

Variable-Length Coding Incremental Redundancy Hybrid Automatic Repeat reQuest (HARQ) over Fading Channels

Saeed R. Khosravirad



Department of Electrical & Computer Engineering
McGill University
Montreal, Canada

April 2015

A dissertation submitted to McGill University in partial fulfillment of the requirements for
the degree of Ph.D.

© 2015 Saeed R. Khosravirad

To my Mom & Dad

Abstract

In order to overcome the adverse effects of fading in wireless telecommunication channels, using a combination of error correcting codes and repeat requests has proven to be a practical solution. Such a combination is indeed offered in Hybrid Automatic Repeat reQuest (HARQ) where retransmissions of encoded bits from the same message packet increase chances of successful decoding. HARQ thus becomes a method for reliable delivery of bursty traffic of data in wireless channels. By exploiting a feedback channel, HARQ targets smaller error rate and higher throughput by retransmitting the failed packets. Conventional HARQ transmission assumes a single-bit feedback message that can acknowledge the success or failure of a transmission attempt to the transmitting node. It is also conventionally accepted to assume fixed transmission parameters throughout the HARQ retransmission process. As a natural extension, the idea of being able to vary transmission parameters from one attempt to another has been of an extensive interest in the recent studies on HARQ. Moreover, the feedback message shows a great performance boost to the conventional HARQ when accompanied by side information about the channel state.

This dissertation aims to answer the question: how beneficial a feedback of more than just one bit message can be to the performance of HARQ, if the transmission parameters such as rate, can be varied? To answer this question, we define the optimal rate control problem to find the maximum achievable throughput of HARQ protocol. We also consider two scenarios: in the first one the transmitter does not have any knowledge about the state of the channel and gets a reliable feedback from the receiver in the conventional way; in the second case we assume that the instantaneous channel state information is not available in transmitter but an independent outdated version of it that can be accessible via the feedback channel can be used by the transmitter to adapt the transmission rate. Assuming a fixed-power transmission, we find the optimal rate policies, which yield the maximum achievable throughput and outage-probability-constrained optimal throughput. We present various simplification approaches to the non-convex optimization problems that can significantly reduce the complexity. Our approach is based on the Markov decision process theoretical framework and the optimization that uses the dynamic programming. Theoretically proven performance bounds for truncated and infinite transmission HARQ are presented for both single-hop and relay channels. We show that a few bits of extra feedback message with truncated HARQ can approach closely the performance limit.

Résumé

Le protocole Hybride ARQ est principalement utilisé pour garantir une livraison fiable de trafic de données dans les canaux sans fil à évanouissements. HARQ permet de réduire la probabilité d'erreurs grâce à l'utilisation d'un canal de retour. En effet, HARQ retransmet correctement les paquets non reçus. Les protocoles HARQ classiques utilisent un message de retour à un seul bit pour déclarer l'échec ou le succès de la transmission. On suppose également que les paramètres de transmission restent fixes tout au long du processus de la retransmission pour le HARQ traditionnel. l'idée de varier les paramètres de HARQ d'une transmission à l'autre a fait récemment l'objet de plusieurs de travaux de recherche. De même, les performances du système HARQ augmentent en améliorant les messages de feedback. Notamment, lorsque ces derniers contiennent une information partielle à propos du canal supposée non connue par l'émetteur. les messages de feedback donnent une idée sur le canal de transmission au récepteur.

Dans cette thèse, on montre l'avantage d'utiliser plusieurs bits dans le canal de retour. On suppose que les paramètres de transmission, notamment le débit, peuvent varier entre les retransmissions dans le protocole HARQ.

Notre objectif est d'optimiser le débit afin de maximiser le débit (throughput) du protocole HARQ. Notre étude couvre les systèmes "point à point" et les systèmes à relais dans un canal à évanouissements par blocs. Nous considérons deux scénarios: le premier cas où l'émetteur n'a pas d'informations sur l'état du canal, mais il reçoit un seul bit de retour. Dans le deuxième cas, l'émetteur n'a aussi pas d'informations instantanées sur l'état du canal, cependant, une information sur l'état précédant est disponible. Pour les deux scénarios, et en supposant que la puissance reste constante tout le long du processus, on détermine le débit qui maximise le throughput avec contraintes sur la probabilité de coupure. Nous prouvons que les solutions obtenues des cas simplifiés sont aussi proches que celles du problème original avec moins de complexité. Pour résoudre le problème d'optimisation, on utilise la théorie sur les processus markoviens, précisément, la programmation dynamique. On détermine aussi les limites des performances du protocole HARQ avec un nombre fini ou infini de retransmissions. Nous montrons qu'uniquement quelques bits de messages de rétroaction supplémentaire suffisent pour atteindre les performances maximales.

Acknowledgments

My thesis advisors, Fabrice Labeau and Leszek Szczecinski, have my sincerest gratitude; both this thesis and my personal development have benefited greatly from their insightful guidance and endless support. Their immense knowledge and inspiration helped me to overcome many difficulties during my research. I feel fortunate to have had the enlightening experience to work with these enthusiastic and talented people during my time at McGill. I want to also thank my thesis committee members, Ioannis Psaromiligkos and Long Le. Their suggestions have helped to make this a better thesis.

I am very thankful for many friends and colleagues in the Telecommunications and Signal Processing group and the INRS-EMT who were sources of inspiration and friendship to me, including Sina Hamidi, Mojtaba Vaezi, Mohammed Jabi, Aata Elhamss, Fabian Sacuto and Vahid Asghari. It has been a privilege to know such a great group of people. I especially thank Sina and Mohammed for always being willingly helpful to me.

Outside of school, I would like to thank my friends Nima Lashkari, Hossen Karimiani, Shahrzad Haddadan, Omid Sanei, Ali Noruzi, Sepideh Zarifinia, Haleh Shahbazi, Mahdi Milanifard and Skye Miner for the amazing time I had in Montreal during the past five years and all the great memories.

Finally, on a personal note, I like to thank my parents, my brothers, and my sisters. I am grateful for their support and encouragements. I only hope that I have made their support worthwhile.

Preface and Contributions

The research works presented in this dissertation are the result of my original work under the supervision of Prof. Fabrice Labeau at the department of Electrical and Computer Engineering, McGill University and Prof. Leszek Szczecinski at INRS-EMT. Some of the contributions in this dissertation are published in the following scholarly articles.

In all of the publications listed below, I was involved in the discussion of the approach, conducting the analysis, implementing the simulations and writing the manuscripts. My supervisors have initiated the discussion on the topics and provided me with the technical advice on the conducted research work.

Journal Papers

- S. R. Khosravirad, L. Szczecinski, F. Labeau, “Opportunistic Relaying without CSI: Optimizing Variable-Rate HARQ,” *submitted to IEEE Trans. Veh Techn.*, Dec. 2014 [arXiv:submit/1138351]
- S. R. Khosravirad, L. Szczecinski, F. Labeau, “Rate Adaptation for Cooperative HARQ,” *IEEE Trans. on Commun.*, vol. 62, no. 5, pp. 1469–1479, May 2014
- L. Szczecinski, S. R. Khosravirad, P. Duhamel, and M. Rahman, “Rate Allocation and Adaptation for Incremental Redundancy Truncated HARQ,” *IEEE Trans. on Commun.*, vol. 61, no. 6, pp. 2580–2590, June 2013

Conference Papers

- S. R. Khosravirad, L. Szczecinski, and F. Labeau, “Rate-Adaptive HARQ in Relay-based Cooperative Transmission,” in *Proc. ICC 2013*, Budapest, Hungary, 9-13 Jun. 2013
- S. R. Khosravirad, L. Szczecinski, F. Labeau, “Rate Allocation for HARQ in Relay-based Cooperative Transmission,” in *Proc. WCNC 2014*, Istanbul, Turkey, 6-10 Apr. 2014

Contents

List of Figures	viii
List of Acronyms	x
Notations	xii
1 Introduction	1
1.1 Motivation and Problem Statement	2
1.1.1 Block-Fading Channel Model	4
1.1.2 Transmission over Fading Channel	4
1.1.3 How HARQ Works	5
1.1.4 Motivation	8
1.1.5 Objectives	11
1.2 Outline and Contributions	12
1.2.1 Contributions	12
1.2.2 Document Organization	14
2 HARQ Resource Control Problem	16
2.1 Channel Model with no Transmitter CSI	17
2.1.1 Single-hop Channel Model	17
2.1.2 Normalized Accumulated Mutual Information	20
2.1.3 Multi-hop Channel Model	22
2.2 Performance Criteria	26
2.2.1 Outage Probability	26
2.2.2 HARQ Throughput	27

2.3	Markov Decision Process Framework	28
2.4	Practical Concerns	30
3	Answers In the Literature – State of the Art	33
3.1	Exploiting Feedback	34
3.2	Optimal Coding Rate	36
3.3	Joint Rate and Power Optimization	38
3.4	Cooperative HARQ	41
3.5	Optimization with Dynamic Programming	42
4	Single-Hop HARQ Transmission Policies	47
4.1	Throughput Optimization	47
4.2	Optimal Rate Adaptation	51
4.3	Optimal Rate Allocation	53
4.3.1	Simplified Optimization Method	57
4.4	Performance Limits	58
4.5	Numerical Results	60
4.6	Summary and Complementary Notes	64
5	Orthogonal Relaying Policies and Multi-Hop Channel	67
5.1	Adaptive-Rate Orthogonal Relaying	68
5.1.1	Optimum Policy	68
5.1.2	Initializing DP Recursion	69
5.1.3	Throughput Calculation	72
5.1.4	Quantization of the Feedback	73
5.2	Rate Allocation in Orthogonal Relaying	74
5.2.1	Closed-Form Representations	76
5.2.2	Approximate Recursive Optimization	78
5.2.3	Simplified One-Dimensional State Optimization	79
5.2.4	Exact Throughput Calculation	80
5.2.5	Remarks on the Complexity	81
5.3	Performance Limits	86
5.4	Numerical Results	87
5.5	Multi-Hop HARQ Transmission	94

5.6	Summary and Complementary Notes	96
6	Conclusion and Future Work	98
6.1	Concluding Remarks	98
6.2	Future Directions	101
6.2.1	Transmission Power	102
6.2.2	Channel Model	102
6.2.3	Coding/Modulation	102
A	Proofs for Propositions and Theorems	103
B	Throughput of Cooperative Variable-Rate HARQ Transmission	106
C	DP Recursive Optimization for Rate Allocation in Relay Channel	109
D	One Dimensional State Optimization	112
	Bibliography	116

List of Figures

1.1	Illustration of a point-to-point communication channel.	2
1.2	Channel gain experience in different channel fading types	3
1.3	Comparison between ARQ and CC-HARQ decoders	5
1.4	variable vs fixed rate HARQ transmission	9
2.1	Block-fading channel model with feedback.	17
2.2	Codeword with variable-length sub-codewords	18
2.3	Decoding codeword with variable-length sub-codewords	20
2.4	Topology of the relay channel.	23
2.5	Flowchart of an MDP model	29
2.6	Example of the structure of frames for variable-rate HARQ transmission. . .	31
4.1	Optimal rate policies for HARQ over single-hop channel	53
4.2	Optimal outage-constrained throughput for $\epsilon = 10^{-2}$	60
4.3	Optimal outage-constrained throughput for $\epsilon = 10^{-4}$	61
4.4	Optimal throughput for HARQ over single-hop fading channel	62
4.5	Outage probability of the optimal rate policies for different scenarios of HARQ.	63
4.6	Throughput of the fixed-rate scenario with respect to redundancy value ρ . .	64
4.7	Throughput of an optimal rate policy for a different SNR	65
4.8	Outage probability of an optimal rate policy for a different SNR	66
5.1	Optimal rate adaptation policy for HARQ over the relay channel model . . .	71
5.2	Topology of the relay channel for experimental analysis.	82
5.3	Throughput for different optimization experiments.	83
5.4	Histogram plot for the normalized frequency of a solution π_r	84

5.5	Optimal rate allocation policy for HARQ over the relay channel model . .	85
5.6	Optimal outage-constrained throughput for HARQ over relay channel . . .	88
5.7	Optimal throughput for HARQ over relay channel	90
5.8	Outage Probability for maximum-throughput-achieving policies	91
5.9	Discretized feedback analysis results	92
5.10	Effect of the relay positioning parameter d , on the optimal throughput . .	93
5.11	Topology of the M -relay network.	95

List of Acronyms

3GPP	3rd Generation Partnership Project
ACK	Acknowledgement
ACMI	ACcumulated Mutual Information
AMC	Adaptive Modulation and Coding
ARQ	Automatic Repeat reQuest
Avg.	average
AWGN	Additive White Gaussian Noise
bpcu	bits per channel use
CC	Chase Combining
cdf	cumulative density function
CDI	Channel Distribution Information
CQI	Channel Quality Indicator
CSI	Channel State Information
DP	Dynamic Programming
EDGE	Enhanced Data GSM Environment

FD	full-duplex
FEC	Forward Error Correction
GSM	Global System for Mobile communications
HARQ	Hybrid Automatic Repeat reQuest
HD	half-duplex
HSPA	High Speed Packet Access
i.i.d.	independent and identically distributed
IEEE	Institute of Electrical and Electronics Engineers
IR	Incremental Redundancy
LATR	long-term average transmission rate
LTE	Long Term Evolution
MAC	Media Access Control
MDP	Markov Decision Process
Med.	median
Min.	minimum
MRC	maximal-ratio combining
NACK	Negative Acknowledgement
NACMI	Normalized ACcumulated Mutual Information
OFDM	Orthogonal Frequency-Division Multiplexing
pdf	probability density function

POMDP	Partially Observable Markov Decision Process
RLC	Radio Link Control
RTR	round transmission rate
SNR	Signal to Noise Ratio
TD	Time Division
TDMA	Time Division Multiple Access
UWB	Ultra Wideband
WiMax	Worldwide interoperability for Microwave access
WLAN	Wireless Local Area Network

Notations

A	action
B	bandwidth
C_0	capacity with outage
I	normalized accumulated mutual information of k transmission attempts
J	minimum average cost
K	maximum number of allowed transmission attempts
L	number of quantization/discretization points
N_s	number of symbols
P	probability function
P_{out}	outage probability
$Q(.)$	Q-function
R	transmission rate
S	state
T_c	channel coherence time
$\tilde{\eta}$	throughput optimized using one-dimensional Gaussian approximation outage probabilities
η	throughput

γ	signal to noise ratio
$\hat{\eta}$	maximum throughput
ι	normalized mutual information of a single transmission attempt
λ	Lagrangian multiplier
$\mathbb{F}$	transmission frame
$\mathbb{I}_k$	accumulated mutual information of k transmission attempts
$\mathbb{P}$	packet of symbols
$\mathbf{C}$	channel codebook
$\mathbf{D}(\cdot)$	decoding function
$\mathbf{X}$	transmit codeword
$\mathbf{Y}$	received codeword
$\mathbf{x}$	transmit sub-codeword
$\mathbf{y}$	received sub-codeword
$\mathbf{z}$	set of dummy symbols
$\mathbb{E}\{\cdot\}$	expectation
$\mathbb{I}(\cdot)$	indicator function
$\mathbb{R}$	set of all real numbers
$\mathbb{R}_0^+$	set of all non-negative real numbers
$\mathfrak{g}(\cdot)$	Lagrangian function
ν	path-loss exponent
ω	disturbance
$\overline{C}$	ergodic capacity

π	policy set
ρ	number of symbols transmitted per bit (redundancy)
σ	variance
F	cumulative distribution function
$\tilde{\eta}$	throughput optimized using two-dimensional Gaussian approximation outage probabilities
ξ	complex Gaussian noise of the channel
h	channel complex coefficient
k	counter for transmission attempt
p	probability distribution function
q	transmit power
x	transmit symbols
y	received symbols

Chapter 1

Introduction

WIRELESS networks provide the users with *anywhere/anytime* services. This feature makes the wireless communication technologies very popular as the number of wireless devices are now larger than the world population. Wireless communication has always been facing an ever-increasing demand on data transmission services; for instance, one can look at the various web-based services that are extremely popular in telecommunication industry today. Bursty sporadic communication from a large population of users in wireless mobile access to the Internet or packet-oriented data transmission that requires instantaneous large data rates and very small error probabilities for a short time [1].

Wireless networks can generally cost less to be built compared to the wired networks. But, what makes it challenging to create wireless networks is the fading characteristics of the wireless channel which increases the chances of packet loss. This way, the random fading process crucially debilitates wireless communication. Reducing the chances of losing a message packet in communication system providing high data rate along with the system performance is the motivation for researchers to introduce advanced techniques like *Hybrid Automatic Repeat reQuest* (HARQ) or *Adaptive Modulation and Coding* (AMC).

Because of the multi-path characteristic of a wireless channel, a signal may appear in multiple constructive and/or destructive versions at the receiver. This may result in various fading levels on the signal which may make the decoding impossible in deep fading situation [2]. As a result, error correction coding is not guaranteed to be successful in message delivery and transmission of a codewords have to be repeated occasionally. This chapter discusses how a fading channel should be dealt with, in order to increase the reliability of

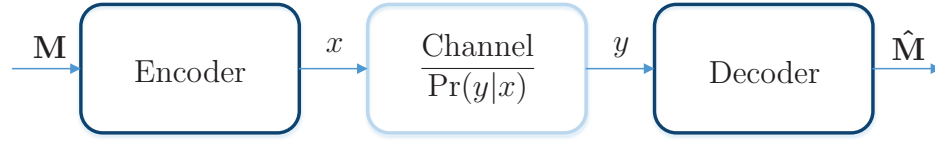


Figure 1.1 Illustration of a point-to-point communication channel.

communications among the parties. This will bring to light a number of questions with regard to the reliable and fast transmission over fading channel and we will talk about our motivations in this research work. At the end of this chapter, a summarized overview of the contributions made in this research work is presented, which will be explained in detail in the following chapters.

1.1 Motivation and Problem Statement

A communication channel without fading, the characteristics of noise (and interference) describes the performance (Figure 1.1). The channel is represented by $\Pr(y|x)$ as the conditional probability density function for an output y given an input x . In a wireless channel, it is not only noise and interference but also changes in channel state over time that may harm a reliable communication. Path-loss and shadowing in a wireless communication link vary with time and, as a result, the wireless signal faces time-variant fading. The *coherence time* of a fading channel is the minimum time required to have a new channel coefficient value (changed in magnitude or phase) which is uncorrelated from the previous value¹. Based on the relation between the number of symbols in a codeword N_s (in channel uses), and the channel coherence time T_c (also in channel uses), three scenarios can happen in a fading channel (Figure 1.2).

- *Fast-fading* - where N_s is greater than the channel coherent time T_c . This is the case for fast-moving users and/or encoders with long codewords. An example for fast-fading experience is in *Ultra Wideband* (UWB) channels [3].
- *Slow-fading* - where the codeword length is considered to be not greater than the channel coherent time.

¹The same definition can be considered in frequency domain.

- *Quasi-static* - where the coherence time is long enough so that the channel remains fixed for several codeword length. This is a useful model for slow-moving users and/or relatively short codewords in networks with general stationary situations.

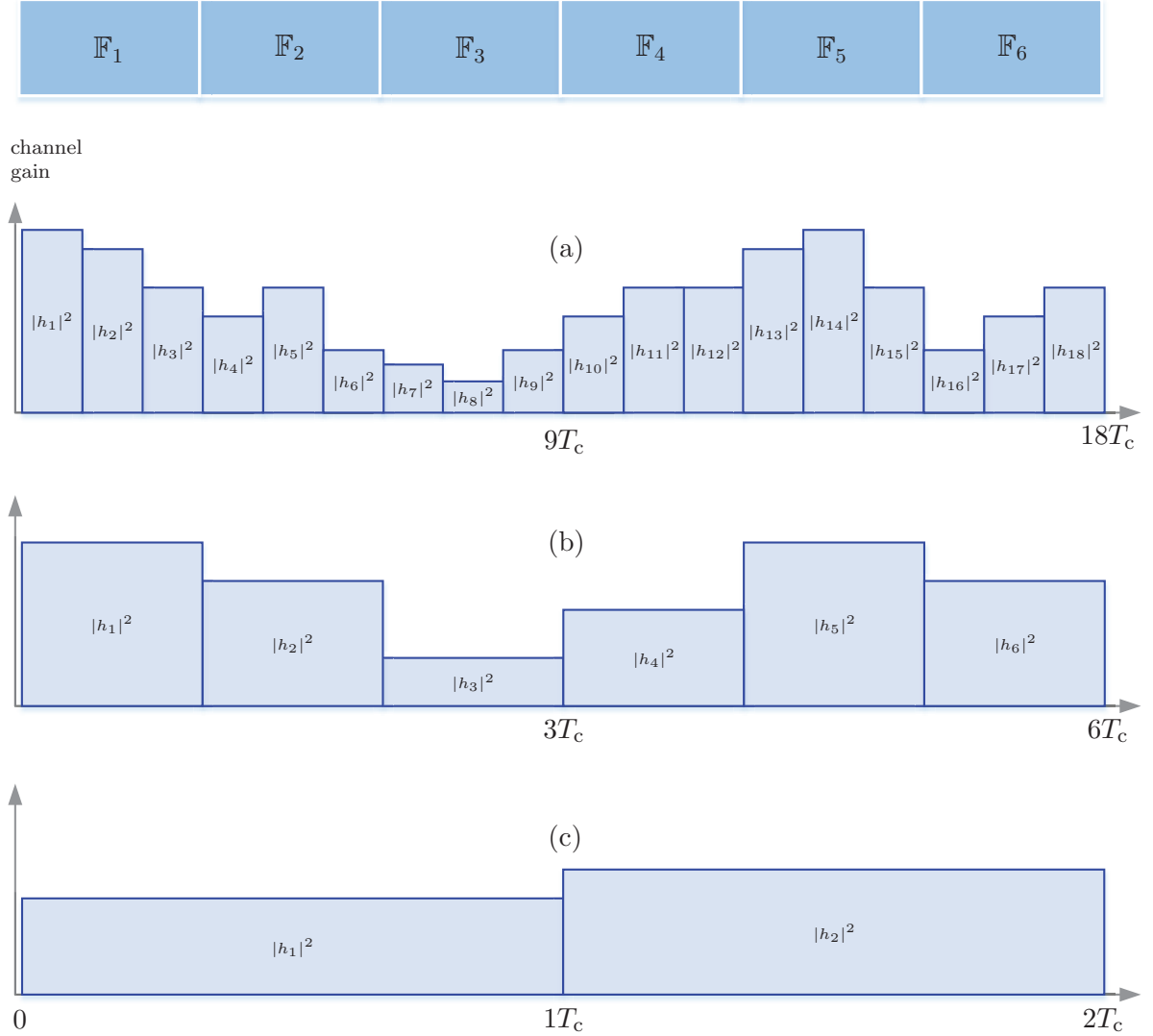


Figure 1.2 Channel gain experience for six consecutive frames of $\mathbb{F}_1, \dots, \mathbb{F}_6$ (i.e., $6 \times N_s$ symbols), with respect to time (in channel uses). The three cases of (a) fast-fading, (b) slow-fading and (c) quasi-static are being compared. The channel fading factor is denoted by a complex value h_i and as a result the channel gain is equal to $|h_k|^2$.

1.1.1 Block-Fading Channel Model

In *block-fading* channel model, an entire encoding frame experiences a constant channel condition. This can be interpreted as both slow-fading and quasi-static models. In case of having a quasi-static situation, (for instance, using *Orthogonal Frequency-Division Multiplexing* (OFDM) [4]) a coherence time slot of the channel can be converted into a set of parallel block-fading channels.

The characteristics of block-fading channel is known to be common in many practical wireless communication setups. One truly interesting example is OFDM which is the core technology in many wireless standards including *Institute of Electrical and Electronics Engineers* (IEEE) 802.11 *Wireless Local Area Network* (WLAN) [5] and IEEE 802.16 *Worldwide interoperability for Microwave access* (WiMax) [6]. Moreover, the *Global System for Mobile communications* (GSM) and the *Enhanced Data GSM Environment* (EDGE), also rely on a block-fading channel model [7]. For further discussions about fading channel we refer the readers to [8].

1.1.2 Transmission over Fading Channel

The two main schemes for channel coding are the *Forward Error Correction* (FEC) and the *Automatic Repeat reQuest* (ARQ). In FEC, error-correcting channel codes are employed to help the receiver node with correcting errors. In ARQ on the other hand, only error detection codes are used. Here, the receiver tries to detect errors in the received packet and, using a reliable feedback channel, notifies the transmitter about the result of decoding procedure. Depending on success or failure of the decoding process, an *Acknowledgement* (ACK) or *Negative Acknowledgement* (NACK) message is sent back to the transmitter respectively. In case of failure, retransmission of the same message is the solution to channel error in ARQ.

Throughput is a well accepted performance criterion for wireless channel analysis. It is defined as the average number of correctly received bits per average number of channel uses necessary for the data delivery and is measured by *bits per channel use* (bpcu). In HARQ, FEC is used along with ARQ to decrease the number of retransmissions and increase channel throughput [9]. In other words, HARQ looks for efficient information delivery in data transmission by combining re-transmitted versions of the same data packet in case of unsuccessful decoding in the first attempt to decode the packet.

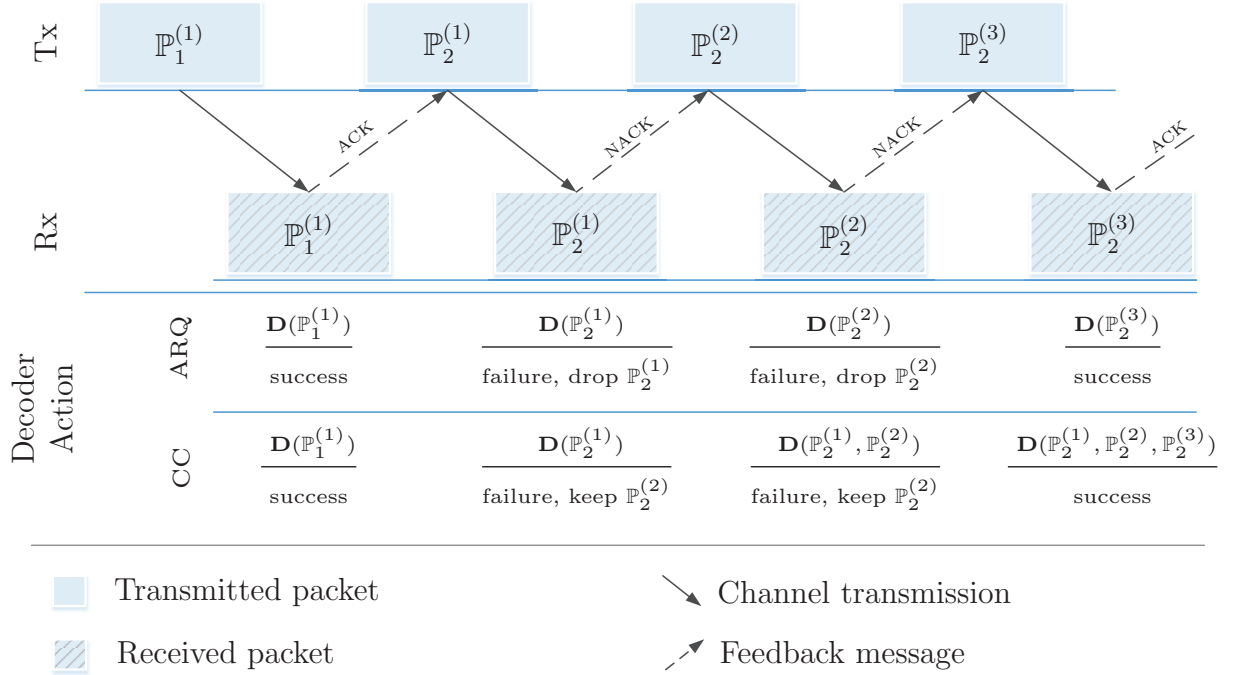


Figure 1.3 A comparison of decoder action between ARQ and CC-HARQ transmission methods. Here, we denote by $\mathbb{P}_n^{(k)}$, the k th transmission attempt for the n th packet. Packet $\mathbb{P}_1$ is successfully decoded after one transmission attempt. Packet $\mathbb{P}_2$ is retransmitted two extra times, due to the decoding failure on its first attempt $\mathbb{P}_2^{(1)}$. The decoder function $\mathbf{D}(\cdot)$, takes only the last received packet for decoding process in ARQ, while in CC-HARQ, all of the received packets of the same packet will be used in the decoding process.

1.1.3 How HARQ Works

HARQ can be divided into two main groups: *Chase Combining* (CC) and *Incremental Redundancy* (IR). In CC-HARQ, every retransmission contains the same packet (Figure 1.3). The receiver uses *maximal-ratio combining* (MRC) to combine the received packets during the ARQ retransmissions. Because all transmissions are identical, CC can be seen as repetition coding. In IR-HARQ, multiple sets of codeword symbols are generated, each representing the same set of information bits. The retransmission typically uses a different set of codeword symbols than the previous transmission, with different redundancy versions generated by puncturing the encoder output. Thus, at every retransmission the receiver

gains extra knowledge [10].

The key idea in HARQ is to keep the received information from the first transmission attempt in case of decoding failure, instead of dropping it. The received information is buffered, not in the form of *hard channel bits* (i.e., observation sequences), but rather in form of *soft channel bits* (i.e., data only known in terms of probabilities). The soft channel bits in the buffer will then be used for soft MRC in CC or, soft combined decoding in IR [11].

In CC-HARQ, because MRC is used at the decoder, transmission rate is fixed throughout all the retransmission attempts. In other words, with R_k denoting the transmission rate at the k th attempt, we have $R_i = R_j, \forall i, j$. On the other hand, variable transmission rates are feasible in IR-HARQ, which gives this setup new degrees of freedom for performance optimization (Figure 1.4). The parameter of the maximum number of transmission attempts for a message governs the number of degrees of freedom in this sense. We denote this parameter by K throughout this dissertation. An HARQ process is called truncated when K is finite.

To have a better idea on how the two different HARQ types work, we can refer to the outage probability of the two. For CC-HARQ transmission, the outage probability has a form as follows.

$$P_{\text{out}}^{\text{CC-HARQ}} = \Pr \left\{ \log(1 + \sum_{k=1}^K \gamma_k) < R \right\}, \quad (1.1)$$

where we assume Gaussian noise and codeword distribution and use γ_k to denote the *Signal to Noise Ratio* (SNR) of the transmission attempt k and R to denote the coding rate. (1.1) shows how CC-HARQ with K transmission attempts and MRC at the decoder can be seen as a single transmission with an SNR equal to $\sum_{k=1}^K \gamma_k$. Using the same notation, IR-HARQ has an outage probability of the following form.

$$P_{\text{out}}^{\text{IR-HARQ}} = \Pr \left\{ \sum_{k=1}^K \log(1 + \gamma_k) < R \right\}, \quad (1.2)$$

where a summation overtime of the mutual information, denoted by $\log(1 + \gamma_k)$, must be less than the transmission rate to have an outage.

HARQ in its simplest version, can be discussed as if the error correcting code is

chosen (possibly differently) based on the operating SNR and the encoded data packets are being transmitted as in conventional ARQ. The error correcting code merely increases the probability of successful transmission, and thus the overall throughput. It is well-agreed that the throughput criterion (along with other criteria such as transmission delay, probability of losing the packet, etc.) is one of the key criteria to exhibit the usefulness of a transmission protocol. For the sake of a better throughput, in HARQ it is important to have as few channel uses as possible for a desired level of reliability to make sure that the bandwidth is employed in an efficient way.

It is not surprising that HARQ is now widely popular in telecommunication industry. In applications such as mobile and satellite packet data transmission, the IR-HARQ achieves high throughput efficiency by adapting the error correcting code redundancy to varying channel conditions [10]. Both CC-HARQ and IR-HARQ are being widely used in contemporary wireless systems such as *High Speed Packet Access* (HSPA), WiMax and *3rd Generation Partnership Project* (3GPP) *Long Term Evolution* (LTE) [12]. Also, HARQ can be used in a single-user channel or multi-user channel for packet transmission. Currently, HARQ

HARQ is also being used for implementing AMC technique in wireless mobile communication. Suppose two users, where one of them is closer to the cell base station of a wireless network than the other and so having a more favorable position. The better user can be assigned a higher order modulation and code rate than the user with poorer position to increase its average throughput. Now, for a single user, when the channel experience changes over time, an adaptive modulation and coding rate can help the user to have a more efficient communication. Estimation of the channel state needs to be accurate enough to make the adaptation useful. Using HARQ, a feedback message from the receiver to the transmitter conveying the estimated channel condition can help the transmitter to adapt its transmission parameters. AMC is being used in HSPA to increase throughput combined with IR-HARQ [13]. If the code or modulation rate of the subsequent transmissions of a given packet is adapted with the channel condition and changes between re-transmissions then we have adaptive HARQ. Otherwise if the transmissions use the same modulation order and code rate scheme, then we have non-adaptive HARQ.

Dealing with HARQ, as the combination of FEC and ARQ, two main parameters show up for design: coding rate and number of retransmissions.

- In FEC, the greater number of redundant bits in a codeword for a well-designed

system, results in better error correcting capability. On the other hand throughput may suffer by decreasing coding rate.

- In ARQ it is obvious that allowing an infinite number of retransmissions gives a zero outage probability, which is the probability of losing the data. However, by truncating the ARQ process to finite K , there is always a possibility of outage which is non-increasing by K increasing.

The tradeoff between transmission rate and the number of retransmissions in an ARQ protocol is the critical point in the design of an HARQ system.

1.1.4 Motivation

The capacity of a fast-fading channel, when only the decoder knows the *Channel State Information* (CSI), is known in general. In a fast-fading channel, an asymptotically long transmitted packet is assumed to experience the whole characteristics of the channel, thus the channel coding rate is limited by the ergodic capacity (the so-called Shannon capacity) [14].

$$\overline{C} = B \int_0^\infty \log_2(1 + \gamma) p(\gamma) d\gamma \quad (1.3)$$

with B being the bandwidth of a fading channel with Gaussian noise (and consequently using Gaussian input) while the distribution of the fading is characterized with $p(\gamma)$.

The capacity $\overline{C}$ in (1.3), is the probabilistic average of the capacity of an *Additive White Gaussian Noise* (AWGN) channel with SNR of γ . One can easily investigate, based on the well-known Jensen's inequality, that the capacity of a fast-fading channel with decoder CSI is less than or at most equal to the capacity of an AWGN channel with SNR of $\bar{\gamma} = \mathbb{E}\{\gamma\}$, where $\mathbb{E}$ denotes the expectation of a random variable.

In block-fading channel model with an idealistic capacity achieving code in each block, the transmitter can adapt the transmission power to the channel conditions if the CSI is available at the transmitter. The result of this adaptive transmission is to reach the Ergodic capacity in the long term [9]. In a practical sense, when CSI is available at the transmitter side, (e.g., the *Channel Quality Indicator* (CQI) signalling in LTE that represents the receiver's desired modulation and coding rate), AMC can be exploited in order to adapt the transmission parameters to the instantaneous channel conditions [15] in order to guarantee a certain level of successful message delivery.

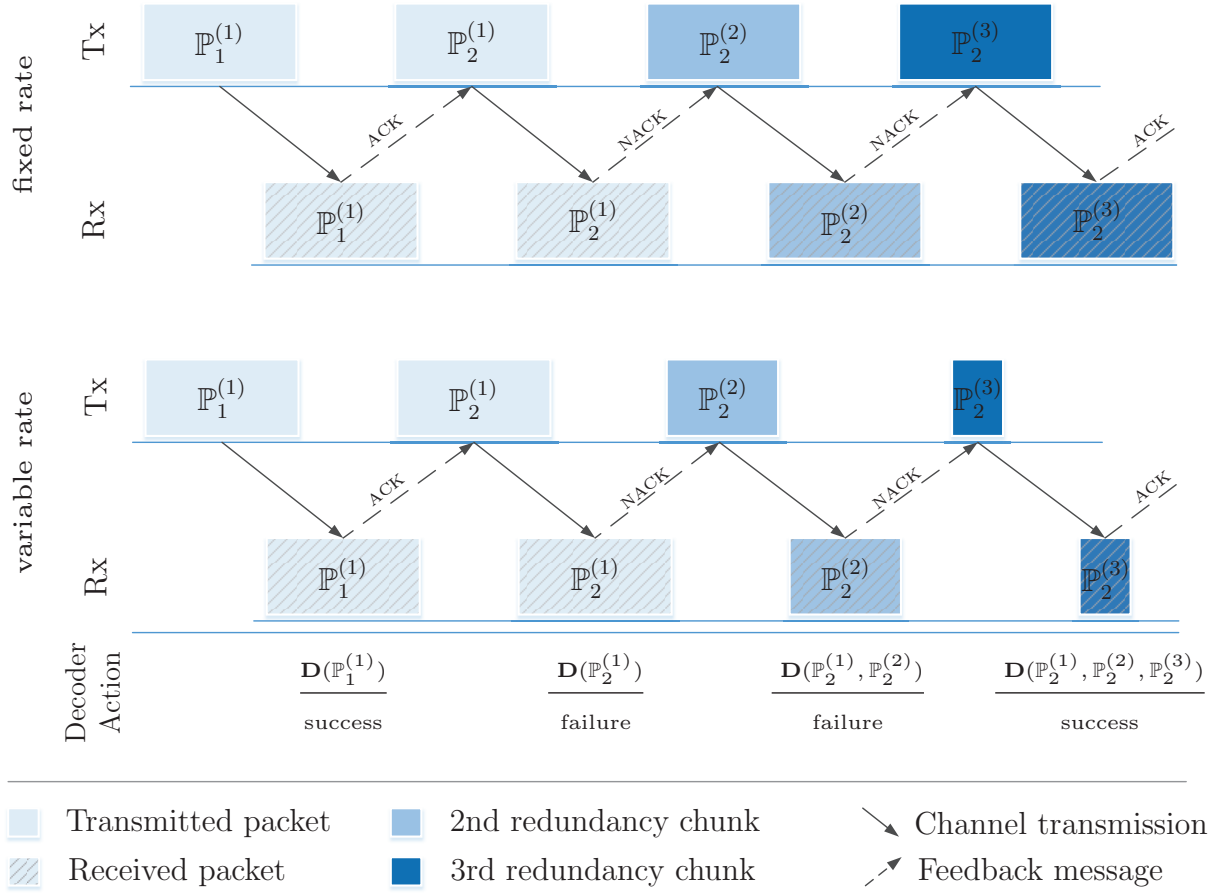


Figure 1.4 Incremental redundancy HARQ transmission with variable/fixed rate. The Retransmission attempts are consisted of new redundancy chunks of symbols, unlike CC-HARQ, where the same packet is retransmitted in case of decoding failure.

