On the Performance of Block Codes over Finite-State Channels in the Rare-Transition Regime

Fatemeh Hamidi-Sepehr, Jean-Francois Chamberland, *Senior Member, IEEE*,
Henry D. Pfister, *Senior Member, IEEE*

***Abstract*—Contemporary wireless networks are tasked with supporting different connection profiles, including real-time traffic and delay-sensitive communications. This creates a need to better understand the fundamental limits of forward error correction in non-asymptotic regimes. This article characterizes the performance of block codes over finite-state channels and evaluates their queueing performance under maximum-likelihood decoding. Classical results from digital communications are revisited in the context of channels with rare transitions, and bounds on the probabilities of decoding failure are derived for random codes. This creates an analysis framework where channel dependencies within and across codewords are preserved. These results are subsequently integrated into a queueing problem formulation.**

***Index Terms*—Block codes, Communication systems, Data communication, Markov processes, Queuing analysis.**

I. Introduction

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 continues to grow rapidly. 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. An important aspect of delay-sensitive traffic stems from the fact that intrinsic delivery requirements may preclude the use of very long codewords. As such, the insights offered by classical information theory and based on Shannon capacity are of limited value in this context. A primary goal of this article is to develop a better understanding of delay-constrained communication, queue-based performance criteria, and service dependencies attributable to channel memory. In particular, we seek to derive performance limits for delay-aware systems operating over channels with memory.

Computing probabilities of decoding failure for specific communication channels and coding schemes is of great interest. This line of work dates back to the early days of information theory [1, p. 233] and it has received considerable attention. One approach that has been successful consists of deriving exponential error bounds on the behavior of asymptotically long codewords. This approach was popularized by Feinstein, Shannon, and Gallager [2], [3], [4]. Such bounding techniques have been applied to both memoryless channels and finite-state channels with memory. However, for finite-state channels where the receiver does not have perfect channel-state information and the transmitter does not get complete feedback of the received sequence, there has been little progress towards a computable expression for the error exponent since [5], [4]. Practically, the Monte Carlo computation of the capacity [6], [7], [8], [9] and, more recently, channel dispersion [10] have had the most impact. It is worth mentioning [11], which requires partial feedback, and [12], which assumes vanishing rates. In general, exponential error bounds are quite accurate for long, yet finite block lengths. The interplay between error probability, rate and blocklength has also gathered attention in papers focusing on the error exponents regime, the normal approximation regime, and the moderate deviations regime [13], [14], [15], [16]. This renewed interest in the performance of coded transmissions points to the timeliness of the topic under investigation [17], [18], [19], [20], [21], [22].

A distinguishing feature of the approach that we wish to develop is the focus on indecomposable channels with memory and state-dependent operation. In many established 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 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. In practice, the tradeoff between performance and delay often encourages system designers to choose block lengths of the

This material is based upon work supported by the National Science Foundation (NSF) under Grants No. 0747363 and No. 0830696. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. This paper was presented in part at the 50th Annual Allerton Conference on Communication, Control, and Computing; and the 46th Annual Conference on Information Sciences and Systems.

This research was conducted while the authors were with the Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX 77843, USA. F. Hamidi-Sepehr is now with Qualcomm Inc. in Santa Clara, CA (f.hamidisepehr@gmail.com). H. D. Pfister is with the Department of Electrical and Computer Engineering, Duke University, Durham, NC 27708, USA (henry.pfister@duke.edu). J.-F. Chamberland is with Texas A&M University (chmbrlnd@tamu.edu).

same order as the channel mixing time.

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 [23], [24]. In this article, we are interested in regimes where the block length is of the same order or smaller than the coherence time of the channel. Informally, 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. Such asymptotic regimes, known as rare-transition or slow-mixing regimes, can be achieved by slowing down the profile of the underlying channel as the block length of the code grows unbounded.

We limit our attention to the case where the transition rate of the finite-state fading channel is slowed so that the expected number of state transitions is finite in each codeword. This model 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 failure. It also differs from block-fading models where the evolution of the channel is independent from block to block. Our proposed framework accounts for scenarios where decoding events are correlated over time.

Rare-transition regimes have been studied in other contexts such as channel estimation, asymptotic filtering, and entropy rate analysis of hidden Markov processes [25], [26], [27], [28]. Herein, we examine probabilities of decoding failure, their distributions and temporal characteristics within the context of rare transitions. 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.

Specifically, this article proposes a methodology to derive 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 necessary mathematical machinery is then provided to develop approximate upper bounds which considerably increase computational efficiency. Numerical results are provided to show that the price to pay for this efficiency is a small loss in accuracy. The ensuing results are compared to the exact probabilities of decoding failure obtained for a Gilbert-Elliott channel under a minimum distance decoder and a maximum-likelihood decision rule [1], [29], [30].

The analysis conducted under the assumption of channel state information available at the receiver may not match the particular system one is interested in. This points to the complication hidden in computing the error exponent in the absence of state information, which makes the problem intractable. One should notice that in [4], the error exponent is computed for finite-state channels when the channel state is known at the receiver; this is the only known tractable case. Our results will be different in the sense that Gallager considers an ergodic regime for which the equations simplify using an eigenvalue characterization, whereas we are interested in the rare-transition regime.

We believe that, in the rare-transition regime, the bounds derived under the assumption of state information at the receiver, result in the right asymptotic behavior when there is no state information. This is due to the fact that, asymptotically, the receiver has an unbounded number of channel observations to estimate the state between consecutive transitions. This assumption provides a valid simplification for deriving upper bounds in this regime. This intuition is supported by the mathematical results in [26], [27].

The potential implications of this framework are then discussed in terms of queueing theory. 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. We exploit the results of the error-probability analysis in the rare-transition regime to evaluate the queueing performance of communication systems that transmit encoded data over channels with correlated behavior over time.

We introduce a novel methodology that employs the derived upper bounds on the probability of decoding failure to bound the queueing performance of the system. We show how stochastic dominance enables a tractable analysis of overall performance. This is an important contribution of the paper as it allows us to justify the analysis we carry on the queueing performance based on the upper bounds on error probabilities. These results are then compared with a performance characterization based on the exact probability of decoding failure for a Gilbert-Elliott channel.

Specifically, Section IV-C focuses on system models with scalable arrival profiles, which are based on Poisson processes, and the finite-state channels with memory analyzed in earlier sections. These assumptions permit the rigorous comparison of system performance for codes with arbitrary block lengths and code rates. Based on the resulting characterizations, it is possible to select the best code parameters for delay-sensitive applications over various channels based on the error bounds in the rare-transition regime. The methodology introduced herein offers a new perspective on the joint queueing-coding analysis of finite-state channels with memory, and it is supported by numerical simulations.

II. Modeling and Exponential Bounds

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

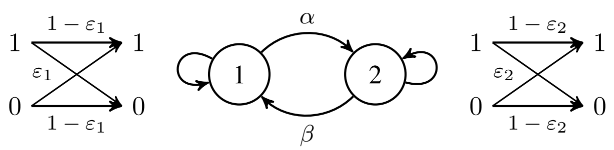

Fig. 1. The Gilbert-Elliott model is the simplest, non-trivial instantiation of a finite-state channel with memory. State evolution over time forms a Markov chain and the input-output relationship is governed by a state-dependent crossover probability.

been used extensively in the literature. Adopting classical notation [4], 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 $P_{\cdot|\cdot}(\cdot|\cdot)$ to denote conditional joint probability mass function, and $P_{\mathrm{e},\cdot|\cdot}(\cdot|\cdot)$ to denote conditional joint probability of decoding error.

