On Performance Analysis of Retransmission Schemes with Fading Channels

PETER LARSSON

Doctoral Thesis
Stockholm, Sweden 2018
Akademisk avhandling som med tillstånd av Kungl Tekniska högskolan framläggdes till offentlig granskning för avläggande av teknologi doktorsexamen i telekommunikation Torsdag 18 Oktober 2018 kl. 09:00 i Kollegiesalen, Kungliga Tekniska Högskolan, Brinellvägen 8, Stockholm.

© Peter Larsson, September 2018
Parts of this thesis have been published under IEEE copyright

Tryck: Universitetsservice US AB
To my father and brother

In loving memory of my mother
Abstract

Future wireless communication systems and services brings increased performance demands, with respect to data-rate(s), reliable communication, (stochastic) real-time guarantees, and more. In this context, not only new communication schemes are needed, but also more capable performance analysis methods are essential. Digital wireless communication systems convey information (digital messages, as data packets) inherently susceptible to errors when communicated. In this respect, fading channels, receiver noise, and interference, are often the main causes of errors. State-of-the-art wireless systems use, e.g., retransmissions (and channel coding) to correct possible remaining errors in communicated message. Retransmissions of erroneous messages is generally known, under the umbrella-term, as automatic repeat request. Adopting a modern terminology (see Chapter 2.2), the three main schemes are here denoted; automatic repeat request (ARQ), repetition redundancy hybrid-ARQ (RR-HARQ), and incremental redundancy hybrid-ARQ (IR-HARQ).

There are at least three factors, that motivate further performance studies of the ARQ-, RR-, and IR-schemes. First, many commercially important and extensively deployed wireless systems, e.g. cellular systems, use those (H)ARQ schemes as core system components. Second, those schemes are often integrated with various (recently invented) communication schemes, such as multiple-antenna systems, which promote the need of further studies. Third, the information theoretically-based performance characterization of those (H)ARQ schemes is, in our view, only in its infancy, and only a few closed-form expressions for very basic (H)ARQ-cases exists in the literature. The thesis deals, on a high level, with the problem of developing performance analysis methods for (H)ARQ schemes, and, on a more detailed level, studying particularly important (H)ARQ-cases, i.e. with respect to (wrt) (H)ARQ-scheme, fading statistics, antenna-scheme, etc. In doing so, the thesis addresses tools and models that support, ease, or strengthen the analysis.

We start our study with a basic throughput analysis of (H)ARQ (Chapter 4). A general throughput expression for HARQ is given in terms of the Laplace-transform (LT) for the probability density function (pdf) of a so called effective-channel. Here, the effective-channel represents the signal-to-noise-ratio (SNR), or mutual information (MI), after signal processing. We then focus on some important (H)ARQ-cases and give closed-form throughput expressions in a general diversity (GD) channel, accounting for space-time-block coding (STC), maximal ratio combining (MRC), and Nakagami-$m$ fading. The throughput of (H)ARQ can, in many cases, be max-
imized by tuning the initial transmission rate. However, analytical throughput optimization has proven challenging to solve even for the simplest (H)ARQ-cases. We propose a parametric optimization approach, based on judiciously chosen parameter, that allows expressions for the optimal throughput, and the optimal rate point, to be given in closed-forms (Chapter 5). The method is demonstrated for several important, but previously not handled, (H)ARQ-cases. An inherent assumption in this thesis, shared with many other works in wireless communication analysis, is the assumption of that the average symbol MI equals the additive white Gaussian noise (AWGN) channel capacity. The underlying assumption is that the communication symbol can be modeled as an independent and identically distributed (iid) complex Gaussian random variable (r.v). However, practical systems use discrete modulation, not a continuous r.v. Quadrature amplitude modulation (QAM) is the most common (discrete) modulation format in communication systems. Unfortunately, QAM exhibits an asymptotic 1.53 dB SNR-gap relative to the AWGN channel capacity. We substantiate the assumption, of modeling the communication signal as iid complex Gaussian, and close the SNR-shaping-gap, by proposing a novel modulation framework inspired from packing arrangements (spiral-phyllotaxis) among plants (Chapter 6). Much work on wireless performance analysis focus on specialized fading channel gain models, such as exponentially- or gamma-distributed-fading. We introduce the idea of a matrix exponential (ME) distributed effective channel SNR (Chapter 7). The ME-distribution is dense on the positive axis, and includes the exponential- and gamma-distribution as special cases. With the ME-distributed channel at hand, we develop an overall ME-distribution-based framework that simplifies the performance analysis and directly express performance measures in the ME-distributed (effective) channel parameters. It has proven hard to analyze (H)ARQ with interference via standard methods, and only special cases have previously been handled successfully. With the ME-distribution-based performance analysis framework, we can now analyze interferers with ME-distributed SNRs. Numerous closed-form throughput expressions are also given in terms of ME-distribution-based channels. Up to this point, the performance measure of choice has been throughput. However, communication systems may impose delay requirements. For this purpose, the effective capacity, giving an indication of communication rate for a given maximum delay and delay violation probability, is a more suitable performance measure. We formulate a very general retransmission system model (allowing for multiple transmissions, multiple communication modes, and multiple rate increments), going beyond classical ARQ-, RR-, and IR-models, and develop a powerful recurrence-based effective capacity performance analysis framework (Chapter 8).

Thus, to summarize on a high-level, we introduce a simplifying LT-based performance analysis framework, develop a powerful auxiliary-parameterized throughput optimization method, propose a novel AWGN channel capacity approaching (golden angle) modulation scheme, introduce the ME-distributed channel, develop the ME-distribution-based performance analysis framework, design a highly general retransmission system model, and propose a recurrence-based (effective capacity)
performance analysis framework. Throughout the thesis, numerous new closed-form performance expressions are given built on the tools and models introduced in the preceding chapters.
Sammanfattning

Framtida trådlösa kommunikationssystem och tjänster för med sig högre prestandakrav m.a.p. data hastigheter, tillförlitlig kommunikation, (stokastiska) realtidskrav, och mer. I detta sammanhang behövs inte bara nya kommunikationsmetoder, men även mer kapabla prestandaanalysmetoder är viktiga. Digitala trådlösa kommunikationssystem förmedlar information, digitala meddelanden (datapaket), vilka är känsliga för fel vid kommunikation. I detta avseende är fädnande kanaler, mottagarbrus, och interferens, oftast huvudsakerna till sådana fel. Moderna kommunikationssystem använder, t.ex., omsändningar (och kanalkodning) för att korrigera eventuella fel i meddelanden. Omsändning av meddelanden med informationsfel är generellt känt, under paraplytermen, såsom automatic repeat request. Vid antagandet av en modern terminologi (kapitel 2.2), de tre huvudmetoderna är här benämnda: automatic repeat request (ARQ), repetition redundancy hybrid-ARQ (RR-HARQ), och incremental redundancy hybrid-ARQ (IR-HARQ). Ätminstone tre faktorer motiverar vidare prestandastudier av ARQ-, RR-, och IR-metoder. För det första, många kommersiellt viktiga och brett utbyggda trådlösa system, t.ex. cellulära system, använder dessa (H)ARQ metoder som kärnkomponenter. För det andra, dessa metoder är ofta integrerade med olika (nyligen framforskade) kommunikationsmetoder, såsom multipelantenennmetoder, vilket motiverar behovet av vidare studier. För det tredje, den informationsbaserade prestandakarakteriseringen av dessa (H)ARQ-metoder är bara i dess början, och endast ett fåtal uttryck på sluten form existerar i litteraturen. Avhandlingen behandlar, på en hög nivå, utmaningen med att utveckla prestandaanalysmetoder för (H)ARQ-metoder och, på en mer detaljerad nivå, studerar särskilt viktiga (H)ARQ-fall m.a.p. (H)ARQ-metoder, fädningsstatistik, antennsystem, o.d. Avhandlingen handlar m.a.o. om verktyg och modeller, vilka stödjer, förenklar, eller stärker prestandaanalysen.

Vi startar vår studie med en grundläggande throughпутanalys av (H)ARQ (Kapitel 4). Ett generellt throughpututtryck för HARQ ges i termer av dess Laplace transform (LT) för sannolikhetstäthetsfunktionen av en s.k. effektiv kanal, där den effektiva kanalen representerar signal-till-brus förhållandet, eller ömsesidig information, efter signalprocessering. Sedan fokuserar vi på viktiga (H)ARQ-fall, och ger throughpututtryck på sluten form för en generell diversitetskanal, beaktande rumtid-block-kodning, maximal-kvot-kombinering, och Nakagami-\(m\) fädnings. IR är också beaktat med LT-ramverket, och nya outage-sannolikhetsgränsvärden ges. Throughput för (H)ARQ kan, i många fall, maximeras genom att anpassa...
nyu slutna prestanda uttryck vilka bygger på verktyg, modeller, och resultat introducerade i tidigare kapitel.
Acknowledgements

As I lay the final words to my thesis, looking back, I reflect on how much this experience has been both challenging and rewarding. Its with joy and pride I see how all hard work paid off in new knowledge, skills, and publications.

I would like to express my deepest gratitude towards my main-advisor, Prof. Lars Kildehøj Rasmussen, and my co-advisor Prof. Mikael Skoglund. With your guidance, ideas, and courses, I’ve learned so much from you. Lars, in addition to sharing your advises and insights, your encouragements, your positive attitude and humor have been of immense support, an absolute inspiration. Whenever I’ve been in moments of doubt, just felt overwhelmed with challenges, you brought balance and joy back. Mikael too, your clever subtle sense of humor, mixed with critical and exploring research questions, seeing research parallels I otherwise would have missed, it has all been beyond invaluable. Lars and Mikael, always available for discussions, reading, commenting, and returning draft manuscripts ahead of real-time. In times of setbacks, you have shown great empathy, kindness, and the best sides of being human. I’m deeply grateful to Assoc. Prof. Tobias Oechtering, Assoc. Prof. Ragnar Thobaben, Assoc. Prof. James Gross, and Ph.D. Hussein Al-Zubaidy, whom have shared their vast expertise in courses taught as well as in many research discussions.

I would also like to thank all my fantastic colleagues at the department of information science and engineering for your great friendship, and for sharing your great knowledge and inspiring ideas in the internal seminars. Thanks also to Raine Tiivel and Irene Kindblom for helping me with all administrative affairs. To prof. Mats Bengtsson, I’m sincerely grateful for proof-reading the thesis.

I wish to thank Prof. Leszek Szczecinsky for taking time to act as opponent for this thesis, and also to Prof. Erik G. Larsson, Prof. Iryna Andriyanova, and Prof. Thomas Eriksson, for participating in the evaluation committee. I’m deeply grateful to Prof. Mats Bengtsson for the quality review of this thesis.

I’m eternally grateful to my beloved late mother, Solbritt, and my beloved father, Jan, who have always been there for me, giving me their unconditional love, support, and believing in me. To my brother Stefan, my dearest friend, so grateful to have you. To you and all I hold dear, with all my love, I dedicate this thesis.

Peter Larsson
Stockholm, September 2018
Imagination will often carry us to worlds that never were. 
But without it we go nowhere. 
Carl Sagan
# Contents

Abstract v  
Sammanfattning ix  
Acknowledgements xiii  
Contents xv  

1 Introduction 1  
  1.1 System Overview 2  
  1.2 Problem Formulation 3  
  1.3 Outline and Contributions of the Thesis 5  
  1.4 Contributions Outside the Thesis 17  
  1.4.1 Papers by the Author 17  
  1.4.2 Patents by the Author 23  
  1.4.3 Contribution Highlights 23  
  1.4.4 (H)ARQ-related Contributions Beyond/Within the Thesis 24  
  1.5 Notation and Acronyms 25  
  1.5.1 Notation 25  
  1.5.2 Parameters, Variables, Perf. Measures, and Constants 25  
  1.5.3 Functions and Expressions 28  
  1.5.4 Acronyms and Abbreviations 28  

2 Preliminaries 33  
  2.1 Communication Theory Fundamentals 33  
  2.1.1 Information Measures 33  
  2.1.2 Antenna-based Communication Concepts 37  
  2.1.3 Special Network Topology Concepts 41  
  2.2 (H)ARQ Fundamentals 42  
  2.2.1 (H)ARQ Operation 42  
  2.2.2 Historical Development and Naming Conventions 45  
  2.3 The Laplace Transform 47  
  2.4 Appendices 47
3 System Model 49
  3.1 Performance Measures 49
  3.2 Time Division Model 49
  3.3 (H)ARQ Models 50
  3.4 Physical Layer Model 51
     3.4.1 Decoding Failure Probability - Based Model 52
     3.4.2 Effective Channel pdf - Based Model 53
     3.4.3 Wireless Communication System - Based Model 53

4 Basic Throughput Analysis of (H)ARQ 57
  4.1 Motivation and Outline 57
  4.2 Performance Analysis 58
     4.2.1 General Throughput Analysis 58
     4.2.2 ARQ and Persistent/Truncated-RR with GD-channel 66
     4.2.3 Persistent/Truncated-IR with Rayleigh Fading Channel 69
  4.3 Numerical Results and Discussions 74
  4.4 Summary and Conclusions 79
  4.5 Appendices 80

5 Parametric Throughput Maximization of (H)ARQ 89
  5.1 Motivation and Outline 89
  5.2 Performance Analysis 91
     5.2.1 Throughput Optimization: Classical Method 91
     5.2.2 Throughput Optimization: Auxiliary-Parameter Method 93
     5.2.3 ARQ and Persistent/Truncated-RR with GD-Channel 96
     5.2.4 Truncated-IR in Rayleigh Fading Channel 97
     5.2.5 Throughput Optimization: Rate-Parameter Method 98
  5.3 Numerical Results and Discussions 102
  5.4 Summary and Conclusions 105
  5.5 Appendices 107

6 Golden Angle Modulation and Rank-1 Lattice Designs 109
  6.1 Motivation and Outline 110
  6.2 Golden Angle Modulation 111
     6.2.1 GAM-framework, Disc-GAM, and Motivation of Framework 111
     6.2.2 Geometric Shaping 121
     6.2.3 Probabilistic Shaping 128
     6.2.4 Joint Probabilistic-Geometric-Shaping 130
  6.3 Rank-1 (Fibonacci) Lattice-based Constellation Design 131
  6.4 GAM-based (H)ARQ 134
     6.4.1 GAM-based HARQ with Mapping Rearrangement 134
  6.5 Numerical Results and Discussions 134
     6.5.1 Mutual Information Performance 135
     6.5.2 Throughput Performance of (H)ARQ with GAM 141
When the electromagnetic field theory was formulated by Maxwell in 1864, and the electromagnetic radio waves were experimentally discovered by Hertz in 1888, no one could envision the profound impact those light-speed propagating waves would come to have up to now, some 130 years later. Today, with e.g. digital-radio and -television, satellites, radio-links, wireless-LAN, and cellular (mobile telephone) systems, radio communication have undoubtedly come to influence everyone’s life. This is particularly true for cellular systems, which have evolved from a mobile telephone voice only service, to a versatile Internet-integrated mobile data service. Cellular systems, as well as many other modern radio communication systems, requires reliable, efficient, but often also delay-constrained, data communication. For reliable and efficient communication, handling e.g. noisy and fading channels, channel coding is generally used to encode data packets (a.k.a. data words, messages, packet data units) into codewords (CWs). Then, a communication rate equal to the capacity of the channel is asymptotically achievable with infinitely long codes. From practical viewpoint, however, codewords have to be limited in length. When the capacity of the channel is unknown, or it varies unpredictably, one strategy is to retransmit a channel codeword representation of a message repeatedly, until the message can be correctly decoded. This can be seen as a dynamic adjustment of the channel code rate, which works without prior knowledge about the channel state, and noise realization. Schemes following this approach are generally called automatic repeat request schemes. The original objective of such automatic repeat request-schemes were to ensure fully reliable communication of data packets in the sense that data packets were persistently retransmitted until they were correctly received/decoded, at least to the extent that a cyclic redundancy check (CRC) could guarantee. If the communication failed for a prolonged time, the connection was terminated. Since the first ARQ-schemes were invented and patented, e.g. [VD40, VD49, vDVD55], more advanced variants have been developed to increase the communication performance, such as throughput, and to handle real-time traffic with delay constraints. In this thesis we are concerned with three modern automatic repeat request-schemes, which, with
a modern naming convention (see Chapter 2.2), will be denoted ARQ, RR-HARQ (RR), and IR-HARQ (IR). For these three modern schemes, collectively referred to as (H)ARQ-schemes henceforth, all are assumed to use channel coding. For HARQ, it is assumed that all transmissions relating to a message are processed jointly when decoding a message, whereas for ARQ, only the last transmission is used in decoding a message. There are (at least) three factors that motivate further studies of ARQ, RR, and IR and their performance. First, many commercially important and extensively deployed wireless systems, such as WiFi, GSM, IS-95, WCDMA, CDMA-2000, and LTE, use ARQ, RR, or IR as fundamental core system component(s) [Mol05, Stü12]. Second, the ARQ, RR, and IR schemes are often integrated with various (recently invented/researched) communication features, e.g. multiple antenna communication schemes, which suggest the need of new studies. Third, the information-theoretically-based performance characterization of ARQ-, RR-, and IR-schemes is, surprisingly, limited. Thus, we argue that, further analytical studies of ARQ-, RR-, and IR-schemes, are important, needed and well justified.

1.1 System Overview

In Fig. 1.1, a schematic high-level view of a (H)ARQ communication system, as studied in this thesis, is shown. The network topology is comprised of a transmitter (TX), and a receiver (RX). The transmitter encode an awaiting data packet into a codeword, and transmit the codeword (or a redundancy block, as for IR) over a channel. The channel is impaired by noise and fades independently for each transmission. Feedback from the receiver, in form of acknowledgements (ACK) indicates which of the data packets are correctly decoded, and does not need to be retransmitted. We expand this high-level system view in Chapter 2, where the fundamentals of (H)ARQ is reviewed, and in Chapter 3, where the detailed (H)ARQ system models, considered in the thesis, are given.
1.2 Problem Formulation

This thesis is primarily concerned with the following two problem areas:

- Developing analytical performance analysis methods, including improved modeling and suitable performance measures, for (H)ARQ systems operating in semi-static fading wireless channels.

- Deriving analytical, preferably closed-form, performance expressions of important, yet unexplored, (H)ARQ-cases, where a (H)ARQ-case is comprised of a (H)ARQ-scheme, fading channel model, antenna scheme, etc.

The objective of the first item is to offer new tools, general approaches, and detailed methods for analysis of (H)ARQ systems. This is, e.g., of interest when (possibly) existing methods are inadequate, or too complicated, to facilitate the derivation of analytical performance expressions of (H)ARQ-cases of interest. For this, we need new theory and methods. The second item aims to offer analytical (information-theoretically optimal) performance expressions that can be used as reference cases, and to which other schemes can be benchmarked against. Those other schemes may be existing, or newly developed, ones, theoretical models or real-world implementations, and characterized by analytical performance expressions or by simulated performance curves. By benchmarking against the information-theoretically-based performance expressions (which we derive in the thesis for different (H)ARQ-cases), a (H)ARQ designer/standardizer/researcher/implementer can be guided in their work, realizing that some considered (H)ARQ-scheme operates poorly, and requires more work, or is sufficiently good, and hence requires no more work. In addition, information-theoretical performance expressions may give deepened insights in the operation, e.g. performance behaviour in the high- and low-SNR ranges, the effects involved, and thereby suggesting directions for improved (H)ARQ designs. Once information-theoretically-based performance expressions have been derived for different (H)ARQ-schemes, this can offer guidance to which (H)ARQ-scheme to choose for a standard or a product. Additional problem areas that are of interest are to develop methods to derive optimized performance expressions, to give analytical expressions for optimized (H)ARQ-cases, to propose new performance measures to use, to propose new, more powerful and easy-to-use, fading channel models suitable for (H)ARQ analysis. Yet another problem area, briefly touched on in the thesis, is to further substantiate some modeling aspects in information-theoretically-based (H)ARQ analysis.

Limitations

No single work can study everything. Thus, we limit the problem studied in this thesis in the following manner: First, only fixed transmit power (H)ARQ systems are considered. This is motivated since; this has not been researched in sufficient detail,
current cellular systems operates with rate adaptation and (essentially) fixed downlink transmit power, and the maximum transmit power (even when power control is available) need to be used for the critical low-SNR users close to the communication boundary, the cell-edge. Second, the performance evaluation is limited to analyzing the performance of a single transmitter-receiver pair, not multiple links, and not complex topologies as (cooperative) relaying, multihop networks, etc. The reason is the fundamentally important topology, a node-pair communicating over a single link, which preferably should be studied in greater detail before advancing to more complicated communication topologies. Third, the performance study focuses primarily on throughput as performance measure, and less on alternative performance measures, such as outage probability, packet loss rate, etc. This is so since, in our view, the (human) user experience is primarily characterized by the throughput, and not on secondary indirect measures, like outage probability. Note also that packet loss rate, the probability of dropping a message, is not directly linked to outage probability, but can be handled by retransmission, or packet level (network) coding, with higher layer functionality. Fourth, we also limit the study to semistatic block-fading channels. This is motivated since; with not to fast and not to slow mobile users, the channel can be block-fading, a scheduler may strive to separate retransmissions with about the coherence time of the channel, and with slow fading channel, adaptive modulation and coding (AMC), rather than (H)ARQ, may be preferred. While a correlated block-fading channel scenario is not the main focus of the thesis, Chapter 8 provides a performance analysis framework (with communication modes) that can be arranged to handle the correlated fading case. Fifth, the study is limited to (long) capacity achieving channel codes. The motivation for this is that it gives a fundamental upper performance limit, and, practically even for short datawords, long codes can be created by merging multiple users’ traffic or by delaying and aggregating individual user’s traffic. Sixth, we do not consider rateless codes, which could be seen as an alternative to (H)ARQ, in this thesis. The reasons are that (H)ARQ is widely deployed in existing wireless communication systems, and a time-slotted system (more suited for (H)ARQ than rateless codes) is more common, as well as considered here. Moreover, in order to motivate alternatives to (H)ARQ, such as rateless codes, it is necessary to know how well (H)ARQ could perform in the best of worlds and then benchmark against rateless codes.

Philosophy Behind Thesis Disposition and Form

"Make everything as simple as possible, but not simpler", a quote attributed to Einstein. This is a leading philosophy, a tradition, in (classical) physics, readily noticed in the formulation of experiments, scientific theories, and the mathematical laws governing the reality. We prescribe to such philosophy, and have therefore strived after a style and form of the thesis making it accessible, focusing on the fundamentals, and hoping for practical adoption. We intentionally strive to avoid untractable analysis and performance expressions, of marginal, or of no, practical (or theoretical) use.
This is achieved in various ways. For example, a judiciously chosen lean notation simplifies the look of the mathematical expressions. By focusing on fundamental, yet important, network topologies, channels, and communication schemes, the number of parameters, and the system complexity, is limited, which yields less convoluted analysis and performance expressions. The organization of the chapters, and their content, has a structure of, basic/classical-to-advanced, narrow-to-general, and a methodological red thread, which should facilitate accessibility. The models and performance analysis methods are deliberated crafted to allow for exact (closed-form) performance expressions, mitigating the need for approximations, bounds, or complicated long expressions. For readability, examples (and counterexamples) are included to build intuition, motivations for various design choices are emphasized, and the vast majority of derivations and proofs are pushed to the appendix.

1.3 Outline and Contributions of the Thesis

We now proceed with a chapter-by-chapter outline of the thesis, mainly including introduction, related work, and contributions sections.

Chapter 2
In this chapter, we review the preliminaries for the thesis - the fundamentals of communication-theory and (H)ARQ-schemes. We also include a historical account for the development of the (H)ARQ area and naming conventions.

Chapter 3
In this chapter, we introduce the system model and assumptions used in the thesis. Briefly, we deploy a time division channel access model for the forward channel, and immediate (error-free) communication for the reverse (feedback) channel. The considered (H)ARQ models are persistent/truncated-ARQ/IR/RR. The main performance measure considered in the thesis is throughput. The performance measure is expressed in one of three physical layer system model levels; i) a probability mass function (pmf), ii) a pdf, or corresponding LT, of the fading channel SNR, or iii) in parameters detailing a specific (H)ARQ-case.

Chapter 4
In this chapter, we initiate the performance study by analyzing some fundamental (H)ARQ-cases. The objective of this chapter is to provide an introduction, a prelude, to subsequent chapters. We seek a general expression allowing for direct computation of the throughput for persistent/truncated-HARQ, whilst omitting computing individual decoding failure probabilities for each retransmission. We look for closed-form throughput expressions for ARQ and RR with respect to various diversity processing cases in Nakagami-$m$- and Rayleigh-fading channels. The
throughput of persistent/truncated-IR, operating in Rayleigh fading channels, is characterized by bounds.

Related Work

The renewal reward theorem, essentially (4.1), was used in [WL83, ZR96] to define and study the throughput performance of HARQ, and go-back-N ARQ, respectively. In [CT01], ARQ (called ALO), RR (called RTD), and IR were studied based on renewal theory in an information-theoretical setting, assuming asymptotically long (capacity achieving) codes, the AWGN channel, and an information outage probability performance analysis framework. We adopt this information-theoretic view henceforth. An analytical throughput expression (with an infinite sum of decoding probabilities in the numerator) was given as [CT01, (24)], and a closed-form throughput expression for ARQ was subsequently derived as [CT01, (28)]. Closed-form throughput expressions for ARQ in Rayleigh block-fading were independently also examined in [BS06, (8)-(9)], [SLF08, (5)-(6)]. For RR, a closed-form throughput expression for Rayleigh block fading was derived in [LSKAT10, (20)], and in [KJSS10, (7)]. No closed-form throughput expression is known for IR, but individual decoding probabilities expressed in generalized upper incomplete Fox’s H-functions have been derived in [CA13, YA09]. In real world wireless systems, (H)ARQ is often combined with various multiple-antenna processing schemes. Some basic multiple-antenna schemes are, e.g., maximal ratio combining and selection diversity combining (SDC) [SA05, Sec. 9.1], Alamouti’s transmit diversity [AT97, Ala98], orthogonal space-time block coding (OSTBC) [AT97, TSC98, LS08], cyclic delay diversity (CDD) [PatL8, Lar98, DK01], and spatial-multiplexed MIMO (SM-MIMO) [LS08]. Multiple-antenna-related (H)ARQ performance studies are, e.g., found in [SF08, SF11] on OSTBC-ARQ, in [SF08, SF11] on CDD-ARQ, and in [GCD05, CGiFRC08, ME14] on MIMO-ARQ. The decoding failure probabilities for IR in Rayleigh fading has been considered in [HL11, CA13].

Since the publication of [LRS14c], several works have considered an LT-based throughput analysis framework, e.g. in [JBSL16, Sec. IIIa] and [SPSP15, Sec. IVc] for RR. In [HSL15, Sec. IV], the authors derived throughput expressions for deterministic multihop routing with RR/IR-HARQ based on a multidimensional extension of the LT-framework. Based on the LT-framework, analytical expressions for persistent/truncated-RR for the Rician fading channel and Nakagami-m fading channels was derived in [JHZW18, Sec. III] and in several related works.

Contributions

This chapter offers two levels of contributions; i) a general throughput expression of HARQ, and ii) closed-form (or analytical) throughput expressions of particular (H)ARQ communication cases. Theorem 4.1 and Corollary 4.1, gives general, and directly computable, throughput expressions for truncated- and persistent-HARQ, respectively. Based on those expressions, we show in Theorem 4.2 that the through-
1.3. Outline and Contributions of the Thesis

put of truncated-HARQ is upper bounded by the throughput of persistent-HARQ. Specifically, closed-form throughput expressions are given, with respect to a general diversity effective channel (incorporating any combination of OSTBC, MRC, and Nakagami-m fading channel with $m^N \in \mathbb{Z}^+$), for persistent-RR with N-fold diversity in Proposition 4.5, persistent-RR with \{2, 3, 4\}-fold diversity in Proposition 4.7, $K$-truncated-RR without diversity in Proposition 4.8. IR is explored for the Rayleigh fading channel, analytical throughput expressions are given and throughput bounds for truncated-IR, with $K = 2$, are given in (4.46)-(4.48). We also give a new expressions for the LT of the MIMO channel capacity.

The material in this chapter is based on the following published paper:


Chapter 5

In this chapter, we study the problem of maximizing the throughput of (H)ARQ systems wrt the initial rate, i.e. the rate of the first transmission. The throughput expressions are assumed to be continuous and differentiable for positive rates, and monotonically increasing with increasing SNR. We develop an analytical parametric optimization framework that can provide closed-form expressions for the maximum throughput and the optimal rate point. This is done for (H)ARQ-cases which can generally not be solved analytically to give closed-form expression with a more standard optimization approach. The developed optimization framework can, as exemplified in Chapter 8, also be applied to other performance measures and problem settings. We focus on throughput optimization for ARQ and RR, and consider approximate optimal throughput expressions for truncated-IR.

Related Work

Using the standard analytical optimization approach\(^1\) for (H)ARQ systems, throughput optimization for Rayleigh fading channel has been considered in [BS06, SLF08] for ARQ, and in [KJSS10] for RR. Expressions for the throughput maxima of ARQ and RR in Rayleigh fading have also been considered in [TC02, Appendix A], but then parameterized in the initial rate $R$ instead of the average SNR as in [BS06, SLF08, KJSS10].\(^2\) Besides those cases, we are not aware of any other closed-

---

\(^1\)I.e. taking the derivative of the throughput with respect to the rate, equate the expression to zero, solve for the optimal rate point, and use the optimal rate point expression as the rate in the throughput expression.

\(^2\)Independently of [TC02, Appendix A], via optimization efforts of throughput expression for ARQ and RR-with in Rayleigh fading in 2009, we found a closed-form throughput expression expressed in terms of the optimal rate point $R^*$ rather than the SNR. Later, we found out that, while some throughput optimization problems could be solved more easily with the rate-parametrization,
form throughput expressions for the (H)ARQ-cases considered here, and hence no related expressions for the maximum throughput.

Since the publication of [LRS14c], the auxiliary parameter optimization framework has subsequently been explored in [KCLN16, Sec. V] for the purpose of analyzing RR/IR in an uplink multicell (massive MIMO) multi-user-SIMO system.

Contributions

In this chapter, the main contribution is the auxiliary parameter (AP) optimization method which, due to its more general formulation, solves a larger class of problems than the classical SNR-parameter-method, as well as the rate-parameter (RP) method in [TC02, Appendix A]. The general solution is given in Theorem 5.1 and Corollary 5.1. Other contributions are new exact expressions for specific cases, such as ARQ and RR in Nakagami-\(m\) fading channel with OSTBC and MRC. For the RP-method, we also give new expressions for ARQ in Nakagami-\(m\) fading channel with OSTBC and MRC, specifically for the Alamouti diversity scheme. The method is also applied to analyze analytical, but approximate, throughput expressions for truncated-IR, where a different optimization parametrization than for ARQ and RR is useful.

The material in this chapter is based on the following published paper:


The AP-optimization framework has subsequently also been applied, and refined, in [LRS14a, LRS16a, LGAZ+16, LRS16b].

Chapter 6

In this chapter, we reexamine the assumption in Section 3.4.3 of modeling the per packet average MI with the AWGN channel capacity. Underlying this in turn is the assumption of using a continuous complex Gaussian r.v. as the communication signal. To support the signal model, and hence certain parts of the performance analysis in this thesis, we develop a novel modulation framework, the *Golden angle modulation*. Under this framework, we develop various modulation formats that, e.g., can approximate the pdf of a complex Gaussian r.v. We show, through numerical computation and simulation that the MI asymptotically approaches the AWGN channel capacity. Peak-to-average-power (PAPR) constraints are also considered in the performance evaluation w/wo optimization.

than with the SNR-parametrization, some problems resisted all attempts. Thus, the more general and powerful auxiliary parametric optimization method in Chapter 5 was developed to handle an even wider class of (H)ARQ throughput optimization problems.
The design of GAM can mathematically be formulated as a, judiciously chosen, spiral of certain kind. There are, basically, two key design features in the GAM framework. The first is separating consecutively indexed constellation points with (modulus $2\pi$) the golden angle, \( \varphi = \pi (3 - \sqrt{5}) \), in the phase domain. The second is allowing the radial and the phase distribution of constellation points to have different forms and dependency wrt the constellation point index. This deliberate design has, as will be seen, many attractive features. A prominent feature is that geometric radial shaping is not only possible, but the constellation expands (almost perfectly) symmetrically around the origin for geometric radial shaping. The inspiration to the GAM design comes from observations in nature, noting the inherent shape-versatility of some plants seed, leaf, and flower petal arrangements, known as spiral phyllotaxis, in disc-, sphere-, and cylinder-like configurations. Vogel [Vog79] proposed a mathematical model for the spiral-based packing arrangement (a.k.a. spiral phyllotaxis) of sunflower seeds. We build on Vogel’s mathematical model, but decouple the radial and phase arrangement of constellation points, but, importantly, target modulation design with certain performance objective of choice.

In this chapter, we also reexamine the assumption in Section 3.4.3 of modeling the accumulated per packet average MI (AMI) for IR as the sum of AWGN channel capacities for each transmission. Classically, IR-designs build on transmitting a punctured codewords based on a low-rate mother code. As an alternative, we consider a joint design of HARQ, with mapping rearrangement (a.k.a. mapping diversity, mapping/constellation rearrangement), and GAM.

Related Work

Many modulation formats have been developed in the literature. Besides basic square/rectangular-quadrature amplitude modulation (QAM) and phase shift keying (PSK), some more elaborate schemes are, e.g., star-QAM [HNKW04], amplitude PSK (APSK) [TWD74], and QAM to circular isomorphic constellation [Kay16]. QAM is, possibly, the most studied of all constellation designs, but also the most widely implemented scheme in commercial communication systems. However, at increasing SNR, QAM exhibits an asymptotic \( \approx 1.53 \text{ dB} \) SNR-gap (a shaping-loss) between its MI-performance and the AWGN capacity [FU98]. Thus, (classical uniform) QAM does not support the notion of modeling the MI as the AWGN channel capacity for increasing constellation sizes, or increasing SNR. Geometric- and probabilistic-shaping have been proposed to overcome this shaping-loss. Some works on geometric-shaping are on asymmetric constellations in [DSY87], on non-uniform-QAM (NU-QAM) in [BCL94], on NU-PAM for AWGN in [SF00], and on NU-PSK/PAM capacity optimization in [BJF07]. Similarly, some works on probabilistic shaping are as a survey in [FGL84], on N-dimensional sphere/cube constellations in [CO90], on Maxwell-Boltzmann distribution based probabilistic shaping in [KP93], and on APSK in [LXPY11, XV13, M15]. While geometric shaping can be applied to QAM, a constellation with circular symmetry (like APSK) is often preferred, as the AWGN capacity achieving (complex Gaussian) distribution is cir-
circular symmetric. Yet, one issue with APSK is that the number of constellation points in each ring depends on the radial shaping.

While geometric-shaped modulation schemes, e.g. NU-QAM and NU-APSK, in principle allow the MI to be modeled with the AWGN channel capacity, we target an improved design, with greater versatility (e.g. in terms of handling different shaping requirements, different number of constellations points), and with improved performance (e.g. in terms of MI and PAPR). In our work, we propose a spiral design of a very specific kind. We note that relatively few prior works, preceding GAM, consider spiral-based designs. Analog spiral-based modulation was proposed in [TMW75, KR06, KR07]. The latter two studied Archimedean spirals with amplitudes $f(x) = xe^{ix}$, $x \geq 0$, in [KR06], and $f(x) = g(x)e^{ig(x)}$ with non-linear stretch, $g(x) \geq 0$, in [KR07]. In [JEMH06, KSPK08] discrete spiral-based modulation was proposed. Whereas the former examined a logarithmic spiral design, the latter proposed a design with four intertwined Archimedean spirals. The GAM framework does not adopt an Archimedean spiral design, simply because it inherently defy the notion of geometric radial shaping, much desired for a versatile, AWGN capacity achieving, or low-PAPR, design. None of the modulation works preceding GAM have considered, nor benefited from, the golden angle based design to achieve a near-ideal uniform distribution in phase.

Modulation with mapping rearrangement was studied in [Met77, Ben92, SDH05], whereas HARQ, combined with mapping rearrangement, was studied in [SB05, SCF09], yet mainly for low order modulation. All those studies consider non-GAM constellations. Primarily deterministic (optimum) mappings are studied, whereas [SDH05] studied random mappings.

While not immediately related to modulation, GAM is built on the notion of spiral phyllotaxis (SP). SP is a certain kind of spiral arrangement of leaves, seeds, or petals, and can be observed among, e.g., mosses (leafy shoot), ferns (leaf-branches), gymnosperm (e.g. cycad- and pine-cones), and angiosperm (e.g. sunflowers and dahlias). A historical source, on the numerous works on SP, is, e.g., [ABJ97]. In [Vog79], Vogel introduced a mathematical model to describe the arrangements of seeds of a sunflower, $x_m = \sqrt{m}e^{i2\pi \varphi}$, $m \in \{1, 2, \ldots, M\}$, with $\varphi$ being the golden ratio (or angle). A number of researchers have observed the SP in nature, and applied Vogel’s mathematical model in their respective fields of research. For example, the SP packing has been utilized for antenna array designs in [Boe00, VTC*09], an irregular multi-layered 3D-cone color pallet design (color quantization and processing) in [MS01a] (and extended to a spherical color pallet design in [PLC09]), ultrasound imaging array designs in [MGGMU10], and orbital angular momenta based designs for light in [LNT*11]. In the historical work on SP, see e.g. [Vog79], the angular distance between consecutive points has been modeled as the ratio of two consecutive Fibonacci numbers, as well as (the fractional part of) the golden ratio. Both approaches has been used in subsequent literature. In contrast to the present work, the above SP-related works do, e.g., not; consider the problem of modulation, exploit the radial shape-versatility of SP for geometric-shaping, consider probabilistic-shaping, target MI-performance maximization, consider an optimized
1.3. Outline and Contributions of the Thesis

Since the publication of [Lar18, Lar17], an Archimedean spiral based phase-noise robust modulation design, without radial geometric-shaping or probabilistic-shaping, has also been considered in [DHC†17].

Contributions

The key contribution of this chapter is proposing a novel, versatile, high-performance modulation format framework - the Golden angle modulation (GAM). This framework, basically, builds on separating consecutively indexed constellation points with the golden angle (allowing for some small margins), and allowing for different forms of the spiral radial and phase functions. For large enough constellations, this discrete, spiral-based, design offers a near-ideal circular symmetric design, a near-ideal uniform phase distribution, and allows for geometric radial shaping that is near-ideal symmetric wrt the origin. The GAM framework is demonstrated with numerous geometric and probabilistic shaping designs, mainly targeting high (or optimal) MI performance w/wo PAPR constraints. In particular, The AWGN channel is considered. Numerical MI-performance results indicate that the designs are (asymptotically) AWGN channel capacity achieving as the number of constellation points increases, thus supporting the assumption in this thesis of modeling the MI with the AWGN channel capacity. Moreover, for sufficiently large number of constellation points, GAM-performance exceed QAM, NU-QAM, APSK etc. For clarity, note that the GAM-design does not claim to be MI-optimal given a fixed number of constellation points, but rather offers a practical, flexible, structured, and generally MI-high-performing design. Also, a modulation design, based on an optimized rank-1 lattice formulation, is proposed and studied.

Another contribution is to propose, and show, that by combining HARQ, with low-complex random mapping rearrangement, we can offer a throughput performance that is nearly identical to the throughput of IR when the AMI is modeled as the sum of AWGN Shannon capacities.

The material in this chapter is based on the following submitted paper:


- [LS18]: P. Larsson, and M. Skoglund, “Golden Angle Modulation: Approaching the AWGN Capacity,” accepted for publication at IEEE 88th Vehicular
Chapter 7

In this chapter, we introduce the notion of an ME-distributed fading channel gain. The resulting SNR, due to various signal processing steps, e.g., MRC, SDC, and OSTBC, is called the effective channel SNR. With an ME-distributed fading channel gain, the effective channel SNR is, wrt standard signal processing operations such as max-, min-, and sum-SNR, also ME-distributed. Closed-form throughput expressions for truncated/persistent-(H)ARQ are derived and expressed in ME-distribution vector- and matrix-parameters. The throughput of ARQ is studied for non-identical and (in)dependent ME-distributed fading signal and interferers. The unified ME-distribution-based performance analysis framework, can also be applied to other performance measures, e.g. the effective capacity as considered in Chapter 8.

Related Work

Various fading channel models, e.g. Rayleigh, Rician, Hoyt, and Log-normal distributed fading, have been reviewed in many works, e.g. [Wil96, PM96, Rap01, Sha12, TV04, TSC98, Ala98, SA05, Mol05, BPS98]. Analysis of the SNR distribution, for different diversity signal processing steps (MRC, SDC, OSTBC) together with different fading channels, are also considered in many works, e.g. [Wil96, PM96, Rap01, TV04]. Works on, and the throughput of (H)ARQ for Rayleigh, Nakagami-m channel w/wo diversity (MRC, SDC, OSTBC) is reviewed and considered in Chapter 4. There are also works focusing on lower physical layer functions, analyzing outage probability and diversity combining for general fading channels, e.g. in [KAS00, JSB12]. The ME-distribution, proposed here for fading channel gain modeling, has been explored in a wide range of fields. Examples of such are: economics (risk and ruin probabilities), control theory (linear ordinary differential equations), queuing theory, see e.g. [BFT08, AB96] and references therein. Only two known works, [MvdLR95, McM95], have considered the ME-distribution in the context of wireless networks, but then only for modeling queueing networks, not fading channels.

Contributions

The main contributions in this chapter are to introduce the ME-distribution for fading channel gain modeling, and to propose an ME-distribution-based performance analysis framework. More specifically, this chapter offers a number of new results,
1.3. Outline and Contributions of the Thesis

expressions, and useful ME-properties. We give closed-form ME-form throughput expressions for ARQ in Theorem 7.4, truncated-HARQ in Theorem 7.5, and persistent-HARQ in Theorem 7.6, wrt ME-distributed wireless (effective) channel SNRs. For illustrative purpose, we also give the sum-throughput for ARQ with network coded bidirectional relaying wrt the ME-distributed fading channel in Theorem 7.7. ARQ, also accounting for dependent (independent) ME-distributed sum-interfering channel, is considered in Theorem 7.8 (7.9). In Example 7.13 and Remark 7.5, a truncated continued fraction approach is proposed which allows MI-pdfs, without a rational LT, to be approximated by a ME-density. Complex cases, like IR with exponentially fading SNR, or ARQ/RR/IR for $N \times N$-MIMO channels with iid fading complex Gaussian matrix entries, can then be handled. Some useful new ME-distribution tools are the integral expressions in Theorem 7.1, Corollary 7.4 and Lemma 7.4, and the expression for the maximum of two ME-distributed r.v.s in Theorem 7.2. Using the new integral expressions in Theorem 7.1, we give a new expression for the maximum of two ME-distributed r.v.s in Theorem 7.2.

The material in this chapter is based on the following published papers:


The ME-distribution performance analysis framework is originally proposed in [LRS16a]. The ideas in [LRS16a] are however refined and expanded in this chapter, which lay the foundation for [LRS16b].


The work [LRS14a] contributes mainly indirectly to this chapter (and hence also [LRS16b]) with the problem formulation involving ARQ and interference, but also a somewhat complicated throughput expressions which illustrate the benefit of ME-distribution-based analysis. In [LRS16a], the idea of ME-distributed interference fading channel gains are mentioned as an area of exploration, which is explored further in this chapter, and in [LRS16b].
Chapter 8

In this chapter, we study the effective capacity performance measure of general retransmission schemes formulated in terms of transition probabilities. The effective capacity gives (asymptotically) the maximum sustainable source rate for a stochastic delay constraint imposed. Specifically, the effective capacity of persistent-/truncated-(H)ARQ is studied, and expressed, e.g., in terms of transition probabilities, and in ME-distributed fading channel parameters. A new method for analyzing the effective capacity of systems with memory is developed. We model the retransmission scheme as a 3D random-walk with given transition probabilities, and express the moment generating function (mgf) on a recurrence relation form. For finite time, the mgf is computed matrix-algebraically, and for infinite time, it is computed through the spectral radius of a certain matrix on companion-form.

Related Work

Based on the effective bandwidth performance measure in [CT95, Kel96], the notion of effective capacity was introduced in [WN03]. The use of AMC with signal-to-noise-ratio (SNR) dependent rate adaptation was analyzed in [WN03, Cho12, TZ07]. The effective capacity of ARQ was considered in [Cho12, LGV15, HZ12, AF15], and HARQ-like schemes were studied in [LGV15, GLCH10, AF15]. On the modeling side, finite state Markov chains (FSMCs) has been used for modeling the effects of channel variations in e.g. [TZ07, GLCH10, AF15]. The work [TZ07] considers such FSMC-modeling for AMC with fading channel, but no retransmission scheme. Assuming a large number of ARQ cycles in each AMC state, FSMC-modeling for joint AMC and ARQ with fading channel was used in [TZ07, FRC09]. The combined effects of fading channel and IR-HARQ (with two transmission attempts) are approximated, using a finite state Markov process, in [GLCH10, Sec. II.B, Fig. 2]. A HARQ-like scheme with truncated transmissions was considered in [AF15] using a Markov modulated process model. The performance analysis framework of this scheme assumed guaranteed correct decoding of the transmitted message, at latest when the number of transmission reached the transmission limit, but did not model a possible failed decoding and packet drop on the last retransmission attempt. For the works [TZ07, FRC09, HZ12, AF15], i.e. those modeling the systems as Markov modulated processes, each define a transition probability matrix $P$ and a diagonal reward matrix $\Phi$, for which the effective capacity is computed from the spectral radius of a matrix $P\Phi$. We note that [WN03]-[AF15] assumed Rayleigh fading channels, whereas [TZ07, MLN12] also considered Nakagami-$m$ fading.

Since the publication of [LGAZ16], the recurrence-relation-based effective capacity analysis framework has also been explored in [LGV16, LGV17].
Contributions

The main contribution of this chapter is a structured and unified effective capacity performance analysis method of the general class of retransmission schemes involving multiple transmissions, multiple rate increments, and multiple communication modes. This class is fully defined by a set of transition probabilities, wrt to a 3D-random walk system model, indexed in terms of number of transmissions, number of rate increments, and communication modes. The effective capacity is determined by solving a certain recurrence relation, yielding a matrix- and a characteristic equation-based solution. This new performance modeling and analysis approach can, in contrast to a (hypothetical) retransmission system modeled as a Markov modulated process with a $P\Phi$-formulation, avoid a (non-trivial and non-unique) case-specific modeling step$^3$, and inherently handles the classical truncated-HARQ operation with a possible packet discard on last transmission if the receiver fail to decode the transmitted message. The proposed method also leads to resulting model matrices of lower dimension (determined only by the number of transmission attempts and communication modes, but independent of the number of reward rates), less complex forms (without, e.g., transition probabilities repeated in multiple matrix entries and ratios of transition probabilities, as encountered in related works), and lower computational complexity compared to a hypothetical Markov modulated approach with its inherent modeling requirement. The methodology is also more intuitive, easier to apply, and facilitate the derivation of closed-form performance expressions. In contrast to the $P\Phi$-formulation, which would require infinitely-sized matrices, the recurrence relation formulation (and associated characteristic equation solution) enables performance analysis of persistent-HARQ.

Numerous closed-form effective capacity expressions are given for truncated- and persistent-HARQ for several physical layer models, e.g. for ME-distributed fading channel gains. One interesting example of particular contributions is Corollary 8.8, which generalizes the classical throughput expression for truncated-HARQ in Lemma , to the effective capacity performance metric. Another example is Corollary , which generalizes the classical throughput expression for persistent-HARQ with exponentially distributed fading gain, (4.27), to ME-distributed fading channel gain and the effective capacity metric. Effective capacity analysis is also done for NC-ARQ (and multilayer-ARQ), exemplifying multiple packet transmissions. Additional contributions are the effective capacity analysis and expressions, for systems, with $t$-time slots, a joint parametrization of target delay and delay violation probability.

$^3$Note that a Markov chain model of a retransmission system, as is needed for a Markov modulated process-based approach, is not unique since it depends on the modeler interpretation of the system. I.e., the same retransmission system can, from performance evaluation point of view, be modeled with different number of states, and with different sets of transitions. The operation in our retransmission scheme is fully characterized by a transition probability $P_{k\nu \tilde{l}l}$, where $k$ represents the number of transmissions needed until successful decoding, $\nu$ is the number of correctly decoded packets, $\tilde{l}$ is the originating communication mode, and $l$ is the final communication mode during a retransmission cycle.
The material in this chapter is based on the following published paper:


**Copyright Notice**

Parts of the material presented in this thesis are partly verbatim based on the thesis author’s joint works which are previously published or submitted to conferences and journals held by or sponsored by the Institute of Electrical and Electronics Engineer (IEEE). IEEE holds the copyright of the published papers and will hold the copyright of the submitted papers if they are accepted. Materials (e.g., figure, graph, table, or textual material) are reused in this thesis with permission.

**High-level Summary of Contributions**

A, simplified, high-level summary of the thesis contributions is:

- New performance analysis tools.
  - Direct Laplace transform-based HARQ throughput analysis.
  - Auxiliary parametric based optimization.
  - Matrix exponential distribution analysis approach and tools.
  - Recurrence-relation-based analysis approach.

- New analytical performance expressions.
  - Throughput/effective capacity of truncated/persistent-ARQ/RR/IR. (wrt $Q_k$, $F(s)$, Rayleigh/Nakagami-$m$/ME-distributed fading, and ARQ with independent/dependent ME-distributed fading interference channels.)
  - Optimized throughput/effective capacity wrt rate. (Rayleigh/Nakagami-$m$/ME-distributed fading channels.)

- New motivation for the AWGN channel capacity MI-model, and complex Gaussian r.v. signal model. New modulation framework.
  - Golden angle modulation.
  - Schemes: Geometric/probabilistic bell-shaped-GAM, disc-GAM, etc.

- New (H)ARQ-schemes.
  - New golden angle modulation-based variant of (H)ARQ with (random) mapping rearrangement.

- New channel model.
1.4 Contributions Outside the Thesis

In this section, we list paper- and patent-contributions by the author that have not been included in the thesis. We include those contributions, not just because it is customary, but also since several concepts studied in the thesis are drawn from those publications, and several (H)ARQ-related works have been performed in those publications. Specifically, in the thesis, we analyze the fundamental ideas (extended with (H)ARQ when needed) contributed by the author in i) [PatL50, PatL57], [LJ06, Lar07a] (NC-ARQ), and ii) [PatL43], [LJS05, LJS06] (NC bidirectional relaying), with the tools developed in the thesis.

In addition to paper- and patent-contributions outside the thesis, the author have submitted a number of standardization contributions (omitted for brevity) that are based on, [PatL7, PatL12, PatL15, PatL52], in IEEE 802.11, and [PatL2] in ETSI Hiperlan2. Several inventions have been included directly, or indirectly, in standards (GSM, WCDMA, LTE, WiMAX, WiFi), demonstrators, and products.

1.4.1 Papers by the Author

The papers by the author can be divided in during or before the PhD studies.

Papers Outside the Thesis, but During the PhD Studies


- [LRS14b]: P. Larsson, L. K. Rasmussen, and M. Skoglund, “Analysis of rate optimized throughput for large-scale MIMO-(H)ARQ schemes,” in Proc. IEEE Global Communications Conference (GLOBECOM’14), Austin, USA, 8-12 Dec. 2014, pp. 3760-3765. (Manuscript submitted for review to (ICC’14), 30 Sept. 2013, not accepted. Same manuscript resubmitted for review to (GLOBECOM’14), 26 Mar. 2014, accepted.)


### Papers Outside the Thesis, but Before the PhD Studies


• [Lar06]: P. Larsson, “A routing metric for floor acquisition oriented medium access schemes,” in Proc. 6th Scandinavian Workshop on Ad Hoc Networks (ADHOC’06), Johanneberg, Sweden, May 2006.


• [Lar01b]: P. Larsson, “Selection diversity forwarding in a multihop packet radio network with fading channel and capture,” in Proc. ACM MobiHoc Ad Hoc Networks (MobiHoc’01), Long beach, USA, Mar. 2001, pp. 279-282. (Best poster award)


• [LL98]: M. Larsson, and P. Larsson, “The DLC layer of the wireless ATM research project (WARP),” in Proc. The 1st Workshop on wireless mobile ATM implementations, Hangzhou, China, Apr. 1998, pp. 6-10.
## 1.4. Contributions Outside the Thesis

<table>
<thead>
<tr>
<th>Contribution</th>
<th>(OR) ARQ</th>
<th>Network coding</th>
<th>Multihop/Ad-hoc</th>
<th>Opportunistic</th>
<th>Relay/Repeater</th>
<th>Antenna systems</th>
<th>IC, Full-duplex</th>
<th>RRM</th>
<th>Cognitive radio</th>
<th>Protocol</th>
<th>Other</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>[PatL1]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[PatL2]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[PatL3]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[PatL4]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[PatL5]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[PatL6]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[PatL7]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[PatL8]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[PatL9]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[PatL10]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[PatL11]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[PatL12]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[PatL13]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[PatL14]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[PatL15]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[PatL16]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[PatL17]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[PatL18]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[PatL19]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[PatL20]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[PatL21]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[PatL22]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[PatL23]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[PatL24]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[PatL25]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[PatL26]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[PatL27]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[PatL28]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[PatL29]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[PatL30]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[PatL31]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[PatL32]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[PatL33]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[PatL34]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[PatL35]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[PatL36]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[PatL37]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[PatL38]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[PatL39]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[PatL40]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[PatL41]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[PatL42]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[PatL43]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[PatL44]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[PatL45]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[PatL46]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[PatL47]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[PatL48]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[PatL49]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[PatL50]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1.1: Coarse categorization of contributions wrt areas.
<table>
<thead>
<tr>
<th>Contribution</th>
<th>(H)ARQ</th>
<th>Network coding</th>
<th>Multi-hop/Ad-hoc</th>
<th>Opportunistic</th>
<th>Relay/Repeater</th>
<th>Antenna systems</th>
<th>IC, Full-duplex</th>
<th>RRH</th>
<th>Cognitve radio</th>
<th>Protocol</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>PatL51</td>
<td>⋆</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PatL52</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PatL53</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PatL54</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PatL55</td>
<td></td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PatL56</td>
<td></td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PatL57</td>
<td></td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td>⋆</td>
<td>⋆</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PatL58</td>
<td></td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PatL59</td>
<td></td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PatL60</td>
<td></td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PatL61</td>
<td></td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PatL62</td>
<td></td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PatL63</td>
<td></td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PatL64</td>
<td></td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PatL65</td>
<td></td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PatL66</td>
<td></td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PatL67</td>
<td></td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PatL68</td>
<td></td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PatL69</td>
<td></td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PatL70</td>
<td></td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PatL71</td>
<td></td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PatL72</td>
<td></td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PatL73</td>
<td></td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PatL74</td>
<td></td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PatL75</td>
<td></td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PatL76</td>
<td></td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PatL77</td>
<td></td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PatL78</td>
<td></td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PatL79</td>
<td></td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PatL80</td>
<td></td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PatL81</td>
<td></td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PatL82</td>
<td></td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PatL83</td>
<td></td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PatL84</td>
<td></td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PatL85</td>
<td></td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PatL86</td>
<td></td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PatL87</td>
<td></td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PatL88</td>
<td></td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PatL89</td>
<td></td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PatL90</td>
<td></td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PatL91</td>
<td></td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PatL92</td>
<td></td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PatL93</td>
<td></td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PatL94</td>
<td></td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PatL95</td>
<td></td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PatL96</td>
<td></td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PatL97</td>
<td></td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PatL98</td>
<td></td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td>⋆</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1.2: Coarse categorization of contributions wrt areas (continued).
1.4.2 Patents by the Author

The author is the inventor of about 250 granted patents, approximately 98 unique patented inventions, 18 (H)ARQ related patents, and the main inventor of 85% of the patented inventions. The complete (in filing order) patent list is given in the Bibliography of Author’s patents at the end of the thesis. Since the patent owners/lawyers often chose non-descriptive names for patent applications, we list each patent in Tab. 1.1 and Tab. 1.2 with a compact description, but also classify each patent wrt to communication concepts involved. It is observed that the patents cover the areas: (H)ARQ, network coding, multihop networks, ad hoc networks, opportunistic channel (gain/SNR/SINR) dependent communication, 2-hop relaying, advanced repeaters (involving multi-antenna, self-IC, full-duplex), multi-antenna systems (involving SM-MIMO, diversity, beamforming, network-MIMO etc.), interference cancelation (IC), radio resource management (RRM) (involving power, rate, and channel), cognitive radio, protocol design. A few other areas are covered too. Already from the first patent, as shown in the tables, cross-layer design-based inventions were strategically targeted.

1.4.3 Contribution Highlights

The conceptually most important contributions (outside the thesis) are (possibly):

- IC of known, or overheard, own signal or interference: [PatL21], I-99, F-03.
  - 2-phase bidirectional relaying (Analogue NC): [PatL21], I-99, F-03.
  - Unidirectional multihop chain w IC of interference from previously transmitted packets: [PatL21], I-99, F-03.
  - General multihop networking w IC of any known/overheard, own signal or interference: [PatL21], I-03, F-03.

- Cyclic delay diversity: [PatL8], I-98, F-00. Also in [Lar98].

- Line-of-Sight MIMO (LoS-MIMO):
  - Linear/square ant. arrays: [PatL28], I-03, F-03.
  - Rectangular/triangular ant. arrays: [PatL42] I-03, F-04. Also in [Lar05].

- Opportunistic routing:
  - Data-ACKs-based: [PatL9], I-99, F-00. Also [Lar01a, Lar01b, Lar01c].
  - CTS-RTSs-(WiFi)-based:
  - [PatL10], I-99, F-01. Data-ACKs-based with MUD: [PatL20], I-03, F-03.
    Also in [LJ03].
  - Probe-CSIs-based: [PatL19], I-03, F-03. Also in [LJ04].
  - OR least-cost routing: [PatL19], I-99, F-03.

- Network-coded bidirectional (two-way) relaying: I-04 (March).
  - Network-coded bidirectional relaying: [PatL43], I-04, F-04. Also in [LJS05, LJS06], and variants in [Lar08b, KALPT09].
(H)ARQ patents outside the scope of the thesis

- [PatL5, PatL9, PatL10, PatL19, PatL20, PatL32, PatL33, PatL38, PatL44, PatL50, PatL57, PatL58, PatL61, PatL62, PatL64, PatL65, PatL66]

(H)ARQ papers outside the scope of the thesis

- [LRS14b, LRS14a, LSKAT13, LRS12, LSKAT10, Lar08b, Lar08a, Lar07a, LJ06, LJ05, LJ04, LJ03, Lar01c, Lar01b, Lar01a]

(H)ARQ papers within the scope of the thesis

- [LRS16b, LRS16a, LGAZ+16, LRS14c, LRS14a], and [Lar18, Lar17]

Table 1.3: (H)ARQ-related contributions outside/within the scope of the thesis.

- PSK-modulation-based NC (w bidirectional relaying and ARQ) [PatL69]: I-06, F-07. Also in [Lar08b].

- Network coded (H)ARQ: I-04 (March).
  - Multiple unicast NC-ARQ: [PatL50] and partly in [PatL57], I-04, F-05. Also in [LJ06, Lar07a], and a variant in [LSKAT12].
  - Optimal multicast NC-ARQ: [PatL61], I-06, F-06. Also in [Lar08a]. The suboptimal XOR-case covered in [PatL58], I-04, F-06.
  - Distributed NC-ARQ in relay system: [PatL64], I-07, F-07.
  - NC-ARQ in one-hop multinode Ad-hoc network: [PatL63], I-06, F-06.

- NC-OR: [PatL56], I-04, F-05.

- Self-IC (Full-duplex) communication:
  - Instantaneous OFDM relaying: [PatL53], I-05, F-05.
  - Full-duplex self-IC MIMO relaying: [PatL59], I-05, F-05. Also in [LP08, LP09].
  - MIMO null-space-projection-based interference-suppression (and self-IC) in relaying: [PatL67], I-06, F-07.
  - Self-IC full-duplex TDD in cellular system: [PatL60]: I-05, F-06.

  - Distributed (iterative) UL Network-MIMO: [PatL30], I-99, F-03.
  - RRM for UL Network-MIMO: [PatL31], I-99, F-03.

1.4.4 (H)ARQ-related Contributions Beyond/Within the Thesis

For the readers convenience, (H)ARQ-related papers and patents outside (and within) the thesis are summarized, here at one place, in Tab. 1.3.
1.5 Notation and Acronyms

1.5.1 Notation

We let \( x(\cdot) \), \( x \), \( X \) denote polynomials, scalars, vectors, and matrices. We write \( \text{vec}(X) \), \( \text{tr}(X) \), \( \|X\|_F \), \( |X| \), \( \rho(X) \) to denote the vectorized form (concatenation of columns), trace, Frobenius norm, determinant and spectral radius of a matrix \( X \) with appropriate dimensions, e.g., being square in the case of trace or determinant.

The convolution, \( k \)-fold convolution, the matrix transpose, Kronecker-product, and Kronecker-sum, are indicated by \( \ast \), \( (\cdot)^{\otimes k} \), \( (\cdot)^T \), \( \oplus \), and \( \otimes \), respectively. Moreover, the optimality, lower bound and upper bound are indicated as \( (\cdot)^* \), \( (\cdot)^\ast \), and \( (\cdot)\)\( \ast \)\( \)\( \). The expectation and the probability of a r.v.
uses the notation \( \mathbb{E}\{\cdot\} \) and \( \mathbb{P}\{\cdot\} \). For complex, real, rational, irrational, natural and integer numbers, the notation \( \mathbb{C}, \mathbb{R}, \mathbb{Q}, \mathbb{I}, \mathbb{N} \in \{1, 2, \ldots\} \), and \( \mathbb{Z} \in \{\ldots, -2, -1, 0, 1, 2, \ldots\} \), are used, respectively. Special constants are the standard basis unit vector \( e_n \) (with a one at the \( n \)th position), the identity matrix \( I \), the shift matrix \( S \) (with all ones on the super-diagonal, otherwise all zero entries). The pdf, the Laplace transform of a pdf, and the cdf of a r.v. \( X \) are written as \( f_X(x) \), \( F(s) \), and \( F_X(x) \). We let \( \simeq \) denote the asymptotic equivalence, i.e. functions \( f \) and \( g \) are asymptotically equivalent if the limit \( \lim_{x \to \infty} f(x)/g(x) \) exists and is equal to 1. \( X \sim f(x) \) means that the random variable \( X \) follows the distribution \( f(x) \). A parameter/variable \( x \) referring to the effective channel, i.e. after signal processing, is indicated as \( \tilde{x} \). Throughout the thesis, we often denote general, unspecified, variables/parameters with the letters \( x \), \( y \), and \( z \).

1.5.2 Parameters, Variables, Perf. Measures, and Constants

Parameters

- \( d \) ME-distr. numerator polynomial degree
- \( B \) Bandwidth
- \( D_{\text{max}} \) Maximum permissible packet delay
- \( E_s \) Symbol energy
- \( J \) Upper limit of a general index \( j \)
- \( J_{\text{MC}} \) Number of Monte-Carlo iterations
- \( J_{\text{MED}} \) Number of products of ME-distr.-polynomials
- \( K \) Maximum number of permissible transmission attempts
- \( L \) Number of retransmission communication modes
- \( m^N \) Nakagami-\( m \) parameter
- \( M \) Number of signal constellation points
- \( M_h \) Highest signal constellation point index
- \( M_l \) Lowest signal constellation point index
- \( N \) Diversity order (or number of antennas)
- \( N_{\text{gd}} \) Diversity order for the gd-channel
- \( N_k \) Number of modulated symbols
$N_{rx}$ Number of receive antennas per receiver
$N_{tx}$ Number of transmit antennas per transmitter
$N_X$ Number of channel uses in a MIMO symbol
$N_0$ Noise power density
$P$ Average transmit power
$P_{hv}$ Random-walk transition probability
$p(s)$ Numerator polynomial of the ME-distr.-channel
$p$ Vector characterizing the ME-distr.-channel
$P$ Matrix characterizing the ME-distr.-channel
$q(s)$ Denominator polynomial of the ME-distr.-channel
$q$ Vector characterizing the ME-distr.-channel
$Q$ Matrix characterizing the ME-distr.-channel
$r_{stc}$ OSTBC rate
$R$ Initial rate (in nats)
$R_b$ Initial rate (in bits)
$R_{gd}$ Effective initial rate (in nats) for the gd-channel
$S$ Average SNR
$S_{gd}$ Average effective SNR for the gd-channel
$U_I$ Number of interference users
$\epsilon_d$ Limit of acceptable delay violation probability
$\eta_d$ Probability that queue is non-empty
$\theta$ QoS-exponent
$\Theta$ Decoding threshold
$\Theta_{gd}$ Decoding threshold for the gd-channel
$\psi$ Effective capacity related QoS-parameter

**Variables**

$A$ Block companion recurrence-relation matrix
$C$ Channel capacity
$D$ Steady state delay of packets in queue
$f_\Xi(\xi)$ Generic pdf
$F_\Xi(\xi)$ Generic cdf
$F(s)$ Matrix with Laplace transforms used for MIMO analysis
$g$ Instantaneous channel gain
$H(s)$ SISO channel gain realization (scalar)
$h$ SIMO channel gain realization (vector)
$H$ MIMO channel gain realization (matrix)
$h(\cdot)$ Differential entropy
$H(\cdot)$ Entropy
$I_{Acc}^k$ Accumulated MI realization for the $k$-th transmission
$j$ General index
$k$ Retransmission index
1.5. Notation and Acronyms

\(l\) Communication mode index
\(m\) Constellation point index
\(n\) Antenna index
\(P_k\) Probability of successful packet decoding at the \(k\)th transmission
\(p_m\) Probability of using signal constellation point \(m\)
\(Q_k\) Probability of failed packet decoding up to the \(k\)th transmission
\(r_m\) Magnitude of signal constellation point \(m\)
\(s\) Laplace variable
\(t\) Transmission time index
\(u\) User index
\(v\) Accumulated service process index
\(V_t\) Accumulated service process at time \(t\)
\(x_m\) Complex value of signal constellation point \(m\)
\(X\) Input baseband signal
\(Y\) Output baseband signal
\(z\) Effective channel realization
\(Z\) Effective channel r.v.
\(W\) General noise signal
\(\lambda\) Matrix eigenvalue, Lagrange multiplier
\(\lambda_+\) The spectral radius of eigenvalues of a matrix
\(\nu\) Number of rate increments
\(\zeta\) GAM constellation point probability parameter
\(\sigma_w^2\) Noise variance

**Performance Measures**

\(C_{\text{eff}}\) Effective capacity
\(C_{\text{out}}\) Outage capacity
\(I(Y; X)\) Mutual information between r.v. \(Y\) and \(X\)
\(\text{SER}\) Symbol-error rate
\(T\) Throughput
\(Q_k\) The \(k\)th decoding failure probability
\(\epsilon_{\text{loss}}\) Packet loss rate
\(\rho\) Average rate
\(\tau_j\) The \(j\)th raw moment of the number of transmissions, \((\tau \triangleq \tau_1)\)

**Constants**

\(e_n\) Basis vector. One on the \(n\)th entry, all other entries are zero
\(I\) Identity Matrix
\(S\) Shift Matrix
\(\varphi\) The golden angle is \(2\pi\varphi\) radians, where \(\varphi \triangleq (3 - \sqrt{5})/2\)
1.5.3 Functions and Expressions

Functions

Exponential integral:
\[ E_1(x) = \int_x^\infty \frac{e^{-t}}{t} \, dt \]

Matrix exponential:
\[ E(tX) = \exp(tX) \]

Modified Bessel func. of the 2nd kind, order \( \nu \):
\[ K_\nu(x) = \int_0^\infty e^{-x \cosh t} \cosh \nu t \, dt \]

Q-function:
\[ Q(x) = (2\pi)^{-1/2} \int_x^\infty e^{-t^2/2} \, dt \]

Lambert’s W-function:
\[ W(x)e^{W(x)} = x \]

Regularized lower incomplete gamma function:
\[ \gamma_\nu(n, x) = 1 - \Gamma_\nu(n, x) \]

Upper incomplete gamma function:
\[ \Gamma(n, x) = \int_x^\infty t^{n-1}e^{-t} \, dt \]

Gamma function:
\[ \Gamma(n) = \Gamma(n, 0) \]

Regularized upper incomplete gamma function:
\[ \Gamma_\nu(n, x) = \Gamma(n, x)/\Gamma(n) \]

Expressions

Definition 1.1. (Closed-form expression) A closed-form expression can be evaluated in a finite number of operations, and may contain constants, variables, basic arithmetic operations (+, −, ×, ÷), and the functions including exponentiation, nth root, logarithm, trigonometric and hyperbolic functions, and their inverses.

Definition 1.2. (Semi-closed-form expression) An expression on semi-closed-form extends the set of closed-form expressions to also include special functions such as Bessel-, Gamma-, and Lambert’s-W-function.

