PERFORMANCE ANALYSIS OF BLOCK CODES OVER FINITE-STATE CHANNELS IN DELAY-SENSITIVE COMMUNICATIONS

A Dissertation

by

FATEMEH HAMIDI SEPEHR

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Henry D. Pfister
Co-Chair of Committee,	Jean-Francois Chamberland
Committee Members,	Srinivas Shakkottai
	N. Sivakumar
Head of Department,	Chanan Singh

May 2014

Major Subject: Electrical Engineering

ABSTRACT

As the mobile application landscape expands, wireless networks are tasked with supporting different connection profiles, including real-time traffic and delay-sensitive communications. Among many ensuing engineering challenges is the need to better understand the fundamental limits of forward error correction in non-asymptotic regimes. This dissertation seeks to characterize the performance of block codes over finite-state channels with memory and also evaluate their queueing performance under different encoding/decoding schemes.

In particular, a fading formulation is considered where a discrete channel with correlation over time introduces errors. For carefully selected channel models and arrival processes, a tractable Markov structure composed of queue length and channel state is identified. This facilitates the analysis of the stationary behavior of the system, leading to evaluation criteria such as bounds on the probability of the queue exceeding a threshold. Specifically, this dissertation focuses on system models with scalable arrival profiles based on Poisson processes, and finite-state memory channels. These assumptions permit the rigorous comparison of system performance for codes with arbitrary block lengths and code rates. Based on this characterizations, it is possible to optimize code parameters for delay-sensitive applications over various channels. Random codes and BCH codes are then employed as means to study the relationship between code-rate selection and the queueing performance of point-to-point data links. The introduced methodology offers a new perspective on the joint queueing-coding analysis for finite-state channels, and is supported by numerical simulations.

Furthermore, classical results from information theory are revisited in the context

of channels with rare transitions, and bounds on the probabilities of decoding failure are derived for random codes. An analysis framework is presented where channel dependencies within and across codewords are preserved. The results are subsequently integrated into a queueing formulation. It is shown that for current formulation, the performance analysis based on upper bounds provides a good estimate of both the system performance and the optimum code parameters. Overall, this study offers new insights about the impact of channel correlation on the performance of delay-aware communications and provides novel guidelines to select optimum code rates and block lengths.

DEDICATION

To my lovely husband, Salman, and my family.

ACKNOWLEDGEMENTS

I would like to express my special thanks to all who supported me throughout my journey to the PhD degree.

First and foremost, I would like to extend my sincerest thanks and appreciations to my advisors, Professor Henry Pfister and Professor Jean-Francois Chamberland. I was honored with their continuous guidance, supervision, assistance, patience and above all encouragement in every single discussion we had since the first day of joining their group. Their strong mathematical background, maturity in thinking big picture, and astonishing intuition through an extremely wide range of research areas have been the most significant help in my career. I learned from them not only about academics and research, but also about graciousness, optimism, flexible discipline, and hard work in my life.

I am also very grateful to Professor Krishna Narayanan for his generous help and thoughtfulness along the way. His valuable comments and suggestions introduced to me new perspectives in my career. Besides his incredible intuition and deep understanding in coding theory and information theory, his great enthusiasm to help and support the graduate students has always amazed me. He has a unique way of following up with students problems and issues, which taught me big lessons.

I express my sincere thanks to Dr. Shakkottai and Dr. Sivakumar for their time in serving on my committee, and carefully going through my dissertation. Their generous support has greatly improved the quality of my research.

I would also like to thank the faculty and staff in the electrical and computer engineering department for making my PhD years at Texas A&M University a great experience. My special thanks go to Tammy Carda, Gayle Travis, and Jeanie

Marshall in ECE department for patiently answering our questions day by day and helping us in any way they could.

I would like to thank all my friends and colleagues in the group, especially Yung-Yih Jian, Arvind Yedla, Santhosh Kumar, Jason Moore, Avinash Vem, Andrew Young, Phong Sy Nguyen, and Byung-Hak Kim. I thank them for all the discussions we had, all the problems we solved together, and all the fun we had together. I wish all of them the bests.

Last but not least, countless thanks to my amazing husband, mother, father, sister, and two brothers for their unconditional love and continuous encouragement. They have believed in me and supported me throughout.

TABLE OF CONTENTS

Page

LIST OF FIGURES

CHAPTER I

INTRODUCTION

I.1 Background Framework

With the ever increasing popularity of advanced mobile devices such as smartphones and tablet personal computers, the demand for low-latency, high-throughput wireless services is growing rapidly. The constantly increasing demand for ubiquitous Internet access and mobile communication is pushing the limits of current communication infrastructures. This demand is fueled, partly, by the growing popularity of real-time applications such as voice-over-internet-protocol (VoIP), video conferencing, gaming, electronic commerce, control and actuation.

As the popularity of real-time applications rises, new paradigms that maximize throughput subject to quality of service (QoS) constraints are becoming highly desirable. Designing wireless systems for delay-sensitive applications with quality of service guarantees is an ambitious task. Contemporary wireless communication systems must be designed to accommodate the wide range of applications that compose today's digital landscape. Modern mobile devices should be able to support heterogeneous data flows with a variety of delay and bandwidth requirements. The shared desire for a heightened user experience, which includes real-time applications and mobile interactive sessions, acts as a motivation for the study of highly efficient communication schemes subject to stringent delay constraints. While point-to-point channels have received much attention in the past, the asymptotic approaches favored by classical information theory offer only limited insights on efficient designs in the context of delay-sensitive communications. Indeed, real-time traffic and live interactive sessions are typically subject to very stringent delay requirements. An

important aspect of delay-sensitive traffic stems from the fact that intrinsic delivery requirements may preclude the use of very long codewords. Such constraints can hardly be captured by asymptotic regimes where block lengths and, consequently, delay become unbounded. As such, the insights offered by classical information theory and based on Shannon capacity are of limited value in this context. This necessitates merging the study of different aspects of the network theory such as queueing, scheduling, throughput, and delay with those of the information theory such as channel capacity, coding, and power control. In fact, joint optimization of both network and physical layers is strongly required in design of such systems.

Queuing theory usually deals with the design of communication links with quality of service by considering an idealized pure queueing system that ignores the physical layer. On the other hand, information theory considers the design and analysis of channel codes with low decoding error probability for time-varying channels. Although cross-layer designs have been proposed to merge these two approaches, they are limited to specific cases. There have been few attempts to combine these two trends to specify the ultimate limit of delay-constrained communications. In the following, we present an elaborate survey of pertinent prior research contributions.

I.2 Literature Review

Forward error-correcting codes have historically played an instrumental role in digital communication systems by providing protection against channel uncertainties. For instance, it is well-known that for rates below capacity, one can improve transmission reliability by increasing the block length of a code. There is a tradeoff between the improvements offered by low-rate codes and the payload reduction associated with an increase in redundancy. Finding a suitable balance between these two intertwined considerations is a fundamental pursuit in coding theory. The

Shannon capacity, for instance, characterizes the maximum achievable throughput a channel can support subject to an asymptotic reliability constraint as block length tends to infinity [1].

Due to the delay requirements of certain modern applications, one may be forced to employ schemes with short codewords. While sometimes necessary, short codes can preclude the concentration of empirical measures for errors and channel state occupancy. This may, in turn, engender excess decoding failures and undetected errors. Furthermore, these undesirable events may be correlated in time, thereby causing queue buildups at the source that induce unacceptable delays at the destination. The latter issue is especially important for channels with memory, as correlation in service is known to exacerbate deviations in queueing systems. This discussion points to the need to carefully explore the tradeoffs between queueing and coding for communication systems with tight delay requirements, giving due consideration to optimal block lengths and code rates.

At this point, it is important to note that several recent contributions to communication theory seek to address the tradeoffs between average power, throughput, rate, distortion, and delay [2–14]. For example, in [9], the problem of delay-limited throughput maximization with a constraint on the expected waiting-time is considered. The trade off between queueing and coding delays is studied and a queue management technique for fading channels is proposed. In [10], Dunn and Laneman contribute a rate-delay trade off region by introducing some information theoretic bounds for the overhead of varying code rate. However, the protocols used in this paper are inherently limited to the BEC with feedback, and cannot be generalized to more general channels.

Many such articles make idealized assumptions about the performance of coded transmissions. These assumptions are often reasonable for long codewords, but

they are not necessarily justified for low-latency communication over channels with memory. Delay-sensitive systems have been studied in the past, leading to several landmark contributions [15–18]. Indeed, [15, 16, 18] are among the first papers in which the importance of a unifying approach to information (coding) and queuing theories has been investigated. In [15], Ephremides and Hajek reviewed several topics that are related to communication networks with an information-theoretic flavor, including multiaccess protocols, timing channels, effective bandwidth of bursty data sources, deterministic constraints on data streams, queuing theory, and switching networks. They also described the relationship between information-theoretic ideas and networking. Before this thorough survey, delay-sensitive communications with bursty traffic had not received much attention in the information theory literature.

The communication of delay-sensitive bits has been addressed recently under various assumptions and settings. In particular, delay analysis for communication over wireless fading channels has gained popularity [2, 6–8, 11, 13, 19]. Often, asymptotic approximations are employed to enable a tractable analysis of the problem. Below, we detail some of the existing work with their corresponding settings.

Several articles in this area, emphasize delay-constrained information capacity [11, 12, 17, 18, 20]. For instance, Anantharam and Verdu, analyze the Shannon capacity of the single server queue in [18]. The capacity is shown to be the lowest under exponential service time distribution. They also prove that this capacity cannot be increased by feedback. For general service time distributions, upper bounds on the Shannon capacity are derived. In [11], Negi and Goel present an approach to obtains a joint QoS exponent by merging the queueing and information-theoretic models and using a streaming code such as a convolutional code. Ideas from large deviation theory (effective capacity [21]) and information

theory (random coding error exponents), are combined. The goal is to find the code-rate allocation that maximizes the decay rate of the asymptotic probability of error for a given asymptotically large delay requirement. Real-time streaming broadcast is considered over the downlink of a single cell in [20]. The authors study the broadcast capacity of the system under deadline constraints for both uncoded and coded wireless broadcast schemes. In [12], the authors present an analysis framework that accounts for codeword retransmission in the performance analysis of fading channels in terms of maximum zero-outage throughput and maximum-throughput. The throughput problem is posed in terms of a single server queue, and it is connected to the transmission rate and the average codeword service time.

When transmitting stochastically arriving data over fading channels, there is an inherent tradeoff between the required average transmission power and the average queueing delay experienced by the data. This tradeoff can be exploited by appropriately scheduling the transmission of data over time [8]. For example, the use of advanced power-control policies can be tailored to the needs of various applications [2, 4, 7, 8, 22]. In many such articles, the emphasis is put on average delay and the optimization objective naturally leads to dynamic programming formulations [2, 3, 7, 23, 24].

For instance, Bettesh and Shamai optimize the system average delay by using combined power/rate control under average power constraints and fixed transmission rate, in [2]. The minimization is performed using dynamic programming and asymptotic analysis has been provided. Efficient power control schemes for delay sensitive communication over fading channels are also studied in [4] with the goal of optimizing the link layer performance under a long-term average power constraint. In [7], Berry and Gallager consider communication over time-varying fading channels with perfect channel state information (CSI). The transmission rate and power are

chosen according to the CSI and the buffer occupancy. Their goal is to control both the long-term average transmission power and the average buffer delay using dynamic programming. The behavior of this tradeoff is then quantified in the regime of asymptotically large delay. In [8], Berry studies the behavior of the optimal power-delay tradeoff for a single user in the regime of asymptotically small delays. Particularly, he lower bounds how much average power is required as a function of the average queueing delay. In [3], optimal policies in terms of transmission power and packet queueing delay over fading channels with feedback are characterized. In [5, 6], Rajan et al. derive optimal delay-bounded schedulers for transmission of constant-rate bursty traffic over finite-state fading channels, with the focus on minimum power requirements.

We stress that the wide applicability of Little's law can be leveraged to simplify the analysis of many systems where average delay is a prime consideration [24,25]. In [25], the average delay performance of a block coded communication over memoryless channels is analyzed. Therein, Swamy and Javidi, model the communication system as a queueing system with bulk service (similar to [26]), and optimize the expected delay using some choice of forward error correction scheme. Similar to [10] and [2], they also consider acknowledgment in the form of a request for retransmission. Unlike [27] their notion of delay includes the queueing delay and accounts for some sort of randomness in the inter-arrival times of bits. In [24], a cross-layer approach for a system with a single-server queue concatenated with a multilayer channel encoder is studied. The focus is on the minimization of the average delay of a packet, from the moment it enters the queue until completion of successful service. Bounds are derived for the expected delay as function of the numbers of coded layers. It is shown that code layering may yield performance gains in terms of average delay.

On a different note, some of the existing articles in the area of delay-sensitive

communications study the effect of feedback on system performance [3, 27]. For example, the advantages of using feedback from an integrated information-theoretic and queuing-theoretic perspective is studied in [3]. It has been shown that the use of feedback yields improvements in the power-delay tradeoff over systems without feedback. In [27], systems with fixed end-to-end delay are studied and it has been shown that feedback generally provides gains in terms of error exponents. An upper bound is provided on the probability of symbol error in a fixed-delay communication system with feedback. In the following, we briefly go through the aforementioned publications and summarize describe their main contributions.

It is also interesting to categorize existing works from a different, yet equally important perspective. In this regard, the first group of works consists of scenarios where the current channel state information (CSI) is assumed to be known at both the transmitter and receiver [6–8, 11]. For example, in [7, 8], Berry and Gallager address the power-delay tradeoff over a Markovian fading channel with CSI both at the transmitter and the receiver. In such a setting, the transmitter dynamically varies power (i.e., the rate) in response to the current queue length and channel state. In [6], Rajan et al. derive optimal delay-bounded schedulers for transmission of constant-rate traffic over finite-state fading channels. Similar to [7, 8], the proposed dynamic code-rate allocation in [11] is in response to the current channel fading and relies on CSI knowledge at the transmitter. Another group of works (e.g., [2, 19]) focuses on scenarios where CSI is unknown to the transmitter, but there is a mechanism for retransmission of codewords when the channel is in outage. In [2], the authors provide asymptotic analysis for the optimal adaptive policies that adjust the transmission rate and/or transmission power in response to the current queue length at the transmitter. In another example, Liu et al. in [19] study the problem of optimal (fixed) transmission rate to maximize the decay rate of the probability

of buffer overflow for on-off channels and Markov-modulated arrivals. The channel is considered off, when an outage occurs. In [13], Kittipiyakul et al., analyze the high-SNR asymptotic performance of outage-limited communications with fading under stringent delay constraint. They use a similar performance measure to [11], namely the decay rate of the asymptotic probability of error. However, they cover scenarios in which CSI is not readily available to the transmitter and there is no retransmission. In such a setting, the variation of the fading channel is combatted via coding over multiple independent fading realizations. While this approach improves the transmission reliability, its longer coding duration increases the end-to-end delay any bit faces, and can potentially increase the probability of delay violation. In other words, in the absence of CSI and retransmission, the transmission reliability as well as the delay violation probability are functions of the coding rate and duration.

On a different note, the recent advent of network coding and the complementary approach of channel coding over networks have been applied to delay-sensitive communications [20,26,28–35]. In [28], the performance gains of optimal transmission strategies in an unreliable packet network setting using random network codes are considered. The strategies are examined in the presence and absence of coding capability at the transmitter and performance is evaluated in delay and throughput. Closed-form as well as asymptotic expressions are provided for the delay performance with and without network coding. It is shown that the network coding capability can lead to better delay performance. In [29], analytical bounds on the completion time and stable throughput for random linear coding are provided across multiple multicast sessions.

In [30], Cogill and Shrader analyze the average queue backlog at a source node serving a single multicast flow consisting of multiple destination nodes. In [32], the performance of random linear network coding for time division duplexing channels

with Poisson arrivals is studied. The system is modeled as a bulk-service queue with variable bulk size. The effect of the allowable bulk sizes on the mean queue length is studied, and it is shown that there is an optimal choice for the allowable range that minimizes the queue length.

In [26], Shrader and Ephremides present a queueing model for a random linear coding scheme with different amount of stochastic traffic at the source node with an infinite-capacity buffer. They mainly seek to characterize the answer to the question, "Does the throughput gain offered by network coding result in a penalty in terms of the delay?". In [33], a lossy communication channel for unicast with zero-delay feedback is considered. This work is an extension of [26], which is obtained by deriving an expression for the delay of random linear coding over a field of infinite size and refining the delay estimates produced in [26]. In [34], the analysis of [33] is extended by comparing delay in protocols using retransmission and packet encoding for bidirectional unicast communication over a satellite channel with high propagation delay. In [35], the queueing delay performance is analyzed when random linear coding is performed over packets for multicast transmission over packet erasure channels.

Many of the works in the in the area of delay-sensitive communications, including several of those adopting the idea of network coding, consider a stream of data (bits or symbols) arriving at and leaving the transmitter. These works do not usually consider block coding over the arriving data. Instead, they employ streaming (i.e., convolutional) codes to protect data. Hence, they do not allow characterization of block code parameters in terms of the delay/throughput performance of the system, which is the main focus of this dissertation (e.g., see [11, 20]).

From a different perspective, many of the proposed schemes, such as [28, 29], seem especially well-suited for packet-loss networks, and the ensuing framework

represents a potential alternative to automatic repeat requests when feedback is slow or error-prone. Although closely related, these contributions differ from our formulation in that the main focus is on the operation of the system at the packet level, whereas we seek to characterize the impact of channel behavior at the symbol level.

I.3 Contributions

In the first chapter of this dissertation, the impact of certain coding strategies on the queueing performance of finite-state channels with and without memory is studied. This is accomplished without resorting to characteristic, simplifying assumptions about the operation of coded transmissions.

Optimum code-rate selection has previously been studied for Gilbert-Elliott erasure channels with Bernoulli arrivals and maximum-likelihood decoding [36]. This prior line of work centers around random codes of fixed lengths, and it offers a distinct approach to assess the performance of communication systems operating over erasure channels. In the first chapter we introduce a significant extension to these existing contributions in that we examine finite-state error channels, we leverage pragmatic coding schemes and we adopt a scalable arrival profile. The first important distinction between the present findings and previously results is the rigorous characterization of queueing behavior for communications over finite-state error channels, as opposed to erasure channels. This is an important and nontrivial extension, which arises through the fact that erasure channels intrinsically pinpoint the location of channel distortion events at the receiver whereas error channels do not. This lack of location information renders the decoding process much more challenging in the latter case. Although technically more demanding, error channels permit the more realistic modeling of practical communication links. For example, in our

analysis, detected and undetected errors both demand appropriate considerations.

In addition to this channel enhancement, we leverage pragmatic coding schemes such as BCH encoding with bounded distance decoding to bring a pragmatic flavor to the analysis. Furthermore, under random coding schemes, we present a novel framework to analyze overall system performance (e.g., probability of buffer overflow) using both optimal decoding and minimum distance decoding. This perspective is very beneficial because the probability of decoding failure plays a crucial role in characterizing packet departures, queue transitions and the stationary behavior of the transmit buffer.

Another prime distinction between contribution presented in this dissertation, and previous work is the adoption of a scalable arrival profile which is formed based on the Poisson process. Among other advantages, the proposed framework allows for the rigorous comparison of coding schemes with different block lengths, something that could not be done before. Indeed, this appears to be the first time one can perform the rigorous optimization of queueing performance over block lengths. In [36], the arrival process is Bernoulli and it is intrinsically tied to codeword transmissions. As such, arrivals are implicitly linked to the block length. By adopting a Poisson (or Markov modulated Poisson) model, we are able to overcome these limitations. We emphasize that the scaling property of the Poisson process is crucial in enabling the fair comparison of systems with different block lengths. The price to pay for this additional flexibility is a slightly more complicated analysis. In particular, we need to employ an advanced version of the matrix-geometric method. Our analysis leads to an enhanced framework for code design and resource allocation in the context of delay-sensitive wireless communications.

One of the challenges in dealing with block codes over finite-state channels with memory is the time dependencies among proximate decoding events. For instance, if

the underlying channel forms a Markov chain, then the decoding process becomes a hidden Markov process as block codes operate over series of channel states. This often entails a difficult analysis of the queue behavior at the source. To make this problem tractable, we use the idea of state augmentation which was also used in [36], where the value of the channel at the onset of a codeword is appended to the queue length. Under this state augmentation, the coded system retains the Markov property, which facilitates the precise characterization of the queueing behavior at the transmitter. This approach is paralleled in the present dissertation, albeit in the general context of error channels.

We review and extend the necessary mathematical machinery to handle error events, as opposed to erasures, starting with the binary memoryless channel. This step is pivotal in better understanding the encoding/decoding analysis of communication links with errors. We then turn to finite-state channels with memory, as originally introduced by Gilbert [37] and Elliott [38]. We leverage the latter abstractions to assess how channel dependencies over time can affect overall performance. It is well-known that correlation in service can significantly alter the behavior of a queueing system or network; such changes should be expected in the present scenario as well. Still, a novel facet of the problem we are considering is the study of how such dependencies affect the selection of optimal design parameters in terms of code rate and block length. Furthermore, the framework presented in this dissertation can be used to derive novel and fundamental bounds on the maximum arrival rate that a wireless system can support when subject to certain quality of service requirements.

As mentioned earlier, a prime goal of this dissertation is to develop a better understanding of delay-constrained communication, queue-based performance criteria and service dependencies attributable to channel memory. In particular, in

the second chapter we seek to derive meaningful performance limits for delay-aware systems operating over channels with memory. The emphasis is put on identifying upper bounds on the probabilities of decoding failure for systems employing short block lengths. This is an essential intermediate step in characterizing the queueing behavior of contemporary communication systems, which is also done in this chapter.

Computing the probability of decoding failure for specific communication channels and particular coding schemes is of great interest. This line of work dates back to the early days of information theory [39] and has received significant attention in the past, with complete solutions in some cases. One approach that has been remarkably successful, consists of deriving exponential error bounds on the behavior of asymptotically long codewords. This approach was popularized by Gallager [1] and these bounding techniques have been applied to memoryless channels as well as finite-state channels with memory. In general, they can be very accurate for long, yet finite block lengths. It is worth mentioning that the subject of error bounds has also appeared in many more recent studies, with the advent of new approaches such as dispersion, the uncertainty-focusing bound, saddlepoint approximation [27, 40–43]. This renewed interest in the performance of coded transmissions points to the timeliness of the topic under investigation.

A distinguishing feature of the approach that we wish to develop is the focus on indecomposable channels with memory and state-dependent operation. In most prevalent asymptotic frameworks, channel parameters are kept constant while the length of the codeword increases to infinity. Although this point of view leads to mathematically appealing characterizations, the resulting bounds on error probabilities do not depend on the initial or final states of the underlying communication channel. This situation can be explained through the fact that, no matter how slow the mixing time of the physical channel is, the duration of a

codeword eventually far exceeds this quantity. Still, in many practical scenarios, the service requirements imposed on a communication link forces the use of relatively short codewords, with no obvious time-scale separation between the duration of a codeword and the mixing time of the channel.

The mismatch between existing techniques and commonly deployed systems, together with the growing popularity of real-time applications on wireless networks, demands a novel approach where the impact of boundary conditions are preserved throughout the analysis. A suitable methodology should be able to capture both the effects of channel memory as well as the impact of the channel state at the onset of a codeword. In the second chapter of this dissertation, we are interested in regimes where the block length is of the same order or smaller than the coherence time of the channel. Formally, we wish to study the scenario where the mixing time of the underlying finite-state channel is similar to the time necessary to transmit a codeword. This leads to two important phenomena. First, the state of the channel at the onset of a transmission has a significant impact on the empirical distribution of the states within a codeword transmission cycle. Second, channel dependencies may extend beyond the boundaries of individual codewords. This is in stark contrast with rapidly mixing channels where initial channel conditions have no effects on the probability of decoding failures, and with block-fading models where the evolution of the channel is independent from block to block. Our proposed framework is rich enough to account for scenarios where decoding events are strongly correlated over time. Dependencies from codeword to codeword give rise to correlation in decoding failure events and can therefore greatly affect perceived service quality.

It is important to note that the standard asymptotic regimes naturally result in concentration of channel quality over a codeword transmission time. This enables the application of the law of large numbers to compute the mutual information,

the convergence rate, etc. However, in non-asymptotic analysis of channels with memory, variations of the channel quality over a codeword transmission time need careful consideration. At the same time, high correlations make the channel behavior more predictable. This in turn, paves the way to implement adaptive opportunistic scheduling schemes. Our proposed methodology, called the rare-transition regime, forms a powerful framework which enables to leverage asymptotic in bits, without transcending asymptotic in time. In other words, the rare-transition regime facilitates non-asymptotic analysis in high-correlation mode of operation.

At this point it is worth mentioning that the rare-transition regime, alternatively called the slow-mixing regime, have been studied in the past, for different frameworks such as channel estimation, asymptotic filtering, and entropy rate of a hidden Markov process [44–46]. In the second chapter of this dissertation, we examine probabilities of decoding failure, their distributions and temporal properties within the context of rare-transition regime. The purpose of deriving upper bounds on the probabilities of decoding failure for rare transitions is to capture overall performance for systems that transmit data using block lengths on the order of the coherence time of their respective channels. More specifically, the second chapter focuses on Gallager-type exponential bounds applied to probabilities of decoding failure in rare-transition regimes. By construction, these bounds depend explicitly on the initial and terminating channel states at the codeword boundaries. The analysis is conducted for the scenario where channel state information is available at the receiver. The ensuing results are subsequently compared to the probabilities of decoding failure obtained for a Gilbert-Elliott channel under a minimum distance decoder and a maximum-likelihood decision rule [37–39].

The potential implications of this novel framework are then discussed in terms of queueing theory. We exploit the results of the error-probability analysis in the

rare-transition regime to evaluate the queueing performance of correlated channels. Particularly, we employ the derived upper bounds on the probability of decoding failure to bound the queueing performance of the system. In that regard, we show how stochastic dominance enables a tractable analysis of overall performance. The results of this evaluation are then compared with a performance characterization based on the exact probability of decoding failure for a Gilbert-Elliott channel under maximum-likelihood decoding.

CHAPTER II

DELAY-SENSITIVE COMMUNICATION OVER FADING CHANNEL: QUEUING BEHAVIOR AND CODE PARAMETER SELECTION

II.1 Introduction

In this chapter, we study the queueing performance of finite-state channels with and without memory. The chapter is organized as follows. The Gilbert-Elliott channel model is described in Section II.2. The arrival and departure processes of data packets at the transmitter, and the effect of feedback information are discussed in Section II.3. The Markov model used to evaluate queueing behavior is constructed in Section II.4. A detailed study of the coding analysis and derivations for the probabilities of decoding failure are given in Section II.5. Issues related to undetected errors are discussed in Section II.6. Numerical results showing the performance of our communication system are presented in Section II.7. Finally, we offer pertinent conclusions in Section II.8.

