 - 1\}$. The other two cases describe the system evolution after an unsuccessful transmission. As described in the third equation, when an unsuccessful transmission occurs at contention stage $i - 1$, the contention stage increases and the backoff counter is initialized with a uniformly chosen value in the range $\{0, 1, \ldots, W_i\}$. Finally, the last case models the fact that the contention stage is not increased in subsequent packet transmissions when the contention window size reaches the maximum.

By modifying the Markov chain model described above, we can take into account packet errors as shown in Fig. 2.3. We denote the error probability of the four kinds of packets in the system as $P_{e,RTS}$, $P_{e,CTS}$, $P_{e,DT}$ and $P_{e,ACK}$. We assume that the channel is memoryless between packets. These probabilities depend on the particular channel, coding, modulation etc (a specific example will be given in Section 2.5). A successful packet transmission requires that the RTS, CTS, DT, and ACK packets are received correctly. Let p_{ce} denote the probability of the complement of this event, i.e., collision in the RTS packet or error in any of the packets. This is also the transition probability from one contention stage to the next in the two-dimensional Markov chain, as shown in Fig. 2.3. Using similar assumptions as in [5] for the packet collision probability p_c, and since the events of packet collision and packet error are

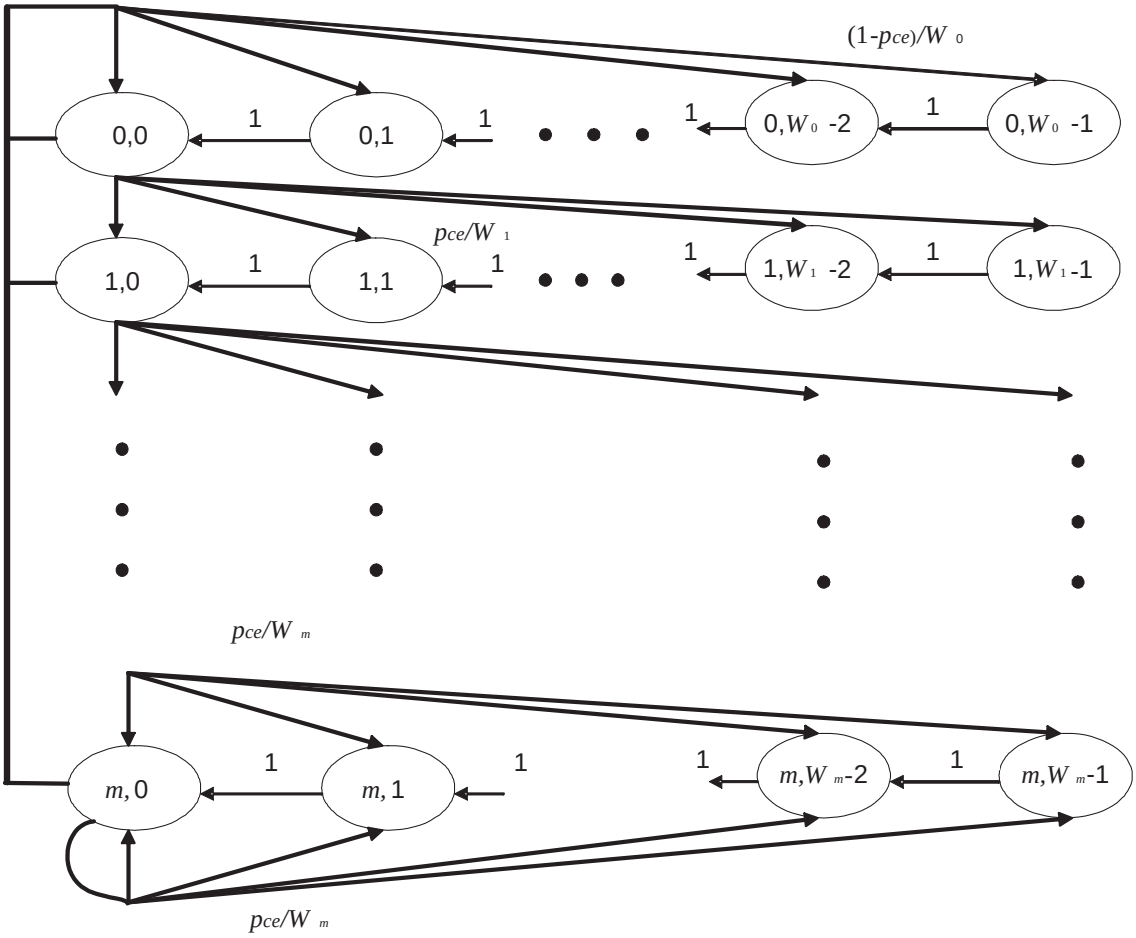

Figure 2.3: Markov chain for backoff counter and contention window stage.

independent, the probability of p_{ce} can be expressed as

$$\begin{aligned} p_{ce} &= p_c + (1-p_c)[P_{e,RTS} + (1-P_{e,RTS})P_{e,CTS} + (1-P_{e,RTS})(1-P_{e,CTS})P_{e,DT} \\ &\quad +(1-P_{e,RTS})(1-P_{e,CTS})(1-P_{e,DT})P_{e,ACK}]. \end{aligned} \tag{2.2}$$

Following the derivation in [5], we can evaluate the stationary probability $P(i,j)$ of each state (i,j) of the Markov chain. Let p_{tx} be the probability of a transmitting station sending an RTS packet during each backoff slots. The transmission probability p_{tx} and the collision probability p_c can be related as in [5]

$$p_{tx} \triangleq \sum_{i=0}^{m} P(i,0) = \frac{2(1-2p_{ce})}{(1-2p_{ce})(W+1) + p_{ce}W(1-(2p_{ce})^m)} \tag{2.3}$$

$$p_c = 1-(1-p_{tx})^{n-1}. \tag{2.4}$$

From the above three nonlinear equations, (2.2)–(2.4), the probabilities p_{ce}, p_{tx} and p_c can be evaluated numerically. Another important probability that will be used in our analysis later is the probability of a transmission of an RTS packet from exactly one of the remaining $n-1$ stations given that at least one of the remaining stations is transmitting. It is denoted by p_{tx_1} and can be expressed as

$$p_{tx_1} = \frac{(n-1)p_{tx}(1-p_{tx})^{n-2}}{p_c}. \tag{2.5}$$

2.3.2 Joint Generating Function of Energy and Delay

Our goal in this section is to obtain the joint generating function of energy and delay for a successful data packet transmission, which can be expressed as

$$G_s(X,Y) = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} Pr(T_d = i\Delta_t, E_t = j\Delta_e)X^iY^j, \tag{2.6}$$

where Δ_t, Δ_e are parameters that determine the resolution of our analysis (we choose $\Delta_t = T_b$ and $\Delta_e = E_c$ in the remaining part of our analysis). We further define the

quantities $N_{SIFS} = T_{SIFS}/\Delta_t$, $N_{DIFS} = T_{DIFS}/\Delta_t$, and $N_\sigma = \sigma/\Delta_t$, and make the additional assumption that N_{SIFS}, N_{DIFS} and N_σ are integers.

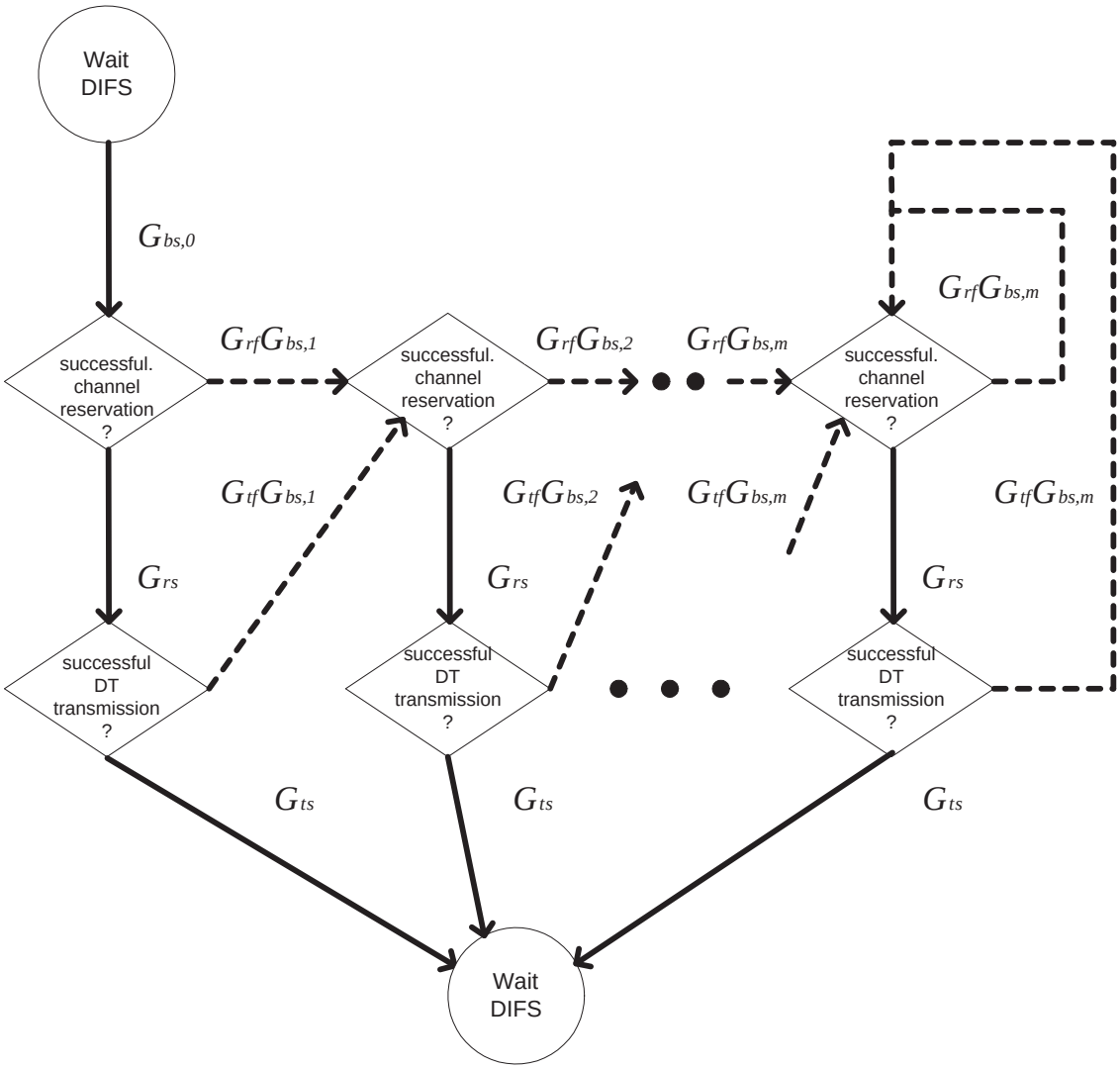

Figure 2.4: State diagram representation of the 802.11 MAC protocol. Transform variables X and Y are omitted for simplicity.

Based on the protocol description in Section 2.2, the state flow diagram shown in Fig. 2.4 can be obtained. Each transition from state A to state B in this diagram is labelled with the conditional joint generating function of the additional energy and delay incurred when the protocol makes a transition from state A to state B. For

instance, the generating function G_{rs} corresponding to a channel reservation success is

$$G_{rs}(X,Y) = (1-p_c)(1-P_{e,RTS})(1-P_{e,CTS})X^{N_{RTS}+N_{CTS}+2N_{SIFS}}Y^{N_{RTS}+N_{CTS}}, \tag{2.7}$$

where $(1-p_c)(1-P_{e,RTS})(1-P_{e,CTS})$ is the probability of not having a collision and having a correct reception of the RTS and CTS packets, $[N_{RTS}+N_{CTS}+N_{SIFS}+T_{DIFS}]\times T_b$ is the delay incurred during this transition, and $[N_{RTS}+N_{CTS}]\times E_c$ is the corresponding energy consumed. Similarly, the generating function G_{ts} corresponding to a data packet transmission success once the channel is reserved can be expressed as

$$G_{ts}(X,Y) = (1-P_{e,DT})(1-P_{e,ACK})X^{N_{DT}+N_{ACK}+N_{SIFS}+N_{DIFS}}Y^{N_{DT}+N_{ACK}}. \tag{2.8}$$

The generating functions associated with failure to either reserve a channel or to transmit a data packet are products of two generating functions. Each product contains a factor that is independent of the particular CW stage, and a factor that depends on the contention stage i. We first describe the factors that are independent of the CW stage. The generating function G_{rf}, corresponding to a channel reservation failure is given by

$$G_{rf}(X,Y) = [p_c+(1-p_c)P_{e,RTS}]X^{N_{RTS}+N_{DIFS}}Y^{N_{RTS}} + (1-p_c)(1-P_{e,RTS})P_{e,CTS}\cdot X^{N_{RTS}+N_{CTS}+N_{SIFS}+N_{DIFS}}Y^{N_{RTS}+N_{CTS}}, \tag{2.9}$$

the meaning of which is that a failure can be due to either a collision/RTS transmission error, or a CTS transmission error. Similarly, after reserving the channel, a data packet transmission failure is either due to a data packet transmission error, or

an acknowledgement packet transmission error, which is captured by the generating function G_{tf} as follows

$$G_{tf}(X,Y) = P_{e,DT}X^{N_{DT}+N_{DIFS}}Y^{N_{DT}} + (1-P_{e,DT})P_{e,ACK} \cdot X^{N_{DT}+N_{ACK}+N_{SIFS}+N_{DIFS}}Y^{N_{DT}+N_{ACK}}. \quad (2.10)$$

We now evaluate the generating functions denoted by $G_{bs,i}$ of the state diagram. This generating function characterizes the delay for the transmitting station from the instant of starting the backoff procedure to the instant that the backoff counter reaches to zero at stage i. We do not need to consider energy consumption here since the reference station is not transmitting during this period. Thus, $G_{bs,i}$ is not a function of the variable Y. The probability of a busy slot due to the transmission of other stations is p_c and this event is independent and identical for each slot from our previous assumptions. At contention stage i, the range of possible backoff slots is from 1 to 2^iW. Let j be the backoff slot chosen uniformly from the above range. The number of the occupied slots in these j slots is binomialy distributed with parameters (j, p_c). Hence, the generating function, $G_{bs,i}$, is

$$G_{bs,i}(X) = \sum_{j=1}^{2^iW} \frac{1}{2^iW} \sum_{k=0}^{j} \binom{j}{k} [(1-p_c)X^{N_\sigma}]^{j-k}(p_c G_{oc}(X))^k. \quad (2.11)$$

where $G_{oc}(X)$ is the generating function of the delay due to an occupied slot. We define an occupied slot as a slot when the transmitting station senses the channel is busy due to the transmission of one of the $n-1$ remaining stations. In an occupied slot, there are four possible cases. The first case is when two or more RTS packets from the remaining $n-1$ stations collide or have a packet error. The second case is when the RTS packet is correctly received without collision but there is an error in the CTS packet transmission. The third case is when the RTS packet is correctly received without collision and there is no error in the CTS packet but there is an error

in the DT packet transmission. The last case is when there is a correct reception of RTS, CTS, and DT and either correct or erroneous reception of the ACK packet. Note that the time duration of successful or unsuccessful ACK transmission are the same. Therefore, the corresponding generating function is

$$\begin{aligned} G_{oc}(X) &= [(1-p_{tx_1})+p_{tx_1}P_{e,RTS}]X^{N_{RTS}+N_{DIFS}} \\ &\quad +p_{tx_1}(1-P_{e,RTS})P_{e,CTS}X^{N_{RTS}+N_{CTS}+N_{SIFS}+N_{DIFS}} \\ &\quad +p_{tx_1}(1-P_{e,RTS})(1-P_{e,CTS})P_{e,DT}X^{N_{RTS}+N_{CTS}+N_{DT}+N_{SIFS}+N_{DIFS}} \\ &\quad +p_{tx_1}(1-P_{e,RTS})(1-P_{e,CTS})(1-P_{e,DT})\cdot \\ &\quad X^{N_{RTS}+N_{CTS}+N_{DT}+N_{ACK}+N_{SIFS}+N_{DIFS}}. \end{aligned} \tag{2.12}$$

From the state diagram and Mason's gain formula [20], we obtain the following backward recursive equations for the generating function G_s.

$$\begin{aligned} G_s(X,Y) &= G_{bs,0}\Big\{1+\sum_{j=1}^{m-1}[G_{rf}+G_{rs}G_{tf}]^j\left(\prod_{i=1}^{j}G_{bs,i}\right)+ \\ &\quad \frac{[G_{rf}+G_{rs}G_{tf}]^m}{1-[G_{rf}+G_{rs}G_{tf}]G_{bs,m}}\left(\prod_{i=1}^{m}G_{bs,i}\right)\Big\}G_{rs}G_{ts}. \end{aligned} \tag{2.13}$$

where i $(1 \le i \le m)$ is the index of the CW stage.

From the joint generating function, the average energy and delay can be easily evaluated as

$$\overline{T}_d = \Delta_t \frac{\partial G_s}{\partial X}\mid_{X=Y=1}, \tag{2.14a}$$

$$\overline{E}_t = \Delta_e \frac{\partial G_s}{\partial Y}\mid_{X=Y=1}. \tag{2.14b}$$

Below, we will utilize the generating function to derive the average energy with an outage delay constraint.

2.3.3 Mean System Energy Consumption and Delay

In this section, we analyze average energy and delay of our system. The average delay analysis is presented first. In each CW stage $i \in \{0, 1, 2, \ldots, m\}$, the initial value of the backoff counter has mean of $(W_i + 1)/2$, so that the average number of deferred slots (include standard slots and interrupted slots) before a retransmission attempt is $(W_i+1)/2$. The average number of consecutive standard slots n_s between two consecutive interrupted slots due to the transmission of the remaining $n-1$ stations can be evaluated as

$$n_s = \sum_{i=0}^{\infty} i(1-p_c)^i p_c = \frac{1}{p_c} - 1. \tag{2.15}$$

We define a renewal cycle as the period between two consecutive transmission of the $n-1$ remaining stations including multiple consecutive standard slots and an interrupted slot as shown in Fig. 2.1. There are four possible cases in an interrupted slot. The first case is that the RTS packet from the remaining $n-1$ stations suffers collision or packet error. The second case is that the RTS packet from the remaining $n-1$ stations is correctly received and collision free but there is an error in the CTS packet transmission. The third case is that the RTS packet from the remaining $n-1$ stations is correctly received and no collision occurs, moreover there is no error in the CTS packet. However, there is an error happened in the DT packet transmission. The last case is to extend case three with a correct DT packet transmission so there is an ACK packet transmission. The average duration of a renewal cycle can be expressed as

$$\begin{aligned} T_{rc} = {} & n_s\sigma + [(1-p_{tx_1}) + p_{tx_1}P_{e,RTS})]T_{s,RTS} \\ & + [p_{tx_1}(1-P_{e,RTS})P_{e,CTS}]T_{s,CTS} \\ & + [p_{tx_1}(1-P_{e,RTS})(1-P_{e,CTS})P_{e,DT}]T_{s,DT} \\ & + [p_{tx_1}(1-P_{e,RTS})(1-P_{e,CTS})(1-P_{e,DT})]T_{s,ACK}, \end{aligned} \tag{2.16}$$

and

$$T_{s,RTS} = N_{RTS}T_b + T_{DIFS} \quad (2.17a)$$

$$T_{s,CTS} = (N_{RTS} + N_{CTS})T_b + T_{SIFS} + T_{DIFS} \quad (2.17b)$$

$$T_{s,DT} = (N_{RTS} + N_{CTS} + N_{DT})T_b + 2T_{SIFS} + T_{DIFS} \quad (2.17c)$$

$$T_{s,ACK} = (N_{RTS} + N_{CTS} + N_{DT} + N_{ACK})T_b + 3T_{SIFS} + T_{DIFS}, \quad (2.17d)$$

where T_b is the time duration to transmit each coded bit, and T_{DIFS}, T_{SIFS} are system delay parameters defined by the standard.

Since the retransmission attempt of a reference station in contention stage i is on the average preceded by $(W_i + 1)/[2(n_s + 1)] = (W_i + 1)p_c/2$ renewal cycles of $n - 1$ remaining stations, the average time duration between two consecutive retransmissions of the reference station is $(W_i + 1)p_cT_{rc}/2$. If the reference station fails to receive the ACK packet correctly, i.e., it is not a successful transmission, there are four possible cases of unsuccessful transmission. The previous three cases are the same as the first three cases of the above interrupted slot. But the last case will be modified as that the reference station receives ACK packet incorrectly. The average duration while the reference station itself occupies the channel during each unsuccessful retransmission attempt, denoted as T_{coe}, is

$$\begin{aligned} T_{coe} = \frac{1}{p_{ce}}[&(p_c + (1 - p_c)P_{e,RTS})T_{s,RTS} \\ &+(1 - p_c)(1 - P_{e,RTS})P_{e,CTS}T_{s,CTS} \\ &+(1 - p_c)(1 - P_{e,RTS})(1 - P_{e,CTS})P_{e,DT}T_{s,DT} \\ &+(1 - p_c)(1 - P_{e,RTS})(1 - P_{e,CTS})(1 - P_{e,DT})P_{e,ACK}T_{s,ACK}]. \end{aligned} \quad (2.18)$$

The average time before the reference station makes its $(i + 1)$-th retransmission

attempt, denoted as $T_{r,i}$, is

$$T_{r,i} = \sum_{k=0}^{i} \Big[\frac{(W_k+1)}{2} p_c T_{rc}\Big] + iT_{coe} \tag{2.19a}$$

$$= \begin{cases} iT_{coe} + (i+1)\frac{p_c T_{rc}}{2} + W(2^{i+1}-1)\frac{p_c T_{rc}}{2} & \text{for } 0 \le i \le m-1 \\ iT_{coe} + (i+1)\frac{p_c T_{rc}}{2} + W(2^{m+1}-1+2^m(i-m))\frac{p_c T_{rc}}{2} & \text{for } m \le i. \end{cases} \tag{2.19b}$$

Finally, the average delay of a successful data packet transmission can be obtained from (2.14) by taking the partial derivative of $G_s(X,Y)$ with respect to X. It is expressed with $T_{s,ACK}$, T_{coe} and T_{rc} as

$$\overline{T}_d = T_{s,ACK} + \sum_{i=0}^{\infty}(1-p_{ce})p_{ce}^i T_{r,i} \tag{2.20a}$$

$$\begin{aligned} &= T_{s,ACK} + T_{coe}\frac{p_{ce}}{1-p_{ce}} \\ &\quad + T_{rc}\Big[\frac{p_c W}{2}\frac{1-2(2p_{ce})^m(1-p_{ce})+p_{ce}^m(1-2p_{ce})}{1-2p_{ce}} + \\ &\quad \frac{p_c W}{2}(2^{m+1}-1+\frac{2^m p_{ce}}{1-p_{ce}})p_{ce}^m + \frac{p_c}{2(1-p_{ce})}\Big]. \end{aligned} \tag{2.20b}$$

The average transmission energy E_t of each data packet is the total energy consumed in the transmitting and receiving stations from the moment the backoff procedure is initiated until the moment the ACK packet is received by the transmitting station correctly. We will consider only the energy spent in packet transmission and assume that the energy consumption for each packet transmission is NE_c for a packet of length N. The four possibilities of consuming energy when the reference station fails to receive the ACK packet correctly are the same in the previous delay analysis. The average energy consumption can be expressed as

$$\begin{aligned} E_{coe} &= \frac{1}{p_{ce}}[(p_c + (1-p_c)P_{e,RTS})E_{s,RTS} \\ &\quad + (1-p_c)(1-P_{e,RTS})P_{e,CTS}E_{s,CTS} \\ &\quad + (1-p_c)(1-P_{e,RTS})(1-P_{e,CTS})P_{e,DT}E_{s,DT} \\ &\quad + (1-p_c)(1-P_{e,RTS})(1-P_{e,CTS})(1-P_{e,DT})P_{e,ACK}E_{s,ACK}], \end{aligned} \tag{2.21}$$

where

$$E_{s,RTS} = N_{RTS}E_c \tag{2.22a}$$

$$E_{s,CTS} = (N_{RTS} + N_{CTS})E_c \tag{2.22b}$$

$$E_{s,DT} = (N_{RTS} + N_{CTS} + N_{DT})E_c \tag{2.22c}$$

$$E_{s,ACK} = (N_{RTS} + N_{CTS} + N_{DT} + N_{ACK})E_c. \tag{2.22d}$$

In the above equations, $E_{s,RTS}$, $E_{s,CTS}$, $E_{s,DT}$, and $E_{s,ACK}$ denote the energy consumed for the four above mentioned cases. Hence, the average transmitting energy for a successful data packet transmission can also be derived from (2.14) by taking the partial derivative of $G_s(X, Y)$ with respect to Y. It can be written with $E_{s,ACK}$ and E_{coe} as

$$\overline{E}_t = E_{s,ACK} + \sum_{i=0}^{\infty} i(1 - p_{ce})p_{ce}^i E_{coe} \tag{2.23a}$$

$$= E_{s,ACK} + E_{coe}\frac{p_{ce}}{1 - p_{ce}}. \tag{2.23b}$$

2.3.4 Average Energy with Delay Constraint

We define the transform, denoted X^+, for a sequence of numbers $(n_0, n_1, n_2, \cdots)$ by

$$X^+(x) \triangleq \sum_{k=1}^{\infty} n_k x^k. \tag{2.24}$$

The average energy with delay constraint γ_d is calculated as follows. Let $n_d = \gamma_d/\Delta_t$ be the normalized delay constraint (which we assume is an integer). Then

$$\begin{aligned}
\mathbf{E}[E_t \mid T_d \le \gamma_d] &= \sum_{j=0}^{\infty} j\Delta_e Pr(E_t = j\Delta_e \mid T_d \le \gamma_d) && (2.25a)\\
&= \Delta_e \frac{\sum_{j=0}^{\infty} j \sum_{i=0}^{n_d} Pr(T_d = i\Delta_t, E_t = j\Delta_e)}{Pr(T_d \le \gamma_d)} && (2.25b)\\
&= \Delta_e \frac{\sum_{i=0}^{n_d} \sum_{j=0}^{\infty} jPr(T_d = i\Delta_t, E_t = j\Delta_e)}{Pr(T_d \le \gamma_d)} && (2.25c)\\
&= \Delta_e \frac{\sum_{i=0}^{n_d} \sum_{j=0}^{\infty} jPr(T_d = i\Delta_t, E_t = j\Delta_e)}{\sum_{i=0}^{n_d} Pr(T_d = i\Delta_t)} && (2.25d)
\end{aligned}$$

In order to evaluate the denominator of (2.25) we proceed as follows. We let $s_l = \sum_{i=0}^{l} Pr(T_d = i\Delta_t)$ and $S(X) = \sum_{l=0}^{\infty} s_l X^l$. Then we have

$$s_l - s_{l-1} = \sum_{j=0}^{\infty} Pr(T_d = l\Delta_t, E_t = j\Delta_e), \qquad l = 1, 2, \ldots. \tag{2.26}$$

Multiplying both sides by X^l, summing over l, and taking into account that $s_0 = \sum_{j=0}^{\infty} Pr(T_d = 0\Delta_t, E_t = j\Delta_e)$, we have

$$\begin{aligned}
S(X)(1-X) &= \sum_{l=0}^{\infty}\sum_{j=0}^{\infty} Pr(T_d = l\Delta_t, E_t = j\Delta_e)X^l\\
&= G_s(X,Y)\mid_{Y=1}. && (2.27)
\end{aligned}$$

Thus,

$$S(X) = \frac{G_s(X,Y)\mid_{Y=1}}{1-X}. \tag{2.28}$$

Therefore, the denominator of (2.25) is the coefficient of term in $S(X)$ that has degree equal to n_d and can be found by inverse transforming X^+.

The numerator of (2.25) can also be expressed with system generating function G_s. By setting $u_l = \sum_{j=0}^{\infty} jPr(T_d = l\Delta_t, E_t = j\Delta_e)$ and taking the transform of $\{u_l\}$ with dummy variable X, we get

$$U(X) = \frac{\partial G_s(X,Y)}{\partial Y}\mid_{Y=1}. \tag{2.29}$$

Following a similar procedure as in (2.26) and (2.27), the numerator becomes

$$\sum_{i=0}^{n_d} u_i = \Delta_e \frac{\frac{\partial G_s(X,Y)}{\partial Y} \mid_{Y=1}}{1-X}. \tag{2.30}$$

Finally, $\mathbf{E}[E_t \mid T_d \leq \gamma_d]$ can be written as

$$\mathbf{E}[E_t \mid T_d \leq \gamma_d] \;=\; \Delta_e \frac{\text{coef}(\frac{\frac{\partial G_s(X,Y)}{\partial Y}|_{Y=1}}{1-X}, X^{n_d})}{\text{coef}(\frac{G_s(X,Y)|_{Y=1}}{1-X}, X^{n_d})}, \tag{2.31}$$

where $\text{coef}(f(X), X^i)$ is the coefficient of the term X^i in the polynomial expansion of $f(X)$. By a similar approach as above, we can also determine the average delay with an energy constraint. The quantity $\mathbf{E}[T_d \mid E_t \leq n_e\Delta_e]$ can be written as

$$\mathbf{E}[T_d \mid E_t < n_e\Delta_e] \;=\; \Delta_t \frac{\text{coef}(\frac{\frac{\partial G_s(X,Y)}{\partial X}|_{X=1}}{1-Y}, Y^{n_e})}{\text{coef}(\frac{G_s(X,Y)|_{X=1}}{1-Y}, Y^{n_e})}. \tag{2.32}$$

2.4 Energy-Delay Optimization

In order to evaluate the energy and delay characteristics, we need to have a relation between error probability, energy and delay for each type of packet in the system. We will use the reliability function bounds for a memoryless channel to determine the packet error probability. Let K be the number of information bits in a packet (specified by the 802.11 standard) and N be the number of coded symbols for this packet. Then there exist an encoder and decoder for which the packet error probability $P_{e,K,N}$ is bounded by

$$P_{e,K,N} \leq 2^{K-NR_0}, \tag{2.33}$$

where R_0 is the cutoff rate. For an additive white Gaussian noise (AWGN) channel using binary input the cutoff rate is $R_0 = 1 - \log_2(1 + e^{-E_c/N_0})$, where E_c/N_0 is the received signal-to-noise ratio per coded symbol. An interesting observation is that there is an optimal N for minimum delay. For small N the packet error probability of

a random code will be large. Large packet error probability will increase the system delay and energy due to the high chance of packet retransmission. On the other hand, if N is large, the packet error probability for a random code will be small but the transmission time will be large. Our goal is to find the best N for each kind of packet to minimize system delay. To achieve this, we fix E_c/N_0, and optimize delay (average or energy constrained) over the corresponding packet lengths N_{RTS}, N_{CTS}, N_{DT} and N_{ACK}. Based on these optimal values, we evaluate the energy consumption (average or delay constrained). Repeating the above procedure for different E_c/N_0, we get a curve quantifying the tradeoff between energy and delay. For the case when the number of information bits is large, we propose an approximate method to determine the energy and delay tradeoff curves mentioned above. The purpose to provide the approximation method to determine the energy and delay curve is that it can reduce the time dramatically in evaluating the energy and delay curve. The numerical results of energy and delay curves will be given in Section 2.5. In this section, we will concentrate on deriving the approximate equations for energy and delay tradeoff curves.

The first step of this approximation is to derive the average delay $\overline{T}_d$ when there is only one transmitting station in the network. Some approximations are made to obtain the estimates of the packet error probabilities that achieve minimum average delay, $\overline{T}_d$. We then evaluate the corresponding packet lengths by solving (2.33) for N (used as an equality) with the optimal packet error probabilities. After getting the optimal packet error probabilities and lengths, the average energy $\overline{E}_t$ and delay $\overline{T}_d$ are evaluated according to Section 2.3.1 and 2.3.3.

We begin by considering the average delay with only one transmitting station that uses the same protocol as described in Section 2.2. This means that there are

no RTS packet collisions in the medium. We need to modify the analysis for the single station case by excluding the effects of packet collision. Therefore, the packet collision probability p_c becomes zero. We still use p_{ce} to represent the probability from one stage to next. An unsuccessful transmission can happen only due to packet errors. Now the probability p_{ce} can be rewritten as

$$\begin{aligned} p_{ce} &= P_{e,RTS} + (1 - P_{e,RTS})P_{e,CTS} + (1 - P_{e,RTS})(1 - P_{e,CTS})P_{e,DT} \\ &\quad +(1 - P_{e,RTS})(1 - P_{e,CTS})(1 - P_{e,DT})P_{e,ACK}. \end{aligned} \tag{2.34}$$

There are four possible cases of an unsuccessful transmission. The first case is packet error of the RTS packet. The second case is that the RTS packet is correctly received but there is an error in the CTS packet transmission. The third case is that the RTS packet is correctly received, there is no error in the CTS packet, but, there is an error happened in the DT packet transmission. The last case is that the RTS, CTS and DT packets are successful but there is an error at the ACK packet. The time duration of each above cases is $T_{s,RTS}, T_{s,CTS}, T_{s,DT}$ and $T_{s,ACK}$. The average duration while the reference station itself occupies the channel during each unsuccessful retransmission attempt, denoted as T_{coe}, is

$$\begin{aligned} T_{coe} &= \frac{1}{p_{ce}}[P_{e,RTS}T_{s,RTS} + (1 - P_{e,RTS})P_{e,CTS}T_{s,CTS} \\ &\quad +(1 - P_{e,RTS})(1 - P_{e,CTS})P_{e,DT}T_{s,DT} \\ &\quad +(1 - P_{e,RTS})(1 - P_{e,CTS})(1 - P_{e,DT})P_{e,ACK}T_{s,ACK}]. \end{aligned} \tag{2.35}$$

The average elapsed time $T_{r,i}$ before the reference station makes its $(i+1)th$ retrans-

mission attempt is

$$\begin{aligned} T_{r,i} &= (\sum_{k=0}^{i} \frac{W_k \sigma}{2}) + iT_{coe} \\ &= \begin{cases} iT_{coe} + \frac{W\sigma(2^{i+1}-1)}{2} & \text{for } 0 \le i \le m-1 \\ iT_{coe} + \frac{W\sigma(2^{m+1}-1+2^m(i-m))}{2} & \text{for } m \le i. \end{cases} \end{aligned} \tag{2.36}$$

Observed that (2.36) is not derived from (2.19) by substituting $p_c = 0$. The reason is that (2.19) is derived under the assumption of large n while (2.36) is derived separately for $n = 1$ since the large n approximation does not hold. Finally, the average system delay for successful transmission is

$$\begin{aligned} \overline{T}_d &= T_{s,ACK} + \sum_{i=0}^{\infty} (1-p_{ce}) p_{ce}^i T_{r,i} \\ &= T_{s,ACK} + T_{coe} \frac{p_{ce}}{1-p_{ce}} + \frac{W\sigma}{2} \Big[\frac{1 - 2(2p_{ce})^m (1-p_{ce}) + p_{ce}^m (1-2p_{ce})}{1-2p_{ce}} + \\ &\quad (2^{m+1} - 1 + \frac{2^m p_{ce}}{1-p_{ce}}) p_{ce}^m \Big]. \end{aligned} \tag{2.37}$$

If we assume that the optimal error probability of each packet is quite small (i.e., $P_{e,RTS} \ll 1$, $P_{e,CTS} \ll 1$, $P_{e,DT} \ll 1$ and $P_{e,ACK} \ll 1$), then p_{ce} will also be very small compared to 1, as can be seen from the (2.34). The validity of this assumption will be verified later. The average system delay in (2.37) can be approximated by keeping the first order terms in P_{ce} as

$$\overline{T}_d \approx \widetilde{T}_d = T_{s,ACK} + T_{coe} \frac{p_{ce}}{1-p_{ce}} + \frac{W\sigma}{2}(1 + 2p_{ce}). \tag{2.38}$$

Using the assumption that the number of information bits K is large for each kind of packet, T_{SIFS} and T_{DIFS} can be ignored in the evaluation of $T_{s,ACK}$ and T_{coe}, resulting in

$$\begin{aligned} T_{s,ACK} &= (N_{RTS} + N_{CTS} + N_{DT} + N_{ACK}) T_b + 3T_{SIFS} + T_{DIFS} \\ &\approx (N_{RTS} + N_{CTS} + N_{DT} + N_{ACK}) T_b, \end{aligned} \tag{2.39}$$

and

$$
\begin{aligned}
T_{coe} &= \frac{1}{p_{ce}}[P_{e,RTS}T_{s,RTS} + (1-P_{e,RTS})P_{e,CTS}T_{s,CTS} \\
&\quad +(1-P_{e,RTS})(1-P_{e,CTS})P_{e,DT}T_{s,DT} \\
&\quad +(1-P_{e,RTS})(1-P_{e,CTS})(1-P_{e,DT})P_{e,ACK}T_{s,ACK}] \\
&= \frac{1}{p_{ce}}[P_{e,RTS}(N_{RTS}T_b + T_{DIFS}) + (1-P_{e,RTS})P_{e,CTS}((N_{RTS}+N_{CTS})T_b + \\
&\quad T_{SIFS} + T_{DIFS}) + (1-P_{e,RTS})(1-P_{e,CTS})P_{e,DT}((N_{RTS}+N_{CTS}+N_{DT})T_b + \\
&\quad 2T_{SIFS} + T_{DIFS}) + (1-P_{e,RTS})(1-P_{e,CTS})(1-P_{e,DT})P_{e,ACK} \cdot \\
&\quad ((N_{RTS}+N_{CTS}+N_{DT}+N_{ACK})T_b + 3T_{SIFS} + T_{DIFS})] \\
&\approx \frac{1}{p_{ce}}[P_{e,RTS}N_{RTS}T_b + (1-P_{e,RTS})P_{e,CTS}(N_{RTS}+N_{CTS})T_b \\
&\quad +(1-P_{e,RTS})(1-P_{e,CTS})P_{e,DT}(N_{RTS}+N_{CTS}+N_{DT})T_b \\
&\quad +(1-P_{e,RTS})(1-P_{e,CTS})(1-P_{e,DT})P_{e,ACK} \cdot \\
&\quad (N_{RTS}+N_{CTS}+N_{DT}+N_{ACK})T_b]. \qquad (2.40)
\end{aligned}
$$

By using (2.39) and (2.38), the approximate average delay of the system becomes

$$
\begin{aligned}
\widetilde{T}_d &= (N_{RTS}+N_{CTS}+N_{DT}+N_{ACK})T_b + \frac{1}{1-p_{ce}}\Big[P_{e,RTS}N_{RTS}T_b \\
&\quad +(1-P_{e,RTS})P_{e,CTS}(N_{RTS}+N_{CTS})T_b \\
&\quad +(1-P_{e,RTS})(1-P_{e,CTS})P_{e,DT}(N_{RTS}+N_{CTS}+N_{DT})T_b \\
&\quad +(1-P_{e,RTS})(1-P_{e,CTS})(1-P_{e,DT})P_{e,ACK} \cdot \\
&\quad (N_{RTS}+N_{CTS}+N_{DT}+N_{ACK})T_b\Big] + \frac{W\sigma}{2}(1+2p_{ce}) \qquad (2.41a) \\
&\approx (N_{RTS}+N_{CTS}+N_{DT}+N_{ACK})T_b + \Big[P_{e,RTS}N_{RTS}T_b \\
&\quad +P_{e,CTS}(N_{RTS}+N_{CTS})T_b + P_{e,DT}(N_{RTS}+N_{CTS}+N_{DT})T_b \\
&\quad +P_{e,ACK}(N_{RTS}+N_{CTS}+N_{DT}+N_{ACK})T_b\Big] \\
&\quad +\frac{W\sigma}{2}(1+2(P_{e,RTS}+P_{e,CTS}+P_{e,DT}+P_{e,ACK})), \qquad (2.41b)
\end{aligned}
$$

where we have used the assumption that the optimal packet error probabilities are small.