The conditional probability distribution governing a finite-state channel can then be written as

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

When state transitions are independent of input symbols, this expression admits the factorization

$$\begin{aligned} &P_{Y_n,S_n|X_n,S_{n-1}}(y_n,s_n|x_n,s_{n-1}) \\ &= P_{S_n|S_{n-1}}(s_n|s_{n-1})P_{Y_n|X_n,S_{n-1}}(y_n|x_n,s_{n-1}). \end{aligned} \tag{2}$$

Throughout, we assume that channel statistics are homogeneous over time and the sequence $\{S_n\}$ forms a Markov chain.

The Gilbert-Elliott model is an example of a channel that possesses the structure described above [29], [30]. This model is governed by a two-state Markov chain, as illustrated in Fig. 1. The state transition probability matrix for the Gilbert-Elliott channel can be expressed as

$$\mathbf{P} = \begin{bmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{bmatrix}, \tag{3}$$

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 crossover probability ε_s, where

$$\begin{aligned} \Pr(x_n = y_n|S_{n-1}=s) &= 1-\varepsilon_s \\ \Pr(x_n \neq y_n|S_{n-1}=s) &= \varepsilon_s. \end{aligned} \tag{4}$$

We adopt a random coding scheme that employs a code ensemble $\mathcal{C}$ with $M = \lceil e^{NR} \rceil$ elements [4]. Variable N denotes the block length of the code and R is the code rate in nats per code bit. Every element in $\mathcal{C}$ corresponds to a sequence of channel inputs. The input sequence associated with the kth codeword, which we denote by $\mathbf{X}(k)$, is determined through the following procedure. Suppose that $Q(\cdot)$ is a distribution on the set of 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 $\mathbf{Q}_N$, and every codeword is selected independently from other elements in $\mathcal{C}$. A message is sent to the destination by first selecting one of the codewords, and then sequentially transmitting its entries over the communication channel.

To begin our examination, we provide an extension to Theorem 5.6.1 in [4, p. 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 1: Suppose that the realization of the channel state 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

$$P_{\mathrm{e}|\mathbf{S}}(\mathbf{s}) \leq e^{-N(E_{0,N}(\rho,\mathbf{Q}_N,\mathbf{s})-\rho R)} \tag{5}$$

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

$$\frac{-1}{N}\ln\prod_{n=1}^{N}\sum_{y_n}\left[\sum_{x_n}Q(x_n)P_{Y_n|X_n,S_{n-1}}(y_n|x_n,s_{n-1})^{\frac{1}{1+\rho}}\right]^{1+\rho}. \tag{6}$$

Proof: The condition $\mathbf{S} = \mathbf{s}$ alters the probability measure governing the input-output relationship of the channel. Applying Theorem 5.6.1 in [4] with $M = \lceil e^{NR} \rceil$, we get

$$P_{\mathrm{e}|\mathbf{S}}(\mathbf{s}) \leq e^{\rho NR}\sum_{\mathbf{y}}\left[\sum_{\mathbf{x}}\mathbf{Q}_N(\mathbf{x})P_{\mathbf{Y}|\mathbf{X},\mathbf{S}}\left(\mathbf{y}|\mathbf{x},\mathbf{s}\right)^{\frac{1}{1+\rho}}\right]^{1+\rho} \tag{7}$$

where $P_{\mathbf{Y}|\mathbf{X},\mathbf{S}}(\mathbf{y}|\mathbf{x},\mathbf{s})$ denotes the conditional distribution of receiving $\mathbf{y}$ given $\mathbf{X} = \mathbf{x}$ and $\mathbf{S} = \mathbf{s}$. ■

A key insight revealed through the proof of Proposition 1 is that $E_{0,N}(\rho,\mathbf{Q}_N,\mathbf{s})$ only depends on $\mathbf{s}$ through its empirical distribution, designated $\mathcal{T}(\mathbf{s})$.

Corollary 1: 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 we can write $E_{0,N}(\rho,\mathbf{Q}_N,\mathbf{s}) = E_{0,N}(\rho,\mathbf{Q}_N,\mathbf{s}') = E_{0,N}(\rho,\mathbf{Q}_N,T)$ with some abuse of notation. The probability of decoding failure at the destination, conditioned on $\mathbf{S} = \mathbf{s}$, is bounded by $P_{\mathrm{e}|\mathbf{S}}(\mathbf{s}) \leq e^{-N(E_{0,N}(\rho,\mathbf{Q}_N,T)-\rho R)}$ for any sequence $\mathbf{s}$ with empirical distribution $\mathcal{T}(\mathbf{s}) = T$ and $\rho \in [0,1]$.

For the problem at hand, we are interested in probabilities of the form $P_{\mathrm{e},S_N|S_0}(s_N|s_0)$; each of these represents the probability of a decoding failure while keeping track of boundary states. In view of Corollary 1, 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 2: 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

$$\begin{aligned} & P_{\mathrm{e},S_N|S_0}(s_N|s_0) \\ & \leq \sum_{T \in \mathcal{T}} P_{\mathcal{T}(\mathbf{S}),S_N|S_0}(T, s_N|s_0) \min_{\rho \in [0,1]} e^{-N(E_{0,N}(\rho, \mathbf{Q}_N, T) - \rho R)} \\ & \leq \min_{\rho \in [0,1]} \sum_{T \in \mathcal{T}} P_{\mathcal{T}(\mathbf{S}),S_N|S_0}(T, s_N|s_0) e^{-N(E_{0,N}(\rho, \mathbf{Q}_N, T) - \rho R)} \end{aligned} \tag{8}$$

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: This demonstration parallels an argument found in [4, Section 5.9]. By partitioning the set of length-N sequences according to their empirical distributions and applying Proposition 1, we obtain

$$\begin{aligned} & P_{\mathrm{e},S_N|S_0}(s_N|s_0) \\ & = \sum_{T \in \mathcal{T}} \sum_{\mathbf{s}:\mathcal{T}(\mathbf{s})=T} \Pr(\mathbf{S} = \mathbf{s}, S_N = s_N | S_0 = s_0) P_{\mathrm{e}|\mathbf{S}}(\mathbf{s}) \\ & \leq \sum_{T \in \mathcal{T}} \sum_{\mathbf{s}:\mathcal{T}(\mathbf{s})=T} \Pr(\mathbf{S} = \mathbf{s}, S_N = s_N | S_0 = s_0) \\ & \quad \times \min_{\rho \in [0,1]} e^{-N(E_{0,N}(\rho, \mathbf{Q}_N, \mathbf{s}) - \rho R)}. \end{aligned} \tag{9}$$

The first inequality in (8) follows from the fact that $E_{0,N}(\rho, \mathbf{Q}_N, \mathbf{s})$ only depends on $\mathbf{s}$ through its empirical distribution $\mathcal{T}(\mathbf{s})$. The second inequality holds because the sum of minimized summands is upper bounded by the minimum of the sums. ∎

Next, we consider a generalized Gilbert-Elliott type channel, where the cardinality of the channel state space is $\mathcal{S}$. For this class of channels, the uniform input distribution achieves capacity [7, Theorem 5.1]. To handle state information at the receiver, one simply enlarges the channel output alphabet to include the channel state. With (or without) this change, it is easy to verify that the channel satisfies the conditions of the cited theorem. Thus, in the remainder of the paper, we assume that $\mathbf{Q}_N$ is the uniform input distribution.