II.2 Gilbert-Elliott Channel Model

At present, the term Gilbert-Elliott channel often refers to a wide class of finite-state fading channels that model communication links with memory. In our study, however, we allude to its original definition and we use the denomination Gilbert-Elliott channel to designate a binary symmetric channel that features two possible states: a *good* state $\mathtt{g}$ with crossover probability $\varepsilon_{\mathtt{g}}$, and a *bad* state $\mathtt{b}$ with crossover probability $\varepsilon_{\mathtt{b}}$. While simple, this model can account for uncertainties associated with transmitting symbols over a noisy channel and correlation over time.

The evolution of the channel is governed by a finite-state Markov chain. We denote the transition probability from b to g by α, and we label the transition probability in the reverse direction by β. Thus, the channel evolution forms a Markov chain with transition probability matrix

$$\mathbf{P} = \begin{bmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{bmatrix}. \tag{II.1}$$

A graphical representation of this channel appears in Fig. II.1. It is worth mentioning that the steady-state probabilities of the good and bad states are $\frac{\alpha}{\alpha+\beta}$ and $\frac{\beta}{\alpha+\beta}$, respectively.

We note that, in defining the matrix $\mathbf{P}$, we have implicitly ordered the states from bad to good. With a slight abuse of notation, we use this bijection between channel states and their numerical indices to refer to specific entries in the matrix. We employ random variable C_n to denote the state of the channel at time n. Then, entry $[\mathbf{P}]_{c,d}$ represents the probability of a channel transition to state d, given that the current state is c. For groups of random variables, we use the common expression $P_{\cdot|\cdot}(\cdot|\cdot)$ to denote conditional joint probability mass functions. Accordingly, we can write $P_{C_{n+1}|C_n}(d|c) = \Pr(C_{n+1} = d|C_n = c)$, where $c, d \in \{\mathtt{b}, \mathtt{g}\}$. In a similar fashion, $P_{C_{n+N}|C_n}(d|c)$ can be obtained by looking at the proper entry of matrix $\mathbf{P}^N$.

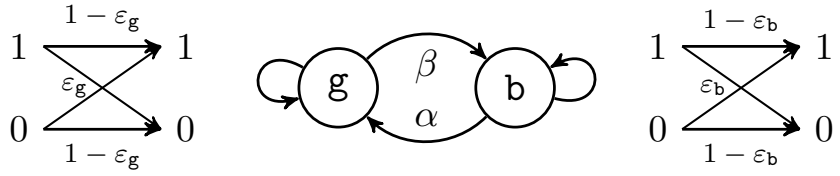

Figure II.1: The Gilbert-Elliott model is one of the simplest non-trivial instantiations of a finite-state channel with memory. State evolution over time forms a Markov chain, and the input-output relationship of this binary channel is governed by a state-dependent crossover probability, as illustrated above.

To proceed, we need a way to compute the conditional distribution of the number of errors that occur during N consecutive uses of the channel. Let E denote the number of errors occurring in a data block. The distribution of E can be obtained using the matrix of polynomials

$$\mathbf{P}_x = \begin{bmatrix} (1-\alpha)(1-\varepsilon_{\mathtt{b}}+\varepsilon_{\mathtt{b}}x) & \alpha(1-\varepsilon_{\mathtt{b}}+\varepsilon_{\mathtt{b}}x) \\ \beta(1-\varepsilon_{\mathtt{g}}+\varepsilon_{\mathtt{g}}x) & (1-\beta)(1-\varepsilon_{\mathtt{g}}+\varepsilon_{\mathtt{g}}x) \end{bmatrix}.$$

Throughout, we employ $[\![x^j]\!]$ to represent the linear functional that maps a polynomial in x to the coefficient of x^j. Using this notation, we get $P_{E,C_{N+1}|C_1}(e,d|c) = \Pr(E=e, C_{N+1}=d|C_1=c) = [\![x^e]\!]\left[\mathbf{P}_x^N\right]_{c,d}$. Eventually, we will use this distribution to compute the conditional probabilities of decoding failure and undetected error. We note that closed-form recursions for these values have been derived a number of times in the past [38, 47].

II.3 Arrivals, Departures, and Feedback

In this section, we describe the elements that compose our queueing system. Suppose that a packet of length L needs to be sent over the Gilbert-Elliott channel to a destination. In the proposed framework, this packet is divided into S segments, each containing K information bits. The last segment is zero padded, if needed, to conform to the prescribed length. A block code (e.g., a BCH or a random code) is used to encode each data segment into a codeword of length N (see Fig. II.2). These codewords are then transmitted over the communication link. Packet arrivals at the source are initially assumed to form an instance of a Poisson process with rate λ packets per channel use (see Fig. II.3). Therefore, the number of packets expected to arrive during an interval of length N is equal to $\lambda_N = \lambda N$. As we will see, our framework can accommodate more general packet arrivals, such as

Markov processes with discrete state spaces [48, 49]. This comes at the expense of additional bookkeeping. For instance, a Markov modulated Poisson process (MMPP) with distinct arrival rates, can be employed to better capture bursty traffic [50] and fluctuations in workload.

Packet sizes are assumed to form a sequence of independent and identically distributed random variables, where each element has a geometric distribution with parameter $\rho \in (0, 1)$. Mathematically, we write $\Pr(L = \ell) = (1 - \rho)^{\ell-1}\rho$ where $\ell \geq 1$. This assumption plays a key role in our study and it has been selected, partly, to facilitate the analysis we wish to carry. In particular, the memoryless property of the geometric distribution makes for a tractable queueing model. Not too surprisingly, having a geometric distribution for the size of packets is commonplace in the literature [51, 52].

We can further relate the packet-length distribution to the progression of coded transmissions. For fixed block length N and code rate R, every successful decoding event reveals exactly RN information bits to the destination. As such, when a data packet contains L bits, one needs to successfully decode $S = \left\lceil \frac{L}{RN} \right\rceil$ codewords to complete the transmission of the entire packet. We note that random variable S possesses a geometric distribution with $\Pr(S = s) = (1 - \rho_r)^{s-1}\rho_r$, where $s \geq 1$ and $\rho_r = 1 - (1 - \rho)^{RN}$. Thus, in the current setting, the number of coded blocks

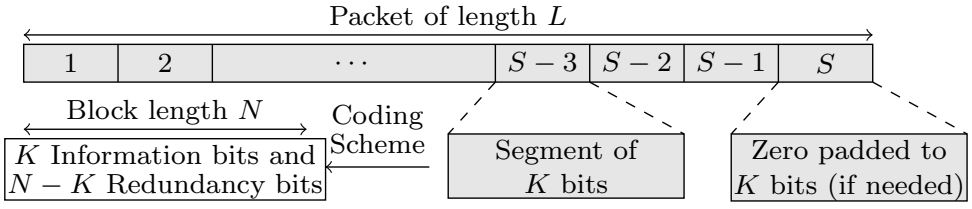

Figure II.2: Each packet is divided into S segments, and a channel encoding scheme is used to encode each segment

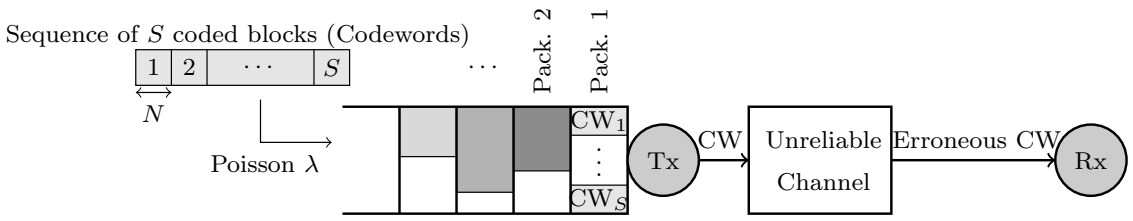

Figure II.3: Coded segments are transmitted over the unreliable communication link. A data packet is discarded from the transmit buffer only when all its codewords are successfully transmitted

per data packet S retains the memoryless property. We emphasize that, in our framework, a data packet is discarded from the transmit buffer if and only if the destination acknowledges reception of the latest codeword and this codeword contains the last parcel of information corresponding to the head packet. The departure process is governed by the parameters of the channel and the coding scheme adopted. Generally, a lower code rate yields smaller probabilities of decoding failure, but it also entails having more data segments to send. Thus, for a given channel and load, it is important to choose the block length and the code rate which give the best overall performance. In Section II.7, we present simulation results for a system with packets of a constant size and we compare its performance to the corresponding system with geometrically distributed packet lengths.

A subtle, yet important aspect associated with automatic repeat request over unreliable connections is the amount of feedback needed by a particular scheme. Using shorter block lengths necessarily entails more frequent feedback messages from the destination. In general, adequately evaluating the costs and benefits of various feedback strategies is a complicated task. Since this is not a prime objective of this dissertation, we circumvent this issue by making simple assumptions. We assume that feedback is instantaneously and faithfully received at the source; this idealized view is frequently found in the literature [53,54]. In contrast, any detailed analysis of

feedback requires making strong assumptions about correlation between the forward and reverse links, the delay associated with receiving feedback, and mechanisms to cope with corrupted messages. Although these issues warrant attention, they are outside the scope of this dissertation. Beyond that, we hypothesize that the price of feedback is captured by having a portion of every data segment dedicated to a header of length h. Of course, this reduces the size of the packet payload to $RN - h$. This crude approximation treats feedback bits as constant overhead, and it is a modest step in better accounting for control messages. Feedback overhead will affect the number of segments contained in a data packet. If h information bits in every segment pertain to the header, then the number of successful codeword transmissions necessary to transfer a packet becomes $S = \left\lceil \frac{L}{K-h} \right\rceil$, a slight variation compared to the original value. Nevertheless, S retains a geometric distribution, albeit with parameter $\rho_r = 1 - (1-\rho)^{K-h}$.

A very important aspect of queueing systems is stability. The Foster-Lyapunov criterion ensures that our simple system remains stable so long as the packet service rate exceeds the arrival rate. To calculate the mean service rate, we recall that a packet leaves the queue whenever a codeword is decoded successfully and this codeword carries the last data segment of the head packet. Let $P_{\mathrm{s}|E}(e)$ and $P_{\mathrm{f}|E}(e)$ denote the conditional probabilities of decoding success and failure, respectively, given the number of errors within a block, $E = e$. By reciprocity, the conditional success probability is equal to $P_{\mathrm{s}|E}(e) = 1 - P_{\mathrm{f}|E}(e)$. Then, the average service rate can be computed as $\mu_N = \rho_r \mathrm{E}\left[P_{\mathrm{s}|E}(e)\right]$ packets per codeword transmission. The stability factor for this system is $\frac{\lambda_N}{\mu_N}$, and the process is stable provided that this ratio is less than unity. Conditional failure probabilities will be computed explicitly in Section II.5 for different channels and coding schemes.

II.4 Queueing Model

We are ready to examine more closely the queueing behavior of our communication link. Throughout, we use Q_s to denote the number of packets waiting in the transmit buffer. The channel state at the same instant is C_{sN+1}. By grouping these two random variables together, we can construct a discrete-time Markov chain (DTMC), which we write $U_s = (C_{sN+1}, Q_s)$. The resulting DTMC is of the M/G/1 type, and there are many established techniques that apply to such systems [36, 55, 56]. We note that for the binary symmetric channel, input-output properties are unchanged over time. In this degenerate case, the queue length Q_s contains all the information relevant to the DTMC, and the random variable U_s is mathematically equivalent to the state of the transmit buffer.

Using the total probability theorem, the transition probabilities for the DTMC $\{U_s\}$ can be decomposed as

$$\Pr(U_{s+1} = (d, q_{s+1}) | U_s = (c, q_s)) = \sum_{e \in \mathbb{N}_0} P_{Q_{s+1}|E,Q_s}(q_{s+1}|e, q_s) P_{E,C_{(s+1)N+1}|C_{sN+1}}(e, d|c).$$

Examining the summands, we need to derive expressions for $P_{Q_{s+1}|E,Q_s}(q_{s+1}|e, q_s)$. Suppose that the current number of packets in the queue is $Q_s = q_s$. Then, admissible values for Q_{s+1} are restricted to the collection $\{q_s - 1, q_s, q_s + 1, \ldots\}$. The corresponding transition probabilities are given by

$$\begin{aligned} P_{Q_{s+1}|E,Q_s}(q_s - 1|e, q_s) &= a_0(1 - P_{\mathrm{f}|E}(e))\rho_r, \\ P_{Q_{s+1}|E,Q_s}(q_s + i|e, q_s) &= a_{i+1}(1 - P_{\mathrm{f}|E}(e))\rho_r \\ &\quad + a_i \left(P_{\mathrm{f}|E}(e) + (1 - P_{\mathrm{f}|E}(e))(1 - \rho_r) \right), \quad i \geq 0 \end{aligned} \tag{II.2}$$

where $a_i = \frac{(\lambda N)^i}{i!} e^{-\lambda N}$ is the probability that i packets arrive during the transmission of one codeword. When the queue is empty, $\{Q_s = 0\}$, the transition probabilities

reduce to $P_{Q_{s+1}|E,Q_s}(q_s + i|e, 0) = a_i$ with $i \geq 0$.

Using these equations, we can get the probability transition matrix of the Markov process $\{U_s\}$. First, we introduce the following convenient notation, where $q \in \mathbb{N}_0$ and $c, d \in \{\mathtt{g}, \mathtt{b}\}$,

$$\mu_{cd}^{i} = \Pr(U_{s+1} = (d, q + i)|U_s = (c, q)) \quad i \geq 1,$$
$$\kappa_{cd} = \Pr(U_{s+1} = (d, q)|U_s = (c, q))$$
$$\xi_{cd} = \Pr(U_{s+1} = (d, q - 1)|U_s = (c, q)).$$

Similarly, when the queue is empty, we write $\mu_{cd}^{i0} = \Pr(U_{s+1} = (d, i)|U_s = (c, 0))$ and $\kappa_{cd}^{0} = \Pr(U_{s+1} = (d, 0)|U_s = (c, 0))$. Possible state transitions are illustrated in Fig. II.4.

Next, we review briefly the matrix-geometric method, an efficient way to compute the stationary distributions of chains with repetitive structures. We can represent the equilibrium distribution of our system as a semi-infinite vector $\boldsymbol{\pi} = (\pi(1), \pi(2), \ldots)$, where $\pi(2q + 1) = \Pr(C = \mathtt{b}, Q = q)$ and $\pi(2q + 2) = \Pr(C = \mathtt{g}, Q = q)$. Alternatively, we can group pairs of states together and write $\boldsymbol{\pi} = [\boldsymbol{\pi}_0 \ \boldsymbol{\pi}_1 \ \boldsymbol{\pi}_2 \ \cdots]$ where $\boldsymbol{\pi}_q$ comprises the stationary probabilities of the qth level of the chain with $\boldsymbol{\pi}_q = [\pi(2q+1) \ \pi(2q+2)]$. Using this notation, one can express the detailed balance equation $\boldsymbol{\pi}\mathbf{T} = \boldsymbol{\pi}$ in terms of the transition probability matrix $\mathbf{T}$, which appears in block-partitioned form below

$$\mathbf{T} = \begin{bmatrix} \hat{\mathbf{A}} & \hat{\mathbf{F}}^{(1)} & \hat{\mathbf{F}}^{(2)} & \hat{\mathbf{F}}^{(3)} & \cdots \\ \mathbf{B} & \mathbf{A} & \mathbf{F}^{(1)} & \mathbf{F}^{(2)} & \cdots \\ \mathbf{0} & \mathbf{B} & \mathbf{A} & \mathbf{F}^{(1)} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

The labels $\mathbf{A}$, $\mathbf{F}$, and $\mathbf{B}$ symbolize local, forward, and backward transition-probability blocks, respectively; the superscript (i) indicates that i additional data packets are present in the buffer at the next time instant; and the hat designates instances where

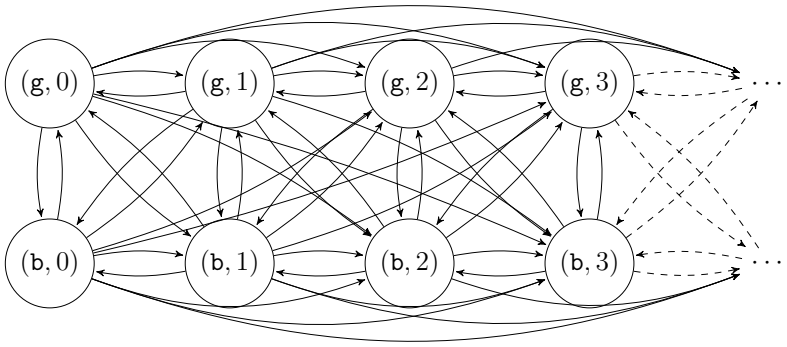

Figure II.4: State space and transition diagram for the aggregate queued process $\{U_s\}$; self-transitions are intentionally omitted.

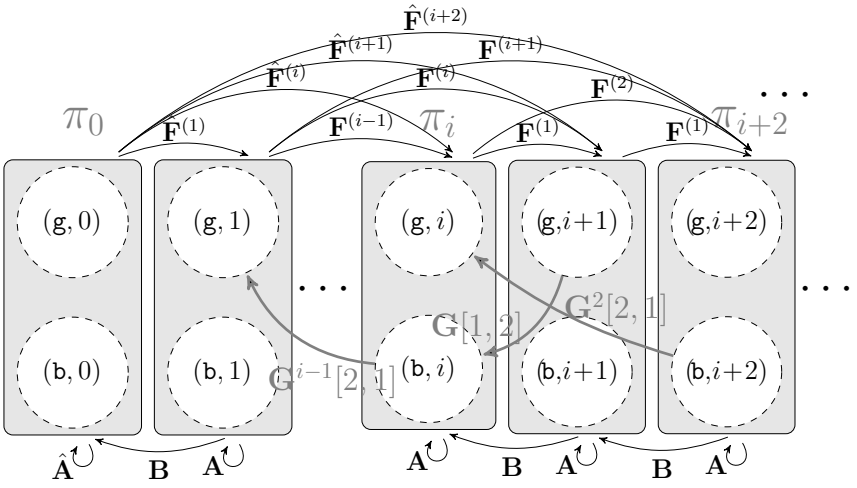

Figure II.5: Level transition diagram and probabilistic interpretation of $\mathbf{G}$

the queue is initially empty. More specifically, we have

$$\mathbf{F}^{(i)} = \begin{bmatrix} \mu^i_{\mathrm{bb}} & \mu^i_{\mathrm{bg}} \\ \mu^i_{\mathrm{gb}} & \mu^i_{\mathrm{gg}} \end{bmatrix}, \ \mathbf{A} = \begin{bmatrix} \kappa_{\mathrm{bb}} & \kappa_{\mathrm{bg}} \\ \kappa_{\mathrm{gb}} & \kappa_{\mathrm{gg}} \end{bmatrix}, \ \mathbf{B} = \begin{bmatrix} \xi_{\mathrm{bb}} & \xi_{\mathrm{bg}} \\ \xi_{\mathrm{gb}} & \xi_{\mathrm{gg}} \end{bmatrix}.$$

For an empty queue, the blocks are

$$\hat{\mathbf{F}}^{(i)} = \begin{bmatrix} \mu^{i0}_{\mathrm{bb}} & \mu^{i0}_{\mathrm{bg}} \\ \mu^{i0}_{\mathrm{gb}} & \mu^{i0}_{\mathrm{gg}} \end{bmatrix} \qquad \hat{\mathbf{A}} = \begin{bmatrix} \kappa^0_{\mathrm{bb}} & \kappa^0_{\mathrm{bg}} \\ \kappa^0_{\mathrm{gb}} & \kappa^0_{\mathrm{gg}} \end{bmatrix}.$$

Figure II.5 shows the possible transitions among the different levels of the system.

Proposition 1. *Let* $\mathbf{G}$ *be the limiting matrix of the recursion*

$$\mathbf{G}_{i+1} = -\mathbf{L}^{-1}(\mathbf{B} + \sum_{j=1}^{\infty} \mathbf{F}^{(j)} \mathbf{G}_i^{j+1}) \tag{II.3}$$

starting from $\mathbf{G}_0 = \mathbf{0}$ *and where* $\mathbf{L} = \mathbf{A} - \mathbf{I}$. *For* $j \geq 1$, *the stationary probability vectors* $\boldsymbol{\pi}_j$ *associated with* $\mathbf{T}$ *are given by*

$$\boldsymbol{\pi}_j = -(\boldsymbol{\pi}_0 \hat{\mathbf{S}}^{(j)} + \sum_{k=1}^{j-1} \boldsymbol{\pi}_k \mathbf{S}^{(j-k)})(\mathbf{S}^{(0)})^{-1},$$

where $\mathbf{F}^{(0)} = \mathbf{L}$, $\hat{\mathbf{S}}^{(j)} = \sum_{l=j}^{\infty} \hat{\mathbf{F}}^{(l)} \mathbf{G}^{l-j}$ *for* $j \geq 1$, *and* $\mathbf{S}^{(j)} = \sum_{l=j}^{\infty} \mathbf{F}^{(l)} \mathbf{G}^{l-j}$ *for* $j \geq 0$. *Vector* $\boldsymbol{\pi}_0$ *is uniquely determined by the normalization condition, and it can be found by solving*

$$\boldsymbol{\pi}_0 [(\hat{\mathbf{L}} - \hat{\mathbf{S}}^{(1)} (\mathbf{S}^{(0)})^{-1} \mathbf{B})^{\diamondsuit} | \mathbf{1}^T - \mathbf{H}\mathbf{1}^T] = [\mathbf{0}|1],$$

where $\mathbf{H} = \sum_{j=1}^{\infty} \hat{\mathbf{S}}^{(j)} (\sum_{j=0}^{\infty} \mathbf{S}^{(j)})^{-1}$, $\hat{\mathbf{L}} = \hat{\mathbf{A}} - \mathbf{I}$, *and the symbol* $\diamondsuit$ *denotes an operator that discards the last column of the corresponding matrix [55].*

Proof. A proof for an equivalent continuous-time formulation is available in [55]; it is based on solving $\boldsymbol{\pi}\tilde{\mathbf{T}} = \mathbf{0}$. The discrete-time case can be obtained by defining $\tilde{\mathbf{T}} = \mathbf{T} - \mathbf{I}$, which leads to a solution for $\boldsymbol{\pi}\mathbf{T} = \boldsymbol{\pi}$, as desired. □

Matrix $\mathbf{G}$ admits a nice interpretation: entry $[\mathbf{G}]_{r,c}$ is the conditional probability that the Markov process first enters level $i-1$ through state c given that it starts at level i, in state r [55]. Such a matrix must naturally satisfy the relation

$$\mathbf{A}\mathbf{G} + \mathbf{B} + \sum_{j=1}^{\infty} \mathbf{F}^{(j)} \mathbf{G}^{j+1} = \mathbf{G},$$

and this equation can be solved recursively, as described above. Figure II.5 illustrates the probabilistic interpretation of $\mathbf{G}$ and its powers. As a side note,

we emphasize that all the matrix equations simplify to scalar computations for the binary symmetric channel.

To conclude this section, we introduce a slight generalization of the arrival process. Consider a two-state discrete-time Markov-modulated Poisson process with arrival rates λ_1 and λ_2, MMPP(λ_1, λ_2). The only elements of our analysis that need to be modified are the blocks in the transition probability matrix $\mathbf{T}$; they become 4×4 matrices to account for the state of the modulating process. Proposition 2 offers a formal description of the quantities involved in making changes.

Proposition 2. *Suppose that T_1 represents the amount of time the arrival process, MMPP(λ_1, λ_2), spends in modulating state one during the transmission of a codeword. The joint probability that i packets arrive during that time interval together with the modulating process transitioning to state A_{N+1}, conditioned on starting state A_1, is*

$$P_{K_a, A_{N+1}|A_1}(i, l|m) = \sum_{t=0}^{N} P_{K_a|T_1}(i|t) P_{T_1, A_{N+1}|A_1}(t, l|m),$$

$l, m \in \{1, 2\}$, where K_a denotes the number of arrivals and $P_{T_1, A_{N+1}|A_1}(t, l|m)$ accounts for the occupation time of the modulating process as well as edge transitions (see Lemma 17).The conditional distribution of arrivals, $P_{K_a|T_1}(i|t)$ becomes

$$\sum_{k_1=0}^{i} \sum_{k_2=0}^{i-k_1} \frac{(\lambda_1 t)^{k_1}}{k_1!} e^{-\lambda_1 t} \frac{(\lambda_2 (N-t))^{k_2}}{k_2!} e^{-\lambda_2 (N-t)}.$$

Collecting these results, we gather that a_i must be replaced by $P_{K_a, A_{N+1}|A_1}(i, l|m)$ in the transition probabilities of the queue, (II.2). This yields 4×4 blocks in the modified transition probability matrix. In the revised formulation,

$$\boldsymbol{\pi}_q = [\pi(4q+1)\ \pi(4q+2)\ \pi(4q+3)\ \pi(4q+4)],$$

which corresponds to having q packets with a specific pair of channel state and

modulating state for the arrival process.

II.5 Probability of Decoding Failure

In the previous section, a general queueing analysis was presented in which, different encoding/decoding schemes can be adopted. In this section, we derive probabilities of decoding failures for various scenarios. We begin with the simpler BSC case, and then we proceed to the Gilbert-Elliott channel. Still, applicability of the queueing analysis does not limit to the schemes we study in this section.

II.5.1 Random Coding with ML/MD Decoding

Consider a coding scheme in which a codebook of size $M = 2^{NR}$ is generated at random. As before, R denotes code rate and N stands for block length. For every index $i \in \{1, \ldots, M\}$, a codeword $\mathbf{X}(i)$ is selected uniformly and independently from the set of length-N binary sequences, $\{0,1\}^N$. The maximum number of information bits encoded in each transmission is $K = \log_2 M$. For performance assessment, we assume that one of the codewords is chosen at random and sent over the communication channel. On the receiver side, a maximum likelihood (ML) decision rule is used to decode the received vector $\mathbf{Y}$; that is, $\hat{\mathbf{X}} = \arg\max_{\mathbf{X}} P_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})$.

II.5.1.1 Binary Symmetric Channel

For our memoryless channel, the ML decoder actually decodes to the closest valid codeword. The ML decision rule is therefore equivalent to the minimum distance (MD) decoder. A subtle, yet important point in analyzing this decoder is that the decoding radius is not fixed in advance; this can be inferred from the following well-known result.

Theorem 3 ([39]). *Random coding is employed to send information over a BSC(p).*

Assume that ML decoding is performed at the receiver, with ties treated as decoding failures. Then, the failure probability for this scenario is given by

$$P_{\mathrm{f}} = \sum_{e=0}^{N} P_E(e) P_{\mathrm{f}|E}(e) \tag{II.4}$$
$$= \sum_{e=0}^{N} \binom{N}{e} p^e (1-p)^{N-e} \left[1 - \left(1 - 2^{-N} \sum_{i=0}^{e} \binom{N}{i} \right)^{M-1} \right].$$

We note this result holds for any forward error correction scheme in which all codewords are equally likely, and that they are pairwise independent (e.g., Shannon random coding or random linear codes). Moreover, the format of (II.4) extends to other encoding strategies. For a BSC(p), random variable E possesses a binomial distribution and a suitable expression for the conditional probabilities of decoding failure should be substituted. In [41], Fano's result is modified to better handle ties, and it is generalized to a wider class of channels. It turns out that, for our purpose, this modification has a negligible effect on performance; and it is therefore disregarded.