Our goal is to find the packet lengths N^*_{RTS}, N^*_{CTS}, N^*_{DT} and N^*_{ACK} that minimize the average system delay $\widetilde{T}_d$. After taking the partial derivative with respect to the four packet lengths and setting them equal to zero, we get four packet error probabilities that achieve minimum average delay. We will derive the optimal packet probability of ACK as an example. The other three optimal packet error probabilities can be obtained in a similar way. We have

$$\begin{aligned}
\frac{\partial \widetilde{T}_d}{\partial N_{ACK}} &= T_b(1+P_{e,ACK}) - W\sigma P_{e,ACK} R_0 \ln(2) \\
&\quad -P_{e,ACK} R_0 \ln(2)(N_{RTS}+N_{CTS}+N_{DT}+N_{ACK}) \\
&= T_b(1+P_{e,ACK}) - W\sigma P_{e,ACK} R_0 \ln(2) \\
&\quad +P_{e,ACK}[\ln(2^{-N_{RTS}R_0}) + \ln(2^{-N_{CTS}R_0}) + \ln(2^{-N_{DT}R_0}) + \ln(2^{-N_{ACK}R_0})] \\
&= T_b(1+P_{e,ACK}) - W\sigma P_{e,ACK} R_0 \ln(2) \\
&\quad +P_{e,ACK}[\ln(2^{-N_{RTS}R_0}2^{K_{RTS}}2^{-K_{RTS}}) + \ln(2^{-N_{CTS}R_0}2^{K_{CTS}}2^{-K_{CTS}}) \\
&\qquad + \ln(2^{-N_{DT}R_0}2^{K_{DT}}2^{-K_{DT}}) + \ln(2^{-N_{ACK}R_0}2^{K_{ACK}}2^{-K_{ACK}})] \\
&= T_b(1+P_{e,ACK}) - W\sigma P_{e,ACK} R_0 \ln(2) + P_{e,ACK}(\ln(P_{e,RTS}2^{-K_{RTS}}) \\
&\quad + \ln(P_{e,CTS}2^{-K_{RTS}}) + \ln(P_{e,DT}2^{-K_{DT}}) + \ln(P_{e,ACK}2^{-K_{ACK}})) \\
&= T_b(1+P_{e,ACK}) - W\sigma P_{e,ACK} R_0 \ln(2) \\
&\quad +P_{e,ACK}[\ln P_{e,RTS} - K_{RTS}\ln(2) + \ln P_{e,CTS} - K_{CTS}\ln(2) \\
&\qquad + \ln P_{e,DT} - K_{DT}\ln(2) + \ln P_{e,ACK} - K_{ACK}\ln(2)] \\
&\approx T_b - W\sigma P_{e,ACK} R_0 \ln(2) - P_{e,ACK}\ln(2)(K_{RTS}+K_{CTS}+K_{DT}+K_{ACK}).
\end{aligned} \tag{2.42}$$

By setting (2.42) equal to zero, we obtain the optimal probability for the ACK packet

as

$$P^*_{e,ACK} \approx \frac{1}{\ln(2)(K_{RTS}+K_{CTS}+K_{DT}+K_{ACK}+W\sigma R_0/T_b)}. \quad (2.43a)$$

Using a similar derivation as above, the other optimal packet error probabilities are given as

$$P^*_{e,RTS} \approx \frac{T_b}{T_b+\ln(2)(T_bK_{RTS}+WR_0)} \quad (2.43b)$$

$$P^*_{e,CTS} \approx \frac{T_b}{T_b+\ln(2)(T_b(K_{RTS}+K_{CTS})+WR_0)} \quad (2.43c)$$

$$P^*_{e,DT} \approx \frac{T_b}{T_b+\ln(2)(T_b(K_{RTS}+K_{CTS}+K_{DT})+WR_0)}. \quad (2.43d)$$

Note that the optimal packet error probabilities in (2.43) are consistent with the assumption that they are small compared to 1 when K_{RTS}, K_{CTS}, K_{DT} and K_{ACK} are large. Furthermore, we can also solve for the optimal packet lengths by using (2.33) as an equality, thus obtaining

$$N^*_{RTS} \approx \frac{K_{RTS}+\log_2(\ln(2)(K_{RTS}+W\sigma R_0/T_b))}{R_0} \quad (2.44a)$$

$$N^*_{CTS} \approx \frac{K_{CTS}+\log_2(\ln(2)((K_{RTS}+K_{CTS})+W\sigma R_0/T_b))}{R_0} \quad (2.44b)$$

$$N^*_{DT} \approx \frac{K_{DT}+\log_2(\ln(2)((K_{RTS}+K_{CTS}+K_{DT})+W\sigma R_0/T_b))}{R_0} \quad (2.44c)$$

$$N^*_{ACK} \approx \frac{K_{ACK}+\log_2(\ln(2)(K_{RTS}+K_{CTS}+K_{DT}+K_{ACK})+W\sigma R_0/T_b))}{R_0}. \quad (2.44d)$$

For the single user case, we can derive the approximate energy-delay values analytically from (2.43) and (2.44). Since p^*_{ce} is evaluated from (2.34) with (2.43) and $T^*_{s,ACK}$, T^*_{coe}, $E^*_{s,ACK}$ and E^*_{coe} can be determined from (2.17), (2.21), (2.22), and (2.35) with (2.44), the energy-delay values can be analytically obtained as:

$$(\widehat{T}^*_d, \widehat{E}^*_t) = (T^*_{s,ACK}+\frac{p^*_{ce}}{1-p^*_{ce}}T^*_{coe}+B^*_{single}, E^*_{s,ACK}+\frac{p^*_{ce}}{1-p^*_{ce}}E^*_{coe}), \quad (2.45)$$

where B^*_{single} is

$$B^*_{single} = \frac{W\sigma}{2}[\frac{1-2(2p^*_{ce})^m(1-p^*_{ce})+p^{*m}_{ce}(1-2p^*_{ce})}{1-2p^*_{ce}}+(2^{m+1}-1+\frac{2^m p^*_{ce}}{1-p^*_{ce}})p^{*m}_{ce}]. \quad (2.46)$$

For the multiuser case, the probabilities of p^*_{ce}, p^*_c, p^*_{tx} and $p^*_{tx_1}$ will be evaluated according to (2.2), (2.3), (2.4) and (2.5) with (2.43). By using the optimal probabilities of p^*_{ce}, p^*_c, p^*_{tx} and $p^*_{tx_1}$ and the optimal packet lengths obtained from (2.44), the approximate energy-delay values for the multiuser case can be obtained in a similar manner to the single user case as

$$(\widehat{T}^*_d, \widehat{E}^*_t) = (T^*_{s,ACK}+\frac{p^*_{ce}}{1-p^*_{ce}}T^*_{coe}+B^*_{multi}T^*_{rc}, E^*_{s,ACK}+\frac{p^*_{ce}}{1-p^*_{ce}}E^*_{coe}), \quad (2.47)$$

where B^*_{multi} is

$$B^*_{multi} = \frac{p^*_c W\sigma}{2}[\frac{1-2(2p^*_{ce})^m(1-p^*_{ce})+p^{*m}_{ce}(1-2p^*_{ce})}{1-2p^*_{ce}}+(2^{m+1}-1+\frac{2^m p^*_{ce}}{1-p^*_{ce}})p^{*m}_{ce}]. \quad (2.48)$$

2.5 Numerical Results

In this section we present and discuss our numerical results for various scenarios and system parameters. Table 3.1 summarizes the system parameters used to obtain the numerical results. In all plots, both energy and delay are normalized with respect to K_{DT} to allow for a fair comparison. In addition, we normalize the energy by N_0 since all results depend on the ratio of energy to noise power spectral density.

In Fig. 2.5, we plot the average energy and delay tradeoff curves for different K_{DT}. The interesting observation in Fig. 2.5 is that while both energy and delay increase with K_{DT}, the normalized energy and delay decrease. The explanation is that the delay or energy consumption of the overhead transmissions, such as RTS, CTS and

Table 2.1: System Parameters for Numerical Results

K_{RTS}	128 bits
K_{CTS}	128 bits
K_{ACK}	128 bits
Channel Bit Rate, $1/T_b$	1 Mbit/s
Slot Duration, σ	50μs
Maximum Contention Stage, m	5
Minimum Contention Window Size, W	8 slots
T_{SIFS}	28μs
T_{DIFS}	128μs

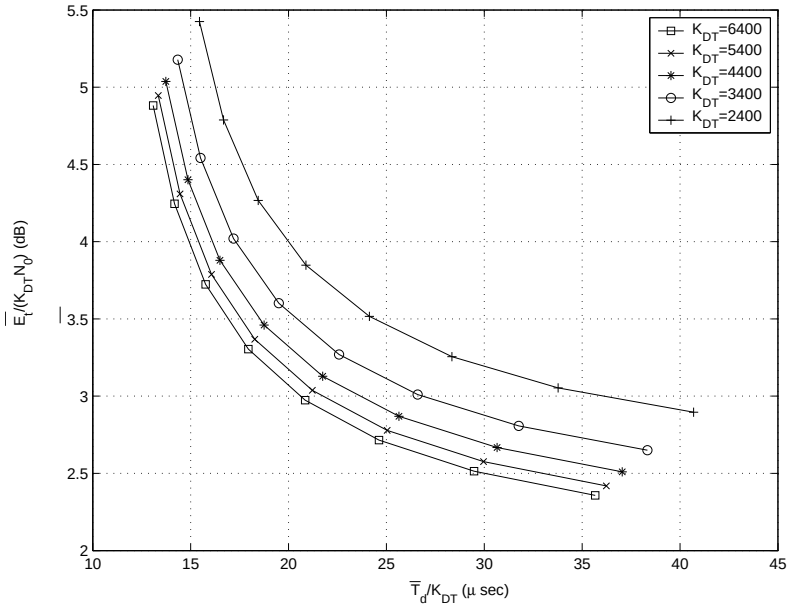

Figure 2.5: Energy-delay curves for different K_{DT} with $n = 10$.

ACK, become less significant for larger data packet lengths. To study the effect of different users, we fix the data packet length as $K_{DT} = 6400$ and plot the average energy-delay curves for different n. We can see from Fig. 2.6 that both energy and delay increase with the number of users. We also plot the single user case for the same protocol (802.11) and the simple ARQ. These two curves are lower bounds to the multiuser curves.

To verify the energy and delay tradeoff curves, we used C++ programming language to write an event-driven custom simulation program for 802.11 MAC protocol. In Fig. 2.7, we plot the average energy and delay tradeoff curves evaluated via our

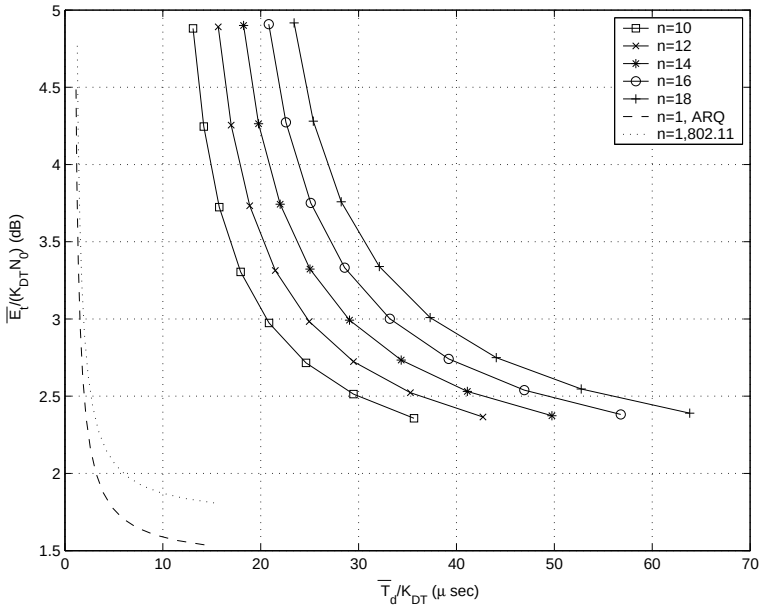

Figure 2.6: Energy-delay curves for different number of users n with $K_{DT} = 6400$.

numerical analysis and via simulation for different K_{DT} and n with parameters described in Table 3.1. We observe that the average energy and delay tradeoff curves obtained from analytical model is very accurate.

The delay variance can be calculated from the generating function as $\sigma_d^2 = \Delta_t^2 \frac{\partial^2 G_s}{\partial X^2} \mid_{X=Y=1} + \Delta_t \overline{T}_d - \overline{T}_d^2$. In Fig. 2.8, we plot the normalized standard deviation of the delay vs. the normalized average delay for different n and K_{DT}. We observe that the normalized standard deviation of the delay decreases as K_{DT} increases. It is also clear from the figure that the normalized standard deviation is grown in proportion to the normalized average delay with a factor of 1.5.

The average energy-delay tradeoff curves with different delay constraints are shown in Fig. 2.9. We fix E_c/N_0 and optimize $\overline{T}_d$ over $N_{RTS}, N_{CTS}, N_{DT}, N_{ACK}$. Let $N_{RTS}^*, N_{CTS}^*, N_{DT}^*, N_{ACK}^*$ be the packet lengths that minimize $\overline{T}_d$. The system generating function with these packet lengths and E_c/N_0 is denoted as G_s^*. If outage delay probability is given, we can evaluate both outage delay γ_d and $\mathbf{E}[E_t \mid T_d < \gamma_d]$

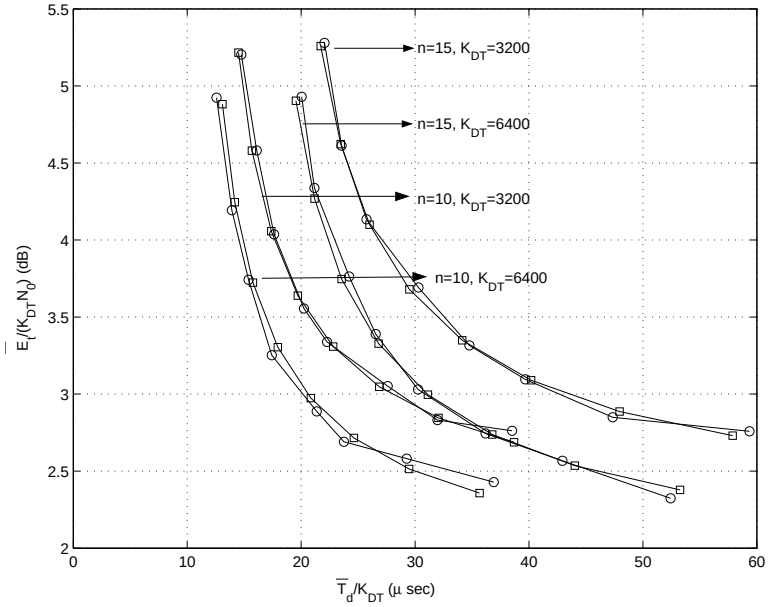

Figure 2.7: Energy delay curves for different number of users and packet sizes. The lines with squares represent numerical results and the lines with circles represent simulation results.

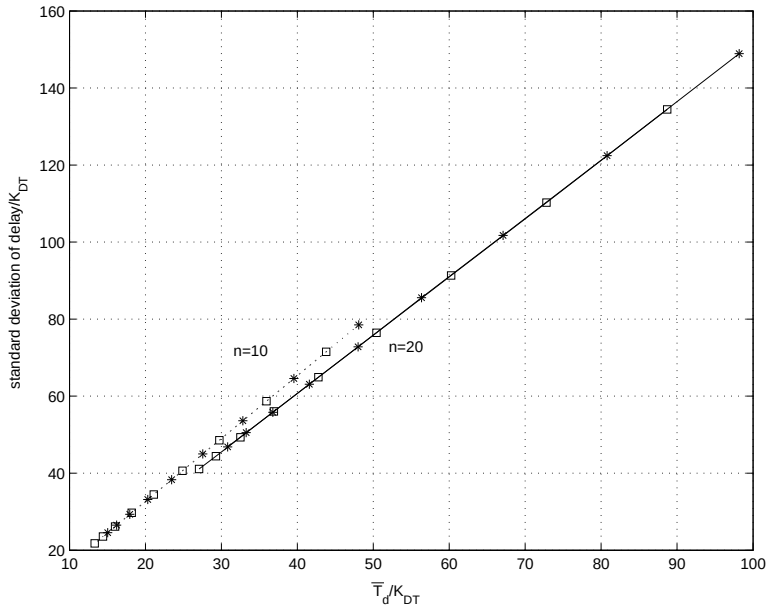

Figure 2.8: Normalized standard deviation of delay vs. average delay curves for different n and K_{DT}. Lines with a square symbol represent the case of $K_{DT} = 4400$ and lines with a star symbol represent the case of $K_{DT} = 2400$.

from G_s^*. Repeating the above procedure for different E_c/N_0, we obtain the average energy-delay tradeoff curve with a delay constraint. Fig. 2.9 shows that when the outage delay probability increases, the range of outage delay decreases and conditional average energy also decreases. For comparison, we also plot the average energy vs. average delay curve. We see that the average energy used can be quite different depending on whether an average delay or a strict delay requirement is imposed.

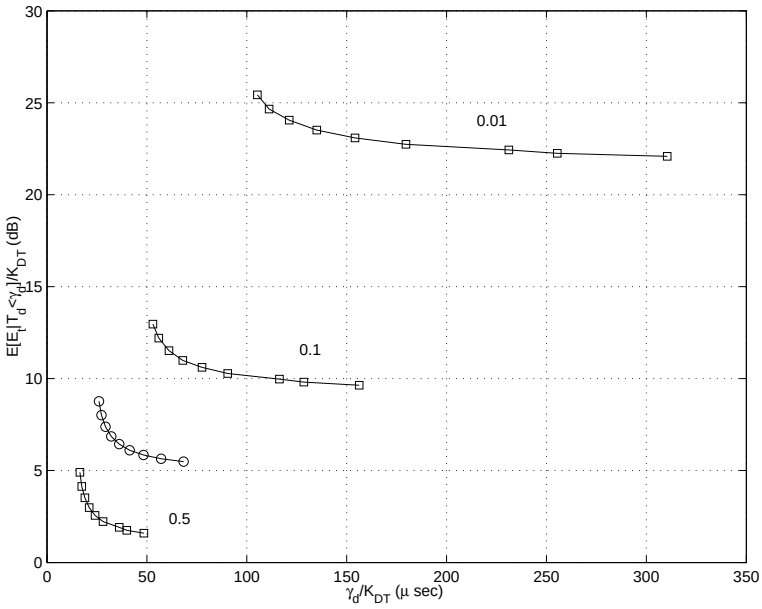

Figure 2.9: Energy-delay curves for various outage delay probabilities and different values for the outage probability $Pr(T_d > \gamma_d)$ $(m = 1, n = 10, W = 8, K_{DT} = 640)$. The average energy and delay curve (circle symbol) is also shown for comparison.

The average energy-delay tradeoff curves with different energy constraints is given in Fig. 2.10. Given an outage energy probability constraint, we can evaluate both the outage energy γ_e and $\mathbf{E}[T_d \mid E_t < \gamma_e]$. Then using the same procedure as above we obtain the average energy-delay tradeoff curve with an energy constraint. The observations made for Fig. 2.9 apply to Fig. 2.10 as well.

In Fig. 2.11, we plot energy-delay curves of under different K_{DT} with $n = 10$ users in the network. It is observed that the approximate method is very accurate over a

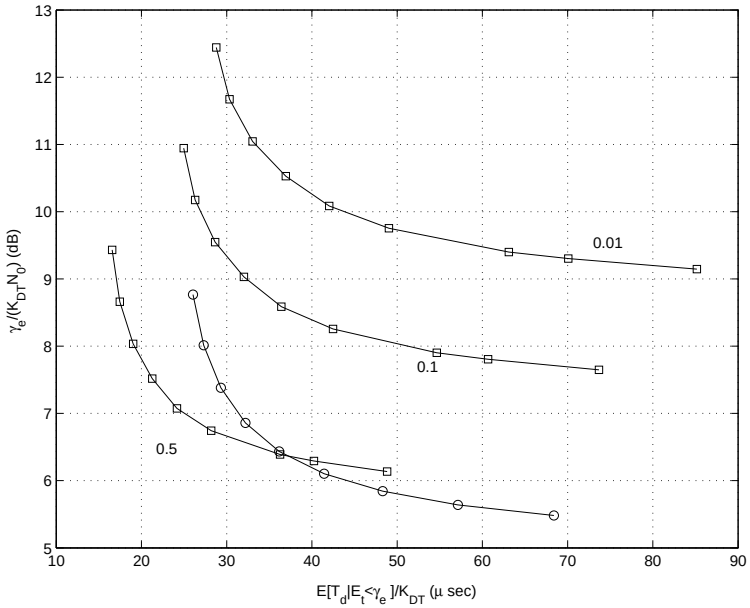

Figure 2.10: Energy-delay curves for various outage energy probabilities and different values for the outage probability $Pr(E_t > \gamma_e)$ $(m = 1, n = 10, W = 8, K_{DT} = 640)$. The average energy and delay curve (circle symbol) is also shown for comparison.

wide range of K_{DT}.

2.6 Conclusion

In this chapter we obtained the joint distribution of energy and delay for packet transmission using RTS/CTS-type MAC protocols taking into account the PHY layer. By representing the protocol using a state diagram we derived the joint generating function of energy and delay. The generating function was used to obtain various statistics such as the average energy and delay, the average energy with a delay constraint, etc. This approach allows us to optimize the performance over the block lengths used for different packets. Finally, a very accurate analytical approximation for the minimum average delay and corresponding delay was derived that enabled us to get the optimal energy-delay tradeoff curves analytically. We believe that our methods can be extended to the analysis of other protocols, including

segmentation of data packets or retransmission of data segments without channel reservation.

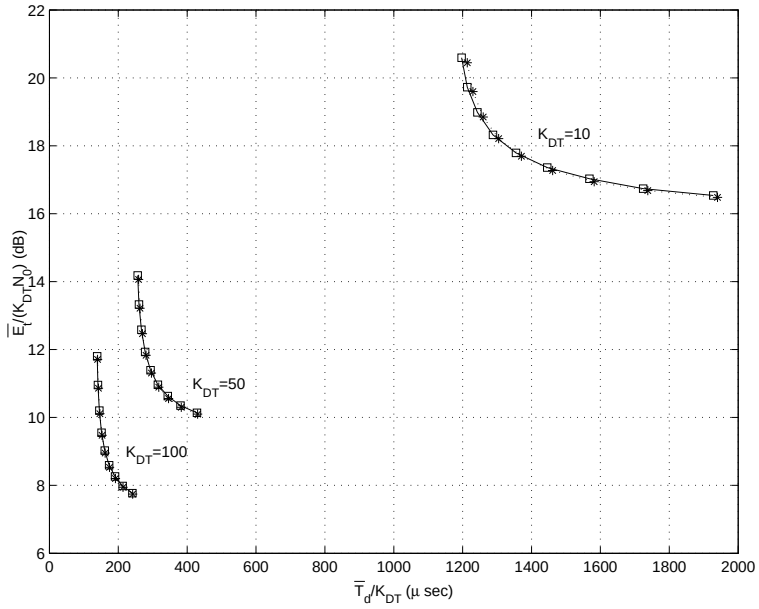

Figure 2.11: Energy-delay tradeoff curves evaluated from the approximation method (star symbol) and the exact energy-delay tradeoff curves (square symbol) with $n = 10$

CHAPTER III

Variations of 802.11 MAC Protocol

In this chapter, we apply our proposed generating function method to the analysis and design of two different wireless network protocols modified from the original IEEE 802.11 protocol. The first wireless network protocol, discussed in Section 3.1, adapts the energy per coded symbol whenever it suffers a failed transmission. The second wireless network protocol related to the application of ARQ to IEEE 802.11 is investigated in Section 3.2. The tradeoff between energy and delay is studied for both proposed protocols. The conclusion of this chapter is stated in Section 3.3.

3.1 Adaptive Energy Scheme for Wireless Network Systems

Wireless networks consists of mobile and sensors devices. Most such devices are battery powered. Since the energy of a battery is a limiting resource it is important to have an energy efficient network design. Higher energy efficiency protocols and signal design may extend the battery lifetime or result in using batteries with smaller required energy capacity. However, wireless channels have fading that typically requires more energy for transmission than an unfaded channel for the same packet error probability. Decreasing the amount of energy used in transmission increases the packet error probability and can increase the delay for a successfully received packet. Decreasing the transmission delay by increasing the transmission rate gen-

erally requires more energy. Hence, there is a fundamental tradeoff between energy and delay in designing a wireless network.

The energy efficiency of wireless networks has been a subject of current research. In order to synchronize different stations in a wireless network, the beacon needs to be sent out repeatedly for some time interval. The duration of this time interval is called beacon interval. In [16], the authors presented a new MAC protocol for energy saving mode in IEEE 802.11 by proposing a quorum-based sleep/wake-up mechanism, which conserves energy by allowing the host to sleep for more than one beacon interval if few transmissions are involved. This new scheme is more energy efficiency compared to the original IEEE 802.11 energy saving mode since the original scheme fails to adjust a station's sleep duration according to its traffic and nearby network topology. In [8], the authors proposed a new scheduling scheme called adaptive Weighted Round Robin (WRR) scheme that provides service differentiation in multiple data class queuing system by considering battery life and load of traffic. As actual energy of the battery becomes low, the proposed scheme gives more weight to high priority queue. So high priority data flows would get more chance to be served when energy in the battery becomes lower. They studied and analyzed the performance of this scheme and compare its performance with the original IEEE 802.11 standard. Their results showed that the proposed scheme improves overall end-to-end throughput as well as support service differentiation over multi-hop wireless networks. The authors in [51, 42] proposed a novel MAC scheme that changes the duration of listen/sleep modes adaptively according to the information about the network, e.g., network topology and network traffic flows. Their analytical and simulation results demonstrated a significant decrease in energy consumption.

In a wireless environment, the packet error probability is an important parame-

ter in overall system performance. However, in all the above mentioned papers, the effects of energy consumption for each kind of packet error probability used in IEEE 802.11 (RTS, CTS, DATA and ACK packets) was not considered. Moreover, none of the above papers consider the energy and delay consumption jointly of transmitting a data packet in the system. The motivation of this section is to understand the energy and delay relation of networks by simultaneously incorporating the packet error probabilities (physical layer) and basic access method (MAC layer) in the performance analysis.

In this section, we propose and analyze an adaptive protocol in which the transmitter will increase the energy level per coded symbol whenever it suffers a failure transmission. The numerical results show that the proposed scheme can improve system performance significantly compared to the original IEEE 802.11 protocol when the channel condition is bad. By using Reed-Solomon codes we can optimize the system performance over the code rate and energy per coded bit. Although we use Reed-Solomon codes over an additive white Gaussian noise (AWGN) channel to represent packet error probabilities, the framework of our analysis can be used for other coding and modulation schemes over various wireless channels. Finally, we also compare the system performance between Reed-Solomon codes and convolutional codes.

The rest of this section is organized as follows. In Subsection 3.1.1, we introduce the system assumptions and give a brief description for the protocol used in our analysis. In Subsection 3.1.2, we introduce the generating function and determine the joint generating function for energy and delay. In Subsection 3.1.3, we present numerical results regarding the throughput, energy efficiency and energy/delay tradeoff with Reed-Solomon codes over an AWGN channel.

3.1.1 System Description

The wireless networks that we analyze here have the following specifications. First, each station with a fixed position can hear (detect and synchronize) the transmission of $n-1$[1] other stations in the network. Second, stations always have a packet ready to transmit. Third, the receivers have no multiple-access capability (i.e., they can only receive one packet at a time). Fourth, all stations have fixed position (no mobility). Fifth, we assume that the propagation loss between transmitter and receiver is one (0 dB) and also assume that the propagation time is negligible. These assumptions are valid for local area networks. Finally, each station uses 802.11 protocol for medium access control (MAC).

In the following we give a brief description of the most salient features of the IEEE 802.11 MAC protocol (more details can be found in [3] and [4]). When a station is ready to transmit a packet, it picks randomly a number j, uniformly distributed in $\{0, 1, \ldots, W_i - 1\}$, where $W_i = 2^i W$ is the contention window (CW) size, i is the contention stage (initially $i = 0$), and W is the minimum CW size. A backoff time counter begins to count down with an initial value j: it decreases by one for every idle slot of duration σ seconds as long as the channel is sensed idle, stops the count down when the channel is sensed busy, and reactivates when the channel is sensed idle again. The station transmits a request-to-send (RTS) packet when the counter counts down to zero. After transmitting the RTS packet, the station will wait for a clear-to-send (CTS) packet from the receiving station. If there is a collision of the RTS packet with other competing stations or a transmission error occurs in the RTS or CTS packet, the transmitting station doubles the CW size (increases the contention stage i by one) and picks a random number j as before. If

[1] n is the density of stations.

there are no collisions or errors in the RTS and CTS packets, the station begins to transmit the data packet and waits for an acknowledgment (ACK) packet. However, if the data or the ACK packet is not successfully received, the CW size will also be doubled (the contention stage i will increase by one) and the transmitting station will join the contention period again. The contention stage is reset (i is set to zero) when the transmitting station receives an ACK correctly. A time diagram indicating the sequence of these events is depicted in Fig. 3.1. It is also noted that there is a maximum CW size (or equivalently, a maximum contention stage, m); when the transmitter is in this maximum stage and needs to join the contention period again, it does not increase further the contention window, but picks a random number in $\{0, 1, \ldots, 2^m W - 1\}$.

The proposed adaptive scheme uses the same MAC protocol as above, however, the transmitter increases the energy level per coded symbol whenever the CW stage increases and let $E_{c,i}$ represent the energy level per coded symbol at transmission with CW stage $i - 1$.

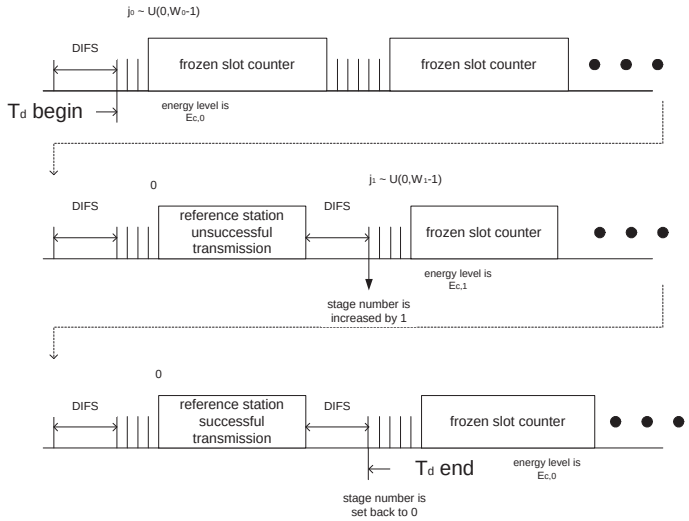

Figure 3.1: Timing diagram for protocol. The j_0 and j_1 are backoff random numbers at CW stage $i = 0$, and $i = 1$, respectively.

3.1.2 System Delay and Energy Analysis

In this section, we analyze the energy and delay characteristics of the wireless networks described above. The delay T_d of each data packet is defined as the time duration from the moment the backoff procedure is initiated until DIFS seconds after the ACK packet is received correctly by the transmitting station, as shown in Fig. 3.1. Similarly, the energy E_t is defined as the energy consumed by both transmitting and receiving stations in the duration of T_d. In this paper, we only consider the energy consumption for packet transmission and omit the energy required for signal processing and channel sensing since the energy consumption for packets transmission requires most energy compared to other energy consumption aspects.

Nonlinear System Equations

As in [5], we define two random processes to characterize the random backoff scheme. The first process b_t represents the value of the backoff counter while it is active. The second random process s_t is used to represent the contention stage $i \in \{0, 1, \ldots, m\}$ of the station. It uses the same time scale as the backoff counter process. The first assumption made in this analysis is that the event of packet collision is independent of past collisions and thus independent of the contention stage. The second assumption is that the packet collision is identical for all states (values of b_t and s_t) of a user. As verified in [5], these two assumptions are extremely accurate when the number of stations in the network is large (say greater then 10). Let p_{cc} be the conditional collision probability, i.e., the probability of a collision given a packet transmitted on the channel. As a result of the above assumptions, we will model the two dimensional process (s_t, b_t) by a discrete-time Markov chain. With the assumption that errors in transmission can occur only due to collision, the one-

step transition probabilities developed in [5] are given by $P\{i,k \mid j,l\} = P\{s_{t+1} = i, b_{t+1} = k \mid s_t = j, b_t = l\}$ with

$$\begin{aligned} P\{i,k \mid i,k+1\} &= 1, && 0 \le k \le W_i - 2, 0 \le i \le m \\ P\{0,k \mid i,0\} &= \tfrac{1-p_c}{W_0}, && 0 \le k \le W_0 - 1, 0 \le i \le m \\ P\{i,k \mid i-1,0\} &= \tfrac{p_c}{W_i}, && 0 \le k \le W_i - 1, 1 \le i \le m \\ P\{m,k \mid m,0\} &= \tfrac{p_c}{W_m}, && 0 \le k \le W_m - 1. \end{aligned} \tag{3.1}$$

The first equation in (3.1) corresponds to the decrement of the backoff counter at the beginning of each time slot. The second equation accounts for the fact that a new packet following a successful packet transmission starts at contention stage $i = 0$, and thus the backoff counter is initially uniformly chosen in the range of $\{0, 1, \ldots, W_0 - 1\}$. The other two cases describe the system evolution after an unsuccessful transmission. As described in the third equation, when an unsuccessful transmission occurs at contention stage $i - 1$, the contention stage increases and the backoff counter is initialized with a uniformly chosen value in the range $\{0, 1, \ldots, W_i\}$. Finally, the last case models the fact that the contention stage is not increased in subsequent packet transmissions when the contention window size reaches the maximum. Fig. 4 in [5] depicts the aforementioned two-dimensional Markov chain.

By modifying the Markov chain model described above, we can take into account packets errors with different energy level per coded symbols. We denote the error probability of the four kinds of packets at CW stage i in the system as $P_{e,RTS,i}$, $P_{e,CTS,i}$, $P_{e,DT,i}$ and $P_{e,ACK,i}$, where i in $\{0, 1, \ldots, m\}$. Packets which are transmitted at stage i will be transmitted with energy level $E_{c,i}$ per coded symbol. We assume that the channel is memoryless between packets. These probabilities depend on the particular channel, coding, modulation etc (a specific example will be given in Section 3.1.3). Let $p_{ce,i}$ represent the transition probability from CW stage i to

the next. This is also the probability of an unsuccessful transmission attempt seen by the transmitting station when its packet is being transmitted on the channel. However, in this paper, an unsuccessful transmission can happen not only due to packet collision, but also due to packet errors. Using similar assumptions as in [5] for the packet collision probability p_c, and since the events of packet collision and the event of packet error are independent, $p_{ce,i}$ can be expressed as

$$\begin{aligned} p_{ce,i} = p_c + (1-p_c)[&P_{e,RTS,i} + (1-P_{e,RTS,i})P_{e,CTS,i} \\ &+ (1-P_{e,RTS,i})(1-P_{e,CTS,i})P_{e,DT,i} \\ &+ (1-P_{e,RTS,i})(1-P_{e,CTS,i})(1-P_{e,DT,i})P_{e,ACK,i}]. \end{aligned} \tag{3.2}$$

The one-step transition probabilities are modified as

$$\begin{aligned} P\{i,k \mid i,k+1\} &= 1, && 0 \le k \le W_i - 2, 0 \le i \le m \\ P\{0,k \mid i,0\} &= \tfrac{1-p_{ce,i}}{W_0}, && 0 \le k \le W_0 - 1, 0 \le i \le m \\ P\{i,k \mid i-1,0\} &= \tfrac{p_{ce,(i-1)}}{W_i}, && 0 \le k \le W_i - 1, 1 \le i \le m \\ P\{m,k \mid m,0\} &= \tfrac{p_{ce,m}}{W_m}, && 0 \le k \le W_m - 1 \end{aligned} \tag{3.3}$$

The packet transmission probability p_{tx} which is the probability of a transmitting station sending an RTS packet during each backoff slot is obtained by summing over the stationary probability of all the states at first column of the two-dimensional Markov diagram as shown in Fig. 4 at [5]. It can be shown as

$$p_{tx} = \frac{2(1 + \sum_{j=1}^{m-1}(\prod_{i=0}^{j-1} p_{ce,i}) + \prod_{i=0}^{m-1} p_{ce,i}/(1-p_{ce,m}))}{D}, \tag{3.4}$$

where D is expressed as

$$\begin{aligned} D = 1 + \sum_{j=1}^{m-1}(\prod_{i=0}^{j-1} p_{ce,i}) + \frac{\prod_{i=0}^{m-1} p_{ce,i}}{1-p_{ce,m}} + \\ W(1 + \sum_{j=1}^{m-1}(\prod_{i=0}^{j-1} 2^{i+1} p_{ce,i}) + \frac{2^m \prod_{i=0}^{m-1} p_{ce,i}}{1-p_{ce,m}}). \end{aligned} \tag{3.5}$$

The collision probability p_c can be expressed as

$$p_c = 1 - (1 - p_{tx})^{n-1}, \tag{3.6}$$

since the collision will happen if there are at least one remaining $n-1$ stations transmitting packets.

From the above nonlinear equations, (3.2), (3.4) and (3.6), one can evaluate the probabilities $p_{ce,i}$, p_{tx} and p_c. Another important probability that will be used in our analysis later is the probability of exactly one transmission of an RTS packet from one of the remaining $n-1$ stations given that at least one of the remaining stations is transmitting. It is denoted by p_{tx_1} and can be expressed as

$$p_{tx_1} = \frac{(n-1)p_{tx}(1-p_{tx})^{n-2}}{p_c}. \tag{3.7}$$

Delay/Energy Joint Generating Function

We assume that stations spend energy only in packet transmission and that the energy consumption for each packet transmission is proportional to the packet length. Then the energy transmission for a packet of length N (normalized to the thermal noise energy level, N_0) is $NE_{c,i}/N_0$, where i is the stage of CW. Similarly, the *transmission* delay of each packet is also proportional to the packet length. Let R_t be the transmitting rate in bits per second. The delay, in seconds, for a packet transmission with packet length N is N/R_t.

Based on the protocol description in Section 3.1.1, the state flow diagram shown in Fig. 3.2 can be built. Solid lines represent successful reservation or transmission, while dotted lines represent unsuccessful reservation or transmission.

Let T_{RTS}, T_{CTS}, T_{DT} and T_{ACK} denote the time duration for the transmission of RTS, CTS, DT and ACK packets, respectively. Similarly, let $E_{RTS,i}$, $E_{CTS,i}$,

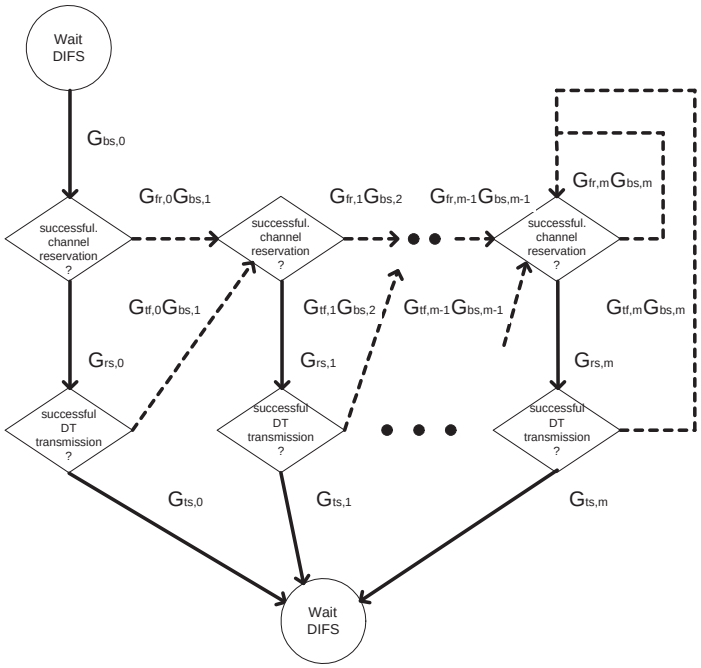

Figure 3.2: State diagram for the protocol.

$E_{DT,i}$ and $E_{ACK,i}$ denote the energy for the transmission of RTS, CTS, DT and ACK packets at CW stage i, respectively. Finally, let T_{DIFS} and T_{SIFS} be system delay parameters defined by the standard. We will now derive the conditional joint generating function of the delay and energy random variables, associated with the transitions in the state diagram in Fig. 3.2. In all the derivations below, functions are of the form $G(X, Y)$, where the variables X, Y are the transform variables of the delay and energy, respectively.

The generating function $G_{rs,i}$ corresponding to successful channel reservation at CW stage i is

$$G_{rs,i}(X,Y) = (1-p_c)(1-P_{e,RTS,i})(1-P_{e,CTS,i}) \\ X^{T_{RTS}+T_{CTS}+2T_{SIFS}} Y^{E_{RTS,i}+E_{CTS,i}}. \tag{3.8}$$

Similarly, the generating function G_{ts} corresponding to successful transmission of a

data packet at CW stage i can be expressed as

$$G_{ts,i}(X,Y) = (1 - P_{e,DT,i})(1 - P_{e,ACK,i}) \\ X^{T_{DT}+T_{ACK}+T_{SIFS}+T_{DIFS}} Y^{E_{DT,i}+E_{ACK,i}}. \quad (3.9)$$

We now develop expressions for the generating functions associated with failure to either reserve a channel or to transmit data. We first derive the generating function $G_{fr,i}$, corresponding to failure to reserve the channel at stage i is given by

$$G_{fr,i}(X,Y) = [p_c + (1-p_c)P_{e,RTS,i}]X^{T_{RTS}+T_{DIFS}} Y^{E_{RTS,i}} \\ + (1-p_c)(1-P_{e,RTS,i})P_{e,CTS,i} \\ X^{T_{RTS}+T_{CTS}+T_{SIFS}+T_{DIFS}} Y^{E_{RTS,i}+E_{CTS,i}}, \quad (3.10)$$

the meaning of which is that a failure can be due to either a collision or RTS transmission error, or a CTS transmission error. Similarly, after reserving the channel, a failure to transmit a data packet at CW stage i is either due to the data packet transmission error, or the acknowledgement packet transmission error, which is captured by the generating function $G_{tf,i}$ as follows

$$\begin{aligned} G_{tf,i}(X,Y) &= P_{e,DT,i} X^{T_{DT}+T_{DIFS}} Y^{E_{DT,i}} + (1 - P_{e,DT}) \\ &\cdot \quad P_{e,ACK} X^{T_{DT}+T_{ACK}+T_{SIFS}+T_{DIFS}} \\ &\cdot \quad Y^{E_{DT,i}+E_{ACK,i}}. \end{aligned} \quad (3.11)$$

We now evaluate the generating functions denoted by $G_{bs,i}$ of the state diagram. This generating function will characterize the delay for the transmitting station from the instant of starting the backoff procedure to the instant that the backoff counter reaches to zero at stage i. We do not need to consider energy consumption here since the transmitting station stops any transmission at this period. The probability of

each slot being sensed busy due to the transmission of other stations is p_c and this event is independent and identical to each slot from our previous assumptions. At backoff stage i, the range of possible backoff slots is from 1 to 2^iW. Let j be the backoff slots randomly chosen uniformly from the above range, then the number of the occupied slots in these j slots is binomialy distributed with parameters (j, p_c). Hence, the generating function, $G_{bs,i}$, of backoff procedure at stage i is

$$G_{bs,i}(X) = \sum_{j=1}^{2^iW} \frac{1}{2^iW} \sum_{k=0}^{j} \binom{j}{k} [(1-p_{cc})X^{\sigma}]^{j-k} (p_c G_{oc}(X))^k. \tag{3.12}$$

where $G_{oc}(X)$ is the generating function for an slot sensed busy by a station. We define an occupied slot as a slot when the transmitting station senses the channel is busy due to the transmission of one of the $n-1$ remaining stations. There are four possible cases of an occupied slot. The first case is when two or more RTS packets from the remaining $n-1$ stations collide or have a packet error. The second case is when the RTS packet is correctly received without collision but there is an error in the CTS packet transmission. The third case is when the RTS packet is correctly received without collision and there is no error in the CTS packet but there is an error in the DT packet transmission. The last case is correct reception of RTS, CTS, and DT and either correct or erroneous reception of the ACK packet. Since there are m possible CW stages of packets transmission from the other $n-1$ stations, the probability of the transmission come from CW stage i given only one remaining $n-1$ station transmission is $b_{i,0}/p_{tx}$, where $b_{i,0}$ is the stationary probability of state that CW stage is i and the counter reaches to zero. Note that the time duration of successful or unsuccessful ACK transmission are same. Therefore, the correspond

generating function for this occupied slot is

$$\begin{aligned} G_{oc}(X) &= (1-p_{tx_1})X^{T_{RTS}+T_{DIFS}} \\ &+ \frac{p_{tx_1}}{p_{tx}} \sum_{l=0}^{l=m} b_{l,0}[P_{e,RTS,l}X^{T_{RTS}+T_{DIFS}} \\ &+ (1-P_{e,RTS,l})P_{e,CTS,l}X^{T_{RTS}+T_{CTS}+T_{SIFS}+T_{DIFS}} \\ &+ (1-P_{e,RTS,l})(1-P_{e,CTS,l})P_{e,DT,l} \\ &\quad X^{T_{RTS}+T_{CTS}+T_{DT}+2T_{SIFS}+T_{DIFS}} \\ &+ (1-P_{e,RTS,l})(1-P_{e,CTS,l})(1-P_{e,DT,l}) \\ &\quad X^{T_{RTS}+T_{CTS}+T_{DT}+T_{ACK}+3T_{SIFS}+T_{DIFS}}]. \end{aligned} \tag{3.13}$$

From the state diagram and Mason's gain formula, we obtain the following backward recursive generating function to characterize the energy and delay of the system. Let G_s be the system generating function. Then the recursion is

$$\begin{aligned} f_m(X,Y) &= \frac{G_{rs,m}G_{ts,m}}{1-G_{fr,m}G_{bs,m}-G_{rs,m}G_{tf,m}G_{bs,m}} \\ f_{i-1}(X,Y) &= G_{rs,i-1}G_{ts,i-1}+(G_{fr,i-1}+G_{rs,i-1}G_{tf,i-1})G_{bs,i}f_i \\ G_s(X,Y) &= G_{bs,0}f_0, \end{aligned} \tag{3.14}$$

where i is the index of the CW stage from 1 to m. Hence, we can evaluate any joint statistics of the energy and delay of a successful packet transmission. In particular, the mean values are given by

$$T_{d,avg} = \frac{\partial G_s}{\partial X} \mid_{X=Y=1}, E_{t,avg} = \frac{\partial G_s}{\partial Y} \mid_{X=Y=1} . \tag{3.15}$$

3.1.3 Numerical Results

In this section we present and discuss our numerical results for various scenarios and system parameters. Table 3.1 summarizes the system parameters used to obtain

Table 3.1: System Parameters for Numerical Results

K_{RTS}	128 bits
K_{CTS}	128 bits
K_{ACK}	128 bits
Channel Bit Rate, $1/T_b$	1 Mbit/s
Slot Duration, σ	50μs
Maximum Contention Stage, m	3
Minimum Contention Window Size, W	8 slots
T_{SIFS}	28μs
T_{DIFS}	128μs

the numerical results. We will use Reed-Solomon codes and convolutional codes to represent packet error probabilities. Let K be the number of information symbols (1 symbol = 8 bits) in a packet (specified by the 802.11 standard) and N be the number of coded symbols for this packet, then the packet error probability $P^{RS}_{e,K,N}$ over an AWGN channel with BPSK modulation can be expressed as

$$P^{RS}_{e,K,N} = \sum_{l=t+1}^{N} \binom{N}{l} P_s^l (1-P_s)^{N-l}, \tag{3.16}$$

where $t = \lfloor \frac{N-K}{2} \rfloor$ and $P_s = 1 - (1 - Q(\sqrt{\frac{2E_bK}{NN_0}}))^8$ is the symbol error probability. Note that E_b is the energy used per information bit.