Suppose that the sequence $\mathbf{s}$ is fixed and recall that $s_n \in \{1, \ldots, \mathcal{S}\}$. Then, by Proposition 1 we obtain

$$\begin{aligned} & E_{0,N}(\rho, \mathbf{Q}_N, \mathbf{s}) \\ & = -\frac{1}{N} \ln \prod_{n=0}^{N-1} \frac{1}{2^\rho} \left(\varepsilon_{s_n}^{\frac{1}{1+\rho}} + (1 - \varepsilon_{s_n})^{\frac{1}{1+\rho}} \right)^{1+\rho} \\ & = -\frac{1}{N} \sum_{i=1}^{\mathcal{S}} n_i \ln \frac{1}{2^\rho} \left(\varepsilon_i^{\frac{1}{1+\rho}} + (1 - \varepsilon_i)^{\frac{1}{1+\rho}} \right)^{1+\rho} \\ & = \sum_{i=1}^{\mathcal{S}} \frac{n_i}{N} b_i(\rho), \end{aligned} \tag{10}$$

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

$$b_i(\rho) = -\ln \frac{1}{2^\rho} \left(\varepsilon_i^{\frac{1}{1+\rho}} + (1 - \varepsilon_i)^{\frac{1}{1+\rho}} \right)^{1+\rho}. \tag{11}$$

For convenience, we assume that the error probabilities in different states are distinct and we order the states such that $\epsilon_1 < \cdots < \epsilon_{\mathcal{S}} \leq 0.5$, which implies $b_1(\rho) > \cdots > b_{\mathcal{S}}(\rho)$.

An important observation from (10) is that the upper bound in (8) can be rewritten as an expectation with respect to the distribution of a weighted sum of the state occupancy times. This significantly reduces the complexity of computing (8). Suppose that random variable N_i denotes the number of visits to channel state i over the duration of a codeword, and let

$$W(\rho) = \frac{1}{N} \sum_{i=1}^{\mathcal{S}} N_i b_i(\rho) \tag{12}$$

designate the weighted sum of the normalized occupancy times. Then, the second upper bound on the error probability in (8) can be rewritten as

$$\min_{\rho \in [0,1]} \sum_{w} P_{W(\rho),S_N|S_0}(w, s_N|s_0) e^{-N(w - \rho R)}, \tag{13}$$

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 $S_0 = s_0$.

The exponent in (8) is averaged with respect to the joint distribution of channel state occupations $P_{\mathcal{T}(\mathbf{S}),S_N|S_0}(T, s_N|s_0)$. Still, 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, $W(\rho)$ is a sufficient statistic to compute the upper bound on the probability of decoding error. This structure improves the computational efficiency of the bounding technique, especially when the number of channel states is large.

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

$$\begin{aligned} & P_{\mathrm{e},S_N|S_0}(s_N|s_0) \\ & \leq \min_{\rho \in [0,1]} \sum_{\mathbf{s}} e^{-N(E_{0,N}(\rho, \mathbf{Q}_N, \mathbf{s}) - \rho R)} \\ & \qquad \times \Pr(\mathbf{S} = \mathbf{s}, S_N = s_N | S_0 = s_0) \\ & = \min_{\rho \in [0,1]} e^{\rho N R} \left(\mathbf{e}_{s_0} \begin{bmatrix} a(1,1) & a(1,2) \\ a(2,1) & a(2,2) \end{bmatrix}^N \mathbf{e}_{s_N}^{\mathrm{T}} \right). \end{aligned} \tag{14}$$

In this equation, $\mathbf{e}_i$ represents the unit vector of length two with a one in the ith position. Matrix entries are defined by $a(i,j) = [\mathbf{P}]_{ij} \, e^{-b_i(\rho)}$, where the transition probability matrix $\mathbf{P}$ is given in (3). In deriving (14), we use

$$\begin{aligned} & \mathbf{e}_{s_0} \begin{bmatrix} a(1,1) & a(1,2) \\ a(2,1) & a(2,2) \end{bmatrix}^N \mathbf{e}_{s_N}^{\mathrm{T}} \\ & = \sum_{\mathbf{s}} \prod_{n=1}^{N} \Pr(S_n = s_n | S_{n-1} = s_{n-1}) e^{-b_{s_{n-1}}(\rho)} \\ & = \sum_{\mathbf{s}} \left(\prod_{n=0}^{N-1} \frac{1}{2^\rho} \left(\varepsilon_{s_n}^{\frac{1}{1+\rho}} + (1 - \varepsilon_{s_n})^{\frac{1}{1+\rho}} \right)^{1+\rho} \right) \\ & \qquad \times \prod_{n=1}^{N} \Pr(S_n = s_n | S_{n-1} = s_{n-1}) \\ & = \sum_{\mathbf{s}} e^{-N(E_{0,N}(\rho, \mathbf{Q}_N, \mathbf{s}))} \Pr(\mathbf{S} = \mathbf{s}, S_N = s_N | S_0 = s_0). \end{aligned} \tag{15}$$

Inequality (14) holds for any $\rho \in [0,1]$ and, hence, the bound can be tightened by minimizing over ρ. The bound in (14) is reminiscent of Gallager's exponential bound for finite-state channels [4, Thm. 5.9.3, p. 185] when the receiver has perfect state information. The main difference is that Gallager considers an ergodic regime for which this equation simplifies using an eigenvalue characterization, whereas we are interested in the rare-transition regime.

III. The Rare-Transition Regime

In a traditional setting where $\mathbf{P}$ is kept constant, the upper bound given in (8) can be refined using the Perron-Frobenius theorem [4, pp. 184–185]. For the problem at hand, we study the case where the state transition probabilities decay as $1/N$. This is a rare-transition regime where the average number of state transitions per block converges to a constant. We define a suitable model for this regime through the sampling of a continuous-time Markov chain (CTMC) $\mathcal{X}(\cdot)$ with infinitesimal generator matrix $\mathcal{Q}$.

A. Construction of Discrete-Time Markov Chains

Following conventional notation, we use Ω to designate the sample space and we represent a generic outcome by ω. Whenever necessary, we use superscript ω to refer to a particular realization. For instance, $\mathcal{X}(\cdot)$ denotes the CTMC whereas $\mathcal{X}^{\omega}(\cdot)$ symbolizes the sample path associated with realization ω. In particular, $\mathcal{X}^{\omega}(\cdot)$ defines a mapping from the time interval $[0,\infty)$ to the state space $\{1,\ldots,\mathcal{S}\}$. The need for this notation will become manifest shortly.