Definition 1.3. (Analytical expression) An expression on analytical-form extends the set of semi-closed-form expression to also include infinite series and continued fractions, but excludes integrals and limits.

1.5.4 Acronyms and Abbreviations

In this thesis we make use of the following nomenclature:

ACK Acknowledgement
ADC Analog digital converter
a.k.a Also known as
AMC Adaptive modulation and coding
AMI Accumulated mutual information
APSK Amplitude PSK
ARQ Automatic repeat request
AP Auxiliary-parameter
AWGN Additive white Gaussian noise
BB Baseband
BER Bit-error rate
BF Beamforming
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BICM</td>
<td>Bit-interleaved coded modulation</td>
</tr>
<tr>
<td>CC-HARQ</td>
<td>Chase combining HARQ</td>
</tr>
<tr>
<td>CDD</td>
<td>Cyclic delay diversity</td>
</tr>
<tr>
<td>CF</td>
<td>Continued fraction</td>
</tr>
<tr>
<td>CoMP</td>
<td>Coordinated multipoint, a.k.a. network-MIMO</td>
</tr>
<tr>
<td>CP</td>
<td>Cyclic prefix</td>
</tr>
<tr>
<td>CRC</td>
<td>Cyclic redundancy check</td>
</tr>
<tr>
<td>CSI</td>
<td>Channel state information</td>
</tr>
<tr>
<td>CTS</td>
<td>Clear-to-send in WiFi</td>
</tr>
<tr>
<td>cu</td>
<td>Channel use</td>
</tr>
<tr>
<td>CW</td>
<td>Codeword</td>
</tr>
<tr>
<td>DAC</td>
<td>Digital analog converter</td>
</tr>
<tr>
<td>dB</td>
<td>deciBell</td>
</tr>
<tr>
<td>DC</td>
<td>Direct current</td>
</tr>
<tr>
<td>Distr.</td>
<td>Distribution</td>
</tr>
<tr>
<td>EC</td>
<td>Error correction</td>
</tr>
<tr>
<td>ED</td>
<td>Error detection</td>
</tr>
<tr>
<td>Erg.</td>
<td>Ergodic</td>
</tr>
<tr>
<td>EtE</td>
<td>End-to-end</td>
</tr>
<tr>
<td>Ex.</td>
<td>Example</td>
</tr>
<tr>
<td>FB</td>
<td>Feedback</td>
</tr>
<tr>
<td>FC</td>
<td>Fading channel</td>
</tr>
<tr>
<td>FEC</td>
<td>Forward error correction</td>
</tr>
<tr>
<td>Fig.</td>
<td>Figure</td>
</tr>
<tr>
<td>FSMC</td>
<td>Finite state Markov chain</td>
</tr>
<tr>
<td>GAM</td>
<td>Golden angle modulation</td>
</tr>
<tr>
<td>GB</td>
<td>Geometrical bell-(shaped)</td>
</tr>
<tr>
<td>GBN-ARQ</td>
<td>Go-back-N ARQ</td>
</tr>
<tr>
<td>GD</td>
<td>General diversity</td>
</tr>
<tr>
<td>GPB</td>
<td>Geometric probabilistic bell-(shaped)</td>
</tr>
<tr>
<td>HARQ</td>
<td>Hybrid-ARQ</td>
</tr>
<tr>
<td>(H)ARQ</td>
<td>ARQ and Hybrid-ARQ</td>
</tr>
<tr>
<td>IC</td>
<td>Interference cancellation</td>
</tr>
<tr>
<td>ID</td>
<td>Identifier</td>
</tr>
<tr>
<td>iid</td>
<td>Independent and identically distributed</td>
</tr>
<tr>
<td>im</td>
<td>Imaginary component</td>
</tr>
<tr>
<td>IR-HARQ</td>
<td>Incremental-redundancy HARQ, or simply IR</td>
</tr>
<tr>
<td>ISI</td>
<td>Inter-symbol-interference</td>
</tr>
<tr>
<td>LHS</td>
<td>Left-hand side</td>
</tr>
<tr>
<td>LoS</td>
<td>Line-of-sight</td>
</tr>
<tr>
<td>LT</td>
<td>Laplace transform</td>
</tr>
<tr>
<td>MAC</td>
<td>Medium access control</td>
</tr>
<tr>
<td>MAS</td>
<td>Multi antenna system</td>
</tr>
<tr>
<td>ME</td>
<td>Matrix exponential</td>
</tr>
</tbody>
</table>
mgf | Moment generating function
---|---
MH | Multihop
MI | Mutual information
MIMO | Multiple-input multiple-output
MISO | Multiple-input single-output
MIXO | MISO or MIMO
Mod. | Modulation
MRC | Maximal ratio combining
MR-HARQ | HARQ with mapping rearrangement
MRM | Multi-resolution-modulation
MSE | Mean-square-error
MU-MIMO | Multi-user MIMO
MUD | Multi-user detection
NACK | Negative-ACK
NC | Network coding
NCBR | Network coded bidirectional relaying
OR | Opportunistic routing
OSTBC | Orthogonal space-time block code
PAM | Pulse amplitude modulation
PAPR | Peak-to-average-power-ratio
PB | Probabilistic bell-(shaped)
pdf | Probability density function
PDU | Packet data unit
PEP | Pairwise-error-probability
Perf. | Performance
pmf | Probability mass function
Prop. | Proposition
PSK | Phase shift keying
QAM | Quadrature amplitude modulation
QoS | Quality-of-service
rads | Radians
re | Real component
RF | Radio frequency
RHS | Right-hand side
RMT | Random matrix theory
RP | Rate-parameter
RQ | Request
RR-HARQ | Repetition-redundancy HARQ, or simply RR
RRM | Radio resource management
RTS | Request-to-send in WiFi
r.v. | Random variable
RW | Random walk
RX | Receiver
SDC | Selection diversity combining
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIMO</td>
<td>Single-input multiple-output</td>
</tr>
<tr>
<td>SINR</td>
<td>Signal-to-interference-plus-noise ratio</td>
</tr>
<tr>
<td>SISO</td>
<td>Single-input single-output</td>
</tr>
<tr>
<td>SM-MIMO</td>
<td>Spatially-multiplexed MIMO</td>
</tr>
<tr>
<td>SNR</td>
<td>Signal-to-noise ratio</td>
</tr>
<tr>
<td>SP</td>
<td>Spiral phyllotaxis</td>
</tr>
<tr>
<td>SR-(H)ARQ</td>
<td>Selective-repeat (H)ARQ</td>
</tr>
<tr>
<td>STC</td>
<td>Space-time code</td>
</tr>
<tr>
<td>SVD</td>
<td>Singular value decomposition</td>
</tr>
<tr>
<td>SW-ARQ</td>
<td>Stop-and-wait ARQ</td>
</tr>
<tr>
<td>Tab.</td>
<td>Table</td>
</tr>
<tr>
<td>TDD</td>
<td>Time division duplexing</td>
</tr>
<tr>
<td>TX</td>
<td>Transmitter</td>
</tr>
<tr>
<td>UL</td>
<td>Uplink</td>
</tr>
<tr>
<td>VAA</td>
<td>Virtual antenna array</td>
</tr>
<tr>
<td>WiFi</td>
<td>Wireless Fidelity (IEEE 802.11 standard)</td>
</tr>
<tr>
<td>WLAN</td>
<td>Wireless local area network</td>
</tr>
<tr>
<td>w/wo</td>
<td>With/without</td>
</tr>
<tr>
<td>wrt</td>
<td>With respect to</td>
</tr>
<tr>
<td>XOR</td>
<td>Exclusive-OR gate</td>
</tr>
<tr>
<td>ZF</td>
<td>Zero-forcing</td>
</tr>
<tr>
<td>ZF-MIMO</td>
<td>Zero-forcing MIMO</td>
</tr>
<tr>
<td>2D</td>
<td>Two-dimensional</td>
</tr>
<tr>
<td>3D</td>
<td>Three-dimensional</td>
</tr>
</tbody>
</table>
In this chapter, we review fundamentals of communication theory and (H)ARQ as needed for the thesis. The Laplace transform is frequently encountered throughout the thesis, and for this reason, also reviewed in this chapter.

2.1 Communication Theory Fundamentals

(H)ARQ does not operate alone, but is generally an integral part of a complex communication system, with interaction to different communication entities and functions. The abstract modeling of such entities and functions, as useful for describing and analyzing (H)ARQ in this thesis, are introduced in the following. This involves information measures, antenna-based communication concepts, and some special network topology concepts.

2.1.1 Information Measures

As a foundation of the performance analysis in the thesis lies information theoretical concepts, such as capacity achieving codes and outage capacity. Below, we discuss the fundamental information measures in information theory, on which those concepts relies upon. For a more extensive treatment, see, e.g., [CT06].

In Fig. 2.1, a basic information theoretical communication system, a discrete memoryless channel, is illustrated. A source generates a message \( M \), uniformly selected from \( |M| \) possible messages. The message is mapped to a sequence of symbols \( X = \{X_1, X_2, \ldots, X_{N_M}\} \), where the symbols, \( X_n, n \in \{1, 2, \ldots, N_M\} \) are (in the following) assumed iid and selected from an alphabet \( \mathcal{X} \). The sequence is sent over a channel introducing errors. The channel is characterized by the conditional error probability \( p(y|x) \). The decoder observes a (possibly) distorted sequence, \( Y = \{Y_1, Y_2, \ldots, Y_{N_M}\} \), after the channel, from which it decodes (estimates) a message \( \hat{M} \).

The code-rate is defined as \( R_{\text{code}} \triangleq \log_2(|M|)/N_M \) bits per channel use [bits/cu]. In a practical wireless communication system, the encoder is often split into a channel
encoder and a modulator, and the decoder is often split into a demodulator and a channel decoder, respectively.

![Discrete (stationary) memoryless channel (DMC).](image)

**Entropy**

The entropy, is a measure of the average amount of information required to describe a discrete r.v. $X$, with an alphabet $\mathcal{X}$. The entropy for a discrete r.v. is always non-negative. It is defined as

**Definition 2.1. (Entropy) The entropy of a discrete r.v., $X \in \mathcal{X}$, with pmf $p(x) \triangleq p(X = x)$, is**

$$H(X) = - \sum_{x \in \mathcal{X}} p(x) \log(p(x)). \quad (2.1)$$

The joint entropy between two discrete r.v.s describes the uncertainty in the joint distribution. More specifically, it is defined as

**Definition 2.2. (Joint entropy) The joint entropy of two discrete r.v.s, $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$, with joint pmf $p(x, y)$, is**

$$H(X, Y) = - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log(p(x, y)). \quad (2.2)$$

The conditional entropy is the uncertainty in the conditional distribution of two r.v.s.

**Definition 2.3. (Conditional entropy) The conditional entropy of a discrete r.v. is**

$$H(Y|X) = - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x)p(y|x) \log(p(y|x)). \quad (2.3)$$
2.1. Communication Theory Fundamentals

Mutual Information

Based on the entropy concept, the mutual information measures the uncertainty of a r.v. $X$ after observing a r.v. $Y$.

**Definition 2.4. (Mutual information)** The mutual information of a discrete r.v. $X \in \mathcal{X}$, when observing a r.v. $Y \in \mathcal{Y}$, is

$$I(X;Y) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x,y) \log \left( \frac{p(x,y)}{p(x)p(y)} \right).$$

(2.4)

The mutual information can, equivalently, be written as

$$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X).$$

(2.5)

The information measures can also be extended to r.v.s with continuous distributions, then denoted differential entropy measures. For example, the differential entropy of a continuous r.v., $X \sim f_X(x)$, is

$$h(X) = -\int f_X(x) \log (f_X(x)) \, dx.$$ The differential entropy can, in contrast to the entropy of a discrete r.v., be negative.

Channel Capacity

The channel capacity is the highest data rate at which information can be communicated without errors. The notion of channel capacity assumes that the codeword length approaches infinity, $N_M \to \infty$, and it describes the amount of information that can be deduced from observing channel output r.v. $Y$ given an input r.v. $X$ with an optimal distribution wrt the channel. The channel capacity is defined as the maximum of the mutual information for a channel when the input probability distribution $p(x)$ is optimized over all possible input distributions.

**Definition 2.5. (Channel capacity)** The channel capacity of a discrete r.v. $X \in \mathcal{X}$, when observing a r.v. $Y \in \mathcal{Y}$, is

$$C = \max_{p(x)} I(X;Y) \quad [\text{b/cu}].$$

(2.6)

The channel capacity is also extendable to the continuous r.v. case. In this thesis, we are primarily concerned with the complex-valued AWGN baseband channel, modeled as

$$y = \sqrt{E_s} x + w,$$

(2.7)

where $y \in \mathbb{C}$ is the output signal, $E_s \in \mathbb{R}_{\geq 0}$ is the symbol energy, $x \in \mathbb{C}$ is the unit-variance input signal, and $w \in \mathbb{C}$ is the unit-variance Gaussian distributed noise. In the AWGN channel, an iid complex Gaussian input distribution with zero mean and fixed variance, maximizes the mutual information, and yields the
channel capacity. The average power-constrained and bandwidth-limited (and here bandwidth-normalized) channel capacity for the AWGN channel, given in [Sha48], is

$$C = \ln (1 + S) \text{ [nats/Hz/s]},$$  \hspace{1cm} (2.8)

where $S = \Bar{P}/BN_0 \in \mathbb{R}_{\geq 0}$ is the signal-to-noise-ratio (SNR), $B$ the bandwidth, and $N_0$ the noise power density. See, e.g., [CT06, pp. 270-273] for a more extensive treatment on the bandwidth limited channel. The AWGN channel capacity for a single transmit- and single receive-antenna system may be generalized to a communication system comprising $N_{tx}$ transmit- and $N_{rx}$ receive-antennas, a so called multiple-input multiple-output (MIMO) system. The baseband model can then be written

$$y = \sqrt{E_s} H x + w,$$ \hspace{1cm} (2.9)

where $y \in \mathbb{C}^{N_{rx} \times 1}$, $H \in \mathbb{C}^{N_{rx} \times N_{tx}}$, $x \in \mathbb{C}^{N_{tx} \times 1}$, and $w \in \mathbb{C}^{N_{rx} \times 1}$. When $w$ is componentwise Gaussian with covariance $R_w = I$, and $x$ is componentwise iid Gaussian distributed with covariance $R_x = E\{xx^H\}$, the mutual information is

$$I(y; x)_{\text{MIMO}} = \ln \det \left( I_{N_{rx}} + \frac{S}{N_{tx}} HR_x H^H \right) \text{ [nats/Hz/s].}$$ \hspace{1cm} (2.10)

The proof is sketched in Appendix 2.A. The channel capacity is achieved when $I(y; x)_{\text{MIMO}}$ is maximized under the constraint $\text{tr}(R_x) \leq \Bar{P}$, where $\Bar{P}$ is the average output power.

**Mutual Information of Block Fading Channel**

Assuming a block fading channel, as for our studied IR-HARQ scenario, the output distribution (generally) change from transmission to transmission of redundancy blocks. For this case, the accumulated (average) mutual information (AMI) is important in the analysis. Assume $K$ redundancy blocks, each block with $N_M$ channel uses. Then, under the iid assumption among symbols, we get the accumulated mutual information

$$I^{\text{Acc}} = \lim_{N_M \to \infty} \frac{I(Y; X)}{KN_M}$$

$$= \lim_{N_M \to \infty} \frac{I(Y_1, Y_2, \ldots, Y_{KN_M}; X_1, X_2, \ldots, X_{KN_M})}{KN_M}$$

$$= \sum_{k=1}^{K} \frac{I(Y_1^{(k)}, Y_2^{(k)}, \ldots, Y_{N_M}^{(k)}; X_1^{(k)}, X_2^{(k)}, \ldots, X_{N_M}^{(k)})}{N_M}$$

$$= \sum_{k=1}^{K} I(Y^{(k)}; X).$$ \hspace{1cm} (2.11)
where $Y^{(k)}$ is the output r.v. for the $k$th redundancy block, and assuming the same input distribution $X = X^{(k)}$ for each redundancy block.

Thus, the accumulated mutual information for $k$ independent codeword transmissions in AWGN, with average SNR $S$, and unit variance fading channel power gain $g_k$, is

$$I_{k}^{\text{Acc}} = \sum_{k'=1}^{k} \ln (1 + S g_{k'}) \text{ [nats/Hz/s]}.$$ (2.12)

### Information Outage

The notion of information outage is often used for wireless communication performance evaluation studies. Assume that a message is encoded into an infinitely long, capacity achieving, codeword of rate $R$, and sent over a noisy channel. Then, if the (accumulated) mutual information of the observed received codeword exceeds the initial code rate (including both the modulation and channel coding), the message can be recovered without any errors. More formally,

**Definition 2.6. (Information outage)** An information outage occurs, i.e. a message can not be decoded without error(s), if the (accumulated) mutual information is less than the rate, i.e.

$$I(Y; X) < R.$$ (2.13)

### 2.1.2 Antenna-based Communication Concepts

Multiple antenna channels, shown in Fig. 2.2, are categorized as single-input single-output (SISO), single-input multiple-output (SIMO), multiple-input single-output (MISO), or multiple-input multiple-output (MIMO) -channels. For the SIMO-channel, receiver processing allows, e.g., for maximal ratio combining, and selection diversity combining. For the MISO-channel, with transmit processing, transmit diversity (TX-diversity) with orthogonal space time block coding and cyclic delay diversity are possible. In contrast to CDD, OSTBC also requires special multiple antenna system signal processing at the receiver side. Many different communication designs have been proposed for the MIMO-channel. Here, we only briefly give some example schemes. We refer to, e.g., [Rap01, TV04, SA05, Mol05, LS08, Sha12] for a more extensive treatment on antenna-based communication schemes.

**MRC**

The baseband (column vector) signal model for MRC is

$$y = \sqrt{E_s} h x + w.$$ (2.14)
Assuming that the noise-components have the same variance, the post-processed scalar MRC signal, with appropriate SNR maximizing MRC weighting, is

\[ y_{\text{mrc}} = w_{\text{mrc}} y, \]  

(2.15)

where \( w_{\text{mrc}} = \frac{h^H}{||h||^2} \). The instantaneous channel capacity for MRC is then

\[ C_{\text{mrc}} = \ln \left( 1 + S \frac{1}{\text{max}} \sum_{n=1}^{N_{\text{rx}}} g_n \right) \quad [\text{nats/Hz/s}], \]  

(2.16)

where the instantaneous channel power gain is \( g_n = |h_n|^2 \).

**SDC**

Similarly, the received vector signal for SDC is

\[ y = \sqrt{E_s} h x + w. \]  

(2.17)

In SDC, the strongest vector signal component is selected, i.e.

\[ y_{\text{sdc}} = \max_n \{ y \}. \]  

(2.18)

The instantaneous channel capacity for SDC is therefore

\[ C_{\text{sdc}} = \ln(1 + S \text{max}_n \{ g_n \}) \quad [\text{nats/Hz/s}]. \]  

(2.19)
OSTBC - Alamouti TX diversity

Orthogonal space time block coding (OSTBC) [TSC98] enables (transmit) diversity with only a single receive antenna, but using multiple transmit antennas. The complex symbols to be transmitted, expressed on vector form, are \( \mathbf{x} = [x_1 \ x_2 \ldots \ x_{N_{\text{sym}}}]^T \). Through linear operations, the entries in \( \mathbf{x} \) are mapped to an OSTBC matrix \( \mathbf{X} \in \mathbb{C}^{N_{\text{tx}} \times N_{\text{x}}} \), with \( N_{\text{tx}} \) transmit antennas and \( N_{\text{x}} \) channel uses. The OSTBC matrix is designed to fulfill the condition \( \mathbf{X}^H \mathbf{X} = \| \mathbf{x} \|^2 \mathbf{I}_{N_{\text{x}}} \), thereby allowing for linear receiver signal processing. The OSTBC rate is defined as \( r_{\text{stc}} = N_{\text{sym}} / N_{\text{x}} \).

It has been shown in [WX03] that \( r_{\text{stc}} \leq 1 \), with equality only when \( N_{\text{tx}} = 2 \). To illustrate the operation, we exemplify with Alamouti’s TX-diversity scheme [AT97, Ala98]. It represents the simplest of OSTBC schemes, and is applicable to \( 2 \times 1 \) MISO antenna system using two symbol transmissions, and has \( r_{\text{stc}} = 1 \). For the Alamouti TX diversity case, the OSTBC matrix is

\[
\mathbf{X} = \frac{1}{\sqrt{2}} \begin{bmatrix} x_1 & x_2^* \\ -x_2 & x_1^* \end{bmatrix}
\]

The received (row) vector signal for the general OSTBC case is

\[
\mathbf{y} = \sqrt{E_s} \mathbf{h} \mathbf{X} + \mathbf{w}.
\]

where \( \mathbf{y} \in \mathbb{C}^{1 \times N_{\text{x}}} \), \( \mathbf{h} \in \mathbb{C}^{1 \times N_{\text{tx}}} \), and \( \mathbf{w} \in \mathbb{C}^{1 \times N_{\text{x}}} \). This expression can, after appropriate elementwise linear operation on \( \mathbf{y} \), be rewritten on an equivalent form

\[
\mathbf{y}_{\text{stc}} = \sqrt{E_s} \mathbf{H}_{\text{stc}} \mathbf{x} + \mathbf{w}_{\text{stc}},
\]

where \( \mathbf{y}_{\text{stc}} \in \mathbb{C}^{N_{\text{x}} \times 1}, \mathbf{H}_{\text{stc}} \in \mathbb{C}^{N_{\text{x}} \times N_{\text{sym}}}, \) and \( \mathbf{w}_{\text{stc}} \in \mathbb{C}^{N_{\text{x}} \times 1} \). Due to the orthogonal OSTBC design, \( \mathbf{H}_{\text{stc}} \) is an orthogonal matrix \( \mathbf{H}_{\text{stc}}^H \mathbf{H}_{\text{stc}} = c_{\text{stc}} \| \mathbf{h} \|^2 \mathbf{I} \), and \( c_{\text{stc}} \) is an OSTBC-rate dependent constant. For the Alamouti-TX diversity example, we have

\[
\begin{bmatrix} y_1 \\ y_2^{*} \end{bmatrix} = \frac{\sqrt{E_s}}{2} \begin{bmatrix} h_1 & h_2 \\ -h_2^* & h_1^* \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} w_1 \\ w_2^{*} \end{bmatrix}.
\]

The received vector \( \mathbf{y}_{\text{stc}} \) is then weighted with \( \mathbf{H}_{\text{stc}}^H \). The processed vector becomes

\[
\tilde{\mathbf{y}}_{\text{stc}} = \mathbf{H}_{\text{stc}}^H \mathbf{y}_{\text{stc}} = c_{\text{stc}} \| \mathbf{h} \|^2 \mathbf{x} + \mathbf{w}_{\text{stc}}^{*},
\]

Again, for Alamouti TX diversity

\[
\frac{1}{\sqrt{2}} \begin{bmatrix} h_1^* & -h_2 \\ h_2 & h_1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2^{*} \end{bmatrix} = (|h_1|^2 + |h_2|^2) \frac{\sqrt{E_s}}{2} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \frac{1}{\sqrt{2}} \begin{bmatrix} h_1^* w_1 - h_2 w_2^{*} \\ h_2 w_1 + h_1^* w_2 \end{bmatrix}.
\]

The instantaneous channel capacity for an ideal OSTBC with MRC (operating in a MIMO-channel \( \mathbf{H} \)) [LS08] is

\[
C_{\text{stc}} = r_{\text{stc}} \ln \left( 1 + \frac{S}{N_{\text{tx}} N_{\text{sym}}} \sum_{n=1}^{N_{\text{tx}}} \sum_{n'=1}^{N_{\text{sym}}} |\mathbf{H}_{n,n'}|^2 \right) \text{ [nats/Hz/s]}.
\]
CDD

Cyclic delay diversity is a transmit-diversity scheme proposed in [PatlL8], [Lar98], and subsequently in [DK01]. It is used in channel-coded OFDM to introduce (artificial) frequency selective fading without causing inter-symbol-interference (ISI) (as linear delay diversity would do). With channel coding over the OFDM-subcarriers, a diversity gain is provided without affecting the OFDM receiver structure. With $V_{\text{ofdm}}$ subcarriers, up to $V_{\text{ofdm}}$ transmit antennas, and hence $V_{\text{ofdm}}$-fold CDD diversity, can be designed. In CDD, the discrete OFDM signal of antenna branch $n \in \{1, 2, \ldots N_{\text{tx}}\}$ is cyclically shifted with delay $(n - 1)T_{\text{ofdm}}/V_{\text{ofdm}}$, $T_{\text{ofdm}}$ being the duration of the OFDM-symbol (excluding the cyclic prefix). Assuming frequency flat channels, the worst channel-fading case, the (pre-cyclic-prefix) per symbol time-domain signal received is

$$y[t] = \sqrt{\frac{E_s}{N_{\text{tx}}}} \sum_{n=1}^{V_{\text{ofdm}}} h_n x[t - (n - 1)T_{\text{ofdm}}/V_{\text{ofdm}}] + w[t], \; t = (0, T_{\text{ofdm}}),$$

(2.27)

where $[\cdot]_{T_{\text{ofdm}}}$ signifies a cyclic shift. Alternatively, the received (post-cyclic-prefix) discrete frequency-domain signal is

$$y[v] = \sqrt{\frac{E_s}{N_{\text{tx}}}} \tilde{h}[v] x[v] + w[v], \; v \in \{1, 2, \ldots V_{\text{ofdm}}\},$$

(2.28)

where the effective channel is

$$\tilde{h}[v] = \sum_{n=1}^{N_{\text{tx}}} h_n e^{-j \frac{2\pi (n-1)(v-1)}{V_{\text{ofdm}}}}.$$  

(2.29)

The channel capacity for CDD is

$$C_{\text{CDD}} = \frac{1}{V_{\text{ofdm}}} \sum_{v=1}^{V_{\text{ofdm}}} \ln(1 + g_v) \; [\text{nats/Hz/s}],$$

(2.30)

with $g_v = |\tilde{h}[v]|^2$. Note that CDD may use different delay structures than with $\{0, 1, \ldots N_{\text{tx}}-1\}$ OFDM baseband samples can be used. However, at closer scrutiny, the channel capacity, under a per antenna branch frequency flat channel assumption, is the same.

MIMO Communication Schemes

In this thesis we are not specifically concerned with MIMO-communication schemes per se, yet we will briefly encounter (H)ARQ performance evaluation for the MIMO channel where the channel capacity is given by (2.10). With this in mind, we just briefly mentioning some examples of well-known MIMO schemes. Such involves; Layered space-time codes (layered-STC), Space-time convolution codes (STCC),
2.1. Communication Theory Fundamentals

NCBR
TRX 1
NCBR
Relay
NCBR
TRX 2
A
B
A\oplus B
A, (A\oplus B)\oplus A \Rightarrow B
B, (A\oplus B)\oplus B \Rightarrow A

Figure 2.3: Network coded bidirectional relaying.

Space-time trellis codes (STTC), space-time turbo codes, unitary-STC (USTC), differential STC (DSTC), and algebraic-STC. Among layered-STCs is the implementation architecture of the channel capacity achieving Bell-labs layered space-time (BLAST) codes, Diagonal-BLAST (D-BLAST) [Fos96] and Vertical-BLAST (V-BLAST) [WFGV98]. When the channel-state is known, beamforming can be used to enhance SNR, or separate the multiplexed streams. Well-known beamforming approaches involves singular-value-decomposition (SVD), or eigen-decomposition of the MIMO channel. For a more extensive treatment of various MIMO-schemes, including the above, see e.g. [GLMZ06, DG07, Gli11].

2.1.3 Special Network Topology Concepts

In this thesis, in addition to the single transmitter – single receiver topology shown in Fig. 1.1, two further special network topologies are encountered, namely network coded bidirectional relaying (NCBR) with ARQ in Chapter 7, and network coded ARQ (NC-ARQ) in Chapter 8.

Network Coded Bidirectional Relaying

In Fig. 2.3, we illustrate NCBR. In traditional relaying between two transceivers (TRX), TRX 1 and TRX 2, using an intermediate relay, four transmissions are required two exchange two messages, A and B. With NCBR, only 3 transmissions are required to exchange messages A and B. This is achieved by letting the relay-node bit-wise XOR-encode message A and B, prior channel coding, and send message A\oplus B. Each TRX, can then, after channel decoding, decode the other TRX’s messages through bit-wise-XOR with its own transmitted (and stored) message. This idea was patented in [PatL43], subsequently presented in [LJS05, LJS06], and independently proposed in [WCK05]. It represents, in retrospect, a wireless channel variant (which offers a broadcast medium) of a degenerated butterfly network in [ACLY00].
Network Coded (Hybrid) ARQ

In Fig. 2.4, we illustrate NC-ARQ for two receiving users. The objective is to send different data to each user. The idea is to exploit the occurrence that users have correctly decoded each other’s data packets, but not their own. The transmitter is informed, via feedback, which packets are decoded by which user. At the instance, when a network coding opportunity arises, the transmitter encode data packets to the users together. For two users, bitwise XOR operation can be used. By using the overheard data packet, a user can decode their own data packet if correctly decoding the XOR-packet. The idea of NC-ARQ for different data-flows were introduced by the author in [PatL50, PatL57], [LJ06, Lar07a]. It was, e.g., shown that if the number of users approaches infinity, the transmitter throughput approaches one successfully decoded data packet per transmission. NC-ARQ can also be extended to RR- and IR-NC-HARQ, as suggested and analyzed in [LSKAT10, LSKAT13].

2.2 (H)ARQ Fundamentals

In the following, we review the practical aspects of (H)ARQ operation, and then discuss the historical development, as well as naming conventions, of (H)ARQ.

2.2.1 (H)ARQ Operation

In Fig. 2.5, a block-diagram of a wireless communication system with (H)ARQ functionality is exemplified. Based on this example system, we now discuss the operation of (H)ARQ from a more practical perspective.

On the transmit side, the TX medium access control (MAC) entity indicates a transmission opportunity to the ARQ entity, which then receives (or polls) a (potentially queued) data packet from the immediate higher layer. The ARQ entity append an ARQ-header containing (among other things) an ARQ packet sequence number. To this, a CRC word is then computed and appended at the end of the data
sequence. The resulting (H)ARQ data packet unit (PDU), is subsequently channel-encoded and modulated (Mod.). Assuming here that some multiple-antenna communication scheme is used, some further scheme-specific signal processing are generally needed. Here, it is indicated as an multiple antenna system (MAS), but the details are beyond the current (H)ARQ description. Next, the resulting (multiple-antenna) baseband (BB) signals are converted, and filtered, from discrete to continuous signal representations, up-mixed to radio frequency (RF) and amplified in the RF entity, and then emitted from the antennas.

At the receiving side, time-controlled by the RX-MAC entity, after the RF-signals propagated through the wireless channel, the antennas receive the signals, which are then down-mixed to baseband frequency. For the specific multiple-antenna scheme used, necessary multiple antenna baseband processing takes place. Then, the resulting baseband signals are demodulated (demod.) and channel-decoded. For RR- and IR-HARQ, the channel-decoding exploit soft information (baseband or log-likelihood-ratio samples) stored from previous transmissions representative of the corresponding data packet. Then, the CRC-syndrome of the estimated (H)ARQ PDU is computed. If the CRC-syndrome indicates pass, i.e. it is an all-zero word, an ACK is generated and returned to the transmitter, and

![Block diagram of (H)ARQ system.](image-url)
the data packet is forwarded to the immediate higher layer. Note that different (H)ARQ schemes may implement different acknowledgement policies. Some (H)ARQ schemes may only send ACKs, whereas some may send negative-ACKs (NACKs), or a combination of both. For SR-ARQ, as considered here, resequencing wrt sequence number generally also takes place prior forwarding data packets to the immediate higher layer. Note further that TX- and/or RX-functions, such as modulation, coding, and antenna schemes may be performed jointly for some communication schemes.

An example of the time-evolution for a retransmission operation is depicted in Fig. 2.6. For data packet 1, only one transmission is needed, before successful decoding, whereas data packet 2 is, in this case, first correctly decoded on the $K$th transmission. For HARQ, the decoding process takes stored information from the previous $k - 1$ transmissions into account in the decoding process, whereas for ARQ, only the last transmission of data packet is considered in the decoding operation. The probability of correctly decoding a data packet on the $k$th transmission, $P_k$, is also indicated in Fig. 2.6. The probability of failing to decode the data packet on the $k$th transmission, given that the previous $k - 1$ transmission attempts failed, is indicated as $Q_k$. Different retransmission policies can be adopted. For example, the transmitter may retransmit codewords (or incremental redundancy blocks for IR) an infinite number of times (persistent), up to $K$ times at most
(retransmission-truncated), or wrt a certain time (time-truncated). We will further
detail the considered (H)ARQ operation in Section 3. Next, we review the historical
development of (H)ARQ.

2.2.2 Historical Development and Naming Conventions

Perhaps not surprisingly, but the naming convention of ARQ-schemes have changed
over time. Hence, a brief historical account on ARQ-scheme development, naming
conventions, and contemporary terminology is useful.

The first electronic retransmission scheme, to the best of the author’s knowl-
edge, was introduced in a patent [VD40] for a telegraph system filed by Van Dureen
in 1940. Its primary intended use, according to Van Duuren, was for radio trans-
mission, countering static and fading disturbances of transmitted signals, but also
for wire transmission systems, e.g. avoiding disturbances from adjacent power lines.
It employed error detection with a form of even parity, by converting a five-bit Bau-
dot code to an eight-bit word with equal number of positive and negative symbols.
Van Duuren also included the retransmission invention in his PhD thesis [VD41].
At this early stage, the invention had not yet been named as ARQ. In a subsequent
patent, [VD49], Van Dureen refined the retransmission idea introduced in [VD40],
and used the term (retransmission) request. In a patent [Can53] by Canfora, au-
tomatic request and repetition telegraph signalling system was considered and the
term automatic request was used. Dalen and Dureen introduced a more modern
(retransmission system related) language and nomenclature, referring to automatic
error detection (ED) and error correction (EC), in the patent [vDVD55]. The
first use of the term automatic repeat request seems to have appeared in [Stu57],
an editorial column in IRE Transactions on Information Theory. In the patent
[LCDP53], the inventors used, possibly for the first time, the abbreviation ARQ,
but with the meaning of automatic request. Moore discusses automatic-RQ, and
the abbreviation ARQ, in [Moo60], and states that automatic-RQ was in service
between New York and Amsterdam already in 1947. One of the earliest perfor-
ance analysis works of ARQ, in 1961, is also by Van Dureen, [VD61]. Thus, ARQ
was invented in the early 40s, but it appears that the terms automatic repeat re-
quest and ARQ cemented first in the early 60s. In parallel with those early, more
practically motivated, works, communication with feedback was studied from an
information theoretical point of view in [Sha56, Cha56] by Shannon and Chang.

During the later part of the 20th century, many ARQ schemes were developed
and analyzed. To provide enhanced performance over the original ARQ approach,
i.e. using an uncoded dataword, hybrid-ARQ, a combination of retransmission
with forward-error-correction (FEC), was developed. It is worth noting that al-
ready in 1961, Metzner and Morgan argued in [MM62] for a combined coded feed-
back approach, essentially introducing hybrid-ARQ. Good overviews of classical
ARQ and classical hybrid-ARQ schemes, as viewed about 1980-2000, are found in
[LC04, CHIW98, LCM84]. In those early years, classical hybrid-ARQ was often cat-
egorized in type-I/II/III hybrid-ARQ. In type-I hybrid-ARQ, data is FEC-encoded
and has error correction, and error detection, capability. In type-II hybrid-ARQ, transmissions alternate between sending systematic data bits together with a CRC. Then, if needed, FEC parity bits is sent in a subsequent transmission that can be decoded together with the systematic bits. In type-III HARQ, self-decodeable packets are sent in each transmission, where each packet can be decoded jointly if needed. Today, *HARQ with soft combining*, is often employed in wireless communication systems. This kind of HARQ scheme has historically been divided in so called Chase combining-HARQ [Sin77, Ben85, Cha85] (CC-HARQ) and incremental-redundancy-HARQ [Man74] (IR-HARQ). IR-HARQ is often designed to be self-decodeable, as for type III-HARQ. Another classical categorization of ARQ schemes is in so called stop-and-wait (SW), go-back-N (GBN), and selective repeat (SR) ARQ, indicating if one packet is transmitted at a time, a group of packets (from the last failed packet) must be retransmitted, or a selection of failed outstanding packets are retransmitted. The original definition of ARQ, i.e. sending a message as an uncoded dataword, does not make much sense in today’s wireless communication systems, as channel coding is (essentially) always used. The division in Type-I and -II HARQ, does also not make much sense in today’s wireless communication system, since self-decodeable codewords can be designed with modern channel coding schemes. Moreover, SW- and GBN-ARQ were primarily developed due to buffer-size constraints at the receiver, but, nowadays, memory-size constraints are seldom a severe practical problem. As SR-(H)ARQ schemes also deliver higher throughput than both SW- and GBN-ARQ schemes, SR-(H)ARQ is generally preferred. For completeness, we also mention that a line of works, from the 80s and onwards, considers truncated-ARQ schemes, which implies accepting packet-losses but trying to limit individual packet delays. In its original version, retransmission-truncated-ARQ, an upper limit on the number of retransmission attempts, a transmission limit, is imposed. This was introduced in [CFC80], and subsequently investigated further in [YB93]. Another variant of truncated-ARQ, is time-truncated-ARQ schemes, imposing either a maximum time limit for the retransmissions, or a source-defined message due-time, after which the message is not retransmitted [PatL5]. In this thesis, we consider only retransmission-truncated-(H)ARQ, and let term truncated-(H)ARQ imply retransmission-truncated-(H)ARQ.

The naming convention for the ARQ schemes studied in this thesis aims to follow a more modern nomenclature. The ARQ-schemes are denoted ARQ, RR-HARQ (RR), and IR-HARQ (IR). In this contemporary notation, all three schemes use channel encoding, selective-repeat operation, and where we let HARQ stand for joint processing of (potentially) multiple received noisy codewords at the receiver. The detailed operation of each scheme are described in the system model section, Chapter 3.3. Note that the modern naming convention of ARQ, i.e. assuming channel encoded messages with no combining at the receiver (often implicitly assumed due to the use of an AWGN channel channel capacity model), is used in several recent works, e.g., [CGiFRC08, SLF08, LRS14c, HGS16]. Note also that, in this thesis, the term RR-HARQ is used instead of CC-HARQ. This is to align the terminology of RR- and IR-HARQ, i.e. both schemes focus on what is sent at the
transmitter side. This can be contrasted with the term CC-HARQ which focus on the combining at the receiver side instead.

### 2.3 The Laplace Transform

The laplace transform of the function \( f(x) \) is defined as

\[
\mathcal{L}_s\{f(x)\} \triangleq \int_0^\infty e^{-sx} f(x) \, dx.
\]

We let \( F(s) \triangleq \mathcal{L}_s\{f(x)\} \). The inverse Laplace is defined as

\[
\mathcal{L}_s^{-1}\{F(s)\} \triangleq \lim_{t \to \infty} \frac{1}{2\pi i} \int_{-i\epsilon}^{+i\epsilon} e^{sx} F(s) \, ds,
\]

where \( \epsilon \) is large enough to have all poles to the left of the line of integration. We then get \( f(x) = \mathcal{L}_s^{-1}\{F(s)\} \). Some basic Laplace transforms are shown in Table 2.1. An extensive table of Laplace transforms is given in [Spi90, pp. 164-173].

<table>
<thead>
<tr>
<th>( F(s) ), ( n \in \mathbb{Z}_+ )</th>
<th>( f(t) ), ( x \geq 0 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( 1 )</td>
<td>( \delta(x) )</td>
</tr>
<tr>
<td>( 1/s )</td>
<td>( 1 )</td>
</tr>
<tr>
<td>( 1/(s + a)^n )</td>
<td>( e^{-ax} x^{n-1}/(n-1)! )</td>
</tr>
<tr>
<td>( s/(s^2 + a^2) )</td>
<td>( \cos(ax) )</td>
</tr>
<tr>
<td>( a/(s^2 + a^2) )</td>
<td>( \sin(ax) )</td>
</tr>
<tr>
<td>( F(s/a)/a )</td>
<td>( f(ax) )</td>
</tr>
<tr>
<td>( F(s-a) )</td>
<td>( e^{-at} f(x) )</td>
</tr>
<tr>
<td>( F(s)^n )</td>
<td>( f(x)^{n}(x) )</td>
</tr>
</tbody>
</table>

Table 2.1: Basic Laplace transforms.

#### 2.4 Appendices

### 2.A Proof of equation (2.10)

**Proof.** The MIMO baseband channel is \( y = \sqrt{E_s} Hx + w \), where we let \( R_{xx} = \mathbb{E}\{XX^H\} = I_{N_{tx}}/N_{tx} \), and \( R_{ww} = \mathbb{E}\{WW^H\} = I_{N_{rx}} \). The mutual information of the output \( y \) wrt to the input \( x \) is

\[
I(y; x) = h(y) - h(y|x) = h(y) - h(w).
\]

The differential entropy (in nats) of the noise vector is

\[
h(w) = \ln \det (\pi e R_{ww}).
\]
The differential entropy of the received signal is maximized if the input distribution of the entries in \( x \) are iid complex Gaussian. The differential entropy is then

\[
h(y) = \ln \det (\pi e R_{yy}),
\]

with

\[
R_{yy} = \mathbb{E} \{ Y Y^H \} = SHR_{xx} H^H + R_{ww} = \frac{S}{N_{tx}} HH^H + I_{N_{rx}}.
\]

Hence, the channel capacity is

\[
C_{\text{MIMO}} = \max_{f(x)} I(y; x) = \ln \det (\pi e R_{yy}) - \ln \det (\pi e R_{ww}) = \ln \det \left( R_{yy} R_{ww}^{-1} \right) = \ln \det \left( I_{N_{rx}} + \frac{S}{N_{tx}} H H^H \right) \text{[nats/Hz/s]}.
\]

\( \square \)
We now consider the performance measure(s) and system model used in the thesis. The overall system model is divided in a time division-, (H)ARQ-, and physical layer-model, as described in this chapter. The physical layer model in turn, is divided in three different abstraction levels. The most detailed one, reflecting actual wireless communications systems, also includes fading channel model(s).

Note that in some chapters of the thesis, certain details of the system model are extended. Specifically, in Chapter 7, a new and more general fading channel model is introduced and analyzed for (H)ARQ, whereas in Chapter 8, the HARQ models are extended to a more general retransmission system. Also, in Sections 7.5.7, 7.5.6, and 8.4.5, the network topology with interferers, 3-phase bidirectional relaying, and two receivers, are considered, respectively.

3.1 Performance Measures

The primary performance measures in the thesis is the throughput, $T$. This metric is selected since it, we argue, best captures the experienced service of interest for a user. To determine the throughput, some secondary performance measures often also need to be determined, such as the average number of transmissions, $\tau$, the mean rate, $\rho$, the $k$th decoding failure probability, $Q_k$, and the packet loss rate, $\epsilon_{\text{loss}}$. In Chapter 8, we go beyond the throughput performance measure, and explore the effective capacity performance measure, $C_{\text{eff}}$, for the case when a stochastic delay constraint is imposed. Conveniently, the throughput can be determined as a special case directly from the effective capacity. The performance measures are formally defined in Chapters 4 and 8, respectively.

3.2 Time Division Model

The communication system is comprised of a single transmitter and a single receiver as shown in Fig. 1.1, unless systems with NC-ARQ, NCBR, or interference are studied. The communication from the transmitter to the receiver takes place
in a forward channel of constant bandwidth which is divided in time slots. The signal, corresponding to a (H)ARQ transmission, fits exactly in such a time slot. While real-world systems would separate retransmissions with some intermediate time slots, we make the simplifying assumption that retransmissions can take place in consecutive time slots. Each such time slot undergoes an iid block fading channel gain. A reverse channel exist from the receiver to the transmitter for feedback, acknowledgement, and negative acknowledgement. It is assumed that the feedback is error-free, and fast enough to be available prior the next transmission. The forward and the reverse channels are arranged to avoid any cross-interference. The bandwidth-normalized information rate per transmission is denoted by $R$, i.e. the amount of information sent (in nats), normalized with the transmission time and bandwidth. The overhead, due to any header and error detection CRC-word, is assumed to be negligible. The error detection operation is also assumed ideal, i.e. omitting the practical aspect of CRC-misdetection. In a multiple access system where multiple transmissions occur and interfere with each other on the same channel, all transmissions are assumed synchronized wrt the time slot structure. The time division model is schematically illustrated in Fig. 3.1.

### 3.3 (H)ARQ Models

In this thesis, we consider ARQ, RR-HARQ, and IR-HARQ. The operations of each scheme is described below, as well as summarized in Tab. 3.1.

- In ARQ, a message is first channel-encoded into a codeword, and then mapped to a sequence of modulation symbols, by the transmitting node. The resulting initial transmission rate is denoted by $R$ and is expressed in the unit [nats/Hz/s]. The same codeword (including the modulation mapping) is (re)transmitted until it can be correctly decoded by the receiving node. Only the last received codeword for a message is channel decoded, and the previously received codewords are not taken into account, nor stored. For performance analysis of ARQ, the MI of only the last transmission is used.
### 3.4 Physical Layer Model

<table>
<thead>
<tr>
<th>ARQ-Schemes:</th>
<th>ARQ</th>
<th>RR-HARQ</th>
<th>IR-HARQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Channel coding:</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>TX signal:</td>
<td>CW repeated</td>
<td>CW repeated</td>
<td>Different IR-blocks</td>
</tr>
<tr>
<td>RX processing:</td>
<td>Most recent CW</td>
<td>MRC of all CWs</td>
<td>Decoding of all IR-blocks</td>
</tr>
</tbody>
</table>

Table 3.1: (H)ARQ models.

- In RR-HARQ, a message is channel-encoded into a codeword, and mapped to modulation symbols, resulting in an initial rate $R$. The codeword is (re)transmitted until it can be correctly decoded. Each received (noisy) codeword for a message are maximal ratio combined before channel decoding. For performance analysis of RR, the accumulated SNR for all the transmissions is utilized.

- In IR-HARQ, a message is channel-encoded into a low-rate codeword, which is then split into several redundancy blocks. Each redundancy block is mapped to a sequence of modulation symbols. The initial rate for the redundancy block is also $R$. For the analysis, however, it is assumed that the transmitter can generate an unlimited number of redundancy blocks, each one assumed self-decodeable. Redundancy blocks are (re)transmitted, and channel-decoded jointly, until the message can be correctly decoded. And finally, for performance analysis of IR, the accumulated MI for all the transmissions is exploited.

The above schemes, are further divided into persistent- and truncated-(H)ARQ.

**Definition 3.1.** Define $K$ as the maximum allowed number of transmissions of a message, after which the message is dropped if the message has not been correctly decoded. When a retransmission scheme has $1 \leq K < \infty$, it is said to be truncated, whereas when $K = \infty$, it is said to be persistent.

### 3.4 Physical Layer Model

How can the physical layer be modeled? In this thesis, the performance measures, i.e. mainly the throughput $T$ and the effective capacity $C_{\text{eff}}$, are expressed in parameters corresponding to one of three physical layer model levels, a decoding failure probability – based level, an effective channel pdf (Laplace transform) – based level, or a wireless communication system – based level. The first level is the most general one and could model any communication system, even non-wireless systems. This is the level where much (H)ARQ analysis in the literature stops at. The second level is also relatively general model, only assuming a general iid block fading channel pdf (or its LT) and a decoding threshold. The third level is the most detailed one, and connects directly to state-of-the art wireless communication systems. This third
level is where just relatively few closed-form (H)ARQ expressions have been given in the literature. With this three-level division, analytical performance expressions can be developed wrt model generality and specialization, i.e. according to need and viability. Below, we discuss the three levels in detail, and summarize those in Fig. 3.2.

### 3.4.1 Decoding Failure Probability - Based Model

The following two definitions are useful.

**Definition 3.2.** The kth (conditional) decoding failure probability, $Q_k$, is the probability that the decoding of the kth (codeword) transmission of a message fail, given that the decoding of the previous $k-1$ (codeword) transmissions failed.

Correspondingly, we also have the following definition.

**Definition 3.3.** The kth decoding success probability, $P_k = Q_{k-1} - Q_k$, is the probability that the decoding of a message succeed on the kth (codeword) transmission.

In the decoding failure probability based-model, a performance measure is expressed in terms of $Q_k$. This is useful as a first step of performance derivation if $Q_k$ is given, or if the fading channel pdf, or the communication system, are unknown. Specifically, the throughput depends only on $R$ and $Q_k$ as

$$T = f_{(H)ARQ}(R, Q_k), \quad k \in \{1, 2, \ldots, K\},$$

where the subscript in $f_{(H)ARQ}(\cdot)$ also signifies the dependence of scheme used.
3.4. Physical Layer Model

3.4.2 Effective Channel pdf - Based Model

For a (H)ARQ system experiencing iid block fading channel, the $k$th decoding failure probability, $Q_k$, can typically be determined using the pdf, $f_G(g)$, of the r.v. fading channel gain $G$, and a decoding threshold $\Theta$. More generally, we assume that $Q_k$ can be computed after possible signal processing, such as diversity processing. We define an effective channel r.v. as follows.

Definition 3.4. The effective channel is the postprocessing SNR (or MI) r.v. $Z$ with pdf $f_Z(z)$, and where $F(s)$ denotes the Laplace transform of $f_Z(z)$.

Remark 3.1. In the sequel, a decoding threshold, $\Theta$, is generally also associated with the effective channel pdf, $f_Z(z)$, or the LT, $F(s)$.

In the effective channel pdf-based model, the performance measures is expressed in terms of $F(s)$, or alternatively in $f_Z(z)$. Specifically, for $F(s)$, the throughput is expressed as a function

$$T = f_{(H)ARQ}(R, S, K, F(s)).$$  \hspace{1cm} (3.2)

We let $f_{(H)ARQ}(\cdot)$ depend on both rate, $R$, and SNR, $S$, since the decoding threshold $\Theta$ could, as will be seen, either be expressed in both $R$ and $S$, or just in rate $R$. It will also be seen that this model is suited for many, though not all, (H)ARQ-cases of interest. Note that in Chapter 7, the effective channel model is specialized to the ME-distributed channel model introduced in Chapter 7.

3.4.3 Wireless Communication System - Based Model

We seek not just a (H)ARQ performance expression, expressed in $Q_k$ or $F(s)$, but also performance expressions for specific wireless communication cases. A relatively general wireless communication system model on this level, is a multiple antenna system comprised of a MIMO block-fading channel and some MAS-scheme of choice. Such general model is shown in Fig. 3.3.
This general multiple antenna scheme is comprised of a transmitter with $N_{tx}$ antennas, and a receiver with $N_{rx}$ antennas, communicating over a channel $H$. Each transmission has an equivalent receiver baseband model, which is

$$Y = \sqrt{E_s} H X + W,$$

where $Y \in \mathbb{C}^{N_{rx} \times N_x}$ is the received signal matrix, $H \in \mathbb{C}^{N_{rx} \times N_{tx}}$ is the channel matrix, $X \in \mathbb{C}^{N_{tx} \times N_x}$ is the transmitted signal matrix, and $W \in \mathbb{C}^{N_{rx} \times N_x}$ is the noise matrix, and $N_x$ is the number of channel uses. For the noise matrix, we assume iid AWGN entries, $w_{ij} \sim \mathcal{CN}(0,1)$. The SISO average SNR is denoted by $S$. The channel matrix $H$ is assumed to remain constant for the duration of a (H)ARQ transmission, and then it changes randomly to the next transmission. It is further assumed that the receiver has, and exploits, full channel-state knowledge. Depending on the design of the signal matrix, $X$, and the corresponding MAS-scheme; SM-MIMO, OSTBC [TSC98], Alamouti’s $2 \times 1$ TX-diversity [AT97, Ala98], MRC, and plain SISO can be handled. It is generally assumed that the modulation and coding scheme is modeled with Gaussian distributed signal(s) with constant average power, and that the channel coding is of sufficient high dimensionality such that the AWGN channel capacity limit can be assumed. This also applies to those entries in $X$ which represent different signal components.

The performance expression for this more detailed communication scenario involves the initial rate, SNR, HARQ scheme with parameters, MAS with parameters, fading channel (FC) model with parameters, and has the generic form

$$T = f_{(H)ARQ,MAS,FC}(R,S,K,N_{tx},N_{rx},\ldots). \quad (3.3)$$

We consider two kinds of communication system models. The first is the general diversity channel, which jointly models OSTBC, MRC, and (integer restricted) Nakagami-$m$ fading, whereas the second (just briefly considered in the thesis) is channel capacity achieving SM-MIMO.

**General Diversity Channel Model – OSTBC, MRC, and Nakagami-$m$**

Orthogonal space time codes (OSTBC), [AT97, Ala98, TSC98], may be, and often is, used in real world wireless systems to counter the effects of fading channels. At the same time, those systems may employ RR (or ARQ). Therefore, it is of interest to analyze RR and OSTBC jointly. An alternative, or complementary diversity scheme that may also be combined with RR, is based on receiver-based maximal ratio combining. A relatively general (amplitude) fading model for wireless channels$^1$, is the Nakagami-$m$ fading channel model [Nak60]. It spans over a wide range of fading conditions through the parameter $m N$. For example, $m N = 1$

$^1$Other (amplitude) fading channel models, like Ricean, Hoyt etc, are possible. However, we limit the study to Nakagami-$m$ here, as an even broader class of fading channels is introduced in Chapter 7.
corresponds to Rayleigh fading, and, in the limit, \( m^N \to \infty \), it converges to a non-fading channel.

The GD-channel is defined below. See also Ex. 4.5 for further motivations.

**Definition 3.5.** (General diversity channel) The GD-channel is characterized by a post-SNR r.v., \( Z_{gd} = S_{gd}Z \), where \( Z \) is a unit-variance r.v. with pdf
\[
f_Z(z) = \frac{z^{N_{gd}-1}}{(N_{gd} - 1)!} e^{-z}, \tag{3.4}
\]
which has LT
\[
F(s) = \left( \frac{1}{1 + s} \right)^{N_{gd}}, \tag{3.5}
\]
where \( N_{gd} \triangleq N_{tx}N_{rx}m^N \) is the GD effective channel diversity order, \( S_{gd} \triangleq S / m^N r_{stc} N_{tx} \) is the GD effective channel SNR, \( R_{gd} \triangleq R/r_{stc} \) is the GD effective channel rate, \( r_{stc} \) is the OSTBC code rate, and \( m^N \) is the integer-restricted Nakagami-\( m \) parameter.

Specifically, the throughput for the GD-channel is expressed as a function
\[
T = f_{(H)ARQ,GD}(R, S, K, r_{stc}, N_{tx}, N_{rx}, m^N). \tag{3.6}
\]

The GD-channel model can capture many interesting special cases, e.g. Alamouti TX diversity in Rayleigh fading when \( r_{stc} = 1, N_{tx} = 2, N_{rx} = 1, m^N = 1 \).

**SM-MIMO Model**

In this thesis, most of the performance analysis deals with the GD-channel. However, we briefly encounter SM-MIMO in Chapters 4 and 7. The instantaneous SM-MIMO MI, as discussed in Chapter 2, can be modeled with the effective channel r.v.
\[
Z_{MIMO} = \ln \det \left( I_{N_{tx}} + \frac{S}{N_{tx}} H H^H \right), \tag{3.7}
\]
where \( H_{ij} \sim CN(0, 1) \). The joint pdf of the unordered eigenvalues of the complex central Wishart matrix, i.e. \( W = H H^H \) if \( N_{tx} < N_{rx} \), or \( W = H^H H \) if \( N_{tx} \geq N_{rx} \), was shown in [Tel99] to be
\[
f(\lambda_1, \ldots, \lambda_N) = K_{N_{tx},N_{rx}} \exp \left( -\sum_{n=1}^{N} \lambda_n \right) \prod_{n=1}^{N} \lambda_n^{N_{max} - N_{min}} \prod_{n<n'} \left( \lambda_n - \lambda_n' \right)^2, \tag{3.8}
\]
where \( N_{max} = \max\{N_{tx}, N_{rx}\} \), \( N_{min} = \min\{N_{tx}, N_{rx}\} \), and \( K_{N_{tx},N_{rx}} \) is a normalizing constant.
Chapter 4

Basic Throughput Analysis of (H)ARQ

In Chapter 1, the overall problem to study was formulated and motivated. Then, in Chapter 2 and 3, the preliminaries and the system model were given. In this chapter, we establish a foundation for performance analysis of fundamental (H)ARQ-cases with basic fading channel model(s). The analysis acts as a stepping stone for subsequent chapters. Specifically, persistent- and truncated-transmission cases for ARQ, RR- and IR-HARQ, are studied. We give analytical throughput performance expressions, as well as expressions for the outage probability and average number of transmissions, in terms of an effective channel, and for Rayleigh and Nakagami-\(m\) fading channels. Transmission schemes including, SISO, MRC, OSTBC, and partially also SM-MIMO, are addressed. For persistent-IR-HARQ accurate throughput bounds are developed. The main outcome of this chapter is a performance analysis framework based on the Laplace transform of the effective channel pdf, as well as a number of new closed-form performance expressions for important (H)ARQ-cases not previously considered.

4.1 Motivation and Outline

Although ARQ, RR and IR, have been studied for a long time, surprisingly few information-theoretically based closed-form throughput expressions appear to exist even under relatively ideal assumptions. Exceptions, where throughput expressions are known, are for ARQ with SISO communication in Rayleigh fading [BS06, (8)-(9)], [SLF08, (5)-(6)], and similarly for RR [LSKAT10, (20)], [KJSS10, (7)]. We note that while a general throughput expression for persistent-HARQ exists in the literature, [CT01, Tun11, MGiAE12], the disadvantage with the expression is that it involves an infinite sum of decoding failure probability terms for the \(k\)th transmission. In addition, each one of the terms need to be known and expressed on an analytical form. Hence, a simpler more direct expression, avoiding the need of summing and computing of all decoding failure probability terms for the \(k\)th transmission, would be beneficial. Moreover, closed-form throughput expression for (H)ARQ-cases involving persistent/truncated-transmission and basic antenna
diversity schemes, are desirable to examine as they are not extensively treated in the literature. A further motivation of this chapter is, as will be seen, to lay the foundation for Chapter 7.

The outline of the chapter is the following. For the performance analysis, we start with the general throughput analysis in Section 4.2.1. A throughput analysis for ARQ and persistent/truncated-RR with GD-channel is given in 4.2.2. In Section 4.2.3, throughput bounds for persistent/truncated-IR with Rayleigh fading channel are considered, as well as the LT of the channel capacity pdf for SISO (persistent-IR) and $N \times N$ MIMO. Towards the end, numerical results and discussions take place in Section 4.3, and in Section 4.4, we summarize and conclude.

4.2 Performance Analysis

We start the study by considering a general HARQ throughput expression, formulated in the Laplace transform of the fading channel pdf.

4.2.1 General Throughput Analysis

In our search for a general HARQ throughput expression, we first consider a known throughput expression, expressed in decoding failure probabilities. For this purpose, it is convenient to start with truncated-HARQ, and then specialize to persistent-HARQ and ARQ. Subsequently, we get rid of the decoding failure probabilities, and base the throughput expression on the Laplace transform of the pdf of a generic channel. This expression is subsequently specialized to a generic diversity channel.

Several works, e.g. [WL83, (11)], [ZR96, (5)], and [CT01, (9)-(11)], have considered a throughput definition of (H)ARQ based on the renewal reward theorem [GS92, (16)]. This is captured in the following definition.

Definition 4.1. ([WL83, (11)], [ZR96, (5)], [CT01, (9)-(11)]) The throughput of (H)ARQ is, when the limit exists, defined as

$$T = \lim_{t \to \infty} \frac{R E \{V_t\}}{t} = \frac{\rho}{\tau},$$

(4.1)

where $R$ is the initial coding rate (hereafter simply called the rate)$^4$, $V_t$ is a r.v. representative of the number of messages correctly decoded at slot $t$, $\rho$ is the average rate per message, and $\tau$ is the average number of transmissions per message. The right-hand side (RHS) equality in (4.1) is considered in (8.11) and Corollary 8.4 with proofs.

$^4$Throughout the thesis, to avoid a factor $\ln(2)$ in the analysis, $T$ and $\rho$ are expressed in the unit [nats/Hz/s]. For convenience, however, the unit [bits/Hz/s] is used for $T$ and $\rho$ in the figures.
Based on the RHS of (4.1), \( T = \rho/\tau \), a general throughput expression for truncated-HARQ, defined below, can be given, in terms of decoding failure probabilities. This is a well-established result, see e.g. [CT01], [SCA10, (6)], which we summarize below in the following lemma.

**Throughput on Decoding Failure Probability-Form**

**Lemma 4.1.** Let \( R \) be the rate, \( K \) the maximum number of allowed transmission attempts per message, and \( Q_k \) is the decoding failure probability for the \( k \)th transmission. Then, the throughput of truncated-HARQ, is

\[
T_K(R, Q_k) = \frac{R(1 - Q_K)}{\sum_{k=0}^{K-1} Q_k},
\]

where, for notational convenience, \( Q_0 \triangleq 1 \).

**Proof.** In terms of decoding probabilities, the average rate at the \( K \)th transmission is

\[
\rho_K(R, Q_K) = R \sum_{k=1}^{K} P_k = R \sum_{k=1}^{K} (Q_{k-1} - Q_k) = R(1 - Q_K),
\]

whereas the corresponding average number of transmissions per message is

\[
\tau_K(Q_k) = \sum_{k=1}^{K} kP_k + KQ_K = \sum_{k=1}^{K} k(Q_{k-1} - Q_k) + KQ_K = \sum_{k=0}^{K-1} Q_k,
\]

where \( P_k = Q_{k-1} - Q_k \) is the probability of successful decoding at the \( k \)th transmission. With (4.3) and (4.4) in (4.1), we get (4.2).

The motivation for truncating the number of transmissions in truncated-HARQ is to (in some sense) limit the delay for each transmission before a message is discarded. It is vital to recognize that this is a mathematical model, whereas in practical systems, messages may instead be discarded after a due-time. The analysis of truncated-HARQ is not without challenges, however. Even for a moderate \( K \), with analytical expressions for \( Q_k \) given, relatively complicated (analysis and) expressions (may) result, providing little, or no, insights. This in turns (may) complicate analytical throughput optimization. For those reasons, and other discussed below, persistent-HARQ is also considered in the thesis.

As suggested above, persistent-HARQ often gives simpler expressions and enables easier analysis than for truncated-HARQ. For example, in the low-SNR range where large numbers of retransmissions are required (large value of \( K \)), the throughput expressions of truncated-HARQ (can) become very complicated. When considering the maximum throughput (wrt the rate) of persistent- and truncated-HARQ, we expect the performance to be nearly identical. The reason for this is that
the average number of transmissions for throughput optimal (H)ARQ are often of roughly one-to-two transmissions per message. We also note that traditional analysis in the literature of (H)ARQ systems assumes infinite $K$, and that the attention to truncated-(H)ARQ is a relatively new focus of HARQ-research. Notably, as will be seen in Chapters 7 and 8, the persistent-HARQ case allow for very compact and general performance expressions. This concludes the motivations for studying the persistent-HARQ case. In the proposition below, the throughput of persistent-HARQ is considered.

**Proposition 4.1.** ([CT01]) Let $R$ be the rate, and $Q_k$, $k = \{1, 2, \ldots \}$, is the decoding failure probability for the $k$th transmission of a message. Then, the throughput of persistent-HARQ, is

$$T_{HARQ}^\infty (R, Q_1, Q_2, \ldots) \triangleq \frac{R}{\sum_{k=0}^\infty Q_k},$$  \hspace{1cm} (4.5)

*Proof.* Let $K \to \infty$ in (4.1), and take note that $Q_K \to 0$, which gives (4.5). $\Box$

At this point, it is instructive to consider truncated-ARQ in the context of the truncated-HARQ throughput expression (4.2).

**Definition 4.2.** Truncated-ARQ is defined as truncated-HARQ without memory.

**Proposition 4.2.** Let $R$ be the rate, and $Q_k$, $k = \{1, 2, \ldots \}$, is the decoding failure probability for the $k$th transmission of a message. Then, the throughput of truncated-ARQ, is

$$T_{ARQ}^K (R, Q_1) = R(1 - Q_1).$$  \hspace{1cm} (4.6)

*Proof.* Due to independence and memoryless processing, the decoding failure probability is geometrically distributed with $Q_k = Q_1^k$. Using this in (4.5), computing the geometric series, then gives (4.6). $\Box$

**Remark 4.1.** We note that the throughput of truncated-ARQ is independent of the maximum number of transmissions $K$. Hence, in the subsequent analysis\(^5\), we do not distinguish between the truncated- and persistent-ARQ-cases.

While the throughput expression of ARQ has a very simple form, it present its own challenges as the outage probability, $Q_1$, can in practice be hard to determine. This is, e.g., the case for ARQ operating with SM-MIMO, see Section 4.2.3, 2 × 1 CDD, see [LRS14c], or in a fading interference environment, see Section 7.5.7.

\(^5\)Note that using $K = 1$, $K = \infty$, or any other $K$, can affect an implementation greatly.
4.2. Performance Analysis

The Effective Channel

If $Q_k$ are given as plain numbers, the analysis do not need to be extended beyond (4.5). However, this is seldom the case. If $Q_k$ are given as functions of other parameters, as is the typical case for wireless system analysis, (4.5) requires further examination. Before illustrating the challenges for analysis of (4.5), we motivate the effective channel based model, introduced in Section 3.4.2, with the following definition and examples.

**Definition 4.3. (Effective channel)** Let $Z_k$ signify the $k$th effective channel iid r.v. with pdf $f_{Z}(z)$, and realization $z_k$. Then, with the effective channel threshold $\Theta$, the $k$th decoding failure probability is

$$Q_k(\Theta, f_{Z}(z)) = P\left\{ \sum_{k'=1}^{k} Z_{k'} < \Theta \right\}. \quad (4.7)$$

We illustrate the effective channel formulation for outage probability analysis by two examples, capacity achieving SISO-RR and SISO-IR in block fading bandwidth-limited AWGN.

**Example 4.1.** For SISO-RR with capacity achieving coding in an AWGN channel, it holds that

$$Q_{k,1}(\Theta, f_{Z}(z)) = P\left\{ \ln \left( 1 + \sum_{k'=1}^{k} G_{k'} \right) < R \right\} = P\left\{ \sum_{k'=1}^{k} Z_{k'} < \Theta \right\}, \quad (4.8)$$

where $Z_k = G_k$, $G_k$ is a unit-average iid channel gain r.v., and the decoding threshold is $\Theta = (e^R - 1)/S$.

**Example 4.2.** For capacity achieving coding for SISO-IR in an AWGN channel, we have

$$Q_{k,1}(\Theta, f_{Z}(z, S)) = P\left\{ \sum_{k'=1}^{k} \ln (1 + SG_{k'}) < R \right\} = P\left\{ \sum_{k'=1}^{k} Z_{k'} < \Theta \right\}, \quad (4.9)$$

where $Z_k = \ln(1 + SG_k)$ is an iid r.v. of the mutual information of the channel with pdf $f_{Z}(z, S)$, and $\Theta = R$. In the following we do not explicitly state the dependence on $S$ and write $f_{Z}(z)$ instead of $f_{Z}(z, S)$.

We note that, in addition to SISO-RR and -IR, many other important HARQ schemes involving summing of iid r.v.s can be analyzed within the effective channel framework. We further note that HARQ schemes operating under the influence of interference, can be handled with the effective channel model. We explore this in Chapter 7.5.7. Due to the sum of iid r.v.s on the RHS of (4.7), the $k$th decoding failure probability (4.7) can be expressed through the $k$-fold convolution,

$$Q_k(\Theta, f_{Z}(z)) = \int_{0}^{\Theta} f_{Z}(z)^{\otimes(k)} \, dz. \quad (4.10)$$
In much of the literature, the throughput analysis stops after giving (4.2), or stops after giving (4.2) together with expressions of $Q_k$, $f_Z(z)^{(k)}$, or $f_Z(z)$, of some HARQ system and fading channel model. However, halting the analysis at this stage, i.e. with equation (4.2) (or for persistent-HARQ (4.5)) and (4.10)), poses a number of problems, such as:

- Every $f_Z(z)^{(k)}$, $k \in \mathbb{N}^+$, need to be analytically computed as a $k$-fold convolution from $f_Z(z)$. This may, or may not, be doable in practice.

- Every $Q_k$, $k \in \mathbb{N}^+$, (may) need to be analytically computed as an integral of $f_Z(z)^{(k)}$. This may, or may not, be doable in practice.

- The average number of transmissions, $\tau_K$, needs to be analytically computed as a truncated sum of $Q_k$. This may, or may not, be doable in practice.

Also, for performance optimization, discussed in Chapter 5, it is desirable to give throughput expressions that do not contain any sums having large (or potentially infinite) number of terms. The average number of transmissions may also be explicitly computed as

$$
\tau_K = \sum_{k=0}^{K-1} Q_k = 1 + \sum_{k=1}^{K-1} \int_0^\Theta f_Z(z)^{(k)} \, dz = 1 + \int_0^\Theta \left( \sum_{k=1}^{K-1} f_Z(z)^{(k)} \right) \, dz,
$$

(4.11)

where the order of summation and integration is changed in the last step. We illustrate the use of (4.11) in the following example.