In Fig. 3.3, the relation between the average delay throughput S_{Td} and the channel SNR with code rate $\frac{1}{2}$ for different K_{DT} of convolutional codes is shown. For high SNR, we can find that the throughput increases with longer information packets because the system with longer information packets has smaller overhead redundant in RTS/CTS MAC protocol. For low SNR, we observe that the throughput decreases with longer information packets since the packet error probability for convolutional codes increases when the length of information packets becomes larger. Moreover, if longer information packets are used, the effect of packet error probability to the S_{Td} will be more significant compared to the effect of less overhead redundant for low channel SNR condition. We expect that there is an optimal value of K_{DT} which achieves the largest throughput for each given channel SNR.

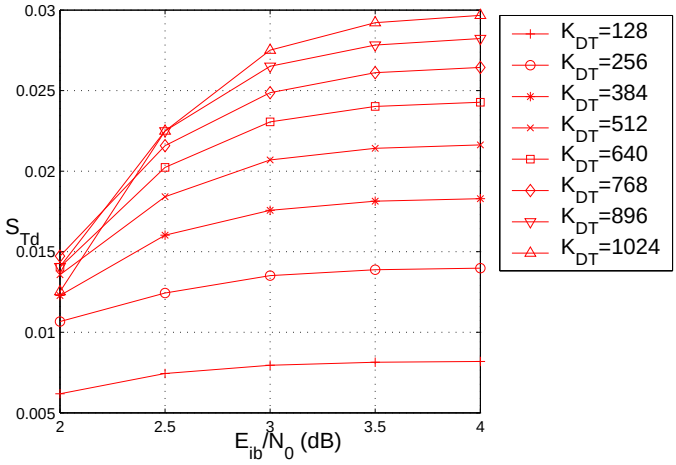

Figure 3.3: The S_{Td} comparison of convolutional codes for different K_{DT} with code rate $\frac{1}{2}$ and $n = 10$.

In Fig. 3.4, the relation between the average delay throughput S_{Td} and the channel SNR with code rate $\frac{1}{2}$ for different K_{DT} of Reed-Solomn codes is shown. For high SNR, we observe that the throughput increases with longer information packets because the packet error probabilities of Reed-Solomn codes decrease with the increase of the length of information packets. For low SNR, the throughput decreases with longer information packets since the packet error probabilities of Reed-Solomn codes increases with the increase of the length of information packets. We also found that the convolutional codes has better performance since it adopts soft decision instead of hard decision used in Reed Solomon codes.

In Fig. 3.5, we compare the S_{Td} of convolutional and Reed-Solomn codes for different rate with $K_{DT} = 128$. For convolutional codes, we can see that the S_{Td} increases with the decrease of code rate since the gain of S_{Td} due to smaller packet error probabilities is more than the loss of S_{Td} due to less information transmission

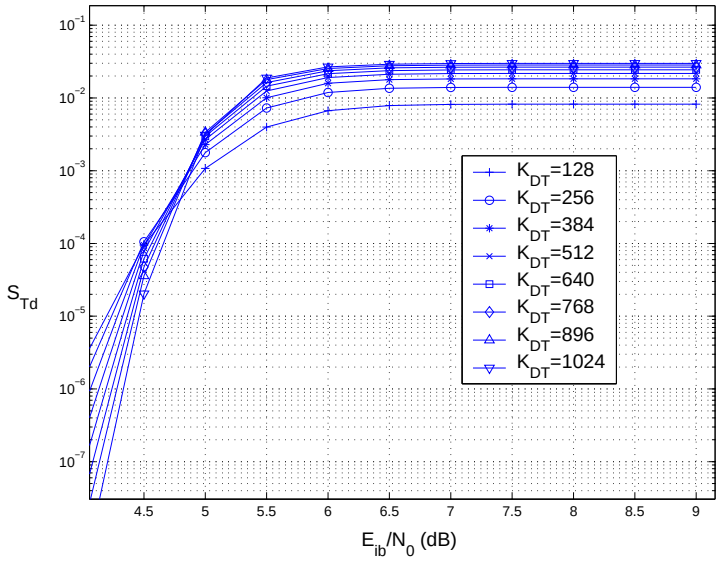

Figure 3.4: The S_{Td} comparison of Reed-Solomn codes for different K_{DT} with code rate $\frac{1}{2}$ and $n = 10$.

efficiency per coded packet when the code rate becomes smaller. For Reed-Solomn codes, we observe that the S_{Td} increases with the increase of code rate since the loss of S_{Td} due to larger packet error probabilities is less than the gain of S_{Td} due to higher information transmission efficiency per coded packet when the code rate becomes larger.

In Fig. 3.6, we plot the S_{Td} of Reed-Solomn codes with the channel SNR for different code rate as $K_{DT} = 128$. The S_{Td} is affected by two factors, packet error probabilities and the information transmission efficiency per coded packet (code rate). For lower code rate, we will have lower information transmission efficiency per coded packet, however, packet error probabilities may not decrease since the energy used per coded bit also decreases with the lower code rate from (3.16). Therefore, the total system throughput S_{Td} is the tradeoff result of these two factors. For higher SNR, since the packet error probabilities are almost zero, the value of S_{Td} is dominated directly by the information transmission efficiency per coded packet (code

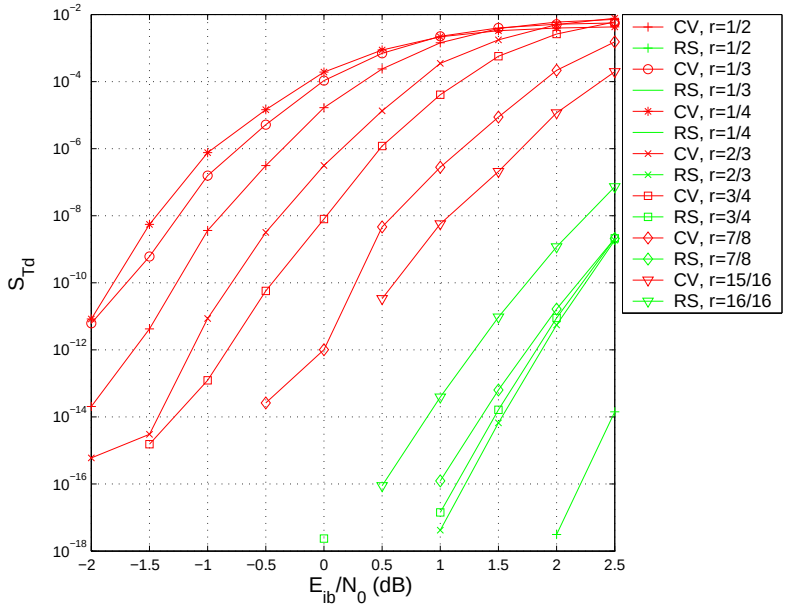

Figure 3.5: The S_{Td} comparison of convolutional and Reed-Solomn codes for different code rate with $K_{DT} = 128$ and $n = 10$.

rate). For lower SNR, since the gain of S_{Td} due to smaller packet error probabilities is more than the loss of S_{Td} due to less information transmission efficiency when the code rate becomes smaller. Therefore, we can find that there is a cross of S_{Td} values for curve with code rate $\frac{7}{8}$ and curve with code rate $\frac{1}{4}$ at 5.6 dB.

An interesting observation is that there is an optimal N. For small N the packet error probability of a random code will be large. Large packet error probability will increase the system delay and energy due to the high chance of packet retransmission. On the other hand, if N is large, the packet error probability for a random code will be small but the transmission time will be large. Our goal is to find the best N for each kind of packet to minimize system delay. We first fix $E_{c,1}/N_0$ and optimize the $T_{d,avg}$ to get the corresponding packet lengths for N_{RTS}, N_{CTS}, N_{DT} and N_{ACK}. From this, we can evaluate both average delay and average energy consumption. Repeating the above argument for different $E_{c,1}/N_0$, we get the energy-delay curve.

In Fig. 3.7, we compare energy and delay tradeoff curves for different values of

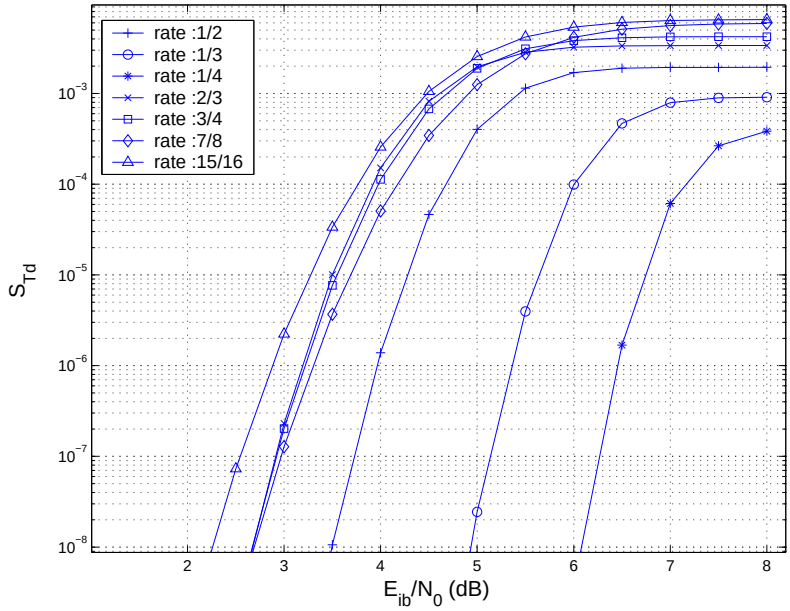

Figure 3.6: The S_{Td} comparison of Reed-Solomn codes for different code rate with $K_{DT} = 128$ and $n = 10$.

Δ. We observe that the adaptation scheme can improve the system performance by reducing both average energy and delay as the channel SNR is small. At good channel condition, although higher value of Δ can reduce the average delay, the system incurs higher energy consumption (compared to the system with smaller value of Δ) to transmit a data packet successfully. This demonstrates that that the system performance was improved significantly when the channel SNR is low and the average delay is reduced insignificantly even consuming more energy at high channel SNR case.

For the effect of different users, we fix the data packet length as $K_{DT} = 64$ symbols and plot energy-delay curve for different n. We can see from Fig. 3.8 that both energy and delay increase with the number of users. The other fact shown in this figure is that both energy and delay increase with the increment of K_{DT} but the normalized (with respect to K_{DT}) energy and delay decrease with increasing of K_{DT}. The explanation of this is that the delay or energy consumption of overhead

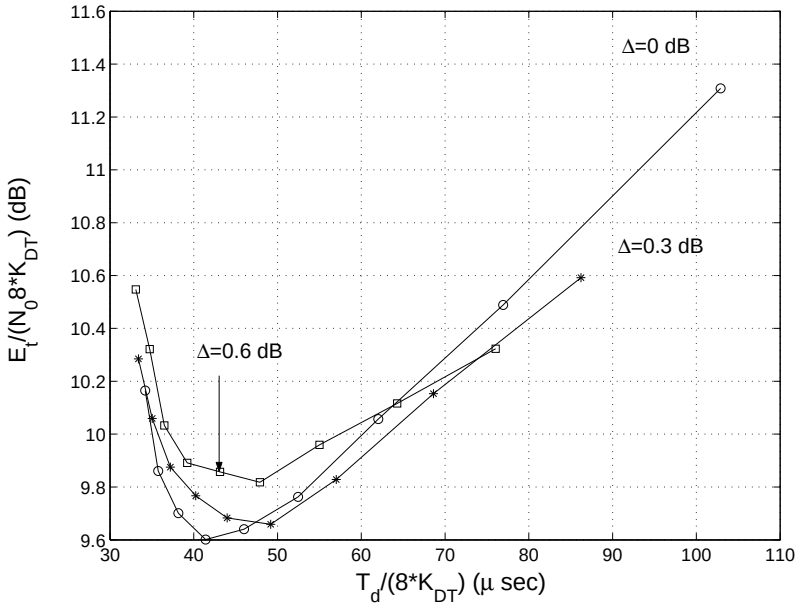

Figure 3.7: Comparison of energy and delay curves for different Δ with $n = 10$ and $K_{DT} = 64$.

transmissions such as RTS, CTS and ACK will become less significant for larger data packet lengths. Hence, when the data packet length is large, we can almost assume that delay or energy consumption comes solely from K_{DT}.

3.2 802.11 Protocol with ARQ

Recently there has been considerable interest in the design and performance evaluation of wireless local area networks (WLANs). According to the OSI protocol layers specification, the physical layer (lowest layer) handles the transmission of bits through a communication link and includes the forward error control (FEC) and modulator/demodulator. The forward error control (FEC) coding technique is used to mitigate the effect of channel noise at the receiver. The error control coding technique reduces the required energy needed for transmission at the expense of increased delay and reduced data rate. The second layer is the data link layer which is responsible for establishing a reliable and secure logical link over the unreliable wireless link.

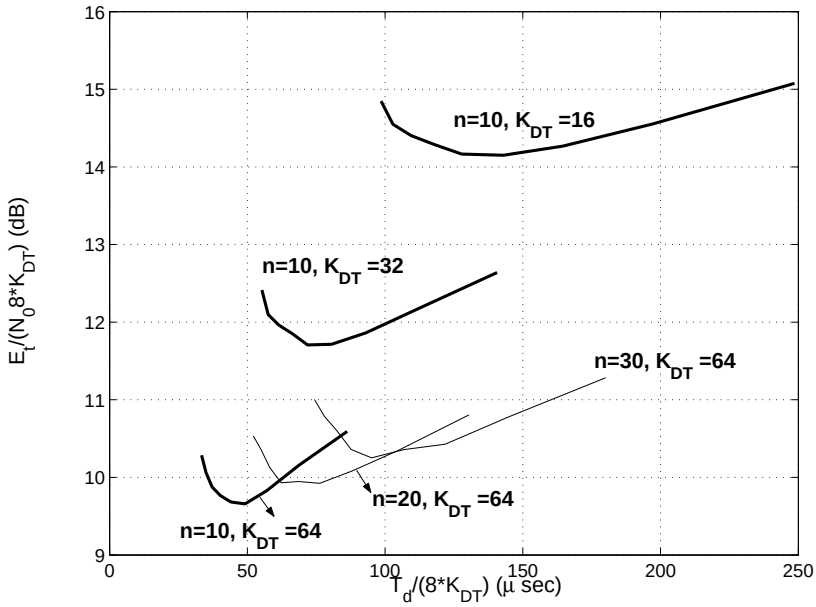

Figure 3.8: Comparison of energy and delay curves for different n and K_{DT} with $\Delta = 0.3(dB)$.

The lower sublayer of data link layer is the medium access control (MAC) protocol layer which resolves conflicts between two users attempting to access the channel. Thus there is some amount of energy and delay spent in reserving the channel. The upper sublayer of the data link layer (DDL) is logical link control (LLC) sublayer that implements error control involving feedback from the receiver. The most common technique to combat errors used in this sublayer is automatic repeat request (ARQ). Given the energy used per data packet, a wireless communication system adopting ARQ scheme can provide higher protection by allowing larger retransmission attempts. However, a system with larger retransmission attempts needs longer delay to transmit the same data packet. If we want to decrease the delay to transmit the same data packet, the system requires higher energy per data packet to achieve the same reliability as a system with lower retransmission attempts. Therefore, there is an energy and delay tradeoff problem happened in both layers (PHY and DLL).

Several papers have investigated the effects of using ARQ techniques in cross-layer

design[2] of wireless networks. In [21], the authors explored the possibilities of cross-layer design as a candidate tool for performance optimization, focusing on problems deriving from the usage of ARQ schemes in today's wireless networks and proposed a novel and general cross-layer interaction scheme between the transport and the link layers. The proposed design achieves the performance in the framework of a IEEE 802.11-based multi-hop wireless network which demonstrates the value of the proposed framework. In [10], a new cross-layer ARQ algorithm for video streaming over 802.11 wireless networks was presented. The algorithm combines application-level information about the perceptual and temporal importance of each packet into a single priority value, which drives packet selection at each retransmission opportunity. Hence, only the most most perceptually important packets are retransmitted, delivering higher perceptual quality and less bandwidth usage compared to the standard 802.11 MAC-layer ARQ scheme. Their results showed that the proposed method consistently outperforms the standard MAC-layer 802.11 retransmission scheme, delivering more than 1.5 dB PSNR gains using approximately half of the retransmission bandwidth. In [45], the authors proposed a class-based adaptive ARQ scheme with QoS support for multimedia traffic to increase the throughput through the reduction of MAC overhead. Different QoS support in the MAC layer is implemented by using service differentiation and traffic class prioritization. An OPNET simulation model is used to show that the proposed ARQ enhancement increases system performance.

In a wireless environment, the value of packet error probability can affect the overall system performance significantly. However, in all above described papers, the issue of energy consumption for each kind of packet error probability used in IEEE 802.11 (RTS, CTS, DATA and ACK packets) is not discussed. Moreover, none of

[2] A network system design methodology which considers at least two different function layers jointly during the design task.

the above papers consider the energy and delay consumption jointly of transmitting a data packet by considering the effects of physical layer and data link layer characteristics. The contribution of this section is that we propose a state diagram to represent the operation of the system that allows us to analyze the energy and delay consumption of the system. We determine the generating function of the state diagram and use it to characterize the joint energy and delay distribution of the system incorporating physical layer and data link layer characteristics. By using a random coding bound we can optimize the system performance over the code rate and energy per coded bit. Although we use random coding for an AWGN channel to represent packet error probabilities, the framework of our analysis can be used for other coding and modulation schemes and various wireless channel.

The rest of this section is organized as follows. In Subsection 3.2.1, we introduce the system assumptions and give a brief description for the protocol used in our analysis. In Subsection 3.2.2, we introduce the generating function and determine the joint generating function for energy and delay. In Subsection 3.2.3, we present numerical results to demonstrate the energy-delay tradeoff with random coding.

3.2.1 System Description

The operation of MAC protocol will be the same as 802.11 standard except the transmission of data packets. Suppose each transmitter has K (fixed) information bits to be transmitted. These K bits will be encoded to form a packet with length N_1 and transmit. If the receiver decodes this packet correctly, it sends an ACK to the transmitter. Otherwise, if decoding is not correct, after the expiration of a timeout the transmitter encodes the same number of information bits K to a packet of length N_2 and transmits it again. This procedure will continue (using length N_l packets for the l-th transmission) until either the transmitter receives the ACK correctly or the

total number of attempts achieves a maximum allowed number, d.

3.2.2 Analysis

In this section, we analyze the energy and delay characteristics of the wireless networks operated according to previous subsection. The delay T_d and E_t of each data packet is defined as before.

We begin by calculating some important event probabilities that are crucial in later analysis. The performance evaluation is based on the Markov chain model described in [5] and [22]. The state of the Markov chain represents the value of the backoff counter and the contention window stage. Our model differs from the model proposed in [5, 22] in that packet error probabilities are included in the CW stage transition probability. By modifying the Markov chain model described in [5, 22], we can incorporate the packet error probabilities into the analysis. We denote the error probability of an RTS packet as $P_{e,RTS}$, a CTS packet as $P_{e,CTS}$ and an ACK packet as $P_{e,ACK}$. The probability P_{e,DT_i} represents the error probability of ith transmission of a data packet, where $1 \leq i \leq d$. We assume that the channel is memoryless between packets. These probabilities depend on the particular channel, channel coding, modulation etc. (a specific example will be given in Section 3.2.3). Let p_{ce} represent the transition probability (as defined in [5]) from one CW stage to the next. This is also the probability of an unsuccessful transmission attempt seen by the transmitting station when its packet is being transmitted on the channel. However, in this work, an unsuccessful transmission can happen not only due to packet collision, but also due to packet errors. There are four possible cases when the channel is sensed busy. The first case is that the RTS packet suffers a collision or packet error. The second case is that the RTS packet is correctly received and collision free but there is an error in the CTS packet transmission. The third case is

that the RTS packet is correctly received and collision free, there is no error in CTS packet, but there is an error in all d transmissions of the DT packet. The last case is that the data packet is correctly received in one of these d transmissions but there is an error in the corresponding ACK packet. Because the event of packet collision and the event of packet error are independent, p_{ce} can be expressed as

$$\begin{aligned} p_{ce} = p_c + (1-p_c)[&P_{e,RTS} + (1-P_{e,RTS})P_{e,CTS} + \\ &(1-P_{e,RTS})(1-P_{e,CTS})\prod_{i=1}^{d} P_{e,DT_i} + (1-P_{e,RTS})\cdot \\ &(1-P_{e,CTS})(\sum_{i=1}^{d}\prod_{j=0}^{i-1} P_{e,DT_j}(1-P_{e,DT_i}))P_{e,ACK}]. \end{aligned} \tag{3.17}$$

where p_c is the packet collision probability and we set $P_{e,DT_0} = 1$. The packet transmission probability p_{tx} which is the probability of a transmitting station sending an RTS packet during each backoff slot, and the collision probability p_c can be expressed as (see [5])

$$p_{tx} = \frac{2(1-2p_{ce})}{(1-2p_{ce})(W+1) + p_{ce}W(1-(2p_{ce})^m)} \tag{3.18}$$

and

$$p_c = 1 - (1-p_{tx})^{n-1}. \tag{3.19}$$

From the above three nonlinear equations, (3.17)–(3.19), one can evaluate the probabilities p_{ce}, p_{tx} and p_c. Another important probability that will be used in our analysis later is the probability of a transmission of an RTS packet from exactly one of the remaining $n-1$ stations given that at least one of the remaining stations is transmitting. It is denoted by p_{tx_1} and can be expressed as

$$p_{tx_1} = \frac{(n-1)p_{tx}(1-p_{tx})^{n-2}}{p_c}. \tag{3.20}$$

We assume that stations spend energy only in packets transmission and that the energy consumption for each packet transmission is proportional to the packet length.

In particular, let αE_c be the energy required to transmit one coded bit, where $\alpha > 1$ is the inverse of the path loss between transmitter and receiver. Then the energy received for a packet of length N (normalized to the thermal noise energy level, N_0) is NE_c/N_0. Similarly, the *transmission* delay of each packet is also proportional to the packet length. Let R_t be the transmitting rate in bits per second. The delay, in seconds, for a packet transmission with packet length N is N/R_t. Based on the protocol description in Section 3.2.1, the state flow diagram shown in Fig. 3.9 can be built.

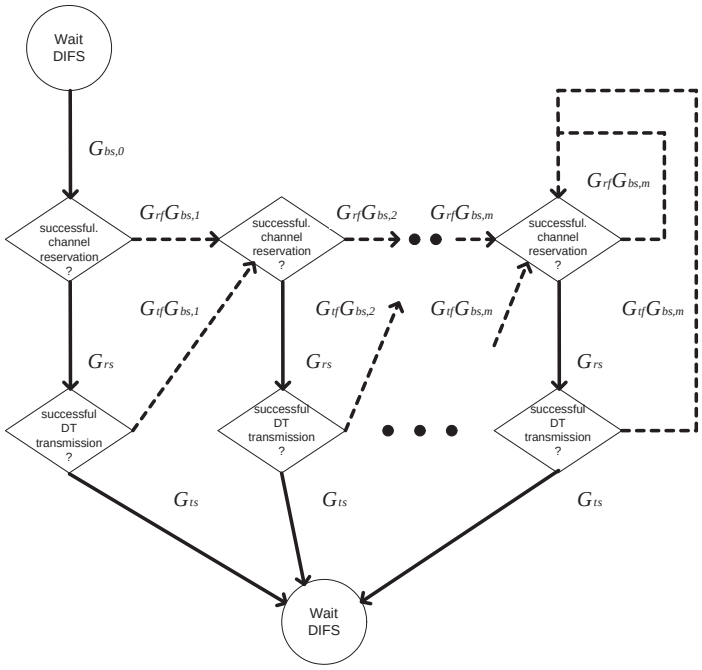

Figure 3.9: State diagram for the protocol. Solid lines represent successful reservation or transmission, while dotted lines represent unsuccessful reservation or transmission.

Let T_{RTS}, T_{CTS}, T_{DT_k} and T_{ACK} denote the time duration for the transmission of RTS, CTS, k-th DT and ACK packets, respectively. Similarly, let E_{RTS}, E_{CTS}, E_{DT_k} and E_{ACK} denote the (received) energy for the transmission of RTS, CTS, k-th DT and ACK packets, respectively. Finally, let T_{DIFS} and T_{SIFS} be system delay parameters defined by the standard. We will now derive the joint generating

function of the delay and energy random variables associated with the transitions in the state diagram in Fig. 3.9. In all the derivations below, functions are of the form $G(X, Y)$, where the variables X, Y are the transform variables of the delay and energy, respectively.

The generating function G_{rs} corresponding to successful channel reservation is

$$G_{rs}(X,Y) = (1-p_c)(1-P_{e,RTS})(1-P_{e,CTS}) \\ X^{T_{RTS}+T_{CTS}+2T_{SIFS}} Y^{E_{RTS}+E_{CTS}}. \tag{3.21}$$

Similarly, the generating function G_{ts} corresponding to successful transmission of a data packet can be expressed as

$$G_{ts}(X,Y) = \sum_{i=1}^{d} (\prod_{j=0}^{i-1} P_{e,DT_j})(1-P_{e,DT_i})(1-P_{e,ACK}) \\ X^{\sum_{k=1}^{i} T_{DT_k}+T_{ACK}+iT_{SIFS}+T_{DIFS}} Y^{\sum_{k=1}^{i} E_{DT_k}+E_{ACK}}. \tag{3.22}$$

We now develop expressions for the generating functions associated with failure to either reserve a channel or to transmit data. As can be seen from the state diagram, these functions are products of two generating functions. Each product contains a factor that is independent of the particular contention stage, and a factor that depends on the contention stage i. We first describe the constant factors. The generating function G_{fr}, corresponding to failure to reserve the channel is given by

$$G_{fr}(X,Y) = [p_c + (1-p_c)P_{e,RTS}] X^{T_{RTS}+T_{DIFS}} Y^{E_{RTS}} \\ + (1-p_c)(1-P_{e,RTS})P_{e,CTS} \\ X^{T_{RTS}+T_{CTS}+T_{SIFS}+T_{DIFS}} Y^{E_{RTS}+E_{CTS}}, \tag{3.23}$$

the meaning of which is that a failure can be due to either a collision/RTS transmission error, or a CTS transmission error. Similarly, after reserving the channel,

a failure to transmit a data packet is either due to the data packets transmission error, or the acknowledgement packet transmission error, which is captured by the generating function G_{tf} as follows

$$G_{tf}(X,Y) = (\prod_{i=1}^{d} P_{e,DT_i}) X^{\sum_{k=1}^{d} T_{DT_k} + (d-1)T_{SIFS} + T_{DIFS}} \cdot$$
$$Y^{\sum_{k=1}^{d} E_{DT_k}} + \sum_{i=1}^{d} (\prod_{j=0}^{i-1} P_{e,DT_j} (1 - P_{e,DT_i}) P_{e,ACK} \cdot$$
$$X^{\sum_{k=1}^{i} T_{DT_k} + T_{ACK} + iT_{SIFS} + T_{DIFS}} Y^{\sum_{k=1}^{i} E_{DT_k} + E_{ACK}}). \tag{3.24}$$

We now evaluate the generating functions denoted by $G_{bs,i}$ of the state diagram. This generating function characterizes the delay for the transmitting station from the instant of starting the backoff procedure to the instant that the backoff counter reaches to zero at stage i. We do not need to consider energy consumption here since the transmitting station stops any transmission during this period. The event of each slot being sensed busy due to the transmission of other stations is independent for each slot and has the same probability, p_c, for each slot (from our previous assumptions). At backoff stage i, the range of possible backoff slots is from 1 to $2^i W$. Let j be the backoff slots randomly chosen uniformly from the above range, then the number of the occupied slots in these j slots is binomially distributed with parameters (j, p_c). Hence, the generating function, $G_{bs,i}$, of backoff procedure at stage i is

$$G_{bs,i}(X) = \sum_{j=1}^{2^i W} \frac{1}{2^i W} \sum_{k=0}^{j} \binom{j}{k} [(1-p_c)X^{\sigma}]^{j-k} (p_c G_{oc}(X))^k, \tag{3.25}$$

where $G_{oc}(X)$ is the generating function for an occupied slot. We define an occupied slot as a slot when the transmitting station senses the channel is busy due to the transmission of one of the remaining $n-1$ stations. Using arguments similar to the

ones used in the derivation of (3.17), we can express $G_{oc}(X)$ as follows

$$\begin{aligned} G_{oc}(X) = {} & [(1-p_{tx_1}) + p_{tx_1}P_{e,RTS}]X^{T_{RTS}+T_{DIFS}} + \\ & p_{tx_1}(1-P_{e,RTS})P_{e,CTS}X^{T_{RTS}+T_{CTS}+T_{SIFS}+T_{DIFS}} + \\ & p_{tx_1}(1-P_{e,RTS})(1-P_{e,CTS})(\prod_{i=1}^{d} P_{e,DT_i})\cdot \\ & X^{T_{RTS}+T_{CTS}+\sum_{k=1}^{d} T_{DT_k}+(d+1)T_{SIFS}+T_{DIFS}} + \\ & p_{tx_1}(1-P_{e,RTS})(1-P_{e,CTS})\sum_{i=1}^{d}\prod_{j=0}^{i-1} P_{e,DT_j}\cdot \\ & (1-P_{e,DT_i})X^{\sum_{k=1}^{i} T_{DT_k}+T_{ACK}+(i+2)T_{SIFS}+T_{DIFS}}. \end{aligned} \tag{3.26}$$

From the state diagram and Mason's gain formula, we obtain the following backward recursive generating function to characterize the energy and delay of the system. Let G_s be the system generating function. Then the recursion is

$$\begin{aligned} f_m(X,Y) &= \frac{G_{rs}G_{ts}}{1-G_{fr}G_{bs,m}-G_{rs}G_{tf}G_{bs,m}} \\ f_{i-1}(X,Y) &= G_{rs}G_{ts} + (G_{fr}+G_{rs}G_{tf})G_{bs,i}f_i \\ G_s(X,Y) &= G_{bs,0}f_0, \end{aligned} \tag{3.27}$$

where i is the index of the CW stage from 1 to m. Hence, we can evaluate any joint statistics of the energy and delay of a successful packet transmission. In particular, the mean values are given by

$$\overline{T}_d = \frac{\partial G_s}{\partial X}\mid_{X=Y=1}, \qquad \overline{E}_t = \frac{\partial G_s}{\partial Y}\mid_{X=Y=1}. \tag{3.28}$$

3.2.3 Numerical Results

We will use channel reliability based bounds for an AWGN channel to estimate the packet error probability. Let K be the number of information bits in a packet (specified by the 802.11 standard) and N be the number of coded bits in a packet. Then

Table 3.2: System Parameters for Numerical Results

K_{RTS}	128 bits
K_{CTS}	128 bits
K_{ACK}	128 bits
Channel Bit Rate, R_t	1 Mbit/s
Slot Time, σ	50μs
T_{SIFS}	28μs
T_{DIFS}	128μs

there exist an encoder and decoder with packet error probability $P_{e,K,N}$ bounded as

$$P_{e,K,N} \leq 2^{K-NR_0}, \tag{3.29}$$

where $R_0 = 1 - \log_2(1 + e^{-E_c/N_0})$ is the cutoff rate determined by SNR.

The energy-delay curves of the IEEE 802.11 protocol with SWARQ-FT ($d = 2$) is evaluated for different number of users and data packet lengths. For comparison, we also present similar curves for the original 802.11 protocol, which is essentially an SWARQ-FT protocol with $d = 1$. For both protocols there is an optimal N for each kind of packets. For small N the packet error probability is large, which increases the system delay and energy due to the high chance of packet retransmission. On the other hand, if N is large, the packet error probability is small but the transmission time is large. Our goal is to find the best N for each kind of packet to minimize system delay. We first fix E_c/N_0 and minimize $\overline{T}_d$ from (3.28) to get the corresponding optimal packet lengths $N^*_{RTS}, N^*_{CTS}, N*_{DT}$ and N^*_{ACK} (note that in the case of $d = 2$ we need to optimize over two data packet lengths, N_{DT1}, and N_{DT2}). Using these values, we can evaluate both average delay and average energy consumption. Repeating the above procedure for different E_c/N_0, we get the energy-delay curves. Table 3.2 summarizes the system parameters used in our numerical evaluations.

We first plot in Fig. 3.10 the energy-delay curves for the original protocol and the IEEE 802.11 MAC Protocol with SWARQ-FT ($d = 2$). We observe that the energy-delay curves for two protocols are almost identical. This shows that if we can select

packet lengths properly, the SWARQ-FT does not give significant improvement. To compare these two protocols with respect to their robustness to SNR, we fix the packet lengths, vary E_c/N_0 and plot energy and delay for both protocols. We observe that both systems are equally and very sensitive to SNR. Essentially, the reduction in delay the SWARQ achieves by immediately retransmitting an erroneously received packet, is compensated by the increased delay needed to access the channel, since other users use the same protocol. While ARQ provides some robustness to SNR uncertainty in conventional single user system, this is not the case in this multiuser scenario.

In Fig. 3.11, the energy-delay tradeoff is plotted for the SWARQ protocol for different values of quantity $N_{DT} - K_{DT}/R_0(E_c/N_0)$. This figure shows that the optimal packet lengths are on the order of 60 bits larger than the minimum length implied by the cutoff rate. For high SNR values as small as 10 will provide almost optimal performance, while for low SNR, the number of redundant bits has to increase.

3.3 Conclusion

In this chapter, we applied our generating function method to characterize the energy and delay of a wireless network for two proposed protocols modified from 802.11 standard. The first proposed protocol is related to the adaptive energy scheme which is implemented by increasing the energy level per coded bit due to previous failure transmission attempts. The numerical results indicated that both throughput and energy efficiency are improved significantly when the channel SNR is low. The second proposed protocol is to incorporate the IEEE 802.11 protocol with SWARQ-FT. Our results demonstrated that both protocols (original IEEE 802.11 and IEEE 802.11 with SWARQ-FT) are sensitive to the knowledge at the transmitter of the

received SNR. This is a consequence of the steep falloff of packet error probability with respect to SNR for random coding. We expect similar results when LDPC or turbo codes are used. Our methods can be extended to the analysis of other protocols or to consider a wireless network system where more function layers are involved by representing the operation of a protocol with a system state diagram. Once the system state diagram is obtained, we shall be able to derive the joint generating function of interesting random variables for the state diagram. In this thesis, we select energy and delay consumption as our interesting variables since we want to determine the energy and delay tradeoff. However, one may select other variables in the generating function for their analysis purpose, such as event's cost, event's reliability, etc. Therefore, the generating function approach allows one to find the entire system performance distribution based on the performance distribution of its elements by using algebraic procedures in a systematic way.

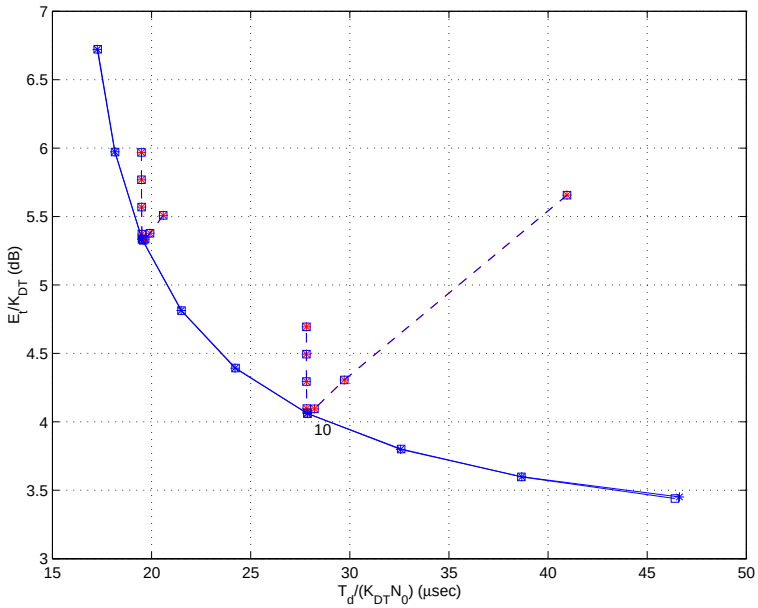

Figure 3.10: Energy-delay curves for $n = 10$ and $K_{DT} = 1400$. Solid lines represent the curves after packets optimization. Dashed lines represent the energy-delay curves for E_c/N_0 of 0 and 3 dB using the optimal packet lengths.

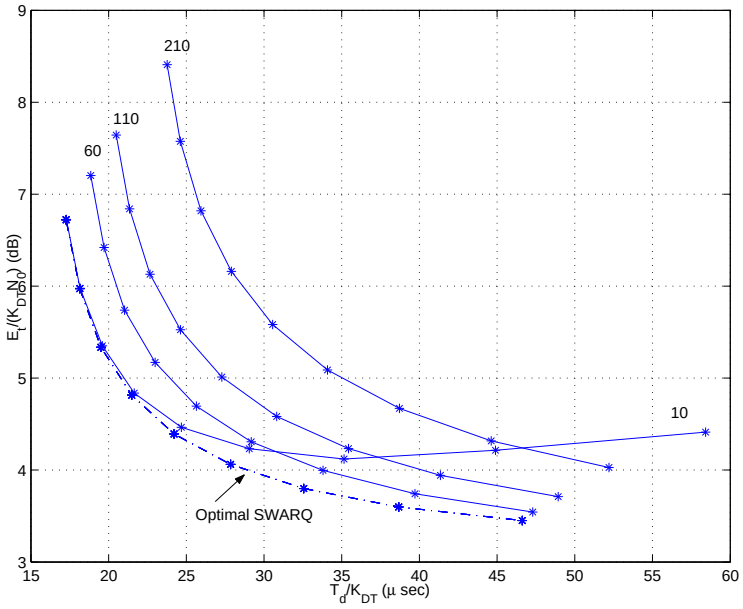

Figure 3.11: The dashed lines represent SWARQ after optimization. The number beside the curve is the redundant bits.

CHAPTER IV

Analysis of Energy and Delay for ARQ Systems over Time Varying Channels

4.1 Introduction

Error control techniques can generally be classified either as forward-error-correction (FEC) or as automatic-repeat-request (ARQ). Although FEC can provide a communication system with a fixed throughput which is equal to the code rate, the disadvantage of a conventional FEC system without feedback is that the transmitter does not adjust the error correction capability according to the channel variations. Thus the code rate must be low enough for the worst-case channel. An ARQ system changes the error correction capability to adapt to the channel variation by using a feedback channel. Therefore, ARQ error control systems are more suitable than FEC systems for error control in data communication systems having time varying channels where a feedback channel is available [35, 47]. The advantages of both FEC and ARQ techniques can be obtained by combining the techniques within a single hybrid-ARQ protocol [34].