II.5.1.2 Gilbert-Elliott Channel

Having gained valuable insight with the BSC(p), we turn to the more challenging case. We derive probabilities of decoding failure for the Gilbert-Elliott channel under ML decoding, and conditioned on the occupancy times. We emphasize that knowing the empirical channel state distribution is key in finding useful expressions for failure probabilities. Let $N_{\mathtt{g}}$ and $N_{\mathtt{b}} = N - N_{\mathtt{g}}$ represent the numbers of visits to each channel state during the transmission of a length-N codeword. These random variables are sometimes collectively called the channel state type [57]. Using the empirical state distribution and the corresponding conditional error probabilities, one can average over all channel types to get the probabilities of decoding failure while accounting

for boundary states,

$$P_{\mathrm{f},C_{N+1}|C_1}(d|c) = \sum_{n_{\mathtt{g}}=0}^{N} P_{\mathrm{f}|N_{\mathtt{g}}}(n_{\mathtt{g}}) P_{N_{\mathtt{g}},C_{N+1}|C_1}(n_{\mathtt{g}}, d|c), \tag{II.5}$$

where $P_{N_{\mathtt{g}},C_{N+1}|C_1}(\cdot,\cdot|\cdot)$ is given by Lemma 17. One can also compute this latter quantity using the N-th power of the matrix generating function of the good state occupation time.

$$\mathbf{G}(x) = \begin{bmatrix} (1-\alpha)x & \alpha x \\ \beta & 1-\beta \end{bmatrix}.$$

We stress that the failure probabilities depend on the initial and final states of the channel through the distribution of $N_{\mathtt{g}}$. Since we are interested in moderate block lengths, on the order of the mixing time of the channel, these boundary states can have a significant impact on the probabilities of decoding failure.

For a specific channel realization, let $\mathbf{X}_c$ and $\mathbf{Y}_c$ be the subvectors of $\mathbf{X}$ and $\mathbf{Y}$ corresponding to time instants when the channel is in state $c \in \{\mathtt{g}, \mathtt{b}\}$. We denote the number of errors in state c by $E_c = d_{\mathrm{H}}(\mathrm{X}_c, \mathrm{Y}_c)$, where $d_{\mathrm{H}}(\cdot,\cdot)$ represents the Hamming distance. Conditional error probability $P_{\mathrm{f}|N_{\mathtt{g}}}(n_{\mathtt{g}})$ can then be written as

$$\sum_{e_{\mathtt{g}}=0}^{n_{\mathtt{g}}} \sum_{e_{\mathtt{b}}=0}^{n_{\mathtt{b}}} P_{\mathrm{f}|N_{\mathtt{g}},E_{\mathtt{g}},E_{\mathtt{b}}}(n_{\mathtt{g}}, e_{\mathtt{g}}, e_{\mathtt{b}}) P_{E_{\mathtt{g}},E_{\mathtt{b}}|N_{\mathtt{g}}}(e_{\mathtt{g}}, e_{\mathtt{b}}|N_{\mathtt{g}}). \tag{II.6}$$

Given the channel type, the numbers of errors in the good and bad states are independent,

$$P_{E_{\mathtt{g}},E_{\mathtt{b}}|N_{\mathtt{g}}}(e_{\mathtt{g}}, e_{\mathtt{b}}|n_{\mathtt{g}}) = P_{E_{\mathtt{g}}|N_{\mathtt{g}}}(e_{\mathtt{g}}|n_{\mathtt{g}}) P_{E_{\mathtt{b}}|N_{\mathtt{g}}}(e_{\mathtt{b}}|n_{\mathtt{g}}), \tag{II.7}$$

where individual distributions are simply given by

$$P_{E_c|N_c}(e_c|n_c) = \binom{n_c}{e_c} \varepsilon_c^{e_c} (1-\varepsilon_c)^{n_c - e_c} \qquad c \in \{\mathtt{g}, \mathtt{b}\}. \tag{II.8}$$

Theorem 4. *When ties are treated as errors, the probability of decoding failure for*

a length-N uniform random code with M codewords, conditioned on the number of symbol errors in each state and the channel state type, is given by

$$P_{\mathrm{f}|N_{\mathtt{g}},E_{\mathtt{g}},E_{\mathtt{b}}}(n_{\mathtt{g}},e_{\mathtt{g}},e_{\mathtt{b}}) = 1-\left(1-2^{-N}\sum_{\mathcal{M}(\gamma e_{\mathtt{g}}+e_{\mathtt{b}})}\binom{n_{\mathtt{g}}}{\tilde{e}_{\mathtt{g}}}\binom{n_{\mathtt{b}}}{\tilde{e}_{\mathtt{b}}}\right)^{M-1}. \tag{II.9}$$

where $\mathcal{M}(d)$ is the set of pairs $(\tilde{e}_{\mathtt{g}},\tilde{e}_{\mathtt{b}})\in\{0,\ldots,N\}^2$ that satisfy $\gamma\tilde{e}_{\mathtt{g}}+\tilde{e}_{\mathtt{b}}\le d$. This holds with $\gamma=\frac{\ln\varepsilon_{\mathtt{g}}-\ln(1-\varepsilon_{\mathtt{g}})}{\ln\varepsilon_{\mathtt{b}}-\ln(1-\varepsilon_{\mathtt{b}})}$ for the ML decision rule, and with $\gamma=1$ for the MD decoder.

Proof. First, we revisit the ML decoding rule for the Gilbert-Elliott channel when channel state information is available at the receiver. Given the state occupation $n_{\mathtt{g}}$, we have

$$\begin{aligned}P_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) &= P_{\mathbf{Y}_{\mathtt{g}}|\mathbf{X}_{\mathtt{g}}}(\mathbf{y}_{\mathtt{g}}|\mathbf{x}_{\mathtt{g}})P_{\mathbf{Y}_{\mathtt{b}}|\mathbf{X}_{\mathtt{b}}}(\mathbf{y}_{\mathtt{b}}|\mathbf{x}_{\mathtt{b}})\\ &= \varepsilon_{\mathtt{g}}^{e_{\mathtt{g}}}(1-\varepsilon_{\mathtt{g}})^{n_{\mathtt{g}}-e_{\mathtt{g}}}\varepsilon_{\mathtt{b}}^{e_{\mathtt{b}}}(1-\varepsilon_{\mathtt{b}})^{n_{\mathtt{b}}-e_{\mathtt{b}}}\end{aligned}$$

Upon receiving word $\mathbf{Y}$, the ML decoder returns the codeword $\mathbf{X}$ that maximizes $\ln P_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})$. Thus, a little algebra shows the decoded message will be

$$\arg\min_{\mathbf{x}\in\mathcal{C}}[\gamma e_{\mathtt{g}}(\mathbf{x})+e_{\mathtt{b}}(\mathbf{x})], \tag{II.10}$$

where $e_{\mathtt{g}}(\mathbf{x})=d_{\mathrm{H}}(\mathbf{x}_{\mathtt{g}},\mathbf{y}_{\mathtt{g}})$ and $e_{\mathtt{b}}(\mathbf{x})=d_{\mathrm{H}}(\mathbf{x}_{\mathtt{b}},\mathbf{y}_{\mathtt{b}})$ are realizations of $E_{\mathtt{g}}$ and $E_{\mathtt{b}}$, respectively. This argument is used to demonstrate the dependency on $\mathbf{x}$. Notice that the term $n_{\mathtt{g}}\ln(1-\varepsilon_{\mathtt{g}})+n_{\mathtt{b}}\ln(1-\varepsilon_{\mathtt{b}})$ in $\ln P_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})$ does not change the ML decision.

Next, we consider the probability of failure for the decoding rule given in (II.10) when random codes are used. In our system, decoding succeeds if and only if the correct codeword is returned as the unique minimizer in (II.10). The failure probability found in (III.22) can be obtained in a few steps. By symmetry, we can

assume that the transmitted codeword $\mathbf{x}$ is the all-zero codeword. The other $M-1$ codewords are drawn independently and uniformly. For any received vector $\mathbf{y}$ that satisfies $E_{\mathtt{g}} = e_{\mathtt{g}}$ and $E_{\mathtt{b}} = e_{\mathtt{b}}$, decoding succeeds when every other codeword produces a strictly larger value for the cost function in (II.10). A straightforward combinatorial argument shows that the number of codewords that meet this requirement is

$$V(n_{\mathtt{g}}, n_{\mathtt{b}}, e_{\mathtt{g}}, e_{\mathtt{b}}) = \sum_{(\tilde{e}_{\mathtt{g}}, \tilde{e}_{\mathtt{b}}) \in \mathcal{M}(\gamma e_{\mathtt{g}} + e_{\mathtt{b}})} \binom{n_{\mathtt{g}}}{\tilde{e}_{\mathtt{g}}} \binom{n_{\mathtt{b}}}{\tilde{e}_{\mathtt{b}}}. \tag{II.11}$$

The probability that a uniformly chosen random vector falls in this set is $q = V(n_{\mathtt{g}}, n_{\mathtt{b}}, e_{\mathtt{g}}, e_{\mathtt{b}})/2^N$. Since codewords are independent, the failure probability is equal to $1-(1-q)^{M-1}$.

One can infer from (II.10) that, for the ML decision rule, errors in the bad state do not affect performance as much as errors in the good state. This is because the decoder gives more weight to symbols that are received while the channel is in its good state, as they are deemed more reliable. The MD decoder, on the other hand, only considers the total number of errors within a block, irrespective of the state they occur in. That is, errors in either state cost the same and $\gamma = 1$. The terms over which the sum is taken need to be modified accordingly. □

In view of Theorem 18, one can substitute the appropriate expressions for decoding performance into (III.19) to get overall probabilities of decoding failure.

As a side note, Vandermonde's convolution identity implies that

$$\sum_{\tilde{e}_{\mathtt{g}}=0}^{e_{\mathtt{g}}+e_{\mathtt{b}}} \sum_{\tilde{e}_{\mathtt{b}}=0}^{e_{\mathtt{g}}+e_{\mathtt{b}}-\tilde{e}_{\mathtt{g}}} \binom{n_{\mathtt{g}}}{\tilde{e}_{\mathtt{g}}} \binom{n_{\mathtt{b}}}{\tilde{e}_{\mathtt{b}}} = \sum_{j=0}^{e_{\mathtt{g}}+e_{\mathtt{b}}} \binom{N}{j},$$

and therefore the volume expression in (II.11) for MD decoding ($\gamma = 1$) reduces to the volume computation associated with the binary symmetric channel. Finally, we

note that

$$\sum_{\mathcal{M}(\gamma e_{\mathsf{g}}+e_{\mathsf{b}})} \binom{n_{\mathsf{g}}}{\tilde{e}_{\mathsf{g}}}\binom{n_{\mathsf{b}}}{\tilde{e}_{\mathsf{b}}} + \sum_{\bar{\mathcal{M}}(\gamma e_{\mathsf{g}}+e_{\mathsf{b}})} \binom{n_{\mathsf{g}}}{\tilde{e}_{\mathsf{g}}}\binom{n_{\mathsf{b}}}{\tilde{e}_{\mathsf{b}}} = 2^N,$$

where $\bar{\mathcal{M}}(c)$ is the set of pairs $(\tilde{e}_{\mathsf{g}}, \tilde{e}_{\mathsf{b}}) \in \{0, \ldots, N\}^2$ that satisfy $\gamma\tilde{e}_{\mathsf{g}} + \tilde{e}_{\mathsf{b}} > c$.

II.5.2 BCH Coding with Bounded Distance Decoding

In this section, we present a more pragmatic facet of our inquiry. We consider a primitive binary BCH code of minimum distance $d_{\min}$, which is capable of correcting up to $t = \left\lfloor \frac{d_{\min}-1}{2} \right\rfloor$ errors. This entails having $N = 2^m - 1$, with $m \geq 2$, and a single optimal K for each $d_{\min}$ [58, p. 486]. We analyze the queueing behavior of the system in terms of the block length N and the code rate $R = K/N$. The goal is to characterize the performance over admissible parameters. At the receiver, the bounded distance decoder either declares a decoding success, or it detects a failure and requests a retransmission. It is important to emphasize that, when the number of errors is greater than t, the decoder may be subject to an undetected error. We discuss this issue in greater depth in Section II.6.

For the binary symmetric channel, the conditional probability of failure in (II.4) is equal to $P_{\mathrm{f}|E}(e) = \mathbb{1}_{\{z\in\mathbb{Z}|z>t\}}(e)$, where $\mathbb{1}_A(\cdot)$ is the standard indicator function of the set A. Similarly, for the Gilbert-Elliott case, the average failure probability is given by

$$\begin{aligned} P_{\mathrm{f}} &= \sum_{c,d\in\{\mathsf{g},\mathsf{b}\}} P_{C_1}(c) P_{\mathrm{f},C_{N+1}|C_1}(d|c) \\ &= \sum_{c,d\in\{\mathsf{g},\mathsf{b}\}} P_{C_1}(c) \sum_{e=1}^{N} P_{E,C_{N+1}|C_1}(e,d|c) P_{\mathrm{f}|E}(e), \end{aligned}$$

where $P_{\mathrm{f}|E}(e)$ appears above. The expected success probability can be computed in an analog fashion, albeit replacing $P_{\mathrm{f}|E}(e)$ by $1 - P_{\mathrm{f}|E}(e)$. The average service rate

can be expressed as $\mu_N = \rho_r P_s$ packets per codeword transmission, thereby implicitly setting a bound for system stability.

II.6 Undetected Errors

A serious issue with pragmatic communication systems is the presence of undetected decoding failures. In the present setting, this occurs when the receiver uniquely decodes to the wrong codeword. For delay-sensitive applications, this problem is especially important because recovery procedures can lead to undue delay. To address this issue, we apply standard techniques that help control the probability of admitting erroneous codewords [59,60]. This, in turn, leads to slight modifications to the performance analysis presented above. The probability of undetected failure is a system parameter that must be set during the design phase of the system.

II.6.1 Random Coding with ML/MD Decoding

Under our aforementioned scheme, information is sent over the channel and the decoder reports the codeword with the minimum (weighted) distance to the received vector, as seen in (II.10). To reduce the probability of undetected error, we revisit the technique established in [59] regarding the error exponents, and introduce a safety margin ν . This scheme and its ramifications are easiest to explain for the binary symmetric channel. Recall that, for this simpler channel model, the ML and MD decision rules coincide. Suppose that $d_{\mathrm{H}}(\hat{\mathbf{x}}, \mathbf{y}) = \hat{e}$, where $\hat{\mathbf{x}}$ is the closest codeword to received vector $\mathbf{y}$. The enhanced decoder only returns $\hat{\mathbf{x}}$ when the distance between $\mathbf{y}$ and the next closest codeword is greater than $\hat{e}+\nu$. If another codeword is present within distance $\hat{e} + \nu$, then the receiver declares a decoding failure.

As before, let e denote the distance between the sent message and the received vector. The performance associated with this procedure can be characterized by

considering balls of radii $e-\nu$, e, and $e+\nu$ centered around the received vector. Notice that, by construction, the transmitted codeword always lies in the last two balls. To analyze the system, consider the list of all codewords contained in the ball of radius $e+\nu$. If there is exactly one codeword on this list, it must be the correct one and it is returned successfully by the decoder. On the other hand, if there are more than one codeword on the list, then a decoding failure (detected or undetected) will occur. One can write the probability of this event as

$$P_{\mathrm{f}|E}(e) = 1 - \left(1 - 2^{-N} \sum_{i=0}^{e+\nu} \binom{N}{i}\right)^{M-1}. \tag{II.12}$$

A detected failure takes place when the decoder elects not to output a candidate codeword. The problem is setup so that the correct codeword is always on the list. As such, an undetected failure can only occur when there is at least one other candidate inside the ball of radius $e-\nu$. Note that this condition is necessary, but not sufficient; multiple incorrect candidates can be found in proximity of the received vector in such a way that a failure is reported. If there are only two codewords in the ball of radius e and one of them is inside the ball of radius $e-\nu$, then the decoder will necessarily return the incorrect one. If there are more than two codewords with the ball of radius e, then detected and undetected failures can occur, although for well-designed systems such events are very rare. Collecting these observations, we can derive an upper bound for the probability of undetected failure,

$$P_{\mathrm{ue}} < \sum_{e=0}^{N} \binom{N}{e} p^e (1-p)^{N-e} \left[1 - \left(1 - \frac{\sum_{i=0}^{e-\nu-1} \binom{N}{i}}{2^N}\right)^{M-1}\right]. \tag{II.13}$$

It may be instructive to point out that ties between the closest codewords are always treated as detected failures. Also, the probability of undetected failure decreases rapidly as ν gets larger. Thus, by choosing an appropriate value for ν, one can manage the level of undetected failures and hence make the decoding process more

robust, at the expense of a higher overall probability of failure. Lastly, since the probability of undetected failure is typically much smaller than the probability of detected failure, we can upper bound the latter by P_{f} with a negligible penalty.

Much of the intuition developed under the binary symmetric channel applies to the Gilbert-Elliott model, with one important distinction related to weighted distance. Indeed, for this more elaborate finite-state channel, the ML decoder picks the codeword that minimizes the weighted distance found in (II.10), $\gamma e_{\mathtt{g}}(\mathbf{x}) + e_{\mathtt{b}}(\mathbf{x})$. Suppose that B is the minimum weighted distance between the received vector and a codeword, and let C be the weighted distance associated with the transmitted codeword. To deal with the probability of undetected failure, the decoder declares a failure if there is another codeword of weighted distance at most $B + \nu$. Otherwise, the best candidate codeword is returned.

Similar to the BSC case, performance can be analyzed by considering three balls, with respect to weighted distance, of radii $C - \nu$, C, and $C + \nu$ centered around the received vector. Again, the transmitted codeword always resides in the last two balls. If there are multiple codewords on the list of codewords in the ball of radius $C + \nu$, then a decoding failure will occur. This happens with probability

$$P_{\mathrm{f}|N_{\mathtt{g}},E_{\mathtt{g}},E_{\mathtt{b}}}(n_{\mathtt{g}}, e_{\mathtt{g}}, e_{\mathtt{b}}) = 1 - \left(1 - 2^{-N} \sum_{\mathcal{M}(\gamma e_{\mathtt{g}} + e_{\mathtt{b}} + \nu)} \binom{n_{\mathtt{g}}}{\tilde{e}_{\mathtt{g}}} \binom{n_{\mathtt{b}}}{\tilde{e}_{\mathtt{b}}}\right)^{M-1}. \tag{II.14}$$

The joint probability of decoding failure and ending in state C_{N+1}, conditioned on starting state C_1, denoted $P_{\mathrm{f},C_{N+1}|C_1}(d|c)$, is upper bounded by

$$\bar{P}_{\mathrm{f},C_{N+1}|C_1}(d|c) = \sum_{n_{\mathtt{g}}=0}^{N} \sum_{e_{\mathtt{g}}=0}^{n_{\mathtt{g}}} \sum_{e_{\mathtt{b}}=0}^{n_{\mathtt{b}}} \binom{n_{\mathtt{g}}}{e_{\mathtt{g}}} \binom{n_{\mathtt{b}}}{e_{\mathtt{b}}} \varepsilon_{\mathtt{g}}^{e_{\mathtt{g}}} (1 - \varepsilon_{\mathtt{g}})^{n_{\mathtt{g}} - e_{\mathtt{g}}} \varepsilon_{\mathtt{b}}^{e_{\mathtt{b}}} (1 - \varepsilon_{\mathtt{b}})^{n_{\mathtt{b}} - e_{\mathtt{b}}}$$

$$\times P_{\mathrm{f}|N_{\mathtt{g}},E_{\mathtt{g}},E_{\mathtt{b}}}(n_{\mathtt{g}},e_{\mathtt{g}},e_{\mathtt{b}})P_{N_{\mathtt{g}},C_{N+1}|C_1}(n_{\mathtt{g}},d|c). \quad \text{(II.15)}$$

In a similar fashion, the joint probability of undetected failure accounting for boundary states, $P_{\mathrm{ue},C_{N+1}|C_1}(d|c)$, is upper bounded by

$$\bar{P}_{\mathrm{ue},C_{N+1}|C_1}(d|c) = \sum_{n_{\mathtt{g}}=0}^{N}\sum_{e_{\mathtt{g}}=0}^{n_{\mathtt{g}}}\sum_{e_{\mathtt{b}}=0}^{n_{\mathtt{b}}}\binom{n_{\mathtt{g}}}{e_{\mathtt{g}}}\binom{n_{\mathtt{b}}}{e_{\mathtt{b}}}\varepsilon_{\mathtt{g}}^{e_{\mathtt{g}}}(1-\varepsilon_{\mathtt{g}})^{n_{\mathtt{g}}-e_{\mathtt{g}}}\varepsilon_{\mathtt{b}}^{e_{\mathtt{b}}}(1-\varepsilon_{\mathtt{b}})^{n_{\mathtt{b}}-e_{\mathtt{b}}}$$
$$\times P_{\mathrm{ue}|N_{\mathtt{g}},E_{\mathtt{g}},E_{\mathtt{b}}}(n_{\mathtt{g}},e_{\mathtt{g}},e_{\mathtt{b}})P_{N_{\mathtt{g}},C_{N+1}|C_1}(n_{\mathtt{g}},d|c), \quad \text{(II.16)}$$

where $P_{\mathrm{ue}|N_{\mathtt{g}},E_{\mathtt{g}},E_{\mathtt{b}}}(n_{\mathtt{g}},e_{\mathtt{g}},e_{\mathtt{b}})$ is equal to

$$1-\left(1-2^{-N}\sum_{\mathcal{M}(\gamma e_{\mathtt{g}}+e_{\mathtt{b}}-\nu)}\binom{n_{\mathtt{g}}}{\tilde{e}_{\mathtt{g}}}\binom{n_{\mathtt{b}}}{\tilde{e}_{\mathtt{b}}}\right)^{M-1}.$$

As before, the probability of undetected decoding failure diminishes as ν increases. Also, for most systems, the probability of detected failure is well approximated by the upper bound $P_{\mathrm{f},C_{N+1}|C_1}(d|c)$ because undetected failures are very unlikely.

II.6.2 BCH Codes with Bounded Distance Decoding

Our BCH codes are decoded using bounded distance decoding. It is possible to devise a safety margin and thereby reduce the probability of undetected decoding failures in this setting as well. In this case, an undetected error occurs when the received vector lies in the decoding region of an incorrect codeword. Therefore, shrinking the decoding regions of admissible codewords can prevent undetected failures. Let ν denote the size of the safety margin, and assume that the desired error-correcting capability of the code is $t-\nu$ errors, where t is defined in Section II.5.2. Under this slight modification, the decoder can detect up to $t+\nu$ symbol errors.

We assume that a codeword is mapped to the channel using a uniform random

interleaver and, as such, all error patterns consisting of e errors are equally probable [38]. This introduces a symmetry in the problem that facilitates analysis. Without loss of generality, one can assume that the zero codeword is transmitted to the destination. For this situation, an undetected error occurs whenever the Hamming distance between the received word and a nonzero codeword is less than $t - \nu$.

We consider the performance of this scheme for the binary symmetric channel first. In [61], the probability of undetected error for bounded distance decoding is computed. Using the enhanced detecting radius $t + \nu$ (instead of t), we can write $P_{\text{ue}} = \sum_{e=t+\nu+1}^{N} W(e)P_E(e)$, where $W(e)$ denotes the conditional decoder failure probability defined as the ratio of the number of weight e error patterns lying within distance $t-\nu$ from a codeword over the total number of weight e words in the entire space. This can be written as

$$W(e) = \frac{\sum_{j=0}^{t-\nu} \sum_{l=e-j}^{e+j} A_l \binom{N-l}{(j+e-l)/2} \binom{l}{(j-e+l)/2}}{\binom{N}{e}}, \tag{II.17}$$

where A_l denotes the number of weight l codewords in a BCH code space, designed to correct up to $t = \left\lfloor \frac{(t-\nu)+(t+\nu)}{2} \right\rfloor$ errors where $(t-\nu)+(t+\nu) = d_{\min} - 1$. In other words, we use the weight distribution of a t error-correcting BCH code in our decoder design; however, by using the lower $t-\nu$ error correcting capability and $t+\nu$ error detecting capability, we get better performance in terms of undetected errors.

Still, a main issue with this expression is that the weight distributions for most BCH codes are not known. Furthermore, when an expression is known [62], it may be too complicated to integrate into our analysis. Nevertheless, one can approximate the weight distribution of a binary primitive BCH code of length $N = 2^m - 1$ and designed distance $d_{\min} = 2t + 1$, where $2t - 1 < 2^{\lceil m/2 \rceil} + 1$, by a binomial-like

distribution as [61],

$$A_l = \begin{cases} 1, & l = 0 \\ 0, & 1 \leq l < d_{\min} \\ 2^{-mt}\binom{N}{l}(1 + E_l), & d_{\min} \leq l \leq \lfloor \frac{N}{2} \rfloor \\ A_{N-l}, & \lfloor \frac{N}{2} \rfloor \leq l \leq N \end{cases} \tag{II.18}$$

where E_l is an error term in the approximation of the weight distribution of the BCH code by a binomial distribution. It has been shown that for moderately large block lengths, E_l is negligible. Consequently, $W(e)$ is well approximated by $2^{-mt}\sum_{j=0}^{t-\nu}\binom{N}{j}$. As a result, the probability of undetected error is approximately

$$P_{\text{ue}} \approx 2^{-mt}\sum_{j=0}^{t-\nu}\binom{N}{j}\sum_{e=t+\nu+1}^{N} P_E(e).$$

This interpretation generalizes to the Gilbert-Elliott channel, and the conditional probability of undetected error is equal to

$$P_{\text{ue},C_{N+1}|C_1}(d|c) = \sum_{e=t+\nu+1}^{N} W(e)P_{E,C_{N+1}|C_1}(e,d|c),$$

where $c, d \in \{\mathtt{g}, \mathtt{b}\}$ and $W(e)$ is unchanged from (II.17). Similar to the BSC case, this function is well approximated by

$$2^{-mt}\sum_{j=0}^{t-\nu}\binom{N}{j}\sum_{e=t+\nu+1}^{N} P_{E,C_{N+1}|C_1}(e,d|c).$$

This result is supported through numerical simulations.

II.7 Performance Evaluation

In this section, we evaluate our proposed methodology using traffic parameters based on a voice over IP (VoIP) application for an EVDO system, a 3G component of CDMA2000 [63]. This system offers an uplink sector capacity of 500 Kb/s with 16 active users per sector [64]. For a VoIP system with more users and lower per-user rates, this is somewhat optimistic. As such, for illustrative purposes, we choose a

total uplink rate of 460 Kb/s per sector; this gives a rate of $R_{\mathrm{b}} = 28.75$ Kb/s for each active user.