**Example 4.3.** Consider RR in Rayleigh fading (without diversity) characterized by a pdf $f_Z(z) = e^{-z}$ and a decoding threshold $\Theta$. The $k$-fold convolution can be determined to be $f_Z(z)^{(k)} = z^{k-1}e^{-z}/(k-1)!$. This gives the decoding failure probability $Q_{k,1}^{RR} = \gamma_k(\Theta)$, and the average number of transmissions is

$$
\tau_{\infty,1}^{RR} = 1 + \int_0^\Theta \sum_{k=0}^{\infty} \frac{z^k e^{-z}}{k!} \, dz = 1 + \Theta.
$$

(4.12)

While we can handle the case of persistent-RR in Rayleigh fading without diversity, we are not aware of any other case in the literature, for which the average number of transmissions has explicitly been computed based either on (4.2) with (4.10), or on (4.11). In the following, we propose a method that avoids the trouble of explicitly determining expressions for each $Q_k$, then summing all $Q_k$, $k \in \{1 \ldots K\}$, and instead determine the average number of transmissions $\tau$ directly. This expression can in many cases simplify the required analytical calculations. This approach is considered next.
4.2. Performance Analysis

Throughput on Effective Channel LT-form

**Theorem 4.1.** (Throughput of truncated-HARQ) Let $R$ be the rate, $Z$ signify the effective channel r.v. with pdf $f_Z(z)$ and Laplace transform $F(s) = \mathcal{L}_s \{f_Z(z)\}$. Then, the throughput of $K$-slot truncated-HARQ is

$$T_K(R, S, F(s)) = R \frac{\mathcal{L}_s^{-1} \left\{ \frac{1-F(s)^K}{s} \right\}}{\mathcal{L}_s^{-1} \left\{ \frac{1}{2} \frac{1-F(s)^K}{1-F(s)} \right\}},$$

(4.13)

where $\mathcal{L}_s^{-1} \{ \cdot \}$ denotes the inverse Laplace transform.

**Proof.** The average rate is

$$\rho_K(R, S, F(s)) = R(1 - Q_K)$$

$$= R \int_0^\Theta \left( \delta(z) - f_Z(z)^\Theta(K) \right) \, dz$$

$$= R \mathcal{L}_s^{-1} \left\{ \frac{1 - F(s)^K}{s} \right\},$$

(4.14)

where we used the (time-)integration rule of the Laplace transform in the last step. The average number of transmissions expressed in $F(s)$ is

$$\tau_K(R, S, F(s)) = \sum_{k=0}^{K-1} Q_k$$

$$= \sum_{k=0}^{K-1} \int_0^\Theta f_Z(z)^\Theta(k) \, dz$$

$$= \sum_{k=0}^{K-1} \int_0^\Theta \mathcal{L}_s^{-1} \left\{ F(s)^k \right\} \, dz$$

$$= \mathcal{L}_s^{-1} \left\{ \frac{1}{s} \frac{1 - F(s)^K}{1 - F(s)} \right\},$$

(4.15)

where we changed the order of integration and summation, determined the geometric series, and used the Laplace transform integration rule in the last step. With (4.14) and (4.15) in (4.2), we get (4.13).

**Remark 4.2.** In order to simplify the analysis, or facilitate numerical computation, it can be useful to express (4.14) on the form $\rho_K(R, S, F(s)) = R \left( 1 - \mathcal{L}_s^{-1} \left\{ \frac{F(s)^K}{s} \right\} \right)$, and (4.15) on the form $\tau_K(R, S, F(s)) = 1 + \mathcal{L}_s^{-1} \left\{ \frac{F(s)}{s} \frac{1-F(s)^K}{1-F(s)} \right\}$.

The following two corollaries give throughput expressions of persistent-HARQ and ARQ in terms of $F(s)$. 

Corollary 4.1. (Throughput of persistent-HARQ) Let \( R \) be the rate, \( \Theta \) be the decoding threshold, and \( F(s) \) be the Laplace transform of the effective channel pdf. Then, the throughput of persistent-HARQ is

\[
T_{\text{HARQ}}(R, S, F(s)) = \frac{R}{\mathcal{L}^{-1}\left\{ \frac{1}{2} \left[ 1 - F(s) \right] \right\}}.
\]  \tag{4.16}

Proof. Following the method of proof as in Theorem 4.1, and using \( \lim_{K \to \infty} Q_K = 0 \), we get (4.16).

As seen, (4.14) and (4.16) enable analytical performance evaluation, often giving closed-form expressions, even without computing infinite (or large) sums of the integrals of \( k \)-fold convolved pdfs. Determining the inverse Laplace transforms have their own challenges, but we propose a systematic approach in Chapter 7 handling this for a wide range of effective channels based on the ME-distribution. It is, of course, very well known that convolutions in the time domain correspond to multiplication of their respective LTs in the Laplace domain. This has, e.g., been exploited in the area of the renewal theory to derive the so called renewal function \([BF01]\) on a Laplace transform form, which with our notation corresponds to \( F(s)/(1 - F(s)) \). We note that for the particular case of persistent-HARQ, but importantly not for truncated-HARQ, the denominator of the throughput expression, (4.16), is related to the Laplace transformed renewal function. A difference is that (4.16) also takes one integration, and \( Q_0 = 1 \), into account. From a more practical point of view, we consider numerical evaluation of (4.13), which can approximate (4.16) for a large \( K \), in Appendix 4.L. Two examples of LTs for effective channel pdfs are now given below.

Example 4.4. Selection diversity of order 2, operating in a unit variance Rayleigh fading channel, has the pdf \( f_Z(z) = 2e^{-z}(1 - e^{-z}) \) with corresponding LT \( F(s) = \frac{2}{1+s} - \frac{2}{2+s} \). On one hand, inserting \( F(s) \) in (4.15) for RR, gives \( T_{\text{SD-RR}} = \frac{9R}{1+6e^{-\Theta}+2e^{-2\Theta}} \). On the other hand, inserting \( F(s) = \frac{2}{1+s} - \frac{2}{2+s} \) in (4.7) for ARQ results in \( T_{\text{SD-ARQ}} = R(2 - e^{-\Theta})e^{-\Theta} \).

In the following theorem, we state that the throughput of truncated-HARQ is upper bounded by the throughput of persistent-HARQ\(^6\). The bound applies to all HARQ-schemes where the decoding failure probability can be written on the form \( Q_k = \mathbb{P}\left\{ \sum_{k'=1}^k Z_{k'} < \Theta \right\} \), where \( Z_k \) is an iid r.v. characterizing the per transmission SNR, or MI.

\(^6\)To the best of our knowledge, this bound has not previously been fully proven. A related, but sharper, problem is to show that \( T_{K+1}(R, S, F(s)) \geq T_K(R, S, F(s)), \forall K \). This problem was formulated in [SJS+17], but not completely proven, but a sufficient condition were given. This latter bound was also conjectured in [CT01], but not proved
4.2. Performance Analysis

**Theorem 4.2.** The throughput of truncated-HARQ is upper bounded by the throughput of persistent-HARQ, i.e.

$$T_\infty(R, S, F(s)) \geq T_K(R, S, F(s)).$$

(4.17)

**Proof.** Let \( f(x) \geq 0, x \geq 0 \) (with \( f(x) = 0, x < 0 \)), and \( g(x) \geq 0, x \geq 0 \) (with \( g(x) = 0, x < 0 \)) be two positive functions on the positive real axis. Then, as shown in Appendix 4.A, the following inequality holds

$$\int_0^a (f * g)(t) \, dt \leq \int_0^a f(x) \, dx \int_0^a g(t) \, dt. \quad (4.18)$$

Subsequently, let \( f(x) = \mathcal{L}_x^{-1}\{F(s)K\}, g(x) = \mathcal{L}_x^{-1}\{\frac{1}{1-F(s)}\}, \) and \( a = \Theta \). With \( F(s) \) being the LT of pdf for per transmission SNR, or ML. Then, using (4.18) gives

$$\int_0^\Theta \mathcal{L}_x^{-1}\{F(s)K\} \ast \mathcal{L}_x^{-1}\{\frac{1}{1-F(s)}\}(t) \, dt$$

$$\leq \int_0^\Theta \mathcal{L}_x^{-1}\{F(s)K\} \, dx \int_0^\Theta \mathcal{L}_x^{-1}\{\frac{1}{1-F(s)}\} \, dt,$$

$$\Rightarrow - \int_0^\Theta \mathcal{L}_x^{-1}\{F(s)K\} \, dx \geq - \int_0^\Theta \mathcal{L}_x^{-1}\{F(s)K\} \, dx \int_0^\Theta \mathcal{L}_x^{-1}\{\frac{1}{1-F(s)}\} \, dt,$$

$$\Rightarrow \int_0^\Theta \mathcal{L}_x^{-1}\{\frac{1}{1-F(s)}\} \, dx - \int_0^\Theta \mathcal{L}_x^{-1}\{\frac{F(s)K}{1-F(s)}\} \, dx$$

$$\geq \int_0^\Theta \mathcal{L}_x^{-1}\{\frac{1-F(s)K}{1-F(s)}\} \, dx - \int_0^\Theta \mathcal{L}_x^{-1}\{F(s)K\} \, dx \int_0^\Theta \mathcal{L}_x^{-1}\{\frac{1}{1-F(s)}\} \, dt,$$

$$\Rightarrow \int_0^\Theta \mathcal{L}_x^{-1}\{\frac{1-F(s)K}{1-F(s)}\} \, dx \geq \int_0^\Theta \mathcal{L}_x^{-1}\{1-F(s)K\} \, dx \int_0^\Theta \mathcal{L}_x^{-1}\{\frac{1}{1-F(s)}\} \, dt,$$

$$\Rightarrow R \int_0^\Theta \mathcal{L}_x^{-1}\{\frac{1}{1-F(s)}\} \, dx \geq R \int_0^\Theta \mathcal{L}_x^{-1}\{1-F(s)K\} \, dx \int_0^\Theta \mathcal{L}_x^{-1}\{\frac{1}{1-F(s)}\} \, dt,$$

$$\Rightarrow T_\infty(R, S, F(s)) \geq T_K(R, S, F(s)), \quad (4.19)$$

where we negate each side and reverse the inequality, as well as merge the convolution integrand on the LHS, in step (a), add identical terms on each side in step (b), incorporate the added terms under the inverse Laplace operator in step (c), and rearrange the terms, as well as multiply each side with rate \( R \), in step (d). \( \square \)

Next, the throughput (on the general LT-form \( F(s) \)) for systems with MRC-like diversity of persistent-HARQ is considered.
Proposition 4.3. Let $F(s) = G(s)^N$, where $N$ is the diversity order, and $G(s)$ is the Laplace transform of a diversity-free (effective) channel pdf. Then, the throughput of persistent-HARQ is

$$T_{\infty,N}^{\text{HARQ}}(R, S, G(s)) = R \mathcal{L}^{-1} \left\{ \frac{1}{N} \sum_{n=1}^{N} \frac{1}{1 - \alpha_n G(s)} \right\},$$

(4.20)

where $\alpha_n \triangleq e^{i2\pi n/N}$.

Proof. We use Corollary 4.1 together with partial fraction expansion of the form $P(x)/Q(x) = \sum_{n=1}^{N} P(\alpha_n)/(Q'(\alpha_n)(x - \alpha_n))$, where we let $P(x) = 1$, $Q(x) = 1 - x^N$ and $x = G(s)$.

Corollary 4.2. (Throughput of ARQ) Let $R$ be the rate, $\Theta$ be the decoding threshold, and $F(s)$ be the Laplace transform of the effective channel pdf. Then, the throughput of ARQ is

$$T^{\text{ARQ}}(R, S, F(s)) = R \mathcal{L}^{-1} \left\{ \frac{1 - F(s)}{s} \right\}.$$  

(4.21)

Proof. Use Theorem 4.1 with $K = 1$.

In the remaining sections of this chapter, we study the throughput performance of several important wireless HARQ communication cases, involving diversity, Rayleigh/Nakagami-$m$-fading, and truncated/persistent-HARQ, by exploring the use of (4.13) and (4.16).

4.2.2 ARQ and Persistent/Truncated-RR with GD-channel

In Sections 3.4.3 and 4.1, we motivated the consideration of the general diversity channel reflecting OSTBC, MRC, and (integer parameter restricted) Nakagami-$m$ fading. In the following, we consider ARQ, as well as persistent- and truncated-RR operating in the GD-channel. We start with a motivating example for RR.

Example 4.5. (General diversity channel) Let the OSTBC code\(^7\) be characterized by the rate $r_{\text{stc}} \leq 1$, the Nakagami-$m$ channel have an integer Nakagami-parameter, $m^N \in \mathbb{N}^+$, and the number of transmit and receive antennas are $N_{\text{tx}}$ and $N_{\text{rx}}$, respectively. Then, for a wireless channel with OSTBC, Nakagami-$m$ fading, and MRC, the decoding failure probability is

$$Q^{\text{RR}}_{k, N_{\text{tx}}}(\Theta_{\text{gd}}) = P \left\{ \sum_{k'=1}^{k} Z_{k'} \leq \Theta_{\text{gd}} \right\} = \gamma_{k}(kN_{\text{tx}}, \Theta_{\text{gd}}).$$

(4.22)

\(^7\)Practically, the OSTBC rate, $r_{\text{stc}}$, need to adhere to real world codes. The maximum rate is $r_{\text{stc}} = 1$ for Alamouti’s code with $N_{\text{tx}} = 2$. It is shown in [WX03] that $r_{\text{stc}} \leq 3/4$ for $N_{\text{tx}} \geq 3$. 
where $Z_k'$ is a chi-square distributed r.v. with unit variance and $2N_{gd}$ degrees of freedom, $N_{gd} \triangleq N_{tx}N_{rx}m^N$ is the GD-channel diversity order, $\Theta_{gd} \triangleq (e^{R_{gd}} - 1)/S_{gd}$ is the GD-channel decoding threshold, $R_{gd} \triangleq R/r_{stc}$ is the GD-channel rate, and $S_{gd} \triangleq S/m^N r_{stc} N_{tx}$ is the GD-channel SNR. The corresponding Laplace transform of the GD-channel pdf is

$$F(s) = \left(\frac{1}{1 + s}\right)^{N_{gd}}.$$  \hspace{1cm} (4.23)

**Proof.** The proof is given in Appendix 4.B. \hfill \Box

Since ARQ correspond to RR with transmit limit $K = 1$, the GD-channel is also applicable to ARQ.

**ARQ**

We first give the throughput expression for ARQ operating in the GD-channel.

**Proposition 4.4.** Let $F(s) = 1/(1 + s)^{N_{stc}}$ be the Laplace transform of the effective channel pdf, $N_{gd}$ denotes the GD-channel diversity order, and $\Theta_{gd}$ denotes the GD-channel decoding threshold. Then, the throughput of ARQ is

$$T_{1,N_{gd}}^{ARQ}(R, \Theta_{gd}) = R(1 - Q_1) = RT_r(N_{gd}, \Theta_{gd}).$$ \hspace{1cm} (4.24)

**Proof.** Using Prop. 4.2 with Ex. 4.5 for $k = 1$, proves Prop. 4.4. \hfill \Box

When $N_{gd} = 1$, (4.27) degenerates to

$$T_{1,1}^{ARQ}(R, \Theta_{gd}) = Re^{-\Theta_{gd}}.$$ \hspace{1cm} (4.25)

**Persistent-RR**

In this section, we consider persistent RR-HARQ. The throughput for an arbitrary diversity order $N_{gd} \in \mathbb{N}^+$, arbitrary decoding threshold $\Theta_{gd}$, when $F(s) = 1/(1 + s)^{N_{stc}}$, is given in the proposition below.

**Proposition 4.5.** Let $F(s) = 1/(1 + s)^{N_{stc}}$ be the GD-channel, $N_{gd} \in \mathbb{N}^+$ denotes the GD-channel diversity order, and $\Theta_{gd}$ denotes the GD-channel decoding threshold. Then, the throughput of persistent-RR is

$$T_{\infty,n_{gd}}^{RR}(R, \Theta_{gd}) = \frac{2N_{gd}R}{1 + N_{gd} + 2\Theta_{gd} + 2 \sum_{n=1}^{N_{gd}-1} e^{-\Theta_{gd}(1-\alpha_n)/(1-\alpha_n^H)}}.$$ \hspace{1cm} (4.26)

where $\alpha_n = e^{2\pi n/N_{stc}}$ is the nth root of unity, and $(\cdot)^H$ denotes the conjugate transpose (complex conjugate for a scalar).

**Proof.** The proof is given in Appendix 4.C. \hfill \Box
When the diversity order $N_{gd} \in \mathbb{N}^+$ is a small integer, we can give (relatively) simple closed-form expressions for the throughput as illustrated in the proposition below. For example, (4.28), with $\Theta_{gd} = (e^R - 1)/S_{gd}$, $S_{gd} = S/2$, corresponds to the throughput of RR with Alamouti-diversity [Ala98] for iid exponentially distributed fading SNR.

**Proposition 4.6.** Let $F(s) = 1/(1+s)^{N_{gd}}$ be the effective channel, $N_{gd} = \{1, 2, 3, 4\}$ are the GD-channel diversity orders, and $\Theta_{gd}$ denotes the GD-channel decoding threshold. Then, the throughput efficiencies of persistent-RR with $N_{gd} = \{1, 2, 3, 4\}$-fold diversity degrees are

\[
T_{RR,1}^\infty(R, \Theta_{gd}) = \frac{R}{1 + \Theta_{gd}}, \quad (4.27)
\]

\[
T_{RR,2}^\infty(R, \Theta_{gd}) = \frac{4R}{3 + 2\Theta_{gd} + e^{-2\Theta_{gd}}}, \quad (4.28)
\]

\[
T_{RR,3}^\infty(R, \Theta_{gd}) = \frac{3R}{2 + \Theta_{gd} + e^{-2\Theta_{gd}} \left( \cos \left( \frac{\sqrt{3}\Theta_{gd}}{2} \right) + \frac{1}{\sqrt{3}} \sin \left( \frac{\sqrt{3}\Theta_{gd}}{2} \right) \right)}, \quad (4.29)
\]

\[
T_{RR,4}^\infty(R, \Theta_{gd}) = \frac{8R}{5 + 2\Theta_{gd} + e^{-2\Theta_{gd}} + 2e^{-\Theta_{gd}} \left( \cos (\Theta_{gd}) + \sin (\Theta_{gd}) \right)} \quad (4.30)
\]

*Proof.* The proofs are given in Appendix 4.D. \qed

**Truncated-RR**

For $K$-truncated-RR, with the GD-channel (OSTBC, MRC, and (integer-restricted) Nakagami-$m$ fading channel), it is not straightforward to express (a simple) closed-form throughput expression. Nevertheless, allowing for a sum of incomplete gamma functions in the denominator, the following proposition gives a throughput expression for the GD-channel.

**Proposition 4.7.** Let $K$ be the maximum allowed number of transmissions, $F(s) = 1/(1+s)^{N_{gd}}$ be the GD-channel, $N_{gd} \in \mathbb{N}^+$ be the GD-channel diversity order, and $\Theta_{gd}$ be the GD-channel decoding threshold. Then, the throughput of $K$-truncated-RR is

\[
T_{RR,K,N_{gd}}^\infty(R, \Theta_{gd}) = \frac{R \Gamma_r(K N_{gd}, \Theta_{gd})}{1 + \sum_{k=1}^{K-1} \Gamma_r(k N_{gd}, \Theta_{gd})}. \quad (4.31)
\]

*Proof.* The proof is given in Appendix 4.E. \qed

---

8The throughput expression (4.27) is known in the literature.

9In Chapter 7, we develop a framework that gives a simple and compact closed-form throughput expression for this case.
As noted, the denominator of (4.31) contains a sum of incomplete gamma functions (with $K - 1$ terms) which we are unable to simplify for GD-channel diversity order $N_{gd} \in \{2,3,\ldots\}$. However, a simplified form of the throughput for transmit limit $K$, and the diversity-free case, $N_{gd} = 1$, is given in the following proposition.

**Proposition 4.8.** Let $K$ be the maximum allowed number of transmissions, $F(s) = 1/(1 + s)$ be the Laplace transform of the effective channel pdf, $N_{gd} = 1$ be the GD-channel diversity order, and $\Theta_{gd}$ be the GD-channel decoding threshold. Then, the throughput of $K$-truncated-RR is

$$T_{RR}^{K,1}(R, \Theta_{gd}) = \frac{R\gamma_r(K, \Theta_{gd})}{1 + \Theta_{gd}\gamma_r(K - 1, \Theta_{gd}) + (K - 1)\gamma_r(K, \Theta_{gd})}. \quad (4.32)$$

**Proof.** The proof is given in Appendix 4.F. 

For $K = 1$, (4.32) degenerates, as expected, to $T_{RR}^{1,1}(R, \Theta_{gd}) = R e^{-\Theta_{gd}}$, i.e. the throughput expression of ARQ.

### 4.2.3 Persistent/Truncated-IR with Rayleigh Fading Channel

So far, we have given general throughput expressions for ARQ and HARQ expressed in the LT of the effective channel pdf, $F(s)$, and specific throughput expressions for ARQ and RR wrt the GD-channel. It is however well-known that IR offers higher throughput than RR and ARQ [CT01, (21)]. This in turn motivates a corresponding throughput analysis of IR, which is therefore the subject of this section. Many works, e.g. [CT01], have attempted to analyze the throughput of IR in wireless fading channels. However, no closed-form throughput expression is known, not even for a simple case comprised of persistent-IR, AWGN channel capacity achieving coding, block Rayleigh fading, and without diversity. Despite the apparently simple communication scenario, it is not an easy problem to solve, and it is not known whether a closed-form throughput expression exists or not. Nevertheless, we hope to make some progress in analyzing this case. With this goal in mind, we adopt the Laplace transform approach developed earlier and give some expressions in inverse Laplace transform forms. Those analytical expressions can either be numerically computed, see e.g. Appendix 4.L, or may be useful in finding closed-form expressions in the future. As an alternative, based on a recent expression of $Q_k$ for IR [CA13, YA09], in terms of the generalized upper incomplete Fox’s H function, we also give a throughput expression expressed in Fox’s H function outgoing from (4.14). When exact analysis does not work, or the complicated functions (such as the generalized upper incomplete Fox’s H function) bring only little insight (or are unavailable in standard math software), bounding the throughput (or $Q_k$) is an alternative. In [CT01, (21)], it was shown that the throughput of IR in iid block Rayleigh fading (without diversity), $T_{IR}^{1,1}$, can be bounded as $C_{erg} \geq T_{IR}^{1,1} \geq T_{RR}^{1,1} \geq T_{ARQ}^{1,1}$, where $C_{erg}$ denotes the ergodic capacity for the studied communication case. Those bounds are in general lose,
especially when $R$ is large. Therefore, we introduce new upper- and lower-bounds for $Q_k$ and the throughput of IR. In addition, using Jensen’s inequality, another (yet known) throughput bound is given. We briefly note that several works, e.g. [Lan03] and [WJ10], have considered approximations (i.e. not bounds) for the outage probabilities of IR. We initiate the study by first analyzing IR with the Laplace transform-based approach.

**Persistent-IR: Laplace Transform-based Method**

The following lemma considers the effective channel mutual information for the SISO-IR case with Rayleigh fading.

**Lemma 4.2.** Let the instantaneous channel gain $g$ be iid exponentially distributed, the effective channel MI be $z = \ln(1 + Sg)$, and $R$ be the effective channel decoding threshold. Then, the effective channel pdf, and its Laplace transform, are

$$f_z(z, S) = \begin{cases} S^{-1} e^{1/S} + z - e^{1/S} & z \geq 0 \\ 0 & z < 0 \end{cases}, \quad (4.33)$$

$$F(s, S) = e^{1/S} S^{-s} \Gamma(-s + 1, 1/S). \quad (4.34)$$

**Proof.** The proof is given in Appendix 4.G.

Next, three variants are given for the average number of transmissions.

**Proposition 4.9.** Let the instantaneous channel gain $g$ be iid exponentially distributed, the effective channel MI be $z = \ln(1 + Sg)$, and $R$ the effective channel decoding threshold. Then, the average number of transmissions per message can be expressed as

$$r_{\infty, IR}^R(S, R) = \int_0^R \frac{1}{1 - e^{1/S} S^{-s} \Gamma(-s + 1, 1/S)} dz, \quad (4.35)$$

$$= \int_0^R \left\{ \frac{1}{s^2 e^{1/S} S^{-s} \Gamma(-s + 1, 1/S)} \right\} dz, \quad (4.36)$$

$$= \frac{1}{S} \frac{d}{dS} \frac{s^2}{1} \frac{1}{s^2} \ln \left( \Gamma\left(-s, 1/S\right) \right), \quad (4.37)$$

**Proof.** The proof is given in Appendix 4.I.

Unfortunately, we can not find a closed-form throughput expression for the IR-case studied. On the other hand, we are not aware of any work that offers a closed-form throughput expression for IR. On the contrary, expressions found in other works are, to the best of our knowledge, limited to bounds and approximations of either information outage or, more seldom, throughput. However, we will see in Chapter 7 that the argument in (4.35) (or (4.36)) can be approximated as a rational polynomial expressions, which has a closed-form expression. Alternatively, we can
numerically compute the average number of transmissions, e.g. with the numerical evaluation method in Appendix 4.L.

Another approach, compared to the direct Laplace transform method (4.14) used here, is to express the average number of transmissions (and hence the throughput) in terms of the generalized upper incomplete Fox’s H function. This, and the shortcomings of this approach, is discussed in Appendix 4.M.

**Laplace Transform for the \( N \times N \)-MIMO-case**

As will be shown here, the LT of the \( N \times N \)-MIMO channel capacity pdf shares similarity in form to the LT of the IR (SISO) channel capacity pdf (4.34). For this reason, we briefly consider the \( N \times N \)-MIMO-case here too.

**Lemma 4.3.** Let the unordered eigenvalues of the complex central Wishart matrix, for a \( N \times N \) MIMO system, be \( f(\lambda_1, \ldots, \lambda_N) = \exp\left(-\sum_{n=1}^{N} \lambda_n\right) \prod_{n'<n} (\lambda_{n'} - \lambda_n)^2\) [Tel99]. Then, the LT of the MIMO channel capacity is

\[
F(s) = \frac{e^{N/\tilde{S}} \tilde{S}^{-N} \det(G(s))}{\prod_{n=1}^{N} (n!)^2},
\]

where \( \tilde{S} \equiv S/N_{tx} \), \( G(s) \) is an \( N \times N \) Hankel matrix with entries \( g_{ij}(s) = \Gamma(-s + i + j - 1, 1/\tilde{S}) \), \( i = \{1, \ldots, N\} \), \( j = \{1, \ldots, N\} \).

**Proof.** The proof is given in Appendix 4.H. \( \square \)

Note that with the upper incomplete Gamma functions, (4.38) differs non-trivially from the form in [WG04, (10)]. The LT for \( 2 \times 2 \) MIMO is simply

\[
F(s) = e^{4/S} (S/2)^{-2s} \left( \Gamma(-s + 3, 2/S) \Gamma(-s + 1, 2/S) - \Gamma(-s + 2, 2/S) \right),
\]

which, with the upper incomplete gamma functions and the \( e^{4/S} (S/2)^{-2s} \) pre-factor, has a clear resemble with the LT for the SISO-IR case in Lemma 4.2. The LT of the MIMO channel capacity, \( F(s) \) in (4.38), can be used for determining the outage probability, \( Q_{MIMO} = \mathcal{L}_R^{-1}(F(s)/s) \), and hence the throughput of MIMO-ARQ, \( T_{MIMO-ARQ} = R(1 - Q_{MIMO}) \), or the throughput of MIMO-IR, \( T_{MIMO-IR} = R/\mathcal{L}_R^{-1}(1/s(1 - F(s))) \). Unfortunately, much like for the SISO-IR-case, we can not determine the exact throughput expression for MIMO-ARQ, nor for MIMO-IR. Nevertheless, the numerical approach, Algorithm 1 in Appendix 4.L, may be used. In Section 7.6, we develop a continued-fraction-based rational approximation to the upper incomplete gamma function that can be used to write \( \det(G(s)) \) in (4.38) on a ME-distribution-based form as developed in Chapter 7. From (4.39), using the inverse LT, we also note that the corresponding pdf of the channel capacity can be interpreted on a form involving convolutions of more basic pdfs. More precisely, this pdf can be expressed as

\[
f_X(x) = f_X^{(2)}(x) * f_X^{(0)}(x) - f_X^{(1)}(x) * f_X^{(1)}(x),
\]

where \( f_X^{(k)}(x) = \)
\[ \tilde{S}^{-(k+1)}e^{1/\tilde{S}}e^{-x/\tilde{S}}x^{(k+1)}, \quad x \geq 0, \text{and the LT of } f^{(k)}_x(x) \text{ is } e^{1/\tilde{S}}\tilde{S}^{-x} \Gamma(-s+1+k, 1/\tilde{S}). \]

The generalization to \( N \times N \)-MIMO, with \( N > 2 \), is straightforward. Building on results from random matrix theory (RMT), we have also studied the throughput performance (and its optimization) of MIMO-ARQ/RR/IR-systems in [LRS14b].

**Truncated-IR: Convolution and Bounding-based Methods**

As noted above, the Laplace transform-based method does not easily offer a closed-form throughput expression for IR with Rayleigh fading. Instead of an exact result, we now look for (closed-form) bound expressions. It would be valuable to offer improved upper and lower bounds for the region of operation when the throughput is close to the rate. This is so, since a system operates in this region if the throughput is optimized wrt the rate. It would also be of interest to provide throughput bounds for truncated-IR with transmit limit \( K = 2 \), to reveal some fundamental properties of truncated-IR. With those objectives in mind, we first consider bounds for the decoding failure probability in the following proposition.

**Proposition 4.10.** Let the instantaneous channel gain, \( g \), be iid exponentially distributed, the SNR be \( S \), the effective channel be \( z = \ln(1+SQ) \), and \( R \) the effective channel decoding threshold. Then, the decoding failure probability, \( Q_{k,1,1} \), can be lower bounded, \( \hat{Q}_{k,1,1} \), and upper bounded, \( \tilde{Q}_{k,1,1} \), as

\[
0 \leq (\hat{Q}_{k,1,1})_+ \leq Q_{k,1,1} \leq (\tilde{Q}_{k,1,1})_+ \leq 1,
\]

where

\[
\hat{Q}_{k,1} = 1 - e^{k/S} I_{k,1},
\]

\[
\tilde{Q}_{k,1} = 1 - I_{k,1},
\]

and

\[
I_{k,1}(R,S) \triangleq \int_0^\infty \cdots \int_0^\infty \prod_{q=1}^{k-1} e^{-x_q - e^{k/S}/x_q} \prod_{q=1}^{k-1} x_q dx_q,
\]

where \( c_k \triangleq e^{R/S}k \).

**Proof.** The proof is given in Appendix 4.J. \( \square \)

However, the integral \( I_{k,1} \) is in general hard to determine exactly. Nevertheless, the integral \( I_{k,1} \) can be given for \( k = \{2, 3\} \). In Appendix 4.K, we show that

\[
I_{2,1}(R,S) = 2\sqrt{c_3}K_1(2\sqrt{c_3}),
\]

and,

\[
I_{3,1}(R,S) = 8c_3 \int_0^\infty e^{-4c_3t^2} K_1(1/t) dt,
\]
4.2. Performance Analysis

where $K_{
u}(\cdot)$ is the modified Bessel function of the second kind and order $\nu$. Now, to bound the throughput of IR, we use the bounds of decoding failure probabilities, (4.41), (4.42), together with (4.2). However, we use the exact values for $Q_{k,1}^{\text{IR}} = 1$, and $Q_{k,1}^{\text{IR}} = 1 - e^{-\Theta}$, with $\Theta = (e^R - 1)/S$, respectively. This gives the throughput bounds

$$\hat{T}_{K,1}^{\text{IR}} \leq T_{K,1}^{\text{IR}} \leq \hat{T}_{K,1}^{\text{IR}},$$

(4.46)

where

$$\hat{T}_{K,1}^{\text{IR}}(R, \hat{Q}_k) \doteq \frac{R(1 - \hat{Q}_k^{\text{IR}})}{\sum_{k=0}^{K-1} \hat{Q}_k^{\text{IR}}},$$

(4.47)

and

$$\hat{T}_{K,1}^{\text{IR}}(R, \hat{Q}_k, 1) \doteq \frac{R(1 - \hat{Q}_k^{\text{IR}})}{\sum_{k=0}^{K-1} \hat{Q}_k^{\text{IR}}}.$$  

(4.48)

A particularly interesting case, revealing some fundamental properties of IR (such as the emergence of multiple optimal rate points as the SNR increases), is for truncated-IR with transmit limit $K = 2$. The throughput has the bounds

$$\hat{T}_{2,1}^{\text{IR}}(R, S) \doteq \frac{R(1 - \hat{Q}_{2,1}^{\text{IR}})}{1 + \hat{Q}_{1,1}^{\text{IR}}} = \frac{R2\sqrt{c_2}K_1(2\sqrt{c_2})}{2 - e^{-\Theta}},$$

(4.49)

and

$$\hat{T}_{2,1}^{\text{IR}}(R, S) \doteq \frac{R(1 - \hat{Q}_{2,1}^{\text{IR}})}{1 + \hat{Q}_{1,1}^{\text{IR}}} = \frac{Rc_2^{1/2}\sqrt{c_2}K_1(2\sqrt{c_2})}{2 - e^{-\Theta}}.$$  

(4.50)

Throughput optimization of (4.49) (and indirectly also (4.50)) is considered in Chapter 5.

**Truncated-IR with GD-channel: Jensen’s Inequality Bound**

It is instructive to compare the bounds above with a throughput bound using a well-known technique involving Jensen’s inequality. For this particular case, there is no additional complexity in considering the GD-channel. We first lower bound the outage probabilities with Jensen’s inequality, i.e. $\varphi(\sum x_i/n) \geq \varphi(\sum x_i)/n$ for a concave function $\varphi(\cdot)$, by

$$Q_{k,N_{gd}}^{\text{IR}}(R, S) = \mathbb{P}\left\{ \sum_{k'=1}^{k} \ln \left( 1 + S_{gd}G_{k'} \right) < R_{gd} \right\} \geq \mathbb{P}\left\{ k \ln \left( 1 + \frac{S_{gd}}{k} \sum_{k'=1}^{k} G_{k'} \right) < R_{gd} \right\}$$
where \( \Theta_{k,N_{gd}}^{Jensen} \equiv \frac{1}{N_{gd}} (e^{R_{gd}/k} - 1) \). The throughput bound is then given by

\[
T_{IR}^{K,N_{gd}}(R,S) \leq \frac{R\Gamma_{r}(K,N_{gd},\Theta_{K,N_{gd}}^{Jensen})}{1 + \sum_{k=1}^{K-1} \gamma_{r}(kN_{gd},\Theta_{k,N_{gd}}^{Jensen})}.
\]  

(4.52)

In Section 4.3, we illustrate that the judiciously chosen throughput bounds (4.49) and (4.50) are significantly tighter than the common Jensen’s inequality based throughput bound for \( K = 2 \).

4.3 Numerical Results and Discussions

In this section, we illustrate the throughput performance of (H)ARQ wrt different parameters. The purpose of the plots are to collect, and visualize, the performance results of the basic, yet important, (H)ARQ-cases, at one common place. This illustrates the overall behaviour wrt parameter choices, but also allows for comparison between the different (H)ARQ-cases. We characterize, in order, the throughput dependency vs. SNR and rate, vs. SNR and wrt diversity order, as well as vs. SNR and wrt different transmit limits. Then, a plot for throughput bounds vs. SNR of truncated-IR is shown. In the final plot, the performance of all three schemes, ARQ, RR and IR, are shown together, enabling ease of comparison.

To start with, in Fig. 4.1 (Fig. 4.2), a contour plot of the specialized throughput expression (4.25) ((4.27)), of ARQ (RR) for the GD-channel, where \( N_{gd} = 1 \), is shown. Moreover, in Fig. 4.3, a contour plot of the throughput of (the approximated) persistent-IR vs. SNR and rate for Rayleigh fading channel is shown. The (I)FFT-based numerical performance evaluation, Algorithm 1 in Appendix 4.L, is used together with (4.33). To approximate the throughput performance of persistent-IR, truncated-IR with a large transmit limit (in this case \( K = 64 \)) is used. It is noted that, for those cases, the throughput increases with the SNR for a given rate. It is observed that for ARQ and RR, the throughput has a maxima for some optimal rate point, \( R_{gd} \), for a given SNR, \( S_{gd} \). IR may also exhibit maxima (i.e. for truncated-IR), but the behaviour for an optima is more complicated. Throughput maximization wrt rate is explored in Chapter 5.

In Fig. 4.4 and Fig. 4.5, the general throughput expressions (4.24) and (4.26) of ARQ and persistent-RR, respectively, are plotted for the GD-channel with parameters \( N_{gd} = \{1,2,4,8\} \), and \( R_{gd} = 4 \) [b/Hz/s]. Moreover, in Fig. 4.6, the throughput (numerically computed with Algorithm 1) of truncated-IR (an accurate approximation for persistent-IR), for diversity order \( N = \{1,2,4,8\} \), \( K = 64 \), and \( R = 4 \) [b/Hz/s] is plotted. For this case, we exploit that the SNR for MRC is Gamma-distributed with pdf \( f_{\zeta}(\zeta) = \frac{\zeta^{N-1}e^{-\zeta/\zeta}}{\Gamma(N)} \). With the effective channel MI mapping \( z = \ln(1 + Sg) \), through variable substitution, we get the MI-pdf \( f_{\zeta}(z) = \frac{(e^{z-1})^{N-1}e^{z-1+S-e^{z}/\zeta}}{S^{N}(N-1)!} \). It is observed, as expected, that increased diversity
4.3. Numerical Results and Discussions

Figure 4.1: Throughput, $T$, vs. GD-channel SNR, $S_{gd}$, and GD-channel rate, $R_{gd}$, of ARQ for GD-channel diversity order $N_{gd} = 1$.

Figure 4.2: Throughput, $T$, vs. GD-channel SNR, $S_{gd}$, and GD-channel rate, $R_{gd}$, of persistent-RR for GD-channel diversity order $N_{gd} = 1$.

Figure 4.3: Throughput, $T$, vs. SNR, $S$, and rate, $R$, of truncated-IR with $K = 64$ (approx. persistent-IR) for Rayleigh fading channel.
order leads to increased throughput and is particularly useful in the low SNR region. Observe that since the GD-channel SNR, $S_{gd}$, is used for the X-axis in Fig. 4.4 and in Fig. 4.5, the plots directly illustrates the performance for $1 \times N_{rx}$ MRC, but not OSTBC. To show the performance for OSTBC, $S_{gd}$ must take $N_{tx}$ into account.

The throughput expression (4.31) (or alternatively (4.32)) of truncated-RR for the GD-channel, with $K_{gd} = \{1, 2, \ldots, 64\}$, $N_{gd} = 1$, and $R = 4 \text{ [b/Hz/s]}$, is plotted in Fig. 4.7. Correspondingly, Fig. 4.8 shows the throughput of truncated-IR with Rayleigh fading channel. The curve is numerically computed with the (I)FFT-based numerical performance evaluation, Algorithm 1, for transmit limit $K = \{1, 2, 4, 8, 16, 32, 64\}$. It is found, as expected, for both cases that increased transmit limit leads to increased throughput. It is further noted that a large transmit limit value, say $K \approx 64$, is required to approach the performance of persistent-HARQ operation in the low SNR region. On the other hand, a relatively small transmit limit value, say $K = \{2, 3, 4\}$, is needed for operation in the high throughput range in order to approximate the performance of persistent-RR and -IR.

Using Algorithm 1, Fig. 4.9 shows the throughput of truncated-IR, with transmit limit $K = 2$, the lower bound (4.49), the upper bound (4.50), and the Jensen’s inequality based bound (4.52). We consider a comparatively high rate, $R = 8 \text{ [b/Hz/s]}$, to reveal the effect on the throughput for the second transmission. This effect is visible as a bump on the throughput curve just below $T = 4 \text{ [b/Hz/s]}$. It is evident that, for this case, the lower and upper bounds, (4.49) and (4.50) respectively, perform better than the Jensen’s inequality bound. Moreover, the upper bound (4.50) is indistinguishable from the exact (albeit numerically computed) throughput curve.

In our final figure, Fig. 4.10, we illustrate the throughput vs. SNR for persistent-IR, truncated-IR with $K = 2$, persistent-RR, truncated-RR with $K = 2$, and ARQ, for rate $R = 4 \text{ [b/Hz/s]}$ and a Rayleigh fading channel. Also, the ergodic capacity vs. SNR, $C_{IR}^{ER} = \int_0^\infty e^{-z} \ln(1 + Sz) \, dz = e^{1/S}E_1(1/S)$, is illustrated for a Rayleigh fading channel. The ergodic capacity is known to be the throughput limit for persistent-IR with an infinite rate, see e.g. [CT01]. An important observation is that IR, as well as the other schemes, exhibits a fairly large SNR-gap to the ergodic capacity in the high throughput region for high rates. For example, in Fig. 4.10, a gap of $\approx 10 \text{ dB}$ is seen for $T = 0.95R$. This is due to that the throughput for IR (but also for RR) is mainly determined by the outage probability for only one transmission in this region, i.e. $T^{IR} \approx R/(1 + Q_1)$. Thus, to achieve high throughput, say $T = 0.95R \text{ [b/Hz/s]}$, with a reasonably small SNR-gap, we would need 2-3 times higher rate. In this case, this implies $R = 8 - 12 \text{ [b/Hz/s]}$, which raises practical complexity concerns. Note that optimization in a more moderate rate range, does (unfortunately) not close the gap. This motivates that more efficient HARQ schemes for this region of operation should be developed by the research community.
4.3. Numerical Results and Discussions

Figure 4.4: Throughput, $T$, vs. GD-channel SNR, $S_{gd}$, of ARQ for GD-channel diversity order $N_{gd} = \{1, 2, 4, 8\}$ and GD-channel rate $R_{gd} = 4$ [b/Hz/s].

Figure 4.5: Throughput, $T$, vs. GD-channel SNR, $S_{gd}$, of persistent-RR for GD-channel diversity order $N_{gd} = \{1, 2, 4, 8\}$ and GD-channel rate $R_{gd} = 4$ [b/Hz/s].

Figure 4.6: Throughput, $T$, vs. SNR, $S$, of truncated-IR with $K = 64$ (approx. persistent-IR) for MRC-diversity order $N = \{1, 2, 4, 8\}$ and rate $R = 4$ [b/Hz/s].
Figure 4.7: Throughput, $T$, vs. GD-channel SNR, $S_{gd}$, of truncated-RR for transmit limit $K = \{1, 2, 4, 8, 16, 32, 64\}$, diversity order $N_{gd} = 1$, and rate $R_{gd} = 4 \,[\text{b/Hz/s}]$.

Figure 4.8: Throughput, $T$, vs. SNR, $S$, of truncated-IR for transmit limit $K = \{1, 2, 4, 8, 16, 32, 64\}$, Rayleigh fading channel, and rate $R = 4 \,[\text{b/Hz/s}]$.

Figure 4.9: Throughput, $T$, (Numerical, (4.49), (4.50), and (4.52)) vs. SNR, $S$, of truncated-IR with $K = 2$ for Rayleigh fading channel, and rate $R = 8 \,[\text{b/Hz/s}]$. 
4.4 Summary and Conclusions

To summarize, we derived a general throughput expression that avoided the infinite sum of outage probabilities in prior work. We then derived closed-form throughput expressions for particular cases of ARQ and truncated/persistent-RR in a GD-channel comprising a combination of OSTBC, MRC, and (Nakagami-parameter integer-restricted) Nakagami-$m$ fading. Throughput bounds for IR in Rayleigh fading were derived, which then resulted in a closed-form throughput bound for truncated-IR with $K = 2$. We presented a numerical method to determine the throughput for an arbitrary effective channel pdf.

To conclude, this chapter indicates that: i) Throughput analysis of HARQ can often (but not always) be conducted directly without computing individual (possibly cumbersome) decoding failure probabilities by using the Laplace transform of the effective channel pdf. ii) Throughput increases for increasing diversity order, $N$, average SNR, $S$, and number of transmissions, $K$. The overall throughput improvement due to exploitation of OSTBC and MRC multiple antenna schemes is moderate. iii) A significant SNR-gap is observed, for all schemes studied, ARQ, RR, and IR, between the ergodic capacity and throughput when operating at a throughput close to the rate $R$. This motivates the development of improved (H)ARQ schemes in this (important) region of operation, or alternatively, if processing retransmission delay constraints and complexity permits, using a rate which is, at least, 2-3 times higher than the ergodic capacity for the average SNR experienced.
4.5 Appendices

4.A Proof of (4.18)

Proof. Expanding the LHS of (4.18), for the case $t > a$, gives

\[
\int_0^a (f \ast g)(t) \, dt = \int_0^a \left( \int_0^t f(x)g(t-x) \, dx \right) \, dt \tag{4.53}
\]

where we have written out the convolution in step (a), rearranged the order of integration in step (b), split the integration interval for the integrand $f(x)$ in step (c), and used the fact that $g(t-x) = 0$, $t < x$, implying that $\int_0^a g(t-x) \, dt \leq \int_0^a g(t) \, dt$. For the case when $t < a$, we get

\[
\int_0^a (f \ast g)(t) \, dt = \int_0^a \left( \int_0^t f(x)g(t-x) \, dx \right) \, dt \\
= \int_0^t f(x) \, dx \int_0^a g(t-x) \, dt \\
\leq \int_0^t f(x) \, dx \int_0^a g(t) \, dt \\
\leq \int_0^a f(x) \, dx \int_0^a g(t) \, dt. \tag{4.54}
\]

The case with $t = a$ is trivial. Thus, $\int_0^a (f \ast g)(t) \, dt \leq \int_0^a f(x) \, dx f_0 g(t) \, dt$. The equality in (4.18) is achieved for $f(x) = \delta(x)$, or $g(x) = \delta(x)$. \qed

4.B Proof of Example 4.5

Proof. Using the mutual information expression for OSTBC in [LS08, (7.4.43)], allows the $k$th decoding failure probability to be expressed as

\[
Q_k = P \left\{ r_{\text{stc}} \ln \left( 1 + \frac{S}{m^S r_{\text{stc}} N_{\text{tx}}} \sum_{k'=1}^k G_{k'} \right) \leq R \right\} \\
= P \left\{ \sum_{k'=1}^k G_{k'} \leq \frac{e^{R/r_{\text{stc}}} - 1}{(S/m^S r_{\text{stc}} N_{\text{tx}})} \right\}. \tag{4.55}
\]
Here, the GD-channel SNR $G_k'$ is iid unit average chi-square distributed r.v. with $2N_{gd} = 2N_{tx}N_{rx}m^N$ degrees of freedom. The LT of the unit-average normalized pdf is $F(s) = 1/(1 + s)^{N_{gd}}$.

4.C Proof of Proposition 4.5

Proof. The average number of transmissions is

$$
\tau^{RR}_{\infty,N_{gd}}(\Theta_{gd}) = \int_{0}^{\Theta_{gd}} \sum_{n=0}^{N_{gd}-1} \left( \frac{1}{1 - F(s)} \right) \frac{N_{gd} - 1}{s^{N_{gd}} - 1} \sum_{n=1}^{N_{gd} - 1} \frac{a_n e^{(a_n - 1)z}}{1 - a_n^2},
$$

where we used

$$
\sum_{n=0}^{N_{gd}-1} \left( \frac{1}{s^{N_{gd}} - 1} \right) = \sum_{n=0}^{N_{gd}-1} \frac{a_n}{(s - a_n)N_{gd}} = \frac{1}{N_{gd}} \sum_{n=0}^{N_{gd}-1} a_n e^{a_n z},
$$

with $a_n = e^{2\pi n/N_{gd}}$.

4.D Proof of Proposition 4.6

Proof. Below, (4.27)-(4.30) are derived.

Derivation of (4.27)

$$
\tau^{RR}_{\infty,1}(\Theta_{gd}) = 1 + \int_{0}^{\Theta_{gd}} e^{-z} \sum_{n=0}^{N_{gd}-1} \left( \frac{1}{s^{N_{gd}} - 1} \right) \frac{N_{gd} - 1}{s^{N_{gd}} - 1} \sum_{n=1}^{N_{gd} - 1} \frac{a_n e^{(a_n - 1)z}}{1 - a_n^2}.
$$

Derivation of (4.28)

$$
\tau^{RR}_{\infty,2}(\Theta_{gd}) = 1 + \int_{0}^{\Theta_{gd}} e^{-z} \sum_{n=0}^{N_{gd}-1} \left( \frac{1}{s^{N_{gd}} - 1} \right) \frac{N_{gd} - 1}{s^{N_{gd}} - 1} \sum_{n=1}^{N_{gd} - 1} \frac{a_n e^{(a_n - 1)z}}{1 - a_n^2}.
$$
Derivation of (4.29)

\[ \tau_{\infty}^{RR}(\Theta_{gd}) = 1 + \int_{0}^{\Theta_{gd}} e^{-z} \left( \frac{1}{s^4 - 1} \right) \, dz = 1 + \int_{0}^{\Theta_{gd}} \frac{e^{-3z/2}}{3} \left( e^{3z/2} - \cos \left( \frac{\sqrt{3}z}{2} \right) - \sqrt{3} \sin \left( \frac{\sqrt{3}z}{2} \right) \right) \, dz \]

\[ = \frac{1}{3} \left( 1 + \Theta_{gd} + e^{-2\Theta_{gd}} \left( \cos \left( \frac{\sqrt{3}\Theta_{gd}}{2} \right) + \frac{1}{\sqrt{3}} \sin \left( \frac{\sqrt{3}\Theta_{gd}}{2} \right) \right) \right). \tag{4.60} \]

Derivation of (4.30)

\[ \tau_{\infty, A}^{RR}(\Theta_{gd}) = 1 + \int_{0}^{\Theta_{gd}} e^{-z} \left( \frac{1}{s^4 - 1} \right) \, dz = 1 + \int_{0}^{\Theta_{gd}} e^{-z} \left( \sinh(z) - \sin(z) \right) / 2 \, dz \]

\[ = \frac{1}{8} \left( 5 + 2\Theta_{gd} + e^{-2\Theta_{gd}} + 2e^{-\Theta_{gd}}(\cos(\Theta_{gd}) + \sin(\Theta_{gd})) \right). \tag{4.61} \]

4.E Proof of Proposition 4.7

Proof. Since \( Q_{k,N_{gd}}^{RR} = \gamma_r(kN_{gd}, \Theta_{gd}) \), the average rate is \( \rho_{k,N_{gd}}^{RR} = Rg_{(K_{gd}, \Theta_{gd})} \), and the average number of transmissions is \( \tau_{k,N_{gd}}^{RR} = \sum_{k=0}^{K} Q_{k} = 1 + \sum_{k=1}^{K} \gamma_r(kN_{gd}, \Theta_{gd}) \), concludes the proof. \( \square \)

4.F Proof of Proposition 4.8

Proof. The average number of transmissions for RR with \( F(s) = 1/(1 + s) \) and \( K \) transmissions at most, is

\[ M_{1,K}^{RR} = \int_{0}^{\Theta} \left( \frac{1 - F(s)^{K}}{1 - F(s)} \right)_z \, dz \]

\[ = \int_{0}^{\Theta} \left( 1 + \frac{1}{s} - \frac{1}{s(1 + s)^{K-1}} \right)_z \, dz \]

\[ = 1 + \Theta - \int_{0}^{\Theta} y^{K-2} e^{-y} \, dy \]

\[ = 1 + \Theta - \int_{0}^{\Theta} \gamma_r(K - 1, z) \, dz \]

4.5. Appendices

\[ = 1 + \Theta \Gamma_r(K - 1, \Theta) + (K - 1) \gamma_r(K, \Theta), \quad (4.62) \]

where we used the identity \( \int_0^x \gamma_r(t) dt = x \gamma_r(s, x) - s \gamma_r(s + 1, x) \), from integration by parts, in the last step.

\[ \square \]

4.G Proof of Lemma 4.2

**Proof.** The pdf is derived by variable transformation.

\[ z = \ln(1 + S g), \text{ with pdf } f_g(g) = e^{-g}, \Rightarrow f_z(z) = S^{-1} e^{S^{-1} z} - S^{-1} e^{z}, \quad z \geq 0. \quad (4.63) \]

The corresponding Laplace transform is given by

\[ F(s) = \int_0^\infty e^{-sz} f_z(z) dz = \int_0^\infty e^{-sz} S^{-1} e^{S^{-1} z} - S^{-1} e^{z} dz, \]

\[ \overset{(a)}{=} e^{S^{-1}} \int_{S^{-1}}^\infty e^{-s \ln(u) S} e^{-u} du = e^{S^{-1}} S^{-1} \Gamma(-s + 1, S^{-1}), \quad (4.64) \]

where the substitution \( u = S^{-1} e^{z} \) is used in step (a).

\[ \square \]

4.H Proof of Lemma 4.3

**Proof.** The LT for \( N \times N \) MIMO channel capacity pdf can be written

\[ F(s) = C_N \int_0^\infty \cdots \int_0^\infty e^{-\sum_{n=1}^N \ln(1+\tilde{S} x_n)} e^{-\sum_{n=1}^N \lambda_n} \prod_{n' > n} (\lambda_{n'} - \lambda_n)^2 d\lambda_1 \cdots d\lambda_N \]

\[ \overset{(a)}{=} C_N e^{N / \tilde{S}} \tilde{S}^{-N} \int_{1 / \tilde{S}}^\infty \cdots \int_{1 / \tilde{S}}^\infty \prod_{n=1}^N x_n^{-1} e^{-\sum_{n=1}^N x_n} \prod_{n' > n} (x_{n'} - x_n)^2 dx_1 \cdots dx_N \]

\[ = C_N e^{N / \tilde{S}} \tilde{S}^{-N} \int_{1 / \tilde{S}}^\infty \cdots \int_{1 / \tilde{S}}^\infty \prod_{n=1}^N x_n^{-1} e^{-\sum_{n=1}^N x_n} \det(V) dx_1 \cdots dx_N \]

\[ = C_N e^{N / \tilde{S}} \tilde{S}^{-N} \det(G(s)), \quad (4.65) \]

where the substitution \( x_n = 1 / \tilde{S} + \lambda_n \) is used in step (a), \( V \) is an \( N \times N \) Vandermonde matrix with entries \( v_{ij} = x_i^{j-1} \), and \( G(s) \) is an \( N \times N \) Hankel matrix with entries

\[ g_{ij}(s) = \int_{1 / \tilde{S}}^\infty x^{-i} e^{-x} x^{i+j-2} dx = \Gamma(-s + i + j - 1, 1 / \tilde{S}). \quad (4.66) \]
To determine the constant $C_N$, we also have

$$F(s = 0) = C_N \int_0^\infty \cdots \int_0^\infty e^{-\sum_{n=1}^{N} \lambda_n} \prod_{n' > n} (\lambda_{n'} - \lambda_n) \, d\lambda_1 \cdots d\lambda_N$$

$$= C_N \int_0^\infty \cdots \int_0^\infty e^{-\sum_{n=1}^{N} \lambda_n} \det(\tilde{V}) \, d\lambda_1 \cdots d\lambda_N$$

$$= C_N \det(U) = C_N \prod_{n=1}^{N-1} (n!)^2,$$

where $\tilde{V}$ is an $N \times N$ Vandermonde matrix with entries $\tilde{v}_{ij} = \lambda_i^{j-1}$, and $U$ is an $N \times N$ Hankel matrix with entries $u_{ij} = (i + j - 2)!$. In [Siv10], an identity for the determinant of a Hankel matrix with factorials numbers is given, which yields $\det(U) = \prod_{n=1}^{N-1} (n!)^2$. With the condition that the density integrate to one, $\lim_{s \to 0} F(s) = 1$, the constant is found as $C_N = 1/\prod_{n=1}^{N-1} (n!)^2$. Hence, in total, the LT is $F(s) = e^{N/S} S^{-Ns} \det(G(s))/\prod_{n=1}^{N-1} (n!)^2$. $\square$

### 4.I Proof of Proposition 4.9

**Proof.** We get (4.35) by inserting (4.33) in (4.14). We get (4.36) by partial integration of the incomplete gamma function in the denominator of (4.35) and using the integration rule of the Laplace transform. The expression (4.37) is found by exploiting partial derivative wrt to $S$ and changing the order of inverse Laplace transform and the derivative. $\square$

### 4.J Proof of Proposition 4.10

The outage probability for $Q_{IR}^{R_k}$ is upper bounded as

$$Q_{IR}^{R_k} = P\left\{ \sum_{q=1}^{k} \ln (1 + SG_q) < R \right\}$$

$$= P\left\{ \sum_{q=1}^{k} X_q < c_k \right\} \cdot \left\{ X_q \triangleq \frac{1}{S} + G_q, c_k' \triangleq \frac{e^{R}}{S^k} \right\},$$

$$= e^{k/S} \int_{x_1 > 1/S} \cdots \int_{x_q > 1/S} e^{-x_q} \, dx_q$$

$$= e^{k/S} \int_{1/S}^{c_1} \int_{1/S}^{c_2} \cdots \int_{1/S}^{c_k} e^{-x_1} \cdots e^{-x_k} \, dx_q.$$
4.5. Appendices

\begin{align*}
&= e^{k/S} \int \int \cdots \int \frac{c_k - 1}{\prod_{q=1}^{k-1} x_{q-1}} \prod_{q=1}^{k-1} e^{-x_q} \left( e^{-1/S} - e^{-c_k/\prod_{q'=1}^{k-1} x_{q'}} \right) dx_q \\
\geq & e^{k/S} \int \int \cdots \int \prod_{q=1}^{k-1} e^{-x_q} \left( e^{-1/S} - e^{-c_k/\prod_{q'=1}^{k-1} x_{q'}} \right) dx_q \\
= & 1 - e^{k/S} \int \int \cdots \int \prod_{q=1}^{k-1} e^{-x_q - c_k/\prod_{q'=1}^{k-1} x_{q'}} dx_q \\
\geq & 1 - e^{k/S} \int \int \cdots \int \prod_{q=1}^{k-1} e^{-x_q - c_k/\prod_{q'=1}^{k-1} x_{q'}} dx_q \\
= & 1 - e^{k/S} I_k. \quad \text{(4.68)}
\end{align*}

The outage probability for $Q_{IR}^1$ is lower bounded as

\begin{align*}
Q_{IR}^1 & = P \left\{ \sum_{q=1}^{k} \ln \left( 1 + SG_q \right) < R \right\} \\
\leq & P \left\{ \prod_{q=1}^{k} X_q < c_k \right\}, \left\{ X_q \triangleq G_q, c_q \triangleq \frac{e^R}{S^q} \right\}, \\
& \prod_{q=1}^{k} e^{-x_q} dx_q \\
= & \int_{x_k > 0, \forall k} \cdots \int_{x_k > 0} \prod_{q=1}^{k} e^{-x_q} dx_q \\
& e^{c_1 x_1} / \prod_{q=1}^{k} x_q \\
= & \int_{0}^{\infty} \int_{0}^{\infty} \cdots \int_{0}^{\infty} \prod_{q=1}^{k} e^{-x_q} dx_q \\
& e^{-c_k \prod_{q=1}^{k-1} x_{q'}} dx_q \\
\leq & \int_{0}^{\infty} \int_{0}^{\infty} \cdots \int_{0}^{\infty} \prod_{q=1}^{k-1} e^{-x_q \left( 1 - e^{-c_k/\prod_{q'=1}^{k-1} x_{q'}} \right)} dx_q \\
& = 1 - \int_{0}^{\infty} \int_{0}^{\infty} \cdots \int_{0}^{\infty} \prod_{q=1}^{k-1} e^{-x_q - c_k/\prod_{q'=1}^{k-1} x_{q'}} dx_q \\
= & 1 - I_k. \quad \text{(4.69)}
\end{align*}
In all, $Q^\text{IR}_k$ is upper and lower bounded as

$$0 \leq (1 - e^{k/S} I_k)_+ \leq Q^\text{IR}_k \leq (1 - I_k)_+ \leq 1$$

(4.70)

4.K Proof of (4.44) and (4.45)

Proof. For $k = 2$, with $c_2 = e^R/S^2$, we evaluate the integral $I_2$ as

$$I_2 = \int_0^\infty e^{-x} e^{-c_2/x} \, dx$$

$$= 2\sqrt{c_2} K_1(2\sqrt{c_2}).$$

(4.71)

For $k = 3$, with $c_3 = e^R/S^3$, we express the double integral $I_3$ as the single integral

$$I_3 = \int_0^\infty \int_0^\infty e^{-x-y-c_3/x/y} \, dx \, dy$$

$$= \int_0^\infty e^{-x} 2\sqrt{\frac{c_3}{x}} K_1(2\sqrt{\frac{c_3}{x}}) \, dx$$

$$= 8c_3 \int_0^\infty e^{-4c_3t^2} K_1(1/t) \, dt,$$  (4.72)

where we defined $\frac{1}{t} \triangleq 2\sqrt{\frac{c_3}{x}}$.

4.L Numerical Performance Analysis Method

Not all HARQ schemes, modulation, channel coding schemes, and fading channel models allow for an analytical throughput expression to be derived. A practical way is to numerically calculate the average rate (4.14), and the average number of transmission (4.15), and use them in the throughput expression (4.13). This can be done by discretizing (4.14) and (4.15), and use the fast (Inverse) Fourier Transform (I)FFT instead of the Laplace transform. For this purpose, we use the expressions for $\rho_K$ and $\tau_K$, as in Remark 4.2. The procedure is given in Algorithm 1, where $f_Z(z)$ is the pdf of the SNR, or alternatively the MI, r.v. The pdf is uniformly sampled over $(0, \Theta)$, with $I$ samples, and normalized such that $\sum f_I \leq 1$. The average rate, $\rho_K$, and the average number of transmissions, $\tau_K$, are then computed, using zero-padding, (I)FFT, and numerical integration operations. Note that the throughput for persistent-HARQ may be approximated by assuming a large enough value of the transmit limit $K$. Other integration methods, like Gaussian quadrature methods, may alternatively be used.

4.M Persistent-IR: Generalized Upper Incomplete Fox’s H Function

Another approach, compared to the direct Laplace transform method (4.14), is to derive analytical expression for all outage probabilities, $Q^\text{IR}_{k,1}, k \in \{1, 2, \ldots, \infty\}$,
Algorithm 1 Numerical calculation of $\tau_K$, $\rho_K$, and $T_K$

$f_{smp}^\text{mp} \leftarrow f_Z(\Theta_i^I), i \in \{0, I-1\} : \text{Sample fading channel pdf}$

$f_{pad}^j \leftarrow [f_{smp}^1, \ldots, f_{smp}^{I-1}, 0, \ldots, 0_{(K-1)I}], j \in \{0, KI\} : \text{Pad}$

$F_j \leftarrow F(f_{pad}^j) : \text{FFT of discretized and padded pdf}$

$\hat{\rho}_K \approx R \left(1 - \sum_{j=0}^I f_{\rho}^j\right) : \text{Numerical estimate of the average rate}$

$\hat{\tau}_K \approx 1 + \sum_{j=0}^I f_{\tau}^j : \text{Numerical estimate of average # of transmissions}$

$T_K \approx \hat{\rho}_K / \hat{\tau}_K : \text{Numerical estimate of throughput}$

separately and then to compute $\tau_{\infty,1}^{\text{IR}} = 1 + \sum_{k=1}^\infty Q_k^{\text{IR}}$. Recently, $Q_k^{\text{IR}}$ has been given in analytical form. In [HL11, (11)-(13)], $Q_k^{\text{IR}}$ is given in a Laplace transform involving the Tricomi confluent hypergeometric function, and in [CA13, (10)], it is given in a generalized upper incomplete Fox’s $H$ function $H_{m,n}^{p,q}[\cdot]$. The latter function was introduced, defined and implemented in [YA09, Appendix A-B]. While [HL11], [CA13] give the exact $Q_k^{\text{IR}}$, determining the average number of transmissions, $\tau_{\infty,1}^{\text{IR}}(R,S)$, from these functions is a daunting task. In line with recent works, expressing IR outage probabilities with generalized upper incomplete Fox’s $H$ function, we give the throughput expression on the same form in the proposition below. However, it appears far easier to numerically evaluate (4.74), rather than (4.73).

**Proposition 4.11.** Let the instantaneous channel gain, $g$, be iid exponentially distributed, $S$ be the SNR, the effective channel be $z = \ln(1 + Sg)$, and $R$ be the decoding threshold. Then, the average number of transmissions per message is

$$\tau_{\infty,1}^{\text{IR}}(R,S) = 1 + \frac{R}{s} \int_0^\infty \int_0^z \mathcal{H}_{0,1}^{1,1} \left[ \ln(t) \left| \begin{array}{c} (0,1,S^{-1}) \\ (1,1,S^{-1}) \end{array} \right. \right] dt \, dz. \quad (4.73)$$

**Proof.** Based on Proposition 4.9, we can write the throughput of IR as

$$\tau_{\infty,1}^{\text{IR}}(R,S) = 1 + \sum_{k=1}^R \left\{ \frac{\Gamma(-s+1, 1/S)}{s^2 \Gamma(-s, 1/S)} \right\}. \quad (4.74)$$

We then use the integration rule to extract out and turn the factor $1/s^2$ of (4.74) into two integrals, and then combine (4.74) with the generalized upper incomplete Fox’s $H$ function in [YA09, Appendix A-B].
In Chapter 4, we analyzed the throughput of HARQ in general, and of ARQ, RR- and IR-HARQ in particular. We gave, among other things, analytical closed-form throughput expressions for many important persistent/truncated-ARQ- and -RR-cases, as well as tight throughput bounds for truncated-IR. In this chapter, we consider analytical throughput optimization of ARQ, RR, and (an approximate throughput optimal solution to) truncated-IR. We highlight shortcomings of the classical (SNR-parameterized) optimization framework, and then develop a new parametric optimization approach, the auxiliary parameter-method (AP-method), for which we give new exact results for many fundamental ARQ and RR communication scenarios. We also discuss another (but less versatile) parametric optimization approach, the rate parameter-method (RP-method), and provide new results for ARQ for the GD-channel introduced in Chapter 4. The proposed AP-method, due to the greater flexibility in parametrization, handles a larger class of (H)ARQ throughput optimization problems compared to the classical and rate-parameter optimization approaches. The AP-method introduced here also lay the foundation to handling a more general channel model, Chapter 7, as well as a more general performance measure, Chapter 8.

5.1 Motivation and Outline

For efficient use of invested resources, good engineering practice suggests to optimize the performance of any engineered system under study. This is also the case for wireless retransmission systems, as studied here. In Chapter 4, we made the case for focusing on the throughput performance measure. Hence, in this chapter, we focus on maximization of this performance measure with respect to the initial rate for fixed transmit power (which translates to fixed SNR). Such problem setting has been, and is, encountered in many practical wireless systems, such as for GSM/WCDMA/LTE (in downlink using fixed transmit power), WiFi, and WiMax,
Parametric Throughput Maximization of (H)ARQ

...[Mol05, Stü12], and remains a fundamental research problem. It is reasonable to ask why a throughput maximum should occur in (H)ARQ systems at all? Clearly, if packets with rate zero are transmitted, zero throughput results. Hence, any positive non-zero rate assignment will at least offer a higher throughput than with rate zero. At the other end of the extreme, when transmitting packets with a rate approaching infinity, zero throughput may also result (and often does). This occurs if the average number of transmissions (which depends on the rate) increases faster than the rate itself. Examples of such cases have been studied in [BS06, SLF08, KJSS10] and involves ARQ and RR. With zero throughput for \( R = 0 \) and \( R = \infty \), we expect the existence of one maximum, or possibly even several throughput maxima, for a positive finite rate. A notable (and well-known) exception, however, is persistent IR-HARQ in Rayleigh fading for which the throughput is maximized when the rate approaches infinity [CT01, (9)-(11)]. In this chapter, for practical and analytical interest, we are only concerned with (H)ARQ communication scenarios where we assume a finite non-zero rate maximum (or maxima).

It is, as suggested above, well-known that performance optimization is used in many practical systems, e.g. LTE and WiFi. This is achieved by adapting the initial transmit rate through adaptation of modulation scheme and channel coding. Likewise, analytical throughput maximization with respect to the rate has been considered in many theoretical works, e.g [BS06, SLF08, KJSS10]. Analytical maximization is motivated by the potential insights that may be gained. For example, by comparing the optimal throughput performance between different (H)ARQ-cases may reveal the best candidate to use in a system, or by benchmarking simulated/measured real systems against the theoretically optimal performance, a (lack of) potential for improvements of a system is indicated. In the following, we focus on optimal throughput expressions that can be put on (semi-)closed-form. Those expressions, in contrast to possible non-(semi-)closed-form counterparts, make comparison and insights easier to attain.