In general, ARQ protocols are classified into three basic schemes: stop-and-wait (SW), go-back-N (GBN), and selective-repeat (SR). In SW-ARQ, the transmitter must receive the ACK of a packet before transmitting the next packet. This scheme preserves the order of packets but results in low channel utilization if the round-

trip delay is large. In GBN-ARQ, packets are transmitted continuously without waiting for ACKs/NACKs. Upto to N subsequent packets can be transmitted before receiving an ACK/NACK for a given packet. If a NACK is received, the transmitter retransmits the negatively acknowledged packet and all subsequent packets regardless of their acknowledgments. The round trip delay is assumed to be smaller than N times the packet duration and the transmitter is capable of storing N packets. In SR-ARQ, packets are transmitted continuously as in GBN, but only negatively acknowledged packets are retransmitted. Of the three schemes, SR-ARQ achieves the highest throughput but requires the most feedback.

For a fixed amount of energy used per information bit, a wireless communication system adopting FEC can provide higher protection by using a lower code rate. However, a system with lower code rate needs longer delay to transmit the same information packet due to the fixed bandwidth. If we want to decrease the delay required to transmit the same information packet, a system with high code rate requires higher energy per information bit to achieve the same reliability as a system with lower code rate. Similarly, an ARQ system requires higher energy to transmit an information packet successfully with fewer trials (shorter delay to transmit a packet). On the other hand, we may use more attempts (longer delay) to transmit the same information packet successfully and use smaller energy. Therefore, there is a fundamental tradeoff between energy and delay in a wireless communication system.

Many adaptive ARQ protocols have been suggested in the literature to improve the system delay (or equivalently throughput). In [9, 37, 50], different block (packet) sizes and multicopy transmission schemes were used as adaptation mechanisms, respectively. On the other hand, in [44, 26, 40], the authors suggested to vary the

FEC code rate to compensate for the variations in the channel conditions. Although the above papers consider the analysis based on time varying channels, they do not consider the energy consumption as part of the system performance. In [38], the authors investigated the energy savings of an ARQ strategy within two different time varying channels; fast Rayleigh fading and log-normally shadowed Rayleigh fading. However, their work does not provide the exact method to evaluate the system delay. Another important issue that previous research has not addressed is the relation of the channel transition probability with the transmission delay of a packet. Most previous papers just assume that the channel transition probability is the same for different packet lengths. However, this assumption is not reasonable if the variation of packet lengths is large.

In this chapter, we propose an analytical method to determine the joint distribution of the energy and delay of automatic repeat request (ARQ) schemes over time varying channels. By extending the traditional Markov chain [13] that models the channel as well as the transmitter, the overall system operation can be represented by a state diagram that takes into account the joint operation of the transmitter, receiver and channel. The model of the transmitter accounts for different packet lengths used by the transmitter and the model of the receiver accounts for the receiver memory [27]. We determine the joint generating function of the energy consumed and delay incurred of an ARQ system from its state diagram. The joint generating function enables us to characterize the joint energy and delay distribution of the system incorporating physical layer characteristics (packet error probability as a function of energy and delay). The effect of packet length on the channel transition probability is also investigated. As the numerical results demonstrate, the time-varying characteristics of the channel have a great influence on system performance especially at

low channel SNR.

The rest of the chapter is organized as follows. In Section 4.2, we model an ARQ system with a finite state machine (FSM) by considering the transmitter, receiver and channel states jointly. A general ARQ system which allows different packet lengths in each retransmission is discussed and analyzed in Section 4.3. Several practical ARQ systems with specific packet lengths are considered in Section 4.4. The detailed derivation of the generating functions for systems discussed in Subsection 4.4.1 and 4.4.2 will be presented in the appendices. Another kind of ARQ system, go-back-N ARQ (GBN-ARQ), will be investigated in Section 4.5 and its joint generating function for energy and delay over time varying channels will be derived. In Section 4.6, the cutoff rate for two kinds of receiver structures that contain memory is derived. In Section 4.7, we present numerical results for the performance by considering the energy and delay relationship with random coding for Gilbert-Elliott channels. The conclusions are given in Section 4.8.

4.2 FSM Model, Channel Model and Assumptions for ARQ Systems

A finite state machine (FSM) or finite automaton is a model of a system which is composed of states, transitions and actions. Originally defined in the automata theory, FSM's are used in the theory of computation [12]. Finite state machines are very widely used in modeling of application behavior, design of hardware digital systems, software engineering, study of computation and languages. A state stores a cumulative information about the past, i.e. it reflects the input changes from the system start to the present moment. The content of a state depends on the characteristics and behaviors of the system that we are interested in. A transition indicates a state change and is described by a condition that would need to be fulfilled

to enable the transition. An action is a description of an activity that is to be performed at a given state. A finite state machine is a sextuple $< \Sigma, \Gamma, S, s_0, \delta, \omega >$, where:

- Σ is the alphabet of the input variables.
- Γ is the alphabet of the output variables.
- S is a finite non empty set of states.
- s_0 is an initial state.
- δ is the state transition function: $\delta : S \times \Sigma \rightarrow S$.
- ω is the output function $\omega : S \times \Sigma \rightarrow \Gamma$.

By extending the traditional Markov chain that models the channel state, the overall ARQ system operation is represented by a modified FSM that also takes into account the states of the transmitter and the receiver. The states of the transmitter can be used to model the different coding or modulation schemes and the states of the receiver can be utilized to model the receiver memory content or different decoding schemes. For our communication system model, the state for the transmitter, channel and receiver we be consider jointly in the system state space S. Also for our purpose the output of modified transition functions δ, i.e., the next state, is a random state with some probability distribution due to the uncertainty of the channel characteristic. To determine the energy and delay generating function of the system, the output at each state is used to characterize the energy and delay consumed at the state. By using the concept of signal flow proposed by Mason [20], the total energy and delay generating function from the initial state to the final state can be obtained from Mason's gain formula.

Gilbert-Elliott Channel Model

To model time-varying channels, the Gilbert-Elliott channel model is widely used [26]. It uses two states to characterize the channel condition, the state for good channel condition is designated by G and the state for bad channel condition is designated by B. In each channel state, we model the channel as AWGN channel and assume that the channel signal-to-noise ratio (SNR)[1] of good state is SNR_G and the channel SNR of bad state is SNR_B. We define r as the ratio of the channel SNR between good and bad states, i.e., $r = \frac{\mathrm{SNR}_G}{\mathrm{SNR}_B}$. We assume that the duration of the good and bad states are exponentially distributed with mean $1/\lambda$ and $1/\mu$, respectively. The value λ is the transition rate from good to bad channel condition and μ is the transition rate from bad to good channel condition. From the above system parameters, we can determine the average channel SNR (SNR_{avg}), i.e., $\mathrm{SNR}_{avg} = \frac{\mu \mathrm{SNR}_G + \lambda \mathrm{SNR}_B}{\lambda + \mu}$. The Gilbert-Elliott channel model can be specified completely if the following four parameters are given : $(r, \mathrm{SNR}_{avg}, \lambda, \mu)$. The crossover probabilities between channel states are determined by the time duration of occupying the current state before the event of next channel transition and the channel transition rates. If the time duration for a packet transmission is T_* where the subscript is used to distinguish different packet types, then the channel transition probability from good state to good (bad) state is $e^{-\lambda T_*}$ $(1 - e^{-\lambda T_*})$. Similarly, the channel transition probability from bad state to bad (good) state is $e^{-\mu T_*}$ $(1 - e^{-\mu T_*})$. The channel model is shown in the Fig. 4.1

General ARQ Systems

A general stop and wait (SW) ARQ protocol is described as follows. Whenever there is an information packet with length K generated, it will be encoded to a

[1]The SNR is the power ratio between a signal (meaningful information) and the background noise.

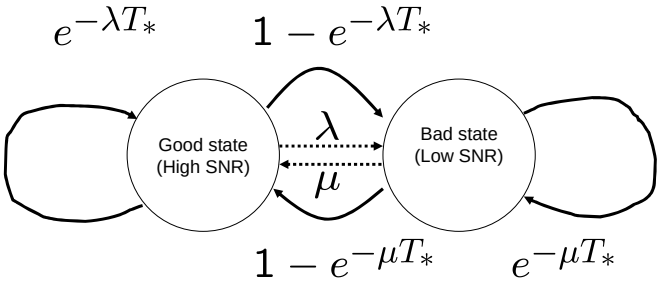

Figure 4.1: Gilbert-Elliott channel model for time varying channel.

data packet with length N_1. The coded bits are passed to the modulator. The modulator maps the input sequence of bits to a signal waveform. We will assume binary phase-shift keying (BPSK) modulation scheme is used by the transmitter. After the modulated waveform is generated, it is propagated through the channel. We use the symbol T_b to represent the time duration needed to transmit a coded bit and the symbol E_c to indicate the energy consumption required to transmit a coded bit.

The demodulator of the receiver is implemented by a maximum a posterior probability (MAP) criterion in determining the transmitted signal. After demodulation, the data packet will be decoded. If the data packet is decoded correctly at the receiver, the receiver will send an ACK packet back to the transmitter. If it is decoded incorrectly at the receiver, a NACK packet will be sent back to the transmitter and the received signal samples corresponding to the erroneously decoded packet will be stored in the receiver memory. Then the transmitter will encode the same information packet again to a data packet with length N_2 and transmit it. After receiving this packet with length N_2, the receiver will decode this new packet jointly with those packets stored in the receiver memory. If the data packet is decoded correctly at the receiver, the receiver will send an ACK packet back to the transmitter. If it is de-

coded incorrectly at the receiver, a NACK packet will be sent back to the transmitter and the new erroneously decoded packet (N_2) will be stored in the receiver memory. After receiving the NACK packet, the transmitter will encode the same information packet with length N_3 and retransmit. The above procedure will be repeated until an ACK is received by the transmitter. Because we allow different packet lengths used in each retransmission, therefore, the sequence of packet lengths in transmission given an information packet has the following form: $\{N_1, N_2, N_3, \cdots\}$. If we use notation $\mathbb{N}$ to represent a set of natural numbers, then the set of $\{N_1, N_2, N_3, \cdots\}$ can be written as $\{N_i | i \in \mathbb{N}\}$. When an ACK is received by the transmitter, a new coded packet is transmitted with length N_1. Let m_R be the maximum number of packets that the receiver can store in its memory. If the number of consecutive packet errors is greater then m_R, the receiver memory will take some action to keep the number of stored packets smaller or equal than m_R. For example, the receiver may update the receiver memory content by storing the newest erroneous packet and discarding the oldest erroneous packet.

We make following assumptions about our system: (1) The ACK and NACK packets are error free, (2) the round-trip delay and the transmission time of the ACK and NACK packets are negligible, (3) delay is incurred only during data packet transmission, (4) energy is consumed only during data packet transmission, the transmission energy of ACK and NACK packets are negligible, (5) the Gilbert-Elliott channel model will be adopted. (6) the event of packet errors is independent of the event of channel transition. However, it is straightforward to consider the cases with ACK/NACK packet error and the round-trip delay under our analytical framework.

4.3 Generating Function Analysis for General ARQ Systems

We will derive the generating function of energy and delay consumption for a general ARQ system in this section. For the notational clarity, the corresponding elements of the FSM for a generalized ARQ system are listed as follows :

- (Input variables) Since the system will change its state when an ACK or NACK packet is received by the transmitter, we use Σ to represent the input space, i.e., $\Sigma = \{\text{ACK}, \text{NACK}\}$.

- (Output variables) For the requirements of the generating function approach for the energy and delay consumption between each state transition (as shown in Chapter 2), the symbol Γ is used to represent the output space, i.e., $\Gamma = \{X^{N_i T_b} Y^{N_i E_c} | i \in \mathbb{N}\}$.

- (State space) Because the transmitter can use different packet lengths for each retransmission, we use S_T to represent the state space for the transmitter, i.e., $S_T = \{N_i | i \in \mathbb{N}\}$. The channel state space for one packet transmission is denoted as S_C and $S_C = \{G, B\}$ since the Gilbert-Elliott model is adopted.

  The state space for the receiver represents the lengths of the packets stored in the memory and the corresponding channel condition. If a memory location does not contain a packet, e.g., at startup, then we represent that by a packet length N_0. The state space for the receiver is denoted by S_R which is a collection of vectors of length $2m_R$. The first pair contains the channel state and packet length of the most recently received erroneous packet. The next pair contains the channel state and packet length of the previous erroneously received packet. For example, the state of the receiver for memory with $m_R = 3$ might be $[G, N_3, B, N_2, G, N_1]$. After receiving a packet with length N_4 trans-

mitted in the bad channel condition, the next state of the receiver memory will be $[B, N_4, G, N_3, B, N_2]$. We will use use a shorthand notation for the state of the receiver memory by representing G_i as a packet of length N_i received when this packet is transmitted in the good channel condition. So the receiver state $[G, N_3, B, N_2, G, N_1]$ can be written compactly as $[G_3, B_2, G_1]$. Therefore, the total state space of the system which contains the information about the channel, the transmitter, and the receiver is denoted as vectors of length $2(m_R + 1)$. The first component of this vector represents the channel state and the second component of this vector represents the transmitter state. The remaining $2m_R$ components represent the receiver memory state. A state of the system, denoted as s, can be written as $s = [s_C, s_T, s_R]$. For notational simplicity, the total system state can also be expressed as $[G_i, s_R]$ ($[B_i, s_R]$) for the case that the transmitter sends a packet with length N_i in good (bad) channel condition given that the receiver state is s_R. In Fig. 4.2, we show an example about the transition of states for a system with $m_R = 2$ with the sliding window operation in the receiver memory. The detail about the sliding window operation in the receiver memory will be presented in subsection 4.4.1.

For later derivation convenience, the state space of packet lengths for the receiver is denoted by $\widetilde{S}_R$ which is a collection of vectors of length m_R. The first component of $\widetilde{S}_R$ contains the packet length of the most recently received erroneous packet. The next component of $\widetilde{S}_R$ contains the packet lengths of the previous erroneously received packets. For example, if the state space for the receiver is $[G, N_3, B, N_2, G, N_1]$, then the the corresponding state space for the packet lengths is $[N_3, N_2, N_1]$. Given an element from $\widetilde{S}_R$, say $\widetilde{s}_R$, and we assume that there are a packet lengths in the vector $\widetilde{s}_R$. In order to consider all possibil-

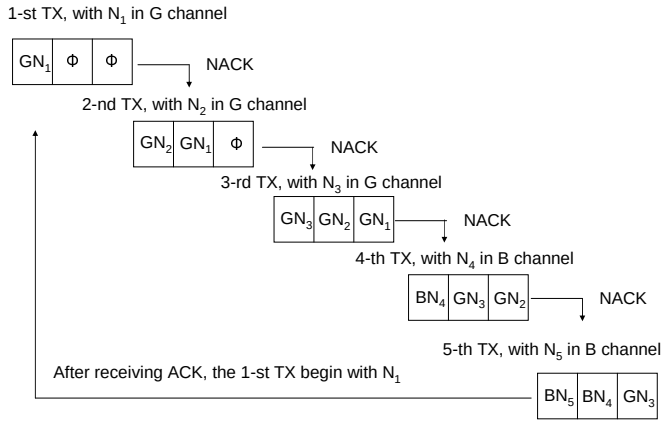

Figure 4.2: An example for a general ARQ system with $m_R = 2$. The first slot represents the state of the channel and transmitter. The second and third slots represent the state of the receiver memory. The second (third) slot is used to represent the first (second) position of the the receiver memory.

ities for the receiver memory states that have these a packet lengths, the index l for $1 \leq l \leq 2^a$, called channel realization index, is used to distinguish different memory states introduced by the randomness of the channel. The channel realization index l is assigned to the set of all channel realizations of these a packets according to the lexicographically order by assuming $G > B$ and the position order of the receiver memory. The symbol $\widetilde{s}_R(l)$ represents a state of the receiver memory that the state of packet lengths for the receiver is $\widetilde{s}_R$ and the channel realization index is l. For example, if $m_R = a = 2$ and $\widetilde{s}_R = [N_3, N_3]$, then all possible states of the memory are $\{[G_3, G_3], [G_3, B_3], [B_3, G_3], [B_3, B_3]\}$. By the above ordering regulation, we have $\widetilde{s}_R(1) = [G_3, G_3]$, $\widetilde{s}_R(2) = [G_3, B_3]$, $\widetilde{s}_R(3) = [B_3, G_3]$ and $\widetilde{s}_R(4) = [B_3, B_3]$. We have two ways to represent the memory content; one is s_R and the other is $\widetilde{s}_R(l)$. Each element of S_R may differ in the number of packets stored in the receiver memory. However, all memory states associated with $\widetilde{s}_R(l)$ have the same number of packets and packet length indices stored in the receiver memory for each channel realization index l.

- (Initial state) The initial state is represented by the symbol s^{ini}. This is a state used at the first time transmission of an information packet. If the memory content is empty at the first transmission, then $s^{ini} = G_1$ if the initial channel condition is good and $s^{ini} = B_1$ if the initial channel condition is bad.

- (State transition function) The state transition function, δ, will define the transition between states for different input. It will map an element of space $S \times \Sigma$ to an element of space S. If the input is a NACK after the k-th transmission, then the state of the system s^k will be mapped into the state s^{k+1} by the transition function δ as $\delta(s^k, \text{NACK}) \rightarrow s^{k+1}$. The k-th transmission is counted from the transmission of the original K information bits with packet length N_1 (1-st transmission). If the input is an ACK, then the state of s^k will be mapped into the state of s^{ini} by function δ as $\delta(s^k, \text{ACK}) \rightarrow s^{ini}$. Since $s = [s_C, s_T, s_R]$, we also have $\delta([s_C^k, s_T^k, s_R^k], \text{NACK}) \rightarrow [s_C^{k+1}, s_T^{k+1}, s_R^{k+1}]$ and $\delta([s_C^k, s_T^k, s_R^k], \text{ACK}) \rightarrow [s_C^{ini}, s_T^{ini}, s_R^{ini}]$. The state transition function δ will be determined by specified protocol and this is often represented by a state diagram.

- (Output functions) The output function will map a system with state s^k to $X^{N_k T_b} Y^{N_k T_c}$ since the packet length used in the k-th transmission is N_k.

Recall that an event is a set of random outcomes to which a probability is assigned. For the energy and delay analysis of ARQ systems, we then define the joint generating function of an event as:

$$G_{event}(X, Y) \quad = \quad Pr(\text{event}) X^{T_{event}} Y^{E_{event}}, \tag{4.1}$$

where $T_{event}(E_{event})$ is the required delay (energy) for this event. If we are interested in evaluating the generating function for a set of disjoint events, then the generating

function for all events in this group has the following form

$$G_{group}(X,Y) = \sum_{i \in group} Pr(event_i) X^{T_{event_i}} Y^{E_{event_i}}, \tag{4.2}$$

where i is the index for events in this group and $T_{event_i}(E_{event_i})$ is the required delay (energy) for this i-th event.

We use symbol $E_{f,G_i|s_R^i}$ to represent the event that the transmitter receives a NACK (failure) after the i-th transmission of a given packet in the good channel condition given that the receiver memory state is s_R^i. Note that the state space of the system can be expressed as $[G_i, s_R^i]$ when this event happens because the packet length used in the i-th transmission is N_i. The generating function of this event is represented by $G_{f,G_i|s_R^i}$ according to (4.1). Similarly, $G_{f,B_i|s_R^i}$ is used to represent the generating function for the event that the transmitter receives a NACK after the i-th transmission in bad channel condition given that the receiver memory state is s_R^i. The probability $P_{e,G_i|s_R^i}$ ($P_{e,B_i|s_R^i}$) is used to represent the packet error probability for the i-th transmission in the good (bad) channel condition given that the memory state is s_R^i. If the memory content is empty at i-th transmission, the probability P_{e,G_i} (P_{e,B_i}) is used to represent the packet error probability for the i-th transmission in the good (bad) channel condition. We will use the shorthand notation, $\overline{\alpha} = 1 - \alpha$, for the following derivation.

Because the i-th (for $i \geq 2$) unsuccessful transmission is preceded by $(i-1)$ unsuccessful transmissions, the recursive formulas for $G_{f,G_i|s_R^i}$ and $G_{f,B_i|s_R^i}$ for ARQ systems can be expressed as

$$\begin{aligned} G_{f,G_i|s_R^i}(X,Y) = \sum_{s_R^{i-1} \in S_R} \Big[& G_{f,G_{i-1}|s_R^{i-1}}(X,Y) e^{-\lambda N_{i-1} T_b} I(\delta([G_{i-1}, s_R^{i-1}], \text{NACK}) = [G_i, s_R^i]) + \\ & G_{f,B_{i-1}|s_R^{i-1}}(X,Y) \overline{e^{-\mu N_{i-1} T_b}} I(\delta([B_{i-1}, s_R^{i-1}], \text{NACK}) = [G_i, s_R^i]) \Big] \cdot \\ & P_{e,G_i|s_R^i} X^{N_i T_b} Y^{N_i E_c} \end{aligned}$$

$$
\begin{aligned}
G_{f,B_i|s_R^i}(X,Y) \;=\; & \sum_{s_R^{i-1}\in S_R} \Big[G_{f,G_{i-1}|s_R^{i-1}}(X,Y)\overline{e^{-\lambda N_{i-1}T_b}} I(\delta([G_{i-1},s_R^{i-1}],\text{NACK})=[B_i,s_R^i]) + \\
& G_{f,B_{i-1}|s_R^{i-1}}(X,Y)e^{-\mu N_{i-1}T_b} I(\delta([B_{i-1},s_R^{i-1}],\text{NACK})=[B_i,s_R^i])\Big] \cdot \\
& P_{e,B_i|s_R^i} X^{N_iT_b} Y^{N_iE_c} \qquad (4.3)
\end{aligned}
$$

where $I(\delta(s^{i-1},\text{NACK})=s^i)$ is an indicator function. The function $I(\delta(s^{i-1},\text{NACK})=s^i)$ is 1 if the inside condition is true and $I(\delta(s^{i-1},\text{NACK})=s^i)=0$ if the inside condition is false.

By assuming that the system begins with the state space $[G_1]$ or $[B_1]$ (the receiver memory content is empty) and the channel has reached a stationary condition, we set $G_{f,G_1|s_R^1}=\frac{\mu}{\lambda+\mu}P_{e,G_1}X^{N_1T_b}Y^{N_1E_c}$ and $G_{f,B_1|s_R^1}=\frac{\lambda}{\lambda+\mu}P_{e,B_1}X^{N_1T_b}Y^{N_1E_c}$. Since the $(i+1)$-th successful transmission is preceded by i unsuccessful transmissions and a successful transmission at the $(i+1)$-th trial, the total generating function for a successful transmission can be written as

$$
\begin{aligned}
G_s(X,Y) \;=\; & \frac{\mu}{\lambda+\mu}\overline{P_{e,G_1}}X^{N_1T_b}Y^{N_1E_c} + \frac{\lambda}{\lambda+\mu}\overline{P_{e,B_1}}X^{N_1T_b}Y^{N_1E_c} + \sum_{i=1}^{\infty}\Big\{ \sum_{s_R^i\in S_R}\sum_{s_R^{i+1}\in S_R} \\
& \Big[G_{f,G_i|s_R^i}e^{-\lambda N_iT_b}\overline{P_{e,G_{i+1}|s_R^{i+1}}} I(\delta([G_i,s_R^i],\text{NACK})=[G_{i+1},s_R^{i+1}]) + \\
& G_{f,B_i|s_R^i}\overline{e^{-\mu N_iT_b}}\,\overline{P_{e,G_{i+1}|s_R^{i+1}}} I(\delta([B_i,s_R^i],\text{NACK})=[G_{i+1},s_R^{i+1}]) + \\
& G_{f,G_i|s_R^i}\overline{e^{-\lambda N_iT_b}}\,\overline{P_{e,B_{i+1}|s_R^{i+1}}} I(\delta([G_i,s_R^i],\text{NACK})=[B_{i+1},s_R^{i+1}]) + \\
& G_{f,B_i|s_R^i}e^{-\mu N_iT_b}\overline{P_{e,B_{i+1}|s_R^{i+1}}} I(\delta([B_i,s_R^i],\text{NACK})=[B_{i+1},s_R^{i+1}])\Big] \cdot \\
& X^{N_{i+1}T_b}Y^{N_{i+1}E_c}\Big\}, \qquad (4.4)
\end{aligned}
$$

where the index i represents the event that the system fails at the i-th transmission but succeeds in the $(i+1)$-th transmission.

The symbol $\widetilde{s}$ is used to the set of packet lengths transmitted and stored in the receiver, i.e., $\widetilde{s}=[s_T,\widetilde{s}_R]$. For the i_j-th transmission, the packet lengths used

for the transmitter state is $s_T^{i_j}$ and the receiver state is $\widetilde{s}_R^{i_j}$. Let m_j be the minimum number (if it exists) such that $s_T^{i_j+m_j} = s_T^{i_j}$ and $\widetilde{s}_R^{i_j+m_j} = \widetilde{s}_R^{i_j}$. We will group all transmission with transmission indices with form as $(i_j + km_j)$ for $k \in \mathbb{Z} = \{\cdots, -2, -1, 0, 1, 2, \cdots\}$ into a group and denote this group as g_j. The number m_j is called the transmission period of this group. The set H contains those indices which are not belonged to any group g_j. If a set C is the union of two sets A and B and $A \bigcap B = \phi$, we then write $C = A \bigoplus B$. We have the following theorem about our general ARQ systems for the set of all transmission indices $\mathbb{N}$, denoted as TI.

Theorem: If the number of different packet lengths used in the transmitter and the receiver memory size are finite, then the set of TI, can be classified as $TI = H \bigoplus \bigoplus_{j=1}^{m} g_j$ according to the state of packet lengths, s_T and $\widetilde{s}_R$. The transmission indices in the same group have same packet length indices set $\widetilde{s}$ and each group has same period m.

Proof :

Let d be the number of different packet lengths used in the system, then the total possible system states for the system is less then $(2(d+1))^{m_R+1}$, which is a finite integer. Suppose the system begins with the initial state $s^{ini} \equiv s^0$, then $s^k = \delta^k(s^0, \text{NACK})$ represents the system state after k failures for $k \in \mathbb{N}$. We cannot require that the system states are always different after $(2(d+1))^{m_R+1}$ failures by the Pigeonhole Principle. Therefore, the number of m_j must be finite.

By examining the states of packet lengths for all transmission indices from $\widetilde{s}^{ini} = \widetilde{s}^1$, if the state $\widetilde{s}^{i_1+m_1}$ is the first repeating state of any one state for packet lengths state in previous $(i_1 + m_1 - 1)$ transmissions, say $\widetilde{s}^{i_1}$. We claim that the state for packet lengths $\widetilde{s}$ is the same for those indices $(i_1 + km_1)$ for $k \in \{0\} \bigcup \mathbb{N}$. Because the state $\widetilde{s}^{i_1+1+m_1}$ will be equal to the state $\widetilde{s}^{i_1+1}$ since they are the outputs of the

same inputs ($\widetilde{s}^{i_1+m_1} = \widetilde{s}^{i_1}$). From the mathematical induction, the state $\widetilde{s}^{i_1+n}$ will be equal to the state $\widetilde{s}^{i_1+m_1+n}$ for any positive integer n. Therefore, if we set $n = k'm_1$ ($k' \in \mathbb{N}$), then $\widetilde{s}^{i_1+k'm_1} = \widetilde{s}^{i_1+m_1+k'm_1}$. Therefore, the state for packet lengths $\widetilde{s}$ is the same for those indices $(i_1 + km_1)$ for $k \in \{0\} \bigcup \mathbb{N}$.

If $m_1 \neq 1$, the state $\widetilde{s}^{(i_1+1)}$ is different from $\widetilde{s}^{i_1}$ since there are no repeating states in

$\{\widetilde{s}^{(i_1+1)}, \widetilde{s}^{i_1+2}, \cdots, \widetilde{s}^{i_1+m_1-1}\}$ for $\widetilde{s}^{i_1}$. Then the system state $\widetilde{s}^{(i_1+1)+m_1}$ will be equal to the system state $\widetilde{s}^{(i_1+1)}$ since they are the outputs of the same inputs ($\widetilde{s}^{i_1+m_1} = \widetilde{s}^{i_1}$) by state transition function δ. Again, by mathematical induction, the state $\widetilde{s}^{(i_1+1)+n}$ will be equal to the state $\widetilde{s}^{(i_1+1)+m_1+n}$ for any positive integer n. Therefore, the state for packet lengths $\widetilde{s}$ is the same for those indices $((i_1 + 1) + km_1)$ for $k \in \{0\} \bigcup \mathbb{N}$.

The states for $\widetilde{s}^{(i_1+j-1)}$ are different from each other for $1 \leq j \leq m_1$ since m_1 is the period of the group g_1 and, for each j, we know that the states for indices $((i_1+j-1)+km_1)$ for $k \in \{0\} \bigcup \mathbb{N}$ are same from mathematical induction argument shown in the previous paragraph. Therefore, for each $j \in \{1, 2, \cdots, m_1\}$, we group the transmission indices $((i_1 + j - 1) + km_1)$ as group g_j. It is easy to see that $g_{k_1} \bigcap g_{k_2} = \{((i_1 + k_1 - 1) + km_1)\} \bigcap \{((i_1 + k_2 - 1) + km_1)\} = \phi$ if $k_1 \neq k_1$, hence, there are m_1 groups since $k_1 \in \{1, 2, \cdots, m_1\}$. And we set $m_1 = m$.

The elements in group H will be $\{1, 2, \cdots, (i_1 - 1)\}$ from above construction since the indices $\{1, 2, \cdots, (i_1 - 1)\}$ do not belong to any groups in g_k for $1 \leq k \leq m$. We can find that the set of $TI = \{1, 2, 3, \cdots\} = H \bigoplus \bigoplus_{j=1}^{m} g_j$.

If $m_1 = 1$, the state with transmission indices $i_1 + km_1 = i_1 + k$ will be same. In this case, the set of TI will become $H \bigoplus g_1$ and $g_1 = \{i_1, i_1 + 1, i_1 + 2, \cdots\}$since there is only one group.

□

If we consider a general ARQ system with finite number of different packet lengths, then the general formula of total generating function shown in (4.4) can be expressed in a more structured way due to the previous theorem. Because the generating function for those unsuccessful transmissions with transmission indices in group g_j can be evaluated systematically with closed form. We define a packet length index assignment function, T, which assigns the p-th transmission to $q = T(p)$ where q is the index for packet length used in the p-th transmission. The function T is determined by the state transition function δ. Then $T(i_j)$ will indicate the packet length used for transmissions in group g_j. The symbol $\widetilde{s}_{R,g_j}(l_j)$ (l_j is the channel realization index) is used to represent a state of the receiver memory in group g_j.

We use symbol $E_{f,G_{T(i_j)}|\widetilde{s}_{R,g_j}(l_j)}$ to represent the group of events that the transmitter receives a NACK after the $(i_j + km)$-th transmission in the good channel state given that the receiver memory state is $\widetilde{s}_{R,g_j}(l_j)$. The generating function of this group is denoted as $G_{f,G_{T(i_j)}|\widetilde{s}_{R,g_j}(l_j)}$, Similarly, $E_{f,B_{T(i_j)}|\widetilde{s}_{R,g_j}(l_j)}$ is used to represent the group of the events that the transmitter receives a NACK after the $(i_j + km)$-th transmission in the bad channel state given that the receiver memory state is $\widetilde{s}_{R,g_j}(l_j)$. Let $P^{i_j+qm}_{f,G_{T(i_j)}|\widetilde{s}_{R,g_j}(l_j)}$ ($P^{i_j+qm}_{f,B_{T(i_j)}|\widetilde{s}_{R,g_j}(l_j)}$) be the probability of the event that the transmitter can not receive an ACK packet after the $(i_j + qm)$-th transmission with good (bad) channel condition given that the receiver memory state is $\widetilde{s}_{R,g_j}(l_j)$. Then, we have

$$G_{f,G_{T(i_j)}|\widetilde{s}_{R,g_j}(l_j)} = \sum_{q=0}^{\infty} P^{i_j+qm}_{f,G_{T(i_j)}|\widetilde{s}_{R,g_j}(l_j)} X^{[(\sum_{k=1}^{i_j} N_{T(k)})+q(\sum_{p=i_j+1}^{i_j+m} N_{T(p)})]T_b} \times Y^{[(\sum_{k=1}^{i_j} N_{T(k)})+q(\sum_{p=i_j+1}^{i_j+m} N_{T(p)})]E_c}$$

$$G_{f,B_{T(i_j)}|\widetilde{s}_{R,g_j}(l_j)} = \sum_{q=0}^{\infty} P^{i_j+qm}_{f,B_{T(i_j)}|\widetilde{s}_{R,g_j}(l_j)} X^{[(\sum_{k=1}^{i_j} N_{T(k)})+q(\sum_{p=i_j+1}^{i_j+m} N_{T(p)})]T_b} \times Y^{[(\sum_{k=1}^{i_j} N_{T(k)})+q(\sum_{p=i_j+1}^{i_j+m} N_{T(p)})]E_c}. \quad (4.5)$$

We should note that it is possible that the packet length used in the k_1-th transmission may equal to the packet used in the k_2-th transmission for $k_1 \neq k_2$.

We have the following recursion formulas for probabilities $P^{i_j+qm}_{f,G_{T(i_j)}|\widetilde{s}_{R,g_j}(l_j)}$ and $P^{i_j+qm}_{f,B_{T(i_j)}|\widetilde{s}_{R,g_j}(l_j)}$ from the state transition diagram specified by any protocol,

$$\begin{bmatrix} P^{i_j+(q+1)m}_{f,G_{T(i_j)}|\widetilde{s}_{R,g_j}(1)} \\ P^{i_j+(q+1)m}_{f,G_{T(i_j)}|\widetilde{s}_{R,g_j}(2)} \\ \vdots \\ P^{i_j+(q+1)m}_{f,G_{T(i_j)}|\widetilde{s}_{R,g_j}(2^{m_{R,g_j}})} \\ P^{i_j+(q+1)m}_{f,B_{T(i_j)}|\widetilde{s}_{R,g_j}(1)} \\ P^{i_j+(q+1)m}_{f,B_{T(i_j)}|\widetilde{s}_{R,g_j}(2)} \\ \vdots \\ P^{i_j+(q+1)m}_{f,B_{T(i_j)}|\widetilde{s}_{R,g_j}(2^{m_{R,g_j}})} \end{bmatrix} = \begin{bmatrix} \left[c^{g_j}_{G(k)G(l)}\right] & \left[c^{g_j}_{B(k)G(l)}\right] \\ \left[c^{g_j}_{G(k)B(l)}\right] & \left[c^{g_j}_{B(k)B(l)}\right] \end{bmatrix} \begin{bmatrix} P^{i_j+qm}_{f,G_{T(i_j)}|\widetilde{s}_{R,g_j}(1)} \\ P^{i_j+qm}_{f,G_{T(i_j)}|\widetilde{s}_{R,g_j}(2)} \\ \vdots \\ P^{i_j+qm}_{f,G_{T(i_j)}|\widetilde{s}_{R,g_j}(2^{m_{R,g_j}})} \\ P^{i_j+qm}_{f,B_{T(i_j)}|\widetilde{s}_{R,g_j}(1)} \\ P^{i_j+qm}_{f,B_{T(i_j)}|\widetilde{s}_{R,g_j}(2)} \\ \vdots \\ P^{i_j+qm}_{f,B_{T(i_j)}|\widetilde{s}_{R,g_j}(2^{m_{R,g_j}})} \end{bmatrix}, \quad (4.6)$$

where the symbol m_{R,g_j} represents the number of packets stored in the receiver memory for transmissions in group g_j. The components of matrix $\left[c^{g_j}_{G(k)G(l)}\right]$ are the transition probabilities for the system which suffers an unsuccessful transmission at the $(i_j + qm)$-th transmission with good channel condition to the system which has an unsuccessful transmission at the $(i_j+(q+1)m)$-th transmission with good channel condition. The component for the position (k, l) of matrix $\left[c^{g_j}_{G(k)G(l)}\right]$ is obtained by considering the m consecutive packet errors and channel transitions from the system

state $[G_{T(i_j)}, \widetilde{s}_{R,g_j}(k)]$ to $[G_{T(i_j)}, \widetilde{s}_{R,g_j}(l)]$. Similar explanations could be applied for components in matrices $\left[c^{g_j}_{G(k)B(l)}\right]$, $\left[c^{g_j}_{B(k)G(l)}\right]$ and $\left[c^{g_j}_{B(k)B(l)}\right]$.

The value of $\left[c^{g_j}_{G(k)G(l)}\right]$ is derived as follows

$$c^{g_j}_{G(k)G(l)} = \overbrace{\sum_{s^{(i_j+m-1)}\in S}\sum_{s^{(i_j+m-2)}\in S}\cdots\sum_{s^{(i_j+1)}\in S}}^{(m-1)\text{ fold}}\Big[(\prod_{k=1}^{m-2} t^{s^{i_j+k+1}}_{s^{i_j+k}})t^{s^{i_j+1}}_{[G_{T(i_j)},\widetilde{s}_{R,g_j}(k)]}t^{[G_{T(i_j)},\widetilde{s}_{R,g_j}(l)]}_{s^{i_j+m-1}}\Big], \tag{4.7}$$

where the summation symbol $\sum_{s^p\in S}$ for $p \in \mathbb{N}$ denotes the sum over all possible system states at the p-th transmission. The symbol $t^{s^{p+1}}_{s^p}$ for $p \in \mathbb{N}$ denotes the state transition probability from a state, $s^p = [s^p_C, s^p_T, s^p_R]$, at the p-th transmission to a state, $s^{p+1} = [s^{p+1}_C, s^{p+1}_T, s^{p+1}_R]$, at the $(p+1)$-th transmission. It is given by

$$t^{s^{(p+1)}}_{s^p} = \begin{cases} e^{-\lambda N_{T(p)}T_b}P_{e,G_{T(p)}|s^p_R}, \text{ if } s^p_C = G, s^{(p+1)}_C = G,\ \delta(s^p, \text{NACK}) = s^{(p+1)} \\ \overline{e^{-\lambda N_{T(p)}T_b}}P_{e,B_{T(p)}|s^p_R}, \text{ if } s^p_C = G, s^{(p+1)}_C = B,\ \delta(s^p, \text{NACK}) = s^{(p+1)} \\ \overline{e^{-\mu N_{T(p)}T_b}}P_{e,G_{T(p)}|s^p_R}, \text{ if } s^p_C = B, s^{(p+1)}_C = G,\ \delta(s^p, \text{NACK}) = s^{(p+1)} \\ e^{-\mu N_{T(p)}T_b}P_{e,B_{T(p)}|s^p_R}, \text{ if } s^p_C = B, s^{(p+1)}_C = B,\ \delta(s^p, \text{NACK}) = s^{(p+1)} \\ 0, \text{ otherwise} \end{cases}, \tag{4.8}$$

where s^p_C and $s^{(p+1)}_C$ are the channel condtions for the p and $(p+1)$-th transmissions. The components in matrix $\left[c^{g_j}_{G(k)B(l)}\right]$, $\left[c^{g_j}_{B(k)G(l)}\right]$ and $\left[c^{g_j}_{B(k)B(l)}\right]$ can be derived similarly as shown in (4.7) by just changing the channel condtions at the i_j-th and (i_j+m)-th transmissions. If we multiply both sides of (4.6) by $X^{[(\sum_{k=1}^{i_j} N_{T(k)})+q(\sum_{p=i_j+1}^{i_j+m} N_{T(p)})]T_b}$

and $Y^{[(\sum_{k=1}^{i_j} N_{T(k)})+q(\sum_{p=i_j+1}^{i_j+m} N_{T(p)})]E_c}$ and sum over q, we obtain

$$\begin{bmatrix} G_{f,G_{T(i_j)}|\widetilde{s}_{R,g_j}(1)} \\ G_{f,G_{T(i_j)}|\widetilde{s}_{R,g_j}(2)} \\ \vdots \\ G_{f,G_{T(i_j)}|\widetilde{s}_{R,g_j}(2^{m_{R,g_j}})} \\ G_{f,B_{T(i_j)}|\widetilde{s}_{R,g_j}(1)} \\ G_{f,B_{T(i_j)}|\widetilde{s}_{R,g_j}(2)} \\ \vdots \\ G_{f,B_{T(i_j)}|\widetilde{s}_{R,g_j}(2^{m_{R,g_j}})} \end{bmatrix} - \begin{bmatrix} A_{f,G_{T(i_j)}|\widetilde{s}_{R,g_j}(1)} \\ A_{f,G_{T(i_j)}|\widetilde{s}_{R,g_j}(2)} \\ \vdots \\ A_{f,G_{T(i_j)}|\widetilde{s}_{R,g_j}(2^{m_{R,g_j}})} \\ A_{f,B_{T(i_j)}|\widetilde{s}_{R,g_j}(1)} \\ A_{f,B_{T(i_j)}|\widetilde{s}_{R,g_j}(2)} \\ \vdots \\ A_{f,B_{T(i_j)}|\widetilde{s}_{R,g_j}(2^{m_{R,g_j}})} \end{bmatrix} =$$

$$\begin{bmatrix} \left[c^{g_j}_{G(k)G(l)}\right] & \left[c^{g_j}_{B(k)G(l)}\right] \\ \left[c^{g_j}_{G(k)B(l)}\right] & \left[c^{g_j}_{B(k)B(l)}\right] \end{bmatrix} \cdot \begin{bmatrix} G_{f,G_{T(i_j)}|\widetilde{s}_{R,g_j}(1)} \\ G_{f,G_{T(i_j)}|\widetilde{s}_{R,g_j}(2)} \\ \vdots \\ G_{f,G_{T(i_j)}|\widetilde{s}_{R,g_j}(2^{m_{R,g_j}})} \\ G_{f,B_{T(i_j)}|\widetilde{s}_{R,g_j}(1)} \\ G_{f,B_{T(i_j)}|\widetilde{s}_{R,g_j}(2)} \\ \vdots \\ G_{f,B_{T(i_j)}|\widetilde{s}_{R,g_j}(2^{m_{R,g_j}})} \end{bmatrix}, \tag{4.9}$$

where

$$\begin{aligned} A_{f,G_{T(i_j)}|\widetilde{s}_{R,g_j}(l_j)} &= G_{f,G_{T(i_j)}|s_R^{T(i_j)}}, \\ A_{f,B_{T(i_j)}|\widetilde{s}_{R,g_j}(l_j)} &= G_{f,B_{T(i_j)}|s_R^{T(i_j)}}. \end{aligned} \tag{4.10}$$

The quanity in (4.10) can be evaluted from (4.3) by setting i in (4.3) as $T(i_j)$ and s_R^i in (4.3) as $\widetilde{s}_{R,g_j}(l_j)$.

