FEEDBACK IN WIRELESS NETWORKS: CROSS-LAYER DESIGN, SECRECY AND RELIABILITY

DISSERTATION

Presented in Partial Fulfillment of the Requirements for

the Degree Doctor of Philosophy in the

Graduate School of The Ohio State University

By

Praveen Kumar Gopala, B.E., M.S.

* * * * *

The Ohio State University

2007

Dissertation Committee:

Hesham El Gamal, Adviser

Philip Schniter

Randolph Moses

Approved by

Adviser

Graduate Program in
Electrical & Computer
Engineering

ABSTRACT

The central theme of this dissertation is the impact of feedback on the performance of wireless networks. Wireless channels offer a multitude of new challenges and opportunities that are uncharacteristic of wireline systems. We reveal the crucial role of feedback in exploiting the opportunities and in overcoming the challenges posed by the wireless medium. In particular, we consider three distinct scenarios and demonstrate the different ways in which feedback helps improve performance.

We first consider cellular multicast channels and show that the availability of feedback allows for the cross-layer design of efficient multicast schedulers. Here we focus on two types of feedback scenarios: perfect channel state information (CSI) feedback and automatic repeat request (ARQ) feedback. We propose low-complexity multicast schedulers that achieve near-optimal asymptotic throughput-delay tradeoffs for both feedback scenarios. We further propose a cooperative multicast scheduler, requiring perfect CSI feedback, that achieves the optimal asymptotic scaling of both throughput and delay with the number of users. For the multiple transmit antenna scenario, we show that the wireless multicast gain dominates the throughput performance of multicast schedulers and demonstrate the near-optimality of the proposed worst user scheduler with a large number of transmit antennas.

Next, we consider fading eavesdropper channels and reveal the importance of feedback in establishing secure communications. We characterize the secrecy capacity

of such channels under the assumptions of full CSI and main (legitimate) channel CSI knowledge at the transmitter, and propose optimal rate and power allocation strategies. Interestingly, we show that the availability of CSI feedback enables one to exploit the time-varying nature of the wireless medium and achieve a perfectly secure non-zero rate even when the eavesdropper channel is more capable than the legitimate receiver channel on the average. We further establish the critical role of rate adaptation, based on the main channel CSI, in facilitating secure communications over slow fading channels. We also propose a low-complexity on/off power allocation strategy and establish its asymptotic optimality. We then consider a minimal ARQ feedback scenario and propose transmission schemes that leverage the ARQ feedback to achieve non-zero perfect secrecy rates even when the eavesdropper has a superior channel on the average. Thereby, we establish the positive impact of feedback on the secrecy capacity of fading channels.

Finally, we consider ARQ channels with strict delay deadline constraints and demonstrate the impact of ARQ feedback on reliability. We show that ARQ feedback can be exploited to significantly improve the achievable error exponents, and propose an Incremental Redundancy ARQ (IR-ARQ) scheme that significantly outperforms the schemes based on memoryless decoding.

To Amma, Achan, and Chinnu, for their love and support.

ACKNOWLEDGMENTS

First, I would like to thank my dear parents, Mr. P. Gopala Kurup and Mrs. Rema G. Kurup, and my brother for their constant love and support. They have always been there behind me, providing me with encouragement and support, whenever I have needed it.

I am very grateful to my advisor, Prof. Hesham El Gamal, for guiding me through the Ph.D. program and molding me into what I am today. I thank him for providing me with motivation and direction, for being confident in my abilities, and for spending time on my research in the midst of his busy schedule. I have always been amazed by his ability to think intuitively, while simultaneously having a clear idea of the bigger picture. That is the one quality that I hope I have imbibed from him during my Ph.D. program.

I would like to thank Prof. Philip Schniter and Prof. Randolph L. Moses for agreeing to be in my dissertation committee, and for providing me with valuable comments and feedback, all through the program. A special thanks to Prof. Andrea Serrani for inspiring me during the initial phase of my graduate life with his exemplary teaching. My first graduate school course was taught by him and I would still rank it as the best course that I have taken in my entire graduate program.

Not thanking all the people in the IPS lab would be a crime at this point. They have been like family to me, and I have enjoyed every minute of my graduate school life that I have spent with them. I will always cherish the innumerable group lunches/dinners, the IPS picnics, the discussions and the wonderful, friendly and symbiotic environment that they offered me throughout my graduate life. I would also like to thank all my friends at Ohio State. Life would not have been much fun without all of you. Last but not in any way least, I thank Ms. Jeri McMichael for being patient with me and helping me out of one too many tough situations. Thanks Jeri for making our lives a lot easier.

VITA

August, 1981	Born - Ernakulam, Kerala, India
2002	B.E. Electronics & Communication, College of Engineering, Guindy, Anna University, Chennai, India
2004	M.S. Electrical Engineering, The Ohio State University, Columbus, Ohio
2004-2007	Graduate Research Associate, The Ohio State University.

PUBLICATIONS

Research Publications

1. Praveen K. Gopala and Hesham El Gamal, "Scheduling for Cellular Multicast: A Cross-Layer Perspective", *Submitted to the IEEE Transactions on Mobile Computing*, Sep. 2007.

2. Praveen K. Gopala, Young-Han Nam and Hesham El Gamal, "On the Error Exponents of ARQ Channels with Deadlines", *To appear in the IEEE Transactions on Information Theory*, Jun. 2007.

3. Praveen K. Gopala, Lifeng Lai and Hesham El Gamal, "On the Secrecy Capacity of Fading Channels", *Submitted to the IEEE Transactions on Information Theory*, Oct. 2006.

4. Young-Han Nam, Praveen K. Gopala and Hesham El Gamal, "ARQ Diversity in Fading Random Access Channels", *Submitted to the IEEE Transactions on Wireless Communications*, Aug. 2006.

5. Arul D. Murugan, Praveen K. Gopala and Hesham El Gamal, "Correlated Sources over Wireless Channels: Cooperative Source-Channel Coding", *IEEE Journal on Selected Areas in Communications*, Vol. 22, No. 6, pp. 988–998, Aug. 2004.

6. Praveen K. Gopala, Lifeng Lai and Hesham El Gamal, "On the Secrecy Capacity of Fading Channels", *Proceedings of ISIT 2007*, June 2007.

7. Young-Han Nam, Praveen K. Gopala and Hesham El Gamal, "Resolving Collisions via Incremental Redundancy: The ARQ Diversity", *Proceedings of IEEE Infocom 2007*, May 2007.

8. Lifeng Lai, Praveen K. Gopala and Hesham El Gamal, "Secure Communication over Wireless Channels", *Information Theory & Applications (ITA) Workshop*, Jan. 2007 **(Invited)**.

9. Stephen F. Bush, Praveen K. Gopala and Orhan Imer, "Enhancing Reliable Multicast Transport to Mitigate the Impact of Blockage", *Proceedings of IEEE CAMAD 2006*, Jun. 2006.

10. Praveen K. Gopala and Hesham El Gamal, "On the Throughput-Delay Tradeoff in Cellular Multicast", *Proceedings of the Symposium on Information Theory in WirelessCom 2005*, Jun. 2005.

11. Praveen K. Gopala and Hesham El Gamal, "Opportunistic Multicasting", *Proceedings of the Asilomar Conf. on Signals, Systems and Computers*, Nov. 2004 **(Invited)**.

12. Praveen K. Gopala and Hesham El Gamal, "On the Scaling Laws of Multimodal Wireless Sensor Networks", *Proceedings of IEEE Infocom'04*, Mar. 2004.

13. Praveen K. Gopala and Hesham El Gamal, "On the Scaling Laws of Dense Wireless Sensor Networks", *Proceedings of ACM Sensys 2003*, UCLA, Nov. 2003 **(Invited)**.

14. N. R. Karthikeyan, Mani Sridhar, G. Praveen Kumar, K. Ramakrishnan, R. Jayaparvathy and S. Srikanth, "Priority Oriented Scheduling in cellular systems with dynamic packet assignment", *Proceedings of the* 9^{th} *National Conf. on Communications*, Aug. 2002.

FIELDS OF STUDY

Major Field: Electrical and Computer Engineering

Studies in:

Communications	Prof. Hesham El Gamal
Signal Processing	Prof. Philip Schniter
Control Theory	Prof. Andrea Serrani

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

Table **Page**

CHAPTER 1

INTRODUCTION

One of the significant advances in communication technology has been the shift from wireline communication systems (like the Public Switched Telephone Network (PSTN) and Ethernet-based Local Area Networks (LANs)) to wireless systems (like Cellular networks (GSM, IS-95, WCDMA), Bluetooth and Wireless LANs (Wi-Fi, 802.11a/b/g/n)). Even though the core infrastructure in some of these new systems is still wireline-based, the segments of the communication link at the end users use wireless technology. The freedom of mobility that these wireless systems provide to the users has been the primary impetus to their popularity. However, this shift to wireless communication also brings with it a multitude of new challenges and opportunities, that require an in-depth understanding of the properties of the wireless medium, and cannot be addressed by mere extensions of wireline solutions. We now present a brief overview of some of the properties of wireless channels that will play an integral role in the rest of the dissertation.

A significant property affecting performance in wireless systems is the phenomenon of channel fading. When a signal is transmitted over a wireless medium, multiple copies of the signal are seen by the receiver due to the reflection and scattering of

the transmitted signal by objects in the vicinity of the transmission. These multiple copies arrive with different delays at the receiver and interfere constructively or destructively with each other, thereby causing random fluctuations in the received signal power, which is referred to as **multi-path fading**. Several methods have been proposed to mitigate fading and improve the reliability of transmissions. The main idea of these diversity techniques is to transmit multiple copies of the same symbol over different channels that experience independent fading, while the receiver uses all its observations jointly to decode the transmitted symbol. This ensures that even if some channels are in a deep fade, the symbol can still be reliably decoded using the observations from other channels. Thus, by taking multiple instances of the channel and averaging out the effects of fading, all these methods, in principle, strive to convert the fading channel into an AWGN channel. Examples of such diversity techniques include time diversity (Time interleaving of coded bits), frequency diversity (Frequency hopping/ Direct sequence spread spectrum) and spatial diversity (Beamforming, Space-time coding, Rake combining).

However, it has been shown recently that averaging out the time-varying nature of fading channels, through diversity techniques, is not optimal when there are a large number of users in the system. One can reap significant performance gains by exploiting a form of diversity, called **multi-user diversity**, that focuses rather on exploiting the time-varying nature of fading channels and is inherent in networks with a large number of users. This contrasting outlook of exploiting fading was first introduced by Knopp and Humblet [1] in the uplink of a cellular system. They showed that when the users experience symmetric fading, the optimal scheme that maximizes system capacity, allows only the user with the best channel to transmit at any given

time[1]. This result can be intuitively explained as follows: When the system has a large number of users that experience independent fading, it is highly likely that there will be at least one user in the system, at any particular instant, whose channel is much better than the average. Thus by always transmitting to such "good" users in every time slot, one can ride the "peaks" of the users' channel variations and achieve significant performance gains. In fact, it was shown in [1] that by exploiting multi-user diversity, the throughput of a system with a large number of fading users can be made significantly higher than that of a Gaussian system with the same average received power, where the users do not experience any fading. Similar results highlighting the throughput gains due to multi-user diversity were shown in [2–4]. However, it should be noted that these throughput gains come at the price[2] of an increased delay and may also lead to fairness issues, especially in asymmetric fading scenarios.

Wireless systems also have an inherent **multicast** property, that is uncharacteristic of wireline systems. Any signal transmitted to a particular user will also reach all the other users in the system *for free*. This property can be exploited to yield performance gains in multicast systems, where one desires to send the same information to all the users (or a group of users) in the system, as will be shown in Chapter 2 of this dissertation. On the other hand, this inherent multicast property also has a detrimental effect in that a user's transmission will now interfere with any other simultaneous transmissions within the system that are in the vicinity of that user. This inter-user **interference** complicates system design by introducing new constraints,

[1]If all the users experience deep fades in a particular time slot, then none of the users transmit in that slot. But the probability of this event is negligible when there are a large number of users.

[2]Note that a feedback link is also necessary for implementing such a scheme, since the transmitter either needs to know the users' channel gains (downlink) or needs to communicate the scheduling decisions to the users (uplink).

and forces the designer to adopt a holistic view of the system while devising communication schemes. For example, it has been shown by Gupta and Kumar [5] that in wireless ad hoc networks, the best strategy to increase the total throughput of the network, is to let a node use the least possible transmission power and communicate with its nearest neighbors (thereby causing least interference to other simultaneous transmissions within the network).

Another advantage of the multicast property of wireless channels is the possibility for user **cooperation**. Since the transmission to any particular user is also received by other users in the system, one can improve the throughput and/or reliability of the transmission by allowing these "relay" users to cooperate with the source and/or destination user. The problem of devising intelligent strategies for user cooperation has received considerable attention in recent years, and several cooperation schemes have been proposed [6–10]. Some examples of cooperative protocols are Amplify-and-Forward, Decode-and-Forward and Compress-and-Forward, wherein the cooperating users amplify/decode/compress their received observations (respectively) before forwarding them to the destination. Another cooperation strategy for a dense sensor network scenario with a central collector node is proposed in [10]. Here cooperation is achieved by allowing each sensor to first transmit its information to its neighbors, and then using a cooperative beamforming (Virtual-MIMO) strategy to improve the throughput and reliability of transmissions to the collector node. In this dissertation, we highlight the advantages of user cooperation for a multicast setting in Chapter 2, where we propose a Cooperative Multicast[3] scheme that is similar in spirit to the Decode-and-Forward scheme. The proposed scheme allows the stronger users in the

[3]A detailed description of this scheme is provided in Section 2.2.3.

system to decode the multicast transmissions and forward them to the other weaker users. In fact, it is shown that this cooperative scheme is asymptotically optimal in both the delay and throughput sense.

Security is also an important issue that arises in wireless systems due to their inherent multicast nature. While transmitting confidential messages to a particular user, it becomes important to ensure that the other users in the system, who also receive the transmission, are unable to decode these messages. The notion of information-theoretic security, where an eavesdropper does not gain any information about the confidential message even after observing all the transmissions, has been studied in [11–14]. Interestingly, it has been shown in [13] that for AWGN channels, one cannot ensure perfect secrecy when the channel of an eavesdropper is better than the channel of the legitimate receiver. In Chapter 3 of this dissertation, we consider a wireless fading scenario and show that one can exploit the time-varying nature of the wireless medium (channel fading) to ensure perfect secrecy, even when the channel of an eavesdropper is better than the legitimate user's channel on the average.

All these different challenges and opportunities offered by the wireless channel suggest that the physical layer of any wireless communication system can no longer be separated from other higher layers (like the network and medium access control (MAC) layers). It necessitates the need for a **cross-layer design** perspective, which takes the properties of the wireless medium into account, while designing efficient routing and scheduling schemes for wireless networks. One example highlighting the importance of cross-layer design at the network layer is the multi-hop nearest-neighbor routing protocol proposed by Gupta and Kumar [5], wherein physical layer issues like inter-user interference are also included in the routing design. Another example that

emphasizes cross-layer design at the MAC layer is the best-user scheduler proposed by Knopp and Humblet [1], wherein the scheduler tries to exploit the gains offered by multi-user diversity to increase system throughput. Other works highlighting the importance of cross-layer design for wireless systems include [15–18]. In Chapter 2 of this dissertation, we consider the design of efficient schedulers for a multicast scenario, which is characterized by a strong interaction between the network, medium access and physical layers. We demonstrate the importance of adopting a cross-layer perspective for this scenario by quantifying the potential gains achieved by exploiting physical layer properties (like multi-user diversity, multicast gain and user cooperation) in the scheduler design.

Throughout this dissertation, we focus on studying the impact of **feedback** on the performance of wireless systems. Feedback plays a crucial role in overcoming the afore-mentioned challenges and in efficiently utilizing all the resources offered by the wireless medium. We focus primarily on two different feedback scenarios:

- Perfect CSI feedback, where perfect channel state information is made available to the transmitter(s). This feedback scenario is idealistic and serves as an upper bound on the performance achieved by other practical feedback scenarios.

- ARQ feedback, which represents the minimal feedback scenario where only one bit (ACK/NACK) indicating the success or failure of a transmission, can be fed back to the transmitter(s).

The impact of these and other feedback mechanisms on the performance of communication systems has been studied extensively by many researchers [19–23]. For example, Shannon proved an interesting negative result in [19] that feedback does not

increase the capacity of discrete memoryless channels. However, it was later shown that feedback does offer significant gains in error performance (reliability) [20, 21] and can greatly simplify the system design for memoryless channels. For example, to achieve the capacity of a Binary Erasure Channel (BEC) without feedback, one needs to use a fairly complex encoding/decoding strategy (The encoder needs to transmit long codeword sequences which result in significant decoding complexity at the decoder). However, the structure of the capacity-achieving encoding/decoding scheme changes significantly with the availability of minimal ARQ feedback. In this case, the receiver feeds back a NACK/ACK bit based on whether it sees an erasure or not, while the transmitter resorts to uncoded transmission and merely repeats each information bit until it gets an ACK from the receiver. This shows that even the presence of minimal ARQ feedback can greatly simplify the system design. It has been shown in [22, 23] that minimal ARQ feedback also has a positive impact on the reliability of wireless channels.

In this dissertation, we reveal the role of feedback in improving the performance of wireless networks. In particular, we consider three distinct scenarios and demonstrate the different ways in which feedback helps improve performance. In Chapter 2, we first consider a cellular multicast scenario and show that the availability of feedback allows for the cross-layer design of efficient multicast schedulers. Specifically, we propose low-complexity multicast schedulers for both the perfect CSI and the ARQ feedback scenarios, that achieve near-optimal throughput-delay tradeoffs. Next, in Chapter 3, we reveal the importance of feedback in establishing secure communication over fading eavesdropper channels. Specifically, we propose transmission schemes for different perfect CSI and ARQ feedback scenarios, that achieve non-zero perfect secrecy rates

even when an eavesdropper's channel is better than the legitimate receiver's channel on the average. Finally, in Chapter 4, we demonstrate the impact of ARQ feedback on reliability. We show that ARQ feedback can be exploited to significantly improve the achievable error exponents over channels with strict delay deadline constraints.

1.1 Contributions and Outline

We now provide a brief outline and the main contributions of each chapter in the dissertation. (The main results of this dissertation are documented in [24–26].)

In **Chapter 2**, we consider the multicast channel in a single cell system, where a common information stream is transmitted by the base station to multiple users. For this scenario, we propose three classes of scheduling algorithms with progressively increasing complexity, viz. Static schedulers with memoryless decoding, Incremental Redundancy (IR) multicast and Cooperative multicast, and evaluate their asymptotic throughput-delay performance with the number of users in the system. The main contributions of this chapter can be summarized as follows:

- We show the existence of a static scheduler with memoryless decoding (Median user scheduler) that achieves near-optimal scaling of both throughput and delay with the number of users in the system.

- IR multicast is shown to achieve a superior throughput-delay tradeoff than the static schedulers with memoryless decoding. Moreover, unlike the other proposed schedulers, IR multicast relies only on minimal ARQ feedback and does not require perfect CSI knowledge at the base station.

- Cooperative multicast is shown to be optimal in the asymptotic sense, i.e., it achieves the optimal scaling of both throughput and delay with the number of users in the system.

- When the base station is equipped with multiple transmit antennas (with limited feedback), it is shown that the wireless multicast gain harnessed by a scheduler dominates its throughput performance (as the number of transmit antennas becomes large).

In **Chapter 3**, we consider the secure transmission of information over an ergodic fading channel in the presence of an eavesdropper. We characterize the secrecy capacity of such a system under different assumptions on the CSI available at the transmitter. We further derive the perfect secrecy rates achievable using ARQ feedback from the legitimate receiver. The main contributions of this chapter can be summarized as follows:

- We characterize the secrecy capacity for the full CSI case (where the transmitter knows the CSI of the legitimate receiver and the eavesdropper) and propose the optimal power and rate allocation strategies that achieve capacity.

- We characterize the secrecy capacity for the main channel CSI case (where the transmitter only knows the CSI of the legitimate receiver) and propose the optimal power and rate allocation strategies. Thereby, we establish the critical role of rate adaptation, based on the main channel CSI, in facilitating secure communication over fading channels.

- For the main channel CSI case, we propose a low-complexity on/off power allocation strategy that achieves near-optimal performance. Moreover, the proposed

on/off scheme is shown to be asymptotically optimal (for large average SNR), and interestingly, is also shown to attain the full-CSI secrecy capacity.

- We propose two transmission schemes based on ARQ feedback, viz. Repetition ARQ and IR-ARQ, and characterize their achievable perfect secrecy rates. Thereby, we reveal the positive impact of minimal ARQ feedback on the secrecy capacity of fading channels.

In **Chapter 4**, we consider communication over Automatic Repeat reQuest (ARQ) memoryless channels with strict delay deadline constraints. The delay constraint is imposed in the form of an upper bound L on the maximum number of ARQ transmission rounds. For this setup, we propose a transmission scheme based on incremental redundancy ARQ with joint decoding at the receiver, and evaluate the achievable error exponent. The main contributions of this chapter are as follows:

- Without delay constraints, Incremental Redundancy ARQ (IR-ARQ) achieves the same error exponent as Forney's memoryless decoding scheme [21].

- Under a deadline constraint, IR-ARQ outperforms Forney's memoryless decoding scheme in terms of the achievable error exponents.

- For the Binary Symmetric Channel (BSC) and the Very Noisy Channel (VNC), choosing $L = 4$ for IR-ARQ is sufficient to ensure the achievability of Forney's feedback exponent, which is typically achievable with memoryless decoding only as $L \to \infty$. Our numerical calculations indicate that this result also holds for the AWGN channel (at least for some range of SNRs).

Acronym	Expansion
ACK	Acknowledgement
ARQ	Automatic Repeat Request
AWGN	Additive White Gaussian Noise
BS	Base Station
BSC	Binary Symmetric Channel
CSI	Channel State Information
DMC	Discrete Memoryless Channel
IR-ARQ	Incremental Redundancy ARQ
KKT	Karush Kuhn Tucker
MAC	Medium Access Control
MRC	Maximal Ratio Combining
NACK	Negative ACK
QoS	Quality of Service
Rep-ARQ	Repetition ARQ
SNR	Signal to Noise Ratio
VNC	Very Noisy Channel

Table 1.1: List of Acronyms used in the dissertation

Finally in **Chapter 5**, we offer some concluding remarks and possible directions for future research. To enhance the flow of the dissertation, we collect all the proofs that do not offer much intuition, in the Appendix.

CHAPTER 2

FEEDBACK FOR CROSS-LAYER SCHEDULING: THE CELLULAR MULTICAST CHANNEL

Wireless networks are becoming increasingly popular primarily due to their ease of installation and the freedom of mobility that they offer to the users. Traditional data link and network layer protocols designed for wireless networks adopt simplified on/off models for the physical layer, and thereby focus on reducing the system to a wireline scenario. This approach of designing algorithms based on the assumption that the wireless channel behaves like a reliable, time-invariant bit-pipe has been shown to be highly sub-optimal, especially for applications with strict Quality of Service (QoS) constraints. Recent years have witnessed a growing interest in cross-layer design approaches for wireless system design. The underlying idea in these approaches is to jointly optimize the physical, data link, and networking layers in order to satisfy the QoS constraints with the minimum expenditure of network resources. Early investigations on cross-layer design have focused on the single user case [27, 28]. These works have shed light on the fundamental tradeoffs in this scenario and devised efficient power and rate control policies that approach these limits. More recent works have considered multi-user cellular networks [29–33]. These studies have enhanced

our understanding of the fundamental limits and the structure of optimal resource allocation strategies.

In this chapter, we generalize this cross-layer approach to the wireless multicast channel, where the same information stream is transmitted by the base station to multiple users within the network. Such multicast scenarios often occur in broadband wireless networks, due to their support for streaming video (Mobile TV) applications using the IP Datacasting (IPDC) and Digital Video Broadcasting-Handheld (DVB-H) frameworks. Moreover, these multicast scenarios are characterized by a strong interaction between the network, medium access, and physical layers. This interaction adds significant complexity to the problem which motivated the adoption of a simplified on/off model for the wireless channel in several of the recent works on wireless multicast [34–36]. In this chapter, we argue that employing more accurate models for the wireless channel allows for valuable opportunities for exploiting the wireless medium to yield performance gains. More specifically, we shed light on the role of the following characteristics of the wireless channel in the design of multicast scheduling strategies: 1) The *multi-user diversity* resulting from the statistically independent channels seen by the different users [1], 2) The *wireless multicast gain* resulting from the fact that any information transmitted over the wireless channel is *overheard* by all users, possibly with different attenuation factors, and 3) The *cooperative gain* resulting from antenna sharing between users [6].

2.1 System Model

We consider the downlink of a single cell system where a base station (BS) serves a group of N users. All the users request the same information from the BS. Unless

otherwise stated, the BS is assumed to be equipped with a single transmit antenna. Each user is assumed to have only a single receive antenna. We consider time-slotted transmission in which the received symbol vector at user i in time slot k is given by

$$\underline{y_i}[k] = h_i[k]\underline{x}[k] + \underline{n_i}[k],$$

where $\underline{x}[k]$ denotes the complex-valued vector of length m transmitted by the BS in slot k, $h_i[k]$ represents the complex flat fading coefficient of the channel between the BS and the i^{th} user in time slot k, and $\underline{n_i}[k]$ represents the zero-mean unit-variance complex additive white Gaussian noise vector at the i^{th} user in slot k. The noise processes are assumed to be circularly symmetric and independent across users. The channel between the BS and each user is assumed to be quasi-static with coherence time T_c. Thus the fading coefficients remain constant throughout an interval of length T_c (or m channel uses) and change independently from one interval to the next. The fading coefficients $\{h_i\}$ are assumed to be independent and identically distributed (i.i.d.) across the users (symmetric fading scenario) and follow a Rayleigh distribution with $\mathbb{E}\left[|h_i[k]|^2\right] = 1$, $\forall i, k$. Each packet transmitted by the BS is assumed to be of constant size S. We further employ the following short-term average power constraint at the base station

$$\frac{1}{m}\,\mathbb{E}\left[\|\underline{x}[k]\|^2\right] \quad \leq \quad P.$$

Clearly, further performance gain may be reaped through a carefully constructed power allocation policy if this short term power constraint is replaced by a long term one. However, in this chapter, we only focus on rate adaptation and scheduling based on the instantaneous channel state available at the BS. Moreover, the proposed

scheduling strategies, except the incremental redundancy scheme[4], assume perfect CSI knowledge at both the transmitter (BS) and receiver. In our throughput analysis, we use capacity expressions for the channel transmission rates. Here we implicitly assume that the BS employs coding schemes that approach the channel capacity which justify our use of the fundamental information theoretic limit of the channel.

In our delay analysis, we consider backlogged queues, and hence, the only meaningful measure of delay is the transmission delay. This leads to the following definitions for throughput and delay that will be adopted in this chapter.

Definition 1. *The* **throughput** *of a scheduling scheme is defined as the sum of the throughputs provided by the base station to each individual user within the system.*

Definition 2. *The* **delay** *of a scheduling scheme is defined as the delay between the instant representing the start of transmission of a packet, and the instant when the packet is successfully decoded by all the users in the system.*

We note here that our notion of delay does not account for the queuing delay experienced by the packets. We adopt this restricted notion to simplify the delay analysis, since significant complexity is added to the queuing delay analysis by the formation of coupled queues[5] in the multicast setting. However, our delay analysis offers a lower bound on the *total* delay which is very tight in several important special cases. Furthermore, this analysis provides a very useful tool for rank-ordering the different classes of scheduling algorithms and sheds light on their structural properties.

[4]For the incremental redundancy scheme, the BS only needs to know when to stop transmission of the current codeword.

[5]This notion of coupled queues will be made clear in the discussion of the best user scheduler in the next section.

To facilitate analytical tractability, we focus only on evaluating the asymptotic scaling laws of the throughput and delay of the proposed schedulers with the number of users in the system. In this analysis, we use the following asymptotic notations:
a) $f(n) = O(g(n))$ iff there are constants c and n_0 such that $f(n) \leq cg(n)\ \forall n > n_0$.
b) $f(n) = \Omega(g(n))$ iff there are constants c and n_0 such that $f(n) \geq cg(n)\ \forall n > n_0$.
c) $f(n) = \Theta(g(n))$ iff there are constants c_1, c_2 and n_0 such that $\forall n > n_0$, we have $c_1 g(n) \leq f(n) \leq c_2 g(n)$.

2.2 Cross-layer Multicast Schedulers

In a non-cooperative setting, wherein the users cannot help each other to decode the multicast transmissions, the throughput-optimal scheme is an N-level superposition coding/successive decoding scheme [37], where N parallel streams (layers) are transmitted simultaneously from the BS (Layered Multicast). While the user with the best channel can successfully decode all the N streams, the user with the worst channel can decode only one of the streams successfully (i.e., with arbitrarily small probability of error). This layered strategy, however, suffers from excessive complexity (especially when there are a large number of users in the system) and might be infeasible to implement in practice, since the mobile nodes have low processing power and tight battery power constraints. This motivates our work where we focus on the throughput and delay achieved by low complexity scheduling schemes, which do not require any superposition coding/successive decoding at the BS and users respectively, and are based only on single stream transmissions by the BS. Interestingly, we identify low complexity schedulers that achieve near-optimal scaling laws of both throughput and delay with the number of users N. Furthermore, we establish the

asymptotic optimality of the proposed cooperative multicast scheduler in terms of both delay and throughput.

2.2.1 Static Schedulers with Memoryless Decoding

In this class of schedulers, the BS always schedules transmissions to a desired fraction of the users with favorable channel conditions (by adjusting the transmission rate accordingly). While the identity of the users change, based on the instantaneous channel conditions, the fraction of users that are able to decode every transmitted packet always remains the same (and hence the name "static"). The memoryless decoding property dictates that the remaining users, who do not succeed in decoding, *flush* their memories and wait for future re-transmissions of the packet. This assumption is imposed to limit the complexity of the encoding/decoding process. In Section 2.2.2, we relax this memoryless decoding assumption and quantify the gains offered by carefully constructed ARQ schemes. Under this class of schedulers, we now study three different schemes of scheduling transmissions to the best, worst and median user in detail. These three schemes highlight the tradeoff between exploiting the multi-user diversity and the wireless multicast gain offered by the wireless medium.

Worst User Scheduler

This scheme maximally exploits the wireless multicast gain by always transmitting to the user with the least instantaneous SNR (worst channel) in the system. Hence at any time instant, the BS chooses its transmission rate to be the rate supported by the worst user in the system. This enables all the users to successfully decode the transmission (since any user with a higher supported rate can decode the transmitted information). Thus any data transmitted by the BS reaches all the users in a single

transmission. However, since the transmission rate supported by the worst user decays as the number of users becomes large, it is clear that the multi-user diversity inherent in the system works against the performance of this scheme and results in a decrease in the individual throughput to any user. The average throughput of the worst user scheduler is given by

$$R_{tot} = N\mathbb{E}\left[\log\left(1 + |h_{\pi(1)}|^2 P\right)\right],$$

where $|h_{\pi(1)}|^2 = \min_{i=1}^{N} |h_i|^2$ is the minimum channel gain among all the N users in the system, whose distribution and density functions are given by

$$F_{|h_{\pi(1)}|^2}(x) = 1 - e^{-Nx} \quad \text{and} \quad f_{|h_{\pi(1)}|^2}(x) = Ne^{-Nx}, \quad x \geq 0.$$

Throughout this chapter, the log(.) function refers to the natural logarithm, and hence, the average throughput is expressed in nats. Since any transmission by the BS reaches all the users in the system, the BS needs to maintain only a single common queue for all the users to implement this scheme.

Theorem 3. *The average throughput and the average delay of the worst user scheduler scale as*

$$R_{tot} = \Theta(1) \qquad \textit{and} \qquad D = \Theta(N), \tag{2.1}$$

with the number of users N.

Proof. Refer Appendix A.1. □

From Theorem 3, it is clear that the average throughput of the worst user scheduler does not scale with the number of users N, while the average delay increases linearly with N.