In a block-fading channel with only decoder CSI, transmissions with even a very low coding rate are not guaranteed to be delivered correctly at the receiver. This is a result of the time-varying nature of a fading channel and the probable times of channel being in deep fade. HARQ transmission can be seen as a means of error control in this situation. The feedback channel of the HARQ protocol can be exploited to report the *outdated*-CSI (i.e., the CSI experienced in the previous attempts by the decoder) to the transmitter and inform the encoder of how close the decoder is to a successful decoding. Capacity can not

be defined for such a channel and instead, capacity with outage is used in analysis [2].

Capacity with outage [2] is represented with a dual of a transmission rate limit and the probability of outage when functioning above that limit. The transmitter fixes a minimum received SNR of $\gamma_{\min}$ and encodes its data with the rate of $\log_2(1 + \gamma_{\min})$. Decoding of asymptotically large codewords will be successful if the instantaneous SNR is greater than or equal to $\gamma_{\min}$ [16]. In case that the experienced SNR is less than $\gamma_{\min}$, the channel will be in fade and outage happens. Therefore, the probability of outage is as follows:

$$P_{\text{out}} = \Pr\{\gamma < \gamma_{\min}\}, \quad (1.4)$$

and the capacity with outage is calculated as

$$C_o = (1 - P_{\text{out}}) \log_2(1 + \gamma_{\min}) \quad (1.5)$$

in bits per channel use.

The relation between P_{out} and C_o as capacity versus outage gives the capacity with outage defined for block fading channels. In general, the performance of a block-fading channel is discussed in the average sense of achievable rate (or throughput) and transmission over such a channel is never guaranteed to be successful. However, we are interested to know if a block-fading channel can achieve the rates in ergodic capacity in the long run, where the outage probability is arbitrarily small.

It is shown in [1] that with an arbitrarily large number of retransmission attempts in HARQ, the throughput achieves the ergodic capacity limit for a single-hop transmission channel. This interests us in finding out the performance limits of a truncated HARQ, since it considers technical limits of a system where time is of severe importance (and the service tolerance to delay is very high) in delivery of data. We want to take advantage of the degrees of freedom in truncated HARQ, such as the adjustable transmission parameters, e.g. transmit power [17] or transmission rate [18]. Since a fixed-rate fixed-power transmission is just a special case of the general problem, it will be interesting to know how much throughput gain can we achieve, in terms of throughput, with variable rate and/or variable power transmission in truncated HARQ.

For the transmitter to wait for twice the propagation delay time (*round-trip delay*) to receive the feedback message from the receiver, it might lose the time resource for

transmission of other message packets in ARQ protocol. The effect of a large propagation delay on decreasing the throughput, which can happen in satellite communications [19], can be ignored by assuming the use of many parallel ARQ processes to fully exploit the available time resources. This idea is being used in the *Media Access Control* (MAC) layer HARQ transmission in the LTE standard.

Moreover, HARQ can be introduced as a practical approach to communication networks comprising relays and cooperative communication [20]. This makes us wonder how beneficial variable transmission parameters can be in a truncated cooperative HARQ. As in the case of the single-hop channel, we are interested in studying the performance limits of channels with relays.

With a relay node lying in between transmitter and receiver nodes, and assuming independent channels between nodes, there is always a higher possibility of decoding success at the relay node compared to the receiver. As a result, the HARQ process, started from the node that has the message, can be continued with the relay node in case it decodes the message successfully. Therefore, if it could be proved optimal, the transmission policy can take this idea into account in order to make the best out of the available resources. Since the capacity of a relay channel is not known in general [21, 22], it is highly intriguing to look for the performance limits of HARQ transmission over relay channel.

1.1.5 Objectives

The main question we look to answer in this research is how to find the performance limits of truncated IR-HARQ in single-hop and relay channels. Our vision is to find the optimal transmission rate policy for HARQ in different channel models. We are interested in finding out how the feedback of an HARQ channel can be exploited in order to adjust transmission parameters in useful way.

In a block fading channel with *independent and identically distributed* (i.i.d.) channel states, HARQ feedback which is transmitted back to the encoder at the end of a transmission attempt, can never help to estimate the instantaneous CSI. Although, it may help to make the best out of the transmission time by giving the transmitter information about how far the decoder is from a successful decoding.

Fixed-outage performance of block fading channel is an interesting topic which has been addressed in the literature very often [2]. This is an as interesting question to be

investigated for HARQ transmission. We look for an optimal resource allocating policy to increase the throughput with respect to the fixed-rate fixed-power conventional HARQ.

The complexity of an optimization problem is in most cases an exponentially growing function of the number of optimization parameters. Using the well-known *Dynamic Programming* (DP) technique is one of the approaches we would like to investigate as the possible simplifying methods to decrease the complexity of optimization problems. Thus, the questions to be addressed in the thesis are

- What is the throughput-optimal transmission rate in a fading channel when using truncated HARQ?
- Can we get better fixed-outage performance in a block-fading channel by applying an optimized HARQ transmission protocol?
- What is the effect of outdated CSI on throughput for HARQ?
- What are the limits of performance for a truncated HARQ?
- How well can a cooperative channel perform in terms of achievable throughput, using HARQ?
- How robust are the optimal policies in fulfilling the achievable throughput with respect to the limitations on the feedback message?

1.2 Outline and Contributions

1.2.1 Contributions

The following summarizes the major contributions of this dissertation:

- This work analyzes optimal rate control policies for single hop and relay block fading channel to meet the best throughput performance of HARQ transmission process. We study two different scenarios: first, where the feedback message in the HARQ process is limited to a single bit of ACK or NACK and second, when there is no limit imposed on the length of this feedback message. We show that the two scenarios require different optimization approaches. In the first scenario we call the optimal control policies *rate allocation* since it shall be allocated before the start of the process to all

the transmission attempts. In the second scenario, the control policy is called *rate adaptation* because the transmission rate can be adapted depending on the information in the feedback message.

- We show how to use the well-known DP methods [23] to optimize the rate adaptation policies for truncated HARQ in order to maximize the throughput under constraints on outage. We study the trade-off between average total number of channel uses and probability of packet dropping. The idea of using DP was already suggested for HARQ optimal design, e.g., in [24] and [18] however to the best of our knowledge, it has not been used for rate adaptation in truncated HARQ.
- For truncated HARQ, the rate allocation scenario is presented for both single hop and the relay channel, we optimize the presented closed-form representation of throughput. While, DP was already suggested to optimize the rate allocation for an infinite number of transmissions in [24], its practical implementation aspects were not shown. We argue that it may be difficult to solve the problem exactly and we propose suitable approximations.
- A dual to the optimization problem in both rate allocation and rate adaptation scenaria is presented. We prove that the solution to the dual problem is a globally optimal solution for the original optimization problem. This helps us to significantly reduce the complexity of the optimization problem.
- We introduce a cooperative rate adaptation scheme based on multi-bit feedback in multi-relay cooperative HARQ transmission with opportunistic relay selection. For the case of single-relay networks, we show that the presented scheme shows significant performance improvement over conventional non-adaptive fixed-rate cooperative HARQ. The rate adaptation process is studied as a *Markov Decision Process* (MDP) statistical model. We emphasize the benefits of such an approach and show how the complex problem of HARQ rate optimization can be reduced to a recursive algorithm which is significantly less complex.
- In the rate adaptation scenario, we study the effect of a discretized feedback message with limited number of bits on the performance of the presented scheme. The results suggest that a very small number of feedback message bits is sufficient to achieve the

maximum achievable throughput. In the rate allocation scenario for relay channel, we introduce the use of the *nested loop dynamic programming* for optimization of the throughput in a recursive problem.

- Finally, we show numerical results for practically interesting models of the wireless channel, comparing the throughput and outage for various rate adaptation/allocation methods. These results indicate that, specially in high SNR regime, a maximum number of retransmissions of $K = 4$ can give an optimal throughput close to that of the performance limit (i.e., ergodic capacity). We show that throughput-optimal rate adaptation may result in packet dropping and terminating the HARQ process before the last transmission is even reached. When an infinite number of transmission attempts are allowed, we present the performance limit for HARQ transmission in single hop and relay channels.

1.2.2 Document Organization

This dissertation is organized as follows:

- Chapter 2 presents the HARQ resource allocation problem and describes the channel under investigation in both cases of single-hop and multi-hop networks. We introduce the normalized accumulated information at the decoder as a measure of the decoder state in HARQ transmission. Then, based on the number of feedback messages being available for each transmission attempt, we categorize the problem into two main branches of *allocation* and *adaptation* studies. We discuss the throughput of HARQ channel and the outage probability and present them as the performance criteria in this dissertation and the limits of performance are stated for all the channel model setups. We show how the resource allocation problem in HARQ fits the definition of a MDP problem and finally, we will briefly discuss some of the practical concerns on the subject of resource allocation.
- In Chapter 3 we review the background on optimizations of transmission parameters over HARQ protocol. We will cover the most critical problems and the answers in this research area in five sections: exploiting feedback, optimal coding rate, joint rate and power optimization, cooperative HARQ and finally dynamic programming. For

each of the topics we will present the approach that we employed to the problem in this research thesis.

- In Chapter 4, we first establish the optimization framework that we use in this dissertation. We prove that the throughput maximization problem in both adaptation/allocation problems can be substituted with a dual problem in order to reduce the optimization complexity and cast it into a recursive minimization problem. Then, for the single-hop channel model the rate adaptation and the rate allocation problems in a fixed-power scenario are studied. Numerical results are presented in both cases, together with simplifying approaches to find the optimal results.
- Chapter 5 studies the orthogonal-relaying channel model. We first cover the rate adaptation problem and show in a case study how a few feedback message bits can be enough for an optimum rate policy to reach the performance limits. Then, for the rate allocation problem in the relay channel, we present two optimization approaches. Then the numerical results are presented with comparisons between the performance of rate allocation and rate adaptation methods.
- Chapter 6 covers all the concluding remarks and future research perspective.

Chapter 2

HARQ Resource Control Problem

ASSUMPTION of a feedback channel is inherent for HARQ protocols and it is used to inform the transmitter node whether if the destination has successfully decoded the previous packet (ACK) or has failed in decoding it (NACK). Figure 2.1 shows fading channel with feedback. However, the amount of information conveyed by the feedback signal may be different based on the capabilities of the feedback channel and the protocol design. We categorize the types of feedback into two categories as follows.

- A one-bit feedback (ACK/NACK signal): This type of feedback is only capable of informing the transmitter node as to whether the decoding has been successful or not.
- A feedback conveying more than one bit of information: Here we can expect the transmitter to become somewhat aware of the state of the decoder, which implies how close the decoder is to a successful decoding.

This dissertation focuses on the assumption of having error-free feedback channels while we note here that there are studies in the literature on the effects of feedback error on the performance of HARQ (See [25–27]). In this chapter, for the two scenarios explained above, we present the resource control problem for HARQ in block-fading channel. First we describe the channel model for single hop and relay transmission and talk about the decoder state. Then, we discuss the performance criteria that we want to focus on in this dissertation and the effects on these performance criteria, of having CSI at the decoder/encoder .

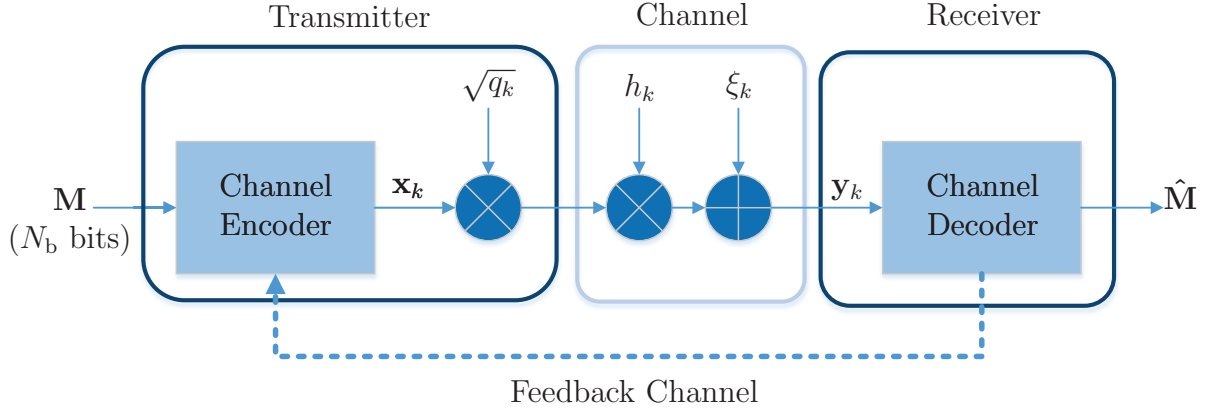


Figure 2.1 Block-fading channel model with feedback.

2.1 Channel Model with no Transmitter CSI

In this dissertation, we analyze the case where CSI is available at the decoder, but not at the encoder. Because of the behavior of block-fading channel, estimation of the CSI at the receiver is relatively simple and results in a negligible loss in the transmission rate [7, 28, 29]. We use the term *outdated-CSI* to describe the case where transmitter CSI is not available but instead, an outdated version of the CSI i.e., some information about the history of the channel state in the HARQ process, is available at the encoder (See [30]). In all cases we assume that the distribution of channel state (i.e., statistics of the channel or *Channel Distribution Information* (CDI) [2]) is known to all parties.

2.1.1 Single-hop Channel Model

In this case, we assume a channel with two nodes: *Source* ($\mathcal{S}$) and *Destination* ($\mathcal{D}$). A packet of N_b bits of information is to be delivered to node $\mathcal{D}$ while it is only known at node $\mathcal{S}$ at time zero. The packet is encoded into a codeword $\mathbf{X}$ of N_s symbols $x_1, x_2, \dots, x_{N_s}$. The codeword is chosen from a codebook which we denote by $\mathbf{C}^{N_s}$ and is generated from random and independent symbols with complex zero-mean unit variance Gaussian distribution and is already revealed to all nodes. We are interested in IR-HARQ so each codeword will be partitioned into disjoint subsets of symbols x_i to create sub-codewords $\mathbf{x}_k$ for $1 \leq k \leq K$.

The lengths of sub-codewords are chosen according to the coding rate policy. This is depicted in Figure 2.2.

Encoding

The encoding function acts from the set of N_b -bit messages to the set of codewords $\mathbf{C}^{N_s}$ generated randomly according to $\prod_{i=1}^{N_s} p_{x_i}(x_i)$ with the *probability density function* (pdf) $p_x(x)$ being a Gaussian distribution.

At the start of the IR-HARQ process, a sub-codeword $\mathbf{x}_1$ of $N_{s,1}$ symbols is transmitted to $\mathcal{D}$. Then, $\mathcal{S}$ receives an error-free feedback from $\mathcal{D}$. In case that the decoding has failed, the transmitter sends the second sub-codeword $\mathbf{x}_2$ composed of $N_{s,2}$ symbols and listens to the new feedback from $\mathcal{D}$. This process will go on until the message is successfully decoded at $\mathcal{D}$ or the maximum number of allowed transmission attempts K has been reached (truncated HARQ).

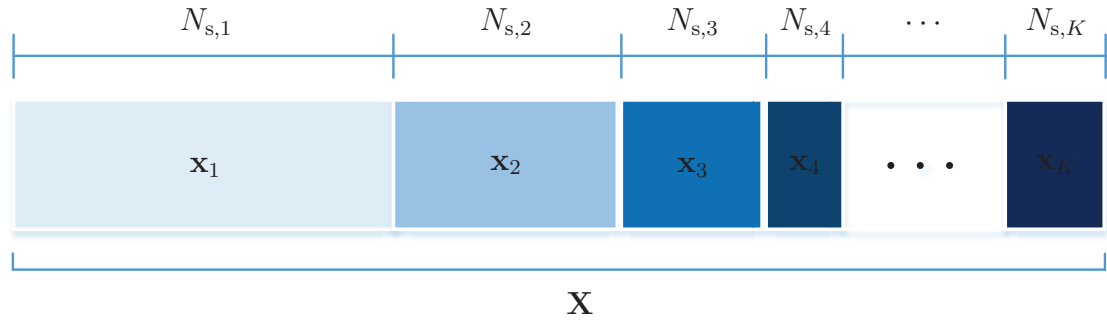


Figure 2.2 A codeword $\mathbf{X}$ and the sub-codewords $\mathbf{x}_k$, for $1 \leq k \leq K$. The length of the sub-codewords can be chosen by the encoder in variable-rate transmission. For fixed-rate transmission though, $N_{s,k} = \frac{N_s}{K}$, $\forall k$.

Channel Parameters

The channel is block-fading and as a result it remains constant during each transmission period. The received signal during the k th transmission attempt ($1 \leq k \leq K$) can be shown

as

$$\mathbf{y}_k = h_k \sqrt{q_k} \mathbf{x}_k + \xi_k, \quad (2.1)$$

where h_k denotes the complex channel gain of the channel and ξ_k is the zero mean unit variance complex Gaussian noise of the channel at k th transmission and q_k is the transmit power (Figure 2.1). Since the message symbols and the noise are assumed to have unit variance the SNR will be $\gamma_k = |h_k|^2 \cdot q_k$. The channel gain is assumed to be perfectly estimated at the receiver but unknown to the transmitter. Moreover, it varies independently from one transmission to another. This corresponds to a practical model where subsequent sub-codewords are sent in non-adjacent time instants and are separated by several channel coherence times.

The transmit power q_k is a parameter to be set by the transmitter, which can be also used besides transmission rate as a means to optimize throughput, and truncated HARQ implies $q_k \equiv 0, k > K$. A constant power transmission discipline sets $q_k \equiv q_1, \forall k$.

Decoding

A code combining approach as shown in Figure 2.3, is assumed at the decoder node based on all the observations that the decoder has of the transmitted sub-codewords up to the decoding time [1]. That is, an observation $\mathbf{Y}$ is generated using all the received symbols y_i up to the decoding time k where $1 \leq i \leq \sum_{l=1}^k N_{s,l}$, and $N_s - \sum_{l=1}^k N_{s,l}$ dummy symbols z_i generated independently from the received signal, for the spots where no symbol has been transmitted yet. The decoder function is then from the set of codewords $\mathbf{C}^{N_s}$ to the set of all N_b bit messages.

Redundancy Variables

From now on in this dissertation, we will use a normalized version of the sub-codeword length $\rho_k = N_{s,k}/N_b$ that is interpreted as *redundancy*. ρ_k is equal to the inverse of transmission rate R_k and is measured by the number of channel uses per information bit. As we will see in the following chapters, the main goal in analyzing the HARQ problem in here would be optimization of throughput for rate allocation/adaptation using ρ_k s. In case of rate adaptations $\rho_k(\cdot)$ is a function of the decoder state of $\mathcal{D}$ and may change with k . On the

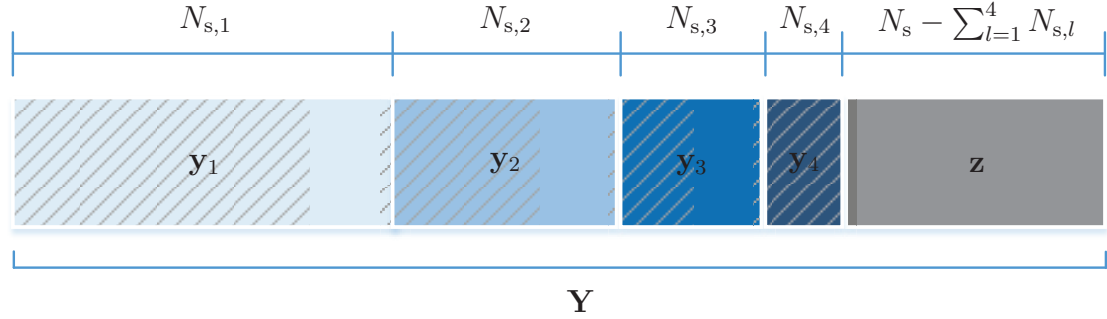


Figure 2.3 A decoding codeword $\mathbf{Y}$ after 4 transmission attempts. The codeword is generated using the observed sub-codewords $\mathbf{y}_k$ for $k = 1, 2, 3, 4$ and the dummy sub-codeword $\mathbf{z}$.

other hand for rate allocation ρ_k s are constant values which may only vary by k . The special case of $\rho_k \equiv \rho_1, \forall k$ is fixed-rate transmission which is already discussed in [1] and [12].

2.1.2 Normalized Accumulated Mutual Information

In the analysis of HARQ the *ACcumulated Mutual Information* (ACMI) is used in most of the research works in the literature as a measure for the decoder state. It is proved to be an appropriate measure for theoretical analysis of HARQ (particularly ACMI is introduced and analyzed for different HARQ types in [1] and [31]). For an information theoretic model with asymptotically long codewords, the ACMI after k transmission attempts with a code combining decoder, denoted with $\mathbf{I}_k$ is defined as follows [1, 18, 31–33].

$$\mathbf{I}_k = \sum_{l=1}^k \frac{N_{s,l}}{N_{s,1}} C_l \quad (2.2)$$

where, for the model described in Section 2.1.1, $C_l = C(\gamma_l) = \log_2(1 + \gamma_l)$ with γ_l being the SNR received at decoder during the l th transmission period.

Decoding is successful if the ACMI is not less than the targeted transmission rate of the

initial transmission (i.e., if $\mathbf{I}_k \geq R_1 = \frac{N_b}{N_{s,1}}$). This can be interpreted as follows.

$$1 \leq \sum_{l=1}^k C_l \frac{N_{s,l}}{N_b} \quad (2.3)$$

or,

$$1 \leq \sum_{l=1}^k C_l \rho_l. \quad (2.4)$$

Using (2.4), we define $I_k = \sum_{l=1}^k C_l \rho_l = \frac{\mathbf{I}_k}{N_b}$ as the *Normalized ACcumulated Mutual Information* (NACMI) after k transmission attempts. This definition makes the analysis simple since the NACMI does not depend on the initial transmission rate and HARQ is successful if and only if $I_k \geq 1$.

The NACMI model can be used for adaptation/allocation analysis as we will show in the following chapters. We note here that the ACMI model has already been used for adaptation analysis in [30] and for allocation analysis in [34–36]. Besides the information theoretic analysis, the ACMI model has also been used in coded modulations in [37].

The NACMI quantity is a weighted sum of the channel capacities of the block fading channel experienced by the decoder in the past blocks. By getting informed of the series of $\{I_l | 1 \leq l \leq k\}$, because the encoder already knows the ρ_k variables, it will be able to find out the history of channel up to time k , i.e., the outdated CSI¹. Thus, choosing the NACMI measure as the feedback message of the HARQ process, we can notify the encoder of all the outdated CSI that has been experienced by decoder in the previous attempts. Clearly, $I_k = 1$ means a successful decoding of the message and any less value of this variable with $I_k \in [0, 1)$ implies decoding failure. A feedback channel which is capable of carrying more than one bit, can be used to transmit the value of I_k to $\mathcal{S}$. This value can be quantized into the arbitrary number of bits depending on the capacity of the feedback channel. The number of bits in the feedback message, as we pointed out in the beginning of this chapter, specifies the category of the rate optimization analysis.

¹Also note that CSI is the only information that the decoder node has it and is worthy to share with the encoder.

Rate Allocation

We will shortly see that rate allocation, as the first scenario, is a special case of the second where I_k is discretized into one single bit. However, the reason why it is a different scenario is that in the case of a single-bit feedback message, the encoder has no side information to adapt the transmission rate based on it and it can only decide between zero or an allocated amount of redundancy to be transmitted. In other words, for a maximum of K transmission attempts, the transmitter can allocate upfront the set of K transmission rates to be used in case it will receive a NACK. Throughout this dissertation we will use the term rate allocation to address optimization of transmission rates for this category of IR-HARQ. On the other hand with a feedback containing more than one bit of information, at the beginning of k th transmission attempt, rate can be adapted with respect to I_{k-1} in a way (which we call it rate adaptation) that might be more beneficial to throughput than allocating the rate set upfront.

Rate Adaptation

A discretized version of I_k can be substituted interchangeably with the CSI experienced by the receiver at the previous attempt. This CSI information, when delivered at the transmitter node, is outdated and independent of the current CSI. However, the transmitter can decide on the transmission rate based on the outdated CSI by inferring the amount of information that the decoder has gathered so far. Using outdated CSI for throughput optimization has been noticed earlier in research works in [17, 30, 31, 34] and we will cover in details in Section 4.2 and Section 5.1 what make our approach different from the previous works.

2.1.3 Multi-hop Channel Model

A cooperative version of the channel model described above has a relay node added to the network which will aid the transmission. In this section we first describe the topology of the cooperative channel and the cooperative protocol of the problem. Then, the incremental redundancy scheme is defined on top of the channel model based code combining.

Relaying Protocol

The communication network considered here includes three nodes: Source ($\mathcal{S}$), Destination ($\mathcal{D}$) and *Relay* ($\mathcal{R}$), each having a single half-duplex radio. We assume a block fading channel and an error free feedback network between all nodes. At the end of each transmission attempt, the receiver node(s), can transmit their feedback signals back to transmitter node.

Node $\mathcal{R}$ is a decode-and-forward communication party so, it can start transmitting to $\mathcal{D}$ only after successful decoding of the message from node $\mathcal{S}$ which is the only node having knowledge of the N_b -bits message at time zero.

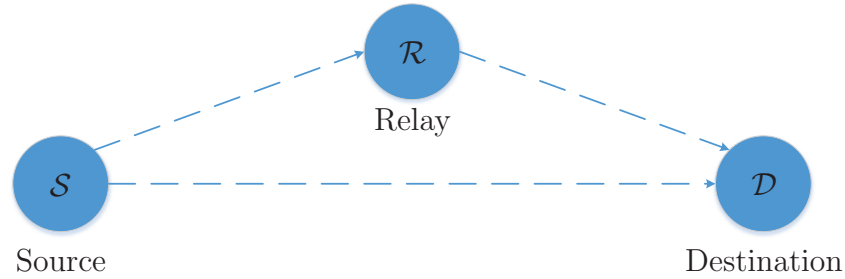


Figure 2.4 Topology of the relay channel.

As depicted in Figure 2.4 node $\mathcal{R}$ is positioned somewhere closer to $\mathcal{D}$ than $\mathcal{S}$, implying that, on average, it has a better channel to $\mathcal{D}$ than what node $\mathcal{S}$ has. The goal of the network is to deliver the message from $\mathcal{S}$ to $\mathcal{D}$ using the cooperation of $\mathcal{R}$.

The transmission starts with $\mathcal{S}$ *broadcasting* the encoded message to all the other nodes in the network. We call this the broadcasting phase. While one node is transmitting the other nodes are only listening and remain silent. The transmission process ends when node $\mathcal{D}$ successfully decodes the message. If node $\mathcal{R}$ decodes the message first, then it relays the message to $\mathcal{D}$ and node $\mathcal{S}$ goes silent. We call this the *relaying* phase.

Adaptive Incremental Redundancy Transmission

Coding and decoding is the same as described in Section 2.1.1: We assume that random channel codes are used through all transmission attempts and a packet of N_b information bits is encoded into N_s random complex symbols $x_1, x_2, \dots, x_{N_s}$. The symbols are samples of a zero-mean unit-variance complex Gaussian random process. The channel code is revealed

to all nodes. Here, K is the maximum total number of transmissions performed by source and relay nodes together.

At each transmission attempt a subset of the N_s symbols is transmitted. At the start of the ARQ process nodes are in the relaying phase in which only $\mathcal{S}$ has the message. So, source has to decide the number of symbols $N_{s,1}^s$ for broadcasting to $\mathcal{D}$ and $\mathcal{R}$. The selected sub-codeword $\mathbf{x}_1$ would be modulated into a signal and transmitted over the block fading channel. Then $\mathcal{S}$ will receive feedback signals from both receiver nodes. In the case that both receivers have failed decoding the message, the same process will be repeated with new decision on $N_{s,2}^s$ based on the knowledge that $\mathcal{S}$ has gathered from feedbacks. This will keep going until either $\mathcal{D}$ or $\mathcal{R}$ decodes the N_b bits message successfully or the maximum number of transmission attempts K has been reached.

If $\mathcal{R}$ can successfully decode the message at any time k ($k < K$) then we go into the broadcasting phase when $\mathcal{S}$ will go silent. In this phase, $\mathcal{R}$ starts transmitting the message by making a decision on the number of symbols $N_{s,k+1}^r$ to create a sub-codeword $\mathbf{x}_{k+1}$. Since incremental redundancy is assumed to be implemented and the sub-codewords at each transmission attempt are meant to be disjoint codes, $\mathcal{R}$ will choose the $N_{s,k+1}^r$ symbols from the subset of symbols in the corresponding codeword that have not been transmitted by $\mathcal{S}$ before. This process will continue for $K - k$ attempts or will stop if $\mathcal{D}$ decodes the message. We assume that N_s can be arbitrarily large which results in an arbitrarily low rate. This will allow us to apply no constraint on the size of sub-codewords.

Here, like in the single-hop channel model, we use normalized sub-codeword length (redundancy) $\rho_k = N_{s,k}/N_b$. Encoders at the transmitter nodes can decide on the amount of redundancy to transmit, based on the knowledge they get from the feedback channels embedded in the network. For the case of rate adaptation, ρ_k s are functions and in rate allocation they are constant values changing with k (variable rate). This makes our work different from that in [38] which analyzes the special case of $\rho_k \equiv \rho_1, \forall k$.

Similar to the channel model in Section 2.1.1 the signal over the k th transmission attempt ($1 \leq k \leq K$) transmitted by node a and received at node b can be written as

$$\mathbf{y}_k^b = h_k^{ab} \sqrt{q_k^a} \mathbf{x}_k^a + \eta_k \quad (2.5)$$

with h_k^{ab} being the complex channel gain of the channel between nodes a and b and η_k being the zero mean unit variance complex Gaussian noise of the channel at the k th transmission.

Again $\gamma_k = |h_k^{ab}|^2 \cdot q_k^a$ is the SNR experienced at node b of this transmission attempt while q_k^a is the transmission power of transmitter node a , during the k th attempt.

Decoder Accumulated Information

Information at the decoder defined in Section 2.1.2 is cumulative over the discrete time and the state of the dynamic system is defined via the pair (I_k^D, I_k^R) , the accumulated information at the decoders of $\mathcal{R}$ and $\mathcal{D}$ nodes. Therefore, the system can be described using the following relations:

$$I_k^D = I_{k-1}^D + \rho_k^S \cdot C_k^{SD} + \rho_k^R \cdot C_k^{RD}, \quad (2.6a)$$

$$I_k^R = I_{k-1}^R + \rho_k^S \cdot C_k^{SR} \quad (2.6b)$$

for $k = 1, 2, \dots, K$ with $(I_0^D, I_0^R) \equiv (0, 0)$, i.e., the destination and the relay have zero information at time zero. The decision variables are (ρ_k^S, ρ_k^R) and the random parameters are $(C_k^{SD}, C_k^{SR}, C_k^{RD})$.

Successful decoding condition in here is the the same as in (2.4). So, we can interpret the broadcasting phase as the state where $I_k^R < 1$ holds and the relaying phase as the state where $I_k^R \geq 1$ (noting that at both phases we have $I_k^D < 1$). Moreover, the transmission process stops when $I_k^D \geq 1$ for some $k < K$ or the maximum number of transmission attempts K is reached.

Transmission Parameters

The redundancy transmitted from $\mathcal{S}$ and $\mathcal{R}$ at attempt $k \in \{1, \dots, K\}$ can be written as

$$\rho_k^S = \begin{cases} \rho_k^S(I_{k-1}^D, I_{k-1}^R) & I_{k-1}^D < 1, \text{ \& } I_{k-1}^R < 1 \\ 0, & \text{o.w.} \end{cases} \quad (2.7a)$$

$$\rho_k^R = \begin{cases} \rho_k^R(I_{k-1}^D) & I_{k-1}^D < 1, \text{ \& } I_{k-1}^R \geq 1 \\ 0, & \text{o.w.} \end{cases} \quad (2.7b)$$

where the $\rho_k^{\{\mathcal{R}, \mathcal{S}\}}()$ are constant functions in the rate allocation scheme. Similarly, the transmit power can generally be a function and it can be written as

$$q_k^{\mathcal{S}} = \begin{cases} q_k^{\mathcal{S}}(I_{k-1}^{\mathcal{D}}, I_{k-1}^{\mathcal{R}}) & I_{k-1}^{\mathcal{D}} < 1, \text{ \& } I_{k-1}^{\mathcal{R}} < 1 \\ 0, & \text{o.w.} \end{cases} \quad (2.8a)$$

$$q_k^{\mathcal{R}} = \begin{cases} q_k^{\mathcal{R}}(I_{k-1}^{\mathcal{D}}) & I_{k-1}^{\mathcal{D}} < 1, \text{ \& } I_{k-1}^{\mathcal{R}} \geq 1 \\ 0, & \text{o.w.} \end{cases} \quad (2.8b)$$

with the $q_k^{\{\mathcal{R}, \mathcal{S}\}}()$ being constant functions in the allocation case.

We use the term *policy* to describe a set of transmission control parameters for a HARQ process and we denote it with π . A control policy in general might include all transmission parameters such as power, rate, modulation, etc. In this dissertation whenever we use the term policy it refers to the set of all redundancy functions. Thus, in case of a single-hop transmission a policy π is: $\pi = \{\rho_k | 1 \leq k \leq K\}$, while for the cooperative transmission it is: $\pi = \{(\rho_k^{\mathcal{S}}, \rho_k^{\mathcal{R}}) | 1 \leq k \leq K\}$. From now on in our analysis we will assume fixed-power transmission which is why the q variables do not show up in the policy set π .

2.2 Performance Criteria

In the channel models described above, an outage happens if decoding fails at node $\mathcal{D}$ after at most K transmission attempts. In this section we consider the two channel models in a similar approach to introduce outage probability and throughput as the performance criteria in our optimization problem.

2.2.1 Outage Probability

Decoding at node $\mathcal{D}$, as it is shown in [1] and discussed more in [18, 33], can asymptotically have an arbitrarily small failure probability at the k th transmission attempt, for $1 \leq k \leq K$ (considering large values of $N_{s,k}$ and as a result large value of N_s), if the code rate is less than the accumulated information about the transmitted codeword up to time k or

$$R < \sum_{l=1}^k I(p_x(x), p_l(y|x)) \quad (2.9)$$

where $p_x(x)$ is the pdf of the code generator and $p_l(y|x)$ is the single letter transmission pdf of the channel.

For a block fading channel, the maximum achievable $I(p_x(x), p_l(y|x))$ during transmission block l is the capacity of the channel C_l using the maximizing $p_x(x)$. For asymptotically large sub-codewords, the probability of decoding failure after k decoding attempts can be arbitrarily small if the normalized accumulated mutual information up to k th transmission is larger than 1. On the contrary, we have an outage if the normalized accumulated mutual information up to k th transmission is smaller than 1, which can be written as follows with I_k defined in (2.3) [12, 38]:

$$\Pr(\text{error} | I_k > 1) = 0, \quad (2.10a)$$

$$\Pr(\text{error} | I_k < 1) = 1. \quad (2.10b)$$

After a truncated HARQ process is finished there is still a possibility that the message is not decoded correctly at the receiver. This is called outage and the probability of an outage happening P_{out} , is equal to the probability that I_K is less than one.

Here P_{out} is equal to the probability that $I_K^{\mathcal{D}}$ is less than one:

$$P_{\text{out}} = \Pr\{I_K^{\mathcal{D}} < 1\}. \quad (2.11)$$

2.2.2 HARQ Throughput

The throughput η of the presented channel is equal to the number of bits decoded successfully divided by the number of channel uses up to time t while $t \rightarrow \infty$. With $c(t)$ denoting the number of channel uses and $b(t)$ counting the number of successfully decoded bits up to time t we have

$$\eta = \lim_{t \rightarrow \infty} \frac{b(t)}{c(t)} \quad (2.12)$$

which, based on the renewal-reward theorem [1], changes into

$$\eta = \frac{\overline{N_b}}{\overline{N_s}}, \quad (2.13)$$

with $\overline{N}_b$ being the expected number of correctly received bits. Also, $\overline{N}_s$ denotes the expected number of channel uses (or transmitted symbols) which can be written as

$$\overline{N}_s = \sum_{k=1}^K \overline{N}_{s,k} = \sum_{k=1}^K \mathbb{E}\{N_{s,k}\} \quad (2.14)$$

while $\overline{N}_{s,k}$ is the expected number of channel uses in the k th transmission attempt.

For a truncated HARQ transmission the number of successfully decoded bits at the end of the process can be zero with probability of P_{out} or N_b with probability $1 - P_{\text{out}}$ which changes (2.13) as follows.

$$\eta = \frac{N_b(1 - P_{\text{out}})}{\overline{N}_s} \quad (2.15)$$

Maximization of throughput η for a fixed-power scenario will be done based on all of the transmission redundancies (ρ variables). We denote the optimal throughput with $\hat{\eta}$.

$$\hat{\eta} = \max_{\pi} \eta(\pi) \quad (2.16)$$

An outage constrained optimal throughput is also defined as follows.

$$\begin{aligned} \hat{\eta}_{\epsilon} &= \max_{\pi} \eta(\pi), \\ \text{s.t. } & P_{\text{out}}(\pi) \leq \epsilon \end{aligned} \quad (2.17)$$

2.3 Markov Decision Process Framework

The problem model described earlier can be cast into MDP model. The general framework of an MDP process, as shown in Figure 2.5, with the discrete time k consists of the following elements [39]:

- The system is measured by its *state* parameter S_k which is a function $f_S(\cdot)$ of the previous state S_{k-1} , action A_{k-1} and disturbance ω_{k-1} ;
- A controller that creates the *action* A_k using the function $f_A(\cdot)$;
- A *cost* (or *reward*) associated to a state transition;

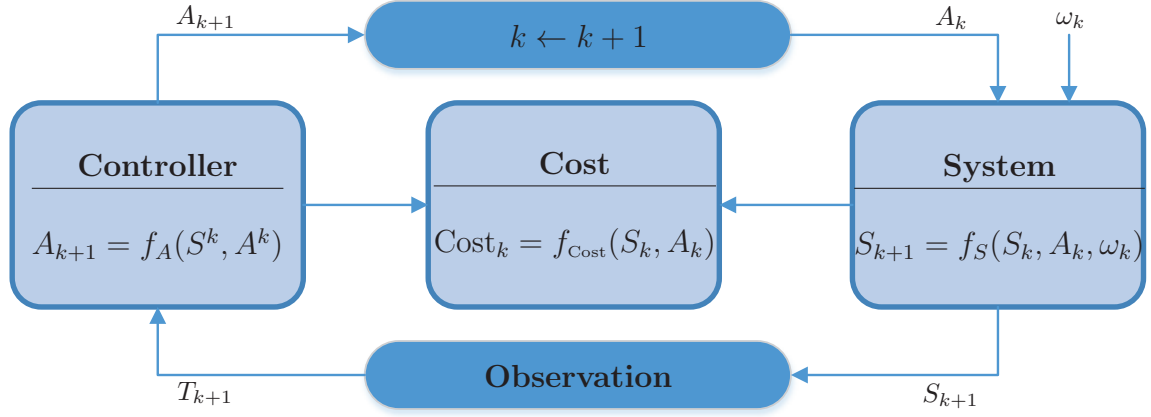


Figure 2.5 Flowchart of an MDP model: Note that $S^k = \{S_1, \dots, S_k\}$ and similarly $A^k = \{A_1, \dots, A_k\}$.