The enhanced variable rate codec (EVRC), used by CDMA2000 systems for low bit-rate speech, generates a voice packet every 20 ms. EVRC features four distinct frame types corresponding to different bit-rates: full rate gives 171 bits, 1/2 rate gives 80 bits, 1/4 rate gives 40 bits, and 1/8 rate gives 16 bits. Hereafter, we adopt the rough estimates of the relative frequencies for the speech coder states published in [63]. Moreover, as the header size for voice packets are usually very large relative to the voice payload, we assume that ROHC compression is employed to reduce overhead to four bytes. Under these parameters, the average size of a voice packet becomes $1/\rho = \sum_i f_i(l_i + \text{overhead}) = 88.55$ bits, where f_i is the relative frequency of state i and l_i denotes the frame size for the same state. The number of header bits in every segment is set to $h = 2$. Throughout the numerical evaluation, packets are assumed to arrive according to a Poisson process. Since packets are generated every 20 ms, we find that $\lambda = 50$ packets per second and we receive an average of $50/R_{\mathrm{b}}$ packets/channel use.

The choice of a Poisson arrival process (or MMPP) allows us to make fair comparisons between codes with different block lengths. In particular, the rate λ in packets per channel use is fixed, and arrivals in the queue correspond to the number of packets produced by the source during the transmission time of one codeword. The marginal distribution of the sampled process is also Poisson with arrival rate λN, in packets per codeword. Given this framework, a prime goal is to minimize the tail probability of the queue over possible values for parameters N and K.

One drawback associated with our closed-form approach is that handling undetected block errors in a realistic manner (e.g., via late detection when the packet CRC fails) is not possible. Therefore, to facilitate the analysis, we assume

the presence of a genie that informs the receiver when an undetected block decoding error occurs. Still, we require that the system maintain a probability of undetected error less than some threshold, and we disregard (N, K) parameter pairs that violate this constraint. Then, we evaluate the tail probability of the queue (the probability that the number of packets in the queue exceeds a prescribed threshold τ) over all admissible values of N, K, and ν satisfying the undetected error probability constraint. More precisely, we perform a two-stage procedure. During the initial phase, the algorithm finds the smallest admissible integer ν corresponding to each pair (N, K), subject to the prescribed upper bound on $P_{\text{ue}}(N, K, \nu)$. Once this is accomplished, the tail probability of the queue $\sum_{i=2\tau+1}^{\infty} \pi(i)$ is evaluated for different (N, K) pairs using the optimum value of ν found in the previous step. We emphasize that distribution $\{\pi(i)\}$ is an implicit function of $\nu(N, K)$ and $P_{\text{f}}(N, K, \nu)$. To perform this procedure, we first evaluate the undetected error probabilities for different rates and $\nu = 0$. In many cases, $\nu = 0$ satisfies the constraint. For rates with high probabilities of undetected error, we increase ν progressively as to reduce the corresponding probabilities of undetected failure. We stress that this necessarily increases the overall probability of decoding failure, as seen in (II.12)–(II.16). Since we are interested in keeping the latter probability as small as possible, we raise ν until the undetected-error requirement is met and then stop. The proper value of ν is generally very small, which makes the task fast and convenient. Note that the initial phase of the procedure can be carried out offline beforehand, whereas the parameters of the coding scheme can be selected based on the current system conditions. Values of N and K for which this procedure gives poor performance are ignored.

For illustrative purposes we present the curves corresponding to the tail probability of the queue versus the code rates, for various block lengths. This effectively helps to understand how the choice of the code parameters significantly

affects the queueing performance. Furthermore, these curves reveal the existence of an optimal code rate associated to each clock length, and an optimal block length over all possible code lengths. As such, one can fairly pick the (N, K) pair which results in the best queueing performance.

While numerically evaluating our proposed methodology, we consider two cases: random coding with ML decoding over the BSC, and BCH coding with bounded distance decoding over Gilbert-Elliott channel. The concise size of this survey is due, primarily, to space limitations. Nonetheless, we believe that the insights offered by these two cases are applicable to other scenarios as well.

II.7.1 Random Codes over the Binary Symmetric Channel

Let the channel bit error rate be $p = 0.1$, which yields a capacity of $C = 0.531$ bits per channel use, and suppose that the constraint on $P_{\text{ue}}(N, K, \nu)$ is 5×10^{-5}. We know that increasing code rate R for a fixed block length decreases redundancy and therefore reduces the error-correcting capability of the code. Thus, the probability of decoding failure found in (II.4) becomes larger. At the same time, changes in code rate affect ρ_r, the probability with which a codeword contains the last parcel of information of a packet. As this rate varies, these two effects alter the transition probabilities and, hence, the stationary distribution of the Markov chain in opposite ways.

Figure II.6 shows the complementary cumulative distribution functions evaluated at $\tau = 10$ packets as functions of K. For each (N, K) pair, ν has been chosen to satisfy the undetected error probability constraint, following the steps outlined above. Each curve corresponds to a different block length and, as seen on the graph, there is a natural tradeoff between the probability of decoding failure and the payload per codeword. For a fixed block length, neither the smallest segment length nor the

largest one delivers optimal performance. Moreover, block length must be selected carefully; longer codewords do not necessarily yield better queueing performance. For our system, optimal parameters are close to $(N, K) = (150, 51)$, for which, the probability of undetected error is 3.67×10^{-5}, and $\nu = 4$.

II.7.2 BCH Codes over the Gilbert-Elliott Channel

The parameters for our Gilbert-Elliott model are selected loosely based on QPSK modulation, a vehicular speed of 20 mph, and a carrier frequency of 2.1 GHz. This gives a normalized Doppler frequency of $f_D T_s = 0.00082$, where f_D represents the Doppler frequency and $T_s = 2/R_{\mathrm{b}}$ is the symbol transmission time. Setting the SNR threshold for transitions between the good and bad states to a common value of $\gamma_{\mathrm{th}} = 2$ dB and the average received SNR to $\bar{\gamma} = 15$ dB, we can apply the formulas given in [47] and get model parameters

$$\alpha = \frac{\rho f_D T_s \sqrt{2\pi}}{e^{\rho^2} - 1} = 0.3938 \qquad \beta = \rho f_D T_s \sqrt{2\pi} = 0.0202$$

where $\rho = 10^{(\gamma_t - \bar{\gamma})/20}$. The probabilities of error in the good and bad states are chosen to be

$$\varepsilon_{\mathtt{g}} = \frac{\alpha + \beta}{\alpha} \int_{\gamma_{\mathrm{th}}}^{\infty} f_\Gamma(\gamma) P_{\mathrm{e-QPSK}}(\gamma) d\gamma = 0.0097,$$
$$\varepsilon_{\mathtt{b}} = \frac{\alpha + \beta}{\beta} \int_{0}^{\gamma_{\mathrm{th}}} f_\Gamma(\gamma) P_{\mathrm{e-QPSK}}(\gamma) d\gamma = 0.3713,$$

where $f_\Gamma(\cdot)$ is the probability distribution of the received SNR and $P_{\mathrm{e-QPSK}}(\gamma) = 1 - (1 - \mathcal{Q}(\sqrt{\gamma}))^2$ is the probability of symbol error for QPSK modulation.

This time, we require that the system features a probability of undetected error no greater than 10^{-5}. Recall that, for a specific (N, K)-BCH code, we can tradeoff the probability of misclassification and the ability to correct errors by changing the value of ν. Hence, we evaluate the tail probability of the queue over all admissible values

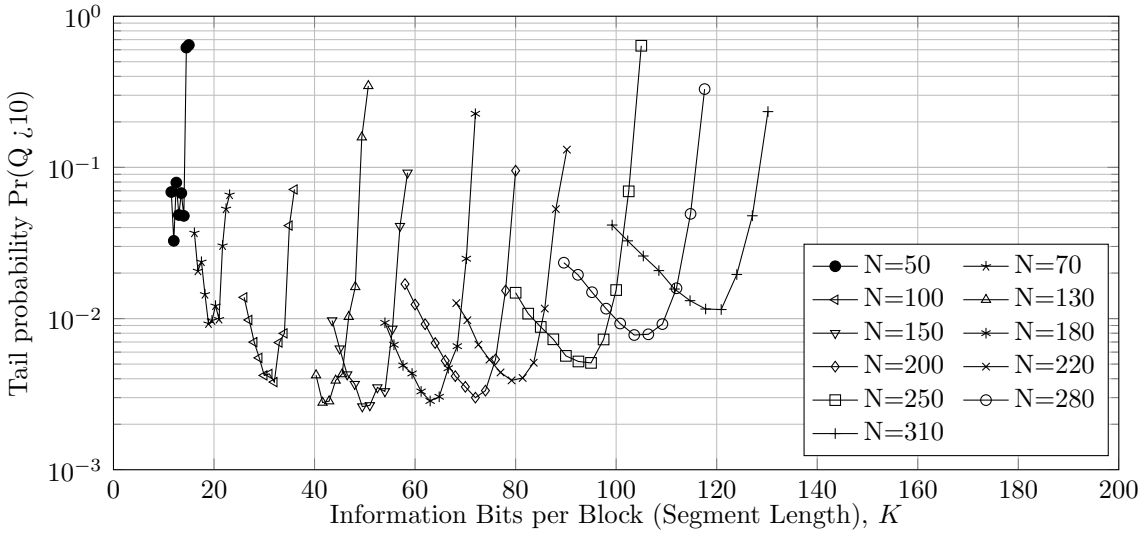

Figure II.6: Probabilities of buffer overflow for random codes over the BSC as functions of K, subject to constraint $P_{\text{ue}} \leq 5 \times 10^{-5}$.

of N, K, and ν satisfying the undetected error probability constraint. To proceed, we first evaluate tail probabilities for admissible values of N and K, with $\nu = 0$ (see Fig. II.7(a)). Then, for the values of K with high probabilities of undetected error, we increase ν progressively as to control misclassifications and meet the desired constraint. Again, values of N and K that lead to inferior performance are discarded. For example, for $N = 63$, the values of $K = 30$, 36, 39, 45 are the ones with high probability of undetected error that are refined by increasing ν (see Fig. II.7(a)-(b)). The values of K greater than 45 associated to $N = 63$ are ignored, since they result in poor performance after meeting the constraint on the undetected errors. Interestingly, for $N = 63$, $K < 30$, the constraint on the undetected errors is met with $\nu = 0$. Similar behavior is also observed for other block lengths.

The results associated with this procedure, in terms of the tail probability of the queue evaluated at $\tau = 5$, are illustrated in Fig. II.7(b). Comparing this graph to Fig. II.7(a), we gather that decreasing the likelihood of undetected error increases

the tail probability of the queue. In fact, because this forces the system to declare a detected error and request a retransmission more often, packets leave the queue less frequently. Accordingly, the probability that the buffer exceeds a certain threshold goes up. Looking at Fig. II.7(b), we see that the optimal code parameters are $(N, K) = (63, 36)$. The corresponding probability of undetected error is 8.78×10^{-6} and $\nu = 1$. We note that the tail probability for $(N, K) = (127, 71)$ is close to this optimal value. This alternate configuration features an undetected error probability of 3.80×10^{-8}, which is achieved with $\nu = 0$. This survey demonstrates the need to adjust the value of ν on a per code basis. Moreover, the results suggest that the proper value of ν is very small relative to N.

Figure II.8 plots the stability factor for the (N, K) pairs found in our previous graph. Systems for which λ_N/μ_N is larger than one are unstable. We note that the tail probability is a good predictor of stability. In general, systems with small stability factors feature good delay profiles as well.

Monte Carlo simulations provide additional empirical evidence for our proposed methodology. This is especially important because our analysis assumes the existence of a genie that reports undetected errors. To understand the effect of the genie, we perform simulations with and without the genie. As expected, the genie-aided simulation results match our analysis almost perfectly. In the absence of a genie, we assume that an undetected decoding error is eventually revealed by the packet CRC. So long as the probabilities of undetected error remain relatively small, our simulations without the genie agree with both the coding and queueing performance predicted by the analytical framework. For instance, Fig. II.7(b) superimposes simulation results for $N = 63$ without the genie (dashed curve). The plotted curves in this case are nearly indistinguishable.

Another important concern pertains to possible modeling inaccuracies related

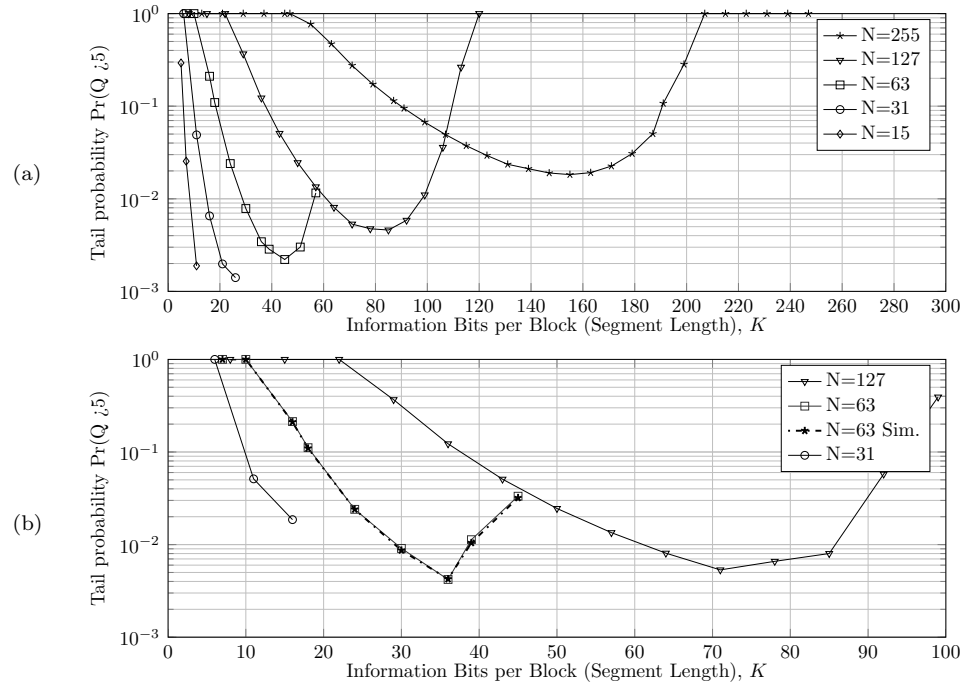

Figure II.7: Probabilities of buffer overflow are displayed for various BCH codes over Gilbert-Elliott channel; (a) when undetected errors are not considered ($\nu = 0$), (b) when the decoding radius in every case is adjusted to meet the constraint on the probability of undetected error $P_{\text{ue}} \leq 10^{-5}$.

to the traffic or the channel. To examine such limitations, we carry Monte Carlo simulations for a system with constant packet lengths, $L = 90$. Figure II.9 demonstrates the results in terms of the complementary cumulative distribution function (CCDF) of the queue occupancy for $N = 63$ and different values of K. We compare the results with those obtained for systems with geometric packet distributions, matching the means. Not surprisingly, reducing variations in the arrival process decreases the tail probability of the queue. That is, it makes the probability of a long queue very small. This behavior should be expected since fixing the packet size precludes the arrival of a very long packet, an event that exacerbates the distribution of the queue. In other words, designing the system using a geometric packet distribution leads to a conservative performance assessment compared to using

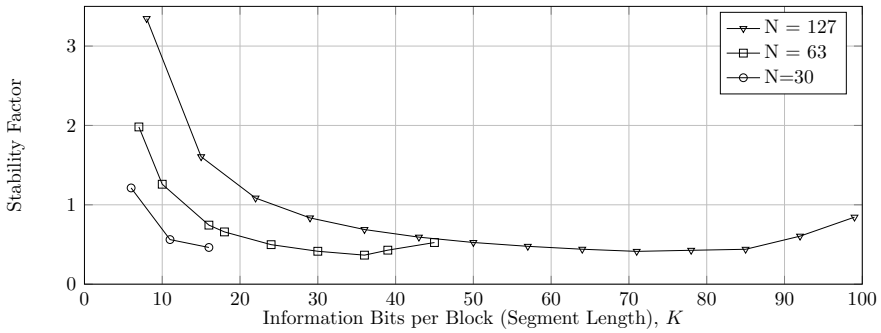

Figure II.8: Stability factors as functions of BCH code parameters; when this factor exceeds one, the system is unstable.

a constant packet length. Empirically, the system performs uniformly better in the latter case. In a similar manner, smoothing the arrival process over time (e.g., periodic arrivals) should lead to a better profile. Simulation results also suggest that the CCDF of the stationary queue length for the system with periodic packet arrivals is dominated by that of the system with Poisson packet arrivals (using the same packet length distribution). The results for $N = 63$ and different block lengths are presented in Fig. II.10.

Performance prediction aside, our analytical framework affords an efficient and accurate means of selecting system parameters. For example, under stated channel conditions and queueing objectives, the optimum values for N and K are the same for constant and geometric packet length distributions. Specifically, the minimum tail probability associated with the abstract model is achieved at $N = 63$ and $K = 36$. Simulation results with constant packet sizes lead to the same operating point, although this latter approach is much more computationally demanding. Altogether, simulation results offer strong support for the proposed methodology.

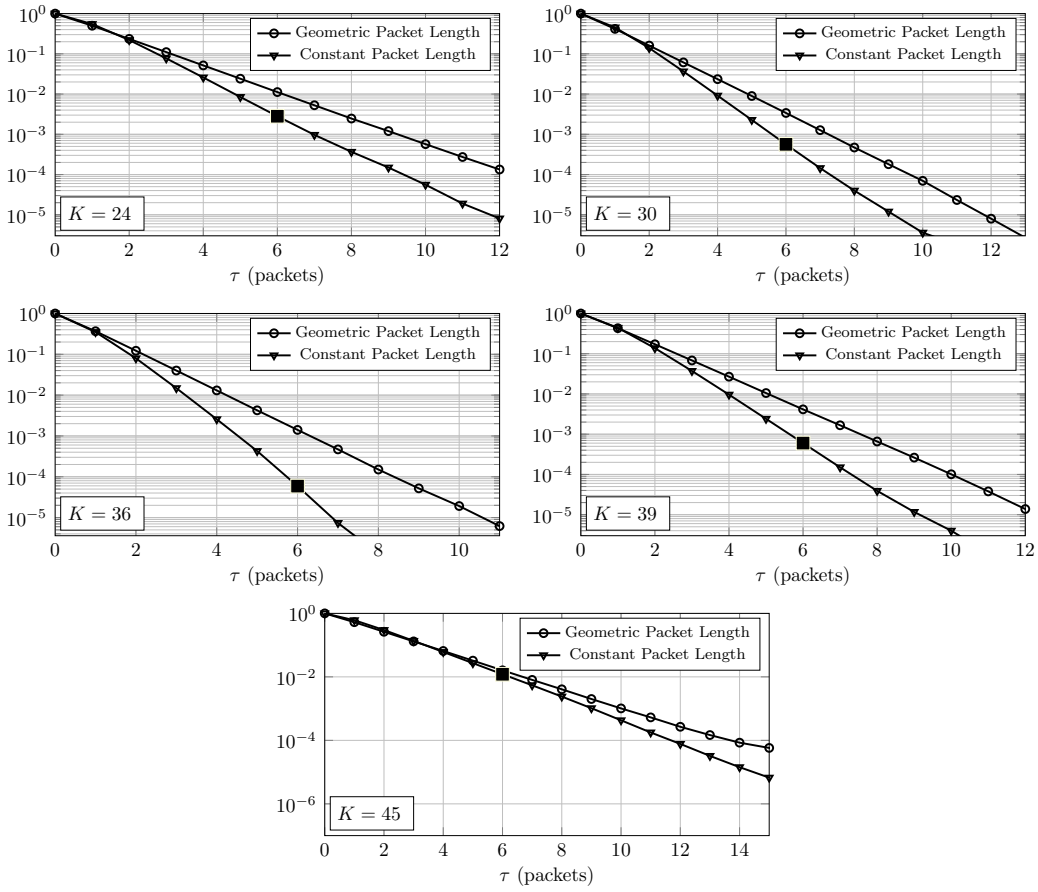

Figure II.9: CCDF of stationary distribution of the queue length ($\Pr(\text{queue length} > \tau)$), is displayed for geometric and constant packet length, $N = 63$. The tail probability of the queue for $\tau = 5$, has been marked with black squares in case of constant packet size.

II.8 Concluding Remarks

In this chapter, we introduced a novel framework to study the queueing behavior of coded wireless communications over finite-state error channels. The proposed methodology applies to both memoryless channels and channels with memory. Careful consideration is given to undetected decoding failures, as they can have a very detrimental impact on the operation of pragmatic systems.

For illustrative purposes, a VoIP application is considered. Channel parameters are derived from the CDMA2000 family of 3G mobile technology standards. The

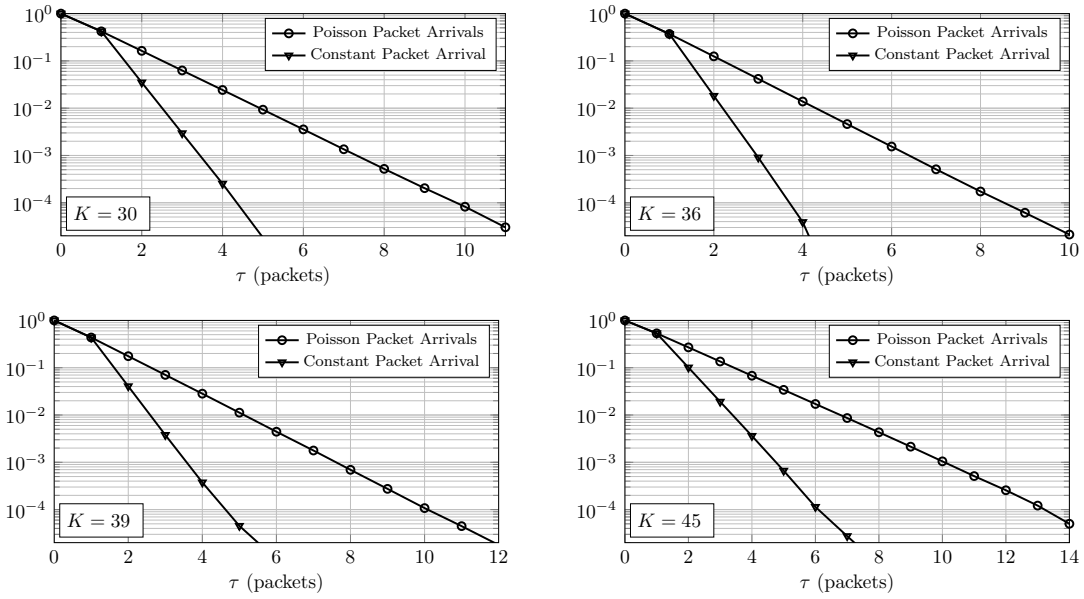

Figure II.10: CCDF of the stationary queue length for Poisson and periodic (constant) packet arrival, $N = 63$

proposed methodology enables the numerical evaluation of the queue distribution and the tail probabilities of the queue. This, in turn, can be employed to find the optimal operating point. Due to the scalable arrival profile, the framework allows for rigorous comparison of coding schemes with different block lengths and code rates. The results and assumptions associated with our methodology are supported by Monte Carlo simulations.

CHAPTER III

ON THE PERFORMANCE OF SHORT-BLOCK RANDOM CODES OVER FINITE-STATE FADING CHANNELS IN THE RARE-TRANSITION REGIME

III.1 Introduction

In this chapter, we study the bounding performance of finite-state channels with memory. The chapter is organized as follows. The channel model, coding strategy and developed exponential bounds on probability of decoding failure are described in Section III.2. In Section III.3, we propose a novel methodology which captures both the effects of channel memory as well as the impact of the channel state at the onset of a codeword. In Section III.4 we briefly review the derivation of exact probability of decoding failure, mainly described in the first chapter. Standard modifications to the upper bounds on decoding failures in order to lower the probability of undetected error are discussed in Section III.5. The remainder of the chapter is devoted to the potential implication of the proposed bounding technique in terms of the queueing theory. In Section III.6, the required adjustments in the queueing model are studied. The stochastic dominance is discussed in Section III.7. Numerical results evaluating the proposed upper bounds and showing the bounding performance of the communication system are presented in Section III.8. Finally, we offer pertinent conclusions in Section III.9.

III.2 Modeling and Exponential Bounds

In this chapter, we consider indecomposable finite-state channels where state transitions are independent of the input symbols. Such channels are often classified

as fading models, and they have been used extensively in the information theory literature. We employ X_n and Y_n, respectively, to denote the input and output symbols at time n. The channel state that determines the channel law at time n is represented by S_{n-1}. We typically reserve capital letters for random variables, whereas lower case letters identify outcomes and values. Boldface letters are used to denote length-N sequences of random variables or outcomes. For groups of random variables, we use the common expression $P_{|\cdot}(\cdot|\cdot)$ to denote the conditional joint probability mass function, and $P_{\mathrm{e},|\cdot}(\cdot|\cdot)$ to denote the conditional joint probability of decoding error.

In general, the conditional probability distribution governing a finite-state channel can be written as

$$P_{Y_n,S_n|X_n,S_{n-1}}\left(y_n, s_n|x_n, s_{n-1}\right) = \Pr\left(Y_n = y_n, S_n = s_n|X_n = x_n, S_{n-1} = s_{n-1}\right).$$

When state transitions are independent of input symbols, this conditional distribution reduces to

$$P_{Y_n,S_n|X_n,S_{n-1}}(y_n, s_n|x_n, s_{n-1}) = P_{S_n|S_{n-1}}(s_n|s_{n-1})P_{Y_n|X_n,S_{n-1}}(y_n|x_n, s_{n-1}). \quad \text{(III.1)}$$

Throughout, we assume that the channel statistics are homogeneous over time and the sequence $\{S_n\}$ forms a Markov chain. When dealing with finite-state channels, it is customary to use integers to denote the possible input and output symbols; in our exposition, we adhere to this convention.

As introduced in the first chapter, the famed Gilbert-Elliott channel is the proverbial example of a channel that possesses the structure described above. This quintessential model is governed by a two-state Markov chain, and it is illustrated in Fig. II.1. The transition probability matrix for the Gilbert-Elliott channel is expressed in (II.1), where $[\mathbf{P}]_{ij} = \Pr(S_n = j|S_{n-1} = i)$. The state-dependent

input-output relationship induced by channel state $s \in \{1, 2\}$ is governed by the crossover probability ε_s, where

$$\Pr(x_n = y_n | S_{n-1} = s) = 1 - \varepsilon_s$$
$$\Pr(x_n \neq y_n | S_{n-1} = s) = \varepsilon_s.$$

For convenience, we order states such that $\varepsilon_1 < \varepsilon_2$.