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, $\mathcal{X}_N(t) = \mathcal{X}(\lfloor Nt \rfloor/N)$. 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 $\mathcal{Q}$ through the equation [31, Thms. 2.1.1 & 2.1.2]

$$\mathbf{P}_N = \exp(\mathcal{Q}/N). \tag{16}$$

We emphasize that $\mathbf{P}_N$ is also the transition probability matrix of the Markov chain $\mathcal{X}(t)$ for a time interval of length $1/N$, i.e., $[\mathbf{P}_N]_{ij} = \Pr(\mathcal{X}(1/N) = j | \mathcal{X}(0) = i)$. As before, we can distinguish between the sampled chain $\mathcal{X}_N$ and its realization $\mathcal{X}_N^{\omega}$ associated with outcome ω.

We turn to a simple example. Consider a two-state Markov process with infinitesimal generator matrix

$$\mathcal{Q} = \begin{bmatrix} -\mu & \mu \\ \xi & -\xi \end{bmatrix} \qquad \mu, \xi > 0. \tag{17}$$

The transition probability matrix of the sampled process then becomes [32, Ch. 6, p. 261]

$$\mathbf{P}_N = \frac{1}{\mu+\xi} \begin{bmatrix} \xi + \mu e^{-\frac{\xi+\mu}{N}} & \mu\left(1 - e^{-\frac{\xi+\mu}{N}}\right) \\ \xi\left(1 - e^{-\frac{\xi+\mu}{N}}\right) & \mu + \xi e^{-\frac{\xi+\mu}{N}} \end{bmatrix}. \tag{18}$$

As seen above, jumps in the discrete chain become less likely when N increases. This should be 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 analysis. The inequalities presented in Section II 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. In particular, we first 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. We then leverage a numerical procedure to compute the distributions of weighted sums of channel state occupations for the CTMC [33], [34]. Collecting these results, we arrive at the desired characterization of channels with memory.

B. Convergence of Measures

For every channel state i, we define the occupation times pathwise through the integrals below,

$$\eta_i^{\omega} = \int_{[0,1]} \mathbf{1}_{\{\mathcal{X}^{\omega}(t)=i\}} dt \tag{19}$$

$$\eta_{N,i}^{\omega} = \int_{[0,1]} \mathbf{1}_{\{\mathcal{X}_N^{\omega}(t)=i\}} dt = \frac{1}{N} \sum_{k=1}^{N} \mathbf{1}_{\{\mathcal{X}_N^{\omega}(k/N)=i\}} \tag{20}$$

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

Proposition 3: The sequence of random vectors given by $\boldsymbol{\eta}_N = (\eta_{N,1}, \ldots, \eta_{N,\mathcal{S}})$ converges almost surely to random vector $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_{\mathcal{S}})$ as N approaches infinity.

Proof: The CTMC $\mathcal{X}(t)$ is time-homogeneous and its state space has finite cardinality. This process is therefore non-explosive [31, Sec. 2.7], which implies that the number of transitions in the interval $[0,1]$ is finite almost surely. We can then write $\Pr(\Omega') = 1$, where

$$\Omega' = \{\omega \in \Omega | \mathcal{X}^{\omega}(t) \text{ has finitely many jumps in } [0,1]\}.$$

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 [35, p. 323] and, as such,

$$\begin{aligned} \eta_i^{\omega} &= \int_{[0,1]} \mathbf{1}_{\{\mathcal{X}^{\omega}(t)=i\}} dt \\ &= \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} \mathbf{1}_{\{\mathcal{X}_N^{\omega}(k/N)=i\}} = \lim_{N\to\infty} \eta_{N,i}^{\omega}. \end{aligned} \tag{21}$$

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

$$\lim_{N\to\infty} d_1\left(\boldsymbol{\eta}_N^{\omega}, \boldsymbol{\eta}^{\omega}\right) = 0 \qquad \forall \omega \in \Omega' \tag{22}$$

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

$$\Pr\left(\omega \in \Omega \middle| \lim_{N\to\infty} d_1\left(\boldsymbol{\eta}_N^{\omega}, \boldsymbol{\eta}^{\omega}\right) = 0\right) = 1. \tag{23}$$

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

Almost sure convergence implies convergence in probability and in distribution [36, Sec. 8.5]. Thus, from Proposition 3, we gather that $\boldsymbol{\eta}_N$ converges to $\boldsymbol{\eta}$ in distribution, which is sufficient for our purpose. That is, the occupation times of a sequence of independent discrete-time Markov chains, each generated according to $\mathbf{P}_N$, converge in distribution to $\boldsymbol{\eta}$.

In the proof above, we make no explicit mention of the underlying probability law on Ω. This arguments apply to the equilibrium distribution of the Markov chain as well as the conditional measure where the Markov process starts in state $i \in \{1, \ldots, \mathcal{S}\}$ 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 joint probabilities,

$$F(r_1, \ldots, r_{\mathcal{S}-1}) = \Pr\left(\eta_1 \leq r_1, \ldots, \eta_{\mathcal{S}-1} \leq r_{\mathcal{S}-1}\right) \tag{24}$$

$$\begin{aligned} &F_{ij}(r_1, \ldots, r_{\mathcal{S}-1}) \\ &\quad = \Pr\left(\eta_1 \leq r_1, \ldots, \eta_{\mathcal{S}-1} \leq r_{\mathcal{S}-1}, S_{\mathrm{f}} = j, S_{\mathrm{i}} = i\right) \end{aligned} \tag{25}$$

$$\begin{aligned} &F_{j|i}(r_1, \ldots, r_{\mathcal{S}-1}) \\ &\quad = \Pr\left(\eta_1 \leq r_1, \ldots, \eta_{\mathcal{S}-1} \leq r_{\mathcal{S}-1}, S_{\mathrm{f}} = j | S_{\mathrm{i}} = i\right). \end{aligned} \tag{26}$$

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 3 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. We can readily apply the results of Proposition 3 to affine combinations of $\eta_1, \ldots, \eta_{\mathcal{S}}$.

Corollary 2: Let ρ be fixed and recall the coefficients $b_i(\rho)$ found in (11). 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.

The expression for $W(\rho)$ above differs from (12) because 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:

$$G^{(\rho)}(w) = \Pr(W(\rho) \leq w), \tag{27}$$

$$G^{(\rho)}_{ij}(w) = \Pr(W(\rho) \leq w, S_{\mathrm{i}} = i, S_{\mathrm{f}} = j), \tag{28}$$

$$G^{(\rho)}_{j|i}(w) = \Pr(W(\rho) \leq w, S_{\mathrm{f}} = j | S_{\mathrm{i}} = i). \tag{29}$$

Similarly, we write $G^{(\rho)}_N$, $G^{(\rho)}_{N,ij}$ and $G^{(\rho)}_{N,j|i}$ for the positive measures associated with $W_N(\rho)$.

Below, we seek to find approximate bounds for conditional probabilities of decoding failure. Define $g_N(w) = \min\left\{1, e^{-N(w-\rho R)}\right\}$ and note that these functions converge pointwise to $g(w) = \mathbf{1}_{[\rho R, \infty)}(w)$. A preliminary result in our characterization is the following lemma.