- Transition *probabilities* $\Pr(S_{k+1} = s' | S_k = s, A_k = a)$.

An MDP state S_k is a Markov function of its history and two other parameters; the action A_k and disturbance ω_k . At time step k , the controller, which might aim at decreasing the long-term total cost of the system, observes the last state measure of the system through an *observation* block. Normally, the observation T_k is equivalent to the state S_k . However, in some examples T_k is a modified version of the state S_k , e.g. a quantized version of it, which changes the model to a *Partially Observable Markov Decision Process* (POMDP).

The analogues of the above elements in our problem model are as follows. The normalized ACMI I_k is the state S_k in Figure 2.5. The action A_k in our problem model is the redundancy parameter ρ_k . The cost being paid at time k is the number of channel uses that a transmission takes or $N_{s,k}$. The transition probability is determined by the channel coefficient.

The observation block in Figure 2.5 determines whether the state of the HARQ system (I_k) or a modified version of it (e.g., a single bit quantized version of it as an ACK/NACK) is being observed at the controller (encoder). In other words, the difference between rate allocation and rate adaptation can be seen in the observation block in Figure 2.5.

The outdated CSI transmitted back to node $\mathcal{S}$ on the feedback channel gives information about the state of the decoder at the destination node. This decoder state can be seen as a random state of the process which is predictable for the future of the process solely by the

action (i.e., redundancy values) and the disturbance (i.e., channel state random process). This makes our problem suitable to be dealt with using DP [23].

2.4 Practical Concerns

In the system-level implementation of the variable-rate HARQ described above, it may be assumed that each transmission contains only one sub-codeword. In such a case, the duration of transmission attempts must vary, which might be a valid approach for a single-user communication when the transmitter and the receiver can negotiate the transmission time for each sub-codeword. On the other hand, it may be a questionable strategy in multi-user communications, where sharing the requirement for a variable-length transmission with all the users is not practical.

However, even if it might be possible to assign the time resources independently of the varying transmission length, it would lead to bandwidth waste (sub-codewords shorter than the assigned transmission time slot) or to collisions (sub-codewords longer than the assigned time slot). Because of this, in the design of the system model described in this chapter, we assume an implementation approach similar to [30], which is described as follows.

We can still assume a communication network with a fixed transmission time (*Time Division Multiple Access* (TDMA)) and let the variable-length packets of our system lie inside the fixed-rate frames of the TDMA channel. To avoid such a conceptual difficulty, we assume that the sub-codewords corresponding to different packets are gathered in frames that have a constant duration of N_F symbols. Such an assumption, also used in [40, 41] allows us to deal with variable-length codewords to fill up the frame and corresponds to TDMA-type communication, where users are provided with a fixed transmission time (i.e., fixed frame length). An example is shown in Figure 2.6.

The structure of four frames sent over channel realizations h_1 to h_4 is shown in Figure 2.6. The sub-codeword packets are denoted by $\mathbb{P}_i$ for $i = 1, \dots, 13$. All the frames have the same length of N_{frame} . The sub-codewords corresponding to different transmission attempts are shown by different colors. A re-transmission for a packet $\mathbb{P}_i$ happens only if decoding has failed in the previous attempts of it and the packet has been transmitted less than K times. The HARQ process for the first five packets takes 4 transmission attempts to terminate while this number for packets $\mathbb{P}_6, \mathbb{P}_7$ is 2. The two packets $\mathbb{P}_8, \mathbb{P}_9$ are transmitted only once. The differences in the number of attempts for different packets might be because of any of

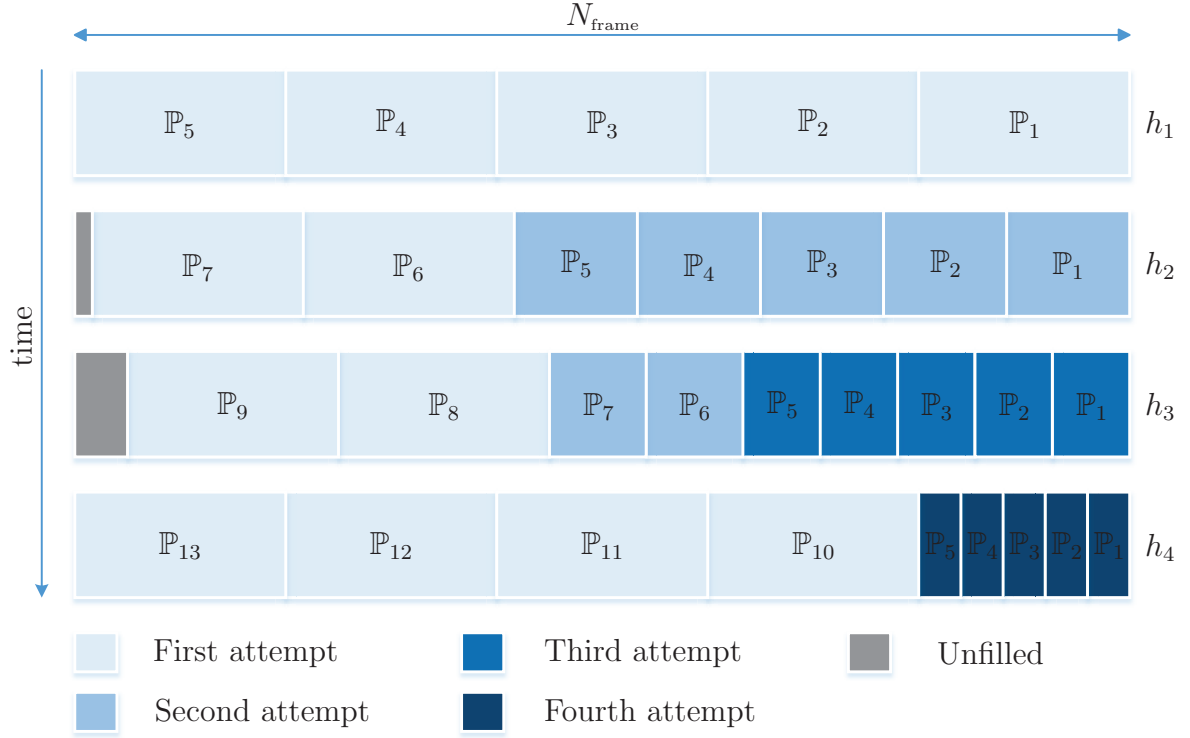


Figure 2.6 Example of the structure of frames for variable-rate HARQ transmission.

the following reasons:

1. The packet is delivered correctly and the feedback message from the decoder stops the encoder from re-attempting the same packet.
2. The controller unit (encoder), decides on discarding (dropping) the packet. This can happen in rate adaptation scenario, if the decoder is in a poor state (i.e., very low ACMI) and continuing on the retransmissions of the same packet is too costly for the objective criteria. We will talk more about the packet droppings in Chapter 4.

The length of the sub-codeword in the k th transmission attempts for $1 < k \leq K$, might vary from frame to frame based on the feedback message, only in rate adaptation scenario. Also, we note that the relative loss due to unfilled space in the frames can be made arbitrarily small, by loading the frame with many sub-codewords.

Due to unavoidable processing and communication delays, the CSI at moment of the k th transmission attempt is not the same as the CSI that may be obtained from the receiver via feedback. In general, if channel coefficients for different transmission periods were correlated, they might be used to predict the instantaneous CSI. However, we assume that the channel gains are i.i.d. random variables. This means that they might not be used to predict the instantaneous CSI and as a result our case can be considered the worst case from the point of view of usefulness of the outdated CSI. This makes our work different from what is presented in [21] and [42] where the channel coefficients are assumed correlated and used to estimate CSI.

Chapter 3

Answers In the Literature – State of the Art

IT is known that different types of *diversity* (spatial/frequency/time) can help in increasing the performance of wireless communication. Truncated HARQ with K maximum transmission attempts serves as a form of retransmission diversity. Such a process has transmission parameters like coding rate and transmit power as degrees of freedom that can be used to improve the overall performance.

As we already explained in Chapter 2, any HARQ process can be seen in general as an MDP process: For a data packet of N_b bits, at the beginning of the HARQ process a decision is made about the transmission parameters. This decision along with a random process of the channel state form the state of the decoder at the receiver node. Then a feedback signal is fed back to the transmitter, informing it about the success/failure of the decoding process. Based on the received feedback signal, the next decision is made by the transmitter for the second attempt. Decision parameters in such an MDP process can be the coding rate, transmission power, etc.

It has been of much interest in the literature to optimize HARQ transmission using as many degrees of freedom as possible. Moreover, a cooperative network of communication nodes, exploiting an HARQ protocol has been of a considerable interest since it shows significant improvements over conventional multi-hop relaying protocols [20, 43–46].

In this chapter we study the state of the art research on HARQ and optimization of transmission control parameters. We will review the cooperative HARQ topic in the

literature and discuss the control policies adaptable to a relay network. At the end of each subsection, we highlight our contribution to the particular sub-area. In the final section, we discuss the DP technique and will show how it can be used in optimizing the throughput of IR-HARQ.

3.1 Exploiting Feedback

A communication system with a feedback channel, as depicted in Figure 2.1, is called a *closed-loop* communication system as opposed to an *open-loop* communication system, that has no feedback channel and relies only on the CSI at the receiver. The idea behind using a feedback is thus informing the transmitter about the CSI. Based on the scheduling discipline that is used in the communication system, the feedback is in one of the following approaches:

- **Quantized instantaneous CSI:** This consists in giving the transmitter a rough measure of the CSI before starting the transmission [17, 47]. With the partial information about instantaneous CSI, the transmitter can adapt the transmission parameters to reduce chances of outage. However, in practice, the reported CSI can never be fed back perfectly and without quantization/compression. In [48, 49], it is shown that a compression on the quantized CSI with a ratio of 5 – 30 will still keep the throughput virtually the same in an adaptive OFDM-based system. A study on the effects of feedback reliability in [50] shows that the assumption of a noiseless feedback line with no limit on the power is not so far-fetched. The work in [50] assumes that the feedback channel is also affected by fading and has a limited transmission power. In this situation, for a high quality feedback channel, virtually the same performance as the noiseless feedback can be expected.
- **Outdated single-bit/multi-bit CSI:** This is the case studied in this dissertation, where an already experienced CSI is fed back to the transmitter. As we discussed earlier, this can be conveyed in a single-bit message of ACK/NACK, e.g. in ARQ and conventional HARQ transmission. It can also be a multi-bit message of a quantized outdated CSI. This is the assumption that is adopted in [30] [51–54] for an uncorrelated channel state. In [55], for temporally-correlated channels, the outdated CSI is used for beamforming purposes. Power allocation mechanisms with the outdated CSI is also studied in the literature [56, 57].

The MDP nature of HARQ process has made it highly desirable in the literature to study the resource allocation strategies that can make the best performance out of the retransmission attempts. One of the challenges here is how to exploit the feedback channel and what are the different ways of informing the encoder about the decoder state.

It is practical to assume that CSI is known (approximated) at the decoder and is not known at the encoder. Therefore, it is not feasible for the transmitter to predict its optimal decision unless a side information is provided to help the encoder to do so. In case of i.i.d. block-fading channel, the transmitter cannot predict the instantaneous channel state based on the information about the outdated CSI. However if the channel gains are correlated for different transmission attempts in the assumed model, partial information about the instantaneous CSI can be generated by the encoder using the feedback signal. A similar idea has been considered in [17] and [58].

The work in [17] studies the idea of using HARQ feedback channel as a means of informing transmitter about CSI to some extent. This is mainly the idea behind rate/power adaptation for ARQ protocols. The intuitive question presented in [17] is on the possibility of gaining benefits out of using HARQ feedback channel in a way different from the classic way. It is customary in HARQ transmission that the feedback conveys only a one bit information at the end of each transmission attempt declaring decoding success/failure at the receiver party. So the question addressed in [17] is: how beneficial would it be to send the feedback at the start of transmission attempt in one (or probably more) bit(s) to inform the transmitter, who is by definition unaware of the instantaneous CSI, of the knowledge that receiver has about state of the channel? As shown in [17], a quantized version of the instantaneous CSI fed back to the transmitter as partial CSI can help to improve the performance of HARQ.

In [58], the conventional ACK/NACK feedback is assumed however, using the control actions and the single-bit feedback messages of several previous time slots, a history record is created to estimate the probability of ACK/NACK for the present time slot. This idea has been studied as a POMDP, over the ARQ.

As already discussed in Chapter 2, the information accumulated at the decoder can be considered as the state of the receiver. It can be assumed that the transmitter has a perfect knowledge of this state using a reliable feedback channel or has a partial knowledge of the state (like a quantized multi-bit or single-bit ACK/NACK version of it). Using this knowledge of the state, the transmitter would decide on its following actions and this goes

on until the state of the receiver decoder gets to a satisfactory level for successful decoding or the maximum number of attempts is achieved. This idea has been addressed in [30] for adaptive IR-HARQ with considering the feedback signal as outdated CSI.

In [30], it is assumed that the encoder does not have the instantaneous CSI, but that instead, it is capable of optimizing the transmission rates, using the outdated version of the CSI experienced by the decoder at the previous attempt. An example of incremental redundancy HARQ channel supports this claim in [30] while it is optimized for throughput with respect to transmission rate.

Our Approach

In this dissertation, we study the performance of the presented system model, in various feedback situations. First we start with an ideal feedback channel that has not limits and no errors (with un-quantized feedback message), to find the performance limit. Then we consider a very limited feedback capacity of only one bit per transmitted packet. This gives us the performance capacity of conventional HARQ. Finally, we investigate the required feedback rate that can reach the upper limit of the performance. This allows to emphasize on the values of the rate/power design with respect to fixed transmission strategies.

The rate adaptation HARQ presented in our research on the HARQ relay channel is inspired by the work in [30] and can be considered as a generalization over the special case of relay being positioned infinitely far from source and destination. We assume a perfect feedback that informs the encoder about the accumulated information at the decoder. This completes the MDP realization of the rate adaptation problem by choosing feedback signal as the state of the process.

3.2 Optimal Coding Rate

A fixed-rate transmission policy is clearly just a special case of HARQ transmission and not necessarily the optimal one. This is already discussed and proved otherwise in [31], showing an example of variable-rate HARQ that gain a better throughput than the fixed-rate IR.

It is best shown in [30] how significantly an optimal variable-rate policy can improve the throughput of IR-HARQ with respect to the fixed-rate case. In [30], a rate adaptation policy based on outdated CSI is introduced where $\rho_k(I_{k-1})$ are set to be functions of the accumulated information at the decoder.

As means of simplification of the problem, transmission rate policies in [30] are heuristically assumed to be linear as $\rho_k(I_{k-1}) = (1 - I_{k-1}) \cdot \bar{\rho}$ with respect to the accumulated information at decoder. The linear function coefficient $\bar{\rho}$ is then found using a simplified recursive optimization technique. It is shown that the optimized linear rate policy functions will result in a significant improvement for throughput compared to the maximum achievable throughput without using the outdated CSI.

In [59], optimization of throughput with respect to coding rate is studied for a constrained transmitter buffer occupancy in type-I HARQ. An MDP problem model is established with the composition of buffer occupancy, incoming traffic and channel state as the state of the process, ARQ feedback as the observation and transmission rate as the decision. DP optimization of the defined MDP is found to be complex to solve, therefore a number of heuristics have been employed to solve the problem. The numerical results presented in [59] are basically the suboptimal throughput and the corresponding average buffer occupancy with respect to the packet arrival probability for specific channel characterization and coding/modulation.

An approximate method is presented in [60], to derive the outage probability of a HARQ process. The combining effect of retransmission over a Rayleigh fading channel is approximated with a sum of Gamma variables. It is shown that the product of two independent Rayleigh distributed random variables could be approximated by a sum of two Gamma variables. This is extended to the case of Nakagami fading, when the involved variables' statistical moments are the same. Using this approximation method, [60] obtains the performance estimate of HARQ with code combining over the ideally interleaved Rayleigh fading channel. In the optimization of allocated rate we use the Gaussian approximation for outage probability to cast the complex problem into an iterative simplified optimization.

Our Approach

We present a general framework for rate optimization in IR-HARQ for both cases of limited and unlimited feedback capacity, based on the renowned DP technique. We present the throughput-optimal rate policies for truncated HARQ as well as the outage-constrained optimal policies, in both rate adaptation/allocation scenarios. In Chapter 4 the analysis is presented for a single-hop channel. Then in Chapter 5, we extend the analysis to a cooperative channel with a single relay node and the a general network with M nodes.

Moreover, we consider a random coding/code combining approach with unlimited buffer size for IR-HARQ therefore, presented results can be considered as the performance limit for the channel model. The heuristic assumption of linear redundancy functions in [30] is not exploited in this dissertation and, as we will show in the following chapters, this will result in finding the global optimal solutions and also prove the optimality of packet-dropping when the previous channel experiences have been poor.

3.3 Joint Rate and Power Optimization

Considering rate and power control parameters for optimization dramatically increases the complexity of the problem, firstly because of the increase in the number of optimization variables and secondly because of various constraints applied on the optimization at the same time. The general optimization of a truncated IR-HARQ transmission has not been studied in the literature so far. However, there are research works presented in the literature that consider both power and rate in their optimization problem to some extent, e.g., [58,61].

Minimization of the amount of transmitted power is usually aimed for one (or several) of the following reasons [4].

- To increase the battery life or allow using smaller batteries on the devices.
- To increase the number of users that a network can serve by reducing the interference between them.
- To reduce the chances of a user/communication line to be detected.

Moreover, energy saving and environmental protection has become a global demands and researchers have shown much interest in energy efficiency oriented design [62]. Moreover, for most applications the acceptable service entails an average rate with small probability of error with constraints relating to the data delivery delay (e.g., average delay constraint or maximum delay constraint).

The problem of power adaptation using the outdated CSI can be solved using the DP technique (as already remarked in [17]) if we change the constraint on the average power to a constraint on the total power. This means that the maximization of throughput can be done subject to the constraint of $\sum_{k=1}^K q_k \leq q_{\text{tot}}$ where q_k is the power policy at k th attempt and K is the maximum number of transmissions allowed. This modification may change

the problem into a classic DP problem. Here, we can assume $q_{\text{tot}} = K \cdot \bar{q}$. This comes from applying the short-term constraint on each transmission attempt separately since the outage probability on each attempt happens with transmission rate exceeding capacity, and capacity for each block depends only on the short-term power constraint on that block.

Moreover, an average power constrained representation of the maximum throughput is presented in [17] without giving the solution on the general case. In [17] author uses the long-term power constraint defined in [63] as a constraint to satisfy the total transmit power. The long-term power constraint is an average of the short-term power constraint when a sequence of transmission attempts occur. The short-term constraint satisfies the necessary condition of capacity problem.

In [58], type-I HARQ, which assumes discarding previously received signals on the same message, is studied in which the same packet is retransmitted upon receiving a NACK from the receiver. The system model comprises two nodes with a fading channel that can be slow fading or fast fading. A limited buffer size at the transmitter node is considered for rate and power optimization. Besides, the constraint on the delay is taken into account in some parts of the analysis.

Authors in [58] assume two different scenarios, one with perfect CSI knowledge at both receiver and transmitter and the other scenario where only receiver knows the CSI. In both scenarios, based on the one bit feedback from receiver at the end of each attempt, transmitter makes a decision on the transmission rate and power in an optimum way to meet the delay and buffer overflow constraints. Optimization is done using DP since the process is an MDP, and results are shown for special cases of channel characteristics and demanded delay/buffer overflow values.

The throughput or outage probability are not explicitly defined as the optimization goal in [58], which instead optimizes the weighted sum of criteria related to the buffer occupancy, overflow probability and the throughput. The system model has a known coding/modulation scheme which makes it different from our work where we assume random coding/code combining. In more general terms, [58] is an extension of the idea presented in [24] with different MDP state definition.

Two different approaches are exploited in [58] for optimization. First one is to add the different average cost functions together with weighting and minimize this summation, second one is to minimize one of the average cost functions subject to the other costs constrained to a certain value. As it is shown in [17] it can be more helpful if the observation

of the MDP process is sent back in the beginning of the attempt to inform transmitter about the instantaneous channel quality or as shown in [30] to use the outdated CSI as the channel state.

In [64], type-I and type-II HARQ transmission over a single-link channel with Rayleigh block fading is studied. It is assumed that the transmissions are fixed-rate fixed-power but there are design parameters available for energy-efficiency and delay trade-off which are the transmission time per bit and the total energy per bit. They present trade-off analysis for the design parameters and conclude that only for IR-HARQ the minimum total energy per bit decreases as well as the transmission time per bit if we increase the number of retransmission attempts.

The work in [61] is basically on the same idea presented in [58] with the same set of analyzation and goals except that this paper is on the IR-HARQ. Again, based on a system model that consists of limited buffer size at transmitter, with known coding/modulation schemes, state of the channel accompanied by the state of the feedback message and the incoming number of packets at transmitter buffer, is chosen as the state of MDP problem. Transition probability matrix for the Markov process is derived and the reward of the process is defined as a combination of three different costs of transmitted power, delivery delay, and buffer overflow with the transmission rate being the decision parameter. Using this model, long term average transmitted power is optimized over decision parameter (transmission rate) with constraints on maximum delay and overflow.

Our Approach

In this research we focus on the effects of variable-rate transmission. The more general problem (i.e., considering power variables in the optimization), provides an upper-bound on the performance of the model we consider here. However, to the authors knowledge, there is no research in the literature addressing this general problem so far which makes it an interesting topic for future work. What makes this work different from [61], is in the absence of outdated CSI at the transmitter node in the channel model of [61]. Moreover, the system model in [61] is defined for a specific coding/modulation technique as opposed to considering the general case to find the performance limits as is the case in this dissertation.

3.4 Cooperative HARQ

Cooperative HARQ is of much interest recently because in the modern wireless networks cooperation seems to be necessary to ensure the reliability and efficiency of transmission. While HARQ is named to be the practical mean of implementing real cooperation between nodes in a communication network [20], there are only a few research studies on optimization of HARQ in cooperative channel though channel models differ from one study to another which makes it difficult for comparison purposes.

The work in [65] discusses the cross-layer analysis of a cooperative channel without fading. On the MAC layer, network stability is defined as whether the maximum of the average overflow of a transmitter node with buffer size of ζ goes to zero by $\zeta \rightarrow \infty$. In the stability situation of a cooperative network, the throughput optimal rate region is analyzed.

In [20] the idea of a practical relay network based on HARQ is studied. A network of K communication node with $K - 2$ relays is assumed and it is shown that using HARQ with retransmission being done from the relays that have successfully decoded the message, outperforms both HARQ with retransmissions from source node and the multi-hop protocols (cascade of point-to-point).

The research in [38] studies cooperative HARQ and shows the throughput for three different protocols. The idea of this paper is basically generalizing the formulation of throughput calculation given in [1] for a relay channel. Rate and transmission power are assumed fixed over a block-fading channel and the only optimization presented in this paper is over the position of the relay node with respect to source and destination and also optimization of throughput over the value of the fixed rate. In this paper, outage probability at each transmission attempt is calculated using a Gaussian approximation. Moreover, [38] uses the state transition diagram to find a closed-form representation of throughput.

The work in [44] presents some bounds on the expected number of transmissions and the average throughput for a general multi-relay network. It assumes fixed-rate fixed-power transmission and the relays collaborate only when the decoding at destination is erroneous.

In [66], for a M -hop channel (i.e., $M - 1$ relay nodes), optimization of *long-term average transmission rate* (LATR) is studied for CC-HARQ for a limited number of retransmissions among whole hops. The work considers an outage-constrained approach using suboptimal search algorithms to find the optimal *round transmission rate* (RTR).

For a HARQ relay channel with one relay node, the throughput has been optimized

in [67] with respect to the transmission power of retransmission attempts. A closed-form of the throughput is approximated for high average values of SNR and with a fixed relay transmission power, and fixed power values for source node (but different from that of relay), and the results show improvement on throughput compared to fixed-power transmission. The total power constraint applied on the optimization process in [67] assumes a fixed total power transmission on each step of the HARQ on the sum of relay and source transmitted power. This totally differs from the long-term power constraint presented in [63].

The achievable rate region for any AWGN channel needs a constraint on the average total power. For a number of retransmissions this results in a long-term average total power on the all transmission attempts. This is well presented in [17] as a constraint on the optimization problem. On the other hand, assuming fixed rate on each node is definitely not the optimal case as implied in [24] or [17].

Our Approach

We study the cooperative channel model described in Chapter 2 for rate optimization. As it is presented in Chapter 5, the rate adaptation cooperative IR-HARQ shows a significant improvement in throughput over fixed-rate transmission. Rate allocation in cooperative channel is also studied in the same chapter and a framework is presented for analysis of the rate allocation problem in a other network realizations. We will show in Chapter 5 a closed-form of the throughput for variable-rate allocation IR-HARQ and present ideas for optimization of the multi-dimensional maximization.

3.5 Optimization with Dynamic Programming

A mathematical optimization problem with objective function $f_0(\cdot)$ and constraint functions $f_i(\cdot)$, can be shown in general in the following form for the optimization parameter vector π of size n .

$$\begin{aligned} & \text{minimize} && f_0(\pi) \\ & \text{subject to} && f_i(\pi) \leq \epsilon_i, \quad i = 1, \dots, m \end{aligned} \tag{3.1}$$

If the optimization problem has the following form

$$\min f_0(\pi), \quad (3.2)$$

and there are no imposed constraints, it is called an *unconstrained* optimization. For various types of optimization problems different techniques are presented as solution methods. However, most of the time, the general optimization problem is surprisingly difficult to solve. Therefore, approaches to solve a general optimization problem therefore involve some kind of compromise either in the computation time or on the possibility of not finding the solution [68]. In this regard, we can categorize optimization problems into the following classes:

- *Linear programming*: If the objective function and the constraint functions are linear functions, i.e., they satisfy

$$f_i(\alpha\pi_1 + \beta\pi_2) = \alpha f_i(\pi_1) + \beta f_i(\pi_2), \quad (3.3)$$

for all $\alpha, \beta \in \mathbb{R}$ and all $\pi_1, \pi_2 \in \mathbb{R}^n$, the optimization problem is called linear programming.

- *Least-squares*: It is an optimization problem where the objective function is a sum of squares of linear terms as follows.

$$\min f_0(\pi) = \sum_{k=1}^K (\alpha_k \pi - \beta_k)^2 \quad (3.4)$$

where $\alpha_i \in \mathbb{R}^n$ and all $\beta_i \in \mathbb{R}$.

- *Convex Optimization*: An optimization problem of the form in (3.1) is convex if $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$ is convex, i.e., it satisfies

$$f_i(\alpha\pi_1 + \beta\pi_2) \leq \alpha f_i(\pi_1) + \beta f_i(\pi_2), \quad (3.5)$$

for all $\alpha, \beta \in \mathbb{R}_0^+$ and all $\pi_1, \pi_2 \in \mathbb{R}^n$, with $\alpha + \beta = 1$.

- *Nonlinear programming*: This class of optimization problems includes all the problems that are neither linear nor least-squares but are not known to be convex.

For the two important and well known classes of least-squares problems and linear programs, there are effective and reliable algorithms that can solve even large problems, with thousands of variables and constraints. To prove that an optimization problem is convex can be really complex. Moreover, there are no general analytical formulas for the solution of convex optimization; however, there are effective methods like the *interior-point* methods that in some cases can provably solve the problem to a specified accuracy. For a nonlinear programming problem however, there is no effective solution method and with number of variables growing, the optimization problem can get intractable.

Convex optimization can still have an important role in solving an optimization problem that is not known to be convex. The trick is to modify the original problem to find an approximate but convex version of its formulation. The approximate convex problem can then be solved and the solution, which essentially sub-optimal compared to the original problem, can be used as starting point to a local optimization approach on the original problem to improve the solution¹.

DP is a well-known technique presented in early 1950s [69] that is commonly used for optimally solving complex problems. In DP, complex problems are broken down into simpler ones, which reduces the complexity dramatically. It is both a mathematical optimization method and a computer programming method which deals with the situations where decisions are made [23].

In the literature, DP is usually introduced using case studies. Therefore, we use our particular problem to present the DP idea and we will then use it in solving our problem in the following chapters.

In [24] the Markov characteristics of a general ARQ protocol are introduced. A dynamic approach to programming the control parameters in ARQ is formulated and the optimization of the protocol for transmission redundancies is presented. It is shown that, using the Bellman equation, an optimal rate allocation control can be searched which optimizes cost-to-go for infinite attempts. The idea presented in [24] is used in the research works in [58], [59] and [61].

The research presented in this dissertation is also based on the Markov representation of ARQ presented in [24]. However, since we are using truncated HARQ in our problem description, it is our interest to optimize the throughput for a limited number of transmission

¹The new solution is still a lower bound to the optimal solution however, it has the potential to improve the solution found by the approximate convex problem.

attempts for different K .

As we will see in the following chapters, in rate allocation we basically ignore the MDP model for our problem and instead a closed-form of the desired criterion (throughput) is subjected to optimization using recursive optimization with respect to allocated rates. However, in rate *adaptation*, our MDP model is the same as in [24] and we use the Bellman equation to find the upper bound to the achievable throughput.

The basic model for DP problem has the main features of a discrete time dynamic system and a cost function which is additive over time. The *state* parameter expresses the evolution of the dynamic system and is a function of the previous state, the decision variable and the random parameter which affects system. The state of the system I can be written as

$$I_k = g_k(I_{k-1}, \rho_k, C_k), \quad (3.6)$$

with ρ_k being the decision variable, C_k denoting a random variable and g_k being a function describing the system.

We assume I_0 as the state of the system at time zero while decision on ρ_1 is needed and C_1 affects the system so that altogether it gets to I_1 .

The additive cost of the system is a summation over the cost at all the steps or

$$\text{Cost}_{\text{tot}} = \sum_{k=1}^{K-1} \text{Cost}_k(I_{k-1}, \rho_k, C_k) + \text{Cost}_K(I_{K-1}), \quad (3.7)$$

while in our problem we simply have $\text{Cost}_k = \rho_k(I_{k-1})$, since it simply represent the number of channel uses per bit. The goal of the optimization is to minimize the whole cost with choosing the optimal control policy. This can be shown as

$$J_1(I_0) = \min_{\rho_1, \dots, \rho_K} \mathbb{E} \left\{ \text{Cost}_{\text{tot}} \right\}, \quad (3.8)$$

assuming that the system is at state I_0 at the start of the time.

The main idea of DP technique is the simple principle of optimality [23]. According to this principle, having an optimal policy like $\hat{\pi} = \{\hat{\rho}_1, \dots, \hat{\rho}_K\}$ for (3.8), and assuming that by using $\hat{\pi}$ a given state I_i happens at time i with positive probability, we have the

truncated $\{\hat{\rho}_{i+1}, \dots, \hat{\rho}_K\}$ as the optimal policy for the subproblem

$$\min_{\rho_{i+1}, \dots, \rho_K} \mathbb{E} \left\{ \sum_{k=i+1}^K \text{Cost}_k \right\} \quad (3.9)$$

if the subproblem is in state I_i at time i .

Based on the simple principle above, the optimization of the total cost can be formed into a recursive representation as follows:

$$J_1(I_0) = \min_{\rho_1} \mathbb{E}_{C_1} \{ \text{Cost}_1(I_0, \rho_1, C_1) + J_2(g_1(I_0, \rho_1, C_1)) \} \quad (3.10a)$$

$$J_2(I_1) = \min_{\rho_2} \mathbb{E}_{C_2} \{ \text{Cost}_2(I_1, \rho_2, C_2) + J_3(g_2(I_1, \rho_2, C_2)) \} \quad (3.10b)$$

...

$$J_K(I_{K-1}) = \min_{\rho_K} \mathbb{E}_{C_K} \{ \text{Cost}_K(I_{K-1}) \} \quad (3.10c)$$

The optimization process then starts by solving the one dimensional minimization of (3.10c) and continues up to (3.10a). The result of the optimization will be the optimal policy $\hat{\pi}$ and $J_1(I_0)$ will give the expected total cost for this optimal policy [23].

Our Approach

We prove in the next chapter (Section 4.1), that the optimal throughput-maximizing rate policy can be found by solving the recursive optimization system in (3.10), considering the final cost to be equal to the minimum number of average channel uses per bit delivery (i.e., inverse of maximum throughput). In other words, setting $\text{Cost}_K(I_{K-1}) = \frac{1}{\eta} \cdot \mathbb{I}(I_K)$ with $\mathbb{I}(\cdot)$ being the indication function, the DP recursive system presented above will give us the optimal rate policy.

Chapter 4

Single-Hop HARQ Transmission Policies

IN this chapter we study rate optimization in the HARQ channel model described in Section 2.1.1 for fixed transmit power. The two classes of feedback signal of HARQ channel introduced in Section 2 are analyzed separately. The function ρ is constant in the case of rate allocation, while, in the case of rate adaptation, it is a function of feedback message, i.e. the NACMI parameter I_k . Thus, optimization of the two cases go through different approaches. Rate adaptation HARQ, as discussed earlier, is a MDP problem and DP can be used to find optimal rate control policies. On the other hand, as we will see rate allocation is a multi-dimensional optimization problem which needs some simplifications to make the recursive optimization feasible.

4.1 Throughput Optimization

In this section we present our approach for solving throughput maximization problems in (2.16) and (2.17). The framework will be used in rate optimization of both single-hop channel in this chapter and the relay channel in Chapter 5.

In general, the problem of maximizing throughput in (2.16) can be represented as follows:

$$\hat{\eta} = \max_{\pi} \eta(\pi). \quad (4.1)$$

We define $\hat{\mathcal{R}}$ as the set of all the solutions to (4.1), and $\check{\mathcal{R}}_{\epsilon}$ as the set of all policies π that

have an outage probability of $P_{\text{out}}(\pi) = \epsilon$.

$$\hat{\mathbf{R}} = \{\pi | \eta(\pi) = \hat{\eta}\} \quad (4.2)$$

$$\check{\mathbf{R}}_\epsilon = \{\pi | P_{\text{out}}(\pi) = \epsilon\} \quad (4.3)$$

Denoting the expected number of channel uses $\bar{N}_s$ from (2.15) by $\mathbf{D}(\pi)$, the problem in (4.1) can be reformulated as follows,

$$\begin{aligned} \hat{\eta} &= \max_{\pi} \frac{1 - P_{\text{out}}(\pi)}{\mathbf{D}(\pi)} \\ &= \max_{\epsilon} \max_{\pi \in \check{\mathbf{R}}_\epsilon} \frac{1 - \epsilon}{\mathbf{D}(\pi)} \\ &= \max_{\epsilon} \frac{1 - \epsilon}{\min_{\pi \in \check{\mathbf{R}}_\epsilon} \mathbf{D}(\pi)} \\ &= \max_{\epsilon} \frac{1 - \epsilon}{\mathbf{D}_\epsilon}, \end{aligned} \quad (4.4)$$

which leaves us with solving the following problem.

$$\begin{aligned} \mathbf{D}_\epsilon &= \min_{\pi \in \check{\mathbf{R}}_\epsilon} \mathbf{D}(\pi) \\ &= \min_{\pi} \mathbf{D}(\pi), \quad \text{s.t.} \quad P_{\text{out}}(\pi) = \epsilon. \end{aligned} \quad (4.5)$$

We formulate the Lagrangian dual problem of (4.5) using a Lagrangian multiplier λ as

$$\mathbf{g}(\lambda, \epsilon) = \min_{\pi} \mathbf{D}(\pi) + \lambda \cdot (P_{\text{out}}(\pi) - \epsilon), \quad (4.6)$$

where due to the non-negative duality gap of Lagrangian, leads to

$$\mathbf{g}(\lambda, \epsilon) \leq \mathbf{D}_\epsilon. \quad (4.7)$$

The policy π that minimizes $\mathbf{g}(\lambda, \epsilon)$, does not depend on ϵ for a given λ thus, for a given λ an equivalent set of solutions of (4.6) can be made by solving the following minimization.

$$J^\lambda = \min_{\pi} \mathbf{D}(\pi) + \lambda \cdot P_{\text{out}}(\pi). \quad (4.8)$$

We denote by $\tilde{R}_\lambda$ the set of all solutions to this minimization problem as follows.

$$\tilde{R}_\lambda = \{\pi | \pi = \arg J^\lambda\}. \quad (4.9)$$

Therefore, $\tilde{R}_\lambda$ is equivalent to the set of all solutions to (4.6) for λ given.

Following the definition in [51] we call a policy π *degenerate* if it leads to zero redundancy transmission (and as a result $P_{\text{out}}(\pi) = 1$). In other words a policy is degenerate if and only if $\mathbf{D}(\pi) = 0$ which happens if and only if $\rho_k^s(S_k = 0) = 0, \forall k$. We also call a policy π *non-degenerate* if it is not degenerate.

One obvious conclusion that follows from (4.8) is that,

$$\mathbf{D}(\pi) + \lambda \cdot P_{\text{out}}(\pi) \leq \lambda, \quad \forall \pi \in \tilde{R}_\lambda \quad \text{and} \quad \forall \lambda > 0, \quad (4.10)$$

which can be proved by putting $\rho_k^s(S_k = 0) = 0, \forall k$.

The following propositions, that are proved in Appendix A, state the degenerate characteristics of $\tilde{R}_\lambda$ for different values of λ .