A coding strategy that has proven exceptionally fruitful in information theory is the use of random codes. Upholding this tradition, we adopt a random coding scheme that employs a code ensemble $\mathcal{C}$ with $M = e^{NR}$ elements. Familiarity with this topic may help because random coding arguments tend to be notationally heavy [1]. Variable N denotes the block length of the code, and R represents its rate. Every element in $\mathcal{C}$ corresponds to a sequence of admissible channel inputs, $\mathbf{x} = (x_1, x_2, \ldots, x_N)$. Moreover, codewords are indexed by $k \in \{1, \ldots, M\}$. The input sequence associated with index k, which we denote by $\mathbf{X}(k)$, is determined through the following procedure. Suppose that $Q(\cdot)$ is a distribution on the set of admissible input symbols. Let $\mathbf{Q}_N(\mathbf{x}) = \prod_{n=1}^{N} Q(x_n)$ be the product measure induced by Q. Codeword $\mathbf{X}(k)$ is selected at random according to distribution $\mathbf{Q}_N$, i.e.,

$$\Pr(\mathbf{X}(k) = \mathbf{x}) = \mathbf{Q}_N(\mathbf{x}) = \prod_{n=1}^{N} Q(x_n).$$

We emphasize that every codeword is selected independently from other elements in $\mathcal{C}$. Once a code ensemble has been generated, a message is sent to the destination by first selecting one of the codewords, and then sequentially transmitting its entries over the communication channel. For the sake of clarity, we summarize our assumptions below; they apply from this point forward, unless otherwise stated.

Assumption 5. *Communication takes place over a finite-state channel that admits the conditional decomposition of* (III.1)*. Information is transmitted using the random*

coding strategy outlined above. On the receiver side, a maximum-likelihood decision rule is used to decode the received sequence. Furthermore, the state of the channel is causally known at the receiver.

At this point, it is worth restating our objective. We want to upper bound the probability that a codeword is decoded erroneously at the receiver. Concurrently, we wish to develop a rare-transition regime that remains true to the fact that the channel state at the onset of the codeword transmission process affects the evolution of the system. Ultimately, this can be achieved in an asymptotic framework by slowing down the transition profile of the underlying channel as the block length of the code grows unbounded. One of the repercussions of this setting is that we have to modify some of the results on error exponents presented by Gallager [1]. In particular, we need the ability to restrict a channel sequence $\mathbf{S}$ to specific events and decompose the error probability accordingly.

Our first formal result is a straightforward extension to Theorem 5.6.1 in [1, pp. 135], which is itself quite general. Since we are interested in finite-state channels with memory in a slow transition regime, we require the ability to track channel realizations explicitly. From an abstract perspective, conditioning on a specific fading realization is equivalent to altering the statistical profile of the underlying channel.

Proposition 6. *Suppose that the realization of the channel over the duration of a codeword is given by* $\mathbf{s}$*. Then, for any* $\rho \in [0, 1]$*, the probability of decoding failure at the destination, conditioned on state sequence* $\mathbf{S} = \mathbf{s}$*, is upper bounded by*

$$P_{\mathrm{e}|\mathbf{S}}(\mathbf{s}) \leq e^{-N(E_{0,N}(\rho, \mathbf{Q}_N, \mathbf{s}) - \rho R)}$$

where the exponent $E_{0,N}(\rho, \mathbf{Q}_N, \mathbf{s})$ *is equal to*

$$-\frac{1}{N}\ln\prod_{n=1}^{N}\sum_{y_n}\left[\sum_{x_n} Q(x_n)P_{Y_n|X_n,S_{n-1}}(y_n|x_n,s_{n-1})^{\frac{1}{1+\rho}}\right]^{1+\rho}.$$

Proof. First, we emphasize that the condition $\mathbf{S} = \mathbf{s}$ simply alters the probability measure governing the input-output relationship of the channel. Applying Theorem 5.6.1 in [1] with $M = e^{NR}$, we immediately get

$$P_{\mathrm{e}|\mathbf{S}}(\mathbf{s}) \leq e^{\rho NR}\sum_{\mathbf{y}}\left[\sum_{\mathbf{x}} \mathbf{Q}_N(\mathbf{x})P_{\mathbf{Y}|\mathbf{X},\mathbf{S}}\left(\mathbf{y}|\mathbf{x},\mathbf{s}\right)^{\frac{1}{1+\rho}}\right]^{1+\rho},$$

where $P_{\mathbf{Y}|\mathbf{X},\mathbf{S}}(\mathbf{y}|\mathbf{x},\mathbf{s})$ represents the conditional distribution of receiving $\mathbf{y}$ given $\mathbf{X} = \mathbf{x}$ and $\mathbf{S} = \mathbf{s}$. Moving forward, the crux of the argument is based on interchanging the order of exhaustive products and sums. Under the channel decomposition introduced in (III.1), the double summation that appears in this upper bound becomes

$$\begin{aligned}&\sum_{\mathbf{y}}\left[\sum_{\mathbf{x}}\prod_{n=1}^{N} Q(x_n)P_{Y_n|X_n,S_{n-1}}(y_n|x_n,s_{n-1})^{\frac{1}{1+\rho}}\right]^{1+\rho}\\ &=\prod_{n=1}^{N}\sum_{y_n}\left[\sum_{x_n} Q(x_n)P_{Y_n|X_n,S_{n-1}}(y_n|x_n,s_{n-1})^{\frac{1}{1+\rho}}\right]^{1+\rho}.\end{aligned}$$

Collecting these various results and using equivalent notation, we obtain the desired proposition. □

A key insight revealed through the proof of Proposition 6 is that $E_{0,N}(\rho, \mathbf{Q}_N, \mathbf{s})$ only depends on $\mathbf{s}$ through its empirical distribution, designated $\mathcal{T}(\mathbf{s})$. We state this fact as a corollary because it will become very useful shortly.

Corollary 7. *Let* T *be the empirical state distribution of a sequence of* N *consecutive channel realizations. If* $\mathbf{s}$ *and* $\mathbf{s}'$ *are two sequences such that* $\mathcal{T}(\mathbf{s}) = \mathcal{T}(\mathbf{s}') = T$*, then* $E_{0,N}(\rho, \mathbf{Q}_N, \mathbf{s}) = E_{0,N}(\rho, \mathbf{Q}_N, \mathbf{s}')$*. Furthermore, the probability of decoding failure at*

the destination, conditioned on $\mathbf{S} \in U$*, is bounded by*

$$P_{\mathrm{e}|\mathbf{S}}(U) \leq e^{-N(E_{0,N}(\rho,\mathbf{Q}_N,T)-\rho R)}$$

for all non-empty subset $U \subset \{\mathbf{s}|\mathcal{T}(\mathbf{s}) = T\}$ *and* $\rho \in [0,1]$*. Note that, with some abuse of notation, we have implicitly defined* $E_{0,N}(\rho, \mathbf{Q}_N, T) = E_{0,N}(\rho, \mathbf{Q}_N, \mathbf{s})$ *where* $\mathbf{s}$ *represents any sequence with empirical distribution* $\mathcal{T}(\mathbf{s}) = T$*.*

For the problem at hand, we are especially interested in probabilities of the form $P_{\mathrm{e},S_N|S_0}(s_N|s_0)$. In some sense, each of these represents the probability of a decoding failure while keeping track of boundary states. In view of Corollary 7, it is natural to upper bound this quantity by partitioning the set of possible sequences according to their empirical distributions. This is accomplished below. In stating our results, we use $\mathcal{T}$ to denote the collection of all admissible empirical channel distributions over sequences of length N.

Proposition 8. *Suppose that a codeword is transmitted over a finite-state channel. The joint probability that decoding fails at the destination and* $S_N = s_N$*, conditioned on initial state* $S_0 = s_0$*, is upper bounded as follows*

$$\begin{aligned} P_{\mathrm{e},S_N|S_0}(s_N|s_0) &\leq \sum_{T\in\mathcal{T}} P_{\mathcal{T}(\mathbf{S}),S_N|S_0}(T, s_N|s_0) \min_{\rho\in[0,1]} e^{-N(E_{0,N}(\rho,\mathbf{Q}_N,T)-\rho R)} \\ &\leq \min_{\rho\in[0,1]} \sum_{T\in\mathcal{T}} P_{\mathcal{T}(\mathbf{S}),S_N|S_0}(T, s_N|s_0) e^{-N(E_{0,N}(\rho,\mathbf{Q}_N,T)-\rho R)} \end{aligned} \tag{III.2}$$

where $P_{\mathcal{T}(\mathbf{S}),S_N|S_0}(T, s_N|s_0)$ *represents the probability that* $\mathcal{T}(\mathbf{S}) = T$ *and* $S_N = s_N$*, given initial state* $S_0 = s_0$*.*

Proof. The progression of this demonstration parallels an argument found in Section 5.9 of [1]. By partitioning the set of length-N sequences according to their

empirical distributions, we can write

$$\begin{aligned}
P_{\mathrm{e},S_N|S_0}(s_N|s_0) &= \sum_{\mathbf{s}} \Pr(\mathbf{S}=\mathbf{s}, S_N = s_N|S_0 = s_0) P_{\mathrm{e}|\mathbf{S}}(\mathbf{s}) \\
&= \sum_{T\in\mathcal{T}} \sum_{\mathbf{s}:\mathcal{T}(\mathbf{s})=T} \Pr(\mathbf{S}=\mathbf{s}, S_N = s_N|S_0 = s_0) P_{\mathrm{e}|\mathbf{S}}(\mathbf{s}) \\
&\leq \sum_{T\in\mathcal{T}} \sum_{\mathbf{s}:\mathcal{T}(\mathbf{s})=T} \Pr(\mathbf{S}=\mathbf{s}, S_N = s_N|S_0 = s_0) \min_{\rho\in[0,1]} e^{-N(E_{0,N}(\rho,\mathbf{Q}_N,\mathbf{s})-\rho R)}.
\end{aligned}$$

The inequality in this expression comes from a direct application of Proposition 6. Following our previous observation that $E_{0,N}(\rho, \mathbf{Q}_N, \mathbf{s})$ only depends on $\mathbf{s}$ through its empirical distribution $\mathcal{T}(\mathbf{s})$, we can rewrite this upper bound as

$$P_{\mathrm{e},S_N|S_0}(s_N|s_0) \leq \sum_{T\in\mathcal{T}} \Pr\left(\mathcal{T}(\mathbf{S})=T, S_N = s_N|S_0 = s_0\right) \min_{\rho\in[0,1]} e^{-N(E_{0,N}(\rho,\mathbf{Q}_N,T)-\rho R)}.$$

The first bound in Proposition 8 is equivalent to this inequality, yet it is expressed using a more concise notation. The second bound holds because the sum of minimums is upper bounded by the minimum of the summands. □

In words, this result is obtained by first grouping channel state sequences according to types, applying an exponential upper bound on the probability of decoding failure to each group, and then taking an expectation over possible empirical distributions. From a large deviations perspective, this decomposition into summands is pertinent because it can be employed to identify the dominating behavior of the system as block length becomes increasingly large. Interestingly, the upper bound on the error probability only depends on the initial and final states of the channel through the empirical distributions.

Next, we consider a Gilbert-Elliott type channel, where the cardinality of the channel state space is $\mathcal{S}$. Let sequence $\mathbf{s}$ be fixed and recall that $s_n \in \{1, 2, \ldots, \mathcal{S}\}$.

Then, by Proposition 6, we get

$$\begin{aligned}
&E_{0,N}(\rho, \mathbf{Q}_N, \mathbf{s}) \\
&= -\frac{1}{N} \ln \prod_{n=0}^{N-1} \frac{1}{2^\rho} \left(\varepsilon_{s_n}^{\frac{1}{1+\rho}} + (1-\varepsilon_{s_n})^{\frac{1}{1+\rho}} \right)^{1+\rho} \\
&= -\frac{1}{N} \sum_{i=1}^{\mathcal{S}} n_i \ln \frac{1}{2^\rho} \left(\varepsilon_i^{\frac{1}{1+\rho}} + (1-\varepsilon_i)^{\frac{1}{1+\rho}} \right)^{1+\rho} \\
&= \sum_{i=1}^{\mathcal{S}} \frac{n_i}{N} b_i(\rho),
\end{aligned} \tag{III.3}$$

where n_i is the number of visits to channel state i in sequence $\mathbf{s}$, $\frac{n_i}{N}$ is the fraction of time spent in state i, and

$$b_i(\rho) = -\ln \frac{1}{2^\rho} \left(\varepsilon_i^{\frac{1}{1+\rho}} + (1-\varepsilon_i)^{\frac{1}{1+\rho}} \right)^{1+\rho}. \tag{III.4}$$

Without loss of generality, we assume the error probabilities in different states are ordered such that

$$\epsilon_1 < \epsilon_2 < \cdots < \epsilon_{\mathcal{S}} \leq \frac{1}{2},$$

which implies

$$b_1(\rho) > b_2(\rho) > \cdots > b_{\mathcal{S}}(\rho). \tag{III.5}$$

A key observation from (III.3) is that the upper bound in (III.2) can be rewritten as an expectation with respect to the distribution of a weighted sum of the state occupation times. This observation is of significant importance, as it reduces the computational complexity of the bound in (III.2) by a great amount. The following remark precisely characterizes the statement.

Remark 9. *Suppose the random variable N_i denotes the number of visits to channel*

state i over the duration of a codeword, and let

$$W(\rho) = \frac{1}{N} \sum_{i=1}^{\mathcal{S}} N_i b_i(\rho), \tag{III.6}$$

designate the weighted sum of the normalized occupancy times. Then, the second upper bound on the error probability in (III.2) *can be rewritten as*

$$\min_{\rho \in [0,1]} \sum_{w} P_{W(\rho),S_N|S_0}(w, s_N|s_0) e^{-N(w-\rho R)}, \tag{III.7}$$

where $P_{W(\rho),S_N|S_0}(w, s_N|s_0)$ represents the probability that $W(\rho) = w$ and $S_N = s_N$, given initial state $S_0 = s_0$. We emphasize that the weights $b_i(\rho)$ are functions of the optimizing parameter ρ and they are deterministic.

We note that the exponential term in (III.2) is averaged with respect to the joint distribution of channel state occupations $P_{\mathcal{T}(\mathbf{S}),S_N|S_0}(T, s_N|s_0)$. However, conditioned on $W(\rho)$, the bound is statistically independent of $\mathcal{T}(\mathbf{S})$. In fact, when the channel state information is available at the receiver, $W(\rho)$ provides sufficient statistics to compute the upper bound on probability of decoding error. The joint distribution of the channel states provides more information than what is needed to derive the bound. This key characterization significantly improves the computational efficiency of the bounding technique, especially when the number of channel states increases.

For illustrative purposes, we derive the upper bounds on $P_{\mathrm{e},S_N|S_0}(s_N|s_0)$ for the two-state Gilbert-Elliott channel. Exploiting the Markov structure of this channel, we get

$$\begin{aligned} P_{\mathrm{e},S_N|S_0}(s_N|s_0) &\leq \min_{\rho \in [0,1]} \sum_{\mathbf{s}} e^{-N(E_{0,N}(\rho,\mathbf{Q}_N,\mathbf{s})-\rho R)} \Pr(\mathbf{S} = \mathbf{s}, S_N = s_N|S_0 = s_0) \\ &= \min_{\rho \in [0,1]} \left(\mathbf{e}_{s_0} \begin{bmatrix} a(1,1) & a(1,2) \\ a(2,1) & a(2,2) \end{bmatrix}^N \mathbf{e}_{s_N}^{\mathrm{T}} \right) e^{\rho N R}. \end{aligned} \tag{III.8}$$

In this equation, $\mathbf{e}_i$ represents the unit vector of length two with a one in the

ith position and zero otherwise. Matrix entries are defined by $a(i,j) = [\mathbf{P}]_{ij}\, e^{b_i}$, where the transition probability matrix $\mathbf{P}$ is given in (II.1). We emphasize that this inequality holds for any $\rho \in [0,1]$ and, hence, the bound can be tightened by minimizing over ρ.

Remark 10. *The bound in* (III.8) *is very similar to Gallager's exponential bound for finite-state channels [1, Thm 5.9.3, pp. 185] when the receiver has perfect state information. The main difference is that Gallager considers the ergodic regime and his equation simplifies to the logarithm of the largest eigenvalue of the matrix. We omit this simplification because we are mainly interested in non-asymptotic regimes.*

III.3 The Rare-Transition Regime

In a traditional setting where $\mathbf{P}$ is kept constant, the upper bound given in (III.2) can be refined using the Perron-Frobenius theorem [1, pp. 184–185]. The more intriguing scenario for our purpose is the rare-transition regime where individual transition probabilities vary with N. We introduce such a dependency through the sampling of a continuous-time Markov chain (CTMC). Consider the CTMC $\mathcal{X}(\cdot)$ defined by the infinitesimal generator matrix $\mathbf{Q}$. Furthermore, suppose $\mathcal{X}(\cdot)$ is sampled at every $1/N$ unit of time. Then, we can construct a continuous-time version of the sampled chain as follows,

$$\mathcal{X}_N(t) = \mathcal{X}\left(\frac{\lfloor Nt \rfloor}{N}\right). \tag{III.9}$$

Let $\mathbf{P}_N$ represent the transition probability matrix of the sampled Markov chain given by $\mathcal{X}_N(n/N)$, $n \in \mathbb{N}$. Matrix $\mathbf{P}_N$ is governed by $\mathbf{Q}$ through the equation [65, Thms 2.1.1 & 2.1.2]

$$\mathbf{P}_N = \exp\left(\frac{\mathbf{Q}}{N}\right). \tag{III.10}$$

We note that $\mathbf{P}_N$ is also the transition probability matrix of the Markov chain $\mathcal{X}(t)$ for a time interval of length $1/N$,

$$[\mathbf{P}_N]_{ij} = \Pr\left(\mathcal{X}\left(\frac{1}{N}\right) = j | \mathcal{X}(0) = i\right).$$

As an example, consider a two-state Markov process. In this case, the infinitesimal generator matrix can be written as

$$\mathbf{Q} = \begin{bmatrix} -\mu & \mu \\ \xi & -\xi \end{bmatrix} \qquad \mu, \xi > 0, \tag{III.11}$$

and, consequently, we can express the transition probability matrix of the sampled process as

$$\mathbf{P}_N = \frac{1}{\mu + \xi} \begin{bmatrix} \xi + \mu e^{-\frac{\xi+\mu}{N}} & \mu\left(1 - e^{-\frac{\xi+\mu}{N}}\right) \\ \xi\left(1 - e^{-\frac{\xi+\mu}{N}}\right) & \mu + \xi e^{-\frac{\xi+\mu}{N}} \end{bmatrix}. \tag{III.12}$$

As seen above, jumps in the discrete chain become less likely as N increases. This is expected because a refined sampling of the CTMC does not alter the character of the underlying process. Furthermore, the roles of boundary states are preserved, a property which is key for our subsequent analysis. At this point, it is instructive to note that the inequalities presented in Section III.2 apply in the context of rare transitions as well, albeit using $\mathbf{P}_N$ rather than a fixed $\mathbf{P}$.

An important benefit of the rare-transition regime is the existence of approximate error bounds that can be computed efficiently. These approximate bounds can be obtained using the following steps. First, we show that the distributions of the occupation times for the sampled Markov chains converge to the distribution of the channel state occupation times of the original CTMC, as the sampling interval $1/N$ decreases to zero. Second, we employ standard results pertaining to the convergence of empirical measures to get approximate upper bounds on the probabilities of

decoding failure at the destination, using the continuous measures. We then leverage a numerical procedure to compute the distributions of weighted sums of channel state occupations for the CTMC [66, 67]. Collecting these results, we arrive at the desired characterization of channels with memory.

Following conventional notation, we use Ω to designate the sample space, and we represent a generic outcome by ω. Whenever needed, we use superscript ω to refer to a particular realization. For instance, $\mathcal{X}$ denotes the CTMC whereas $\mathcal{X}^\omega$ symbolizes the sample path associated with realization ω. In a similar fashion, we can make a distinction between the sampled chain $\mathcal{X}_N$ and the realization $\mathcal{X}_N^\omega$ associated with outcome ω. The need for this elaborate notation will become manifest shortly.

Recall the CTMC $\mathcal{X}(t)$ and its sampled variants $\mathcal{X}_N(t)$, as introduced in (III.9). For every channel state i, we define the occupation times pathwise through the integrals below,

$$\eta_i^\omega = \int_0^1 \mathbf{1}_{\{\mathcal{X}^\omega(t)=i\}} dt$$
$$\eta_{N,i}^\omega = \int_0^1 \mathbf{1}_{\left\{\mathcal{X}_N^\omega(t)=i\right\}} dt = \frac{1}{N}\sum_{k=1}^{N} \mathbf{1}_{\left\{\mathcal{X}_N^\omega(k/N)=i\right\}}$$

where $\mathbf{1}_{\{\cdot\}}$ represents the standard indicator function. Having specified the values of the occupation times for every possible outcome ω, these equations unambiguously define random variables η_i and $\eta_{N,i}$.

Proposition 11. *The sequence of random vectors given by*

$$\overline{\eta}_N = (\eta_{N,1}, \ldots, \eta_{N,\mathcal{S}}), \qquad N = 1, 2, \ldots$$

converges almost surely to random vector $\overline{\eta} = (\eta_1, \ldots, \eta_{\mathcal{S}})$ *as* N *approaches infinity.*

Proof. The CTMC $\mathcal{X}(t)$ is time-homogeneous and its state space has finite cardinality. It is therefore non-explosive [65, Sec. 2.7]. This implies that the number

of transitions in the interval $[0, 1]$ is finite almost surely; we can then write $\Pr(\Omega') = 1$, where

$$\Omega' = \{\omega \in \Omega | \mathcal{X}^\omega(t) \text{ has finitely many jumps}\}.$$

For any $\omega \in \Omega'$, the function $\mathcal{X}^\omega$ is bounded and continuous almost everywhere on $[0, 1]$. It therefore fulfills Lebesgue's criterion for Riemann integrability [68, pp. 323] and, as such,

$$\begin{aligned}\eta_i^\omega &= \int_0^1 \mathbf{1}_{\{\mathcal{X}^\omega(t)=i\}} dt \\ &= \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} \mathbf{1}_{\{\mathcal{X}_N^\omega(k/N)=i\}} = \lim_{N\to\infty} \eta_{N,i}^\omega.\end{aligned}$$

Since the number of channel states is finite, this result readily extends to vectors,

$$\lim_{N\to\infty} d_1\left(\overline{\eta}_N^\omega, \overline{\eta}^\omega\right) = 0 \qquad \forall \omega \in \Omega'$$

where $d_1(\cdot, \cdot)$ is the ℓ_1 distance on $\mathbb{R}^{\mathcal{S}}$. Equivalently, we can write

$$\Pr\left(\omega \in \Omega \middle| \lim_{N\to\infty} d_1\big(\overline{\eta}_N^\omega, \overline{\eta}^\omega\big) = 0\right) = 1.$$

That is, $\overline{\eta}_N$ converges to $\overline{\eta}$ almost surely, as desired. □

Almost sure convergence implies convergence in probability and in distribution [69, Sec. 8.5]. Thus, from Proposition 11, we gather that $\overline{\eta}_N$ converges to $\overline{\eta}$ in distribution, which is sufficient for our later purpose. This further implies that the occupation times of a sequence of independent discrete-time Markov chains, each generated according to $\mathbf{P}_N$, converge in distribution to $\overline{\eta}$. That is, the sampling of the CTMC described above is a powerful mathematical tool to get the result we seek.

In the discussion above, we are intentionally vague about the probability laws on Ω. The careful reader will note that our arguments apply to the equilibrium

distribution of the Markov chain as well as the conditional measure where the Markov process starts in state $i \in \{1, 2, \ldots, \mathcal{S}\}$, at time zero. Moreover, by extension, these findings apply to probabilities where the final channel state is taken into account. To distinguish between these different scenarios, we introduce a shorthand notation for these joint probabilities,

$$\begin{aligned} F(r_1, r_2, \ldots, r_{\mathcal{S}-1}) &= \Pr\left(\eta_1 \leq r_1, \cdots, \eta_{\mathcal{S}-1} \leq r_{\mathcal{S}-1}\right) \\ F_{ij}(r_1, r_2, \ldots, r_{\mathcal{S}-1}) & \\ &= \Pr\left(\eta_1 \leq r_1, \cdots, \eta_{\mathcal{S}-1} \leq r_{\mathcal{S}-1}, S_{\mathrm{f}} = j, S_{\mathrm{i}} = i\right) \\ F_{j|i}(r_1, r_2, \ldots, r_{\mathcal{S}-1}) & \\ &= \Pr\left(\eta_1 \leq r_1, \cdots, \eta_{\mathcal{S}-1} \leq r_{\mathcal{S}-1}, S_{\mathrm{f}} = j | S_{\mathrm{i}} = i\right). \end{aligned}$$

In our labeling, S_{i} identifies the initial state of the channel and S_{f} specifies its final value. We can define F_N, $F_{N,ij}$ and $F_{N,j|i}$ in an analogous manner. It is immediate from Proposition 11 that $dF_N \Rightarrow dF$, $dF_{N,ij} \Rightarrow dF_{ij}$, and $dF_{N,j|i} \Rightarrow dF_{j|i}$ as N grows to infinity, where the symbol $\Rightarrow$ denotes convergence in distribution. As a direct consequence of Proposition 11, we can apply these results to affine combinations of $\eta_1, \ldots, \eta_{\mathcal{S}}$.

Corollary 12. *Let ρ be fixed and recall the coefficients $b_i(\rho)$ found in* (III.4)*. The sequence of random variables given by*

$$W_N(\rho) = \sum_{i=1}^{\mathcal{S}} \eta_{N,i} b_i(\rho)$$

converges in distribution to random variable

$$W(\rho) = \sum_{i=1}^{\mathcal{S}} \eta_i b_i(\rho)$$

as N approaches infinity.

We point out that the expression for $W(\rho)$ above differs slightly from (III.6), as it reflects the notation developed for the current asymptotic setting. Again, this result is valid for the probability laws associated with F, F_{ij} and $F_{j|i}$. The weighted sum $W(\rho)$ is of such importance in our impending discussion that we introduce a convenient notation for its corresponding probability laws as well,

$$\begin{aligned} G(w) &= \Pr(W(\rho) \leq w) \\ G_{ij}(w) &= \Pr(W(\rho) \leq w, S_{\mathrm{i}} = i, S_{\mathrm{f}} = j) \\ G_{j|i}(w) &= \Pr(W(\rho) \leq w, S_{\mathrm{f}} = j | S_{\mathrm{i}} = i). \end{aligned}$$

In addition, we write G_N, $G_{N,ij}$ and $G_{N,j|i}$ for the measures associated with $W_N(\rho)$. It may be helpful to point out that we introduce a slight abuse of notation in establishing these quantities; the dependence of these probability laws on ρ is implicit. This is intentional as the alternative makes the notation overly cumbersome and confusing. In the limiting case, the measures G, G_{ij}, and $G_{j|i}$ are continuous almost everywhere. Wherever needed, we can emphasize the dependence on ρ by writing

$$\begin{aligned} dG(w) &= f_{W(\rho)}(w)dw \\ dG_{ij}(w) &= f_{W(\rho),S_{\mathrm{i}},S_{\mathrm{f}}}(w,i,j)dw \\ dG_{j|i}(w) &= f_{W(\rho),S_{\mathrm{f}}|S_{\mathrm{i}}}(w,j|i)dw \end{aligned}$$

where $f(\cdot)$ is our generic representation of a probability density, possibly with weighted Dirac delta components.