From (4.9), we can solve the generating functions of $G_{f,G_{T(i_j)}|\widetilde{s}_{R,g_j}(l_j)}$ and $G_{f,B_{T(i_j)}|\widetilde{s}_{R,g_j}(l_j)}$ for $1 \leq l_j \leq 2^{m_{R,g_j}}$ in closed form. The symbol $s_{R,g_j,G}$ represents the state of the

memory for the system state generated by $\delta([G_{T(i_j)}, \widetilde{s}_{R,g_j}(l_j)], \text{NACK})$. Similarly, the symbol $s_{R,g_j,B}$ represents the state of the memory for the system state generated by $\delta([B_{T(i_j)}, \widetilde{s}_{R,g_j}(l_j)], \text{NACK})$. Therefore, the total generating function is composed of two parts. The first part is the generating function for those successful transmissions which do not belong to any indices of the form $(i_j + km + 1)$. We use $G_i(X, Y)$ to represent the generating function for the i-th successful transmission. This part of the generating function can be evaluated from (4.3). The second part is the generating function for those successful transmissions with transmission indices form $(i_j + km + 1)$. Hence, we can express the total generating function, G_s, for the system as

$$\begin{aligned} G_s(X,Y) &= \sum_{i=1}^{i_1} G_i(X,Y) + \sum_{j=1}^{m} \Big\{ \sum_{l_j=1}^{2^{m_{R,g_j}}} \Big[G_{f,G_{T(i_j)}|\widetilde{s}_{R,g_j}(l_j)} e^{-\lambda N_{T(i_j)} T_b} \overline{P_{e,G_{T(i_j+1)}|s_{R,g_j,G}}} + \\ & G_{f,B_{T(i_j)}|\widetilde{s}_{R,g_j}(l_j)} \overline{e^{-\mu N_{T(i_j)} T_b}\, \overline{P_{e,G_{T(i_j+1)}|s_{R,g_j,B}}}} + \\ & G_{f,G_{T(i_j)}|\widetilde{s}_{R,g_j}(l_j)} \overline{e^{-\lambda N_{T(i_j)} T_b}\, \overline{P_{e,B_{T(i_j+1)}|s_{R,g_j,G}}}} + \\ & G_{f,B_{T(i_j)}|\widetilde{s}_{R,g_j}(l_j)} e^{-\mu N_{T(i_j)} T_b} \overline{P_{e,B_{T(i_j+1)}|s_{R,g_j,B}}} \Big] X^{N_{T(i_j+1)} T_b} Y^{N_{T(i_j+1)} E_c} \Big\} \end{aligned} \quad (4.11)$$

where

$$\begin{aligned} G_i(X,Y) &= \sum_{s_R^{i-1} \in S_R} \sum_{s_R^{i} \in S_R} \Big[\\ & G_{f,G_{T(i-1)}|s_R^{i-1}} e^{-\lambda N_{T(i-1)} T_b} \overline{P_{e,G_{T(i)}|s_R^i}} I(\delta([G_{T(i-1)}, s_R^{i-1}], \text{NACK}) = [G_{T(i)}, s_R^i]) + \\ & G_{f,B_{T(i-1)}|s_R^{i-1}} \overline{e^{-\mu N_{T(i-1)} T_b}\, \overline{P_{e,G_{T(i)}|s_R^i}}} I(\delta([B_{T(i-1)}, s_R^{i-1}], \text{NACK}) = [G_{T(i)}, s_R^i]) + \\ & G_{f,G_{T(i-1)}|s_R^{i-1}} \overline{e^{-\lambda N_{T(i-1)} T_b}\, \overline{P_{e,B_{T(i)}|s_R^i}}} I(\delta([G_{T(i-1)}, s_R^{i-1}], \text{NACK}) = [B_{T(i)}, s_R^i]) + \\ & G_{f,B_{T(i-1)}|s_R^{i-1}} e^{-\mu N_{T(i-1)} T_b} \overline{P_{e,B_{T(i)}|s_R^i}} I(\delta([B_{T(i-1)}, s_R^{i-1}], \text{NACK}) = [B_{T(i)}, s_R^i]) \Big] \cdot \\ & X^{N_{T(i)} T_b} Y^{N_{T(i)} E_c}. \end{aligned} \quad (4.12)$$

4.4 Some Practical ARQ Systems

In this section, we will specialize the general ARQ systems to two special categories. In the first category, we will consider ARQ systems which have the sequence of packets with the following form : $\{N_1, N_2, \cdots, N_m, N_m, \cdots\}$. We call this version of ARQ as nonrepeating ARQ (NR-ARQ). This category will be discussed in Subsection 4.4.1. In the second category, we will consider ARQ systems which have the sequence of packets with the following form :
$\{N_1, N_2, \cdots, N_m, N_1, N_2, \cdots, N_m, N_1, N_2, \cdots, N_m, \cdots\}$. We call this repeating ARQ (R-ARQ). This category will be discussed in Subsection 4.4.2.

4.4.1 SW-ARQ over Time-Varying Channels: No Repetition

In this section, we will first consider a NR-ARQ system which the receiver does not contain any memory. Then, we will discuss a NR-ARQ system with memory in the receiver.

Memoryless Receiver

The protocol of SW-ARQ with incremental redundancy is described as follows. Whenever there is an information packet with length K generated, it will be encoded to a data packet with length N_1. If the data packet is decoded correctly at the receiver, the receiver will send an ACK packet back to the transmitter. If it is decoded incorrectly at the receiver, a NACK packet will be sent back to the transmitter. Then the transmitter will encode the same information packet again to a data packet with length N_2 and retransmit it. If the data packet is decoded correctly at the receiver, the receiver will send an ACK packet back to the transmitter. If it is decoded incorrectly at the receiver, a NACK packet will be sent back to the transmitter. Let m be the number of different packet lengths adopted by the transmitter. The coded

packet lengths have the following relation : $N_1 < N_2 < \cdots < N_m$.

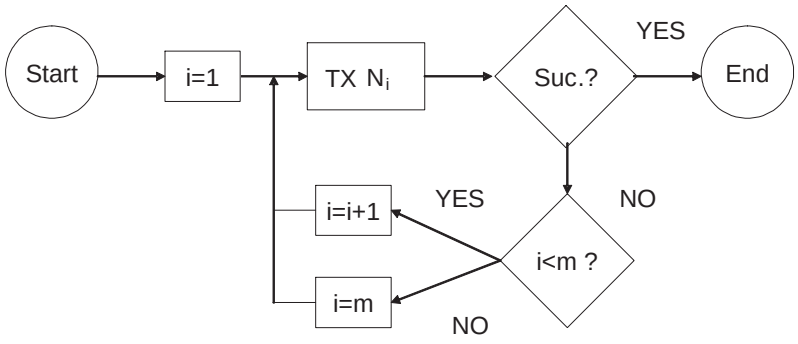

Figure 4.3: Stop-and-Wait ARQ protocol with increment redundancy.

We can use a FSM to model this system. The set of Σ is composed of $\{\text{ACK}, \text{NACK}\}$. The set Γ, which indicates the output functions of energy and delay for packets transmission, is composed of $\Gamma = \{X^{N_1 T_b} Y^{N_1 E_c}, X^{N_2 T_b} Y^{N_2 E_c}, \cdots, X^{N_m T_b} Y^{N_m T_b}\}$ where T_b is the transmission time of a coded bit and the E_c is the transmission energy of a coded bit. The transmitter/channel state space, $[S_C, S_T]$, consists of $\{G_1, B_1, G_2, B_2, \cdots, G_m, B_m\}$. The symbol $G_i(B_i)$ for $(1 \leq i \leq m)$ represents a packet of packet length index i ready to be transmitted and the channel condition is good (bad) in transmitting this packet. We do not assign states to the receiver because a receiver without memory only decodes the latest packet sent from the transmitter for the whole ARQ procedure. The initial state is G_1 or B_1 depending on whether the channel condition is in a good or bad state. The channel state transition probability from a good state to a bad state after transmitting a packet with length N_i is $1 - e^{-\lambda N_i T_b}$ due to the exponential distribution of the channel state duration. Similarly, $e^{-\lambda N_i T_b}$, $e^{-\mu N_i T_b}$ and $1 - e^{-\mu N_i T_b}$ are state transition probabilities from G-G, B-B, and B-G, respectively. According to our protocol shown in Fig. 4.3, we can summarize the FSM states in Fig. 4.4. The state transition probabilities and output functions are given in Table 4.1. Note that $\overline{P} = 1 - P$ for a probability

Table 4.1: Output functions and state transition probabilities for SWARQ

previous state	next state	transition prob.	output function
G_i $1 \leq i \leq m-1$	G_{i+1}	$e^{-\lambda N_i T_b} P_{e,G_i}$	$X^{N_i T_b} Y^{N_i E_c}$
	B_{i+1}	$\overline{e^{-\lambda N_i T_b}} P_{e,G_i}$	$X^{N_i T_b} Y^{N_i E_c}$
	G_1	$e^{-\lambda N_i T_b} \overline{P_{e,G_i}}$	$X^{N_i T_b} Y^{N_i E_c}$
	B_1	$\overline{e^{-\lambda N_i T_b}}\,\overline{P_{e,G_i}}$	$X^{N_i T_b} Y^{N_i E_c}$
G_m	G_m	$e^{-\lambda N_m T_b} P_{e,G_m}$	$X^{N_m T_b} Y^{N_m E_c}$
	B_m	$\overline{e^{-\lambda N_m T_b}} P_{e,G_m}$	$X^{N_m T_b} Y^{N_m E_c}$
	G_1	$e^{-\lambda N_m T_b} \overline{P_{e,G_m}}$	$X^{N_m T_b} Y^{N_m E_c}$
	B_1	$\overline{e^{-\lambda N_m T_b}}\,\overline{P_{e,G_m}}$	$X^{N_m T_b} Y^{N_m E_c}$
B_i $1 \leq i \leq m-1$	G_{i+1}	$\overline{e^{-\mu N_i T_b}} P_{e,B_i}$	$X^{N_i T_b} Y^{N_i E_c}$
	B_{i+1}	$e^{-\mu N_i T_b} P_{e,B_i}$	$X^{N_i T_b} Y^{N_i E_c}$
	G_1	$\overline{e^{-\mu N_i T_b}}\,\overline{P_{e,B_i}}$	$X^{N_i T_b} Y^{N_i E_c}$
	B_1	$e^{-\mu N_i T_b} \overline{P_{e,B_i}}$	$X^{N_i T_b} Y^{N_i E_c}$
B_m	G_m	$\overline{e^{-\mu N_m T_b}} P_{e,B_m}$	$X^{N_m T_b} Y^{N_m E_c}$
	B_m	$e^{-\mu N_m T_b} P_{e,B_m}$	$X^{N_m T_b} Y^{N_m E_c}$
	G_1	$\overline{e^{-\mu N_m T_b}}\,\overline{P_{e,B_m}}$	$X^{N_m T_b} Y^{N_m E_c}$
	B_1	$e^{-\mu N_m T_b} \overline{P_{e,B_m}}$	$X^{N_m T_b} Y^{N_m E_c}$

quantity P.

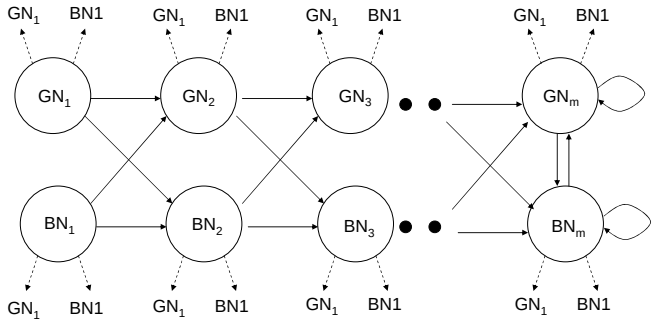

Figure 4.4: FSM representation for SW-ARQ protocol with increment redundancy.

Memory Receiver

Memory ARQ (MARQ) is a method of improving the performance of ARQ over a non-stationary channel by storing and using received erroneous packets to aid decoding which are discarded by a memoryless ARQ protocol. We begin with the description of our SW-MARQ protocol.

The MARQ protocol is the same as the SW-ARQ protocol described in subsec-

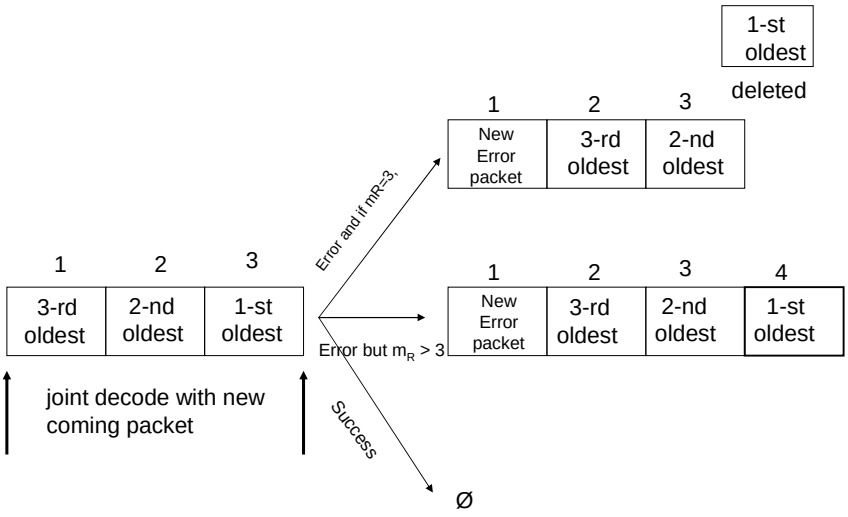

Figure 4.5: Operation of sliding window.

tion 4.4.1 except that the receiver will store upto m_R previously erroneous packets. If the receiver receives an erroneous packet after the memory is full, the receiver will update the receiver memory content by storing the newest erroneous packet at the position 1 of the receiver memory and discard the oldest erroneous packet originally stored at the position m_R. The receiver will clear all stored packets in its memory when it receives a correct packet. In Fig. 4.5, the operation of the receiver memory is illustrated.

The FSM model for a SW-MARQ system is the same as a SW-ARQ system except the receiver state. We need to add to the receiver state space S_R to characterize the memory content. A state to describe SW-MARQ system consists of three parts. For example, a state $[G_3, [G_2, B_1]]$ represents that the transmitter transmits with packet length N_3 with good (G) channel condition and the memory content of the receiver is two erroneous packets: the first position of the receiver memory stores a packet with length N_2 which received when the channel condition is good (G) and the second

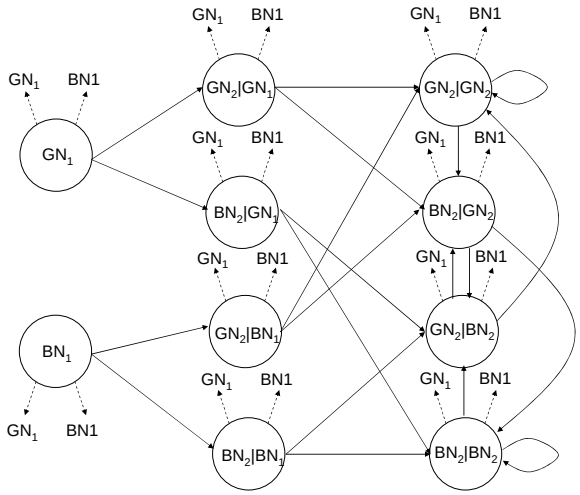

Figure 4.6: FSM representation for Stop-and-Wait ARQ protocol with $m = 2$ and $m_R = 1$.

position of the receiver memory stores a packet with length N_1 which is received when the channel condition is bad (B). The packet error probability corresponding to this state is $P_{e,G_3|G_2B_1}$. In Fig. 4.6, the FSM diagram for SW-MARQ with $m = 2$ and $m_R = 1$ is shown.

For general values of m and m_R, we can summarize the FSM states, state transition probabilities, and output functions (denoted by energy and delay generating function) in table 4.2. In the following tables of this section, we only write those components of the system state vector, e.g., G_1 will be equal to a system state $[G_1, \overbrace{\phi, \cdots, \phi}^{m_R \text{terms}}]$. The symbol R_i represents the memory content of some position in the receiver memory, the subscript i is used to distinguish the different position of the receiver memory. In the columns of the "previous state" and "probability", R_i represents the memory content of the i-th position in the receiver memory. However, in the column of the "next state", R_i represents the memory content of the $(i+1)$-th position in the receiver memory. The symbol R_i could be G_j or B_j where j is the index of the packet length.

Table 4.2: Output functions and state transition probabilities for SWMARQ

previous state; output	next state	probability
G_1 $X^{N_1T_b}Y^{N_1E_c}$	G_1	$e^{-\lambda N_1 T_b}\overline{P_{e,G_1}}$
	B_1	$\overline{e^{-\lambda N_1 T_b}}\overline{P_{e,G_1}}$
	$[G_2, G_1]$	$e^{-\lambda N_1 T_b}P_{e,G_1}$
	$[B_2, G_1]$	$\overline{e^{-\lambda N_1 T_b}}P_{e,G_1}$
B_1 $X^{N_1T_b}Y^{N_1E_c}$	G_1	$\overline{e^{-\mu N_1 T_b}}\overline{P_{e,B_1}}$
	B_1	$e^{-\mu N_1 T_b}\overline{P_{e,B_1}}$
	$[G_2, B_1]$	$\overline{e^{-\mu N_1 T_b}}P_{e,B_1}$
	$[B_2, B_1]$	$e^{-\mu N_1 T_b}P_{e,B_1}$
$[G_i, R_1, R_2, \cdots, R_k]$ $i < m, k < m_R$; $X^{N_iT_b}Y^{N_iE_c}$	G_1	$\overline{e^{-\lambda N_i T_b}}\overline{P_{e,G_i\|R_1R_2\cdots R_k}}$
	B_1	$\overline{e^{-\lambda N_i T_b}}\overline{P_{e,G_i\|R_1R_2\cdots R_k}}$
	$[G_{i+1}, G_i, R_1, \cdots, R_k]$	$e^{-\lambda N_i T_b}P_{e,G_i\|R_1R_2\cdots R_k}$
	$[B_{i+1}, G_i, R_1, \cdots, R_k]$	$\overline{e^{-\lambda N_i T_b}}P_{e,G_i\|R_1R_2\cdots R_k}$
$[B_i, R_1, R_2, \cdots, R_k]$ $i < m, k < m_R$; $X^{N_iT_b}Y^{N_iE_c}$	G_1	$\overline{e^{-\mu N_i T_b}}\overline{P_{e,B_i\|R_1R_2\cdots R_k}}$
	B_1	$e^{-\mu N_i T_b}\overline{P_{e,B_i\|R_1R_2\cdots R_k}}$
	$[G_{i+1}, B_i, R_1, \cdots, R_k]$	$\overline{e^{-\mu N_i T_b}}P_{e,B_i\|R_1R_2\cdots R_k}$
	$[B_{i+1}, B_i, R_1, \cdots, R_k]$	$e^{-\mu N_i T_b}P_{e,B_i\|R_1R_2\cdots R_k}$
$[G_m, R_1, R_2, \cdots, R_k]$ $k < m_R$; $X^{N_mT_b}Y^{N_mE_c}$	G_1	$\overline{e^{-\lambda N_m T_b}}\overline{P_{e,G_m\|R_1R_2\cdots R_k}}$
	B_1	$\overline{e^{-\lambda N_m T_b}}\overline{P_{e,G_m\|R_1R_2\cdots R_k}}$
	$[G_m, G_m, R_1, \cdots, R_k]$	$e^{-\lambda N_m T_b}P_{e,G_m i\|R_1R_2\cdots R_k}$
	$[B_m, G_m, R_1, \cdots, R_k]$	$\overline{e^{-\lambda N_m T_b}}P_{e,G_m\|R_1R_2\cdots R_k}$
$[B_m, R_1, R_2, \cdots, R_k]$ $k < m_R$; $X^{N_mT_b}Y^{N_mE_c}$	G_1	$\overline{e^{-\mu N_m T_b}}\overline{P_{e,B_m\|R_1R_2\cdots R_k}}$
	B_1	$e^{-\mu N_m T_b}\overline{P_{e,B_m\|R_1R_2\cdots R_k}}$
	$[G_m, B_m, R_1, \cdots, R_k]$	$\overline{e^{-\mu N_m T_b}}P_{e,B_m\|R_1R_2\cdots R_k}$
	$[B_m, B_m, R_1, \cdots, R_k]$	$e^{-\mu N_m T_b}P_{e,B_m\|R_1R_2\cdots R_k}$
$[G_i, R_1, R_2, \cdots, R_{m_R}]$ $i < m$; $X^{N_iT_b}Y^{N_iE_c}$	G_1	$\overline{e^{-\lambda N_i T_b}}\overline{P_{e,G_i\|R_1R_2\cdots R_{m_R}}}$
	B_1	$\overline{e^{-\lambda N_i T_b}}\overline{P_{e,G_i\|R_1R_2\cdots R_{m_R}}}$
	$[G_{i+1}, G_i, R_1, \cdots, R_{m_R-1}]$	$e^{-\lambda N_i T_b}P_{e,G_i\|R_1R_2\cdots R_{m_R}}$
	$[B_{i+1}, G_i, R_1, \cdots, R_{m_R-1}]$	$\overline{e^{-\lambda N_i T_b}}P_{e,G_i\|R_1R_2\cdots R_{m_R}}$
$[B_i, R_1, R_2, \cdots, R_{m_R}]$ $i < m$; $X^{N_iT_b}Y^{N_iE_c}$	G_1	$\overline{e^{-\mu N_i T_b}}\overline{P_{e,B_i\|R_1R_2\cdots R_{m_R}}}$
	B_1	$e^{-\mu N_i T_b}\overline{P_{e,B_i\|R_1R_2\cdots R_{m_R}}}$
	$[G_{i+1}, B_i, R_1, \cdots, R_{m_R-1}]$	$\overline{e^{-\mu N_i T_b}}P_{e,B_i\|R_1R_2\cdots R_{m_R}}$
	$[B_{i+1}, B_i, R_1, \cdots, R_{m_R-1}]$	$e^{-\mu N_i T_b}P_{e,B_i\|R_1R_2\cdots R_{m_R}}$
$[G_m, R_1, R_2, \cdots, R_{m_R}]$ $X^{N_mT_b}Y^{N_mE_c}$	G_1	$\overline{e^{-\lambda N_m T_b}}\overline{P_{e,G_m\|R_1R_2\cdots R_{m_R}}}$
	B_1	$\overline{e^{-\lambda N_m T_b}}\overline{P_{e,G_m\|R_1R_2\cdots R_{m_R}}}$
	$[G_m, G_m, R_1, \cdots, R_{m_R-1}]$	$e^{-\lambda N_m T_b}P_{e,G_m\|R_1R_2\cdots R_{m_R}}$
	$[B_m, G_m, R_1, \cdots, R_{m_R-1}]$	$\overline{e^{-\lambda N_m T_b}}P_{e,G_m\|R_1R_2\cdots R_{m_R}}$
$[B_m, R_1, R_2, \cdots, R_{m_R}]$ $X^{N_mT_b}Y^{N_mE_c}$	G_1	$\overline{e^{-\mu N_m T_b}}\overline{P_{e,B_m\|R_1R_2\cdots R_{m_R}}}$
	B_1	$e^{-\mu N_m T_b}\overline{P_{e,B_m\|R_1R_2\cdots R_{m_R}}}$
	$[G_m, B_m, R_1, \cdots, R_{m_R-1}]$	$\overline{e^{-\mu N_m T_b}}P_{e,B_m\|R_1R_2\cdots R_{m_R}}$
	$[B_m, B_m, R_1, \cdots, R_{m_R-1}]$	$e^{-\mu N_m T_b}P_{e,B_m\|R_1R_2\cdots R_{m_R}}$

4.4.2 SW-ARQ over Time-Varying Channels: Repetition

In this section, we will consider an ARQ systems with repetition packet length. There are three kinds of repetition ARQ systems analyzed in this section; the first ARQ system does not contain any memory in the receiver, the second ARQ system equips the receiver with memory and uses a sliding window scheme as mentioned in Section 4.4.1, the third ARQ system also equips the receiver with memory but the receiver memory contents will be updated to be empty after the receiver cannot decode correctly when the memory buffer is full.

Memoryless Receiver

In this subsection, we will consider ARQ systems with a memoryless receiver. The protocol, named repetition SW-ARQ with memoryless receiver (R-SW-ARQ-ML), operates as follows. At the transmitter, whenever there is an information packet with length K generated, it will be encoded to a data packet with length N_1. If the data packet is decoded correctly at the receiver, the receiver will send an ACK packet back to the transmitter. If a detectable but uncorrectable error occurs at the receiver, a NACK packet will be sent back to the transmitter. Then the transmitter will encode the same information packet to a data packet with length N_2 and transmit the packet. If the data packet is decoded correctly at the receiver, the receiver will send an ACK packet back to the transmitter. If a detectable but uncorrectable error occurs at the receiver, a NACK packet will be sent back to the transmitter. Let m be the number of different packet lengths used by the transmitter. Therefore, if the transmitter can not receive an ACK by sending a data packet with length N_m at m-th transmission, the transmitter will encode the information packet to a data packet with length N_1 again. If it is decoded in error at the receiver, the

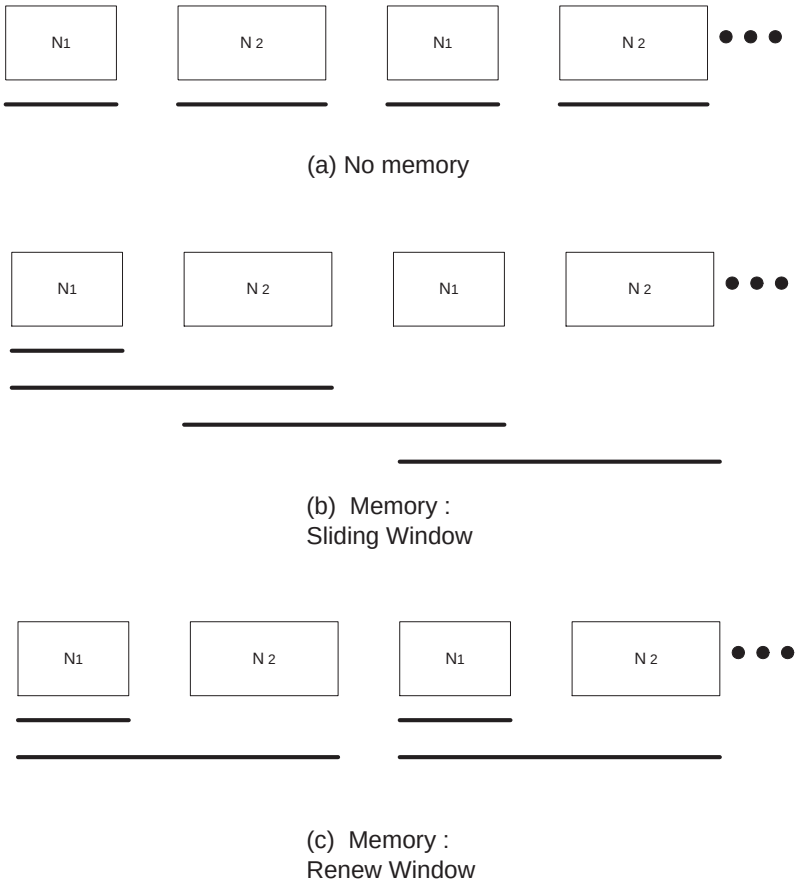

Figure 4.7: Three different repetition ARQ protocols with $m = 2$ considered in this section.

transmitter will encode the same information packet to a data packet with length N_2 again. If the transmitter can not receive an ACK by sending a data packet with length N_m at $2m$-th transmission, it will encode the information packet to a data packet with length N_1 at $2m + 1$-th transmission. The repetition procedure of these m different packet lengths will be terminated if an ACK packet is received. The operation of this protocol is shown in part (a) of Fig. 4.7. Each piece of the bold line in Fig. 4.7 indicates the packets in the receiver memory used to decode jointly when the transmitter has transmitted the last packet underscored by each piece of the bold line.

By using the FSM model described in Section 4.2, we can summarize the state

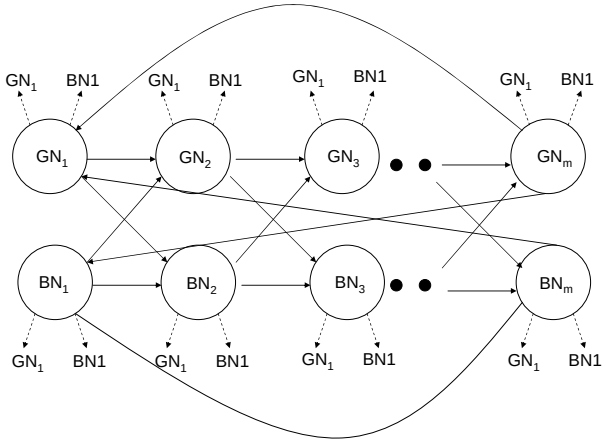

Figure 4.8: State diagram for R-SW-ARQ-ML.

transition probabilities and output functions (denoted by energy and delay generating function) in Table 4.3.

Memory Receiver: Sliding Window

The protocol discussed in this subsection is the same as the repetition SW-ARQ protocol described in 4.4.2 except that the receiver will store up to m_R previously erroneous packets. We set $m_R = m - 1$ be the maximum number of packets that the receiver can store in its memory. If the receiver receives an erroneous packet after the memory is full, the receiver will update its memory content by storing the newest erroneous packet and discarding the oldest erroneous packet. The receiver will clear all stored packets in its memory when it receives a correct packet.

The FSM model for repetition SW-ARQ systems with sliding window memory (R-SW-ARQ-SWM) is extended to a repetition SW-ARQ system by including the receiver state. We need to add the receiver state S_R to characterize the memory content. A state space to describe R-SW-ARQ-SWM system consists of three parts. The first part of the state represents the channel condition. The second part is used to represent the packet lengths for transmission. The third part denotes the

Table 4.3: Table for transition probabilities and output functions of R-SW-ARQ-ML

previous state	next state	transition prob.	output function
G_i $1 \le i \le m-1$	G_{i+1}	$e^{-\lambda N_i T_b} P_{e,G_i}$	$X^{N_i T_b} Y^{N_i E_c}$
	B_{i+1}	$\overline{e^{-\lambda N_i T_b}} P_{e,G_i}$	$X^{N_i T_b} Y^{N_i E_c}$
	G_1	$e^{-\lambda N_i T_b} \overline{P_{e,G_i}}$	$X^{N_i T_b} Y^{N_i E_c}$
	B_1	$\overline{e^{-\lambda N_i T_b}}\,\overline{P_{e,G_i}}$	$X^{N_i T_b} Y^{N_i E_c}$
G_m	G_1	$e^{-\lambda N_m T_b} P_{e,G_m}$	$X^{N_m T_b} Y^{N_m E_c}$
	B_1	$\overline{e^{-\lambda N_m T_b}} P_{e,G_m}$	$X^{N_m T_b} Y^{N_m E_c}$
	G_1	$e^{-\lambda N_m T_b} \overline{P_{e,G_m}}$	$X^{N_m T_b} Y^{N_m E_c}$
	B_1	$\overline{e^{-\lambda N_m T_b}}\,\overline{P_{e,G_m}}$	$X^{N_m T_b} Y^{N_m E_c}$
B_i $1 \le i \le m-1$	G_{i+1}	$\overline{e^{-\mu N_i T_b}} P_{e,B_i}$	$X^{N_i T_b} Y^{N_i E_c}$
	B_{i+1}	$e^{-\mu N_i T_b} P_{e,B_i}$	$X^{N_i T_b} Y^{N_i E_c}$
	G_1	$\overline{e^{-\mu N_i T_b}}\,\overline{P_{e,B_i}}$	$X^{N_i T_b} Y^{N_i E_c}$
	B_1	$e^{-\mu N_i T_b} \overline{P_{e,B_i}}$	$X^{N_i T_b} Y^{N_i E_c}$
B_m	G_1	$\overline{e^{-\mu N_m T_b}} P_{e,B_m}$	$X^{N_m T_b} Y^{N_m E_c}$
	B_1	$e^{-\mu N_m T_b} P_{e,B_m}$	$X^{N_m T_b} Y^{N_m E_c}$
	G_1	$\overline{e^{-\mu N_m T_b}}\,\overline{P_{e,B_m}}$	$X^{N_m T_b} Y^{N_m E_c}$
	B_1	$e^{-\mu N_m T_b} \overline{P_{e,B_m}}$	$X^{N_m T_b} Y^{N_m E_c}$

contents of the receiver memory. For example, state $[G_3, G_2, B_1]$ represents that the transmitter transmits with packet length N_3 with good (G) channel condition and the memory contents of the receiver is two erroneous packets: the first packet with length N_1 is received when the channel condition is bad (B) and the second packet with length N_2 is received when the channel condition is good (G). The packet error probability corresponding to this state is $P_{e,G_3|G_2B_1}$. In Fig. 4.9, the FSM diagram for SW-MARQ with $m = 2$ and $m_R = 1$ is shown.

For general values of m and m_R, we summarize the FSM states, state transition probabilities, and output functions (denoted by energy and delay generating function) in Table 4.4.

Memory Receiver: Non Overlap

In this section, we consider the case that $m_R = m - 1$ is the maximum number of packets that the receiver can store in its memory. The protocol is the same as the repetition SW-ARQ protocol with the receiver memory described in 4.4.2 except

Table 4.4: Table for transition probabilities and output functions of R-SW-ARQ-SWM

previous state; output	next state	probability
G_1	G_1	$e^{-\lambda N_1 T_b}\overline{P_{e,G_1}}$
	B_1	$\overline{e^{-\lambda N_1 T_b}}\overline{P_{e,G_1}}$
$X^{N_1 T_b} Y^{N_1 E_c}$	$[G_2, G_1]$	$e^{-\lambda N_1 T_b} P_{e,G_1}$
	$[B_2, G_1]$	$\overline{e^{-\lambda N_1 T_b}} P_{e,G_1}$
B_1	G_1	$\overline{e^{-\mu N_1 T_b}}\overline{P_{e,B_1}}$
	B_1	$e^{-\mu N_1 T_b}\overline{P_{e,B_1}}$
$X^{N_1 T_b} Y^{N_1 E_c}$	$[G_2, B_1]$	$\overline{e^{-\mu N_1 T_b}} P_{e,B_1}$
	$[B_2, B_1]$	$e^{-\mu N_1 T_b} P_{e,B_1}$
$[G_i, R_1, R_2, \cdots, R_{i-1}]$	G_1	$\overline{e^{-\lambda N_i T_b}}\overline{P_{e,G_i\|R_1 R_2 \cdots R_{i-1}}}$
	B_1	$\overline{e^{-\lambda N_i T_b}}\overline{P_{e,G_i\|R_1 R_2 \cdots R_{i-1}}}$
$i < m$	$[G_{i+1}, G_i, R_1, \cdots, R_{i-1}]$	$e^{-\lambda N_i T_b} P_{e,G_i\|R_1 R_2 \cdots R_{i-1}}$
$X^{N_i T_b} Y^{N_i E_c}$	$[B_{i+1}, G_i, R_1, \cdots, R_{i-1}]$	$\overline{e^{-\lambda N_i T_b}} P_{e,G_i\|R_1 R_2 \cdots R_{i-1}}$
$[B_i, R_1, R_2, \cdots, R_{i-1}]$	G_1	$\overline{e^{-\mu N_i T_b}}\overline{P_{e,B_i\|R_1 R_2 \cdots R_{i-1}}}$
	B_1	$e^{-\mu N_i T_b}\overline{P_{e,B_i\|R_1 R_2 \cdots R_{i-1}}}$
$i < m$	$[G_{i+1}, B_i, R_1, \cdots, R_{i-1}]$	$\overline{e^{-\mu N_i T_b}} P_{e,B_i\|R_1 R_2 \cdots R_{i-1}}$
$X^{N_i T_b} Y^{N_i E_c}$	$[B_{i+1}, B_i, R_1, \cdots, R_{i-1}]$	$e^{-\mu N_i T_b} P_{e,B_i\|R_1 R_2 \cdots R_{i-1}}$
$[G_i, R_1, R_2, \cdots, R_{m_R}]$	G_1	$\overline{e^{-\lambda N_i T_b}}\overline{P_{e,G_i\|R_1 R_2 \cdots R_{m_R}}}$
	B_1	$\overline{e^{-\lambda N_i T_b}}\overline{P_{e,G_i\|R_1 R_2 \cdots R_{m_R}}}$
$1 \le i \le m$;	$[G_{\overline{i+1}}, G_i, R_1, \cdots, R_{m_R-1}]$	$e^{-\lambda N_i T_b} P_{e,G_i\|R_1 R_2 \cdots R_{m_R}}$
$X^{N_i T_b} Y^{N_i E_c}$	$[B_{\overline{i+1}}, G_i, R_1, \cdots, R_{m_R-1}]$	$\overline{e^{-\lambda N_i T_b}} P_{e,G_i\|R_1 R_2 \cdots R_{m_R}}$
$[B_i, R_1, R_2 \cdots R_{m_R}]$	G_1	$\overline{e^{-\mu N_i T_b}}\overline{P_{e,B_i\|R_1 R_2 \cdots R_{m_R}}}$
	B_1	$e^{-\mu N_i T_b}\overline{P_{e,B_i\|R_1 R_2 \cdots R_{m_R}}}$
$1 \le i \le m$;	$[G_{\overline{i+1}}, B_i, R_1, \cdots R_{m_R-1}]$	$\overline{e^{-\mu N_i T_b}} P_{e,B_i\|R_1 R_2 \cdots R_{m_R}}$
$X^{N_i T_b} Y^{N_i E_c}$	$[B_{\overline{i+1}}, B_i, R_1, \cdots R_{m_R-1}]$	$e^{-\mu N_i T_b} P_{e,B_i\|R_1 R_2 \cdots R_{m_R}}$

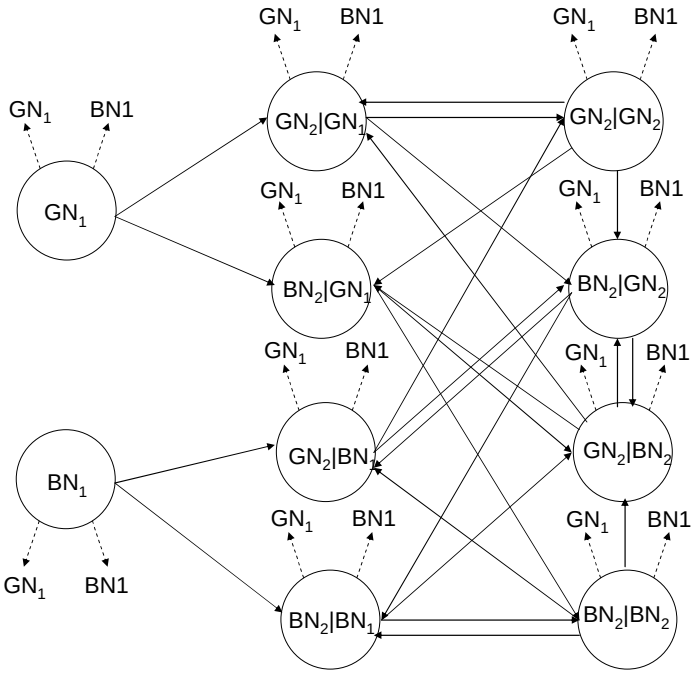

Figure 4.9: FSM representation for repetition SW-ARQ systems with sliding window memory with $m = 2$ and $m_R = 1$.

that the receiver will also refresh its memory contents to be empty after the capacity of the receiver memory reaches m_R as shown in part (c) of Fig. 4.7. In Fig. 4.10, the FSM diagram for SW-ARQ systems with non-overlapping window memory (R-SW-ARQ-NOM) with $m = 2$ and $m_R = 1$ is shown. For general values of m and m_R, we can summarize the FSM states, state transition probabilities, and output functions in Table 4.5.

4.5 Go-Back-N ARQ

We begin this section by describing the operation of Go-Back-N ARQ (GBN-ARQ) protocol. The transmitter sends packets with a sequence number contiguously. Once the transmitter receives a negative acknowledge (NACK) packet after one round-trip propagation time, the transmitter retransmits the received erroneously decoded packet and succeeding packets. If the receiver decodes a packet correctly, it will send

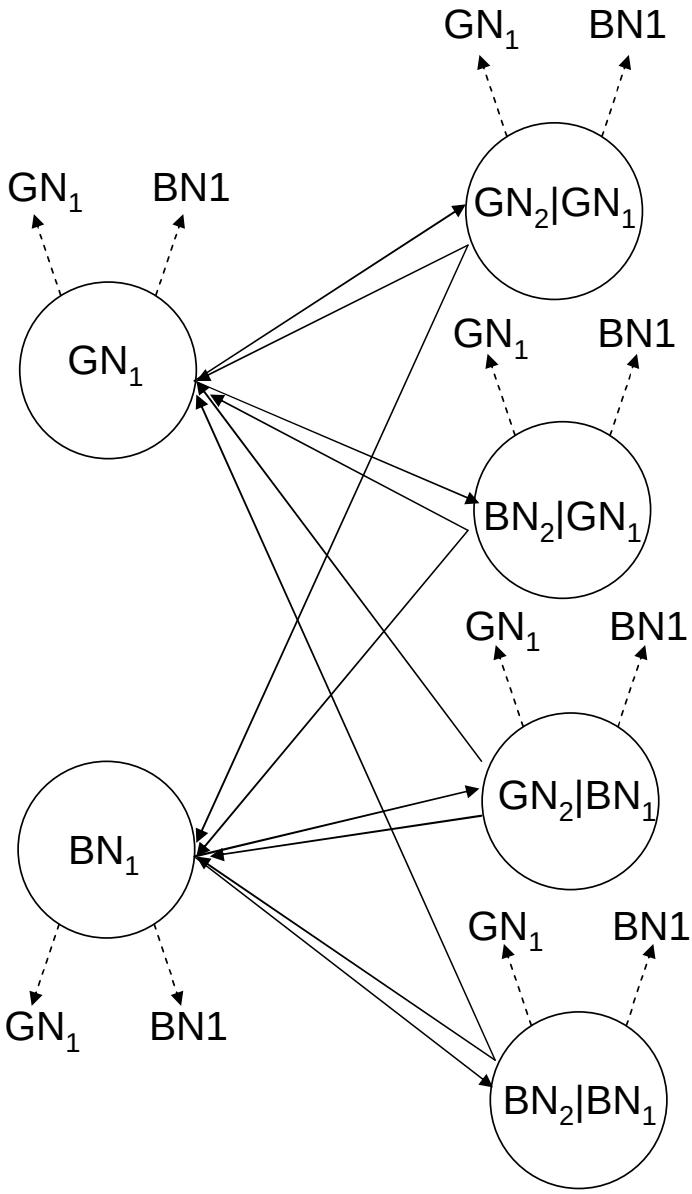

Figure 4.10: FSM representation for repetition SW-ARQ systems with non-overlapping window memory with $m = 2$ and $m_R = 1$.