Proposition 1. (Degenerate Policies) *For all $\lambda < 1/\hat{\eta}$, $\tilde{R}_\lambda$ is a set of degenerate policies.*

Proposition 2. (Non-Degenerate Policies) *For all $\lambda > 1/\hat{\eta}$, all the policies in the set $\tilde{R}_\lambda$ are non-degenerate policies.*

We denote by $\lambda_{\text{th}} = 1/\hat{\eta}$ the threshold that separates the degenerate and non-degenerate policies to (4.8).

Proposition 3. (Separating Point) *The set $\tilde{R}_{\lambda_{\text{th}}}$ contains both degenerate and non-degenerate policies.*

The following theorem, which is proved in Appendix A, asserts the characteristics of the set of solution for (4.8) for given $\lambda = \lambda_{\text{th}}$.

Theorem 1. (Maximum Achievable Throughput) *The throughput (2.15) is maximized by any of the policies π in the non-degenerate subset of $\tilde{R}_{\lambda_{\text{th}}}$ (denoted by $\tilde{R}_{\lambda_{\text{th}}}^{\text{non-deg}}$).*

As the result of Theorem 1, the maximization problem (4.1) is equivalent to finding λ_{th} for (4.8), which is the smallest value of λ where a non-degenerate solution for J^λ can be found. Note that this solution is found by solving (4.8) instead of (4.6), because they have

the same sets of solutions for a given λ . As a result, the smallest value of λ that gives us a non-degenerate policy, can be found without involving ϵ parameter in the optimization process. However, the policy π , found for λ_{th} from (4.8), is the throughput maximizing policy.

The approach to solve (2.16), based on Theorem 1, is to solve (4.8) for a number of λ values and then apply any root-finding method (e.g., *bisection method*) until the λ_{th} value is derived.

To solve the constrained optimization in (2.17), we use the same approach as described above noting that there is a one-to-one correspondence between outage probability values ϵ and the λ values in (4.8). To prove this statement, note that the outage probability $P_{out}(\pi)$ for $\pi \in \tilde{R}_\lambda$ decreases monotonically with increasing λ . This is because for any $\lambda \geq \lambda_{th}$ there is one and only one ϵ where $\tilde{R}_\lambda \subset \tilde{R}_\epsilon$.

We may equivalently prove that $\tilde{R}_\lambda = \mathring{R}_{D_\epsilon}$, where

$$\mathring{R}_{D_\epsilon} = \{\pi | \pi = \arg D_\epsilon\}. \quad (4.11)$$

We denote by $\hat{\epsilon}$ the outage for the found $\hat{\pi} \in \hat{R}$. In the following two propositions (proved in Appendix A) we assert that $\tilde{R}_\lambda = \mathring{R}_{D_\epsilon}$ for any ϵ where $\epsilon \in (0, \hat{\epsilon}]$.

Proposition 4. *For a given λ , if there is at least one policy π where $\pi \in \tilde{R}_\lambda$, and $P_{out}(\pi) = \epsilon$, then $\mathring{R}_{D_\epsilon} \subset \tilde{R}_\lambda$.*

Proposition 5. *For an arbitrary λ where $\lambda \geq \lambda_{th}$, there is one and only one $\epsilon \in (0, \hat{\epsilon}]$ where $\tilde{R}_\lambda \subset \mathring{R}_{D_\epsilon}$.*

Finally, for an arbitrary outage probability $\epsilon \leq \hat{\epsilon}$, the optimal outage constrained throughput is the solution $\eta(\pi)$ of (4.8) that gives out $P_{out}(\pi) = \epsilon$. Also, for $\epsilon > \hat{\epsilon}$ the optimal outage constrained throughput is $\hat{\eta}$.

$$\hat{\eta}_\epsilon = \begin{cases} \hat{\eta}, & \text{if } \hat{\epsilon} < \epsilon \\ \frac{1-\epsilon}{D_\epsilon}, & \text{o.w.} \end{cases}. \quad (4.12)$$

4.2 Optimal Rate Adaptation

We start from (2.14) where the number of channel uses in the k th transmission is the random variable $N_{s,k}$.

$$\overline{N}_{s,k} = \mathbb{E}\{N_{s,k}\} = \mathbb{E}\{N_b \cdot \rho_k\}, \quad (4.13)$$

with $\mathbb{E}$ being the expectation over all channel gain variables (this will be the case throughout this dissertation).

This gives the representation for throughput of adaptive rate single-hop HARQ channel $\eta_{\text{AD-S}}$, with the help of (2.15), as follows:

$$\eta_{\text{AD-S}} = \frac{1 - P_{\text{out}}}{\mathbb{E}\{\sum_{k=1}^K \rho_k\}}. \quad (4.14)$$

From (2.11) we have

$$P_{\text{out}} = \Pr\{I_K < 1\} = \mathbb{E}\{\mathbb{I}(I_K < 1)\} \quad (4.15)$$

where $\mathbb{I}(x)$ is equal one if x is true and is equal to zero otherwise.

In order to find the maximum throughput,

$$\hat{\eta}_{\text{AD-S}} = \max_{\rho_1, \dots, \rho_K} \eta_{\text{AD-S}}, \quad (4.16)$$

we use the Lagrangian form of it by substituting (4.13) in (4.8) to form the recursive equations for any particular value of λ as follows.

$$J_1^\lambda(I_0) = \min_{\rho_1, \dots, \rho_K} \mathbb{E}\left\{\sum_{k=1}^K \rho_k\right\} + \lambda \cdot P_{\text{out}}, \quad (4.17)$$

where the expectation is over all channel gains similar to (4.15) which gives us

$$J_1^\lambda(I_0) = \min_{\rho_1, \dots, \rho_K} \mathbb{E}\left\{\sum_{k=1}^K \rho_k + \lambda \cdot \mathbb{I}(I_K < 1)\right\}, \quad (4.18)$$

and in a recursive form it appears as

$$J_1^\lambda(I_0) = \min_{\rho_1} \mathbb{E}\{\rho_1 + J_2^\lambda(I_0 + \rho_1 \cdot C_1)\} \quad (4.19a)$$

$$J_2^\lambda(I_1) = \min_{\rho_2} \mathbb{E}\{\rho_2 + J_3^\lambda(I_1 + \rho_2 \cdot C_2)\} \quad (4.19b)$$

...

$$J_K^\lambda(I_{K-1}) = \min_{\rho_K} \mathbb{E}\{\rho_K + \lambda \cdot \mathbb{I}(I_{K-1} + \rho_K \cdot C_K < 1)\} \quad (4.19c)$$

where $I_0 \equiv 0$ and is the zero state at time zero.

The optimization starts from (4.19c) and goes backward to (4.19a). In each step, the state I has to be discretized to L points over the interval $[0, 1)$. Then, for a given I_{k-1} value we can optimize the value of the function $\rho_k(I_{k-1})$ provided the function $J_{k+1}^\lambda(I_k)$ is known. So, as discussed in Section 3.5, the global optimization of the possibly non-convex function $J(\lambda)$ over the space of L^K values is reduced to $K \cdot L$ one-dimensional optimizations. The optimization starts with solving (4.19c) which can be written as

$$J_K^\lambda(I_{K-1}) = \min_{\rho_K} \left\{ \rho_K + \lambda \cdot F_C\left(\frac{1-I_{K-1}}{\rho_K}\right) \right\}. \quad (4.20)$$

In (4.20), $F_C(\cdot)$ is the *cumulative density function* (cdf) of the random variable C . With the first and second derivative test we can easily see that only one minimizing solution of ρ_K exists for (4.20) (We leave this analysis to the reader, however in Section 5.1.2 we will go into the details of it in the more general case of multi-hop channel.). After finding $\rho_K(I_{K-1})$ the recursive optimization continues with $J_{K-1}^\lambda(I_{K-2})$ and goes on up to (4.19a).

This process is assumed for a given value of λ and for the optimal policy π derived from this DP process, we have $\pi \in \tilde{\mathbf{R}}_\lambda$. In case of π being a degenerate solution, then we know that $\lambda < 1/\hat{\eta}$, otherwise, $\lambda \geq 1/\hat{\eta}$. As we discussed earlier, following a root-finding method and repeating these steps, λ_{th} can be found.

In Figure 4.1 the maximum achieving policy for $K = 4$ in average SNR of 10 dB is shown. At the start of the HARQ process a redundancy of ρ_1 would be the best start. However, based on the feedback provided at the end of the first transmission, a zero redundancy transmission (*packet dropping*) for the following attempts may be decided if the normalized information accumulated at the decoder is below a certain threshold (as it appears in Figure 4.1 to be around 0.06, 0.2 and 0.41 for the second, third and fourth transmission

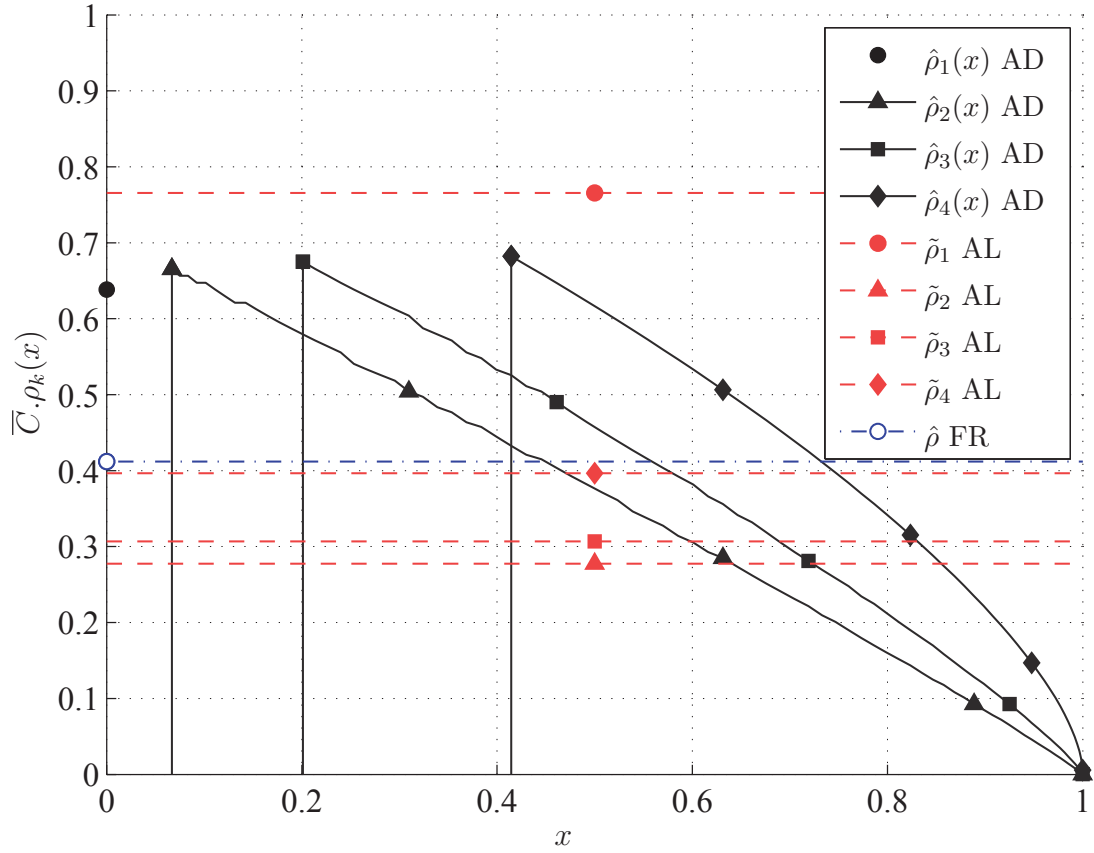


Figure 4.1 Throughput maximizing redundancy policies for three different rate control approaches compared together. The channel exhibits Rayleigh-fading with $\bar{\gamma} = 10$ dB and a truncated HARQ process with $K = 4$ and $L = 100$ is assumed. Redundancy values are multiplied by the ergodic capacity $\bar{C}$ of the same channel. The optimal policies shown are: $\rho_k(x)$ (solid lines) for rate adaptation (AD), ρ_k for rate allocation (AL) and fixed rate (FR) (dashed lines).

respectively.). This shows us that an optimal throughput HARQ may drop the packet at some level if the previous attempts of the same packet have been *disappointing*.

4.3 Optimal Rate Allocation

Rate allocation HARQ is associated with the single-bit feedback case where the signal over feedback channel is only capable of informing the transmitter whether the decoding has

been successful or not. Here, we start with the failure probability of the k th transmission for $1 \leq k \leq K$. We define f_k as the probability of failure at k th decoding. Therefore, from (2.11) we have

$$\begin{aligned} P_1^{SD} &= \Pr\{I_k < 1\} \\ &= \Pr\left\{\sum_{l=1}^k C_l \cdot \rho_l < 1\right\} \\ &= \int_0^1 p_{I_k}(x) dx \end{aligned} \quad (4.21)$$

where $P_K^{SD} = P_{\text{out}}$ and $p_{I_k}(x)$ is the pdf of the random variable $I_k = \sum_{l=1}^k C_l \cdot \rho_l$. Since this random variable is a sum of random variables $\nu_l = C_l \cdot \rho_l$, each with pdf of $p_{\nu_l}(x)$, we have

$$p_{I_k}(x) = (p_{\nu_1} * \dots * p_{\nu_k})(x) \quad (4.22)$$

with $*$ being the convolution operator. On the other hand, in (2.14) we have

$$\begin{aligned} \overline{N}_s &= \sum_{k=1}^K \mathbb{E}\{N_{s,k}\} \\ &= \sum_{k=1}^K N_{s,k} \cdot P_{k-1}^{SD} \\ &= N_b \cdot \sum_{k=1}^K \rho_k \cdot P_{k-1}^{SD} \end{aligned} \quad (4.23)$$

with $P_0^{SD} = 1$ by definition.

Then according to (2.15) throughput for the single-hop HARQ rate allocation $\eta_{\text{AL-S}}$ is

$$\eta_{\text{AL-S}} = \frac{1 - P_K^{SD}}{\sum_{k=1}^K \rho_k \cdot P_{k-1}^{SD}} \quad (4.24)$$

and casted in the Lagrangian form of (4.8), the optimization problem is the following:

$$\min_{\rho_1, \dots, \rho_K} \left\{ \sum_{k=1}^K \rho_k \cdot P_{k-1}^{SD} + \lambda \cdot P_K^{SD} \right\}. \quad (4.25)$$

However, DP recursive representation of Section 3.5 is not feasible for (4.25), since P_k^{SD} depends on all of the optimization parameters ρ_i with i from 1 up to k . However, for it to be cast into DP we need it to be only dependent on a state parameter and the current optimization parameter at each step. This observation leads us to the conclusion that the rate allocation problem cannot be solved exactly via DP method and approximations are necessary.

Here we use the Gaussian approximation introduced in several papers like [12], to approximate computation of (4.22). From the *central limit theorem*, a summation over n i.i.d. random variables is a random variable with *normal* distribution if n goes to infinity. This has been proved also for a weighted sum of i.i.d. random variables in [70]. Therefore, we can use the Gaussian approximation proposed in [15] to approximate $p_{V_k}(x)$ in (4.22) since it is a summation of random variables $v_k = \rho_l \cdot C_l$. As a result, P_k^{SD} can be approximated as shown below.

$$\begin{aligned} P_k^{SD} &= \Pr\left\{\sum_{l=1}^k C_l \cdot \rho_l < 1\right\} \\ &\approx Q\left(\frac{\bar{C} \cdot X_k - 1}{\sigma_C \cdot \sqrt{Y_k}}\right) \end{aligned} \quad (4.26)$$

where $\bar{C} = \mathbb{E}_C\{C\}$ and $\sigma_C^2 = \mathbb{E}_C\{C^2\} - \bar{C}^2$. Also, $X_k = \sum_{l=1}^k \rho_l$, $Y_k = \sum_{l=1}^k \rho_l^2$ and $Q(x)$ is the *Q-function* defined as

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp\left(-\frac{\tau^2}{2}\right) d\tau. \quad (4.27)$$

Using this approximation we substitute P_k^{SD} in (4.25) with $\tilde{P}_k^{SD}$, where

$$\tilde{P}_k^{SD} = \begin{cases} F_C\left(\frac{1}{\rho_k}\right), & k = 1 \\ Q\left(\frac{\bar{C} \cdot X_k - 1}{\sigma_C \cdot \sqrt{Y_k}}\right), & \text{otherwise.} \end{cases} \quad (4.28)$$

$\tilde{P}_k^{SD}$ is a function of X_k and Y_k so, we use the two dimensional real valued vector (X_k, Y_k) as the state of a new optimization formulation. Multiplying (4.25) with $\bar{C}$ and using the

approximations we have

$$J_1^\lambda(X_0, Y_0) = \min_{\rho_1, \dots, \rho_K} \left\{ \rho_1 + \rho_2 \cdot \tilde{P}_1^{SD} + \sum_{k=3}^K \rho_k \cdot \tilde{P}_{k-1}^{SD} + \lambda \cdot \tilde{P}_K^{SD} \right\} \quad (4.29)$$

where we can use the relation between successive states as $X_{k+1} = X_k + \rho_k$ and $Y_{k+1} = Y_k + \rho_k^2$.

The formulation in (4.29) can be solved recursively in K steps and reduce the complexity of a K -dimensional optimization into K one-dimensional subproblems. Furthermore, we use the approximation of P_k^{SD} by $\tilde{P}_k^{SD}$ only for $k \geq 2$ and $\tilde{P}_1^{SD} = P_1^{SD}$ is calculated in exact form without harming the recursive optimization.

In a recursive form (4.29) appears as follows:

$$J_1^\lambda(X_0, Y_0) = \min_{\rho_1} \left\{ \rho_1 + J_2^\lambda(X_0 + \rho_1, \sqrt{Y_0^2 + \rho_1^2}) \right\} \quad (4.30a)$$

$$J_2^\lambda(X_1, Y_1) = \min_{\rho_2} \left\{ \rho_2 \cdot \tilde{P}_1^{SD} + J_3^\lambda(X_1 + \rho_2, \sqrt{Y_1^2 + \rho_2^2}) \right\} \quad (4.30b)$$

...

$$J_K^\lambda(X_{K-1}, Y_{K-1}) = \min_{\rho_K} \left\{ \rho_K \cdot \tilde{P}_{K-1}^{SD} + \lambda \cdot \tilde{P}_K^{SD}(X_{K-1} + \rho_K, \sqrt{Y_{K-1}^2 + \rho_K^2}) \right\} \quad (4.30c)$$

The recursive optimization process starts with (4.30c) and finds the optimal two dimensional function $\rho_K(x, y)$ for all the state space (X_K, Y_K) (discretized into L^2 points) and goes on backward until it finds the optimal function $\rho_1(x, y)$. Then, setting $(X_0, Y_0) \equiv (0, 0)$ we find $\rho_1(0, 0)$ as the optimal allocated first transmission redundancy ρ_1 for the given particular λ value. Using ρ_1 we then search for $\rho_2(x, y)$ at point $(0 + \rho_1, 0 + (\rho_1)^2)$ to find optimal ρ_2 and continue until all the optimal allocatee redundancies for the given λ are found. The heuristic limit for the two dimensional state space is from $(\bar{C} \cdot K, \bar{C} \cdot K)$ since we don't expect ρ to be greater than $1/\bar{C}$. The optimal π will be used in the exact formulation of throughput and the resulting $P_{\text{out}}(\pi)$ and $\eta(\pi)$ are linked to the given λ in a table.

This process repeats for all the given λ values and among all the results we are able to look for maximum throughput $\hat{\eta}_{\text{AL-S}}$ as well as finding the optimal outage-constrained throughput $\hat{\eta}_{\text{AL-S}, \epsilon}$. The optimal π for $K = 4$ and an average channel SNR of 10 dB is shown in Figure 4.1. We show them as constant functions with respect to the normalized accumulated information at decoder, because the rates are allocated before the transmission

process begins.

4.3.1 Simplified Optimization Method

The optimization process discussed above can be simplified by getting rid of the two-dimensional state and using a one dimensional state. This needs another level of approximation in (4.28) by letting $\sqrt{Y_k} \approx X_k$ as follows.

$$\check{P}_k^{SD}(X_k) = \begin{cases} F_C\left(\frac{1}{\rho_k}\right), & k = 1 \\ Q\left(\frac{\bar{C} \cdot X_{k-1}}{\sigma_C \cdot X_k}\right), & \text{otherwise.} \end{cases} \quad (4.31)$$

This approximation will reduce the complexity of the recursive minimization. However, as it is showed in [51], the throughput of a policy π in (4.24), calculated with the approximate outage probabilities $\check{P}_k^{SD}$, is a lower bound to the throughput calculated with the same policy π but using the outage probabilities $\tilde{P}_k^{SD}$ in (4.28).

$$\tilde{\eta}_{\text{AL-S}}(\pi) = \frac{1 - \tilde{P}_K^{SD}}{\sum_{k=1}^K \rho_k \cdot \tilde{P}_{k-1}^{SD}} \geq \check{\eta}_{\text{AL-S}}(\pi) = \frac{1 - \check{P}_K^{SD}}{\sum_{k=1}^K \rho_k \cdot \check{P}_{k-1}^{SD}} \quad (4.32)$$

As a result of (4.32), maximization of throughput using the new approximate outage probabilities in (4.31) will result in maximizing the lower bound to $\tilde{\eta}_{\text{AL-S}}$. Moreover, maximization of $\check{\eta}_{\text{AL-S}}$ does not need to follow the Lagrangian function in (4.25). Instead, this simplification allows us to solve the following optimization problem.

$$\max_{\pi} \check{\eta}_{\text{AL-S}}(\pi) = \max_X \frac{1 - \check{P}_K^{SD}(X)}{\mathbf{D}_K(X)} \quad (4.33)$$

where

$$\begin{aligned} \mathbf{D}_K(X) &= \min_{\substack{\rho_1, \dots, \rho_k \\ \sum_{l=1}^k \rho_l = X}} \rho_1 + \sum_{l=2}^k \rho_l \cdot \check{P}_{l-1}^{SD}(X_{l-1}) \\ &= \min_{0 \leq \rho_k \leq X} \min_{\substack{\rho_1, \dots, \rho_{k-1} \\ \sum_{l=1}^{k-1} \rho_l = X - \rho_k}} \rho_1 + \sum_{l=2}^{k-1} \rho_l \cdot \check{P}_{l-1}^{SD}(X_{l-1}) + \rho_k \cdot \check{P}_{k-1}^{SD}(X - \rho_k) \\ &= \min_{0 \leq \rho \leq X} \mathbf{D}_{k-1}(X - \rho) + \rho \cdot \check{P}_{k-1}^{SD}(X - \rho). \end{aligned} \quad (4.34)$$

The optimization results are stored as $\rho_k(X) = \arg \min_{\rho} \mathbf{D}_{k-1}(X - \rho) + \rho \cdot \check{P}_{k-1}^{SD}(X - \rho)$. Once the functions $\mathbf{D}_k(X)$ are obtained, we can recover the solution $\hat{\rho}_k$ as follows letting $\hat{X}_K = \arg_X \max \frac{1 - \check{P}_K^{SD}(X)}{\mathbf{D}_K(X)}$.

1. $\hat{\rho}_K = \rho_K(\hat{X}_K)$
2. for $k : K - 1 \rightarrow 2$
 - $\hat{X}_k = \hat{X}_{k+1} - \hat{\rho}_{k+1}$
 - $\hat{\rho}_k = \rho_k(\hat{X}_k)$
3. $\hat{\rho}_1 = \hat{X}_2 - \hat{\rho}_2$.

The results obtained through the simplified formulation were very close to those obtained using the two dimensional state space optimization. This reduces the complexity at every step of the recursive optimization, by the number of discretization points over the state parameter. In Chapter 5 we will discuss the complexity of optimization for rate allocation and compare it to the alternative approaches.

4.4 Performance Limits

The obvious lower bound of one transmission $K = 1$ for both the single-hop channel model and the cooperative channel model (which will be studied in the next chapter) is the direct transmission from $\mathcal{S}$ to $\mathcal{D}$ (also known as lower bound for decode-and-forward in multi-hop) averaged on the channel state. In this case $P_{\text{out}}(\rho) = \Pr\{C(\gamma) \cdot \rho < 1\} = F_{C(\gamma)}(\frac{1}{\rho})$ and lower bound for maximum achievable throughput is derived as

$$\hat{\eta}_0 = \max_{\rho} \left\{ \frac{1 - P_{\text{out}}(\rho)}{\rho} \right\} \quad (4.35)$$

An upper bound of the single-hop model would happen when an infinite number of transmission attempts is allowed ($K \rightarrow \infty$). For an infinite number of transmissions $P_{\text{out}} \rightarrow 0$. As a result, maximization of throughput is the same as minimization of the denominator in (2.15). The steady average total redundancy function for infinite transmissions is the expectation on the amount of redundancy (number of channel uses divided by N_b) in infinite

time and we denote it by J_∞ .

$$\begin{aligned}
 J_\infty &= \lim_{K \rightarrow \infty} \frac{\overline{N}_s}{N_b} \\
 &= \lim_{K \rightarrow \infty} \frac{\sum_{k=1}^K \mathbb{E}\{N_{s,k}\}}{N_b} \\
 &= \lim_{K \rightarrow \infty} \sum_{k=1}^K \mathbb{E}\{\rho_k\}
 \end{aligned} \tag{4.36}$$

According to Bellman's theorem [23, Chap. 3] we can find the minimum of J_∞ as follows.

$$\hat{J}_\infty(S) = \min_{\rho} \mathbb{E}\{\rho + \hat{J}_\infty(\acute{S})\} \tag{4.37}$$

where ρ is the redundancy transmitted in the infinite time by the active transmitter and S is the state of the process as we defined in Section 2.3 and $\acute{S} = f_S(S, \rho, C)$ is the new state which is a function of the previous state, redundancy ρ and the channel random parameter C . Therefore, the maximum achievable throughput can be calculated as

$$\hat{\eta}_\infty = \frac{1}{\hat{J}_\infty} \tag{4.38}$$

as soon as J_∞ for the desired scenario is known. It is shown in [1] that for fixed-power fixed-rate transmission, when the allowed number of transmission attempts grows, $K \rightarrow \infty$, the throughput of a well designed HARQ approaches the ergodic channel capacity given by [21, 42].

$$\overline{C} = \int_0^\infty c \cdot p_C(c) dc \tag{4.39}$$

A fixed-rate HARQ is obviously a particular case of variable rate HARQ however, since the channel is memoryless the block-wise feedback does not change the capacity of the channel [1] and as a result the achievable region for variable-rate (adaptation and allocation scheme) is limited by the ergodic capacity (4.39).

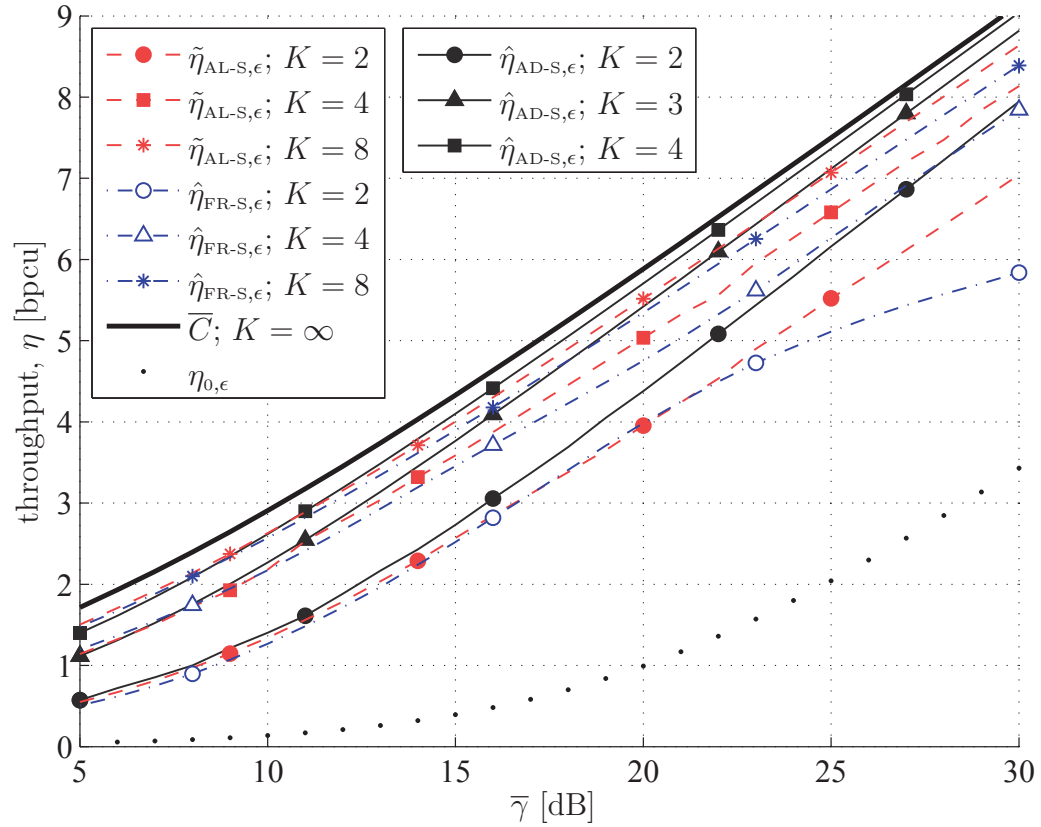


Figure 4.2 Optimal outage-constrained throughput for $\epsilon = 10^{-2}$. The solid black line shows ergodic capacity $\bar{C}$ and black dots show $\eta_{0,\epsilon}$ (or $K = 1$).

4.5 Numerical Results

In this section we present numerical examples for the two problems solved in Section 4.2 and Section 4.3. Results will be compared with the fixed-rate case ($\eta_{\text{FR-S}}$) which is a special case of rate allocation HARQ. The channel is assumed to be Rayleigh block fading and, as a result, SNR is characterized by the exponential probability distribution function

$$p_\gamma = \frac{1}{\bar{\gamma}} \exp\left(-\frac{\gamma}{\bar{\gamma}}\right), \quad (4.40)$$

with $\bar{\gamma}$ denoting the average SNR.

The results shown in this section for the optimal throughput, are compared to the maximum achievable throughput $\hat{\eta}_\infty$ and the throughput lower bound $\hat{\eta}_0$, that we described

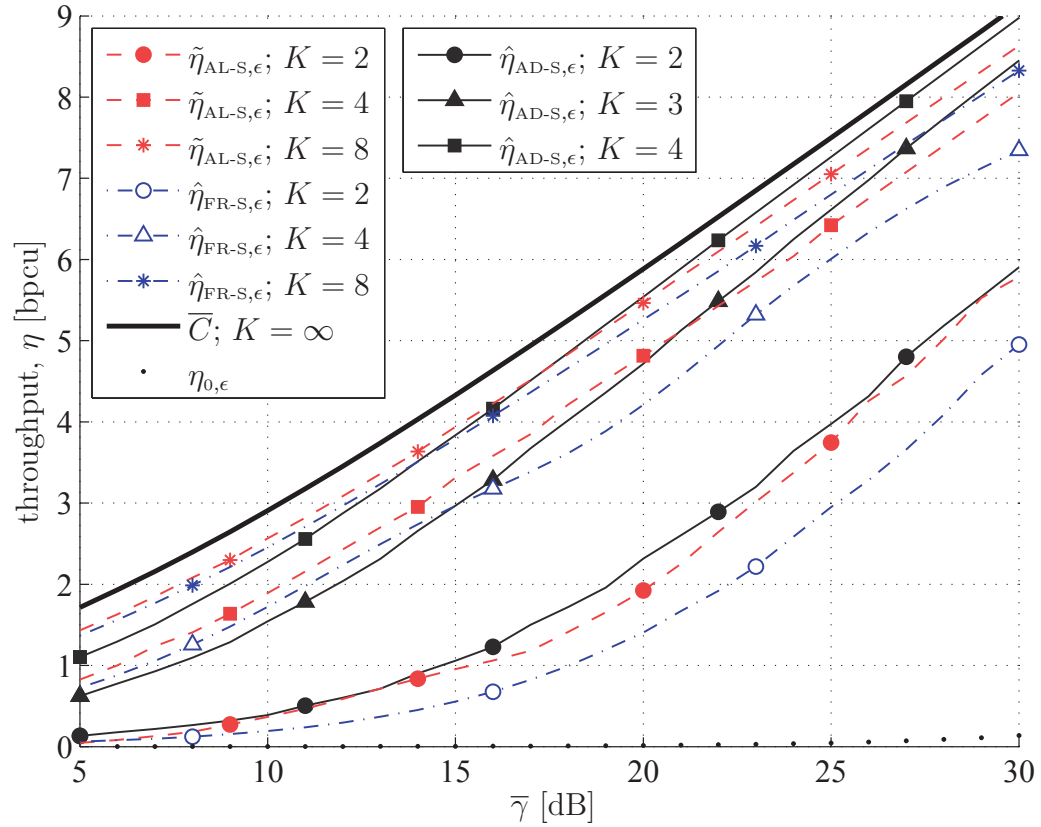


Figure 4.3 Optimal outage-constrained throughput for $\epsilon = 10^{-4}$. The solid black line shows ergodic capacity $\bar{C}$ and black dots show $\eta_{0,\epsilon}$ (or $K = 1$).

in Section 4.4. The maximum throughput is achievable when $K \rightarrow \infty$ is equal to the ergodic capacity $\bar{C}$. This is also confirmed in our numerical analysis by solving the Bellman's equation in (4.37). With K increasing, the number of channel uses (i.e., the redundancy value) in each transmission frame for a particular message decreases. As a result, the observation $\mathbf{Y}$ defined in Section 2.1.1 can be seen as the output of a fast-fading channel to a long codeword $\mathbf{X}$. From the capacity of a channel with state [71] we know that a coding rate bounded with the ergodic capacity can be delivered with an arbitrary small outage probability using an arbitrary long codeword.

First we show the results for outage constrained throughput optimization which is of much interest in trade-off analysis in system design [72]. It can be computed according to (4.12) and using the table of λ values associated with the corresponding η and P_{out} .

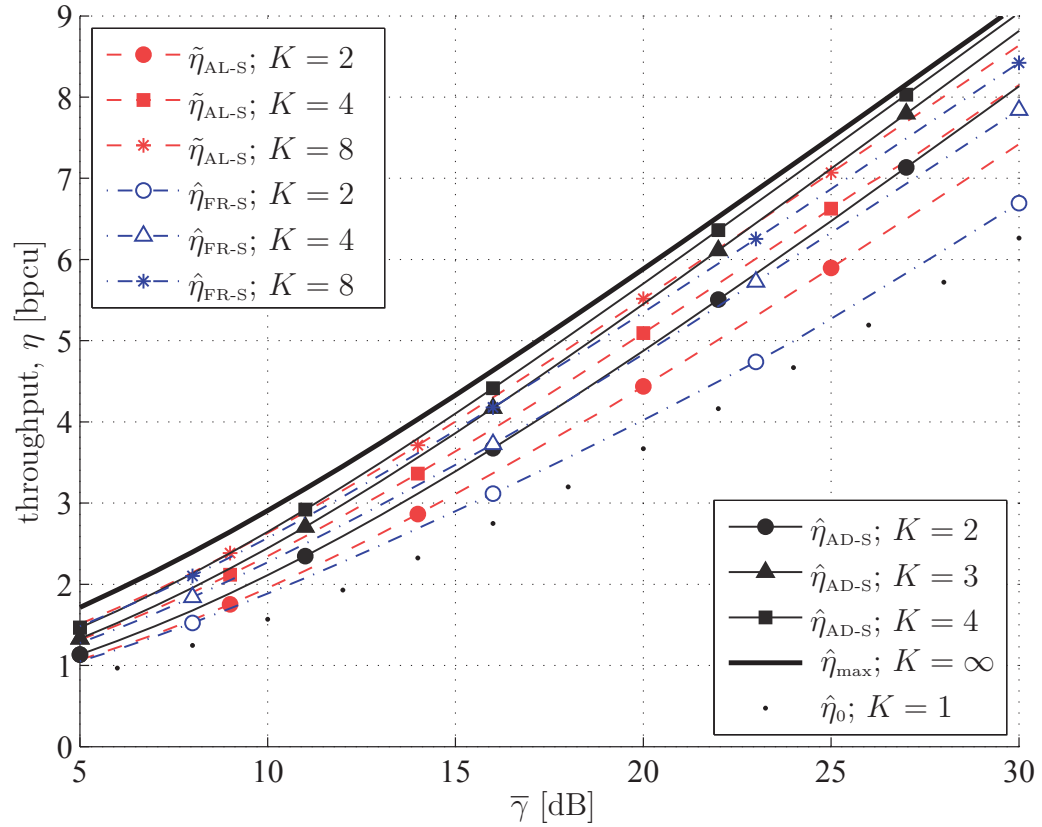


Figure 4.4 Optimal throughput for different scenarios of HARQ. The solid black line shows the ergodic capacity $\bar{C}$ and black dots show $\hat{\eta}_0$ (or $K = 1$).

The optimum outage constrained throughput $\hat{\eta}_\epsilon$, for two different values of $\epsilon = 10^{-2}$ and $\epsilon = 10^{-4}$, is shown respectively in Figure 4.2 and Figure 4.3 for the three cases of rate adaptation $\hat{\eta}_{\text{AD-S},\epsilon}$ for $K = 2, 3, 4$, rate allocation $\hat{\eta}_{\text{AL-S},\epsilon}$ and fixed-rate $\hat{\eta}_{\text{FR-S},\epsilon}$ for $K = 2, 4, 8$. We choose to compare $K = 4$ of rate adaptation with $K = 8$ of rate allocation and fixed-rate to show the advantage of the rate adaptation.