Best User Scheduler

This scheme maximally exploits the multi-user diversity available in the system. At any instant, the BS chooses its transmission rate to be the rate supported by the best user in the system. Since the transmission rate is adjusted based on the user with the maximum instantaneous SNR, none of the other users in the system will be able to successfully decode the transmission. Thus only one user is targeted in every transmission and the scheme fails to exploit any wireless multicast gain. Hence every packet must eventually be *repeated* N times before it reaches all the users. The average throughput of the best user scheduler is given by

$$R_{tot} = \mathbb{E}\left[\log\left(1 + |h_{\pi(N)}|^2 P\right)\right],$$

where $|h_{\pi(N)}|^2 = \max_{i=1}^{N} |h_i|^2$ is the maximum channel gain among all the N users in the system, whose distribution function is given by

$$F_{|h_{\pi(N)}|^2}(x) = \left(1 - e^{-x}\right)^N, \quad x \geq 0.$$

To implement this scheme, the BS needs to maintain a set of N *coupled* queues, one for each user in the system. These queues are coupled in the sense that any packet that needs to be transmitted enters all the N queues simultaneously (to ensure that the packet reaches all the users), and the BS serves only one of these N queues (the queue corresponding to the best user) at any time. Thus the delay in transmitting a particular packet to all the users is given by the delay in transmitting that packet from all of the N queues at the BS. In our analysis, we benefit from the concept of worst case delay proposed in [4] for analyzing the delay in unicast networks. In [4], the authors characterized the worst case delay by restating their problem as the "coupon collector problem" which has been studied extensively in the mathematics literature [38, 39].

In the coupon collector problem, the users are assumed to have coupons and the transmitter is the collector that selects one of the users randomly (with a uniform distribution) and collects his coupon. The problem is to characterize the average number of trials required to ensure that the collector collects m coupons from all the users. Our queuing problem is analogous to the coupon collector problem with the only *fundamental* difference being that the size of the coupons is time-varying in our problem due to rate adaptation (the detailed analysis is presented in the proofs). The following theorem establishes the average throughput and delay achieved by the best user scheduler.

Theorem 4. *The average throughput and the average delay of the best user scheduler scale as*

$$R_{tot} \;=\; \Theta(\log\log N) \qquad \textit{and} \qquad D \;=\; \Omega\left(N \log N\right), \tag{2.2}$$

with the number of users N.

Proof. Refer Appendix A.2. □

From Theorems 3 and 4, one can conclude that *maximally* exploiting the multi-user diversity yields higher throughput gains than *maximally* exploiting the wireless multicast gain, when there are a large number of users in the system. This throughput gain, however, is obtained at the expense of a higher delay. This observation motivates the investigation of other static schedulers that achieve a better throughput-delay tradeoff.

Median User Scheduler

This scheme strikes a balance between exploiting the multi-user diversity and the wireless multicast gain offered by the system. At any instant, the BS chooses its

transmission rate such that the better half of the users in the system can successfully decode each transmission, i.e., the rate is adjusted based on the user whose instantaneous SNR is the median among the SNRs of all users. Thus $(N/2)$ users are targeted in each transmission and therefore, unlike the best user scheduler, this scheduler benefits from the wireless multicast gain. Moreover, unlike the worst user scheduler, the inherent multi-user diversity does not degrade the performance of this scheduler (since the instantaneous SNR of the median user does not degrade with N). Once the BS starts transmitting a packet, it keeps on repeating the same packet until it is successfully decoded by all the users in the system. Since the BS is assumed to have perfect channel knowledge, it can easily keep track whether or not the transmitted packet has been decoded by all the users in the system. Once the current packet reaches all the N users, the BS immediately starts transmitting the next packet in the same fashion. The BS needs to maintain only a single common queue that caters to all the users in the system.

One drawback of this scheme is that some of the users may receive redundant copies of the same packet (since any user who has decoded the packet has to wait until all the other users have decoded that packet). This redundancy leads to a reduction in the effective throughput achieved by this scheme. However, it is interesting to note that the median user scheduler achieves near-optimal scaling laws of both throughput and delay, as shown in the following theorem.

Theorem 5. *The average throughput and the average delay of the median user scheduler scale as*

$$R_{tot} \;=\; \Theta\left(\frac{N}{\log N}\right) \qquad \textit{and} \qquad D \;=\; \Theta\left(\log N\right), \tag{2.3}$$

with the number of users N.

Proof. Refer Appendix A.3. □

By comparing the results in Theorems 3, 4 and 5, it is clear that the proposed median user scheduler achieves a much superior throughput-delay tradeoff than the best and worst user schedulers, by striking a balance between exploiting multi-user diversity and multicast gain.

An upper bound on the average throughput of **any** scheduling scheme is given by

$$R_{tot} \;\leq\; \mathbb{E}\left[\sum_{i=1}^{N} \log\left(1+|h_i|^2 P\right)\right] \;=\; N\mathbb{E}\left[\log\left(1+|h_1|^2 P\right)\right] \;=\; \Theta(N). \tag{2.4}$$

Moreover, the average delay of **any** scheduling scheme can also be easily lower bounded as

$$D \;=\; \Omega(1). \tag{2.5}$$

Comparing these bounds with the results in Theorem 5, we find that the proposed median user scheduler achieves near-optimal scaling laws of both throughput and delay. In fact, the loss in both delay and throughput scaling laws, compared to the optimal values, is only a factor of $(\log N)$. However, this scheme requires perfect knowledge of the channel gains of all the users (CSI) at the BS, and hence entails a significant feedback requirement. We next propose a scheduling scheme that does not require perfect CSI at the BS and entails very minimal feedback.

2.2.2 Incremental Redundancy Multicast

In this section, we relax the memoryless decoding requirement and propose a scheme that employs a higher complexity incremental redundancy encoding/decoding strategy to achieve a better throughput-delay tradeoff than the median user scheduler. The proposed scheme is an extension of the incremental redundancy scheme given by

Caire *et al* in [22]. An information sequence of b bits is encoded into a codeword of length LM, where M refers to the rate constraint. The first L bits of the codeword are transmitted in the first attempt. If a user is unable to successfully decode the transmission, it sends back an ARQ request to the BS. If the BS receives an ARQ request from any of the users, it transmits the next L bits of the same codeword in the next attempt. After each transmission attempt, the users try to decode the transmitted information sequence using the received sequence in that attempt *jointly* with the received sequences in all previous ARQ attempts. This process continues until either all N users successfully decode the information sequence or the rate constraint M is violated[6]. Then the codeword corresponding to the next b information bits is transmitted in the same fashion.

In this scheme, similar to the median user scheduler proposed earlier, even if some of the users successfully decode the packet in very few attempts, they still have to wait until all the N users successfully receive the packet before any new packet is transmitted to them by the BS. This sub-optimality of the proposed schemes results in significant complexity reduction by avoiding the use of superposition coding and successive decoding. However, unlike the median user scheme, this scheme does not require the knowledge of perfect CSI at the BS. The BS only needs to know when to stop transmission of the current codeword. Hence the feedback required is minimal. The following result establishes the superior throughput-delay tradeoff achieved by this scheme compared with the median user scheduler.

[6]In our analysis, we consider the unconstrained case, where $M \to \infty$.

Theorem 6. *The average throughput and the average delay of the incremental redundancy multicast scheduler scale as*

$$R_{tot} \;=\; \Theta\left(\frac{N \log\log N}{\log N}\right) \qquad \textit{and} \qquad D \;=\; \Theta\left(\frac{\log N}{\log\log N}\right), \tag{2.6}$$

with the number of users N.

Proof. Refer Appendix A.4. □

Thus, from Theorem 6 and the throughput and delay bounds in (2.4) and (2.5), it is clear that incremental redundancy multicast achieves near-optimal scaling laws of both throughput and delay. The loss in both delay and throughput scaling laws, compared to the optimal values, is only a factor of $(\log N/\log\log N)$. In this approach, the BS again needs to maintain only a single queue that serves all the users in the system. This approach, however, entails added complexity in the incremental redundancy encoding and the storage and joint decoding of all the observations.

2.2.3 Cooperative Multicast

In this section, we demonstrate the benefits of user cooperation and quantify the tremendous gains that can be achieved by allowing the users to cooperate with each other. In particular, we propose a cooperation scheme that minimizes the delay while achieving the optimal scaling law of the throughput. This scheme is similar in spirit to the Decode-and-Forward scheme [8] and is divided into two stages. In the first half of each time slot, the BS transmits the packet to one half of the users in the system (i.e., the median user scheduler). During the next half of the slot, the BS remains silent. Meanwhile all the users that successfully decoded the packet in the first half of the slot cooperate with each other and transmit the packet to the other

$(N/2)$ users in the system. This is equivalent to a transmission from a transmitter equipped with $(N/2)$ transmit antennas to the worst user in a group of $(N/2)$ users. If R_{s1} and R_{s2} are the rates supported in the first and second stage respectively, then the actual transmission rate is chosen to be $\min\{R_{s1}, R_{s2}\}$ in both stages of the cooperation scheme. Note that the rate R_{s2} is chosen such that the information can be successfully decoded even by the worst of the remaining $(N/2)$ users. Here, we note that this scheme requires the BS to know the CSI of the inter-user channels. The scheme, however, does not require the users to have such transmitter CSI (i.e., in the second stage the users cooperate blindly by using i.i.d. random coding). The average throughput of the proposed cooperation scheme is thus given by

$$R_{tot} = \left(\frac{N}{2}\right) \mathbb{E}\left[\min\{R_{s1}, R_{s2}\}\right].$$

The following theorem establishes the optimality of the proposed scheme, in terms of both delay and throughput scaling laws.

Theorem 7. *The proposed cooperative multicast scheduler achieves the optimal scaling laws of both delay and throughput. In particular, the average throughput and the average delay of this scheduler scale as*

$$R_{tot} \;=\; \Theta(N) \qquad \textit{and} \qquad D \;=\; \Theta(1), \tag{2.7}$$

with the number of users N. Here we assume that the inter-user channels have the same fading statistics as the channels between the base station and users, and the **total** *transmitted power is upper bounded by P.*

Proof. Refer Appendix A.5. □

Scheme	Average Throughput R_{tot}	Average Delay D
Worst User Scheduler	$\Theta(1)$	$\Theta(N)$
Best User Scheduler	$\Theta(\log\log N)$	$\Omega(N\log N)$
Median User Scheduler	$\Theta\left(\frac{N}{\log N}\right)$	$\Theta(\log N)$
IR Multicast	$\Theta\left(\frac{N\log\log N}{\log N}\right)$	$\Theta\left(\frac{\log N}{\log\log N}\right)$
Cooperative Multicast	$\Theta(N)$	$\Theta(1)$

Table 2.1: Comparison of the throughput-delay tradeoffs achieved by the proposed multicast schedulers

The price for this optimal performance is the added complexity needed to 1) equip every user terminal with a transmitter, 2) decode/re-encode the information at each cooperating user terminal, and 3) inform the BS with perfect CSI of the inter-user channels. In Table 2.1, we provide a comparison of the throughput-delay tradeoffs achieved by each of the proposed multicast schedulers.

2.3 Multi-Transmit Antenna Gain

The performance of the best and worst user schedulers proposed in Section 2.2.1 depends on the spread of the fading distribution of the users' channels. For exploiting

significant multi-user diversity gains, the distribution needs to be well-spread out [3]. The lower the spread of the distribution, the lesser the multi-user diversity gain (or loss as shown in the following). To illustrate this point, we consider a scenario where the BS is equipped with L transmit antennas. We assume that the BS has knowledge of only the total effective SNR at any particular user and does not know the individual channel gains from each transmit antenna to that user. Under this assumption, the BS just distributes the available power equally among all the L transmit antennas. Thus the effective fading power gains follow a normalized Chi-square distribution with $2L$ degrees of freedom. Note that the fading power gains were exponentially distributed (Chi-square with 2 degrees of freedom) in the single transmit antenna case. We now characterize the asymptotic throughput scaling laws of the best and worst user schedulers for this multi-transmit antenna scenario. Note that all the results in this section are derived for the case where L is a constant and does not scale with N.

2.3.1 Worst User Scheduler

The average throughput of the worst user scheduler is given by

$$R_{tot} = N\mathbb{E}\left[\log\left(1+|\chi_{min}|^2 P\right)\right],$$

where $|\chi_{min}|^2 = \min_{i=1}^{N} |\chi_i|^2$, and $|\chi_i|^2$ corresponds to the effective fading power gain at the i^{th} user that follows a normalized Chi-square distribution with $2L$ degrees of freedom and whose distribution function is given by

$$F(x) = 1 - e^{-Lx}\left(\sum_{k=0}^{L-1}\frac{(Lx)^k}{k!}\right), \quad x \geq 0. \tag{2.8}$$

Lemma 8. *When the base station is equipped with L transmit antennas, the average throughput of the worst user scheduler scales as*

$$R_{tot} \;=\; \Theta\left(N^{\left(\frac{L-1}{L}\right)}\right). \tag{2.9}$$

Proof. Refer Appendix A.6. □

From (2.9), it is clear that the throughput scaling law of the worst user scheduler improves as L increases. This throughput improvement is expected since the performance of the worst user scheduler is known to be *degraded* by the tail of the fading distribution. Hence as L increases, the spread of the fading distribution decreases, and consequently, the inherent multi-user diversity has a reduced effect on the performance of the scheduler. This leads to a rise in the average throughput of the worst user scheduler from $\Theta(1)$ for the single transmit antenna case to almost $\Theta(N)$ for large values of L. Thus the worst user scheduler achieves a near-optimal throughput scaling for large values of L.

2.3.2 Best User Scheduler

The average throughput of the best user scheduler is given by

$$R_{tot} = \mathbb{E}\left[\log\left(1+|\chi_{max}|^2 P\right)\right],$$

where $|\chi_{max}|^2 = \max_{i=1}^{N} |\chi_i|^2$.

Lemma 9. *When the base station is equipped with L transmit antennas, the average throughput of the best user scheduler scales as*

$$R_{tot} \;=\; \Theta\left(\log\left(1+\frac{\log N+(L-1)\log\log N}{L}\right)\right). \tag{2.10}$$

Proof. Refer Appendix A.7. □

Since the best user scheduler leverages multi-user diversity to enhance the throughput, one can see from (2.10) that its throughput decreases as L increases.

It is interesting to note that even when the BS is equipped with only 2 transmit antennas ($L = 2$), the throughput of the worst user scheduler is significantly higher than that of the best user scheduler. This is contrary to our conclusion in Section 2.2.1 for the single transmit antenna case, where we showed that the best user scheduler outperforms the worst user scheduler. Since the multi-user diversity gain decreases as L increases (due to channel hardening [40]), the wireless multicast gain starts to dominate the achieved throughput for large values of L, which accounts for the better performance of the worst user scheduler.

2.4 Numerical Results

We now present simulation results that validate our theoretical claims. Our results were obtained through Monte-Carlo simulations and were averaged over 10000 iterations. The power constraint P is taken to be unity. A comparison of the throughput of all the scheduling schemes proposed in Section 2.2 is presented in Fig. 2.1 for increasing values of N. Although the incremental redundancy scheme outperforms the cooperative multicast scheme for small values of N, it is clear that the latter eventually outperforms the former for large values of N ($N > 45$). The corresponding delay-comparison of the proposed schedulers is presented in Fig. 2.2. It is clear that the simulation results follow the same trends that were predicted by our asymptotic analysis. In Fig. 2.3, we present a comparison of the throughput of the best and worst

user schedulers for the multi-transmit antenna case discussed in Section 2.3. As predicted by our analysis, the worst user scheduler outperforms the best user scheduler even for the $L = 2$ case. It is also clear that the throughput scaling of the worst user scheduler is almost linear for large values of L ($L \geq 10$). Finally, we observe that the utility of our asymptotic analysis is manifested in its accurate predictions even with the relatively small number of users used in our simulations (i.e., in the order of $N = 10$).

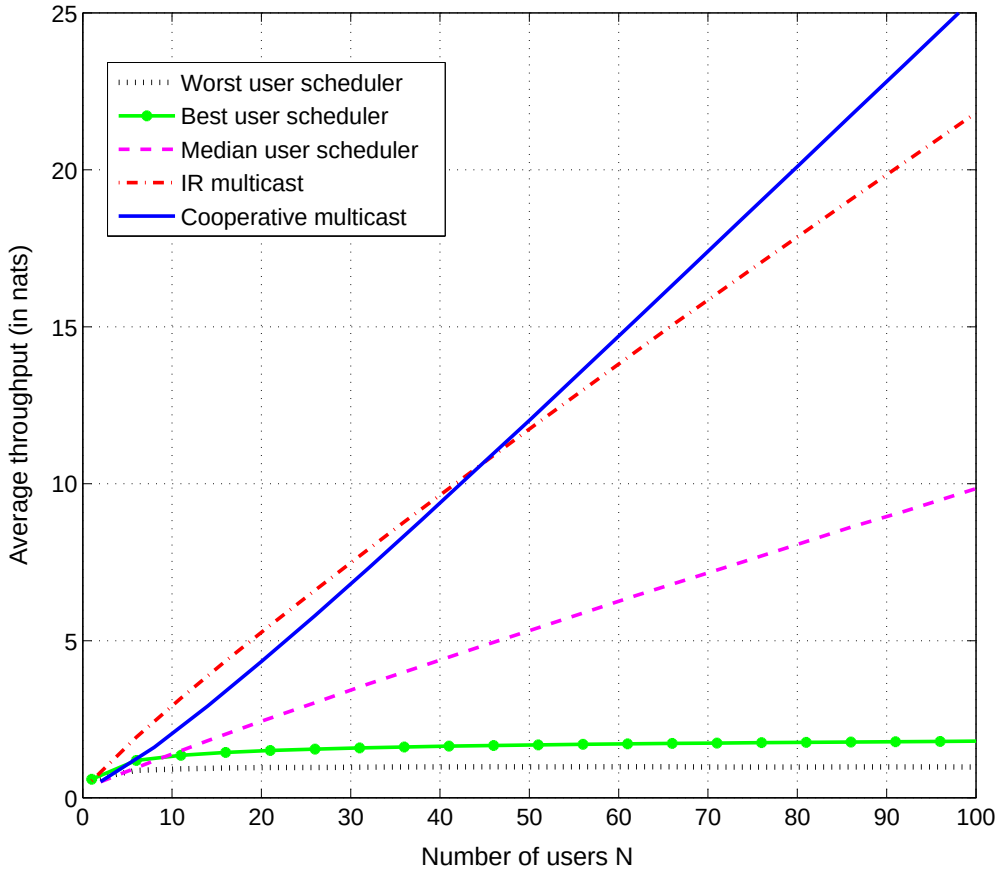

Figure 2.1: Comparison of the average throughput of the proposed static multicast schedulers (best, worst and median user schemes), incremental redundancy multicast and cooperative multicast

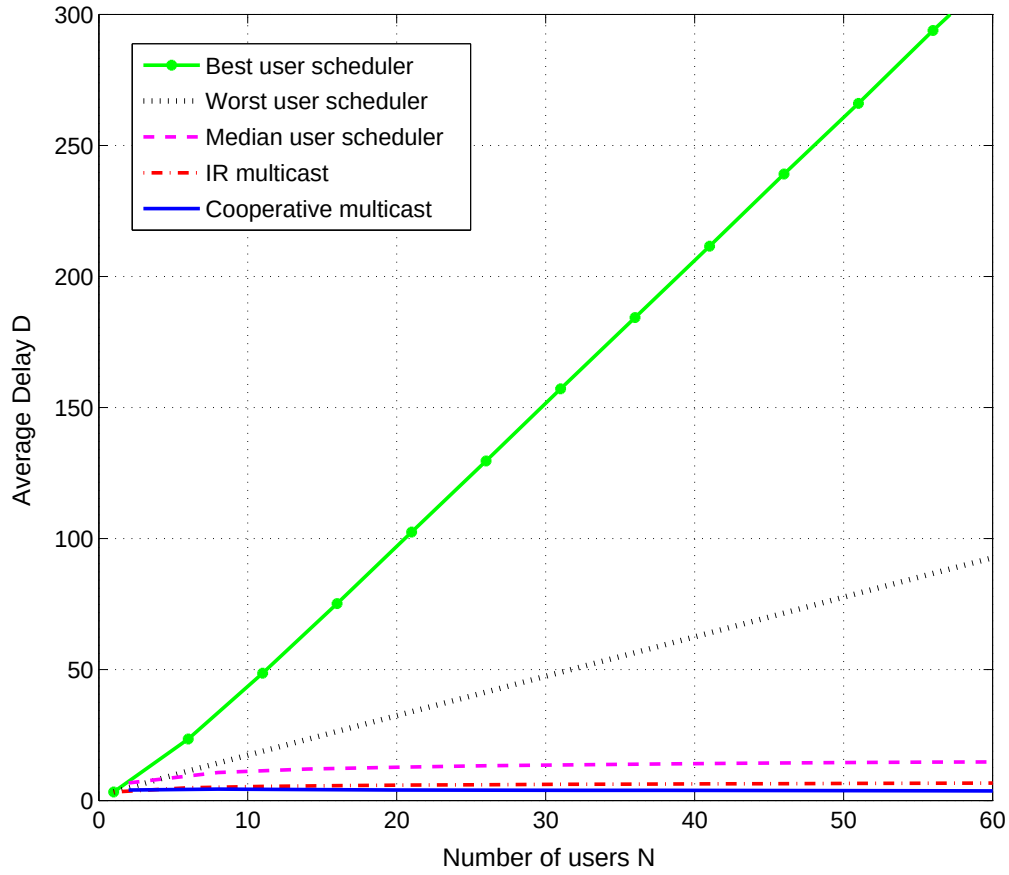

Figure 2.2: Comparison of the average delay of the proposed static multicast schedulers (best, worst and median user schemes), incremental redundancy multicast and cooperative multicast

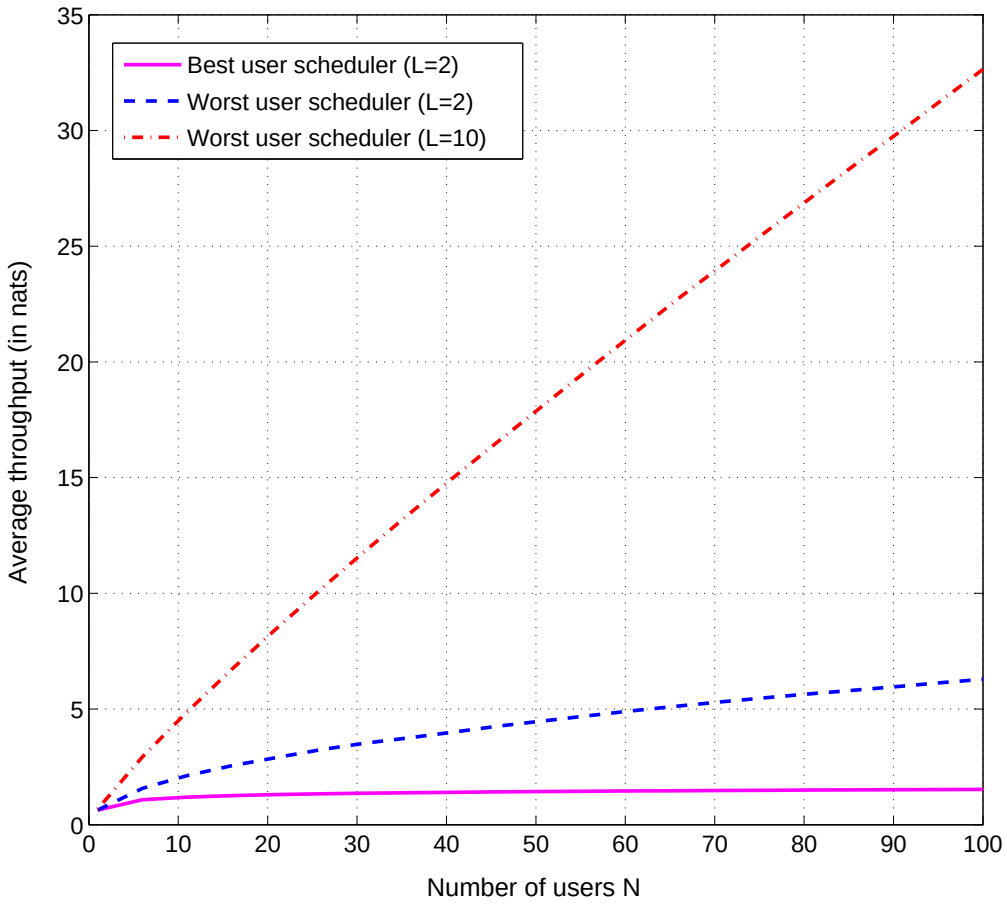

Figure 2.3: Comparison of the throughput of the best and worst user schedulers for the multi-transmit antenna scenario, with L transmit antennas at the base station

CHAPTER 3

FEEDBACK FOR SECRECY: THE FADING EAVESDROPPER CHANNEL

In this chapter, we reveal the importance of feedback in facilitating secure communication over wireless channels. We consider the secure transmission of information over an ergodic fading channel in the presence of an eavesdropper. We adopt the notion of information-theoretic secrecy, which was first introduced by Shannon in [11]. This strong notion of secrecy does not rely on any assumptions on the computational resources of the eavesdropper. More specifically, perfect information-theoretic secrecy requires that $I(W;Z) = 0$, i.e., the signal Z received by the eavesdropper does not provide any additional information about the transmitted message W.

Shannon considered a scenario where both the legitimate receiver and the eavesdropper have direct access to the transmitted signal [11]. Under this model, he proved a negative result implying that the achievability of perfect secrecy requires the entropy of the private key K, used to encrypt the message W, to be larger than or equal to the entropy of the message itself (i.e., $H(K) \geq H(W)$ for perfect secrecy). However, it was later shown by Wyner in [12] that this negative result was a consequence of the over-restrictive model used in [11]. Wyner introduced the wiretap channel which accounts for the difference in the two noise processes, as observed by the destination

and wiretapper. In this model, the wiretapper has no computational limitations and is assumed to know the codebook used by the transmitter. Under the assumption that the wiretapper's signal is a degraded version of the destination's signal, Wyner characterized the tradeoff between the information rate to the destination and the level of ignorance at the wiretapper (measured by its equivocation), and showed that it is possible to achieve a non-zero secrecy capacity. This work was later extended to non-degraded channels by Csiszár and Körner [14], where it was shown that if the main channel is less noisy or more capable than the wiretapper channel, then it is possible to achieve a non-zero secrecy capacity.

More recently, the effect of slow fading on the secrecy capacity was studied in [41, 42]. In these works, it is assumed that the fading is quasi-static which leads to an alternative definition of outage probability, wherein secure communications can be guaranteed only for the fraction of time when the main channel is stronger than the channel seen by the eavesdropper. This performance metric appears to have an operational significance only in delay sensitive applications with full Channel State Information (CSI). The absence of CSI sheds doubt on the operational significance of outage-based secrecy since it limits the ability of the source to know which parts of the message are decoded by the eavesdropper.

In this chapter, we focus on delay-tolerant applications which allow for the adoption of an ergodic version of the slow fading channel, instead of the outage-based formulation. Quite interestingly, we show that, under this model, one can achieve a perfectly secure non-zero rate even when the eavesdropper channel is more capable than the legitimate channel on the average.

3.1 System Model

The system model is illustrated in Fig. 3.1. The source S communicates with a destination D (legitimate receiver) in the presence of an eavesdropper E. During any coherence interval i, the signal received by the destination and the eavesdropper are given by, respectively

$$\begin{aligned} y(i) &= g_M(i)x(i) + w_M(i), \\ z(i) &= g_E(i)x(i) + w_E(i), \end{aligned}$$

where $g_M(i), g_E(i)$ are the channel gains from the source to the legitimate receiver (main channel) and the eavesdropper (eavesdropper channel) respectively, and $w_M(i)$, $w_E(i)$ represent the i.i.d additive Gaussian noise with unit variance at the destination and the eavesdropper respectively. We denote the fading power gains of the main and eavesdropper channels by $h_M(i) = |g_M(i)|^2$ and $h_E(i) = |g_E(i)|^2$ respectively. We assume that both channels experience block fading, where the channel gains remain constant during each coherence interval and change independently from one coherence interval to the next. The fading process is assumed to be ergodic with a bounded continuous distribution. Moreover, the fading coefficients of the destination and the eavesdropper in any coherence interval are assumed to be independent of each other. We further assume that the number of channel uses n_1 within each coherence interval is large enough to allow for invoking random coding arguments (this assumption is instrumental in our achievability proofs).

The source wishes to send a message $W \in \mathcal{W} = \{1, 2, \cdots, M\}$ to the destination. An (M, n) code consists of the following elements: 1) a stochastic encoder $f_n(.)$ at the

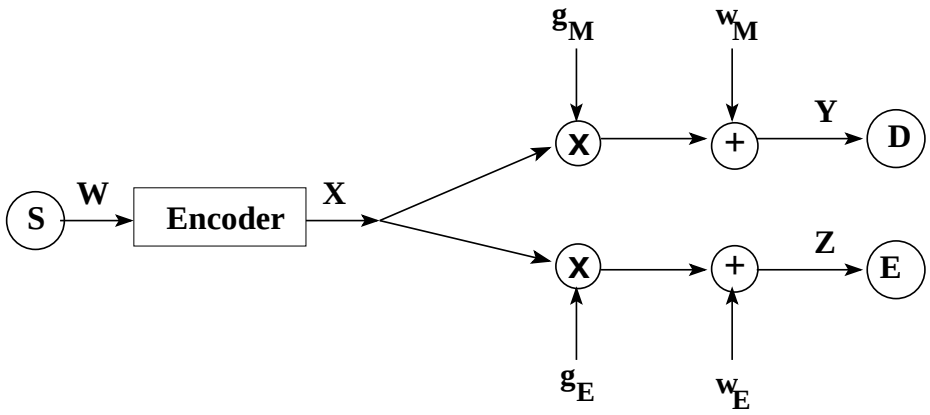

Figure 3.1: The Fading Channel with an Eavesdropper

source that maps the message[7] w to a codeword $x^n \in \mathcal{X}^n$, and 2) a decoding function ϕ: $\mathcal{Y}^n \to \mathcal{W}$ at the legitimate receiver. The average error probability of an (M, n) code at the legitimate receiver is defined as

$$P_e^n = \sum_{w \in \mathcal{W}} \frac{1}{M} \Pr(\phi(y^n) \neq w | w \text{ was sent}). \tag{3.1}$$

The equivocation rate R_e at the eavesdropper is defined as the entropy rate of the transmitted message conditioned on the available CSI and the channel outputs at the eavesdropper, i.e.,

$$R_e \triangleq \frac{1}{n} H(W | Z^n, h_M^n, h_E^n) \ , \tag{3.2}$$

where $h_M^n = \{h_M(1), \cdots, h_M(n)\}$ and $h_E^n = \{h_E(1), \cdots, h_E(n)\}$ denote the channel power gains of the legitimate receiver and the eavesdropper in n coherence intervals, respectively. It indicates the level of ignorance of the transmitted message W at the eavesdropper. The perfect secrecy rate R_s is said to be achievable if for any $\epsilon > 0$,

[7]The realizations of the random variables W, X, Y, Z are represented by w, x, y, z respectively.

there exists a sequence of codes $(2^{nR_s}, n)$ such that for any $n \geq n(\epsilon)$, we have

$$\begin{aligned} P_e^n &\leq \epsilon, \\ R_e &= \frac{1}{n} H(W|Z^n, h_M^n, h_E^n) \geq R_s - \epsilon. \end{aligned}$$

The secrecy capacity C_s is defined as the maximum achievable perfect secrecy rate, i.e.,

$$C_s \triangleq \sup_{P_e^n \leq \epsilon} R_s \ . \tag{3.3}$$

Throughout this chapter, we assume that the CSI is known at the destination perfectly. Based on the available CSI, the transmitter adapts its transmission power and rate to maximize the perfect secrecy rate subject to a long-term average power constraint $\bar{P}$.

3.2 Full CSI Feedback

Here we assume that at the beginning of each coherence interval, the transmitter knows the channel states of the legitimate receiver and the eavesdropper perfectly. When h_M and h_E are both known at the transmitter, one would expect the optimal scheme to allow for transmission only when $h_M > h_E$, and to adapt the transmitted power according to the instantaneous values of h_M and h_E. The following result formalizes this intuitive argument.