Table 4.5: Table for transition probabilities and output functions of R-SW-ARQ-NOM

previous state; output	next state	probability
G_1; $X^{N_1T_b}Y^{N_1E_c}$	G_1	$e^{-\lambda N_1 T_b}\overline{P_{e,G_1}}$
	B_1	$\overline{e^{-\lambda N_1 T_b}}\overline{P_{e,G_1}}$
	$[G_2, G_1]$	$e^{-\lambda N_1 T_b}P_{e,G_1}$
	$[B_2, G_1]$	$\overline{e^{-\lambda N_1 T_b}}P_{e,G_1}$
B_1; $X^{N_1T_b}Y^{N_1E_c}$	G_1	$\overline{e^{-\mu N_1 T_b}}\overline{P_{e,B_1}}$
	B_1	$e^{-\mu N_1 T_b}\overline{P_{e,B_1}}$
	$[G_2, B_1]$	$\overline{e^{-\mu N_1 T_b}}P_{e,B_1}$
	$[B_2, B_1]$	$e^{-\mu N_1 T_b}P_{e,B_1}$
$[G_i, R_1, R_2, \cdots, R_{i-1}]$ $i < m$ $X^{N_iT_b}Y^{N_iE_c}$	G_1	$\overline{e^{-\lambda N_i T_b}}\overline{P_{e,G_i\|R_1R_2\cdots R_{i-1}}}$
	B_1	$\overline{e^{-\lambda N_i T_b}}\overline{P_{e,G_i\|R_1R_2\cdots R_{i-1}}}$
	$[G_{i+1}, G_i, R_1, \cdots, R_{i-1}]$	$e^{-\lambda N_i T_b}P_{e,G_i\|R_1R_2\cdots R_{i-1}}$
	$[B_{i+1}, G_i, R_1, \cdots, R_{i-1}]$	$\overline{e^{-\lambda N_i T_b}}P_{e,G_i\|R_1R_2\cdots R_{i-1}}$
$[B_i, R_1, R_2, \cdots, R_{i-1}]$ $i < m$ $X^{N_iT_b}Y^{N_iE_c}$	G_1	$\overline{e^{-\mu N_i T_b}}\overline{P_{e,B_i\|R_1R_2\cdots R_{i-1}}}$
	B_1	$e^{-\mu N_i T_b}\overline{P_{e,B_i\|R_1R_2\cdots R_{i-1}}}$
	$[G_{i+1}, B_i, R_1, \cdots, R_{i-1}]$	$\overline{e^{-\mu N_i T_b}}P_{e,B_i\|R_1R_2\cdots R_{i-1}}$
	$[B_{i+1}, B_i, R_1, \cdots, R_{i-1}]$	$e^{-\mu N_i T_b}P_{e,B_i\|R_1R_2\cdots R_{i-1}}$
$[G_m, R_1, R_2, \cdots, R_{m_R}]$ $X^{N_mT_b}Y^{N_mE_c}$	G_1	$e^{-\lambda N_m T_b}\overline{P_{e,G_m\|R_1R_2\cdots R_{m_R}}}$
	B_1	$\overline{e^{-\lambda N_m T_b}}\overline{P_{e,G_m\|R_1R_2\cdots R_{m_R}}}$
	G_1	$e^{-\lambda N_m T_b}P_{e,G_m\|R_1R_2\cdots R_{m_R}}$
	B_1	$\overline{e^{-\lambda N_m T_b}}P_{e,G_m\|R_1R_2\cdots R_{m_R}}$
$[B_m, R_1, R_2, \cdots, R_{m_R}]$ $X^{N_mT_b}Y^{N_mE_c}$	G_1	$\overline{e^{-\mu N_m T_b}}\overline{P_{e,B_m\|R_1R_2\cdots R_{m_R}}}$
	B_1	$e^{-\mu N_m T_b}\overline{P_{e,B_m\|R_1R_2\cdots R_{m_R}}}$
	G_1	$\overline{e^{-\mu N_m T_b}}P_{e,B_m\|R_1R_2\cdots R_{m_R}}$
	B_1	$e^{-\mu N_m T_b}P_{e,B_m\|R_1R_2\cdots R_{m_R}}$

an acknowledge (ACK) back to the transmitter. The receiver will send a NACK packet back to the transmitter if a packet can not be decoded correctly. From the described protocol of GBN-ARQ, we can incorporate the previous SW-ARQ with incremental redundancy and the receiver memory scheme into GBN-ARQ protocol by using different packet lengths and modifying the decoding structure (from a non-memory receiver to a memory receiver).

Let $\mathbf{T}(N_i)$ be the channel transition probability matrix after transmitting a packet with length N_i. It can be written as:

$$\mathbf{T}(N_i) = \begin{vmatrix} e^{-\lambda N_i T_b} & (1-e^{-\lambda N_i T_b}) \\ (1-e^{-\mu N_i T_b}) & e^{-\mu N_i T_b} \end{vmatrix}. \tag{4.13}$$

Then the transition probability matrix after N_{rt} transmissions of a packet with length N_i is

$$\begin{aligned} \mathbf{T}(N_{rt}N_i) &= [\mathbf{T}(N_i)]^{N_{rt}} \\ &= \frac{1}{2-e^{-\lambda N_i T_b}-e^{-\mu N_i T_b}}\left(\begin{vmatrix} (1-e^{-\mu N_i T_b}) & (1-e^{-\lambda N_i T_b}) \\ (1-e^{-\mu N_i T_b}) & (1-e^{-\lambda N_i T_b}) \end{vmatrix} + \right. \\ &\quad \left. (e^{-\lambda N_i T_b}+e^{-\mu N_i T_b}-1)^{N_{rt}} \begin{vmatrix} (1-e^{-\lambda N_i T_b}) & -(1-e^{-\lambda N_i T_b}) \\ -(1-e^{-\mu N_i T_b}) & (1-e^{-\mu N_i T_b}) \end{vmatrix} \right), \end{aligned} \tag{4.14}$$

where the N_{rt} is the number of packets in a round trip delay. From this transition probability matrix, the transition probability from G to G after one round trip time of transmitting packets with length N_i, denoted as $T_{rt}^{G_iG_i}$, is

$$T_{rt}^{G_iG_i} = 1-(1-e^{-\mu N_i T_b})\frac{1-(e^{-\lambda N_i T_b}+e^{-\mu N_i T_b}-1)^{N_{rt}}}{2-e^{-\lambda N_i T_b}-e^{-\mu N_i T_b}}$$

Similarly, we can obtain the transition probability from G to B, B to G and B to B.

They are

$$\begin{aligned} T_{rt}^{G_iB_i} &= (1-e^{-\mu N_i T_b})\frac{1-(e^{-\lambda N_i T_b}+e^{-\mu N_i T_b}-1)^{N_{rt}}}{2-e^{-\lambda N_i T_b}-e^{-\mu N_i T_b}} \\ T_{rt}^{B_iG_i} &= (1-e^{-\lambda N_i T_b})\frac{1-(e^{-\lambda N_i T_b}+e^{-\mu N_i T_b}-1)^{N_{rt}}}{2-e^{-\lambda N_i T_b}-e^{-\mu N_i T_b}} \\ T_{rt}^{B_iB_i} &= 1-(1-e^{-\lambda N_i T_b})\frac{1-(e^{-\lambda N_i T_b}+e^{-\mu N_i T_b}-1)^{N_{rt}}}{2-e^{-\lambda N_i T_b}-e^{-\mu N_i T_b}} \end{aligned} \tag{4.15}$$

We will give an example about calculating the generating function for a GBNARQ system with $m = 2$ and $m_R = 0$ by using the approach presented in subSection 4.4.1 and the channel transition probabilities (4.15). The generating function is

$$\begin{aligned} G_s &= P_G(1-P_{e,G_1})X^{N_{rt}N_1T_b}Y^{N_{rt}N_1E_c} + P_B * (1-P_{e,B_1})X^{N_{rt}N_1T_b}Y^{N_{rt}N_1E_c} + \\ &\quad (P_GP_{e,G_1}T_{rt}^{G_iG_i} + P_BP_{e,B_1}T_{rt}^{B_1G_1})X^{N_{rt}N_1T_b}Y^{N_{rt}N_1E_c}G_G + \\ &\quad (P_GP_{e,G_1}T_{rt}^{G_iB_i} + P_BP_{e,B_1}T_{rt}^{B_1B_1})X^{N_{rt}N_1T_b}Y^{N_{rt}N_1E_c}G_B, \end{aligned} \tag{4.16}$$

where

$$\begin{aligned} G_G &= \frac{(X^{N_2T_b}Y^{N_2E_c}(1-P_{e,B_2}-[(1-P_{e,B_2})P_{e,G_2}T_{rt}^{B_2B_2}-P_{e,B_2}(1-P_{e,G_2})T_{rt}^{G_2B_2}]Z))}{1-[P_{e,B_2}T_{rt}^{G_2G_2}+P_{e,G_2}T_{rt}^{B_2B_2}]Z+P_{e,B_2}P_{e,G_2}(T_{rt}^{G_2G_2}+T_{rt}^{B_2B_2}-1)Z^2} \\ G_B &= \frac{(X^{N_2T_b}Y^{N_2E_c}(1-P_{e,G_2}-[(1-P_{e,G_2})P_{e,B_2}T_{rt}^{G_2G_2}-P_{e,G_2}(1-P_{e,B_2})T_{rt}^{B_2G_2}]Z))}{1-[P_{e,G_2}T_{rt}^{B_2B_2}+P_{e,B_2}T_{rt}^{G_2G_2}]Z+P_{e,G_2}P_{e,B_2}(T_{rt}^{B_2B_2}+T_{rt}^{G_2G_2}-1)Z^2} \end{aligned} \tag{4.17}$$

and $Z = X^{N_{rt}N_2T_b}Y^{N_{rt}N_2E_c}$. We also can obtain the generating function of GBN-ARQ with the receiver memory scheme by using the same method as we derived before. The only change is to modify the channel transition probabilities and multiply the round trip factor N_{rt} in energy and delay components at each transmission.

4.6 Cutoff Rate for Memory Receiver over Memoryless Channels

Let K be the number of information bits in a packet and N be the number of coded bits for this packet. From the random coding argument, it is well known that

there exist an encoder and decoder that has packet error probability $P_{e,K,N}$ bounded as

$$P_{e,K,N} \leq 2^{K-NR_0}, \tag{4.18}$$

where R_0 is the cutoff rate determined by the channel SNR and other channel characteristics. Because the channel SNR and packet length used for packets transmission are two crucial components in determining the energy and delay tradeoff of an ARQ system. By using the cutoff rate to represent the packet error probability, we can relate the channel SNR and packet length N with system performance.

The cutoff rate of two different kinds of receiver structure will be derived in this section. The first cutoff rate is evaluated by assuming that the receiver has all the channel state information. The second cutoff rate is calculated according to the imperfect channel state information by using estimated channel transition probability of each signal. The third cutoff rate is derived by the same assumption as second cutoff rate, however, the receiver with memory will utilize more information (the failure information) to decode.

4.6.1 Random Coding Bound for Memory Receiver in AWGN channel- Part I: Perfect Channel Knowledge

Let m be the total number of packets combined at the receiver which are used to decode jointly. Let N_j be the number of coded symbols used at j-th packet at the receiver memory, where $j \in \{1, 2, \cdots, m\}$. Here we consider a set of M-ary coded signals constructed from $\sum_{j=1}^{m} N_j$-dimensional q symbols. Let $\boldsymbol{s_i}$ be the signal vector received at the receiver end. It is

$$s_i = [\overbrace{s_{i_{11}} s_{i_{12}} \cdots s_{i_{1N_1}}}^{\text{packet 1}} \overbrace{s_{i_{21}} s_{i_{22}} \cdots s_{i_{2N_m}}}^{\text{packet 2}} \cdots \overbrace{s_{i_{m1}} s_{i_{m2}} \cdots s_{i_{mN_m}}}^{\text{packet } m}], \tag{4.19}$$

where $s_{i_{j,k}}$ are selected from the signal set $\{a_1, a_2, \cdots a_q\}$ of q elements.

According to the random coding argument, there are $B = q^{M\sum_{j=1}^{m} N_j}$ possible sets of M signals since there are $\sum_{j=1}^{m} N_j$ positions can be chosen from q symbols. Let $P_{e,i}(k)$ be the error probability of signal set k given the i-th message was transmitted. The average error probability given the transmission of the i-th message is $\overline{P}_{e,i} = \sum_{k=1}^{B} P_{e,i}(k)$. By the union-Bhattacharyya bound for an AWGN channel, the average error probability of transmitting i-th message over the ensemble of codewords sets is

$$\overline{P}_{e,i} \leq \sum_{i \neq j} E[e^{-\frac{\|s_i - s_j\|^2}{4N_0}}]. \tag{4.20}$$

If we can show that $\overline{P}_{e,i} \leq \alpha$, then there exists a signal set with $P_{e,i} \leq \alpha$. Let $d_{l,m} = E_k \parallel a_l - a_m \parallel^2$, where E_k is the received energy level of packet k and $\parallel a_l - a_m \parallel$ is the distance between signal a_l and a_m. If we assume that the selection of symbols $\{a_1, a_2, \cdots a_q\}$ is according to probabilities $\{p_1, p_2, \cdots p_q\}$. Then we have

$$\overline{P}_{e,i} \leq \sum_{i \neq j} \prod_{k=1}^{m} (\sum_{l_k=1}^{q} \sum_{m_k=1}^{q} p_{l_k} p_{m_k} e^{-\frac{d^2_{l_k m_k}}{4N_0}})^{N_k} \tag{4.21}$$

Therefore, if we set $R_{0,k} = -\log_2(\sum_{l_k=1}^{q} \sum_{m_k=1}^{q} p_{l_k} p_{m_k} e^{-\frac{d^2_{l_k m_k}}{4N_0}})$ we can further express the right hand side of (4.21) as

$$\begin{aligned} \overline{P}_{e,i} &\leq (M-1) \prod_{k=1}^{m} (\sum_{l_k=1}^{q} \sum_{m_k=1}^{q} p_{l_k} p_{m_k} e^{-\frac{d^2_{l_k m_k}}{4N_0}})^{N_k} \\ &< M \prod_{k=1}^{m} 2^{-N_k R_{0,k}} \\ &= 2^{-(\sum_{k=1}^{m} N_k)(R_{0,MRX} - R)}, \end{aligned} \tag{4.22}$$

where $R = \frac{\log_2 M}{\sum_{k=1}^{m} N_k}$ is the code rate after packets combination and $R_{0,MRX}$ is

$$R_{0,MRX} = \frac{N_1}{\sum_{k=1}^{m} N_k} R_{0,1} + \frac{N_2}{\sum_{k=1}^{m} N_k} R_{0,2} + \cdots + \frac{N_m}{\sum_{k=1}^{m} N_k} R_{0,m}. \tag{4.23}$$

We can interpret $R_{0,MRX}$ as the cutoff rate by combining m packets according to their packet length ratio to the total length of m packets at the receiver. We only

show that there exist a signal, say i-th signal, for which the average error probability can be made arbitrary small as packets length becomes large. We will show that we can find a single codeword set such that the error probability of each signal of this codeword set approaches to zero as packets length approaches to infinity.

Choosing a code with $M = 2 \cdot 2^{R_{0,MRX}(\sum_{k=1}^{m} N_k)}$ codewords that the average error probability of this code is less then $\epsilon/2$ for large $\sum_{k=1}^{m} N_k$. There are at least $M/2$ of the codewords with error probability less then ϵ (otherwise the average error probability of this code will greater then $\epsilon/2$). By deleting those bad codewords (their error probability is greater then ϵ), we obtain a code with (at least) $2^{R_{0,MRX}(\sum_{k=1}^{m} N_k)}$ codewords with error probability small as $\sum_{k=1}^{m} N_k \longrightarrow \infty$.

4.6.2 Random Coding Bound for Memory Receiver - Part II : Imperfect Channel Knowledge

In this section, we consider that the receiver uses the channel transition probabilities $p^*(y|x)$ as if they were the true transition probabilities. We will derive random coding bound of the receiver with memory with m packets. When the channel condition is good, the receiver end has signal-to-noise ratio E_G/N_0 but the receiver uses the incorrect signal-to-noise ratio E_G^*/N_0 to decode. Similarly, the receiver end has signal-to-noise ratio E_B/N_0 when the channel condition is bad but the receiver uses the incorrect signal-to-noise ratio E_B^*/N_0 to decode.

Let $\boldsymbol{s}_i$ be the signal vector received at the receiver end. It is

$$s_i = [\overbrace{s_{i_{11}} s_{i_{12}} \cdots s_{i_{1N_1}}}^{\text{packet 1}} \overbrace{s_{i_{21}} s_{i_{22}} \cdots s_{i_{2N_m}}}^{\text{packet 2}} \cdots \overbrace{s_{i_{m1}} s_{i_{m2}} \cdots s_{i_{mN_m}}}^{\text{packet } m}], \tag{4.24}$$

where $s_{i_{j,k}} \in \{a_1, a_2, \cdots, a_q\}$.

The decision region are

$$R_i^* = \{\mathbf{y} : p^*(\mathbf{y}|\mathbf{s}_i, \overline{s}_R) > p^*(\mathbf{y}|\mathbf{s}_j, \overline{s}_R) \text{for all } j \neq i\}, \tag{4.25}$$

where $\overline{s}_R$ is the receiver memory state. Let $\overline{R}_i^*$ be the region where signal i is not decided. Then

$$\overline{R}_i^* = \bigcup_{j \neq i} R_{i,j}^*, \tag{4.26}$$

where $R_{i,j}^* = \{\mathbf{y} : p^*(\mathbf{y}|\mathbf{s}_i) < p^*(\mathbf{y}|\mathbf{s}_j)\}$. Then we have

$$\begin{aligned} P_{e,i} &= P(\overline{R}_i^*|H_i) \\ &\leq \sum_{j \neq i} P(R_{i,j}^*). \end{aligned} \tag{4.27}$$

For each term of above summand, we have

$$P(R_{i,j}^*) = \int_{\mathbf{y} \in R_{i,j}^*} p(\mathbf{y}|\mathbf{s}_i, \overline{s}_R). \tag{4.28}$$

Since

$$\frac{p^*(\mathbf{y}|\mathbf{s}_j, \overline{s}_R)}{p^*(\mathbf{y}|\mathbf{s}_i, \overline{s}_R)} > 1, \, \mathbf{y} \in R_{i,j}^*, \tag{4.29}$$

we can further bound (4.28) as follows

$$P(R_{i,j}^*) < \int_{\mathbf{y} \in R_{i,j}^*} p(\mathbf{y}|\mathbf{s}_i, \overline{s}_R) \Big[\frac{p^*(\mathbf{y}|\mathbf{s}_j, \overline{s}_R)}{p^*(\mathbf{y}|\mathbf{s}_i, \overline{s}_R)}\Big]^{\lambda}, \tag{4.30}$$

where $\lambda > 0$. By relaxing the integral domain, we can upper bound (4.30) as

$$P(R_{i,j}^*) < \int_{\mathbf{y} \in \mathbf{R}^{\sum_{k=1}^m N_k}} p(\mathbf{y}|\mathbf{s}_i, \overline{s}_R) \Big[\frac{p^*(\mathbf{y}|\mathbf{s}_j, \overline{s}_R)}{p^*(\mathbf{y}|\mathbf{s}_i, \overline{s}_R)}\Big]^{\lambda}. \tag{4.31}$$

Thus, the error probability $P_{e,i}$ is upper bounded by

$$\begin{aligned}
P_{e,i} &< \sum_{j\neq i}\int_{\mathbf{y}\in\mathbf{R}^{\sum_{k=1}^{m}N_k}} p(\mathbf{y}|\mathbf{s}_i,\overline{s}_R)\Big[\frac{p^*(\mathbf{y}|\mathbf{s}_j,\overline{s}_R)}{p^*(\mathbf{y}|\mathbf{s}_i,\overline{s}_R)}\Big]^{\lambda} \\
&= \sum_{j\neq i}\int_{\mathbf{y}\in\mathbf{R}^{\sum_{k=1}^{m}N_k}} \Big(\prod_{l_1=1}^{N_1} p(y_{l_1}|s_{i_{1l_1}},R_1)\Big[\frac{p^*(y_{l_1}|s_{j_{1l_1}},R_1)}{p^*(y_{l_1}|s_{i_{1l_1}},R_1)}\Big]^{\lambda}\Big)\times \\
&\quad \Big(\prod_{l_2=1}^{N_2} p(y_{l_2}|s_{i_{2l_2}},R_2)\Big[\frac{p^*(y_{l_2}|s_{j_{2l_2}},R_2)}{p^*(y_{l_2}|s_{i_{2l_2}},R_2)}\Big]^{\lambda}\Big)\cdots \\
&\quad \Big(\prod_{l_m=1}^{N_m} p(y_{l_m}|s_{i_{ml_m}},R_m)\Big[\frac{p^*(y_{l_m}|s_{j_{ml_m}},R_m)}{p^*(y_{l_m}|s_{i_{ml_m}},R_m)}\Big]^{\lambda}\Big) \\
&= \sum_{j\neq i}\Big(\prod_{l_1=1}^{N_1}\int_{y_{l_1}\in\mathbf{R}} p(y_{l_1}|s_{i_{1l_1}},R_1)\Big[\frac{p^*(y_{l_1}|s_{j_{1l_1}},R_1)}{p^*(y_{l_1}|s_{i_{1l_1}},R_1)}\Big]^{\lambda}\Big)\times \\
&\quad \Big(\prod_{l_2=1}^{N_2}\int_{y_{l_2}\in\mathbf{R}} p(y_{l_2}|s_{i_{2l_2}},R_2)\Big[\frac{p^*(y_{l_2}|s_{j_{2l_2}},R_2)}{p^*(y_{l_2}|s_{i_{2l_2}},R_2)}\Big]^{\lambda}\Big)\cdots \\
&\quad \Big(\prod_{l_m=1}^{N_m}\int_{y_{l_m}\in\mathbf{R}} p(y_{l_m}|s_{i_{ml_m}},R_m)\Big[\frac{p^*(y_{l_m}|s_{j_{ml_m}},R_m)}{p^*(y_{l_m}|s_{i_{ml_m}},R_m)}\Big]^{\lambda}\Big). \qquad (4.32)
\end{aligned}$$

By taking the error probability over all possible codes and let $p(\mathbf{s}_i)=\prod_{k=1}^{m}(\prod_{l_k=1}^{N_k}p(s_{i_{kl_k}}))$.

Then

$$\begin{aligned}
\overline{P}_{e,i} &< \sum_{j\neq i}\sum_{\mathbf{s}_i}\sum_{\mathbf{s}_j}\Big(\prod_{l_1=1}^{N_1} p(s_{i_{1l_1}})p(s_{j_{1l_1}})\int_{y_{l_1}\in\mathbf{R}} p(y_{l_1}|s_{i_{1l_1}},R_1)\Big[\frac{p^*(y_{l_1}|s_{j_{1l_1}},R_1)}{p^*(y_{l_1}|s_{i_{1l_1}},R_1)}\Big]^{\lambda}\Big)\times \\
&\quad \Big(\prod_{l_2=1}^{N_2} p(s_{i_{2l_2}})p(s_{j_{2l_2}})\int_{y_{l_2}\in\mathbf{R}} p(y_{l_2}|s_{i_{2l_2}},R_2)\Big[\frac{p^*(y_{l_2}|s_{j_{2l_2}},R_2)}{p^*(y_{l_2}|s_{i_{2l_2}},R_2)}\Big]^{\lambda}\Big)\cdots \\
&\quad \Big(\prod_{l_m=1}^{N_m} p(s_{i_{ml_m}})p(s_{j_{ml_m}})\int_{y_{l_m}\in\mathbf{R}} p(y_{l_m}|s_{i_{ml_m}},R_m)\Big[\frac{p^*(y_{l_m}|s_{j_{ml_m}},R_m)}{p^*(y_{l_m}|s_{i_{ml_m}},R_m)}\Big]^{\lambda}\Big) \\
&= \sum_{j\neq i}\Big(\prod_{l_1=1}^{N_1}\sum_{s_{i_{1l_1}}}\sum_{s_{j_{1l_1}}} p(s_{i_{1l_1}})p(s_{j_{1l_1}})\int_{y_{l_1}\in\mathbf{R}} p(y_{l_1}|s_{i_{1l_1}},R_1)\Big[\frac{p^*(y_{l_1}|s_{j_{1l_1}},R_1)}{p^*(y_{l_1}|s_{i_{1l_1}},R_1)}\Big]^{\lambda}\Big)\times \\
&\quad \Big(\prod_{l_2=1}^{N_2}\sum_{s_{i_{2l_2}}}\sum_{s_{j_{2l_2}}} p(s_{i_{2l_2}})p(s_{j_{2l_2}})\int_{y_{l_2}\in\mathbf{R}} p(y_{l_2}|s_{i_{2l_2}},R_2)\Big[\frac{p^*(y_{l_2}|s_{j_{2l_2}},R_2)}{p^*(y_{l_2}|s_{i_{2l_2}},R_2)}\Big]^{\lambda}\Big)\cdots \\
&\quad \Big(\prod_{l_m=1}^{N_m}\sum_{s_{i_{ml_m}}}\sum_{s_{j_{ml_m}}} p(s_{i_{ml_m}})p(s_{j_{ml_m}})\int_{y_{l_m}\in\mathbf{R}} p(y_{l_m}|s_{i_{ml_m}},R_m)\Big[\frac{p^*(y_{l_m}|s_{j_{ml_m}},R_m)}{p^*(y_{l_m}|s_{i_{ml_m}},R_m)}\Big]^{\lambda}\Big) \\
&= \sum_{j\neq i}[E[J^*_{\lambda,1}(X_1,X_2)]]^{N_1}[E[J^*_{\lambda,2}(X_1,X_2)]]^{N_2}\cdots[E[J^*_{\lambda,m}(X_1,X_2)]]^{N_m} \\
&= (M-1)[E[J^*_{\lambda,1}(X_1,X_2)]]^{N_1}[E[J^*_{\lambda,2}(X_1,X_2)]]^{N_2}\cdots[E[J^*_{\lambda,m}(X_1,X_2)]]^{N_m} \qquad (4.33)
\end{aligned}$$

where $E[J^*_{\lambda,n}(X_1,X_2)]$ is $\sum_{s_{i_{nl_n}}}\sum_{s_{j_{nl_n}}} p(s_{i_{nl_n}})p(s_{j_{nl_n}})\int_{y_{nl_n}\in\mathbf{R}} p(y_{nl_n}|s_{i_{nl_n}},R_n)\Big[\frac{p^*(y_{nl_n}|s_{j_{nl_n}},R_n)}{p^*(y_{nl_n}|s_{i_{nl_n}},R_n)}\Big]^{\lambda}$ for the n-th packets in the receiver memory. The random variable X_1 represents source signals $s_{i_{lk}}$ and X_2 represents source signals $s_{j_{lk}}$. If let $R^*_{0,k}(\lambda)$ as

$$\begin{aligned}
R^*_{0,k}(\lambda) &= \max_{p(s_{i_{kl_k}})} -\log_2\Big\{\sum_{s_{i_{kl_k}}}\sum_{s_{j_{kl_k}}} p(s_{i_{kl_k}})p(s_{j_{kl_k}})\int_{ky_{l_k}\in\mathbf{R}} p(y_{kl_k}|s_{i_{kl_k}},R_k)\Big[\frac{p^*(y_{kl_k}|s_{j_{kl_k}},R_k)}{p^*(y_{kl_k}|s_{i_{kl_k}},R_k)}\Big]^{\lambda}\Big\}, \\
&\quad \text{and } s_{i_{kl_k}}\in\{a_1,a_2,\cdots a_q\}. \qquad (4.34)
\end{aligned}$$

Then we can further upper bound of (4.33) as

$$\overline{P}_{e,i} < 2^{-(\sum_{k=1}^m N_k)(R^*_{0,MRX}-R)}, \qquad (4.35)$$

where $R=\frac{\log_2 M}{\sum_{k=1}^m N_k}$ is the code rate after packets combination. $R^*_{0,MRX}$ is

$$R^*_{0,MRX} = \max_{\lambda}\Big[\frac{N_1}{\sum_{k=1}^m N_k}R^*_{0,1}(\lambda)+\frac{N_2}{\sum_{k=1}^m N_k}R^*_{0,2}(\lambda)+\cdots+\frac{N_m}{\sum_{k=1}^m N_k}R^*_{0,m}(\lambda)\Big]. \qquad (4.36)$$

The channel model used in the following cutoff rate evaluation is Gilbert-Elliott channel model described at the beginning of subsection 4.4.1. We assume that the first packet with length N_1 is decoded erroneously after receiving over the good channel condition. Therefore, it is stored in the receiver memory. The cutoff rate is evaluated by considering the first erroneous packet stored in the receiver memory and the second incoming packet with length N_2 from bad channel condition jointly. The cutoff rate is plotted with respect to good channel SNR for the memory receiver with perfect channel knowledge described in Subsection 4.6.1 (indicated by symbol I) and the memory receiver with imperfect channel knowledge described in Subsection 4.6.2 (represented by symbol II) in Fig.4.11. We observe that the cutoff rate of the memory receiver structure I is higher than the cutoff rate of the memory receiver structure II because the receiver of structure I has much more information about the channel condition. Since the cutoff rate with parameters $N_1 = 2N_2$ has higher chance for the joint packet (the packet combining the first erroneous one and the second incoming one) to stay at good channel condition compared to the cutoff rate with parameters $N_2 = 2N_1$, the cutoff rate represented by the solid lines is lower then the cutoff rate represented by the dashed lines. The cutoff rate with symbol + is higher then the cutoff rate with symbol o because the lines with symbol + have higher bad channel SNR compared to the lines with symbol o.

4.6.3 Existence of Codes for Memory Receiver

In the previous derivation of cutoff rate for the receiver with memory, we know that there is an ensemble random codes of m blocks, $(\mathcal{C}_1, \mathcal{C}_2, \cdots, \mathcal{C}_m)$ satisfying the random coding bound if there $m-1$ packets stored in the receiver memory. However, it does not guarantee that the same code ensemble can satisfy the random coding bound for any usage of previous k codes for $k < m$, i.e., there are only $k-1$ packets

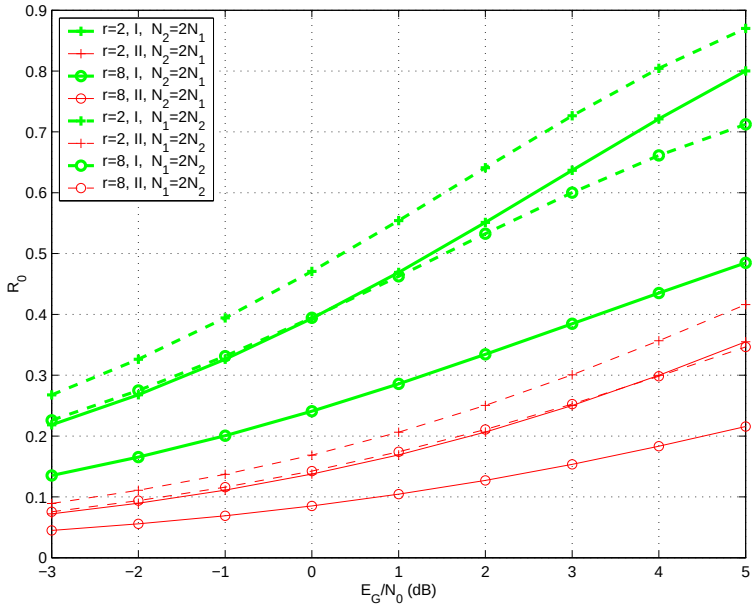

Figure 4.11: Cutoff rate comparison for two different receiver structures.

stored in the receiver memory. The following theorem indicates such ensemble codes exist [49].

Theorem : There exists an ensemble of codes $(\mathcal{C}_1, \mathcal{C}_2, \cdots, \mathcal{C}_m)$ **satisfying** $P_e(\mathcal{C}_1, \mathcal{C}_2, \cdots, \mathcal{C}_k) \leq 2^{-(\sum_{l=1}^{k})(R_{0,MRX}^{k} - R)}$ **for** $1 \leq k \leq m$**.**

Proof : By applying Markov inequality, we have

$$Pr(\mathbf{P}_{e,k} > \beta 2^{-(\sum_{l=1}^{k})(R_{0,MRX}^{k} - R)}) \leq \frac{1}{\beta}, \qquad \text{for } 1 \leq k \leq m. \tag{4.37}$$

where $\mathbf{P}_{e,k}$ is the probability of error for previous k codes. Note that $\mathbf{P}_{e,k}$ is a random number since we use random coding selection for codes construction. Hence,

$$\begin{aligned} Pr(\bigcap_{k=1}^{m}\{\mathbf{P}_{e,k} > \beta 2^{-(\sum_{l=1}^{k})(R_{0,MRX}^{k} - R)}\}) &= 1 - Pr(\bigcup_{k=1}^{m}\{\mathbf{P}_{e,k} > \beta 2^{-(\sum_{l=1}^{k})(R_{0,MRX}^{k} - R)}\}) \\ &\geq 1 - \sum_{k=1}^{m} Pr(\mathbf{P}_{e,k} > \beta 2^{-(\sum_{l=1}^{k})(R_{0,MRX}^{k} - R)}) \\ &\geq 1 - \frac{m}{\beta} > 0 \end{aligned} \tag{4.38}$$

by choosing a suitable β, say $\beta = 2m$. This implies that there exist at least an

ensemble of m blocks, $(\mathcal{C}_1^*, \mathcal{C}_2^*, \cdots, \mathcal{C}_m^*)$ with

$$P_{e,k}^* \leq \beta 2^{-(\sum_{l=1}^{k})(R_{0,MRX}^k - R)}), \qquad \text{for } 1 \leq k \leq m. \tag{4.39}$$

Here we can see that the factor β has no effect on the cutoff rate. □

4.6.4 Cutoff Rate Optimization

In the previous derivation of cutoff rate, we need to optimize the following cutoff rate per waveform given λ

$$R_0(\lambda) \triangleq \max_{P_j} \left\{ -\log \left[\sum_j \sum_k P_j P_k \int_z f(z|k)[f^*(z|j) f^*(z|k)]^\lambda dz \right] \right\}, \tag{4.40}$$

where

1. P_j is the a priori probability of the j-th symbol selected by the transmitter.
2. z denotes a received signal.
3. j, k are the index for symbols.
4. $f(z|j)$ is the true channel transition probability, i.e., the probability density function of the receiver output, assuming the j-th symbol was transmitted. The superscript symbol $*$ indicates the incorrect channel transition probability used by the receiver to decode.

Denote $D_{j,k}$ by the quantity $\int_z f(z|k)[f^*(z|j) f^*(z|k)]^\lambda dz$. If $\lambda = 1/2$, $f(z|k) = f^*(z|k)$ and the system is operated in an AWGN channel, the value $\int_z f(z|k)[f^*(z|j) f^*(z|k)]^\lambda dz$ is equal to $e^{-d_B(j,k)}$ where $d_B(j,k)$ is Bhattachryya distance. We want to find the priori probability of the input symbol P_j in two cases; (1) there is no constraint of P_j except that its components are sum up to one and (2) there is an energy constraint of input symbols.

We will consider the case (1) first. The optimization problem of this case can be formulated as follows:

$$\begin{aligned} R_0 &= \max_{P_j}\{-\log[\sum_j \sum_k P_j P_k D_{j,k}]\} \\ &= \max_{P_j}\{-\log[\mathbf{P}^T\mathbf{D}\mathbf{P}]\}, \end{aligned} \tag{4.41}$$

under constraint

$$\sum_j P_j = 1. \tag{4.42}$$

From the signal constellation, the matrix $\mathbf{D}$ has the following form

$$\begin{bmatrix} 1 & D_{1,2} & \cdots & D_{1,M} \\ D_{2,1} & 1 & \cdots & D_{2,M} \\ \vdots & \vdots & \cdots & \vdots \\ D_{M,1} & D_{M,2} & \cdots & 1 \end{bmatrix}, \tag{4.43}$$

and $D_{i,j} = D_{j,i}$ due to the symmetry of the Bhattachryya distance function. From this, we can write the term $\mathbf{P}^T\mathbf{D}\mathbf{P}$ as

$$\mathbf{P}^T\mathbf{D}\mathbf{P} = \sum_{i=1}^{M} P_i^2 + 2 \sum_{1\le i<j\le M} D_{i,j} P_i P_j. \tag{4.44}$$

The quantity $\mathbf{P}^T\mathbf{D}\mathbf{P}$ is a quadratic function of $\mathbf{P}$, therefore, it is a convex $\bigcup$ function. From Kuhn-Tucker theorem, the necessary and sufficient conditions on the probability vector $\mathbf{P}$ that minimizes $\mathbf{P}^T\mathbf{D}\mathbf{P}$ are

$$P_i + \sum_{s=1,\neq i}^{M} D_{s,i} P_s \ge \frac{-2\det\mathbf{D}}{\sum_{s=1}^{M} \det[\mathbf{D}, \mathbf{1_s}]}; \qquad \text{for all } i \tag{4.45}$$

where $[\mathbf{D}, \mathbf{1_s}]$ is matrix obtained by replacing the s-th column of matrix $\mathbf{D}$ by one vector. The equality will hold for all i such that $P_i > 0$. Because the probability vector obtained in (4.45) minimizes $\mathbf{P}^T\mathbf{D}\mathbf{P}$, it will also maximize the cutoff rate R_0.

According to the previous derivation, we can calculate the cutoff rate for optimal a priori probability input signals. In Fig.4.12, we plot the cutoff rates for uniform and optimal priori probability input signals for PAM modulation scheme with $M = 3$ and $M = 4$. The amplitude levels are assigned equally spaced over the interval $[-\sqrt{E_c}, +\sqrt{E_c}]$. We observe that there is about 40% increase of cutoff rate at -10 dB for $M = 3$ and about 60% increase of cutoff rate at -10 dB for $M = 4$. This indicates that the suitable selection of a priori probability for the input signals is crucial especially for the bad channel condition. We also note that when the channel SNR is low, the cutoff rate of smaller input signals is higher then the larger input signals due to larger distance between signals. However, when the channel SNR is high, the cutoff rate of smaller input signals is lower then the larger input signals due to high bandwidth efficiency.

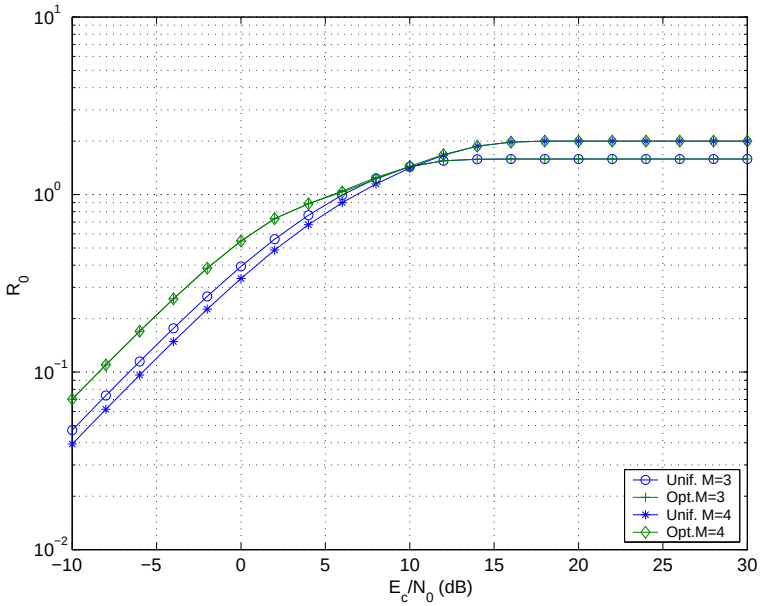

Figure 4.12: The comparison of cutoff rate for uniform and optimal priori probability for input signals.

Now, we will consider the case (2). The optimization problem of this case can be formulated as follows:

$$\begin{aligned} R_0 &= \max_{P_j}\{-\log[\sum_j \sum_k P_j P_k D_{j,k}]\} \\ &= \max_{P_j}\{-\log[\mathbf{P}^T\mathbf{D}\mathbf{P}]\}, \end{aligned} \tag{4.46}$$

under constraint

$$\sum_j P_j = 1, \qquad \text{and} \qquad \sum_j P_j E_j = \mathbf{P}^T\mathbf{E} = E_{av}, \tag{4.47}$$

where E_j is the energy used in j-th signal and E_{av} is the average energy constraint.