Optimal throughput and the corresponding outage probability for the three cases are also shown, respectively, in Figure 4.4 and Figure 4.5. As shown, a significant improvement on the throughput is achievable by adaptive incremental redundancy even if the transmitter does not know the instantaneous CSI. An upper bound for the maximum achievable throughput is shown too which is the ergodic capacity $\bar{C}$. The outage probability corresponding to a maximum throughput is shown in Figure 4.5. Even if the outage is varying, we assume that

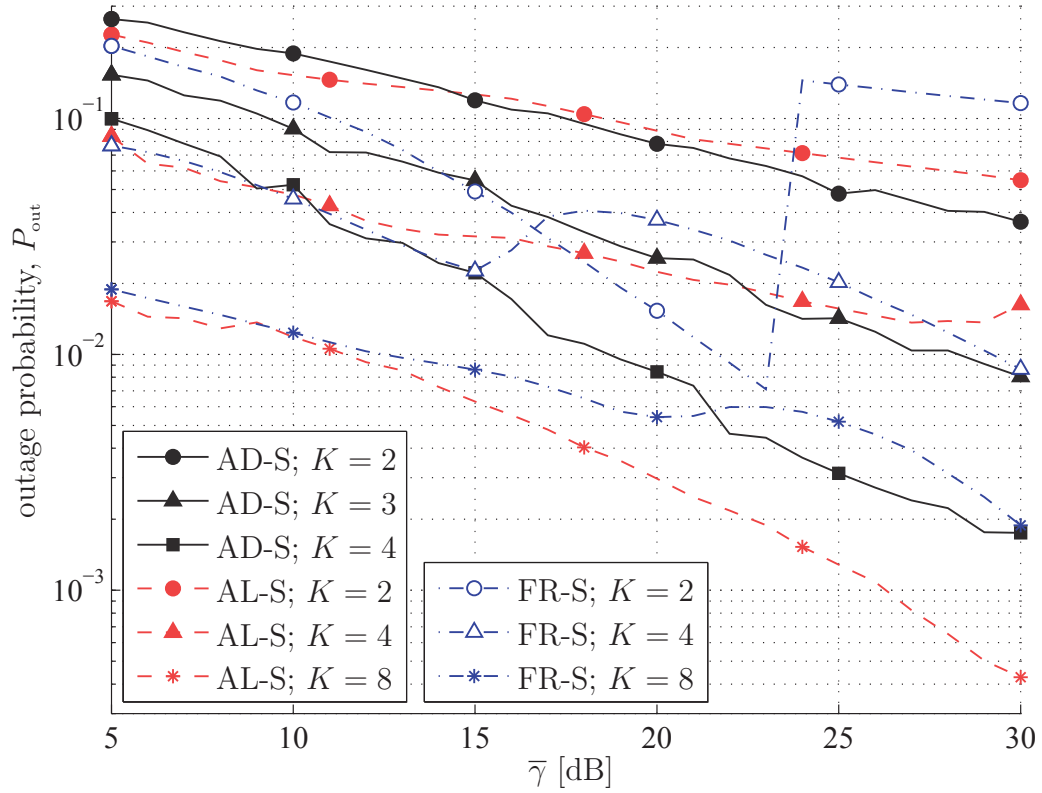


Figure 4.5 Outage probability of the optimal rate policies for different scenarios of HARQ.

the suitable level of reliability can be obtained via additional retransmission mechanisms, implemented at the upper layers. This idea is adopted in the LTE standard and the outage probability of the MAC layer HARQ transmission is controlled by an upper layer ARQ protocol embedded in the *Radio Link Control* (RLC) layer.

The outage probabilities of fixed-rate HARQ, shown in Figure 4.5, follow a non-smooth behavior which we explain in Figure 4.6 for $K = 2$. As shown in Figure 4.6, the throughput value $\eta_{\text{FR-S}}(\rho)$, as a function of ρ , has a global and a local maximum which switch places while increasing $\bar{\gamma}$ (from 23 to 24 dB). For $K = 2$ according to (4.24), throughput follows $\eta_{\text{FR-S}}(\rho) = \frac{1-P_2^{sd}}{\rho \cdot (1+P_1^{sd})}$, which for small values of ρ goes to $\frac{1-P_2^{sd}}{2\rho}$ and for large values of ρ it goes to $\frac{1}{\rho \cdot (1+P_1^{sd})}$ (respectively depicted by dashed and dotted lines in Figure 4.6). By increasing the SNR value, the argument $\hat{\rho}$ for which the global maximum (indicated with $\times$ in Figure 4.6) occurs, may change abruptly. This makes the outage probability (P_2^{sd}) to

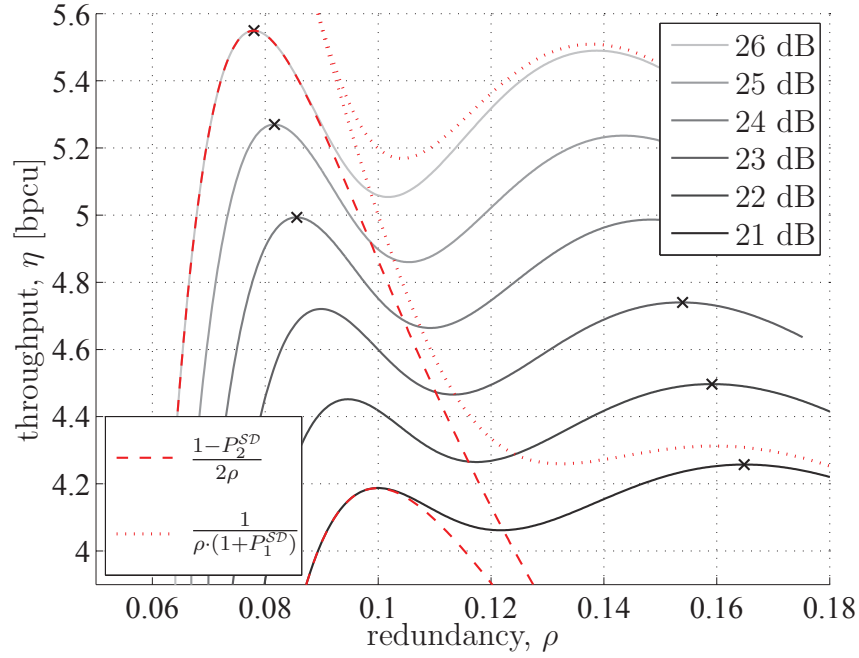


Figure 4.6 Throughput of the fixed-rate scenario with respect to redundancy value ρ for different $\bar{\gamma}$ values and $K = 2$. The $\times$ marker indicates the point where the globally optimum throughput is attained.

increase abruptly as shown in Figure 4.5. The same analysis can be followed for $K = 4$ and 8.

The policy is optimized for a given channel SNR for both rate allocation/adaptation cases as we saw earlier. As a result we expect the same optimal policy to have a different performance if the channel SNR varies. Figure 4.7 shows an example of this matter, where optimal policies for $\bar{\gamma} = 10$ dB and $\bar{\gamma} = 20$ dB are being used as transmission policy for a rang of different SNR values. Figure 4.8 shows the corresponding outage probability values.

4.6 Summary and Complementary Notes

In this chapter we discussed two main problems of rate adaptation and rate allocation IR-HARQ along with the fixed-rate transmission to show the effect of variable-rate transmission in a fixed-power scenario. Even though the channel is assumed to be block-fading and channel gains for different blocks are assumed independent, the information fed back from

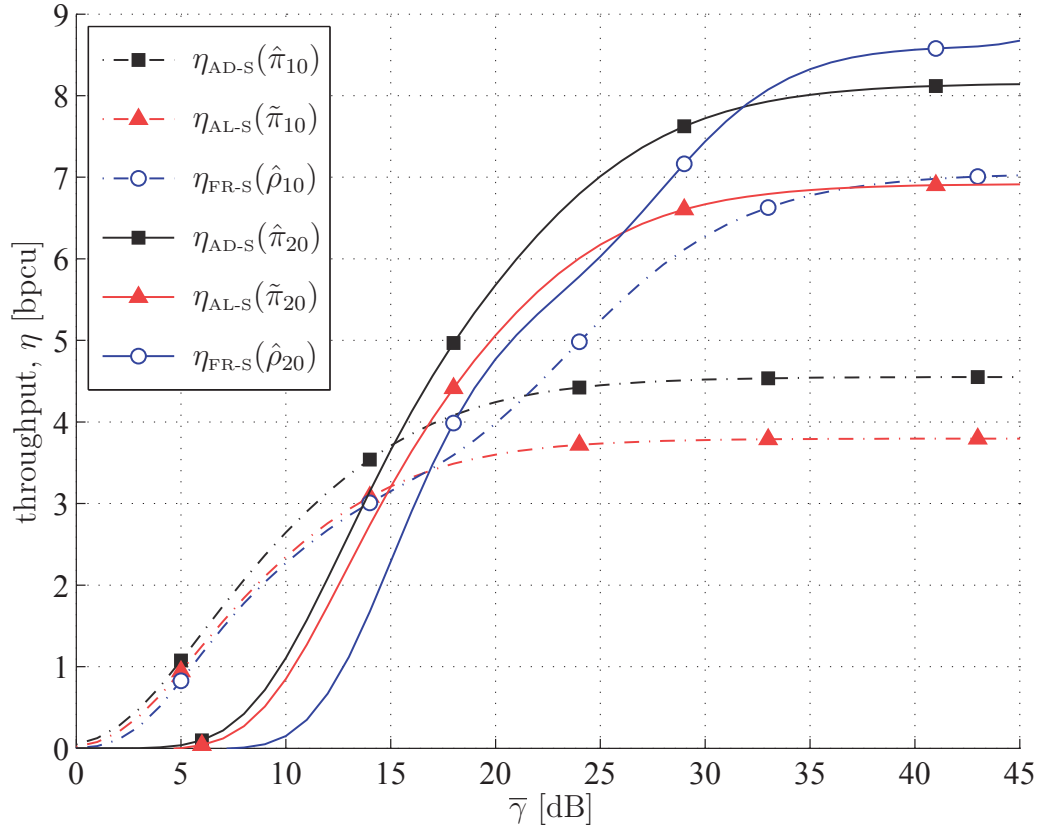


Figure 4.7 Throughput gained by using an optimal policy calculated for different average channel gain. The results are shown for the following policies: $\hat{\pi}_{\bar{\gamma}}$ as the optimal rate adaptation policy, $\tilde{\pi}_{\bar{\gamma}}$ as the optimal rate allocation policy and $\hat{\rho}_{\bar{\gamma}}$ as the optimal fixed-rate redundancy value. Optimal policies found for two average SNR values of $\bar{\gamma} = 10$ and $\bar{\gamma} = 20$ dB are used for transmission in different average SNR channels for $K = 4$.

receiver to transmitter is shown to be useful to increase the throughput.

The rate adaptation IR-HARQ performs very close to the ergodic capacity limit even with a small number of retransmissions (e.g. for truncated HARQ with $K = 4$). Outage constrained results show that very small values of outage probability (e.g. $P_{\text{out}} = 10^{-4}$) can be achieved with acceptable throughput values using optimized variable rate HARQ.

A multi-bit feedback for rate adaptation is considered to convey the actual value of I_k in (2.4) (i.e., the discretization of I_k is not explicitly considered). This may not seem practical as the information has to be conveyed using a limited number of bits. First, we

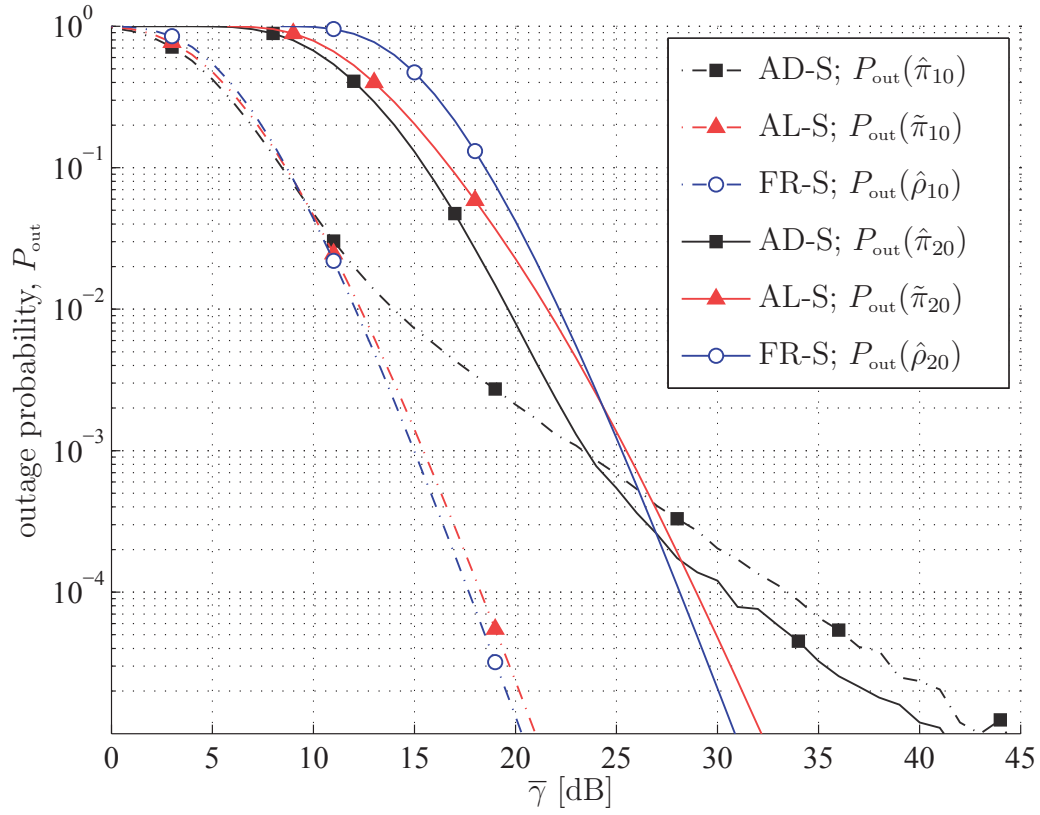


Figure 4.8 Outage probability corresponding to the use of optimal rate policies of $\hat{\pi}_{\bar{\gamma}}$, $\tilde{\pi}_{\bar{\gamma}}$ and $\hat{\rho}_{\bar{\gamma}}$, for $\bar{\gamma} = 10$ and $\bar{\gamma} = 20$ dB in different channel average SNR situation for $K = 4$.

need to note that all numerical calculations imply that the signal are discretized. Following this line of thought we would need to convey the number of bits necessary for floating point representation of the numbers, i.e., 64 bits. While it is possible to reformulate the problem to take explicitly into account the discretization levels of I_k , such a problem may be difficult to solve optimally (we note that [56] made an intent in this direction in the context of power adaptation). Instead, we may opt for a sub-optimal approach, choosing the uniform discretization and adopting the solutions obtained with "infinite" discretization offered by the numerical tools we used. In the next chapter we will discuss this for the relay channel in more details and we will see that only a small number of feedback bits can actually do the work.

Chapter 5

Orthogonal Relaying Policies and Multi-Hop Channel

THIS chapter studies the rate optimization problem in the cooperative HARQ channel model described in Section 2.1.3 for fixed transmit power. The node $\mathcal{R}$ in the network operates an orthogonal relaying and has no information of its own to deliver to the destination [73, 74].

In the first section, we present the problem solution to rate adaptation in cooperative IR-HARQ and investigate the performance limits for the throughput of the proposed channel model, which is important since the capacity of such a channel is not known in general [71]. A lower bound of the performance is the optimal transmission from $\mathcal{S}$ to $\mathcal{D}$ which is known as *Decode-and-Forward* direct transmission. Rate adaptation HARQ as an MDP problem and can be solved straightforwardly using DP which is similar to the solution proposed in Section 4.2.

Next, we discuss the rate allocation scenario and present the formulation on the throughput as well as introducing a novel approach for optimization of rate allocation problem. Similar to the rate allocation problem in Chapter 4, we start with solving an approximated version of the problem. Then we discuss the complexity of the proposed approach compared to that of optimizing the original non-convex problem. Then we investigate the possibility of improving the solution of the proposed approximate optimization.

Performance limits of the relay channel model and the numerical results of rate allocation/adaptation HARQ are also presented in this chapter. Finally we extend the problem

set-up to a general network with M relay nodes and discuss the rate adaptation problem for such a network.

5.1 Adaptive-Rate Orthogonal Relaying

In this section, we study the achievable throughput of truncated HARQ transmission over the cooperative channel model. We assume that the NACMI I_k is being fed back to the transmitter node(s) at the end of each transmission attempt and the encoder is capable of adapting the transmission rate based on the feedback message.

5.1.1 Optimum Policy

To find the optimum decision policy for K transmission attempts we start from (4.8), and look for λ_{th} by finding the smallest value of λ where a non-degenerate solution for J^λ exists. The minimization problem in (4.8) can be cast into a DP problem and be solved in a recursive manner, starting from last stage to find the optimal decision policy for a given λ . The policy found in this way would correspond to transmitting the smallest amount of redundancy for the given λ value. So, (4.8) can be rewritten as,

$$\begin{aligned} J^\lambda &= \min_{\pi} \mathbf{D}(\pi) + \lambda \cdot P_{\text{out}}(\pi) \\ &= \min_{\substack{\rho_1^S, \dots, \rho_K^S \\ \rho_1^R, \dots, \rho_K^R}} \mathbb{E} \left\{ \sum_{k=1}^K \rho_k^S + \rho_k^R + \lambda \cdot \mathbb{I}(I_K^D < 1) \right\}. \end{aligned} \quad (5.1)$$

The recursive optimization problem, using (2.6a) and (2.6b), can be shown as follows in (5.2). The MDP state of this problem is a two-tuple $S_k = (I_k^D, I_k^R)$.

$$J^\lambda = J_1^\lambda(I_0^D, I_0^R) = \min_{\rho_1^S, \rho_1^R} \mathbb{E} \left\{ \rho_1^S + \rho_1^R + J_2^\lambda(I_0^D + \rho_1^S \cdot C_1^{SD} + \rho_1^R \cdot C_1^{RD}, I_0^R + \rho_1^S \cdot C_1^{SR}) \right\} \quad (5.2a)$$

$$J_2^\lambda(I_1^D, I_1^R) = \min_{\rho_2^S, \rho_2^R} \mathbb{E} \left\{ \rho_2^S + \rho_2^R + J_3^\lambda(I_1^D + \rho_2^S \cdot C_2^{SD} + \rho_2^R \cdot C_2^{RD}, I_1^R + \rho_2^S \cdot C_2^{SR}) \right\} \quad (5.2b)$$

...

$$J_K^\lambda(I_{K-1}^D, I_{K-1}^R) = \min_{\rho_K^S, \rho_K^R} \mathbb{E} \left\{ \rho_K^S + \rho_K^R + \lambda \cdot \mathbb{I}(I_{K-1}^D + \rho_K^S \cdot C_K^{SD} + \rho_K^R \cdot C_K^{RD} < 1) \right\} \quad (5.2c)$$

The minimization process starts with (5.2c) by solving

$$J_K^\lambda(I_{K-1}^D, I_{K-1}^R) = \min_{\rho_K^S, \rho_K^R} \left\{ \rho_K^S + \rho_K^R + \lambda \cdot \Theta(I_{K-1}^D, I_{K-1}^R) \right\}, \quad (5.3)$$

where

$$\Theta(I_{K-1}^D, I_{K-1}^R) = \begin{cases} F_{CSD}\left(\frac{1-I_{K-1}^D}{\rho_K^S}\right) & I_{K-1}^R < 1 \\ F_{CRD}\left(\frac{1-I_{K-1}^R}{\rho_K^R}\right) & I_{K-1}^R \geq 1 \end{cases}. \quad (5.4)$$

and $F_X(x) = \Pr\{X < x\}$ is the cdf.

The optimization is done *off-line* before the start of the transmission process. However, in real-time it is possible to optimize the remaining transmission attempts of an HARQ process. For instance, if 2 (not necessarily optimal) transmission attempts are already done and the decoder state is S_2 and maximum number of attempts is 5, we can always find the optimum policy for the remaining 3 attempts. In this example we need to solve the recursive system in (5.2) for $K = 3$ and setting $(I_0^D, I_0^R) = S_2$ in (5.2a) (instead of choosing $(I_0^D, I_0^R) = (0, 0)$ when we optimize the process off-line). This way we can find an optimal policy for the remaining 3 transmission attempts in real-time.

5.1.2 Initializing DP Recursion

A policy π , as a set of ρ functions, has $2K$ elements which we find one by one in a recursive optimization process. This can be seen as finding $2K - 1$ functions ρ that optimize the throughput of the transmission process (we already know that $\rho_1^R = 0$). Solving this problem directly leads to a complexity which grows exponentially in the number of points required to represent the function ρ and exponential in K as well. But, thanks to the recursive optimization using DP, the complexity of the $(2K - 1)$ -dimensional minimization problem reduces significantly (An example of an optimal redundancy policy for adaptive-rate transmission is shown in Figure 5.1).

The recursive optimization looks for $K - 1$ of 2-dimensional ρ^S functions, plus $K - 1$ of 1-dimensional ρ^R functions, which is significantly less complex than the main problem (ρ_1^S is a scalar of transmission redundancy at the start of the process and not a function therefore, complexity of optimizing it is negligible compared to the whole problem).

The optimization is done point by point over the discretized values of the state variables

(I_{k-1}^p, I_{k-1}^r) and the overall computational complexity grows linearly with K increasing¹. As a result, it is possible to optimize the policies even for large values of K and increase the number of discretization points as desired. As defined in (2.7a) the $\rho_k^s(I_{k-1}^p, I_{k-1}^r)$ is a 2-dimensional function and is non-zero only where $0 \leq I_{k-1}^p < 1$ & $0 \leq I_{k-1}^r < 1$; $\rho_k^r(I_{k-1}^p)$ is one-dimensional function and is used only in the Relaying phase.

The minimization process starts with (5.3) and the first step can be done analytically by finding the zero-crossings of the first derivative of cost in (5.4), with respect to the optimization parameters ρ . Since at each time only one node can be transmitting in the system, we are looking for two one-dimensional functions in this stage (This is because the ρ_K^s is fixed with respect to I_{K-1}^r). As a result, we find the solution to the following two equalities as the optimal ρ values

$$p_{C^{SD}}\left(\frac{1 - I_{K-1}^p}{\rho_K^s}\right) \cdot \frac{1 - I_{K-1}^p}{\rho_K^s \cdot \rho_K^s} = \frac{1}{\lambda}, \quad I_{K-1}^r < 1 \quad (5.5a)$$

$$p_{C^{RD}}\left(\frac{1 - I_{K-1}^p}{\rho_K^r}\right) \cdot \frac{1 - I_{K-1}^p}{\rho_K^r \cdot \rho_K^r} = \frac{1}{\lambda}, \quad I_{K-1}^r \geq 1. \quad (5.5b)$$

We discuss the first equation in (5.5a) while the second one follows the same discussion. Following the first equation in (5.5a), the right part of the below equation is non-negative ($I_{K-1}^r < 1$)

$$\frac{1}{\lambda} = p_{C^{SD}}\left(\frac{1 - I_{K-1}^p}{\rho_K^s}\right) \cdot \frac{1 - I_{K-1}^p}{\rho_K^s \cdot \rho_K^s}. \quad (5.6)$$

For a Rayleigh fading channel, like most of the other common wireless channel models, the pdf decays exponentially which is faster than ρ^{-2} . This tells us that the right side of (5.6) has a maximum $\max_{I_{K-1}^p, \rho_K^s} \left\{ p_{C^{SD}}\left(\frac{1 - I_{K-1}^p}{\rho_K^s}\right) \cdot \frac{1 - I_{K-1}^p}{\rho_K^s \cdot \rho_K^s} \right\} = \frac{1}{\lambda_c}$, where λ_c varies with I_{K-1}^p and therefore, we denote it as $\lambda_c(I_{K-1}^p)$. As a result, if $\lambda < \lambda_c(I_{K-1}^p)$ there is no solution to (5.6) which yields $\rho_K^s = 0$ and $J_K^\lambda = \lambda$. On the other hand, for $\lambda > \lambda_c(I_{K-1}^p)$ there are two solutions to (5.6) and we chose the one for which the second derivative is positive. The chosen solution and the solution equal to zero, will then be tested to find the optimum. The same way we can solve the second equation in (5.5a) for ρ_K^r . Having the $J_K^\lambda(I_{K-1}^p, I_{K-1}^r)$ the recursive minimization continues numerically to finally solve (5.2a) and finding the policy

¹This means that for 100 discretization points, the number of point-by-point minimizations to be solved is: $(K - 1) \times (100)^2 + 1 + (K - 1) \times 100$ which becomes 30301 minimizations for $K = 4$.

for the K -stage process. It is worth noting here that for $K = 1$ one can easily show that $\lambda_{\text{th}} = \lambda_c(0)$ while for $K > 1$ we have $\lambda_{\text{th}} < \lambda_c(0)$.

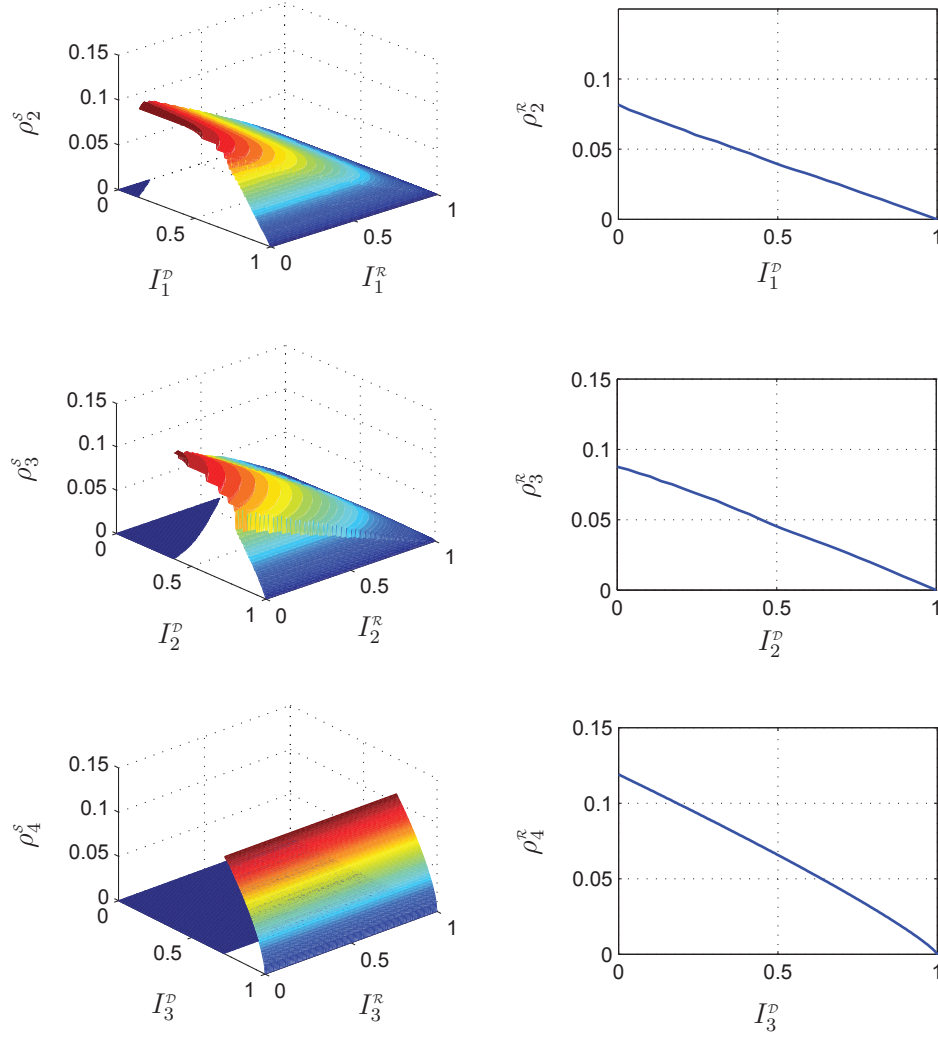


Figure 5.1 Optimal rate adaptation policy for HARQ over the relay channel model: $\rho_k^s(I_{k-1}^s, I_{k-1}^p)$ and $\rho_k^r(I_{k-1}^p)$ shown for $k = 2, 3, 4$ ($K = 4$) and $\bar{\gamma}^{sd} = 10$ dB. The packet dropping regions, where the redundancy value is equal to zero, is shown in blue for ρ_k^s .

5.1.3 Throughput Calculation

For a given policy π , we can anticipate the events of transmission stages by finding the joint probability distribution of I_k^D and I_k^R , starting from $k = 1$ and going recursively up to $k = K$. Due to the independency of channels for $k = 1$ (for given ρ_1^s) we have

$$p_{I_1^D, I_1^R}(x, y) = p_{I_1^D}(x) \cdot p_{I_1^R}(y) \quad (5.7)$$

with $p_{I_k^D, I_k^R}(x, y)$ being the joint pdf of I_k^D and I_k^R . For $k > 1$ the joint pdf can recursively be calculated as shown in (5.8).

$$p_{I_k^D, I_k^R}(x, y) = \int_0^x \int_0^y p_{I_{k-1}^D, I_{k-1}^R}(\alpha, \beta) \cdot \Pr\{I_k^D = x, I_k^R = y | I_{k-1}^D = \alpha, I_{k-1}^R = \beta\} d\beta d\alpha \quad (5.8)$$

It easily follows from (5.8) that, for $x, y < 1$, we have

$$p_{I_k^D, I_k^R}(x, y) = \int_0^x \int_0^y p_{I_{k-1}^D, I_{k-1}^R}(\alpha, \beta) \cdot p_{C^{SD}}\left(\frac{x - \alpha}{\rho_k^s(\alpha, \beta)}\right) \cdot p_{C^{SR}}\left(\frac{y - \beta}{\rho_k^s(\alpha, \beta)}\right) d\beta d\alpha. \quad (5.9)$$

Also, for $x < 1$ & $y \geq 1$, (5.8) yields the following

$$\begin{aligned} p_{I_k^D, I_k^R}(x, y) &= \int_0^x p_{I_{k-1}^D}^*(\alpha) \cdot p_{C^{RD}}\left(\frac{x - \alpha}{\rho_k^r(\alpha)}\right) d\alpha \\ &+ \int_0^x \int_0^1 p_{I_{k-1}^D, I_{k-1}^R}(\alpha, \beta) \cdot p_{C^{SD}}\left(\frac{x - \alpha}{\rho_k^s(\alpha, \beta)}\right) \cdot \left(1 - F_{C^{SR}}\left(\frac{1 - \beta}{\rho_k^s(\alpha, \beta)}\right)\right) \cdot \rho_k^s(\alpha, \beta) d\beta d\alpha, \end{aligned} \quad (5.10)$$

where,

$$p_{I_{k-1}^D}^*(\alpha) = \int_1^\infty p_{I_{k-1}^D, I_{k-1}^R}(\alpha, \beta) d\beta. \quad (5.11)$$

Then, the outage probability P_{out} , as the probability of I_K^D being less than 1, can be found as follows.

$$P_{\text{out}} = \int_0^1 \int_0^\infty p_{I_K^D, I_K^R}(\alpha, \beta) d\beta d\alpha. \quad (5.12)$$

The process of finding the optimal policy that gives the maximum achievable throughput for K transmission attempts can be summarized as the following steps:

- For a given λ value we first solve (4.8) in the recursive manner depicted in (5.2a) through (5.2c). This process starts with (5.3) and goes backward in k to find a policy $\pi \in \tilde{\mathbf{R}}_\lambda$.
- If $\rho_1^s(0, 0) = 0$ then the found policy π is degenerate, otherwise it is non-degenerate.
- We need to find at least one degenerate and one non-degenerate solution first and their corresponding λ values, λ_{deg} and $\lambda_{\text{non-deg}}$.
- Using the Bisection method (or any other root-finding method) we look for λ_{th} which is somewhere in $\lambda_{\text{th}} \in (\lambda_{\text{deg}}, \lambda_{\text{non-deg}}]$. The accuracy of the solution obtained can be arbitrarily chosen by changing the number of Bisection iterations.
- The throughput $\hat{\eta}$ is then

$$\hat{\eta} = \frac{1}{\lambda_{\text{th}}}. \quad (5.13)$$

Finally, for an arbitrary outage probability $\epsilon \leq \hat{\epsilon}$, the optimal outage constrained throughput is the solution $\eta(\pi)$ of the DP process described in Section 5.1.2 that gives $P_{\text{out}}(\pi) = \epsilon$. Also, for $\epsilon > \hat{\epsilon}$ the optimal outage constrained throughput is $\hat{\eta}$.

The computation of the outage probability for a given policy π can be done using the joint probability distribution of I_k^p and I_k^r as shown in (5.12). We note here that, because of the complexity of the integration process due to the discontinuities in the policies π , in Section 5.4, we use the Monte-Carlo method instead to calculate outage probabilities.

5.1.4 Quantization of the Feedback

In practice, the feedback message is supposed to be a limited-length packet of bits. To complete our analysis on rate adaptation, we assume an index of the discretized version of I_k^p and I_k^r , denoted respectively by ℓ_k^p and ℓ_k^r , being sent back through the feedback channels. For N_f bits of feedback message and a uniform discretization over $I_k \in [0, 1)$ we have $L + 1 = 2^{N_f}$ possible messages as feedback, or

$$\ell_k \in \{0, 1, \dots, L\} \quad (5.14)$$

where N_f can be chosen differently for any of the feedback channels. The uniform discretization can then be formalized as

$$\ell_k = \begin{cases} \lfloor I_k \cdot L \rfloor & 0 \leq I_k < 1 \\ L & I_k \geq 1 \end{cases}, \quad (5.15)$$

where $\lfloor \cdot \rfloor$ gives the largest integer smaller than the argument.

We assume that the encoders will still decide on the number of symbols (or, equivalently, on the amount of redundancy) to transmit, using the optimum policies $\pi \in \hat{\mathcal{R}}$ as found through the dynamic programming optimization. However, this decision may be made in different ways. We discuss here three different approaches of *minimum* (Min.), *median* (Med.) and *average* (Avg.) as defined below since they are simple to analyze and to implement in practice.

The encoder first finds an approximation of the actual value of I_k , denoted by $\mathring{I}_k$ in one of the three approaches as follows

$$\mathring{I}_k = \begin{cases} \ell_k \cdot \Delta & \text{Min.} \\ (\ell_k + \frac{1}{2}) \cdot \Delta & \text{Med.} \\ \bar{I}_k & \text{Avg.} \end{cases} \quad (5.16)$$

where $\Delta = \frac{1}{L}$ and $\bar{I}_k$ is the expectation of I_k given the received feedback messages. For the encoder of $\mathcal{S}$ this means that: $\bar{I}_k^{\mathcal{D}} = \mathbb{E}\{I_k^{\mathcal{D}} | \ell_k^{\mathcal{D}} = l_1, \ell_k^{\mathcal{R}} = l_2\}$ and $\bar{I}_k^{\mathcal{R}} = \mathbb{E}\{I_k^{\mathcal{R}} | \ell_k^{\mathcal{D}} = l_1, \ell_k^{\mathcal{R}} = l_2\}$, while for the encoder at node $\mathcal{R}$ it means: $\bar{I}_k^{\mathcal{D}} = \mathbb{E}\{I_k^{\mathcal{D}} | \ell_k^{\mathcal{D}} = l_1\}$. The encoders then respectively encode $\rho_k^{\mathcal{D}}(\mathring{I}_k^{\mathcal{D}}, \mathring{I}_k^{\mathcal{R}})$ and $\rho_k^{\mathcal{R}}(\mathring{I}_k^{\mathcal{D}})$ redundancies.

Based on the definitions presented here, we will show in the numerical results in Section 5.4, that using a few feedback message bits $N_f^{\mathcal{D}} = \log_2(L^{\mathcal{D}} + 1)$ and $N_f^{\mathcal{R}} = \log_2(L^{\mathcal{R}} + 1)$, the HARQ process can achieve the maximum throughput.

5.2 Rate Allocation in Orthogonal Relaying

The rate allocation policy π for the proposed relaying protocol is composed of K (the maximum number of transmission attempts) sets.

$$\pi = \{\varrho^{\mathcal{S}}, \varrho_l^{\mathcal{R}}\}, \quad 1 \leq l < K \quad (5.17)$$

The set of transmission rates for node $\mathcal{S}$, denoted by ϱ^s , is a vector of length K . Moreover, depending on the time l , at which the relay node decodes the message successfully and the system transits from broadcasting phase to relaying phase, a different set of transmission rates ϱ_l^r for the $K - l$ remaining transmissions from relay node are employed. Hence, altogether $\frac{K(K+1)}{2}$ input parameters (degrees of freedom) would be available for the optimization problem. So, inputs to this optimization problem are a set of transmission rates described as follows.

$$\varrho^s : \{\rho_k^s | 1 \leq k \leq K\} \quad (5.18a)$$

$$\varrho_l^r : \{\rho_{l,k}^r | l < k \leq K\}, \quad 1 \leq l < K \quad (5.18b)$$

We denote the normalized mutual information exchanged between nodes $\mathcal{S}$ and $\mathcal{D}$ at the k th transmission attempt by $\nu_k^{s\mathcal{D}} = C(\gamma_k^{s\mathcal{D}}) \cdot \rho_k^s$. Where, for the Gaussian distributed symbols of the coding method used here, the mutual information per channel use (symbol) gained by the decoder of node $\mathbf{b}$ from the node $\mathbf{a}$ encoder is equal to $C_k^{\mathbf{ab}} = \log_2(1 + \gamma_k^{\mathbf{ab}})$. We also denote the normalized mutual information exchanged between nodes $\mathcal{R}$ and $\mathcal{D}$ at the k th transmission attempts, with the successful decoding at relay happening in l th attempt, by $\nu_{l,k}^{r\mathcal{D}} = C(\gamma_{l,k}^{r\mathcal{D}}) \cdot \rho_{l,k}^r$.

Following from these definitions, the NACMI at decoder of $\mathcal{D}$ at the end of the k th transmission for two example cases is as follows.

- in case that relay node decodes the message at the l th attempt $I_k^p = \sum_{t=1}^l \nu_t^{s\mathcal{D}} + \sum_{t=l+1}^k \nu_{l,t}^{r\mathcal{D}}$,
- in case that relay node doesn't decodes the message up to the time k then, $I_k^p = \sum_{t=1}^k \nu_t^{s\mathcal{D}}$.

In broadcasting phase $I_k^r < 1$, and in relaying phase $I_k^r \geq 1$ (noting that at both phases $I_k^p < 1$). Also, $\gamma^{r\mathcal{D}} \equiv 0$ in the broadcasting phase while $\gamma^{s\mathcal{D}} \equiv 0$ and $\gamma^{s\mathcal{R}} \equiv 0$ during the relaying phase. The transmission process stops as soon as $I_k^p \geq 1$ or $k = K$.