A well-known alternative to analytical optimization, is numerical optimization. The latter approach is good for practical algorithms aiming at real world systems. However, numerical optimization does not give the same level of insight as analytical, or closed-form, expressions do. It also opposes a mathematical tradition of simplifying results as much as possible, i.e. if a closed-form expression can be derived, we generally do so.

The main theme, throughput optimization is addressed in Section 5.2. More specifically, the classical, the proposed AP-method, and the RP-method are considered in Sections 5.2.1, 5.2.2 and 5.2.3, respectively. Optimal throughput expressions for important (H)ARQ-cases are also given in each subsection. Numerical results and discussions are treated in Section 5.3, and the chapter ends with summary and conclusions in Section 5.4.
5.2 Performance Analysis

We start this section by considering classical throughput optimization. After presenting some results for ARQ and RR in Rayleigh fading, we exemplify the problem with the classical optimization approach. We then present the main contribution of this chapter, the auxiliary-parameter method. This method is applied to more general ARQ- and RR-cases, as well as to truncated-IR for $K = 2$. We then look at the less versatile RP-method, and formalize it in the context of ARQ and RR. Despite the shortcomings of the RP-method, we offer some new expressions for the optimal throughput of ARQ for the GD-channel.

5.2.1 Throughput Optimization: Classical Method

The optimization problem often encountered in the analysis of (H)ARQ, and also the subject of interest here, is to maximize the throughput with respect to the rate for a given SNR. This can be formalized as solving the problem

$$\begin{align*}
\max_R & \quad T(R, S) \\
\text{subject to} & \quad R \geq 0,
\end{align*}$$

(P1)

where $T$ is the throughput, $R$ is the rate, and $S$ is the given average SNR. We denote the (globally) maximum throughput value as $T^*$, and the (globally) optimal rate point as $R^*$. The latter is defined through $T^* \triangleq T(R^*, S)$. For a meaningful (and generally practically encountered) problem setting, we consider only HARQ-cases where the throughput $T(R, S)$ is a continuous function of the rate $R$, a continuous and monotonically increasing function of the SNR $S$, and has (at least) one local maximum for the rate $R \geq 0$. The first two assumptions makes sense from a physical point of view, i.e. if we increase the SNR $S$, we expect the throughput $T(R, S)$ to increase monotonically, and if we increase the rate $R$, we expect the throughput $T(R, S)$ to also change smoothly (decrease or increase). For the third assumption, it is the HARQ communication problem under study that dictates whether no maximum, a maximum, or several optima (with local maxima and minima) exist. The standard optimization approach now tell us to solve the equation $dT(R, S)/dR = 0$ for the (globally) optimal rate point $R^*$. The resulting equation from $dT(R, S)/dR = 0$ is said to be a necessary optimality criterion. For a throughput maximum $T^*$, it is also required that $d^2T(R, S)/dR^2|_{R^*} < 0$. In the sequel, we check for a global maximum for given explicit expressions, but we omit elaborating on such details henceforth. Note that whenever multiple throughput optima (both maxima and minima) exist for $R \geq 0$ (which is known to be the case for truncated-IR), we then talk about an optimal rate set, $R^* = \{R^*_1, R^*_2, \ldots\}$, where each optimal rate point having a corresponding optimal throughput.

\footnote{Hence, it is vital to investigate the nature of the function being maximized to guarantee the study of a global maximum.}
While the main contribution of this chapter is found in Section 5.2.2, and some new results are given in Section 5.2.3, we start by considering throughput optimization of ARQ and RR in Rayleigh fading without diversity (i.e. $N_{tx} = 1$, $N_{rx} = 1$, $m^N = 1$) with the classical optimization method. Those results are included for completeness, and to illustrate the limited set of known results, for the classical optimization framework.

**Classical Method: ARQ**

The throughput of ARQ in Rayleigh fading without diversity was given in (4.22). Based on the classical optimization method (described above), the maximum throughput is given in the following proposition.

**Proposition 5.1.**(ARQ with Rayleigh fading and no diversity, [BS06, (9)], [SLF08, (6)]). Let, $R$ be the rate and $S$ the SNR. The maximum throughput value, the optimal rate point, and the parameter SNR, for ARQ in Rayleigh fading (without diversity), are

$$T_{K,1}^{\text{ARQ}}(S) = W_0(S)e^{1/S-1/W_0(S)}, \quad (5.1)$$

$$R_{K,1}^{\text{ARQ}}(S) = W_0(S), \quad (5.2)$$

$$S \in [0, \infty),$$

where $W_0(\cdot)$ denotes the main branch of Lambert’s $W$ function.

**Proof.** The proof is given in Appendix 5.A. \qed

We note that the equivalent of (5.1)-(5.2) has also been shown in [BS06, (9)], and also (in a slightly more complicated form) in [SLF08, (6)]. The optimality expressions, (5.1)-(5.2), use the implicitly formulated Lambert’s W function. Since the Lambert W function is implicit, it is not a true closed-form expression. However, since Lambert’s W function is defined with the exponential, an elementary function, we refer to expressions involving only Lambert’s W function, and other elementary operations and functions, as a semi-closed-form expression.

**Classical Method: Persistent-RR**

Following a similar procedure as for ARQ, we consider throughput optimization of persistent-RR in the proposition below.

**Proposition 5.2.**(RR with Rayleigh fading and no diversity) Let, $R$ be the rate and $S$ the SNR. The maximum throughput value, the optimal rate point, and the parameter SNR, for RR in Rayleigh fading without diversity, are

$$R_{\infty,1}^{\text{RR}}(S) = \frac{S}{S-1}W_0\left(\frac{S-1}{e}\right), \quad (5.3)$$
\[ R_{\infty,1}^{RR}(S) = 1 + W_0 \left( \frac{S - 1}{e} \right), \quad S \in [0, \infty). \]  

(5.4)

**Proof.** The proof is given in Appendix 5.B.

We also note that in [KJSS10, (7)], an optimum SNR decoding threshold is derived for RR operating in Rayleigh fading. With the threshold translated into an optimum rate point, it would be equivalent to (5.4).

**A Problem Case**

We now turn our attention to one simple example where the classical optimization method fails to deliver not just a closed-form, but even a semi-closed-form, expression.

**Example 5.1.** Consider RR in Rayleigh fading with \( N = 2 \) branch MRC. The throughput was derived in (4.28) and is \( T = 4R/(3 + 2\Theta + e^{-2\Theta}) \), with \( \Theta = (e^R - 1)/S \). For this case, the necessary optimality criterion is

\[
\frac{1}{R^*} - \frac{2e^{R^*}}{S} \frac{1 - e^{2(e^{R^*} - 1)/S}}{3 + 2(e^{R^*} - 1)/S + e^{-2(e^{R^*} - 1)/S}} = 0. \quad (5.5)
\]

It is observed that the optimality criterion (5.5) can not be solved for \( R^* \) into a closed-form expression.

We note two important limitations with the classical optimization framework when dealing with (H)ARQ. First, the classical optimization method does not always yield an optimality criterion that can be expressed in a closed-form (or a semi-closed-form), but instead often (or as typically observed in practise) gives a complicated (unsolvable) equation. Second, we are not aware of any other (H)ARQ-cases than the ones given above, i.e. than for ARQ and persistent-RR in Rayleigh fading with no diversity, where the classical method yields (semi-)closed-form expressions. With those severe limitations, it is crucial to develop an alternative, more versatile, optimization framework.

**5.2.2 Throughput Optimization: Auxiliary-Parameter Method**

Why does the classical optimization method fail? The answer is evident once i) realizing that there is a problem, and ii) recognizing what the solution is. For the analysis of (H)ARQ, we are often dealing with fairly complicated throughput expressions, for which there are simply no way to express \( R^*(S) \) in a (semi-)closed-form. This leads to the question if there is a way to recast the problem such that a solution on a (semi-)closed-form can be found. In the following, we reformulate the objective of the optimization problem (somewhat). Instead of finding explicit
expressions for \( T^*(S) \) and \( R^*(S) \), we note that researchers in the field are, e.g., often interested in plotting, or studying the asymptotes of, the optimal throughput \( T^* \) vs. the SNR \( S \), or the optimal rate \( R^* \). With such goals in mind, we develop a generalized parametric optimization method below which handles a large class\(^2\) of practically interesting (H)ARQ systems.

The basic idea is to introduce an auxiliary parameter that is judiciously selected to ensure (semi-)closed-form mappings to both the optimal rate point \( R^* \) and the corresponding SNR. Thus, neither a mapping from \( S \) to \( R^* \), nor a mapping from \( R^* \) to \( S \) require a (semi-)closed-form expression. To exemplify a well-known function that is expressed in a closed-form, but does not have an inverse with a closed-form wrt elementary functions, consider (Kepler’s) equation \( f(x) = x - c \sin(x) \). The inverse \( f^{-1}(x) \) is, however, known to be expressible as an infinite (but not always converging) polynomial series.

We now specialize this general idea to throughput optimization to a class of retransmission schemes that includes ARQ and RR in the following theorem.

**Theorem 5.1.** Let \( T = R/f_\Theta \), and assume that \( f_\Theta(\Theta) \) has an argument of the form \( \Theta(R,S) = f_R(R)/f_S(S) \). Then, the optimal rate point, the corresponding SNR, and the optimal throughput are

\[
R^*(\Theta) = g^{-1}_R (g_\Theta), \tag{5.6}
\]

\[
S(\Theta) = f^{-1}_S \left( \frac{f_R(R^*)}{\Theta} \right), \tag{5.7}
\]

\[
T^*(\Theta) = \frac{R^*}{f_\Theta(\Theta)}, \tag{5.8}
\]

where

\[
g_R(R) \triangleq \frac{R f'_R(R)}{f_R(R)}, \tag{5.9}
\]

\[
g_\Theta(\Theta) \triangleq \frac{f_\Theta(\Theta)}{f'_{\Theta}(\Theta) \Theta}, \tag{5.10}
\]

\( \Theta \in [0, \infty) \).

**Proof.** The proof is given in Appendix 5.C. \( \square \)

We noted in Chapter 4 that some practical (H)ARQ schemes, RR and ARQ have a threshold of the form \( \Theta_{gd} = (e^{R_{gd}} - 1)/S_{gd} \) (GD-channel easily included), and the throughput has the form \( T(R,S) = R/f_\Theta(\Theta_{gd}) \). Under those conditions, the solution to the throughput optimization problem is given by the following corollary.

\(^2\)While a large class of (H)ARQ-cases, can be handled, e.g. all RR- and ARQ-cases where the throughput \( T = R/f_\Theta(\Theta) \) has an argument \( \Theta \) on the form \( \Theta(R,S) = f_R(R)/f_S(S) \), all cases can not be handled. Exceptions are, e.g., RR-MIMO and persistent/truncated-IR in general.
Corollary 5.1. Let \( f_R(R_{gd}) = e^{R_{gd}} - 1 \), \( f_S(S_{gd}) = S_{gd} \), and \( \Theta_{gd}(R_{gd}, S_{gd}) = f_R(R_{gd})/f_S(S_{gd}) \). The optimal rate point, the optimal throughput, and the SNR are parametrically given (in the auxiliary parameter \( \Theta_{gd} \)) by

\[
T^*(\Theta_{gd}) = \frac{R^*}{\Theta_{gd}}, \quad (5.11)
\]

\[
S_{gd}(\Theta_{gd}) = \frac{e^{R^*_0} - 1}{\Theta_{gd}}, \quad (5.12)
\]

\[
R^*(\Theta_{gd}) = g_{\Theta} + W_0(-g_{\Theta}e^{-g_{\Theta}}), \quad (5.13)
\]

where

\[
g_{\Theta}(\Theta_{gd}) = \frac{f_{\Theta}}{\Theta_{gd}f'_{\Theta}}, \quad (5.14)
\]

\[\Theta_{gd} \in [0, \infty).\]

The use of Corollary 5.1 is exemplified below.

Example 5.2. Consider RR in GD-channel with \( N_{gd} = 2 \), i.e. a slightly more general channel than in Example 5.1. The throughput is

\[
T = \frac{4R}{(3+2\Theta_{gd}+e^{-2\Theta_{gd}})} \quad (4.16),
\]

which means that

\[
ge_{\Theta}(\Theta_{gd}) = \frac{3 + 2\Theta_{gd} + e^{-2\Theta_{gd}}}{2\Theta_{gd}(1 - e^{-2\Theta_{gd}})}, \quad (5.15)
\]

and the solution is given with Corollary 5.1.

In Fig. 5.1, we illustrate the AP optimization method given in Corollary 5.1. It is natural to briefly discuss the form of the optimality solution in Corollary 5.1, especially (5.13). Since it is required that \( R^* > 0 \), (5.13) implies that \( g_{\Theta}(\Theta) > 1 \). Hence, in order for a throughput maximum to exist, the criterion \( f_{\Theta}(\Theta) > \Theta f'_{\Theta}(\Theta) \) for \( \Theta \geq 0 \), must be fulfilled. We also note from (5.13) that the optimal rate point \( R^* \) is monotonically increasing in \( \Theta \).

A lingering question that remains to be answered is how the result can be used in practice? The auxiliary parametric method can for example be used to parametrically plot optimal throughput curves, e.g. the optimal throughput vs. SNR, \( (x, y) = (10\log_{10}(S(\Theta)), T^*(\Theta)) \), or the optimal rate point vs. SNR, \( (x, y) = (10\log_{10}(S(\Theta)), R^*(\Theta)) \), or the optimal throughput vs. the optimal rate point, \( (x, y) = (R^*(\Theta), T^*(\Theta)) \). The AP-method can also be used to study or plot the slope of the optimized throughput vs., e.g., SNR by using the chain rule \( \frac{dT^*(S)}{dS} = \frac{dT^*(\Theta)}{d\Theta} \frac{d\Theta}{dS} \). Also, asymptotic characteristics for low and high SNRs, e.g. for \( T^*(S) \) or \( dT^*(S)/dS \), can be examined by letting \( \Theta^* \to 0 \) and \( \Theta^* \to \infty \) respectively in the parameterized expressions. In the limit, it is often possible to combine the parameterized expressions into a non-parameterized closed-form expression.
5.2.3 ARQ and Persistent/Truncated-RR with GD-Channel

In the following, we leverage the AP-method and study several important (H)ARQ communication scenarios for which the classical optimization approach fail to deliver closed-form solutions.

**AP-Method: ARQ**

An important HARQ communication case, due to its basic nature, is ARQ with some form of diversity. We consider throughput optimization for the GD-channel, i.e. OSTBC, MRC and/or integer restricted Nakagami-$m$ channel, in the following proposition.

**Proposition 5.3.** (ARQ in GD-channel). Let (4.22) be the throughput of ARQ with $N_{gd}$-fold diversity, and $\Theta_{gd}$ be the decoding threshold. Then, the maximum throughput is given by Corollary 5.1 with

$$g_\Theta(\Theta_{gd}) = \Gamma(N_{gd}, \Theta_{gd})\Theta_{gd}^{-N_{gd}}e^{-\Theta_{gd}}.$$  \hfill (5.16)

**Proof.** The proof is given in Appendix 5.D. \hfill \Box

Whereas the classical method could not handle the ARQ case for $N_{gd} \neq 1$, we note that with the AP-method, we easily handle diversity of any order.

**AP-Method: Persistent-RR**

Another important HARQ communication case is persistent-RR in the GD-channel.

**Proposition 5.4.** (Persistent-RR in GD-channel). Let (4.23) be the throughput of persistent-RR with $N_{gd}$-fold diversity, and $\Theta_{gd}$ be the decoding threshold. Then, the maximum throughput is given by Corollary 5.1 and

$$g_\Theta(\Theta_{gd}) = \frac{(N_{gd} + 1)/2 + \Theta_{gd} + \sum_{n=1}^{N_{gd}-1} a_n e^{-\Theta_{gd}b_n}/b_n^H}{\Theta_{gd}(1 + \sum_{n=1}^{N_{gd}-1} a_n e^{-\Theta_{gd}b_n})}.$$  \hfill (5.17)
5.2. Performance Analysis

Proof. The proof is given in Appendix 5.E.

More explicit expressions for $g_{d}(\Theta_{d})$ for persistent-RR with $N_{gd} = \{2, 3, 4\}$ can be derived based on (4.28)-(4.30).

5.2.4 Truncated-IR in Rayleigh Fading Channel

It has been shown, e.g. in [CT01], that for persistent-IR in Rayleigh fading, the optimal rate point is infinite and the optimal throughput is the ergodic channel capacity. Does this imply that no interesting throughput optimization problem exists for IR? It is easy to see that the optimal rate point must be finite for truncated-IR as a zero and an infinite rate both have zero throughput. A non-zero throughput is then expected in between, for a finite rate. The case with transmission limit $K = 1$ is trivial and simply corresponds to ARQ. However, a transmission limit with $K = 2$ offers the opportunity for throughput optimization, and could also give some insight to the case of an arbitrary $K$. The choice of $K = 2$ can be further motivated since we expect that mainly one or two transmissions are used when the rate is selected to maximize the throughput (except at extremely low SNRs). We therefore consider throughput optimization for truncated-IR with $K = 2$. The exact throughput expression for $K = 2$ truncated-IR (without diversity) is $T_{2,1}^{IR} = R(1 - Q_{1,2}^{IR})/(1 + Q_{1,2}^{IR})$. The exact throughput in a Rayleigh fading channel is more complicated since no simple expression for $Q_{1,2}^{IR}$ is known. However, we derived an upper bound in (4.51), and a lower bound in (4.52), which are tight with increasing SNR. We consider optimization of the upper throughput bound, as an approximation. In a first attempt, we compute $d \ln(T_{1,2}^{IR})/d R = 0$, and get the necessary optimality criterion

$$\frac{1}{R^*} = e^{R^*/2} K_0 \left( \frac{2e^{R^*/2}}{S} \right) - \frac{e^{R^*}}{S} \frac{1}{2e^{(e^{R^*} - 1)/2} - 1} = 0,$$  \hspace{1cm} (5.18)

However, (5.18) is not easily solvable into a closed-form expression, not even with the auxiliary parametrization trick. This is not an artefact of the approximation, but due to that IR does not have a decoding threshold of the form $\Theta(R, S) = f_{d}(R)/f_{a}(\Theta)$, a prerequisite for the AP-method. Nevertheless, by considering further (yet accurate) approximations, the AP-method remains useful. To proceed, we split the problem into two approximations, one for the low- and one for the high-rate region.

High-rate Maximum

We note that the higher-rate maximum emerge when $Q_{1,1}^{IR} \rightarrow 1$. The approximate throughput in the high-rate regime is then

$$T_{2,1}^{IR} \approx R(1 - Q_{2,1}^{IR}) = \frac{R e^{2R/S} 2 \sqrt{c} K_1(2 \sqrt{c})}{2}.$$

(5.19)
To illustrate that the high-rate approximation makes sense, we illustrate it together with the numerically computed throughput of truncated-IR in Fig. 5.7. With the high-rate approximation, we now get the (slightly modified) necessary optimality criterion

\[
\frac{1}{R} - \frac{e^{R/2}}{S} K_0 \left( \frac{2e^{R/2}}{S} \right) K_1 \left( \frac{2e^{R/2}}{S} \right) = 0, \quad (5.20)
\]

which can be parameterized. We set \( x \triangleq 2e^{R/2}/S \) as the auxiliary parameter and solve for the optimal rate point

\[
R_{IR}^{*2,1} \approx \frac{2K_1(x)}{xK_0(x)}. \quad (5.21)
\]

We insert (5.21) into the definition of \( x \), and solve for the SNR

\[
S_{IR}^{*2,1} \approx \frac{K_1(x)}{x}. \quad (5.22)
\]

Subsequently, (5.21), (5.22) are inserted into (5.19), and we get

\[
T_{IR}^{*2,1} \approx \frac{2K_1(x)^2}{K_0(x)} e^{x} e^{-\frac{K_1(x)}{xK_0(x)}}. \quad (5.23)
\]

What does this result tell us? It tells us, e.g., that there exist optimization problems that has another parametrization (here \( x = 2e^{R/2}/S \) rather than \( \Theta = (e^R - 1)/S \)) that works well with the AP-method.

**Low-rate Optima**

For the lower-rate (LR) local maximum, a sensible approximation, since \( Q_{2,1} \to 0 \), is

\[
T_{IR}^{*2,1} \approx \frac{R}{1 + Q_{IR}^{*2,1}} = \frac{R}{2 - e^{-\Theta}}, \quad (5.24)
\]

where \( \Theta = (e^R - 1)/S \). For optimization of (5.24), we use the AP-method, with \( f_\Theta(\Theta) = 2 - e^{-\Theta} \) which gives \( g_\Theta(\Theta) = (2e^\Theta - 1)/\Theta \). We also show the low-rate approximation in Fig. 5.7. It is interesting to note that the AP-method is able to plot both the two maxima and the minimum.

**5.2.5 Throughput Optimization: Rate-Parameter Method**

In this section, we address the optimization problem from an alternative viewpoint. Compared with the auxiliary parameter method, this method gives a more direct result, i.e. excluding the need of an auxiliary parameter, but as we will see handles a smaller class of (H)ARQ-cases. We start with a motivating example below, and then formalize the method.
Example 5.3. Consider ARQ in GD-channel with $N_{gd} = 2$. The throughput is given by (4.21), which can be written as $T_{1,2}^\text{ARQ} = R e^{-\Theta_{gd}} (1 + \Theta_{gd})$, with $\Theta_{gd} = (e^{R_{gd}} - 1)/S_{gd}$. With $d \ln(T(R, S))/dR = 0$, the necessary optimality criterion is

$$\frac{1}{R^*} - \frac{e^{R_{gd}^*}}{S_{gd}} + \frac{e^{R_{gd}^*} + 1}{e^{R_{gd}^*} - 1} = 0. \quad (5.25)$$

As noted in this case, it is hard to solve for $R^*(S)$ analytically, and the classical method for finding the optimal rate point does not give a closed-form expression.

This example hints at an alternative solution to the classical- and the AP-method. Assuming a one-to-one mapping between the optimal rate point $R^*$ and the SNR $S$, we can solve for $S$ as function of $R^*$. We now formalize the RP-method with the following theorem.

Theorem 5.2. Let $T = R/f_\Theta$, and assume that $f_\Theta(\Theta)$ has an argument of the form $\Theta(R, S) = f_R(R)/f_S(S)$. Then, the optimal throughput and SNR are parametrically given in the rate parameter $R$ by

$$T^*(R^*) = \frac{R^*}{f_\Theta(\Theta)^\prime}, \quad (5.26)$$

$$S(R^*) = f_S^{-1}\left(\frac{f_R(R^*)}{\Theta(R^*)}\right), \quad (5.27)$$

$$R \in [0, \infty),$$

where

$$\Theta(R^*) = g_\Theta^{-1}(g_R), \quad (5.28)$$

$$g_\Theta(\Theta) = \frac{f_\Theta(\Theta)}{f_\Theta(\Theta)^\prime}, \quad (5.29)$$

$$g_R(R^*) = \frac{R^* f_R(R^*)}{f_R(R^*)}. \quad (5.30)$$

Similar to the auxiliary parameter method, we can now, e.g., plot curves for the maximum throughput vs. SNR as $(x, y) = (10 \log_{10}(S(R^*)), T^*(R^*))$, or maximum throughput vs. the optimal rate point as $(x, y) = (R^*, T^*(R^*))$.

It also becomes evident that the RP-method faces two severe problems and limitations. First, we are not always able to analytically solve for the inverse $g_\Theta^{-1}(\Theta)$. Second, we require different solutions, one for every (H)ARQ case, as $g_\Theta(\Theta)$ differ from case to case. In contrast, the proposed AP-method in Section 5.2.2, does not exhibit such shortcomings.

Corollary 5.2. Let $f_R(R) = e^R - 1$, $f_S(S) = S$, and $\Theta(R, S) = f_R(R)/f_S(S)$. The optimal rate point, the optimal throughput, and the SNR are parametrically given (in the rate-parameter $R^*$) by

$$T^*(R^*) = \frac{R^*}{f_\Theta(\Theta)^\prime}. \quad (5.31)$$
Figure 5.2: The RP-method in Corollary 5.2.

\[
S(R^*) = \frac{e^{R^*} - 1}{\Theta}, \\
R \in [0, \infty),
\]

where

\[
\Theta(R^*) = g_{\Theta}^{-1}(g_R), \\
g_{\Theta}(\Theta) = \frac{f_{\Theta}(\Theta)}{\Theta f'_{\Theta}(\Theta)}, \\
g_R(R^*) = \frac{R^* e^{R^*}}{(e^{R^*} - 1)}. 
\]

We illustrate Corollary 5.2 in Fig. 5.2.

The basic idea of expressing the optimal throughput and SNR parameterized in the optimal rate point\(^3\) has, however, also been suggested in [TC02] for slotted-Aloha throughput optimization. However, Corollary 5.2 is a tailored systematic approach to solve throughput optimization of (H)ARQ, i.e. identifying the function \(g_{\Theta}(\Theta)\), and using its inverse \(g_{\Theta}^{-1}(\Theta)\). We extend the results expressed in the optimal rate point, and handle new HARQ-cases, in the following.

We now consider ARQ and Persistent-RR with the GD-channel. The following cases handles, e.g., the important case of Alamouti-diversity with ARQ.

**RP-Method: ARQ in GD-Channel**

We first consider the general case of \(N_{gd}\)-fold diversity, and then specialize to the interesting case \(N_{gd} = 2\), which includes Alamouti diversity and 2-branch MRC in Rayleigh fading, as well as the case where \(N_{gd} = 1\).

\(^3\)This idea arose when trying to solve the throughput optimization problem for ARQ and RR in Rayleigh fading channel which was hard to solve for \(R^*\) in \(S\), but simple to solve for \(S\) in \(R^*\).
Proposition 5.5. (ARQ in GD-channel) Consider ARQ with $N_{gd}$-fold diversity, where $T = R\Gamma_1(N_{gd}, \Theta_{gd})$. Then, the maximum throughput is given by Corollary 5.2 with

$$g_\Theta(\Theta_{gd}) = \frac{(N_{gd} - 1)!}{\Theta_{gd}^{N_{gd}}} \sum_{n=0}^{N_{gd}-1} \frac{\Theta_{gd}^n}{n!}. \quad (5.36)$$

Proof. The proof is given in Appendix 5.F. \qed

From (5.33) and (5.36), we see that the polynomial equation to solve for $\Theta_{gd}$ is

$$\frac{\Theta_{gd}^N}{N_{gd}} - \frac{(1 - e^{-R*_{gd}})}{N_{gd} R*_{gd}} \sum_{n=0}^{N_{gd}-1} \frac{\Theta_{gd}^n}{n!} = 0.$$ 

The solution to this equation has only one positive real root $\Theta$. This can e.g. be seen by Descartes’ rule of signs. To the extent an $N_{gd}$-degree polynomial has a known closed-form solution, we can thus always give a parametric closed-form expression for the maximum throughput for ARQ with the GD-channel. To exemplify, two important special cases are for $N_{gd} = 2$ and $N_{gd} = 1$. We consider $N_{gd} = 2$ in the following proposition.

Proposition 5.6. Let $N_{gd} = 2$, then the maximum throughput for ARQ is

$$T^\text{ARQ}_1(R_{gd}^*) = R^* e^{-\Theta^\text{ARQ}_{gd}} (1 + \Theta^\text{ARQ}_{gd}), \quad (5.37)$$

$$\Theta^\text{ARQ}_{gd}(R^*) = e^{R^*_{gd}} - 1, \quad (5.38)$$

where

$$\Theta^\text{ARQ}_{gd}(R^*) = 2 \left( \sqrt{1 + \frac{4R^*_{gd}}{1 - e^{-R^*_{gd}}} - 1} \right)^{-1}. \quad (5.39)$$

Proof. The proof is given in Appendix 5.G. \qed

This result is, e.g., useful for Alamouti transmit diversity (then with the GD SNR $S_{gd} = S/2$), and 2-branch MRC. In a similar manner, but left out here for brevity, we give the exact analytical expression for the maximum throughput with the RP-method for $N_{gd} = 3$ in [LRS14c].

An interesting special case is for $N = 1$, no diversity, and Rayleigh fading. Using, Prop. 5.5 for $N = 1$ in Corollary 5.2 allows the maximum throughput to be written on the parametric form

$$T^\text{ARQ}_{1,1}(R^*) = R^* e^{(e^{-R^*} - 1)/R^*}, \quad (5.40)$$
\[ S_{1,1}^{\text{ARQ}}(R^*) = R^* e^{R^*}, \quad (5.41) \]
\[ R^* \in [0, \infty). \]

For this special case of ARQ, limited to Rayleigh fading and no diversity, Tuninetti have reported the above result in [Tun02, (3.17)]. In contrast, (5.36) allows for general diversity order to be handled for ARQ with GD-channel, whereas (5.37)-(5.39) specifically gives closed-form expression for \( N_{gd} = 2 \)-fold diversity order.

We also note that we can use all three approaches, the classical optimization method, (5.1)-(5.2), the AP-method, (5.11)-(5.13) with (5.16) where \( N = 1, \Theta = (e^{R} - 1)/S \), as well as the RP-method, (5.40)-(5.41), to express the maximum throughput for ARQ in Rayleigh fading without diversity.

**RP-Method: Persistent-RR in GD-channel**

We saw in (5.17), using the proposed AP-method, that Persistent-RR in the GD-channel could be handled. In contrast, the RP-method does not seem possible to apply on any diversity case for RR, i.e. when \( N_{gd} > 1 \) and \( F(s) = 1/(1 + s)^{N_{gd}} \) for any \( \Theta_{gd} \). A simple example, illustrating the problem, is for RR with \( N_{gd} = 2 \). The throughput is given by (4.31), which gives \( g_d(\Theta_{gd}) = (3 + 2\Theta_{gd} + e^{-2\Theta_{gd}})/(1 - e^{-2\Theta_{gd}})/(2\Theta_{gd}) \), for which the inverse \( g_d^{-1}(g_R) \) is hard to determine. In fact, we have not found any HARQ communication case, except the one discussed below, that allows the RP-method to be used. It is for this reason, this limitation of the RP-method, that the AP-method was developed.

Nevertheless, it is straightforward to handle the special case of persistent-RR for \( N = 1 \), no diversity, and Rayleigh fading. Using the optimality criterion in Prop. 5.2, and solving for \( S \) we get the parametric form

\[ T_{\infty,1}^{\text{RR}}(R^*) = e^{-R^*} + R^* - 1, \quad (5.42) \]
\[ S_{\infty,1}^{\text{RR}}(R^*) = e^{R^*} (R^* - 1) + 1, \quad (5.43) \]
\[ R^* \in [0, \infty). \]

Like for the special case of ARQ, Tuninetti have reported this result for the special case of RR with \( N = 1 \) (no diversity) in [Tun02, (3.18)].

Like the ARQ case, we note that we can use all three approaches, the classical optimization method, (5.3)-(5.4), the AP-method, (5.11)-(5.13) with (5.17) where \( N = 1, \Theta = (e^{R} - 1)/S \), as well as the RP-method, (5.42)-(5.43), to express the maximum throughput for ARQ in Rayleigh fading without diversity.

### 5.3 Numerical Results and Discussions

The objective in this section is to demonstrate the use of the AP-method, show that it works as desired, and also to show the optimal throughput vs. SNR, and the rate, for two (H)ARQ communication examples.
First, in Fig. 5.3, we plot the throughput vs. SNR for ARQ, where Alamouti’s TX diversity scheme is deployed together with Rayleigh fading channel for rates $R = \{2, 4, 6, 8\}$ [b/Hz/s]. Additionally, the maximum throughput, as well as the optimal rate point, are also plotted vs. the SNR. A corresponding plot for RR is also shown in Fig. 5.4. The reason to consider Alamouti’s diversity scheme here is that the RP-method can not be used for RR. Yet another reason is that, while the RP-method can be used for ARQ with Alamouti’s diversity (5.33)-(5.35), no closed-form expression for the maximum throughput has been given previously in
the literature. We observe that the curves for the maximum throughput gently touches the curves for the discrete rates, as expected.

Next, in Fig. 5.8, we plot the throughput vs. rate for ARQ using Alamouti’s TX diversity scheme in a Rayleigh fading channel for rates \( S = \{0, 10, 20, 30\} \) [dB]. Moreover, the maximum throughput vs. the optimal rate point is plotted. A corresponding plot for RR is also shown in Fig. 5.6. We observe, again as expected, that the maximum throughput curves traverses through throughput maximum of the curves corresponding to discrete SNRs.

Now, in Fig. 5.7, we plot \( T^*_2 \), Algorithm 1 in Appendix 4.L, and the bound
5.4 Summary and Conclusions

In this chapter, we have considered analytical maximization of the throughput with respect to the rate. We noted that finding a closed-form maximum throughput expression using a classical optimization framework, i.e., expressing the maximum throughput and the optimal rate point wrt to the SNR, was unfeasible even for a simple HARQ case such as RR with two-branch MRC and Rayleigh fading (Example 5.1). We proposed a parametric throughput maximization method, using (for the problem) a well chosen auxiliary-parameter, that was able to handle a large class of (H)ARQ schemes. We also formalized, and discussed, a rate-parameter based parametric throughput maximization method, where we gave a new expression for the maximum throughput of ARQ in the GD-channel. The overall philoso-

Figure 5.7: Throughput, $T$, (Numerical and bound (4.49)) vs. rate, $R$, of truncated-IR with $K = 2$ for Rayleigh fading channel and SNR $S = \{0, 5, \ldots, 40\}$ [dB]. Also showing, the maximum throughput, (5.23), vs. optimal rate, (5.21), for the high-rate local maximum, as well as (5.24), vs. optimal rate, for the low-rate local optima.

$\hat{T}^\text{IR}_{2,1}$, (4.49), vs. the rate. We observe an excellent match between the curves, the appearance of a second local maximum, and how the higher-rate local maximum becomes the global maximum, as the SNR increases. This transient behaviour of the optimal rate point has, for instance, been discussed in [SCA10, pp. 2588]. In Fig. 5.7, the maximum throughput vs. the optimal rate point for the AP-method with (5.23) vs. (5.21) is also plotted. We find a good match with the higher-rate local maximum. For the lower-rate local maximum (and minimum), we also plot (5.24) vs. the rate. Also here is a good match with the local optima observed. Hence, the approximations (5.21) (5.23), and (5.24) appears to be good approximations for the local optima in the high SNR range.
Figure 5.8: Basic philosophies behind the classical SNR-parameterized, the rate-parameterized, and the auxiliary-parameterized optimization methods.

Philosophies behind the classical-, AP-, and RP-optimization methods are summarized and sketched in Fig. 5.8. We note that while the classical and the RP-method appear simpler (in the sense that they do not introduce any additional variable), the AP-method formulation is less constrained and can include the classical and the RP-method as special cases.

We finally note some limitations of the presented optimization schemes. There are known cases for which the presented optimization approaches, i.e. the classical, AP-, and RP-method, does not handle. Clearly, since no analytical closed-form expressions are known for spatially multiplexed MIMO for ARQ, RR, and IR, the exact optimal throughput cannot be determined analytically. The same applies for truncated-IR in Rayleigh fading too. However, we showed that the AP-method was useful for approximative solutions for truncated-IR with $K = 2$, giving insights about a transient behaviour of the global optimal rate point. Whether the AP-method is applicable to (H)ARQ-cases with the throughput expression on other forms, need to be assessed case by case.

In Chapter 7 and 8, we extend the AP-method to retransmission systems employing a more general channel model, the matrix exponentially distributed channel, and to a more general performance measure, the effective capacity, respectively.
5.5 Appendices

5.A Proof of Proposition 5.1

Proof. The throughput expression is given by (4.22). The necessary optimality criterion is found to be \( Re^R = S \), which is solved for \( R^* \) with Lambert’s \( W \) function. Inserting the expression for \( R^* \) in (4.22), and using the fact that \( e^{W_0(S)}/S = 1/W_0(S) \) yields the optimal throughput expression. A check also confirms that the optima is a maximum, and that it is global.

5.B Proof of Proposition 5.2

Proof. The throughput expression is given by (4.27). The necessary optimality criterion is found to be \( 1/R - e^{R^*}/(S + e^{R^*} - 1) = 0 \), which after some rearrangements can be solved with Lambert’s \( W \) function. Inserting \( R^* \) in (4.27), and noting that \( e^{W_0((S-1)/e)} = ((S-1)/e)/W_0((S-1)/e) \) gives (5.3).

5.C Proof of Theorem 5.1

Proof. Consider \( T = \frac{R^f_{\Theta}}{f(R)f(S)} \). Using \( d\ln(T(R, S))/dR = 0 \), since \( \ln(x) \) is a monotonically increasing function in \( x \), we get

\[
\frac{1}{R} \frac{f'_{\Theta}(\Theta)}{f_{\Theta}(\Theta)} = 0 \\
\Rightarrow \frac{1}{R} \frac{f'_{\Theta}(\Theta)}{f_{\Theta}(\Theta)} = 0 \\
\Rightarrow \frac{Rf'_{\Theta}(R)}{f_{\Theta}(R)} = \frac{f_{\Theta}(\Theta)}{\Theta f'_{\Theta}(\Theta)}. \tag{5.44}
\]

We define the LHS and the RHS as \( g_R(R) \) and \( g_{\Theta}(\Theta) \) respectively.

5.D Proof of Proposition 5.3

Proof. The throughput expression (4.22) implies \( f_{\Theta}(\Theta gd) = 1/\Gamma_r(N_{gd}, \Theta_{gd}) \), which is used in (5.14) to determine \( g_{\Theta}(\Theta_{gd}) \).

5.E Proof of Proposition 5.4

Proof. We get \( f_{\Theta}(\Theta_{gd}) \) from (4.26).
5.F Proof of Proposition 5.5

Proof. We have, $f_{\Theta} = 1/\Gamma_r(N_{gd}, \Theta_{gd})$. Then, using the well-known identity

$\Gamma_r(N_{gd}, \Theta_{gd}) = e^{-\Theta_{gd}} \sum_{n=0}^{N_{gd}} \Theta_{gd}^n/n!$, the derivative of the upper incomplete gamma function $\Gamma'_r(N_{gd}, \Theta_{gd}) = -\Theta_{gd}^{N_{gd}-1} e^{-\Theta_{gd}}/(N_{gd} - 1)!$, gives (5.36). The rest follows from Corollary 5.2. □

5.G Proof of Proposition 5.6

Proof. We use $g_{\Theta}(\Theta_{gd})$ and $T_{1,ARQ} = R \Gamma_r(N_{gd}, \Theta_{gd})$ from Prop. 5.5 for $N_{gd} = 2$, solve for $\Theta_{gd}$ from $g_{\Theta}(\Theta_{gd})$, and insert in Corollary 5.2. □
In Chapter 3, we assumed, and modeled, the communication signal as an AWGN channel capacity achieving time-discrete iid complex Gaussian distributed r.v. Under this assumption, in Chapters 4 and 5, we analyzed the throughput, and its optimization, for different (H)ARQ-cases together with the GD-channel.

In this chapter, we note that such an assumption is not realistic in view of widely deployed modulation formats, e.g. (square) quadrature amplitude modulation (QAM). For this reason, we introduce a new, near circular-symmetric, signal constellation denoted Golden Angle Modulation (GAM). The constellation design was originally inspired by a certain (leaf, flower, and seed) packing arrangement, known as spiral phyllotaxis, which we came to recognize as tailored for radial geometric shaping in modulation schemes. The basic idea of GAM is to arrange consecutively indexed constellation points with a phase difference equal to the golden angle, and where successive points have an increasing (or decreasing) magnitude. We propose a number of geometrical-, probabilistic-, and joint geometric-probabilistic GAM-based shaping-designs for which the mutual information (MI) performance approaches the AWGN channel capacity as the constellation size (cardinality) increases. We also explore joint GAM-(H)ARQ designs and confirm that the throughput performance agree well with earlier chapters analytical performance expressions where the per (re)transmission MI is modeled as the instantaneous AWGN channel capacity. Inspired by the GAM design, we also propose an alternative modulation design comprising of cartesian-to-polar transformed modulus-1 rank-1 lattices. Specifically we consider a rank-1 Fibonacci-lattice based design, for which the phase difference between consecutive constellation points asymptotically converges to golden angle, as in GAM, with increasing constellation size.
6.1 Motivation and Outline

A (possible) shortcoming of many information theoretically-based (H)ARQ studies is the assumption of modeling the MI (as used to determine outage events) with the AWGN channel capacity without considering (or at least further motivating) its feasibility wrt practical modulation schemes. Up to now, this has also been the assumption in this thesis. However, such assumption of the MI-model requires a more solid motivation, which is one objective of this chapter. We approach this here, in this chapter, by also considering practical modulation. We are interested in a modulation format with MI-performance approaching the AWGN channel capacity, but we also have an interest in its numerical throughput performance together with (H)ARQ. The two approaches, the MI-model used up to now, and the MI-model used in this chapter, are sketched in Fig. 6.1.

Many modulation formats, signal constellation designs, have been proposed since the founding of digital communication. Examples are PAM, Square- and Rectangular-QAM, Star-QAM [HNKW04], PSK, and APSK [TWD74]. Square-QAM (hereon simply referred to as QAM) is the de-facto-standard in existing wireless communication systems. However, at high SNR, QAM experiences an SNR shaping-loss of ≈ 1.53 dB compared with the AWGN channel capacity [FU98]. This is attributed to the square-shape and the uniform discrete distribution of the QAM-signal constellation points. The MI-performance of PSK is even worse than QAM. Geometric- and probabilistic-shaping techniques have been proposed to overcome the shaping-loss [FU98]. An early work on geometric-shaping is nonuniform-QAM in [BCL94]. Correspondingly, the papers on trellis shaping, [For92], and shell-mapping, [KK93], are early works on probabilistic-shaping. More recent works in this direction are, e.g., [SAGB06, MDSM11, XV13, BISH17]. Nevertheless, for some modulation schemes, e.g. non-uniform QAM (NU-QAM) [SF00, (1)-(2)], and

![Figure 6.1: MI-models for (H)ARQ performance analysis in previous chapters, and in this chapter.](image)
6.2 Golden Angle Modulation

As mentioned earlier, the proposed signal constellation design was inspired by a certain spiral arrangement of, e.g., leaves, seeds, and flower petals, found among many plant species. Such arrangement, or packing, is known as spiral phyllotaxis, see e.g. [Vog79, ABJ97]. The archetypical spiral phyllotaxis packing can, e.g., be seen for the near circular-symmetric packing of sunflower seeds, the near cylindrical-symmetric packing of scales on a cycad cone, or the near spherical-symmetric packing of seeds on a thistle seed head. We observed that this packing offered an interesting shape-versatility, i.e. points could be stretched and shaped into a disc, a cylinder, a sphere, or (more importantly) offer a non-uniform packing in the radial domain. And yet, the angular distribution of the points remained virtually uniform. We came to recognize that this remarkable shape-versatility, with retained uniform angular distribution, and near circular-symmetry, could be useful for geometric shaping of signal constellations when targeting AWGN channel capacity achieving performance. In the following, we examine this bio-inspired modulation framework, as well as some more detailed designs. See also Section for a discussion on the use of SP in other areas.

6.2.1 GAM-framework, Disc-GAM, and Motivation of Framework

Here, in this section, we introduce the overall GAM-framework, give an approximately uniform disc-shaped GAM constellation, and motivate the overall GAM-framework design.
GAM-Framework Design

We start by giving the basic form, the framework, of GAM in Definition 6.1.

**Definition 6.1.** (Golden angle modulation – Basic form) The complex amplitude of the $m$th constellation point, of GAM, has the general form

$$x_m = r_m e^{i2\pi\varphi_m}, \quad m \in \{1, 2, \ldots, M\},$$

where $r_m$ is the radius of constellation point $m$, $r_{m+1} \geq r_m$, and $2\pi\varphi$ is the golden angle in radians, where $\varphi = (3 - \sqrt{5})/2$. The probability of exciting the $m$th point is denoted by $p_m$, which is dependent on index $m$ for probabilistic shaping, or otherwise equiprobable, $p_m = 1/M$.

**Remark 6.1.** Hence, in Definition 6.1, a constellation point, is located (the irrational number) $\varphi \approx 0.381$ turns (or $\approx 137.5^\circ$) relative to the previous constellation point. Letting $x_m = r_m e^{i2\pi\alpha_m}$, equivalent constellation designs results for $\alpha = k \pm (1 + \sqrt{5})/2$, $k \in \mathbb{Z}$ since the (complementary) fractional part is identical to the fractional part of $\varphi$. A special case is $\alpha = (1 + \sqrt{5})/2 \approx 1.618$, the golden ratio. Without limitation, we use the golden angle, $\varphi$, in the following. Note that in a practical implementation, with limited numerical precision, this leads to that $\alpha = \varphi \pm \epsilon$, where $\epsilon$ is assumed sufficiently small. Henceforth, it is assumed that $r_{m+1} \geq r_m$ for an increasing spiral design.

**Remark 6.2.** We note that GAM has a small complex-valued DC-component since, in general, $\frac{1}{M} \sum_{m=1}^{M} x_m \neq 0$. If this is considered a practical problem, this could, e.g., be handled by, negating every second transmitted symbol, or alternatively, the DC-component could be subtracted directly.

In Fig. 6.2, one example of a GAM constellation, (a variant of disc-GAM with a constellation point at (0, 0), see Section 6.2.1), is illustrated. The golden angle phase difference is here illustrated between the third and fourth constellation point. It may be noted that the phases for any two constellation points, $m$ and $m'$, are always non-identical. More precisely, $2\pi m\varphi \mod 1 \neq 2\pi m'\varphi \mod 1$, $\forall m \neq m'$. This can be proven by noting that $m\varphi \mod 1 - m'\varphi \mod 1 = (m - m')\varphi \mod 1 \neq 0$, and that $(m - m')\varphi$ has a non-zero fractional part since $\varphi$ is irrational. In spiral phyllotaxis, what gives the impression of, spiraling spokes can be noted. Such spiral spokes are called parastichies in phyllotaxis taxonomy. Three left-winding spiral spokes are shown in Fig. 6.2, and five right-winding (opposing) spiral spokes can, at closer scrutiny, also be identified. Note that we are often interested in approximately circular-symmetric constellation shapes. For this reason, GAM requires sufficiently large constellation sizes, say with at least $M \geq 16$.

**Remark 6.3.** The mathematical design in Definition 6.1 is inspired from the work by Vogel, [Vog79], who described an idealized growth pattern for the sunflower seeds, $x_m \propto \sqrt{m}e^{i2\pi\varphi_m}$ (in our notation). A key aspect, enabling the approximation of,
6.2. Golden Angle Modulation

-1.5 -1 -0.5 0 0.5 1 1.5
-1.5
-1
-0.5
0
0.5
1
1.5
-1.5
-1
-0.5
0
0.5
1
1.5

Figure 6.2: Example of a GAM signal constellation (Disc-GAM), where $M = 16$, $r_m = c_{\text{disc}}\sqrt{m-1}$, $m \in \{1, 2, \ldots, 16\}$, and $c_{\text{disc}} = \sqrt{2\bar{P}/(M - 1)}$.

e.g., a complex Gaussian pdf, is to not restrict the radial function $r_m$ to Vogel’s (main) radial function

\[ r_m \propto \sqrt{m}, \quad m \in \mathbb{R}, \]

and further to allow for general non-identical probabilities $p_m$, and average power scaling. In addition, our design differs in terms of application and performance objective(s).

**Disc-GAM**

Before exploring GAM with shaping, we first introduce the disc-GAM constellation. The objective here is to design a constellation with discrete probabilities $p_m = 1/M, \forall m$, that approximates a continuous uniform distribution constrained to a disc. This will then be the baseline design for the probabilistic- and radial geometric-shaping designs. The magnitudes $r_m$ are, as discussed in Appendix 6.A, determined by means of inverse sampling. Briefly, in inverse sampling, one equates a desired cdf $F(\tilde{r})$ with a continuous uniformly distributed r.v., say $U \in (0, 1)$, and solve for $\tilde{r}$. Generating $U$ uniformly on $(0, 1)$, then gives $\tilde{r}$ with pdf $f(\tilde{r})$. For large number of constellation points, we can approximate the continuous uniformly

\[ 1 \text{Vogel also discussed two other radial functions, } r_m \propto m^\delta, \delta \in \mathbb{R}, \text{ and } r_m \propto e^{m^\delta}, \delta \in \mathbb{R}. \]
distributed r.v. $U$ with a uniformly distributed discrete r.v. $U_M$ with probabilities $1/M$ at $(m - 1/2)/M$, $m \in \{1, 2, \ldots, M\}$, or alternatively at $(m - 1)/M$. Thus, we have $F(\tilde{r}_m) = (m - 1/2)/M$, from which we solve for $\tilde{r}_m$ and then normalize to unit average power. The approximately uniform disc-shaped design, with approximately same-sized Voronoi-cells, is analogous to classical square-shaped QAM. Now, the disc-GAM constellation is given in the following definition.

**Theorem 6.1.** (Disc-GAM) The disc-GAM format, with average power $\bar{P}$, $p_m = 1/M$, and a lower offset $M_l \geq -1/2$, is characterized by

$$r_m = c_{\text{disc}} \sqrt{m - 1/2 + M_l}, \ m \in \{1, 2, \ldots, M\}, \ M_l \geq -1/2,$$

$$c_{\text{disc}} \triangleq \sqrt{\frac{2\bar{P}}{M + 2M_l}}. \ (6.3)$$

**Proof.** The proof is given in Appendix 6.A. □

In the theorem above, to enable control of PAPR, and for generality, we have allowed for a lower offset $M_l$. While this yields a circular whole in the disc, a millstone-shaped appearance, the disc-GAM nomenclature is kept for simplicity. When $M_l = 0$, the expression for $r_m$ simplifies as follows.

**Proposition 6.1.** The disc-GAM format, with average power $\bar{P}$ and a lower offset $M_l = 0$ has the magnitudes

$$r_m = \sqrt{\frac{2\bar{P}(m - 1/2)}{M}}, \ m \in \{1, 2, \ldots, M\}. \ (6.4)$$

**Proof.** This is seen directly in Theorem 6.1. □

In Fig. 6.3, we illustrate the disc-GAM constellation for $M = 2^{10}$, $M_l = 0$. In the following remarks, some basic (easily derived or known) facts of disc-GAM, in comparison to QAM, are compiled.

**Remark 6.4.** The entropy is $H_{\text{disc}} = \log_2(M)$.

**Remark 6.5.** The PAPR is defined as $\text{PAPR} \triangleq \frac{r_m^2}{\sum_{m=1}^{M} p_m r_m^2}$. The PAPR for disc-GAM is then $\text{PAPR}_{\text{disc}} = (2M - 1 + 2M_l)/(M + 2M_l)$. When $M \to \infty$, $\text{PAPR}_{\text{disc}} \simeq 2 \ (\simeq 3\,\text{dB})$. QAM, in contrast, has a corresponding asymptotical PAPR of $\text{PAPR}_{\text{QAM}} = 4.8 \,\text{dB}$.

**Remark 6.6.** Letting the number of constellation points $M$ grow towards infinity, and assuming (approximately) the same constellation point density, QAM asymptotically requires $10\log_{10}(\pi/3) \ (\approx 0.2 \,\text{dB})$ higher average power relative to disc-QAM.

**Remark 6.7.** Letting the number of constellation points $M$ grow towards infinity, and assuming (approximately) the same constellation point density, QAM asymptotically requires $10\log_{10}(\pi/2) \ (\approx 1.96 \,\text{dB})$ higher peak power relative to disc-GAM.
Motivation for the GAM-framework

In the following, we aim to further motivate the design choices of GAM, in particular the phase factor $e^{i2\pi\varphi m}$. First, we focus on examples and intuition, and then, we outline some more mathematically inclined arguments. We start by discussing spiral-based constellation designs in general, and then GAM with the golden angle-based phase-factor in particular. We then consider alternative signal constellations that have, or can be truncated to, (approximately) circular-symmetric outer boundaries. Note that classical uniform square-shaped QAM has already been considered in Section 6.2.1, and is therefore not discussed further here.

A first question is regarding the form of the GAM expression in Definition 6.1. We have settled for the model $x_m = r_m e^{i2\pi\varphi m}$, where $r_m$ can be adapted independently of the phase factor $e^{i2\pi\varphi m}$, and where the phase argument, $2\pi\varphi m$, is linear in $m$ and uses the golden angle $\varphi$ as scaling factor. The proposed expression is generic for any $M$, simple enough for implementation, and (as will be seen) yields good MI-performance. In terms of alternative spiral designs, this could be contrasted against a more general spiral-based constellation, $x_m = r_m e^{i2\pi\alpha m}$, where $\alpha_m$ is an unspecified, possibly non-linear, phase function that depends on $m$ and $M$. With this choice, both $r_m$ and $\alpha_m$ can be tuned independently for performance optimization. However, it is non-trivial to jointly determine the optimal $r_m$ and $\alpha_m$. 

Figure 6.3: Disc-GAM signal constellation with $M = 2^{10}$ and $M_1 = 0$. 

![Diagram of Disc-GAM signal constellation]
Figure 6.4: Signal constellations for rational, and irrational, $\alpha$, where $x_m \propto \sqrt{m} e^{i2\pi \alpha m}$, $m \in \{1, 2, \ldots, 32\}$, with $\alpha = \{1/3, 1/4, \pi, e, \sqrt{2}, \sqrt{3}\}$.

and the possible performance benefit over a simpler linear phase argument $2\pi \varphi_m$ is questionable. Nevertheless, a reasonable more general form to the GAM-design is $x_m = r_m e^{i2\pi \alpha m}$, where an arbitrary $\alpha$, is allowed for.

Now, given a model $x_m = r_m e^{i2\pi \alpha m}$, $0 < \alpha \mod 1 < 1$, what is a “good” choice of $\alpha$? This may depend on the considered channel, performance measure, constellation size $M$ in question. Hence, there may not exist a unique single optimal $\alpha$ for all cases. However, for a circular-symmetric complex AWGN channel, and performance measures, such as MI or symbol error rate (SER), it is (intuitively) desirable to target a circular-symmetric design, trying to maximize the minimum Euclidean distances between the constellation points, and ensuring that constellation points are uniformly distributed in the angular (phase) domain. It is noted, by inspecting trivial counter examples, that certain values of $\alpha$ do not offer such design. In Fig. 6.4, we illustrate spiral-based constellations for different values of $\alpha$. First, we consider the choice of a rational $\alpha \in \mathbb{Q}$, e.g. $\alpha = \{1/3, 1/4\}$, with (disc-GAM’s) $r_m \propto \sqrt{m}$. For those cases, 3 and 4 distinct ‘spokes’ results. More generally, a rational $\alpha = p/q$, where $p \in \mathbb{N}$ and $q \in \mathbb{N}$ are coprime, and $M \gg q$, generates $q$ radially extending ‘spokes’. It is evident that for $\alpha = p/q$ and $M \gg q$, the
6.2. Golden Angle Modulation

Figure 6.5: Signal constellations, where $x_m \propto \sqrt{m} e^{2\pi i m}$, $m \in \{1, 2, \ldots, 32\}$, with $\alpha = \{0.98\varphi, \varphi, 1.04\varphi\}$.

constellation points do not distribute as uniform on a disc as desired. For this reason, it is then natural to consider an irrational $\alpha \in \mathbb{I}$. In Fig. 6.4, we plot the signal constellations for $\alpha = \{\pi, e\}$, or rather their fractional parts. Also here, similar, but slightly curved, spokes develop, with poor minimal constellation point distances. This is due to the fractional parts of $\pi, e$, which are relatively well approximated by rational number $1/7$, and $5/7$, respectively. Thus, due to the numerator $q = 7$, 7 curved spokes results. However, as seen in Fig. 6.4 where signal constellations for the irrational numbers $\alpha = \{\sqrt{2}, \sqrt{3}\}$ are plotted, the disk is filled approximately uniformly.

Another question is to which precision a practical implementation of $\alpha = \varphi$ needs to be? As can be seen in Fig. 6.5, very minor phase deviations from the $\alpha = \varphi$ destroys the, relatively, approximately uniform packing of constellations points. As a rule of thumb, based on inspecting the CDF of the minimum Euclidean distances between constellation points, it is suggested that, say, $\alpha = \varphi \pm \epsilon\varphi$, $\epsilon\varphi \approx 10^{-4}$.

Yet another question is if the choice of $\alpha$ has any impact on constellation shape, and performance, when considering (radial) geometric shaping? In Fig. 6.6, we illustrate two geometric shaping cases, one with consecutive points separated with $\alpha = \varphi$, and another one with $\alpha = 0.15\varphi \approx 0.0573$ (or $\approx 0.9427$) turns. It is observed that a more circular-symmetric uniform constellation design results for $\alpha = \varphi$ compared to when $\alpha$ is relatively close to an integer, say 0 or 1. From the above, examples and intuition tell us that rational $\alpha$, $\alpha = p/q$, at least for small coprime $p$ and $q$, when $M \gg q$, are generally undesirable, certain irrational $\alpha$ are undesirable, to small, or to large, fractional (ir)rational $\alpha$ are undesirable, whereas other certain (medium-sized) irrational $\alpha$ appear desirable. The question is which irrational $\alpha$ (or $\alpha$s) is (are) the best choice(s)?

The choice of an irrational $\alpha$, and specifically the choice $\alpha = \varphi$, can be argued more formally for by considering the notion of uniformly distributed modulo 1 sequences. As we have seen above, it is generally desirable to have constellation points uniformly distributed in phase. Below, we review some useful results, mainly
from [KN74], for uniformly distributed sequences. We will subsequently also use this to show that we GB-GAM approaches the channel capacity as the constellation size (the cardinality) increases. We first consider the definition for a sequence which is uniformly distributed modulo 1, abbreviated u.d. mod 1. For generality, we consider the multidimensional case in the following. Let $d \in \mathbb{N}$ be the dimension, and let $a = (a_1, a_2, \ldots, a_d)$ and $b = (b_1, b_2, \ldots, b_d)$ be two vectors with real components; $a, b \in \mathbb{R}^d$. The set of points $a \leq z < b$ is denoted $[a, b)$. The function $A([a, b); N)$ denotes the number of points in a sequence $z_n, 1 \leq n \leq N$ that lie in $[a, b)$.

**Definition 6.2.** (u.d. mod 1, based on [KN74, Def. 6.1]) The sequence $(z_n), n = \{1, 2, \ldots\}$, is said to be u.d. mod 1 in $\mathbb{R}^d$ if

$$
\lim_{N \to \infty} \frac{A([a, b); N)}{N} = \prod_{j=1}^{d} (b_j - a_j)
$$

(6.5)

for intervals $[a, b) \in [0, 1)^d$, where $[0, 1)^d$ denotes the unit $d$-dimensional cube.

A criterion to test whether a sequence is u.d. mod 1 is given by Weyl’s theorem.

**Theorem 6.2.** (Weyl’s criterion, [KN74, Thm. 6.2]) A sequence $(z_n), n \in \{1, 2, \ldots, N\}$, is u.d. mod 1 in $\mathbb{R}^d$ iff every lattice point $h \in \mathbb{Z}^d, h \neq 0$,

$$
\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} e^{i2\pi < h, z_n>} = 0.
$$

(6.6)

**Proof.** The proof is given in [KN74].

If the sum of a function of a $d$-dimensional sequence is u.d. mod 1, the summation may be replaced with an integration over the unit $d$-dimensional cube.
Theorem 6.3. (Kinchin Theorem, based on [KN74, Thm. 6.1]) A sequence \((z_n), n = \{1, 2, \ldots N\}\), is u.d. mod 1 in \(\mathbb{R}^d\) iff for every complex-valued function \(f\) on \([0, 1)^d\), the following relation holds:
\[
\lim_{N \to \infty} \frac{1}{N} \sum_{k=1}^{N} f(z_n \mod 1) = \int_{[0,1)^d} f(z) \, dz.
\] (6.7)

Proof. The proof is given in [KN74].

Theorem 6.4. (Equidistribution theorem, based on [KN74, Thm. 6.1]) \(z_n = na \mod 1, n = \{1, 2, \ldots N\}\), is uniformly distributed on the circle \(\mathbb{R}/\mathbb{Z}\) when \(a\) is an irrational number.

Proof. Given directly from Weyl's criterion, Theorem 6.2.

Corollary 6.1. The sequence \((a_m, b_m), m \in \{1, 2, \ldots M\}\), where \(a_m = m\varphi \mod 1\), \(b_m = (m - 1/2)/M\), is 2D u.d. mod 1

Proof. The proof is given in Appendix 6.B.

Corollary 6.1 ensures, e.g., that Kinchin's theorem may be used when studying performance limits of GAM. We illustrate the 2D-sequence \((m\varphi, b_m)\) for GAM in Fig. 6.7. It is seen that the 2D point set uniformly covers the unit-square.

Since \(\varphi\) is irrational, the phase values of the GAM constellation points are u.d. mod 1. Now, let GAM's phase argument as \(2\pi a_m\), where \(a_m = m\varphi\), or \(a_m = m\varphi \mod 1\), and let the magnitude \(r_m\) be a mapping from \(b_m = (m - 1/2)/M\), \(m \in \{1, 2, \ldots M\}\). In the corollary below, based on Weyl's criterion, we show that the 2D-sequence \((a_m, b_m)\) for GAM is 2D u.d. mod 1.

Corollary 6.1. The sequence \((a_m, b_m), m \in \{1, 2, \ldots M\}\), where \(a_m = m\varphi \mod 1\), \(b_m = (m - 1/2)/M\), is 2D u.d. mod 1

Proof. The proof is given in Appendix 6.B.

It turns out that the golden ratio, golden angle, or any other irrational number with an identical continued fraction (CF) tail, is the worst case in Hurwitz's theorem in the sense that the bound becomes increasingly tighter for increasing \(q\). Independently, a similar argument has also been given in [MS01a]. Thus, selecting \(\alpha\) as the golden ratio (or the related golden angle) has the least tendency to develop...
Figure 6.7: 2D-sequence \((m\varphi \mod 1, (m - 1/2)/M), m \in \{1, 2, \ldots M\}, M = 64\).

Remark 6.8. However, if the golden ratio (and all other irrational numbers with the same CF tail) is excluded, the bound in Hurwitz’s theorem becomes \(|a - p/q| < 1/\sqrt{8q^2}\). This bound is met for \(a = \sqrt{2}\). Repeating the same procedure, if \(a = \sqrt{2}\) is excluded instead, the following bound results \(|a - p/q| < 5/\sqrt{21q^2}\), which is met for \(a = \sqrt{5}\).

Hence, it is also of interest to study the performance of GAM by testing other irrational numbers than \(\alpha = \varphi\). In the numerical MI-performance evaluation, we find that using the golden angle performs better than any other tested irrational number.

Now, not limiting the discussion to a spiral-based design, GAM can also be compared with alternative constellation designs, specifically to square- and hexagonal-lattices with an approximately circular-symmetric outer boundary, as well as APSK. At closer scrutiny, it can be noted that the approximately circular-symmetric outer boundary for square- and hexagonal-lattices do not retain the circular-symmetry under radial geometric shaping. Moreover, for APSK, a radial geometric transform targeting a complex Gaussian-like distribution, gives a sparse angular density for
the outermost points in one variant of APSK, APSK-a, but dense in another variant, APSK-b. This is discussed, and illustrated, in more detail Appendix 6.C. We return to the APSK-based constellations in the section for numerical evaluation.

6.2.2 Geometric Shaping

In this section, we explore GAM with geometric shaping. The objective is to determine the magnitudes \( r_m \) such that the MI-performance measure tends to be "as large as" possible.

Geometric Bell-shaped GAM

A first idea, and studied below, is to let the equiprobable GAM constellation points approximate a continuous complex Gaussian pdf. This is motivated since the continuous complex Gaussian input distribution is known to achieve AWGN channel capacity. Again, we use the discretized version of the inverse sampling method to determine the magnitudes \( r_m \), see Appendix 6.D. Alternatively, the design can also be derived based on results from high-rate quantization theory, which is also discussed in Appendix 6.E. The design, assuming a pmf with support at \( m \in \{1, 2, \ldots, M\} \) for the inverse sampling method, is given by the following theorem.

**Theorem 6.6.** (Geometric-bell-GAM) Let \( M \) be the number of constellation points, each point be equiprobable and indexed as \( m \in \{1, 2, \ldots, M\} \), and \( \bar{P} \) be the average power constraint. Then, the magnitude of the \( m \)th constellation point for the geometric-bell-GAM (GB-GAM) is

\[
r_m = c_{gb}\sqrt{\frac{\ln \left(\frac{M}{M-m+1}\right)}{M}},
\]

\[ p_m = \frac{1}{M}, \text{ where } \]

\[
c_{gb} \triangleq \frac{M\bar{P}}{M\ln M - \ln(M!)}.
\]

**Proof.** The proof is given in Appendix 6.D for the inverse sampling method, and complementary also for the high-rate quantization approach in Appendix 6.E.

We show the geometric-bell-GAM signal constellation in Fig. 6.8, and note that it is densest at its center, i.e. where the pdf for the complex Gaussian r.v. peaks.

**Remark 6.9.** When \( M \to \infty \), since \( \lim_{M \to \infty} \ln \left(\frac{M^M}{(M!)}\right)/M = 1 \), we get \( r_m \simeq \sqrt{\bar{P}\ln \left(\frac{M}{(M-m+1)}\right)} \).

**Remark 6.10.** The entropy is \( H_{gb} = \log_2(M) \).
Remark 6.11. The PAPR is \( \text{PAPR}_{gb} = \ln(M)\bar{P}/(M \ln(M) - \ln(M!)) \approx \bar{P} \ln(M) \), which tends to infinity with \( M \). This is expected as the PAPR of a Gaussian distributed r.v. is infinite.

Remark 6.12. If we instead use \( r_m = c_{gb} \sqrt{\ln(M/(M - m + 1/2))} \), i.e. the pmf used with the inverse sampling method has support at \( \{1/2, 3/2, \ldots, M - 1/2\} \), we show in Appendix 6.D that \( c_{gb} \equiv \sqrt{M \bar{P}/(M \ln M - \ln(\Gamma(M + 1/2)/\sqrt{\pi}))} \).

In the next theorem, we now show that GB-GAM is capacity achieving. In fact, this theorem holds for any spiral-based modulation where \( \sqrt{-\log(1 - a_m) e^{i2\pi b_m}} \), when \( (a_m, b_m) \) are 2D u.d. mod 1 on the unit square \( (0, 1]^2 \).

Theorem 6.7. (GB-GAM is capacity achieving) Let \( X \sim \mathcal{CN}(0, 1) \), and \( Y \sim \mathcal{CN}(0, 1 + 1/S) \), where \( S \) is the SNR, such that \( I(Y; X|S) = \ln(1 + S) = C(S) \) [b/Hz/s], and \( C(S) \) is the complex AWGN channel capacity. Further let \( X_M \), with \( \mathbb{E}\{|X_M|^2\} = 1 \), be the input distribution defined by the \( M \) equiprobable GB-GAM constellation points, Theorem 6.6, and \( Y_M = X_M + W \), be the output distribution with noise \( W \sim \mathcal{CN}(0, 1/S) \). Then, the gap between the complex AWGN channel capacity \( I(Y; X|S) \) and the MI \( I(Y_M; X_M|S) \) of GB-GAM vanishes with increasing...
constellation size $M$, i.e.

$$\lim_{{M \to \infty}} I(Y; X|S) - I(Y_M; X_M|S) = 0. \quad (6.11)$$

**Proof.** The proof is given in Appendix 6.F.

---

**Truncated Geometric Bell-shaped GAM**

It can be recognized that GB-GAM and disc-GAM can be seen as resulting from two extremes of a truncated Gaussian input distribution. GB-GAM would then approximate a (non-truncated) Gaussian pdf, and disc-GAM would approximate a truncated Gaussian pdf when the radius approaches zero. We can approach this line of reasoning more formally. The overall procedure is to first find a continuous input distribution that (closely) approximates the MI-maximizing distribution, and then use this distribution together with the GAM framework and inverse sampling to determine constellation point magnitudes $r_m$. First, the MI between a continuous input signal $X$, and a continuous output signal $Y$, can be lower bounded as follows,

$$I(Y; X) = h(Y) - h(Y|X) = h(Y) - h(W) \geq h(X) - h(W),$$

where the fact that conditioning reduces entropy, $h(Y) \geq h(Y|W) = h(X)$, is used in the last step. Hence, instead of finding an input distribution for $X$ that maximize the entropy $h(Y)$, the entropy $h(X)$ is maximized, thereby giving a lower bound to $I(Y; X)$.

Since GAM is near circular symmetric, i.e. for a sufficiently large $M$, we target a continuous circular symmetric input distribution $f(u, v)$ that maximizes $h(X)$ for an average power constraint. For notational simplicity and clearness below, we use the notation $f(u, v)$ instead of $f(x_{re}, x_{im})$. Moreover, for generality, and control over the PAPR, we let the circular symmetric density $f(u, v)$ be nonzero for magnitudes less than an outer radius, $\tilde{r}_o$, but also greater than an inner radius, $\tilde{r}_i$. More precisely, the pdf is assumed to fulfill

$$f(u, v) \geq 0, \quad (u, v) \in A,$$

$$f(u, v) = 0, \quad \text{Otherwise}, \quad (6.12)$$

where $A \in \{u, v : \tilde{r}_i^2 \leq u^2 + v^2 \leq \tilde{r}_o^2\}$.