Lemma 1: Consider integrals of the functions $\{g_N(w)\}$. This sequence converges to

$$\begin{aligned} &\lim_{N \to \infty} \int_{[b_{\mathcal{S}}(\rho), b_1(\rho)]} \min\left\{1, e^{-N(w-\rho R)}\right\} dG^{(\rho)}_{j|i}(w) \\ &\quad = \int_{[b_{\mathcal{S}}(\rho), b_1(\rho)]} g \, dG^{(\rho)}_{j|i}. \end{aligned} \tag{30}$$

Proof: We note that $|g_N(w)| \leq 1$ for all values of N and w. Since $G^{(\rho)}_{j|i}$ is a bounded measure, then (30) holds by Lebesgue's Dominated Converge Theorem. ■

We derive asymptotic upper bounds next.

Proposition 4: Suppose that a message is transmitted over a generalized Gilbert-Elliott type fading channel with $\mathcal{S}$ states using the random coding scheme of Section II and uniform prior distribution. Given $\epsilon > 0$, there exists N_ϵ such that

$$\begin{aligned} &P_{\mathrm{e}, S_N | S_0}(j|i) \\ &\leq \int_{[b_{\mathcal{S}}(\rho), b_1(\rho)]} \min\left\{1, e^{-N(w-\rho R)}\right\} dG^{(\rho)}_{j|i}(w) + \epsilon \end{aligned} \tag{31}$$

for every $N > N_\epsilon$.

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

$$P_{\mathrm{e}|\mathbf{S}}(T) \leq e^{-N(E_{0,N}(\rho, \mathbf{Q}_N, T) - \rho R)} = e^{-N(w-\rho R)} \tag{32}$$

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. By taking the expectation over $W(\rho)$, we get

$$\begin{aligned} &P_{\mathrm{e}, S_N | S_0}(j|i) \\ &\leq \int_{[b_{\mathcal{S}}(\rho), b_1(\rho)]} \min\left\{1, e^{-N(w-\rho R)}\right\} dG^{(\rho)}_{N,j|i}(w) \\ &= \sum_{w} P_{W_N(\rho), S_N | S_0}(w, j|i) \min\left\{1, e^{-N(w-\rho R)}\right\}. \end{aligned} \tag{33}$$

Below, we show that

$$\lim_{N \to \infty} \int_{[b_{\mathcal{S}}(\rho), b_1(\rho)]} g_N \, dG^{(\rho)}_{N,j|i} = \int_{[b_{\mathcal{S}}(\rho), b_1(\rho)]} g \, dG^{(\rho)}_{j|i}. \tag{34}$$

From Corollary 2, we know that the sequence of measures $\left\{G^{(\rho)}_{N,j|i}\right\}$ converges in distribution to $G^{(\rho)}_{j|i}$. For converging sequence $w_N \in [b_{\mathcal{S}}(\rho), b_1(\rho)]$ with $\lim_{N \to \infty} w_N = w \neq \rho R$, we have $g_N(w_N) \to g(w)$. This is pertinent because the limiting function $G^{(\rho)}_{j|i}(w)$ is continuous at $\rho R \in (b_{\mathcal{S}}(\rho), b_1(\rho))$ and, consequently, the event $\{w = \rho R\}$ has probability zero. Collecting these observations, we can can apply [37, Thm. 5.5] and thereby establish the validity of (34).

In view of this result and Lemma 1, we get

$$\begin{aligned} \lim_{N \to \infty} \int_{[b_{\mathcal{S}}(\rho), b_1(\rho)]} g_N \, dG^{(\rho)}_{N,j|i} &= \int_{[b_{\mathcal{S}}(\rho), b_1(\rho)]} g \, dG^{(\rho)}_{j|i} \\ &= \lim_{N \to \infty} \int_{[b_{\mathcal{S}}(\rho), b_1(\rho)]} g_N \, dG^{(\rho)}_{j|i}. \end{aligned} \tag{35}$$

Thus, for any $\epsilon > 0$, there exists N_ϵ such that

$$\left| \int_{[b_{\mathcal{S}}(\rho), b_1(\rho)]} g_N \, dG^{(\rho)}_{N,j|i} - \int_{[b_{\mathcal{S}}(\rho), b_1(\rho)]} g_N \, dG^{(\rho)}_{j|i} \right| < \epsilon \tag{36}$$

for all $N > N_\epsilon$. Equation (31) follows. ■

In view of Proposition 4, we can write an approximate upper bound for $P_{\mathrm{e}, S_N | S_0}(j|i)$;

$$\begin{aligned} &P_{\mathrm{e}, S_N | S_0}(j|i) \\ &\lesssim \int_{[b_{\mathcal{S}}(\rho), b_1(\rho)]} \min\left\{1, e^{-N(w-\rho R)}\right\} dG^{(\rho)}_{j|i}(w). \end{aligned} \tag{37}$$

This approximation is justified when the code length is large enough.

C. Methodology to Compute the Distribution of $W(\rho)$

Equation (37) provides a computationally efficient way to select system parameters. 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 $W(\rho)$ using an iterative approach. In contrast, the standard procedure associated with (33) entails computing $G^{(\rho)}_{N,j|i}$ explicitly for multiple values of N, a cumbersome task. Indeed, Proposition 5 offers a numerical method to compute the distribution of $W(\rho)$. This method is adapted from [33], [34] and, as such, it is presented without a 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 $\mathcal{A} = \mathbf{I} + \mathcal{Q}/\sigma$, where $\mathbf{I}$ is the identity matrix, $\sigma \geq \max_k |\mathcal{Q}_{kk}|$, $k \in \{1, \ldots, \mathcal{S}\}$, is a constant and $\{\mathcal{Q}_{kk}\}$ are the diagonal elements of $\mathcal{Q}$ defined earlier in this section. Also, define matrix $\mathbf{G}(w)$ by $[\mathbf{G}(w)]_{ij} = G^{(\rho)}_{j|i}(w)$ where $i, j \in \{1, \ldots, \mathcal{S}\}$.

Proposition 5: 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

$$\mathbf{G}(w) = \sum_{n=0}^{\infty} e^{-\sigma} \frac{\sigma^n}{n!} \sum_{l=0}^{n} \binom{n}{l} w_k^l (1-w_k)^{n-l} \mathbf{C}^{(k)}(n,l) \quad (38)$$

where $w_k = \frac{w - b_k(\rho)}{b_{k-1}(\rho) - b_k(\rho)}$. Matrices $\left\{\mathbf{C}^{(k)}(n,l)\right\}$ are specified by $\left[\mathbf{C}^{(k)}(n,l)\right]_{cd} = C^{(k)}_{cd}(n,l)$ with $c, d \in \{1, \ldots, \mathcal{S}\}$, and individual entries in each of these matrices are given by the following two recurrence relations. For $k \leq c \leq \mathcal{S}$ and $1 \leq d \leq \mathcal{S}$,

$$\begin{aligned} C^{(k)}_{cd}(n,l) &= \frac{b_k(\rho) - b_c(\rho)}{b_{k-1}(\rho) - b_c(\rho)} C^{(k)}_{cd}(n, l-1) \\ &+ \frac{b_{k-1}(\rho) - b_k(\rho)}{b_{k-1}(\rho) - b_c(\rho)} \sum_{e=1}^{\mathcal{S}} [\mathcal{A}]_{ce} C^{(k)}_{ed}(n-1, l-1) \end{aligned} \quad (39)$$

where $1 \leq l \leq n$. For $n \geq 0$ and $k > 1$, we apply the boundary conditions $C^{(1)}_{cd}(n,0) = 0$ and $C^{(k)}_{cd}(n,0) = C^{(k-1)}_{cd}(n,n)$. Similarly, for $1 \leq c \leq k-1$ and $1 \leq d \leq \mathcal{S}$,

$$\begin{aligned} C^{(k)}_{cd}(n,l) &= \frac{b_c(\rho) - b_{k-1}(\rho)}{b_c(\rho) - b_k(\rho)} C^{(k)}_{cd}(n, l+1) \\ &+ \frac{b_{k-1}(\rho) - b_k(\rho)}{b_k(\rho) - b_c(\rho)} \sum_{e=1}^{\mathcal{S}} [\mathcal{A}]_{ce} C^{(k)}_{ed}(n-1, l) \end{aligned} \quad (40)$$