5.2.1 Closed-Form Representations

A failure happens in the truncated HARQ process only if after K transmission attempts $I_K^p < 1$. This can be the resulting from any of the K disjoint events bellow:

$$E_l^* = \left\{ \sum_{k=1}^{l-1} \nu_k^{sR} < 1 \wedge \sum_{k=1}^l \nu_k^{sR} > 1 \wedge \sum_{k=1}^l \nu_k^{sD} + \sum_{k=l+1}^K \nu_{l,k}^{rD} < 1 \right\}, \quad 1 \leq l \leq K-1 \quad (5.19a)$$

$$E_K^* = \left\{ \sum_{k=1}^{K-1} \nu_k^{sR} < 1 \wedge \sum_{k=1}^K \nu_k^{sD} < 1 \right\} \quad (5.19b)$$

There are also several events that lead to the decoding success in the HARQ process. We categorize them into two groups:

- Success events that happen in the broadcasting phase which we denote by E_l . In these events, decoding at $\mathcal{D}$ is done only based on the information from node $\mathcal{S}$.
- Success events that happen following a transition to the relaying phase at transmission attempt l and we denote them by $E_{l,k}$. In this group of events, $\mathcal{R}$ has succeeded in decoding at some time l and therefore, the destination node has some mutual information from the relay node too.

We formulate these two events as follows:

$$E_l = \left\{ \sum_{i=1}^{l-1} \nu_i^{sD} < 1 \wedge \sum_{i=1}^{l-1} \nu_i^{sR} < 1 \wedge \sum_{i=1}^l \nu_i^{sD} > 1 \right\}, \quad 1 \leq l \leq K \quad (5.20a)$$

$$E_{l,k} = \left\{ \sum_{i=1}^{l-1} \nu_i^{sR} < 1 \wedge \sum_{i=1}^l \nu_i^{sD} + \sum_{i=l+1}^{k-1} \nu_{l,i}^{rD} < 1 \right. \\ \left. \wedge \sum_{i=1}^l \nu_i^{sR} > 1 \wedge \sum_{i=1}^l \nu_i^{sD} + \sum_{i=l+1}^k \nu_{l,i}^{rD} > 1 \right\}, \quad 1 \leq l < k \leq K \quad (5.20b)$$

To be able to compute the probability of the events defined above, we introduce three

probability functions as follows:

$$P_k^{SD} \triangleq \Pr\left\{\sum_{i=1}^k \nu_i^{SD} < 1\right\}, \quad (5.21a)$$

$$P_k^{SR} \triangleq \Pr\left\{\sum_{i=1}^k \nu_i^{SR} < 1\right\}, \quad (5.21b)$$

$$P_{l,k}^{SRD} \triangleq \Pr\left\{\sum_{i=1}^l \nu_i^{SD} + \sum_{i=l+1}^k \nu_i^{RD} < 1\right\}. \quad (5.21c)$$

Here, P_k^{SD} and P_k^{SR} are the probabilities of decoding failure, respectively at the destination and the relay, after k transmission attempts by the source node $\mathcal{S}$. Moreover, $P_{l,k}^{SRD}$ is the probability of decoding failure at the destination after l transmission attempts from the source node followed by $k - l$ transmission attempts from the relay node.

We use the definitions in (5.21a), (5.21b) and (5.21c) to calculate the throughput of the relaying protocol, which follows from (2.13) as:

$$\eta = \frac{N_b \cdot (1 - P_{\text{out}})}{\overline{N}_s} \quad (5.22)$$

The following corollary to the throughput theorem in [1] presents a closed-form for the outage probability and the throughput of the presented variable-rate transmission, based on the policy set π in (5.17).

Corollary 1. (Throughput Formula) *Throughput of the cooperative HARQ protocol described in Section 2.1.3 with variable-rate transmission can be calculated as*

$$\eta = \frac{1 - P_{\text{out}}}{\mathbf{D}}, \quad (5.23)$$

where P_{out} and $\mathbf{D}$ are given as follows.

$$P_{\text{out}} = P_K^{SD} \cdot P_{K-1}^{SR} + \sum_{i=1}^{K-1} [P_{i-1}^{SR} - P_i^{SR}] \cdot P_{i,K}^{SRD} \quad (5.24)$$

$$\mathbf{D} = \sum_{i=1}^K \rho_i^s \cdot P_{i-1}^{SD} \cdot P_{i-1}^{SR} + \sum_{i=1}^{K-1} [P_{i-1}^{SR} - P_i^{SR}] \cdot \left[\sum_{l=i+2}^K \rho_{i,l}^R \cdot P_{i,l-1}^{SRD} + \rho_{i,i+1}^R \cdot P_i^{SD} \right] \quad (5.25)$$

The proof is given in Appendix B.

5.2.2 Approximate Recursive Optimization

The throughput of HARQ for K retransmissions as introduced in Corollary 1 has $K(K+1)/2$ optimization variables. In this section we present a dual optimization problem inspired by [51, 53] and solve the optimization problem in a recursive manner which reduces the complexity of problem.

The first step here is to substitute the original problem with its dual in (4.8). However, we need to have the MDP state S_k which meets two conditions: first, the probability of failure events at the end of k th transmission must be computed knowing S_k ; second, knowing the k th optimization parameter (transmitted redundancy) and S_k , the new state S_{k+1} should be obtained [24]. However, for calculation of the probability of any failure event we need to know the optimization parameters for all the steps up to that stage, which implies that whatever the state S_k we choose, it cannot satisfy the first condition.

Following this observation, in the same approach as in Section 4.3, we do some modifications to the problem. As already suggested in [12, 51] we choose to approximate the probability of failure events using a Gaussian approximation [15]. For instance for P_k^{SD} in (5.21a) we use $\tilde{P}_k^{SD}$ where

$$\tilde{P}_k^{SD} = \begin{cases} F_{C^{SD}}\left(\frac{1}{\rho_k^s}\right), & k = 1 \\ Q\left(\frac{\bar{C}^{SD} \cdot X_k - 1}{\sigma_{C^{SD}} \cdot \sqrt{Y_k}}\right), & \text{otherwise.} \end{cases} \quad (5.26)$$

In (5.26), $\bar{C}^{ab} = \mathbb{E}_{C^{ab}}\{C^{ab}\}$ and $\sigma_{C^{ab}}^2 = \mathbb{E}_{C^{ab}}\{C^{ab^2}\} - \bar{C}^{ab^2}$. Also, $X_k = \sum_{l=1}^k \rho_l^s$, $Y_k = \sum_{l=1}^k \rho_l^{s^2}$ and $Q(x)$ defined as in (4.27).

We can define $\tilde{P}_k^{SR}$ in the same way for the channel with $\mathbf{ab} = \mathcal{SR}$. Moreover, we use $\tilde{P}_{l,k}^{SRD}$ defined below, instead of $P_{l,k}^{SRD}$

$$\tilde{P}_{l,k}^{SRD} = \begin{cases} F_{C^{SD}}\left(\frac{1}{\rho_k^s}\right), & k = 1 \\ Q\left(\frac{\bar{C}^{SD} \cdot X_k + \bar{C}^{RD} \cdot x'_k - 1}{\sigma_{C^{SD}} \cdot \sqrt{Y_k} + \sigma_{C^{RD}} \cdot \sqrt{y'_k}}\right), & \text{otherwise} \end{cases}, \quad (5.27)$$

where $x'_k = \sum_{i=l+1}^k \rho_{l,i}^{\mathcal{R}}$ and $y'_k = \sum_{i=l+1}^k \rho_{l,i}^{\mathcal{R}^2}$.

In this case, instead of J^λ we find $\tilde{J}^\lambda$, which is calculated with substituting the Gaussian approximated outage probabilities in (4.8).

$$\tilde{J}^\lambda = \tilde{J}_1^\lambda(0, 0), \quad (5.28)$$

We show in Appendix C how to find $\tilde{J}_1^\lambda(X_0, Y_0)$. It is also shown how to find the optimal $\tilde{\pi} = \pi(\hat{X}_K, \hat{Y}_K)$. The throughput of the maximizing policy $\hat{\pi}$ will then be denoted as $\tilde{\eta} = \eta(\tilde{\pi})$, with the exact calculation as discussed in Section 5.2.4.

5.2.3 Simplified One-Dimensional State Optimization

A simplified version of the proposed optimization can be obtained by modifying the problem in a way that the MDP state is only one dimensional or $S_k = X_k$. The state elements in (5.31) have be discretized into L points and for a two dimensional space, which would create L^2 minimizations at each step. Therefore, reducing the state space to a one dimensional state will immediately decrease the complexity of the optimization process by reducing the number of minimizations from L^2 to L .

We discuss the one dimensional state in this section using Gaussian approximation as presented in Section 4.3 and approximating the state elements as: $\sqrt{Y_k} \approx X_k$ and $\sqrt{y'_k} \approx x'_k$.

The failure probabilities P_k^{SD} (and similarly P_k^{SR}) and $P_{l,k}^{SRD}$ when approximated as functions of X_k and x'_k , are presented as follows.

$$P_k^{SD} \approx \tilde{P}_k^{SD}(X_k) = \begin{cases} F_{C^{SD}}\left(\frac{1}{\rho_k^{\mathcal{S}}}\right), & k = 1 \\ Q\left(\frac{\bar{C}^{SD} \cdot X_k - 1}{\sigma_{C^{SD}} \cdot X_k}\right) & \text{otherwise} \end{cases} \quad (5.29)$$

$$P_{l,k}^{SRD} \approx \tilde{P}_{l,k}^{SRD}(X_l, x'_k) = \begin{cases} F_{C^{SD}}\left(\frac{1}{\rho_k^{\mathcal{S}}}\right), & k = 1 \\ Q\left(\frac{\bar{C}^{SD} \cdot X_l + \bar{C}^{RD} \cdot x'_k - 1}{\sigma_{C^{SD}} \cdot X_l + \sigma_{C^{RD}} \cdot x'_k}\right), & \text{otherwise} \end{cases} \quad (5.30)$$

As a result, to maximize the throughput using one-dimensional Gaussian approximation probabilities, we look for $\tilde{J}^\lambda$ instead of J^λ with substituting the outage probabilities in (4.8)

with the approximated version. Then, the goal is to find the following.

$$\check{J}^\lambda = \check{J}_K^\lambda(\check{X}_K), \quad (5.31)$$

where $\check{X}_K = \arg_X \min J_K^\lambda(X)$ and $\check{J}_K^\lambda$ is presented in Appendix D. After $\check{X}_K$ is found, the solution set $\check{\pi} = \pi(\check{X}_K)$ is created and $\check{\eta} = \eta(\check{\pi})$ is computed using the exact throughput calculation in Section 5.2.4.

5.2.4 Exact Throughput Calculation

The next step after finding a solution π is to calculate the exact probability of failure P_k^{SD} , P_k^{SR} and $P_{l,k}^{SRD}$ as introduced respectively in (5.21a), (5.21b) and (5.21c) and then calculate $\eta(\pi)$ using (5.24) and (5.25). In general we have the representation of a failure probability as

$$P_k = \Pr\left\{\sum_{i=1}^k \nu_i < 1\right\}, \quad (5.32)$$

where $\nu_i = \rho_i \cdot C_i$ is a random variable with cdf as follows:

$$\begin{aligned} F_{\nu_i}(x) &= \Pr\{\nu_i < x\} \\ &= \Pr\left\{C_i < \frac{x}{\rho_i}\right\} \\ &= \int_0^{2^{\frac{x}{\rho_i}} - 1} p_{\gamma_i}(\gamma) d\gamma. \end{aligned} \quad (5.33)$$

Moreover, the pdf of the random variable ν_i , using (5.33), can be shown as follows.

$$\begin{aligned} p_{\nu_i}(x) &= \frac{d}{dx} F_{\nu_i}(x) \\ &= p_{\gamma_i} \left(2^{\frac{x}{\rho_i}} - 1 \right) \frac{2^{\frac{x}{\rho_i}}}{\rho_i} \ln(2). \end{aligned} \quad (5.34)$$

As discussed briefly in Section 4.3, the pdf of the random variable $I_k = \sum_{i=1}^k \nu_i$, since all ν_i s are independent for $1 \leq i \leq k$, would be the convolution of the individual pdf functions

as follows.

$$p_{I_k}(x) = p_{\nu_1} * p_{\nu_2} * \cdots * p_{\nu_k}(x), \quad (5.35)$$

which can be found numerically for known $p_{\nu_i}(x)$ s using, e.g., discrete Fourier transform. The failure probability P_k can then be found as

$$\begin{aligned} P_k &= \Pr\{I_k < 1\} \\ &= \int_0^1 p_{I_k}(x) dx. \end{aligned} \quad (5.36)$$

5.2.5 Remarks on the Complexity

It is difficult to prove that the optimization function in (4.1) is convex. As we discussed so far, for the rate adaptation, thanks to DP, we can easily transform the exponentially complex non-linear problem into a set of recursive optimizations, which can be globally optimized without major complexity concerns.

In case of the rate allocation problem, we discussed that a recursive optimization on the original problem will not be feasible. As a result, we generated an approximate version of the throughput formulation using approximate outage probabilities presented in (4.28) and (5.27) respectively for the single-hop and relay channel models. We also presented a more simplified version by introducing the one dimensional approximated outage probabilities in (4.31) and (5.29) respectively for single-hop and relay channel models. The modified problem can then be solved as a recursive optimization problem with a global optimization at each recursion. The solution found in this way, as noted earlier, is a lower bound to the optimal solution.

Here, we want to use a convex programming optimization method on the rate allocation throughput maximization problem. The question is: Can we get a better solution by locally optimizing the original problem and using the solution of the approximate problem $\tilde{\pi}$ as the starting point?

To answer this question we run a set of experiments using the “`fminsearch`” function in MATLAB which is an interior-point optimization function. We focus on the relay channel model rate allocation problem because it is the more complex problem with a significantly large number of optimization parameters. However, we note that the same experiments

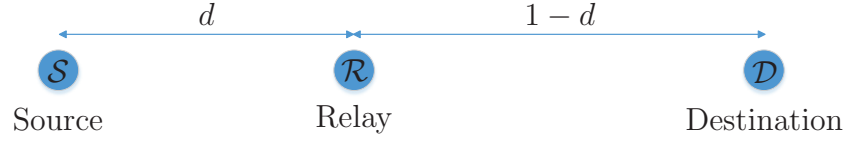


Figure 5.2 Topology of the relay channel for experimental analysis.

can be done for the single-hop channel model. The experiments are on optimizing the original rate allocation throughput representation in (5.23), using different starting points, as follows:

1. Set the starting point at 0.1 for all the optimization parameters (i.e., the redundancy values). We denote the result of this experiment by π_o .
2. Optimization using $\tilde{\pi}$ (i.e., the solution to the two-dimensional approximated version of the problem) as the starting point. We denote the result of this experiment by $\tilde{\pi}_o$.
3. Starting point being set at $\tilde{\pi}$ (i.e., the solution to the one-dimensional approximated version of the problem) with the result of this experiment being denoted by $\tilde{\pi}_o$.

We run the tests for the channel characteristics as follows. We assume a channel with normalized distance of one between source and destination and a relay node positioned with a distance of $0 \leq d \leq 1$ from source on the line between source and relay (Figure 5.2). Therefore, the relation of the average long-term channel gain of the links between the nodes will be

$$\bar{\gamma}^{SR} = \frac{1}{d^\nu} \bar{\gamma}^{SD} \quad (5.37a)$$

$$\bar{\gamma}^{RD} = \frac{1}{(1-d)^\nu} \bar{\gamma}^{SD}, \quad (5.37b)$$

with ν the path-loss exponent. For a Rayleigh fading channel, SNR is characterized by the exponential probability distribution function

$$p_{\gamma^{ab}}(\gamma) = \frac{1}{\bar{\gamma}^{ab}} \exp\left(-\frac{\gamma}{\bar{\gamma}^{ab}}\right), \quad (5.38)$$

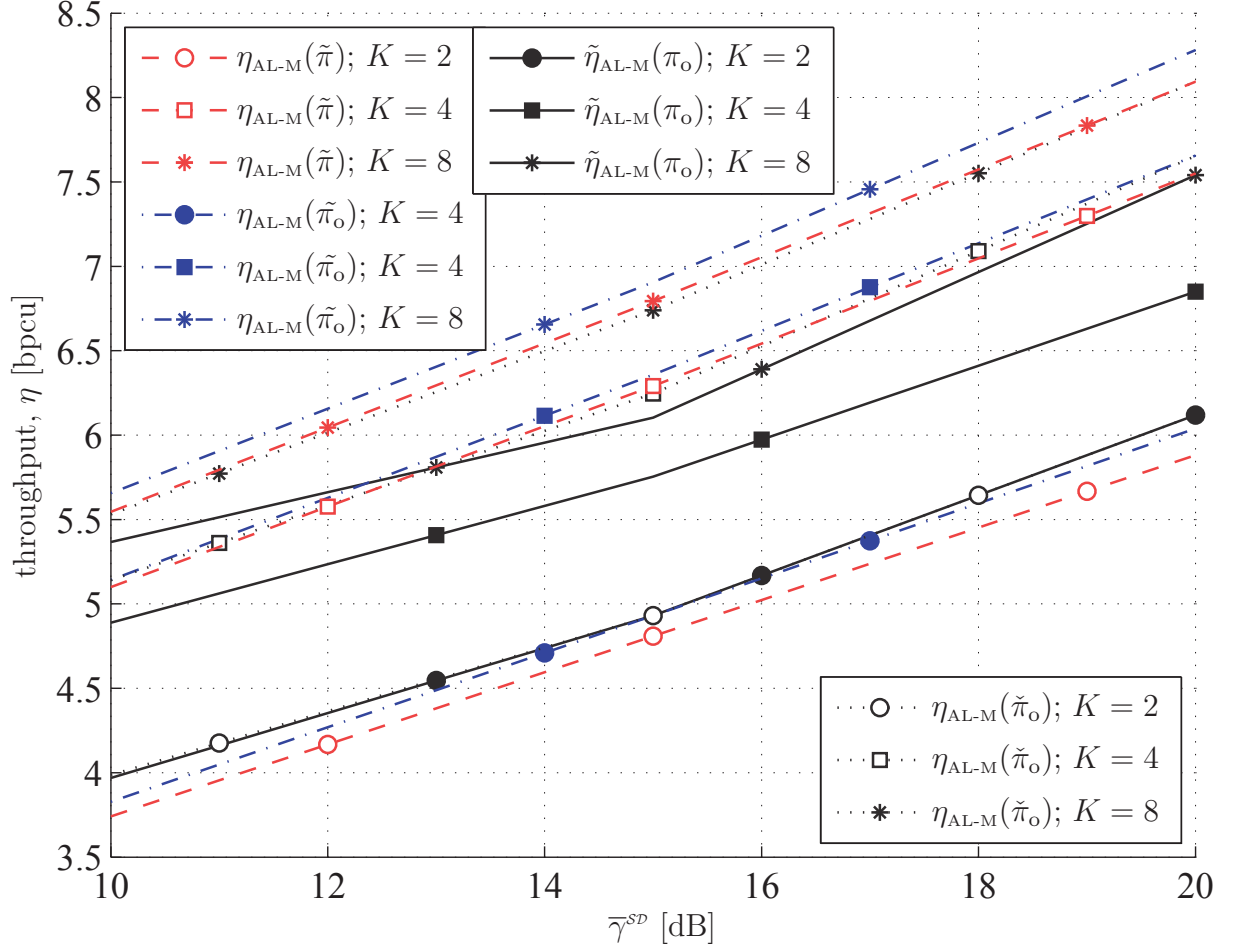


Figure 5.3 Throughput for different optimization experiments.

where $\bar{\gamma}^{ab}$ is the average SNR. Unless otherwise specified, for all the numerical results we assume that $d = 0.5$ and we set $\nu = 4$ for the path-loss exponent².

The results of the maximum achieved throughput with each of the above experiments are shown in Figure 5.3. The optimization experiments result in a slightly improved throughput value in all the cases except for the first experiment where a random point is given to the optimization algorithm as an starting point. This underlines the importance of the starting point in a non-linear optimization problem. The optimal policy found using both

²Any desired changes can be made to the topology of the relay channel and/or the path-loss exponent, which may change the optimum policy but will not affect the approach for the optimization process.

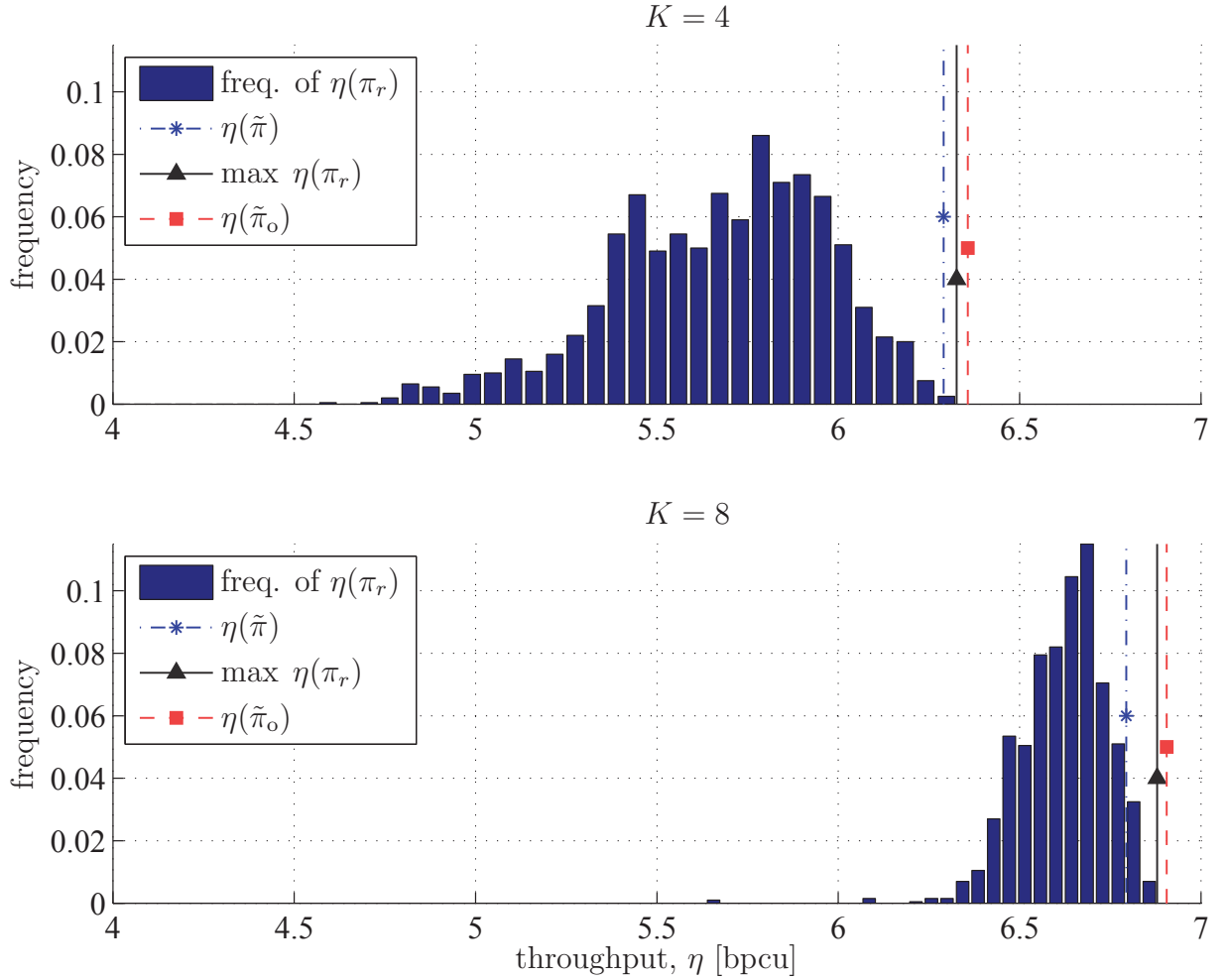


Figure 5.4 Histogram plot for the normalized frequency of a solution π_r that is found using the MATLAB optimization function with randomly generated starting point for $\bar{\gamma}^{sd} = 15$ dB. The maximum throughput found in the random starting point experiment and the maximum throughput found using the proposed optimization method ($\eta(\tilde{\pi})$) is shown for comparison.

2-dimensional and simplified one-dimensional approaches are compared to $\tilde{\pi}_o$ for $\bar{\gamma}^{sd} = 15$ dB and $K = 4$ in Figure 5.5.

Simulation results for the second test that we run are shown in Figure 5.4. In this test we try to globally optimize the throughput using randomly generated starting points π_r . We repeated the test for 2000 randomly generated starting points. For $K = 4$, 96.4 % of the tests converged to a solution with the solution values depicted in Figure 5.4, while only

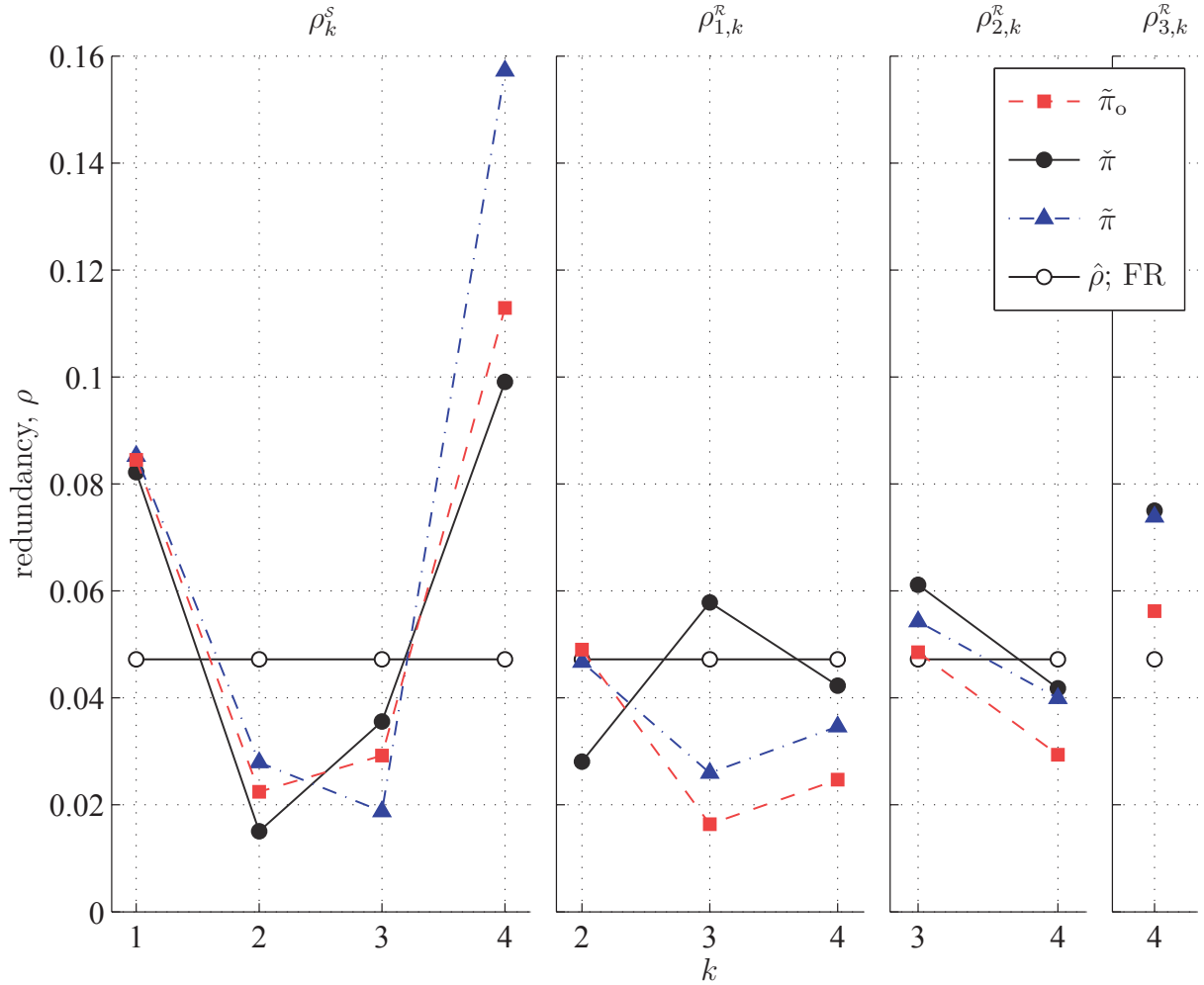


Figure 5.5 Optimal rate allocation policy for HARQ over the relay channel model found for $\bar{\gamma}^{sd} = 15$ dB and $K = 4$. Solution of the two recursive optimizations methods presented with 2-dimensional state space denoted by $\tilde{\rho}$ and with one dimensional state space denoted by $\tilde{\rho}$ are compared to the optimized version of $\tilde{\rho}$ which is denoted by $\tilde{\rho}_o$. The back dotted line show the optimum fixed-rate (FR) redundancy value.

0.15 % of the results are in the range of $\eta(\tilde{\rho})$ or larger. For $K = 8$ the convergence rate is only 64.9 %.

The optimal result of this test is less than the result of optimization result when the starting point is set to $\tilde{\rho}$ which is shown as $\tilde{\rho}_o$ in Figure 5.4. This is despite the fact that finding $\tilde{\rho}$ and then $\tilde{\rho}_o$ takes at most a few hours of time on a regular personal computer for

$K = 8$ while the random starting point test above takes time in order of weeks on the same machine.

5.3 Performance Limits

In [30], it is shown that for the HARQ single-hop transmission channel, $\hat{\eta}_K$ grows with K increasing. As shown in the same paper, and proved in [1], with infinite transmission attempts the maximum achievable throughput reaches the ergodic capacity of the fading channel. Here, we expect $\hat{\eta}_K$ to grow with K however, knowing the throughput of a channel is always bounded with the capacity, there should be certain maximum achievable throughput for $K \rightarrow \infty$.

The capacity of the relay channel with input X_1 , relay input X_2 , output Y and relay output Y_1 for an arbitrary channel given by $p(y, y_1|x_1, x_2)$ and a feedback from (Y, Y_1) to both X_1 and X_2 is given by [?, Theorem 17.3]

$$C = \max_{p(x_1, x_2)} \min \left\{ I(X_1, X_2; Y), I(X_1; Y, Y_1|X_2) \right\}, \quad (5.39)$$

where $I(\cdot)$ is the mutual information function. For the *full-duplex* (FD) relay transmission the capacity of AWGN channel becomes [75]

$$C_{\text{FD}} = \max_{\beta} \min \left\{ C(\bar{\beta}(\gamma^{sR} + \gamma^{sD})), C(\gamma^{sD} + \gamma^{rD} + 2\sqrt{\beta\gamma^{sD}\gamma^{rD}}) \right\}, \quad (5.40)$$

where $\bar{\beta} = 1 - \beta$, $C(x) = \frac{1}{2} \log(1 + x)$ and the ergodic capacity is $C_{\text{FD-erg}} = \mathbb{E}\{C_{\text{FD}}\}$.

However a full-duplex transmission in the relay node is not fully practical. In case of a *half-duplex* (HD) relay we can assume a *Time Division* (TD) manner over the relay node as suggested in [76]. We suppose that relay node listens (and does not transmit) in α portion of the time ($0 \leq \alpha \leq 1$) and transmits in the remaining $\bar{\alpha} = 1 - \alpha$ portion of the time. Under this assumptions the terms in (5.39) can be represented as

$$I(X_1, X_2; Y) = \alpha I(X_1; Y) + \bar{\alpha} I(X_1, X_2; Y), \quad (5.41a)$$

$$I(X_1; Y, Y_1|X_2) = \alpha I(X_1; Y, Y_1) + \bar{\alpha} I(X_1; Y|X_2). \quad (5.41b)$$

Since the relay node uses all its energy in $\bar{\alpha}$ portion of time and the source node can allocate

κ fraction of its energy ($0 \leq \kappa \leq 1$) in the first α portion of time and the remaining $\bar{\kappa} = 1 - \kappa$ fraction in the remaining $\bar{\alpha}$ portion of time, (5.41b) becomes (5.42a) and (5.42b).

$$\begin{aligned} C_{\text{HD-1}} &= I(X_1, X_2; Y) \\ &= \alpha C\left(\frac{\kappa}{\alpha}(\gamma^{sR} + \gamma^{sD})\right) + \bar{\alpha} C\left(\bar{\beta}\gamma^{sR}\frac{\bar{\kappa}}{\bar{\alpha}}\right) \end{aligned} \quad (5.42a)$$

$$\begin{aligned} C_{\text{HD-2}} &= I(X_1; Y, Y_1|X_2) \\ &= \alpha C\left(\frac{\kappa}{\alpha}\gamma^{sD}\right) + \bar{\alpha} C\left(\frac{\bar{\kappa}}{\bar{\alpha}}\gamma^{sD} + \frac{1}{\bar{\alpha}}\gamma^{rD} + 2\sqrt{\beta\frac{\bar{\kappa}}{(\bar{\alpha})^2}\gamma^{sD}\gamma^{rD}}\right) \end{aligned} \quad (5.42b)$$

Then the capacity is

$$C_{\text{HD}} = \max_{\beta, \alpha, \kappa} \min \left\{ C_{\text{HD-1}}, C_{\text{HD-2}} \right\}. \quad (5.43)$$

Since we are assuming fixed-power transmission in the problem definition of our HARQ channel, we can relax κ in the maximization in (5.43) by choosing $\kappa = \alpha$ in (5.42b). Moreover, for the particular case where only one transmitter node can be active at a time, we study $\beta = 0$ as well. Finally, the ergodic form of the half-duplex capacity can be obtained as $C_{\text{HD-erg}} = \mathbb{E}\{C_{\text{HD}}\}$.

Even though we cannot compare the upper limit of our multi-hop channel model with a known capacity region (e.g., upper bound on it are introduced in the literature [71]), it can be still accepted as an achievable region (perhaps close to the unknown capacity). The performance upper bound in this case happens when infinite number of transmission attempts is allowed ($K \rightarrow \infty$). The steady average cost for the cooperative channel can be found by putting $\rho = \rho^s + \rho^r$ in (4.37), noting that $J_\infty = \lim_{K \rightarrow \infty} \sum_{k=1}^K \mathbb{E}\{\rho_k^s + \rho_k^r\}$.

5.4 Numerical Results

In this section we present some numerical results for the described relaying schemes with IR-HARQ for both rate adaptation/allocation. In case of rate allocations, we use the optimized version of the 2-D state space solution for maximum throughput in Section 5.2.5 denoted by $\tilde{\pi}_o$, as the maximum throughput achieved with variable rate allocation as introduced in and we denote it by $\hat{\eta}_{\text{AL-M}} = \eta(\tilde{\pi}_o)$. We use the same channel characteristics as described in Section 5.2.5.

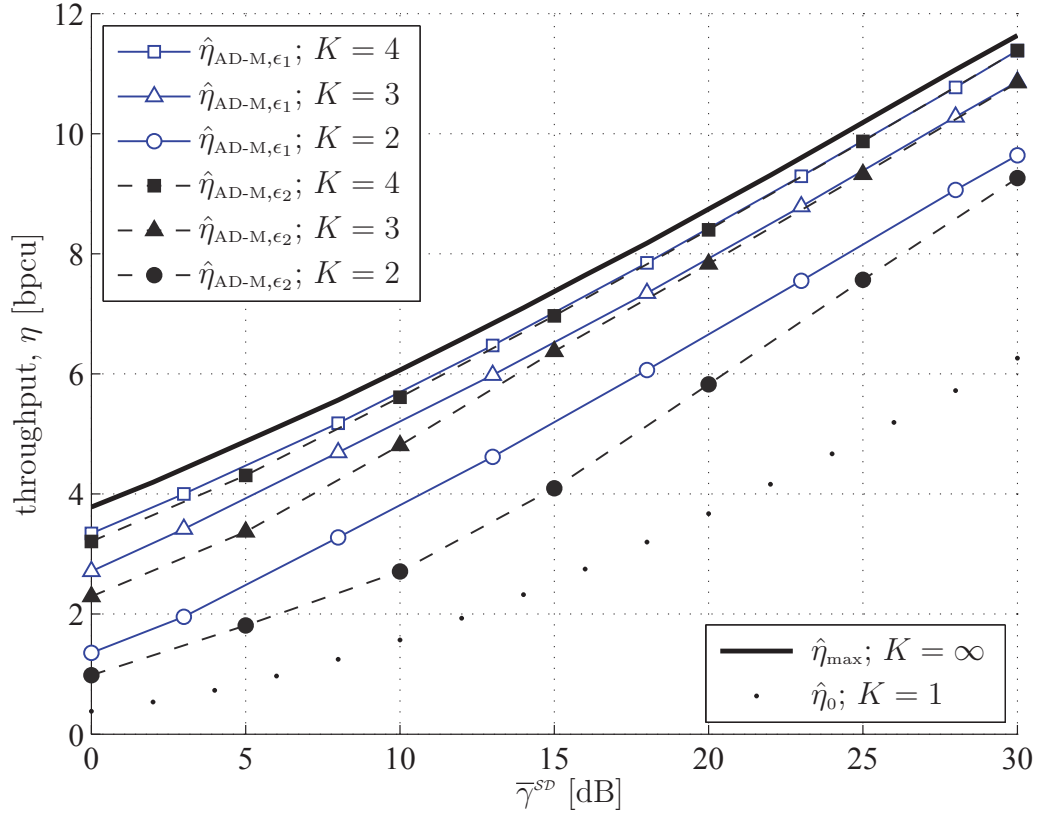


Figure 5.6 Throughput for $K = 2, 3, 4$; for constrained outage probability of $\epsilon_1 = 10^{-2}$ (red) and $\epsilon_2 = 10^{-4}$ (blue). The results are compared with the maximum achievable throughput of the relaying protocol (thick black) and the direct transmission lower bound (dotted black).

As stated in (2.7a), the optimum rate adaptation policy $\hat{\pi}_{AD-M}$ for the transmission attempt k consists of two functions: a two parameter function $\rho_k^s(I_{k-1}^p, I_{k-1}^r)$ which can be non-zero only at $0 \leq I_{k-1}^p, I_{k-1}^r < 1$ and a one parameter function $\rho_k^r(I_{k-1}^p)$ which can be non-zero only at $0 \leq I_{k-1}^p < 1$. As an example in Figure 5.1 for average SNR of $\bar{\gamma}^{SD} = 10$ dB, we can see the optimal policy for maximum $K = 4$ transmission attempts.