The optimization problem can now be formulated as

$$\maximize_{f(u, v)} - \int \int_A f(u, v) \ln(f(u, v)) \, du \, dv,$$

subject to

$$\int \int_A (u^2 + v^2) f(u, v) \, du \, dv = 1,$$

$$\int \int_A f(u, v) \, du \, dv = 1. \quad (6.13)$$

This problem can, due to convexity of the entropy, be solved by classical Lagrangian optimization. The solution is straightforward and is simply a truncated 2-dim Gaussian pdf. Writing it on the standard bi-variate Gaussian form, with an
undetermined normalization constant $c_1$, we have $f(u, v) = \frac{c_1}{\pi \sigma^2} e^{-\frac{(u^2+v^2)}{\sigma^2}}$, where $(u, v) \in A$. Since $\tilde{r}_i$ and $\tilde{r}_o$ can be tuned to any value, without loss of generality, we let $\sigma^2 = 1$. As the discrete GAM magnitudes $r_m$ are sought after, it useful to transform $f(u, v)$ to a marginal density $f(\tilde{r})$ with respect to a magnitude $\tilde{r}$. Via Euclidean-to-polar coordinate variable substitution, $u = \tilde{r} \sin(\phi)$ and $v = \tilde{r} \cos(\phi)$, and integrating over a uniform distribution in phase, we get

$$f(\tilde{r}) = \begin{cases} \frac{c_1}{\pi} e^{-\tilde{r}^2} 2\pi \tilde{r}, & \tilde{r}_i \leq \tilde{r} \leq \tilde{r}_o, \\ 0, & \text{Otherwise.} \end{cases}$$  \hspace{1cm} (6.14)$$

Integrating the pdf, $\int_{\tilde{r}_i}^{\tilde{r}_o} f(r) \, dr = 1$, yields the constant $c_1 = 1/(e^{-\tilde{r}_i^2} - e^{-\tilde{r}_o^2})$.

The corresponding cdf is then

$$F(\tilde{r}) = \begin{cases} 0, & 0 \leq \tilde{r} \leq \tilde{r}_i, \\ \frac{e^{-\tilde{r}_i^2} - e^{-\tilde{r}^2}}{e^{-\tilde{r}_i^2} - e^{-\tilde{r}_o^2}}, & \tilde{r}_i \leq \tilde{r} \leq \tilde{r}_o, \\ 1, & \tilde{r}_o \leq \tilde{r} < \infty. \end{cases}$$  \hspace{1cm} (6.15)$$

As seen, this scheme relates to a truncated Gaussian input distribution for the AWGN channel capacity. A number of prior works exist. In [Smi71], the channel capacity for a scalar AWGN channel, with a peak-power limited input distribution, was studied. It was found that the optimal distribution had to be discrete. This channel was recently revisited in [HY16]. The problem of the peak-power constrained optimal input distribution for the complex channel was studied in [SBD95]. Here, it was found that the input distribution had to be discrete in the radial domain, but uniformly continuously distributed in the phase domain. As a sub-optimal input distribution, they derived and considered (among other things) a peak-power (truncated) complex Gaussian distribution based on the same line of reasoning as above. However, no actual modulation scheme was proposed. Moreover, the result on a uniformly continuous distribution in the phase domain is incompatible with the notion of discrete signal constellation points. Even if a modulation scheme had been proposed, say APSK-based (which was conceivable at the time), it is not obvious how to combine APSK with the truncated complex Gaussian input distribution assumption in a simple and well-performing manner. The designer of an APSK scheme must carefully split the total number of constellation points among the rings, and arrange the distance among rings. The constellations must be designed, and optimized, for every distinct $M$. Certain values of $M$, e.g. prime numbered $M$, give APSK constellations with asymmetries. GAM, on the other hand, together with the method of inverse sampling, is versatile enough to seamlessly adapt to a desired truncated complex Gaussian input distribution and any integer $M$. To more flexibly control PAPR, we have allowed for $0 \leq \tilde{r}_i < \tilde{r}_o$.

As for previous GAM constellation designs, we use the inverse sampling method.

The magnitudes (without power normalization) are then

$$\tilde{r}_m = \sqrt{-\ln \left( e^{-\tilde{r}_i^2} - \frac{m}{M} \left( e^{-\tilde{r}_i^2} - e^{-\tilde{r}_o^2} \right) \right)}, \quad m \in \{1, 2, \ldots, M\}. \hspace{1cm} (6.16)$$
Letting \( r_m = c_{tgb} \tilde{r}_m \) denote the power-normalized magnitude, where \( \bar{P} \) is the average power constraint and \( c_{tgb} = \sqrt{MP/\sum_{m=1}^{M} \tilde{r}_m^2} \), we get

\[
r_m = \frac{\bar{P} \ln \left( e^{-\tilde{r}_i^2} - \frac{m}{M} \left( e^{-\tilde{r}_i^2} - e^{-\tilde{r}_o^2} \right) \right)}{\frac{1}{M} \sum_{m'=1}^{M} \ln \left( e^{-\tilde{r}_i^2} - \frac{m'}{M} \left( e^{-\tilde{r}_i^2} - e^{-\tilde{r}_o^2} \right) \right)}, \quad m \in \{1, 2, \ldots M\}. \tag{6.17}
\]

Note that since the constellation is unit-power normalized, the resulting PAPR is simply \( \text{PAPR} = r_M \).

As \( r_m \) depends only on two parameters \( \tilde{r}_i \) and \( \tilde{r}_o \), a simplified optimization problem can be formulated compared to optimizing all \( M \) magnitude values independently. However, the performance is reduced somewhat. The optimization problem, including an optional PAPR constraint \( \text{PAPR}_0 \), can be written

\[
\begin{align*}
\text{maximize} & \quad I(Y; X), \\
\text{subject to} & \quad \tilde{r}_i \geq 0, \\
& \quad \tilde{r}_o \geq \tilde{r}_i, \\
& \quad r_M^2 \leq \text{PAPR}_0.
\end{align*} \tag{6.18}
\]

This problem is hard to solve analytically, but, advantageously, large-sized constellations can easily be handled with numerical optimization software.

In Fig. 6.9(a)-(c), we plot the signal constellations of optimized TGB-GAM (6.17)+(6.18), for \( \tilde{r}_i = 0 \), \( M = 256 \), and \( S = \{10, 22.5, 35\} \) dB. We note that the signal constellation change in shape from an approximate discrete complex Gaussian-like design, over an intermediate design, to a disc-shaped design. In Fig. 6.10(a)-(c), we plot the magnitude distributions for the TGB-GAM together with GB-GAM and disc-GAM for the same SNRs as in Fig. 6.9(a)-(c). We observe how the signal constellation magnitudes have nearly the same distribution as GB-GAM for low SNRs, and nearly the same distribution as disc-GAM for high SNR. I.e., the constellation approximates a complex Gaussian r.v. at low SNR, and a uniform disc r.v. at high SNR.

To simplify the expression for \( r_m \) further, we may chose an approximate form that converges to GB- and disc-GAM for \( S = 0 \) and \( S = \infty \), respectively. First let \( \tilde{r}_i = 0 \). Inspecting (6.17), and knowing that \( \tilde{r}_o \) diminishes with increasing SNR, we propose the simple, yet well-performing, form

\[
r_m = \sqrt{\frac{\ln \left( 1 - \frac{m}{S + 37} \right)}{\frac{1}{M} \sum_{m'=1}^{M} \ln \left( 1 - \frac{m'}{S + 37} \right) \left( e^{-\tilde{r}_i^2} - e^{-\tilde{r}_o^2} \right) \left( e^{-\tilde{r}_i^2} - e^{-\tilde{r}_o^2} \right)}}, \tag{6.19}
\]

It is easy to see that when \( S \to 0 \), (6.19) converges to the GB-GAM. By Taylor-expanding (6.19) for \( S \to \infty \), it can be seen that (6.19) converges to the disc-GAM.
While MI-performance is the prime performance metric of interest, GB-GAM does not specifically aim to maximize the MI, but rather approximate a complex Gaussian r.v., with a finite number of constellation points $M$. We also note that the larger magnitudes for GB-GAM appears "unnecessarily large" for moderate SNRs. For those reasons, we now explore an MI-optimization formulation for GAM.

In this first method, we let $p_m = 1/M$, and vary $r_m$ in order to maximize the MI, for a desired SNR $S$. The optimized signal constellation points are $x_m^* = r_m e^{j2\pi \phi_m}$. In the following optimization problems, it is convenient to normalize the transmit power to unit-power, and let the receiver noise be the inverse of the SNR, $\sigma_w^2 = 1/S$. More formally, allowing for a complex valued output r.v. $Y$, and a complex valued
6.2. Golden Angle Modulation

(discrete modulation) input r.v. $X$, the optimization problem is

$$
\text{maximize } \quad I(Y; X) \\
\text{subject to } \quad r_{m+1} \geq r_m, \ m = \{1, 2, \ldots M\}, \\
r_1 \geq 0, \\
\sum_{m=1}^{M} p_m r_m^2 = 1, \\
$$

(6.20)

with $p_m = 1/M$, and $\sigma_w^2 = 1/S$.

**Remark 6.13.** In some cases, a PAPR-criterion is of interest. For this reason, we simply add a PAPR inequality constraint, $\text{PAPR} \leq \text{PAPR}_0$, to the above optimization problem, where $\text{PAPR}_0$ is the target PAPR. This can be expressed as

$$
\left( \frac{1}{\text{PAPR}_0} - p_M \right) r_M^2 - \sum_{m=1}^{M-1} p_m r_m^2 \leq 0. \\
$$

(6.21)

When the PAPR inequality constraints is included, the constellation design thin out in the center, and concentrate as a ring. Note that this PAPR inequality constraint could be included in any given GAM-optimization formulation.

In the above, the MI can be expressed in terms of the differential entropies $h(Y)$, and conditional differential entropy $h(Y|X)$, as $I(Y; X) = h(Y) - h(Y|X) = h(Y) - h(W)$, where (for the AWGN channel) $h(W) = \ln(\pi e \sigma_w^2)$, and $h(Y) = -\int_C f_Y \ln(f_Y) \, dy$. The pdf for the output r.v. $Y$ is then

$$
f_Y = \sum_{m=1}^{M} f(y|x_m) p_m = \frac{1}{\pi \sigma_w^2} \sum_{m=1}^{M} p_m e^{-\frac{|y-x_m|^2}{\sigma_w^2}}, \\
$$

(6.22)

where $x_m = r_m e^{2\pi\varphi_m}$. Note that the integral for $h(Y)$ is over the complex domain.

As this problem involves inequality constraints, the corresponding Karush-Kuhn-Tucker conditions need to be fulfilled. However, even just by considering the partial derivatives of the non-linear objective MI-function, involving the pdf (6.22), it is seen that an analytical can not be found. Hence, in Section 6.5, we resort to a numerical optimization solver\(^2\) for $M = 16$ (due to computational complexity with greater $M$) to illustrate the improvement over GB-GAM.

A problem with the optimization formulation-G1 is that $M$ undetermined variables $r_m$ need to be solved for. The numerical optimization of $M$ undetermined variables $r_m$ is computationally intensive and time consuming. As an alternative to formulation-G1, and in addition to the truncated GB-GAM case in Section 6.2.2, we

\(^2\)This optimization problem can, e.g., be solved with *fmincon.m* in [Matlab]\(^6\). A suitable initialization for the optimization solver is disc-GAM with magnitudes $r_m = c_{\text{disc}} \sqrt{m - 1/2}$. 
have also proposed a reduced-complexity approach in [Lar17], denoted formulation-G2, where we let a (polynomial) function describes a spiral, and then numerically optimize the relatively few parameters to maximize the MI. Very large constellations can then be handled, but the resulting optimal MI is, due to the constrained form of the (polynomial) function, lower than for G1-formulation.

6.2.3 Probabilistic Shaping

We now turn the attention to probabilistic shaping. Here, we assume the disc-GAM signal constellation design with \( r_m = c_{pb} \sqrt{m} \) and assign \( p_m \) such that the pmf, in some sense, tends to approximate a bell-shaped complex Gaussian pdf. We denote such scheme probabilistically shaped bell-GAM (PB-GAM). We analytically develop an PB-GAM signal constellation which is SNR-independent, and also formulate MI-optimized SNR-dependent variants.

GAM with Minimum SNR and Entropy-constraint

Our first take on this is to optimize \( p_m \) such that an entropy constraint, \( H_{pbse} \), is fulfilled, while minimizing the average transmit power (or equivalently the SNR). Hence, the optimization problem is

\[
\begin{align*}
\text{minimize} & \quad \sum_{m=1}^{M} p_m r_m^2, \quad m = \{1, 2, \ldots M\} \\
\text{subject to} & \quad - \sum_{m=1}^{M} p_m \log_2(p_m) = H_{pbse}, \\
& \quad \sum_{m=1}^{M} p_m = 1,
\end{align*}
\]

(6.23)

where we let \( r_m = c_{pb} \sqrt{m} \). Using Lagrangian optimization, the solution to the optimization problem is given by the following theorem. The solution to this problem is known in general as the Maxwell-Boltzmann distribution [KP93]. As seen below, disc-GAM with \( r_m \propto \sqrt{m}, m \in \{1, 2, \ldots M\} \), offers a particularly neat design with a near ideal discrete circular-symmetric complex-Gaussian constellation having geometrically distributed probabilities \( p_m \).

**Theorem 6.8.** (Probabilistic-bell-GAM with minimum SNR and an entropy constraint) Let \( M \) be the number of constellation points, \( p_m \) is the probability that constellation point \( m \in \{1, 2, \ldots M\} \) is used, \( H_{pbse} \) is the entropy constraint, and \( \bar{P} \) is average power. Then, PG-GAM is characterized by

\[
\begin{align*}
r_m &= c_{pbse} \sqrt{m}, \ m \in \{1, 2, \ldots M\}, \\
c_{pbse} &\triangleq \sqrt{\bar{P} \left( \frac{1}{1-\zeta} - \frac{M \zeta^M}{1-\zeta^M} \right)^{-1}},
\end{align*}
\]

(6.24) (6.25)
Figure 6.11: Minimum SNR and entropy-constrained PB-GAM illustrated in a 3D-view, $M = 2^{10}$.

\begin{align*}
p_m &= \frac{1 - \zeta}{1 - \zeta^M} \zeta^{m-1}, \quad (6.26) \\
H_{\text{pbse}} &= -\ln \left( \frac{1 - \zeta}{1 - \zeta^M} \right) + \left( \frac{\zeta}{1 - \zeta} + \frac{M\zeta^M}{1 - \zeta^M} \right) \ln (\zeta), \quad (6.27)
\end{align*}

where $\zeta$ is a constant determined from (6.27) for a desired entropy $H_{\text{pbse}}$ in nats.

\textit{Proof.} The proof is given in Appendix 6.G. \hfill $\Box$

\textbf{Remark 6.14.} When $M \to \infty$, the following asymptotic results holds: $r_m \simeq \sqrt{P(1 - \zeta)\sqrt{m}}$, $p_m \simeq (1 - \zeta)\zeta^{m-1}$, and $H_{\text{pbse}} \simeq -\ln (1 - \zeta) + \left( \frac{\zeta}{1 - \zeta} \right) \ln (\zeta)$.

In Fig. 6.11, we illustrate the PB-GAM constellation with probabilistic shaping in form of a 3D-sideview. We observe the expected discretized bell-shape, with the larger discrete probability values at its center, where the pdf for a continuous complex Gaussian r.v. peaks.
MI-optimization of PB-GAM: Formulation-P1

In this method we optimize \( p_m \) to maximize the MI for a given SNR, while we also assume that \( r_m \) is given. In the following, we assume the disc-GAM format \( r_m = c_{\text{pbis}} \sqrt{m} \), again \( \sigma_w^2 = 1/S \), and the transmit power is unit-normalized. The optimization problem becomes

\[
\text{maximize } I(Y; X) \\
\text{subject to } \sum_{m=1}^{M} p_m = 1, \quad m = \{1, 2, \ldots M\}, \\
\sum_{m=1}^{M} p_m c_{\text{pbis}}^2 m = 1.
\] (6.28)

One can assume \( c_{\text{pbis}} \) given, or it can be seen as a variable for optimization. A constraint \( p_m + 1 < p_m \), which make sense for approximating a complex Gaussian r.v., could also be added.

At a first look, analytical optimization may appear straightforward, as this optimization problem only involves equality constraints. The Lagrangian is simply

\[
\Lambda = I(Y; X) + \lambda_1 \left( 1 - \sum_{m=1}^{M} p_m c_{\text{pbis}}^2 m \right) + \lambda_2 \left( 1 - \sum_{m=1}^{M} p_m \right).
\] (6.29)

However, the MI-expression, \( I(Y; X) \), here is nonlinear and does not (as far as can be seen) allow for an analytical optimization approach\(^3\).

6.2.4 Joint Probabilistic-Geometric-Shaping

In the preceding sections, we have proposed to use either geometrically-shaped GAM, or probabilistically-shaped GAM. In the following, we briefly point out the possibility for joint probabilistic-geometric shaping.

MI-optimization of GPB-GAM: Formulation-GP1

Here, we propose a joint optimization of \( r_m \) and \( p_m \) to maximize the MI under an average unit-power (and alternatively also with a PAPR) constraint. The \( m \)th constellation point is \( x_m = r_m e^{i2\pi m} \) with probability \( p_m \). The optimization problem,

\(^3\)The challenge to solve the problem is seen at a closer look. The partial derivative of the Lagrangian wrt \( p_m \), equating to zero, is \( \frac{dI(Y)}{dp_m} - \lambda_1 c_{\text{pbis}}^2 m - \lambda_2 = -1 - \int_f f_{y|x_m} \ln(f_y) \, dy \),

\[
f_{y|x_m} = \frac{1}{\pi \sigma_w} e^{-\frac{|y-x_m|^2}{\sigma_w^2}} , \quad f_y = \sum_{m=1}^{M} p_m f_{y|x_m} , \quad x_m = c_{\text{pbis}} \sqrt{m} e^{i2\pi m}.
\]
with $\sigma^2_w = 1/S$, is then
\[
\begin{align*}
\text{maximize} & \quad I(Y; X) \\
\text{subject to} & \quad \sum_{m=1}^{M} p_m = 1, \ m = \{1, 2, \ldots M\} \\
& \quad r_{m+1} > r_m, \\
& \quad r_1 \geq 0, \\
& \quad \sum_{m=1}^{M} p_m r_m^2 = 1.
\end{align*}
\] (6.30)

One may optionally also add the, intuitively-based and esthetic, constraint $p_{m+1} < p_m$. Again, an analytical solution is untractable. In Section 6.5, we opt for numerical optimization with non-linear objective function, and non-linear constraints.

### 6.3 Rank-1 (Fibonacci) Lattice-based Constellation Design

Up to now, we have assumed, and argued, for the phase difference between consecutive constellation points where $\alpha = \varphi$. However, when the cardinality $M$ is given, which is generally the case, it would be desirable if the $M$ constellation points were not just 2D u.d. mod 1, but uniformly distributed with equidistant phase values with support at, say, $2\pi m/M, \ m = \{0, 1, \ldots M - 1\}$. Such design can be based on a rank-1 lattice designs [SJ94, Ch. 4], which is proposed in this section.

**Definition 6.3.** (Rank-1 lattice) Let $p \in \mathbb{N}, q \in \mathbb{N}, 1 < p < q$, with $p$ and $q$ coprime, then a rank-1 Lattice can be defined as the sequence $(mp/q \mod 1, m/q)$, $m \in \{0, 1, \ldots q - 1\}$.

The corresponding (GAM-like) constellation, based on the rank-1 lattice design, is defined as follows.

**Definition 6.4.** (GAM-like constellation based on rank-1 lattice) The complex amplitude of the $m$th constellation point, of rank-lattice-based GAM, has the form
\[
x_m = r_m e^{i2\pi mp/q}, \ m \in \{1, 2, \ldots q\}.
\] (6.31)

We may, e.g., choose $r_m \propto F^{-1}((m - 1/2)/q)$, where $F^{-1}(\cdot)$ is the desired inverse sampling cdf.

However, not all rank-1 lattices are (so called) good lattices, i.e. where points are uniformly distributed over $[0, 1]^d$, which prompts searching for generating vectors $(1, p)$ yielding good lattice designs for each constellation size, $M = q$. One approach is to formulate it as an optimization problem, i.e. determining the generating vector that maximizes MI for a given SNR and $M$. The optimization problem is
Definition 6.5. (Rank-1 Fibonacci-lattice form) The complex amplitude of the \( m \)th constellation point, of Fibonacci-lattice-based GAM, has the form

\[
x_m = r_m e^{i2\pi m F_{n-1}/F_n}, \quad m \in \{1, 2, \ldots, F_n\}.
\]  

(6.32)

Note that since \( \lim_{n \to \infty} F_{n-1}/F_n = (1 + \sqrt{5})/2 = 2 + \varphi \), the Fibonacci-lattice-based design converges to the GAM-design, where \( \alpha = \varphi \).

Remark 6.15. Note that the considered rank-1 lattice design, including the Fibonacci-lattice design, can be used as basis for signal space diversity, a.k.a. mapping rearrangement, e.g. together with HARQ. This offers a variant of the \( q \)-PAM based signal space diversity schemes for two transmissions introduced in [BC92], to \( q \)-PAM, \( q \in \mathbb{N} \). The \( q \)-PAM signal constellation in the first transmission is then \( x_{m}^{(1)} = s_{m}^{(1)} - E\{s_{m}^{(1)}\} \), \( s_{m}^{(1)} = \{0, 1/q, \ldots, (q - 1)/q\} \), and in the second transmission, then \( x_{m}^{(2)} = s_{m}^{(2)} - E\{s_{m}^{(2)}\} \), \( s_{m}^{(2)} = \{1\}p/q \bmod 1 \). Note that by symmetry, the constellation point at \((0,0)\) in the rank-1 lattice can be discarded, allowing for a \( q - 1 \)-PAM design instead. Similarly, the rank-1 lattice can be extended with a constellation point at \((1,1)\), which allows for \( q + 1 \)-PAM. Signal-space diversity have been treated in many other works. For example, multidimensional rotated QAM was introduced in [BV98]. In contrast, the rank-1 design avoids the need of any (judiciously chosen) rotation of a cubic lattice.

Note that the term Fibonacci-lattice differs in, e.g., [SJ94, Ch. 4] compared to its use in [MS01a, MS01b, PLC09]. In [SJ94, Ch. 4], \( \alpha = F_{n-1}/F_n \) is rational and the lattice is plotted in Euclidean coordinates, whereas in [MS01a, MS01b] ([PLC09]), \( \alpha = (\sqrt{5} - 1)/2 \) is the golden mean, and the points are plotted in polar (spherical) coordinates. Here, we follow the convention in [SJ94, Ch. 4]. Strictly speaking, but practically irrelevant, using the irrational value \( \alpha = (\sqrt{5} - 1)/2 \) does not fulfill the translation-criteria of a lattice in Euclidean coordinates on \((0,1)^2\). Note also that each layer of the 3-D color codebook in [MS01a, MS01b] define (truncated sets of) points on the form \( x_m = \sqrt{m} e^{i2\pi \tau \delta} \), and in the most general formulation as \( x_m = m^{\delta} e^{i2\pi \tau \delta} \), \( \tau, \delta \in \mathbb{R} \). In contrast, the modulation framework here is not limited to, e.g., \( r_m \propto \sqrt{m} \), but can, e.g., be given by GB-GAM, Theorem 6.6, or optimized for desired performance metric.
6.3. Rank-1 (Fibonacci) Lattice-based Constellation Design

Figure 6.12: Rank-1 Fibonacci-lattices \((m F_{n-1}/F_n \mod 1, m/F_n), m \in \{0, 1, \ldots, F_n - 1\}\). (a) \(n = 9, F_n = 34\); (b) \(n = 8, F_n = 21\).
6.4 GAM-based (H)ARQ

While designing a fully-fledged (H)ARQ scheme with practical (de-)modulation and channel (de-)coding designs goes beyond the scope of the thesis, it is still of interest to motivate the system models and analytical throughput expressions in the foregoing chapters with a joint design of GAM and (H)ARQ. ARQ, as well as RR/IR-HARQ, are straightforward to integrate with GAM for a given channel code and constellation mapping\(^5\). Hence, we merely study the numerical throughput performance using a joint design of GAM and ARQ and RR. However, we also discuss (and numerically study) an alternative HARQ-design to IR below.

6.4.1 GAM-based HARQ with Mapping Rearrangement

In this thesis, and most other modern (H)ARQ literature, (H)ARQ throughput analysis for IR-HARQ is done under the assumption of that the AMI can be modeled by the sum of AWGN channel capacities for each (re)transmission. It is of interest to sketch a simple practical (H)ARQ design exploiting GAM that makes the AMI-assumption, used in the thesis, practically plausible. We consider HARQ with mapping rearrangement (MR-HARQ), see e.g. [SB05, SCF09].

Specifically, we consider a joint design between GAM and MR-HARQ (GMR-HARQ) with random mapping rearrangement. The motivation for the GMR-HARQ scheme is to simplify the HARQ design, and reduce the en/decoding complexity, compared to the more complex (low-rate mother code-based) IR-HARQ design. At the same time we want to retain the performance benefits of IR-HARQ where one assume that the AMI can be modeled with the AWGN channel capacity. The rationale behind the random mapping is that a mapping design that maximizes, e.g., the MI-performance is not straightforward to determine, and that the random mapping, based on numerical studies, indicates MI-performance virtually as good as IR-HARQ with the MI modeled as the AWGN channel capacity. As HARQ with mapping rearrangement as such is already reported in the literature, we do not pursue the discussion further here.

6.5 Numerical Results and Discussions

In this section, the MI-performance for some of the presented GAM-schemes, and the throughput for the GMR-HARQ scheme, are studied.

---

\(^5\)To be more precise, if the alphabet size of the channel coder with equiprobable symbols is of the same as the modulation alphabet, a particular mapping is not required. However, using a binary channel code alphabet suggest that an optimal mapper design exists. The optimal design for such mapper is unknown. One possible design for a mapper, between a binary channel code to GAM constellation, could be built on Gray-coding the Cartesian-coordinate 2D-lattice, see e.g. Fig. 6.7, prior transformation to polar-coordinates.
6.5.1 Mutual Information Performance

To exactly analyze the MI of GAM (with its irregular cell-shapes and different cell-sizes) is untractable. Except for the numerically optimized constellations, which use numerical integration, we settle for Monte-Carlo simulations wrt the MI-performance results. The MI-estimator used is

\[ \hat{I}(Y; X) = \frac{1}{J_{MC}} \sum_{j=1}^{J_{MC}} \log_2 \left( \frac{p(y^{(j)}|x^{(j)})}{\sum_{m=1}^{M} p(y^{(j)}|x_m)p(x_m)} \right), \tag{6.33} \]

where \( J_{MC} \) is the number of iterations, using the baseband samples \( y^{(j)} = x^{(j)} + w^{(j)} \) and the bivariate conditional pdf is \( p(y|x) = \frac{1}{\pi \sigma_w^2} e^{-\frac{(y-x)^2}{\sigma_w^2}} \). For each iteration \( j \), \( x \) is randomly selected from \( x_m \) with probabilities \( p(x_m) = p_m \). The average power of the modulation signal is unit-normalized, and the noise variance is \( \sigma_w^2 = 1/S \).

For the non-linear MI-maximization problems, involving inequality (as well as equality) constraints, we use the \texttt{fmincon.m} routine in Matlab\textsuperscript{®} together with numerical integration of the objective function.

Channel Capacity Gap

We set out to propose a modulation method that asymptotically approached the AWGN channel capacity with increasing constellation size. For this reason, we plot the capacity gap, i.e. the gap between the AWGN channel capacity and the MI-performance vs. cardinality \( M \) at fixed SNR, \( S = 20 \) dB, for several variants of GAM, but also other modulation schemes, in Fig. 6.13. The studied schemes are Disc-GAM, GB-GAM, MI-optimized TGB-GAM, SNR-based TGB-GAM, APSK-a, APSK-b, QAM, and NU-QAM. All schemes, except QAM and Disc-GAM with approximately uniform distribution of constellation points, exhibits a demising capacity-gap with increasing cardinality, \( M \). NU-QAM and APSK-b perform worse than the GAM-variants. The same applies to APSK-a, except for \( M \gtrsim 10^4 \). Both GB-GAM and optimized TGB-GAM, where the latter performs best, have approximately a unit negative slope, implying \( C - I(Y_M; X_M) \approx \text{Const}/M \) for large \( M \).

To test which irrational \( \alpha \) to use, we consider the gap between optimized TGB-GAM with \( \alpha = \varphi \), and optimized TGB-GAM with other \( \alpha = \{\sqrt{2}, \sqrt{3}, \pi, e\} \), in Fig. 6.14. We note that the other choices, than \( \alpha = \varphi \), leads to a performance loss, which also depends on \( M \).

Geometric Bell-GAM

In Fig. 6.15, we illustrate the MI-performance for GB-GAM together with the AWGN channel capacity. As expected, for larger constellation size \( M \), a greater overlap with the AWGN channel capacity is seen. The MI-approximation of GAM
Figure 6.13: Capacity-to-MI gap for Disc-GAM, GB-GAM, MI-optimized TGB-GAM, SNR-based TGB-GAM, APSK-a, APSK-b, QAM and NU-QAM, vs. signal constellation cardinality, $M$, and $S = 20$ dB.

Figure 6.14: MI-to-MI gap for MI-optimized TGB-GAM with $\alpha = \varphi$, and MI-optimized TGB-GAM with irrational $\alpha = \{\sqrt{2}, \sqrt{3}, \pi, e\}$, vs. signal constellation cardinality, $M$, and $S = 20$ dB.
6.5. Numerical Results and Discussions

<table>
<thead>
<tr>
<th>SNR</th>
<th>AWGN capacity</th>
<th>GB</th>
<th>G1</th>
</tr>
</thead>
<tbody>
<tr>
<td>≈ 4.8 dB</td>
<td>2</td>
<td>1.921</td>
<td>1.961</td>
</tr>
<tr>
<td>≈ 11.8 dB</td>
<td>4</td>
<td>3.440</td>
<td>3.549</td>
</tr>
<tr>
<td>15 dB</td>
<td>≈ 5.03</td>
<td>3.828</td>
<td>3.926</td>
</tr>
</tbody>
</table>

Table 6.1: MI in [b/Hz/s] for GB-GAM, and G1-formulation, with $M = 16$.

appears good when $MI < H - 2$ [b/Hz/s]. Naturally, the MI is limited by the entropy of the signal constellation. We observe an intermediate SNR region (which depends on $M$), where the MI does not coincide with $\min(\log_2(1 + S), \log_2(M))$, and further constellation optimization is of interest, e.g. as in formulation-G1.

In Tab. 6.1, the MI of GB-GAM and the G1-formulations are given. As expected, the optimized scheme, G1, performs slightly better than GB-GAM.

Truncated GB-GAM

In Fig. 6.16, we illustrate the MI-performance for the SNR-dependent TGB-GAM (6.19) together with QAM and the AWGN channel capacity. It is observed that the MI-performance of TGB-GAM is higher than QAM, and approaches the AWGN channel capacity as SNR decreases. For $M = 16$, TGB-GAM and QAM performs about the same. A possible misconception could, potentially, be that since 16-QAM and 16-GAM have almost the same MI vs. SNR performance, there is no point in favoring GAM over QAM for relatively small constellation sizes, e.g. $M = 16$. Such an idea would be based on a certain design paradigm. Let us consider an example. Say that a rate $R = 3$ [b/Hz/s] is desired. Then, the idea could be to use a constellation size $M = 2^4$ and a channel code rate $r = 3/4$. Zooming in around $C = 3$ [b/Hz/s], Fig. 6.17 shows that the SNR-gaps for 16-QAM and 16-TGB-GAM (6.19) to the AWGN channel capacity at $C = 3$ [b/Hz/s] are, respectively, $\approx 0.85$ dB.
Figure 6.16: MI of TGB-GAM (6.19), QAM, and the AWGN channel capacity vs. SNR, $S$, for $H = \log_2(M) = \{4, 6, 8, 10\}$.

Figure 6.17: MI of TGB-GAM (6.19), QAM, and the AWGN channel capacity vs. SNR, $S$, where $C \approx 3 \text{ [b/Hz/s]}$. 
6.5. Numerical Results and Discussions

<table>
<thead>
<tr>
<th>SNR [dB]</th>
<th>MI and Shannon capacity [b/Hz/s]</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>MI opt. TGB-GAM, (6.17)+(6.18)</td>
</tr>
<tr>
<td>8.2</td>
<td>MI GB-GAM, (6.9)+(6.10)</td>
</tr>
<tr>
<td>8.4</td>
<td>MI Disc-GAM (6.2)+(6.3)</td>
</tr>
<tr>
<td>8.6</td>
<td>MI QAM</td>
</tr>
<tr>
<td>8.8</td>
<td>MI NU-QAM</td>
</tr>
<tr>
<td>9</td>
<td>C = \log_2 (1 + S)</td>
</tr>
<tr>
<td>9.2</td>
<td>MI TGB-GAM, (6.19)</td>
</tr>
<tr>
<td>9.4</td>
<td></td>
</tr>
<tr>
<td>9.6</td>
<td></td>
</tr>
<tr>
<td>9.8</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
</tr>
</tbody>
</table>

Figure 6.18: MI of opt. TGB-GAM (6.17)+(6.18), GB-GAM (6.9)+(6.10), disc-GAM (6.2)+(6.3), QAM, NU-QAM, and TGB-GAM (6.19) with \(M = 2^{10}\), and the AWGN channel capacity vs. SNR, \(S\).

and \(\approx 0.76\) \(\text{dB}\). (Note that the MI can be improved for GAM using (6.17)+(6.18) or with optimization of \(r_m\)). However, this is not necessarily the best design approach, as the required SNR for a desired rate \(R\) is unnecessarily high. Let us instead consider the MI-performance for very large \(M\), say \(M = 2^{10}\). Then, the SNR-gaps from 1024-QAM and 1024-TGB-GAM (6.19) to \(C = 3\) [b/Hz/s] are, respectively, \(\approx 0.5\) \(\text{dB}\) and \(\approx 0.01\) \(\text{dB}\). So a designer could, on one hand choose 16-QAM with a \(r = 3/4\) code, or on another hand choose 1024-TGB-GAM with a rate \(r = 3/10\) code. The latter choice reduces the SNR-gap with 0.84 dB. This has implication for modem design. Instead of implementing many different constellation sizes, one may implement just the largest desired constellation size, say \(M = 2^{10}\), and then adapt the code rate \(r\) to achieve the desired rate \(R\). Of course, this reasoning assumes that other costs, e.g. any possible extra energy consumption, are of less concern. In Fig. 6.18, we compare the MI of TGB-GAM (both for (6.17)+(6.18), and (6.19)), with GB-GAM, disc-GAM, QAM, and non-uniform QAM (NU-QAM) [SF00, (1)-(2)]. As aimed for, TGB-GAM (both variants) performs better than all reference-cases. NU-QAM performs quite poorly when \(MI \approx H\). Disc-GAM is, as expected due its disc-shape, about 0.2 dB better than QAM.

**Probabilistic Bell-GAM**

Similar conclusion, as for the GB-GAM, Fig. 6.15, is drawn for PB-GAM (min SNR - entropy constrained) shown in Fig. 6.19. However, the MI for PB-GAM schemes is substantially higher, almost with a near ideal shape. But keep in mind, the scheme requires a larger \(M\) for the same entropy as PG-GAM (high-rate), due to non-identical \(p_m\). For the optimization formulation-P2, we show the MI-performance in Fig. 6.20. Here, \(H = \log_2(M)\), and the probabilistic-shaping is optimized wrt the
Figure 6.19: MI of PB-GAM (min SNR - entropy constrained) with $\log_2(M) = \{5, 7, 9, 11\}$, QAM asymptote, and the AWGN channel capacity vs. SNR, $S$.

Figure 6.20: MI of PB-GAM (Formulation-P2) with $\log_2(M) = \{4, 6, 8, 10\}$, QAM asymptote, and the AWGN channel capacity vs. SNR, $S$.

SNR. This is slightly better than GB-GAM, but it requires a shaper which adapt $p_m$ to the SNR.

**Overall Comparison**

In Fig. 6.21, we compare four different modulation schemes, disc-GAM, GB-GAM, PB-GAM (min SNR - entropy constrained), and QAM, together with the AWGN channel capacity for high $M$. We note the SNR-gap of QAM and disc-GAM to the AWGN channel capacity, and the $\approx 0.2$ dB SNR gap between disc-GAM and QAM. We observe that when the MI approaches the entropy, disc-GAM, and QAM,
Figure 6.21: MI of disc-GAM, GB-GAM, PB-GAM (min SNR - entropy constrained), QAM, and AWGN channel capacity vs. SNR, \( S \).

perform better than GB-GAM scheme. Thus, GB-GAM, while essentially optimal at lower MI, is suboptimal when we have \( MI \approx H \).

In Fig. 6.22, the focus is shifted toward the MI-optimized scheme, but includes non-optimized schemes for reference. GB-GAM, which is not based on any optimization, performs the worst. GB-GAM with optimization formulation-G1 performs significantly better than GB-GAM when the MI is of the same order as the entropy, but is still the second worst. PB-GAM optimization formulation-P2, tuning the probabilities, comes next, and, overall, performs better than geometric-GAM in general. We also investigated PB-GAM optimization formulation P1, and it matches the MI-performance for P2. Hence, not shown here. When geometric and probabilistic shaping is combined, and optimized, as in GPB-GAM formulation GP1, it performs (as expected) better than either of GB-, and PB-GAM. The MI PB-GAM (min SNR - entropy constraint) performs the best here, at the cost of twice as many constellation points as the other schemes.

**Rank-1 (Fibonacci) Lattice-based Constellation Design**

In Fig. 6.23, we show the gap between MI for optimized TGB-GAM using \( \alpha = \phi \) for any \( M \), and the MI for optimized TGB-GAM with a rank-1 lattice using \( \alpha = p^*/q \) for any \( M = q \). Here \( p^* \) indicates that the MI-optimizing \( p \). In Fig. 6.24, we illustrate the particular case of the Fibonacci-lattice. We note that the performance differences are relatively minor when optimization with respect to \( p \) is considered.

### 6.5.2 Throughput Performance of (H)ARQ with GAM

We now study the throughput performance of (H)ARQ complemented with GAM and compare with the throughput performance of (H)ARQ where the MI is ap-
Golden Angle Modulation and Rank-1 Lattice Designs

Figure 6.22: MI of GB-, PB-, and GPB-GAM variants, vs. SNR, S, for \( M = 16 \).

Figure 6.23: MI-to-MI gap for MI-optimized TGB-GAM with \( \alpha = \varphi \), and MI-optimized TGB-GAM with rational \( \alpha = p^*/q \), vs. signal constellation cardinality, \( M = q \), and \( S = 20 \) dB.
proximated with the AWGN channel capacity. For simplicity, while not offering the optimal performance, we consider the GB-GAM formulation in Theorem 6.6.

The throughput of the (H)ARQ-schemes with GAM are Monte-Carlo simulated for iid block Rayleigh fading channels. We use the MI-estimator in (6.33) to check whether the MI (AMI) for ARQ (RR) exceeds the rate \( R \) such that the message can be decoded. For the random mapping rearrangement based HARQ case, the MI-estimator in (6.33) is extended to \( k \) transmissions. The conditional pdf for \( k \) transmissions-case is now \( p(y|x) = \prod_{k'=1}^{k} \frac{1}{\sigma_{y_k'}} \exp \left( - \frac{|y_{k'} - x_{k'}|^2}{\sigma_{y_k'}^2} \right) \). We chose to keep the average power of the constellation constant, and equal to unity, whereas the noise power \( \sigma_{y_k'}^2 = 1/S_k \) varies between transmissions depending on the instantaneous fading channel gain.

In Fig. 6.25(a)-(c), we show the throughput performances for ARQ, RR, and MR-HARQ with GB-GAM, as well as with the AWGN channel capacity modeled MI, respectively. We consider an iid block Rayleigh fading channel and the rates \( R = \{1, 2, 3, 4, 5\} \) [b/Hz/s]. We note that the HARQ cases with GB-GAM performs as well as the (H)ARQ cases with AWGN channel capacity modeled MI, provided that the initial channel code rate \( R \) is somewhat smaller than the entropy, \( H = \log_2(M) \). Specifically, for the studied case with \( M = 64 \), the performance curves essentially overlap when the rate is \( R \approx 3 \) [b/Hz/s], or half the entropy. For higher rates up to the entropy, the performance for GB-GAM-based (H)ARQ is slightly smaller than for (H)ARQ with AWGN channel capacity modeled MI. This is because the MI of any discrete signal constellation, including GAM, is upper bounded by the signal constellation entropy, whereas the MI of IR-HARQ is unbounded with the AWGN channel capacity MI model. Thus, the AWGN channel capacity based analysis, and related analytical performance expressions, seem

Figure 6.24: MI-to-MI gap for MI-optimized TGB-GAM with \( \alpha = \varphi \), and MI-optimized TGB-GAM with \( \alpha = F_{n-1}/F_n \), \( n = \{7, 8, \ldots, 16\} \), vs. signal constellation cardinality, \( M = F_n \), and \( S = 20 \) dB.
practically well-founded by a modulation scheme such as GB-GAM.

6.6 Summary and Conclusions

In this chapter, we introduced a new, AWGN channel capacity-achieving, near-circular-symmetric, high-order modulation format that with geometric-shaping and probabilistic-shaping can approximate a circular-symmetric pdf of choice. We illustrated such approximations for the pdf of a continuous complex Gaussian r.v. We gave several optimization frameworks, maximizing the MI under an average SNR-equality constraint, and optionally also a PAPR-inequality constraint. The best MI-performance was achieved when both constellation point-geometry and -probabilities where optimized, and the second best appears to be when only the constellation point probabilities were optimized. We showed that, with this practical modulation format, the iid complex Gaussian communication signal model assumption, as used for performance analysis throughout the thesis, is well founded. Relating directly to HARQ, the theme of the thesis, MR-HARQ build on GB-GAM was considered. We found that this scheme offered classical IR-HARQ-like throughput performance for which MI was approximated with the AWGN channel capacity. Likewise, ARQ and RR with GB-GAM offered near-identical throughput-performance to the case where MI is modeled with the AWGN channel capacity. It is believed that GAM may find applications in transmitter-resource-limited links, such as space probe-to-earth-, satellite-, or mobile-to-basestation-based communication. This is so since given a desired, and preferably a high, spectral-efficiency, power- and energy- resources are vital to conserve. However, higher decoding complexity can be acceptable at the receiver side. In a broader perspective, the demonstrated AWGN channel capacity achieving performance of GAM is believed to be of interest for any communication system of choice, wireless, optical, or wired. Numerous promising variants and extensions can be envisioned, and some further research directions are outlined in Appendix 6.H.
Figure 6.25: Monte Carlo simulated throughput of (H)ARQ with GB-GAM, and with the AWGN channel capacity modeled MI, for $M = 64$, and $R = \{1, 2, 3, 4, 5\}$ [b/Hz/s]: (a) ARQ with GB-GAM; (b) RR-HARQ with GB-GAM; (c) MR-HARQ with GB-GAM.
6.7 Appendices

6.A Proof in Theorem 6.1

Proof. The angular distribution for the constellation points for GAM is already given by the phase-factor, $e^{i2\pi \phi_m}$. However, the magnitudes $r_m$ need to be determined. For a versatile design, the arrangement of constellation points target an approximately uniform distribution on a wheel-shaped region with an inner and outer circular boundary. When there is no inner boundary, the shape is fully disc-like. For the purpose, we first consider a complex r.v. uniformly distributed with a millstone/wheel-shaped support having inner radius $\tilde{r}_i$ and outer radius $\tilde{r}_o$. We use the method of inverse sampling. I.e., a r.v. $X$, with the cdf $F_X(x)$, is generated by $X = F_X^{-1}(U)$, where $U$ assumes a uniform continuous r.v. on $(0, 1)$. To map $M$ discrete constellation points, we modify the inverse sampling method and approximate $U \sim \text{Unif}(0, 1)$ with a discrete r.v. with probability $1/M$ and support at $(\frac{m - 1/2}{M}, m \in \{1, 2, \ldots, M\})$. The cdf over the wheel-shaped continuous pdf is

$$F(\tilde{r}_m) = \int_{\tilde{r}_i}^{\tilde{r}_m} f_R(r)dr = \int_{\tilde{r}_i}^{\tilde{r}_m} \frac{1}{\pi(\tilde{r}_o^2 - \tilde{r}_i^2)} 2\pi r dr = \frac{\tilde{r}_m^2 - \tilde{r}_i^2}{\tilde{r}_o^2 - \tilde{r}_i^2}. \quad (6.34)$$

We now assume that a hypothetical number of constellation points, $M_l$ (a lower offset), inside the circle of radius $r_1$ is proportional to the circle’s area. Without limitation, due to subsequent power normalization, we can set $\tilde{r}_i^2 = M_l$. Similarly, we let $\tilde{r}_o^2 = M + M_l$, so the number of constellation points on the wheel-shaped pdf, between $r_i$ and $r_o$, is $M$. We now set $F(\tilde{r}_m) = (m - 1/2)/M$, and solve for $\tilde{r}_m$.

The tentative, not-yet-power-normalized, magnitudes are then

$$\tilde{r}_m = \sqrt{m - 1/2 + M_l}. \quad (6.35)$$

Hence, the magnitudes of the signal constellation has the form $r_m = c_{\text{disc}} \tilde{r}_m$. The constant $c_{\text{disc}}$ is determined from the average power constraint $\bar{P}$, as follows

$$\bar{P} = \sum_{m=1}^{M} p_m r_m^2 = c_{\text{disc}}^2 \sum_{m=1}^{M} \frac{m - 1/2 + M_l}{M} = c_{\text{disc}}^2 \frac{M + 2M_l}{2}$$

$$\Rightarrow c_{\text{disc}} = \sqrt{\frac{2\bar{P}}{M + 2M_l}}. \quad (6.35)$$

In all, the magnitude $m$th constellation point is

$$r_m = \sqrt{\frac{2\bar{P}(m - 1/2 + M_l)}{M + 2M_l}}. \quad (6.36)$$

When $M_l = 0$, the magnitude expression simplifies to $r_m = \sqrt{2\bar{P}(m - 1/2)/M}$. Allowing for a constellation point at the origin, $(0, 0)$, which implies $M_l = -1/2$, an alternative magnitude expression is $r_m = \sqrt{2\bar{P}(m - 1)/(M - 1)}$. □
6.7. Appendices

6.B Proof in Corollary 6.1

Proof.

\[
\lim_{M \to \infty} \frac{1}{M} \sum_{m=1}^{M} e^{i2\pi \langle h, x_m \rangle} = \lim_{M \to \infty} \frac{1}{M} \sum_{m=1}^{M} e^{i2\pi (h_1 m/M + h_2 \phi_m)} \\
= \lim_{M \to \infty} \frac{1}{M} \frac{1 - e^{i2\pi (M+1) (h_1/m + h_2 \phi)}}{1 - e^{i2\pi (h_1/M + h_2 \phi)}} \\
< \lim_{M \to \infty} \frac{1}{M} \frac{2}{|1 - e^{i2\pi (h_1/M + h_2 \phi)}|} \\
= 0. \quad (6.37)
\]

6.C Alternative Constellation Designs

Square- and Hexagonal-lattices

Now, not limiting the discussion to a spiral-based design, two cases to compare GAM with are square- and hexagonal-lattices with an approximately circular-symmetric outer boundary. This is shown in Fig. 6.26. The motivation to consider the hexagonal-lattice is that it offers the densest packing on an infinite plane. Thus, the uniformly distributed disc-GAM can not offer higher packing density than the uniformly distributed hexagonal-lattice on the infinite plane. However, other features, such as the ease of radial geometric shaping, design flexibility, and AWGN channel capacity approaching performance could motivate the GAM-design. Now, on the LHS in Fig. 6.26, we find constellations, depicted as Voronoi diagrams, where the cell areas of each constellation point are of the same sizes for the square and hexagonal lattice, and approximately the same for GAM. On the RHS, we have the same constellations, under the radial transformation \( r'_m = c \sqrt{-\ln(1 - r^2_m)} \), with retained phase. It is clear that the initial circular symmetry for square(hexagonal)-lattice, due to the 2(3)-fold axis-symmetry, gives a non-circular outer boundary. In contrast, GAM retains the approximately circular outer boundary under geometric shaping in the radial domain. The latter fact was also the key observation made when observing various plants. A problem with the radial transformations for square- and hexagonal-lattices is that the cell-shapes are deformed with increasing magnitude. Hence, the favourable packing properties for, in particular, the hexagonal-lattice is lost. Other square- and hexagonal-lattice-based designs with circular-symmetric outer boundaries can also be envisioned. For example, non-linear geometric shaping with respect to the symmetry-axis retains the square- and hexagonal cell-shapes under transformation. Unfortunately, much like in Fig. 6.26, the (desirable) approximately circular boundary is not retained under radial geometric shaping.
Figure 6.26: Uniform and Gaussian-shaped (circular-trunc.) constellations: (a) Uniform GAM (disc-GAM) $M = 256$; (b) Gaussian-shaped GAM $M = 256$; (c) Uniform (circular-trunc.) square lattice $M = 253$; (d) Gaussian-shaped (circular-trunc.) square $M = 253$; (e) Uniform (circular-trunc.) hexagonal lattice $M = 253$; and, (f) Gaussian-shaped (circular-trunc.) hexagonal $M = 253$. 
APS K and Star-QAM.

APS K and Star-QAM, have, similarly to GAM, a circular outer boundary, and radial symmetry. We show two different APSK designs, APSK-a (Star-QAM) and APSK-b, in Fig. 6.27. In APSK-a, which is based on the design [M15], we let
\[ x_m \propto r_m e^{i2\pi m/M}, \quad m_r = \{1, 2, \ldots, M_r\}, \quad m_\alpha = \{1, 2, \ldots, M_\alpha\}, \]
where \( M_r = M_\alpha = \sqrt{M} \), and \( M \) is a squared integer. In APSK-b, for an approximately uniform disc design, we let
\[ x_m \propto r_m e^{i2\pi m/M}, \quad r_{m,r} = r_{m,r+1}, \quad m_r = \{1, 2, \ldots, M_r\}, \quad M_{\alpha,r} = 6m_r. \]
We also have \( x_{0,0} = 0 \). Hence
\[ M = 1 + 3M_r(M_r + 1). \]
While both APSK-a and APSK-b have radial symmetry, a radial transform
\[ r'_r = \frac{c}{\sqrt{-\ln(1 - r_{m,r}^2)}}, \]
gives a sparse angular density for the outermost points in APSK-a, but dense in APSK-b. We consider those constellations as reference cases in the numerical evaluation section.

6.D Proof in Theorem 6.6

Proof. Similar as for the proof of disc-GAM in Section 6.A, we use the method of inverse sampling. A complex Gaussian r.v. with variance \( \sigma^2 \) is initially considered. Since only the shape is important for the pdf, and power normalization is done for the constellation, we can assume \( \sigma^2 = 1 \). The cdf is then
\[
F(\tilde{r}_m) = \int_0^{\tilde{r}_m} f_R(r)dr = \int_0^{\tilde{r}_m} e^{-\frac{r^2}{\sigma^2}} \frac{2\pi r}{\pi\sigma^2} dr = 1 - e^{-\frac{\tilde{r}_m^2}{\sigma^2}}. \quad (6.38)
\]
Setting \( F(\tilde{r}_m) = (m - 1/2)/M \), and solving for \( \tilde{r}_m \), yields
\[
\tilde{r}_m = \sqrt{\ln \left( \frac{M}{M - m + 1/2} \right)}. \quad (6.39)
\]
Thus, the general solution for the magnitude has the form \( r_m = c_{gb} \sqrt{\ln \left( \frac{M}{M - m + 1/2} \right)} \).

From the power constraint, \( \bar{P} \), the constant \( c_{gb} \) is given by
\[
\bar{P} = \sum_{m=1}^{M} p_m r_m^2 = \frac{c_{gb}^2}{M} \sum_{m=1}^{M} \ln \left( \frac{M}{M - m + 1/2} \right),
\]
\[
\Rightarrow c_{gb} = \sqrt{\frac{M\bar{P}}{\sum_{m=0}^{M-1} \ln \left( \frac{M}{M - m + 1/2} \right)}} = \sqrt{\frac{M\bar{P}}{M \ln M - \ln (\Gamma(M + 1/2)/\sqrt{\pi})}}. \quad (6.40)
\]
Figure 6.27: APSK constellations: (a) Uniform APSK-a with $M = 253$; (b) Gaussian-shaped APSK-a with $M = 253$; (c) Uniform APSK-b with $M = 289$; (d) Gaussian-shaped APSK-b with $M = 289$.

If we instead consider the support $\{0, 1, \ldots, M-1\}$ for the pmf used in inverse sampling, rather than the support $\{\frac{1}{2}, \frac{3}{2}, \ldots, M - \frac{1}{2}\}$, $r_m$ simplifies to

$$r_m = c_{gb} \sqrt{\ln \left( \frac{M}{M-m+1} \right)} \quad (6.41)$$

$$c_{gb} = \sqrt{\frac{MP}{M \ln M - \ln (M!)}} \quad (6.42)$$
6.7. Appendices

6.8 Alternative Proof in Theorem 6.6

Proof. As an alternative proof to determine \( r_m \)'s dependence on \( m \), we use high-rate quantization theory, valid for large number of quantization points. The radius \( r_m \) is determined by integrating the asymptotically optimal quantization point density, equating to the number of sampling points, \( m \), within the circles radius, and then solving for the \( r_m \). The asymptotical optimal quantization point density, with a MSE distortion measure, is

\[
\lambda(x^k) = M \frac{f(x^k)^{1/2}}{\int_{\mathbb{R}^k} f(x^k)^{1/2} \, dx^k},
\]

where \( f(x^k) \) is the pdf, \( k \) the dimension, and \( r \) the distortion norm. For the bivariate case, with pdf \( f(x, y) = \frac{1}{\pi \sigma^2} e^{-x^2 + y^2 \sigma^2} \), and MSE distortion measure, we have \( k = 2 \) and \( r = 2 \). This gives

\[
\lambda(x, y) = M \frac{f(x, y)^{1/2}}{\int_{\mathbb{R}^2} f(x, y)^{1/2} \, dx \, dy} = M \frac{1}{\sqrt{\pi \sigma^2}} e^{-\frac{x^2 + y^2}{2 \sigma^2}} \int_{\mathbb{R}^2} \frac{1}{\sqrt{\pi \sigma^2}} e^{-\frac{x^2 + y^2}{2 \sigma^2}} \, dx \, dy = M \frac{1}{2\pi \sigma^2} e^{-\frac{x^2 + y^2}{2 \sigma^2}}. \tag{6.44}
\]

From the definition of the point density function, and assuming a circular-symmetric pdf with radius \( r_m \) to point \( m \), the condition \( \int_{\|x\| \leq r_m} \lambda(x^k) \, dx^k = m - 1/2 \) needs to be fulfilled. Specifically, for the studied case, we get

\[
\int_0^{\tilde{r}_m} M \frac{1}{2\pi \sigma^2} e^{-\frac{r^2}{2 \sigma^2}} 2\pi r \, dr = m - 1/2
\]

\[
M(1 - e^{-\frac{\tilde{r}_m^2}{2 \sigma^2}}) = m - 1/2
\]

\[
\tilde{r}_m = \sqrt{2\sigma} \sqrt{\ln \left( \frac{M}{M - m + 1/2} \right)} \tag{6.45}
\]

Thus, the signal constellation has the form \( r_m = c_{gb} \sqrt{\ln \left( \frac{M}{M - m + 1/2} \right)} \). The constant \( c_{gb} \) is determined as in Appendix 6.D.

\[\Box\]
6.F Proof in Theorem 6.7

Proof. First, the channel capacity gap is upper bounded the relative entropy.

\[ I(Y;X|S) - I(Y_M;X_M|S) = (h(Y|S) - h(Y|X,S)) - (h(Y_M|S) - h(Y_M|X_M, S)) \]

\[ = h(Y|S) - h(W, S)) - (h(Y_M|S) - h(W, S)) \]

\[ = \left( - \int_{S_Y} f_Y \log(f_Y) \, dy \right) - \left( - \int_{S_Y} f_{Y_M} \log(f_{Y_M}) \, dy \right) \]

\[ \leq - \int_{S_Y} f_Y \log(f_Y/f_{Y_M}) \, dy \] (6.46)

where \( h(\cdot) \) is the differential entropy, and \( f_Y \) and \( f_{Y_M} \) are the output distributions for r.v.s \( Y \) and \( Y_M \), respectively. Hence, we can prove \( \lim_{M \to \infty} I(Y;X|S) - I(Y_M;X_M|S) = 0 \), by proving that \( \lim_{M \to \infty} f_{Y_M} = f_Y \). The output density for \( Y_M \) is known to be \( f_{Y_M} = \frac{1}{\pi \sigma^M} \sum_{m=1}^{M} e^{-\frac{|y-x_m|^2}{\sigma^2}} \). As the constellation size increases, we expect that \( f_{Y_M} \) to converges to the output density of r.v. \( Y \), namely \( f_Y = \frac{1}{\pi \sigma^2} \sum_{m=1}^{M} e^{-\frac{|y|^2}{\sigma^2}} \). We first give a generic expression for \( f_{Y_M} \) as \( M \to \infty \), which is not restricted to GB-GAM but allows for any continuous radial pdf \( f(x) \), \( F^{-1}(0) \leq x \leq F^{-1}(1) \). We can then write

\[ \lim_{M \to \infty} f_{Y_M} = \lim_{M \to \infty} \frac{S}{\pi M} \sum_{m=1}^{M} e^{-S|y-x_m|^2} \]

\[ \leq \lim_{M \to \infty} \frac{S}{\pi M} \sum_{m=1}^{M} e^{-S\rho^2 - \sigma^2 + 2\sigma r_m \rho \cos(2\pi \varphi_m - \alpha_y)} \]

\[ \leq \lim_{M \to \infty} \frac{S}{\pi M} e^{-S\rho^2} \sum_{m=1}^{M} e^{-S(F^{-1}(\frac{m-1}{M}))^2 + 2SF^{-1}(\frac{m-1}{M}) \rho \cos(2\pi \varphi_m - \alpha_y)} \]

\[ \leq \frac{S}{\pi} e^{-S\rho^2} \int_{0}^{1} \int_{0}^{1} e^{-S(F^{-1}(v))^2 + 2SF^{-1}(v) \rho \cos(2\pi u) \, du \, dv} \]

\[ = \frac{S}{\pi} \int_{0}^{1} e^{-S(F^{-1}(v))^2} I_0(2SF^{-1}(v) \rho) \, dv \]

\[ \leq \frac{S}{\pi} \int_{F^{-1}(0)}^{F^{-1}(1)} e^{-S(r_x^2 + x^2)} I_0(2SRx \rho) f(x) \, dx, \] (6.47)

where we in step (a) expressed \( r_m \) as the inverse of the desired inverse sampling cdf \( F^{-1}(\cdot) \), in step (b) we used the fact that we have a 2D u.d. mod 1 sequence on the unit square, and in step (c) we adopted the substitution \( x = F^{-1}(v), F(x) = v \), and defined the (continuous) input pdf \( f(x) \triangleq \frac{dF(x)}{dx} \) in the radial domain.
For the GB-GAM, and assuming average unit-power normalization, we consider the inverse sampling radial cdf
\[ F(x) = 1 - e^{-x^2}, \]
and corresponding pdf \( f(x) = 2xe^{-x^2}. \) Inserting \( f(x) \) in (6.47) yields
\[
\lim_{M \to \infty} f_{Y_M} = \frac{S}{\pi} e^{-Sr_y^2} \int_{F^{-1}(0)}^{F^{-1}(1)} e^{-Sx^2} I_0(2Sxr_y) 2xe^{-x^2} \, dx
\]
\[
= \frac{2S}{\pi} e^{-Sr_y^2} \int_0^{\infty} x e^{-(S+1)x^2} I_0(2Sxr_y) \, dx
\]
\[
= \frac{2S}{\pi} e^{-Sr_y^2} \frac{(2Sx)^2}{2(S+1)}
\]
\[
= \frac{1}{\pi S + 1} e^{-\frac{4Sr_y^2}{2(S+1)}}
\]
\[
= \frac{1}{\pi(1+1/S)} e^{-\frac{r_y^2}{1+1/S}}
\]
(6.48)
where we used the integral identity \( \int_0^\infty x e^{-ax^2} I_0(bx) \, dx = e^{b^2/2a} \) in [GR07, 6.633.4]. We see that \( \lim_{M \to \infty} f_{Y_M} = f_Y, \) which proves the theorem.

**Remark 6.16.** As an illustrative example, we repeat the same procedure for disc-GAM. We then consider an input distribution where the density is modeled as a continuous disc with average unit-power. This corresponds to a pdf for the radial component \( f(x) = x \) for \( 0 \leq x < \sqrt{2}, \) but otherwise zero. The asymptotic radial pdf is then
\[
f_{Y_M} \sim \frac{S}{\pi} e^{-Sr_y^2} \int_{F^{-1}(0)}^{F^{-1}(1)} e^{-Sx^2} I_0(2Sxr_y) f(x) \, dx
\]
\[
= \frac{S}{\pi} e^{-Sr_y^2} \int_0^{\sqrt{2}} e^{-Sx^2} I_0(2Sxr_y) x \, dx
\]
\[
= \frac{S}{\pi} \int_0^{\sqrt{2}} e^{-S(x^2 + r_y^2)} I_0(2Sxr_y) x \, dx
\]
\[
= \frac{S}{\pi} \int_0^{2\sqrt{2}} e^{-\frac{(\sqrt{2}\alpha + \sqrt{2}\beta)^2}{2}} I_0(\sqrt{2\alpha \beta}) x \, dx
\]
\[
= \frac{1}{2\pi} \int_0^{2\sqrt{2}} e^{-\frac{z^2}{2}} I_0(z) z \, dz
\]
\[
= 1 - Q_M(\sqrt{2Sr_y}, 2\sqrt{S}), \quad 0 \leq r_y, 0 \leq \alpha_y < 2\pi.
\]
(6.49)
where \( Q_M(\cdot) \) is the Marcum Q-function.
6.G Proof of Theorem 6.8

Proof. The Lagrange dual function is

$$\Lambda = \sum_{m=1}^{M} p_m m + \lambda_1 \left( 1 - \sum_{m=1}^{M} p_m \right) + \lambda_2 \left( H_{\text{pse}} - \sum_{m=1}^{M} p_m \ln(p_m) \right), \quad (6.50)$$

where $\lambda_1$ and $\lambda_2$ are Lagrange parameters to be determined.

Taking the derivative of the Lagrange function, equating to zero, $\frac{d\Lambda}{dp_m} = 0$, yields the optimality condition

$$m - \lambda_1 - \lambda_2 (\ln(p_m) + 1) = 0,$$

$$\Rightarrow p_m = e^{\frac{m - \lambda_1}{\lambda_2} - 1} \Rightarrow p_m = c_p \zeta^m, \quad (6.51)$$

where $c_p$ and $\zeta$ are parameters, related to $\lambda_1$ and $\lambda_2$, to be determined.

The constant $c_p$ is found from the unit-sum probability condition.

$$1 = \sum_{m=1}^{M} p_m = \sum_{m=1}^{M} c_p \zeta^m = c_p \zeta \frac{1 - \zeta^M}{1 - \zeta}$$

$$\Rightarrow c_p = \frac{1}{\zeta} \frac{1 - \zeta}{1 - \zeta^M} \Rightarrow p_m = \frac{1}{\zeta} \frac{1 - \zeta}{1 - \zeta^M} \zeta^m. \quad (6.52)$$

The constant $c_{\text{pse}}$ is found by using the unit-power normalization condition.

$$\bar{P} = \sum_{m=1}^{M} p_m^2 = c_{\text{pse}} \sum_{m=1}^{M} \frac{1 - \zeta}{1 - \zeta^M} \zeta^m m = c_{\text{pse}}^2 \left( \frac{1}{1 - \zeta} - \frac{M \zeta^M}{1 - \zeta^M} \right)$$

$$\Rightarrow c_{\text{pse}} = \sqrt{\bar{P} \left( \frac{1}{1 - \zeta} - \frac{M \zeta^M}{1 - \zeta^M} \right)^{-1}}. \quad (6.53)$$

Using $p_m$, the entropy is

$$H_{\text{pse}} = -\sum_{m=1}^{M} p_m \ln(p_m)$$

$$= -\sum_{m=1}^{M} \frac{1}{\zeta} \frac{1 - \zeta}{1 - \zeta^M} \zeta^m \ln \left( \frac{1}{\zeta} \frac{1 - \zeta}{1 - \zeta^M} \zeta^m \right)$$

$$= -\sum_{m=1}^{M} \frac{1}{\zeta} \frac{1 - \zeta}{1 - \zeta^M} \zeta^m \ln \left( \frac{1}{\zeta} \frac{1 - \zeta}{1 - \zeta^M} \right) - \sum_{m=1}^{M} \frac{1}{\zeta} \frac{1 - \zeta}{1 - \zeta^M} \zeta^m \ln \left( \zeta^m \right)$$
\[= - \ln \left( \frac{1 - \zeta}{1 - \zeta M} \right) - \ln(\zeta) \frac{1 - \zeta}{1 - \zeta M} \sum_{m=1}^{M} m \zeta^m\]

\[= - \ln \left( \frac{1 - \zeta}{1 - \zeta M} \right) - \left( \frac{1}{1 - \zeta} - \frac{M \zeta^M}{1 - \zeta M} \right) \ln(\zeta)\]

\[= - \ln \left( \frac{1 - \zeta}{1 - \zeta M} \right) + \left( \frac{\zeta}{1 - \zeta} + \frac{M \zeta^M}{1 - \zeta M} \right) \ln(\zeta). \quad (6.54)\]

6.7 Appendices

6.8 Future Research Directions and Extensions

It is recognized that the GAM format encompasses several promising features allowing for interesting extensions and applications. Below, we list a number of ideas possible as future research directions.

- Develop channel coding and decoding methods for GAM. Consider, e.g., i) CM with GF(M), where M is not necessarily an integer power of two, and ii) BICM.

- Design polar-coordinate-based Gray-coded modulation, with a set of bits Gray-coded in the radial-domain, and the other set of bits Gray-coded in the phase-domain. Based on the Gray-coded GAM, develop related (preferable reduced complexity) channel decoding schemes. E.g., regions where a set of constellation points assume the same Gray-bit value can be grouped together and modeled as a continuous distribution in corresponding region.

- Develop blind channel estimation methods based on only observing a set of GAM-symbols. This is doable since the GAM-constellation is unique in both the rotation and the radial domains.

- Study the MI-performance for low-complexity non-coherent detection, when only the amplitude component is considered.

- Design GAM level network coding schemes, e.g., for NCBR [LJS05, LJS06], NC-ARQ [LJ06], or NC-HARQ [LSKAT13]. Let \( m_1 \) and \( m_2 \) represent the GAM indices for a symbol for message 1 and 2. Then, the GAM indices for the symbol of the NC message is a bijective map \( f_{NC} \), \( m_{NC} = f_{NC}(m_1, m_2) \), e.g. \( m_{NC} = \text{mod}(\mu_1(m_1) + \mu_2(m_2), M) \), where \( \mu_1(\cdot) \) and \( \mu_2(\cdot) \) are mapping functions of choice, and \( \text{mod}(\cdot, \cdot) \) is the modulus function. The decoding of \( m_1 \) and \( m_2 \) includes the baseband signal for those indices, as well as for \( m_{NC} \).

- Design of an elliptical-symmetric GAM-signal constellation, e.g. as \( x_m = r_m(c_{re} \cos(2\pi \varphi m) + ic_{im} \sin(2\pi \varphi m)) \), where \( c_{re} \neq c_{im} \).

- Extend the phyllotaxis packing inspired modulation principle to space-time-coded antenna schemes.
Chapter 7

Matrix Exponential Distribution: Fading and Analysis

In Chapter 4, we gave general and case-specific throughput expressions for ARQ and HARQ, whereas in Chapter 5, we considered analytical throughput optimization of ARQ and HARQ in general terms, but also some particular communication cases of interest. In this chapter, we generalize the channel-specific results in earlier chapters by introducing the matrix exponential distribution for the wireless fading (effective) channel gain/SNR modeling, as well as wireless system performance evaluation. The strength of the proposed ME-distribution framework is the unified matrix-algebraic bottom-up performance evaluation framework, for a wide class of fading channels, which gives very compact performance expressions. We explore the ME-distribution approach in the context of ARQ and HARQ, generalize the results given in Chapters 4 and 5, and develop new analytical tools for the analysis of wireless systems. Moreover, we demonstrate that the introduced ME-distribution approach is able to analyze (the otherwise seemingly untractable problem of) ARQ with ME-distributed signal- and sum-interfering-channel SNRs.