In order to minimize $\mathbf{P}^T\mathbf{D}\mathbf{P}$ with these two constraints, we write the constrained minimization using Lagrange multipliers as the minimization of

$$J(\mathbf{p}) = \sum_{i=1}^{M} P_i^2 + 2 \sum_{1\le i<j\le M} D_{i,j} P_i P_j + \lambda_1(\sum_{i=1}^{M} P_i - 1) + \lambda_2(\sum_{i=1}^{M} P_i E_i - E_{av}) \tag{4.48}$$

Differentiating with respect to P_i, we get

$$\frac{\partial J}{\partial P_i} = 2P_i + \sum_{s=1,\neq i}^{M} D_{s,i} P_s + \lambda_1 + \lambda_2 E_i. \tag{4.49}$$

Setting the derivative to 0, we obtain a system of linear equations of P_i

$$2P_i + \sum_{s=1,\neq i}^{M} D_{s,i} P_s = -\lambda_1 - \lambda_2 E_i, \qquad \text{for } 1 \le i \le M. \tag{4.50}$$

Solving this, the P_i can be expressed as

$$\begin{aligned} P_i &= \frac{\det[\widetilde{\mathbf{D}}; \begin{bmatrix} -\lambda_1 - \lambda_2 E_1 \\ -\lambda_1 - \lambda_2 E_2 \\ \vdots \\ -\lambda_1 - \lambda_2 E_M \end{bmatrix}_i]}{\det\widetilde{\mathbf{D}}} \\ &= \frac{-\lambda_1 \det[\widetilde{\mathbf{D}}; \mathbf{1}_i] - \lambda_2 \det[\widetilde{\mathbf{D}}; \mathbf{E}_i]}{\det\widetilde{\mathbf{D}}}, \end{aligned} \tag{4.51}$$

where $\widetilde{D}_{i,j} = D_{i,j}$ for $i \neq j$ and $\widetilde{D}_{i,i} = 2$ for $1 \leq i \leq M$. The matrix $[\widetilde{\mathbf{D}}; \mathbf{E}_i]$ is obtained by replacing its i-th column by the energy vector $\mathbf{E} = [E_1, E_2, \cdots, E_M]^T$. Substituting (4.51) into the two constraints, we can solve for λ_1 and λ_2. The solutions are

$$\lambda_1 = \frac{\det \begin{bmatrix} 1 & \mathbf{b}^T\mathbf{1} \\ E_{av} & \mathbf{b}^T\mathbf{E} \end{bmatrix}}{\det \begin{bmatrix} \mathbf{a}^T\mathbf{1} & \mathbf{b}^T\mathbf{1} \\ \mathbf{a}^T\mathbf{E} & \mathbf{b}^T\mathbf{E} \end{bmatrix}}, \qquad \lambda_2 = \frac{\det \begin{bmatrix} \mathbf{a}^T\mathbf{1} & 1 \\ \mathbf{a}^T\mathbf{E} & E_{av} \end{bmatrix}}{\det \begin{bmatrix} \mathbf{a}^T\mathbf{1} & \mathbf{b}^T\mathbf{1} \\ \mathbf{a}^T\mathbf{E} & \mathbf{b}^T\mathbf{E} \end{bmatrix}} \tag{4.52}$$

where

$$\mathbf{a} = \begin{bmatrix} \frac{-\det[\widetilde{\mathbf{D}};\mathbf{1}_1]}{\det\widetilde{\mathbf{D}}} \\ \frac{-\det[\widetilde{\mathbf{D}};\mathbf{1}_2]}{\det\widetilde{\mathbf{D}}} \\ \vdots \\ \frac{-\det[\widetilde{\mathbf{D}};\mathbf{1}_M]}{\det\widetilde{\mathbf{D}}} \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} \frac{-\det[\widetilde{\mathbf{D}};\mathbf{E}_1]}{\det\widetilde{\mathbf{D}}} \\ \frac{-\det[\widetilde{\mathbf{D}};\mathbf{E}_2]}{\det\widetilde{\mathbf{D}}} \\ \vdots \\ \frac{-\det[\widetilde{\mathbf{D}};\mathbf{E}_M]}{\det\widetilde{\mathbf{D}}} \end{bmatrix} \tag{4.53}$$

From the Kuhn-Tucker theorem, the necessary and sufficient conditions on the probability vector $\mathbf{P}$ that minimizes $\mathbf{P}^T\mathbf{D}\mathbf{P}$ are

$$P_i + \sum_{s=1,\neq i}^{M} D_{s,i}P_s \geq \frac{-\lambda_1 - \lambda_2 E_i}{2}; \qquad \text{for all } i \tag{4.54}$$

According to the previous derivation, we can calculate the cutoff rate for a optimal priori probability with average energy constraints. In Fig.4.13, we plot the cutoff rate for different average energy constraints for 8 QAM.

4.7 Numerical Results

In this section, we will compare the system performance of SW-ARQ ,SW-MARQ and repetition ARQ over time varying channels. The modulation adopted in the transmitter is BPSK. The channel model used for the numerical evaluation is the

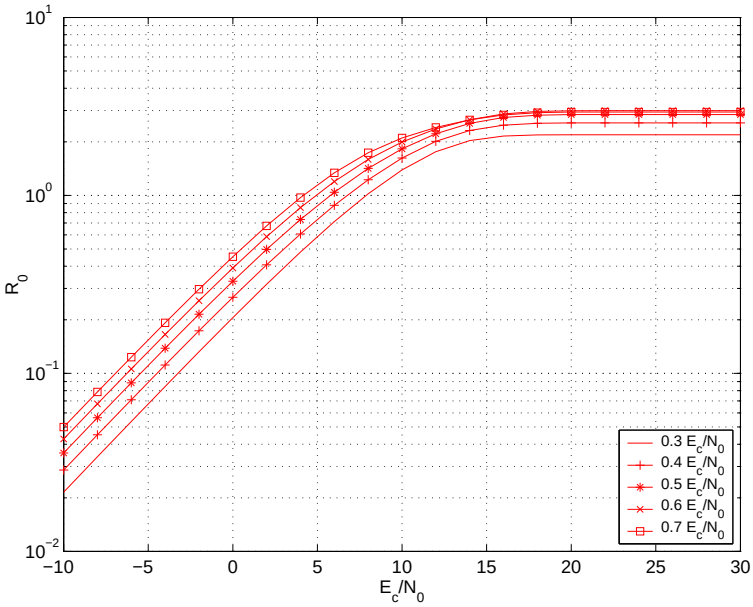

Figure 4.13: The comparison of cutoff rate for different average energy constraints for 8 QAM

two state Gilbert-Elliott model. All packet error probabilities are evaluated with the random coding bound derived in Section 4.6 by assuming that the receiver has perfect knowledge of the channel state information. Let K be the number of information bits in a packet and N be the number of coded bits for this packet, then there exist an encoder and decoder that has packet error probability $P_{e,K,N}$ bounded by

$$P_{e,K,N,\{G,B\}} \leq 2^{K-NR_{0,\{G,B\}}}, \tag{4.55}$$

where $R_{0,\{G,B\}} = 1 - \log_2(1 + e^{-E_{c,\{G,B\}}/N_0})$ is the cutoff rate determined by the channel SNR which will depend on the packet is transmitted in the good or bad channel state. For the receiver with memory, the packet error probability can be written as:

$$P_{e,K,N,\{G,B\};\mathbf{S}_R} \leq 2^{K-NR_{0,\{GN,BN\};\mathbf{S}_R}}, \tag{4.56}$$

where $R_{0,N,\{GN,BN\};\mathbf{S}_R}$ is the cutoff rate determined by the channel SNR of transmitting the describe packet with packet length N and the receiver memory state

$\mathbf{S}_R$ as derived in Section 4.6. For example, the packet error probability of transmitting a packet with length N_2 through the good channel when the memory content is BN_1 (the first packet with length N_1 is transmitted through the bad channel and decoded incorrectly), is $P_{e,K,GN_2;BN_1} = 2^{K-(N_1+N_2)R_{0,GN_2;BN_1}}$ where $R_{0,GN_2;BN_1} = \frac{N_1 R_{0,B} + N_2 R_{0,G}}{N_1+N_2}$.

From the packet error probability formulas (4.55) and (4.56), we observe that there is an optimal N for each information packet that minimizes the average system delay. The reason is that if N is small, then the packet error probability becomes large. This increases the system delay and energy due to the high possibility of packet retransmission. On the other hand, if N is large, although the packet error probability is small, the transmission time per packet is large. Our goal is to find the best N_i to minimize the average delay. We first fix E_c/N_0 and optimize the average delay to get the corresponding packet lengths for N_i. From this, we can evaluate both average delay and average energy consumption. Repeating the above argument for different E_c/N_0, we get the energy-delay curve. All the following figures show the energy-delay curves when the bad channel SNR varies from -3 to 3 dB.

We plot SW-ARQ and SW-MARQ for different K in Fig.4.14. The system uses the following parameters, $m = 2$, $m_R = 1$, $\lambda = \mu = 0.0001$ and $r = 2$. We observe that the average delay of MARQ at low channel SNR is smaller by 20% compared to ARQ. At high channel SNR, the two systems have almost the same performance because the transmitter has a higher chance to get an ACK on the first transmission with higher channel SNR. The effect of the receiver memory improvement to the system performance goes away due to the lack of the receiver memory. Therefore, MARQ outperforms ARQ since MARQ can decode more efficiently by utilizing the previous erroneous packets. We also note that the normalized average delay and

energy of systems decreases as K increases. In [15], we showed the optimal error probability that achieves the minimum average delay is approximated as $\frac{1}{1+K\ln 2}$ when the packet error probability represented by (4.55). From this, we can express the normalized average delay with respect to the number of information bits K as

$$T_{d,avg}/K = \frac{1+\frac{\log_2(\ln K)}{K}}{R_0}, \tag{4.57}$$

for a fixed channel condition with cutoff rate R_0. From (4.57), we know that the normalized delay decreases as K increases for any channel condition (cutoff rate). Because the average performance (delay or energy) of ARQ systems over time varying channels could be treated as a mix of performance of ARQ systems over a good channel and a bad channel with some proportion to each, hence, this explains why the normalized average values decrease as K increases as indicated in the figure.

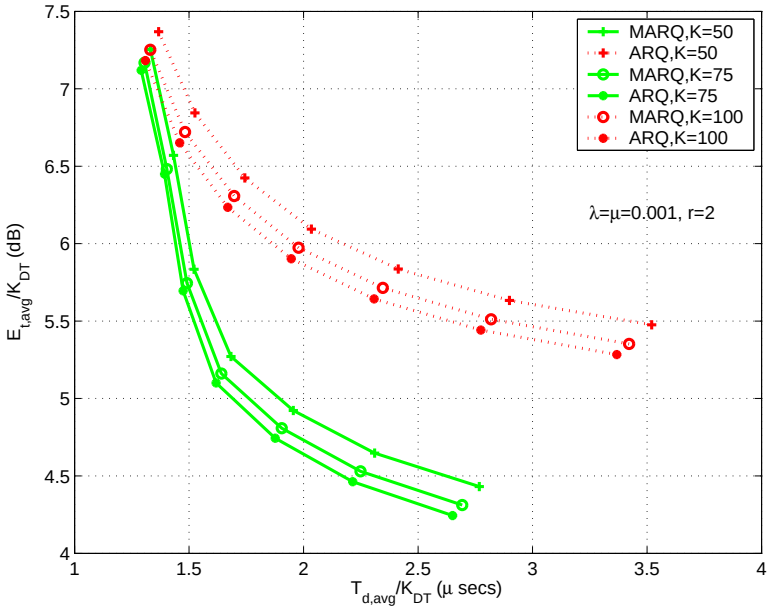

Figure 4.14: Energy-Delay curves for different K_{DT}.

We plot energy-delay curves of SW-ARQ and SW-MARQ for different λ in Fig.4.15 with a fixed average channel SNR. The systems use following parameters, $m = 2$, $m_R = 1$, $K = 100$ and $r = 2(3dB)$. We observe that the average delay and energy

decreases with the increase of λ. The reason is that the average delay and energy values are the function of λ and they decrease as λ increases as shown in Fig.4.16. The effects of λ and μ to the average energy and delay become less significant as the channel SNR increases. When the channel SNR becomes large, the systems have higher chance to transmit a packet successfully on the first attempt, therefore the average values affected by λ and μ through channel transition probabilities become less significant.

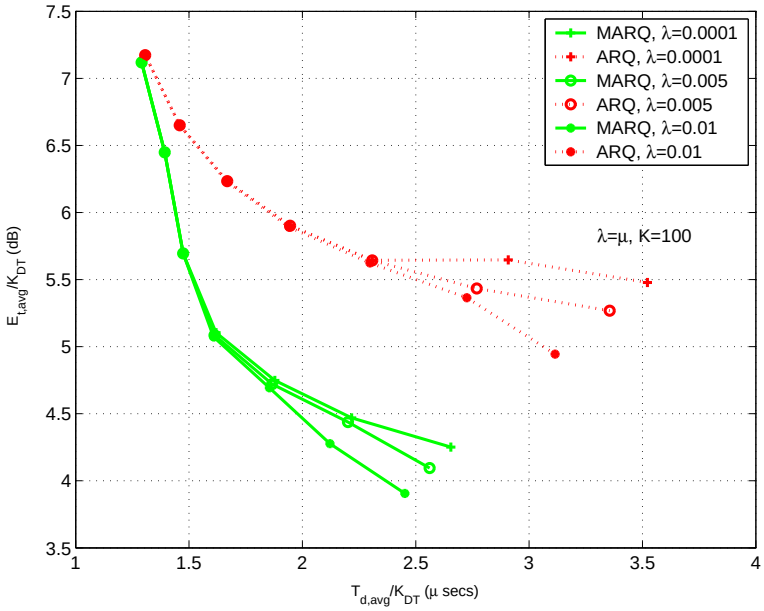

Figure 4.15: Energy-Delay curves for different λ with fixed average SNR.

Fig. 4.17 shows the energy-delay curves for different average channel SNRs. The systems use following parameters, $m = 2$, $m_R = 1$, $K = 100$ and $\lambda = 0.001$. We find that the energy-delay curves have largest values for $\lambda = 4\mu$ because the channel has a probability of $\frac{4}{5}$ to stay in the bad channel state compared to $\frac{1}{2}$ for $\lambda = \mu$ and $\frac{1}{3}$ for $\mu = 2\lambda$. We also observe that MARQ systems are more sensitive to changes of parameters (λ and μ) of time varying channels because the MARQ systems can improve the system performance significantly compared to ARQ systems due to

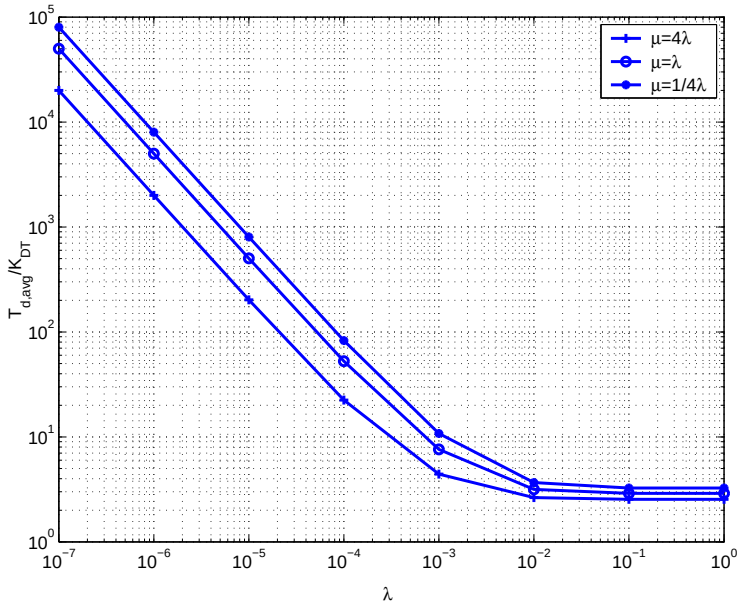

Figure 4.16: Normalized average delay relations with λ.

better error correction capability. Therefore, MARQ has better performance but is more sensitive than ARQ.

The comparison of three different repetition ARQ schemes is given in Fig. 4.18. The systems uses the following parameters, $m = 2$, $m_R = 1$, $K = 100$ and $\lambda = 0.0001$. From Fig. 4.18, we see that the system has worst performance when it does not utilize the previous error packet information (memoryless at the receiver). For the receiver with memory, the R-ARQ-SWM has better performance compared to R-ARQ-NOW because R-ARQ-NOW refreshes the memory buffer with the repetition period of packet lengths even the previous packet is erroneous received. Accordingly, it cannot utilize all the information of previous erroneous packets.

The effect of different N_{rt} for GBN-ARQ is shown in Fig. 4.18. The systems use following parameters, $m = 2$, $m_R = 1$, $K = 100$ and $\lambda = 0.001$. It shows that the average delay and energy consumed by the systems are proportional to the quantity N_{rt} since this is a go-back-N_{rt} protocol. Therefore, we need to retransmit all the

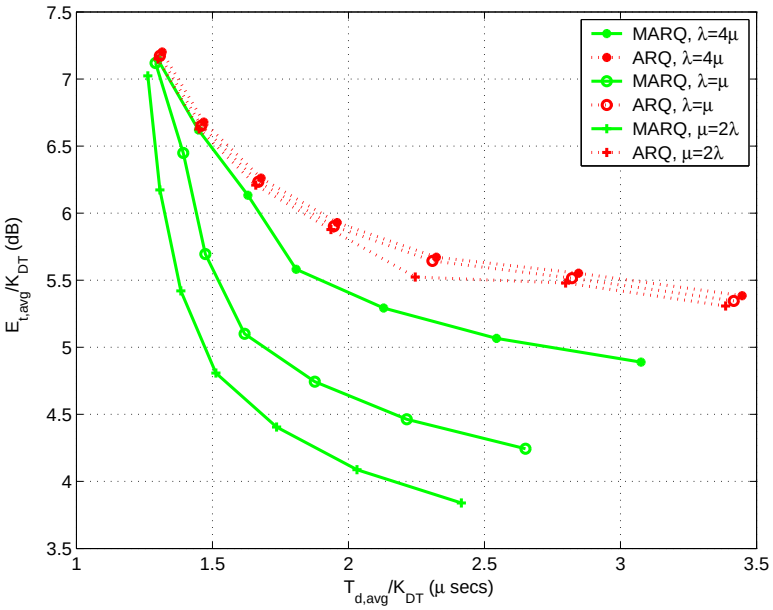

Figure 4.17: Energy-Delay curves for different average channel SNRs.

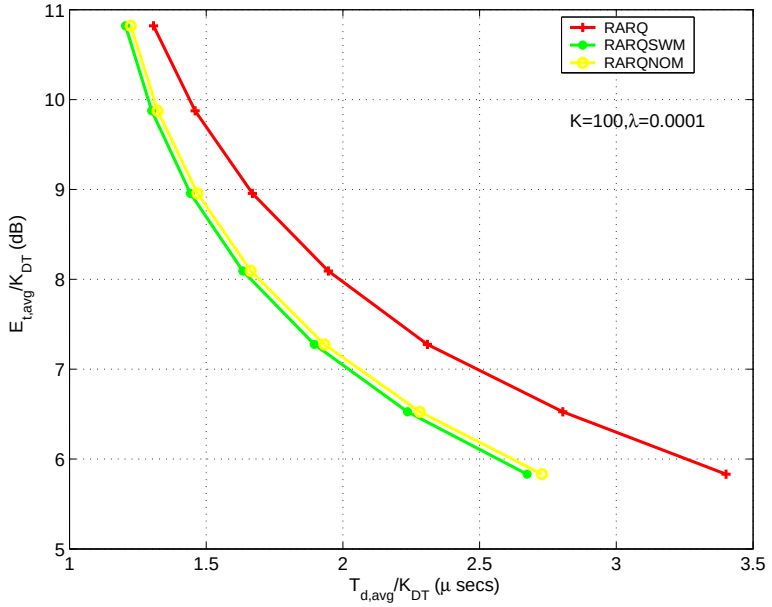

Figure 4.18: Energy-Delay curves for repetition ARQ.

following N_{rt} packets when the transmitter receives a NACK packet. In other words, the system needs to spend N_{rt} times in delay and energy for each packet transmission.

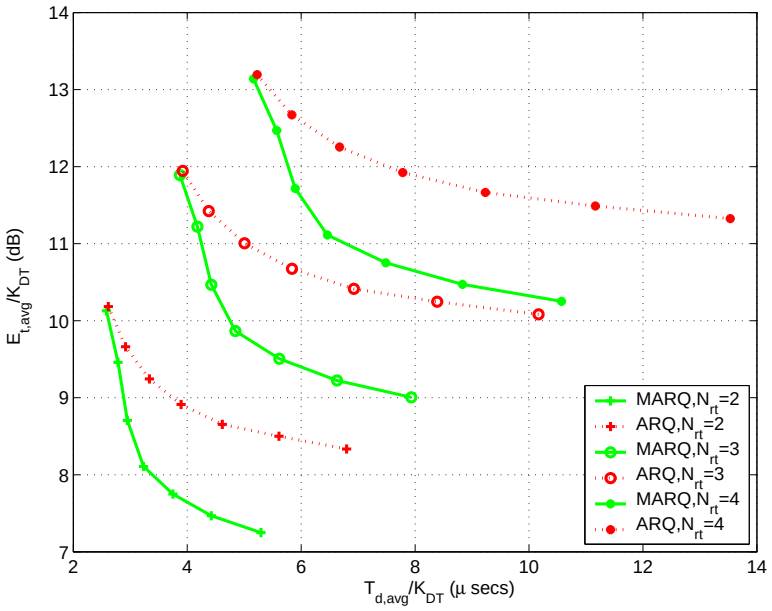

Figure 4.19: Energy-Delay curves for GBNARQ with different N_{rt}.

4.8 Conclusion

In this chapter we modeled an ARQ system with a FSM to obtain the joint distribution of energy and delay for packet transmission over time varying channels. According to the state diagram constructed from the FSM, we derived the joint generating function of energy and delay. The generating function was used to obtain the energy and delay tradeoff curves of an ARQ system. This comprehensive approach allows us to analyze the overall system by taking into account the states of the transmitter, receiver and channel. Finally, the numerical results demonstrated the time-varying characteristics of the channel have a great influence on the system performance.

CHAPTER V

Conclusion and Future Research

In this chapter, we conclude this dissertation by summarizing the content of the thesis and proposing possible future research directions.

5.1 Contributions

The most important contribution of this thesis is to represent the operation of a system by a state diagram and use the generating function approach to derive the statistics of the energy and delay consumption. We investigated the energy and delay tradeoff of a wireless communication system. In the first part (Chapter 2 and 3) of this thesis, we analyzed the energy and delay for a communication network. In the second part (Chapter 4) of this thesis, we studied an ARQ communication system.

In Chapter 2, we began by extending the approach proposed in [22] to determine average energy and delay of a successful packet transmission by taking into consideration the dependence of packet error probability on energy and delay. Thereupon, we proposed a state diagram representation of the operation of the MAC layer to obtain the joint generating function of the energy and delay by considering the effects of the PHY and MAC layer jointly. Furthermore, channel coding was incorporated in the analysis by estimating the packet error probability using error-exponent-based bounds under an AWGN channel. Using these analysis tools, the code rate and the

signal-to-noise ratio per dimension were optimized to achieve minimum average delay per information bit. By changing the signal-to-noise ratio, the energy-delay tradeoff curves for minimum delay were obtained. Finally, an approximation method to express the energy-delay tradeoff curves analytically was provided. The comparison of the energy-delay tradeoff curves evaluated from this approximation method with the exact energy-delay tradeoff curves (from numerical optimization) demonstrated that this approximation method is extremely accurate especially when the number of information bits per packet is large.

In Chapter 3, we applied our proposed generating analysis method to analyze and design other wireless network protocols. We first designed an energy adaptation scheme with original IEEE 802.11 protocol in which the transmitter increases the energy level per coded symbol whenever it suffers an unsuccessful transmission. The numerical results showed that the proposed protocol can improve the energy and delay throughput of the system significantly when the channel condition is bad. By using Reed-Solomon codes we optimized the system performance over the code rate and energy per coded bit. Although we used Reed-Solomon codes over AWGN channel to represent packet error probabilities, the framework of our analysis can be used for other coding and modulation schemes over various wireless channels. The second proposed protocol is an ARQ scheme for data packets transmission with 802.11 protocol. The purpose of this analysis is to demonstrate that the generating function approach can be applied to analyze more protocol layers jointly by including the analysis of the logical link control (LLC) sublayer into the original protocol (MAC and PHY layers only). The numerical results showed that the IEEE 802.11 original protocol and the proposed one have almost identical performances and are equally sensitive to the knowledge of the channel quality, say channel SNR, at the transmit-

ter. We also observed that the reduction in delay of a SWARQ-FT system achieved by immediately retransmitting an erroneously received packet, is compensated by the increased delay needed to access the channel, since other users use the same protocol. While ARQ provides some robustness to SNR uncertainty in a conventional single user system, this is not the case in this multiuser scenario.

In Chapter 4, we modeled an ARQ system with a FSM to obtain the joint distribution of energy and delay for successful packet transmission over a time varying channels. According to the state diagram constructed from FSM, we derived the joint generating function of energy and delay. The generating function was used to obtain the energy and delay tradeoff curves of an ARQ system. This comprehensive approach allows us to analyze the overall system taking into account the states of transmitter, receiver and channel. Finally, the numerical results demonstrated the time-varying characteristics of the channel have a great influence on the system performance. The cutoff rate for two different memory receiver structures was derived and the optimization of the cutoff rate under average energy constraint was discussed.

5.2 Future Research

In this thesis, packet error probabilities are represented by random coding bounds. One extension is to consider the packet error probability for specific coding schemes, such as LDPC or turbo codes. The other research direction that can be explored is to relax the assumption of the saturation traffic by analyzing the non-saturation traffic. For a different traffic model, we can further consider the delay spent in a buffer. By regulating the desired quality of services (QoS), we also can investigate the best energy allocation strategy for transmitting stations.

Although some research in [22, 39, 14, 15] considered the effects of packet errors in their analysis, they all assumed that all stations in a wireless network share an identical channel condition. However, this assumption is not realistic in practice, the wireless communication-links between any pair of source and destination stations may differ from each other due to the different physical objects in between. Therefore, one possible future research is to relax the restriction of the homogeneous channel condition for all the contending stations by considering the channel inhomogeneity incurred among different users. We believe that the proposed generating function method described in this thesis could be applied to other MAC protocols with more complicated channel conditions.

The energy considered in the thesis is just transmission energy. We also can consider other energy consumption, such as hardware processing energy, antenna switching energy between transmitting and receiving modes. In commercial wireless networks, mobility is an important issue, however, we do not consider mobility issue in this thesis. Future research is needed to quantify how the energy and delay consumption changes when stations are mobile.

For ARQ systems, we only considered error free case for the ACK and NACK packets in this thesis. One future research direction is to analyze an ARQ system by taking into consideration the nonzero ACK and NACK packet error probabilities and the energy and delay consumption for ACK and NACK packets. The spatial dimension provided by multiple antennas is known to dramatically increase system capacity. As a result, combining multiple-input multiple-out (MIMO) with ARQ systems has become an important and popular research subject. In particular, the combination of MIMO spatial diversity with temporal diversity has led to widespread development of space−time coding (STC) schemes. Many approaches have been

proposed for STC, for example, space−time trellis coding, space−time bit-interleaved coded modulation (BICM), and space−time block coding. Another future research direction is to develop some new ARQ protocols for MIMO systems which adapt the signal constellation of each ARQ retransmission.

APPENDICES

APPENDIX A

Generating Function Analysis for NR-ARQ Systems

A.1 Memoryless Receiver

In this section, we will determine the joint generating function of the energy consumed and delay incurred for a NR-ARQ system without the receiver memory according to the general formulas provided in Chapter 4.3. In order to illustrate the meaning for those formulas, we consider a system with some specific parameters. We assume that the sequence of packet lengths used for each transmission is $\{N_1, N_2, N_2, \cdots\}$. For all appendices, we use $\overline{P} = 1 - P$ for an event with probability value P.

From the state transition diagram shown in Fig. 4.3, we can group all transmissions with transmission indices from as $(2 + k \cdot 1)$ for $k \in \{0, 1, 2, \cdots\}$ into a group since these transmissions use same packet length N_2 to transmission. We can see that there is only one group, g_1, in this system and $T(i_1) = T(2) = 2$.

We first derive the first part of the generating function for those successful transmissions which are not belonged to $(2 + k \cdot 1 + 1)$. Therefore, we need to evaluate the generating functions for $G_1(X, Y)$ and $G_2(X, Y)$. The generating function for the event that the packet is transmitted successfully in the first transmission is

$$G_1(X, Y) = (P_G(1 - P_{e,G_1}) + P_B(1 - P_{e,B_1}))X^{N_1 T_b} Y^{N_1 E_c}, \tag{A.1}$$

where P_G and P_B are the probabilities that the initial transmission was in good and bad states, respectively. We assume that P_G is equal to stationary probability of good channel ($\frac{\mu}{\lambda+\mu}$) and P_B is equal to stationary probability of bad channel ($\frac{\lambda}{\lambda+\mu}$) if the system has operated for a long time period already. Because $\min i_1 = 2$, we need to evaluate the generating function of G_2 by using the recursion formula shown in (4.12) and obtain

$$\begin{aligned} G_2(X,Y) &= [P_G P_{e,G_1} e^{-\lambda N_1 T_b} + P_B P_{e,B_1} \overline{e^{-\mu N_1 T_b}})]\overline{P_{e,G_2}} X^{(N_1+N_2)T_b} Y^{(N_1+N_2)E_c} + \\ & \quad [P_G P_{e,G_1} \overline{e^{-\lambda N_1 T_b}} + P_B P_{e,B_1} e^{-\mu N_1 T_b}]\overline{P_{e,B_2}} X^{(N_1+N_2)T_b} Y^{(N_1+N_2)E_c}. \end{aligned} \quad \text{(A.2)}$$

For the second part of the generating function for those successful transmissions with transmission indices form as $(2+k\cdot 1+1)$ for $k \in \{0,1,2,\cdots\}$, we need to obtain the generating functions of G_{f,G_2} and G_{f,B_2} by solving a set of linear equations. We omit the memory content parts in G_{f,G_2} and G_{f,B_2} for further notational simplicity since this is a memoryless system. From the systems of equations shown in (4.9), the generating functions of G_{f,G_2} and G_{f,B_2} will be determined if the transition probability matrix for the system states and the initial conditions are given. For transition matrix, it is

$$\begin{bmatrix} P_{e,G_2} e^{\lambda N_2 T_b} & P_{e,G_2}(1 - e^{\lambda N_2 T_b}) \\ P_{e,B_2}(1 - e^{\mu N_2 T_b}) & P_{e,B_2} e^{\mu N_2 T_b}) \end{bmatrix}, \quad \text{(A.3)}$$

and the initial conditions are

$$\begin{aligned} A_{f,G_2} &= [P_G P_{e,G_1} e^{-\lambda N_1 T_b} + P_B P_{e,B_1}(1 - e^{-\mu N_1 T_b})] P_{e,G_2} X^{(N_1+N_2)T_b} Y^{(N_1+N_2)E_c} \\ A_{f,B_2} &= [P_G P_{e,G_1}(1 - e^{-\lambda N_1 T_b}) + P_B P_{e,B_1} e^{-\mu N_1 T_b}] P_{e,B_2} X^{(N_1+N_2)T_b} Y^{(N_1+N_2)E_c}. \end{aligned} \quad \text{(A.4)}$$

Therefore, the total generating function for this system is

$$\begin{aligned} G_s &= G_1 + G_2 + [G_{f,G_2} e^{-\lambda N_2 T_b}(1 - P_{e,G_2}) + G_{f,B_2}(1 - e^{-\mu N_2 T_b})(1 - P_{e,G_2}) + \\ & G_{f,G_2}(1 - e^{-\lambda N_2 T_b})(1 - P_{e,B_2}) + G_{f,B_2} e^{-\mu N_{2T_b}}(1 - P_{e,B_2})] X^{N_2 T_b} Y^{N_2 E_c} \end{aligned} \quad \text{(A.5)}$$

A.2 Memory Receiver

In this section, we will determine the joint generating function of the energy and delay required for a NR-ARQ system with the receiver memory. We assume that the sequence of packet lengths used for each transmission is $\{N_1, N_2, N_2, \cdots\}$.

From the state transition diagram shown in Fig. 4.6, we can group all transmissions with transmission indices from as $(3 + k \cdot 1)$ for $k \in \{0, 1, 2, \cdots\}$ into a group since these transmissions use same packet length N_2 to transmission. We can see that there is only one group, g_1, in this system and $T(i_1) = T(3) = 2$.

We first derive the first part of the generating functions for those successful transmissions which are not belonged to indices $(3 + k \cdot 1)$. Since $i_1 = 3$, we need to evaluate G_1, G_2, and G_3 recursively from (4.12). They are

$$\begin{aligned} G_1(X,Y) &= (P_G(1 - P_{e,G_1}) + P_B(1 - P_{e,B_1})) X^{N_1 T_b} Y^{N_1 E_c} \\ G_2(X,Y) &= [P_G P_{e,G_1}(e^{-\lambda N_1 T_b}(1 - P_{e,G_2|G_1}) + (1 - e^{-\lambda N_1 T_b})(1 - P_{e,B_2|G_1})) + \\ & \quad P_B P_{e,B_1}((1 - e^{-\mu N_1 T_b})(1 - P_{e,G_2|B_1}) + e^{-\mu N_1 T_b}(1 - P_{e,B_2|B_1}))] \cdot \\ & \quad X^{(N_1+N_2)T_b} Y^{(N_1+N_2)E_c} \\ G_3(X,Y) &= [P_G P_{e,G_1}(e^{-\lambda N_1 T_b} P_{e,G_2|G_1} e^{-\lambda N_2 T_b} \overline{P_{e,G_2|G_2}} + e^{-\lambda N_1 T_b} P_{e,G_2|G_1} \overline{e^{-\lambda N_2 T_b}} \overline{P_{e,B_2|G_2}} + \\ & \quad \overline{e^{-\lambda N_1 T_b}} P_{e,B_2|G_1} \overline{e^{-\mu N_2 T_b}} \overline{P_{e,G_2|B_2}} + \overline{e^{-\lambda N_1 T_b}} P_{e,B_2|G_1} e^{-\mu N_2 T_b} \overline{P_{e,B_2|B_2}}) + \\ & \quad P_B P_{e,B_1}(\overline{e^{-\mu N_1 T_b}} P_{e,G_2|B_1} e^{-\lambda N_2 T_b} \overline{P_{e,G_2|G_2}} + \overline{e^{-\mu N_1 T_b}} P_{e,G_2|B_1} \overline{e^{-\lambda N_2 T_b}} \overline{P_{e,B_2|G_2}} + \\ & \quad e^{-\mu N_1 T_b} P_{e,B_2|B_1} \overline{e^{-\lambda N_2 T_b}} \overline{P_{e,G_2|B_2}} + e^{-\mu N_1 T_b} P_{e,B_2|B_1} e^{-\mu N_2 T_b} \overline{P_{e,B_2|B_2}})] \cdot \\ & \quad X^{(N_1+N_2+N_3)T_b} Y^{(N_1+N_2+N_3)E_c}, \end{aligned} \quad \text{(A.6)}$$

For the second part of the generating function for those successful transmissions with transmission indices form as $(3 + k \cdot 1 + 1)$ for $k \in \{0, 1, 2, \cdots\}$, we need to obtain the generating functions of $G_{f,G_2|G2}$, $G_{f,B_2|G2}$, $G_{f,G_2|B2}$ and $G_{f,B_2|B2}$ by solving a set of linear equations. From the systems of equations shown in (4.9), the generating functions of $G_{f,G_2|G2}$, $G_{f,B_2|G2}$, $G_{f,G_2|B2}$ and $G_{f,B_2|B2}$ will be determined if the transition probability matrix for the system states and the initial conditions are given. For transition matrix, it is

$$\begin{bmatrix} P_{e,G_2|G_2}e^{\lambda N_2 T_b} & 0 & P_{e,G_2|G_2}\overline{e^{\lambda N_2 T_b}} & 0 \\ P_{e,G_2|B_2}e^{\lambda N_2 T_b} & 0 & P_{e,G_2|B_2}\overline{e^{\lambda N_2 T_b}} & 0 \\ 0 & P_{e,B_2|G_2}\overline{e^{\mu N_2 T_b}} & 0 & P_{e,B_2|G_2}e^{\mu N_2 T_b} \\ 0 & P_{e,B_2|B_2}\overline{e^{\lambda N_2 T_b}} & 0 & P_{e,B_2|B_2}e^{\lambda N_2 T_b} \end{bmatrix}, \tag{A.7}$$

and the initial conditions are

$$\begin{aligned} A_{f,G_2|G_2} &= (P_G P_{e,G_1} e^{-\lambda N_1 T_b} P_{e,G_2|G_1} + P_B P_{e,B_1} \overline{e^{-\mu N_1 T_b}} P_{e,G_2|B_1}) e^{-\lambda N_2 T_b} P_{e,G_2|G_2} \cdot \\ &\quad X^{(N_1+2N_2)T_b} Y^{(N_1+2N_2)E_c} \\ A_{f,G_2|B_2} &= (P_G P_{e,G_1} \overline{e^{-\lambda N_1 T_b}} P_{e,B_2|G_1} + P_B P_{e,B_1} e^{-\mu N_1 T_b} P_{e,B_2|B_1}) \overline{e^{-\mu N_2 T_b}} P_{e,G_2|B_2} \cdot \\ &\quad X^{(N_1+2N_2)T_b} Y^{(N_1+2N_2)E_c} \\ A_{f,B_2|G_2} &= (P_G P_{e,G_1} e^{-\lambda N_1 T_b} P_{e,G_2|G_1} + P_B P_{e,B_1} \overline{e^{-\mu N_1 T_b}} P_{e,G_2|B_1}) \overline{e^{-\lambda N_2 T_b}} P_{e,B_2|G_2} \cdot \\ &\quad X^{(N_1+2N_2)T_b} Y^{(N_1+2N_2)E_c} \\ A_{f,B_2|B_2} &= (P_G P_{e,G_1} \overline{e^{-\lambda N_1 T_b}} P_{e,B_2|G_1} + P_B P_{e,B_1} e^{-\mu N_1 T_b} P_{e,B_2|B_1}) e^{-\mu N_2 T_b} P_{e,B_2|B_2} \cdot \\ &\quad X^{(N_1+2N_2)T_b} Y^{(N_1+2N_2)E_c} \end{aligned} \tag{A.8}$$

Therefore, the total generating function for this system is

$$\begin{aligned} G_s &= G_1 + G_2 + G_3 + [G_{f,G_2|G_2} e^{-\lambda N_2 T_b}(1 - P_{e,G_2|G_2}) + G_{f,G_2|G_2}\overline{e^{-\lambda N_2 T_b}}(1 - P_{e,B_2|G_2}) + \\ & G_{f,G_2|B_2} e^{-\lambda N_2 T_b}(1 - P_{e,G_2|G_2}) + G_{f,G_2|B_2}\overline{e^{-\lambda N_2 T_b}}(1 - P_{e,B_2|G_2}) + \\ & G_{f,B_2|G_2} e^{-\mu N_2 T_b}(1 - P_{e,B_2|B_2}) + G_{f,B_2|B_2}\overline{e^{-\mu N_2 T_b}}(1 - P_{e,G_2|B_2}) + \\ & G_{f,B_2|B_2} e^{-\mu N_2 T_b}(1 - P_{e,B_2|B_2}) + G_{f,B_2|B_2}\overline{e^{-\mu N_2 T_b}}(1 - P_{e,G_2|B_2})] X^{N_2 T_b} Y^{N_2 E_c}. \quad \text{(A.9)} \end{aligned}$$

APPENDIX B

Generating Function Analysis for R-ARQ Systems

B.1 Memoryless Receiver

In this section, we will determine the joint generating function of the energy consumed and delay incurred for a R-ARQ system without the receiver memory according to the general formulas provided in Chapter 4.3. We consider a system with some specific parameters by assuming that the sequence of packet lengths used for each transmission is $\{N_1, N_2, N_1, N_2, N_1, N_2, \cdots\}$.

From the state transition diagram shown in Fig. 4.8, we can group all transmissions with transmission indices from as $(1 + k \cdot 2)$ for $k \in \{0, 1, 2, \cdots\}$ into a group g_1 and $(2 + k \cdot 1)$ for $k \in \{0, 1, 2, \cdots\}$ into another group g_2. There are two groups, g_1, g_2, in this system and $T(i_1) = T(1) = 1, T(i_2) = T(2) = 2$.

We first derive the first part of the generating function. Since $i_1 = 1, i_2 = 2$, the only term in the first part of the generating function is $G_1(X, Y)$. The generating function for the event that the packet is transmitted successfully in the first transmission is

$$G_1(X, Y) = (P_G(1 - P_{e,G_1}) + P_B(1 - P_{e,B_1}))X^{N_1 T_b} Y^{N_1 E_c}, \tag{B.1}$$

where P_G and P_B are the probabilities that the initial transmission was in good and bad states, respectively. We assume that P_G is equal to stationary probability of

good channel ($\frac{\mu}{\lambda+\mu}$) and P_B is equal to stationary probability of bad channel ($\frac{\lambda}{\lambda+\mu}$) if the system has operated for a long time period already.

For the second part of the generating function for those successful transmissions with transmission indices form as $(1 + k \cdot 2 + 1)$ for $k \in \{0, 1, 2, \cdots\}$, we need to obtain the generating functions of G_{f,G_1} and G_{f,B_1} by solving a set of linear equations. Similarly, for those successful transmissions with transmission indices form as $(2+k\cdot 2+1)$ for $k \in \{0, 1, 2, \cdots\}$, we need to obtain the generating functions of G_{f,G_2} and G_{f,B_2} by solving a set of linear equations.