For the first transmission, as depicted in Figure 5.1, there is no information accumulated at decoders so, the source node starts with transmitting $\rho_1^s(0, 0) = 0.0923$ ($\rho_1^r = 0$). The second decision might be made over ρ_2^s or ρ_2^r depending on the decoder states. In the last transmission attempt in case of still being in broadcasting phase, source is having a last try, as we can appreciate in the left, bottom figure of Figure 5.1, but this time only to deliver

the message at node $\mathcal{D}$ and thus ρ_4^s does not change by I_3^R . It is noticeable that the accurate small redundancy values presented here imply relatively long packets for transmission. We based our analysis on asymptotically long codewords where the arbitrarily small outage probability can be presumed for the transmission rate being in the capacity region of the channel. This means that for relatively short codewords, the presented results in can be optimistic.

As shown in Figure 5.1, the adaptation policies $\rho_k^s(I_{k-1}^p, I_{k-1}^R)$ for $1 < k \leq K$ may become zero for small values of the arguments. This is an interesting result similar to what we noted in Section 4.5 for single-hop transmission which means that in order to maximize the throughput, if the receiver has not accumulated enough of mutual information, we should rather stop the transmission and proceed with the HARQ process of the next available packet.

For the rate allocation problem, Figure 5.5 shows the maximum throughput achieving set $\tilde{\pi}_{\text{AL-M}}$ for the case of two dimensional state space optimization for the SNR of $\bar{\gamma}^{sp} = 15$ dB. Following this result, $\mathcal{S}$ starts the transmission process by choosing a subset of $N_{s,1}^s = \rho_1^s \cdot N_b$ number of symbols from the generated codeword $\mathbf{x}$ and broadcasts it to the other two nodes. Retransmissions from $\mathcal{S}$ (or $\mathcal{R}$ in case of decoding success at relay) will then be pursued using $N_{s,k} = \rho_k \cdot N_b$ new symbols from the same codeword until $\mathcal{D}$ decodes the message successfully or a maximum $K = 4$ transmission chances are over.

It is also depicted in Figure 5.5, the optimal policy found using the one-dimensional state space optimization $\tilde{\pi}_{\text{AL-M}}$, described in Section 5.2.3. Since it can be found through a much simpler minimization process, it is still an interesting result for us, not only because it gives a better performance than fixed-rate transmission but also because it can be used as an easy-to-find starting point for any global optimization method used to solve (5.23).

In Figure 5.6 the maximum achievable throughput for a constrained ϵ outage probability $\hat{\eta}_\epsilon$ is shown for two different values of $\epsilon = 10^{-2}$ and $\epsilon = 10^{-4}$. In this figure, we assume the rate adaptation HARQ is being used and as expected, tighter constraint on the outage decreases the optimal throughput. As depicted in Figure 5.6, a tighter constraint on outage does not affect the throughput at high SNR values as the K value increases. This follows the fact that the outage probability of the optimal solution in the case of high SNR and large K is already less than or very close to the constrained outage value and as a result, the constraint does not affect the throughput as much as it does for lower SNR values.

Figure 5.7 shows the optimal throughput with respect to the average SNR for different

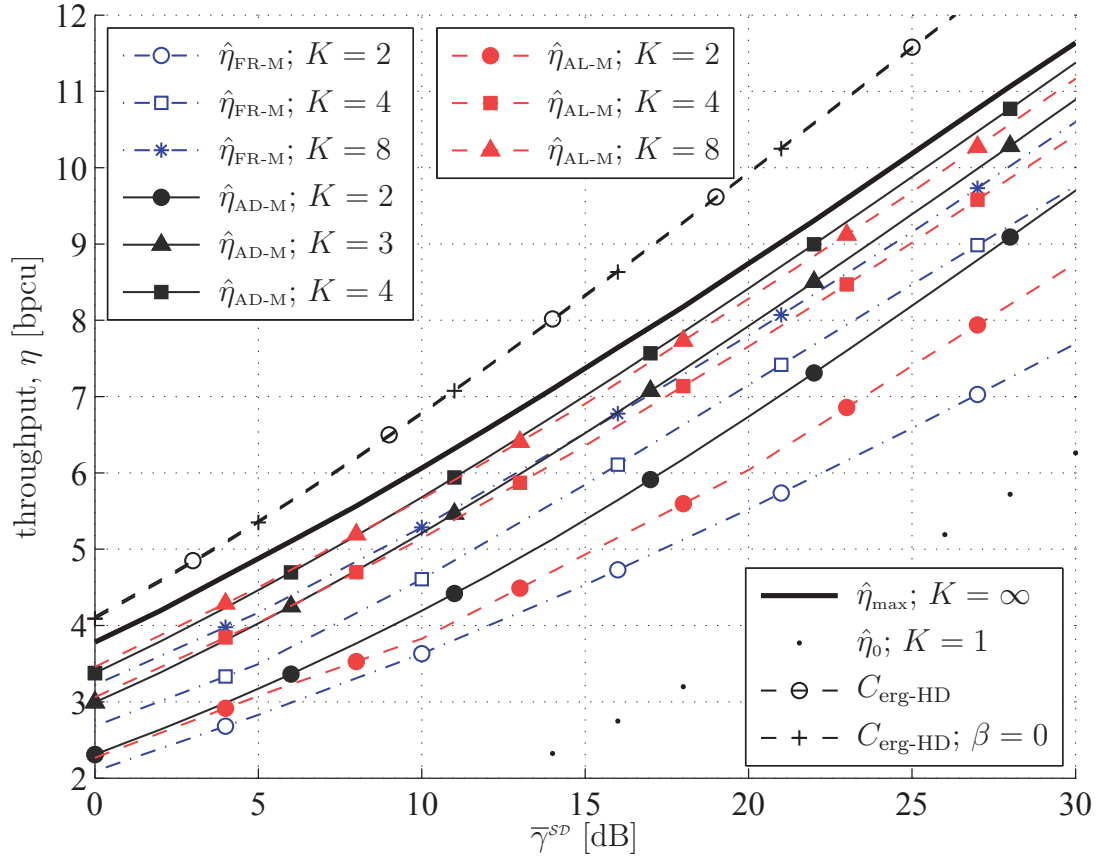


Figure 5.7 Optimal throughput for $K = 2, 3, 4$ of our problem of adaptive-rate (AR) transmission compared with rate allocation (AL) and optimal fixed-rate (FR) throughput; maximum achievable throughput ($K = \infty$) and obvious lower bound of $K = 1$ are shown as well for comparison.

values of K . The results are shown for three different HARQ transmission protocols of:

- Maximum throughput for a multi-hop channel which is achievable with rate adaptation (AD): $\hat{\eta}_{\text{AD-M}}$
- Maximum throughput achievable with fixed-rate (FR) transmission: $\hat{\eta}_{\text{FR-M}}$
- The optimal rate allocation (AL) throughput found using the 2-dimensional state space optimization method presented in Section 5.2 and optimized using `fminsearch` as discussed in Section 5.2.5: $\hat{\eta}_{\text{AL-M}}$

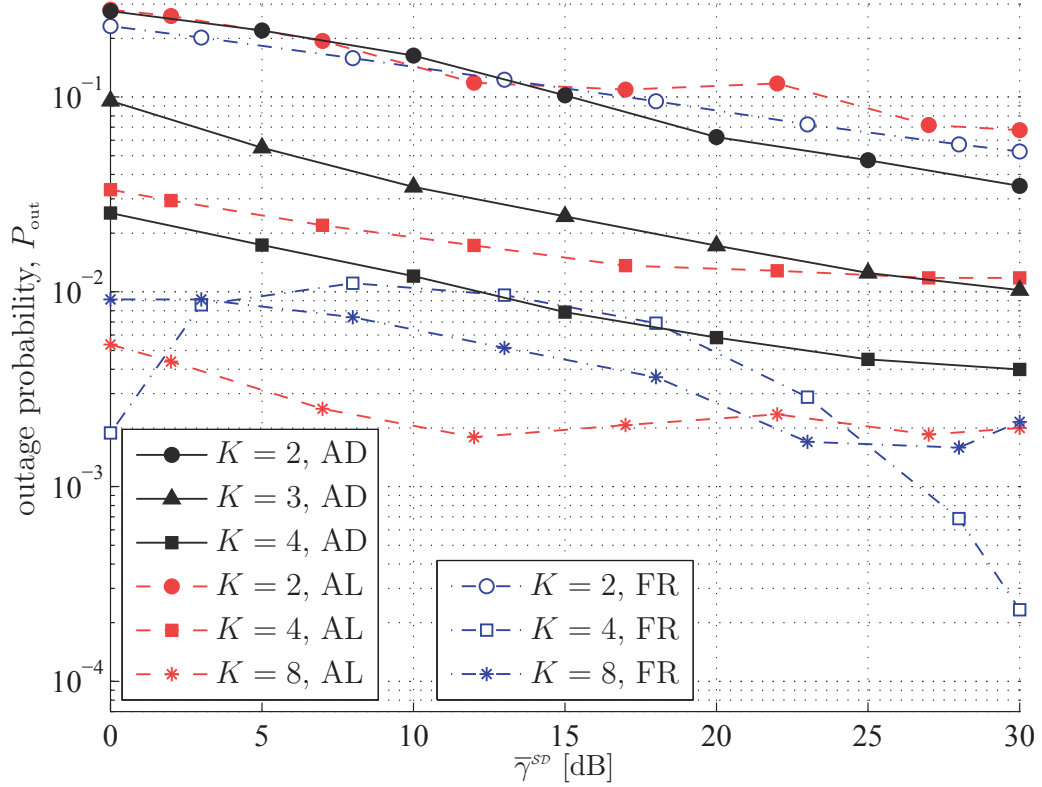


Figure 5.8 Outage Probability for optimal-throughput-achieving policies for $K = 2, 3, 4$ of rate adaptation (AD), rate allocation (AL) and fixed-rate (FR) HARQ transmission.

As shown in Figure 5.7, a significant improvement on the throughput is achievable by rate adaptation HARQ, with transmitting back the outdated CSI on a multi-bit feedback channel.

The upper bound for the maximum achievable throughput $\hat{\eta}_{\max}$, as introduced in Section 5.1 is also shown in Figure 5.7 which is the upper bound to the non-adaptive case as well. One of the most important points depicted in this figure is how fast we can reach to the maximum achievable throughput $\hat{\eta}_{\max}$ by increasing K in truncated HARQ transmission. With $K = 4$ number of rate-adaptive transmissions in high SNR region we get pretty close to the performance limit while this number gets bigger than 8 for both rate allocation and fixed-rate transmission.

The capacity bounds of the relay channel studied in [76], averaged over all channel states

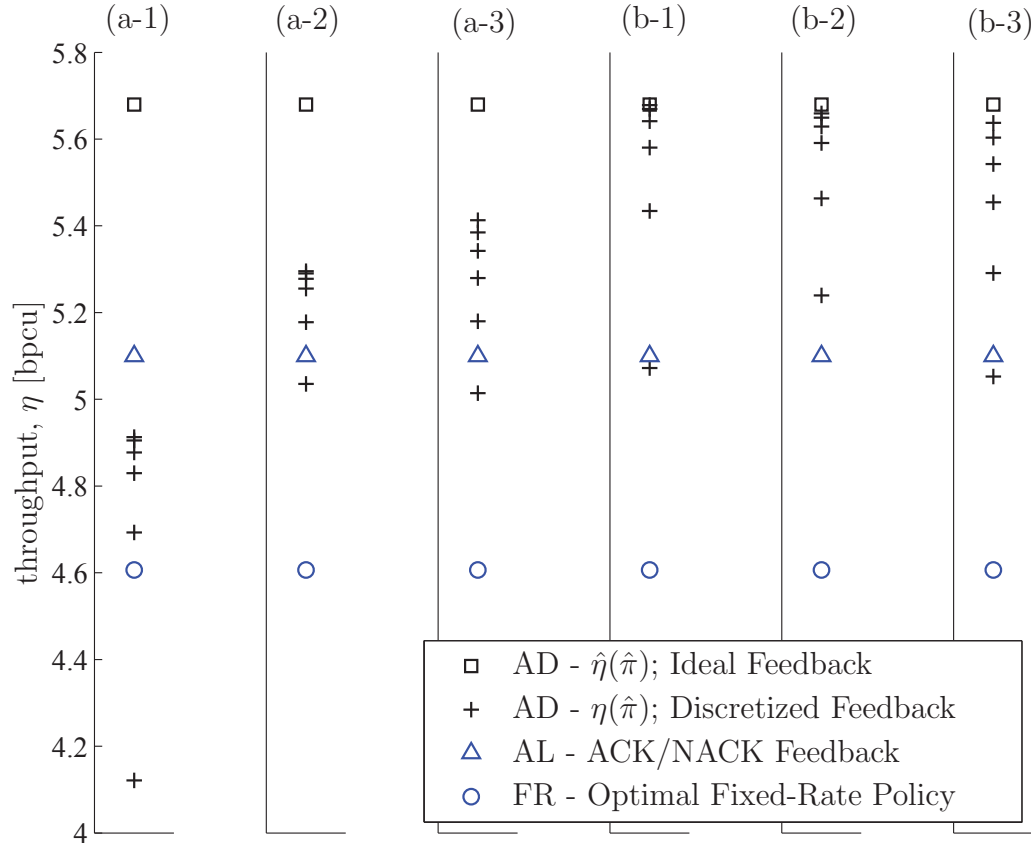


Figure 5.9 Discretized feedback analysis results: throughput of the optimal rate adaptation (AD) policy $\hat{\pi}$ for $\bar{\gamma}^{sd} = 10$ dB and $K = 4$ when the feedback message is discretized with $N_f^p : 2 \rightarrow 7$ (Throughput $\eta(\hat{\pi})$ with discretized feedback increases with N_f^p growing) in 1. Min.; 2. Med.; and 3. Avg. modes of decision making with (a): $N_f^r = 1$ and (b): $N_f^r = N_f^p$. The results are compared with the optimal rate adaptation throughput value for ideal feedback $\hat{\eta}(\hat{\pi})$; the maximum rate allocation (AL) throughput $\tilde{\eta}$; and the throughput of the fixed-rate (FR) transmission.

is also shown in Figure 5.7. A comparison between the low and high SNR regimes in this figure shows that the proposed HARQ transmission scheme has a maximum achievable throughput which is comparable to the capacity of the channel at low SNR. On the other hand, for higher SNR the maximum achievable throughput gap to the capacity increases. Moreover, with the number of retransmissions increasing, the achievable throughput increases significantly as the throughput of the proposed transmission scheme for $K = 3$ is significantly

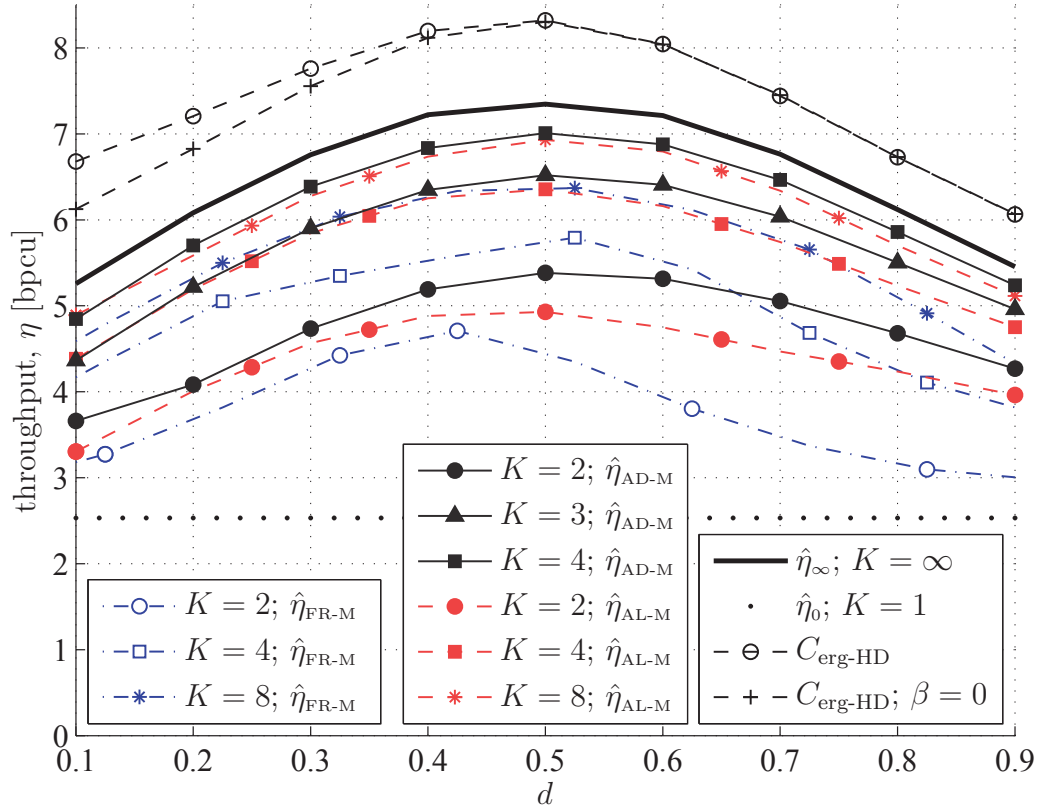


Figure 5.10 Effect of the relay positioning parameter d , on the optimal throughput performed by rate adaptation $\hat{\eta}_{\text{AD-M}}$ and rate allocation $\hat{\eta}_{\text{AL-M}}$ for truncated HARQ. The results are compared to that of the performance limit $\hat{\eta}_{\infty}$ and the ergodic half-duplex capacity limit.

greater than the throughput of an optimal fixed-rate transmission even for $K = 4$. The outage probability corresponding to the optimal policy for different SNR values is shown in Figure 5.8 which shows the SNR gained with increasing K for both rate adaptation and allocation.

Figure 5.9 shows the achievable throughput with limited number of bits of feedback message as discussed in Section 5.1.4. The results are shown for different decision making modes and for different values of N_f^r and N_f^p where by increasing N_f^p the throughput increases. We can see that with $N_f^r = N_f^p = 7$ all the three proposed decision making modes can reach the optimal value of the throughput.

Moreover, in case when the relay node cannot send back more than a single bit of

feedback message, we can still get a significant increase on the throughput with respect to fixed-rate transmission by only a few number of bits fed back from $\mathcal{D}$. With $N_f^R = 1$ the Avg. approach performs comparably better which is the result of packet dropping regions of the state space as shown in Figure 5.1.

Because of the single-bit feedback message in fixed-rate transmission it may not seem a fair comparison to put the optimal throughput in Figure 5.7 next to the fixed-rate maximum throughput. However, as we can see in Figure 5.9, with very few bits of feedback message which is technically feasible, we can still have a higher throughput compared to fixed-rate transmission.

The effect of the relay position on the maximum achievable throughput is also analyzed by changing the position of the relay on the line that connects source node to the destination. The maximum achievable throughput for different number of allowed retransmissions as well as the capacity of the relay channel with respect to the parameter d is shown in Figure 5.10 for a fixed average SNR of $\bar{\gamma}^{SD} = 15$ dB.

5.5 Multi-Hop HARQ Transmission

In a more general form, the communication network may consist of M relay nodes $\mathcal{R}_m$ for $1 \leq m \leq M$ other than the destination $\mathcal{D}$ and the source node denoted as $\mathcal{R}_0 \equiv \mathcal{S}$. In this section we discuss how the throughput analysis can be extended to such a case. The assumption we adopt on top of the system model in Section 2.1.3 is the use of an opportunistic relaying (selection relaying) strategy [77], where at each transmission time k a node with the best channel (closest) to $\mathcal{D}$ will be selected among the nodes that have the message as the next transmitter.

Without loss of generality we assume that $\bar{\gamma}^{R_m \mathcal{D}} < \bar{\gamma}^{R_n \mathcal{D}}$ if and only if $m < n$. Assuming that every node is aware of the distribution of its channel (i.e., physical distance) to all the other nodes in the network, the opportunistic relaying strategy can be managed in either of the following manners [78].

- By having a network of feedback channels among all the nodes: at the end of each transmission attempt, the transmitting node $\mathcal{R}_m$ ($0 \leq m \leq M$) and the receiver relay nodes $\mathcal{R}_n$, $m < n$ will be notified about the state of the decoder of all the receiving nodes (i.e., $\mathcal{R}_n$, $m < n$ and $\mathcal{D}$).

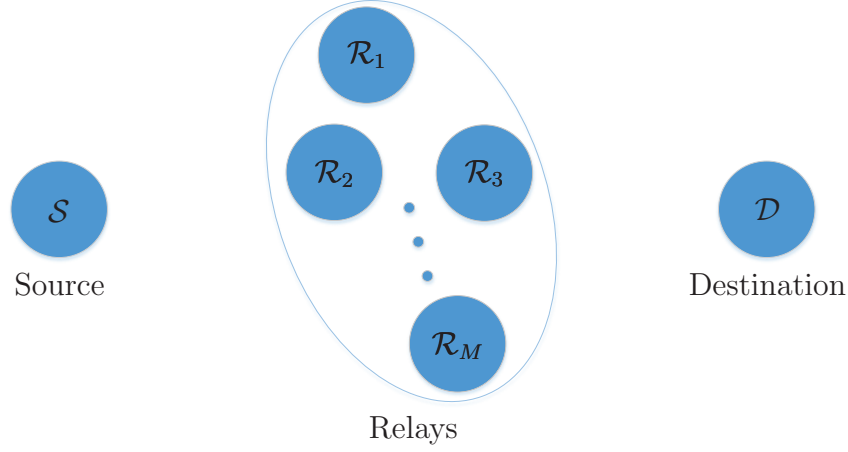


Figure 5.11 Topology of the M -relay network.

- In a centralized manner by node $\mathcal{D}$ assuming that it is being informed about the state of all the decoders.
- By adopting a timer factor t_m for each relay node $\mathcal{R}_m$ where t_m is proportional to the inverse of $h^{\mathcal{R}_m \mathcal{D}}$. At the end of each transmission attempt, if a node $\mathcal{R}_m$ has successfully decoded the message and wants the other nodes to go silent for the next transmission attempt, it has to broadcast an announce message at time t_m except for if it is put to silent earlier by a node with smaller timer factor.

We use the simplified notation of $I_k^{\mathcal{R}_m \rightarrow n} = \{I_k^{\mathcal{R}_m}, I_k^{\mathcal{R}_m+1}, \dots, I_k^{\mathcal{R}_n}\}$ for $1 \leq m < n \leq M$ to represent a set of variables. The redundancy functions can then be shown as follows

$$\rho_k^{\mathcal{S}} = \begin{cases} \rho_k^{\mathcal{S}}(I_{k-1}^{\mathcal{D}}, I_{k-1}^{\mathcal{R}_1 \rightarrow M}) & I_{k-1}^{\mathcal{D}} < 1 \text{ \& } I_{k-1}^{\mathcal{R}_m} < 1 \ \forall m \\ 0 & \text{otherwise} \end{cases} \quad (5.44a)$$

$$\rho_k^{\mathcal{R}_m} = \begin{cases} \rho_k^{\mathcal{R}_m}(I_{k-1}^{\mathcal{D}}, I_{k-1}^{\mathcal{R}_m+1 \rightarrow M}) & I_{k-1}^{\mathcal{D}} < 1 \text{ \& } I_{k-1}^{\mathcal{R}_m} \geq 1 \\ & \text{\& } I_{k-1}^{\mathcal{R}_n} < 1 \text{ for } n > m \ . \\ 0 & \text{otherwise} \end{cases} \quad (5.44b)$$

The optimization process for the M -relay network can follow a similar approach as in (5.2a), (5.2b) and (5.2c). However, every step of this process will have a $M + 1$ dimensional cost function to be minimized with respect to $M + 1$ redundancy variables in (5.44a) and (5.44b). We do not carry out the numerical analysis for $M > 1$ and we note here that with M increasing, the complexity of this optimization problem grows rapidly however, for small M it is still possible to solve the optimization problem using DP.

It might seem possible to simplify the optimization problem by choosing the feedback of all the relay nodes to be a single-bit message, and choose to adapt the transmission rates only based on I^p . However, it should be noted that by assuming the redundancies in (5.44a) and (5.44b) to be one dimensional functions of I^p , the optimization problem does not fit into DP framework any more and as a result it will be a complex optimization. To be able to cast the problem into DP, the state of the process, which is $S_k = (I_k^p, I_k^{r_{1 \rightarrow M}})$ at time k , should be obtained from the state at time $k - 1$ and the decision policy π [24]. The Gaussian approximation that we used in Section 5.2 removes this characteristic from the problem set-up and makes the optimization impossible with DP.

5.6 Summary and Complementary Notes

The feedback network assumed between the three communication parties is supposed to be error-free in all the analysis in this dissertation. However, we note here that this assumption is idealistic in the sense that it needs an infinite amount of information being transmitted back from each receiver node to all the transmitter nodes at the end of each attempt, due to the continuous nature of the accumulated information at the receiver. We refined this into a technically feasible system model by assuming a discretization over the feedback message. In this sense the accumulated information at the receiver would be discretized so that it can be conveyed on a feedback message with a limited number of bits. Decision making on the transmitter side will be then based on a discretized state space.

In this chapter we presented a study on different feedback scenarios. We presented the rate adaptation optimal throughput $\hat{\eta}_{\text{AD-M}}$ as the performance limit that any HARQ transmission protocol is upper bounded by in the same channel model. A general framework was established to find the performance limits of truncated/non-truncated HARQ protocol. Next, we presented the analysis of the channel in a practically more realistic situation of limited feedback messages. It is shown that very small number of bits for the feedback

messages can reach the performance limit discussed above. Moreover, we studied the conventional ACK/NACK feedback case to find the maximum throughput $\tilde{\eta}_{\text{AL-M}}$ we can get by variable rate allocation.

We presented a framework for throughput and outage probability analysis of variable-rate allocation in different channel setups. A closed-form representation of the outage probability and the throughput was presented for truncated HARQ with single-bit feedback message.

Chapter 6

Conclusion and Future Work

In this thesis, we established the groundwork for rate control policies for HARQ in single-hop and cooperative transmission over block fading channels with using outdated CSI and considering fixed-power. As it was depicted in the numerical results, a significant improvement on throughput is achieved over conventional fixed-rate transmission.

We presented a closed-form representation of throughput for the rate allocation scenario where the CSI knowledge (whether outdated or not) is completely absent at the encoder node. Using simplifying approximations the formulations were cast into the recursive optimization to substitute the K -dimensional maximization with several lower dimensioned optimization problems. The presented results showed a dramatic improvement over fixed-rate transmission in both the single-hop and the cooperative channel .

6.1 Concluding Remarks

The objectives of this research, as are listed in Section 1.1.5, were addressed and achieved through the analytical and numerical approaches which are presented in Chapter 4 and Chapter 5. In this section we go through the questions that were raised in Section 1.1.5 and briefly summarize the answers given in this thesis.

What is the throughput-optimal transmission rate in a fading channel when using HARQ protocol?

What is the effect of outdated CSI on throughput of an HARQ channel?

The conventional fixed-rate HARQ transmission is a special case of a general setup which allows for variation of the transmission parameters. Variable-rate transmission provides the encoder with new degrees of freedom to improve the overall performance of the HARQ protocol. Two different scenarios can be considered for a variable-rate transmission: In the first one, the encoder is only provided with the conventional single-bit ACK/NACK feedback. We called this scenario as *rate allocation* throughout this thesis. In the second scenario, a side information is also available at the encoder via the feedback channel. This scenario was addressed as *rate adaptation* in this thesis.

The rate allocation scenario takes advantage of the variable-rate transmission assumption by performing an a priori statistical analysis to find an optimal set of transmission rate values for the truncated HARQ process. The result of this analysis, even with fixed-power transmission, is a significant increase on the overall throughput, in both single-hop and relay channel models. The variable-length coded packets of different messages could be used to fill up a relatively bigger (but fixed-size) frame in order to make the protocol suitable for fixed-frame-size TDMA transmission.

In the rate adaptation scenario, the only information that the encoder can be provided with is the history of the channel. Having this knowledge about the outdated channel experiences, the encoder is able to measure how close the decoder is to a successful decoding. Optimal rate adaptation increases the overall throughput significantly compared to the rate allocation in all the channel models we studied.

The optimal rate policy shows the much interesting conclusion that it is better in terms of throughput to discard a packet in the middle of an HARQ transmission, if the first channel experiences are all “very poor”. We found threshold levels for the decoder state that tells whether a packet is worth more retransmission attempts or is best to be discarded. However, this can never be the case for rate allocation since the encoder only gets informed about the decoding success/failure.

How well a can cooperative channel perform in terms of achievable throughput?

An asymptotically large number of allowed retransmission attempts provides the limit of performance for an HARQ protocol. In a relay channel model, since the capacity is not known in general, this upper limit for achievable throughput is an interesting result that can be found through the rate adaptation scenario. For a point-to-point transmission, the throughput of an infinite number of allowed transmission attempts reaches the ergodic capacity. In the relay channel, depending on the relaying protocol, the maximum achievable throughput found in the rate adaptation scenario, is the upper limit of the fixed-power transmission performance.

Can we get better fixed-outage performance in a block-fading channel by applying optimized HARQ transmission protocol?

Outage constrained optimization of the transmission rate can be exploited to depict a trade-off between the outage probability (error rate) and the channel throughput, where by improving one the other worsens. For values less than outage probability of the optimal rate policy, outage-constrained maximum throughput decreases with the constraint decreasing.

What are the limits of performance for a truncated HARQ?

The throughput of truncated HARQ with K maximum transmission attempts increases with K and, if the incremental redundancy is used, for $K \rightarrow \infty$ the throughput reaches the ergodic capacity. Even a limited number of transmission attempts, e.g., $K = 4$ in the single-hop channel, approaches close to this limit in the rate adaptation scenario. Because of the significant throughput difference between the variable-rate scenarios and the fixed-rate transmission in the proposed channel model, it is easier to approach the capacity with the variable-rate transmission. In other words, a relatively smaller number of transmissions is necessary with variable-rate HARQ.

How robust are the optimal policies in fulfilling the achievable throughput with respect to the limitations on the feedback message?

The exact value of the decoder state has to be fed back to the encoder in the rate adaptation scenario. However, this is only necessary in the theoretical analysis and when it comes

to the numerical optimization, the presented results have already taken into account the discretization of the feedback message. This is because all the numbers on a digital machine are discretized to a limited number of bits and the decoder state parameter in our analysis goes through this discretization too. Moreover, as we showed in Section 5.1.4, only a small number of feedback bits (7 bits in one example of a relay channel) is sufficient for an encoder that follows an optimal rate adaptation policy, to reach the maximum achievable throughput. In other words, even the discretization precision of a 64bit processor computer machine is more than enough for the rate adaptation scenario and the assumption of an infinite bit feedback can be easily relaxed.

Complexity of optimization

Non-convex optimization problems have a complexity that grows exponentially with the number of optimization parameters. This might result in very time-consuming optimization. While there are no general methods to overcome the complexity of non-convex optimization, one idea is to approximate the problem with an approximation which is simple to solve and use the solution as a starting point to globally optimize the original problem. In this dissertation, as presented in Section 5.2.5, we showed that a solution to the approximated version of the throughput optimization problem can be simply optimized using the interior-point optimization tool `fminsearch`. We showed that this optimized solution which can be found in less than one hour of time on a normal machine for $K = 8$ transmission attempts, can not be reached even by trying 2000 random starting points in the same optimization tool which may take several weeks to be done.

6.2 Future Directions

There are several avenues for future work following this dissertation. The perspectives driven from this research work mostly revolve around three main topics:

- Adding the power parameter to the optimization problems
- Analyzing HARQ transmission over other channel models
- Considering practical coding/modulation techniques

6.2.1 Transmission Power

As already mentioned in Section 3.3, the optimization problem with variable transmission rate and power parameters, is computationally more complex. The problem has been addressed in the literature [58,61] for particular cases however, to the best of our knowledge the performance limits of truncated HARQ with variable rate and power parameters is still unknown. By adding the power parameters, new constraints will show up in solving the non-linear throughput optimization problem due to the limits on the average total transmitted power.

6.2.2 Channel Model

The problem of rate adaptation for a communication network with M number of relay nodes was introduced in Section 5.5. We showed that the optimization problem, even though it can be formulated easily, is very difficult to solve. However, exploiting simplifying assumptions both in the relaying protocol and in outage probability calculation, might be helpful in reformulating the optimization problem.

The orthogonal-relaying channel model in Chapter 2 assumes that the relay node is fully dedicated to assist the source node with delivering its' data to the destination. A different approach to the channel model is to assume that all the cooperating nodes have their very own data that they want to deliver to the destination node. This problem is studied in [79] for the case of two cooperating transmitter nodes. It is assumed that each node can use superposition modulation to send their own data along with relaying the other transmitter node's data. The work in [79] finds the optimal superposition ratio for two HARQ transmission attempts however, the performance limits of such a network is still an open problem.

6.2.3 Coding/Modulation

Cross-layer optimization of HARQ transmission is a popular topic in the literature [59,61, 80–82]. The goal is to exploit the retransmission characteristics of the HARQ protocol in designing a communication system where a desirable latency and buffer overflow rate is guaranteed for the least use of the bandwidth. The idea of variable transmission parameters is a suitable idea to increase the performance of the HARQ protocol in such a design system.

Appendix A

Proofs for Propositions and Theorems

Proof of Proposition 1

Assume that for λ , where $\lambda < 1/\hat{\eta}$, there exist a non-degenerate policy π that $\pi \in \tilde{R}_\lambda$. Then according to (4.10) we have,

$$\mathbf{D}(\pi) + \lambda \cdot P_{\text{out}}(\pi) \leq \lambda \tag{A.1}$$

where, since from the assumption we have that $\mathbf{D}(\pi) > 0$, this inequality results in the following.

$$\frac{1}{\lambda} \leq \frac{1 - P_{\text{out}}(\pi)}{\mathbf{D}(\pi)} \leq \hat{\eta}, \tag{A.2}$$

which contradicts with the assumption.

Proof of Proposition 2

From (4.6) and (4.7) we can conclude that

$$\mathbf{D}(\pi) + \lambda \cdot (P_{\text{out}}(\pi) - \epsilon) \leq \mathbf{D}_\epsilon, \quad \forall \pi \in \tilde{R}_\lambda, \quad \forall \epsilon \in [0, 1). \tag{A.3}$$

Now, assume that for λ , where $\lambda > 1/\hat{\eta}$, there exist a degenerate policy π that $\pi \in \tilde{R}_\lambda$.

From (A.3), we have

$$\lambda \cdot (1 - \epsilon) \leq \mathbf{D}_\epsilon, \quad \forall \epsilon \in [0, 1) \quad (\text{A.4})$$

and,

$$\frac{1 - \epsilon}{\mathbf{D}_\epsilon} \leq \frac{1}{\lambda} < \hat{\eta}, \quad \forall \epsilon \in [0, 1), \quad (\text{A.5})$$

which contradicts with $\frac{1-\hat{\epsilon}}{\mathbf{D}_{\hat{\epsilon}}} = \hat{\eta}$. This completes the proof to the proposition.

Proof of Proposition 3

From Proposition 1 and (4.10) one can easily arrive at the conclusion that

$$\mathbf{D}(\pi) + \lambda \cdot P_{\text{out}}(\pi) = \lambda, \quad \forall \pi \in \tilde{\mathbf{R}}_\lambda, \quad \text{and} \quad \forall \lambda < \lambda_{\text{th}} \quad (\text{A.6})$$

First we want to show that (A.6) holds also for $\lambda = \lambda_{\text{th}}$. Assume that $\mathbf{D}(\pi) + \lambda_{\text{th}} \cdot P_{\text{out}}(\pi) < \lambda_{\text{th}}$ for any arbitrary non-degenerate policy π (this assumption is never valid for a degenerate π). This will result in $\frac{1-P_{\text{out}}(\pi)}{\mathbf{D}(\pi)} > 1/\lambda_{\text{th}} = \hat{\eta}$ which contradicts with the definition of $\hat{\eta}$. So, following (4.10) we have

$$\mathbf{D}(\pi) + \lambda_{\text{th}} \cdot P_{\text{out}}(\pi) = \lambda_{\text{th}}, \quad \forall \pi \in \tilde{\mathbf{R}}_{\lambda_{\text{th}}} \quad (\text{A.7})$$

From (A.7), any degenerate policy π can be in $\tilde{\mathbf{R}}_{\lambda_{\text{th}}}$. Also, any $\pi \in \hat{\mathbf{R}}$ satisfies (A.7) and as a result it can be a solution to $J^{\lambda_{\text{th}}}$.

Proof of Theorem 1

From Proposition 3, we have $\hat{\mathbf{R}} \subset \tilde{\mathbf{R}}_{\lambda_{\text{th}}}^{\text{non-deg}}$, and

$$\eta(\pi) = 1/\lambda_{\text{th}} = \hat{\eta}, \quad \forall \pi \in \tilde{\mathbf{R}}_{\lambda_{\text{th}}}^{\text{non-deg}} \quad (\text{A.8})$$

therefore, $\tilde{\mathbf{R}}_{\lambda_{\text{th}}}^{\text{non-deg}} \subset \hat{\mathbf{R}}$ which yields $\tilde{\mathbf{R}}_{\lambda_{\text{th}}}^{\text{non-deg}} \equiv \hat{\mathbf{R}}$.

Proof of Proposition 4

Since $\pi \in \tilde{R}_\lambda$, from (4.7) we have

$$\mathbf{D}(\pi) + \lambda \cdot (P_{\text{out}}(\pi) - \epsilon) \leq \mathbf{D}_\epsilon \quad (\text{A.9})$$

or,

$$\mathbf{D}(\pi) \leq \mathbf{D}_\epsilon \quad (\text{A.10})$$

which follows to $\mathbf{D}(\pi) = \mathbf{D}_\epsilon$ according to (4.5). Therefore, since for any $\pi^* \in \mathring{R}_{\mathbf{D}_\epsilon}$, $\mathbf{D}(\pi^*) = \mathbf{D}_\epsilon$ then, $\pi^* \in \tilde{R}_\lambda$.

Proof of Proposition 5

J^λ increases monotonically by growing λ (strictly positive first derivative with respect to λ). So, according to the Proposition 4, for any arbitrary $\lambda \geq \lambda_{\text{th}}$ there is one and only one $\epsilon \in (0, \hat{\epsilon}]$ where $\tilde{R}_\lambda \subset \mathring{R}_{\mathbf{D}_\epsilon}$, because otherwise there is $\hat{\pi} \in \tilde{R}_\lambda$ where $P_{\text{out}}(\hat{\pi}) = \hat{\epsilon} \neq \epsilon$ which results in $\mathring{R}_{\mathbf{D}_\epsilon} \subset \tilde{R}_\lambda$ and $J^\lambda = J^{\hat{\lambda}}$ where $\lambda \neq \hat{\lambda}$ and this is a contradiction.