7.1 Motivation and Outline

The literature is rich of wireless channel models, such as Hoyt, log-normal, Nakagami-m, Rayleigh, Ricean, etc., for which the performance of different wireless systems has been studied. Throughput performance analysis of ARQ and HARQ have, as seen in the related work to Chapter 4, mainly focused on Rayleigh and Nakagami-m fading. In Chapter 4, we considered the notion of an effective channel, accounting for the wireless fading channel, signal processing (if available), communication scheme, and some performance metrics. We specialized the analysis towards Rayleigh and Nakagami-m fading. Each combination of wireless channel and (H)ARQ-case, had to be treated on a per-case basis, and gave rise to throughput expressions of different forms. However, it would be desirable to have a structured performance analysis framework that could give throughput expressions on a sim-
ple, compact, and unified form for a large class of wireless channels and different communication cases. We have recognized that the ME-distribution has a very general form that can, e.g., incorporate, fading channels like Rayleigh/Nakagami-$m$ fading channels, signal processing operations like OSTBC/MRC/SDC, but also performance analysis expressions on ME-form in one common unified analytical framework. This chapter proposes the ME-distribution approach for performance analysis of retransmission systems.

In Section 7.2, we review the ME-function and -distribution. In Section 7.3, we motivate the introduction of the ME-distribution for wireless system analysis. We introduce the overall performance analysis framework, the ME-distributed wireless channel model, and some new useful ME-properties in Section 7.4. The throughput performance analysis of ARQ, truncated/persistent-HARQ, ARQ with interference, and IR-HARQ, takes place in Section 7.5. Some illustrative numerical examples are given in Section 7.6. The chapter is summarized and concluded in Section 7.7.

7.2 Preliminaries

We start by reviewing some basic properties of the ME-function and the ME-distribution.

7.2.1 Matrix Exponential

Consider the square complex valued matrix $X \in \mathbb{C}^{d \times d}$, where $d \in \mathbb{N}^+$. Then, the matrix exponential can be defined as

$$e^{\xi X} \triangleq \sum_{k=0}^{\infty} \frac{(\xi X)^k}{k!},$$  \hspace{1cm} (7.1)

where $X^0 \triangleq I$, and $\xi$ is an arbitrary scalar. The ME-function is, e.g., also possible to write as the limit

$$e^{\xi X} = \lim_{k \to \infty} \left( 1 + \frac{\xi X}{k} \right)^k.$$  \hspace{1cm} (7.2)

Using the RHS of (7.1), it is seen that the derivative of the ME-function is

$$\frac{d}{d\xi} e^{\xi X} = X e^{\xi X}.$$  \hspace{1cm} (7.3)

Further, using the RHS of (7.1), it is noted that $X e^{\xi X} = e^{\xi X} X$ commute. The integral of the ME-function, which is a scalar integral in $\xi$ with a matrix parameter $X$, can be expressed as

$$\int_a^b e^{\xi X} \, d\xi = X^{-1} \left( e^{\xi X} - I \right) \big|_a^b,$$  \hspace{1cm} (7.4)
given that $X$ is non-singular. The integral (7.4) is easily proven by using the RHS of (7.1). Note also that $X^{-1}e^{X} = e^{X}X^{-1}$.

A good overview of *Nineteen dubious ways to compute the exponential of a matrix* is found in [ML03]. The ME-function is also surveyed in [Hig08]. More practically, the ME-function is, e.g., implemented in MATLAB®, Mathematica, and Maple. Note that in MATLAB®, the matrix- and scalar-exponential commands differ, and are expm($X$), exp($x$), respectively.

### 7.2.2 Matrix Exponential Distribution

Next, we review the ME-distribution and several well-known properties which are often presented in the literature, [BFT08, AB96, Fac03, AO04, RC13].

The cdf of an ME-distributed r.v. $\Xi$ is commonly written as

$$F_{\Xi}(\xi) = 1 + xe^{\xi}Y^{-1}z, \quad \xi \geq 0,$$

(7.5)

where in general $x \in C^{1 \times d}$, $Y \in C^{d \times d}$, and $z \in C^{d \times 1}$. The only requirement on $x$, $Y$, and $z$ are that $F_{\Xi}(\xi)$ corresponds to a cdf, i.e. non-decreasing, right-continuous, $\lim_{\xi \to 0} F_{\Xi}(\xi) = 0$, and $\lim_{\xi \to \infty} F_{\Xi}(\xi) = 1$. From (7.5), the pdf is found to be

$$f_{\Xi}(\xi) = xe^{\xi}Yz, \quad \xi \geq 0,$$

(7.6)

which is known to correspond to a sum of exponential-polynomial-trigonometric terms [AO04]. The moments, easily derived via partial integration, are

$$E\{\Xi^k\} = (-1)^{k+1}k!xY^{-(k+1)}z.$$

(7.7)

Note that the form of (7.5) and (7.6) differ from the form of the scalar exponential distribution wrt negative signs.\(^2\) The Laplace-Stieltje’s transform (LST) of (7.5), corresponds to the Laplace transform of (7.6), which is

$$F(s) = x(sI - Y)^{-1}z.$$

(7.8)

Eq. (7.8) is also known (as discussed below) to correspond to a ratio of two polynomials expressed in the Laplace variable $s$. It has been shown in [AB96] that the class of rational LSTs is equivalent to the class of ME-distributions. It may be noted that phase-type distributions, introduced by Neuts and treated in detail in [Neu81], have the same form as the ME-distribution, but phase-type distributions have certain parameter constraints and allows for a probabilistic interpretation [Fac03]. Neuts [Neu81] states that the phase-type distribution is dense on $[0, \infty)$, and Ruiz-Castro

\(^1\)Sometimes, the ME-distribution class includes a point-mass at zero. However, continuous real-world wireless channels, as considered here, do not have such property. Therefore, any point-mass at zero is omitted in the following.

\(^2\)The scalar exponential distribution is generally defined to have cdf on the form $F_{\xi}(\xi) = 1 - e^{-\xi}$, and pdf $f_{\xi}(\xi) = ye^{-y}$. The analogous ME-cdf form would be $F_{\Xi}(\xi) = 1 - xe^{-\xi}Yz$, with ME-pdf $f_{\Xi}(\xi) = xe^{-\xi}Yz$. The pdf would, however, have a more complicated LT than (7.8)
in [RC13] extends this statement to the wider class of ME-distributions. However, Neuts [Neu81] also points out that the value of this theorem as an approximation theorem is largely illusory. No general approximation results are, in fact, known.

As discussed after (7.5), the ME-distribution allows for a flexible parameter choice of $x$, $Y$, and $z$. A convenient and often occurring real-valued companion matrix-based parametrization, originally given in [AB96], has been treated in, e.g., [AO04] and [BFT08, Theorem 2.1]. With our notation, this translates to

$$Y \triangleq \begin{bmatrix}
0 & 1 & 0 & \cdots & 0 & 0 \\
0 & 0 & 1 & \ddots & \cdots & 0 \\
0 & 0 & 0 & \ddots & \ddots & 0 \\
\vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & 0 & 1 \\
-y_1 & -y_2 & -y_3 & \cdots & -y_{d-1} & -y_d
\end{bmatrix}, \quad (7.9)$$

or equivalently $Y = S - zy$, where $S$ is a shift matrix of appropriate dimension, and

$$x \triangleq [x_1 \ x_2 \ \ldots \ x_{d-1} \ x_d] \in \mathbb{R}^{1 \times d}, \quad (7.10)$$

$$y \triangleq [y_1 \ y_2 \ \ldots \ y_{d-1} \ y_d] \in \mathbb{R}^{1 \times d}, \quad (7.11)$$

$$z \triangleq [0 \ 0 \ \ldots \ 0 \ 1]^T \in \mathbb{R}^{d \times 1}, \quad (7.12)$$

with the corresponding parametrization of the rational LT,

$$F(s) = \frac{x(s)}{y(s)}, \quad (7.13)$$

$$x(s) = x_ds^{d-1} + x_{d-1}s^{d-2} + \ldots + x_2s^1 + x_1, \quad (7.14)$$

$$y(s) = s^d + y_ds^{d-1} + y_{d-1}s^{d-2} + \ldots + y_2s^1 + y_1. \quad (7.15)$$

Using the final- and initial-value theorem, it can be shown, as e.g. in [LRS16a], that necessary, but not sufficient, conditions for $F(s)$ to correspond to a pdf $f_Z(z)$, without a point mass at zero, are $\deg(x(s)) < \deg(y(s))$, and $x_1 = y_1$.

Two works that inspired us in [LRS16a] to consider the closure of convolutions for the ME-distribution class are [RC13] and [Neu81]. In fact, it is well-known that the class of ME-distributions is closed under many different operations, such as the convolution, maximum and minimum of two r.v.s. An excellent overview of various closure properties for the ME-distribution class is found in [RC13]. For phase-type distributions, which have the same form as ME-distributions, closure properties are given in [Neu81, Section 2.2]. We review those three cases, convolution, maximum and minimum below, and refer the interested reader to the literature for further details.\(^3\)

\(^3\)Convolution (e.g. for MRC, truncated-HARQ, and SDC) and the maximum operator (e.g. for SDC) were used for signal processing and performance analysis in [LRS16a], which motivates reviewing those properties here.
Proposition 7.1. (Convolution of two ME-distributed r.v.s. [RC13, Proposition 3.1]) Let the r.v.s. $\Xi_j, j = \{1, 2\}$ have pdfs $f^{(j)}_\Xi(\xi) = x_je^{\xi Y_j z_j}$. Then, $\Xi = \Xi_1 + \Xi_2$ has the pdf
\[ f_\Xi(\xi) = x e^{\xi Y z}, \quad (7.16) \]
where
\[ x = [x_1 \ 0], \quad (7.17) \]
\[ Y = \begin{bmatrix} Y_1 & z_1 x_2 \\ 0 & Y_2 \end{bmatrix}, \quad (7.18) \]
\[ z = \begin{bmatrix} 0 \\ z_2 \end{bmatrix}. \quad (7.19) \]
In the above, and henceforth, vectors/matrices indicated as $0$ are for notational convenience, with appropriate dimensions given by the problem.

Proof. The proof is reviewed in Appendix 7.A.

Proposition 7.2. (Maximum of two ME-distributed r.v.s. [RC13, Proposition 3.5], [BFT08]) Consider the ME-distributions $F^{(j)}_\Xi(\xi) = 1 + x_je^{\xi Y_j Y_j^{-1} z_j}, \xi \geq 0, j \in \{1, 2\}$. Then, $\Xi = \max\{\Xi_1, \Xi_2\}$ has the ME-distribution
\[ F_\Xi(\xi) = 1 + x e^{\xi Y Y^{-1} z}, \quad (7.20) \]
where
\[ x = [x_1 \otimes x_2 \ x_1 \ x_2], \quad (7.21) \]
\[ Y = \begin{bmatrix} Y_1 \otimes Y_2 & 0 & 0 \\ 0 & Y_1 & 0 \\ 0 & 0 & Y_2 \end{bmatrix}, \quad (7.22) \]
\[ z = \begin{bmatrix} (Y_1^{-1} \oplus Y_2^{-1})(z_1 \otimes z_2) \\ z_1 \\ z_2 \end{bmatrix}. \quad (7.23) \]

Proof. Proof by [RC13, Proposition 3.5].

Proposition 7.3. (Minimum of two ME-distributed r.v.s. [RC13, Proposition 3.6], [BFT08]) Consider the ME-distributions $F^{(j)}_\Xi(\xi) = 1 + x_je^{\xi Y_j Y_j^{-1} z_j}, \xi \geq 0, j \in \{1, 2\}$. Then, $\Xi = \min\{\Xi_1, \Xi_2\}$ has the ME-distribution
\[ F_\Xi(\xi) = 1 + x e^{\xi Y Y^{-1} z}, \quad (7.24) \]
where
\[ x = x_1 \otimes x_2, \quad (7.25) \]
\[ Y = Y_1 \oplus Y_2, \quad (7.26) \]
\[ z = -(Y_1^{-1} \oplus Y_2^{-1})(z_1 \otimes z_2). \quad (7.27) \]
Bivariate pdf  
\[ f_h(h) = \frac{1}{\pi \Omega} e^{-h^2/r + h^2/i} \]  
Amplitude pdf  
\[ f_{|h|}(|h|) = \frac{1}{|h|^2} e^{-|h|^2/\sigma^2} \]  
SNR pdf  
\[ f_G(g) = \frac{1}{S} e^{-g/S} \]  
LT of SNR pdf  
\[ F(s) = 1 + e^{-s/S} \]  
SNR cdf  
\[ F_G(g) = \frac{1}{\Gamma(m)} g^{m-1} e^{-g/S} \]  

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(Unnamed distr.) 2/3</td>
<td>(Unnamed distr.) 2/3</td>
<td>(Unnamed distr.) 2/3</td>
<td>(Unnamed distr.) 2/3</td>
</tr>
<tr>
<td>(Unnamed distr.) 2/3</td>
<td>(Unnamed distr.) 2/3</td>
<td>(Unnamed distr.) 2/3</td>
<td>(Unnamed distr.) 2/3</td>
</tr>
</tbody>
</table>

Table 7.1: Comparison of pdfs (and cdfs) for unprocessed fading wireless channels SNRs. The following notation is used: The instantaneous SNR is \( g \equiv |h|^2 P/\sigma^2 \), where \( |h| \) is the channel amplitude gain, \( P \) is the received power, \( \sigma^2 \) is the receiver noise power. The average SNR is \( S \equiv \mathbb{E} \{G\} \). The complex amplitude gain is \( h \equiv h_r + ih_i \), and \( \Omega \equiv \mathbb{E} \{|H|\} \).

\[ \frac{1}{\Gamma(m)} g^{m-1} e^{-g/S} \]

\[ 1 + e^{-s/S} \]

\[ \frac{1}{\Gamma(m)} g^{m-1} e^{-g/S} \]

\[ 1 + e^{-s/S} \]

Proof. Proof by [RC13, Proposition 3.6].

Good surveys of the class of ME-distributions, its use, applications and properties, are e.g. found in, [BFT08, AB96, AO04, RC13]. Finally note also that the ME-distribution has the same analytical form as the phase type-distribution [Neu81], but the parameters in \( x, Y, \) and \( z \) are, in contrast to the phase type-distribution, not restricted to have a probabilistic interpretation. Hence, much of the known properties in the literature of phase type-distributions carry over to the ME-distribution class.

7.3 ME-distribution in Wireless Communications

In this section, we present some observations\(^4\) that motivate us to consider the ME-distribution as a basis for performance analysis of wireless communication systems with fading.

7.3.1 Unprocessed Wireless Channel SNR

We start with the following example.

Example 7.1. The Nakagami-\( m \) channel is a relatively versatile channel model, with Rayleigh fading as a special case when the Nakagami-\( m \) parameter \( m^N = 1 \), and no fading when \( m^N \to \infty \). The Laplace transform for the (gamma-distributed) SNR pdf is \( F(s) = 1/(1 + sS/m^{N})^{m^N} \), where \( S \) denotes the average SNR of the channel, and \( m^N \geq \frac{1}{2} \). For the special case \( m^N \in \mathbb{N}^+ \), \( F(s) \) is on a rational LT...

---

\(^4\)Some observations where given already in [LRS16a], but here, we give a more structured and detailed treatment.
form, and thus belongs to the ME-distribution class. This example correspond to the GD-channel introduced in Section 3.4.3.

To show the versatility of the ME-distribution, going beyond the simplicity of Ex. 7.1, consider the following due to [AB96].

**Example 7.2.** The pdf $f_{\Xi}(\xi) = (1 + 7^{-2})(1 - \cos(7\xi))e^{-\xi}$ has the rational LT $F(s) = 50/(s^3 + 3s^2 + 52s + 50)$ and is ME-distributed. With the oscillatory decaying nature for the pdf, it is noted that the ME-distribution-form can also capture relatively complex behaviors already with low degree LTs.

**Example 7.3.** In [LRS16a, Sec. IV.F.4], we also proposed modeling the unprocessed channel SNR pdf with an ME-density, $f_{\Omega}(g) = P e^{gQ}$, with a corresponding rational LT $F(s) = p(s)/q(s)$ where the $p(s)$ and $q(s)$ are polynomials. The motivation is that this general LT form have the potential to model (exactly, or approximately) the statistical characteristics of many different fading channel SNRs. This is so since the ME-distribution is dense on $(0, \infty]$, [RC13, Neu81].

---

5This is more clearly seen when writing the LT as $F(s) = (1/\hat{S})m^N/(1/\hat{S} + s)^m$, and then on a rational polynomial form as in (7.13).
Some examples of ME-pdfs, of various shapes and degrees, are shown in Fig. 7.1. Of course, not all SNR pdfs of well-known wireless channel models are in the ME-distribution class, and therefore do not have a rational LT. Examples of non-rational functions are, e.g., $1/\sqrt{1+s}$ and $e^{-s}$, which require polynomials of infinite degrees. Other examples are Rician and log-normal fading, which do not have ME-distributed SNRs, and thus also no rational LTs. Hence, using the ME-distribution as an approximation to model wireless channel SNR fading is an interesting option.

The idea of approximating given pdfs, having non-rational LTs, with ME-pdfs, having rational LTs, has been studied extensively in the literature. An excellent overview of state-of-the-art techniques, and review of related works, for approximating a pdf with phase type- or ME-distributions is given in [Fac03]. A detailed review is outside the scope of this work, but the main principles are generally built on norm minimization, either wrt a pdf or its LT. Not only unprocessed wireless channel SNR pdfs can be approximated with pdfs on ME-distribution-form, but also, if desired, the SNR pdfs after SNR processing. More explicitly, we showed in [LRS16a] that the average number of transmissions (and hence the throughput) of persistent-IR operating in a Rayleigh fading channel could be arbitrarily well approximated with an ME-distribution-form using a truncated continued fraction form. Moreover, in [LRS16a] we mentioned, the possibility of using a continuous least squares approximation in the pdf-domain. However, this is hard to solve explicitly. An alternative idea, also proposed in [LRS16a], was to approximate pdfs, with non-rational LTs, by using a rational Padé approximation in the LT domain. A new idea, briefly mentioned in [LRS16b], is to consider fitting an ME-distribution, or -density, directly to measured fading channel gains. In this way, the channel model would be formulated directly as an ME-distribution. We leave this interesting research opportunity for future research to take on.

In Tab. 7.1, we illustrate mathematical expressions for the proposed ME-distributed wireless channel SNR alongside with Gamma-distributed fading, and exponentially distributed fading SNR. The pdfs, the LTs of the pdfs, and the cdfs, are shown from the middle to the rightmost columns. In the amplitude domain, the familiar Rayleigh and Nakagami-$m$ pdfs are shown. Through variable substitution, we also introduce the corresponding (hitherto unnamed) amplitude pdf for the ME-distribution case (second left column). It is well-known that the Rayleigh distribution can be derived from the bivariate Gaussian distribution. In an analogous manner, using the same variable substitutions, we generalize the ME-distribution to a (hitherto unnamed) bivariate pdf expressed in ME-distribution matrix-, and vector-, parameters (left column). Note that this generalized bivariate-pdf degenerates to the bivariate Gaussian pdf for the scalar case. The ME-distribution generalization are treated further in Section 7.7.

### 7.3.2 Processing and Effective Channel SNR

Below, we illustrate the connection between the ME-distribution and signal processing in wireless communication with four motivating examples. The first two,
receiver-MRC and -SDC, illustrate diversity-based signal processing facilitated by antenna hardware capability only. The other two, OSTBC and spatially-multiplexed zero-forcing (ZF) MIMO (ZF-MIMO), shows signal processing at both transmitter- and receiver-side which also involves special signal-design and -processing.

Example 7.4. (MRC) Consider a receiver with $N_{rx}$ antennas, exponentially distributed SNRs $Z_n, n \in \{1, 2, \ldots, N_{rx}\}$, each with average SNR $S_n$. The LT of the MRC SNR, $Z = \sum_{n=1}^{N_{rx}} Z_n$, is then $F(s) = 1/\prod_{n=1}^{N_{rx}} (1 + sS_n)$, which is on a rational form, and hence corresponds to an ME-distribution.

In SDC, the signal with the greatest SNR is selected. The following example on SDC is considered in [LRS16a].

Example 7.5. (SDC) Consider $N$-fold SDC, with effective SNR $Z = \max(Z_1, Z_2, \ldots, Z_N)$, where the SNRs $Z_n, n \in \{1, 2, \ldots, N\}$ are iid exponentially distributed, with average SNR $S$, and cdfs $F_Z(z) = 1 - e^{-zS^{-1}}$. For this case, the pdf is $f_Z(z) = \frac{d}{dz} F_Z(z)^N = Nf_Z(z)F_Z(z)^{N-1} = NS^{-1}e^{-zS^{-1}}(1 - e^{-zS^{-1}})^{N-1}$. The Laplace transform of $f_Z(z)$ can be written as $F(s) = N!/\prod_{n=1}^{N}(n + sS)$ (the proof is given in Appendix 7.B), which is on a rational form and also represents a special case of an ME-distribution. Interestingly, note that the product-form of $F(s)$ can be interpreted as $N$ convolutions, corresponding to a summation of $N$ iid exponentially distributed r.v.s with SNRs $S/n, n \in \{1, 2, \ldots, N\}$. Thus, SDC of $N$ iid exponentially distributed r.v.s, each with SNR $S$, can be seen as MRC of $N$ iid exponentially distributed r.v.s with SNRs $\{S, S/2, S/3, \ldots, S/N\}$.

Example 7.6. (OSTBC+MRC) We now consider a multi-antenna OSTBC+MRC channel, with $N_{tx} (N_{rx})$ transmit (receive) antennas, diversity order $N = N_{tx}N_{tx}$, and OSTBC code rate $r_{stc}$. We have previously discussed this channel in [LRS14c] and found that the LT is $F(s) = 1/(1 + sS/r_{stc}N_{tx})^N$. Hence, as this is on a rational polynomial form, the effective SNR for OSTBC+MRC in Rayleigh fading channel is also a special case of an ME-distribution. This also corresponds to the GD-channel, while omitting Nakagami-m fading.

7.4 Performance Analysis Framework, (Effective) SNR-model, Tools

Encouraged by the observations reviewed in Section 7.3, we now formalize the overall performance analysis framework, generalize the ME-distributed wireless channel SNR model, and consider some new mathematical tools.
7.4.1 Performance Analysis Framework

The performance analysis framework\(^6\) is shown in Fig. 7.2. At the bottom level, the unprocessed SNR channel r.v. \(G\) is modeled with pdf \(f_G(g) = p e^{gQr}\). At the middle level, a performance analysis system model, accounting for various processing steps in the communication system model of interest, translates the unprocessed channel SNR r.v. \(G\) into an effective channel r.v. \(Z\) with pdf \(f_Z(z) = \tilde{p} e^{z\tilde{Q}\tilde{r}}\). At the top level, a performance expression, for some metric of choice, is derived and expressed in the unprocessed channel SNR parameters, \(p, Q\) (and \(r\)). Alternatively, if only the performance evaluation step is of interest, the performance metric may be directly evaluated and expressed in the effective channel parameters, \(\tilde{p}, \tilde{Q}\) (and \(\tilde{r}\)). This system modeling abstraction is reflected in Fig. 7.3, where a more complex communication system model (on the LHS) has a corresponding effective channel model (on the RHS). The studied communication system may, e.g., range from physical layer performance evaluation of a SISO-system to link- or network-layer performance evaluation of a multi-node MIMO-system. The only requirement is that the effective channel is, or can be approximated as, ME-distributed. If no particular processing of the received signal(s) takes place, we simply set \(f_Z(z) = f_G(g)\). Various mathematical tools, suitable for the ME-distribution, can be used at the different performance analysis levels\(^7\), e.g., to reflect the communication system operation at the middle level, and enable computation of the performance metric at the top level. It should be emphasized that the analysis framework is exemplified here for the univariate ME-distribution case, but may be generalized to a multivariate ME-distribution case, as e.g., in Section 7.5.7. Various performance measures can be expressed in the ME-distribution polynomial pairs, \((p(s), q(s))\), or equivalently, the ME-distribution vector/matrix pairs \((p, Q)\). Here, we consider,\(^6\)

\(^6\)This framework was used in [LRS16a], but here, the overall presentation and details of the framework have been reworked. With some guidance to various sections in [LRS16a], we hope to make this clearer. A full bottom-up perspective was, e.g., considered in [LRS16a, Sec. IV.F.2-4]. There, the unprocessed wireless channel SNR r.v. passed through the system level (diversity) signal processing, and performance metrics, such as throughput, average number of transmissions, and packet loss rate, where determined and expressed directly in the unprocessed wireless channel SNR parameters, \(p, Q\) (or equivalently as \(p, q\)). The top-level performance analysis view, taking only the effective channel r.v. (albeit then assuming \(Z = G\)) as input, was studied in [LRS16a, Sec. IV.D]. The middle level, discussing the effective channel SNR due to various signal processing schemes, communication schemes, hardware configurations (\(N_{rx}\)-MRC, \(N_{tx}\) \times \(N_{tx}\) OSTBC-MRC, \(2 \times 1\) Alamouti-TX diversity, \(N_{rx}\)-SDC), were addressed in [Sec IV.D]. The bottom level, with unprocessed fading channel SNR, was discussed in [LRS16a, Sec. IV.D] for Nakagami-\(m\) and Rayleigh fading as instances of ME-distributed channels, whereas the fully general ME-distributed channel, with \(F(s) = p(s)/q(s)\) were handled in [LRS16a, Sec. IV.F.2-4]. In [LGAZ+16], we used the ME-distribution framework for yet another performance measure, the logarithmic moment-generating-function (log-mgf) (the effective capacity) for (H)ARQ systems.\(^7\)

\(^7\)Various ME-distribution tools were introduced and used in [LRS16a], such as the closure of the convolution (for ME-distributions with different parameters, MRC, multiple transmissions) [LRS16a, Sec. IV.F1], k-fold convolution [LRS16a, Sec. IV.F1], integration (for outage probability) [LRS16a, Sec. IV.F5], derivation (for performance optimization) [LRS16a, Sec. IV.F5], the notion of ME-distribution approximations [LRS16a, Sec. IV.D], and implicitly also the closure of the maximum operation (for SDC) [LRS16a, Sec. IV.E].
7.4. Performance Analysis Framework, (Effective) SNR-model, Tools

7.4.2 Unprocessed ME-distributed Wireless Channel SNR

In [LRS16a], we assumed that the LT of the wireless channel SNR pdf was on the form $F(s) = p(s)/q(s)$, i.e. a ratio of a numerator polynomial and a denominator
This form agrees with the companion-form, (7.9)-(7.15), introduced in [AB96]. However, it is sometimes convenient to express the LT of the fading channel SNR in a product form, such as for Nakagami-$m$ fading channel (7.28). To handle such cases, but also the polynomial case, we consider a more general product-polynomial-form in the following corollary.

**Corollary 7.1. (Unprocessed wireless channel SNR pdf with rational LT on polynomial-product-form).** Let the LT of a unprocessed ME-distributed wireless channel SNR pdf $f_G(g)$ have the form

$$
F(s) = \prod_{j=1}^{J_{MED}} \frac{p_j(s)}{q_j(s)},
$$

(7.28)

$$
p_j(s) \triangleq p_{d_j,j} s^{d_j-1} + p_{d_j-1,j} s^{d_j-2} + \cdots + p_{1,j},
$$

(7.29)

$$
q_j(s) \triangleq s^{d_j} + q_{d_j,k} s^{d_j-1} + q_{d_j-1,j} s^{d_j-2} + \cdots + q_{1,j}.
$$

(7.30)

Then, the ME-distribution pdf of the unprocessed SNR is on the form

$$
f_G(g) = p e^{g Q}, z \geq 0,
$$

(7.31)

where

$$
Q \triangleq \begin{bmatrix}
Q_1 & 0 & \cdots & 0 & 0 \\
0 & Q_2 & \cdots & 0 & 0 \\
0 & 0 & Q_3 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & Q_{J_{MED}-1} \\
0 & 0 & 0 & \cdots & 0 & Q_{J_{MED}}
\end{bmatrix} \in \mathbb{R}^{d \times d},
$$

(7.32)

$$
p \triangleq [p_1 \ 0 \ \cdots \ 0] \in \mathbb{R}^{1 \times d},
$$

(7.33)

$$
r \triangleq e_d \in \mathbb{R}^{d \times 1},
$$

(7.34)

$$
d \triangleq \sum_{j=1}^{J_{MED}} d_j,
$$

(7.35)

$$
P_j \triangleq r_{j-1} P_j \in \mathbb{R}^{d_{j-1} \times d_j}, \ j \in \{2,3,\ldots,J_{MED}\},
$$

(7.36)

$$
Q_j \triangleq S_{d_j} - r_j q_j \in \mathbb{R}^{d_j \times d_j}, \ j \in \{1,2,\ldots,J_{MED}\},
$$

(7.37)

$$
p_j \in \mathbb{R}^{1 \times d_j}, \ j \in \{1,2,\ldots,J_{MED}\},
$$

(7.38)

$$
q_j \in \mathbb{R}^{1 \times d_j}, \ j \in \{1,2,\ldots,J_{MED}\},
$$

(7.39)

In Section 7.3, it was also seen that Rayleigh fading channel with OSTBC, MRC and SDC, are readily expressed in a product-form. In Section 7.5.7, we will see that sum-interference SNR has a rational LT on a product-form.
7.4. Performance Analysis Framework, (Effective) SNR-model, Tools

\[ \dot{r}_j \triangleq \dot{e}_d, \quad j \in \{1, 2, \ldots, J_{\text{MED}}\}, \quad (7.40) \]

and the matrices \(0\) are of appropriate dimensions.

**Proof.** The proof is given in Appendix 7.C.

Thus, Corollary 7.4.2 gives a more flexible form than (7.9)-(7.15), and can be adapted to different channel models, purposes, and scenarios as needed.

**Remark 7.1.** Note that the unprocessed SNR r.v. \( G \) has expectation value \( E\{G\} = S \). Often, it is convenient to consider the r.v. \( G_{\text{um}} \) with unit-mean (um) SNR \( E\{G_{\text{um}}\} = 1 \), ME-distribution parameters \( (p_{\text{um}}, Q_{\text{um}}, r_{\text{um}}) \), and work with \( G = S G_{\text{um}} \). By simple variable substitution, it is straightforward to show that \( f_Z(z) = p e^{z Q_{\text{um}}} r_{\text{um}} \), which implies \( Q = S^{-1} Q_{\text{um}}, \quad p = S^{-1} p_{\text{um}}, \quad \text{and} \quad r = r_{\text{um}} \). In the LT-domain, we have \( F(s) = p(s) / q(s) = p_{\text{um}}(sS) / q_{\text{um}}(sS) \), which is shown by the LT of \( S^{-1} p_{\text{um}} e^{z S^{-1} Q_{\text{um}} r_{\text{um}}} \) with variable substitution \( \xi = z / S \).

### 7.4.3 New ME-distribution Properties

In the following, we develop a number of new closed-form expressions that are useful in the analysis.

**New Expression for the Integral of the ME-density**

In (7.4), we showed the standard approach for integrating ME-functions. The cdf in (7.5) is an example where this integration approach is used, giving a somewhat complicated expression. Moreover, the integration approach in (7.4) also requires that \( Y \) is non-singular. To handle singular matrices, which arise in practical analysis, we would like to put the integral expression on a more compact, easy-to-manipulate, and tidy form. This is the role of the next theorem.

**Theorem 7.1.** (Integration of ME-function on ME-pdf form). The integral of \( f(\xi) = x e^{Y z} \), with intervals \((0, b)\) can be expressed as

\[
\int_{0}^{b} x e^{Y z} d\xi = E_{4, d^t}, \quad (7.41)
\]

where

\[
E \triangleq e^{Y d}, \quad (7.42)
\]

\[
d^t = d + 1, \quad (7.43)
\]

\[
Y^t = \begin{bmatrix} 0 & x \\ 0 & Y \end{bmatrix}. \quad (7.44)
\]

**Proof.** The proof is given in Appendix 7.D.
Example 7.7. If \( f_Z(z) = \text{pe}^{z\mathbf{Q}\mathbf{r}} \) is the pdf of an ME-distributed channel SNR r.v. Then, the cdf can be expressed as \( F_Z(z) = \mathbf{E}_{1,d^1} \), where \( \mathbf{E} \triangleq e^{z\mathbf{Q}^1} \), \( d^1 = d + 1 \), \( \mathbf{Q}^1 = \begin{bmatrix} 0 & \mathbf{P} \\ 0 & 0 \end{bmatrix} \).

Remark 7.2. In [LRS16a, (25),(46)], we introduced the - Singular matrix integration by matrix augmentation - idea, but on a different form than in Theorem 7.1. The form in Theorem 7.1, as will be seen, allows for even simpler expressions and analysis.

New Expression(s) for the Maximum (and Minimum) of ME-distributed r.v.s

In [LRS16a, (35), (36)], and the derivation in Ex. 7.5, we considered \( N \)-branch SDC, i.e. selecting the signal with maximum SNR, for exponentially distributed r.v.s. We also showed that the pdf of the max SNR is an ME-distribution. More generally, the characterization of the maximum (and the minimum) of ME-distributed r.v.s is well-known and have been considered in e.g. [BFT08], [RC13], as well as reviewed in Prop. 7.2 (and Prop. 7.3). The expressions (and the derivations) in Prop. 7.2 and 7.3 are somewhat inconvenient, and may discourage practical use. Using the integration idea in Theorem 7.1, we introduce a simpler, more tractable, expression (and derivation) for the maximum operation in the theorem below.

Theorem 7.2. (Maximum of two ME-distributed r.v.s). Let \( \Xi_j, j \in \{1, 2\} \) be ME-distributed r.v.s with pdf \( f_{\Xi}^{(j)}(\xi) = x_j e^{\xi\mathbf{Y}_j} z_j \), and degree \( d_j \). Then, the CDF of the ME-distribution r.v. \( \Xi = \max(\Xi_1, \Xi_2) \) can be expressed as

\[
F_{\Xi}^{\max}(\xi) = \mathbf{E}_{1, d_1^1 + d_2^1}, \tag{7.45}
\]

where

\[
\mathbf{E} \triangleq e^{\xi(\mathbf{Y}_1^1 \oplus \mathbf{Y}_2^1)}, \tag{7.46}
\]

\[
\mathbf{Y}_j^1 = \begin{bmatrix} 0 & x_j \\ 0 & 0 \end{bmatrix}, \tag{7.47}
\]

Proof. The proof is given in Appendix 7.E.

Remark 7.3. (Maximum of two ME-distributed r.v.s - Alternative expression). It is also seen in Theorem 7.2, that from step (a), the cdf can directly be written

\[
F_{\Xi}^{\max}(\xi) = \mathbf{E}_{1, d_1^1}^{(1)} \mathbf{E}_{1, d_2^1}^{(2)}, \tag{7.48}
\]

where

\[
\mathbf{E}^{(j)} \triangleq e^{\xi\mathbf{Y}_j^1}, \tag{7.49}
\]
\[
\begin{bmatrix}
0 \\
0
\end{bmatrix}
\begin{bmatrix}
Y_1 \\
Y_2
\end{bmatrix}.
\]

While this form is built on the well-known fact that \( F_{Y_1}^{\max}(\xi) F_{Y_2}^{(1)}(\xi) F_{Y_2}^{(2)}(\xi) \), it illustrates the simplicity with which the ME distribution can be used. The extension to more than two r.v.s is straightforward.

**Remark 7.4.** (Minimum of two ME-distributed r.v.s - Alternative expression) The minimum of two r.v.s, i.e. \( \Xi = \min\{\Xi_1, \Xi_2\} \), can be derived analogously to Theorem 7.2 and Remark 7.3, based on \( F_{\Xi}^{\min}(\xi) = 1 - (1 - F_{\Xi}^{(1)}(\xi))(1 - F_{\Xi}^{(2)}(\xi)) \). With \( E_{1, d_1}^{(3)} \), as in Remark 7.3, the cdf is simply

\[
F_{\Xi}^{\min}(\xi) = 1 - \left(1 - E_{1, d_1}^{(1)}\right) \left(1 - E_{1, d_1}^{(2)}\right),
\]

which is easily extended to more than two r.v.s.

Note that while the new expressions for the maximum and the minimum of two ME-distributed r.v.s are convenient and easy to use, they do not, in contrast to Prop. 7.2 and 7.3, illustrate closure properties.

### 7.5 Performance Analysis with ME-distributed Fading Channel

We now turn our attention to performance analysis of various (H)ARQ-cases with ME-distributed fading channels. We first address the effective channel SNR processing, e.g. where the MRC- and SDC-cases in [LRS16a] are generalized. Subsequently, we treat ARQ, truncated-HARQ and persistent-HARQ wrt throughput in detail. The throughput performance of NCBR with ARQ is also characterized. After this, we introduce the bivariate ME-distribution and analyze ARQ where the signal and interfering channels are all ME-distributed. We also explore the outage probability, by means of the bivariate ME-distribution, of \( 2 \times 2 \) SM-MIMO.

#### 7.5.1 Effective Channel SNR Processing

One important insight used in [LRS16a, Sec. IV.F] was that the closure property of the convolution of ME-distributed r.v.s, Prop. 7.1, is a powerful tool for wireless system modeling and performance analysis. We now explore this, as well as other closure properties discussed in Section 7.2.2, [RC13], below. We start by generalizing the MRC case in [LRS16a, Sec. IV.F] for iid to non-identical independent ME-distributed r.v.s. We illustrate the basic ideas below for two r.v.s, but the results are trivially extendable to more than two r.v.s.

**Example 7.8.** (MRC of two non-identical independent ME-distributed r.v.s) Consider two ME-distributed r.v.s \( z_n \), with pdfs \( f_{G_n}(g) = p_n e^{g^2 r_n}, n \in \{1, 2\} \). Then,
the effective SNR, \( Z = G_1 + G_2 \), is also ME-distributed with pdf \( f_Z(z) = \tilde{p} e^{z \tilde{Q} \tilde{r}} \), and parameters

\[
\tilde{Q} = \begin{bmatrix} Q_1 & P_2 \\ 0 & Q_2 \end{bmatrix}, \tag{7.52}
\]

\[
\tilde{p} = \begin{bmatrix} p_1 \\ 0 \end{bmatrix}, \tag{7.53}
\]

\[
\tilde{r} = \begin{bmatrix} 0 \\ r_2 \end{bmatrix}^T, \tag{7.54}
\]

where \( Q_1 = S - r_1 q_1 \), \( Q_2 = S - r_2 q_2 \), and \( P_2 = r_1 p_2 \). The above follows from Corollary 7.4.2.

**Example 7.9. (Sum-interference)** For two ME-distributed interfering signals, with non-identical independent SNR r.v.s. \( G_1 \) and \( G_2 \), the sum-interference is \( Z_I = G_1 + G_2 \). Thus, due to the closure of the convolution for the ME-distribution class, \( Z_I \) is also ME-distributed and has the same form as for the MRC case in Ex. 7.8.

Moving on to SDC, where for the special case with iid exponentially distributed SNRs, a convolution view is applicable.

**Example 7.10. (Effective channel of SDC and Rayleigh fading)** The pdf of the SDC effective channel with (unit-mean) exponentially distributed fading SNRs has, as given by Ex. 7.5, \( LT_F(s) = N!/\prod_{n=1}^{N}(n+s) \), which gives

\[
\tilde{Q}_{um} = \begin{bmatrix} -1 & 1 & \cdots & 0 \\ 0 & -2 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & -N \end{bmatrix}, \tag{7.55}
\]

\[
\tilde{p}_{um} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{7.56}
\]

\[
\tilde{r}_{um} = \begin{bmatrix} 0 & \cdots & 0 & 1 \end{bmatrix}^T. \tag{7.57}
\]

**Example 7.11. (SDC of 2 non-identical independent ME-distributed r.v.s)** For the more general case of selecting the maximum of non-identical independent ME-distributed r.v.s \( Z = \max\{G_1, G_2\} \), Proposition 7.2, or the more compactly formulated Theorem 7.2, gives the cdf.

A different multi-antenna arrangement, compared with OSTBC+MRC which was considered in [LRS16a], is ZF-MIMO.

**Example 7.12. (Zero-forcing MIMO)** For a zero-forcing \( N_{tx} \times N_{tx} \)-antenna MIMO system, \( N_{rx} \geq N_{tx} \), with a complex Gaussian channel matrix with iid entries and average SNR \( S \), the per stream SNR is gamma-distributed, [GHP02], with \( N_{tx} - N_{tx} + 1 \) degrees of freedom and the corresponding LT is \( F(s) = 1/(1+sS)^{N_{ts}-N_{tx}} \). Thus, the ZF-MIMO per stream SNR r.v. is ME-distributed.
Consider, e.g., the mapping \( Z = \ln(1 + G) \), where \( G \) is an ME-distributed SNR r.v., and \( Z \) is the mutual information (MI) for a Gaussian distributed signal in additive white Gaussian noise (AWGN). With this mapping, the effective channel MI r.v. has a non-rational LT and is not ME-distributed. For such cases, an ME-distribution can approximate the distribution of the MI r.v. This is exemplified below with a truncated continued-fraction-based approximation.

**Example 7.13.** With the effective channel MI, \( Z = \ln(1 + G) \), where \( G \) is Rayleigh fading, the MI pdf is \( f_Z(z) = S^{-1} e^{s^{-1} z} \Gamma(-s+1, S^{-1}) \), \( (4.33) \), which has the Laplace transform \( F(s) = e^{S^{-1} s} \Gamma(-s+1, S^{-1}) \), \( (4.34) \). Legendre’s CF of the upper incomplete gamma function is \( e^{z^2} \Gamma(-s, z) = e^{-1} \frac{1}{1+2z} \frac{1}{1+2z^2} \frac{1}{1+2z^3} \cdots \) \( [AS70, 6.5.31] \). We observe from the structure of this CF that it allows \( F(s) \) to be expressed recursively as

\[
F(s) = e^{S^{-1} s} \Gamma(-s+1, S^{-1}) = 1/f_1(s, S),
\]

where \( f_1(s, S) \) can be computed through the backward recurrence

\[
f_j(s, S) = 1 + \frac{s + j - 1}{S^{-1} + \frac{f_{j+1}(s, S)}{f_j(s, S)}}, \quad j = \{1, 2, \ldots \infty\}. \tag{7.59}
\]

Now, let \( f_j(s, S) = \tilde{p}_j(s, S)/\tilde{q}_j(s, S) \), where \( \tilde{p}_j(s, S) \) and \( \tilde{q}_j(s, S) \) are the numerator and denominator polynomials respectively. Rewriting (7.59) on a rational form, we then express the backward recurrence on the matrix form

\[
\begin{bmatrix}
\tilde{p}_j(s, S) \\
\tilde{q}_j(s, S)
\end{bmatrix} = \begin{bmatrix}
S^{-1} + s + j - 1 & j \\
S^{-1} & j
\end{bmatrix} \begin{bmatrix}
\tilde{p}_{j+1}(s, S) \\
\tilde{q}_{j+1}(s, S)
\end{bmatrix}. \tag{7.60}
\]

For a rational form, the recurrence is truncated at the \( J \)th step. Let \( \tilde{p}_{J+1}(s, S) = 1 \) and \( \tilde{q}_{J+1}(s, S) = 0 \), so the final rational Laplace transform is a monomial. However, Legendre’s CF is known to converge faster when \( |z/s| \) is large. Hence, in the studied case, convergence is faster the lower the SNR is. In (7.60) it can be seen that \( \tilde{p}_1(s, S) \) and \( \tilde{q}_1(s, S) \) also involves a factorial term \( J! \). Hence, for large \( J \), as required for higher SNR, overflow and numerical issues may result due to the numerical precision of the matrix exponential function.

**Example 7.14.** With random channel \( H \in \mathbb{C}^{N_r \times N_t} \), the MIMO channel capacity \( [Tel99] \), \( C = \ln \det(\mathbf{I} + S_{N_t}^{-1} \mathbf{H}^\dagger \mathbf{H}) \), corresponds to a scalar r.v. Since the ME-distribution is dense on \( (0, \infty) \), we conjecture that the pdf of the MIMO channel capacity can (in principle) be approximated with an ME-distributed r.v., \( Z \), with pdf \( f_Z(z) = \tilde{p} e^{Qx} \).

**Remark 7.5.** Similarly, as in Ex. 7.13, the LT of the MI-pdf for the \( N \times N \)-MIMO channel derived in Lemma 4.39 can be approximated by a truncated-CF. Based on the rational CF-approximation, the approximate MI-pdf may then be expressed as a ME-distribution. For example, for the \( 2 \times 2 \)-MIMO case (4.39), with \( F(s) = e^{S/2} (\Gamma(-s+1, 2/S) - \Gamma(-s+2, 2/S)^2) \), the backward recurrence then requires 4-states and a \( 4 \times 4 \)-matrix.
Effective Channel Algebra

Perhaps one of the more significant aspects of the proposed framework is that the SNR processing operations may also be mixed, e.g. \( Z = G_1 + \max(G_2, G_3) \) (SDC of branch 2 and 3, and then MRC with branch 1), or \( Z = \max(G_1 + G_2, G_1 + G_3, G_2 + G_3) \) (MRC of the two strongest branches), etc. As long as the r.v.s are ME-distributed, and the operations are closed, the effective channel r.v. \( Z \) will also be ME-distributed. We refer to such operations, with closure properties on ME-distributed r.v.s, as an **Effective channel algebra**. Note that such effective channel algebra naturally handles non-identical ME-distributed r.v.s.

7.5.2 Outage Probability Analysis

After the ME-distributed effective channel SNR (or effective channel MI) has been characterized, the outage probability is a performance metric of interest to consider.

**Theorem 7.3.** (Outage probability for the ME-distributed effective channel) Let the effective channel pdf be \( f_Z(z) = \tilde{p}e^{\tilde{Q}\tilde{r}}, \tilde{p} \in \mathbb{R}^{1 \times \tilde{d}}, \tilde{Q} \in \mathbb{R}^{\tilde{d} \times \tilde{d}}, \tilde{r} = [0 \ldots 0 1]^T \in \mathbb{R}^{\tilde{d} \times 1} \). Then, the outage probability, with decoding threshold \( \Theta \), is

\[
Q_{\text{out}} = E_{1, d^1},
\]

where

\[
E = e^{\Theta^1 \tilde{Q}^1},
\]

\[
d^1 = \tilde{d} + 1,
\]

\[
Q^1 = \begin{bmatrix} 0 & \tilde{p} \\ 0 & \tilde{Q} \end{bmatrix}.
\]

**Proof.** The proof is given in Appendix 7.F.

**Example 7.15.** (OSTBC-MRC with exponential fading SNR order \( \tilde{N} \) – Product form) The effective channel has \( F(s) = 1/(1 + s)^{\tilde{N}} \), with threshold \( \tilde{\Theta} = (e^{\tilde{R}} - 1)/\tilde{S} \). This is handled with Theorem 7.3 and with parameters as in Ex. 7.8.

**Example 7.16.** (Effective OSTBC-channel of order \( \tilde{N} \) – Polynomial form) The Effective OSTBC-channel is as in Ex. 7.12. Since, \( \tilde{q}(s) = (1 + s)^{\tilde{N}} = \sum_{n=0}^{\tilde{N}} \binom{\tilde{N}}{n} s^n \), and \( \tilde{p}(s) = 1 \), we use the companion form (7.9) with \( \tilde{Q} = S - \tilde{r}\tilde{q}, \tilde{p} = [1 0 \ldots 0], \tilde{q} = [\binom{\tilde{N}}{0} \binom{\tilde{N}}{1} \ldots \binom{\tilde{N}}{N-1}], \) and \( \tilde{r} = [0 \ldots 0 1]^T \).

As this form is more complicated than the form in Ex. 7.12, it illustrates the benefit of judiciously choosing the best representation for the problem studied.

**Example 7.17.** (SDC of order \( N \) with exponentially-distributed wireless channel - Product form) Use Theorem 7.7 with \( p, Q, \) and \( r, \) as in Ex. 7.10, with \( \Theta = (e^{\tilde{R}} - 1)/\tilde{S} \).
7.5. Performance Analysis with ME-distributed Fading Channel

A related performance measure to the outage probability, is the outage capacity, $C_{\text{out}}$, defined by

$$\Pr \{ \ln(1 + Z) < C_{\text{out}} \} = Q_{\text{out}},$$

where $Q_{\text{out}}$ is the desired outage probability. For the ME-distributed effective SNR, we get

$$e_1^T e^{(C_{\text{out}} - 1)Q_{\text{out}}} e_2^T = Q_{\text{out}},$$

(7.65)

which is implicit in $C_{\text{out}}$, but can be solved numerically.

As discussed in Section 7.3, the ME-distribution can be used to approximate non-ME-distributed r.v.s. To exemplify this, the MIMO outage probability at high SNR is considered next.

**Example 7.18.** It has been shown in [WG04, Thm. 1] that the LT of the AWGN $N_{\text{rx}} \times N_{\text{tx}}$-MIMO channel capacity (with iid equal power zero-mean complex Gaussian signals) can be written

$$F(s) = B^{-1} \det(G(s)),$$

with matrix $G(s)$ entries

$$g_{ij}(s) = \frac{\lambda_{i+j} e^{-\lambda}}{(i+j)!}$$

and $B = \prod_{n=1}^{\min(N_{\text{rx}}, N_{\text{tx}})} \Gamma(d+n) / \prod_{n=1}^{\min(N_{\text{rx}}, N_{\text{tx}}) + d+1} \frac{1}{s^n}$. This is a rational expression in $s$, and so is the asymptotic $F(s)$.

For example, for $2 \times 2$-MIMO, $F(s) \sim 1/(s-1)(s-2)^2(s-3)\tilde{S}^4$. Thus, the high SNR asymptotic outage probability for $2 \times 2$-MIMO can be expressed as $Q_{\text{out}} \sim e^T \tilde{S}^4 e$, with parameter $Q_{\text{I}}$.

$$Q^1 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 2 & 1 & 0 \\ 0 & 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 0 & 3 \end{bmatrix}.$$  

(7.66)

The above procedure can be extended to higher-order MIMO systems. Determining $F(s)$ with the rational approximation to $g_{ij}(s)$, when $N = N_{\text{rx}} = N_{\text{tx}}$, we observe that $F(s) \sim \tilde{S}^{-N^2} \prod_{n=0}^{N-1} \prod_{i=1}^{\min(N, n+1)} \frac{1}{(s+i)^{N-n}}$. The first order rational approximation of $g_{ij}(s)$ used here, could be further refined with a higher order approximation, albeit at the cost of increased complexity.

7.5.3 ARQ Throughput Analysis

In our analysis of retransmission schemes, we consider ARQ first. Here, ARQ operates under the assumption that a data packet can be decoded if the mutual information, for a transmission, exceeds the information rate of the data packet. ARQ, under such assumptions, has been studied in many works, e.g., [CT01, LRS14c, BS06, SLF08]. ARQ is, as seen below, quite straightforward to analyze for the ME-distributed channel, but, due to its fundamental nature, it is important to include.
Theorem 7.4. (Outage probability and ARQ throughput for the ME-distributed effective channel) Let the effective channel pdf be \( f_Z(z) = \tilde{p}e^{\tilde{Q}z} \). Then, the throughput of ARQ is

\[
T_{\text{ARQ}} = R(1 - E_{1,d^*}),
\]

with \( E_{1,d^*} \) given by Theorem 7.3.

Proof. The proof is given in Appendix 7.G.

Remark 7.6. Note that (7.67), in contrast to [LRS16a, (32)], uses the integration form in Theorem 7.1. It also aligns better with the throughput expression for truncated-HARQ in (7.69) (for which we provide a refined expression of), and is more clearly formulated in the effective channel SNR parameters.

Remark 7.7. Alternatively, the throughput of ARQ can be expressed as \( T_{\text{ARQ}} = R P_{\text{ARQ}} \), where, due to (7.4), the probability of a successful transmission is \( P_{\text{ARQ}} = \int_0^\infty \tilde{p}e^{\tilde{Q}z}dz = -\tilde{p}e^\Theta \tilde{Q}^{-1} \). However, the form in (7.67) aligns better with the throughput expression for truncated-HARQ.

Corollary 7.2. (Optimal ARQ throughput for the ME-distributed effective channel) Let the ARQ throughput be defined as in Theorem 7.4. Then, the optimal throughput can be determined with the function

\[
g_{\Theta}(\Theta) = \frac{1 - E_{1,d^*}}{\Theta \tilde{p} e^{\Theta \tilde{Q}}}.
\]

Proof. The proof is given in Appendix 7.H.

7.5.4 Truncated-HARQ Throughput Analysis

In this section, we consider truncated-HARQ where the number of transmission is limited to \( K \) transmissions. Equipped with the powerful idea of an ME-distributed effective channel, and using Theorem 7.1, the throughput expression can be given on a particularly simple form, as shown next.

Theorem 7.5. (Truncated-HARQ throughput for the ME-distributed effective channel) Let the effective channel density be \( f_Z(z) = \tilde{p}e^{\tilde{Q}z} \). Then, the throughput of truncated-HARQ, with a maximum of \( K \) transmissions and decoding threshold \( \Theta \), is

\[
T_{\text{HARQ}}^K = \frac{R(1 - E_{1,(dK+1)})}{1 + \sum_{k=1}^{K-1} E_{1,(dk+1)}}.
\]
7.5. Performance Analysis with ME-distributed Fading Channel

where

\[ E = e^{\Theta Q \Theta} , \quad (7.70) \]

and the ME-parameters are

\[ p_{K\otimes} = e_1 \in \mathbb{R}^{1 \times d}, \quad (7.71) \]
\[ Q_{K\otimes} = \begin{bmatrix} 0 & p_{K\otimes} \\ 0 & Q_{K\otimes} \end{bmatrix} \in \mathbb{R}^{d \times d}, \quad (7.72) \]
\[ r_{K\otimes} = e_d \in \mathbb{R}^{d \times 1}, \quad (7.73) \]
\[ d^1 = \tilde{d}K + 1, \quad (7.74) \]
\[ p_{K\otimes} = [\tilde{p} \ 0] \in \mathbb{R}^{1 \times dK}, \quad (7.75) \]
\[ Q_{K\otimes} = \begin{bmatrix} \tilde{Q} & \tilde{P} & 0 & \cdots & 0 & 0 \\ 0 & \tilde{Q} & \tilde{P} & \cdots & 0 & 0 \\ 0 & 0 & \tilde{Q} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \tilde{Q} & \tilde{P} \\ 0 & 0 & 0 & \cdots & 0 & \tilde{Q} \end{bmatrix} \in \mathbb{R}^{dK \times dK}, \quad (7.76) \]
\[ r_{K\otimes} = e_{dK} \in \mathbb{R}^{dK \times 1}, \quad (7.77) \]
\[ \tilde{Q} = S - \tilde{r}\tilde{q} \in \mathbb{R}^{\tilde{d} \times \tilde{d}}, \quad (7.78) \]
\[ \tilde{P} = \tilde{r}\tilde{p} \in \mathbb{R}^{\tilde{d} \times \tilde{d}}. \quad (7.79) \]

Proof. The proof is given in Appendix 7.1.

Remark 7.8. Observe that (7.69) is, in contrast to [LRS16a, (30)-(31)] which uses vector convolution [LRS16a, (28)], expressed on a more structured, simpler and more intuitive form thanks to the convolution formulation of (7.76). Another difference is that (7.69) only requires one ME computation, whereas [LRS16a, (30)-(31)] requires two, i.e. in addition to the vector convolution.

Remark 7.9. Similar to ARQ, the optimal throughput for truncated-HARQ can be determined with the auxiliary parametric optimization method in Chapter 5. Then, \( g_\Theta(\Theta) \triangleq f_\Theta(\Theta)/\Theta f'_\Theta(\Theta) \) with \( f_\Theta(\Theta) = (1 + \sum_{k=1}^{\tilde{K}-1} E_{1,(dK+1)})/(1 - E_{1,(dK+1)}) \). While the expression is easy to determine, it is not tidy enough to be given here.

7.5.5 Persistent-HARQ Throughput Analysis

We now consider persistent-HARQ with no upper retransmission limit, i.e. \( K = \infty \). In HARQ, if a data packet can not be decoded, all previous transmissions for the same packets are combined. We assume that if the accumulated mutual
information exceeds the initial information rate $R$, the packet can be correctly decoded. This operation is also the foundation in many modern works on HARQ, e.g., [CT01, LRS14c, WJ10, SKDR13]. For truncated-HARQ, the dimension of the ME-vectors and matrices grows linearly with $K$. Clearly, this is a problem when $K \to \infty$, as for persistent-HARQ. Despite this apparent issue, the ME-distribution framework can, as shown below, handle the persistent-HARQ-case too. In contrast to the ARQ- and truncated-HARQ-performance analysis, the effective channel on a polynomial rational LT form, $F(s) = \tilde{p}(s)/\tilde{q}(s)$, is preferred in the following analysis.

**Theorem 7.6.** (Persistent-HARQ throughput for the ME-distributed effective channel) Let the effective channel pdf $f_Z(z) = \tilde{p}e^{\tilde{Q}z}$ have LT $F(s) = \tilde{p}(s)/\tilde{q}(s)$. Then, the throughput, with decoding threshold $\Theta$, is

$$T_{\text{HARQ}}^\infty = \frac{R}{1 + E_{1,\delta^s}},$$

(7.80)

where

$$E = e^{\Theta Q^I},$$

(7.81)

$$Q^I = \begin{bmatrix} 0 \\ S - \tilde{\mathbf{r}}(\tilde{q} - \tilde{p}) \end{bmatrix}. $$

(7.82)

**Proof.** The proof is given in Appendix 7.J. □

**Remark 7.10.** Observe that (7.80) differs wrt [LRS16a, (21)], in a similar way as discussed in Remark 7.6. In addition, we have now generalized (7.80) to include arbitrary diversity order $N$ in Corollary 7.3. In [LRS16a, Sec. IV.F.4], only $N = 2$ for the ME-distributed channel $F(s) = p(s)/q(s)$ is handled.

For persistent-HARQ, it is preferred to use the polynomial LT-form for the effective channel. This cater for that the average number of transmission metric is easy to express on an ME-form. However, an exception that can also be handled is for diversity, where the LT of the effective channel SNR takes the form $F(s) = (p(s)/q(s))^N$.

**Corollary 7.3.** (Persistent-HARQ throughput for the ME-distributed wireless channel and $N$-fold diversity) Let the effective channel pdf $f_Z(z) = \tilde{p}e^{\tilde{Q}z}$ have LT $F(s) = \tilde{p}(s)/\tilde{q}(s) = (p(s)/q(s))^N, N \in \mathbb{N}^+$. Then, the throughput, with decoding threshold $\Theta$, is

$$T_{\text{HARQ}}^\infty = \frac{R}{1 + E_{1,\delta^s}},$$

(7.83)

where

$$E = e^{\Theta Q^I_{N^s}},$$

(7.84)
and the ME-parameters are

\[ p_{IN}^I = e_1 \in \mathbb{R}^{1 \times d}, \]  
\[ Q_{IN}^I = \begin{bmatrix} 0 & p_{N\otimes} \\ 0 & Q_{N\otimes} \end{bmatrix} \in \mathbb{R}^{d \times d}, \]  
\[ r_{IN}^I = e_{d'} \in \mathbb{R}^{d' \times 1}, \]  
\[ d^d = 1 + \tilde{d}, \]  
\[ p_{N\otimes} = [p \ 0] \in \mathbb{R}^{1 \times \tilde{d}}, \]  
\[ Q_{N\otimes} \triangleq \begin{bmatrix} \tilde{Q}_1 & P & 0 & \cdots & 0 & 0 \\ 0 & \tilde{Q}_2 & P & \cdots & 0 & 0 \\ 0 & 0 & \tilde{Q}_3 & \ddots & \vdots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \tilde{Q}_{N-1} & P \\ 0 & 0 & 0 & \cdots & 0 & \tilde{Q}_N \end{bmatrix} \in \mathbb{R}^{d \times d}, \]  
\[ r_{N\otimes} = e_{d'} \in \mathbb{R}^{d' \times 1}, \]  
\[ \tilde{d} = dN, \]  
\[ P = rp \in \mathbb{R}^{d \times d}, \]  
\[ \tilde{Q}_n = S - r \left( q - pe^{\frac{2\pi in}{N}} \right) \in \mathbb{R}^{d \times d}, n \in \{0, 1, \ldots, N-1\}. \]

**Proof.** The proof is given in Appendix 7.K. \[\square\]

**Example 7.19.** (Persistent-HARQ with diversity order 2) Consider Corollary 7.3 with \( N = 2 \). Then,

\[ Q_{2\otimes}^I = \begin{bmatrix} 0 & p & 0 & \cdots & 0 & 0 \\ 0 & S - r(q - p) & rp & \cdots & 0 & 0 \\ 0 & 0 & S - r(q + p) & \cdots & 0 & 0 \end{bmatrix}, \]  

since

\[ E_{\Theta}^{-1} \left\{ \frac{1}{s} \frac{1}{1 - (p(s)/q(s))^2} \right\} = 1 + E_{\Theta}^{-1} \left\{ \frac{1}{s} \frac{p(s)}{q(s) - p(s) q(s) + p(s)} \right\}. \]

The following example formulate an alternative expression for an OSTBC-MRC Nakagami-\( m \) channel on a simple form.

**Example 7.20.** (Diversity order \( \tilde{N} \) - Alternative form) Consider an effective channel with \( F(s) = \frac{1}{1+s} \tilde{N}, \tilde{N} \in \mathbb{N}^+ \), and threshold \( \Theta \). Then, the throughput can be compactly expressed as

\[ T_{\infty}^{\text{HARQ}} = \frac{R}{E_{\tilde{d}', \tilde{d}'},} \]
where

\[ E = e^{\hat{\Theta}Q}, \]  
(7.97)

and the ME-parameters are

\[ p^I = e_{d^I} \in \mathbb{R}^{1 \times d^I}, \]  
(7.98)

\[ Q^I = S - r^Iq^I - I \in \mathbb{R}^{d^I \times d^I}, \]  
(7.99)

\[ q^I = [-1 \ 1 \ 0 \ldots 0 \ 1] \in \mathbb{R}^{1 \times d^I}, \]  
(7.100)

\[ r^I = e_{d^I} \in \mathbb{R}^{d^I \times 1}, \]  
(7.101)

\[ d^I = \hat{\mathcal{N}}. \]  
(7.102)

Proof. The proof is given in Appendix 7.L.

\[ \square \]

**Corollary 7.4.** (Optimal persistent-HARQ throughput for the ME-distributed effective channel) Let the persistent-HARQ throughput be defined as in Corollary 7.3. Then, the optimal throughput, using the auxiliary parametric optimization method in Chapter 5, can be determined with

\[ g_\Theta(\Theta) = 1 + \frac{E_{1,d^I}}{\Theta p_{N \Theta}e^{\hat{\Theta}Q}r_{N \Theta}}. \]  
(7.103)

Proof. The proof is given in Appendix 7.M.

\[ \square \]

From Section 7.5.3-7.5.5, we conclude that the analysis framework handles the truncated/persistent (H)ARQ schemes for any ME-distributed effective channel.

### 7.5.6 3-phase Network Coded Bidirectional Relaying ARQ

To exemplify another use case of the ME-distribution approach, we consider 3-phase network coded bidirectional relaying (NCBR) [PatL43], [LJS05, LJS06], with ErE ARQ, as schematically depicted in Fig. 7.4. The system has 3 nodes, \( u_1, u_2, u_3 \), where \( u_1 (u_2) \) wants to communicate data \( A (B) \) to \( u_2 (u_1) \), via a relay node \( u_3 \). The relay node \( u_3 \) perform network coding, schematically illustrated as an XOR-operation of \( A \) and \( B \), and the receiving nodes, \( u_1 \) and \( u_2 \), perform corresponding decoding given knowledge of \( A \) and \( B \), respectively. Data \( A \) and \( B \) are exchanged in only three phases thanks to the network coding operation. The network coding approach is assumed to allow for the data rates \( R_{12} \) and \( R_{21} \) to differ. We further assume that each links effective SNR (after any potential processing that may differ between the links) are iid ME-distributed.

**Theorem 7.7.** Consider the 3-phase NCBR model with nodes \( u_1, u_2, u_3 \), in Fig. 7.4. Let each link between a node pair \( \{u_i, u_j\} \), \( \{ij\} = \{13, 32, 23, 31\} \), be characterized by the effective channel SNR r.v. \( Z_{ij} \) with pdf \( f^{(ij)}(z) = p_{ij}e^{\hat{Q}_{ij}r_{ij}} \),
Figure 7.4: 3-phase network coded bidirectional relaying ARQ.

\[
\tilde{p}_{ij} \in \mathbb{R}^{1 \times \tilde{d}_{ij}}, \quad \tilde{Q}_{ij} \in \mathbb{R}^{\tilde{d}_{ij} \times \tilde{d}_{ij}}, \quad \tilde{r}_{ij} = [0 \ldots 0 1]^T \in \mathbb{R}^{\tilde{d}_{ij} \times 1}, \quad \text{and the decoding threshold}\n\]
\[
\Theta^{(12)} = e^{R^{(12)}} - 1, \quad \Theta^{(21)} = e^{R^{(21)}} - 1. \quad \text{Then, the EtE sum-throughput is}\n\]
\[
T_{\text{NCBR}} = \frac{R^{(12)}(1 - Q^{(12)}) + R^{(21)}(1 - Q^{(21)})}{3}, \quad (7.104)\n\]

where

\[
Q^{(ij)} = 1 - (1 - \mathbf{E}^{(3j)}_{1, \tilde{d}_{ij}})(1 - \mathbf{E}^{(3j)}_{1, \tilde{d}_{ij}}), \quad (7.105)\n\]
\[
\mathbf{E}^{(ij)} \triangleq e^{\Theta^{(ij)}}, \quad (7.106)\n\]
\[
\mathbf{Q}^{ij} = \begin{bmatrix} 0 & \tilde{p}_{ij} \\ 0 & \tilde{Q}_{ij} \end{bmatrix}, \quad (7.107)\n\]

\[\text{Proof.}\] The proof is given in Appendix 7.N. \[\square\]

Note that (7.105) corresponds to the new form of the minimum operator given in Remark 7.4. With symmetry, \(f_{Z}^{(13)}(z) = f_{Z}^{(23)}(z), f_{Z}^{(31)}(z) = f_{Z}^{(32)}(z),\) and \(R^{(12)} = R^{(21)},\) the throughput simplifies to \(T_{\text{NCBR}} = 2R^{(13)}(1 - \mathbf{E}^{(13)}_{1, \tilde{d}_{ij}})(1 - \mathbf{E}^{(32)}_{1, \tilde{d}_{ij}})/3.\)

7.5.7 Throughput Analysis of ARQ with ME-distributed Signal and Interferers

In many wireless systems, such as cellular systems, communication under the influence of interference is the normal state of operation, rather than an interference-free state considered so far. This problem, ARQ performance in presence of interference, has not been studied in many works. The authors studied this problem in [LRS14a], where the analytical framework only gave outage probabilities for the case when the desired signal was exponentially distributed, and the interference signals where Nakagami-\(m\) distributed, or vice versa. In this section, we generalize
the problem in [LRS14a], incorporate the ME-distribution framework, and consider the case where the signal of interest and the interferes are all ME-distributed.

System Model of ME-distributed Signal and Interferers

To handle interference, we need to revise the system model somewhat. The probability of successful decoding, with the signal of interest affected by interference, is

\[ P_{\text{ARQ}} = \Pr\{ \ln (1 + Z/(1 + Z_l)) > R \} = \Pr\{ Z \leq \Theta (1 + Z_l) \} \]

where \( Z \) is the SNR r.v. of the signal of interest, \( Z_l \) is the sum-interference SNR, and \( \Theta = e^{R_1 - 1} \). The sum-interference SNR is \( Z_l = \sum_{u=1}^{U} Z_u \), where the interfering users \( u \) have SNR pdf \( f_{Z}^{(u)}(z) = p_u e^{zQ_r} r_u \), each with a rational LT \( F^{(u)}(s) = p_u(s)/q_u(s) \). This scenario was discussed in Ex. 7.9 where the ME-parameters were given for two interfering users. Thus, we have in total, \( f_{Z_l}(z_l) = \prod_{u=1}^{U} p_u e^{zQ_r} r_u \) with rational LT \( F_l(s) = \prod_{u=1}^{U} p_u(s)/q_u(s) \). The signal of interest is also ME-distributed with density \( f_Z(z) = p e^{zQ_r} \). We illustrate the ARQ with interferers system model in Fig. 7.5. Later in this section, we generalize this model with independent signal and sum-interferer fading, to a joint density \( f_{Z_l,Z}(z_l,z) = p e^{zQ_p} P_{12} e^{zQ_r} \), where \( P_{12} \) is a matrix, allowing for dependencies between the signal and sum-interference SNRs.

Throughput Analysis

For the analysis, the following lemma is helpful.

**Lemma 7.1.** (Integral of product of independent ME-densities)

\[
\int_0^\infty x_1 e^{x_1^2 z_1} x_2 e^{x_2^2 z_2} d\xi
\]
7.5. Performance Analysis with ME-distributed Fading Channel

\[- \langle x_1 \otimes x_2 \rangle (Y_1 \otimes Y_2)^{-1} (z_1 \otimes z_2). \tag{7.108}\]

Proof. The proof is given in Appendix 7.O. \qed

Using Lemma 7.1, the throughput of ARQ with ME-distributed signal and sum-interference is now given.