Theorem 10. *When the channel gains of both the legitimate receiver and the eavesdropper are known at the transmitter, the secrecy capacity is given by*

$$C_s^{(F)} = \max_{P(h_M,h_E)} \int_0^\infty \int_{h_E}^\infty \left[\log\left(\frac{1+h_M P(h_M,h_E)}{1+h_E P(h_M,h_E)}\right)\right] f(h_M) f(h_E) \mathrm{d}h_M \mathrm{d}h_E, \tag{3.4}$$

$$\textit{such that} \qquad \mathbb{E}\{P(h_M,h_E)\} \leq \bar{P}. \tag{3.5}$$

Proof. A detailed proof of achievability and the converse is provided in Appendix B.1. Here, we outline the scheme used in the achievability part. In this scheme, the source transmits information only when $h_M > h_E$, and uses the power allocation policy $P(h_M, h_E)$ that satisfies the average power constraint (3.5). Moreover, the codeword rate at each instant is set to be $\log\left(1 + h_M P(h_M, h_E)\right)$, which varies according to the instantaneous channel gains. The achievable perfect secrecy rate at any instant is then given by $[\log\left(1 + h_M P(h_M, h_E)\right) - \log\left(1 + h_E P(h_M, h_E)\right)]^+$. Averaging over all fading realizations, we get the average achievable perfect secrecy rate to be

$$\begin{aligned} R_s^{(F)} &= \iint \left[\log\left(\frac{1 + h_M P(h_M, h_E)}{1 + h_E P(h_M, h_E)}\right)\right]^+ f(h_M) f(h_E) \mathrm{d}h_M \mathrm{d}h_E \\ &= \int_0^\infty \int_{h_E}^\infty \left[\log\left(\frac{1 + h_M P(h_M, h_E)}{1 + h_E P(h_M, h_E)}\right)\right] f(h_M) f(h_E) \mathrm{d}h_M \mathrm{d}h_E \ . \end{aligned}$$

One can then optimize over all feasible power control policies $P(h_M, h_E)$ to maximize the perfect secrecy rate. □

We now derive the optimal power allocation policy that achieves the secrecy capacity under the full CSI assumption. It is easy to check that the objective function is concave in $P(h_M, h_E)$, and hence, by using the Lagrangian maximization approach for solving (3.4), we get the following optimality condition

$$\frac{\partial R_s^{(F)}}{\partial P(h_M, h_E)} = \frac{h_M}{1 + h_M P(h_M, h_E)} - \frac{h_E}{1 + h_E P(h_M, h_E)} - \lambda = 0,$$

whose solution is

$$P(h_M, h_E) = \frac{1}{2}\left[\sqrt{\left(\frac{1}{h_E} - \frac{1}{h_M}\right)^2 + \frac{4}{\lambda}\left(\frac{1}{h_E} - \frac{1}{h_M}\right)} - \left(\frac{1}{h_M} + \frac{1}{h_E}\right)\right]. \quad (3.6)$$

If for some (h_M, h_E), the value of $P(h_M, h_E)$ obtained from (3.6) is negative, then it follows from the concavity of the objective function w.r.t. $P(h_M, h_E)$ that the optimal

value of $P(h_M, h_E)$ is 0. Thus the optimal power allocation policy at the transmitter is given by

$$P(h_M, h_E) = \frac{1}{2}\left[\sqrt{\left(\frac{1}{h_E} - \frac{1}{h_M}\right)^2 + \frac{4}{\lambda}\left(\frac{1}{h_E} - \frac{1}{h_M}\right)} - \left(\frac{1}{h_M} + \frac{1}{h_E}\right)\right]^+, \qquad (3.7)$$

where $[x]^+ = \max\{0, x\}$, and the parameter λ is a constant that satisfies the power constraint in (3.5) with equality. The secrecy capacity is then determined by substituting this optimal power allocation policy for $P(h_M, h_E)$ in (3.4).

3.3 Main Channel CSI Feedback

In this section, we assume that at the beginning of each coherence interval, the transmitter only knows the CSI of the main channel (legitimate receiver).

3.3.1 Optimal Power Allocation

We first characterize the secrecy capacity under this scenario in the following theorem.

Theorem 11. *When only the channel gain of the legitimate receiver is known at the transmitter, the secrecy capacity is given by*

$$C_s^{(M)} = \max_{P(h_M)} \iint \left[\log\left(\frac{1 + h_M P(h_M)}{1 + h_E P(h_M)}\right)\right]^+ f(h_M) f(h_E) \mathrm{d}h_M \mathrm{d}h_E \,, \qquad (3.8)$$

$$\textit{such that} \qquad \mathbb{E}\{P(h_M)\} \le \bar{P}. \qquad (3.9)$$

Proof. A detailed proof is provided in Appendix B.2. We use the following **variable rate** transmission scheme to show achievability. During a coherence interval with main channel fading state h_M, the transmitter transmits codewords at rate $\log(1 + h_M P(h_M))$ with power $P(h_M)$. This variable rate scheme relies on the assumption of

large coherence intervals and ensures that when $h_E > h_M$, the mutual information between the source and the eavesdropper is upper bounded by $\log(1 + h_M P(h_M))$. When $h_E \leq h_M$, this mutual information will be $\log(1 + h_E P(h_M))$. Averaging over all the fading states, the average rate of the main channel is given by

$$\iint \log\left(1 + h_M P(h_M)\right) f(h_M) f(h_E) \mathrm{d}h_M \mathrm{d}h_E,$$

while the information accumulated at the eavesdropper is

$$\iint \log\left(1 + \min\{h_M, h_E\} P(h_M)\right) f(h_M) f(h_E) \mathrm{d}h_M \mathrm{d}h_E.$$

Hence for a given power control policy $P(h_M)$, the achievable perfect secrecy rate is given by

$$R_s^{(M)} \;=\; \iint \left[\log\left(\frac{1 + h_M P(h_M)}{1 + h_E P(h_M)}\right)\right]^+ f(h_M) f(h_E) \mathrm{d}h_M \mathrm{d}h_E. \tag{3.10}$$

One can then optimize over all feasible power control policies $P(h_M)$ to maximize the perfect secrecy rate. Here we note that our secure message is *hidden* across the different fading states. □

We now derive the optimal power allocation policy that achieves the secrecy capacity under the main channel CSI assumption. Similar to Theorem 10, the objective function under this case is also concave, and using the Lagrangian maximization approach for solving (3.8), we get the following optimality condition.

$$\frac{\partial R_s^{(M)}}{\partial P(h_M)} \;=\; \frac{h_M \Pr\left(h_E \leq h_M\right)}{1 + h_M P(h_M)} - \int_0^{h_M} \left(\frac{h_E}{1 + h_E P(h_M)}\right) f(h_E) \mathrm{d}h_E - \lambda \;=\; 0,$$

where λ is a constant that satisfies the power constraint in (3.9) with equality. For any main channel fading state h_M, the optimal transmit power level $P(h_M)$ is determined from the above equation. If the obtained power level turns out to be negative, then the

optimal value of $P(h_M)$ is equal to 0. This follows from the concavity of the objective function in (3.8) w.r.t. $P(h_M)$. The solution to this optimization problem depends on the distributions $f(h_M)$ and $f(h_E)$. In the following, we focus on the Rayleigh fading scenario with $\mathbb{E}\{h_M\} = \overline{\gamma}_M$ and $\mathbb{E}\{h_E\} = \overline{\gamma}_E$ in detail. With Rayleigh fading, the objective function in (3.8) simplifies to

$$\begin{aligned} C_s^{(M)} &= \max_{P(h_M)} \int_0^\infty \Big[\left(1 - e^{-(h_M/\overline{\gamma}_E)}\right) \log\left(1 + h_M P(h_M)\right) - \\ &\qquad \int_0^{h_M} \log\left(1 + h_E P(h_M)\right) \frac{1}{\overline{\gamma}_E} e^{-(h_E/\overline{\gamma}_E)} \mathrm{d}h_E \Big] \frac{1}{\overline{\gamma}_M} e^{-(h_M/\overline{\gamma}_M)} \mathrm{d}h_M \\ &= \max_{P(h_M)} \int_0^\infty \Bigg[\log\left(1 + h_M P(h_M)\right) - \exp\left(\frac{1}{\overline{\gamma}_E P(h_M)}\right) \left(\mathrm{Ei}\left(\frac{1}{\overline{\gamma}_E P(h_M)}\right) - \right. \\ &\qquad \left. \mathrm{Ei}\left(\frac{h_M}{\overline{\gamma}_E} + \frac{1}{\overline{\gamma}_E P(h_M)}\right) \right) \Bigg] \frac{1}{\overline{\gamma}_M} e^{-(h_M/\overline{\gamma}_M)} \mathrm{d}h_M, \end{aligned} \tag{3.11}$$

where

$$\mathrm{Ei}(x) = \int_x^\infty \frac{e^{-t}}{t}\, \mathrm{d}t\ .$$

Specializing the optimality conditions to the Rayleigh fading scenario, it can be shown that the power level of the transmitter at any fading state h_M is obtained by solving the equation

$$\begin{aligned} &\left(1 - e^{-(h_M/\overline{\gamma}_E)}\right) \left(\frac{h_M}{1 + h_M P(h_M)}\right) - \lambda - \frac{\left(1 - e^{-(h_M/\overline{\gamma}_E)}\right)}{P(h_M)} + \\ &\qquad \frac{\exp\left(\frac{1}{\overline{\gamma}_E P(h_M)}\right)}{\overline{\gamma}_E (P(h_M))^2} \left[\mathrm{Ei}\left(\frac{1}{\overline{\gamma}_E P(h_M)}\right) - \mathrm{Ei}\left(\frac{h_M}{\overline{\gamma}_E} + \frac{1}{\overline{\gamma}_E P(h_M)}\right) \right] = 0. \end{aligned}$$

If there is no positive solution to this equation for a particular h_M, then we set $P(h_M) = 0$. The secrecy capacity is then determined by substituting this optimal power allocation policy for $P(h_M)$ in (3.11).

We observe that, unlike the traditional ergodic fading scenario, achieving the optimal performance under a security constraint relies heavily on using a variable

rate transmission strategy. This can be seen by evaluating the performance of a constant rate strategy where a single codeword is interleaved across infinitely many fading realizations. This interleaving will result in the eavesdropper *gaining more information*, than the destination, when its channel is better than the main channel, thereby yielding a perfect secrecy rate that is strictly smaller than that in (3.10). It is easy to see that the achievable perfect secrecy rate of the constant rate scheme, assuming a Gaussian codebook, is given by

$$\max_{P(h_M)} \iint \left[\log\left(\frac{1+h_M P(h_M)}{1+h_E P(h_M)}\right)\right] f(h_M) f(h_E) \mathrm{d}h_M \mathrm{d}h_E \ ,$$

$$\text{such that} \qquad \mathbb{E}\{P(h_M)\} \leq \bar{P}.$$

Unlike the two previous optimization problems, the objective function in this optimization problem is not a concave function of $P(h_M)$. Using the Lagrangian formulation, we only get the following *necessary* Karush-Kuhn-Tucker (KKT) conditions for the optimal point.

$$\begin{aligned} P(h_M)\left[\lambda - \frac{h_M}{1+h_M P(h_M)} + \int \left(\frac{h_E}{1+h_E P(h_M)}\right) f(h_E)\mathrm{d}h_E\right] &= 0, \\ \lambda \ \geq \ \frac{h_M}{1+h_M P(h_M)} - \int \left(\frac{h_E}{1+h_E P(h_M)}\right) f(h_E)\mathrm{d}h_E, & \\ \mathbb{E}\{P(h_M)\} \ = \ \bar{P}. & \end{aligned} \tag{3.12}$$

3.3.2 On/Off Power Control

We now propose a transmission policy wherein the transmitter sends information only when the channel gain of the legitimate receiver h_M exceeds a pre-determined constant threshold $\tau > 0$. Moreover, when $h_M > \tau$, the transmitter always uses the same power level P. However, it is crucial to adapt the rate of transmission

instantaneously as $\log(1+Ph_M)$ with h_M. It is clear that for an average power constraint $\bar{P}$, the constant power level used for transmission will be

$$P \;=\; \frac{\bar{P}}{\Pr(h_M > \tau)} \;.$$

Using a similar argument as in the achievable part of Theorem 11, we get the perfect secrecy rate achieved by the proposed scheme, using Gaussian inputs, as

$$R_s^{(CP)} \;=\; \int_0^\infty \int_\tau^\infty \left[\log\left(\frac{1+h_M P}{1+h_E P}\right)\right]^+ f(h_M) f(h_E) \mathrm{d}h_M \mathrm{d}h_E \;.$$

Specializing to the Rayleigh fading scenario, we get

$$P \;=\; \frac{\bar{P}}{\Pr(h_M > \tau)} \;=\; \bar{P} e^{(\tau/\overline{\gamma}_M)} \;,$$

and the achievable perfect secrecy rate simplifies to

$$R_s^{(CP)} \;=\; \int_\tau^\infty \int_0^{h_M} \left[\log\left(\frac{1+h_M \bar{P} e^{(\tau/\overline{\gamma}_M)}}{1+h_E \bar{P} e^{(\tau/\overline{\gamma}_M)}}\right)\right] \frac{1}{\overline{\gamma}_M} e^{-(h_M/\overline{\gamma}_M)} \frac{1}{\overline{\gamma}_E} e^{-(h_E/\overline{\gamma}_E)} \mathrm{d}h_E \mathrm{d}h_M \;,$$

which then simplifies to

$$\begin{aligned} R_s^{(CP)} \;=\; & e^{-(\tau/\overline{\gamma}_M)} \log\left(1+\tau \bar{P} e^{(\tau/\overline{\gamma}_M)}\right) + \exp\left(\frac{1}{\overline{\gamma}_M \bar{P} e^{(\tau/\overline{\gamma}_M)}}\right) \mathrm{Ei}\left(\frac{\tau}{\overline{\gamma}_M} + \frac{1}{\overline{\gamma}_M \bar{P} e^{(\tau/\overline{\gamma}_M)}}\right) \\ & + \exp\left(\frac{1}{\overline{\gamma}_E \bar{P} e^{(\tau/\overline{\gamma}_M)}} - \frac{\tau}{\overline{\gamma}_M}\right) \left[\mathrm{Ei}\left(\frac{\tau}{\overline{\gamma}_E} + \frac{1}{\overline{\gamma}_E \bar{P} e^{(\tau/\overline{\gamma}_M)}}\right) - \mathrm{Ei}\left(\frac{1}{\overline{\gamma}_E \bar{P} e^{(\tau/\overline{\gamma}_M)}}\right)\right] \\ & - \exp\left(\frac{\left[\frac{1}{\overline{\gamma}_M} + \frac{1}{\overline{\gamma}_E}\right]}{\bar{P} e^{(\tau/\overline{\gamma}_M)}}\right) \mathrm{Ei}\left(\left[\frac{1}{\overline{\gamma}_M} + \frac{1}{\overline{\gamma}_E}\right]\left[\tau + \frac{1}{\bar{P} e^{(\tau/\overline{\gamma}_M)}}\right]\right) \;. \end{aligned}$$

One can then optimize over the threshold τ to get the maximum achievable perfect secrecy rate.

Finally, we establish the asymptotic optimality of this on/off scheme as the available average transmission power $\bar{P} \to \infty$. For the on/off power allocation policy, we have

$$R_s^{(CP)} = \lim_{\bar{P}\to\infty} \int_{\tau^*}^\infty \int_0^{h_M} \log\left(\frac{1+h_M P}{1+h_E P}\right) f(h_M) f(h_E) \mathrm{d}h_E \mathrm{d}h_M.$$

Taking $\tau^* = 0$, we get $P = \bar{P}$ and

$$\begin{aligned} R_s^{(CP)} &\geq \lim_{\bar{P}\to\infty} \int_0^\infty \int_0^{h_M} \log\left(\frac{(1/\bar{P})+h_M}{(1/\bar{P})+h_E}\right) f(h_M)f(h_E)\mathrm{d}h_E\mathrm{d}h_M \\ &\overset{(a)}{=} \int_0^\infty \int_0^{h_M} \lim_{\bar{P}\to\infty} \log\left(\frac{(1/\bar{P})+h_M}{(1/\bar{P})+h_E}\right) f(h_M)f(h_E)\mathrm{d}h_E\mathrm{d}h_M \\ &= \int_0^\infty \int_0^{h_M} \log\left(\frac{h_M}{h_E}\right) f(h_M)f(h_E)\mathrm{d}h_E\mathrm{d}h_M \\ &= \mathbb{E}_{\{h_M>h_E\}}\left\{\log\left(\frac{h_M}{h_E}\right)\right\}, \end{aligned} \tag{3.13}$$

where (a) follows from the Dominated Convergence Theorem, since

$$\left|\log\left(\frac{(1/\bar{P})+h_M}{(1/\bar{P})+h_E}\right)\right| \leq \left|\log\left(\frac{h_M}{h_E}\right)\right|, \qquad \forall \bar{P} \quad \text{when } h_M > h_E,$$

$$\text{and} \qquad \int_0^\infty \int_0^{h_M} \log\left(\frac{h_M}{h_E}\right) f(h_M)f(h_E)\mathrm{d}h_E\mathrm{d}h_M < \infty,$$

since $\mathbb{E}\{h_M\} < \infty$, $\left|\int_0^1 \log x \ \mathrm{d}x\right| = 1 < \infty$ and $f(h_M), f(h_E)$ are continuous and bounded.

Now under the full CSI assumption, we have

$$C_s^{(F)} = \mathbb{E}_{\{h_M>h_E\}}\left\{\log\left(\frac{\frac{1}{P(h_M,h_E)}+h_M}{\frac{1}{P(h_M,h_E)}+h_E}\right)\right\} \leq \mathbb{E}_{\{h_M>h_E\}}\left\{\log\left(\frac{h_M}{h_E}\right)\right\}. \tag{3.14}$$

From (3.13) and (3.14), it is clear that the proposed on/off power allocation policy that uses only the main channel CSI achieves the secrecy capacity under the full CSI assumption as $\bar{P} \to \infty$. Thus the absence of eavesdropper CSI at the transmitter does not reduce the secrecy capacity at high SNR values.

3.4 ARQ Feedback

In Sections 3.2 and 3.3, we assumed the presence of perfect CSI feedback to the transmitter. This is an idealistic feedback scenario whose performance serves as an upper bound on the performance of other practical feedback scenarios. In this

section, we assume that the transmitter does not know the CSI of both the main and eavesdropper channels. We focus on a simple ARQ feedback scenario and quantify the significant fraction of gains obtained though minimal ARQ feedback, as compared to that possible with perfect CSI feedback. We assume that the legitimate receiver is capable of perfectly feeding back ARQ bits to the source, which convey whether its decoding was successful or not. We further assume that the feedback channel is public and hence the ARQ bits transmitted by the legitimate receiver are also received by the eavesdropper.

We propose two transmission schemes, viz. Repetition ARQ (Rep-ARQ) and Incremental Redundancy ARQ (IR-ARQ), that exploit the ARQ feedback to facilitate secure communication. In the Rep-ARQ scheme, the transmitter repeats codewords until it receives an ACK from the legitimate receiver, while the destination employs Maximal Ratio Combining (MRC) to decode the transmitted message from the observations accumulated in all the feedback rounds. The destination feeds back an ACK only when its decoding is successful and feeds back a NACK otherwise. In the IR-ARQ scheme, the transmitter, upon receiving a NACK, generates a new codebook and transmits new redundancy bits instead of merely repeating the codeword. The destination employs joint decoding across the ARQ rounds to recover the transmitted message. It is important to note here that the transmissions in each ARQ round also *help* the eavesdropper to gain more information about the transmitted message, since it can also employ MRC or joint decoding techniques. However, while the ARQ feedback ensures that the decoding is **always** successful at the destination, the decoding at the eavesdropper might not always be successful. This *important* observation can be leveraged by the destination to gain an advantage over the eavesdropper and

achieve non-zero perfect secrecy rates, even when the eavesdropper channel is better than the main channel on the average.

We now characterize the perfect secrecy rates achieved by the proposed IR-ARQ and Rep-ARQ schemes in the following theorem.

Theorem 12. *The proposed IR-ARQ scheme achieves the perfect secrecy rate*

$$R_s^{(IR)} = \frac{\mathbb{E}\left[\left(R - \sum_{k=1}^{L} \log\left(1 + h_{E,k}P\right)\right)^+\right]}{\mathbb{E}[L]}, \tag{3.15}$$

while the Rep-ARQ scheme achieves

$$R_s^{(Rep)} = \frac{\mathbb{E}\left[\left(R - \log\left(1 + \sum_{k=1}^{L} h_{E,k}P\right)\right)^+\right]}{\mathbb{E}[L]}, \tag{3.16}$$

where L is a random variable denoting the number of transmission rounds required for the successful decoding of any particular block at the destination.

Proof. Refer Appendix B.3. □

We now focus on evaluating the achievable perfect secrecy rate for the Rep-ARQ scheme when the main and eavesdropper channel gains are Rayleigh distributed with $\mathbb{E}[h_M] = 1$ and $\mathbb{E}[h_E] = 1$. The probability distribution of the number of ARQ rounds L required for successful decoding is given by

$$\begin{aligned}
&\Pr(\text{No. of ARQ rounds} = \ell) \\
&= \Pr\left(\log\left(1 + \sum_{i=1}^{\ell-1} h_{M,i}P\right) < R \leq \log\left(1 + \sum_{i=1}^{\ell} h_{M,i}P\right)\right) \\
&= \Pr\left(\sum_{i=1}^{\ell-1} h_{M,i} < \frac{e^R - 1}{P} \leq \sum_{i=1}^{\ell} h_{M,i}\right) \\
&= \int_0^{C_1} f^{(\ell-1)}(x) \Pr\left(h_{M,\ell} \geq C_1 - x\right) \mathrm{d}x \ ,
\end{aligned}$$

where $C_1 = (e^R - 1)/P$, and $f^{(\ell)}(x)$ represents the Chi-square distribution of the random variable $\sum_{i=1}^{\ell} h_{M,i}$ with 2ℓ degrees of freedom, which is given by

$$f^{(\ell)}(x) \;=\; \frac{e^{-x}x^{(\ell-1)}}{(\ell-1)!}, \qquad x > 0.$$

Hence

$$\Pr(L=\ell) \;=\; \int_0^{C_1} \frac{e^{-x}x^{(\ell-2)}}{(\ell-2)!}\left(e^{-(C_1-x)}\right)\mathrm{d}x \;=\; \frac{e^{-C_1}C_1^{(\ell-1)}}{(\ell-1)!}\,.$$

The expected number of required ARQ rounds is thus given by

$$\mathbb{E}[L] \;=\; \sum_{\ell=1}^{\infty} \ell\left(\frac{e^{-C_1}C_1^{(\ell-1)}}{(\ell-1)!}\right) \;=\; 1 + C_1 \;=\; 1 + \left(\frac{e^R-1}{P}\right). \tag{3.17}$$

We now need to characterize the quantity

$$S \;=\; \mathbb{E}\left[\left(R - \log\left(1 + \sum_{i=1}^{L} h_{E,i}P\right)\right)^{+}\right],$$

where the expectation is taken with respect to both the number of ARQ rounds L and the eavesdropper channels $\{h_{E,i}\}$. Since L depends only on the distribution of the main channel, which is independent of the eavesdropper channel, we get

$$\begin{aligned} S \;&=\; \sum_{\ell=1}^{\infty} \Pr(L=\ell)\;\; \mathbb{E}_{\{h_E\}}\left[\left(R - \log\left(1 + \sum_{i=1}^{\ell} h_{E,i}P\right)\right)^{+}\right] \\ &=\; \sum_{\ell=1}^{\infty} \left(\frac{e^{-C_1}C_1^{(\ell-1)}}{(\ell-1)!}\right)\left[\int_0^{C_1} (R - \log(1+xP))\,\frac{e^{-x}x^{(\ell-1)}}{(\ell-1)!}\mathrm{d}x\right] \\ &=\; \sum_{\ell=1}^{\infty} \left(\frac{e^{-C_1}C_1^{(\ell-1)}}{(\ell-1)!}\right)\left\{R\left[1 - e^{-C_1}\left(\sum_{j=0}^{\ell-1}\frac{C_1^j}{j!}\right)\right] - \right. \\ &\qquad\qquad \left. \int_0^{C_1} \frac{\log(1+xP)e^{-x}x^{(\ell-1)}}{(\ell-1)!}\mathrm{d}x\right\}. \end{aligned} \tag{3.18}$$

Thus, from (3.16), the perfect secrecy rate achieved by the Rep-ARQ scheme is given by

$$R_s^{(Rep)} \;=\; \frac{S}{\mathbb{E}[L]}, \tag{3.19}$$

which can be obtained from (3.17) and (3.18).

3.5 Numerical Results

As an additional benchmark, we first obtain the performance when the transmitter does not have any knowledge of both the main and eavesdropper channels (only receiver CSI). In this scenario, the transmitter is unable to exploit rate/power adaptation and always transmits with power $\bar{P}$. It is straightforward to see that the achievable perfect secrecy rate in this scenario (using Gaussian inputs) is given by

$$\begin{aligned} R_s^{(R)} &= \left[\int_0^\infty \int_0^\infty \log\left(\frac{1+h_M\bar{P}}{1+h_E\bar{P}}\right) f(h_M)f(h_E)\mathrm{d}h_M\mathrm{d}h_E\right]^+ \\ &= \left[\int_0^\infty \log\left(1+h_M\bar{P}\right) f(h_M)\mathrm{d}h_M - \int_0^\infty \log\left(1+h_E\bar{P}\right) f(h_E)\mathrm{d}h_E\right]^+, \end{aligned}$$

which reduces to the following for the Rayleigh fading scenario

$$R_s^{(R)} = \left[\exp\left(\frac{1}{\overline{\gamma}_M\bar{P}}\right)\mathrm{Ei}\left(\frac{1}{\overline{\gamma}_M\bar{P}}\right) - \exp\left(\frac{1}{\overline{\gamma}_E\bar{P}}\right)\mathrm{Ei}\left(\frac{1}{\overline{\gamma}_E\bar{P}}\right)\right]^+.$$

Thus when $\overline{\gamma}_E \geq \overline{\gamma}_M$, $R_s^{(R)} = 0$. The results for the Rayleigh normalized-symmetric case ($\overline{\gamma}_M = \overline{\gamma}_E = 1$), under the different transmitter CSI assumptions, are presented in Fig. 3.2. It is clear that the performance of the on/off power control scheme is very close to the secrecy capacity (with only main channel CSI) for a wide range of SNRs and, as expected, approaches the secrecy capacities, under both the full CSI and main channel CSI assumptions, at high values of SNR. The performance of the constant rate scheme is much worse than the other schemes that employ rate adaptation. Here we note that the performance curve for the constant rate scheme might be a lower bound to the secrecy capacity (since the KKT conditions are necessary but not sufficient for non-convex optimization). We then consider an asymmetric scenario, wherein the eavesdropper channel is more capable than the main channel, with $\overline{\gamma}_M = 1$ and $\overline{\gamma}_E = 2$. The performance results for this scenario are plotted in Fig. 3.3. Again

it is clear from the plot that the performance of the on/off power control scheme is optimal at high values of SNR, and that rate adaptation schemes yield higher perfect secrecy rates than constant rate transmission schemes.

We now present simulation results for the proposed schemes that rely only on minimal ARQ feedback. We again consider a symmetric Rayleigh fading scenario with $\overline{\gamma}_M = \overline{\gamma}_E = 1$. A comparison of the perfect secrecy rates achieved by the proposed IR-ARQ and Rep-ARQ schemes, for different average SNR values, is provided in Fig. 3.4. At each SNR level, we first compute the perfect secrecy rates in (3.15) and (3.16) for different first round rates R, and then pick the optimal first round rate R^* that yields the highest perfect secrecy rate for each scheme. It is clear from the plot that IR-ARQ outperforms Rep-ARQ at all considered SNR values. We note that this fact is *not obvious* since the IR-ARQ transmissions, which perform better than Rep-ARQ transmissions for the scenario without secrecy constraints, help both the legitimate receiver and the eavesdropper. The achievable perfect secrecy rates for the asymmetric scenario ($\overline{\gamma}_M = 1$, $\overline{\gamma}_E = 2$) are provided in Fig. 3.5. Again it is clear from the plot that IR-ARQ is superior to Rep-ARQ for the considered range of SNRs. More importantly, both the proposed schemes achieve positive perfect secrecy rates even when the eavesdropper channel is better than the main channel on the average, thereby highlighting the positive impact of minimal ARQ feedback on the secrecy capacity of slow fading channels.