Proposition 13. *Suppose that a message is transmitted over a memory channel with $\mathcal{S}$ possible states, using the random coding scheme introduced in Section III.2. An*

approximate upper bound for the error probability $P_{\mathrm{e},S_N|S_0}(j|i)$ *is given by*

$$P_{\mathrm{e},S_N|S_0}(j|i) \lesssim \int_{b_{\mathcal{S}}(\rho)}^{b_1(\rho)} \min\left\{1, e^{-N(w-\rho R)}\right\} \times f_{W(\rho),S_{\mathrm{f}}|S_{\mathrm{i}}}(w, j|i)dw. \tag{III.13}$$

The approximation in (III.13) *reflects its potential use in selecting efficient coding schemes. The precise mathematical meaning underlying this equation is described in the proof.*

Proof. Let $\rho \in (0,1)$ be fixed. We know from Corollary 7 that, for channel type T, the error probability is bounded by

$$P_{\mathrm{e}|\mathbf{S}}(T) \leq e^{-N\left(E_{0,N}(\rho,\mathbf{Q}_N,T)-\rho R\right)} = e^{-N(w-\rho R)}.$$

where $w = \sum_{i=1}^{\mathcal{S}} \frac{n_i}{N} b_i(\rho)$ is determined by the channel type. We can readily tighten this bound to

$$P_{\mathrm{e}|\mathbf{S}}(T) \leq \min\left\{1, e^{-N(w-\rho R)}\right\}$$

because individual probabilities cannot exceed one. It is useful to point out that the expression $w - \rho R$ is an affine, strictly increasing function of w. For the purpose of exposition, let $g_N(w)$ be defined by

$$\begin{aligned} g_N(w) &= \min\left\{1, e^{-N(w-\rho R)}\right\} \\ &= \begin{cases} e^{-N(w-\rho R)} & w < \varrho \\ 1 & w \geq \varrho \end{cases}, \end{aligned}$$

where the threshold $\varrho = \rho R$. The sequence of functions $\{g_N(w)\}$ converges pointwise

to

$$g(w) = \mathbf{1}_{[\varrho,\infty)}(w) = \begin{cases} 0 & w < \varrho \\ & \\ 1 & w \geq \varrho \end{cases},$$

which is uniformly continuous on the set $w \in [0, b_1(\rho)] \backslash \{\varrho\}$. By taking an expectation over $W(\rho)$, we get

$$\begin{aligned} P_{\mathrm{e},S_N|S_0}(j|i) &\leq \int_{b_{\mathcal{S}}(\rho)}^{b_1(\rho)} \min\left\{1, e^{-N(w-\rho R)}\right\} dG_{N,j|i}(w) \\ &= \sum_{w} P_{W_N(\rho),S_N|S_0}(w, j|i) \min\left\{1, e^{-N(w-\rho R)}\right\}. \end{aligned} \tag{III.14}$$

While this upper bound can be computed numerically if the distribution of $W_N(\rho)$ is known, we are also interested in approximations that provide good intuition as well as computational efficiency at the expense of a little accuracy.

As a next step, we will establish that

$$\lim_{N\to\infty} \int_{b_{\mathcal{S}}(\rho)}^{b_1(\rho)} g_N \, dG_{N,j|i} = \int_{b_{\mathcal{S}}(\rho)}^{b_1(\rho)} g \, dG_{j|i}. \tag{III.15}$$

From Corollary 12, we know that the set of measures $\{G_{N,j|i}\}$ converges in distribution to $G_{j|i}$. Also, for any converging sequence $w_N \in [b_{\mathcal{S}}(\rho), b_1(\rho)]$ with $\lim_{N\to\infty} w_N = w \neq \varrho$, we have $g_N(w_N) \to g(w)$. This is pertinent because the limiting measure $G_{j|i}(w)$ is continuous at $\varrho \in (b_{\mathcal{S}}(\rho), b_1(\rho))$ and, consequently, the event $\{w = \varrho\}$ has probability zero. Collecting these observations, we can can apply [70, Thm. 5.5] and thereby establish the validity of (III.15).

In view of these results, we can write

$$\int_{b_{\mathcal{S}}(\rho)}^{b_1(\rho)} g_N \, dG_{N,j|i} \approx \int_{b_{\mathcal{S}}(\rho)}^{b_1(\rho)} g_N \, dG_{j|i} \approx \int_{b_{\mathcal{S}}(\rho)}^{b_1(\rho)} g \, dG_{j|i} \tag{III.16}$$

for large enough values of N. Since the first integral in (III.16) provides an upper bound on $P_{\mathrm{e},S_N|S_0}(j|i)$, the approximate upper bound of (III.13) immediately follows.

That is, if the code length is large enough, then the approximation in Proposition 13 is justified. □

Using (III.13) to select system parameters gives an alternate, computationally efficient way to design good systems. We emphasize that one does not necessarily need to compute a sequence of distributions for $\{W_N(\rho)\}$ to follow this solution path. Rather, it is possible to accurately approximate the distribution of $W(\rho)$ directly using an iterative approach. We review one such method below; it leverages numerical techniques introduced in [66, 67]. This method applies to channels with arbitrary, yet finite numbers of states. In contrast, the standard approach associated with (III.14) entails computing $G_{N,j|i}$ explicitly for multiple values of N, a cumbersome task.

Proposition 14 presents a numerical method to compute the distribution of $W(\rho)$. This method is adapted from [66] and, as such, it is presented without detailed proof. In practice, the infinite sum needs to be truncated according to an appropriate criterion. To present this result, we need to introduce relevant notations. Let $\mathcal{A}$ be the transition matrix given by

$$\mathcal{A} = \mathbf{I} + \frac{\mathbf{Q}}{\sigma}$$

where $\mathbf{I}$ is the identity matrix, $\sigma \geq \max_k |\mathbf{Q}_{kk}|$, $k \in \{1, 2, \cdots, \mathcal{S}\}$, is a constant and $\{\mathbf{Q}_{kk}\}$ are the diagonal elements of $\mathbf{Q}$. Also, define matrix $\mathbf{G}(w)$ by

$$[\mathbf{G}(w)]_{ij} = G_{j|i}(w)$$

where $i, j \in \{1, 2, \ldots, \mathcal{S}\}$.

Proposition 14. *Let ρ be fixed and suppose that the channel is initially in state $S_{\mathrm{i}} = i$. The probabilities of the events $\{W(\rho) \leq w, S_{\mathrm{f}} = j | S_{\mathrm{i}} = i\}$ as functions*

of w are continuous almost everywhere, and they have at most $\mathcal{S}$ discontinuities, with possible locations $b_{\mathcal{S}}(\rho), \ldots, b_1(\rho)$. Furthermore, for $w \in [b_k(\rho), b_{k-1}(\rho))$ and $2 \leq k \leq \mathcal{S}$, we have

$$\mathbf{G}(w) = \sum_{n=0}^{\infty} e^{-\sigma} \frac{\sigma^n}{n!} \sum_{l=0}^{n} \binom{n}{l} w_k^l (1 - w_k)^{n-l} \mathbf{C}^{(k)}(n, l),$$

where

$$w_k = \frac{w - b_k(\rho)}{b_{k-1}(\rho) - b_k(\rho)}.$$

The matrices $\left\{\mathbf{C}^{(k)}(n,l)\right\}$ are defined component-wise by

$$\left[\mathbf{C}^{(k)}(n,l)\right]_{cd} = C_{cd}^{(k)}(n,l) \qquad c, d \in \{1, 2, \ldots, \mathcal{S}\},$$

and the individual entries in each of these matrices are given by the following two recurrence relations.

For $k \leq c \leq \mathcal{S}$ and $0 \leq d \leq \mathcal{S}$,

$$\begin{aligned} C_{cd}^{(k)}(n,l) = {} & \frac{b_k(\rho) - b_c(\rho)}{b_{k-1}(\rho) - b_c(\rho)} C_{cd}^{(k)}(n, l-1) \\ & + \frac{b_{k-1}(\rho) - b_k(\rho)}{b_{k-1}(\rho) - b_c(\rho)} \sum_{e=0}^{\mathcal{S}} [\mathcal{A}]_{ce} \, C_{ed}^{(k)}(n-1, l-1) \end{aligned}$$

where $1 \leq l \leq n$. For $n \geq 0$ and $k > 1$, we apply the boundary conditions $C_{cd}^{(1)}(n,0) = 0$ and $C_{cd}^{(k)}(n,0) = C_{cd}^{(k-1)}(n,n)$.

Similarly, for $0 \leq c \leq k-1$ and $0 \leq d \leq \mathcal{S}$,

$$\begin{aligned} C_{cd}^{(k)}(n,l) = {} & \frac{b_c(\rho) - b_{k-1}(\rho)}{b_c(\rho) - b_k(\rho)} C_{cd}^{(k)}(n, l+1) \\ & + \frac{b_{k-1}(\rho) - b_k(\rho)}{b_k(\rho) - b_c(\rho)} \sum_{e=0}^{\mathcal{S}} [\mathcal{A}]_{ce} \, C_{ed}^{(k)}(n-1, l) \end{aligned}$$

where $0 \leq l \leq n-1$. In this case, for $n \geq 0$ and $k < \mathcal{S}$, we can write the boundary conditions $C_{cd}^{(\mathcal{S})}(n,n) = [\mathcal{A}^n]_{cd}$ and $C_{cd}^{(k)}(n,n) = C_{cd}^{(k+1)}(n,0)$.

Proof. We emphasize, again, that this proposition is adapted from a general technique found in [66]. In paralleling the argument presented therein, the weights $b_1(\rho), \ldots, b_{\mathcal{S}}(\rho)$ play the role of reward rates. The possible discontinuities in $G_{j|i}(w)$ have to do with the non-vanishing probabilities that the chain does not visit certain states during time interval $[0, 1)$. □

The following corollary is a consequence of Proposition 14, and it specifies density functions for the smooth parts of matrix $\mathbf{G}$.

Corollary 15. *For $w \in (b_k(\rho), b_{k-1}(\rho))$ and $1 \leq k \leq \mathcal{S}$, we can write*

$$d\mathbf{G}(w) = \frac{\sigma e^{-\sigma}}{b_{k-1}(\rho) - b_k(\rho)} \sum_{n=0}^{\infty} \frac{\sigma^n}{n!} \sum_{l=0}^{n} \binom{n}{l} w_k^l (1 - w_k)^{n-l} \times \left[\mathbf{C}^{(k)}(n+1, l+1) - \mathbf{C}^{(k)}(n+1, l) \right].$$

The discrete-time Markov chain whose probability transition matrix is given by $\mathcal{A}$ is called a uniformized chain [71, 72]. This chain can be paired to a Poisson process with rate σ to form a continuous-time Markov chain. The resulting chain is stochastically equivalent to $\mathcal{X}(t)$ and, as such, it possesses the same invariant probability distribution [66]. Furthermore, the matrix $\mathbf{P}_N$ can be written as

$$\mathbf{P}_N = e^{-\frac{\sigma}{N}} \exp\left(\frac{\sigma \mathcal{A}}{N}\right),$$

or, alternatively,

$$[\mathbf{P}_N]_{ij} = e^{-\frac{\sigma}{N}} \sum_{k=0}^{\infty} \frac{1}{k!} \left(\frac{\sigma}{N}\right)^k \left[\mathcal{A}^k\right]_{ij}.$$

While the uniformized chain and the sampled chain are both derived from $\mathbf{Q}$, there remains an important difference. By construction, the uniformized chain never misses a transition in its corresponding continuous-time Markov chain; whereas the sampled chain with its periodic structure can overlook jumps associated with fast transitions. This makes the uniformized chain a more suitable object in describing Proposition 14.

As a special case of Proposition 14, we turn to the situation where $\mathcal{S} = 2$. Not surprisingly, occupation times for this simple scenario have been studied in the past, and explicit expressions for their distributions exist [73–75]. Still, the distributions provided therein only account for an initial state, and they do not specify a final state. We must therefore modify these results slightly to match the needs of our current framework.

Lemma 16. *Consider a continuous-time Markov chain whose generator matrix is given by* (III.11)*. The joint distributions governing occupation times and the final state, conditioned on the initial state, can be written as follows*

$$\begin{aligned}
f_{\eta_1,S_\mathrm{f}|S_\mathrm{i}}(r,1|1) &= e^{-\mu r-\xi(1-r)} \\
&\quad\times\left(\delta(1-r)+\sqrt{\frac{\mu\xi r}{1-r}}I_1\left(2\sqrt{\mu\xi r(1-r)}\right)\right) \\
f_{\eta_1,S_\mathrm{f}|S_\mathrm{i}}(r,2|1) &= \mu e^{-\mu r-\xi(1-r)}I_0\left(2\sqrt{\mu\xi r(1-r)}\right) \\
f_{\eta_1,S_\mathrm{f}|S_\mathrm{i}}(r,1|2) &= \xi e^{-\mu r-\xi(1-r)}I_0\left(2\sqrt{\mu\xi r(1-r)}\right) \\
f_{\eta_1,S_\mathrm{f}|S_\mathrm{i}}(r,2|2) &= e^{-\mu r-\xi(1-r)} \\
&\quad\times\left(\delta(r)+\sqrt{\frac{\mu\xi(1-r)}{r}}I_1\left(2\sqrt{\mu\xi r(1-r)}\right)\right)
\end{aligned}$$

where $I_0(\cdot)$ and $I_1(\cdot)$ represent modified Bessel functions of the first kind.

Proof. See the appendix. □

The corresponding expressions for the sampled chain are present in the following Lemma.

Lemma 17. *Consider a two-state channel whose transition probability matrix is given by* (II.1)*. Assume that the number of visits to each state is recorded for a period spanning N consecutive channel realizations. The joint distributions governing*

the channel type and its final state, conditioned on the initial state, can be written in terms of the Gaussian hypergeometric function ${}_2F_1(\cdot,\cdot;\cdot;\cdot)$. *In particular, for* $m = 1, 2, \ldots, N-1$, *we get*

$$\begin{aligned}
P_{N_1,S_N|S_0}(m,1|1) &= (1-\alpha)^m(1-\beta)^{N-m} \\
&\quad \times \Big({}_2F_1(-N+m,-m;1;\lambda) \\
&\quad\quad - {}_2F_1(-N+m+1,-m;1;\lambda)\Big) \\
P_{N_1,S_N|S_0}(m,2|1) &= \frac{(1-\alpha)^{m-1}(1-\beta)^{N-m+1}\alpha}{(1-\beta)} \\
&\quad \times {}_2F_1(-N+m,-m+1;1;\lambda) \\
P_{N_1,S_N|S_0}(m,1|2) &= \frac{(1-\alpha)^{m+1}(1-\beta)^{N-m-1}\beta}{(1-\alpha)} \\
&\quad \times {}_2F_1(-N+m+1,-m;1;\lambda) \\
P_{N_1,S_N|S_0}(m,2|2) &= (1-\alpha)^m(1-\beta)^{N-m} \\
&\quad \times \Big({}_2F_1(-N+m,-m;1;\lambda) \\
&\quad\quad - {}_2F_1(-N+m,-m+1;1;\lambda)\Big)
\end{aligned}$$

where $\lambda = \frac{\alpha\beta}{(1-\alpha)(1-\beta)}$. *Special consideration must be given to extremal cases. In particular, we have*

$$\begin{aligned}
P_{N_1,S_N|S_0}(0,\cdot|1) = P_{N_1,S_N|S_0}(N,\cdot|2) &= 0 \\
P_{N_1,S_N|S_0}(0,2|2) &= (1-\beta)^N \\
P_{N_1,S_N|S_0}(N,1|1) &= (1-\alpha)^N.
\end{aligned}$$

Proof. See the appendix. □

We can relate these two results through Proposition 11 and the definition of $\mathbf{P}_N$ in (III.12). Let α and β be the constants defined in Lemma 17, and consider the

assignment

$$\alpha = \frac{\mu}{\mu+\xi}\left(1-e^{-\frac{\xi+\mu}{N}}\right)$$
$$\beta = \frac{\xi}{\mu+\xi}\left(1-e^{-\frac{\xi+\mu}{N}}\right).$$

Then, the discrete distributions specified above converge to the occupation times described in Lemma 16, under the asymptotic scaling $\frac{N_1}{N} \to \eta$, as N grow unbounded. As a side observation, we emphasize that two of the limiting measures in Lemma 16 are not absolutely continuous with respect to the Lebesgue measure. Still, these distributions are well-defined positive measures [69, Chap. 8]. The discontinuities can be explained by the fact that, in the rare-transition limit, there is a positive probability of staying in the initial channel state for the whole block.

Using the limiting distributions in Lemma 16, one can compute the approximate upper bound found in (III.13) for the two-state case. We stress that, for this simple channel, $W(\rho) = (b_1(\rho) - b_2(\rho))\eta + b_2(\rho)$ is an affine function of η. Hence, the distribution of $W(\rho)$ can be derived in terms of η. For this simple case, it is straightforward to compute the approximate bound using the empirical distribution of the state occupancies,

$$P_{\mathrm{e},S_N|S_0}(j|i) \lesssim \int_0^1 \min\left\{1, e^{-N((b_1(\rho)-b_2(\rho))r+b_2(\rho)-\rho R)}\right\} f_{\eta,S_{\mathrm{f}}|S_{\mathrm{i}}}(r,j|i)dr. \qquad \text{(III.17)}$$

Yet, as the number of states increases, dealing with the joint distribution of the state occupation times and integrating over multiple variables is an increasingly complex task. This difficulty is bypassed when we use the distribution of the weighted sum of occupation times since, in this latter case, we are dealing with a single random variable as opposed to a random vector.

In general, getting the distributions of the occupation times for the discrete chains is not needed to apply the result of Proposition 13. However, for the

two-state channel, the distributions are available for both the continuous-time and the sampled chains. It is then instructive to compute the exact upper bound in (III.14) using the distributions given by Lemma 17, and compare it to the approximate bound presented in (III.17). Numerical results for this comparison are presented in Section III.8, thereby offering supporting evidence for our proposed methodology. We note that, even for the simple two-state case, computing the approximate upper bound in (III.17) is considerably more efficient than calculating (III.14). As we will see in Section III.8, the price to pay for this computational efficiency is a little accuracy.

From an engineering point of view, we are interested in cases where N is dictated by the code length of a practical coding scheme. The approximate upper bound can be used to perform a quick survey of good parameters. Then, the exact expression based on the hypergeometric function can be employed for fine tuning locally. As a final note on this topic, we emphasize that these upper bounds can be tightened by optimizing over $\rho \in [0, 1]$. This task entails repeated computations of the bounds, which partly explains our concerns with computational efficiency.

A significant benefit in dealing with the Gilbert-Elliott channel model is its tractability. The remaining of this chapter is devoted to the analysis of error probability and the queueing behavior of systems built around this two-state channel. The next sections are dedicated to the derivation of exact probabilities of detected and undetected decoding failure. This is an intermediate step to characterize system performance, and it allows for fair evaluation of the proposed bounding technique. Furthermore, we also show how to bound the probability of undetected errors with slight modifications to the approximated upper bounds. In Sections III.6 and III.7, we explore the queueing performance of the system, using exact expressions and upper bounds on the detected and undetected probabilities of decoding failure.

III.4 Exact Derivation of Probability of Decoding Failures

It is possible to compute exact probabilities of decoding failure under various decision schemes for the two-state Gilbert-Elliott channel, due to its simplicity. Consequently, in this case, we can assess how close the bounds and the true probabilities of error are from one another. The hope is that, if the Gilbert-Elliott bounds are reasonably tight, then the upper bounds for general finite-state channels will also be good. Of course, it may be impractical to compute exact probabilities of error for more elaborate channels. Even for Gilbert-Elliot type channels with more than two states, deriving and computing exact expressions for the probability of decoding failure rapidly becomes intractable. In such situations, the use of upper bounds for performance evaluation is inevitable.

In [76], the authors study data transmission over a Gilbert-Elliott channel using random coding when the state is known at the receiver. Two different decoding schemes are considered: a minimum-distance decoder and a maximum-likelihood decision rule. For the sake of completeness, we briefly review these results. When channel state information is available at the destination, the empirical distribution of the channel sequence provides enough information to determine the probability of decoding failure. Using the measure on N_1 and the corresponding conditional error probabilities, one can average over all possible types to get the probability of decoding failure,

$$P_{\mathrm{e},S_N|S_0}(s_N|s_0) = \sum_{T\in\mathcal{T}} P_{\mathrm{e}|\mathcal{T}(\mathbf{S})}(T)\Pr(\mathcal{T}(\mathbf{S}) = T, S_N = s_N|S_0 = s_0). \tag{III.18}$$

The conditional probability of decoding failure, given type $T = (n_1, N - n_1)$, is examined further below. The probability distributions governing different channel types can be found in Lemma 17.

Consider the channel realization over the span of a codeword. Suppose $\mathbf{X}_i$ and

$\mathbf{Y}_i$ represent the subvectors of $\mathbf{X}$ and $\mathbf{Y}$ corresponding to time instants when the channel is in state i. We can denote the number of errors that occur in each state using random variables E_1 and E_2, where $E_i = d_{\mathrm{H}}(\mathbf{X}_i, \mathbf{Y}_i)$ and $d_{\mathrm{H}}(\cdot,\cdot)$ represents the Hamming distance. The conditional error probabilities can then be written as

$$\begin{aligned} P_{\mathrm{e}|\mathcal{T}(\mathbf{S})}(T) = \sum_{e_1=0}^{n_1}\sum_{e_2=0}^{n_2} & P_{\mathrm{e}|\mathcal{T}(\mathbf{S}),E_1,E_2}(T,e_1,e_2) \\ & \times P_{E_1,E_2|\mathcal{T}(\mathbf{S})}(e_1,e_2|T), \end{aligned} \tag{III.19}$$

where $n_2 = N - n_1$. Given the channel type, the numbers of errors in the good and bad states are independent,

$$P_{E_1,E_2|\mathcal{T}(\mathbf{S})}(e_1,e_2|T) = P_{E_1|\mathcal{T}(\mathbf{S})}(e_1|T)P_{E_2|\mathcal{T}(\mathbf{S})}(e_2|T). \tag{III.20}$$

Furthermore, E_1 and E_2 have binomial distributions

$$P_{E_i|\mathcal{T}(\mathbf{S})}(e_i|T) = P_{E_i|N_i}(e_i|n_i) = \binom{n_i}{e_i}\varepsilon_i^{e_i}(1-\varepsilon_i)^{n_i-e_i}. \tag{III.21}$$

This mathematical structure leads to the following theorem.

Theorem 18. *When ties are treated as errors, the probability of decoding failure for a length-N uniform random code with M codewords, conditioned on the number of symbol errors in each state and the channel state type, is given by*

$$\begin{aligned} & P_{\mathrm{e}|\mathcal{T}(\mathbf{S}),E_1,E_2}(T,e_1,e_2) \\ & = 1 - \left(1 - 2^{-N}\sum_{(\tilde{e}_1,\tilde{e}_2)\in\mathcal{M}(\gamma e_1+e_2)}\binom{n_1}{\tilde{e}_1}\binom{n_2}{\tilde{e}_2}\right)^{M-1} \end{aligned} \tag{III.22}$$

where $\mathcal{M}(d)$ is the set of pairs $(\tilde{e}_1,\tilde{e}_2) \in \{0,\ldots,N\}^2$ that satisfy $\gamma\tilde{e}_1 + \tilde{e}_2 \le d$. This expression holds with

$$\gamma = \frac{\ln\varepsilon_1 - \ln(1-\varepsilon_1)}{\ln\varepsilon_2 - \ln(1-\varepsilon_2)}$$

for the maximum-likelihood decision rule, and with $\gamma = 1$ *for minimum-distance decoding.*

Proof. See [76]. □

Random codes paired with a maximum-likelihood decoding rule form a permutation invariant scheme. The decoding performance is then determined by the number of symbol errors in each state within a codeword, and not by their order or locations. We point out that the methodology introduced herein can potentially be extended to other permutation invariant encoding/decoding schemes to analyze probability of decoding failure.

III.5 Undetected Errors

As we mentioned in the first chapter, a serious matter with pragmatic communication systems is the presence of undetected decoding failures. In the current setting, this occurs when the receiver uniquely decodes to the wrong codeword. For delay-sensitive applications, this problem is especially important because recovery procedures can lead to undue delay. To address this issue, we apply techniques that help control the probability of admitting erroneous codewords [59,60]. This safeguard, in turn, leads to slight modifications to the performance analysis presented above. In applying these techniques, the probability of undetected failure is a system parameter that must be set during the design phase of the system.

III.5.1 The Exact Approach

In [76], the authors show that the probability of decoding failure (including detected errors, undetected errors, and ties) is given by the equations (III.18)-(III.21)

and substituting

$$P_{\mathrm{e}|\mathcal{T}(\mathbf{S}),E_1,E_2}(T,e_1,e_2) = 1-\left(1-2^{-N}\sum_{(\tilde{e}_1,\tilde{e}_2)\in\mathcal{M}(\gamma e_1+e_2+\nu)}\binom{n_1}{\tilde{e}_1}\binom{n_2}{\tilde{e}_2}\right)^{M-1}, \tag{III.23}$$

where ν is a non-negative parameter that specifies the size of the safety margin for undetected errors. The joint probability of undetected error with ending state S_N, conditioned on starting in state S_0, is $P_{\mathrm{ue},S_N|S_0}(s_N|s_0)$ and can be upper bounded by

$$\bar{P}_{\mathrm{ue},S_N|S_0}(s_N|s_0) = \sum_{T\in\mathcal{T}} \bar{P}_{\mathrm{ue}|\mathcal{T}(\mathbf{S})}(T)\Pr(\mathcal{T}(\mathbf{S})=T, S_N=s_N|S_0=s_0) \tag{III.24}$$

where

$$\bar{P}_{\mathrm{ue}|\mathcal{T}(\mathbf{S})}(T) = \sum_{e_1=0}^{n_1}\sum_{e_2=0}^{n_2}\bar{P}_{\mathrm{ue}|\mathcal{T}(\mathbf{S}),E_1,E_2}(T,e_1,e_2)P_{E_1,E_2|\mathcal{T}(\mathbf{S})}(e_1,e_2|T), \tag{III.25}$$

and

$$\bar{P}_{\mathrm{ue}|\mathcal{T}(\mathbf{S}),E_1,E_2}(T,e_1,e_2) = 1-\left(1-2^{-N}\sum_{(\tilde{e}_1,\tilde{e}_2)\in\mathcal{M}(\gamma e_1+e_2-\nu)}\binom{n_1}{\tilde{e}_1}\binom{n_2}{\tilde{e}_2}\right)^{M-1}. \tag{III.26}$$

Since the probability of an undetected error is typically much smaller than that of a detected error, one can upper bound the probability of detected error by $P_{\mathrm{e},S_N|S_0}(s_N|s_0)$ with a negligible penalty.

III.5.2 Exponential Bound

With slight modifications to the derived exponential upper bound, one can get a similar bound on the probability of undetected error.