**Theorem 7.8.** (ARQ throughput for iid ME-distributed signal and interferers) Let the signal and sum-interferer be given by the system model. Then, the throughput is

\[
T_{\text{ARQ}} = R \left( p_1 \otimes p \right) \left( (Q_{\text{I}} \oplus \Theta Q)(I \otimes Qe^{-\Theta Q}) \right)^{-1} (r_1 \otimes r). \tag{7.109}\]

Proof. The proof is given in Appendix 7.P. \qed

**Remark 7.11.** Using the integration approach in Theorem 7.1, we could alternatively express the throughput as

\[
T_{\text{ARQ}} = R(1 - Q_{\text{ARQ}}^{\text{Int}}), \text{ where we now get the integral } Q_{\text{ARQ}}^{\text{Int}} = \int_0^\infty p \frac{e^{zi(Q_{\text{I}} - \Theta)}r_1 e^z}{e^{\Theta Q} e^z Q I e^z d_2}. Q^{\text{I}} = [0:0 \ Q], \text{ which can then be determined by means of Lemma 7.1.}
\]

Let us consider some basic examples.

**Example 7.21.** (Exponentially distributed signal SNR and ME-distributed sum-interference) Consider an ARQ system as defined in the system model, where we now assume that the SNR of the signal of interest is exponentially distributed, implying \( p_{\text{um}} = 1, Q_{\text{um}} = -1, r_{\text{um}} = 1 \), and \( \Theta = (e^R - 1)/S \). Then, (7.135), which proves Theorem 7.8, reduces to

\[
P_{\text{ARQ}}^{\text{Int}} = e^{-\Theta} \int_0^\infty p e^{zi(Q_{\text{I}} - \Theta)}r_1 e^z d_2,
\]

which results in

\[
P_{\text{ARQ}}^{\text{Int}} = e^{-\Theta} p_{\text{I}}(\Theta I - Q_{\text{I}})^{-1} r_1
\]

This gives the probability for successful decoding expression, \( P_{\text{ARQ}}^{\text{Int}} = e^{-\Theta} / \prod_{u=1}^U (1 + \Theta S_u/m_u)^{m_u} \), which we showed in [LRS14a, (3)], from \( U \) iid Nakagami-m interferers only, to ME-distributed interfering signals. A special case of (7.110) is for the exponentially distributed interferer with average SNR \( S_I \), implying \( p_I = S_I^{-1}, Q_I = -S_I^{-1}, r_I = 1 \), which gives \( P_{\text{ARQ}}^{\text{Int}} = e^{-\Theta}/(1 + S_I \Theta) \).

**Example 7.22.** (ME-distributed signal SNR and exponentially-distributed interference) Consider an ARQ system as defined in the system model, where we now assume that the SNR of the signal of interest is ME-distributed, whereas the sum-interference is exponentially distributed with average SNR \( S_I \), implying \( p_I = S_I^{-1}, Q_I = -S_I^{-1}, r_I = 1 \), and \( \Theta = e^R - 1 \). The integral (7.135) then reduces to

\[
P_{\text{ARQ}}^{\text{Int}} = -S_I^{-1} \int_0^\infty p Q^{-1} e^{z Q} e^z (Q - I S_I)^{-1} r e^z d_2
\]

This results in

\[
P_{\text{ARQ}}^{\text{Int}} = p Q^{-1} e^{z Q} (I - \Theta S_I Q)^{-1} r.
\]

(7.112)
The throughput expression in Theorem 7.8 can be optimized with respect to the rate, but the resulting expression becomes somewhat complicated. Instead, we treat the throughput optimization problem, for a more general case, in Corollary 7.7.

We note that when the upper integration interval in Lemma 7.1 is finite, the following corollary gives a new, more compact, expression. This integral property may be useful in different situations, e.g. if the integrand below is a pdf and its cdf is sought after.

**Corollary 7.5.** (New integral of independent bivariate ME-form) The integral of
\[ f(\xi) = x_1 e^{\xi Y_1} z_1 x_2 e^{\xi Y_2} z_2 \]
where \( x_j \in \mathbb{R}^{1 \times d_j}, \ Y_j \in \mathbb{R}^{d_j \times d_j}, \ z_j = \begin{bmatrix} 0 & \cdots & 0 & 1 \end{bmatrix}^T \in \mathbb{R}^{d_j \times 1}, \ j = \{1, 2\} \), and with interval \((0, b)\), is
\[ \int_0^b x_1 e^{\xi Y_1} z_1 x_2 e^{\xi Y_2} z_2 \, d\xi = \mathcal{E}_{1, d_1 d_2+1}, \quad (7.113) \]

where
\[ \mathcal{E} \equiv e^{bQ}, \quad (7.114) \]
\[ Q_l = \begin{bmatrix} 0 & x_1 \otimes x_2 \\ 0 & Y_1 \oplus Y_2 \end{bmatrix}, \quad (7.115) \]

*Proof.* The proof is given in Appendix 7.Q. \[ \square \]

In the following remark, we now digress briefly from ARQ and illustrate the challenge with analyzing HARQ with ME-distributed-signal and -interferers.

**Remark 7.12.** We note that for persistent-HARQ with interference, the LT for the SNR pdf is required, which can be expressed as
\[ F(s) = \int_0^\infty e^{-sz} \left( \int_0^\infty (1 + z_l) p_1 e^{z_l Q_l 1_l p_1 e^{s(1+z_l)} Q_l 1_l d_{z_l}} \right) \, dz l \]
\[ = \int_0^\infty (1 + z_l) p_1 e^{z_l Q_l 1_l p(sI - 1 + z_l)Q_l^{-1} r d_{z_l}} \]
\[ = \int_1^\infty p_1 e^{-Q_l e^{sQ_l 1_l p(sI - xQ_l^{-1})} Q_l^{-1} r d_{z_l}} \]
\[ = -\int_1^\infty p_1 e^{-Q_l e^{sQ_l 1_l pQ_l^{-1}(sQ_l^{-1}x^{-1})} Q_l^{-1} r d_{z_l}} \]
\[ = 1 + \int_1^\infty p_1 e^{-Q_l e^{sQ_l 1_l pQ_l^{-1}(sQ_l^{-1})} Q_l^{-1} r d_{z_l}}. \quad (7.116) \]

This illustrates the reason why we do not analyze the throughput of persistent-HARQ with interference. It is due to that the corresponding LT is not easily (if even possible) to put on a rational form.

Next, we explore an alternative, perhaps less straightforward, but more general, solution approach to the ARQ problem with ME-distributed signal and interferers. This section is also motivated by that the more generally formulated bivariate ME-density, and associated analysis, may find applications beyond the studied case.
7.5. Performance Analysis with ME-distributed Fading Channel

Throughput Analysis - Dependent Signal and Interference

In the previous section, we analyzed the throughput of ARQ when the signal and interference SNR r.v.s where independent. This section gives an alternative analysis framework for this case, but, by introducing a more general joint pdf, also enables analysis of ARQ throughput when the signal and interference SNR r.v.s may be dependent. The dependent case may be less common, but could occur if, e.g., both the signal and interferer(s) experience a common varying channel attenuation, or a common diffractive object whilst the receiver is moving. Regardless of this, for performance optimization, this framework offers an attractive solution. Yet another motivation for this section is that this, more general, joint pdf, and associated analytical framework, should be amenable for analyzing other wireless communication problems involving two dependent r.v.s. We start with the generalization of the joint pdf below, and then proceed with the analysis.

**Definition 7.1. (Bivariate ME-distribution)** We define the joint ME-density of the wireless channel SNR r.v.s $(Z_1, Z_2)$, as $f_{Z_1, Z_2}(z_1, z_2) = P_1 e^{z_1 Q_1} P_2 e^{z_2 Q_2} r_2$, $z_1 \geq 0$, $z_2 \geq 0$, where $P_1 \in \mathbb{R}^{d_1 \times d_1}$, $Q_1 \in \mathbb{R}^{d_1 \times d_1}$, $P_{12} \in \mathbb{R}^{d_1 \times d_2}$, $Q_2 \in \mathbb{R}^{d_2 \times d_2}$, $r_2 \in \mathbb{R}^{d_2 \times 1}$. The parameters defining the joint density are, in a similar manner as for the univariate-ME-distribution, assumed selected to have a corresponding bivariate CDF fulfilling necessary characteristics, e.g. $0 \leq F_{Z_j}(z_j) \leq 1$, $F_{Z_1, Z_2}(z_1 = 0, z_2 = 0) = 0$, $F_{Z_1, Z_2}(z_1 \to \infty, z_2 \to \infty) = 1$, and $\frac{\partial}{\partial z_j} F_{Z_1, Z_2}(z_1, z_2) \geq 0$, $j = \{1, 2\}$.

**Remark 7.13.** The bivariate-ME-distribution in Definition 7.1 is a generalization of the case with independent fading SNRs, with pdf $f_{Z_1, Z_2}(z_1, z_2) = P_1 e^{z_1 Q_1} r_1 p_2 e^{z_2 Q_2} r_2$, considered in the previous Section. This is so since $P_{12}$ may have full rank, but $r_1 p_2$ has rank one. The joint density of this form has also been considered in the literature, e.g. in [BN10, BHT08].

**Remark 7.14.** The LT of the bivariate-ME-density in Definition 7.1 is $F(s_1, s_2) = \int_0^\infty \int_0^\infty e^{-s_1 z_1 - s_2 z_2} f_{Z_1, Z_2}(z_1, z_2) d z_1 d z_2 = P_1 (Q_1 - s_1 I_{d_1})^{-1} P_{12} (Q_2 - s_2 I_{d_2})^{-1} r_2$. Expanding the inverses, we get the rational form $F(s_1, s_2) = p(s_1, s_2)/q_1(s_1) q_2(s_2)$. Due to the product of polynomials, $q_1(s_1) q_2(s_2)$, in the denominator, rather than a more general polynomial $q(s_1, s_2)$, it is clear that the considered bivariate ME-distribution is not on the most general form possible.

To find the information-outage probability for ARQ with ME-distributed signal and interferers, the integral of the bivariate ME-density is of interest, and hence considered next.

**Lemma 7.2. (Integral of dependent bivariate ME-form - Sylvester Equation, [WAG14, Lemma 3])** Consider the function $f(\xi) = x_1 e^{\xi Y_1} x_2 e^{\xi Y_2} z_2$. Then, the integral of $f(\xi)$, with intervals $(a, b)$ is

$$
\int_a^b x_1 e^{\xi Y_1} x_2 e^{\xi Y_2} z_2 d \xi = x_1 X z_2,
$$

(7.117)
where $X$ is given by solving Sylvester equation
\[ Y_1 X + XY_2 = e^{bY_1} X_{12} e^{bY_2} - e^{aY_1} X_{12} e^{aY_2}, \]
and $X \triangleq \int_a^b e^{\xi Y_2} X_{12} e^{\xi Y_1} d\xi$.

Proof. The proof is given in Appendix 7.R. \qed

We note that the integral expression in Lemma 7.2 is well-known in control theory [WAG14, Lemma 3], but, to the best of our knowledge, it has not been used in the context of ME-distribution analysis, nor in the context of outage probability analysis for wireless communication systems.

Remark 7.15. When the interval is $(a, b) = (0, \infty)$, it is well-known that no eigenvalues of $Y_1$, and $-Y_2$ can be the same for a unique solution to Sylvester equation. If all eigenvalues are located in the open left half-plane, and $(a, b) = (0, \infty)$, then the RHS of (7.118) is simply $-X_{12}$.

Remark 7.16. Sylvester equation, $Y_1 X + XY_2 = -X_{12}$, can be numerically solved in MATLAB\textsuperscript{®} by the command $X = \text{sylvester}(Y_1, Y_2, -X_{12})$.

Theorem 7.9. (ARQ throughput solution with Sylvester equation) Let the signal of interest, and the sum-interference, SNRs have joint density $f_{Z, I}(z, I) = p_{I} e^{z I} Q_{I} P_{12} e^{z Q_{r}}$. Then, the throughput is
\[ T_{\text{Int}}^{\text{ARQ}} = R p_{I} X r, \]
where $X$ is given by the solution $X$ to the Sylvester equation
\[ Q_{I} X + X \Theta Q_{2} = -\bar{P}_{12}, \]
with
\[ \bar{P}_{12} \triangleq -P_{12} Q_{2}^{-1} e^{\Theta Q_{r}}, \]
and $\Theta = e^{R} - 1$.

Proof. The proof is given in Appendix 7.S. \qed

Eq. (7.119) can be expressed more explicitly, as done in Corollary 7.6 by means of the well-known vectorization solution in the following lemma.

Lemma 7.3. Sylvester equation, $Y_1 X + XY_2 = -X_{12}$, has solution
\[ \text{vec}(X) = - \left( Y_{2}^T \oplus Y_{1} \right)^{-1} \text{vec}(X_{12}). \]

Proof. The proof is given in Appendix 7.T. \qed
Corollary 7.6. (Explicit ARQ throughput solution with Sylvester equation) Let the throughput be defined as in Theorem 7.9. Then, the throughput can be written (more) explicitly as

\[ T_{\text{ARQ int}} = R(r^T \otimes p_1) (\Theta Q^T \oplus Q_t)^{-1} \text{vec}(P_{12} Q^{-1} e^{iQ}) . \] (7.123)

**Proof.** The proof is given in Appendix 7.U.

Inspired by the vectorization approach in Corollary 7.6, an alternative (and potentially useful) integral expression to Lemma 7.2 with \( a = 0 \) can also be given.

**Lemma 7.4.** The integral of \( f(\xi) = x_1 e^{iY_1} X_{12} e^{iY_2} z_2 \) with integration interval \((0, b)\), is

\[ \int_0^b x_1 e^{iY_1} X_{12} e^{iY_2} z_2 \, d\xi = E_{1..}, \text{vec}(X_{12}) , \] (7.124)

where

\[ E \triangleq e^{iQ^T} , \] (7.125)

\[ Q^T = \begin{bmatrix} 0 & z_2^T \otimes x_1 \\ 0 & Y_2^T \oplus Y_1 \end{bmatrix} . \] (7.126)

**Proof.** The proof is given in Appendix 7.V.

Next, we consider throughput optimization when the throughput is based on a solution to a Sylvester equation. To be able to benefit from the auxiliary parameter method [LRS14c], we assume \( \Theta = (e^R - 1)/S \), and work with the unit-mean ME-parameters for the signal of interest.

Corollary 7.7. (ARQ Optimal throughput solution with Sylvester equation) Consider the ARQ throughput expression (7.119) with (7.120). Then, the auxiliary parameter function \( g_0(\Theta) \) in [LRS14c], for the optimal throughput, is

\[ g_0(\Theta) = \frac{p_1 X_r}{\Theta p_1 X'_o X_r} . \] (7.127)

where \( X \) is the solution to the Sylvester equation

\[ Q_1 X + X \Theta Q_{\text{um}} = -\bar{P}_{12} , \] (7.128)

\[ \bar{P}_{12} \triangleq -P_{12} Q^{-1} e^{iQ_{\text{um}}} . \] (7.129)

and with \( X \) known, \( X'_o \) is then the solution to the Sylvester equation

\[ Q_1 X'_o + X'_o \Theta Q_{\text{um}} = -(\bar{P}_{12} + X) Q_{\text{um}} . \] (7.130)

**Proof.** The proof is given in Appendix 7.W.
Remark 7.17. (Integral of product of two MEs – Van Loan method) An alternative way of computing the integral in Lemma 7.2, and then used for Corollary 7.7, is via the method by Van Loan [Loa78]. With interval \((0, b)\), the integral is
\[
\int_0^b x_1 e^{Y_1 X_{12}} e^{Y_2 Z_2} d\xi = x_1 Y_{11}^{-1} Y_{12} Z_2, \tag{7.131}
\]
where
\[
Y_I = \begin{bmatrix}
-X_1 & X_{12} \\
0 & Y_2
\end{bmatrix}, \tag{7.132}
\]
\[
\begin{bmatrix}
Y_{11} & Y_{12} \\
0 & Y_{22}
\end{bmatrix} = e^{b Y_I}. \tag{7.133}
\]
This method has less computational complexity than the vectorization approaches. However, as \(b\) is finite, \(b\) has to be chosen "large enough" to give a good approximation for the case when \(b \to \infty\). The proof is by Van Loan [Loa78].

Remark 7.18. Pertinent to results on the bivariate integrals presented here, if \(X_{12}\) is square and invertible, we may also rewrite the integral as
\[
\int x_1 e^{Y_1 X_{12}} e^{Y_2 Z_2} d\xi = \int x_1 X_{12} e^{Y_1 X_{12}} Y_{12} e^{Y_2 Z_2} d\xi. \tag{7.134}
\]
If the matrix commutation \([Y_2, X_{12} Y_{12}] = 0\) holds, we may further simplify (7.134) to
\[
\int x_1 X_{12} e^{Y_1 X_{12}} Y_{12} e^{Y_2 Z_2} d\xi.
\]

From this section, it can be concluded that an ME-distributed signal under the influence of ME-distributed interference can be handled in an efficient structured manner, whereas a more traditional (non-ME-distribution-based) analysis would be untractable.

7.6 Numerical Results and Discussions

Since the throughput performance depends on the particular ME-distribution polynomials \((p(s), q(s))\), or equivalently the vector parameters \((\tilde{p}, \tilde{Q})\), which characterize the (effective) channel SNR, it makes little sense to demonstrate numerical results for every possible ME-distribution parameter setting. Not the least since their are infinite to the number. Instead, we chose to demonstrate the convenience and viability of the ME-distribution approach with a few examples and corresponding plots with performance curves.
7.6. Numerical Results and Discussions

Figure 7.6: Throughput, $T$, vs. SNR, $S$, of persistent-RR for Rayleigh fading channel, diversity-order $N = 3$, and rate $R = \{2, 4, 6, 8\}$ [b/Hz/s]. Optimal throughput value $T^*$ and optimal rate point $R^*$ vs. mean SNR $S$.

**RR (and ARQ) with 3-branch MRC**

In a first example, we consider RR and 3-fold diversity. In (4.29), we showed that the mean number of transmissions where $\tau_{RR}^\infty, 3 = (2 + \Theta + e^{-\frac{1}{2}\Theta}(\cos(\sqrt{3}\Theta) + \frac{1}{\sqrt{3}} \sin(\sqrt{3}\Theta)))/3$, where $\Theta \triangleq (e^R - 1)/S$. The corresponding throughput expression is $T_{RR}^\infty, 3 = R/\tau_{RR}^\infty, 3$. This is a slightly, but certainly not the worst example of a complicated expression. Nevertheless, it serves our purpose here.

With the proposed ME-distribution approach, the performance can now be conveniently expressed as follows. For 3-fold MRC with exponential distributed channel SNR, we have $F(s) = 1/(1 + s)^3$. This implies $\tilde{p}(s) = 1$, $\tilde{q}(s) = s^3 + 3s^2 + 3s + 1$, and the corresponding ME-distribution vectors/matrices are

$$\tilde{p} = [1 \ 0 \ 0],$$
$$\tilde{q} = [1 \ 3 \ 3],$$
$$\tilde{r} = [0 \ 0 \ 1]^T,$$
$$Q^I = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -3 & -3 \end{bmatrix}.$$

Using Theorem 7.6 with the above ME-distribution parameters, we plot the throughput for $R = \{2, 4, 6, 8\}$ [b/Hz/s], as well as the optimal throughput value and optimal rate point in Fig. 7.6. We could also have used the expressions in Corollary 7.3, or in Ex. 7.20, to achieve the same result. For ARQ, with the same ME-distribution
Persistent-IR

In Fig. 7.7, we consider the throughput of persistent-IR with exponentially fading SNR. The throughput performance based on the truncated CF approach in Ex. 7.13 is compared with throughput performance based on a corresponding Monte Carlo-simulation. We find a good agreement between the respective curves, and when the polynomial degree increases, we see an improved convergence for the truncated CF method. However, the numerical precision of the ME limits the polynomial degree. In Fig. 7.7, the polynomial degrees used are 6 \((S < 0 \text{ dB})\) and 12 \((S \geq 0 \text{ dB})\), respectively.

ARQ with Interferers

In the following, we give a simple hands-on example on how to determine the throughput for ARQ with ME-distributed signal and interferers. The parameters chosen here are somewhat arbitrary, and just for illustration purpose. First, the ME-vectors chosen for the signal of interest are

\[ p = [1 \ 0 \ 0], \quad p = S^{-1} p^{\text{in}}, \]
Figure 7.8: Throughput, $T$, vs. SNR, $S$, and rate, $R = \{2, 4, 6, 8\}$ [b/Hz/s], of ARQ with interference, $S_1 = -13$ dB and $S_2 = -15$ dB.

Figure 7.9: Throughput, $T$, vs. SNR, $S$, and rate, $R = \{2, 4, 6, 8\}$ [b/Hz/s], of ARQ with interference, $S_1 = -13$ dB and $S_2 = 0$ dB.

Figure 7.10: Throughput, $T$, vs. SNR, $S$, and rate, $R = \{2, 4, 6, 8\}$ [b/Hz/s], of ARQ with interference, $S_1 = 0$ dB and $S_2 = 0$ dB.
\[ q = [1 \ 3 \ 3], \quad q = q^{um}, \]
\[ r = [0 \ 0 \ 1]^T, \quad r = r^{um}, \]
\[ Q = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & -3 & -3 \end{bmatrix}, \quad Q = S^{-1}Q^{um}. \]

This corresponds to a pdf with unit mean LT \( F(s) = 1/(1+s)^3 \), i.e. 3-fold diversity with exponentially distributed SNR. The SNR is \( S \), and the threshold is \( \Theta = e^{R-1} \).

We then select the ME-vectors for the first interferer, with SNR \( S_1 \), as
\[ p_{1}^{um} = [1 \ 0], \quad p_1 = S_1^{-1} p_{1}^{um}, \]
\[ q_{1}^{um} = [1 \ 2], \quad q_1 = q_{1}^{um}, \]
\[ r_{1}^{um} = [0 \ 1]^T, \quad r_1 = r_{1}^{um}, \]
\[ Q_{1}^{um} = \begin{bmatrix} 0 & 1 \\ -1 & -2 \end{bmatrix}, \quad Q_1 = S_2^{-1} Q_{1}^{um}. \]

This corresponds to a pdf with unit-mean LT \( F(s) = 1/(1+s)^2 \). Lastly, the ME-vectors for the second interferer, with SNR \( S_2 \), are here selected as
\[ p_{2}^{um} = [2 \ 0], \quad p_2 = S_2^{-1} p_{2}^{um}, \]
\[ q_{2}^{um} = [2 \ 3], \quad q_2 = q_{2}^{um}, \]
\[ r_{2}^{um} = [0 \ 1]^T, \quad r_2 = r_{2}^{um}, \]
\[ Q_{2}^{um} = \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix}, \quad Q_2 = S_2^{-1} Q_{2}^{um}. \]

This corresponds to a pdf with unit-mean LT \( F(s) = 2/(1+s)(2+s) \), incidentally the same as for 2-branch SDC. The ME-distribution vectors for the sum-interference case are then
\[ p_I = \begin{bmatrix} p_1 & 0 \end{bmatrix}, \]
\[ r_I = \begin{bmatrix} 0 & r_2^T \end{bmatrix}^T, \]
\[ Q_I = \begin{bmatrix} Q_1 & P_2 \\ 0 & Q_2 \end{bmatrix}, \]
\[ P_2 = r_1 p_2. \]

In Fig. 7.8-7.10, we plot the throughput vs. SNR for ARQ with interference together with an interference-free system. As expected, when the interference SNR levels \( S_1 \) and \( S_2 \) increases, the overall throughput is reduced. Note that we could also have used Theorem 7.9, including solving a Sylvester equation, where \( P_{12} = r_1 p \), to compute the throughput.
7.7 Summary and Conclusions

In this chapter, we proposed to use the ME-distributions for modeling fading channel SNRs. The broad class of ME-distributions, which is dense on the positive line, incorporated exponential and Nakagami-\(m\) distributed fading channel SNRs as special cases. We saw that many standard signal processing techniques, like MRC and SDC, gave rise to an ME-distributed effective channel SNR. We referred to this as ME-distribution based effective channel algebra. Moreover, the performance measure (in this case the throughput) could be easily, and directly, expressed in the ME-distribution parameters describing the fading ME-distributed channel SNR. Powerful closed-form throughput expressions for ARQ and HARQ with ME-distributed (effective) channel SNRs were derived. We similarly gave closed-form throughput expressions for ARQ with spatially uncorrelated (and correlated) ME-distributed (ME-distributed related expressions of) own signal and interferers. Besides analyzing several (H)ARQ-cases, new tools were derived (integration), or refined (max and min operation), and new ME-distribution-related fading channel models were introduced. We also exemplified the ME-distribution use in the context of 3-phase NCBR and SM-MIMO. As for extending the proposed framework, it is straightforward to use other performance measures, e.g. the effective capacity as in Chapter 8. The ME-distribution, and the ME-distribution framework, could also have new applications in related areas, e.g. in information theory (computing the entropy of the ME-distributed r.v.), or to describe signals expressed on complicated mathematically forms more compactly than classically possible. While we have already proposed, and to some extent studied, such directions in \[LRS16b\], it has been omitted here as it goes beyond a strict (H)ARQ performance analysis.
7.8 Appendices

7.A Proof of Proposition 7.1

Proof.
\[ x_1 e^{\xi Y_1} z_1 * x_2 e^{\xi Y_2} z_2 = \mathcal{L}^{-1}_{-1} \left( x_1 (sI - Y_1)^{-1} z_1 \cdot x_2 (sI - Y_2)^{-1} z_2 \right) \]
\[ = \mathcal{L}^{-1}_{-1} \left( \begin{bmatrix} x_1 & 0 \\ Y_1 - sI & z_1 x_2 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ z_2 \end{bmatrix} \right) \]
\[ = \begin{bmatrix} x_1 \\ 0 \end{bmatrix} e^{\xi Y_1} \begin{bmatrix} 0 \\ z_1 x_2 \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ z_2 \end{bmatrix}. \]

Proof by [RC13, Proposition 3.1].

7.B Proof of Example 7.5

Proof. The Laplace transform is
\[ F(s) = \int_0^\infty e^{-sz} \left( NS^{-1}e^{-zS^{-1}}(1 - e^{-zS^{-1}})^{N-1} \right) \, dz \]
\[ = \frac{e^{-z(S^{-1}+s)}}{S^{-1}+s} \frac{NS^{-1}(1 - e^{-zS^{-1}})^{N-1}}{0 + \frac{NS^{-1}}{S^{-1}+s}} \]
\[ \times \int_0^\infty e^{-sz} \left( (N-1)S^{-1}e^{-2zS^{-1}}(1 - e^{-zS^{-1}})^{N-2} \right) \, dz \]
\[ \overset{(a)}{=} \frac{N!}{\prod_{n=1}^N (n + sS)} = \frac{1}{\prod_{n=1}^N (1 + sS/n)}, \]
where partial integration is used repeatedly in step (a).

7.C Proof of Corollary 7.4.2

Proof. Eq. (7.27) is a product of rational LTs. This corresponds to a rational LT (7.12), with parameters (7.8)-(7.12), and convolution operations as in Prop. 7.1.

7.D Proof of Theorem 7.1

Proof. Integration corresponds to convolution with a step function that has LT $1/s$. Using Prop. 7.1 gives
\[ \int_0^b \Sigma_{\xi}^{-1} \left\{ \frac{x(s)}{y(s)} \right\} d\xi = \Sigma_b^{-1} \left\{ \frac{1}{s} \frac{x(s)}{y(s)} \right\} = e_1^T e^{\lambda Y_1} e_d = E_{1,d}. \]
7.8. Appendices

7.E Proof of Theorem 7.2

Proof.

\[ F_{\text{max}}(\xi) = F_{\text{max}}^{(1)}(\xi) F_{\text{max}}^{(2)}(\xi) = \left( \int_0^\xi x_1 e^{a_1 Y_1 z_1} du \right) \left( \int_0^\xi x_2 e^{a_2 Y_2 z_2} du \right) \]

\[ \stackrel{(a)}{=} \left( e_1^{(1)} T e_1^{(1)} \right) \left( e_1^{(2)} e_1^{(1)} \right) \left( e_2^{(1)} T e_2^{(2)} \right) \]

\[ \stackrel{(b)}{=} \left( e_1^{(1)} \otimes e_1^{(2)} \right)^T e_1^{(1)} e_1^{(1)} \left( e_2^{(1)} \otimes e_2^{(2)} \right) \]

\[ = E_{1,v_1+d_2}, \quad E \triangleq e^{\xi(Y_1 \oplus Y_2)}, \]

where we used the integration idea in step (a), and the Kronecker product identities

\[ (X_1 \otimes Y_1)(X_2 \otimes Y_2) = (X_1 X_2) \otimes (Y_1 Y_2), \quad X \oplus Y = X \otimes I_n + I_m \otimes Y, \quad \text{and} \quad e^{X \otimes Y} = e^X \otimes e^Y, \]

with rearrangement in step (b).

7.F Proof of Theorem 7.3

Proof. From Theorem 7.1, the outage probability can be directly computed as

\[ Q_{\text{out}} = P \{ Z \leq \Theta \} = \int_0^\Theta \tilde{p} e^{z \Theta Q} dz = e^{(1)} e^{(1)} \Theta \tilde{Q} d_2 = E_{1,v_1+d_2}. \]

7.G Proof of Theorem 7.4

Proof. The throughput is simply \( T_{ARQ} = R(1 - Q_{out}) \), and we then use the outage probability in Theorem 7.3.

7.H Proof of Corollary 7.2

Proof. In [LRS14c], we define \( g_\Theta(\Theta) \triangleq f_\Theta(\Theta)/\Theta f_\Theta'(\Theta) \). We have \( f_\Theta(\Theta) = 1/(1 - E_{1,v_1+d_2}) \). It is also noted that \( \frac{d}{d\Theta} E_{1,v_1+d_2} = \frac{d}{d\Theta} \int_0^\Theta \tilde{p} e^{z \Theta Q} dz = \tilde{p} e^{\Theta \tilde{Q}}. \)

7.I Proof of Theorem 7.5

Proof. The throughput for truncated-HARQ, expressed on a LT-form rather than outage-probability form, is

\[ T_{HARQ}^K = R \frac{1 - L_\Theta^{-1} \{ s^{-1} F(s)^K \}}{1 + L_\Theta^{-1} \{ \sum_{k=1}^{K-1} s^{-1} F(s)^k \}}. \]

For the numerator, the K-fold convolution is needed, whereas for the denominator, the k-fold convolutions, \( k \in \{1, 2, \ldots, K-1\} \), are needed. Thanks to the upper triangular block-structure of (7.76), we propose to compute all required convolutions
at the same time when the $K$-fold convolution is computed. To see this, consider first $e^{X_1} = Y_1$, where $X_1 = X_{11}$, and $Y_1 = Y_{11}$. Then, extend the $X$-matrix to $X_2 = \begin{bmatrix} X_{11} & X_{12} \\ 0 & X_{22} \end{bmatrix}$. Then, the $Y$-matrix is $Y = \begin{bmatrix} Y_{11} & Y_{12} \\ 0 & Y_{22} \end{bmatrix}$. Hence, by computing $e^{X_2}$, one gets $e^{X_1} = Y_1 = Y_{11}$ at the same time. By induction, this procedure can be extended to $K$-fold convolution.

7.J Proof of Theorem 7.6

Proof. The average number of transmissions is

$$\mathcal{L}_\Theta^{-1} \left\{ \frac{1}{s} \frac{1}{1 - F(s)} \right\} = \mathcal{L}_\Theta^{-1} \left\{ \frac{1}{s} \frac{1}{1 - \tilde{p}(s)/\tilde{q}(s)} \right\}$$

$$= 1 + \mathcal{L}_\Theta^{-1} \left\{ \frac{1}{s} \tilde{p}(s)/\tilde{q}(s) - \tilde{p}(s) \right\}$$

$$= 1 + e_1^T e^{\Theta Q} e_d$$

$$= 1 + E_{1,d'}, \quad E \triangleq e^{\Theta Q}.$$ 

7.K Proof of Corollary 7.3

Proof. The average number of transmissions is

$$\mathcal{L}_\Theta^{-1} \left\{ \frac{1}{s} \frac{1}{1 - F(s)} \right\} = 1 + \mathcal{L}_\Theta^{-1} \left\{ \frac{1}{s} \frac{p(s)^N}{q(s)^N - p(s)^N} \right\}$$

$$\overset{(a)}{=} 1 + \mathcal{L}_\Theta^{-1} \left\{ \frac{1}{s} \prod_{n=0}^{N-1} \frac{p(s)}{q(s) - p(s) e^{2\pi i n/N}} \right\}$$

$$= 1 + e_1^T e^{\Theta Q N} e_d$$

$$= 1 + E_{1,d'}, \quad E \triangleq e^{\Theta Q N}.$$ 

Note that the expression at step (a) is expressed on a product form, which suggests the use of Prop. 7.1 (Convolution).

7.L Proof of Example 7.20

Proof. The average number of transmissions is

$$\mathcal{L}_\Theta^{-1} \left\{ \frac{1}{s} \frac{1}{1 - F(s)} \right\} = \mathcal{L}_\Theta^{-1} \left\{ \frac{(1 + s)^N}{s((1 + s)^N - 1)} \right\}$$

$$\overset{(a)}{=} e^{-\Theta} \mathcal{L}_\Theta^{-1} \left\{ \frac{s^N}{s^{N+1} - s^N - s + 1} \right\},$$
where we simplified the numerator and denominator by using the frequency shift property of the Laplace transform, \( L^{-1}\{G(s + x)\} = e^{-x}L^{-1}\{G(s)\} \) in step (a). We then express the rational LT on the ME-distribution form.

7.M Proof of Corollary 7.4

Proof. We have \( f_\Theta(\Theta) = 1 + E_{1,d}\). In [LRS14c] we defined \( g_\Theta(\Theta) \triangleq f_\Theta(\Theta)/\Theta f'_\Theta(\Theta) \).

We also note that\( f'_\Theta(\Theta) = \frac{d}{d\Theta} \Theta 0 p_N \ast e^z Q_N \ast r_N \ast dz = p_N \ast e^z Q_N \ast r_N \ast dz \).

7.N Proof of Theorem 7.7

Proof. The ET E outage probability is
\[
Q_{ij} = 1 - P\{\ln(1 + \min(Z_{i3}, Z_{3j})) > R_{ij}\}
= 1 - P\{\min(Z_{i3}, Z_{3j}) > \Theta_{ij}\}
= 1 - P\{Z_{i3} > \Theta_{ij}\}P\{Z_{3j} < \Theta_{ij}\}
= 1 - (1 - P\{Z_{i3} < \Theta_{ij}\})(1 - P\{Z_{3j} < \Theta_{ij}\})
= 1 - (1 - E_{1,d\mid d\mid}^{(i)})(1 - E_{1,d\mid d\mid}^{(j)}),
\]
with \( E^{(ij)} \) given by (7.106).

7.O Proof of Lemma 7.1

Proof. Using the same properties as in Theorem 7.2, the integral is computed as
\[
\int_0^\infty x_1 e^{Y_1} z_1 x_2 e^{Y_2} z_2 d\xi = \int_0^\infty x_1 e^{Y_1} z_1 \otimes x_2 e^{Y_2} z_2 d\xi
= \int_0^\infty (x_1 e^{Y_1} \otimes x_2 e^{Y_2}) (z_1 \otimes z_2) d\xi
= \int_0^\infty (x_1 \otimes x_2) (e^{Y_1} \otimes e^{Y_2}) (z_1 \otimes z_2) d\xi
= \int_0^\infty (x_1 \otimes x_2) (e^{Y_1 \oplus Y_2}) (z_1 \otimes z_2) d\xi
= -(x_1 \otimes x_2) (Y_1 \oplus Y_2)^{-1} (z_1 \otimes z_2).
\]

7.P Proof of Theorem 7.8

Proof. The throughput is \( T_{\text{ARQ}} = R P_{\text{ARQ}} \), where \( P_{\text{ARQ}} = P\{Z > \Theta(1 + Z_1)\} \) is determined via the integral
\[
P_{\text{ARQ}} = \int_0^\infty \int_0^\infty p_1 e^{z_1} Q_{1} r_1 p e^{z} Q r d z_1 d z.
\]
\[
E = -\int_0^\infty pIe^{zQ}e^{z\Theta Q}d\xi_1
\]  
(7.135)

\[ (p_1 \otimes (pQ^{-1}e^{\Theta Q})) (Q_1 \oplus \Theta Q)^{-1} (r_1 \otimes r) \]

\[ (p_1 \otimes p) (I \otimes Q^{-1}e^{\Theta Q}) (Q_1 \oplus \Theta Q)^{-1} (r_1 \otimes r) \]

\[ (p_1 \otimes p) (I \otimes Q^{-1}e^{\Theta Q})^{-1} (Q_1 \oplus \Theta Q)^{-1} (r_1 \otimes r) \]

\[ (p_1 \otimes p) ((Q_1 \oplus \Theta Q)(I \otimes Q^{-1}e^{\Theta Q}))^{-1} (r_1 \otimes r), \]

where Lemma 7.1 is used in step (a), and the identities \((X_1 \otimes Y_1)(X_2 \otimes Y_2) = (X_1X_2) \otimes (Y_1Y_2), (X \otimes Y)^{-1} = (X^{-1} \otimes Y^{-1}), \) and \(X^{-1}Y^{-1} = (YX)^{-1}, \) are used in step (b)-(d), respectively.

**7.Q Proof of Corollary 7.5**

Proof. The integral is

\[
\int_0^b x_1 e^{\xi Y_1} x_2 e^{\xi Y_2} d\xi = \int_0^b (x_1 \otimes x_2) (e^{\xi (Y_1 \oplus Y_2)}) (z_1 \otimes z_2) d\xi = e_1 e^{Q^t} e_d d_{z_1} = E_1 d_{z_1},
\]

where the rearrangements in Lemma 7.1 is combined with Theorem 7.1.

**7.R Proof of Lemma 7.2**

Proof. It is straightforward to check that \(X \triangleq \int_0^b e^{\xi Y_1} x_1 e^{\xi Y_2} d\xi \) fulfills Sylvester equation (7.120), just by inserting the former expression in the latter.

**7.S Proof of Theorem 7.9**

Proof. The throughput is \(T_{\text{ARQ}}^{\text{Int}} = R P_{\text{ARQ}}^{\text{Int}}, \) where, analogously to Theorem 7.8, the decoding probability is

\[
P_{\text{Int}}^{\text{ARQ}} = \int_0^\infty \int_0^\infty pIe^{zQ}P_{12}e^{z\Theta Q}d\xi_1 dz_1 = -\int_0^\infty pIe^{zQ}P_{12}Q^{-1}e^{z(1+\Theta)Q}d\xi_1
\]

\[ = \int_0^\infty pIe^{zQ} \tilde{P}_{12}e^{z\Theta Q}d\xi_1, \quad \tilde{P}_{12} \triangleq -P_{12}Q^{-1}e^{\Theta Q},
\]

\[ = p_1Xr, \quad X \triangleq \int_0^\infty e^{zQ} \tilde{P}_{12}e^{z\Theta Q}d\xi_1.
\]

Based on Lemma 7.2, we then solve for \(X \) in the Sylvester equation (7.120).
7.7 Proof of Lemma 7.3

Proof. Vectorization of the LHS gives $Y_1X + XY_2 = (Y_T \oplus Y_1) \text{vec}(X)$, and then $\text{vec}(X)$ is solved for. \qed

7.8 Proof of Corollary 7.6

Proof.

$$\mathcal{P}_{\text{ARQ Int}} = p_I X r$$

\begin{align*}
\overset{(a)}{=} & \text{vec}(p_I X r) \\
\overset{(b)}{=} & (r^T \otimes p_I) \text{vec}(X) \\
\overset{(c)}{=} & -(r^T \otimes p_I) (\Theta Q^T \oplus Q_1)^{-1} \text{vec} \left( \hat{P}_{12} \right).
\end{align*}

The scalar is vectorized in step (a), the vectorization identity $X^T \otimes X \text{vec}(X)$ is used in step (b), and Lemma 7.3 is invoked in step (c). An alternative, more direct, proof is to vectorize the integral in the proof of Theorem 7.9 directly.

$$\text{vec} \left( \int_0^\infty p_I e^{z^T Q \hat{P}_{12} e^{z^T Q} d z} \right) = (r^T \otimes p_I) \left( \int_0^\infty e^{z^T Q \hat{P}_{12} e^{z^T Q} d z} \right) \text{vec} \left( \hat{P}_{12} \right)$$

and in the last steps, Theorem 7.1 is used. \qed

7.9 Proof of Lemma 7.4

Proof. From the proof of Corollary 7.6, we see that the integral can be written

$$\int_0^b (z_2^T \otimes x_1) \left( e^{z^T (Y_T \oplus Y_1)} \right) \text{vec}(X_{12}) d \xi = e_1^T e^Q \text{vec}(X_{12})$$

and in the last steps, Theorem 7.1 is used. \qed

7.10 Proof of Corollary 7.7

Proof. Taking the implicit derivative of (7.128) wrt $\Theta$ gives $Q_1 X_\Theta' + X_\Theta' Q_{\text{um}} + X Q_{\text{um}} = -\hat{P}_{12} Q_{\text{um}}$, which rearranged gives (7.130). We then have $\mathcal{P}_{\text{ARQ Int}} = p_I X r$ and $\frac{d}{d \Theta} \mathcal{P}_{\text{ARQ Int}} = p_I X_\Theta r$ which is inserted in the definition for $g_\Theta(\Theta)$. \qed
7.X Proof of Example 7.18

Proof. The entries are expressed as

\[ g_{ij}(s) = \int_0^\infty \frac{\lambda^{i+j+d} e^{-\lambda}}{(1 + S\lambda)^s} d\lambda \]

\[ \stackrel{(a)}{=} \frac{1}{\Gamma(s)} \int_0^\infty \int_0^\infty u^{s-1} e^{-u(1+\lambda)} \lambda^{i+j+d} e^{-\lambda} d\lambda du \]

\[ = \frac{1}{\Gamma(s)} \int_0^\infty \int_0^\infty u^{s-1} e^{-u} \lambda^{i+j+d} e^{-\lambda(1+S\lambda)} d\lambda du \]

\[ \stackrel{(b)}{=} \frac{1}{\Gamma(s)} \int_0^\infty \int_0^\infty \frac{u^{s-1} e^{-u}}{(1 + S\lambda)^{i+j+d+1}} \lambda^{i+j+d} e^{-\lambda} dv du \]

\[ = \frac{(i+j+d)!}{\Gamma(s)} \int_0^\infty \frac{u^{s-1} e^{-u}}{(1 + S\lambda)^{i+j+d+1}} dv \]

\[ \sim \frac{(i+j+d)!}{S^{i+j+d+1}\Gamma(s)} \int_0^\infty u^{s-i-j-d-2} e^{-u} du \]

\[ = \frac{(i+j+d)!\Gamma(s - i - j - d - 1)}{S^{i+j+d+1}\Gamma(s)} \]

\[ \stackrel{(c)}{=} \frac{(i+j+d)!}{S^{i+j+d+1}} \prod_{n=1}^{i+j+d+1} \frac{1}{s - n} \]

where an integral representation, a variable substitution, and \( \Gamma(1 + x) = x\Gamma(x) \) were used in step (a), (b), and (c). \( \square \)
Chapter 8

Effective Capacity Analysis of General Retransmission Schemes

We considered basic (H)ARQ throughput analysis in Chapter 4, and parametric (H)ARQ throughput optimization in Chapter 5. Chapter 6 gave new modulation design(s) that provided support to the AWGN channel capacity based MI-model used in those two chapters. In Chapter 7, we introduced the ME-distributed fading (effective) channel and the related performance analysis framework. In this chapter, we go beyond the throughput measure treated in the previous chapters, and consider a more general, QoS-considerate, performance measure, the effective capacity. We derive and present a new, unified, effective capacity performance analysis approach, built on a recurrence relation formulation, that handles any retransmission system with memory, multiple transmission modes, and multiple reward rates per transmissions. Moreover, effective capacity analysis and new closed-form expressions are derived for numerous special (H)ARQ-cases, for example persistent/truncated-(H)ARQ in terms of the ME-distributed fading channel SNR, NC-ARQ in Rayleigh fading, and more. We believe that the proposed performance analysis framework enables analysis of a substantial number of important (H)ARQ-related systems with respect to the effective capacity.

8.1 Motivation and Outline

Traditionally, (H)ARQ systems are (nearly) always analyzed with respect to the throughput performance measure, e.g in [BS06, LRS16a]. Whilst giving a coarse indication of the experienced service, this performance measure is not completely without shortcomings. For some data services, such as video-streaming, QoS-guarantees in terms of bounded delays are often more desirable. An alternative performance metric for (queueing) systems with varying service rates, e.g., data transmission over fading channels, and delay targets, is the concept of effective capacity. This metric, introduced in [WN03], was inspired by the large deviations principle and the concept of effective bandwidth, [CT95, Kel96]. The objective of
the concept of effective capacity is to quantify the maximum sustainable throughput under stochastic QoS-guarantees with probabilistic delay limits and varying server rate. However, no works (prior [LGAZ+16]) have analyzed and given the exact effective capacity of persistent/truncated-HARQ and NC-ARQ, and no structured effective capacity methodology for handling the general class of multi-transmission, rate increments, and communication modes has been proposed. Moreover, closed-form effective capacity expression are rare, and analytical effective capacity optimization results are not known, and existing studies are (almost) exclusively limited to Rayleigh fading.

The chapter is organized as follows. In Section 8.2, we review the effective capacity performance measure. A system model for the proposed performance analysis framework is described in Section 8.3. In Section 8.4, we first derive general effective capacity expression(s), and then specialize to truncated/persistent-HARQ wrt the three physical layer system model levels in Section 3.4, as well as the ME-distributed channel in Chapter 7. Other examples of treated cases include NC-ARQ, and two-mode ARQ. Numerical and simulation results are presented in Section 8.5. In Section 8.6, we summarize and conclude. Apart from giving the analysis, and expressions, in terms of the three system model levels, decoding failure probability - based model (Section 3.4.1), effective Channel pdf - based model, (Section 3.4.2), and wireless communication scenario - based model (Section 3.4.3), we also develop an even more general transmission mode, i.e. compared to previous chapters. Results are also given for the general channel model with ME-distributed channel gains/SNRs that we developed in Chapter 7.

8.2 Preliminaries

8.2.1 Effective Capacity

Using the effective bandwidth framework in [CT95], the concept of effective capacity was proposed in [WN03]. The effective capacity is, when the limit exists, given by the definition

\[ C_{\text{eff}} \triangleq - \lim_{t \to \infty} \frac{1}{\theta t} \ln \left( \mathbb{E} \{ e^{-\theta V_t} \} \right), \]  

(8.1)

where \( V_t \) is the accumulated service process at discrete time \( t \), and \( \theta \) is the so called QoS-exponent. In the above definition, the accumulated service process \( V_t \) can assume either continuous or discrete values. However, for (H)ARQ analysis, only discrete values are of interest. A more general definition of the effective bandwidth, for a finite time \( t \), has also been considered in [Kel96]. This suggests an alternative effective capacity definition

\[ C_{\text{eff},t} \triangleq - \frac{1}{\theta t} \ln \left( \mathbb{E} \{ e^{-\theta V_t} \} \right), \]  

(8.2)
which reflects a system with a \( t \)-time slots long window for communication\(^1\). The notion of effective capacity also allows for determining the maximum fixed source rate under a statistical QoS-constraint, \( P\{D > D_{\text{max}}\} \leq \epsilon_d \), where \( D \) is the steady state delay r.v. of packets in a considered source queue, \( D_{\text{max}} \) is the limit of the acceptable delay, and \( \epsilon_d \) is the limit of the acceptable delay violation probability.

Those physical parameters connect more directly back to our original motivation on QoS-enabled performance metrics. In [HZ12], it was shown that, when \( D_{\text{max}} \to \infty \), the following holds

\[
P\{D > D_{\text{max}}\} \simeq \eta_q e^{-\theta C_{\text{eff}}(\theta) D_{\text{max}}},
\]

where \( \eta_q \) is the probability that the queue is non-empty. Combining the QoS-constraint, (8.3), and rearranging, we find the QoS-exponent \( \theta^* \) of interest by solving

\[
\theta^* C_{\text{eff}}(\theta^*) = \frac{\ln \left( \eta_q/\epsilon_d \right)}{D_{\text{max}}} \triangleq \psi.
\]

Note here that we also define a QoS-parameter \( \psi \), that jointly reflects the delay target, the delay violation probability, and the probability of a non-empty queue, in (8.4). Thus, the maximum source rate under a statistical QoS-constraint, \( P\{D > D_{\text{max}}\} \leq \epsilon_d \), is \( C_{\text{eff}}(\theta^*) \). For HARQ, either a packet is in error, or a packet of rate \( R \) is communicated error-free. Therefore, (8.1) becomes

\[
C_{\text{eff}}^{\text{HARQ}} \triangleq -\lim_{t \to \infty} \frac{1}{\theta t} \ln \left( \mathbb{E} \left\{ e^{-\theta RV_t} \right\} \right) = -\lim_{t \to \infty} \frac{1}{\theta t} \ln \left( \sum_{v} e^{-\theta Rv} P\{V_t = v\} \right),
\]

where \( V_t \) is the accumulated service process in number of correctly decoded packets at discrete time \( t \). We illustrate two example service processes in Fig. 8.1 for truncated-HARQ. The service process reflects the sequence of successful transmission events, or simply called transmission event sequences. We note that at time \( t \), the service process can terminate, or pass by, at time \( t \). It is, in part, this aspect that makes the analysis challenging. For the case where each single transmission is independent from any other transmission, as is the case for ARQ, the effective capacity expression is straightforward and well-known,

\[
C_{\text{eff}}^{\text{ARQ}} \triangleq -\frac{1}{\theta} \ln \left( \mathbb{E} \left\{ e^{-\theta RV_1} \right\} \right) = -\frac{1}{\theta} \ln \left( Q_1 + e^{-\theta R} P_1 \right),
\]

where \( P_1 = P\{V_1 = 1\} \) and \( Q_1 = P\{V_1 = 0\} = 1 - P_1 \).

---

\(^1\)Later, we show that this can also be motivated with respect to throughput measured over a \( t \)-time slots long window.
Figure 8.1: Examples, A and B, of transmission event sequences, for truncated-HARQ with $K = 2$.

### 8.3 Generalized Retransmission System Model

In order to analyze this more general retransmission system, i.e. concurrently allowing for multiple (re)transmissions, multiple rate increments per transmission cycle, and multiple communication modes, we first need to extend the system model described in Chapter 3, more specifically to generalize the decoding failure probability-based model in Section 3.4.1.

A useful depiction of the proposed retransmission system model is as a (constrained) three-dimensional random-walk (3D-RW) model with varying step sizes on a grid. Each random 3D-step represents a transmission cycle and is characterized by a transition probability $P_{k\nu\tilde{l}l}$, where $k$ represents the number of transmissions needed until successful decoding, $\nu$ is the number of correctly decoded packets, $\tilde{l}$ is the originating communication mode, and $l$ is the final communication mode during a retransmission cycle. This intuitive, and naturally formulated, model, with values for $P_{k\nu\tilde{l}l}$, fully characterizes (essentially) any retransmission scheme and, indirectly, its performance. With communication modes, we mean that a retransmission scheme can shift between different (re-)transmission strategies or communication conditions. For example, in NC-ARQ, Section 8.4.5, the sender alternates between sending regular and network-coded packets, and for 2-mode ARQ, Section 8.4.6, the channel statistics alternates. More formally, we assume that we start a transmission cycle at a transmission event state $\{t, \tilde{v}, \tilde{l}\}$, where $\tilde{l}$ is the current time slot, $\tilde{v}$ is the current number of correctly decoded packets, and $\tilde{l}$ is the current communication mode. At the end of the transmission cycle, we assume that the transmission event state will be $\{t, v, l\}$. Hence, a transmission cycle lasts $k \triangleq t - \tilde{l}$ transmissions to complete, $\nu \triangleq v - \tilde{v}$ number of error-free packets (or rate increments) are communicated, and the communication mode changes from $\tilde{l}$ to $l$. We limit the model to the ranges $k = \{1, 2, \ldots K\}$, $\nu = \{0, 1, \ldots \nu_{\text{max}}\}$, $l = \{1, 2, \ldots L\}$, and $\tilde{l} = \{1, 2, \ldots L\}$. As before, $K$ is the maximum number of
transmission attempts, \( \nu_{\text{max}} \) is the maximum number of packets (each of rate \( R \)) that can be communicated in a transmission cycle, and \( L \) is the number of communication mode layers. We note that the sum of all transition probabilities exiting a transmission event state \( \{ t, v, l \} \) adds to one for each \( l \). We let \( \pi_{t,v,l} \) denote the recurrence termination probability of all transmission event sequences terminating in \( \{ t, v, l \} \). We further let \( P\{ V_t = v, L_t = l \} \) signify the state probability of all transmission event sequences either terminating in, or passing through, \( \{ t, v, l \} \). As before, we still assume ideal retransmission operation with error-free feedback, negligible protocol overhead, ideal error detection with zero probability of misdetection, and other standard simplifying assumptions.

We illustrate the model with some examples. Fig. 8.1 shows two realization of service curves for truncated-HARQ with \( K = 2 \). In Fig. 8.2, we illustrate the basic 3D-RW directions based on, multiple communication mode layers, -transmissions and -rate increments.

8.4 Performance Analysis

In this section, we consider the effective capacity of a general class of truncated-retransmission schemes (with potential packet discards on the last transmission attempt), allowing for multiple transmissions, multiple modes, and multiple rate increments. Subsequently, we specialize the overall analysis and results to HARQ, illustrating the case of multiple-transmissions, and then (for example) to NC-ARQ, illustrating the case of multiple modes and multiple rate increments. For the initial general analysis, we start by studying the effective capacity of a finite \( t \)-slotted retransmission system in Theorem 8.1, and then proceed to develop the analysis for the limiting case, with infinite \( t \).
8.4.1 General Analysis of Retransmission Schemes

**Theorem 8.1.** The effective capacity of a retransmission scheme, with \( t \)-time slots, \( K \) as transmission limit, \( L \) communication mode layers, \( \nu \in \{0,1,\ldots,\nu_{\text{max}}\} \) rate increments per transmission cycle, and the probability of successful decoding on the \( k \)th attempt \( P_{k\nu} \), is

\[
C_{\text{Retr. eff.}} = -\frac{1}{\theta t} \ln (b^T A^{t-K} f_K), \quad t \geq K,
\]

where \( b = [1^L \ 0^{1 \times L(K-1)}]^T \), \( f_K \) is an initial vector\(^2\) of size \( KL \times 1 \) containing the \( K \) first expectation values,

\[
A = \begin{bmatrix}
A_1 & A_2 & \ldots & A_{K-1} & A_K \\
I & 0 & \ldots & 0 & 0 \\
0 & I & \ldots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \ldots & I & 0
\end{bmatrix}
\]

is a block companion matrix of size \( KL \times KL \),

\[
A_k = \begin{bmatrix}
a_{k11} & a_{k12} & \ldots & a_{k1L} \\
a_{k21} & a_{k22} & \ldots & a_{k2L} \\
\vdots & \vdots & \ddots & \vdots \\
a_{KL1} & a_{KL2} & \ldots & a_{KL_L}
\end{bmatrix}
\]

is a size \( L \times L \) submatrix, \( k \in \{1,2,\ldots,K\} \), and

\[
a_{kli} = \sum_{\nu=0}^{\nu_{\text{max}}} P_{k\nu} \tilde{c}_{k\nu} e^{-\theta R_{\nu}},
\]

is a scalar entry for submatrix \( A_k \), for \( \tilde{l} \in \{1,2,\ldots,L\} \), and \( l \in \{1,2,\ldots,L\} \).

**Proof.** The proof is given in Appendix 8.A. \(\square\)

Eq. (8.7) reveals that the effective capacity for a \( t \)-time slots retransmissions system is fully determined by the block companion matrix \( A \), and the initial vector \( f_K \). We now see that, in contrast to related work on effective capacity, [CT95], [WN03]-[AF15], that considers the limiting case \( t \to \infty \), Theorem 8.1 gives an algebraic matrix expression of the log-mgf (and the effective capacity) for any \( t \).

\(^2\)It is also possible to express (8.7) on the form \( C_{\text{Retr. eff.}} = -\frac{1}{\theta t} \ln (b^T f_t) \), where \( f_t = Af_{t-1} + c_{t-1}, t \geq 1 \). \( f_0 \) represents the initial vector, e.g. \( f_0 = [1 \ 0]^T \). We find that \( c_{t\tilde{l}} = \sum_{k=1}^{K} \sum_{l=1}^{L} \sum_{\nu=0}^{\nu_{\text{max}}} P_{k\nu l} \), where \( \tilde{c}_t = [c_{t,1} \ c_{t,2} \ \ldots \ c_{t,L}]^T \), and \( c_t = [\tilde{c}_t \ 0 \ \ldots \ 0]^T \). Note that \( c_t \) is all zero for \( t \geq K \).
8.4. Performance Analysis

An important aspect to note from Theorem 8.1, is that the theorem enables us to also compute the \((t\text{-time slots})\) throughput. This is so since

\[
\lim_{\theta \to 0} C_{\text{eff},t}^{\text{Retr}} = \frac{\mathbb{E}\{V_t\}}{t} \triangleq T_{t}^{\text{Retr}}, \tag{8.11}
\]

which is proved in Appendix 8.B. In Appendix 8.C, we also show that the idea of a recurrence relation formulation, as used in Theorem 8.1, can be used to analyze the \(\kappa\)-moment, \(\mathbb{E}\{V_t^\kappa\}\). This in turn, allows the throughput (using the first-moment \(\mathbb{E}\{V_t\}\)), to be determined directly rather than as a limit in (8.11). From now on, we will focus on the limiting case, \(t \to \infty\), of the effective capacity, (8.1). With this in mind, we first give the effective capacity for infinite \(t\) in the corollary below.

**Corollary 8.1.** The effective capacity of a retransmission scheme, with \(t \to \infty\), \(K\) as transmission limit, \(L\) communication mode layers, \(\nu \in \{0,1,\ldots,\nu_{\text{max}}\}\) rate increments per transmission cycle, and the transition probabilities \(P_{k\nu \tilde{\ell}}\), is

\[
C_{\text{eff}}^{\text{Retr}} = -\ln \left(\lambda_+ / \theta\right), \tag{8.12}
\]

where \(\lambda_+ = \max\{\mid \lambda_1\mid, \mid \lambda_2\mid, \ldots, \mid \lambda_{KL}\mid\}\) is the spectral radius of the block companion matrix \(A\), with eigenvalues \(\{\lambda_1, \lambda_2, \ldots, \lambda_{KL}\}\), given in Theorem 8.1.

**Proof.** The proof is given in Appendix 8.D.

To illustrate the usefulness of this unified approach, Theorem 8.1 and Corollary 8.1, we now consider and analyze practically interesting schemes, for example truncated-HARQ in Section 8.4.2, and NC-ARQ in Section 8.4.5. In Sections 8.4.3 and 8.4.4, we leave the (finite size) matrix formulation in Theorem 8.1 and Corollary 8.1, and instead consider a characteristic equation form allowing for an infinite transmission limit \(K\).

### 8.4.2 General Analysis of HARQ

In this section, we focus on truncated-HARQ where the probabilities, \(P_k, k \in \{1,2,\ldots,K\}\) and \(Q_K\), are assumed given. For this case, we have only one communication mode, \(L = 1\), and a packet of rate \(R\) is either delivered at latest on the \(K\)th transmit attempt, or is not delivered at all. The corollary below enables the corresponding effective capacity to be computed.

**Corollary 8.2.** The effective capacity of truncated-HARQ, with transmission limit \(K\), \(\nu \in \{0,1\}\) packet per transmission cycle, and the probabilities of successful decoding on the \(k\)th attempt \(P_k\), \(Q_K = 1 - \sum_{k=1}^{K} P_k\), is given by Theorem 8.1 for
finite $t$ (or Corollary 8.1 for infinite $t$), together with the companion matrix

$$A = \begin{bmatrix} a_1 & a_2 & \ldots & a_{K-1} & a_K \\ 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & 0 \end{bmatrix},$$

(8.13)

where

$$a_k \triangleq \begin{cases} P_k e^{-\theta R}, & k \in \{1, 2, \ldots, K-1\}, \\ P_K e^{-\theta R} + Q_K, & k = K. \end{cases}$$

(8.14)

Proof. The proof is given in Appendix 8.E. \hfill \Box

While a full derivation for a general retransmission scheme is performed in Theorem 8.1, it is nevertheless instructive and useful to highlight some of the key-expressions for truncated-HARQ. Simplifying the notation from Theorem 8.1, with $f_t \triangleq f_{t,1}$ and $a_k \triangleq a_{k,1}$, we see that the recurrence relation (8.45), for one communication mode, can be written as the homogeneous recurrence relation

$$f_t = \sum_{k=1}^{K} a_k f_{t-k},$$

(8.15)

where $f_t = \mathbb{E}\{e^{-\theta RV_t}\}$. Alternatively, (8.15) can be written on the matrix recurrence relation form

$$f_t = A f_{t-1},$$

(8.16)

where $f_t = [f_t, f_{t-1}, \ldots, f_{t-K}]^T$, and $A$ is given by (8.13). An alternative to the spectral radius of (8.13), is to determine the largest root of the characteristic equation of (8.15). Setting $f_t = \lambda^t$ in (8.15), the characteristic equation is found to be

$$\lambda^K - \sum_{k=1}^{K} a_k \lambda^{K-k} = 0.$$  

(8.17)

This formulation, (8.17), is (as will be seen) vital for validation purpose and extending the analysis to persistent-HARQ, and is also the basis for Corollary 8.3-8.13. For convenience to the reader, we also show the simpler, and more tractable, derivation of the recurrence relation for truncated-HARQ in Appendix 8.F.

Here, it can be noted that matrix $A$, (8.13), with entries (8.14), is not on the $\mathbf{P}\Phi$-form (which is due to the Markov modulated process modeling) as in [TZ07]-[AF15], and discussed in Section 1.3, but on a more compact form. It is easy to see that this
also holds true more generally for (8.8) with (8.9). Observe also that Corollary 8.2, with the expressions (8.13) and (8.14) differs from [AF15, (9)]. This is so for two important reasons. The first reason is that the mathematical model in the latter implies that the packet is always delivered before or at the last transmit attempt, whereas in this work, a data packet is discarded on the last transmit attempt if decoding fails. The second reason is that the Markov modulated process modeling gives a more complicated matrix-form, with e.g. ratios of transition probabilities and transition probabilities repeated in multiple entries. The RW-modeling and recurrence relation framework, not only give (8.13) on a simple form, but also (8.17) which enables the derivation of Corollary 8.3-8.9.

We now continue with three subsections. The first two give validation and application examples, whereas the third frame the effective capacity more directly in real-world QoS-parameters, i.e. the delay target $D_{\text{max}}$ and the delay violation probability $\epsilon_d$, rather than in the QoS-exponent $\theta$.

### Validation Examples

This section serves to validate the soundness of the HARQ analysis above. A first aspect to investigate is that the effective capacity is in the range $(0, \infty)$, which is done in the following corollary.

**Corollary 8.3.** The characteristic equation, $\lambda^K - \sum_{k=1}^{K} a_k \lambda^{K-k} = 0$, $a_k > 0$, has only one positive root, and it lies in the interval $[0, 1]$.

**Proof.** The proof is given in Appendix 8.G.3

Thus, this confirms that $0 \leq C_{\text{HARQ}}^{\text{eff}} \leq \infty$, since $0 \leq \lambda_+ \leq 1$, and $C_{\text{HARQ}}^{\text{eff}} = -\ln(\lambda_+)/\theta$.

The next corollary verifies that the effective capacity, Corollary 8.1 with (8.13) and (8.14), converges to the well-known throughput expression (4.2) of truncated-HARQ.

**Corollary 8.4.** The effective capacity of truncated-HARQ converges to the throughput of truncated-HARQ as $\theta \to 0$,

$$
\lim_{\theta \to 0} C_{\text{HARQ}}^{\text{eff}} = \frac{R(1 - Q_K)}{\sum_{k=1}^{K} kP_k + KQ_K} \triangleq T_{\text{HARQ}}^{\text{trunc.}}. \tag{8.18}
$$

**Proof.** The proof is given in Appendix 8.H.

---

3We note that the same goal as for Corollary 8.3 has also been considered in [AF15, Thm2]. However, Corollary 8.3 with proof differs. First, the characteristic equation differs from [AF15, (4)] since in the present analytical model, a packet may be discarded when reaching the transmission limit. Second, a slightly different approach, exploiting Descartes’ rule of signs, is used to show that only one positive root exists. Third, but not considered in [AF15, Thm2], it is shown that the positive root lies in the interval $[0,1]$ which is required for a real and positive effective capacity.
Note that when $K \to \infty$, i.e. $Q_K \to 0$, (8.18) converges to the throughput of persistent-HARQ (4.5). For the case with $\theta \to \infty$, the characteristic equation yields $\lambda_+ = \frac{Q_1}{K}$, which gives $\lim_{\theta \to \infty} C_{\text{HARQ}}^{\text{Geom}} = -\lim_{\theta \to \infty} \frac{\ln(Q_K)}{\theta K} = 0$.

To further validate Corollary 8.2, and the forms of (8.13), (8.14), we now show that the effective capacity of truncated-HARQ degenerates to the effective capacity of ARQ (8.6) when $P_k$ is geometrically distributed. This need to be the case since this, the assumed transmission independence, implies that information from earlier transmissions is not exploited.

**Corollary 8.5.** The effective capacity for truncated-HARQ with $P_k$ geometrically distributed, $P_k = P_1 Q_k^{k-1}$, $P_1 = 1 - Q_1$, $k \in \{1, 2, \ldots, K\}$, and $Q_K = 1 - \sum_{k=1}^K P_k = Q_1^K$ is

$$C_{\text{HARQGeom}}^{\text{Geom}} = -\frac{1}{\theta} \ln \left( Q_1 + P_1 e^{-\theta R} \right).$$

(8.19)

**Proof.** The proof is given in Appendix 8.I.

As expected, we note that (8.19) has exactly the same form as the effective capacity of ARQ (8.6). We now turn the attention to two application examples of the truncated-HARQ analysis, specifically of (8.17).

**Application Examples**

Here, we apply the truncated-HARQ analysis to the simplest truncated-HARQ system imaginable, i.e. with a maximum of $K = 2$ transmissions\(^4\), and also give an approximative effective capacity expression that is valid for small $\theta$. We start with this simple truncated-HARQ case. The following corollary gives a closed-form expression for the effective capacity of such system.

**Corollary 8.6.** The effective capacity for truncated-HARQ with transmission limit $K = 2$, and probabilities $P_1$, $P_2$, of successful decoding, and probability $Q_2 = 1 - P_1 - P_2$ of failed decoding, is

$$C_{\text{HARQ}}^{\text{HARQ}} = R - \frac{1}{\theta} \ln \left( \frac{P_1 + \sqrt{P_1^2 + 4(P_2 e^{\theta R} + Q_2 e^{\theta^2 R^2})}}{2} \right).$$

(8.20)

**Proof.** The proof is given in Appendix 8.J.

We now turn our attention to the approximation of the effective capacity in terms of the first and second moments of the number of transmissions per packet.

\(^4\)Another interesting case is for $K \to \infty$, but this requires some more assumptions and is therefore handled in Section 8.4.3.
Corollary 8.7. The effective capacity for persistent-HARQ can, for small $\theta$, be approximated as

$$C_{\text{HARQ}}^{\text{eff}} \approx \frac{c_1 + \sqrt{c_1^2 + 4c_2}}{2},$$

(8.21)

where $c_1 = \frac{2\tau_1 (1 - \theta R)}{\theta (\tau_1^2 + \tau_2)}$, $c_2 = \frac{R(2 - \theta R)}{\theta (\tau_1^2 + \tau_2)}$.

(8.22)

where $\tau_1 = \sum_{k=1}^{\infty} kP_k$ signify the first raw moment (the mean number of transmissions), and $\tau_2 = \sum_{k=1}^{\infty} k^2 P_k$ is the second raw moment.

Proof. The proof is given in Appendix 8.K.

We note that Corollary 8.7 offers a new effective capacity approximation of HARQ in the first- and second-order raw moments. This is similar in spirit to the approximation $C_{\text{HARQ}}^{\text{eff}} \approx R/\mu_{\text{retr}} - R^2\sigma_{\text{retr}}^2/2\mu_{\text{retr}}^3$, valid for small $\theta$, proposed in [LGV15].