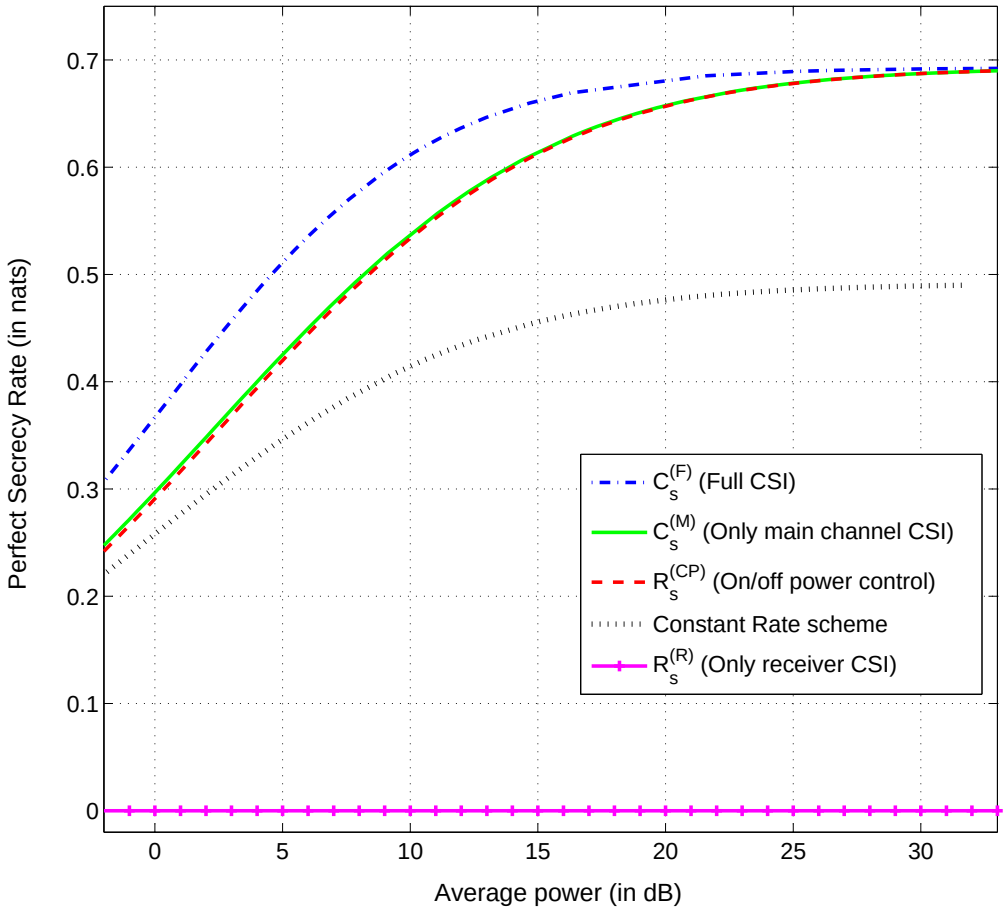

Figure 3.2: Comparison of the perfect secrecy rates achieved by the proposed schemes (under different assumptions on the available transmitter CSI) for the symmetric scenario $\overline{\gamma}_M = \overline{\gamma}_E = 1$

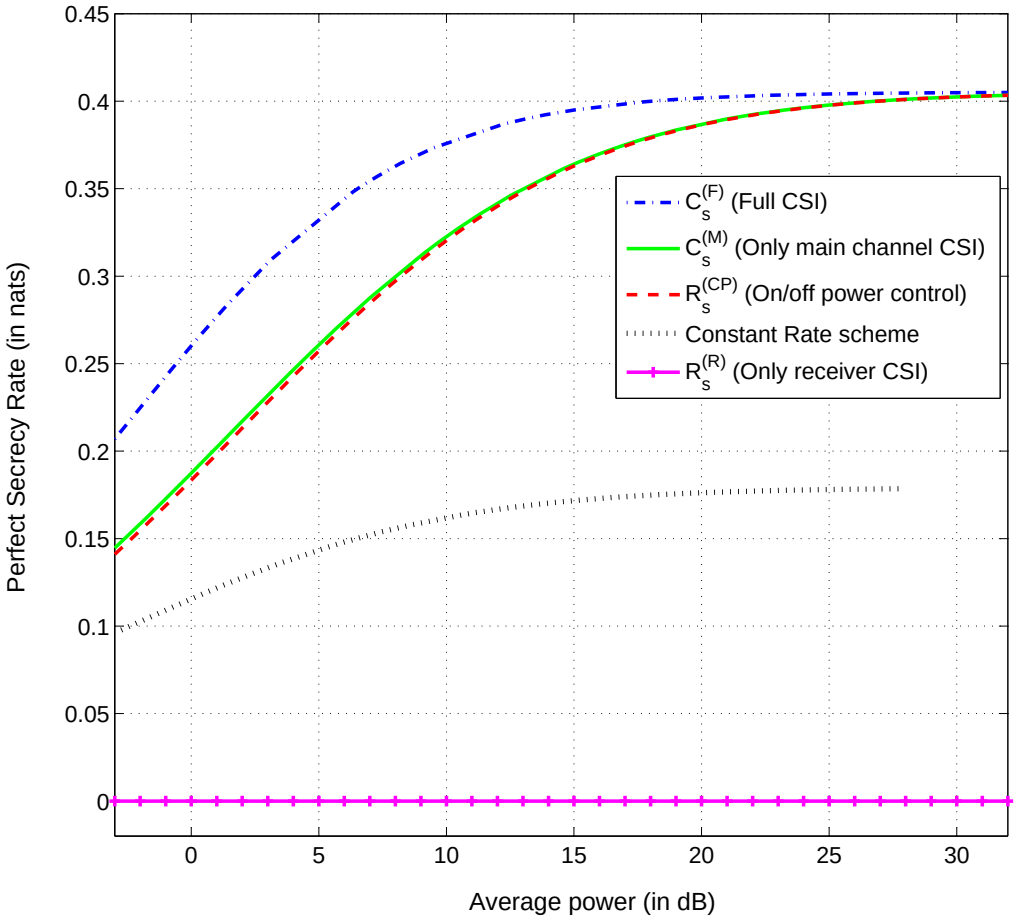

Figure 3.3: Comparison of the perfect secrecy rates achieved by the proposed schemes (under different assumptions on the available transmitter CSI) for the asymmetric scenario $\overline{\gamma}_M = 1$ and $\overline{\gamma}_E = 2$

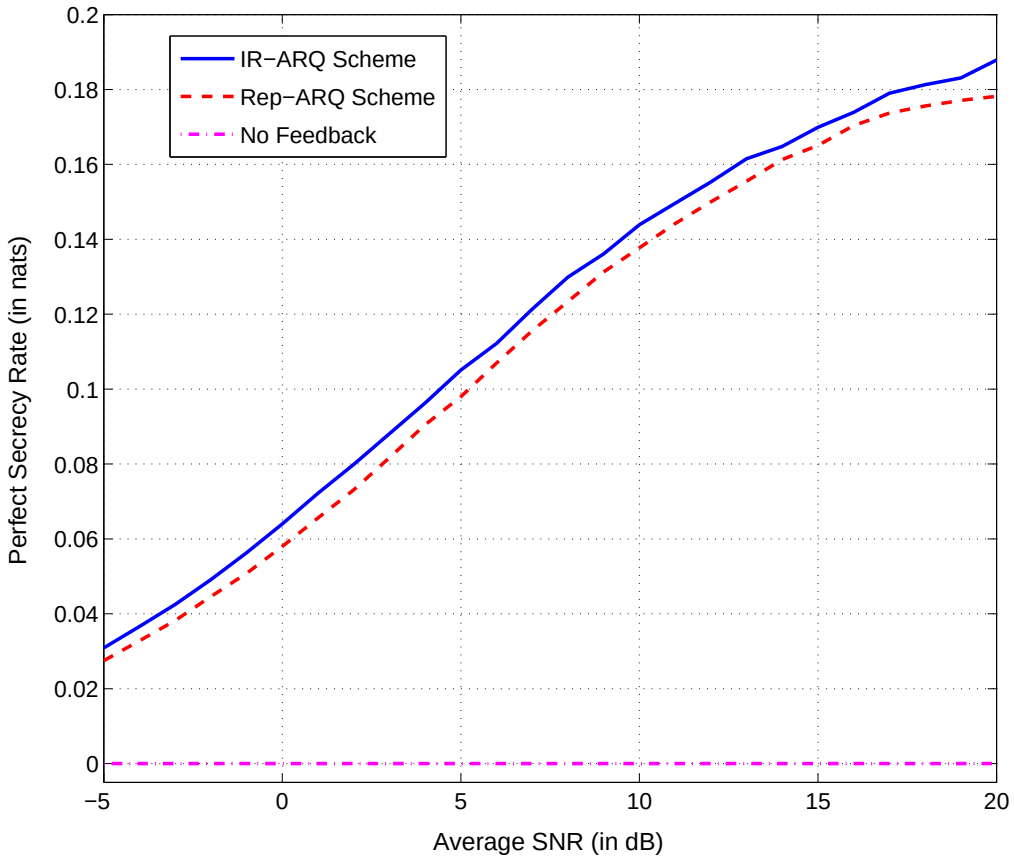

Figure 3.4: Comparison of the perfect secrecy rates achieved by the proposed IR-ARQ and Rep-ARQ schemes for the symmetric scenario $\overline{\gamma}_M = \overline{\gamma}_E = 1$

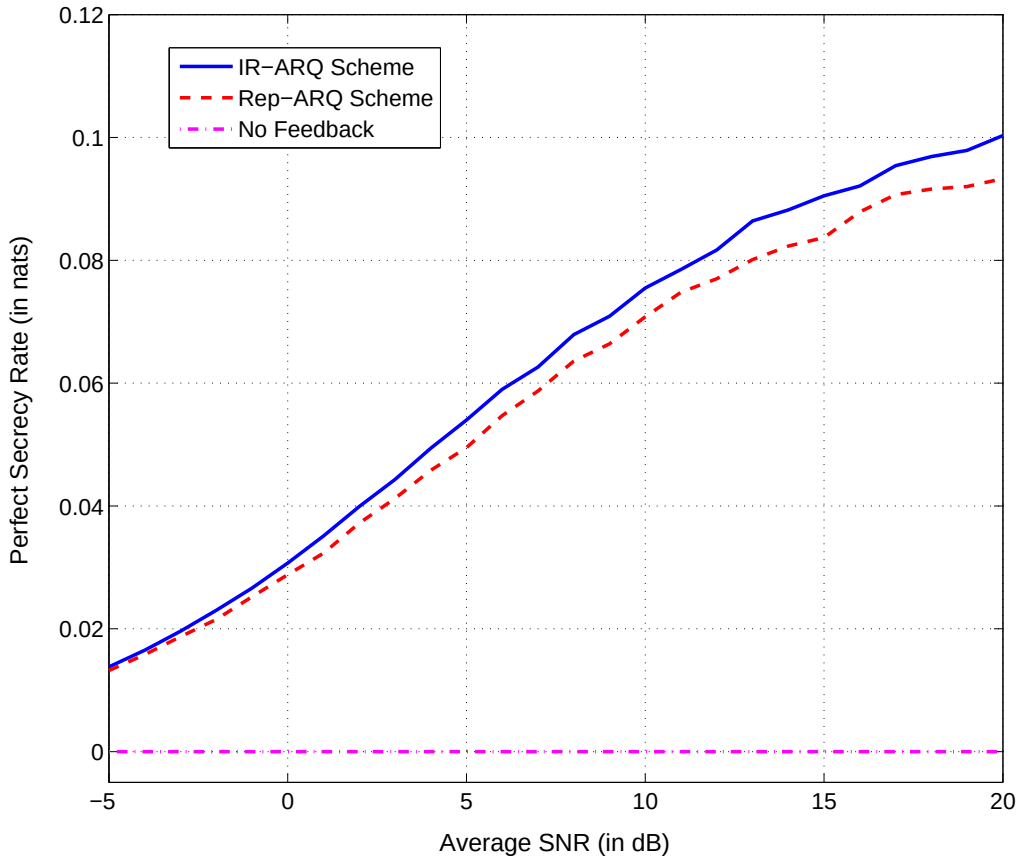

Figure 3.5: Comparison of the perfect secrecy rates achieved by the proposed IR-ARQ and Rep-ARQ schemes for the asymmetric scenario $\overline{\gamma}_M = 1$ and $\overline{\gamma}_E = 2$

CHAPTER 4

FEEDBACK FOR RELIABILITY: THE ARQ CHANNEL WITH DELAY DEADLINES

In this chapter, we demonstrate the impact of feedback on the reliability (in terms of the achievable error exponents) of communication protocols. We consider channels with strict delay deadline constraints and quantify the gains offered by the availability of minimal ARQ feedback.

Shannon proved a negative result in [19] that the capacity of a discrete memoryless channel cannot be increased by using feedback. However, it was later shown by Burnashev [20] that feedback does increase the achievable error exponents. In [20], Burnashev characterized the maximum error exponent achievable over discrete memoryless channels (DMCs) in the presence of perfect output feedback. Interestingly, Forney later showed in [21] that the presence of even one bit of feedback can increase the error exponent significantly. He proposed a memoryless decoding scheme, based on the erasure decoding principle, which achieves a significantly higher error exponent than that achievable through maximum likelihood (ML) decoding without feedback [43]. In Forney's scheme, the transmitter sends codewords of block length N. After receiving each block of N symbols, the receiver uses a reliability-based

erasure decoder and feeds back one ACK/NACK bit indicating whether it has accepted/erased the received block, respectively. If the transmitter receives a NACK message, it then re-transmits the same N-symbol codeword. After each transmission round, the receiver attempts to decode the message using **only** the latest N received symbols, and discards the symbols received previously. This process is repeated until the receiver decides to accept the latest received block and transmits an ACK message back to the transmitter. It is intuitive to expect a better performance from schemes that do not allow for discarding the previous observations at the decoder, as compared with memoryless decoding.

In this chapter, we consider one variant of such schemes, i.e., Incremental Redundancy ARQ (IR-ARQ) [22] and characterize its achievable error exponents under a strict delay deadline constraint, which is imposed in the form of an upper bound L on the maximum number of ARQ rounds.

4.1 The ARQ Channel

We first give a brief overview of the memoryless decoding scheme proposed by Forney in [21]. The transmitter sends a codeword $\mathbf{x}_m$ of length N, where $m \in \{1, \cdots, M\}$. Here M represents the total number of messages at the transmitter, each of which is assumed to be equally likely. The transmitted codeword reaches the receiver after passing through a memoryless channel with transition probability $p(y|x)$. We denote the received sequence as $\mathbf{y}$. The receiver uses an erasure decoder which decides that the transmitted codeword was $\mathbf{x}_m$ iff $\mathbf{y} \in \mathcal{R}_m$, where

$$\mathcal{R}_m \;=\; \left\{ \mathbf{y} \;:\; \frac{p(\mathbf{y}|\mathbf{x}_m)}{\sum_{k \neq m} p(\mathbf{y}|\mathbf{x}_k)} \;\geq\; e^{NT} \right\}, \tag{4.1}$$

where $T \geq 0$ is a controllable threshold parameter. If (4.1) is not satisfied for any $m \in \{1, \cdots, M\}$, then the receiver declares an erasure and sends a NACK bit back to the transmitter. On receiving a NACK bit, the transmitter repeats the codeword corresponding to the same information message. We call such a retransmission as an ARQ round. The decoder discards the earlier received sequence and uses only the latest received sequence of N symbols for decoding (memoryless decoding). It again applies the condition (4.1) on the newly received sequence and again asks for a retransmission in the case of an erasure. When the decoder does not declare an erasure, the receiver transmits an ACK bit back to the transmitter, and the transmitter starts sending the next message. It is evident that this scheme allows for an infinite number of ARQ rounds. This scheme can also be implemented using only one bit of feedback (per codeword) by asking the receiver to only send back ACK bits, and asking the transmitter to keep repeating continuously until it receives an ACK bit. Since the number of needed ARQ rounds for the transmission of a particular message is a random variable, we define the error exponent of this scheme as follows.

Definition 13. *The error exponent $E(R)$ of a variable-length coding scheme is defined as*

$$E(R) \;=\; \limsup_{N\to\infty} \; -\frac{\log \;\Pr(E)}{\overline{\tau}} \;, \tag{4.2}$$

where $\Pr(E)$ denotes the average probability of error, R denotes the average transmission rate, and $\overline{\tau} = (\ln M/R)$ is the average decoding delay of the scheme, when codewords of block length N are used in each ARQ transmission round.

The probability of error of the decoder in (4.1), after each ARQ round, is given by [21]

$$\Pr(\varepsilon) \;=\; \sum_{m} \sum_{k \neq m} \sum_{\mathbf{y} \in R_k} p(\mathbf{y}, \mathbf{x}_m) \;,$$

and the probability of erasure is given by

$$\Pr(X) \;=\; \left(\sum_{m} \sum_{\mathbf{y} \notin R_m} p(\mathbf{y}, \mathbf{x}_m) \right) \;-\; \Pr(\varepsilon) \;.$$

It is shown in [21] that these probabilities satisfy

$$\Pr(X) \;\leq\; e^{-NE_1(R_1,T)} \quad \text{and} \quad \Pr(\varepsilon) \;\leq\; e^{-NE_2(R_1,T)} \;, \tag{4.3}$$

where $R_1 = (\ln M/N)$ denotes the rate of the first transmission round,

$$E_2(R_1, T) \;=\; E_1(R_1, T) + T, \tag{4.4}$$

and $E_1(R_1, T)$ is given at high rates by [21]

$$E_1(R_1, T) \;=\; \max_{0 \leq s \leq \rho \leq 1, \mathbf{p}} \; E_o(s, \rho, \mathbf{p}) - \rho R_1 - sT, \tag{4.5}$$

$$E_o(s, \rho, \mathbf{p}) \;=\; -\log \int \left(\int p(x) p(y|x)^{(1-s)} \mathrm{d}x \right) \left(\int p(x) p(y|x)^{(s/\rho)} \mathrm{d}x \right)^{\rho} \mathrm{d}y, \tag{4.6}$$

and at low rates by

$$E_1(R_1, T) \;=\; \max_{0 \leq s \leq 1, \rho \geq 1, \mathbf{p}} \; E_x(s, \rho, \mathbf{p}) - \rho R_1 - sT, \tag{4.7}$$

$$E_x(s, \rho, \mathbf{p}) \;=\; -\rho \log \iint p(x) p(x_1) \left(\int p(y|x)^{(1-s)} p(y|x_1)^s \mathrm{d}y \right)^{(1/\rho)} \mathrm{d}x \; \mathrm{d}x_1, \tag{4.8}$$

where $\mathbf{p} = \{p(x), \forall x\}$ denotes the input probability distribution (We note that for discrete memoryless channels, the integrals in (4.6) and (4.8) are replaced by summations). The average decoding delay $\overline{\tau}$ of the memoryless decoding scheme is given

by

$$\begin{aligned}\overline{\tau} &= \sum_{k=1}^{\infty} kN \Pr(\text{Transmission stops after } k \text{ ARQ rounds}) \\ &= \sum_{k=1}^{\infty} kN \left[\Pr(X)\right]^{(k-1)}[1-\Pr(X)] = \frac{N}{1-\Pr(X)},\end{aligned}$$

which implies that the average effective transmission rate is given by

$$R = \frac{\ln M}{\overline{\tau}} = \left(\frac{\ln M}{N}\right)[1-\Pr(X)] = R_1[1-\Pr(X)]. \tag{4.9}$$

It is clear from (4.3) and (4.9) that $R \to R_1$ as $N \to \infty$ if $E_1(R_1, T) > 0$. The overall average probability of error can be now computed as

$$\Pr(\mathrm{E}) = \sum_{k=1}^{\infty} [\Pr(X)]^{(k-1)} \Pr(\varepsilon) = \Pr(\varepsilon)\,[1+o(1)], \tag{4.10}$$

where the second equality follows from (4.3) when $E_1(R_1, T) > 0$. It is, therefore, clear that the error exponent achieved by the memoryless decoding scheme is

$$E(R) = \limsup_{N\to\infty} -\frac{\log\left(\Pr(\varepsilon)[1+o(1)]\right)}{\overline{\tau}} \geq E_2(R,T).$$

It is shown in [21] that choosing the threshold T such that $E_1(R_1, T) \to 0$ maximizes the exponent $E_2(R_1, T)$ while ensuring that $R \to R_1$ as $N \to \infty$. This establishes the fact that the memoryless decoding scheme achieves the feedback error exponent $E_F(R)$ defined as

$$E_F(R) \triangleq \lim_{E_1(R,T)\to 0} E_2(R,T) = \lim_{E_1(R,T)\to 0} T. \tag{4.11}$$

At this point, it is interesting to investigate whether a better error exponent can be achieved by employing more complex receivers which exploit observations from previous ARQ rounds in decoding (instead of discarding such observations as in memoryless decoding). Unfortunately, it is easy to see that this additional complexity

does not yield a better exponent in the original setup considered by Forney [21]. The reason is that, as shown in (4.10), the overall probability of error in this setup is dominated by the probability of error $\Pr(\varepsilon)$ in the first transmission round. So, while our more complex decoding rule might improve the probability of error after subsequent rounds, this improvement does not translate into a better error exponent. In the following section, however, we show that in scenarios where a strict deadline is imposed on the maximum number of feedback rounds, significant gains in the error exponent can be reaped by properly exploiting the received observations from previous ARQ rounds (along with the appropriate encoding strategy).

4.2 ARQ with a Deadline

In many practical systems, it is customary to impose an upper bound L on the maximum number of ARQ rounds (in our notation, $L \geq 2$ since we include the first round of transmission in the count). Such a constraint can be interpreted as a constraint on the maximum allowed decoding delay or a deadline constraint. With this constraint, it is obvious that the decoder can no longer use the rule in (4.1) during the L^{th} ARQ round. Therefore, at the L^{th} round, the decoder employs the maximum likelihood (ML) decoding rule to decide on the transmitted codeword. We denote the probability of error of the ML decoder by $\Pr^{(ML)}(\varepsilon)$.

4.2.1 Memoryless Decoding

The following theorem characterizes lower and upper bounds on the error exponent achieved by the memoryless decoding scheme, under the deadline constraint L.

Theorem 14. *The error exponent $E_{MD}(R,L)$ achieved by memoryless decoding, under a deadline constraint L, satisfies*[8] *(for $0 \leq R \leq C$)*

$$E_r(R)+(L-1)\left[\max_{0\leq s\leq\rho\leq 1,\mathbf{p}}\left(\frac{E_o(s,\rho,\mathbf{p})-\rho R-sE_r(R)}{1+s(L-2)}\right)\right] \;\leq\; E_{MD}(R,L) \;\leq\; LE_{sp}(R)\,, \tag{4.12}$$

where $E_r(R)$ and $E_{sp}(R)$ denote the random coding and sphere packing exponents of the memoryless channel, and $E_o(s,\rho,\mathbf{p})$ is as given in (4.6).

Proof. The average decoding delay of memoryless decoding is given by

$$\begin{aligned}\overline{\tau} &= \left(\sum_{k=1}^{L-1} kN\;[\Pr(X)]^{(k-1)}[1-\Pr(X)]\right)+LN[\Pr(X)]^{(L-1)}\\ &= \left(\sum_{k=0}^{L-1}(k+1)N[\Pr(X)]^k\right)-\left(\sum_{k=1}^{L-1}kN[\Pr(X)]^k\right)\\ &= N\left(\sum_{k=0}^{L-1}\;[\Pr(X)]^k\right) \;=\; N\,[1+o(1)]\;, \end{aligned}\tag{4.13}$$

where the last equality follows from (4.3) when $E_1(R_1,T)>0$. Thus the average effective transmission rate is given by

$$R \;=\; \frac{\ln M}{\overline{\tau}} \;=\; \frac{\ln M}{N[1+o(1)]} \;\to\; R_1,$$

as $N\to\infty$ when $E_1(R_1,T)>0$. The average probability of error is given by

$$\begin{aligned}\Pr_{MD}(\mathrm{E}) &= \sum_{k=1}^{L-1}[\Pr(X)]^{(k-1)}\;\Pr(\varepsilon)\;+\;[\Pr(X)]^{(L-1)}\;\Pr^{(ML)}(\varepsilon)\\ &= \Pr(\varepsilon)\,[1+o(1)]\;+\;[\Pr(X)]^{(L-1)}\;\Pr^{(ML)}(\varepsilon) \qquad (4.14)\\ &\leq\; e^{-N[E_1(R_1,T)+T]}\,[1+o(1)]\;+\;e^{-N[E_r(R_1)+(L-1)E_1(R_1,T)]}\,, \qquad (4.15)\end{aligned}$$

[8]We note that a tighter lower bound may be obtained by using the expurgated exponent $E_{ex}(R)$ instead of the random coding exponent $E_r(R)$ at low rates. This observation will be used when generating numerical results.

where the inequality follows from (4.3) and the random coding upper bound on the ML decoding error probability [43]. Letting $E_1(R_1, T) \to 0$ and maximizing T as before, we get the following error exponent

$$E_{MD}(R, L) \;=\; \limsup_{N\to\infty} \; -\frac{\ln \Pr_{MD}(\mathrm{E})}{\overline{\tau}} \;\geq\; \min\{E_F(R), E_r(R)\} \;=\; E_r(R),$$

since the feedback exponent $E_F(R)$ is known to be greater than the random coding exponent $E_r(R)$. Thus by setting $E_1(R_1, T) \to 0$, as suggested by intuitive reasoning, we find that memoryless decoding does not give any improvement over ML decoding without feedback. However, we can get better performance by optimizing the expression in (4.15) w.r.t T without letting $E_1(R_1, T) \to 0$. From (4.15), it is clear that the optimal value of the threshold T^* is the one that yields

$$\begin{aligned} E_1(R_1, T^*) + T^* \;&=\; E_r(R_1) + (L-1)E_1(R_1, T^*) \\ \Rightarrow \quad T^* \;&=\; E_r(R_1) + (L-2)E_1(R_1, T^*)\;. \end{aligned} \tag{4.16}$$

Using this optimal value of T^* in (4.5) and solving for $E_1(R_1, T^*)$, we get

$$E_1(R_1, T^*) \;=\; \max_{0\leq s\leq\rho\leq 1, \mathbf{p}} \left(\frac{E_o(s, \rho, \mathbf{p}) - \rho R_1 - s E_r(R_1)}{1 + s(L-2)} \right)\;. \tag{4.17}$$

Since $E_F(R_1) > E_r(R_1)$, we have $E_1(R_1, T^*) > 0$ and hence $R \to R_1$ as $N \to \infty$. Thus the error exponent of memoryless decoding is lower bounded by

$$\begin{aligned} E_{MD}(R, L) \;&\geq\; E_2(R, T^*) \;=\; E_1(R, T^*) + T^* \;=\; E_r(R) + (L-1)E_1(R, T^*) \\ &=\; E_r(R) + (L-1)\left[\max_{0\leq s\leq\rho\leq 1, \mathbf{p}} \left(\frac{E_o(s, \rho, \mathbf{p}) - \rho R - s E_r(R)}{1 + s(L-2)} \right) \right]. \end{aligned} \tag{4.18}$$

Since $E_1(R, T^*) > 0$, it is clear that the optimal threshold T^* satisfies $0 \leq T^* < E_F(R)$ and thus the lower bound on $E_{MD}(R, L)$ in (4.18) is smaller than the feedback exponent $E_F(R)$.

We now derive an upper bound on $E_{MD}(R, L)$ from (4.14) as follows.

$$\begin{aligned} \Pr_{MD}(\mathrm{E}) &= \Pr(\varepsilon)\,[1 + o(1)] \;+\; [\Pr(X)]^{(L-1)}\ \Pr^{(ML)}(\varepsilon) \\ &\geq [\Pr(X)]^{(L-1)}\ \Pr^{(ML)}(\varepsilon) \\ &\geq [\Pr(X)]^{(L-1)}\ \left(e^{-NE_{sp}(R_1)}\right)\ , \end{aligned} \tag{4.19}$$

where the last inequality follows from the sphere-packing lower bound on the ML decoding error probability [43]. It is easy to see that the probability of erasure $\Pr(X)$ of the decoder in (4.1) decreases when the threshold parameter T is decreased. Thus the probability of erasure $\Pr(X)|_{T=0}$ serves as a lower bound on $\Pr(X)$ for any $T > 0$. In [44], upper and lower bounds on the erasure and error probabilities are derived using a theorem by Shannon *et al.* in [45]. From eqns. (10) and (11) in [44], we have

$$\begin{aligned} \frac{1}{4M}\sum_{m=1}^{M}\exp\left[\mu_m(s) - s\mu_m^{'}(s) - s\sqrt{2\mu_m^{''}(s)}\right] &< \Pr(X) + \Pr(\varepsilon) \\ &\leq \frac{1}{M}\sum_{m=1}^{M}\exp\left[\mu_m(s) - s\mu_m^{'}(s)\right]\ , \end{aligned}$$

and

$$\begin{aligned} \frac{1}{4M}\sum_{m=1}^{M}\exp\left[\mu_m(s) + (1-s)\mu_m^{'}(s) - (1-s)\sqrt{2\mu_m^{''}(s)}\right] &< \Pr(\varepsilon) \\ &\leq \frac{1}{M}\sum_{m=1}^{M}\exp\left[\mu_m(s) + (1-s)\mu_m^{'}(s)\right]\ , \end{aligned}$$

where

$$\mu_m(s) \;=\; \ln\int p(\mathbf{y}|\mathbf{x}_m)^{(1-s)}\left[\sum_{m_1\neq m} p(\mathbf{y}|\mathbf{x}_{m_1})\right]^{s}\mathrm{d}\mathbf{y}\ .$$

It is clear from equation (8) in [44] that the threshold parameter T is related to the parameter $\mu_m(s)$ by $\mu_m^{'}(s) = -NT$. Thus the condition $T = 0$ corresponds to the condition $\mu_m^{'}(s) = 0$. Moreover, it is shown in [44] that $\mu_m(s)$ and $\mu_m^{''}(s)$ are

also proportional to N. Using this fact and the condition $\mu_m^{'}(s) = 0$ in the above expressions for the upper and lower bounds on $\Pr(X)$ and $\Pr(\varepsilon)$, we get

$$\frac{1}{4M}\sum_{m=1}^{M}\exp\left[\mu_m(s)\left(1+o\left(\frac{1}{\sqrt{N}}\right)\right)\right] \;<\; \Pr(X)+\Pr(\varepsilon) \;\leq\; \frac{1}{M}\sum_{m=1}^{M}\exp\left[\mu_m(s)\right] \;, \tag{4.20}$$

and

$$\frac{1}{4M}\sum_{m=1}^{M}\exp\left[\mu_m(s)\left(1+o\left(\frac{1}{\sqrt{N}}\right)\right)\right] \;<\; \Pr(\varepsilon) \;\leq\; \frac{1}{M}\sum_{m=1}^{M}\exp\left[\mu_m(s)\right] \;. \tag{4.21}$$

It is clear from (4.20) and (4.21) that when $T = 0$, the exponents of the upper and lower bounds coincide as $N \to \infty$, and more importantly, the exponent of the erasure probability $\Pr(X)$ is the same as that of the error probability $\Pr(\varepsilon)$. These exponents are further equal to the exponent of the ML decoding error probability since $\Pr(\varepsilon) \leq \Pr^{(ML)}(\varepsilon) \leq \Pr(\varepsilon) + \Pr(X)$. Using this fact and the sphere-packing lower bound on the ML decoding error probability in (4.19), we get

$$\Pr_{MD}(\mathrm{E}) \;\geq\; e^{-NLE_{sp}(R_1)} \quad\Rightarrow\quad E_{MD}(R,L) \;\leq\; LE_{sp}(R) \;,$$

since $R \to R_1$ as $N \to \infty$. □

From Theorem 14, it is clear that ARQ with memoryless decoding does not achieve Forney's error exponent $E_F(R)$ when the maximum number of ARQ rounds L is constrained, at least at high rates for which $LE_{sp}(R) < E_F(R)$. As expected, when $L \to \infty$, the lower bound on the error exponent in (4.12) becomes

$$\lim_{L\to\infty} E_{MD}(R,L) \;\geq\; \max_{0\leq s\leq\rho\leq 1,\mathbf{p}}\left(\frac{E_o(s,\rho,\mathbf{p})-\rho R}{s}\right) \;=\; E_F(R).$$

4.2.2 Incremental Redundancy ARQ

We now derive a lower bound on the error exponent of incremental redundancy ARQ. In IR-ARQ, the transmitter, upon receiving a NACK message, transmits N

new coded symbols (derived from the same message). Since our results hinge on random coding arguments, these new symbols are obtained as i.i.d. realizations from the probability distribution that maximizes the error exponents for erasure decoding[9]. The decoder does not discard the received observations in the case of an erasure and uses the received sequences of all the ARQ rounds jointly to decode the transmitted message. The following erasure decoding rule is employed by the receiver: After the k^{th} ARQ round, the decoder decides on codeword $\mathbf{x}_m$ iff $\mathbf{y} \in \mathcal{R}_m^{'}$, where

$$\mathcal{R}_m^{'} \;=\; \left\{ \mathbf{y} \;:\; \frac{p(\mathbf{y}|\mathbf{x}_m)}{\sum_{i \neq m} p(\mathbf{y}|\mathbf{x}_i)} \;\geq\; e^{kNT_k} \right\} \;, \tag{4.22}$$

and $\mathbf{y}$, $\{\mathbf{x}_i\}$ are vectors of length kN, which contain the received sequences and transmitted codewords (respectively) corresponding to the k ARQ rounds. If no codeword satisfies the above condition, then an erasure is declared by the decoder. It is clear that our formulation allows for varying the threshold T_k as a function of the number of ARQ rounds k. Using thresholds $\{T_k\}$ that decrease with the number of ARQ rounds k makes intuitive sense since the probability of error will be dominated by small values of k (initial ARQ rounds), and hence, one needs to use higher thresholds for these k values to reduce the overall probability of error. We let E_k denote the event that the decoder declares an erasure during *all* the first k ARQ rounds. We also let $E_0 = \phi$ (the empty set). The probability of erasure and error of the decoder in the k^{th} ARQ round will thus be denoted by $\Pr_{(k)}(X|E_{(k-1)})$ and $\Pr_{(k)}(\varepsilon|E_{(k-1)})$, respectively[10]. Here the subscript (k) is used to highlight the fact that the decoder

[9]Note that this optimal error exponent distribution for erasure decoding might not be optimal for ML decoding. However, for the BSC, VNC and AWGN channels considered in the next section, the optimal distributions for erasure decoding and ML decoding coincide.

[10]It follows from our notations that $\Pr_{(1)}(X|E_0) = \Pr(X)$ and $\Pr_{(1)}(\varepsilon|E_0) = \Pr(\varepsilon)$.

uses a received sequence of length kN for decoding in the k^{th} ARQ round. We are now ready to state our main result in this section.

Theorem 15. *The error exponent $E_{IR}(R, L)$ achieved by IR-ARQ, under a deadline constraint L, is given by*[11]

$$E_{IR}(R, L) \;\geq\; \min\left\{E_F(R)\;,\; LE_r^*(R/L)\right\}\;,\quad 0 \leq R \leq C, \tag{4.23}$$

where $E_r^*(.)$ denotes the error exponent achieved by ML decoding under the probability distribution that is optimal for erasure decoding.