Lemma 19. *The exponential upper bounds on $P_{\mathrm{e},S_N|S_0}$, and $\bar{P}_{\mathrm{ue},S_N|S_0}$ can be written as*

$$\tilde{P}_{\mathrm{ue},S_N|S_0}(j|i) = \min_{0\le v\le\rho\le 1} \sum_{n_1=0}^{N} \min\left\{1, e^{-N(E_{0,N}(\rho,\mathbf{Q}_N,n_1)-\rho R-v\tau)}\right\} P_{N_1,S_N|S_0}(n_1,j|i)$$
$$\lesssim \min_{0\le v\le\rho\le 1} \int_0^1 \min\left\{1, e^{-N((b_1(\rho)-b_2(\rho))r+b_2(\rho)-\rho R-v\tau)}\right\} f_{\eta,S_\mathrm{f}|S_\mathrm{i}}(r,j|i)dr \tag{III.27}$$

and

$$\tilde{P}_{\mathrm{e},S_N|S_0}(j|i) = \min_{0\le v\le\rho\le 1} \sum_{n_1=0}^{N} \min\left\{1, e^{-N(E_{0,N}(\rho,\mathbf{Q}_N,n_1)-\rho R-v\tau+\tau)}\right\} P_{N_1,S_N|S_0}(n_1,j|i)$$
$$\lesssim \min_{0\le v\le\rho\le 1} \int_0^1 \min\left\{1, e^{-N((b_1(\rho)-b_2(\rho))r+b_2(\rho)-\rho R-v\tau+\tau)}\right\} f_{\eta,S_\mathrm{f}|S_\mathrm{i}}(r,j|i)dr \tag{III.28}$$

where τ controls the tradeoff between detected and undetected errors and is used to decrease the incidence of undetected errors, in a manner similar to ν for the exact case.

Proof. Following the same approach as in [59, 60], the results are achieved. □

We emphasize that the rare-transition regime, the bounds and the approximation methodologies proposed in this dissertation have a wide range of applications. In fact, the introduced upper bounds can be adopted in various analysis frameworks. The following two sections are dedicated to the potential implications of the proposed bounding techniques in terms of queueing theory. We exploit these results to evaluate the queueing performance of systems built around correlated channels. In particular, we show how stochastic dominance can be combined with these tools to make performance analysis tractable.

III.6 Queueing Model

One of the distinguishing features of the methodology introduced above is the ability to vary block length N and code rate R in a rigorous fashion. This enables

us to account for dependence within and across codewords. This last observation is especially pertinent for queueing systems, as correlation in service is known to exacerbate the distribution of a queue. We describe the components of our queueing model below.

Packets are generated at the source according to a Poisson process with arrival rate λ, measured in packets per channel use. The number of bits per packet forms a sequence of independent geometric random variables, each with parameter $\rho \in (0, 1)$. On arrival, a packet is divided into segments of size RN, where RN is assumed to be an integer. The total number of segments associated with a packet of length L is given by $M = \lceil \frac{L}{RN} \rceil$. As before, a random coding scheme is used to protect the transmitted data while it is transmitted over the channel.

On the receiving end, one must successfully decode all M codewords to recover the corresponding packet. Once this is achieved, this packet is discarded from the queue. We note that random variable M has a geometric distribution with $\Pr(M = m) = (1 - \rho_r)^{m-1}\rho_r$, where $m \geq 1$ and $\rho_r = 1 - (1 - \rho)^{RN}$. Consequently, the number of coded blocks per data packet possesses the memoryless property, a highly desirable attribute for the purpose of analysis.

We emphasize that, in the current framework, a data packet is discarded from the transmit buffer if and only if the destination acknowledges reception of the latest codeword and this codeword contains the last parcel of information corresponding to the head packet. Packet departures are then determined by the channel realizations and the coding scheme. In particular, the code rate R has a major impact on performance. Generally, a lower code rate will have a small probability of decoding failure. However, it also needs more channel uses to complete the transmission of one data packet. Thus, for a fixed channel profile, we can vary the block length and code rate to find optimal system parameters. This natural tradeoff reflects the tension

between the probability of a successful transmission and the size of its payload.

Let Q_s denote the number of data packets waiting in the transmitter queue after s codeword transmission intervals. The channel state at the same time instant is represented by C_{sN+1}. Notice that the channel state evolves more rapidly than events taking place in the queue. This explains the discrepancy between the indices. Based on these quantities, it is possible to define a Markov chain $U_s = (C_{sN+1}, Q_s)$ that captures the joint evolution of the queue and the channel over time. The ensuing transition probabilities from U_s to U_{s+1} are equal to

$$\Pr(U_{s+1} = (d, q_{s+1}) | U_s = (c, q_s)) =$$
$$\sum_{n_1=0}^{N} P_{Q_{s+1}|N_1,Q_s}(q_{s+1}|n_1, q_s) P_{N_1,C_{(s+1)N+1}|C_{sN+1}}(n_1, d|c)$$

where the second term in the summand is given in Lemma 17. We can rewrite $P_{Q_{s+1}|N_1,Q_s}(q_{s+1}|n_1, q_s)$ as

$$\begin{aligned}
&\sum_{e_1=0}^{n_1} \sum_{e_2=0}^{n_2} P_{Q_{s+1},E_1,E_2|N_1,Q_s}(q_{s+1}, e_1, e_2|n_1, q_s) \\
&= \sum_{e_1=0}^{n_1} \sum_{e_2=0}^{n_2} P_{Q_{s+1}|E_1,E_2,N_1,Q_s}(q_{s+1}|e_1, e_2, n_1, q_s) P_{E_1,E_2|N_1,Q_s}(e_1, e_2|n_1, q_s) \\
&= \sum_{e_1=0}^{n_1} \sum_{e_2=0}^{n_2} \binom{n_1}{e_1} \binom{n_2}{e_2} \varepsilon_1^{e_1} (1-\varepsilon_1)^{n_1-e_1} \varepsilon_2^{e_2} (1-\varepsilon_2)^{n_2-e_2} \\
&\qquad \times P_{Q_{s+1}|E_1,E_2,N_1,Q_s}(q_{s+1}|e_1, e_2, n_1, q_s).
\end{aligned}$$

Suppose that the number of packets in the queue is $Q_s = q_s$, where $q_s > 0$. Then, admissible values for Q_{s+1} are restricted to the set $\{q_s - 1, q_s, q_s + 1, \ldots\}$. The

transition probabilities for $q_s > 0$ and $i \geq 0$ are given by

$$\begin{aligned} &P_{Q_{s+1}|E_1,E_2,N_1,Q_s}(q_s + i|e_1, e_2, n_1, q_s) \\ &\quad = a_i P_{\mathrm{e}|E_1,E_2,N_1}(e_1, e_2, n_1) \\ &\quad + a_i(1 - P_{\mathrm{e}|E_1,E_2,N_1}(e_1, e_2, n_1))(1 - \rho_r) \\ &\quad + a_{i+1}(1 - P_{\mathrm{e}|E_1,E_2,N_1}(e_1, e_2, n_1))\rho_r \end{aligned} \tag{III.29}$$

and the probability of the queue decreasing is

$$\begin{aligned} &P_{Q_{s+1}|E_1,E_2,N_1,Q_s}(q_s - 1|e_1, e_2, n_1, q_s) \\ &\quad = a_0 \left(1 - P_{\mathrm{e}|E_1,E_2,N_1}(e_1, e_2, n_1)\right) \rho_r. \end{aligned} \tag{III.30}$$

The queue can only become smaller when there are no arrivals, a codeword is successfully received at the destination, and the decoded codeword contains the last piece of data associated with a packet. Above, $P_{\mathrm{e}|E_1,E_2,N_1}(e_1, e_2, n_1)$ is the conditional probability of decoding failure which appears in (III.23). The terms a_i denotes the probability that i packets arrive within the span of a codeword transmission. Since arrivals form a Poisson process, we have $a_i = \frac{(\lambda N)^i}{i!} e^{-\lambda N}$ for $i \geq 0$. When the queue is empty, $q_s = 0$, (III.29) applies for cases where $i \geq 1$. However, for $i = 0$, the conditional transition probability reduces to

$$\begin{aligned} &P_{Q_{s+1}|E_1,E_2,N_1,Q_s}(0|e_1, e_2, n_1, 0) \\ &\quad = a_0 + a_1 \left(1 - P_{\mathrm{e}|E_1,E_2,N_1}(e_1, e_2, n_1)\right) \rho_r. \end{aligned}$$

The overall profile of this system can be categorized as an M/G/1-type queue. The repetitive structure enables us to employ the matrix geometric method to compute the characteristics of this system and subsequently obtain its stationary distribution [55, 76].

III.7 Stochastic Dominance

When the number of channel states is large, it may be impractical to employ exact probabilities of decoding failure. Even for memoryless channels, finding explicit expressions for different encoding/decoding schemes can be difficult. In the face of such a challenge, it is customary to turn to upper bounds on the probabilities of decoding failure to provide performance guarantees. Furthermore, one can employ such upper bounds to assess the queueing performance of the system through stochastic dominance. We emphasize that this type of argument is not tied to our proposed bounds. Rather, it applies to any upper bound on the probability of decoding failure.

The evolution of the queue length is governed by the Lindley equation,

$$\begin{aligned} Q_{s+1} &= (Q_s + A_s - D_s)^+ \\ &\triangleq \max\{0, Q_s + A_s - D_s\} \end{aligned} \tag{III.31}$$

where A_s is the number of arrivals that occurred during time interval s, and D_s is an indicator function for the potential completion of a packet transmission within the same time period. In this queueing model, the only inherent effect of replacing the probability of decoding failure by an upper bound is a potential reduction in the value of D_s. Using an upper bound on the failure probability naturally gives rise to a new random process $\tilde{Q}_s$ defined by

$$\tilde{Q}_{s+1} = \left(\tilde{Q}_s + A_s - \tilde{D}_s\right)^+ \tag{III.32}$$

where $\tilde{D}_s$ is drawn according to the distribution implied by the upper bound. We wish to show that, conditioned on starting in the same state and given a shared channel trace, the distribution of $\tilde{Q}_s$ offers a conservative estimate of Q_s. To make this statement precise, we turn to an establish concept in probability.

Definition 20. *A random variable Z is stochastically dominated by another random variable $\tilde{Z}$, a relation which we denote by $Z \preceq \tilde{Z}$, provided that*

$$\Pr(Z > z) \leq \Pr(\tilde{Z} > z) \tag{III.33}$$

for all $z \in \mathbb{R}$. This relation extends to conditional probability laws. Suppose that

$$\Pr(Z > z|A) \leq \Pr(\tilde{Z} > z|A) \tag{III.34}$$

for all $z \in \mathbb{R}$. Then, we say that Z is stochastically dominated by $\tilde{Z}$, given A. We write this relation as $Z \preceq_A \tilde{Z}$.

A comprehensive discussion of stochastic dominance can be found in [77, 78]. In order to formulate the results we are interested in, we need to start by introducing two lemmas.

Lemma 21. *The stochastic order defined in* (III.33) *is preserved under the positive part operation, $(\cdot)^+ = \max\{0, \cdot\}$.*

Proof. From [77, Sec 1.A.1], we know that given random variables X and Y, $X \preceq Y$ if and only if

$$\mathbb{E}[\varphi(X)] \leq \mathbb{E}[\varphi(Y)]$$

for all increasing functions $\varphi(\cdot)$ for which the expectations exists. In particular, $\varphi(\cdot) = \max\{0, \cdot\}$ is an increasing function. Thus, if $X \preceq Y$ then one can conclude that $X^+ \preceq Y^+$, as desired. □

We explore the structure of potential packet completion events below. In establishing stochastic dominance, we will condition on a specific channel trace,

$$\vec{C} = \{C_1, C_{N+1}, C_{2N+1}, \ldots\}$$

This ensures that the two stochastic processes experience a same level of difficulty at every step while trying to decode codewords. Note that, for our purposes, the dependence of Q_s and $\tilde{Q}_s$ on $\vec{C}$ is only through decoding attempts and, as such, this dependence is localized in time.

Lemma 22. *Using upper bounds on the probabilities of decoding failure leads to stochastic dominance in potential packet completions,* $\tilde{D}_s \preceq_{\vec{C}} D_s$.

Proof. A potential completion occurs when a codeword is decoded successfully at the destination, and the data it contains is the last segment of a packet. The probability of the latter event is ρ_r, and it is common to both D_s and $\tilde{D}_s$. However, the conditional probabilities of decoding failure differ, with $P_{\mathrm{e}|\vec{C}=\vec{c}} \leq \tilde{P}_{\mathrm{e}|\vec{C}=\vec{c}}$. This, in turn, gives

$$\begin{aligned}\Pr\left(\tilde{D}_s = 1|\vec{C}=\vec{c}\right) &= \left(1-\tilde{P}_{\mathrm{e}|\vec{C}=\vec{c}}\right)\rho_r \\ &\leq \left(1-P_{\mathrm{e}|\vec{C}=\vec{c}}\right)\rho_r = \Pr\left(D_s = 1|\vec{C}=\vec{c}\right).\end{aligned}$$

Since $D_s, \tilde{D}_s \in \{0,1\}$, this equation is enough to establish stochastic ordering. □

Collecting these results, we can turn to the behavior of the system. For a fair comparison, we assume that the two queues have the same number of packets at the onset of the communication process.

Proposition 23. *Suppose that* $Q_0 = \tilde{Q}_0$. *Then, the process* Q_s *is stochastically dominated by* $\tilde{Q}_s$.

Proof. As a first step, we assume that channel trace $\vec{C} = \vec{c}$ is fixed. We prove conditional dominance through mathematical induction. The base case follows from the condition of the theorem. As an inductive hypothesis, suppose that $Q_s \preceq_{\vec{C}} \tilde{Q}_s$. By Lemma 22, we have $\tilde{D}_s \preceq_{\vec{C}} D_s$. Since negation reverses stochastic dominance [77,

p. 9], we deduce that

$$-D_s \preceq_{\vec{C}} -\tilde{D}_s.$$

This, in turn, yields the relation

$$Q_s + A_s - D_s \preceq_{\vec{C}} \tilde{Q}_s + A_s - \tilde{D}_s.$$

The last step leverages the closure property of stochastic orders under convolutions [77, Thm 1.A.3 (b)]. Applying the increasing function $\varphi(\cdot) = \max\{0, \cdot\}$ to both sides, we immediately get

$$(Q_s + A_s - D_s)^+ \preceq_{\vec{C}} \left(\tilde{Q}_s + A_s - \tilde{D}_s\right)^+$$

from Lemma 21. That is, $Q_{s+1} \preceq_{\vec{C}} \tilde{Q}_{s+1}$, thereby establishing our inductive step.

At this point, the statement of the proposition can be obtain by taking expectations over channel traces. Since the channel states are independent of queue sizes and code generation, the probabilistic weighing is the same for both Q_s and $\tilde{Q}_s$. This guarantees that the stochastic ordering is preserved. In other words, the relation $Q_s \preceq \tilde{Q}_s$ holds at all times. □

There is a subtle distinction in the argument presented above. The random variable D_s indicates potential completion. Actual departures from the queue only take place when there are packets awaiting transmission. Mathematically, the distinction is resolved through the positive part operation. Conceptually, when the queue is empty, the source attempts to send a virtual packet with no physical meaning. This object is created for mathematical convenience.

In view of the aforementioned results, we can also consider the interplay between stochastic dominance and the stationary distributions of the queues. When the Markov chains U_s and $\tilde{U}_s$ are positive recurrent, the corresponding queueing processes

Q_s and $\tilde{Q}_s$ are stable. In such cases, the stationary distribution associated with $\tilde{Q}_s$ dominates the stationary distribution of Q_s. In particular, for any integer q, we have $\Pr(Q_s > q) \leq \Pr\left(\tilde{Q}_s > q\right)$ and, hence, in the limit we obtain

$$\begin{aligned}\Pr(Q > q) &= \lim_{s\to\infty} \Pr(Q_s > q) \\ &\leq \lim_{s\to\infty} \Pr\left(\tilde{Q}_s > q\right) = \Pr\left(\tilde{Q} > q\right)\end{aligned}$$

That is, $Q \preceq \tilde{Q}$ where Q and $\tilde{Q}$ denote the stationary distributions of the two queueing processes listed above.

From a more intuitive point of view, one can argue that, pathwise, increasing the probability of failure can only result in fewer departures and, as such, there will remain at least as many packets waiting in the queue. In other words, when comparing two queueing systems with a same arrival process, a same underlying channel, and a same code generator, more decoding failures can only exacerbate the size of the queue. This observation holds in some generality and can be employed when the exact decoding error probability are not known or difficult to compute. This approach allows one to provide performance guarantees for a queueing system using bounds on the probabilities of decoding failure.

III.8 Numerical Results

In this section, we present numerical results for probabilities of decoding error and we compared them to the derived upper bounds. We also evaluate queueing performance using the exact error probabilities and their upper bounds derived in the rare-transition regime.

III.8.1 Comparison of Exponential Upper Bounds

We consider a communication system which transmits data over a Gilbert-Elliott channel with crossover probabilities $\varepsilon_1 = 0.01$ and $\varepsilon_2 = 0.1$. Figure III.1 shows the approximate upper bounds of (III.13) as functions of block length and code rate, and compares them to the standard Gallager-type bounds of (III.8). Each curve shows the value of the bound averaged over all possible state transitions. Although the block lengths are relatively short, the approximate bounds are very close to the standard Gallager-type bounds. Furthermore, the difference becomes more negligible as N grows larger.

In Fig. III.2, we plot the probabilities of decoding failure for the maximum-likelihood and minimum-distance decoders given by (III.18)-(III.22), against the bounds provided in (III.13). As anticipated, the maximum-likelihood decision rule outperforms the minimum distance decoder. For fixed N, there is a roughly constant ratio between the approximate upper bounds and the exact probabilities of error under maximum-likelihood decoding. This is not too surprising as similar statements can be made about the accuracy of Gallager-type bounds. Still, as block length increases, the approximate bounds get progressively closer to the exact values. We note that the figure features particularly short block lengths, as it is difficult to compute exact performance for channels with memory.

III.8.2 Evaluation of Queueing Performance

We turn to the evaluation of overall performance and we consider a situation where, on average, packets are generated every 20 msec. This yields a rate of $\lambda = 50$ packets per second for the arrival process. The symbol rate for our binary channel is set to 28.75 Kb per second, which leads to an expected $1/575$ packets per channel use. The cross-over probabilities of the Gilbert-Elliott channel are set to $\varepsilon_1 = 0.01$

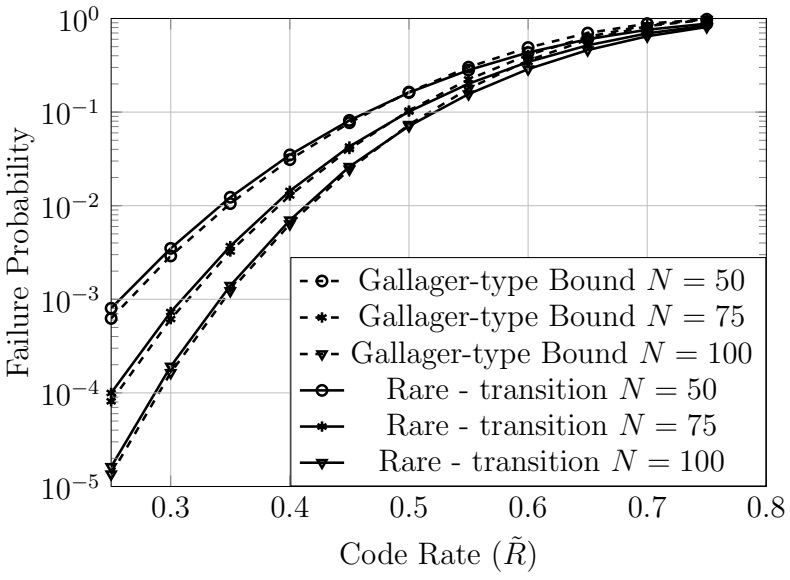

Figure III.1: Comparison of the approximate upper bound (III.13) with the exact bound (III.8) in the rare-transition regime with $N[\mathbf{P}]_{12} \simeq 4$ and $N[\mathbf{P}]_{21} \simeq 6$.

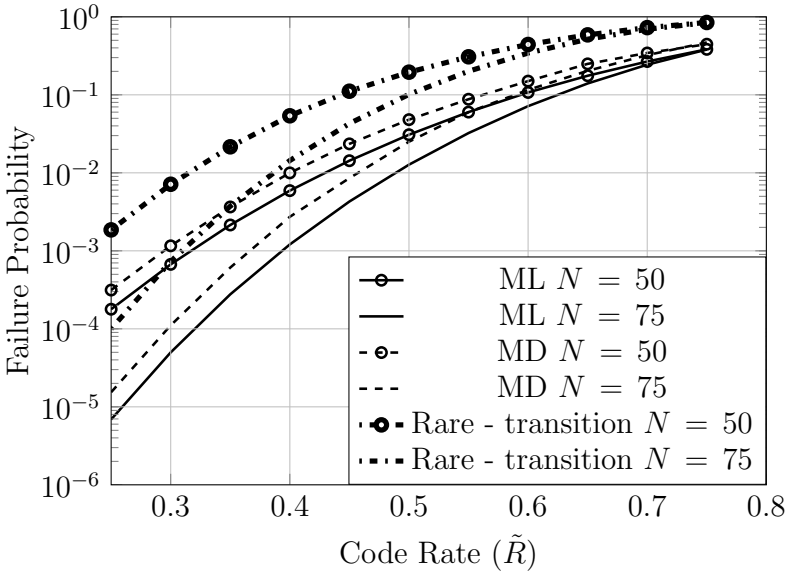

Figure III.2: Comparison of the approximate upper bound (III.13) with the exact probabilities of decoding failure under maximum-likelihood (ML) and minimum-distance (MD) decoding for $[\mathbf{P}]_{12} = 0.0533$ and $[\mathbf{P}]_{21} = 0.08$.

and $\varepsilon_2 = 0.1$, and its state transition probabilities are $\alpha = 0.0533$ and $\beta = 0.08$. Shannon capacity when the state is known at the receiver is therefore equal to 0.764 bits per channel use.

Increasing code rate R for a fixed block length decreases redundancy and therefore reduces the error-correcting capability of the code. Thus, the probability of decoding failure becomes larger. At the same time, changes in code rate affect ρ_r, the probability with which a codeword contains the last parcel of information of a packet. As code rate varies, these two phenomena alter the transition probabilities and, hence, they influence the stationary distribution of the Markov system in opposite ways.

The choice of a Poisson arrival process allows us to make fair comparisons between codes with different block lengths. In particular, the rate λ in packets per channel use is fixed, and arrivals in the queue correspond to the number of packets produced by the source during the transmission time of one codeword. The marginal distribution of the sampled process is Poisson with arrival rate λN, in packets per codeword. This formulation is new, and it bridges coding decision to queueing behavior in a rigorous manner. To examine overall system performance, we assume the existence of a genie which informs the receiver when an undetected decoding error occurs. Undetected errors are intrinsic to error channels, and this approach is standard when it comes to analysis. Still, for consistency, we require the system to feature a very low probability of undetected error, e.g., less than 10^{-5}, by a proper choice of the safety margin. That is, we only consider (N, R) pairs that meet this additional constraint.

Given this framework, a prime goal is to minimize the tail probability of the queue over all admissible values of N, R, and τ or ν that satisfy the constraint on undetected error. To perform this optimization using the approximate exponential bound, we first evaluate the bound on undetected error probabilities for different rates and for

$\tau = 0$ in (III.27). Then, for rates with high probability of undetected error, we increase τ so that the bound on probability of undetected error is decreased. Recall that this increases the probability of decoding failure. As we are also interested in minimizing the latter probability, we increase τ until the system meets the error-detecting condition and then stop. The values of N and R for which this procedure gives poor performance are ignored. A similar approach is used for system evaluation with exact error probabilities by changing the value of ν in (III.23), (III.26).

Figure III.3 shows the approximate probability of the queue exceeding a threshold as a function of system parameters. The constraint on the number of packets in the queue is set to five, which reflects our emphasis on delay-sensitive communication. We have chosen τ in (III.27)-(III.28) such that $\max_{i,j} \tilde{P}_{\mathrm{ue},S_N|S_0}(j|i)$ remains below 10^{-5}. The code rate considered vary from 0.25 to 0.75, with a step size of 0.05. Each curve corresponds to a different block length. As seen on the graph, there is a natural tradeoff between the probability of decoding failure and the payload per codeword. For a fixed block length, neither the smallest segment length nor the largest one delivers optimal performance. Moreover, block length must be selected carefully; longer codewords do not necessarily yield better queueing performance as they may result in large decoding delays. As such, the tail probability has a minimum over all rates and block lengths. Therefore, there are interior optimum points for both N and R. We see in Fig. III.3 that the optimum code parameters are close to $(N, R) = (170, 0.5)$. For this particular set of code parameters, we have $\tau = 0.048$.

Figure III.4 offers similar plots for the exact failure probability. Again, the optimum code parameters are near $(N, R) = (170, 0.5)$. In this case, $\nu = 8$ is the smallest value of ν that keeps $\max_{i,j} \bar{P}_{\mathrm{ue},S_N|S_0}(j|i)$ below the 10^{-5} threshold. As we can see by comparing the results, performance evaluation based on the

bound gives very good estimates for optimum coding parameters and overall system performance. Not only does the approximate bound give a good estimate of performance, it accurately predicts ideal system parameters for code block as small as 125. In addition, since the approximate bounds are slightly pessimistic, they produce conservative estimate of overall performance. Empirically, the systems perform better than predicted by the approximate error bounds.

III.9 Concluding Remarks

In this chapter we introduced the rare-transition regime to characterize communication systems where the block length is of the same order as the channel memory. As such, we have derived an approximate upper bound which is particularly appropriate for the rare-transition regime. We also employed expressions for the exact error probability for a system that uses random codes, under maximum-likelihood and minimum-distance decoding over Gilbert-Elliott channels. In fact, comparison with the exact derivations strongly supports that the bounding technique provides meaningful results. Next, the queueing performance of the system is studied using the proposed upper bounds.

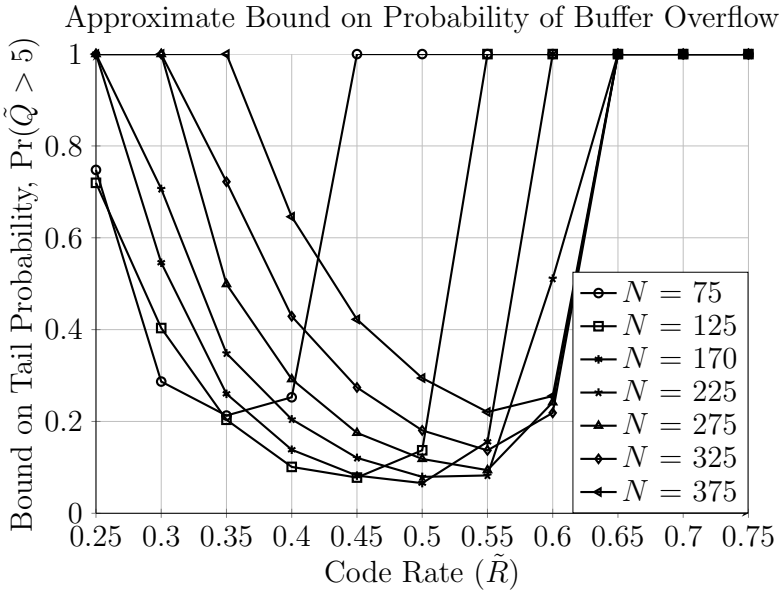

Figure III.3: Approximate bounds on the probability of the queue exceeding a threshold as functions of block length N and code rate R. The system parameters considered above are subject to $\max_{i,j} \tilde{P}_{\mathrm{ue},S_N|S_0}(j|i) \leq 10^{-5}$.