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

Sketch of proof: We reiterate that this methodology is adapted from a general technique found in [33], [34]. In paralleling the argument presented therein, the weights $b_1(\rho), \ldots, b_{\mathcal{S}}(\rho)$ play the role of reward rates and $W(\rho)$ represents the total continuous reward over the interval $[0, 1)$. The possible discontinuities in $G^{(\rho)}_{j|i}(w)$ have to do with the non-vanishing probabilities that the chain does not visit certain states during time interval $[0, 1)$. ■

The discrete-time Markov chain whose probability transition matrix is given by $\mathcal{A}$ is called a uniformized chain [38], [39]. 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 $\mathcal{X}(\cdot)$ and, as such, it possesses the same probability distribution [33]. Furthermore, the matrix $\mathbf{P}_N$ in (16) can be written as $\mathbf{P}_N = e^{-\frac{\sigma}{N}\mathbf{I}} \exp\left(\frac{\sigma \mathcal{A}}{N}\right)$ or, alternatively,

$$[\mathbf{P}_N]_{ij} = e^{-\frac{\sigma}{N}\mathbf{I}} \sum_{k=0}^{\infty} \frac{1}{k!} \left(\frac{\sigma}{N}\right)^k \left[\mathcal{A}^k\right]_{ij}. \quad (41)$$

While the uniformized chain and the sampled chain are both derived from $\mathcal{Q}$, there remains an important distinction. In the process of constructing the uniformized chain, the corresponding continuous-time Markov chain is scaled by the fastest transition rate σ, so that transitions occur at the same rate irrespective of state. In other words, the transition rate out of every state is increased to σ, and a transition from state k is a *dummy* transition to itself (or self-jump) with probability $1 - |\mathcal{Q}_{kk}|/\sigma$. The uniformized chain accounts for all the transitions associated with a CTMC; whereas the sampled chain with its periodic structure can overlook jumps associated with fast transitions. This makes the uniformized chain a suitable object in describing Proposition 5.

D. Study of Two-State Case

As a special case of Proposition 5, we turn to the situation where $\mathcal{S} = 2$. Occupation times for this simple scenario have been studied in the past, and explicit expressions for their distributions exist [40], [41], [42]. Yet, the distributions provided therein only account for an initial state, and they do not specify a final state.

For a codeword of length N, the channel inputs are $\mathbf{x} = (x_1, \ldots, x_N)$. The input-output relationship is then governed by $P_{Y_n|X_n,S_{n-1}}(y_n|x_n, s_{n-1})$. The channel states that influence this codeword are $S_0, \ldots, S_{N-1}$. Consequently, the number of visits to the first channel state is $N_1 = \sum_{n=0}^{N-1} \mathbf{1}_{\{S_n=1\}}$. In particular, N_1 includes state S_0, the first state in the sequence. This is different from the model considered in [40], [41] where one conditions on the state before the sequence starts. In particular, the sequence excludes the initial state, while the distributions are derived conditioned on the initial state. We must therefore modify existing results slightly to match our current needs.

Lemma 2: Consider a continuous-time Markov chain whose generator matrix is given by (17). The joint distributions governing occupation times and the final state, given an initial state, can be written as follows. For any Lebesgue measurable subset $\mathcal{I}$ of $[0, 1]$, we have

$$\begin{aligned} F_{\eta_1, S_{\mathrm{f}}|S_{\mathrm{i}}}(\mathcal{I}, 1|1) &= \mathbf{1}_{\{1 \in \mathcal{I}\}} e^{-\mu} + \int_{\mathcal{I}} e^{-\mu r - \xi(1-r)} \\ &\times \sqrt{\frac{\mu \xi r}{1-r}} I_1\left(2\sqrt{\mu \xi r(1-r)}\right) dr \end{aligned} \quad (42)$$

$$F_{\eta_1,S_\mathrm{f}|S_\mathrm{i}}(\mathcal{I},2|1) = \int_{\mathcal{I}} \mu e^{-\mu r-\xi(1-r)} I_0\left(2\sqrt{\mu\xi r(1-r)}\right) dr \quad (43)$$

$$F_{\eta_1,S_\mathrm{f}|S_\mathrm{i}}(\mathcal{I},1|2) = \int_{\mathcal{I}} \xi e^{-\mu r-\xi(1-r)} I_0\left(2\sqrt{\mu\xi r(1-r)}\right) dr \quad (44)$$

$$F_{\eta_1,S_\mathrm{f}|S_\mathrm{i}}(\mathcal{I},2|2) = \mathbf{1}_{\{0\in\mathcal{I}\}} e^{-\xi} + \int_{\mathcal{I}} e^{-\mu r-\xi(1-r)} \times \sqrt{\frac{\mu\xi(1-r)}{r}} I_1\left(2\sqrt{\mu\xi r(1-r)}\right) dr \quad (45)$$

where $I_0(\cdot)$ and $I_1(\cdot)$ represent modified Bessel functions of the first kind [41, Lem. 2, p. 386].

Proof: See appendix. ■

The corresponding expressions for the sampled chain are presented in the following Lemma. In reporting these results, we use the Gaussian hypergeometric function, which is defined by [41, Lem. 1, p. 383]

$$_2F_1(\tilde{m},\tilde{n};\tilde{o};\psi) = \sum_{k=0}^{\infty} \frac{(\tilde{m})_k(\tilde{n})_k}{(\tilde{o})_n} \frac{\psi^k}{k!} \quad (46)$$

where $(\tilde{m})_k$ is the rising Pochhammer symbol: $(\tilde{m})_0 = 1$ and $(\tilde{m})_k = \tilde{m}(\tilde{m}+1)\cdots(\tilde{m}+k-1)$ for $k > 0$.

Lemma 3: Consider a two-state channel whose transition probability matrix is given by (3). 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. These distributions for $m = 1, \ldots, N-1$ are

$$P_{N_1,S_N|S_0}(m,1|1) = (1-\alpha)^m(1-\beta)^{N-m} \times \big({}_2F_1(-N+m,-m+1;1;\psi) - {}_2F_1(-N+m+1,-m+1;1;\psi)\big) \quad (47)$$

$$P_{N_1,S_N|S_0}(m,2|1) = \alpha(1-\alpha)^{m-1}(1-\beta)^{N-m} \times {}_2F_1(-N+m+1,-m+1;1;\psi) \quad (48)$$

$$P_{N_1,S_N|S_0}(m,1|2) = (1-\alpha)^m\beta(1-\beta)^{N-m-1} \times {}_2F_1(-N+m+1,-m+1;1;\psi) \quad (49)$$

$$P_{N_1,S_N|S_0}(m,2|2) = (1-\alpha)^m(1-\beta)^{N-m} \times \big({}_2F_1(-N+m+1,-m;1;\psi) - {}_2F_1(-N+m+1,-m+1;1;\psi)\big) \quad (50)$$

where N_1 is the number of visits to the first state, and $\psi = \frac{\alpha\beta}{(1-\alpha)(1-\beta)}$. Special consideration must be given to extremal cases: $P_{N_1,S_N|S_0}(0,\cdot|1) = P_{N_1,S_N|S_0}(N,\cdot|2) = 0$, $P_{N_1,S_N|S_0}(0,2|2) = (1-\beta)^N$, and $P_{N_1,S_N|S_0}(N,1|1) = (1-\alpha)^N$.