Proof. The average decoding delay for IR-ARQ is given by

$$\begin{aligned}
\overline{\tau} &= \sum_{k=1}^{L} kN \Pr(\text{Transmission stops after } k \text{ ARQ rounds}) \\
&= \sum_{k=1}^{L-1} kN \left(\prod_{i=1}^{k-1} \Pr_{(i)}(X|E_{(i-1)})\right)\left[1 - \Pr_{(k)}(X|E_{(k-1)})\right] \\
&\qquad + LN\left(\prod_{i=1}^{L-1} \Pr_{(i)}(X|E_{(i-1)})\right) \\
&= \sum_{k=0}^{L-1}(k+1)N\left(\prod_{i=1}^{k} \Pr_{(i)}(X|E_{(i-1)})\right) \;-\; \sum_{k=1}^{L-1} kN\left(\prod_{i=1}^{k} \Pr_{(i)}(X|E_{(i-1)})\right) \\
&= N\left[1 + \sum_{k=1}^{L-1}\left(\prod_{i=1}^{k} \Pr_{(i)}(X|E_{(i-1)})\right)\right] \\
&\leq N\left[1 + \sum_{k=1}^{L-1} \Pr(X)\right] \;\leq\; N\left[1 + L\Pr(X)\right].
\end{aligned} \tag{4.24}$$

Since $\Pr(X) \leq e^{-NE_1(R_1,T)}$, it follows that $\overline{\tau} \to N$ (and hence the average effective transmission rate $R \to R_1$) as $N \to \infty$ when $E_1(R_1, T) > 0$. The average probability

[11]Replacing the random coding exponent $E_r(R)$ by the expurgated exponent $E_{ex}(R)$ may yield a tighter lower bound at low rates.

of error of IR-ARQ is given by

$$\begin{aligned}
\Pr_{IR}(\mathrm{E}) &= \sum_{k=1}^{L} \Pr(\text{error in the } k^{th} \text{ ARQ round}) \\
&= \sum_{k=1}^{L-1} \Pr_{(k)}(\varepsilon, E_{(k-1)}) + \Pr_{(L)}^{(ML)}(\varepsilon, E_{(L-1)}) \\
&\leq \sum_{k=1}^{L-1} \Pr_{(k)}(\varepsilon) \; + \; \Pr_{(L)}^{(ML)}(\varepsilon) \; ,
\end{aligned}$$

where $\Pr_{(k)}(\varepsilon)$ refers to the probability of error when the decoder always waits for kN received symbols before decoding. Following the derivation in [21], it can easily be seen that for the thresholds $\{T_k\}$ used in the decoding rule (4.22), we have

$$\Pr_{(k)}(X) \; \leq \; e^{-kNE_1(R_1/k,T_k)} \quad \text{and} \quad \Pr_{(k)}(\varepsilon) \; \leq \; e^{-kN[E_1(R_1/k,T_k)+T_k]} \; . \tag{4.25}$$

Using this and the fact that $\Pr_{(L)}^{(ML)}(\varepsilon) \leq e^{-LNE_r^*(R_1/L)}$, since $E_r^*(.)$ is the error exponent achieved by ML decoding under the probability distribution that maximizes the error exponent for erasure decoding, we can upper bound the average probability of error of IR-ARQ by

$$\Pr_{IR}(\mathrm{E}) \; \leq \; \sum_{k=1}^{L-1} e^{-kN[E_1(R_1/k,T_k)+T_k]} \; + \; e^{-LNE_r^*(R_1/L)} \; . \tag{4.26}$$

Thus the error exponent achieved by IR-ARQ is lower bounded by

$$\begin{aligned}
E_{IR}(R,L) &= \limsup_{N\to\infty} \; -\frac{\ln \Pr_{IR}(\mathrm{E})}{\overline{\tau}} \\
&\geq \min\Big(LE_r^*(R/L), \; \{ \; k[E_1(R/k,T_k)+T_k] \; \}_{k=1}^{L-1}\Big) \, .
\end{aligned}$$

Taking $T_k = (T/k)$, $\forall k \in \{1, \cdots, (L-1)\}$, we get

$$\begin{aligned}
E_{IR}(R,L) &\geq \min\Big(LE_r^*(R/L), \; \{ \; kE_1(R/k,T/k)+T \; \}_{k=1}^{L-1}\Big) \\
&= \min\left(LE_r^*(R/L), \; E_1(R,T)+T\right) ,
\end{aligned}$$

where the last equality follows from the fact that $E_1(R/k, T/k)$ is an increasing function of k. Letting $E_1(R,T) \to 0$ and maximizing T, we get

$$E_{IR}(R,L) \;\geq\; \min\left(LE_r^*(R/L), E_F(R)\right) \;.$$

□

From Theorem 15, it is clear that if the deadline constraint L is large enough to satisfy

$$LE_r^*(R/L) \;\geq\; E_F(R) \;, \tag{4.27}$$

then IR-ARQ achieves the feedback exponent $E_F(R)$ at rate R. In the following section, we quantify the gains achieved by IR-ARQ, as compared with memoryless decoding, for specific channels.

4.3 Examples

4.3.1 The Binary Symmetric Channel

Here, we compare the error exponents achievable by memoryless decoding and IR-ARQ over a BSC with crossover probability ϵ. The bounds on the error exponents in (4.12) and (4.23) are plotted for a BSC with $\epsilon = 0.15$ in Figs. 4.1 and 4.2 for $L = 2$ and $L = 4$, respectively. The ML decoding error exponent (corresponding to the case $L = 1$) and the feedback exponent $E_F(R)$ are also plotted for comparison purposes. From Fig. 4.1, we find that when $L = 2$, memoryless decoding achieves an error exponent that is strictly sub-optimal to the feedback exponent $E_F(R)$ for all $R \geq 0.006$. On the other hand, IR-ARQ achieves $E_F(R)$ for $0.18 \leq R \leq C$. Moreover, it performs strictly better than memoryless decoding for all $R \geq 0.057$. When $L = 4$, from Fig. 4.2, we find that the error exponent for the memoryless

decoder is strictly sub-optimal, as compared with $E_F(R)$, for $R \geq 0.141$, while IR-ARQ achieves $E_F(R)$ for all rates below capacity. Finally, we note that even when $L = 100$, memoryless decoding is still strictly sub-optimal, as compared with IR-ARQ, for all rates $0.38 \leq R \leq C = 0.39$.

Now, we elaborate on our observation from Fig. 4.2 that $L = 4$ is sufficient to achieve $E_F(R)$ with IR-ARQ when $\epsilon = 0.15$. In particular, we wish to investigate the existence of a finite value for L such that $E_F(R)$ is achieved by IR-ARQ universally (i.e., for all $0 \leq \epsilon \leq 0.5$ and all rates below capacity). Towards this end, we derive an upper bound on the *minimum* required deadline constraint L_{req} for a given BSC(ϵ). From (4.23), it is clear that L_{req} is upper bounded by the minimum value of L required to satisfy[12] $LE_r(R/L) \geq E_F(R)$ for all $0 \leq R \leq C$. We first prove the following result.

Lemma 16. *A sufficient condition for ensuring that $LE_r(R/L) \geq E_F(R)$ for all rates $0 \leq R \leq C$ for a BSC is given by $LE_r(0) \geq E_F(0)$.*

Proof. It has been shown in [43] that both the random coding exponent $E_r(R)$ and the feedback exponent $E_F(R)$ are decreasing functions of R. Since

$$LE_r(R/L) \;=\; \max_{0 \leq \rho \leq 1} \{LE_o(\rho) - \rho R\} \;, \tag{4.28}$$

its slope at a given rate R is given by (following the steps in equations (5.6.28–5.6.33) in [43])

$$\frac{\partial \left(LE_r(R/L)\right)}{\partial R} \;=\; -\rho^*(R) \;\geq\; -1 \;,$$

where $\rho^*(R)$ is the value of ρ that maximizes the RHS of (4.28) for rate R. For a BSC, it is shown in [21] that the feedback exponent can be expressed as

$$E_F(R) \;=\; (C - R) \;+\; \max_{\rho \geq 0} \{E_o(\rho) - \rho R\} \;. \tag{4.29}$$

[12]Note that $E_r^*(.) = E_r(.)$ for the BSC, VNC and AWGN channels, since the optimal distributions for ML decoding and erasure decoding coincide.

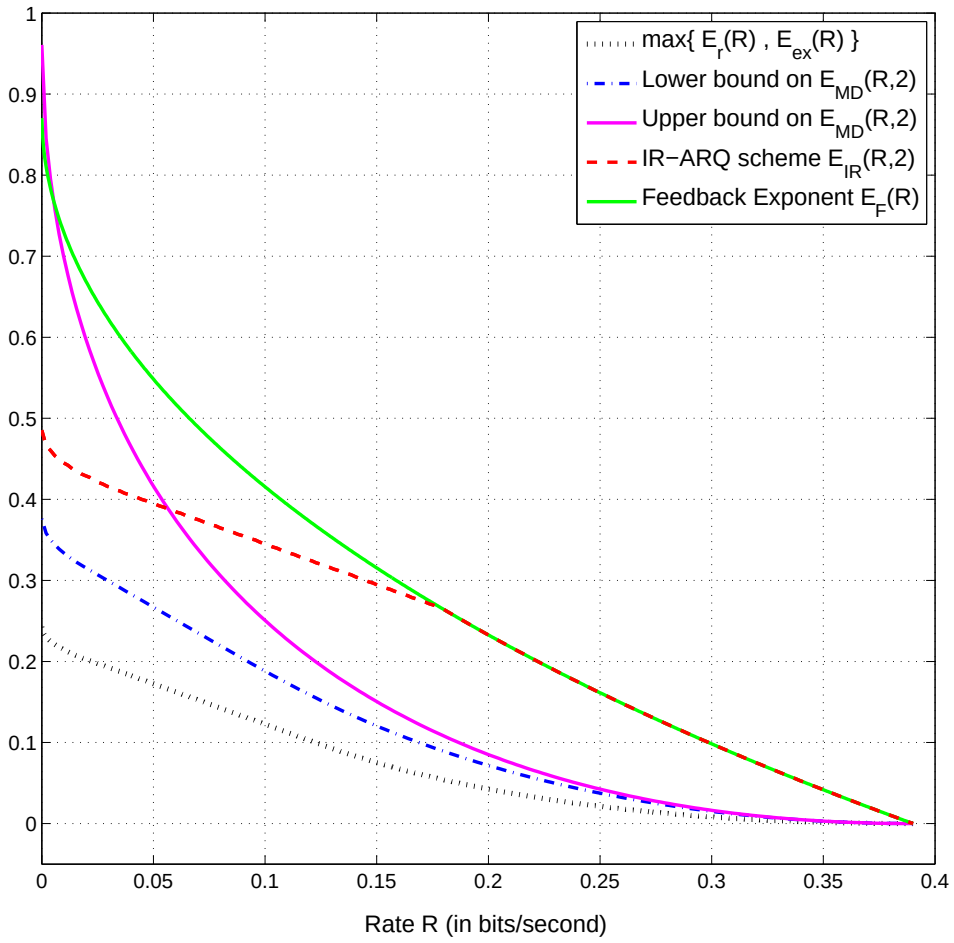

Figure 4.1: Comparison of the error exponents for a Binary Symmetric Channel (BSC) with cross-over probability $\epsilon = 0.15$ and maximum number of ARQ rounds $L = 2$

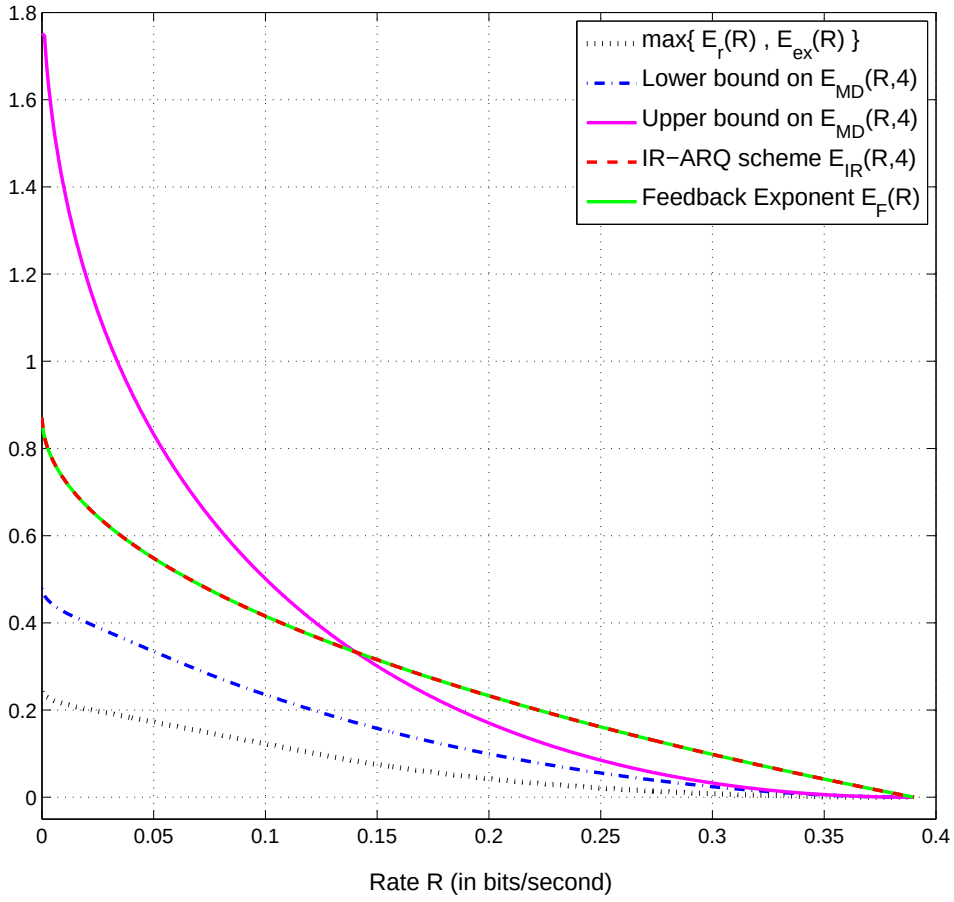

Figure 4.2: Comparison of the error exponents for a Binary Symmetric Channel (BSC) with cross-over probability $\epsilon = 0.15$ and maximum number of ARQ rounds $L = 4$

Hence the slope of $E_F(R)$ at a given rate R is given by

$$\frac{\partial E_F(R)}{\partial R} \;=\; -\left(1+\rho^{'}(R)\right) \;\leq\; -1\;,$$

where $\rho^{'}(R)$ is the value of ρ that maximizes the RHS of (4.29) for rate R. Hence it is clear that for any value of R, the rate of decrease of the feedback exponent $E_F(R)$ is higher than that of $LE_r(R/L)$. It is shown in [21] that $E_F(C) = E_r(C) = 0$. Since $E_r(R)$ is a decreasing function of R, we know that $E_r(C/L) > E_r(C) = 0$. Thus, when $R = C$, we have $LE_r(C/L) > E_F(C)$. Now, if the value of L is chosen such that $LE_r(0) > E_F(0)$, it is clear that the curve $LE_r(R/L)$ lies strictly above the curve $E_F(R)$ in the range $0 \leq R \leq C$. This directly follows from the fact that the feedback exponent $E_F(R)$ decreases faster than $LE_r(R/L)$. Hence the condition $LE_r(0) \geq E_F(0)$ is sufficient to guarantee that $LE_r(R/L) \geq E_F(R)$ for all $0 \leq R \leq C$. □

The above lemma shows that for any BSC(ϵ), an upper bound on L_{req} depends only on the values of $E_F(R)$ and $E_r(R)$ at $R = 0$. From the results in [43], it can be shown that

$$E_r(0) \;=\; \ln 2 - \ln\left(1+2\sqrt{\epsilon(1-\epsilon)}\right) \quad \text{and} \quad E_F(0) \;=\; C - \ln 2 - \ln\left(\sqrt{\epsilon(1-\epsilon)}\right)\;. \tag{4.30}$$

Using Lemma 16 and (4.30), we find that a deadline constraint of $L = 4$ is enough to achieve the feedback exponent $E_F(R)$ at all rates below capacity for *any* BSC with crossover probability $0.05 \leq \epsilon \leq 0.5$. However, the upper bound on L_{req}, derived using Lemma 16, becomes loose as $\epsilon \to 0$. To overcome this limitation, we use the expurgated exponent $E_{ex}(R)$ [43] instead of the random coding exponent $E_r(R)$ at low rates. Using numerical results, we find that the actual value of the minimum

required deadline constraint is $L_{req} = 3$ for all BSCs with $\epsilon \leq 0.025$, and $L_{req} = 4$ otherwise.

4.3.2 The Very Noisy Channel

As noted in [43], a channel is very noisy when the probability of receiving a given output is almost independent of the input, i.e., when the transition probabilities of the channel are given by

$$p_{jk} \;=\; \omega_j \left(1 + \epsilon_{jk}\right) \;,$$

where $\{\omega_j\}$ denotes the output probability distribution, and $\{\epsilon_{jk}\}$ are such that $|\epsilon_{jk}| \ll 1$ for all j and k, and $\sum_j \omega_j \epsilon_{jk} = 0$, $\forall k$. We plot the bounds on the error exponents given in (4.12) and (4.23), derived from the results in [21], in Figs. 4.3 and 4.4 for a VNC with capacity $C = 1$ for $L = 2$ and $L = 4$ respectively. From the plots, it is clear that memoryless decoding is strictly sub-optimal to IR-ARQ for all rates $R \geq 0.12$ (with $L = 2$) and $R \geq 0.25$ (with $L = 4$). Moreover, it is evident that $L = 4$ is sufficient for IR-ARQ to achieve the feedback exponent $E_F(R)$ for all rates below capacity. This observation motivates the following result.

Lemma 17. *For the very noisy channel, a deadline constraint of $L = 4$ is enough for the proposed incremental redundancy scheme to achieve the feedback exponent $E_F(R)$ for all rates $0 \leq R \leq C$.*

Proof. For a VNC, the random coding exponent is given by [21]

$$E_r(R) \;=\; \begin{cases} \left(\frac{C}{2} - R\right) \;, & 0 \leq R \leq \frac{C}{4} \\ (\sqrt{C} - \sqrt{R})^2 \;, & \frac{C}{4} \leq R \leq C \end{cases} \;. \tag{4.31}$$

Thus, under the deadline constraint $L = 4$, we have

$$4E_r(R/4) \;=\; 4\left(\frac{C}{2} - \frac{R}{4}\right) \;=\; 2C - R \;, \qquad 0 \leq R \leq C.$$

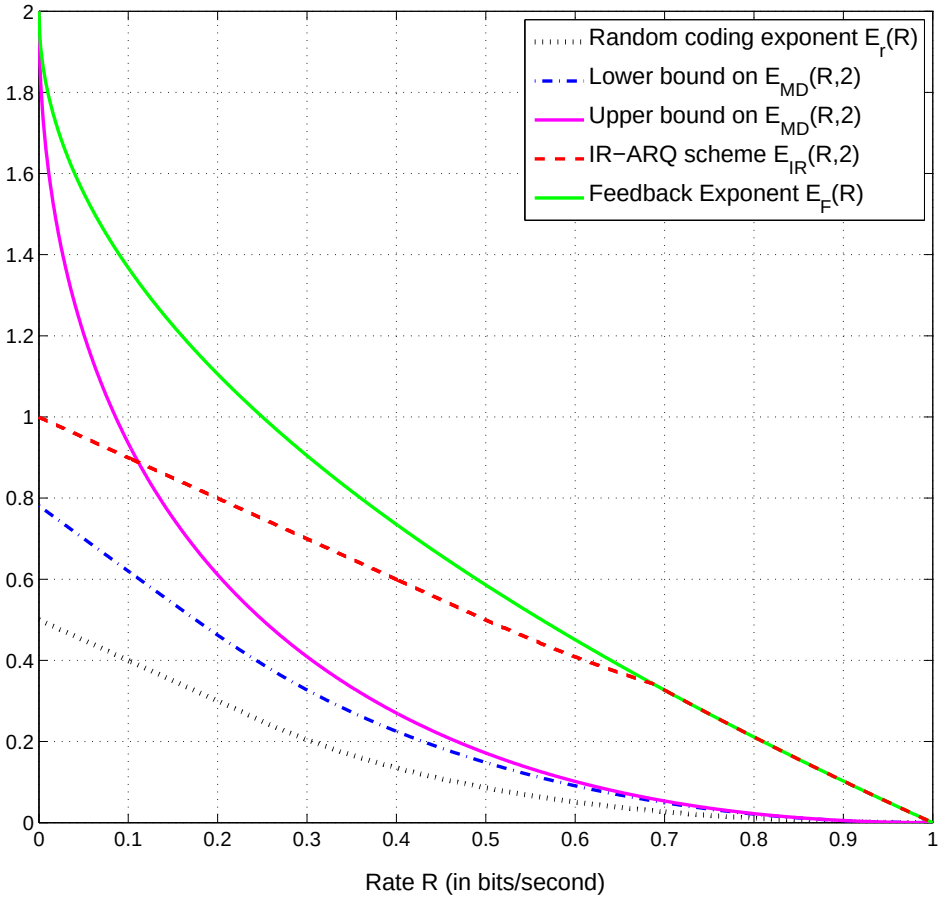

Figure 4.3: Comparison of the error exponents for a Very Noisy Channel (VNC) with capacity $C = 1$ and maximum number of ARQ rounds $L = 2$

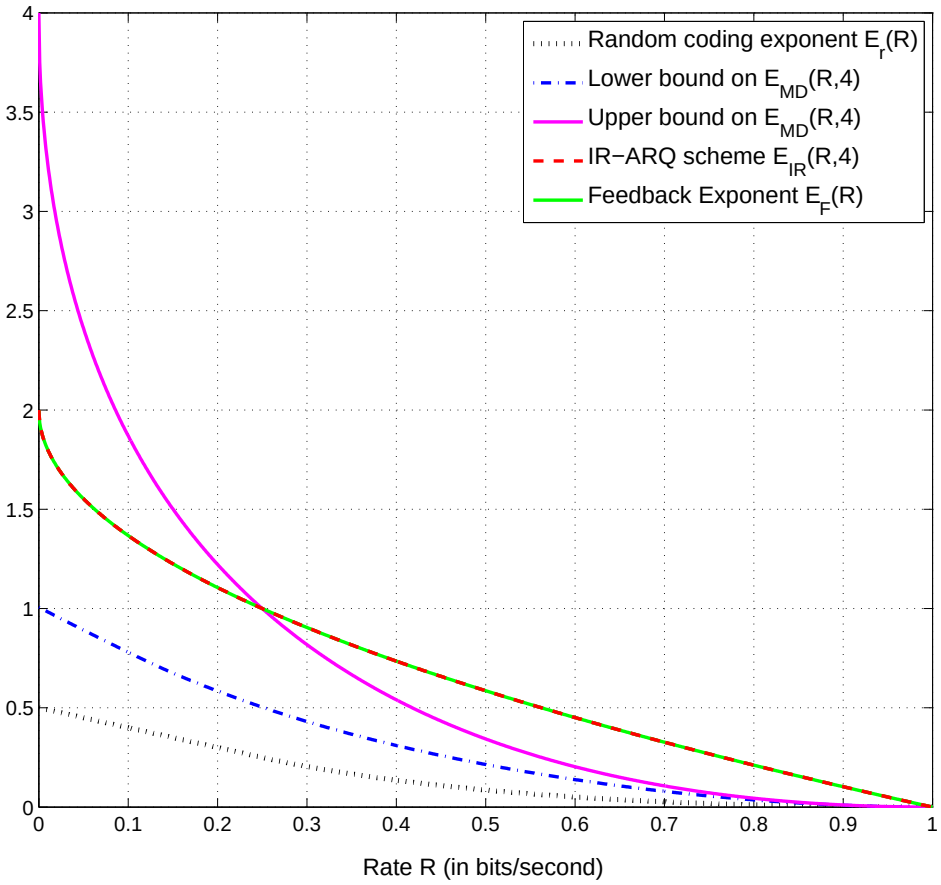

Figure 4.4: Comparison of the error exponents for a Very Noisy Channel (VNC) with capacity $C = 1$ and maximum number of ARQ rounds $L = 4$

Also

$$E_F(R) \;=\; (C-R)+(\sqrt{C}-\sqrt{R})^2 \;\leq\; (C-R)+(\sqrt{C})^2 \;=\; 4E_r(R/4)\;.$$

Putting $L=4$ in (4.23), the error exponent of IR-ARQ is given by

$$E_{IR}(R,4) \;\geq\; \min\left\{E_F(R)\;,\;4E_r(R/4)\right\} \;=\; E_F(R)\;. \tag{4.32}$$

Thus, for a VNC, it is clear that a deadline constraint of $L=4$ is enough for IR-ARQ to achieve the feedback exponent $E_F(R)$ at all rates below capacity. □

4.3.3 The Additive White Gaussian Noise Channel

The random coding and expurgated exponents for an AWGN channel with a Gaussian input of power A and unit noise variance, are given in [43]. The sphere-packing exponent of the AWGN channel is derived in [46–48]. The parameter $E_o(s,\rho,\mathbf{p})$ in the lower bound in (4.12) is replaced by $E_o(s,\rho,t)$ which, following the steps in the derivation of the random coding exponent in [43], is given by

$$\begin{aligned} E_o(s,\rho,t) \;=\; (1+\rho)tA+\left(\frac{1}{2}\right)\log(1-2tA)+\left(\frac{\rho}{2}\right)\log\left(1-2tA+\frac{sA}{\rho}\right) \\ +\;\left(\frac{1}{2}\right)\log\left(1+\frac{sA\left(1-s-\frac{s}{\rho}\right)}{1-2tA+\frac{sA}{\rho}}\right)\;. \end{aligned}$$

The feedback exponent for the AWGN channel is then given by [21, 43]

$$E_F(R) = \max_{\substack{0\leq s\leq\rho\leq 1\\ t\geq 0}} \left(\frac{E_o(s,\rho,t)-\rho R}{s}\right)\;.$$

We plot the bounds on the error exponents, given in (4.12) and (4.23), in Figs. 4.5 and 4.6 for an AWGN channel with signal-to-noise ratio $A=3$ dB for the deadline constraints $L=2$ and $L=4$ respectively. The plots clearly indicate that memoryless

decoding is strictly sub-optimal to IR-ARQ for all rates $R \geq 0.19$ (with $L = 2$) and $R \geq 0.46$ (with $L = 4$). Moreover, when $L = 4$, the proposed IR-ARQ scheme achieves the feedback exponent $E_F(R)$ for all rates below capacity. But we do not have a rigorous proof that this observation holds universally for any general SNR.

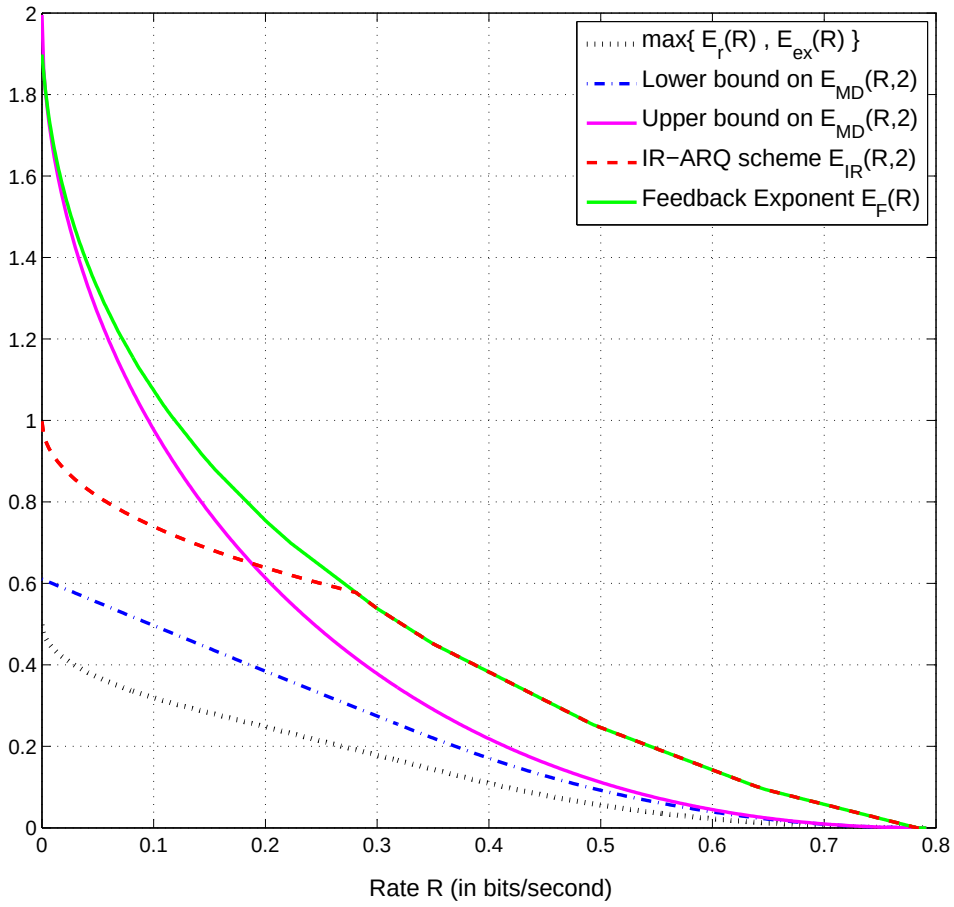

Figure 4.5: Comparison of error exponents for an Additive White Gaussian Noise (AWGN) channel with Signal-to-Noise Ratio SNR = 3 dB and maximum number of ARQ rounds $L = 2$

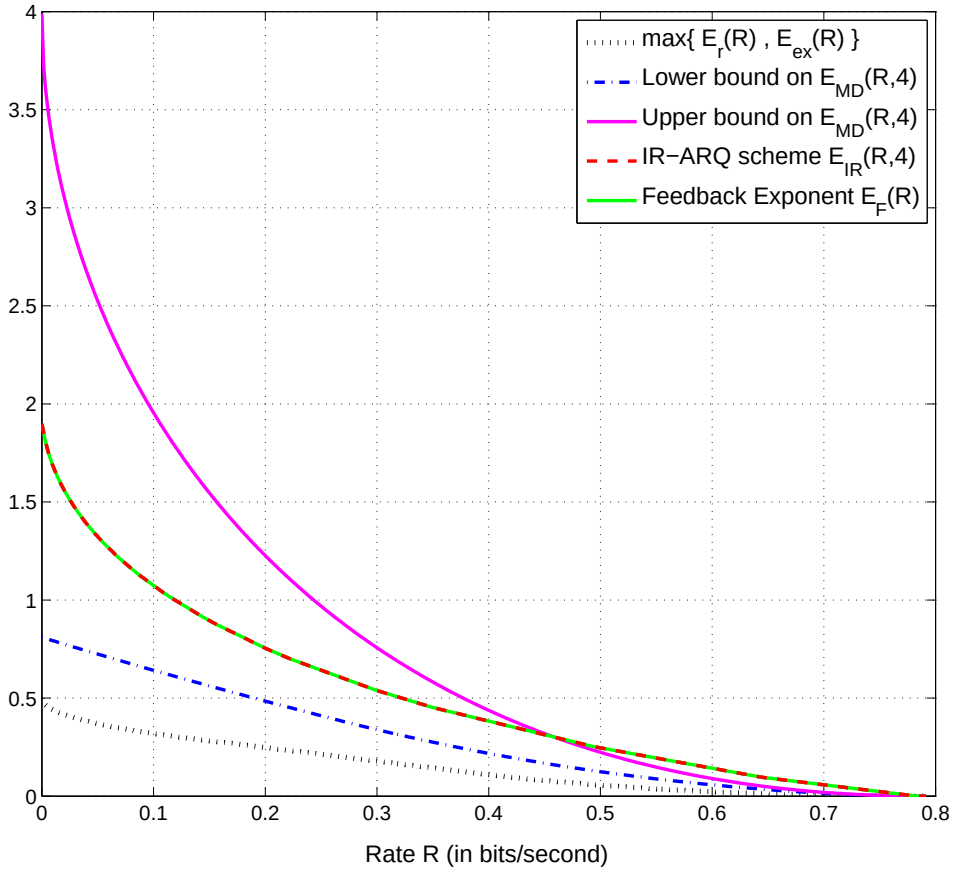

Figure 4.6: Comparison of error exponents for an Additive White Gaussian Noise (AWGN) channel with Signal-to-Noise Ratio SNR = 3 dB and maximum number of ARQ rounds $L = 4$

CHAPTER 5

CONCLUSIONS

The central goal of this work was to highlight the importance of feedback in wireless communications, and to identify many different ways in which feedback improves the performance of wireless systems. Towards this end, we considered three different scenarios and characterized the impact of feedback on each of them. We first considered cellular multicast channels and showed that the availability of feedback allows for the cross-layer design of efficient multicast schedulers. We proposed low-complexity multicast schedulers that achieve near-optimal scaling of both throughput and delay with the user population, for both the perfect CSI feedback and the ARQ feedback scenarios. We further proposed a cooperative multicast scheduler, requiring perfect CSI feedback, that achieves the optimal asymptotic throughput-delay tradeoff. For the multiple transmit antenna scenario, we showed that the throughput performance of multicast schedulers is dominated by the amount of multicast gain they harness, and demonstrated the near-optimality of the worst user scheduler with a large number of transmit antennas.