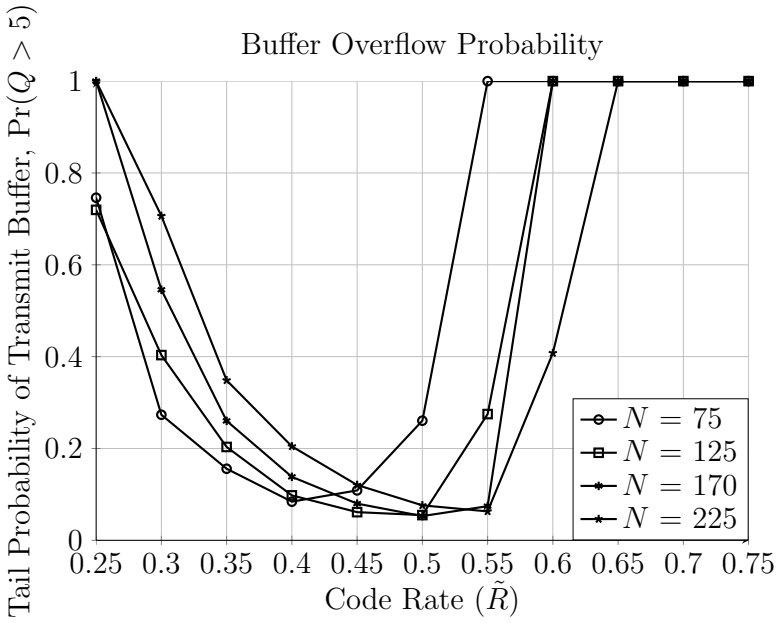

Figure III.4: Exact probability of the queue exceeding a threshold as functions of block length N and code rate R. These system parameters meet the constraint $\max_{i,j} \bar{P}_{\mathrm{ue},S_N|S_0}(j|i) \leq 10^{-5}$.

CHAPTER IV

CONCLUSIONS AND FUTURE WORK

This dissertation aims to develop a better understanding of delay-sensitive communication of coded data, queue-based performance criteria and service dependencies associated to channel memory. Performance evaluation has been done through both exact and bounding characterization of the coding schemes.

In the first chapter, we introduce a novel framework to study the queueing behavior of coded wireless communications over finite-state error channels. Through this framework, it is possible to select the optimal the block length and code rate of the encoding scheme based on the requirements of the system. This is especially useful in the context of delay-sensitive applications for which long block lengths are inadequate. The proposed methodology applies to both memoryless channels and channels with memory. Due attention is given to detected and undetected decoding failures. We dedicated independent sections in the first chapter to analyze each of these decoding failures. As undetected decoding failures can have a very detrimental impact on the operation of pragmatic systems, careful consideration should be given to them in the system design phase. Specifically, by using a safety margin, one can limit the likelihood of such events and thereby ensure adequate performance.

The proposed methodology enables the numerical evaluation of the equilibrium queue distribution. This, in turn, can be employed to compute the tail probabilities of the queue occupancy and, subsequently, find the optimal operating point. Our framework supports the rigorous comparison of coding schemes with different block lengths and code rates. This possibility arises due to the scalable arrival profile adopted in our framework. This study suggests that, for fixed conditions, optimal

system parameters are essentially unaffected by small variations in the buffer overflow threshold. The results and assumptions associated with our methodology are supported by Monte Carlo simulations. This technique can be employed to facilitate adaptive modulation schemes that take into account both the channel profile and the requirements of the underlying traffic. The optimization task can be carried out offline beforehand, whereas the parameters of the coding scheme can be selected based on current system conditions. Possible avenues of future research in this direction, include better accounting for feedback and extending this type of analysis to multi-user environments.

In the second chapter, we mostly focus on the rare-transition regime. This regime introduces a powerful methodology to characterize communication systems where the block length is of the same order or smaller than the coherence time of the channel. This mode of operation is common in many practical implementations. This fact serves as a motivation for the proposed framework. To estimate the probability of decoding error, we have derived an approximate upper bound specifically tailored to the rare-transition regime. A key property of the proposed methodology is that the dependency on the initial and final states is retained in the analysis. This bound is also numerically efficient to compute and can be employed for parameter selection and performance analysis in communication links with queueing constraints. We provided supportive evidence for the accuracy of the bounding methodology by deriving exact expressions for the Gilbert-Elliott channel. Both maximum-likelihood decoding and a minimum-distance decision rule were considered. A numerical comparison between exact and approximate results validates our approach, showcasing the predictive power of the approximate bounding techniques. The numerical study focused on a two-state Markov channel formulation where state information is available at the destination.

The methodology was subsequently extended to performance criteria based on the queueing behavior of the system. We provided a practical method to choose the block length and code rate as to minimize the probability that the transmit buffer exceeds a prescribed threshold. This is especially pertinent for communication links that support delay-sensitive traffic, yet it applies to general data stream as delay is known to negatively affect the performance of flow control and congestion control protocols. Numerical studies suggest that, for fixed conditions, optimal system parameters are essentially unaffected by small variations in the buffer overflow threshold.

In many coded communication systems, the exact decoding failure probability is not known or very complicated to compute. Therefore, a stochastic dominance argument is also used to compare the performance of the system computed using an upper bound on decoding failure probability with the exact performance of the system. For such systems, our results imply that upper bounds on the decoding failure probability naturally imply upper bounds on the performance of the queueing system. Finally, our numerical results imply that, for random coding on the Gilbert-Elliott channel, the performance analysis using the bound on failure probability gives a good estimate of both the system performance and the optimum code parameters.

The methodology and results are developed for finite-state Markov channels, but can be generalized to more intricate channels with memory, with or without symmetry property. In addition, the performance characterization of random codes over finite-state channels may extend to more practical schemes, such as iterative decoding of the Low Density Parity-Check (LDPC) codes. Possible avenues of future research further include the performance analysis of the rare-transition regime in the absence of side information at the receiver. The conjecture is that in the limiting case, the channel sojourn time in each state is long enough for the receiver to estimate

the state. For instance, for a channel with high correlation, patterns of errors with the same number of errors within a block are not equally likely. In fact, the system is more prone to burst of errors when the channel quality is poor. In other words, long channel memory enables the receiver to predict the channel quality. Hence, one might expect similar performance when the state information is not provided at the receiver in the rare-transition regime. Moreover, the correct form of the missing scale factor in front of the proposed exponential bounds is yet unknown. In fact, even for the Gallager type bounds for channels with memory, the scale factor is yet unknown. Derivation of this factor would be a significant improvement in the context of decoding error probabilities.

REFERENCES

[1] R. G. Gallager, *Information Theory and Reliable Communication.* New York, NY: Wiley, 1968.

[2] I. Bettesh and S. Shamai, "Optimal power and rate control for minimal average delay: The single-user case," *IEEE Trans. Inform. Theory*, vol. 52, no. 9, pp. 4115–4141, 2006.

[3] J. Cao and E. M. Yeh, "Power-delay tradeoff analysis for communication over fading channels with feedback," in *Proc. IEEE Int. Symp. Inform. Theory*, Toronto, ON, 2008, pp. 614–618.

[4] X. Li, X. Dong, and D. Wu, "Queue length aware power control for delay-constrained communication over fading channels," *Int. J. Wireless Communications and Mobile Computing*, vol. 12, no. 10, pp. 901–909, 2010.

[5] D. Rajan, A. Sabharwal, and B. Aazhang, "Delay and rate constrained transmission policies over wireless channels," in *Proc. IEEE Global Telecom. Conf.*, vol. 2, San Antonio, TX, 2001, pp. 806–810.

[6] D. Rajan, A. Sabharwal, and A. Aazhang, "Delay-bounded packet scheduling of bursty traffic over wireless channels," *IEEE Trans. Inform. Theory*, vol. 50, no. 1, pp. 125–144, 2004.

[7] R. A. Berry and R. G. Gallager, "Communication over fading channels with delay constraints," *IEEE Trans. Inform. Theory*, vol. 48, no. 5, pp. 1135–1149, 2002.

[8] R. Berry, "Optimal power/delay trade-offs in fading channels-small delay asymptotics," *IEEE Trans. Inform. Theory*, vol. 59, no. 6, pp. 3939–3952, 2013.

[9] N. Ahmed and R. G. Baraniuk, "Throughput maximization for ARQ-like systems in fading channels with coding and queuing delay constraints," in *Proc. Asilomar Conf. on Signals, Systems & Computers*, vol. 2, Pacific Grove, CA, 2004, pp. 1463–1467.

[10] B. P. Dunn and J. N. Laneman, "Rate-delay tradeoffs for communicating a bursty source over an erasure channel with feedback," in *Proc. IEEE International Zurich Seminar on Communications*, Zurich, Switzerland, 2008, pp. 136–139.

[11] R. Negi and S. Goel, "An information-theoretic approach to queuing in wireless channels with large delay bounds," in *Proc. IEEE Global Telecom. Conf.*, vol. 1, Dallas, TX, 2004, pp. 116–122.

[12] N. Ahmed and R. G. Baraniuk, "Throughput measures for delay-constrained communications in fading channels," in *Proc. Annual Allerton Conf. on Commun., Control, and Comp.*, vol. 41, no. 3, Monticello, Ill, 2003, pp. 1496–1505.

[13] S. Kittipiyakul, P. Elia, and T. Javidi, "High-SNR analysis of outage-limited communications with bursty and delay-limited information," *IEEE Trans. Inform. Theory*, vol. 55, no. 2, pp. 746–763, 2009.

[14] X. Yu, J. W. Modestino, and D. Fan, "Joint multiple description coding and FEC for delay-constrained applications in congested networks," in *Proc. IEEE Inform. Theory Workshop*, Tahoe City, CA, 2007, pp. 547–552.

[15] A. Ephremides and B. Hajek, "Information theory and communication networks: An unconsummated union," *IEEE Trans. Inform. Theory*, vol. 44, no. 6, pp. 2416–2434, 1998.

[16] I. E. Telatar and R. G. Gallager, "Combining queueing theory with information theory for multiaccess," *IEEE J. Select. Areas Commun.*, vol. 13, no. 6, pp. 963–969, 1995.

[17] C. Comaniciu and V. H. Poor, "On the capacity of mobile ad hoc networks with delay constraints," *IEEE Trans. Wireless Commun.*, vol. 5, no. 8, pp. 2061–2071, 2006.

[18] V. Anantharam and S. Verdu, "Bits through queues," *IEEE Trans. Inform. Theory*, vol. 42, no. 1, pp. 4–18, 1996.

[19] L. Liu, P. Parag, J. Tang, W.-Y. Chen, and J. F. Chamberland, "Resource allocation and quality of service evaluation for wireless communication systems using fluid models," *IEEE Trans. Inform. Theory*, vol. 53, no. 5, pp. 1767–1777, 2007.

[20] X. Li, C. C. Wang, and X. Lin, "Throughput and delay analysis on uncoded and coded wireless broadcast with hard deadline constraints," in *Proc. IEEE Conf. on Computer Commun, INFOCOM*, San Diego, CA, 2010, pp. 1–5.

[21] D. Wu and R. Negi, "Effective capacity: a wireless link model for support of quality of service," *IEEE Trans. Wireless Commun.*, vol. 2, no. 4, pp. 630–643, 2003.

[22] J. S. Harsini and F. Lahouti, "Adaptive transmission policy design for delay-sensitive and bursty packet traffic over wireless fading channels," *IEEE Trans. Wireless Commun.*, vol. 8, no. 2, pp. 776–786, 2009.

[23] A. Fu, E. Modiano, and J. Tsitsiklis, "Optimal energy allocation for delay-constrained data transmission over a time-varying channel," in *Proc. IEEE Conf. on Computer Commun, INFOCOM*, vol. 2, San Francisco, CA, 2003, pp. 1095–1105.

[24] A. Steiner and S. Shamai, "On queueing and multilayer coding," *IEEE Trans. Inform. Theory*, vol. 56, no. 5, pp. 2392–2415, 2010.

[25] R. N. Swamy and T. Javidi, "Delay analysis of block coding over a noisy channel with limited feedback," in *Proc. Asilomar Conf. on Signals, Systems & Computers*, Pacific Grove, CA, 2008, pp. 1431–1435.

[26] B. Shrader and A. Ephremides, "A queueing model for random linear coding," in *Proc. IEEE Military Commun. Conf.*, Orlando, FL, 2007, pp. 1–7.

[27] A. Sahai, "Why do block length and delay behave differently if feedback is present?" *IEEE Trans. Inform. Theory*, vol. 54, no. 5, pp. 1860–1886, 2008.

[28] A. Eryilmaz, A. Ozdaglar, M. Médard, and E. Ahmed, "On the delay and throughput gains of coding in unreliable networks," *IEEE Trans. Inform. Theory*, vol. 54, no. 12, pp. 5511–5524, 2008.

[29] R. Cogill, B. Shrader, and A. Ephremides, "Stable throughput for multicast with random linear coding," *IEEE Trans. Inform. Theory*, vol. 57, no. 1, pp. 267–281, 2011.

[30] R. Cogill and B. Shrader, "Queue length analysis for multicast: Limits of performance and achievable queue length with random linear coding," in *Proc. Annual Allerton Conf. on Commun., Control, and Comp.*, Monticello, Ill, 2009, pp. 462–468.

[31] S. Bhadra and S. Shakkottai, "Looking at large networks: Coding vs. queueing," in *Proc. IEEE Conf. on Computer Commun, INFOCOM*, Barcelona, Catalunya, Spain, 2006, pp. 1–12.

[32] D. E. Lucani, M. Medard, and M. Stojanovic, "Random linear network coding for time-division duplexing: queueing analysis," in *Proc. IEEE Int. Symp. Inform. Theory*, Seoul, Korea, 2009, pp. 1423–1427.

[33] M. Ravanbakhsh, A. I. Barbero, and O. Ytrehus, "Improved delay estimates for a queueing model for random linear coding for unicast," in *Proc. IEEE Int. Symp. Inform. Theory*, Seoul, Korea, 2009, pp. 1413–1417.

[34] A. I. Barbero and Ø.Ytrehus, "Queuing aspects of packet coding based bidirectional communication over satellite channels," in *Proc. 5th IEEE Advanced Satellite Multimedia Systems Conf., 2010*, Cagliari, Italy, 2010, pp. 38–45.

[35] B. Shrader and A. Ephremides, "Queueing delay analysis for multicast with random linear coding," *IEEE Trans. Inform. Theory*, vol. 58, no. 1, pp. 421–429, 2012.

[36] P. Parag, J.-F. Chamberland, H. D. Pfister, and K. R. Narayanan, "Code-rate selection, queueing behavior, and the correlated erasure channel," *IEEE Trans. Inform. Theory*, vol. 59, no. 1, pp. 397–407, Jan. 2013.

[37] E. N. Gilbert, "Capacity of a burst-noise channel," *The Bell System Technical Journal*, vol. 39, pp. 1253–1265, September 1960.

[38] E. O. Elliott, "Estimates of error rates for codes on burst-noise channels," *The Bell System Technical Journal*, vol. 42, pp. 1977–1997, September 1963.

[39] R. M. Fano, *Transmission of Information.* Cambridge, MA: The M.I.T. Press, 1961.

[40] A. Barg and G. D. Forney, "Random codes: Minimum distances and error exponents," *IEEE Trans. Inform. Theory*, vol. 48, no. 9, pp. 2568–2573, 2002.

[41] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inform. Theory*, vol. 56, no. 5, pp. 2307–2359, 2010.

[42] ——, "Dispersion of the Gilbert-Elliott channel," *IEEE Trans. Inform. Theory*, vol. 57, no. 4, pp. 1829–1848, 2011.

[43] A. Martinez, "Saddlepoint approximation of random-coding bounds," in *Proc. Annual Workshop on Inform. Theory and its Appl.*, San Diego, CA, 2011, pp. 1–6.

[44] C. Nair, E. Ordentlich, and T. Weissman, "Asymptotic filtering and entropy rate of a hidden Markov process in the rare transitions regime," in *Proc. IEEE Int. Symp. Inform. Theory*, Adelaide, Australia, 2005, pp. 1838–1842.

[45] Y. Peres and A. Quas, "Entropy rate for hidden Markov chains with rare transitions," *arXiv preprint arXiv:1012.2086*, 2010.

[46] M. Asadi, R. P. Torghabeh, and N. P. Santhanam, "Estimation in slow mixing, long memory channels," *arXiv preprint arXiv:1301.6798*, 2013.

[47] L. Wilhelmsson and L. B. Milstein, "On the effect of imperfect interleaving for the Gilbert-Elliott channel," *IEEE Transactions on Communications*, vol. 47, no. 5, pp. 681–688, May 1999.

[48] S. K. Das, S. K. Sen, and R. Jayaram, "Call admission and control for quality-of-service provisioning in cellular networks," in *Proc. IEEE International Conf. on Universal Personal Commun.*, San Diego, CA, 1997, pp. 109–113.

[49] S. H. Kang and D. K. Sung, "Two-state MMPP modeling of ATM superposed traffic streams based on the characterization of correlated interarrival times," in *Proc. IEEE Global Telecom. Conf.*, vol. 2, Singapore, 1995, pp. 1422–1426.

[50] H. Lee and D. H. Cho, "Capacity improvement and analysis of VoIP service in a cognitive radio system," *IEEE Trans. Vehicular Technology*, vol. 59, no. 4, pp. 1646–1651, 2010.

[51] K. Sriram and W. Whitt, "Characterizing superposition arrival processes in packet multiplexers for voice and data," *IEEE J. Select. Areas Commun.*, vol. 4, no. 6, pp. 833–846, 1986.

[52] H. Heffes and D. Lucantoni, "A markov modulated characterization of packetized voice and data traffic and related statistical multiplexer performance," *IEEE J. Select. Areas Commun.*, vol. 4, no. 6, pp. 856–868, 1986.

[53] R. Fantacci, "Queuing analysis of the selective repeat automatic repeat request protocol wireless packet networks," *IEEE Trans. Vehicular Technology*, vol. 45, no. 2, pp. 258–264, 1996.

[54] L. Lugand, D. Costello, and R. Deng, "Parity retransmission hybrid ARQ using rate 1/2 convolutional codes on a nonstationary channel," *IEEE Trans. Commun.*, vol. 37, no. 7, pp. 755–765, 1989.

[55] A. Riska and E. Smirni, "Exact aggregate solutions for M/G/1-type Markov processes," in *Proc. of the ACM SIGMETRICS Int. Conf. on Meas. and Modeling of Comp. Syst.*, Marina Del Rey, CA, 2002, pp. 86–96.

[56] S. Karlin and H. Taylor, *A first course in stochastic processes.* New York, NY: Academic Press, 1975.

[57] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, ser. Wiley Series in Telecommunications. Wiley, 1991.

[58] S. B. Wicker, *Error Control Systems for Digital Communication and Storage.* Englewood Cliffs, NJ: Prentice Hall, 1995.

[59] D. Forney, "Exponential error bounds for erasure, list, and decision feedback schemes," *IEEE Trans. Inform. Theory*, vol. 14, no. 2, pp. 206–220, 1968.

[60] E. Hof, I. Sason, and S. Shamai, "Performance bounds for erasure, list, and decision feedback schemes with linear block codes," *IEEE Trans. Inform. Theory*, vol. 56, no. 8, pp. 3754–3778, 2010.

[61] M. G. Kim and J. H. Lee, "Undetected error probabilities of binary primitive BCH codes for both error correction and detection," *IEEE Trans. Commun.*, vol. 44, no. 5, pp. 575–580, 1996.

[62] T. Kasami, "Weight distributions of Bose-Chaudhuri-Hocquenghem codes," in *Proc. Conf. on Combinatorial Mathematics and Its Appl.*, Chapel Hill, NC, 1969, pp. 335–357.

[63] S. Ahson and M. Ilyas, *VoIP handbook: Applications, technologies, reliability, and security*. Boca Raton, Fl: Taylor & Francis, 2008.

[64] N. Bhushan, C. Lott, P. Black, R. Attar, Y. C. Jou, M. Fan, D. Ghosh, and J. Au, "CDMA2000 1x EV-DO revision A: A physical layer and MAC layer overview," *IEEE Commun. Magazine*, vol. 44, no. 2, pp. 37–49, 2006.

[65] J. R. Norris, *Markov chains*. Cambridge, MA: Cambridge University Press, 1998, no. 2008.

[66] B. Sericola, "Occupation times in Markov processes," *Stochastic Models*, vol. 16, no. 5, pp. 479–510, 2000.

[67] M. Bladt, B. Meini, M. F. Neutsl, and B. Sericola, "Distributions of reward functions on continuous-time Markov chains," *Matrix-analytic methods: theory and application*, pp. 39–62, 2002.

[68] W. Rudin, *Principles of Mathematical Analysis (International Series in Pure & Applied Mathematics)*. New York, NY: McGraw-Hill, 1976.

[69] S. I. Resnick, *A probability path*. Boston, MA: Birkhäuser, 1999.

[70] P. Billingsley, *Convergence of probability measures*, ser. Wiley Series in Probability and Mathematical Statistics. New York, NY: Wiley, 1968.

[71] S. M. Ross, *Stochastic Processes*. New York, NY: John Wiley and Sons, 1983.

[72] P. Bremaud, *Markov chains: Gibbs fields, Monte Carlo simulation, and queues*. New York, NY: Springer, 1999, vol. 31.

[73] K. R. Gabriel, "The distribution of the number of successes in a sequence of dependent trials," *Biometrika*, vol. 46, no. 3, pp. 454–460, 1959.

[74] P. J. Pedler, "Occupation times for two-state Markov chains," *Journal of Applied Probability*, pp. 381–390, 1971.

[75] Y. Kovchegov, N. Meredith, and E. Nir, "Occupation times and bessel densities," *Statistics & Probability Letters*, vol. 80, no. 2, pp. 104–110, 2010.

[76] F. Hamidi-Sepehr, J. F. Chamberland, and H. D. Pfister, "Delay-sensitive communication over fading channel: queueing behavior and code parameter selection," *arXiv preprint arXiv:1309.3307*, 2013.

[77] M. Shaked and J. G. Shanthikumar, *Stochastic Orders and Their Applications*. New York, NY: Springer, 2006.

[78] D. Stoyan, *Comparison methods for queues and other stochastic models*. New York, NY: John Wiley & Sons, 1983.

APPENDIX A

PROOF OF LEMMA 16

Distributions of the occupancy times for two-state discrete-time and continuous-time Markov chains have been studied previously. These distributions can be derived using bivariate generating functions and two-dimensional Laplace transforms, respectively [74]. Herein, we show how to adapt these approaches to derive the conditional distributions needed in our work.

The matrix of two-dimensional Laplace transforms for the distribution of the time spent in the first state over the time interval $[0, 1]$ is given by

$$\left[-\left(\mathbf{Q} - \begin{bmatrix} \theta & 0 \\ 0 & 0 \end{bmatrix} - \phi \mathbf{I} \right) \right]^{-1}.$$

For example, the first entry in the matrix is equal to

$$\frac{1}{u} + \frac{\mu\xi}{u(uv - \mu\xi)}$$

where $u = \phi + \theta + \mu$ and $v = \phi + \xi$. The inverse two-dimensional Laplace transform of this entry gives the conditional distribution $f_{\eta_1, S_{\mathrm{f}}|S_{\mathrm{i}}}(\cdot, 1|1)$. After this step, Lemma 2 in [74] can be employed to get the desired format in terms of modified Bessel functions. These are the expressions presented in Lemma 16.

APPENDIX B

PROOF OF LEMMA 17

Let a and b be the numbers of transitions into and out of the initial state, respectively. Then, we can write $c = a + b$ to denote the total number of transitions that occur up to time N. From [74], we gather that

$$P_{N_1|S_0}(m|1) = (1-\alpha)^m(1-\beta)^{N-m} \times \sum_{c=0}^{c_1} \binom{m}{a}\binom{N-m-1}{b-1}\left(\frac{\alpha}{1-\beta}\right)^b\left(\frac{\beta}{1-\alpha}\right)^a$$

where

$$c_1 = \begin{cases} N + \frac{1}{2} - \left|2m - \frac{1}{2} + N\right|, & m < N \\ 0, & m = N. \end{cases}$$

We can split the summation into two parts, one for odd and one for even values of c. If $c = 2k$, then $a = b = k$, and the corresponding sum represents $P_{N_1,S_N|S_0}(m, 1|1)$. If $c = 2k+1$, then $a = k$, $b = k+1$, and the resulting sum is $P_{N_1,S_N|S_0}(m, 2|1)$. As such, we can write

$$\begin{aligned} P_{N_1|S_0}(m|1) &= (1-\alpha)^m(1-\beta)^{N-m} \\ &\times \left(\sum_{k=0}^{\infty} \binom{m}{k}\binom{N-m-1}{k-1}\left(\frac{\alpha}{1-\beta}\right)^k\left(\frac{\beta}{1-\alpha}\right)^k \right. \\ &\left. + \sum_{k=0}^{\infty} \binom{m}{k}\binom{N-m-1}{k}\left(\frac{\alpha}{1-\beta}\right)^{k+1}\left(\frac{\beta}{1-\alpha}\right)^k \right). \end{aligned}$$

We can set the upper and lower limits on k to 0 and ∞, respectively, since all other terms are automatically zero. From the definition of ${}_2F_1(-N+m+1, -m; 1; \lambda)$,

we see that

$$\left(\frac{\alpha}{1-\beta}\right)\sum_{k=0}^{\infty}\binom{m}{k}\binom{N-m-1}{k}\left(\frac{\alpha}{1-\beta}\right)^{k}\left(\frac{\beta}{1-\alpha}\right)^{k}$$
$$=\left(\frac{\alpha}{1-\beta}\right){}_2F_1(-N+m+1,-m;1;\lambda).$$

Collecting these results, we obtain

$$P_{N_1,S_N|S_0}(m,2|1)=(1-\alpha)^m(1-\beta)^{N-m}\left(\frac{\alpha}{1-\beta}\right)$$
$$\times\,{}_2F_1(-N+m+1,-m;1;\lambda)$$

for $m=1,2,\ldots,N-1$. Clearly, for $m=0$ and $m=N$, this conditional probability is equal to zero. Leveraging [74] and observing that

$$P_{N_1|S_0}(m|1)=P_{N_1,S_N|S_0}(m,1|1)+P_{N_1,S_N|S_0}(m,2|1),$$

we can write the simplified equation

$$P_{N_1,S_N|S_0}(m,1|1)=(1-\alpha)^m(1-\beta)^{N-m}$$
$$\times({}_2F_1(-N+m,-m;1;\lambda)-{}_2F_1(-N+m+1,-m;1;\lambda))$$

for $m=1,2,\ldots,N-1$. Moreover, $P_{N_1,S_N|S_0}(0,1|1)=0$ and $P_{N_1,S_N|S_0}(N,1|1)=(1-\alpha)^N$. The remaining conditional probabilities can be derived in the same manner.