Proof: See appendix. ■

We can relate these two results through Proposition 3 and the definition of $\mathbf{P}_N$ in (18). Let α and β be the constants defined in Lemma 3, and consider the assignment

$$\alpha = \frac{\mu}{\mu+\xi}\left(1-e^{-\frac{\xi+\mu}{N}}\right) \quad \beta = \frac{\xi}{\mu+\xi}\left(1-e^{-\frac{\xi+\mu}{N}}\right). \quad (51)$$

Then, the discrete distributions specified above converge to the occupation times described in Lemma 2, under the asymptotic scaling $N_1/N \to \eta$ as N grows unbounded. Two of the limiting measures in Lemma 2 are not absolutely continuous with respect to the Lebesgue measure; nevertheless, these distributions are well-defined positive measures [36, Chap. 8].

Using the limiting distributions in Lemma 2, one can compute the approximate upper bound found in (37) 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, one can compute the approximate bound using the empirical distribution of the state occupancies,

$$P_{\mathrm{e},S_N|S_0}(j|i) \lesssim \int_{[0,1]} \min\left\{1, e^{-N((b_1(\rho)-b_2(\rho))r+b_2(\rho)-\rho R)}\right\} \times dF_{\eta,S_\mathrm{f}|S_\mathrm{i}}(r,j|i). \quad (52)$$

As the number of states increases, dealing with the joint distribution of the state occupation times and integrating over multiple variables is an increasingly intricate 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 4. 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 (33) using the distributions given by Lemma 3, and compare it to the approximate bound presented in (52). Numerical results for this comparison are presented in Section V, offering supporting evidence for our proposed methodology. We note that, even for the simple two-state case, computing the approximate upper bound in (52) is considerably more efficient than calculating (33). As we will see in Section V, the price to pay for this computational efficiency is a small loss in 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 if needed, 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.

IV. Error Analysis and Queue Characterization

The primary goal of this section is to link the error analysis we have developed so far, to the queueing characterization of delay-sensitive communication systems. This gives rise to a unified framework, which enables a joint assessment of the coding and queueing performance of the system in the rare-transition regime. Applications of stochastic dominance are common in queueing theory. This paper presents a novel methodology which connects the coding bounds and queueing performance through an application of stochastic dominance. We note that developing upper bounds on error probability that capture the dependencies on the initial and final channel states is crucial to enable the present queueing analysis. Besides, the computational efficiency of the developed bounds results in faster performance evaluation of the queueing model.

In our framework, the system dynamics and fading process, including channel correlation, are treated from a symbol-level perspective, while a rigorous analysis is conducted at the block level. That is, the block-level behavior is induced from the channel parameters in a consistent fashion. This enables the fair comparison of the queueing performance of communication systems with different block lengths and code rates, and then finding the best coding parameters.

One of the challenges in dealing with block codes over finite-state channels with memory is the statistical dependence between decoding events that are closely spaced in time. 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 augment the state space by appending the value of the channel at the onset of a codeword to the queue length. Under this state augmentation, the coded system retains the Markov property, which facilitates the characterization of the queueing behavior at the transmitter.

We review the necessary mathematical machinery to handle error events for the finite-state channels with memory, as originally introduced by Gilbert [29] and Elliott [30]. We use these models to assess how channel dependencies over time can affect overall performance. A significant benefit in dealing with the Gilbert-Elliott channel model is its tractability. The remainder of this article is devoted to the analysis of error probability and the queueing behavior of systems built around the two-state channel. Sections IV-A and IV-B are dedicated to exact derivation of 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. In Section IV-C, we provide performance analysis of the system from the queueing point of view. In Section IV-D, we show how to bound queueing performance by using upper bounds on the probability of decoding failure, instead of the exact expressions.

A. Exact Probabilities of Decoding Failure

It is possible to compute exact probabilities of decoding failure under various decision schemes for the two-state Gilbert-Elliott channel. Consequently, in this case, we can assess how close the bounds and the true probabilities of error are from one another. It may be impractical to compute exact probabilities of error for more elaborate channels. Even for Gilbert-Elliott 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 may be inevitable.

In [43], the authors study data transmission over a Gilbert-Elliott channel using random coding. 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,

$$\begin{aligned} P_{\mathrm{e},S_N|S_0}(s_N|s_0) &= \sum_{T\in\mathcal{T}} P_{\mathrm{e}|\mathcal{T}(\mathbf{S})}(T) \\ &\quad \times \Pr(\mathcal{T}(\mathbf{S}) = T, S_N = s_N | S_0 = s_0). \end{aligned} \tag{53}$$

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 3.

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 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 $d_{\mathrm{H}}(\cdot,\cdot)$ is the Hamming distance. The conditional error probabilities can then be written as

$$\begin{aligned} P_{\mathrm{e}|\mathcal{T}(\mathbf{S})}(T) &= \sum_{e_1=0}^{n_1}\sum_{e_2=0}^{n_2} P_{\mathrm{e}|\mathcal{T}(\mathbf{S}),E_1,E_2}(T, e_1, e_2) \\ &\quad \times P_{E_1,E_2|\mathcal{T}(\mathbf{S})}(e_1, e_2|T) \end{aligned} \tag{54}$$

where $n_2 = N - n_1$. Given the channel type, the numbers of errors in the first and second states are independent, $P_{E_1,E_2|\mathcal{T}(\mathbf{S})}(e_1, e_2|T) = P_{E_1|\mathcal{T}(\mathbf{S})}(e_1|T) P_{E_2|\mathcal{T}(\mathbf{S})}(e_2|T)$. Furthermore, E_1 and E_2 have binomial distributions

$$\begin{aligned} P_{E_i|\mathcal{T}(\mathbf{S})}(e_i|T) &= P_{E_i|N_i}(e_i|n_i) \\ &= \binom{n_i}{e_i} \varepsilon_i^{e_i} (1-\varepsilon_i)^{n_i - e_i}. \end{aligned} \tag{55}$$

This mathematical structure leads to the following result.