We then considered fading eavesdropper channels and demonstrated the importance of feedback in establishing secure communications. We characterized the secrecy capacity under the assumptions of full CSI and main channel CSI knowledge

at the transmitter, and proposed optimal rate and power allocation strategies. Quite interestingly, we showed that the availability of feedback enables one to exploit the time-varying nature of the wireless medium and achieve a perfectly secure non-zero rate even when an eavesdropper channel is more capable than the legitimate channel on the average. We further established the critical role of rate adaptation, based on the main channel CSI, in facilitating secure communications over slow fading channels. We also proposed a low-complexity on/off power allocation strategy and established its asymptotic optimality. This optimality shows that the presence of eavesdropper CSI at the transmitter does not offer additional gains in the secrecy capacity for slow fading channels, at high enough SNR levels. We then considered an ARQ feedback scenario and proposed transmission schemes that leverage the ARQ feedback to achieve non-zero perfect secrecy rates when the eavesdropper has a superior channel on the average. Thereby, we established the positive impact of feedback on the secrecy capacity of fading channels.

Finally, we considered ARQ channels with strict delay deadline constraints to study the impact of ARQ feedback on reliability. We proposed a transmission scheme based on incremental redundancy ARQ with joint decoding at the receiver, and showed that it outperforms Forney's memoryless decoding scheme in terms of the achievable error exponents. Moreover, the proposed IR-ARQ scheme was shown to achieve the Forney's feedback exponent $E_F(R)$ when the delay constraint satisfies $L \geq 4$ for any BSC and VNC channel, and also for at least some range of SNRs for AWGN channels. This translates into a significant improvement in reliability since Forney's memoryless decoding scheme typically achieves the feedback exponent $E_F(R)$ only as $L \to \infty$.

5.1 Possible Future Work

Some of the promising directions for future research are listed below.

- In Chapter 2, we characterized the throughput-delay tradeoff achieved by the proposed set of schedulers. However, characterizing the optimal throughput-delay tradeoff in cellular multicast channels is still an open problem. Some related work for wireless point-to-point and broadcast channels can be found in [27, 28, 49].

- It will be interesting to extend the multicast schemes proposed in Chapter 2 to the general multi-group multicast scenario, where only the users within a particular group want the same information from the base station. One can harness throughput gains in such a scenario by exploiting the multi-user diversity available across the groups. But this comes at the price of an increased feedback requirement and a higher average delay. Some preliminary results for this scenario can be found in [50].

- We assumed backlogged queues in the throughput-delay analysis of Chapter 2 and hence ignored queuing delay in our delay definition. An interesting open problem would be to analyze the throughput-delay tradeoff achieved under a queuing model with random arrivals. In such cases, it would be advantageous to incorporate the instantaneous queue lengths (states) into the scheduler design problem [27, 28]. Also characterizing the queuing delay for coupled queues (like in the best user scheduler) is a challenging open problem.

- Extending the throughput-delay analysis in Chapter 2 to the asymmetric fading scenario is an open problem. The asymmetric setting will lead to fairness issues between the different users, which need to be taken into account during the scheduler design. Some notions of fairness and the corresponding optimal schedulers for broadcast channels are developed in [3, 51].

- In Chapter 3, we characterized a set of achievable perfect secrecy rates for the system with ARQ feedback. Characterizing the secrecy capacity with ARQ feedback is still an open problem.

- Another interesting venue for research is information-theoretic secrecy for delay-limited channels. Here the challenge is to come up with the correct outage formulation that is meaningful. Some preliminary efforts in this area are given in [41, 42], but these seem to be meaningful only for the full transmitter CSI scenario.

- Characterizing the secrecy capacity for a fast fading scenario (where a codeword sees almost infinite channel realizations) is still an open problem. The perfect secrecy rate achieved by the constant-rate power control scheme proposed in Chapter 3 is a lower bound on the secrecy capacity. It would be interesting to study if this scheme is in fact optimal.

- It would be interesting to investigate whether ARQ feedback from the receiver would improve the secrecy capacity of AWGN eavesdropper channels.

- Burnashev characterized the exact error exponent for DMCs with perfect output feedback in [20]. However, for DMCs with ARQ feedback (one-bit feedback),

only the achievability of the Forney exponent $E_F(R)$ has been proved. Hence only a lower bound on the actual error exponent is known. Thus characterizing the maximum achievable error exponent for DMCs with ARQ feedback still remains an open problem.

- We showed in Chapter 4 that a delay deadline of $L = 4$ is enough for the proposed IR-ARQ scheme to achieve the Forney exponent $E_F(R)$ for any BSC or VNC. Moreover, we showed numerically that this result also holds for AWGN channels at least for the considered range of SNRs. It will be interesting to analytically investigate whether this result holds for any general SNR in AWGN channels.

APPENDIX A

THROUGHPUT-DELAY ANALYSIS FOR CELLULAR MULTICAST

A.1 Worst User Scheduler (Theorem 3)

The average throughput of the worst user scheduler is given by

$$R_{tot} \;=\; N\mathbb{E}\left[\log\left(1+|h_{\pi(1)}|^2P\right)\right].$$

Since the $\{|h_i|^2\}$ are i.i.d. and exponentially distributed with unit mean, it is evident that the random variable $|h_{\pi(1)}|^2 = \min_i |h_i|^2$ also follows an exponential distribution, and hence

$$R_{tot} \;=\; N\int_0^\infty \log(1+xP)Ne^{-Nx}\mathrm{d}x \;=\; -Ne^{\left(\frac{N}{P}\right)}Ei\left(-\frac{N}{P}\right), \qquad \text{(A.1)}$$

where $Ei(x) = \int_{-\infty}^{x}(e^t/t)\mathrm{d}t$. For large values of x, we have

$$Ei(-x) \;=\; \int_{-\infty}^{-x}\frac{e^t}{t}\mathrm{d}t \;=\; -\frac{e^{-x}}{x}\left(1+\epsilon\right),$$

where $\epsilon \to 0$ as $x \to \infty$. Using this fact in (A.1), we get

$$R_{tot} \;=\; P(1+\epsilon) \;=\; \Theta(1).$$

We now calculate the average delay of the worst user scheduler. The BS maintains a single common queue for all the users in the system. We consider each coherence interval of length T_c as a time slot. The service time X is defined as

$$X \;=\; kT_c, \quad \text{when} \quad T_c\left(\sum_{i=1}^{k-1} R_i\right) <\; S \;\leq T_c\left(\sum_{i=1}^{k} R_i\right) \quad (k \in \{1,2,\ldots\}). \quad \text{(A.2)}$$

Here S denotes the size of each packet and R_i represents the service rate in the i^{th} time slot, which is given by $R_i = \log(1+|h^i_{\pi(1)}|^2 P)$. We let $C = (S/T_c)$ in the sequel. We consider the sequence of random variables $\{R_i\}$ and define a stopping instant τ as follows:

$$\tau \;=\; \min\left\{k : \sum_{i=1}^{k} R_i \;\geq\; C\right\}.$$

Using the stopping rule property [52], we get $\mathbb{E}[\tau]\mathbb{E}[R] = \mathbb{E}[\hat{C}] = \mathbb{E}[C+\tilde{C}]$, where $\tilde{C}$ represents the overshoot of the sum of R_i's with respect to the threshold C (Hence $\mathbb{E}[\tilde{C}] \leq \mathbb{E}[R]$). Thus the mean stopping time is given by

$$\frac{C}{\mathbb{E}[R]} \;\leq\; \mathbb{E}[\tau] \;\leq\; 1+\frac{C}{\mathbb{E}[R]} \qquad \Rightarrow \;\; \mathbb{E}[\tau] \;=\; \Theta\left(1+\frac{C}{\mathbb{E}[R]}\right).$$

Thus the average service time is given by

$$\bar{X} \;=\; \mathbb{E}[\tau]T_c \;=\; \Theta\left(T_c+\frac{S}{\mathbb{E}[R]}\right). \quad \text{(A.3)}$$

Since for large values of N, the average service rate $\mathbb{E}[R] = (R_{tot}/N) = \Theta(1/N)$, the average delay of the worst user scheduler scales as

$$D \;=\; \bar{X} \;=\; \Theta(T_c + NS) \;=\; \Theta(N).$$

A.2 Best User Scheduler (Theorem 4)

The average throughput of the best user scheduler is given by

$$R_{tot} \;=\; \mathbb{E}\left[\log\left(1+|h_{\pi(N)}|^2 P\right)\right] \;=\; \int_0^\infty \log(1+xP)\mathrm{d}F(x),$$

where $|h_{\pi(N)}|^2 = \max_i |h_i|^2$ has the distribution $F(x) = (1 - e^{-x})^N$, $x \geq 0$. Integrating by parts and simplifying, we get

$$R_{tot} \;=\; \sum_{i=1}^{N} \binom{N}{i} (-1)^i e^{\left(\frac{i}{P}\right)} Ei\left(\frac{-i}{P}\right). \tag{A.4}$$

It has been shown in [53] that the average throughput in (A.4) scales as

$$R_{tot} \;=\; \Theta(\log\log N) \tag{A.5}$$

with the number of users N.

For calculating the average delay of the best user scheduler, we follow the approach used in [53]. Here the BS maintains N queues, one for each user in the system. These queues are coupled, in the sense that any packet that needs to be transmitted enters all the N queues (since it needs to be transmitted to all the users). Moreover, the BS serves only one of these N queues during any particular time slot. We first calculate the average service time $\bar{X}$ required for transmitting a packet from a queue when the BS always serves that particular queue. The average service rate of the best user scheduler is given by $\mathbb{E}[R] = R_{tot}$. Thus following the argument in the earlier proof and using (A.5), it can be shown that (refer (A.3))

$$\bar{X} \;=\; \Theta\left(T_c + \frac{S}{\mathbb{E}[R]}\right) \;=\; \Theta\left(T_c + \frac{S}{\log\log N}\right) \;=\; \Theta(1). \tag{A.6}$$

We are interested in determining the delay involved in successfully transmitting a particular packet from all of the N coupled queues. The actual delay, as defined in Section 2.1, is the time between the start of transmission of a packet and the instant when the packet reaches all the N users in the system. In our analysis, we assume that the packet of interest is at the head of all the N queues during the start of transmission. This assumption thus results in a lower bound on the actual delay.

We characterize the delay based on the observation that our queuing problem is equivalent to the well-known "coupon collector" problem. A similar observation was made earlier in [4] where the authors characterized the delay of the throughput-optimal broadcast scheme. They assumed that the server (BS) offers a constant rate of service, which is independent of the instantaneous channel gains of the users. In our analysis, however, we incorporate the effects of rate adaptation. Let $X_1, X_2, \cdots, X_N$ denote the service times (assuming continuous service) required for transmitting a packet from each of the N queues. Then the delay of the scheduler is directly proportional to the minimum number of trials required to ensure that the first queue is served at least (X_1/T_c) times by the base station, the second queue is served at least (X_2/T_c) times and so on ...

We lower bound the average delay by calculating the minimum number of trials N_t required to ensure that all the N queues are served at least (X_{min}/T_c) times by the BS, where $X_{min} = \min\{X_1, X_2, \cdots, X_N\}$. We determine the average number of such required trials $\mathbb{E}[N_t|X_{min}]$ using the results derived in [4]. Since the BS serves only one of the N queues at any time and since the fading is symmetric across users, there is an equal probability that the BS serves any one of the queues. Thus the probabilities $\{p_j\}$ of the server choosing the j^{th} queue for service are given by $p_1 = \cdots = p_N = (1/N)$. These probabilities $\{p_j\}$ remain constant through all time slots and are not functions of the instantaneous service rates $\{R_i\}$ provided by the BS. The Moment Generating Function (MGF) of the number of trials required is given by [4]

$$F_{N_t|X_{min}}(z) \;=\; \sum_{i=0}^{\infty} z^i \Pr(N_t > i) \;=\; \sum_{i=0}^{\infty} z^i b_i,$$

where b_i is the probability of failure of sending a packet to all the users in i channel uses. The value of b_i is equal to the polynomial $(x_1 + x_2 + \cdots + x_N)^i/N^i$ evaluated at $x_1 = \cdots = x_N = 1$ after removing all terms that have all x_i's with exponent larger than or equal to (X_{min}/T_c) (denoted by the operator $\{.\}$) [38]. Thus the MGF of the number of trials required is given by

$$F_{N_t|X_{min}}(z) \;=\; \sum_{i=0}^{\infty} \frac{z^i}{N^i} \left\{ (x_1 + \cdots + x_N)^i \right\}.$$

Using the identities [38]

$$\frac{z^i}{N^i} \;=\; \frac{N}{i!z} \int_0^{\infty} e^{\frac{-Nt}{z}} t^i \mathrm{d}t \qquad \text{and}$$

$$\sum_{i=0}^{\infty} \frac{\{(x_1 + \cdots + x_N)^i\}}{i!} \;=\; \left\{ e^{(x_1+\cdots+x_N)} \right\} \;=\; e^{(x_1+\cdots+x_N)} - \prod_{i=1}^{N} \left(e^{x_i} - S_{(\frac{X_{min}}{T_c})}(x_i) \right),$$

where $S_m(t) = \sum_{i=0}^{m-1} (t^i/i!)$, we get

$$F_{N_t|X_{min}}(z) \;=\; \frac{N}{z} \int_0^{\infty} e^{-\frac{Nt}{z}} \left(e^{Nt} \left[1 - \left(1 - S_{(\frac{X_{min}}{T_c})}(t) e^{-t} \right)^N \right] \right) \mathrm{d}t.$$

Hence the average number of trials required $\mathbb{E}[N_t|X_{min}]$ is given by [4]

$$\begin{aligned} \mathbb{E}[N_t|X_{min}] \;&=\; F_{N_t|X_{min}}(1) \;=\; N \int_0^{\infty} \left[1 - \left(1 - S_{(\frac{X_{min}}{T_c})}(t) e^{-t} \right)^N \right] \mathrm{d}t \\ &=\; N\mathbb{E} \left[\max_{1 \le i \le N} Y_i \right], \end{aligned}$$

where the Y_i's are i.i.d random variables that follow a Chi-square distribution with $(2X_{min}/T_c)$ degrees of freedom. Using the results in [4], it can be shown that for such a sequence of random variables $\{Y_i\}$,

$$\mathbb{E} \left[\max_{1 \le i \le N} Y_i \right] \;=\; \max \left\{ \Theta(\log N), \Theta \left(\frac{X_{min}}{T_c} \right) \right\}. \tag{A.7}$$

Thus the average number of trials required is given by

$$\mathbb{E}[N_t|X_{min}] \;=\; \max \left\{ \Theta \left(N \log N \right), \Theta \left(\frac{N X_{min}}{T_c} \right) \right\}.$$

Hence the average delay of the best user scheduler can be lower bounded by

$$D \;\geq\; \mathbb{E}_{X_{min}}\left[\mathbb{E}[N_t|X_{min}]T_c\right] \;=\; \mathbb{E}_{X_{min}}\left[\max\left\{\Theta\left(NT_c\log N\right),\Theta\left(NX_{min}\right)\right\}\right].$$

Since $\mathbb{E}\left[\max\{Z_1,Z_2\}\right] \geq \max\left\{\mathbb{E}[Z_1],\mathbb{E}[Z_2]\right\}$, we have

$$D \;=\; \max\left\{\Omega\left(NT_c\log N\right),\Omega\left(N\mathbb{E}[X_{min}]\right)\right\}.$$

By observing that $\mathbb{E}[X_{min}] \leq \bar{X}$ and using (A.6), we get

$$D \;=\; \Omega\left(NT_c\log N\right) \;=\; \Omega\left(N\log N\right). \tag{A.8}$$

A.3 Median User Scheduler (Theorem 5)

In the median user scheduler, the BS keeps on repeating the same packet to $(N/2)$ users in each time slot, until all the N users receive it successfully. Due to this repetition, some of the users receive redundant information (multiple copies of the same packet). Hence the average throughput of this scheduler **cannot** be specified as

$$R_{tot} \;=\; \left(\frac{N}{2}\right)\mathbb{E}\left[\log\left(1+|h_{\pi(\frac{N}{2}+1)}|^2P\right)\right],$$

where $|h_{\pi(\frac{N}{2}+1)}|^2$ is the median of the channel gains among all the N users in the system. However, the average throughput can be easily calculated using the following renewal theory argument. Consider the renewal process wherein the successful reception of a packet of size S by all the N users is taken to be the renewal event. Since the average inter-renewal time is given by the average delay D, it is straightforward to show, using the renewal reward theorem, that the average throughput of the scheduler is

$$R_{tot} \;=\; \frac{NS}{D}.$$

Thus we first need to characterize the average delay D of the median user scheduler. The average service rate provided to any user is given by

$$\mathbb{E}[R] \;=\; \mathbb{E}\left[\log\left(1+|h_{\pi(\frac{N}{2}+1)}|^2 P\right)\right].$$

We first characterize the scaling of $\mathbb{E}[R]$ with N. Now suppose that the BS does not repeat the same packet after one transmission. Then the average throughput obtained is $T = (N/2)\mathbb{E}[R]$. From the results on central order statistics in [54] (Theorem 8.5.1), we know that the sample median of N i.i.d. exponential random variables converges in distribution to a normal random variable with mean θ and variance $(1/N)$, where $\theta = \log 2$ is the median of the underlying exponential distribution. Hence

$$\left(|h_{\pi(\frac{N}{2}+1)}|^2 - \theta\right)\sqrt{N} \quad\rightarrow\quad W \quad \text{in distribution,} \tag{A.9}$$

where W is a standard normal random variable. Using Chebyshev's inequality, we get $\forall \epsilon > 0$,

$$\begin{aligned}\Pr\left(\left||h_{\pi(\frac{N}{2}+1)}|^2-\theta\right|>\epsilon\right) &= \Pr\left(\sqrt{N}\left||h_{\pi(\frac{N}{2}+1)}|^2-\theta\right|>\epsilon\sqrt{N}\right)\\ &< \frac{\mathbb{E}\left[W^2\right]+\delta}{N\epsilon^2}\rightarrow 0 \quad \text{as } N\rightarrow\infty.\end{aligned}$$

Thus $|h_{\pi(\frac{N}{2}+1)}|^2 \rightarrow \theta$ in probability. Since the $\log(.)$ function is continuous,

$$\log\left(1+|h_{\pi(\frac{N}{2}+1)}|^2 P\right) \quad\rightarrow\quad \log(1+\theta P) \quad \text{in probability.} \tag{A.10}$$

We now derive a lower bound on T. We recall the following property of positive random variables. Let (X_n) be a set of positive random variables converging to a constant A in probability. Then $\forall \epsilon > 0$, $\Pr\left(|X_n - A| \geq \epsilon\right) < \delta$, for some small $\delta > 0$. Now

$$\mathbb{E}[X_n] \;=\; \int_0^\infty t f_{X_n}(t)\mathrm{d}t \quad\geq\quad \int_{A-\epsilon}^{A+\epsilon} t f_{X_n}(t)\mathrm{d}t \quad\geq\quad (A-\epsilon)(1-\delta).$$

Taking the limit as $n \to \infty$, we get $\lim_{n\to\infty} \mathbb{E}[X_n] \geq A$. Using this property in (A.10), we get

$$\lim_{N\to\infty} \mathbb{E}\left[\log\left(1+|h_{\pi(\frac{N}{2}+1)}|^2 P\right)\right] \;\geq\; \log(1+\theta P) = \Theta(1) \quad \Rightarrow \quad T \;=\; \Omega(N). \tag{A.11}$$

Combining this with the upper bound on T in (2.4), we get

$$T \;=\; \Theta(N) \qquad \Rightarrow \quad \mathbb{E}[R] \;=\; \mathbb{E}\left[\log\left(1+|h_{\pi(\frac{N}{2}+1)}|^2 P\right)\right] \;=\; \Theta(1). \tag{A.12}$$

Thus the average service rate in the median user scheduler does not scale with N. We now consider an extension of the coupon collector problem, where the users are assumed to have coupons and the transmitter is the collector that selects $(N/2)$ different users randomly (with a uniform distribution) in each trial, and collects one coupon from each of them. We characterize the average number of trials $\overline{N_t}$ required to ensure that the collector collects at least one coupon from all the N users. An upper bound can be easily derived by considering a *weaker* modified scheme, where in each trial, the $(N/2)$ users are chosen with replacement by the collector from the set of N users. Thus any user can be selected multiple times within the same trial, and hence the average number of trials required for this weaker scheme will be greater than that for the original scheme. It is easy to see that this weaker scheme is in fact the original coupon collector problem [38] with $(N/2)$ independent coupons collected at each instant. Thus

$$\overline{N_t} \;=\; O\left(\frac{N\log N}{(N/2)}\right) \;=\; O(\log N).$$

A lower bound on $\overline{N_t}$ is derived as follows. During the k^{th} trial, the probability that coupon i has not been collected is $(1/2)^k$. The expected number of coupons that have not been collected until the k^{th} trial is given by $E_{N,k} = N(1/2)^k$ [52]. We find

the number of trials k_δ required to ensure that Pr(collecting coupons from all N users within k_δ trials) $> 1-\delta$, for some small $\delta > 0$. This requires that $E_{N,k} < \epsilon$ for some small $\epsilon > 0$.

$$\Rightarrow \quad \frac{N}{2^{k_\delta}} \; < \; \epsilon \quad \Rightarrow \quad k_\delta \; > \; \log_2\left(\frac{N}{\epsilon}\right) \quad \Rightarrow \quad k_\delta \; = \; \Omega(\log N).$$

Since ensuring that coupons have been collected from all users is stronger than the condition Pr(collecting coupons from all N users within k_δ trials) $> 1-\delta$, the value k_δ serves as a lower bound for $\overline{N_t}$. Thus $\overline{N_t} = \Theta(\log N)$. Using (A.12) and the property (A.9), it can be shown that the average delay D of the median user scheduler scales as

$$D \; = \; \Theta\left(\overline{N_t}\left(T_c + \frac{S}{\mathbb{E}[R]}\right)\right) \; = \; \Theta(\log N).$$

The average throughput of the median user scheduler is hence given by

$$R_{tot} \; = \; \frac{NS}{D} \; = \; \Theta\left(\frac{N}{\log N}\right).$$

A.4 Incremental Redundancy Multicast (Theorem 6)

Let A_i denote the event that a packet is successfully decoded by all the N users in the system in i transmission attempts. Following the notation in [22], we define

$$q(m) \; = \; \Pr(\overline{A_1}, \ldots, \overline{A_{m-1}}, A_m) \; = \; p(m-1) - p(m),$$

where

$$p(m) \; = \; \Pr(\overline{A_1}, \ldots, \overline{A_{m-1}}, \overline{A_m}) \; = \; 1 - \sum_{l=1}^{m} q(l),$$

with $p(0) = 1$. The rate $\bar{R}$ is defined as $\bar{R} = (b/L)$. We define the random variable τ to be the number of transmission attempts made between the instant when the codeword is generated and the instant when its transmission is stopped (Transmission

is stopped either when the packet is successfully decoded by all the N users or the number of transmission attempts exceeds the rate constraint M). The probability distribution of τ is given by

$$f_\tau(m) \;=\; \begin{cases} 0, & m = 0 \\ q(m), & 1 \le m \le M-1 \\ q(M)+p(M), & m = M \end{cases} .$$

We define the random reward $\mathcal{R}$ as follows: $\mathcal{R} = N\bar{R}$ if transmission stops because of successful decoding and $\mathcal{R} = 0$ if transmission stops because of the rate constraint violation. Hence

$$\mathbb{E}[\mathcal{R}] \;=\; N\bar{R}\sum_{m=1}^{M} q(m) \;=\; N\bar{R}[1-p(M)].$$

The mean inter-renewal time is given by

$$\begin{aligned} \mathbb{E}[\tau] \;&=\; \sum_{m=1}^{M} m f_\tau(m) \;=\; \sum_{m=1}^{M} mq(m) + Mp(M) \\ &=\; \sum_{m=1}^{M} m[p(m-1)-p(m)] + Mp(M) \;=\; \sum_{m=0}^{M-1} p(m). \end{aligned}$$

Applying the renewal-reward theorem, we obtain the average throughput of the proposed scheme as $R_{tot} = (\mathbb{E}[\mathcal{R}]/\mathbb{E}[\tau])$ with probability 1. Hence

$$R_{tot} \;=\; \frac{N\bar{R}\,[1-p(M)]}{1+\sum_{m=1}^{M-1} p(m)}.$$

The average delay D of the scheme is given by the mean inter-renewal time. Hence $D = \mathbb{E}[\tau]$. The unconstrained throughput and delay are obtained by letting $M \to \infty$ and are given by

$$R_{tot} \;=\; \frac{N\bar{R}}{\sum_{m=0}^{\infty} p(m)} \quad \text{and} \quad D \;=\; \sum_{m=0}^{\infty} p(m). \tag{A.13}$$

From the earlier definitions, we have

$$p(m) = \Pr(\overline{A_1}, \ldots, \overline{A_{m-1}}, \overline{A_m}) = \Pr(\overline{A_m}) = \Pr\left(\min_{i=1}^{N} \sum_{k=1}^{m} I(X; Y_{ik}) \leq \bar{R}\right)$$
$$= 1 - \left[1 - \Pr\left(\sum_{k=1}^{m} I(X; Y_{1k}) \leq \bar{R}\right)\right]^N. \quad \text{(A.14)}$$

Now for a Gaussian input distribution, we have

$$\sum_{k=1}^{m} I(X; Y_{1k}) = \sum_{k=1}^{m} \log(1 + |h_k|^2).$$

We know that

$$\log\left(1 + \sum_{k=1}^{m} |h_k|^2\right) \leq \sum_{k=1}^{m} \log(1 + |h_k|^2) \leq \sum_{k=1}^{m} |h_k|^2.$$

Hence

$$\Pr\left(\sum_{k=1}^{m} |h_k|^2 \leq (e^{\bar{R}} - 1)\right) \geq \Pr\left(\sum_{k=1}^{m} \log(1 + |h_k|^2) \leq \bar{R}\right) \geq \Pr\left(\sum_{k=1}^{m} |h_k|^2 \leq \bar{R}\right).$$

Since both $\bar{R}$ and $(e^{\bar{R}} - 1)$ are constants, substituting both the lower and upper bounds in (A.14) will yield the same scaling with N. So we consider only the lower bound on $p(m)$. Let

$$s(m) = 1 - \left[1 - \Pr\left(\sum_{k=1}^{m} |h_k|^2 \leq \bar{R}\right)\right]^N.$$

Hence $\sum_{m=0}^{\infty} p(m) = \Theta\left(\sum_{m=0}^{\infty} s(m)\right)$ w.r.t N. The random variable $\sum_{k=1}^{m} |h_k|^2$ has a $2m$-dimensional Chi-square distribution with the density and distribution functions given by

$$f(x) = \frac{e^{-x} x^{m-1}}{(m-1)!} \quad \text{and} \quad F(x) = 1 - e^{-x} \left(\sum_{l=0}^{m-1} \frac{x^l}{l!}\right), \quad x \geq 0.$$

Hence

$$s(m) = 1 - \left[e^{-\bar{R}} \left(\sum_{l=0}^{m-1} \frac{\bar{R}^l}{l!}\right)\right]^N.$$

From Taylor's theorem, we know that (for some $0 < \theta < 1$)

$$e^{\bar{R}} = \sum_{l=0}^{m-1} \frac{\bar{R}^l}{l!} + \frac{e^{\theta\bar{R}}\bar{R}^m}{m!} \quad \Rightarrow \quad \sum_{l=0}^{m-1} \frac{\bar{R}^l}{l!} = e^{\bar{R}} - \frac{e^{\theta\bar{R}}\bar{R}^m}{m!}$$

$$\Rightarrow \quad s(m) = 1 - \left(1 - \frac{e^{-(1-\theta)\bar{R}}\bar{R}^m}{m!}\right)^N.$$

To find the scaling of $\sum_{m=0}^{\infty} s(m)$ w.r.t N, we first derive a lower bound by finding the value of m until which $s(m) \to 1$ as $N \to \infty$. Now

$$s(m) \to 1 \quad \Rightarrow \quad \left(1 - \frac{e^{-(1-\theta)\bar{R}}\bar{R}^m}{m!}\right)^N \to 0 \quad \Rightarrow \quad \frac{e^{-(1-\theta)\bar{R}}\bar{R}^m}{m!} > \Theta\left(\frac{1}{N}\right).$$

Using Stirling's approximation, we have

$$\frac{e^{-(1-\theta)\bar{R}}\bar{R}^m}{\sqrt{2\pi m}e^{-m}m^m} > \frac{k}{N}, \quad \forall \quad \text{constant } k.$$

Taking log on both sides, we get

$$(1-\theta)\bar{R} - m + m\log\left(\frac{m}{\bar{R}}\right) + \frac{1}{2}\log(2\pi m) < \log N - \log k, \quad \forall k.$$

For large N, this equation can be reduced to $m \log m < \log N$. This equation is satisfied by all values of m such that

$$m < \Theta\left(\frac{\log N}{\log\log N}\right).$$

Since $s(m) \to 1$ as $N \to \infty$ for all values of m that satisfy the above equation, the sum of $s(m)$'s can be lower bounded as

$$\sum_{m=0}^{\infty} s(m) \geq \Theta\left(\frac{\log N}{\log\log N}\right). \tag{A.15}$$

Similarly an upper bound on $\sum_{m=0}^{\infty} s(m)$ can be derived by finding the value of m from which $s(m) \to 0$ as $N \to \infty$. Following the same procedure as before, we find that $s(m) \to 0$ when $m > \Theta(\log N / \log\log N)$. This yields the following upper bound

$$\sum_{m=0}^{\infty} s(m) \leq \Theta\left(\frac{\log N}{\log\log N}\right).$$

Combining this with the lower bound in (A.15), we get

$$\sum_{m=0}^{\infty} s(m) \;=\; \Theta\left(\frac{\log N}{\log\log N}\right).$$

Thus the average delay is given by

$$D \;=\; \sum_{m=0}^{\infty} p(m) \;=\; \Theta\left(\sum_{m=0}^{\infty} s(m)\right) \;=\; \Theta\left(\frac{\log N}{\log\log N}\right).$$

The average throughput of the incremental redundancy scheme is then given by

$$R_{tot} \;=\; \frac{N\bar{R}}{\sum_{m=0}^{\infty} p(m)} \;=\; \frac{N\bar{R}}{D} \;=\; \Theta\left(\frac{N\log\log N}{\log N}\right).$$

A.5 Cooperative Multicast (Theorem 7)

The average throughput in the first stage of the cooperation scheme is given by

$$\left(\frac{N}{2}\right)\mathbb{E}[R_{s1}] \;=\; \left(\frac{N}{2}\right)\mathbb{E}\left[\log\left(1+|h_{\pi(\frac{N}{2}+1)}|^2 P\right)\right].$$

It is shown in (A.12) that $\mathbb{E}[R_{s1}] = \Theta(1)$. We now characterize the average throughput in the second stage of the cooperation scheme. As noted earlier, the cooperative transmission by the users in the second stage is equivalent to the transmission of packets from a transmitter equipped with $(N/2)$ transmit antennas to the worst ser in a group of $(N/2)$ users. Hence the average transmission rate during the cooperative stage is given by

$$\mathbb{E}[R_{s2}] \;=\; \mathbb{E}\left[\min_{i=1,\ldots,(N/2)} \log\left(1+\frac{|h_{1i}|^2+\cdots+|h_{(N/2)i}|^2}{(N/2)}P\right)\right],$$

where the $|h_{ki}|^2$'s are i.i.d and exponentially distributed and represent the inter-user fading coefficients.