Appendix B

Throughput of Cooperative Variable-Rate HARQ Transmission

All the success events in (5.20a) and (5.20b) are mutually exclusive. This can be proved easily if we notice that the chance of two success events happening is zero. The same way we can show that the success events and the failure events in (5.19a) and (5.19b) are disjoint too.

Probability of a failure event E_K^* can be represented using (5.21a)–(5.21c). For instance from (5.19b) we have the probability of event E_K^* as follows.

$$\begin{aligned}\Pr\{E_K^*\} &= \Pr\left\{\sum_{k=1}^{K-1} \nu_k^{SR} < 1\right\} \cdot \Pr\left\{\sum_{k=1}^K \nu_k^{SD} < 1\right\} \\ &= P_K^{SD} \cdot P_{K-1}^{SR}\end{aligned}\tag{B.1}$$

For any two random events A and B we know that $P(A \cap B) = P(A) - P(A \cap B^c)$, where B^c is the complement the event B (not B , i.e., the event that B does not occur). This gives us the $\Pr\{E_l^*\}$ for $1 \leq l < K$ as,

$$\begin{aligned}\Pr\{E_l^*\} &= \Pr\left\{\sum_{i=1}^{l-1} \nu_i^{SR} < 1 \wedge \sum_{i=1}^l \nu_i^{SD} + \sum_{i=l+1}^K \nu_{l,i}^{\mathcal{RD}} < 1\right\} \\ &\quad - \Pr\left\{\sum_{i=1}^l \nu_i^{SR} < 1 \wedge \sum_{i=1}^l \nu_i^{SD} + \sum_{i=l+1}^K \nu_{l,i}^{\mathcal{RD}} < 1\right\},\end{aligned}\tag{B.2}$$

which results in the following.

$$\Pr\{E_l^*\} = [P_{l-1}^{SR} - P_l^{SR}] \cdot P_{l,K}^{SRD} \quad (\text{B.3})$$

The same way, we can find the probability of the success events as follows.

$$\Pr\{E_k\} = [P_{k-1}^{SD} - P_k^{SD}] \cdot P_{k-1}^{SR} \quad (\text{B.4})$$

$$\Pr\{E_{l,k}\} = \begin{cases} [P_{l-1}^{SR} - P_l^{SR}] \cdot [P_{k-1}^{SD} - P_{l,k}^{SRD}] & k = l + 1 \\ [P_{l-1}^{SR} - P_l^{SR}] \cdot [P_{l,k-1}^{SRD} - P_{l,k}^{SRD}] & k > l + 1 \end{cases} \quad (\text{B.5})$$

An outage in message delivery in the transmission process can happen due to any of the failure events $E_1^*, \dots, E_K^*$. Therefore, the outage probability can be shown as follows.

$$P_{\text{out}} = \Pr\{\cup_{k=1}^K E_k^*\} \quad (\text{B.6})$$

Because the failure events are mutually exclusive, (B.6) can be shown as,

$$P_{\text{out}} = \sum_{k=1}^K \Pr\{E_k^*\}. \quad (\text{B.7})$$

Substituting (B.1) and (B.3) in (B.7) gives us (5.24).

The expected number of channel uses $\bar{N}_s$ in (2.15), is the expectation over the number of channel uses of all the possible events. Thus it can be shown as follows.

$$\bar{N}_s = N_b \cdot \left(\sum_{k=1}^K \Pr\{E_k\} \cdot p_k + \sum_{l=1}^{K-1} \sum_{k=l+1}^K \Pr\{E_{l,k}\} \cdot p_{l,k} + \sum_{k=1}^{K-1} \Pr\{E_k^*\} \cdot p_{k,M} \right), \quad (\text{B.8})$$

where,

$$p_k = \sum_{i=1}^k \rho_k^s \quad (\text{B.9})$$

and,

$$p_{l,k} = \sum_{i=1}^l \rho_i^S + \sum_{i=l+1}^k \rho_{l,i}^R. \quad (\text{B.10})$$

Substituting (B.1), (B.3), (B.5) and (B.5) in (B.8) gives us (5.25).

One can easily investigate the fact that all success and failure events create a set of disjoint events where the sum of their probabilities equals 1. This is shown in the following.

$$\sum_{k=1}^K (\Pr\{E_k\} + \Pr\{E_k^*\}) + \sum_{l=1}^{K-1} \sum_{k=l+1}^K \Pr\{E_{l,k}\} = 1. \quad (\text{B.11})$$

Appendix C

DP Recursive Optimization for Rate Allocation in Relay Channel

Using the approximate failure probabilities, the minimization problem in (4.8) can be follows.

$$\begin{aligned}\tilde{J}^\lambda &= \min_{\pi} \{ \tilde{\mathbf{D}}(\pi) + \lambda \cdot \tilde{P}_{\text{out}}(\pi) \} \\ &= \min_{\pi} \left\{ \sum_{i=1}^{K-1} [\rho_i^s \cdot \tilde{P}_{i-1}^{SD} \cdot \tilde{P}_{i-1}^{SR}] + \tilde{f}_i \cdot \tilde{g}_i^\lambda + \lambda \cdot \tilde{P}_K^{SD} \cdot \tilde{P}_{K-1}^{SR} + \rho_K^s \cdot \tilde{P}_{K-1}^{SD} \cdot \tilde{P}_{K-1}^{SR} \right\},\end{aligned}\quad (\text{C.1})$$

where,

$$\tilde{f}_i = \tilde{P}_{i-1}^{SR} - \tilde{P}_i^{SR} \quad (\text{C.2})$$

and,

$$\tilde{g}_i^\lambda = \lambda \cdot \tilde{P}_{i,K}^{SRD} + \sum_{l=i+2}^K \rho_{i,l}^R \cdot \tilde{P}_{i,l-1}^{SRD} + \rho_{i,i+1}^R \cdot \tilde{P}_i^{SD}. \quad (\text{C.3})$$

It is noticeable here that any solution to (C.1) is not necessarily the optimal solution to (4.8). However, the fact that (C.1) can be solved via DP, makes it a interesting problem for the purpose of throughput optimization.

The function $\tilde{g}_i^\lambda$ is important here because we can optimize it based on the set of

parameters ϱ_i^R , if the two summations of $\sum_{k=1}^i \rho_k^S$ and $\sum_{k=1}^i (\rho_k^S)^2$ are given. Therefore, we start with the term $\tilde{g}_i^\lambda$ which is nested inside of the optimization function in (C.1), and define the minimization function of $\tilde{g}_i^\lambda$ as follows.

$$V^{\lambda,i}(\alpha, \beta) = \min_{\substack{\rho_{i,l}^R \in \varrho_i^R \\ \sum_{k=1}^i \rho_k^S = \alpha, \sum_{k=1}^i (\rho_k^S)^2 = \beta}} \{ \tilde{g}_i^\lambda \}, \quad (\text{C.4})$$

$V^{\lambda,i}(\alpha, \beta)$ can be solved recursively as we will show later, and stored as a pre-stage for the optimization of (C.1). But first we show how the results of the minimization in (C.4) can be used in a *nested loop* minimization problem of (C.1). Using (C.4) we can rewrite (C.1) as in (C.5).

$$\begin{aligned} \tilde{J}^\lambda = \min_{\pi} \bigg\{ & \sum_{i=1}^{K-1} [\rho_i^S \cdot \tilde{P}_{i-1}^{SD} \cdot \tilde{P}_{i-1}^{SR}] + f_i \cdot V^{\lambda,i} \left(\sum_{k=1}^i \rho_k^S, \sum_{k=1}^i (\rho_k^S)^2 \right) \\ & + \lambda \cdot \tilde{P}_K^{SD} \cdot \tilde{P}_{K-1}^{SR} + \rho_K^S \cdot \tilde{P}_{K-1}^{SD} \cdot \tilde{P}_{K-1}^{SR} \bigg\}. \end{aligned} \quad (\text{C.5})$$

The minimization problem in (C.5) can be solved recursively using a two-dimensional state $S_k = (X_k, Y_k)$ and finding $\tilde{J}^\lambda = \tilde{J}_1^\lambda(X_0, Y_0)|_{(X_0, Y_0)=(0,0)}$, where,

$$\tilde{J}_1^\lambda(X_0, Y_0) = \min_{\rho_1^S} \{ \tilde{J}_2^\lambda(X_0 + \rho_1^S, Y_0 + (\rho_1^S)^2) + \rho_1^S \} \quad (\text{C.6})$$

$$+ f_1 \cdot V^{\lambda,1}(X_0 + \rho_1^S, Y_0 + (\rho_1^S)^2) \}, \quad (\text{C.7})$$

and $J_k^\lambda(X_{k-1}, Y_{k-1})$ for $1 < k < K$ and for $k = K$, are shown respectively in (C.8a) and (C.8b) below.

$$\begin{aligned} \tilde{J}_k^\lambda(X_{k-1}, Y_{k-1}) = \min_{\rho_k^S} \bigg\{ & \tilde{J}_{k+1}^\lambda(X_{k-1} + \rho_k^S, Y_{k-1} + (\rho_k^S)^2) + \rho_k^S \cdot \tilde{P}_{k-1}^{SD} \cdot \tilde{P}_{k-1}^{SR} \\ & + f_k \cdot V^{\lambda,k}(X_{k-1} + \rho_k^S, Y_{k-1} + (\rho_k^S)^2) \bigg\} \end{aligned} \quad (\text{C.8a})$$

$$\tilde{J}_K^\lambda(X_{K-1}, Y_{K-1}) = \min_{\rho_K^S} \{ \rho_K^S \cdot \tilde{P}_{K-1}^{SD} \cdot \tilde{P}_{K-1}^{SR} + \lambda \cdot \tilde{P}_K^{SD} \cdot \tilde{P}_{K-1}^{SR} \} \quad (\text{C.8b})$$

The recursive optimization starts with (C.8b) to find the function $\tilde{J}_K^\lambda$ and continues going backward on k up to $k = 1$. After finding $\tilde{J}_1^\lambda$, the optimal ϱ^S will be found starting with ρ_1^S as follows with $(\hat{X}_0, \hat{Y}_0) = (0, 0)$.

1. $\tilde{\rho}_1^s = \arg_{\rho} \tilde{J}_1^{\lambda}(\hat{X}_0, \hat{Y}_0)$
2. for $k = 2 \rightarrow K$
 - $\hat{X}_{k-1} = \hat{X}_{k-2} + \tilde{\rho}_{k-1}^s$ and $\hat{Y}_{k-1} = \hat{Y}_{k-2} + (\tilde{\rho}_{k-1}^s)^2$
 - $\tilde{\rho}_k^s = \arg_{\rho} \tilde{J}_k^{\lambda}(\hat{X}_{k-1}, \hat{Y}_{k-1})$

All the steps for this recursive optimization can be assuming given $V^{\lambda,k}$ for $1 \leq k \leq K-1$. On the other hand, calculation of $V^{\lambda,k}$ in (C.4) can be done recursively using a nested state of $\mathbf{s}_i = (x'_i, y'_i)$. This can be shown as follows:

$$V^{\lambda,i}(\alpha, \beta) = V_{i+1}^{\lambda,i}(x'_{i+1}, y'_{i+1}, \alpha, \beta)|_{(x'_{i+1}, y'_{i+1})=(0,0)},$$

where $V_{i+k}^{\lambda,i}(x'_{i+k}, y'_{i+k}, \alpha, \beta)$ for $k = 1, 1 < k < K-i$ and $k = K-i$ are shown respectively in (C.9a), (C.9b) and (C.9c).

$$V_{i+1}^{\lambda,i}(x'_{i+1}, y'_{i+1}, \alpha, \beta) = \min_{\rho_{i,i+1}^{\mathcal{R}}} \left\{ \rho_{i,i+1}^{\mathcal{R}} \cdot \tilde{P}_i^{SD} + V_{i+2}^{\lambda,i}(x'_{i+1} + \rho_{i,i+1}^{\mathcal{R}}, y'_{i+1} + (\rho_{i,i+1}^{\mathcal{R}})^2, \alpha, \beta) \right\} \quad (\text{C.9a})$$

$$V_{i+k}^{\lambda,i}(x'_{i+k}, y'_{i+k}, \alpha, \beta) = \min_{\rho_{i,i+k}^{\mathcal{R}}} \left\{ \rho_{i,i+k}^{\mathcal{R}} \cdot \tilde{P}_{i,i+k-1}^{SRD} + V_{i+k+1}^{\lambda,i}(x'_{i+k} + \rho_{i,i+k}^{\mathcal{R}}, y'_{i+k} + (\rho_{i,i+k}^{\mathcal{R}})^2, \alpha, \beta) \right\} \quad (\text{C.9b})$$

$$V_K^{\lambda,i}(x'_K, y'_K, \alpha, \beta) = \min_{\rho_{i,K}^{\mathcal{R}}} \left\{ \rho_{i,K}^{\mathcal{R}} \cdot \tilde{P}_{i,K-1}^{SRD} + \lambda \cdot \tilde{P}_{i,K}^{SRD} \right\} \quad (\text{C.9c})$$

This will be solved starting from (C.9c) and ending with (C.9a) considering $\sum_{k=1}^i \rho_k^s = \alpha$ and $\sum_{k=1}^i (\rho_k^s)^2 = \beta^2$. Then the set of $\rho_{i,l}^{\mathcal{R}}$ $i < l \leq K$ will be found starting with $\rho_{i,i+1}^{\mathcal{R}}$ using (C.9a) with $(x'_{i+1}, y'_{i+1}) = (0, 0)$ and going up to $\rho_{i,K}^{\mathcal{R}}$ in (C.9c) recursively.

Appendix D

One Dimensional State Optimization

The minimization problem in (4.8), using the approximation probabilities in (5.29) and (5.30), becomes the following.

$$\begin{aligned} \check{J}^\lambda &= \min_{\pi} \{ \check{\mathbf{D}}(\pi) + \lambda \cdot \check{P}_{\text{out}}(\pi) \} \\ &= \min_{\pi} \left\{ \sum_{i=1}^{K-1} [\rho_i^S \cdot \check{P}_{i-1}^{SD} \cdot \check{P}_{i-1}^{SR}] + \check{f}_i \cdot \check{g}_i^\lambda + \lambda \cdot \check{P}_K^{SD} \cdot \check{P}_{K-1}^{SR} + \rho_K^S \cdot \check{P}_{K-1}^{SD} \cdot \check{P}_{K-1}^{SR} \right\}, \end{aligned} \quad (\text{D.1})$$

where,

$$\check{f}_i = \check{P}_{i-1}^{SR} - \check{P}_i^{SR} \quad (\text{D.2})$$

and,

$$\check{g}_i^\lambda = \lambda \cdot \check{P}_{i,K}^{SRD} + \sum_{l=i+2}^K \rho_{i,l}^R \cdot \check{P}_{i,l-1}^{SRD} + \rho_{i,i+1}^R \cdot \check{P}_i^{SD}. \quad (\text{D.3})$$

With the same approach as in Appendix C, we first start with minimizing $\check{g}_i^\lambda$ as follows.

$$U^{\lambda,i}(X, x') = \min_{\substack{\rho_{i,k}^R \in \mathcal{Q}_i^R \\ \sum_{l=i+1}^K \rho_{i,l}^R = x' \\ \sum_{k=1}^i \rho_k^S = X}} g_i^\lambda \quad (\text{D.4})$$

This will be in order to find the following function for different X values:

$$U^{\lambda,i}(X, \check{x}'), \quad (\text{D.5})$$

where

$$\check{x}' = \arg_{x'} \min_{x'} U^{\lambda,i}(X, x'). \quad (\text{D.6})$$

The minimization in (D.4) can be done as follows.

$$U^{\lambda,i}(X, x') = U_K^{\lambda,i}(X, x') = \min_{\substack{\rho_{i,i+1}^R, \dots, \rho_{i,K}^R \\ \sum_{l=i+1}^K \rho_{i,l}^R = x' \\ \sum_{k=1}^i \rho_k^S = X}} g_i^\lambda \quad (\text{D.7})$$

$$= \min_{\rho_{i,K}^R} \min_{\substack{\rho_{i,i+1}^R, \dots, \rho_{i,K-1}^R \\ \sum_{l=i+1}^{K-1} \rho_{i,l}^R = x' - \rho_{i,K}^R \\ \sum_{k=1}^i \rho_k^S = X}} g_i^\lambda \quad (\text{D.8})$$

$$= \min_{\rho = \rho_{i,K}^R} U_{K-1}^{\lambda,i}(X, x' - \rho) + \rho \cdot \check{P}^{SRD}(X, x' - \rho) + \lambda \cdot \check{P}^{SRD}(X, x'), \quad (\text{D.9})$$

where for $i + 3 \leq k \leq K - 1$ we have (D.10) as

$$U_k^{\lambda,i}(X, x') = \min_{\rho} U_{k-1}^{\lambda,i}(X, x' - \rho) + \rho \cdot \check{P}^{SRD}(X, x' - \rho), \quad (\text{D.10})$$

and for $k = i + 2$, we have (D.11) as

$$U_{i+2}^{\lambda,i}(X, x') = \min_{\rho} (x' - \rho) \cdot \check{P}^{SD}(X) + \rho \cdot \check{P}^{SRD}(X, x' - \rho). \quad (\text{D.11})$$

The minimization process starts with (D.11) and then goes on with (D.10) and ends with (D.9). The optimization results are stored as $\rho_{i,k}^R(X, x') = \arg_{\rho} U_k^{\lambda,i}(X, x')$. Thus, after finding $\check{x}'$ according to (D.6), we find the optimal set of $\check{\rho}_{i,k}^R(X)$, step-by-step as follows.

1. $\check{\rho}_{i,K}^R(X) = \rho_{i,K}^R(\check{x}')$
2. for $k : K - 1 \rightarrow i + 2$
 - $\check{x}' \leftarrow (\check{x}' - \check{\rho}_{i,k+1}^R)$

- $\check{\rho}_{i,k}^R = \rho_{i,k}^R(\check{x}'_k)$

3. $\check{\rho}_{i,i+1}^R = \check{x}' - \check{\rho}_{i,i+2}^R$

The next step is to find $\check{J}^\lambda$ where $\check{J}^\lambda = \check{J}_K^\lambda(\check{X}_K)$ and,

$$\check{X} = \arg_X \min J_K^\lambda(X). \quad (\text{D.12})$$

To find the $\check{J}_K^\lambda(\check{X}_K)$ function, we have the following.

$$J_K^\lambda(X) = \min_{\substack{\rho_1^S, \dots, \rho_K^S \\ \sum_{k=1}^K \rho_k^S = X}} \left\{ \sum_{i=1}^{K-1} [\rho_i^S \cdot \check{P}_{i-1}^{SD} \cdot \check{P}_{i-1}^{SR}] + \check{f}_i \cdot U^{\lambda,i}(X, \check{x}') \right. \\ \left. + \lambda \cdot \check{P}_K^{SD} \cdot \check{P}_{K-1}^{SR} + \rho_K^S \cdot \check{P}_{K-1}^{SD} \cdot \check{P}_{K-1}^{SR} \right\} \quad (\text{D.13})$$

This can be made into a recursive form as follows.

$$J_K^\lambda(X) = \min_{\rho_K^S} \min_{\substack{\rho_{i,1}^S, \dots, \rho_{K-1}^S \\ \sum_{k=1}^{K-1} \rho_k^S = X - \rho_K^S}} \left\{ \sum_{i=1}^{K-1} [\rho_i^S \cdot \check{P}_{i-1}^{SD} \cdot \check{P}_{i-1}^{SR}] + \check{f}_i \cdot U^{\lambda,i}(X, \check{x}') \right. \\ \left. + \lambda \cdot \check{P}_K^{SD} \cdot \check{P}_{K-1}^{SR} + \rho_K^S \cdot \check{P}_{K-1}^{SD} \cdot \check{P}_{K-1}^{SR} \right\} \\ = \min_{\rho = \rho_K^S} J_{K-1}^\lambda(X - \rho) + \rho \cdot \check{P}^{SD}(X - \rho) \cdot \check{P}^{SR}(X - \rho) + \lambda \cdot \check{P}^{SD}(X) \cdot \check{P}^{SR}(X - \rho) \quad (\text{D.14})$$

For $3 \leq k \leq K-1$ we (D.16) as follows,

$$J_k^\lambda(X) = \min_{\rho} J_{k-1}^\lambda(X - \rho) + \rho \cdot \check{P}^{SD}(X - \rho) \cdot \check{P}^{SR}(X - \rho) + \check{f}(X, \rho) \cdot U^{\lambda,k}(X, \check{x}'), \quad (\text{D.16})$$

and for $k=2$ we have the following.

$$J_2^\lambda(X) = \min_{\rho} \left\{ (X - \rho) + [1 - \check{P}_{X-\rho}^{SR}] \cdot U^{\lambda,1}(X - \rho, \check{x}') \right. \\ \left. + \rho \cdot \check{P}^{SD}(X - \rho) \cdot \check{P}^{SR}(X - \rho) + \check{f}(X, \rho) \cdot U^{\lambda,2}(X, \check{x}') \right\}, \quad (\text{D.17})$$

and according to (D.2), $\check{f}(X, \rho) = \check{P}^{SR}(X - \rho) - \check{P}^{SR}(X)$.

The minimization process starts with (D.17) and then goes on with (D.16) and ends

with (D.15). The optimization results are stored as $\rho_k^s(X) = \arg_{\rho} J_k^{\lambda}(X)$. Then, to find the optimal set of $\check{\rho}_k^s$ we go through the following steps.

1. $\check{\rho}_K^s = \rho_K^s(\check{X})$
2. for $k : K - 1 \rightarrow 2$
 - $\check{X} \leftarrow (\check{X} - \check{\rho}_{k+1}^s)$
 - $\check{\rho}_k^s = \rho_k^s(\check{X})$
3. $\check{\rho}_1^s = \check{X} - \check{\rho}_2^s$.

Bibliography

- [1] G. Caire and D. Tuninetti, “The throughput of hybrid-ARQ protocols for the Gaussian collision channel,” *IEEE Trans. Inf. Theory*, vol. 47, pp. 1971–1988, July 2001.
- [2] A. Goldsmith, *Wireless Communications*. Cambridge University Press, Aug. 2007.
- [3] P. Pagani and P. Pajusco, “Characterization and modeling of temporal variations on an ultrawideband radio link,” *IEEE Trans. Antennas Propag.*, vol. 54, pp. 3198–3206, Nov. 2006.
- [4] R. A. Berry, *Power and Delay Trade-offs in Fading Channels*. PhD thesis, Massachusetts Institute of Technology, Cambridge, USA, June 2000.
- [5] S. Nanda, R. Walton, J. Ketchum, M. Wallace, and S. Howard, “A high-performance MIMO OFDM wireless LAN,” *IEEE Commun. Mag.*, vol. 43, pp. 101–109, Feb. 2005.
- [6] A. Ghosh, D. Wolter, J. Andrews, and R. Chen, “Broadband wireless access with WiMax/802.16: current performance benchmarks and future potential,” *IEEE Commun. Mag.*, vol. 43, pp. 129–136, Feb. 2005.
- [7] K. D. Nguyen, *Adaptive transmission for block-fading channels*. PhD thesis, University of South Australia, Australia, Dec. 2009.
- [8] E. Biglieri, J. Proakis, and S. Shamai, “Fading channels: information-theoretic and communications aspects,” *IEEE Trans. Inf. Theory*, vol. 44, pp. 2619–2692, Oct. 1998.
- [9] S. Lin and D. J. Costello, *Error Control Coding*. Pearson-Prentice Hall, 2004.
- [10] C. Lott, O. Milenkovic, and E. Soljanin, “Hybrid-ARQ: theory, state of the art and future directions,” *IEEE Inf. Theory Workshop*, pp. 1–5, July 2007.

-
- [11] N. Gopalakrishnan, *Achievable Rates and Rate Selection Algorithms for Incremental Redundancy (IR) Hybrid ARQ (HARQ) Wireless Systems*. PhD thesis, Purdue University, 2008.
 - [12] P. Wu and N. Jindal, “Performance of hybrid-ARQ in block-fading channels: A fixed outage probability analysis,” *IEEE Trans. Commun.*, vol. 58, no. 4, pp. 1129–1141, 2010.
 - [13] M. Y. Rhee, *Mobile Communication Systems and Security*. Wiley, 2009.
 - [14] R. McEliece and W. Stark, “Channels with block interference,” *IEEE Trans. Inf. Theory*, vol. 30, pp. 44 – 53, Jan. 1984.
 - [15] S. Sesia, G. Caire, and G. Vivier, “Incremental redundancy hybrid-ARQ schemes based on low-density parity-check codes,” *IEEE Trans. Commun.*, vol. 52, pp. 1311 – 1321, Aug. 2004.
 - [16] W. Root and P. Varaiya, “Capacity of classes of Gaussian channels,” *SIAM Journal on Applied Mathematics*, vol. 16, no. 6, pp. 1350–1393, 1968.
 - [17] D. Tuninetti, “On the benefits of partial channel state information for repetition protocols in block fading channels,” *IEEE Trans. Inf. Theory*, vol. 57, pp. 5036 –5053, Aug. 2011.
 - [18] N. Gopalakrishnan and S. Gelfand, “Achievable rates for adaptive IR hybrid-ARQ,” in *IEEE Sarnoff Symp.*, pp. 1–6, 2008.
 - [19] S. Lin and P. Yu, “An effective error control scheme for satellite communications,” vol. 28, pp. 395–401, Mar. 1980.
 - [20] B. Zhao and M. C. Valenti, “Practical relay networks: A generalization of hybrid-ARQ,” *IEEE J. Sel. Areas Commun.*, vol. 23, pp. 7–18, Jan. 2005.
 - [21]
 - [22] A. E. Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge University Press, 2012.

- [23] D. P. Bertsekas, *Dynamic Programming and Optimal Control*, vol. 1. Athena Scientific, 2005.
- [24] E. Visotsky, V. Tripathi, and M. Honig, "Optimum ARQ design: a dynamic programming approach," in *Proc. IEEE Int. Symp. on Inf. Theory*, p. 451, June 2003.
- [25] E. Malkamaki and H. Leib, "Performance of truncated type-II hybrid-ARQ schemes with noisy feedback over block fading channels," *IEEE Trans. Commun.*, vol. 48, pp. 1477–1487, Sept. 2000.
- [26] M. Zorzi and R. Rao, "Throughput analysis of Go-Back-N ARQ in Markov channels with unreliable feedback," in *IEEE Int. Conf. on Commun. 1995 (ICC'95)*, vol. 2, (Seattle, US), pp. 1232–1237 vol.2, June 1995.
- [27] X. Zhaoji and B. Sebire, "Impact of ACK/NACK signalling errors on high speed uplink packet access (HSUPA)," in *IEEE Int. Conf. on Commun. 2005 (ICC'05)*, vol. 4, (Seoul, Korea), pp. 2223–2227 Vol. 4, May 2005.
- [28] T. T. Kim, *Limited feedback information in wireless communications: Transmission schemes and performance bounds*. PhD thesis, KTH School of Electrical Engineering, Stockholm, Sweden, May 2008.
- [29] B. Makki and T. Eriksson, "On hybrid-ARQ and quantized CSI feedback schemes in quasi-static fading channels," *IEEE Trans. Commun.*, vol. 60, pp. 986–997, Apr. 2012.
- [30] L. Szczecinski, P. Duhamel, and M. Rahman, "Adaptive incremental redundancy for HARQ transmission with outdated CSI," in *IEEE Global Telecommun. Conf.*, pp. 1 – 6, Dec. 2011.
- [31] J.-F. Cheng, Y.-P. Wang, and S. Parkvall, "Adaptive incremental redundancy [WCDMA systems]," in *IEEE 58th Veh. Technol. Conf.*, vol. 2, pp. 737 – 741, Oct. 2003.
- [32] J.-F. Cheng, "On the coding gain of incremental redundancy over Chase combining," in *IEEE Global Telecom. Conf. 2003. (GLOBECOM'03)*, vol. 1, pp. 107–112, Dec. 2003.
- [33] E. Malkamaki and H. Leib, "Coded diversity on block-fading channels," *IEEE Trans. Inf. Theory*, vol. 45, pp. 771 – 781, Mar. 1999.

-
- [34] S. Pfletschinger and M. Navarro, "Adaptive HARQ for imperfect channel knowledge," in *Int. ITG Conf. on Source and Channel Coding (SCC'10)*, pp. 1–6, Jan. 2010.
 - [35] S. Stiglmayr, M. Bossert, and E. Costa, "Adaptive coding and modulation in OFDM systems using BICM and rate-compatible punctured codes," in *European Wireless*, pp. 1999–2003, Apr. 2007.
 - [36] Y. Li and W. Ryan, "Mutual-information-based adaptive bit-loading algorithms for LDPC-coded OFDM," *IEEE Trans. Wireless Commun.*, vol. 6, pp. 1670–1680, May 2007.
 - [37] J.-F. Cheng, "Coding performance of hybrid-ARQ schemes," *IEEE Trans. Commun.*, vol. 54, pp. 1017–1029, June 2006.
 - [38] I. Byun and K. S. Kim, "Cooperative hybrid-ARQ protocols: Unified frameworks for protocol analysis," *CoRR*, vol. abs/0812.2301, 2008.
 - [39] R. Bellman, "On the theory of dynamic programming," in *Proc. of the National Academy of Sciences of the United States of America*, vol. 38, p. 716, 1952.
 - [40] Q. Liu, S. Zhou, and G. Giannakis, "Cross-layer combining of adaptive modulation and coding with truncated ARQ over wireless links," *IEEE Trans. Wireless Commun.*, vol. 3, pp. 1746–1755, Sept. 2004.
 - [41] X. Wang, Q. Liu, and G. Giannakis, "Analyzing and optimizing adaptive modulation coding jointly with ARQ for QoS-guaranteed traffic," *IEEE Trans. Veh. Technol.*, vol. 56, pp. 710–720, Mar. 2007.
 - [42] A. Goldsmith and P. Varaiya, "Capacity of fading channels with channel side information," *IEEE Trans. Inf. Theory*, vol. 43, pp. 1986–1992, Nov. 1997.
 - [43] A. Agustin, J. Vidal, and O. Munoz, "Hybrid turbo FEC/ARQ systems and distributed space-time coding for cooperative transmission," *Int. J. of Wireless Inf. Networks*, vol. 12, pp. 263–280, Dec. 2005.
 - [44] I. Stanojev, O. Simeone, and Y. Bar-Ness, "Performance analysis of collaborative hybrid-ARQ incremental redundancy protocols over fading channels," in *IEEE 7th*

- Workshop on Signal Proc. Advances in Wireless Commun. 2006 (SPAWC'06)*, pp. 1–5, July 2006.
- [45] S. Tomasin, M. Levorato, and M. Zorzi, “Analysis of outage probability for cooperative networks with HARQ,” in *IEEE Int. Symp. on Inf. Theory*, pp. 2716–2720, June 2007.
- [46] K. Zheng, L. Hu, W. Wang, and L. Huang, “Performance analysis of HARQ transmission in cooperative DF relaying systems,” *Wireless Personal Commun.*, vol. 55, pp. 441–455, Nov. 2010.
- [47] B. Makki, *Data Transmission in the Presence of Limited Channel State Information Feedback*. PhD thesis, Chalmers University of Technology, Gothenburg, Sweden, Nov. 2013.
- [48] T. Eriksson and T. Ottosson, “Compression of feedback in adaptive OFDM-based systems using scheduling,” *IEEE Commun. Lett.*, vol. 11, pp. 859–861, Nov. 2007.
- [49] T. Eriksson and T. Ottosson, “Compression of feedback for adaptive transmission and scheduling,” *Proceedings of the IEEE*, vol. 95, pp. 2314–2321, Dec. 2007.
- [50] S. Ekbatani, F. Etemadi, and H. Jafarkhani, “Transmission over slowly fading channels using unreliable quantized feedback,” in *Data Compression Conference (DCC'07)*, pp. 353–362, Mar. 2007.
- [51] L. Szczecinski, S. R. Khosravirad, P. Duhamel, and M. Rahman, “Rate allocation and adaptation for incremental redundancy truncated HARQ,” *IEEE Trans. Commun.*, vol. 61, pp. 2580–2590, June 2013.
- [52] S. R. Khosravirad, L. Szczecinski, and F. Labeau, “Rate adaptation for cooperative HARQ,” *IEEE Trans. Commun.*, vol. 62, pp. 1469–1479, May 2014.
- [53] S. R. Khosravirad, L. Szczecinski, and F. Labeau, “Rate-adaptive HARQ in relay-based cooperative transmission,” in *IEEE Int. Conf. on Commun. (ICC'13)*, pp. 5328–5333, 2013.
- [54] S. R. Khosravirad, L. Szczecinski, and F. Labeau, “Rate allocation for HARQ in relay-based cooperative transmission,” in *IEEE Wireless Commun. & Networking Conf. (WCNC'14)*, (Istanbul, Turkey), Apr. 2014.

- [55] K. Huang, R. Heath, and J. Andrews, "Limited feedback beamforming over temporally-correlated channels," *IEEE Trans. Signal Process.*, vol. 57, pp. 1959–1975, May 2009.
- [56] M. Jabi, L. Szczecinski, M. Benjillali, and F. Labeau, "Outage-optimal power adaptation and allocation for truncated HARQ," in *IEEE Int. Conf. on Commun. (ICC'14)*, pp. 1890–1896, June 2014.
- [57] R. Tajan, *Mécanismes de retransmission Hybrid-ARQ en radio-cognitive*. PhD thesis, Université de Cergy Pontoise, 2013.
- [58] D. Djonin, A. Karmokar, and V. Bhargava, "Joint rate and power adaptation for type-I hybrid-ARQ systems over correlated fading channels under different buffer-cost constraints," *IEEE Trans. Veh. Technol.*, vol. 57, pp. 421–435, Jan. 2008.
- [59] A. Karmokar, D. Djonin, and V. Bhargava, "POMDP-based coding rate adaptation for type-I hybrid-ARQ systems over fading channels with memory," *IEEE Trans. Wireless Commun.*, vol. 5, pp. 3512–3523, Dec. 2006.
- [60] Q. Chen and P. Fan, "Performance analysis of hybrid-ARQ with code combining over interleaved rayleigh fading channel," *IEEE Trans. Veh. Technol.*, vol. 54, pp. 1207–1214, May 2005.
- [61] A. Karmokar, D. Djonin, and V. Bhargava, "Cross-layer rate and power adaptation strategies for IR-HARQ systems over fading channels with memory: A SMDP-based approach," *IEEE Trans. Commun.*, vol. 56, pp. 1352–1365, Aug. 2008.
- [62] Y. Chen, S. Zhang, S. Xu, and G. Li, "Fundamental trade-offs on green wireless networks," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 30–37, 2011.
- [63] G. Caire, G. Taricco, and E. Biglieri, "Optimum power control over fading channels," *IEEE Trans. Inf. Theory*, vol. 45, pp. 1468–1489, July 1999.
- [64] I. Stanojev, O. Simeone, Y. Bar-Ness, and D. Kim, "On the energy efficiency of hybrid-ARQ protocols in fading channels," in *IEEE Int. Conf. on Commun. (ICC'07)*, pp. 3173–3177, June 2007.
- [65] E. Yeh and R. Berry, "Throughput optimal control of cooperative relay networks," *IEEE Trans. Inf. Theory*, vol. 53, pp. 3827–3833, Oct. 2007.

- [66] S. Kim, S. Lee, and D. Sung, "HARQ rate selection schemes in a multihop relay network with a delay constraint," *IEEE Trans. Veh. Technol.*, 2014.
- [67] S. Lee, W. Su, S. Batalama, and J. Matyjas, "Cooperative decode-and-forward ARQ relaying: Performance analysis and power optimization," *IEEE Trans. Wireless Commun.*, vol. 9, pp. 2632–2642, Aug. 2010.
- [68] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2009.
- [69] R. Bellman, *Dynamic Programming*. Princeton University Press, 1972.
- [70] I. Ibragimov and I. Linnik, *Independent and Stationary Sequences of Random Variables*. Wolters-Noordhoff, 1971.
- [71] A. E. Gamal and Y.-H. Kim, "Lecture notes on network information theory," *CoRR*, vol. abs/1001.3404, 2010.
- [72] N. Prasad and M. Varanasi, "MIMO outage capacity in the high SNR regime," in *Int. Symp. on Inf. Theory 2005 (ISIT'05)*, pp. 656–660, Sept. 2005.
- [73] J. Laneman, D. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. Inf. Theory*, vol. 50, pp. 3062–3080, Dec. 2004.
- [74] J. Laneman and G. W. Wornell, "Distributed space-time-coded protocols for exploiting cooperative diversity in wireless networks," *IEEE Trans. Inf. Theory*, vol. 49, pp. 2415–2425, Oct. 2003.
- [75] A. E. Gamal and Y.-H. Kim, "Lecture notes on network information theory," *CoRR*, vol. abs/1001.3404, 2010.
- [76] A. Host-Madsen and J. Zhang, "Capacity bounds and power allocation for wireless relay channels," *IEEE Trans. Inf. Theory*, vol. 51, pp. 2020–2040, June 2005.
- [77] A. Bletsas, A. Khisti, D. Reed, and A. Lippman, "A simple cooperative diversity method based on network path selection," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 3, pp. 659–672, 2006.

-
- [78] B. Maham, A. Behnad, and M. Debbah, “Analysis of outage probability and throughput for half-duplex hybrid-ARQ relay channels,” *IEEE Trans. Veh. Technol.*, vol. 61, no. 7, pp. 3061–3070, 2012.
 - [79] T. Chaitanya and E. Larsson, “Superposition modulation-based symmetric relaying with hybrid-ARQ: Analysis and optimization,” *IEEE Trans. Veh. Technol.*, vol. 60, pp. 3667–3683, Oct. 2011.
 - [80] W. Rui and V. Lau, “Combined cross-layer design and HARQ for multiuser systems with outdated channel state information at transmitter (CSIT) in slow fading channels,” *IEEE Trans. Wireless Commun.*, vol. 7, pp. 2771–2777, July 2008.
 - [81] Y.-H. Kwon, *Novel Minipacket Design and Decoding Performance Improvement With Additional Parity Information In Wireless Hybrid ARQ Systems*. Thesis, 2005.
 - [82] J. Ramis and G. Femenias, “Cross-layer design of adaptive multirate wireless networks using truncated HARQ,” *IEEE Trans. Veh. Technol.*, vol. 60, pp. 944–954, Mar. 2011.