Effective Capacity Expressed in QoS-Parameter $\psi$

So far, we have considered the effective capacity for a given $\theta$. Taking on, more of, an engineering point of view, we are also interested in its dependency on $D_{\text{max}}$ and $\epsilon_d$. A different, but more severe problem when exact performance expressions are desired, is that the characteristic equation (8.17) can not be solved for arbitrarily large polynomial degrees, and the spectral radius of the companion matrix $A$ of arbitrarily large dimensions must be computed numerically. Thus, as $\lambda_+$ can not always be given in closed-form, neither can $C_{\text{eff}}$. The main idea suggested here, as a remedy to both problems, is to first consider the effective capacity definition $C_{\text{eff}} = -\ln(\lambda_+)/\theta$, then note from the characteristic equation (8.17) that $\lambda_+ = \lambda_+(\theta)$, and subsequently from (8.4) use the substitution $\psi = \theta C_{\text{eff}}$, which leads to $\psi/\theta = -\ln(\lambda_+(\psi/C_{\text{eff}}))/\theta$. Solving the resulting expression for $C_{\text{eff}}$ then yields $C_{\text{eff}} = \psi/\lambda_+^{-1}(e^{-\psi})$, where $\lambda_+^{-1}(\cdot)$ is the inverse function of $\lambda_+(\cdot)$. This approach is considered for HARQ in the corollary below.

Corollary 8.8. The effective capacity of truncated-HARQ, with transmission limit $K$, $\nu \in \{0, 1\}$, probabilities of successful decoding on the $k$th attempt $P_k$, $Q_K \triangleq 1 - \sum_{k=1}^{K} P_k$, and $\psi \triangleq \ln((\eta_q/\epsilon_d)/D_{\text{max}})$, is

$$C_{\text{HARQ}}^{\text{eff}}(\psi) = \frac{R\psi}{\ln\left(\sum_{k=1}^{K} P_k e^{\psi k}\right) - \ln\left(1 - Q_K e^{\psi K}\right)}$$

(8.23)

Proof. The proof is given in Appendix 8.L.

---

5 An approximation for truncated-HARQ can also be found, but is then, in addition to $\mu_{\text{retr}}$ and $\sigma_{\text{retr}}^2$, also expressed with $Q_K \neq 0$. 
We note that the effective capacity can now, when $C_{\text{HARQ}}^{\text{eff}}(\psi)$ is a function of $\psi$, be given in closed-form. Albeit $C_{\text{HARQ}}^{\text{eff}}(\theta)$ may not have a closed-form expression, we can still plot $C_{\text{HARQ}}^{\text{eff}}(\theta)$ vs. $\theta$ parametrically, as $(\psi/C_{\text{HARQ}}^{\text{eff}}(\psi), C_{\text{HARQ}}^{\text{eff}}(\psi))$, with a closed-form expression for $C_{\text{HARQ}}^{\text{eff}}(\psi)$. We also see that (8.23) reduces to the classical truncated-HARQ throughput expression (4.2) if $\psi \to 0$. Moreover, persistent-HARQ has the simple and compact expression

$$C_{\text{HARQ}}^{\text{eff}}(\psi) = R \psi^{-1} \ln \left( \sum_{k=1}^{\infty} P_k e^{\psi k} \right).$$  (8.24)

Observe that (8.24) is a generalization of the well-known persistent-HARQ throughput expression (4.5). Note also that $\psi \leq -\ln (Q_K)/K$ for a real denominator in (8.23). Eq. (8.23) is also interesting since it suggest, and we conjecture, that $C_{\text{eff}}(\psi) = \lim_{v \to \infty} R \psi / \ln \left( E \{ e^{\psi T_v} \} \right)$, where $T_v$ is a r.v. for the number of time slots used to deliver $v$ rate increments (or packets). Applying the RW idea, and letting $f_v = E \{ e^{\psi T_v} \}$, we can formulate the recurrence

$$f_v = \left( \sum_{k=1}^{K} P_k e^{k \psi} \right) f_{v-1} + Q_K e^{K \psi} f_v.$$  

This recurrence has the characteristic equation solution $\lambda_+ = \left( 1 - Q_K e^{K \psi} \right)^{-1} (\sum_{k=1}^{K} P_k e^{k \psi})$, which agrees with (8.23).

### 8.4.3 HARQ with Effective Channel

In Theorem 8.1, we gave an effective capacity expression for general retransmission schemes in terms of transition probabilities. In Section 8.4.2, we specialized this result to HARQ. Now, this specialized expression is used to derive an effective capacity expression where the effective channel is described by a pdf $f_Z(z)$, its Laplace transform $F(s)$, and a decoding threshold $\Theta$. A second fundamental result of this chapter is given by the following theorem.

**Theorem 8.2.** For persistent-HARQ schemes, characterized by an effective channel pdf $f_Z(z)$ and threshold $\Theta$, the spectral radius, $\lambda_+$, in (8.12), is implicitly given by

$$e^{\theta R} = L_\Theta^{-1} \left( \frac{1}{s} \left( \frac{1 - F(s)}{\lambda_+ - F(s)} \right) \right).$$  (8.25)

**Proof.** The proof is given in Appendix 8.M. \qed

**Remark 8.1.** It is, in principle, possible to extend and generalize the idea behind Theorem 8.2 to the wider scope of Theorem 8.1. This could, e.g., allow $L \geq 2$, with multiple effective channels, to be handled as $K \to \infty$. However, this goes beyond the scope of the chapter, and is omitted in the following.

A similar derivation for truncated-HARQ is possible, but yields the (somewhat less appealing) form

$$e^{\theta R} = \left( \frac{1}{s} \left( 1 - F(s)^K \lambda_+^{-K} \right) \right)^{-1} L_\Theta^{-1} \left( \frac{1}{s} \left( \frac{1 - F(s)}{\lambda_+ - F(s)} \right) \left( 1 + \frac{F(s)^K}{\lambda_+} \right) \right).$$  (8.26)
Also for the effective channel case, it is of interest to express $C_{\text{eff}}$ in the QoS-parameter $\psi$. This is done in Corollary 8.9 where we focus on the persistent-HARQ case.

**Corollary 8.9.** The effective capacity of persistent-HARQ, characterized by an effective channel pdf $f_Z(z)$, threshold $\Theta$, and $\psi \triangleq \ln(\eta/\epsilon_d)/D_{\text{max}}$, is

$$C_{\text{HARQ}}^{\text{eff}}(\psi) = \frac{R}{\psi^{-1} \ln \left( \sum_{\Theta}^{-1} \frac{e^{-\psi} \left( 1 - F(s) \right)}{1 - F(s) e^{\psi}} \right)}.$$  \hfill (8.27)

*Proof.* The proof is given in Appendix 8.N.

Thus, with $F(s)$ given, we can either use (8.25) and solve for $\lambda_+$ to compute the effective capacity in terms of $\theta$, or we can use (8.27), to get the effective capacity in terms of $\psi$. The benefit of the latter is the closed-form expression and relating the effective capacity directly to delay target and delay violation probability. We now turn our attention to the lowest, the third, system model level and consider fading channel models. The case with truncated-HARQ in (8.26) can be handled in a similar manner.

### 8.4.4 HARQ with ME-distributed- and Rayleigh-Fading Channels

We observe that Theorem 8.2 and Corollary 8.9, expressed in $F(s)$, make them suitable to integrate with the compact and powerful ME-distribution-based effective channel framework for wireless channels in Chapter 7. We start by studying the persistent-HARQ case in the following Corollary. In this chapter, we formulate the performance expressions on a slightly different ME-distribution-based form, which involves the expression $ae^{\theta B}c$, where the parameters $a$, $B$, and $c$ depends on the ME-distributed effective-channel parameters $\tilde{p}$ and $\tilde{q}$ that corresponds to the SNR (or MI) after possible signal processing, or alternatively, the SNR without any signal processing.

**Corollary 8.10.** The effective capacity of persistent-HARQ, characterized by an effective channel ME-distribution-based pdf $f_Z(z) = \tilde{p} e^{\tilde{q} \tilde{r}}$, $\tilde{Q} = S - \tilde{r} \tilde{q}$, threshold $\Theta$, and $\psi \triangleq \ln(\eta/\epsilon_d)/D_{\text{max}}$, is

$$C_{\text{HARQ}}^{\text{eff}}(\psi) = \frac{R}{\psi^{-1} \ln \left( ae^{\theta B}c \right)}.$$  \hfill (8.28)

where $a = [0 (\tilde{q} - \tilde{p}) e^{\psi}]$, $B = S - c[0 (\tilde{q} - \tilde{p} e^{\psi})]$, $c = [0; \tilde{r}]$.

*Proof.* The proof is given in Appendix 8.O.
It is worth reflecting over the wide range of applicability of (8.28). This uni-
fying performance expression handles all possible effective channels (constituted
by HARQ schemes, fading channels, diversity-processing, etc.) that the matrix
exponential distribution can model, and gives the effective capacity (with the
throughput as a special case) as a simple closed-form expression. While this ex-
pression is parameterized in $\psi$, it can also be plotted wrt to the parameter
$\theta$, as $(\theta(\psi) \triangleq \psi/C_{HARQ}(\psi), C_{HARQ}(\psi))$.

Many works in the literature consider throughput optimization for (H)ARQ.
Likewise, it is of interest to find the maximum effective capacity of (8.28) wrt the
rate $R$, and the optimal rate point $R^\star$. The classical optimization approach is
to consider $dC_{eff}/dR = 0$, and solve for $R^\star(\Sigma)$. This approach is (generally) not
possible here, since a closed-form expression for the optimal rate-point is hard (or
impossible) to find. We therefore resort to the auxiliary parametric (AP) optimization
method in Chapter 5, and show below that it also handles effective capacity
optimization problems.

**Corollary 8.11.** The optimal effective capacity of persistent-HARQ, characterized
by an effective channel ME-distribution-based pdf $f_Z(z) = \tilde{p}e^z\tilde{Q}\tilde{r}$, threshold $\Theta$, and
$\psi \triangleq \ln(\eta/\epsilon_d)/D_{max}$, is

\[
\begin{align*}
g_\psi(\psi, \Theta) & \triangleq \frac{ae^{Bc}}{ae^{Bc}\Theta c} \ln (ae^{Bc}), \\
R^\star(\psi, \Theta) & = g_\psi + W_0(-g_\psi e^{-g_\psi}), \\
S(\psi, \Theta) & = e^{R^\star - 1}/\Theta, \\
C_{HARQ}^\star(\psi, \Theta) & = R^\star \frac{\psi}{\psi - 1 \ln (ae^{Bc})},
\end{align*}
\]

where $0 \leq \Theta < \infty$ is the auxiliary parameter.

**Proof.** The proof is given in Appendix 8.P.

Corollary 8.11 handles the optimization of the effective capacity expressed in
$\psi$ and the ME-distribution-based channel. Due the implicit structure of (8.25),
i.e. the need of solving for $\lambda_\psi$, it is often hard to solve (8.25), even for the ME-
distribution-based effective channel case. Nevertheless, in the following Corollary,
we show how to solve (8.25) for RR operating in a Rayleigh fading channel. This
also applies to the GD-channel in Definition 3.5.

**Corollary 8.12.** The effective capacity of RR in block Rayleigh fading, with $\tilde{\Theta} = (e^R - 1)/S$, is

\[
C_{RR}^\star = R - \frac{1}{\tilde{\Theta}} \left( W_0 \left( \tilde{\Theta} e^{\tilde{\Theta} + \theta R} \right) - \tilde{\Theta} \right),
\]

(8.33)
and the spectral radius, $\lambda_+$, is
\begin{equation}
\lambda_+ = \frac{\tilde{\Theta}}{W_0 \left( \Theta e^\tilde{\Theta} - \Theta_0 \right)}.
\end{equation}

\textbf{Proof.} The proof is given in Appendix 8.Q. \hfill \Box

\textbf{Remark 8.2.} Alternatively, we may (from the proof in Appendix 8.Q) solve $e^{\tilde{\Theta} R} = \lambda^{-1} e^{-\Theta_0 (1-\lambda^{-1})}$ parametrically for the SNR $S$. By using (8.12), we get the (more appealing) closed-form expression
\begin{equation}
S(C_{\text{eff}}^{\text{RR}}) = \frac{(e^R - 1)(e^{\tilde{\Theta} C_{\text{eff}}^{\text{RR}} - 1})}{\theta(R - C_{\text{eff}}^{\text{RR}})}.
\end{equation}
Using (8.35), we fix $R$ and can parametrically plot $C_{\text{eff}}^{\text{RR}}$ vs. $S$ with coordinates $(x,y) = (10 \log_{10}(S(C_{\text{eff}}^{\text{RR}})), C_{\text{eff}}^{\text{RR}})$, for $0 \leq C_{\text{eff}}^{\text{RR}} \leq R$.

It is however more interesting, and easier, to consider the effective capacity for RR in a Rayleigh fading channel with respect to the QoS-parameter $\psi$. This is done in Corollary 8.13.

\textbf{Corollary 8.13.} The effective capacity of persistent-RR in block Rayleigh fading channel, with $\tilde{\Theta} = (e^R - 1)/S$, and $\psi \triangleq \ln(\eta/\epsilon_d)/D_{\text{max}}$, is
\begin{equation}
C_{\text{eff}}^{\text{RR}}(\psi) = \frac{R}{1 + \tilde{\Theta}(e^\psi - 1)}.
\end{equation}
\textbf{Proof.} The proof is given in Appendix 8.R. \hfill \Box

We observe that (8.36), which takes the delay constraint $D_{\text{max}}$ and the delay violation probability $\epsilon_d$ into account via $\psi$, only involves an extra factor $(e^\psi - 1)/\psi$ compared to the throughput for RR in Rayleigh fading, $T^{\text{RR}} = R/(1 + \tilde{\Theta})$, [LRS14c, (12)]. Since $\tilde{\Theta} \triangleq (e^R - 1)/S$, (8.36) is the same as the throughput [LRS14c, (12)], but operating with a scaled SNR.

It is interesting to be able to compare (8.36) with the effective capacity of ARQ expressed in $\psi$. The following corollary gives the effective capacity of ARQ.

\textbf{Corollary 8.14.} The effective capacity of ARQ in block Rayleigh fading channel, with $\tilde{\Theta} = (e^R - 1)/S$, and $\psi \triangleq \ln(\eta/\epsilon_d)/D_{\text{max}}$, is
\begin{equation}
C_{\text{eff}}^{\text{ARQ}}(\psi) = \frac{R}{1 + \psi^{-1} \ln \left( \frac{1-Q_1}{1-Q_1 e^{\psi}} \right)}, \quad Q_1 = 1 - e^{-\tilde{\Theta}}.
\end{equation}
\textbf{Proof.} The proof is given in Appendix 8.S. \hfill \Box
At closer scrutiny, the RHS of \((8.37)\) converges to 
\[ R(1 - Q_1), \]
the throughput of ARQ, as \(\psi \to 0\). Moreover, with a ME-distributed fading channel, \(f_Z(z) = \tilde{p} e^{\theta Q_1}, \)
instead of Rayleigh fading, \(Q_1\) in \((8.37)\) can be directly written as 
\[ Q_1 = a e^{\theta B} c, \]
where \(a = [0 \tilde{p}], B = S - c[0 \tilde{q}],\) and \(c = [0; \tilde{r}].\)

### 8.4.5 Network-Coded ARQ

So far, we have not yet demonstrated an example where the number of rate increments (or packets) communicated during a transmission exceeds one, nor have we illustrated the use of multiple communication modes. In the following, we consider a three-mode example, where a reward of rate \(2R\) occurs. Namely, we study the case of NC-ARQ for two users, user A and B, as described in Section 2.1.3, or [LJ06]. The operation of NC-ARQ, suitable for the analysis, is reviewed next. Note that due to the assumption of the users having identical decoding probabilities, \(P_1\), it is sufficient to consider a three-mode operation for the analysis. The communication modes are indexed with \(l = \{1, 2, 3\}\). We start in communication mode \(l = 1\), which represents a normal ARQ operation. We enter mode \(l = 2\) when a data packet intended for user A is correctly decoded by user B, but not by user A. This occurs with probability \(P_1 Q_1\). In mode \(l = 2\), we transmit packets (in a normal ARQ fashion) for user B until user A, but not user B, decodes the packet correctly. At such event, we then enter mode \(l = 3\). This occurs with probability \(P_1 Q_1\). In mode \(l = 3\), we send the network-coded packet until it is either decoded by one of the users, and we then enter mode \(l = 2\) again with probability \(2P_1 Q_1\), or it is decoded by both users, and we then enter mode \(l = 1\) with probability \(P_1^2\). For a more detailed description of NC-ARQ, e.g. with more users, we refer to [LJ06]. The main result is given in Corollary 8.15.

**Corollary 8.15.** The effective capacity for 2-user NC-ARQ, with identical decoding probabilities \(P_1, Q_1 \triangleq 1 - P_1\), and a common transmit queue, is given by Theorem 8.1 for finite \(t\) (or Corollary 8.1 for infinite \(t\)) with

\[
A = \begin{bmatrix}
Q_1^2 + P_1 e^{-\theta R} & 0 & 0 \\
Q_1 P_1 & Q_1^2 + P_1 e^{-\theta R} & 2P_1 Q_1 e^{-\theta R}
\end{bmatrix}.
\]  

(8.38)

**Proof.** The proof is given in Appendix 8.T.

Since, the framework handles all aspects of multiple transmissions, multiple communications modes, and multiple packet transmissions, we note that it is directly applicable to RR- and IR-based NC-HARQ described in [LSKAT13] if the transition probabilities, \(P_{k\ell}\), are known. In Appendix 8.W, we analyze yet another retransmission scheme, Layered-ARQ, based on superposition-coding and ARQ (extendible to HARQ), that can transfer multiple-packets concurrently.
8.5 Numerical Results and Discussions

8.4.6 Two-mode ARQ and the Block Gilbert-Elliot Channel

In this section, we illustrate that the developed framework also allows for analysis of a retransmission scheme with a channel model that can change over time. Specifically, we consider two-mode ARQ, and operating in a block Gilbert-Elliot channel. As a first initial step, we analyze a more general system that alters between two modes with transition probabilities $P_{1|\nu|l}, l = \{1, 2\}$.

The effective capacity is then given by the following Corollary.

**Corollary 8.16.** The effective capacity for two-mode-ARQ, with transition probabilities $P_{1|\nu|l}, l = \{1, 2\}$, is computed with Corollary 8.1 for infinite $t$ and

$$
\lambda_+ = \frac{(a_{111} + a_{122}) + \sqrt{(a_{111} - a_{122})^2 + 4a_{121}a_{112}}}{2},
$$

(8.39)

where

$$
a_{1|l\nu|l} = \sum_{\nu=0}^{n_{\max}} P_{1|\nu|l} e^{-\theta R}.
$$

(8.40)

**Proof.** The proof is given in Appendix 8.U.

Specifically note the general form of $a_{1|l\nu|l}$, (8.40), i.e. a sum of several transition probabilities times a rate-dependent function, and again observe that this is more general than the entries in the $P\Phi$-form discussed in [TZ07]-[AF15], and under section 1.3.

The Gilbert-Elliot channel has been a popular tool to study ARQ system performance in channels with burst errors on symbol level. Here, we model a correlated block fading channel with packet (instead of symbol) errors. This could model a communication link that randomly (and with some correlation) shifts, e.g., communication media/frequency and hence propagation conditions. The probability of changing channel modes are denoted $\pi_{gg}, \pi_{gb} = 1 - \pi_{gg}, \pi_{bb}, \pi_{bg} = 1 - \pi_{bb}$ for good-to-good, good-to-bad, bad-to-bad, and bad-to-good, respectively. We further assume that when operating in the good channel mode, the successful decoding probability is $p_g$, and the decoding failure probability is $q_g = 1 - p_g$. Similarly, in the bad channel mode, we have $p_b$, and $q_b = 1 - p_b$, respectively. The effective capacity for the block Gilbert-Elliot channel is, due to independence between the transmission process and communication mode process, then given by $a_{111} = \pi_{gg}(q_g + p_g e^{-\theta R}), a_{121} = \pi_{gb}(q_g + p_g e^{-\theta R}), a_{122} = \pi_{bb}(q_b + p_b e^{-\theta R})$, and $a_{112} = \pi_{bg}(q_b + p_b e^{-\theta R})$ inserted in (8.39).

8.5 Numerical Results and Discussions

In this section, we illustrate the preceding analysis with numerical results for several of the important (H)ARQ-cases discussed in this chapter.
First, in Fig. 8.3, we plot the effective capacity expression (8.28) vs. SNR for four different rates, Rayleigh fading and Alamouti-transmit diversity. We first note that the ME-distribution-based channel model works fine in conjunction with the effective capacity metric. The characteristics shown in Fig. 8.3 are as expected, i.e. the effective capacity increases from 0 at $S = -\infty$ dB to $R$ at $S = \infty$ dB. We can also verify that (8.30)-(8.32), allows the maximum effective capacity and the optimal rate point to be plotted.

We then compare the effective capacity of persistent-RR, truncated-RR with $K = 2$, and ARQ. For simplicity, we consider a Rayleigh fading channel. In Fig. 8.4, we plot the effective capacities of (8.33), (8.20), and (8.6), for $\theta = \{0, 0.5, 1\}$ and $R = 4$ [b/Hz/s]. It is noted that when the effective capacity is low, persistent-RR is more robust wrt $\theta$ than truncated-HARQ with $K = 2$, as well as ARQ. However, when effective capacity approaching the rate $R$, the schemes are similar in robustness to changes in $\theta$. This behavior is, at low SNR, due to maximal ratio combining over many transmissions for persistent-RR (and not the other two), and that all three schemes use nearly just one transmission when $C_{\text{eff}} \approx R$. The same schemes are shown in Fig. 8.5, where the effective capacities are plotted vs. $\psi$ for $S = \{20, 30, 40\}$ dB and $R = 4$ [b/Hz/s]. We note, as expected, that the effective capacity decreases with $\theta$. Persistent-HARQ does not suffer the same performance loss when the QoS-requirement increases (with increasing $\theta$), since persistent-HARQ can benefit more from maximal ratio combining of multiple transmissions compared to the $K = 2$ case. In Fig. 8.6, the same cases are studied, but for the effective capacities of (8.36), (8.20) with (8.4), and (8.37) vs. $\psi$. More interestingly, (8.5) with (8.4) is Monte-Carlo simulated and compared to the analytical results, wherein we
8.5. Numerical Results and Discussions

Figure 8.4: The effective capacity vs. SNR for ARQ, truncated-RR with $K = 2$, and persistent-RR (with $K = \infty$), for $\theta = \{0, 0.5, 1\}$, and $R = 4 \text{ [b/Hz/s]}$.

Figure 8.5: The effective capacity vs. $\theta$ for ARQ, truncated-RR with $K = 2$, and persistent-RR (with $K = \infty$), for $S = \{20, 30, 40\} \text{ dB}$, and $R = 4 \text{ [b/Hz/s]}$.

Figure 8.6: The effective capacity vs. $\psi$ for ARQ, truncated-RR with $K = 2$, and persistent-RR (with $K = \infty$), for $S = 20 \text{ dB}$, and $R = 4 \text{ [b/Hz/s]}$. Analytical and Monte-Carlo simulated results are shown.
Figure 8.7: The effective capacity for NC-ARQ and ARQ with $R = 4$ [b/Hz/s] for $\theta = \{0, 0.1, 0.5, 1\}$. We now plot the effective capacity for the NC-ARQ matrix (8.38) and ARQ expression (8.37) vs. SNR for different values of $\theta$ in Fig. 8.7. We note that NC-ARQ is less sensitive to an increase in $\theta$ (tougher QoS-requirements), and provides higher effective capacity, than ARQ. This can be explained by the increased degree of diversity provided by NC-ARQ over ARQ. We expect this effect to increase with increasing number of users for NC-ARQ. The same phenomenon has been experimentally verified for effective channels with $F(s) = 1/(1+s)^N$, where $N$ is the degree of diversity. An obvious idea, based on the above, is to extend the analysis to more advanced joint network coding - retransmission schemes. However, it has proven hard (since [LJ06]) to design a structured, simple, and capacity achieving, scheme for NC-ARQ with more than 2-users. Nevertheless, with such design, the effective capacity can be determined with the framework proposed in this chapter.

Overall, we observe that the effective capacity plots asymptotically approaches the corresponding throughput curves, when $\theta \to 0$, or $\psi \to 0$. Note that the throughput plots are based on the analytical expressions derived with the methods given in Chapter 4. We also observe, as expected, that the effective capacity is less than the throughput. We noted that diversity, here in form of NC-ARQ, reduces the sensitivity of the effective capacity on $\theta$.

8.6 Summary and Conclusions

In this chapter, we have studied the effective capacity of general retransmission schemes, with multi-transmissions, communication modes, and rate increments. We modeled such schemes as a (constrained) random walk with transition probabilities, and the effective capacity could be compactly formulated and efficiently
analyzed through a system of recurrence relations. From this, a matrix, and a characteristic equation, approach were developed to solve the recurrence relation(s). The characteristic equation method (with its simple form) turned out to be particularly useful to gain new insights in many important cases, e.g. for truncated- and persistent-HARQ. This led to that many useful results could be formulated, in terms of general transition probabilities, general effective channel functions, or in specific wireless fading channels. With results expressed in a real QoS-metric dependent parameter, ψ, and the ME-distributed channel, we could enhance the practical relevance even further. An interesting finding is that diversity, e.g. due to MRC or NC, or for that matter channels with small variance to mean, such as would be the case for spatially multiplexed MIMO, reduces the sensitivity of the effective capacity to θ (or to ψ).
Appendices

8.A Proof of Theorem 8.1

Proof. We first show that \( \mathbb{E}\{e^{-\theta RV_t}\} \) can be expanded into a system of recurrence relations, which we will arrive to in (8.48). Following that, we rewrite the system of recurrence relations on a matrix recurrence form. Starting with the LHS of the recurrence first, we have

\[
\mathbb{E}\{e^{-\theta RV_t}\} = \sum_{l=1}^{L} \sum_{v} e^{-\theta Rv} \mathbb{P}\{V_t = v, L_t = l\}.
\]  

(8.41)

For the RHS of said recurrence relation, we get the expression

\[
\mathbb{E}\{e^{-\theta RV_t}\} = \sum_{l=1}^{L} \sum_{v} e^{-\theta Rv} \mathbb{P}\{V_t = v, L_t = l\}
\]

\[= \sum_{l=1}^{L} \sum_{v} e^{-\theta Rv} \sum_{k=1}^{K} \sum_{\nu=0}^{\nu_{\text{max}}} \sum_{l'=1}^{\tilde{l}} P_{k\nu l'} \mathbb{P}\{V_{t-k} = v - \nu, L_{t-k} = l\}
\]

\[= \sum_{k=1}^{K} \sum_{\nu=0}^{\nu_{\text{max}}} \sum_{l'=1}^{\tilde{l}} P_{k\nu l'} e^{-\theta R\nu}
\]

\[\times \left( \sum_{l=1}^{L} \sum_{v} e^{-\theta R(v-\nu)} \mathbb{P}\{V_{t-k} = v - \nu, L_{t-k} = l\} \right),
\]  

(8.42)

where we used the relation

\[
\mathbb{P}\{V_t = v, L_t = l\} = \sum_{k=1}^{K} \sum_{\nu=0}^{\nu_{\text{max}}} \sum_{l'=1}^{\tilde{l}} P_{k\nu l'} \mathbb{P}\{V_{t-k} = v - \nu, L_{t-k} = l\}
\]

(8.43)

in step (a), and multiplied with \(e^{\theta R(v-\nu)}\), a dummy one, and changed the summation order in step (b). We now define \( f_{t,l} \triangleq \sum_{v} e^{-\theta Rv} \mathbb{P}\{V_t = v, L_t = l\} \), and let

\[
a_{k\nu l'} \triangleq \sum_{\nu=0}^{\nu_{\text{max}}} P_{k\nu l'} e^{-\theta R\nu}.
\]

(8.44)

Using the above definitions, and equating the last expression in (8.41), and the last expression in (8.42), we find that

\[
\sum_{l=1}^{L} f_{t,l} = \sum_{l=1}^{L} \sum_{l'=1}^{\tilde{l}} \sum_{k=1}^{K} a_{k\nu l'} f_{t-k,l'},
\]  

(8.45)
where \( E \{ e^{-\theta RV_t} \} = \sum_{l=1}^{L} f_{t,l} \). A structured strategy\(^6\) to solve (8.45) is to first reformulate it into a matrix recurrence relation, while omitting the summation over \( l \). We then get the order-\( k \) linear homogeneous matrix recurrence relation

\[
\tilde{f}_t = \sum_{k=1}^{K} A_k \tilde{f}_{t-k}, \tag{8.46}
\]

where

\[
\tilde{f}_t = \begin{bmatrix} f_{t,1} & f_{t,2} & \cdots & f_{t,L} \end{bmatrix}^T, \tag{8.47}
\]

and \( A_k \) as in (8.9). Subsequently, (8.46) is rewritten as a first order homogeneous linear matrix recurrence relation

\[
f_t = A f_{t-1}, \tag{8.48}
\]

where

\[
f_t = \begin{bmatrix} \tilde{f}_t & \tilde{f}_{t-1} & \cdots & \tilde{f}_{t-K} \end{bmatrix}^T, \tag{8.49}
\]

and \( A \) is the block companion matrix in (8.8).

We can now finalize the proof of the effective capacity expression by noting that

\[
C_{\text{eff}, t}^{\text{Retr.}} = -\frac{1}{\theta t} \ln \left( E \{ e^{-\theta RV_t} \} \right)
= -\frac{1}{\theta t} \ln \left( \sum_{l=1}^{L} f_{t,l} \right)
\overset{(a)}{=} \frac{1}{\theta t} \ln (b^T f_t)
\overset{(b)}{=} -\frac{1}{\theta t} \ln (b^T A^{t-K} f_K), \quad t \geq K, \tag{8.50}
\]

where the sum is put on a vector-matrix form in step (a), and we use \( f_t = A f_{t-1} \)

repeatedly in step (b).

\[\Box\]

**8.B Proof of Eq. (8.11)**

*Proof.* The effective capacity (for \( t \)-time slots) converges to the throughput (for \( t \)-time slots) when \( \theta \to 0 \) since

\[
\lim_{\theta \to 0} C_{\text{eff}, t}^{\text{Retr.}} = -\lim_{\theta \to 0} \frac{1}{\theta t} \ln \left( E \{ e^{-\theta RV_t} \} \right)
\]

\[\footnote{Another strategy is to rewrite the (8.45) into a single homogeneous recurrence relation (with a resulting longer memory), by some strategic variable substitutions, and to solve the corresponding characteristic equation. Unfortunately, this is often tricky for a system of recurrence relations.}\]
Effective Capacity Analysis of General Retransmission Schemes

\[
\begin{align*}
\overset{(a)}{=} \lim_{\theta \to 0} \frac{1}{t} \left( R \sum_{l=1}^{L} \sum_{v_0} v \mathbb{P}\{V_l = v, L_t = l\} + O(\theta) \right) \\
= \frac{\text{RE}\{V_l\}}{t} \triangleq \tau_{\text{Retr}},
\end{align*}
\]  

(8.51)

where we in step (a) used

\[
\mathbb{E}\{e^{-\theta RV_l}\} = \sum_{l=1}^{L} \sum_{v_0} e^{-\theta Rv} \mathbb{P}\{V_l = v, L_t = l\} \\
= \sum_{l=1}^{L} \sum_{v_0} (1 - \theta Rv + O(\theta^2)) \mathbb{P}\{V_l = v, L_t = l\} \\
= 1 - \theta R \sum_{l=1}^{L} \sum_{v_0} v \mathbb{P}\{V_l = v, L_t = l\} + O(\theta^2),
\]

and exploited that \(-\ln(1 - x) \simeq x\) for small \(x\). \(\square\)

8.C Throughput and \(\kappa\)-Moment of Finite-\(t\) Truncated-HARQ

In this section, we consider a truncated-HARQ system with \(t\)-time slots, and analyze the throughput in the corollary below. Specifically, in the proof, we show that the recurrence relation idea is also applicable to compute the \(\kappa\)-raw-moment \(\mathbb{E}\{V_t^\kappa\}\).

**Corollary 8.17.** The \(t\)-time slot limited throughput, \(T_t \triangleq \text{RE}\{V_l\}/t\), of truncated-HARQ has the form

\[
\tau_{\text{trunc.}}^{\text{HARQ}} = \tau_{\text{trunc.}}^{\text{HARQ}} + \frac{1}{t} \sum_{k=1}^{K} c_k \lambda_k^t, 
\]

(8.52)

where \(\lambda_k\) are the solutions to the characteristic equation \(\lambda^K = Q_K + \sum_{k=1}^{K} P_k \lambda^{K-k}\), and the constants \(c_k\) are determined from initial conditions.

**Proof.** First, a recurrence relation for the \(\kappa\)-moment can be expanded as

\[
\mathbb{E}\{V_t^\kappa\} = \sum_{v} v^\kappa \mathbb{P}\{V_l = v\} \\
= \sum_{v} v^\kappa Q_K \mathbb{P}\{V_{t-K} = v\} + \sum_{v} ((v-1)+1)^\kappa \sum_{k=1}^{K} P_k \mathbb{P}\{V_{t-k} = v-1\} \\
= Q_K \mathbb{E}\{V_{t-K}^\kappa\} + \sum_{v} \sum_{k=1}^{K} P_k \sum_{\beta=0}^{\kappa} \binom{\kappa}{\beta} (v-1)^\beta \mathbb{P}\{V_{t-k} = v-1\}
\]
\[ Q_K \mathbb{E}\{V_t^K\} + \sum_{k=1}^{K} \sum_{\beta=1}^{\kappa} P_k(\beta) \mathbb{E}\{V^\beta_{t-k}\} + 1 - Q_K. \] (8.53)

Now, consider (8.53) with \( \kappa = 1 \), multiply both sides with \( R \), extend the \((t-k)\)th mean-terms with \((t-k)/(t-k)\), and use the definition \( T_t = \mathbb{E}\{V_t\}/t \). We then arrive at the recurrence relation \( tT_t = Q_K(t-K)T_{t-K} + \sum_{k=1}^{K} P_k(t-k)T_{t-k} + R(1-Q_K) \). Now, insert \( T_t = T_{t.trunc} + t \), with \( T_{t.trunc} \) from (8.18), in the recurrence relation which is then transformed into the homogeneous form \( tT_t = Q_K(t-K)T_{t-K} + \sum_{k=1}^{K} P_k(T_{t-k}) \). Then with \( h_t = tT_t \), we get \( h_t = Q_K h_{t-K} + \sum_{k=1}^{K} P_k h_{t-k} \), which has the general solution \( h_t = \sum_{k=1}^{K} c_k \lambda_k^t \) if all roots are unique.

Note that higher moments from (8.53) can be useful elsewhere. For example, the approximate effective capacity, for \( t \)-time slots and small \( \theta \), is \( C_{\text{HARQ}}^{\text{eff}} = -\frac{1}{\theta} \ln \left( 1 - \frac{1}{\theta} \mathbb{E}\{V_t\} + \theta^2 \mathbb{E}\{V^2_t\}/2 + \mathcal{O}(\theta^3) \right) \approx T_{t.trunc}^{\text{HARQ}} - \theta^2 \mathbb{E}\{V^2_t\}/2t \).

**8.D Proof of Corollary 8.1**

**Proof.** From Theorem 8.1, we get the expression

\[
C_{\text{Retr.}}^{\text{eff}} = -\lim_{t \to \infty} \frac{1}{\theta t} \ln \left( b^T A^{t-K} f_K \right)
\]

\[
\overset{(a)}{=} -\lim_{t \to \infty} \frac{1}{\theta t} \ln \left( b^T Q A^{t-K} Q^{-1} f_K \right)
\]

\[
\overset{(b)}{=} -\lim_{t \to \infty} \frac{1}{\theta t} \ln \left( (\lambda^t_+ f_K) b^T Q (\lambda_+)^{t-K} Q^{-1} f_K \right)
\]

\[
\overset{(c)}{=} -\frac{\ln (\lambda_+)}{\theta},
\] (8.54)

where the eigendecomposition \( A = Q \Lambda Q^{-1} \) is exploited in step (a), the largest absolute eigenvalues of \( A \), \( \lambda_+ \), is expanded for in step (b), and \( \lambda_+ \) is shown to dominate as \( t \to \infty \) in step (c). We note that Corollary 8.1 holds even if the transmission limit \( K \) increases with time \( t \), as long as \( K \) increases slower than linearly with \( t \).

**8.E Proof of Corollary 8.2**

**Proof.** Corollary 8.2 follows directly from Theorem 8.1 with \( L = 1 \), with \( P_k \neq 0 \) for \( k \in \{1, 2, \ldots, K\} \) and \( Q_K \neq 0 \), where we introduced and defined the decoding success probabilities \( P_k \triangleq P_{k11} \) and the decoding failure probability \( Q_K \triangleq P_{K01} \). Since Corollary 8.1 followed from Theorem 8.1, we can use the spectral radius of (8.13) with entries (8.14) also for Corollary 8.2.
8.F Effective Capacity Analysis of Truncated-HARQ

For the readers convenience, we give the (more tractable) analysis of the effective capacity for truncated-HARQ. The corresponding recurrence relation is given by

\[
E \{ e^{-\theta RV} \} = \sum_{v} e^{-\theta RV} P \{ V_t = v \} \\
\hspace{1cm} = \sum_{v} e^{-\theta RV} \left( \sum_{k=1}^{K} P_k P \{ V_{t-k} = v-1 \} + Q_K P \{ V_{t-K} = v \} \right) \\
\hspace{1cm} = \sum_{v} \sum_{k=1}^{K} P_k e^{-\theta R(1-v)} P \{ V_{t-k} = v-1 \} \\
\hspace{1cm} + Q_K \sum_{v} e^{-\theta RV} P \{ V_{t-K} = v \} \\
\hspace{1cm} = \sum_{k=1}^{K} P_k e^{-\theta R} E \{ e^{-\theta RV_{t-k}} \} + Q_K E \{ e^{-\theta RV_{t-k}} \}, \tag{8.55}
\]

where the relation \( P \{ V_t = v \} = \sum_{k=1}^{K} P_k P \{ V_{t-k} = v-1 \} + Q_K P \{ V_{t-K} = v \} \) is used in step (a). With \( f_{t-k} = E \{ e^{-\theta RV_{t-k}} \} \), we can then write

\[
f_t = \sum_{k=1}^{K} P_k e^{-\theta R} f_{t-k} + Q_K f_{t-K}. \tag{8.56}
\]

8.G Proof of Corollary 8.3

Proof. Descartes’ rule of signs in algebra states that the largest number of positive roots to a polynomial equation with real coefficients and ordered by descending variable exponent is upper limited by the number of sign changes. Since all \( a_k > 0 \), we only have one sign change, and thus just one positive root. In algebra, it is known that if a function \( f(x) \) has at least one positive root in the interval \([a, b] \), then \( f(a)f(b) < 0 \). Let \( f(\lambda) \) be the characteristic polynomial, we then have \( f(0)f(1) = (-a_K)(1-e^{-\theta R})\sum_{k=1}^{K} P_k < 0 \). Thus, we have one positive root, and it lies in the interval \([0, 1] \).

8.H Proof of Corollary 8.4

Proof. We rewrite (8.12) as \( \lambda_+ = e^{-\theta C_{HARQ}^{\text{eff}}} \), which is inserted into the characteristic equation (8.17), that then gives

\[
e^{-K \theta C_{HARQ}^{\text{eff}}} = e^{-\theta R} \sum_{k=1}^{K} P_k e^{-(K-k)\theta C_{HARQ}^{\text{eff}}} + Q_K
\]
\[
\Rightarrow e^{\theta R} = \sum_{k=1}^{K} P_k e^{k \theta C_{\text{HARQ}}^{\text{eff}}} + Q_K e^{(R + K C_{\text{HARQ}}^{\text{eff}})} \\
\Rightarrow (1 + \theta R) \approx \sum_{k=1}^{K} P_k (1 + k \theta C_{\text{HARQ}}^{\text{eff}}) + Q_K (1 + \theta R + K \theta C_{\text{HARQ}}^{\text{eff}}) \\
\Rightarrow R(1 - Q_K) = \sum_{k=1}^{K} k P_k C_{\text{HARQ}}^{\text{eff}} + Q_K K C_{\text{HARQ}}^{\text{eff}} \\
\Rightarrow \lim_{\theta \to 0} e^{K C_{\text{HARQ}}^{\text{eff}}} = \frac{R(1 - Q_K)}{\sum_{k=1}^{K} k P_k + K Q_K}.
\]

For (8.57), in step (a) we expanded for small \(\theta\), and in step (b), we used the fact that \(\sum_{k=1}^{K} P_k + Q_K = 1\).

8.1 Proof of Corollary 8.5

**Proof.** The characteristic equation, with the truncated geometric probability mass function inserted, is solved for \(\lambda_+\) below. We find that

\[
\lambda^K = \sum_{k=1}^{K} P_1 Q_1^{k-1} e^{-\theta R} \lambda^{K-k} + Q_1^K \\
\Rightarrow \left(\frac{\lambda}{Q_1}\right)^K - 1 = \frac{P_1 e^{-\theta R}}{Q_1} \sum_{k=1}^{K} \left(\frac{\lambda}{Q_1}\right)^{K-k} \\
\Rightarrow (\lambda/Q_1)^K - 1 = \frac{P_1 e^{-\theta R}}{Q_1} \left(\frac{\lambda}{Q_1}\right)^K - 1 \\
\Rightarrow \lambda_+ = P_1 e^{-\theta R} + Q_1, \quad (8.58)
\]

where the step (a) is due to that only one root exists.

8.8 Proof of Corollary 8.6

**Proof.** The recurrence relation for this system is \(f_t = a_1 f_{t-1} + a_2 f_{t-2}\). This is a recurrence relation for a bivariate Fibonacci-like polynomial\(^7\) [Haz01, pp. 152-154], with \(a_1 = P_1 e^{-\theta R}\), \(a_2 = P_2 e^{-\theta R} + Q_2\). The characteristic equation is \(\lambda^2 - a_1 \lambda - a_2 = 0\), with the largest positive root

\[
\lambda_+ = \frac{1}{2} \left( a_1 + \sqrt{a_1^2 + 4a_2} \right). \quad (8.59)
\]

Inserting (8.59) in (8.12), and then bringing out \(R\), gives (8.20).

---

\(^7\)The Fibonacci numbers are given with \(a_1 = a_2 = 1\), \(f_0 = 0\), and \(f_1 = 1\).
8.K Proof of Corollary 8.7

Proof. Consider (8.17), divide by $\lambda^K$, insert $\lambda = e^{-\theta C_{HARQ}^{\text{eff}}}$, and let $K \to \infty$. We then get

$$1 = \sum_{k=1}^{\infty} P_k e^{\theta(kC_{HARQ}^{\text{eff}} - R)}$$

$$(a) \approx \sum_{k=1}^{\infty} P_k (1 + \theta(kC_{HARQ}^{\text{eff}} - R) + \frac{\theta^2}{2} (kC_{HARQ}^{\text{eff}} - R)^2)$$

$$(b) \Rightarrow (C_{HARQ}^{\text{eff}})^2 + c_1 C_{HARQ}^{\text{eff}} - c_2 = 0,$$ \hspace{1cm} (8.60)

where we Taylor-expanded for small $\theta$ in the step (a), and rewrote the expression as a quadratic equation in the step (b) with $c_1, c_2$ being defined as in Corollary 8.7. We then solve the quadratic equation for $C_{HARQ}^{\text{eff}}$.

8.L Proof of Corollary 8.8

Proof. We start the proof by considering the characteristic equation for truncated-HARQ (8.17), which is $\lambda^K - \sum_{k=1}^{K} a_k \lambda^{K-k} = 0$, with variables $a_k$ from (8.14). The resulting expression is first rearranged into

$$e^{\theta R} = \sum_{k=1}^{K} P_k \lambda^{-k}.$$ \hspace{1cm} (8.61)

From the relation $C_{\text{eff}} = -\ln(\lambda_+)/\theta$, we rearrange and get $\lambda_+ = e^{-\theta C_{\text{eff}}}$. The expression for $\lambda_+$ is inserted in the first rearranged expression for $\lambda$. Next, we use the substitution $\psi = \theta(\psi) C_{\text{eff}}(\psi)$ from (8.4), which is inserted in the last expression, and we get

$$e^{R\psi/C_{\text{eff}}(\psi)} = \sum_{k=1}^{K} P_k e^{k\psi}$$ \hspace{1cm} (8.62)

We then solve this last expression for $C_{\text{eff}}(\psi)$ and finally get

$$C_{\text{eff}}(\psi) = \frac{R\psi}{\ln(\sum_{k=1}^{K} P_k e^{k\psi}) - \ln(1 - Q_K e^{K\psi})},$$ \hspace{1cm} (8.63)

which is the desired expression with the effective capacity now expressed in $\psi$.

8.M Proof of Theorem 8.2

Proof. Divide (8.17) by $\lambda^K$, and then let $t \to \infty$. This allows for $K \to \infty$, as long as $K$ increases less than linearly with $t$, and the resulting characteristic equation
for persistent-HARQ becomes
\[ 1 = \sum_{k=1}^{\infty} P_k e^{-\theta R} \lambda^{-k}. \] (8.64)

We rewrite this characteristic equation as
\[ e^{\theta R} = (a) \sum_{k=1}^{\infty} (Q_{k-1} - Q_k) \lambda^{-k} \]
\[ = (b) \sum_{k=1}^{\infty} \left( \int_{0}^{\Theta} L^{-1}_z \{ F(s)^{k-1} - F(s)^k \} dz \right) \lambda^{-k} \]
\[ = (c) \int_{0}^{\Theta} L^{-1}_z \left\{ \sum_{k=1}^{\infty} \frac{1}{\lambda F(s)} \left( \frac{F(s)}{\lambda} \right)^k \left( \frac{F(s)}{\lambda} \right)^{1-k} \right\} dz \]
\[ = (d) \int_{0}^{\Theta} L^{-1}_z \left\{ \frac{1 - F(s)}{\lambda F(s)} \right\} dz = (e) L^{-1}_\Theta \left\{ \frac{1 - F(s)}{s \lambda F(s)} \right\}, \]
(8.65)
where we used \( P_k = Q_{k-1} - Q_k \) in step (a), \( Q_k = \int_{0}^{\Theta} L^{-1}_z \{ F(s)^k \} dz \) in step (b), changed the sum and the integration order in step (c), computed the geometric series in step (d), and applied the Laplace transform integration rule in step (e). \( \square \)

8.N Proof of Corollary 8.9

Proof. We use Theorem 8.2 with \( \lambda_+ = e^{-\psi}, C_{\text{eff}}(\psi) = \psi/\theta(\psi) \), and then rearrange the expression. \( \square \)

8.O Proof of Corollary 8.10

Proof. Using Corollary 8.9, with \( F(s) = \bar{p}(s)/\bar{q}(s) \), gives the argument \( L^{-1}_\Theta \left\{ \frac{e^{\psi} \bar{q}(s) - \bar{p}(s)}{\bar{q}(s) - \bar{p}(s) e^{\psi}} \right\} \) of the logarithm. We then let \( a(s) = e^{\psi}(\bar{q}(s) - \bar{p}(s)) \) and \( b(s) = s(\bar{q}(s) - \bar{p}(s) e^{\psi}) \) and insert the coefficients in the ME-distribution-form. \( \square \)

8.P Proof of Corollary 8.11

Proof. We use the AP-method, Corollary 5.1, with \( f_\Theta = \psi^{-1} \ln (\alpha e^{\Theta} b c) \), which yields \( f'_\Theta = \psi^{-1} \ln (\alpha B e^{\Theta} b c) \). From this, \( g_\Theta = f_\Theta/\Theta f'_\Theta \) is determined. \( \square \)

8.Q Proof of Corollary 8.12

Here, we prove the effective capacity expression for RR-HARQ in Rayleigh fading expressed in \( \theta \).
Proof. Using Theorem 8.2 with \( F(s) = 1/(1 + s) \), we get the expression
\[
e^{\theta R} = \mathcal{L}^{-1}_{\hat{\Theta}} \left\{ \frac{1}{s(\lambda - 1/s)} \right\},
\]
\[\Rightarrow e^{\theta R} = \frac{1}{\lambda} e^{-\theta(1 - \frac{1}{s})},\]
\[\Rightarrow \lambda_+ = \frac{\hat{\Theta}}{W_0(\Theta e^{\hat{\Theta} + \theta R})}. \tag{8.66}\]

Taking the logarithm of \( \lambda_+ \), a more compact form is
\[
\ln(\lambda_+) = \ln \left( \frac{\hat{\Theta}}{W_0(\Theta e^{\hat{\Theta} + \theta R})} \right)
\]
\[\overset{(a)}{=} \ln \left( \frac{e^{W_0(\Theta e^{\hat{\Theta} + \theta R})}}{e^{\Theta + \theta R}} \right)
\]
\[= W_0(\Theta e^{\hat{\Theta} + \theta R}) - \hat{\Theta} - \theta R, \tag{8.67}\]
which is inserted in (8.12). For step (a), we used the definition of Lambert’s-W function, \( x/W_0(x) = e^{W_0(x)} \).

8.R Proof of Corollary 8.13

Proof. Using Corollary 8.10, together with \( F(s) = 1/(1 + s) \), allow us to write \( a(s) = e^\psi \) and \( b(s) = s + (1 - e^\psi) \). For the ME-distribution vector-matrix-form, this corresponds to \( a = e^\psi, B = 1 - e^\psi, c = 1 \), which leads to (8.36).

8.S Proof of Corollary 8.4.4

Proof. Solve (8.6) for \( \theta \) and use (8.4). (Or use Corollary 8.8 with \( P_k = P_1 Q_1^{k-1} \)).
For Rayleigh fading, we also have \( Q_1 = 1 - e^{-\hat{\Theta}} \).  

8.T Proof of Corollary 8.15

Proof. Following the described NC-ARQ operation in Section 8.4.5, the evolution of the communication modes is described by the following system of recurrence relations
\[
f_{t,1} = (Q_1^2 + P_1 e^{-\theta R})f_{t-1,1} + P_1^2 e^{-2\theta R}f_{t-1,3}, \tag{8.68}
f_{t,2} = P_1 Q_1 f_{t-1,1} + (Q_1^2 + P_1 e^{-\theta R})f_{t-1,2} + 2P_1 Q_1 e^{-\theta R}f_{t-1,3}, \tag{8.69}
f_{t,3} = P_1 Q_1 f_{t-1,2} + Q_1^2 f_{t-1,3}. \tag{8.70}\]
By ordering (8.68)-(8.70) as a matrix recurrence relation, the matrix $A$ in Corollary 8.15 is readily identified.

8.11 Proof of Corollary 8.16

Proof. The two-mode (or two-state) matrix $A$ is

$$A = \begin{bmatrix} a_{111} & a_{121} \\ a_{112} & a_{122} \end{bmatrix}.$$  

(8.71)

We solve the characteristic equation of $A$, $\lambda^2 - \lambda(a_{111} + a_{122}) + (a_{111}a_{122} - a_{121}a_{112}) = 0$, for $\lambda$.

8.12 Optimized Effective Capacity of ARQ in Rayleigh Fading

In several past works, e.g. [BS06], [SKDR13]-[LRS16a], the rate-optimized throughput was examined for block Rayleigh fading, having $P_1 = e^{-\Theta}$, $\Theta = (e^R - 1)/S$. In contrast, although the exact effective capacity expression is known for ARQ (8.6), the rate-optimized effective capacity expression is not. The (likely) reason for this is that the problem is (seemingly) intractable. Adopting the RP-method for parametric optimization in Section 5.2, we now illustrate its use for effective capacity optimization below.

Corollary 8.18. The maximum effective capacity of ARQ in block Rayleigh fading, vs. the optimal rate $R^*$, is

$$S(R^*) = \frac{e^{RC} (e^R - 1)}{\theta},$$  

(8.72)
232 Effective Capacity Analysis of General Retransmission Schemes

\[ P_1(R^*) = e^{-\Theta}, \quad \Theta(R^*) = \frac{e^{R^*} - 1}{S}, \quad (8.73) \]

\[ C_{\text{eff}}^{\text{ARQ}^*}(R^*) = -\frac{1}{\theta} \ln \left( Q_1 + P_1 e^{-\theta R^*} \right), \quad Q_1 \triangleq 1 - P_1. \quad (8.74) \]

**Proof.** We see in (8.6) that to maximize \( C_{\text{eff}}^{\text{ARQ}^*} \), we can instead minimize \( P_1(e^{-\theta R} - 1) \). Taking the derivative with respect to \( R \), equating to zero, it is easy to show that the optimality criterion is \( S\theta e^{-R} - e^{\theta R} + 1 = 0 \), and additional checks ensure a minimum for positive \( R \). While it is hard to solve for \( R^*(S) \) into a closed-form expression, it is straightforward to solve for \( S(R^*) \) instead. This is doable if a (monotonic) one-to-one mapping exists between \( R^* \) and \( S \). We then get \( S(R^*) = e^{R^*}(e^{\theta R^*} - 1)/\theta \), which is inserted in \( \Theta = (e^R - 1)/S \), which then gives \( P_1 = e^{-\Theta} \) and \( Q_1 \). The latter two are subsequently inserted in \( C_{\text{eff}}^{\text{ARQ}^*} \). □

The effective capacity expression of ARQ is plotted vs. SNR in Fig. 8.8 with \( R = \{2, 4, 6, 8\} \) [b/Hz/s] and \( \theta = 0.5 \). We also show the parameterized maximum effective capacity value and the optimal rate point (8.72)-(8.74) for ARQ and confirm their correctness.

### 8.W Multilayer-ARQ

With NC-ARQ in Section 8.4.5, we presented a retransmission protocol which may concurrently communicate multiple packets in a single transmission. Another multipacket retransmission approach is multilayer-(H)ARQ\(^8\) [PatL33], [SS08]. The idea is to concurrently transmit multiple super-positioned codewords, each with the SNR level \( x_{lS} \), fractional power allocation \( x_{rl} \), average SNR \( S \), rate \( r_l \) for level \( l \in \{1, 2, \ldots, L\} \), over a block fading channel and decode as many codewords as possible. With proper power and rate allocation, it has been found that the throughput of such schemes exceeds traditional (single-layer) ARQ, [SS08, LRS12]. The effective capacity of multilayer-ARQ is simply

\[ C_{\text{eff}}^{\text{LARQ}} = -\frac{1}{\theta} \ln \left( q + \sum_{l=1}^{L} p_l e^{-\theta R_l} \right) \]

\[ = -\frac{1}{\theta} \ln \left( 1 + \sum_{l=1}^{L} p_l \left( e^{-\theta R_l} - 1 \right) \right), \quad (8.75) \]

where \( q = 1 - \sum_{l'=1}^{L} p_{l'} \) is the probability of decoding no packet, and \( R_l = \sum_{l'=1}^{l} r_{l'} \) is the cumulative rate split. The probabilities of successful decoding, up to \( l \)th layer,

---

\(^8\) A broadcast strategy, designed with multiple (and infinite number of) decodeable layers, was investigated in [Sha97, SS03] under the assumption of a Rayleigh fading channel. It was found that the average rate increased with increasing number of layers. In a strict sense, those works did not consider (H)ARQ, as feedback was specifically excluded.
are
\[ p_l = P \left\{ \log_2 \left( 1 + \frac{(1 - X_{l+1}) S g}{1 + X_{l+1} S g} \right) > R_l \right\} = P \{ g > \Theta_l Y_l \}, \] (8.76)

where \( X_l \triangleq \sum_{l'=1}^L x_{l'} \) is the cumulative fractional power split, \( Y_l \triangleq 1/(1 - 2^{R_l} X_{l+1}) \) is a function of the power split, and \( \Theta_l \triangleq (2^{R_l} - 1)/S \). For block Rayleigh fading, we have \( p_l = e^{-\Theta_l Y_l} \). For multilayer-ARQ, exactly like for the throughput metric, analytical joint power-and-rate optimization of the effective capacity is intractable. Limiting the per layer rate to \( r_l = r, \forall l \), multilayer-HARQ, and joint network coding and multilayer-ARQ [LRS12], should be possible to handle within the analytical framework of this chapter by taking advantage of multiple communication modes.

We also note that while (8.75) has a similar form as [TZ07, (25),(26)], the latter does not consider multilayer-ARQ, but changes constellation sizes based on the current SNR in a block-fading channel.
Chapter 9

Conclusions

9.1 Summary

In this thesis, we set out to develop new models and methods for (H)ARQ performance analysis, and to offer closed-form throughput expressions for new, yet important and useful, (H)ARQ-cases not previously reported in the literature.

In Chapter 4, we derived a general throughput expression for HARQ in terms of the LT of the fading effective channel SNR pdf (or, alternatively, the pdf of the effective channel MI). We also gave closed-form throughput expressions for ARQ, and persistent/truncated-RR, for the GD-channel, as well as new analytical throughput expressions for persistent-IR, and new tight bounds for truncated-IR. An interesting finding was that ARQ, RR, and IR, all have a relatively large SNR-gap between the throughput (when throughput approaches the rate) and the ergodic capacity, which suggests the need of developing new retransmission schemes.

In Chapter 5, we offered a new, more general, parametric optimization method, the auxiliary parameter method. It was noted that it handled a wider class of analytical throughput optimization problems than the classical SNR-parameter method, as well as the rate-parameter method. We also gave new, rate-parameterized, closed-form optimal throughput expression for ARQ for MRC/OSTBC-diversity. Results from Chapter 4 were incorporated in the chapter and built upon.

In Chapter 6, we suggested a new, versatile, modulation framework, the Golden angle modulation (GAM) format, that could overcome the 1.53 dB shaping-loss of QAM and thereby practically substantiated the AWGN channel capacity modeling used throughout in the thesis, as well as in other related works. Moreover, a rank-1 lattice based design, akin to GAM, was introduced. We also considered a joint design between GAM and HARQ with random mapping rearrangement, which achieved IR-HARQ-like performance.

In Chapter 7, we proposed the matrix exponential distribution as a new, more general, fading channel model for performance analysis of (H)ARQ systems in particular, and wireless systems more generally. Numerous, both general and specific, (H)ARQ-cases were investigated, such as for interferers with ME-distributed SNRs,
for which closed-form throughput expressions were given. Results from Chapters 4-6 were integrated in the chapter. An insight gained was that by choosing a suitable mathematical formulation for the fading channel model, here the ME-distribution, the performance analysis could be unified and simplified.

In Chapter 8, we generalized the (H)ARQ assumption, and modeled a substantially more general retransmission system accounting for multiple transmissions, multiple communication modes, and multiple rate increments. We gave a new closed-form effective capacity expression, which includes throughput as a degenerated case, where a compact and easy-to-design companion matrix, comprising sums of transition-probabilities, played a central role. Numerous (H)ARQ-cases were investigated, and closed-form effective capacity expressions were given, e.g. in terms of the LT of the effective channel SNR (or MI) and ME-distribution-based channel parameters. Results from Chapters 4-7 were incorporated in the chapter.

It is instructive to organize the ideas developed in the thesis wrt performance analysis models and methods, as well as the communication mechanisms developed. This is illustrated in Table 9.1, where the proposed concepts are contrasted with more classical approaches encountered in the literature. Thus, in a single sentence, the thesis; offers an effective capacity (and throughput) performance analysis framework, handles general retransmission schemes (beyond ARQ, RR, and IR), models iid fading channel gains with the broad-class of ME-distribution-based pdf, expresses the performance metric in ME-distribution-based channel parameters, provides parametric closed-form performance optimality expressions, backs-up the underlying AWGN channel capacity model assumption for the more detailed performance analysis, and suggests a reduced complexity IR-performance-like HARQ-scheme. In Fig. 9.1, an illustrative, one-page, summary of the thesis is sketched.

9.2 Future Research

In this thesis, we have already tried to advance many research items, wrt models and methods, where a need and opportunity for development could be identified. However, some short-term extensions, directly related to the thesis, and some more long-term ideas, outside the scope of the thesis, can nevertheless be envisioned.

9.2.1 Short-term Research Items

We first consider future, short-term, research items and arrange them in terms of performance analysis models and methods, as well as communication mechanisms.

Models

The performance analysis models can be extended as follows:

- Performance metric: For more realistic modeling, the effective capacity metric, and related analysis, could be revised to allow for varying, rather than
9.2. Future Research

<table>
<thead>
<tr>
<th>Categories</th>
<th>Items</th>
<th>Classical use</th>
<th>Thesis concepts</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Performance analysis models.</strong></td>
<td>Performance measure.</td>
<td>Throughput.</td>
<td>Effective capacity. E.g. able to handle truncated-HARQ.</td>
</tr>
<tr>
<td></td>
<td>Retransmission scheme model.</td>
<td>Scheme-specific.</td>
<td>General transition probability $P_{k\nu l\tilde{l}}$.</td>
</tr>
<tr>
<td></td>
<td>Fading channel model.</td>
<td>E.g. Rayleigh or Nakagami-$m$.</td>
<td>ME-distr. channel, $f_C(g) = pe^{gQ}r$.</td>
</tr>
<tr>
<td></td>
<td>Communication signal model.</td>
<td>2D-Gaussian r.v., unclear modulation.</td>
<td>2D-Gaussian r.v., GB-GAM model.</td>
</tr>
<tr>
<td><strong>Communication mechanisms.</strong></td>
<td>Basic HARQ performance analysis.</td>
<td>Summing of outage probabilities.</td>
<td>Direct LT-based.</td>
</tr>
<tr>
<td></td>
<td>Retransmission scheme.</td>
<td>IR-HARQ: low-rate punctured code.</td>
<td>MR-HARQ with GAM.</td>
</tr>
<tr>
<td></td>
<td>Modulation.</td>
<td>QAM, APSK, non-uniform QAM.</td>
<td>GAM-framework, Rank-1 lattice.</td>
</tr>
</tbody>
</table>

Table 9.1: Proposed thesis concepts wrt models, methods, and mechanisms.

fixed, source rate. Alternative performance measure, to effective capacity and throughput, could also be explored.

- Retransmission model: The retransmission model in Chapter 8, defining the operation with transition probabilities $P_{k\nu l\tilde{l}}$, is very general, but does not account, e.g., for (H)ARQ adaptation based on CSI. Modification of the performance analysis model, and the method, may be needed.

- Fading channel model: (H)ARQ performance analysis for correlated block-fading is generally of interest. For this purpose, the communication modes occurring in the transition probabilities in Chapter 8 could be used and represent different fading channel statistics. Alternatively, if possible, the ME-distribution-based channel model may be extended to cater for time-dependent correlation.

- MI-model: In the thesis, the MI is modeled as the AWGN channel capacity. This relies on infinitely long channel codes, which is unrealistic. For more
realism, the MI could also be modeled wrt finite size channel codewords, discrete modulation constellations, and non-Gaussian noise-sources.

Methods

The performance analysis methods can be extended as follows:

- Retransmission performance analysis method: In some instances, the recurrence relation-based performance analysis method in Chapter 8 may require changes. This may be the case if, e.g., the performance metric is changed, the CSI affects the transition probabilities, or time-truncated-HARQ with packets having source-stamped due times is used.

- Performance analysis framework wrt fading: The throughput analysis of HARQ with ME-distributed signal and interferers is of general interest to solve, but could not be solved in the thesis. Hence, further development of the ME-distribution-based performance analysis framework wrt interference is desired. It is also of interest to advance the ME-distribution-based channel model, and performance analysis framework, for more (H)ARQ-cases.

- Basic HARQ performance analysis: It would be desirable to determine closed-form throughput (or effective capacity) expressions for some basic (H)ARQ-cases, such as IR, CDD-ARQ/RR/IR, and MIMO-ARQ/RR/IR, all with iid complex Gaussian channel amplitude (entries). However, it is not known, nor disproved, if closed-form throughput expressions exist for those cases.

Mechanisms

One immediate communication mechanism is identified.

- GMR-HARQ: In Chapter 6, MR-HARQ with GB-GAM was considered. While numerical MI-based simulation demonstrated IR-HARQ-like throughput performance, channel en/decoding (and mapping) methods need to be developed.

9.2.2 Long-term Research Items

We now shift the focus to ideas on long-term research for performance analysis models and methods, but also for retransmission mechanisms.

Models

- Energy-efficient (H)ARQ: Future wireless communication systems need to consider energy efficiency and suitable performance metric(s), and optimization framework(s), should be developed. For example, energy consumption could be minimized under a throughput (or eff. capacity) constraint. It would be desirable if the recurrence-relation method could be reused.
9.2. Future Research

- (H)ARQ in advanced network topologies: The (H)ARQ performance analysis models and methods, given in the thesis, could be extended to one-to-many, many-to-one, and cellular system topologies. Such (H)ARQ models and methods could also be developed for other topologies involving opportunistic routing, network coding, relaying, multihopping, cooperative communication.

Methods

- Advanced performance analysis framework wrt fading: The current ME-distribution-based channel model determines a scalar (exponential-polynomial-trigonometric) function by a matrix exponential and two constant vector parameters. This basic idea could be tweaked to model even more complicated scalar functions. To exemplify, consider a form like $f_C(c) = \mathbf{p}\exp(\exp(c\mathbf{Q}))\mathbf{r}$, with two matrix exponential functions. This may allow the channel capacity pdf of more complex cases, perhaps even $N \times N$-MIMO with complex Gaussian channel matrix entries, to be modeled in closed-form.

- Time-truncated HARQ: Time-truncated HARQ, where each packet is given a due-time by the source, could be studied. Such system would require new queueing-based modeling and analysis to be developed.

- MIMO-(H)ARQ: It is of interest to develop improved techniques that allows determination (if possible) of exact outage probability $Q_1$, decoding failure probability $Q_m$, or transition probability $P_{k\nu\ell}$ expressions for MIMO-(H)ARQ. Alternatively, approximation techniques and expressions, such as our RMT-based (H)ARQ MI-approximation in [LRS14b], could be considered. The emerging area of massive-MIMO, offers accurate MI-approximations.

Mechanisms

- MIMO-(H)ARQ schemes: The practical design of improved, low-complexity, near channel capacity MIMO-ARQ, -RR, and -IR scheme are of general interest, and particularly so for massive-MIMO. As the number of antennas increases, the channel hardens, and the capacity becomes more deterministic and predictable. It could be so, as we suggest in [LRS14b], that the trade-off between performance and complexity speaks for MIMO-ARQ, rather than MIMO-RR, or -IR.

- Alternatives to (H)ARQ: The SNR-gap between the throughput and the ergodic capacity seen in Chapter 4 in the high throughput range could be reduced if the time-slotted frame structure is avoided. For example, multiple HARQ packets could be allowed to share the same time-slot, or rateless coding without a time-slot structure could be explored. Another alternative, albeit with a time-slotted structure, could be a time-staggered superposition coding of multiple packets with feedback-based power control. Some works already exist in those directions, but could be extended.
Ch. 4: LT-approach and GD-analysis
a) Basic throughput performance analysis
b) A new (direct) LT-based throughput expression
c) New closed-form throughput expressions
\[ T_{\infty}^{\text{HARQ}} = R \left( \frac{1}{s} \right) \left( \frac{1}{1 - F(s)} \right) \]

Ch. 5: Auxiliary parametric opt.
a) A more general parametric optimization method
b) New closed-form optimality throughput expressions

Ch. 6: Golden Angle Modulation
a) A novel modulation framework
b) AWGN channel capacity model motivation
c) A GAM-based HARQ-scheme w. IR-like perf.
\[ \varphi = \frac{3 - \sqrt{5}}{2} \]
\[ \sigma_m = r_m e^{2\sigma \varphi m} \]

Ch. 7: Matrix Exponential Distr.
a) A novel (ME-dist.) fading channel pdf modeling
b) A novel ME-dist.-based perf. analysis method
c) New closed-form expressions

Ch. 8: Recurrence-relation analysis
a) A general retransmission system model
b) A novel eff. capacity perf. analysis method
c) New closed-form eff. capacity expressions

Figure 9.1: Sketch of the basic ideas proposed in Chapters 4-to-8.


Bibliography: Author’s Patents


259


[PatL67] P. Larsson and M. Prytz, “Method and system of communications,”

[PatL68] ——, “Method for allocation of communication parameters in a multiuser
wireless communications system,” Granted Patent 8 200 266 B2, Jun 29,


280-081-558-026

[PatL71] P. Larsson and N. Johansson, “Forwarding node in a wireless

[PatL72] L. Peter, “Interference-considerate scheduling in a wireless communica-
Available: https://lens.org/081-337-528-543-057

[PatL73] P. Larsson and H. Sahlin, “Methods and arrangements in a wireless
communication system for producing signal structure with cyclic prefix,”

770

[PatL75] N. Johansson, G. Mildh, P. Frenger, and P. Larsson, “Methods and
arrangements for frequency selective repetition,” Granted Patent 8 718 540
28X

https://lens.org/021-805-554-744-897

[PatL77] J. Gan, P. Larsson, and Q. Miao, “Filter or amplifier adaptation by an
intermediate device in a multi-hop system,” Granted Patent 8 570 889 B2,

[PatL78] P. Frenger, P. Larsson, and N. Johansson, “Insertion of signals by an


In the list above, only one granted patent (or patent application) is listed per family, unless the family has been deemed consisting of multiple inventions. If a granted US patent exists, it is given, or else the granted EP patent is given. If neither a granted US, nor EP, patent exist in the family, and despite that a granted non-US/EP patent may exist, the WO patent application is given. In some cases, the inventor has invented different inventions at different times, but the patent owner has requested priority from an earlier filed patent application, and placed the new invention and patent application in the same patent family for priority date reasons.