$$\Rightarrow \mathbb{E}[R_{s2}] \;=\; \mathbb{E}\left[\log\left(1+\min_{i=1,\ldots,M}\frac{|\chi^i_{2M}|^2}{M}P\right)\right], \tag{A.16}$$

where $M = (N/2)$ and $|\chi^i_{2M}|^2$'s are Chi-square random variables with $2M$ degrees of freedom whose distribution function is given by $F(x) = 1 - e^{-x}\left(\sum_{j=0}^{M-1}(x^j/j!)\right)$, $x \geq 0$. Using the results on extreme order statistics in [54] (Theorems 8.3.2-8.3.6), it can be shown that the random variable $\left(\min_{i=1}^{M}|\chi^i_{2M}|^2\right)/b_M \rightarrow W$ in distribution as $M \rightarrow \infty$, where W is a Weibull type random variable and b_M satisfies $F(b_M) = (1/M)$. Now

$$F(b_M) \;=\; \frac{1}{M} \quad \Rightarrow \quad 1 - e^{-b_M}\left(\sum_{j=0}^{M-1}\frac{b_M^j}{j!}\right) \;=\; \frac{1}{M}.$$

Using Taylor's theorem, we get for some $0 < \beta_M < 1$

$$1 - e^{-b_M}\left(e^{b_M} - \frac{e^{\beta_M b_M} b_M^M}{M!}\right) \;=\; \frac{1}{M} \quad \Rightarrow \quad \frac{e^{-(1-\beta_M)b_M} b_M^M}{M!} \;=\; \frac{1}{M}.$$

Using Stirling's approximation, we have

$$\frac{e^{-(1-\beta_M)b_M} b_M^M}{\sqrt{2\pi M} M^M e^{-M}} \;=\; \frac{1}{M}.$$

Taking log(.) on both sides, we get

$$(1-\beta_M)b_M - M\log b_M \;=\; M - \left(M - \frac{1}{2}\right)\log M + C.$$

Since $\beta_M \rightarrow 0$ as $M \rightarrow \infty$, we get $b_M = \Theta(M)$. Thus $\left(\min_{i=1}^{M}|\chi^i_{2M}|^2\right)/M \rightarrow kW$ in distribution, for some constant $k > 0$. Since the log(.) function is continuous, we have

$$\log\left(1 + \frac{\min_{i=1}^{M}|\chi^i_{2M}|^2}{M}P\right) \;\rightarrow\; \log(1+kWP) \quad \text{in distribution, as } M \rightarrow \infty.$$

Now, we know

$$\log\left(1 + \frac{\min_{i=1}^{M}|\chi^i_{2M}|^2}{M}P\right) \;\leq\; \frac{\left(\min_{i=1}^{M}|\chi^i_{2M}|^2\right)P}{M} \;\leq\; \frac{|\chi^1_{2M}|^2 P}{M}.$$

Since

$$\mathbb{E}\left[\left(\frac{|\chi^1_{2M}|^2 P}{M}\right)^2\right] \;=\; \frac{\mathbb{E}[(|\chi^1_{2M}|^2)^2]P^2}{M^2} \;=\; \left(1 + \frac{1}{M}\right)P^2 \;\leq\; 2P^2 \;<\; \infty \quad \forall M,$$

the sequence $\{(|\chi_{2M}^1|^2 P)/M; M \geq 1\}$ is uniformly integrable.

$$\Rightarrow \quad \left\{ \log\left(1 + \frac{\min_{i=1}^{M} |\chi_{2M}^i|^2}{M} P\right);\ M \geq 1 \right\} \quad \text{is uniformly integrable.}$$

It is shown in [55] that if a sequence of random variables (X_n) is uniformly integrable and $X_n \to X$ in distribution as $n \to \infty$, then $\mathbb{E}X_n \to \mathbb{E}X$ as $n \to \infty$. Thus

$$\mathbb{E}\left[\log\left(1 + \frac{\min_{i=1}^{M} |\chi_{2M}^i|^2}{M} P\right)\right] \quad \to \quad \mathbb{E}[\log(1 + kWP)] \ = \ \Theta(1).$$

Hence the average transmission rate of the second stage is given by $\mathbb{E}[R_{s2}] = \Theta(1)$ w.r.t N. Since both $\mathbb{E}[R_{s1}]$ and $\mathbb{E}[R_{s2}]$ do not scale with N and since the minimum is taken over only two positive quantities, we have $\mathbb{E}\left[\min\{R_{s1}, R_{s2}\}\right] = \Theta(1)$. Thus the average throughput of the cooperation scheme is given by

$$R_{tot} \ = \ \left(\frac{N}{2}\right) \mathbb{E}\left[\min\{R_{s1}, R_{s2}\}\ \right] \ = \ \Theta(N).$$

We now determine the average delay of the cooperation scheme. We note that the BS needs to maintain only a single queue that caters to all the N users in the system. The information transmitted by the BS in the first half of each time slot reaches all the N users at the end of that time slot. Hence the average delay is equal to the average service time required for transmitting a packet of size S from the queue. Following the steps in Appendix A.1, the average delay D for transmitting a packet in the cooperation scheme is given by (refer equation (A.3))

$$D \ = \ \Theta\left(T_c + \frac{S}{\mathbb{E}[\min\{R_{s1}, R_{s2}\}]}\right) \ = \ \Theta(1).$$

A.6 Multi-Transmit Antenna Worst User Scheduler (Theorem 8)

From the results on extreme order statistics in [54], we know that $(|\chi_{min}|^2/b_N) \to W$ in distribution, where W has a Weibull type distribution and b_N satisfies $F(b_N) =$

$(1/N)$, which implies

$$1 - e^{-Lb_N}\left(\sum_{k=0}^{L-1}\frac{(Lb_N)^k}{k!}\right) \;=\; \frac{1}{N}.$$

Using Taylor's theorem, we get for some $0 < \gamma_N < 1$

$$1 - e^{-Lb_N}\left(e^{Lb_N} - \frac{e^{\gamma_N Lb_N}(Lb_N)^L}{L!}\right) \;=\; \frac{1}{N} \quad\Rightarrow\quad \frac{e^{-(1-\gamma_N)Lb_N}(Lb_N)^L}{L!} \;=\; \frac{1}{N}.$$

Taking log(.) on both sides, we get

$$(1-\gamma_N)Lb_N - L\log b_N \;=\; \log N + L\log L - \log(L!).$$

Since $|\chi_{min}|^2 \leq |\chi_1|^2 = \Theta(1)$, we know that $b_N = O(1)$ and hence the $\log b_N$ term dominates the left hand side of the above expression. Thus we have $b_N = \Theta\left(N^{-\left(\frac{1}{L}\right)}\right)$.

$$\Rightarrow \quad N^{\left(\frac{1}{L}\right)}|\chi_{min}|^2 \;\rightarrow\; kW \quad \text{in distribution, for some constant } k > 0.$$

Since $\mathbb{E}\left[|\chi_{min}|^2\right] \leq \mathbb{E}\left[|\chi_1|^2\right] < \infty$, we can use the result in Theorem 2.1 of [56] to conclude that $N^{\left(\frac{1}{L}\right)}\mathbb{E}\left[|\chi_{min}|^2\right] \rightarrow k\mathbb{E}[W] = \Theta(1)$. Thus $\mathbb{E}\left[|\chi_{min}|^2\right] = \Theta\left(N^{-\left(\frac{1}{L}\right)}\right)$. The average throughput of the worst user scheme can now be upper bounded using Jensen's inequality as follows

$$R_{tot} \;=\; N\mathbb{E}\left[\log\left(1+|\chi_{min}|^2P\right)\right] \;\leq\; N\log\left(1+\mathbb{E}\left[|\chi_{min}|^2\right]P\right)$$

$$\Rightarrow \quad R_{tot} \;=\; O\left(N^{\left(\frac{L-1}{L}\right)}\right). \tag{A.17}$$

We lower bound the average throughput of the worst user scheme as follows

$$R_{tot} \;=\; N\int_0^\infty \log(1+xP)\mathrm{d}F_{min}(x) \;\geq\; N\int_{b_N}^\infty \log(1+xP)\mathrm{d}F_{min}(x).$$

$$\Rightarrow \quad R_{tot} \;\geq\; N\log\left(1+b_NP\right)\left[1-F_{min}(b_N)\right],$$

where $F_{min}(x) = 1-(1-F(x))^N$. Using the fact that $F(b_N) = (1/N)$, we get

$$F_{min}(b_N) \;=\; 1-\left(1-\frac{1}{N}\right)^N \;=\; 1-e^{N\log\left(1-\frac{1}{N}\right)} \;=\; 1-e^{-1}\left(1+O\left(\frac{1}{N}\right)\right).$$

$$\Rightarrow \quad R_{tot} \;\geq\; N\log\left(1+b_N P\right)\left[e^{-1}+O\left(\frac{1}{N}\right)\right]$$
$$= \;\Theta\left(N\log\left(1+N^{-\frac{1}{L}}P\right)\right) \;=\; \Theta\left(N^{\left(\frac{L-1}{L}\right)}\right).$$

Combining this with the upper bound in (A.17), we get $R_{tot}=\Theta\left(N^{\left(\frac{L-1}{L}\right)}\right)$.

A.7 Multi-Transmit Antenna Best User Scheduler (Theorem 9)

From the results on extreme order statistics in [54], we know that

$$\left(\frac{|\chi_{max}|^2-a_N}{b_N}\right) \;\rightarrow\; W \qquad \text{in distribution,}$$

where W has a Gumbel distribution and a_N and b_N satisfy $F(a_N)=1-(1/N)$ and $b_N=(1/Nf(a_N))$, where $f(.)$ denotes the probability density function obtained from (2.8). Now

$$F(a_N) \;=\; 1-\frac{1}{N} \quad \Rightarrow \quad \frac{e^{-La_N}(La_N)^{(L-1)}}{(L-1)!}\left(1+O\left(\frac{1}{a_N}\right)\right) \;=\; \frac{1}{N}.$$

Taking log(.) on both sides and simplifying, we get

$$La_N-(L-1)\log a_N \;=\; \log N+(L-1)-\frac{1}{2}\log(L-1)+K.$$
$$\Rightarrow a_N \;=\; \frac{\log N+(L-1)\log\log N}{L}+O(\log\log N).$$

Since

$$f(a_N) \;=\; \frac{Le^{-La_N}(La_N)^{(L-1)}}{(L-1)!} \;=\; \Theta\left(\frac{1}{N}\right),$$

we have $b_N=C=\Theta(1)$. Thus

$$|\chi_{max}|^2-\left(\frac{\log N+(L-1)\log\log N}{L}+O(\log\log N)\right) \;\rightarrow\; CW \quad \text{in distribution.}$$

Using Chebyshev's inequality, it is easy to show that

$$\frac{|\chi_{max}|^2}{\left(\frac{\log N+(L-1)\log\log N}{L}\right)} \;\rightarrow\; 1 \quad \text{in probability.}$$

Since any Chi-squared random variable with $2L$ degrees of freedom can be expressed as the sum of L exponential i.i.d random variables, we have

$$\mathbb{E}\left[|\chi_{max}|^2\right] \;=\; \mathbb{E}\left[\max_{i=1}^{N}\;\left\{\frac{Z_1^i+\cdots+Z_L^i}{L}\right\}\right] \;\leq\; \mathbb{E}\left[\max_{i=1}^{N} Z_1^i\right],$$

where Z_j^i's are exponential random variables with unit mean. Hence

$$\mathbb{E}\left[\frac{|\chi_{max}|^2}{\left(\frac{\log N+(L-1)\log\log N}{L}\right)}\right] \leq \frac{\mathbb{E}\left[\max_{i=1}^{N} Z_1^i\right]}{\left(\frac{\log N+(L-1)\log\log N}{L}\right)} \leq \frac{k\log N}{\left(\frac{\log N+(L-1)\log\log N}{L}\right)} \leq kL < \infty.$$

Thus we can apply the Dominated Convergence Theorem to get

$$\mathbb{E}\left[\frac{|\chi_{max}|^2}{\left(\frac{\log N+(L-1)\log\log N}{L}\right)}\right] \rightarrow 1 \quad \Rightarrow \mathbb{E}\left[|\chi_{max}|^2\right] = \Theta\left(\frac{\log N+(L-1)\log\log N}{L}\right).$$

Using Jensen's inequality, we get

$$R_{tot} \;=\; \mathbb{E}\left[\log\left(1+|\chi_{max}|^2 P\right)\right] \;\leq\; \log\left(1+\mathbb{E}\left[|\chi_{max}|^2\right]P\right).$$

$$\Rightarrow \quad R_{tot} \;=\; O\left(\log\left(1+\frac{\log N+(L-1)\log\log N}{L}\right)\right). \tag{A.18}$$

The average throughput of the best user scheme can be lower bounded as follows

$$R_{tot} \;=\; \int_0^\infty \log(1+xP)\mathrm{d}F_{max}(x) \;\geq\; \int_{a_N}^\infty \log(1+xP)\mathrm{d}F_{max}(x).$$

$$\Rightarrow \quad R_{tot} \;\geq\; \log\left(1+a_N P\right)\left[1-F_{max}(a_N)\right],$$

where $F_{max}(x) = (F(x))^N$. Using the fact that $F(a_N) = 1-\frac{1}{N}$, we get

$$F_{max}(a_N) \;=\; (F(a_N))^N \;=\; \left(1-\frac{1}{N}\right)^N \;=\; e^{-1}\left(1+O\left(\frac{1}{N}\right)\right).$$

$$\Rightarrow \quad R_{tot} \;\geq\; \log\left(1+a_N P\right)\left[1-e^{-1}+O\left(\frac{1}{N}\right)\right] \;=\; \Theta\left(\log\left(1+a_N P\right)\right).$$

$$\Rightarrow R_{tot} \;=\; \Omega\left(\log\left(1+\frac{\log N+(L-1)\log\log N}{L}\right)\right).$$

Combining this with the upper bound in (A.18), we get

$$R_{tot} \;=\; \Theta\left(\log\left(1+\frac{\log N+(L-1)\log\log N}{L}\right)\right).$$

APPENDIX B

PERFECT SECRECY RATES FOR FADING EAVESDROPPER CHANNELS

B.1 Full CSI at the Transmitter (Theorem 10)

We first prove the achievability of (3.4) by showing that for any perfect secrecy rate $R_s < C_s^{(F)}$, there exists a sequence of $(2^{nR_s}, n)$ block codes with average power $\bar{P}$, equivocation rate $R_e > R_s - \epsilon$, and probability of error $P_e^n \to 0$ as $n \to \infty$. Let $R_s = C_s^{(F)} - 3\delta$ for some $\delta > 0$. We quantize the main channel gains $h_M \in [0, M_1]$ into uniform bins $\{h_{M,i}\}_{i=1}^{q_1}$, and the eavesdropper channel gains $h_E \in [0, M_2]$ into uniform bins $\{h_{E,j}\}_{j=1}^{q_2}$. The channels are said to be in state s_{ij} ($i \in [1, q_1]$, $j \in [1, q_2]$), if $h_{M,i} \leq h_M < h_{M,(i+1)}$ and $h_{E,j} \leq h_E < h_{E,(j+1)}$, where $h_{M,(q_1+1)} = M_1, h_{E,(q_2+1)} = M_2$. We also define a power control policy for any state s_{ij} by

$$P(h_{M,i}, h_{E,j}) \;=\; \inf_{h_{M,i} \leq h_M < h_{M,(i+1)}, h_{E,j} \leq h_E < h_{E,(j+1)}} P(h_M, h_E)\ , \tag{B.1}$$

where $P(h_M, h_E)$ is the optimal power allocation policy in (3.7) that satisfies $P(h_M, h_E) = 0$ for all $h_M \leq h_E$, and the power constraint

$$\int_0^\infty \int_{h_E}^\infty P(h_M, h_E) f(h_M) f(h_E) \mathrm{d}h_M \mathrm{d}h_E \;\leq\; \bar{P}. \tag{B.2}$$

Consider a time-invariant AWGN channel with channel gains $h_M \in [h_{M,i}, h_{M,(i+1)})$ and $h_E \in [h_{E,j}, h_{E,(j+1)})$. It is shown in [13, 57] that for this channel, we can develop a sequence of $(2^{n_{ij}(R_s)_{ij}}, n_{ij})$ codes with codeword rate $\log(1 + h_{M,i}P(h_{M,i}, h_{E,j}))$ and perfect secrecy rate

$$(R_s)_{ij} \;=\; \Big[\log(1 + h_{M,i}P(h_{M,i}, h_{E,j})) - \log\big(1 + h_{E,(j+1)}P(h_{M,i}, h_{E,j})\big)\Big]^+, \quad \text{(B.3)}$$

such that the average power is $P(h_{M,i}, h_{E,j})$ and with error probability $P_e^{ij} \to 0$ as $n_{ij} \to \infty$, where

$$n_{ij} \;=\; n \Pr\big(h_{M,i} \le h_M < h_{M,(i+1)}, h_{E,j} \le h_E < h_{E,(j+1)}\big)$$

for sufficiently large n. Note that the expression in (B.3) is obtained by considering the worst case scenario $h_M = h_{M,i}, h_E = h_{E,(j+1)}$ that yields the smallest perfect secrecy rate.

For transmitting the message index $w \in \{1, \cdots, 2^{nR_s}\}$, we first map w to the indices $\{w_{ij}\}$ by dividing the nR_s bits which determine the message index into sets of $n_{ij}(R_s)_{ij}$ bits. The transmitter uses a multiplexing strategy and transmits codewords $\{x_{w_{ij}}\}$ at codeword rate $\log(1 + h_{M,i}P(h_{M,i}, h_{E,j}))$ and perfect secrecy rate $(R_s)_{ij}$, when the channel is in state s_{ij}. As $n \to \infty$, this scheme achieves the perfect secrecy rate (using the ergodicity of the channel),

$$R_s \;=\; \sum_{i=1}^{q_1}\sum_{j=1}^{q_2}\left[\log\left(\frac{1 + h_{M,i}P(h_{M,i}, h_{E,j})}{1 + h_{E,(j+1)}P(h_{M,i}, h_{E,j})}\right)\right]^+ \Pr\begin{pmatrix} h_{M,i} \le h_M < h_{M,(i+1)}, \\ h_{E,j} \le h_E < h_{E,(j+1)} \end{pmatrix}.$$

Thus for a fixed δ, we can find a sufficiently large n such that

$$R_s \;\ge\; \sum_{i=1}^{q_1}\sum_{j=1}^{q_2}\left[\log\left(\frac{1 + h_{M,i}P(h_{M,i}, h_{E,j})}{1 + h_{E,(j+1)}P(h_{M,i}, h_{E,j})}\right)\right]^+ \Pr\begin{pmatrix} h_{M,i} \le h_M < h_{M,(i+1)}, \\ h_{E,j} \le h_E < h_{E,(j+1)} \end{pmatrix} - \delta. \quad \text{(B.4)}$$

For asymptotically large n, using the ergodicity of the channel, the average power of the multiplexing scheme satisfies

$$\sum_{i=1}^{q_1}\sum_{j=1}^{q_2} P(h_{M,i},h_{E,j})\int_{h_{M,i}}^{h_{M,(i+1)}}\int_{h_{E,j}}^{h_{E,(j+1)}} f(h_M)f(h_E)\mathrm{d}h_M\mathrm{d}h_E$$
$$\overset{(a)}{\leq} \int_0^\infty\int_0^\infty P(h_M,h_E)f(h_M)f(h_E)\mathrm{d}h_M\mathrm{d}h_E \overset{(b)}{\leq} \bar{P},$$

where (a) follows from the definition of $P(h_{M,i},h_{E,j})$ in (B.1) and (b) follows from (B.2). Moreover, the error probability of the multiplexing scheme is upper bounded by

$$P_e^n \;\leq\; \sum_{i=1}^{q_1}\sum_{j=1}^{q_2} P_e^{ij} \;\to\; 0, \quad \text{as } n\to\infty.$$

Now since

$$\begin{aligned} C_s^{(F)} &= \int_0^\infty\int_0^\infty \left[\log\left(\frac{1+h_MP(h_M,h_E)}{1+h_EP(h_M,h_E)}\right)\right]^+ f(h_M)f(h_E)\mathrm{d}h_M\mathrm{d}h_E \\ &\leq \int_0^\infty\int_{h_E}^\infty \log\left(\frac{h_M}{h_E}\right) f(h_M)f(h_E)\mathrm{d}h_M\mathrm{d}h_E \;<\; \infty, \end{aligned}$$

(because $\mathbb{E}\{h_M\}<\infty$, $\left|\int_0^1 \log x\ \mathrm{d}x\right| = 1 < \infty$ and $f(h_M), f(h_E)$ are continuous and bounded), there exist M_1 and M_2 for a fixed δ such that

$$\begin{aligned} &\int_0^{M_1}\int_{M_2}^\infty \left[\log\left(\frac{1+h_MP(h_M,h_E)}{1+h_EP(h_M,h_E)}\right)\right]^+ f(h_M)f(h_E)\mathrm{d}h_M\mathrm{d}h_E \;<\; \frac{\delta}{3}\,, \\ &\int_{M_1}^\infty\int_0^{M_2} \left[\log\left(\frac{1+h_MP(h_M,h_E)}{1+h_EP(h_M,h_E)}\right)\right]^+ f(h_M)f(h_E)\mathrm{d}h_M\mathrm{d}h_E \;<\; \frac{\delta}{3}\,, \qquad \text{(B.5)} \\ &\int_{M_1}^\infty\int_{M_2}^\infty \left[\log\left(\frac{1+h_MP(h_M,h_E)}{1+h_EP(h_M,h_E)}\right)\right]^+ f(h_M)f(h_E)\mathrm{d}h_M\mathrm{d}h_E \;<\; \frac{\delta}{3}\,. \end{aligned}$$

Moreover, for fixed M_1 and M_2, the dominated convergence theorem implies that

$$\lim_{(q_1,q_2)\to\infty} \sum_{i=1}^{q_1}\sum_{j=1}^{q_2} \left[\log\left(\frac{1+h_{M,i}P(h_{M,i},h_{E,j})}{1+h_{E,(j+1)}P(h_{M,i},h_{E,j})}\right)\right]^+ \Pr\left(\begin{array}{l} h_{M,i}\leq h_M < h_{M,(i+1)}, \\ h_{E,j}\leq h_E < h_{E,(j+1)} \end{array}\right)$$

$$= \lim_{(q_1,q_2)\to\infty} \sum_{i=1}^{q_1}\sum_{j=1}^{q_2} \int_{h_{M,i}}^{h_{M,(i+1)}}\int_{h_{E,j}}^{h_{E,(j+1)}} \left[\log\left(\frac{1+h_{M,i}P(h_{M,i},h_{E,j})}{1+h_{E,(j+1)}P(h_{M,i},h_{E,j})}\right)\right]^+ f(h_M)f(h_E)\mathrm{d}h_M\mathrm{d}h_E$$

$$= \int_0^{M_1}\int_0^{M_2}\left[\log\left(\frac{1+h_M P(h_M,h_E)}{1+h_E P(h_M,h_E)}\right)\right]^+ f(h_M)f(h_E)\mathrm{d}h_M\mathrm{d}h_E. \quad \text{(B.6)}$$

Choosing M_1, M_2 that satisfy (B.5) and combining (B.5) and (B.6), we see that for a given δ, there exist sufficiently large q_1, q_2 such that

$$\sum_{i=1}^{q_1}\sum_{j=1}^{q_2}\left[\log\left(\frac{1+h_{M,i}P(h_{M,i},h_{E,j})}{1+h_{E,(j+1)}P(h_{M,i},h_{E,j})}\right)\right]^+ \Pr\left(\begin{array}{l} h_{M,i}\le h_M < h_{M,(i+1)}, \\ h_{E,j}\le h_E < h_{E,(j+1)} \end{array}\right)$$

$$\ge \int_0^{\infty}\int_0^{\infty}\left[\log\left(\frac{1+h_M P(h_M,h_E)}{1+h_E P(h_M,h_E)}\right)\right]^+ f(h_M)f(h_E)\mathrm{d}h_M\mathrm{d}h_E \;-\; 2\delta. \quad \text{(B.7)}$$

Combining (B.4) and (B.7), we get the desired result.

We now prove the converse part by showing that for any perfect secrecy rate R_s with equivocation rate $R_e > R_s - \epsilon$ and error probability $P_e^n \to 0$ as $n \to \infty$, there exists a power allocation policy $P(h_M, h_E)$ satisfying the average power constraint, such that

$$R_s \;\le\; \iint\left[\log\left(\frac{1+h_M P(h_M,h_E)}{1+h_E P(h_M,h_E)}\right)\right]^+ f(h_M)f(h_E)\mathrm{d}h_M\mathrm{d}h_E.$$

Consider any sequence of $(2^{nR_s}, n)$ codes with perfect secrecy rate R_s and equivocation rate R_e, such that $R_e > R_s - \epsilon$, with average power less than or equal to $\bar{P}$ and error probability $P_e^n \to 0$ as $n \to \infty$. Let $N(h_M, h_E)$ denote the number of times the channel is in fading state (h_M, h_E) over the interval $[0, n]$. Also let $P^n(h_M, h_E) = \mathbb{E}\big\{\sum_{i=1}^n |x_w(i)|^2 \mathbf{1}_{\{h_M(i)=h_M, h_E(i)=h_E\}}\big\}$, where $\{x_w\}$ are the codewords corresponding to the message w and the expectation is taken over all codewords. We note that the equivocation $H(W|Z^n, h_M^n, h_E^n)$ only depends on the marginal distribution of Z^n, and thus does not depend on whether $Z(i)$ is a physically or stochastically degraded version of $Y(i)$ or vice versa. Hence we assume in the following derivation that for any fading state, either $Z(i)$ is a physically degraded version of $Y(i)$ or vice versa

(since the noise processes are Gaussian), depending on the instantaneous channel state. Thus we have

$$
\begin{aligned}
nR_e &= H(W|Z^n, h_M^n, h_E^n) \\
&\overset{(a)}{\leq} H(W|Z^n, h_M^n, h_E^n) - H(W|Z^n, Y^n, h_M^n, h_E^n) + n\delta_n \\
&= I(W; Y^n|Z^n, h_M^n, h_E^n) + n\delta_n \\
&\overset{(b)}{\leq} I(X^n; Y^n|Z^n, h_M^n, h_E^n) + n\delta_n \\
&= H(Y^n|Z^n, h_M^n, h_E^n) - H(Y^n|X^n, Z^n, h_M^n, h_E^n) + n\delta_n \\
&= \sum_{i=1}^{n} \left[H(Y(i)|Y^{i-1}, Z^n, h_M^n, h_E^n) - H(Y(i)|Y^{i-1}, X^n, Z^n, h_M^n, h_E^n)\right] + n\delta_n \\
&\overset{(c)}{\leq} \sum_{i=1}^{n} \left[H(Y(i)|Z(i), h_M(i), h_E(i)) - H(Y(i)|X(i), Z(i), h_M(i), h_E(i))\right] + n\delta_n \\
&= \sum_{i=1}^{n} I(X(i); Y(i)|Z(i), h_M(i), h_E(i)) + n\delta_n \\
&= \sum_{i=1}^{n} \iint I(X; Y|Z, h_M, h_E) \mathbf{1}_{\{h_M(i)=h_M, h_E(i)=h_E\}} \mathrm{d}h_M \mathrm{d}h_E + n\delta_n \qquad \text{(B.8)} \\
&= \iint I(X; Y|Z, h_M, h_E) N(h_M, h_E) \mathrm{d}h_M \mathrm{d}h_E + n\delta_n \\
&\overset{(d)}{\leq} \iint N(h_M, h_E) \left[\log\left(\frac{1 + h_M P^n(h_M, h_E)}{1 + h_E P^n(h_M, h_E)}\right)\right]^+ \mathrm{d}h_M \mathrm{d}h_E + n\delta_n.
\end{aligned}
$$

In the above derivation, (a) follows from the Fano inequality, (b) follows from the data processing inequality since $W \to X^n \to (Y^n, Z^n)$ forms a Markov chain, (c) follows from the fact that conditioning reduces entropy and from the memoryless property of the channel, (d) follows from the fact that given h_M and h_E, the fading channel reduces to an AWGN channel with channel gains (h_M, h_E) and average transmission power $P^n(h_M, h_E)$, for which

$$
I(X; Y|Z, h_M, h_E) \;\leq\; \left[\log(1 + h_M P^n(h_M, h_E)) - \log(1 + h_E P^n(h_M, h_E))\right]^+ \;,
$$

as shown in [13, 57]. Since the codewords satisfy the power constraint, we have

$$\iint P^n(h_M, h_E) \left(\frac{N(h_M, h_E)}{n} \right) \mathrm{d}h_M \mathrm{d}h_E \ \leq \ \bar{P}.$$

For any h_M, h_E such that $f(h_M, h_E) \neq 0$, $\{P^n(h_M, h_E)\}$ are bounded sequences in n. Thus there exists a subsequence that converges to a limit $P(h_M, h_E)$ as $n \to \infty$. Since for each n, the power constraint is satisfied, we have

$$\iint P(h_M, h_E) f(h_M) f(h_E) \mathrm{d}h_M \mathrm{d}h_E \ \leq \ \bar{P}. \tag{B.9}$$

Now, we have

$$R_e \ \leq \ \iint \frac{N(h_M, h_E)}{n} \left[\log \left(\frac{1 + h_M P^n(h_M, h_E)}{1 + h_E P^n(h_M, h_E)} \right) \right]^+ \mathrm{d}h_M \mathrm{d}h_E + \delta_n.$$

Taking the limit along the convergent subsequence and using the ergodicity of the channel, we get

$$R_e \ \leq \ \iint \left[\log \left(\frac{1 + h_M P(h_M, h_E)}{1 + h_E P(h_M, h_E)} \right) \right]^+ f(h_M) f(h_E) \mathrm{d}h_M \mathrm{d}h_E + \delta_n.$$

The claim is thus proved.