From the systems of equations shown in (4.9), the generating functions of G_{f,G_1} and G_{f,B_1} will be determined if the transition probability matrix, M_{g_1}, for the system states and the initial conditions are given. For transition matrix, it is

$$M_{g_1} = \begin{bmatrix} c_{11}^{g_1} & c_{12}^{g_1} \\ c_{21}^{g_1} & c_{22}^{g_1} \end{bmatrix}, \tag{B.2}$$

where

$$\begin{aligned} c_{11}^{g_1} &= P_{e,G_1}P_{e,G_2}e^{-\lambda N_1 T_b}e^{-\lambda N_2 T_b} + P_{e,G_1}P_{e,B_2}\overline{e^{-\lambda N_1 T_b}e^{-\mu N_2 T_b}} \\ c_{12}^{g_1} &= P_{e,G_1}P_{e,G_2}e^{-\lambda N_1 T_b}\overline{e^{-\lambda N_2 T_b}} + P_{e,G_1}P_{e,B_2}\overline{e^{-\lambda N_1 T_b}}e^{-\mu N_2 T_b} \\ c_{21}^{g_1} &= P_{e,B_1}P_{e,G_2}\overline{e^{-\mu N_1 T_b}}e^{-\lambda N_2 T_b} + P_{e,B_1}P_{e,B_2}e^{-\mu N_1 T_b}\overline{e^{-\mu N_2 T_b}} \\ c_{22}^{g_1} &= P_{e,B_1}P_{e,G_2}\overline{e^{-\mu N_1 T_b}e^{-\lambda N_2 T_b}} + P_{e,B_1}P_{e,B_2}e^{-\mu N_1 T_b}e^{-\mu N_2 T_b}. \end{aligned} \tag{B.3}$$

And the initial conditions are

$$\begin{aligned} A_{f,G_1} &= P_G P_{e,G_1} X^{N_1 T_b} Y^{N_1 E_c} \\ A_{f,B_1} &= P_B P_{e,B_1} X^{N_1 T_b} Y^{N_1 E_c}. \end{aligned} \tag{B.4}$$

Similarly, the generating functions of G_{f,G_2} and G_{f,B_2} will be determined if the transition probability matrix, M_{g_2}, for the system states and the initial conditions

are given. For transition matrix, it is

$$M_{g_2} = \begin{bmatrix} c_{11}^{g_2} & c_{12}^{g_2} \\ c_{21}^{g_2} & c_{22}^{g_2} \end{bmatrix}, \tag{B.5}$$

where

$$\begin{aligned} c_{11}^{g_2} &= P_{e,G_2}P_{e,G_1}e^{-\lambda N_2 T_b}e^{-\lambda N_1 T_b} + P_{e,G_2}P_{e,B_1}\overline{e^{-\lambda N_2 T_b}e^{-\mu N_1 T_b}} \\ c_{12}^{g_2} &= P_{e,G_2}P_{e,G_1}e^{-\lambda N_2 T_b}\overline{e^{-\lambda N_1 T_b}} + P_{e,G_2}P_{e,B_1}\overline{e^{-\lambda N_2 T_b}}e^{-\mu N_1 T_b} \\ c_{21}^{g_2} &= P_{e,B_2}P_{e,G_1}\overline{e^{-\mu N_2 T_b}}e^{-\lambda N_1 T_b} + P_{e,B_2}P_{e,B_1}e^{-\mu N_2 T_b}\overline{e^{-\mu N_1 T_b}} \\ c_{22}^{g_2} &= P_{e,B_2}P_{e,G_1}\overline{e^{-\mu N_2 T_b}e^{-\lambda N_1 T_b}} + P_{e,B_2}P_{e,B_1}e^{-\mu N_2 T_b}e^{-\mu N_1 T_b}. \end{aligned} \tag{B.6}$$

And the initial conditions are

$$\begin{aligned} A_{f,G_2} &= [P_G P_{e,G_1}e^{-\lambda N_1 T_b} + P_B P_{e,B_1}(1-e^{-\mu N_1 T_b})]P_{e,G_2}X^{(N_1+N_2)T_b}Y^{(N_1+N_2)E_c} \\ A_{f,B_2} &= [P_G P_{e,G_1}(1-e^{-\lambda N_1 T_b}) + P_B P_{e,B_1}e^{-\mu N_1 T_b}]P_{e,B_2}X^{(N_1+N_2)T_b}Y^{(N_1+N_2)E_c}. \end{aligned} \tag{B.7}$$

Therefore, the total generating function for this system is

$$\begin{aligned} G_s = {} & G_1 + [G_{f,G_1}(e^{-\lambda N_1 T_b}\overline{P_{e,G_2}} + \overline{e^{-\lambda N_1 T_b}}\,\overline{P_{e,B_2}}) \\ & G_{f,B_1}(\overline{e^{-\mu N_1 T_b}}\,\overline{P_{e,G_2}} + e^{-\mu N_1 T_b}\overline{P_{e,B_2}})]X^{N_2 T_b}Y^{N_2 E_c} + \\ & [G_{f,G_2}(e^{-\lambda N_2 T_b}\overline{P_{e,G_1}} + \overline{e^{-\lambda N_2 T_b}}\,\overline{P_{e,B_1}}) \\ & G_{f,B_2}(\overline{e^{-\mu N_2 T_b}}\,\overline{P_{e,G_1}} + e^{-\mu N_2 T_b}\overline{P_{e,B_1}})]X^{N_1 T_b}Y^{N_1 E_c} \end{aligned} \tag{B.8}$$

B.2 Memory Receiver : Sliding Window

In this section, we will determine the generating function for a R-ARQ system with the receiver memory which updates its memory contents by the sliding window operation. We assume that the sequence of packet lengths used for each transmission is $\{N_1, N_2, N_1, N_2, N_1, N_2, \cdots\}$.

From the state transition diagram shown in Fig. 4.9, we can group all transmissions with transmission indices from as $(2 + k \cdot 2)$ for $k \in \{0, 1, 2, \cdots\}$ into a group g_1 and $(3 + k \cdot 2)$ for $k \in \{0, 1, 2, \cdots\}$ into another group g_2. There are two groups, g_1, g_2, in this system and $T(i_1) = T(2) = 2, T(i_2) = T(3) = 1$.

We first derive the first part of the generating function. Since $i_1 = 2, i_2 = 3$, the terms in the first part of the generating function are G_1, G_2. They are

$$\begin{aligned} G_1(X,Y) &= (P_G(1 - P_{e,G_1}) + P_B(1 - P_{e,B_1}))X^{N_1 T_b} Y^{N_1 E_c} \\ G_2(X,Y) &= [P_G P_{e,G_1}(e^{-\lambda N_1 T_b}\overline{P_{e,G_2|G_1}} + \overline{e^{-\lambda N_1 T_b}}\,\overline{P_{e,B_2|G_1}}) + \\ &\quad P_B P_{e,B_1}(e^{-\mu N_1 T_b}\overline{P_{e,B_2|B_1}} + \overline{e^{-\mu N_1 T_b}}\,\overline{P_{e,G_2|B_1}})]X^{(N_1+N_2)T_b} Y^{(N_1+N_2)E_c}, \end{aligned} \tag{B.9}$$

where P_G and P_B are the probabilities that the initial transmission was in good and bad states, respectively. We assume that P_G is equal to stationary probability of good channel $(\frac{\mu}{\lambda+\mu})$ and P_B is equal to stationary probability of bad channel $(\frac{\lambda}{\lambda+\mu})$ if the system has operated for a long time period already.

For the second part of the generating function for those successful transmissions with transmission indices form as $(2 + k \cdot 2 + 1)$ for $k \in \{0, 1, 2, \cdots\}$, we need to obtain the generating functions of $G_{f,G_2|G_1}$, $G_{f,G_2|B_1}$, $G_{f,B_2|G_1}$, and $G_{f,B_2|B_1}$ by solving a set of linear equations. Similarly, for those successful transmissions with transmission indices form as $(3 + k \cdot 2 + 1)$ for $k \in \{0, 1, 2, \cdots\}$, we need to obtain the generating functions of $G_{f,G_1|G_2}$, $G_{f,G_1|B_2}$, $G_{f,B_1|G_2}$, and $G_{f,B_1|B_2}$ by solving a set of linear equations.

From the systems of equations shown in (4.9), the generating functions of $G_{f,G_2|G_1}$, $G_{f,G_2|B_1}$, $G_{f,B_2|G_1}$, and $G_{f,B_2|B_1}$ will be determined if the transition probability matrix, M_{g_1}, for the system states and the initial conditions are given. For transition matrix,

it is

$$M_{g_1} = \begin{bmatrix} c_{11}^{g_1} & c_{12}^{g_1} & c_{13}^{g_1} & c_{14}^{g_1} \\ c_{21}^{g_1} & c_{22}^{g_1} & c_{23}^{g_1} & c_{24}^{g_1} \\ c_{31}^{g_1} & c_{32}^{g_1} & c_{33}^{g_1} & c_{34}^{g_1} \\ c_{41}^{g_1} & c_{42}^{g_1} & c_{43}^{g_1} & c_{44}^{g_1} \end{bmatrix}, \tag{B.10}$$

where

$$\begin{aligned}
c_{11}^{g_1} &= P_{e,G_2|G_1} P_{e,G_1|G_2} e^{-\lambda N_2 T_b} e^{-\lambda N_1 T_b}, \quad c_{12}^{g_1} = P_{e,G_2|G_1} P_{e,B_1|G_2} \overline{e^{-\lambda N_2 T_b} e^{-\mu N_1 T_b}} \\
c_{13}^{g_1} &= P_{e,G_2|G_1} P_{e,G_1|G_2} e^{-\lambda N_2 T_b} \overline{e^{-\lambda N_1 T_b}}, \quad c_{14}^{g_1} = P_{e,G_2|G_1} P_{e,B_1|G_2} \overline{e^{-\lambda N_2 T_b}} e^{-\mu N_1 T_b} \\
c_{21}^{g_1} &= P_{e,G_2|B_1} P_{e,G_1|G_2} e^{-\lambda N_2 T_b} e^{-\lambda N_1 T_b}, \quad c_{22}^{g_1} = P_{e,G_2|B_1} P_{e,B_1|G_2} \overline{e^{-\lambda N_2 T_b} e^{-\mu N_1 T_b}} \\
c_{23}^{g_1} &= P_{e,G_2|B_1} P_{e,G_1|G_2} e^{-\lambda N_2 T_b} \overline{e^{-\lambda N_1 T_b}}, \quad c_{24}^{g_1} = P_{e,G_2|B_1} P_{e,B_1|G_2} \overline{e^{-\lambda N_2 T_b}} e^{-\mu N_1 T_b} \\
c_{31}^{g_1} &= P_{e,B_2|G_1} P_{e,G_1|B_2} \overline{e^{-\mu N_2 T_b}} e^{-\lambda N_1 T_b}, \quad c_{32}^{g_1} = P_{e,B_2|G_1} P_{e,B_1|B_2} e^{-\mu N_2 T_b} \overline{e^{-\mu N_1 T_b}} \\
c_{33}^{g_1} &= P_{e,B_2|G_1} P_{e,G_1|B_2} \overline{e^{-\mu N_2 T_b} e^{-\lambda N_1 T_b}}, \quad c_{34}^{g_1} = P_{e,B_2|G_1} P_{e,B_1|B_2} e^{-\mu N_2 T_b} e^{-\mu N_1 T_b} \\
c_{41}^{g_1} &= P_{e,B_2|B_1} P_{e,G_1|B_2} \overline{e^{-\mu N_2 T_b}} e^{-\lambda N_1 T_b}, \quad c_{42}^{g_1} = P_{e,B_2|B_1} P_{e,B_1|B_2} e^{-\mu N_2 T_b} \overline{e^{-\mu N_1 T_b}} \\
c_{43}^{g_1} &= P_{e,B_2|B_1} P_{e,G_1|B_2} \overline{e^{-\mu N_2 T_b} e^{-\lambda N_1 T_b}}, \quad c_{44}^{g_1} = P_{e,B_2|B_1} P_{e,B_1|B_2} e^{-\mu N_2 T_b} e^{-\mu N_1 T_b}
\end{aligned}$$

. (B.11)

And the initial conditions are

$$\begin{aligned}
A_{f,G_2|G_1} &= P_G P_{e,G_1} P_{e,G_2|G_1} e^{-\lambda N_1 T_b} X^{(N_1+N_2)T_b} Y^{(N_1+N_2)E_c} \\
A_{f,G_2|B_1} &= P_B P_{e,B_1} P_{e,G_2|B_1} \overline{e^{-\mu N_1 T_b}} X^{(N_1+N_2)T_b} Y^{(N_1+N_2)E_c} \\
A_{f,B_2|G_1} &= P_G P_{e,G_1} P_{e,B_2|G_1} \overline{e^{-\lambda N_1 T_b}} X^{(N_1+N_2)T_b} Y^{(N_1+N_2)E_c} \\
A_{f,B_2|B_1} &= P_B P_{e,B_1} P_{e,B_2|B_1} e^{-\mu N_1 T_b} X^{(N_1+N_2)T_b} Y^{(N_1+N_2)E_c}.
\end{aligned} \tag{B.12}$$

Similarly, for those successful transmissions with transmission indices form as $(3 + k \cdot 2 + 1)$ for $k \in \{0, 1, 2, \cdots\}$, we need to obtain the generating functions of $G_{f,G_1|G_2}$, $G_{f,G_1|B_2}$, $G_{f,B_1|G_2}$, and $G_{f,B_1|B_2}$ by solving a set of linear equations.

From the systems of equations shown in (4.9), the generating functions of $G_{f,G_1|G_2}$, $G_{f,G_1|B_2}$, $G_{f,B_1|G_2}$, and $G_{f,B_1|B_2}$ will be determined if the transition probability matrix, M_{g_2}, for the system states and the initial conditions are given. For transition matrix, it is

$$M_{g_2} = \begin{bmatrix} c_{11}^{g_2} & c_{12}^{g_2} & c_{13}^{g_2} & c_{14}^{g_2} \\ c_{21}^{g_2} & c_{22}^{g_2} & c_{23}^{g_2} & c_{24}^{g_2} \\ c_{31}^{g_2} & c_{32}^{g_2} & c_{33}^{g_2} & c_{34}^{g_2} \\ c_{41}^{g_2} & c_{42}^{g_2} & c_{43}^{g_2} & c_{44}^{g_2} \end{bmatrix}, \tag{B.13}$$

where

$$\begin{aligned}
c_{11}^{g_2} &= P_{e,G_1|G_2} P_{e,G_2|G_1} e^{-\lambda N_1 T_b} e^{-\lambda N_2 T_b}, \quad c_{12}^{g_2} = P_{e,G_1|G_2} P_{e,B_2|G_1} \overline{e^{-\lambda N_1 T_b} e^{-\mu N_2 T_b}} \\
c_{13}^{g_2} &= P_{e,G_1|G_2} P_{e,G_2|G_1} e^{-\lambda N_1 T_b} \overline{e^{-\lambda N_2 T_b}}, \quad c_{14}^{g_2} = P_{e,G_1|G_2} P_{e,B_2|G_1} \overline{e^{-\lambda N_1 T_b}} e^{-\mu N_2 T_b} \\
c_{21}^{g_2} &= P_{e,G_1|B_2} P_{e,G_2|G_1} e^{-\lambda N_1 T_b} e^{-\lambda N_2 T_b}, \quad c_{22}^{g_2} = P_{e,G_1|B_2} P_{e,B_2|G_1} \overline{e^{-\lambda N_1 T_b} e^{-\mu N_2 T_b}} \\
c_{23}^{g_2} &= P_{e,G_1|B_2} P_{e,G_2|G_1} e^{-\lambda N_1 T_b} \overline{e^{-\lambda N_2 T_b}}, \quad c_{24}^{g_2} = P_{e,G_1|B_2} P_{e,B_2|G_1} \overline{e^{-\lambda N_1 T_b}} e^{-\mu N_2 T_b} \\
c_{31}^{g_2} &= P_{e,B_1|G_2} P_{e,G_2|B_1} \overline{e^{-\mu N_1 T_b}} e^{-\lambda N_2 T_b}, \quad c_{32}^{g_2} = P_{e,B_1|G_2} P_{e,B_2|B_1} e^{-\mu N_1 T_b} \overline{e^{-\mu N_2 T_b}} \\
c_{33}^{g_2} &= P_{e,B_1|G_2} P_{e,G_2|B_1} \overline{e^{-\mu N_1 T_b} e^{-\lambda N_2 T_b}}, \quad c_{34}^{g_2} = P_{e,B_1|G_2} P_{e,B_2|B_1} e^{-\mu N_1 T_b} e^{-\mu N_2 T_b} \\
c_{41}^{g_2} &= P_{e,B_1|B_2} P_{e,G_2|B_1} \overline{e^{-\mu N_1 T_b}} e^{-\lambda N_2 T_b}, \quad c_{42}^{g_2} = P_{e,B_1|B_2} P_{e,B_2|B_1} e^{-\mu N_1 T_b} \overline{e^{-\mu N_2 T_b}} \\
c_{43}^{g_2} &= P_{e,B_1|B_2} P_{e,G_2|B_1} \overline{e^{-\mu N_1 T_b} e^{-\lambda N_2 T_b}}, \quad c_{44}^{g_2} = P_{e,B_1|B_2} P_{e,B_2|B_1} e^{-\mu N_1 T_b} e^{-\mu N_2 T_b}.
\end{aligned} \tag{B.14}$$

And the initial conditions are

$$
\begin{aligned}
A_{f,G_1|G_2} &= (P_G P_{e,G_1} P_{e,G_2|G_1} e^{-\lambda N_1 T_b} e^{-\lambda N_2 T_b} + P_B P_{e,B_1} P_{e,G_2|B_1} \overline{e^{-\mu N_1 T_b}} e^{-\lambda N_2 T_b}) P_{e,G_1|G_2} \cdot \\
&\quad X^{(2N_1+N_2)T_b} Y^{(2N_1+N_2)E_c} \\
A_{f,G_1|B_2} &= (P_G P_{e,G_1} P_{e,B_2|G_1} \overline{e^{-\lambda N_1 T_b} e^{-\mu N_2 T_b}} + P_B P_{e,B_1} P_{e,B_2|B_1} e^{-\mu N_1 T_b} \overline{e^{-\mu N_2 T_b}}) P_{e,G_1|B_2} \cdot \\
&\quad X^{(2N_1+N_2)T_b} Y^{(2N_1+N_2)E_c} \\
A_{f,B_1|G_2} &= (P_G P_{e,G_1} P_{e,G_2|G_1} e^{-\lambda N_1 T_b} \overline{e^{-\mu N_2 T_b}} + P_B P_{e,B_1} P_{e,G_2|B_1} \overline{e^{-\mu N_1 T_b} e^{-\lambda N_2 T_b}}) P_{e,B_1|G_2} \cdot \\
&\quad X^{(2N_1+N_2)T_b} Y^{(2N_1+N_2)E_c} \\
A_{f,B_1|B_2} &= (P_G P_{e,G_1} P_{e,B_2|G_1} \overline{e^{-\lambda N_1 T_b}} e^{-\mu N_2 T_b} + P_B P_{e,B_1} P_{e,B_2|B_1} e^{-\mu N_1 T_b} e^{-\mu N_2 T_b}) P_{e,B_1|B_2} \cdot \\
&\quad X^{(2N_1+N_2)T_b} Y^{(2N_1+N_2)E_c}. \qquad \text{(B.15)}
\end{aligned}
$$

Therefore, the total generating function for this system is

$$
\begin{aligned}
G_s &= G_1(X,Y) + G_2(X,Y) + [(G_{f,G_2|G_1} + G_{f,G_2|B_1})(e^{-\lambda N_2 T_b} \overline{P_{e,G_1|G_2}} + \overline{e^{-\lambda N_2 T_b}}\,\overline{P_{e,B_1|G_2}}) + \\
&\quad (G_{f,B_2|G_1} + G_{f,B_2|B_1})(\overline{e^{-\mu N_2 T_b}}\,\overline{P_{e,G_1|B_2}} + e^{-\mu N_2 T_b} \overline{P_{e,B_1|B_2}})] X^{N_1 T_b} Y^{N_1 E_c} + \\
&\quad [(G_{f,G_1|G_2} + G_{f,G_1|B_2})(e^{-\lambda N_1 T_b} \overline{P_{e,G_2|G_1}} + \overline{e^{-\lambda N_1 T_b}}\,\overline{P_{e,B_2|G_1}}) + \\
&\quad (G_{f,B_1|G_2} + G_{f,B_1|B_2})(\overline{e^{-\mu N_1 T_b}}\,\overline{P_{e,G_2|B_1}} + e^{-\mu N_1 T_b} \overline{P_{e,B_2|B_1}})] X^{N_2 T_b} Y^{N_2 E_c}. \qquad \text{(B.16)}
\end{aligned}
$$

B.3 Memory Receiver : Non Overlap

In this section, we will determine the generating function of energy and delay to transmit a data packet successfully for a R-ARQ system with the receiver memory. We assume that the sequence of packet lengths used for each transmission is $\{N_1, N_2, N_1, N_2, N_1, N_2, \cdots\}$.

From the state transition diagram shown in Fig. 4.10, we can group all transmissions with transmission indices from as $(1 + k \cdot 2)$ for $k \in \{0, 1, 2, \cdots\}$ into a group g_1 and $(2 + k \cdot 2)$ for $k \in \{0, 1, 2, \cdots\}$ into another group g_2. There are two groups, g_1, g_2, in this system and $T(i_1) = T(1) = 1, T(i_2) = T(2) = 2$.

We first derive the first part of the generating function. Since $i_1 = 1, i_2 = 2$, the only term in the first part of the generating function is $G_1(X, Y)$. It is

$$G_1(X,Y) = (P_G(1 - P_{e,G_1}) + P_B(1 - P_{e,B_1}))X^{N_1 T_b} Y^{N_1 E_c}. \quad \text{(B.17)}$$

where P_G and P_B are the probabilities that the initial transmission was in good and bad states, respectively. We assume that P_G is equal to stationary probability of good channel ($\frac{\mu}{\lambda+\mu}$) and P_B is equal to stationary probability of bad channel ($\frac{\lambda}{\lambda+\mu}$) if the system has operated for a long time period already.

For the second part of the generating function for those successful transmissions with transmission indices form as $(1 + k \cdot 2 + 1)$ for $k \in \{0, 1, 2, \cdots\}$, we need to obtain the generating functions of $G_{f,G1}$ and G_{f,B_1} by solving a set of linear equations. Similarly, for those successful transmissions with transmission indices form as $(2+k\cdot 2+1)$ for $k \in \{0, 1, 2, \cdots\}$, we need to obtain the generating functions of $G_{f,G_2|G_1}$, $G_{f,G_2|B_1}$, $G_{f,B_2|G_1}$, and $G_{f,B_2|B_1}$ by solving a set of linear equations.

From the systems of equations shown in (4.9), the generating functions of G_{f,G_1} and G_{f,B_1} will be determined if the transition probability matrix, M_{g_1}, for the system states and the initial conditions are given. For transition matrix, it is

$$M_{g_1} = \begin{bmatrix} c_{11}^{g_1} & c_{12}^{g_1} \\ c_{21}^{g_1} & c_{22}^{g_1} \end{bmatrix}, \quad \text{(B.18)}$$

where

$$\begin{aligned}
c_{11}^{g_1} &= P_{e,G_1} P_{e,G_2|G_1} e^{-\lambda N_1 T_b} e^{-\lambda N_2 T_b} + P_{e,G_1} P_{e,B_2|G_1} \overline{e^{-\lambda N_1 T_b} e^{-\mu N_2 T_b}} \\
c_{12}^{g_1} &= P_{e,G_1} P_{e,G_2|G_1} e^{-\lambda N_1 T_b} \overline{e^{-\lambda N_2 T_b}} + P_{e,G_1} P_{e,B_2|G_1} \overline{e^{-\lambda N_1 T_b}} e^{-\mu N_2 T_b} \\
c_{21}^{g_1} &= P_{e,B_1} P_{e,G_2|B_1} \overline{e^{-\mu N_1 T_b}} e^{-\lambda N_2 T_b} + P_{e,B_1} P_{e,B_2|B_1} e^{-\mu N_1 T_b} \overline{e^{-\mu N_2 T_b}} \\
c_{22}^{g_1} &= P_{e,B_1} P_{e,G_2|B_1} \overline{e^{-\lambda N_1 T_b} e^{-\lambda N_2 T_b}} + P_{e,B_1} P_{e,B_2|B_1} e^{-\mu N_1 T_b} e^{-\mu N_2 T_b} \quad \text{(B.19)}
\end{aligned}$$

And the initial conditions are

$$\begin{aligned} A_{f,G_1} &= P_G P_{e,G_1} X^{N_1 T_b} Y^{N_1 E_c} \\ A_{f,B_1} &= P_B P_{e,B_1} X^{N_1 T_b} Y^{N_1 E_c}. \end{aligned} \tag{B.20}$$

Similarly, for those successful transmissions with transmission indices form as $(2 + k \cdot 2 + 1)$ for $k \in \{0, 1, 2, \cdots\}$, we need to obtain the generating functions of $G_{f,G_2|G_1}$, $G_{f,G_2|B_1}$, $G_{f,B_2|G_1}$, and $G_{f,B_2|B_1}$ by solving a set of linear equations.

From the systems of equations shown in (4.9), the generating functions of $G_{f,G_2|G_1}$, $G_{f,G_2|B_1}$, $G_{f,B_2|G_1}$, and $G_{f,B_2|B_1}$ will be determined if the transition probability matrix, M_{g_2}, for the system states and the initial conditions are given. For transition matrix, it is

$$M_{g_2} = \begin{bmatrix} c_{11}^{g_2} & c_{12}^{g_2} & c_{13}^{g_2} & c_{14}^{g_2} \\ c_{21}^{g_2} & c_{22}^{g_2} & c_{23}^{g_2} & c_{24}^{g_2} \\ c_{31}^{g_2} & c_{32}^{g_2} & c_{33}^{g_2} & c_{34}^{g_2} \\ c_{41}^{g_2} & c_{42}^{g_2} & c_{43}^{g_2} & c_{44}^{g_2} \end{bmatrix}, \tag{B.21}$$

where

$$\begin{aligned} c_{11}^{g_2} &= P_{e,G_2|G_1} P_{e,G_1} e^{-\lambda N_2 T_b} e^{-\lambda N_1 T_b}, \quad c_{12}^{g_2} = P_{e,G_2|G_1} P_{e,B_1} \overline{e^{-\lambda N_2 T_b} e^{-\mu N_2 T_b}} \\ c_{13}^{g_2} &= P_{e,G_2|G_1} P_{e,G_1} e^{-\lambda N_2 T_b} \overline{e^{-\lambda N_1 T_b}}, \quad c_{14}^{g_2} = P_{e,G_2|G_1} P_{e,B_1} \overline{e^{-\lambda N_2 T_b}} e^{-\mu N_1 T_b} \\ c_{21}^{g_2} &= P_{e,G_2|B_1} P_{e,G_1} e^{-\lambda N_2 T_b} e^{-\lambda N_1 T_b}, \quad c_{22}^{g_2} = P_{e,G_2|B_1} P_{e,B_1} \overline{e^{-\lambda N_2 T_b} e^{-\mu N_1 T_b}} \\ c_{23}^{g_2} &= P_{e,G_2|B_1} P_{e,G_1} e^{-\lambda N_2 T_b} \overline{e^{-\lambda N_1 T_b}}, \quad c_{24}^{g_2} = P_{e,G_2|B_1} P_{e,B_1} \overline{e^{-\lambda N_2 T_b}} e^{-\mu N_1 T_b} \\ c_{31}^{g_2} &= P_{e,B_2|G_1} P_{e,G_1} \overline{e^{-\lambda N_2 T_b}} e^{-\lambda N_1 T_b}, \quad c_{32}^{g_2} = P_{e,B_2|G_1} P_{e,B_1} e^{-\mu N_2 T_b} \overline{e^{-\mu N_2 T_b}} \\ c_{33}^{g_2} &= P_{e,B_2|G_1} P_{e,G_1} \overline{e^{-\mu N_2 T_b} e^{-\lambda N_1 T_b}}, \quad c_{34}^{g_2} = P_{e,B_2|G_1} P_{e,B_1} e^{-\mu N_2 T_b} e^{-\mu N_1 T_b} \\ c_{41}^{g_2} &= P_{e,B_2|B_1} P_{e,G_1} \overline{e^{-\mu N_2 T_b}} e^{-\lambda N_1 T_b}, \quad c_{42}^{g_2} = P_{e,B_2|B_1} P_{e,B_1} e^{-\mu N_2 T_b} \overline{e^{-\mu N_2 T_b}} \\ c_{43}^{g_2} &= P_{e,B_2|B_1} P_{e,G_1} \overline{e^{-\mu N_2 T_b} e^{-\lambda N_1 T_b}}, \quad c_{44}^{g_2} = P_{e,B_2|G_1} P_{e,B_1} e^{-\mu N_2 T_b} e^{-\mu N_1 T_b}. \end{aligned} \tag{B.22}$$

And the initial conditions are

$$\begin{aligned}
A_{f,G_2|G_1} &= P_G P_{e,G_1} P_{e,G_2|G_1} e^{-\lambda N_1 T_b} X^{(N_1+N_2)T_b} Y^{(N_1+N_2)E_c} \\
A_{f,G_2|B_1} &= P_B P_{e,B_1} P_{e,G_2|B_1} \overline{e^{-\mu N_1 T_b}} X^{(N_1+N_2)T_b} Y^{(N_1+N_2)E_c} \\
A_{f,B_2|G_1} &= P_G P_{e,G_1} P_{e,B_2|G_1} \overline{e^{-\lambda N_1 T_b}} X^{(N_1+N_2)T_b} Y^{(N_1+N_2)E_c} \\
A_{f,B_2|B_1} &= P_B P_{e,B_1} P_{e,B_2|B_1} e^{-\mu N_1 T_b} X^{(N_1+N_2)T_b} Y^{(N_1+N_2)E_c}. \qquad \text{(B.23)}
\end{aligned}$$

Therefore, the total generating function for this system is

$$\begin{aligned}
G_s = {} & G_1 + [G_{f,G_1}(e^{-\lambda N_1 T_b}\overline{P_{e,G_2|G_1}} + \overline{e^{-\lambda N_1 T_b}}\,\overline{P_{e,B_2|G_1}}) + \\
& G_{f,B_1}(\overline{e^{-\mu N_1 T_b}}\,\overline{P_{e,G_2|B_1}} + e^{-\mu N_1 T_b}\overline{P_{e,B_2|B_1}})]X^{N_2 T_b} Y^{N_2 E_c} + \\
& [(G_{f,G_2|G_1} + G_{f,G_2|B_1})(e^{-\lambda N_2 T_b}\overline{P_{e,G_1}} + \overline{e^{-\lambda N_2 T_b}}\,\overline{P_{e,B_1}}) + \\
& (G_{f,B_2|G_1} + G_{f,B_2|B_1})(\overline{e^{-\mu N_2 T_b}}\,\overline{P_{e,G_1}} + e^{-\mu N_2 T_b}\overline{P_{e,B_1}})]X^{N_1 T_b} Y^{N_1 E_c}. \qquad \text{(B.24)}
\end{aligned}$$

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] *Forum, wireless applicaion protocol architecture specification (2000)*, http://www.wapforum.org/what/technical.html.

[2] *Irda sir data specification (2000)*, http://www.irda.org/standards/specifications.asp.

[3] *IEEE standard for 802.11 wireless lan medium access control (MAC) and physical layer (phy) specifications*, Nov. 1997.

[4] *IEEE standard for 802.11 wireless lan medium access control (MAC) and physical layer (phy) specifications*, Nov. 1999.

[5] G. Bianchi, *Performance analysis of the IEEE 802.11 distributed coordination function*, IEEE J. Select. Areas Commun. **18** (2000), no. 3, 535–547.

[6] G. Bianchi, L. Fratta, and M. Oliveri, *Performance analysis of IEEE 802.11 CSMA/CA medium access control protocol*, Proc. IEEE Intn'l Symposium on Personal, Indoor, and Mobile Radio Communications (Taipei, Taiwan), Oct. 1996, pp. 407–411.

[7] G. Bianchi and I. Tinnirello, *Analysis of priority mechanisms based on differentiated inter frame spacing in csma-ca*, Proc. Vehicular Tech. Conf., Oct. 2003, pp. 1401–1405.

[8] M. Brahma, K. W. Kim, M. E. Hachimi, A. Abouaissa, and P. Lorenz, *A buffer and energy based scheduling in mobile ad hoc networks over link layer*, Proc. Advanced Intn'l. Conf. on Telecom. and Web Applications and Services, Mar. 2006, pp. 19–25.

[9] H. Bruneel and M. Moeneclaey, *On throughput performance of some continuous ARQ strategies with repeated transmissions*, IEEE Trans. Communications **34** (1986), 244–249.

[10] P. Bucciol, G. Davini, E. Masala, E. Filippi, and J. C. De Martin, *Cross-layer perceptual ARQ for h.264 video streaming over 802.11 wireless networks*, Proc. Globecom Conf., Dec. 2004, pp. 3027–3031.

[11] F. Cali, M. Conti, and E. Gregori, *IEEE 802.11 wireless LAN: capacity analysis and protocol enhancement*, Proc. INFOCOM. (San Francisco, CA), Mar. 1998, pp. 142–149.

[12] X. R. Cao and Y. C. Ho, *Models of discrete event dynamic systems*, IEEE Control Systems. Mag. **10** (1990), 69–76.

[13] S. S. Chakraborty and M. Liinaharja, *Performance analysis of an adaptive SR ARQ scheme for time-varying rayleigh fading channels*, Proc. Intn'l Communications Conf. (Helsinki, Finland), June 2001, pp. 2478–2482.

[14] S. Y. Chang, A. Anastaspoulos, and W. E. Stark, *Energy-delay analysis of wireless systems with random coding*, Proc. Globecom Conf. (Dallas, USA), Dec. 2004, pp. 3270 – 3274.

[15] S. Y. Chang, W. E. Stark, and A. Anastasopolous, *Energy-delay analysis of MAC protocols in wireless networks*, IEEE Trans. Communications (2005), (Submitted).

[16] C. M Chao, J. P. Sheu, and I. C. Chou, *An adaptive quorum-based energy conserving protocol for IEEE 802.11 ad hoc networks*, IEEE Trans. Mobile Computing **5** (2006), 560–570.

[17] H. S. Chhaya and S. Gupta, *Performance modeling of asynchronous data transfer methods of IEEE 802.11 MAC protocol*, Wireless Networks **3** (1997), no. 3, 217–234.

[18] B. P. Crow, *Performance evaluaion of the IEEE 802.11 wireless local area network protocol*, Master's thesis, University of Arizona, Tucson, AZ, 1996.

[19] P. E. Engelstad and O. N. Osterbo, *Non-saturation and saturation analysis of IEEE 802.11e edca with starvation prediction*, Proc. MSWiM (Montreal, Quebec, Canada), Oct. 2005, pp. 224–233.

[20] G. F. Franklin, J. D. Powell, and A. Emami-Naeini (eds.), *Feedback control of dynamic systems*, Addision Wesely, 1994.

[21] F. Granelli and D. Kliazovich, *Cross-layering for performance improvement in multi-hop wireless networks*, Proc. Intn'l Symposium on Parallel Architectures, Algorithms and Networks, Dec. 2005, pp. 1–6.

[22] Z. Hadzi-Velkov and B. Spasenovski, *Saturation throughput-delay analysis of IEEE 802.11 DCF in fading channel*, Proc. Intn'l Communications Conf., vol. 1, Sept. 2004, pp. 121–126.

[23] T. S. Ho and K. C. Chen, *Performance evaluation and enhancement of the CAMA/CA MAC protocol for 802.11 wireless LANs*, Proc. IEEE Intn'l Symposium on Personal, Indoor, and Mobile Radio Communications (Taipei, Taiwan), Oct. 1996, pp. 392–396.

[24] R. P. F. Hoefel, *A cross-layer saturation goodput analysis for IEEE 802.11a networks*, Proc. Vehicular Tech. Conf. (Stockholm, Sweden), May 2005, pp. 2051–2055.

[25] L. Huang and T. H. Lai, *On the scalability of IEEE 802.11 ad hoc networks*, Proc. MOBIHOC (Lausanne, Switzerland), June 2002, pp. 173–182.

[26] S. Kallel, *Analysis of memory and incremental redundancy ARQ schemes over a nonstationary channel*, IEEE Trans. Communications **40** (1992), 1474–1480.

[27] S. Kallel, R. Link, and S. Bakhtiyari, *Throughput performance of memory ARQ schemes*, Proc. Vehicular Tech. Conf. **48** (1999), 891–899.

[28] V. Kanodia, C. Li, A. Sabharwal, B. Sadeghi, and E. Knightly, *Distributed multi-hop scheduling and medium access with delay and throughput constraints*, Proc. SIGMOBILE (Rome, Italy), July 2001, pp. 200–209.

[29] ———, *Distributed priority scheduling and medium access in ad hoc networks*, Wireless Networks **8** (2002), 455–466.

[30] ———, *Modeling the 802.11 distributed coordination function in non-saturated conditions*, Communications Lett. **9** (2005), 715–717.

[31] H. Kim and J. C. Hou, *Improving protocol capacity with model-based frame scheduling in IEEE 802.11-operated wlans*, Proc. MobiCom (San Diego, California, USA), Sept. 2003, pp. 14–19.

[32] ———, *Modeling single-hop wireless networks under rician fading channels*, Mar. 2004, pp. 219–224.

[33] L. Kleinrock and F. Tobagi, *Packet switching in radio channels, part II-the hidden terminal problem in carrier sense multiple access and the busy tone solution*, IEEE Trans. Communications **33** (1975), no. 12, 1417–1433.

[34] S. Lin and D. J. Costello (eds.), *Error control coding: Fundamentals and applications*, 1983.

[35] S. Lin, D. J. Costello, and M. J. Miller, *Automatic-repeat-request rrror control scheme*, IEEE Commun. Mag. **22** (1984), no. 12, 5–17.

[36] M. H. Manshaei, G. R. Cantieni, C. Barakat, and T. Turletti, *Performance analysis of the IEEE 802.11 MAC and physical layer protocol*, Proc. Intn'l Symposium on World of Wireless Mobile and Multimedia Networks, June 2005, pp. 88–97.

[37] A. C. Martins and J. C. Alves, *ARQ protocols with adaptive block size perform better over a wide range of bit error rates*, IEEE Trans. Communications **38** (1990), 737–739.

[38] J. J. Metzner and J. M. Chung, *Efficient energy utilization with time constraints in mobile time varying channels*, IEEE Trans. Veh. Tech. **49** (2000), 1169–1177.

[39] Qiang Ni, *Performance analysis and enhancements for IEEE 802.11e wireless networks*, IEEE Network **19** (2005), 21–27.

[40] M. Rice and S. B. Wicker, *A sequential scheme for adaptive error control over slowly varying channels*, IEEE Trans. Communications **42** (1994), 1533–1543.

[41] K. Sakakibara, S. Chikada, and J. Yamakita, *Analysis of unsaturation throughput of IEEE 802.11 dcf*, Proc. Intn'l Conf. on Personal Wireless Communications, Jan. 2005, pp. 134–138.

[42] C. Suh and Y. B. Ko, *A traffic aware, energy efficient MAC protocol for wireless sensor networks*, Proc. Intn'l Symposium on Circuits and Systems, May 2005, pp. 2975–2978.

[43] I. Tinnirello and S. Choi, *Efficiency analysis of burst transmissions with block ack in contention-based 802.11e wlans*, Proc. Intn'l Communications Conf. (Seoul, Korea), May 2005, pp. 3455–3460.

[44] B. Vucetic, *An adaptive coding scheme for time-varying channels*, IEEE Trans. Communications **39** (1991), 653–663.

[45] J. Wall and J. Y. Khan, *An adaptive ARQ enhancement to support multimedia traffic using 802.11 wireless lans*, Proc. Globecom Conf., Dec. 2004, pp. 3037–3041.

[46] L. C. Wang, S. Y. Huang, and A. Chen, *On the throughput performance of csma-based wireless local area network with directional antennas and capture effect: A cross-layer analytical approach*, Mar. 2004, pp. 1879–1884.

[47] Y. M. Wang and S. Lin, *A modified selective-repeat type-II hybrid ARQ system and its perofrmance analysis*, IEEE Trans. Communications **31** (1988), 593–608.

[48] J. Weinmiller, M. Schlager, A. Festag, and A. Wolisz, *Performance study of access control in wireless LANs IEEE 802.11 DFWMAC and ETSI RES 10 HIPERLAN*, Mobile Networks and Applications **2** (1997), no. 1, 55–67.

[49] Lie Hua Wong, *Error exponent regions for multi-user channels*, Ph.D. thesis, Department of Elec. and Comp. Engin., University of Michigan, Ann Arbor, 2005.

[50] Y. D. Yao, *An effective go-back-N ARQ scheme for variable-errorrate channels*, IEEE Trans. Communications **43** (1995), 20–23.

[51] W. Ye, J. Heidemann, and D. Estrin, *Medium access control with coordinated adaptive sleeping for wireless sensor networks*, IEEE/ACM Trans. Networking **12** (2004), 493–506.

[52] J. Yun and S. Bahk, *Parallel contention algorithm with csmaka for ofdm based high speed wireless lans*, Proc. IEEE Intn'l Symposium on Personal, Indoor, and Mobile Radio Communications, Sept. 2003, pp. 2581–2585.

ABSTRACT

Generating Function Analysis of Wireless Networks and ARQ Systems

by

Shihyu Chang

Co-Chair: Professor Wayne E. Stark

Co-Chair: Associate Professor Achilleas Anastasopoulos

In this thesis, there are two main themes. The first part of the thesis is to study the tradeoff between energy and delay for wireless networks. A network using a request-to-send (RTS) and clear-to-send (CTS) type medium access control (MAC) protocol is considered. We first determine the average delay incurred and the average energy consumed when the effects of an imperfect channel are incorporated in the model. In order to incorporate the relation between packet error probability, energy, and delay, we use the reliability function bounds for different channel models on the error probability of a coded system. Further, we present a generic framework that allows us to obtain the joint statistics of energy and delay through their joint generating function. Several important design tradeoffs are studied from the joint generating function, such as the average energy with an outage delay constraint. This framework allows us to optimize over the system parameters for various objective

functions, such as average delay. An approximation method is also proposed to calculate the average energy and average delay analytically. This approximation is found to be quite accurate for a wide range of lengths. The inhomogeneous and time-varying effects for the tradeoff between energy and delay are also studied.

The second part of the thesis is to propose an analytical method to determine the joint distribution of the energy and delay of automatic repeat request (ARQ) protocols over time varying channels. A finite state machine (FSM) is used to model the transmitter, receiver and channel. From the state transition diagram of the FSM, the generating function of energy and delay consumption can be evaluated according to the Manson's gain formula while incorporating physical layer characteristics (packet error probability as a function of energy and delay). We also consider a receiver containing memory of previously received samples and derive the cutoff rate for three different receiver structures. As the numerical results demonstrate, the time-varying characteristic of the channel have a great influence on the system performance.