B.2 Main Channel CSI at the Transmitter (Theorem 11)

Let $R_s = C_s^{(M)} - \delta$ for some small $\delta > 0$. Let $n = n_1 m$, where n_1 represents the number of symbols transmitted in each coherence interval, and m represents the number of coherence intervals over which the message W is transmitted. Let $R = \mathbb{E}\{\log(1 + h_M P(h_M))\} - \epsilon$. We first generate all binary sequences $\{\mathbf{V}\}$ of length nR and then independently assign each of them randomly to one of 2^{nR_s} groups, according to a uniform distribution. This ensures that any of the sequences are equally likely to be within any of the groups. Each secret message $w \in \{1, \cdots, 2^{nR_s}\}$

is then assigned a group $\mathbf{V}(w)$. To encode a particular message w, the stochastic encoder randomly selects a sequence $\mathbf{v}$ from the corresponding group $\mathbf{V}(w)$, according to a uniform distribution. This sequence $\mathbf{v}$ consisting of nR bits is then subdivided into independent blocks $\{\mathbf{v}(1), \cdots, \mathbf{v}(m)\}$, where the block $\mathbf{v}(i)$ consists of $n_1 \left[\log\left(1 + h_M(i)P(h_M(i))\right) - \epsilon\right]$ bits, and is transmitted in the i^{th} coherence interval ($i \in \{1, \cdots, m\}$). As $m \to \infty$, using the ergodicity of the channel, we have

$$\begin{aligned} & \lim_{m\to\infty} \sum_{i=1}^{m} n_1 \left[\log\left(1 + h_M(i)P(h_M(i))\right) - \epsilon\right] \\ = \; & n_1 m \left[\mathbb{E}\{\log\left(1 + h_M P(h_M)\right)\} - \epsilon\right] \;=\; nR. \end{aligned}$$

We then generate i.i.d. Gaussian codebooks $\{X^{n_1}(i) : i = 1, \cdots, m\}$ consisting of $2^{n_1[\log(1+h_M(i)P(h_M(i)))-\epsilon]}$ codewords, each of length n_1 symbols. In the i^{th} coherence interval, the transmitter encodes the block $\mathbf{v}(i)$ into the codeword $x^{n_1}(i)$, which is then transmitted over the fading channel. The legitimate receiver receives $y^{n_1}(i)$ while the eavesdropper receives $z^{n_1}(i)$ in the i^{th} coherence interval. We denote vectors of the form $\{X^{n_1}(1), \cdots, X^{n_1}(m)\}$ as $X^{n_1}(1:m)$. The equivocation rate at the eavesdropper can then be lower bounded as follows.

$$\begin{aligned} nR_e &= H(W|Z^{n_1}(1:m), h_M^n, h_E^n) \\ &= H(W, Z^{n_1}(1:m)|h_M^n, h_E^n) - H(Z^{n_1}(1:m)|h_M^n, h_E^n) \\ &= H(W, Z^{n_1}(1:m), X^{n_1}(1:m)|h_M^n, h_E^n) - H(Z^{n_1}(1:m)|h_M^n, h_E^n) \\ &\qquad - \underbrace{H(X^{n_1}(1:m)|W, Z^{n_1}(1:m), h_M^n, h_E^n)}_{A} \\ &= H(X^{n_1}(1:m)|h_M^n, h_E^n) + H(W, Z^{n_1}(1:m)|X^{n_1}(1:m), h_M^n, h_E^n) \\ &\qquad - H(Z^{n_1}(1:m)|h_M^n, h_E^n) - A \end{aligned}$$

$$
\begin{aligned}
&\geq H(X^{n_1}(1:m)|h_M^n, h_E^n) + H(Z^{n_1}(1:m)|X^{n_1}(1:m), h_M^n, h_E^n) \\
&\qquad\qquad -H(Z^{n_1}(1:m)|h_M^n, h_E^n) - A \\
&= H(X^{n_1}(1:m)|h_M^n, h_E^n) - I(Z^{n_1}(1:m); X^{n_1}(1:m)|h_M^n, h_E^n) - A \\
&= H(X^{n_1}(1:m)|Z^{n_1}(1:m), h_M^n, h_E^n) - A \\
&\overset{(a)}{=} \sum_{i=1}^{m} H(X^{n_1}(i)|Z^{n_1}(i), h_M(i), h_E(i)) - A \\
&\overset{(b)}{\geq} \sum_{i\in\mathcal{N}_m} H(X^{n_1}(i)|Z^{n_1}(i), h_M(i), h_E(i)) - A \\
&= \sum_{i\in\mathcal{N}_m} [H(X^{n_1}(i)|h_M(i), h_E(i)) - I(X^{n_1}(i); Z^{n_1}(i)|h_M(i), h_E(i))] - A \\
&\geq \sum_{i\in\mathcal{N}_m} n_1 \left[\log\left(1 + h_M(i)P(h_M(i))\right) - \log\left(1 + h_E(i)P(h_M(i))\right) - \epsilon\right] - A \\
&\geq \sum_{i=1}^{m} n_1 \left\{ \left[\log\left(\frac{1 + h_M(i)P(h_M(i)}{1 + h_E(i)P(h_M(i)}\right)\right]^+ - \epsilon \right\} - A \\
&\overset{(c)}{=} nC_s^{(M)} - A - n\epsilon. \qquad \text{(B.10)}
\end{aligned}
$$

In the above derivation, (a) follows from the memoryless property of the channel and the independence of the $X^{n_1}(i)$'s, (b) is obtained by removing all those terms which correspond to the coherence intervals $i \notin \mathcal{N}_m$, where the set $\mathcal{N}_m$ is defined as $\mathcal{N}_m = \{i \in \{1, \cdots, m\} : h_M(i) > h_E(i)\}$, and (c) follows from the ergodicity of the channel as $m \to \infty$.

Now we show that the term $A = H(X^{n_1}(1:m)|W, Z^{n_1}(1:m), h_M^n, h_E^n)$ vanishes as $m, n_1 \to \infty$ by using a list decoding argument. In this list decoding, at coherence interval i, the eavesdropper first constructs a list $\mathcal{L}_i$ such that $x^{n_1}(i) \in \mathcal{L}_i$ if $(x^{n_1}(i), z^{n_1}(i))$ are jointly typical. Let $\mathcal{L} = \mathcal{L}_1 \times \mathcal{L}_2 \times \cdots \times \mathcal{L}_m$. Given w, the eavesdropper declares that $\hat{x}^n = (x^{n_1}(1), \cdots, x^{n_1}(m))$ was transmitted, if $\hat{x}^n$ is the only codeword such that $\hat{x}^n \in B(w) \cap \mathcal{L}$, where $B(w)$ is the set of codewords corresponding to the message w. If the eavesdropper finds none or more than one such sequence,

then it declares an error. Hence, there are two type of error events: 1) $\mathcal{E}_1$: the transmitted codeword x_t^n is not in $\mathcal{L}$, 2) $\mathcal{E}_2$: $\exists x^n \neq x_t^n$ such that $x^n \in B(w) \cap \mathcal{L}$. Thus the error probability $\Pr(\hat{x}^n \neq x_t^n) = \Pr(\mathcal{E}_1 \cup \mathcal{E}_2) \leq \Pr(\mathcal{E}_1) + \Pr(\mathcal{E}_2)$. Based on the AEP, we know that $\Pr(\mathcal{E}_1) \leq \epsilon_1$. In order to bound $\Pr(\mathcal{E}_2)$, we first bound the size of $\mathcal{L}_i$. We let

$$\phi_i(x^{n_1}(i)|z^{n_1}(i)) \;=\; \begin{cases} 1, & \text{when } (x^{n_1}(i), z^{n_1}(i)) \text{ are jointly typical,} \\ 0, & \text{otherwise.} \end{cases} \tag{B.11}$$

Now

$$\begin{aligned} \mathbb{E}\{\|\mathcal{L}_i\|\} &= \mathbb{E}\left\{\sum_{x^{n_1}(i)} \phi_i(x^{n_1}(i)|z^{n_1}(i))\right\} \\ &\leq \mathbb{E}\left\{1 + \sum_{x^{n_1}(i)\neq x_t^{n_1}(i)} \phi_i(x^{n_1}(i)|z^{n_1}(i))\right\} \\ &\leq 1 + \sum_{x^{n_1}(i)\neq x_t^{n_1}(i)} \mathbb{E}\left\{\phi_i(x^{n_1}(i)|z^{n_1}(i))\right\} \\ &\leq 1 + 2^{n_1[\log(1+h_M(i)P(h_M(i)))-\log(1+h_E(i)P(h_M(i)))-\epsilon]} \\ &\leq 2^{n_1\left([\log(1+h_M(i)P(h_M(i)))-\log(1+h_E(i)P(h_M(i)))-\epsilon]^+ + \frac{1}{n_1}\right)}. \end{aligned} \tag{B.12}$$

Hence

$$\begin{aligned} \mathbb{E}\{\|\mathcal{L}\|\} &= \prod_{i=1}^{m} \mathbb{E}\{\|\mathcal{L}_i\|\} \\ &\leq 2^{\sum_{i=1}^{m} n_1\left([\log(1+h_M(i)P(h_M(i)))-\log(1+h_E(i)P(h_M(i)))-\epsilon]^+ + \frac{1}{n_1}\right)}. \end{aligned} \tag{B.13}$$

Thus

$$\begin{aligned} \Pr(\mathcal{E}_2) &\leq \mathbb{E}\left\{\sum_{x^n\in\mathcal{L}, x^n\neq x_t^n} \Pr(x^n \in B(w))\right\} \\ &\overset{(a)}{\leq} \mathbb{E}\left\{\|\mathcal{L}\| 2^{-nR_s}\right\} \end{aligned}$$

$$
\begin{aligned}
&\leq \; 2^{-nR_s} 2^{\sum_{i=1}^{m} n_1 \left([\log(1+h_M(i)P(h_M(i))) - \log(1+h_E(i)P(h_M(i))) - \epsilon]^+ + \frac{1}{n_1} \right)} \\
&\leq \; 2^{-n\left(R_s - \frac{1}{m} \sum_{i=1}^{m} \left([\log(1+h_M(i)P(h_M(i))) - \log(1+h_E(i)P(h_M(i))) - \epsilon]^+ + \frac{1}{n_1} \right) \right)} \\
&= \; 2^{-n\left(R_s - \frac{1}{m} \sum_{i=1}^{m} \left([\log(1+h_M(i)P(h_M(i))) - \log(1+h_E(i)P(h_M(i)))]^+ + \frac{1}{n_1} \right) + \frac{|\mathcal{N}_m|\epsilon}{m} \right)},
\end{aligned}
$$

where (a) follows from the uniform distribution of the codewords in $B(w)$. Now as $n_1 \to \infty$ and $m \to \infty$, we get

$$
\Pr(\mathcal{E}_2) \;\leq\; 2^{-n(C_s - \delta - C_s + c\epsilon)} \;=\; 2^{-n(c\epsilon - \delta)},
$$

where $c = \Pr(h_M > h_E)$. Thus, by choosing $\epsilon > (\delta/c)$, the error probability $\Pr(\mathcal{E}_2) \to 0$ as $n \to \infty$. Now using Fano's inequality, we get

$$
A \;=\; H(X^{n_1}(1:m)|W, Z^{n_1}(1:m), h_M^n, h_E^n) \;\leq\; n\delta_n \qquad \to 0 \qquad \text{as } n \to \infty.
$$

Combining this with (B.10), we get the desired result.

For the converse part, consider any sequence of $(2^{nR_s}, n)$ codes with perfect secrecy rate R_s and equivocation rate R_e, such that $R_e > R_s - \epsilon$, with average power less than or equal to $\bar{P}$ and error probability $P_e^n \to 0$ as $n \to \infty$. We follow the same steps used in the proof of the converse in Theorem 10 with the only difference that now the transmission power $P^n(.)$ only depends on h_M. From (B.8), we get

$$
\begin{aligned}
nR_e &\leq \sum_{i=1}^{n} \iint I(X;Y|Z,h_M,h_E) \mathbf{1}_{\{h_M(i)=h_M, h_E(i)=h_E\}} \mathrm{d}h_M \mathrm{d}h_E + n\delta_n \\
&= \iint I(X;Y|Z,h_M,h_E) N(h_M,h_E) \mathrm{d}h_M \mathrm{d}h_E + n\delta_n \\
&\leq \iint N(h_M,h_E) \left[\log\left(\frac{1 + h_M P^n(h_M)}{1 + h_E P^n(h_M)} \right) \right]^+ \mathrm{d}h_M \mathrm{d}h_E + n\delta_n.
\end{aligned}
$$

This follows from the fact that given h_M and h_E, the fading channel reduces to an AWGN channel with channel gains (h_M, h_E) and average transmission power $P^n(h_M)$, for which Gaussian inputs are known to be optimal [13, 57].

Similar to the proof of Theorem 10, we take the limit over the convergent subsequence and use the ergodicity of the channel to obtain

$$R_e \leq \iint \left[\log\left(\frac{1+h_M P(h_M)}{1+h_E P(h_M)}\right)\right]^+ f(h_M) f(h_E) \mathrm{d}h_M \mathrm{d}h_E + \delta_n, \tag{B.14}$$

where $\mathbb{E}\{P(h_M)\} \leq \bar{P}$. The claim is thus proved.

B.3 ARQ Feedback to the Transmitter (Theorem 12)

Since the transmitter does not have any knowledge of the main and eavesdropper channels, we adopt a transmission strategy with constant rate R and power P. Let the achievable secrecy rate be R_s. We first generate all binary sequences $\{\mathbf{V}\}$ of length nR and then independently assign each of them randomly to one of 2^{nR_s} groups, according to a uniform distribution. This ensures that any of the sequences are equally likely to be within any of the groups. Each secret message $w \in \{1, \cdots, 2^{nR_s}\}$ is then assigned a group $\mathbf{V}(w)$. To encode a particular message w, the stochastic encoder randomly selects a sequence $\mathbf{v}$ from the corresponding group $\mathbf{V}(w)$, according to a uniform distribution. This sequence $\mathbf{v}$ consisting of nR bits is then sub-divided into m independent blocks $\mathbf{v}(1), \cdots, \mathbf{v}(m)$, where each block $\mathbf{v}(i)$ consists of $n_1 R$ bits. Here n_1 is the length of a coherence interval and $n = n_1 m$.

From this point, the encoding scheme differs depending on whether we want to use Incremental Redundancy ARQ or Repetition ARQ. We first consider the encoding scheme for the IR-ARQ case. To transmit a given $\mathbf{v}(i)$ in this case, we first generate an i.i.d Gaussian codebook consisting of $2^{n_1 R}$ codewords of length n_1, and transmit the codeword $X_1^{n_1}(i)$ (the subscript here represents the transmission round) corresponding to $\mathbf{v}(i)$. The observations received at the destination and the eavesdropper

are denoted by $Y_1^{n_1}(i)$ and $Z_1^{n_1}(i)$ respectively. Let $h_{M,k}(i)$ and $h_{E,k}(i)$ denote the fading power gains of the main and eavesdropper channels during the k^{th} transmission round for $\mathbf{v}(i)$. If the transmission rate R is less than the instantaneous capacity of the main channel $\log(1 + h_{M,1}(i)P)$, then the destination can successfully decode $\mathbf{v}(i)$ and the transmitter proceeds with the transmission of $\mathbf{v}(i+1)$. Otherwise, a NACK is fed back to the transmitter. On receiving a NACK, the transmitter forms another i.i.d Gaussian codebook of size $2^{n_1 R}$, independent from the first one, and transmits the new codeword $X_2^{n_1}(i)$ corresponding to $\mathbf{v}(i)$. At the destination, the decoding for $\mathbf{v}(i)$ is performed jointly using both $Y_1^{n_1}(i)$ and $Y_2^{n_1}(i)$. Again if the rate R is less than the new mutual information $\sum_{k=1}^{2} \log(1 + h_{M,k}(i)P)$, the destination can successfully decode $\mathbf{v}(i)$, otherwise it sends back another NACK bit to the transmitter. A similar procedure is followed until the destination successfully decodes $\mathbf{v}(i)$, i.e., when $R \leq \sum_{k=1}^{L_i} [\log(1 + h_{M,k}(i)P)]$, where L_i denotes the number of transmission rounds required for successfully decoding $\mathbf{v}(i)$. Then the same IR-ARQ strategy is applied for transmitting the next block $\mathbf{v}(i+1)$.

For the Rep-ARQ scheme, once a NACK bit is fed back to the transmitter, it merely repeats the codeword $X_1^{n_1}(i)$ in each transmission round (instead of generating a new i.i.d Gaussian codebook). This repetition procedure continues until the destination successfully decodes $\mathbf{v}(i)$, i.e., when $R \leq \log\left[1 + \sum_{k=1}^{L_i} h_{M,k}(i)P\right]$. We note that the eavesdropper will be able to successfully decode the block $\mathbf{v}(i)$ only if $R \leq \sum_{k=1}^{L_i} [\log(1 + h_{E,k}(i)P)]$ and $R \leq \log\left[1 + \sum_{k=1}^{L_i} h_{E,k}(i)P\right]$ for the IR-ARQ and Rep-ARQ schemes respectively.

We now introduce the following notations: The number of transmission rounds for the m different blocks are denoted by $\{L_1, L_2, \cdots, L_m\}$. For any block $\mathbf{v}(i)$, let

$\mathbf{X}^{n_1}(i) = \{X_1^{n_1}(i), \cdots, X_{L_i}^{n_1}(i)\}$ be the corresponding independent codewords transmitted until successful decoding at the destination. Let $\mathbf{Y}^{n_1}(i) = \left\{Y_1^{n_1}(i), \cdots, Y_{L_i}^{n_1}(i)\right\}$ and $\mathbf{Z}^{n_1}(i) = \left\{Z_1^{n_1}(i), \cdots, Z_{L_i}^{n_1}(i)\right\}$ be the corresponding received sequences at the destination and eavesdropper respectively. Also let $\mathbf{h}_M(i) = \{h_{M,1}, \cdots, h_{M,L_i}\}$ and $\mathbf{h}_E(i) = \{h_{E,1}, \cdots, h_{E,L_i}\}$ denote the main and eavesdropper channel power gains during the transmission of block $\mathbf{v}(i)$, and $\mathbf{h}_M = \{\mathbf{h}_M(1), \cdots, \mathbf{h}_M(m)\}$ and $\mathbf{h}_E = \{\mathbf{h}_E(1), \cdots, \mathbf{h}_E(m)\}$ denote the entire vector of channel power gains at the destination and eavesdropper respectively. Since a message w is mapped to m blocks and each block $\mathbf{v}(i)$ is transmitted for L_i transmission rounds, the total number of channel uses required for transmitting the message is given by $n_1(L_1 + \cdots + L_m)$. We now calculate the equivocation rate at the eavesdropper as follows:

$$
\begin{aligned}
& n_1(L_1 + L_2 + \cdots + L_m)R_e \\
&= H(W|\mathbf{Z}^{n_1}(1:m), \mathbf{h}_M, \mathbf{h}_E) \\
&= H(W, \mathbf{Z}^{n_1}(1:m)|\mathbf{h}_M, \mathbf{h}_E) - H(\mathbf{Z}^{n_1}(1:m)|\mathbf{h}_M, \mathbf{h}_E) \\
&= H(W, \mathbf{Z}^{n_1}(1:m), \mathbf{X}^{n_1}(1:m)|\mathbf{h}_M, \mathbf{h}_E) - H(\mathbf{Z}^{n_1}(1:m)|\mathbf{h}_M, \mathbf{h}_E) \\
&\qquad - \underbrace{H(\mathbf{X}^{n_1}(1:m)|W, \mathbf{Z}^{n_1}(1:m), \mathbf{h}_M, \mathbf{h}_E)}_{B} \\
&= H(\mathbf{X}^{n_1}(1:m)|\mathbf{h}_M, \mathbf{h}_E) + H(W, \mathbf{Z}^{n_1}(1:m)|\mathbf{X}^{n_1}(1:m), \mathbf{h}_M, \mathbf{h}_E) \\
&\qquad - H(\mathbf{Z}^{n_1}(1:m)|\mathbf{h}_M, \mathbf{h}_E) - B \\
&\geq H(\mathbf{X}^{n_1}(1:m)|\mathbf{h}_M, \mathbf{h}_E) + H(\mathbf{Z}^{n_1}(1:m)|\mathbf{X}^{n_1}(1:m), \mathbf{h}_M, \mathbf{h}_E) \\
&\qquad - H(\mathbf{Z}^{n_1}(1:m)|\mathbf{h}_M, \mathbf{h}_E) - B \\
&= H(\mathbf{X}^{n_1}(1:m)|\mathbf{h}_M, \mathbf{h}_E) - I(\mathbf{Z}^{n_1}(1:m); \mathbf{X}^{n_1}(1:m)|\mathbf{h}_M, \mathbf{h}_E) - B \\
&= H(\mathbf{X}^{n_1}(1:m)|\mathbf{Z}^{n_1}(1:m), \mathbf{h}_M, \mathbf{h}_E) - B
\end{aligned}
$$

$$
\begin{aligned}
&\overset{(a)}{=} \sum_{i=1}^{m} H(\mathbf{X}^{n_1}(i)|\mathbf{Z}^{n_1}(i), \mathbf{h_M}(i), \mathbf{h_E}(i)) - B \\
&\overset{(b)}{\geq} \sum_{i\in\mathcal{N}_m} H(\mathbf{X}^{n_1}(i)|\mathbf{Z}^{n_1}(i), \mathbf{h_M}(i), \mathbf{h_E}(i)) - B \\
&= \sum_{i\in\mathcal{N}_m} [H(\mathbf{X}^{n_1}(i)|\mathbf{h_M}(i), \mathbf{h_E}(i)) - I(\mathbf{X}^{n_1}(i); \mathbf{Z}^{n_1}(i)|\mathbf{h_M}(i), \mathbf{h_E}(i))] - B \\
&= \sum_{i\in\mathcal{N}_m} n_1 \left[R - \sum_{k=1}^{L_i} \log\left(1 + h_{E,k}(i)P\right)\right] - B \\
&\geq \sum_{i=1}^{m} n_1 \left[R - \sum_{k=1}^{L_i} \log\left(1 + h_{E,k}(i)P\right)\right]^+ - B.
\end{aligned}
$$

In the above derivation, (a) follows from the memoryless property of the channel and the independence of the $\mathbf{X}^{n_1}(i)$'s, and (b) is obtained by removing all those terms which correspond to the blocks which can be succesfully decoded by the eavesdropper, i.e.,

$$
\mathcal{N}_m = \left\{ i \in \{1, \cdots, m\} : R > \sum_{k=1}^{L_i} \log\left(1 + h_{E,k}(i)P\right) \right\}.
$$

The term $B = H(\mathbf{X}^{n_1}(1:m)|W, \mathbf{Z}^{n_1}(1:m), \mathbf{h}_M, \mathbf{h}_E)$ can be shown to vanish as $m, n_1 \to \infty$ using a list decoding argument similar to the one in Appendix B.2. Thus, as $m, n_1 \to \infty$, we get

$$
n_1(L_1 + L_2 + \cdots + L_m)R_e \geq \sum_{i=1}^{m} n_1 \left[R - \sum_{k=1}^{L_i} \log\left(1 + h_{E,k}(i)P\right)\right]^+ - \epsilon.
$$

Since $m \to \infty$, using the ergodicity of the main and eavesdropper channels, we get

$$
n\mathbb{E}[L]R_e \geq n\mathbb{E}\left[\left(R - \sum_{k=1}^{L} \log\left(1 + h_{E,k}P\right)\right)^+\right] - \epsilon,
$$

where the expectation on the right is taken over both the number of transmission rounds L and the wiretapper channel gains $\{h_{E,k}\}$. Thus

$$
R_e \geq \frac{\mathbb{E}\left[\left(R - \sum_{k=1}^{L} \log\left(1 + h_{E,k}P\right)\right)^+\right]}{\mathbb{E}[L]} - \epsilon'.
$$

The perfect secrecy rate achieved by the Rep-ARQ scheme can also be derived in a similar manner, with the only difference being that for the Rep-ARQ scheme we have

$$I(\mathbf{X}^{n_1}(i); \mathbf{Z}^{n_1}(i) | \mathbf{h_M}(i), \mathbf{h_E}(i)) \;=\; \log \left[1 + \sum_{k=1}^{L_i} h_{E,k}(i) P \right].$$

Using this fact in the above derivation yields (3.16), which completes the proof of the theorem.

BIBLIOGRAPHY

[1] R. Knopp and P. Humblet, "Information capacity and power control in single cell multiuser communications," in *IEEE International Computer Conference (ICC'95)*, (Seattle, WA), June 1995.

[2] D. N. C. Tse, "Optimal power allocation over parallel gaussian channels," in *International Symposium on Information Theory*, (Ulm, Germany), June 1997.

[3] P. Viswanath, D. N. C. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Transactions on Information Theory*, vol. 48, pp. 1277–1294, June 2002.

[4] M. Sharif and B. Hassibi, "Delay considerations for opportunistic scheduling in broadcast fading channels," *To appear in IEEE Transactions on Wireless Communications*, 2006.

[5] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Transactions on Information Theory*, vol. 46, March 2000.

[6] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity - Part I: System description," *IEEE Transactions on Communications*, vol. 51, pp. 1927–1938, November 2003.

[7] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity - Part II: Implementation aspects and performance analysis," *IEEE Transactions on Communications*, vol. 51, pp. 1939–1948, November 2003.

[8] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Transactions on Information Theory*, vol. 50, pp. 3062–3080, December 2004.

[9] K. Azarian, H. El Gamal, and P. Schniter, "On the achievable diversity-multiplexing tradeoff in half-duplex cooperative channels," *IEEE Transactions on Information Theory*, vol. 51, pp. 4152–4172, December 2005.

[10] P. K. Gopala and H. El Gamal, "On the scaling laws of multi-modal wireless sensor networks," in *Proceedings of IEEE Infocom 2004*, pp. 558–563, March 2004.

[11] C. E. Shannon, "Communication theory of secrecy systems," *Bell System Technical Journal*, vol. 28, pp. 656–715, October 1949.

[12] A. D. Wyner, "The wire-tap channel," *Bell System Technical Journal*, vol. 54, no. 8, pp. 1355–1387, 1975.

[13] S. K. Leung-Yan-Cheong and M. E. Hellman, "The gaussian wire-tap channel," *IEEE Transactions on Information Theory*, vol. 24, pp. 451–456, July 1978.

[14] I. Csiszar and J. Korner, "Broadcast channels with confidential messages," *IEEE Transactions on Information Theory*, vol. 24, pp. 339–348, May 1978.

[15] I. E. Telatar and R. G. Gallager, "Combining queueing theory with information theory for multiaccess," *IEEE Journal on Selected Areas in Communications*, vol. 13, pp. 963–969, August 1995.

[16] A. Ephremides and B. Hajek, "Information theory and communication networks: An unconsummated union," *IEEE Transactions on Information Theory*, vol. 44, pp. 2416–2434, October 1998.

[17] G. Dimic, R. Zhang, and N. D. Sidiropoulos, "Medium access control - physical cross-layer design," in *IEEE Signal Processing Magazine*, vol. 21, pp. 40–58, September 2004.

[18] R. Berry and E. Yeh, "Cross-layer wireless resource allocation," in *IEEE Signal Processing Magazine*, vol. 21, pp. 59–68, September 2004.

[19] C. E. Shannon, "The zero-error capacity of a noisy channel," *IRE Transactions on Information Theory*, vol. 2, pp. 8–19, 1956.

[20] M. V. Burnashev, "Data transmission over a discrete channel with feedback: Random transmission time," *Problems of Information Transmission*, vol. 12, no. 4, pp. 250–265, 1976.

[21] G. D. Forney Jr., "Exponential error bounds for erasure, list, and decision feedback schemes," *IEEE Transactions on Information Theory*, vol. 14, pp. 206–220, March 1968.

[22] G. Caire and D. Tuninetti, "The throughput of hybrid-ARQ protocols for the gaussian collision channel," *IEEE Transactions on Information Theory*, vol. 47, pp. 1971–1988, July 2001.

[23] H. El Gamal, G. Caire, and M. O. Damen, "The MIMO ARQ channel: Diversity-multiplexing-delay tradeoff," *IEEE Transactions on Information Theory*, vol. 52, August 2006.

[24] P. K. Gopala and H. El Gamal, "Scheduling for cellular multicast: A cross-layer perspective," *Submitted to IEEE Transactions on Mobile Computing*, 2007.

[25] P. K. Gopala, L. Lai, and H. El Gamal, "On the secrecy capacity of fading channels," *Submitted to IEEE Transactions on Information Theory*, 2006.

[26] P. K. Gopala, Y. H. Nam, and H. El Gamal, "On the error exponents of ARQ channels with deadlines," *To appear in IEEE Transactions on Information Theory*, 2007.

[27] R. Berry and R. Gallager, "Communication over fading channels with delay constraints," *IEEE Transactions on Information Theory*, vol. 48, pp. 1135–1149, May 2002.

[28] D. Rajan, A. Sabharwal, and B. Aazhang, "Delay bounded packet scheduling of bursty sources over wireless channels," *IEEE Transactions on Information Theory*, vol. 50, pp. 125–144, January 2004.

[29] E. M. Yeh and A. S. Cohen, "Information theory, queueing, and resource allocation in multi-user fading communications," in *38th Annual Conference on Information Sciences and Systems*, March 2004.

[30] J. Zhang and D. Zheng, "Ad hoc networking over fading channels: Multi-channel diversity, MIMO signaling, and opportunistic medium access control," in *41st Allerton Conference on Communications, Control, and Computing*, October 2003.

[31] P. Liu, R. Berry, and M. Honig, "Delay-sensitive packet scheduling in wireless networks," in *IEEE WCNC 2003*, March 2003.

[32] S. Shakkottai and A. Stolyar, "Scheduling for multiple flows sharing a time-varying channel: The exponential rule," *American Mathematical Society Translations, Series 2*, vol. 207, 2002.

[33] M. Airy, S. Shakkottai, and R. Heath Jr., "Spatially greedy scheduling in multi-user MIMO wireless systems," in *IEEE Asilomar Conf. on Signals, Systems, and Computers*, November 2003.

[34] P. Chaporkar and S. Sarkar, "On-line optimal wireless multicast," in *2nd Workshop On Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks*, (Cambridge, England), pp. 282–291, March 2004.

[35] C. Wu and Y. Tay, "Amris: A multicast protocol for ad hoc wireless networks," in *Proceedings of IEEE MILCOM 99*, (Atlantic City, NJ), November 1999.

[36] J. Garcia-Luna-Aceves and E. Madruga, "The core-assisted mesh protocol," *IEEE Journal on Selected Areas in Communications*, vol. 17, pp. 1380–1394, August 1999.

[37] T. Cover and J. Thomas, *Elements of Information Theory*. New York: John Wiley Sons, Inc., 1991.

[38] D. J. Newman and L. Shepp, "The double dixie cup problem," *Amer. Math. Monthly*, vol. 67, pp. 58–61, January 1960.

[39] W. Feller, *An introduction to probability theory and its applications*. John Wiley and Sons, Inc., 1967.

[40] B. M. Hochwald, T. L. Marzetta, and V. Tarokh, "Multiple-antenna channel hardening and its implications for rate feedback and scheduling," *IEEE Transactions on Information Theory*, vol. 50, no. 9, pp. 1893–1909, 2004.

[41] P. Parada and R. Blahut, "Secrecy capacity of SIMO and slow fading channels," in *Proceedings of ISIT 2005*, pp. 2152–2155, September 2005.

[42] J. Barros and M. R. D. Rodrigues, "Secrecy capacity of wireless channels," in *Proceedings of ISIT 2006*, July 2006.

[43] R. G. Gallager, *Information Theory and Reliable Communication*. John Wiley and Sons, Inc., 1968.

[44] A. J. Viterbi, "Error bounds for the white gaussian and other very noisy memoryless channels with generalized decision regions," *IEEE Transactions on Information Theory*, vol. 15, pp. 279–287, March 1969.

[45] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, "Lower bounds to error probability for coding on discrete memoryless channels I," *Information and Control*, vol. 10, pp. 65–103, January 1967.

[46] C. E. Shannon, "Probability of error for optimal codes in a gaussian channel," *Bell Systems Technical Journal*, vol. 38, pp. 611–656, 1959.

[47] A. Valembois and M. P. C. Fossorier, "Sphere-packing bounds revisited for moderate block lengths," *IEEE Transactions on Information Theory*, vol. 50, pp. 2998–3014, December 2004.

[48] S. Dolinar, D. Divsalar, and F. Pollara, "Code performance as a function of block size," in *TMO Progress Report*, pp. 42–133, May 1998.

[49] D. Rajan, A. Sabharwal, and B. Aazhang, "Power efficient broadcast scheduling with delay deadlines," in *Proceedings of Broadnets*, 2004.

[50] P. K. Gopala and H. El Gamal, "On the throughput-delay tradeoff in cellular multicast," in *Proceedings of the Symposium on Information Theory in WirelessCom 2005*, June 2005.

[51] X. Liu, E. K. P. Chong, and N. B. Shroff, "Opportunistic transmission scheduling with resource-sharing constraints in wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 10, 2001.

[52] R. G. Gallager, *Discrete Stochastic Process*. Kluwer Academic Publishers, 1996.

[53] M. Sharif and B. Hassibi, "On the capacity of MIMO broadcast channels with partial channel state information," *IEEE Transactions on Information Theory*, vol. 51, no. 2, 2005.

[54] B. C. Arnold, N. Balakrishnan, and H. N. Nagaraja, *A first course in order statistics*. New York: John Wiley Sons, Inc., 1992.

[55] R. Durrett, *Probability: Theory and Examples*. California: Duxbury Press, Inc., 1996.

[56] J. Pickands, "Moment convergence of sample extremes," *The Annals of Mathematical Statistics*, vol. 39, no. 3, pp. 881–889, 1968.

[57] Y. Liang and H. V. Poor, "Generalized multiple access channels with confidential messages," *Submitted to IEEE Transactions on Information Theory*, 2006.