Theorem 1: 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

$$\begin{aligned} &P_{\mathrm{e}|\mathcal{T}(\mathbf{S}),E_1,E_2}(T, e_1, e_2) \\ &= 1 - \left(1 - 2^{-N} \sum_{(\tilde{e}_1,\tilde{e}_2)\in\mathcal{M}(\gamma e_1 + e_2)} \binom{n_1}{\tilde{e}_1}\binom{n_2}{\tilde{e}_2} \right)^{M-1} \end{aligned} \tag{56}$$

where $\mathcal{M}(d)$ is the set of pairs $(\tilde{e}_1, \tilde{e}_2) \in \{0, \ldots, N\}^2$ that satisfy $\gamma\tilde{e}_1 + \tilde{e}_2 \leq d$. 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 appendix and proof in [43]. ∎

If channel state information is available at the receiver, then random codes paired with a maximum-likelihood decoding rule form a permutation invariant scheme. The performance is then determined by the number of symbol errors in each state within a codeword, and not by their order or locations.

B. Undetected Errors

A serious matter with 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 [44], [45]. This safeguard, in turn, leads to slight modifications to the performance analysis presented above.

1) The Exact Approach: In [43], the authors show that the probability of decoding failure (including detected errors, undetected errors, and ties) is given by the equations (53)–(55) and substituting

$$\begin{aligned} &P_{\mathrm{e}|\mathcal{T}(\mathbf{S}),E_1,E_2}(T,e_1,e_2) \\ &= 1 - \left(1 - 2^{-N} \sum_{(\tilde{e}_1,\tilde{e}_2)\in\mathcal{M}(\gamma e_1 + e_2 + \nu)} \binom{n_1}{\tilde{e}_1}\binom{n_2}{\tilde{e}_2} \right)^{M-1} \end{aligned} \tag{57}$$

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 $P_{\mathrm{ue},S_N|S_0}(s_N|s_0)$. It can be upper bounded by

$$\begin{aligned} \bar{P}_{\mathrm{ue},S_N|S_0}(s_N|s_0) &= \sum_{T\in\mathcal{T}} \bar{P}_{\mathrm{ue}|\mathcal{T}(\mathbf{S})}(T) \\ &\quad \times \Pr(\mathcal{T}(\mathbf{S}) = T, S_N = s_N | S_0 = s_0) \end{aligned} \tag{58}$$

where the bounded component in the summand is given by

$$\begin{aligned} \bar{P}_{\mathrm{ue}|\mathcal{T}(\mathbf{S})}(T) &= \sum_{e_1=0}^{n_1}\sum_{e_2=0}^{n_2} \bar{P}_{\mathrm{ue}|\mathcal{T}(\mathbf{S}),E_1,E_2}(T,e_1,e_2) \\ &\quad \times P_{E_1,E_2|\mathcal{T}(\mathbf{S})}(e_1,e_2|T) \end{aligned} \tag{59}$$

and its associated term is

$$\begin{aligned} &\bar{P}_{\mathrm{ue}|\mathcal{T}(\mathbf{S}),E_1,E_2}(T,e_1,e_2) \\ &= 1 - \left(1 - 2^{-N} \sum_{(\tilde{e}_1,\tilde{e}_2)\in\mathcal{M}(\gamma e_1 + e_2 - \nu)} \binom{n_1}{\tilde{e}_1}\binom{n_2}{\tilde{e}_2} \right)^{M-1}. \end{aligned} \tag{60}$$

Since the probability of undetected error is typically much smaller than that of 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. Additional details for this proof are available in Section C of the appendix.

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

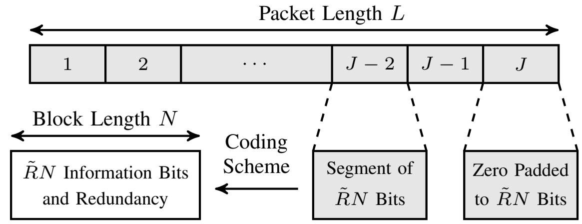

Fig. 2. Each packet is divided into J segments, and a channel encoding scheme is employed to encode each segment.

Lemma 4: 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

$$\begin{aligned} &\tilde{P}_{\mathrm{ue},S_N|S_0}(j|i) \\ &= \min_{0\le\rho\le1} \sum_{n_1=0}^{N} \min\left\{1, e^{-N\left(E_{0,N}(\rho,\mathbf{Q}_N,n_1) - \rho R - \frac{\rho\tau}{1+\rho}\right)}\right\} \\ &\qquad \times P_{N_1,S_N|S_0}(n_1,j|i) \\ &\approx \min_{0\le\rho\le1} \int_{[0,1]} \min\left\{1, e^{-N\left((b_1(\rho)-b_2(\rho))r + b_2(\rho) - \rho R - \frac{\rho\tau}{1+\rho}\right)}\right\} \\ &\qquad \times dF_{\eta,S_{\mathrm{f}}|S_{\mathrm{i}}}(r,j|i) \end{aligned} \tag{61}$$

$$\begin{aligned} &\tilde{P}_{\mathrm{e},S_N|S_0}(j|i) \\ &= \min_{0\le\rho\le1} \sum_{n_1=0}^{N} \min\left\{1, e^{-N\left(E_{0,N}(\rho,\mathbf{Q}_N,n_1) - \rho R + \frac{\tau}{1+\rho}\right)}\right\} \\ &\qquad \times P_{N_1,S_N|S_0}(n_1,j|i) \\ &\approx \min_{0\le\rho\le1} \int_{[0,1]} \min\left\{1, e^{-N\left((b_1(\rho)-b_2(\rho))r + b_2(\rho) - \rho R + \frac{\tau}{1+\rho}\right)}\right\} \\ &\qquad \times dF_{\eta,S_{\mathrm{f}}|S_{\mathrm{i}}}(r,j|i) \end{aligned} \tag{62}$$

where $\tau \ge 0$ 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: One can obtain these expressions by paralleling the approach found in [44], [45]. See appendix. ■

The rare-transition regime, the performance bounds, and the approximation methodologies proposed in this article have a wide range of applications. The next two sections are dedicated to the potential implications of the proposed bounding techniques in terms of queueing theory.

C. Queueing Model

Consider a queueing system in which 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 $\varrho \in (0,1)$. On arrival, a packet is divided into segments of size $RN/\ln 2$ bits, where N denotes the block length and R is the code rate in nats per code bit, as defined in Section II (see Fig. 2). The number of information bits per segment, $RN/\ln 2$, is assumed to be an integer. Since units of rate differ between the exponential upper bounds and conventional coding theory, we use notation

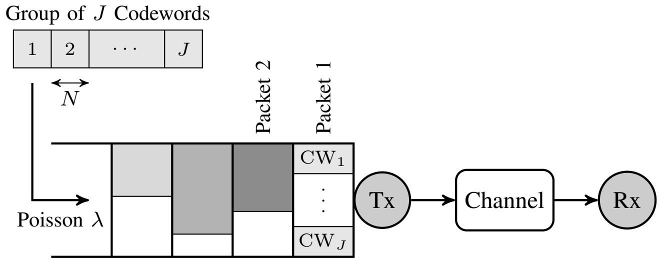

Fig. 3. Coded segments are transmitted over the unreliable communication link. A data packet is discarded from the transmit buffer only when all its codewords are successfully received at the destination.

$\tilde{R} = R/\ln 2$ to convert rates from nats to bits. The total number of segments associated with a packet of length L is given by $J = \left\lceil L/\tilde{R}N \right\rceil$. As before, a random coding scheme is used to protect the data while it is transmitted over the correlated channel.

On the receiving end, one must successfully decode all J codewords to recover the corresponding packet. Once this is achieved, this packet is discarded from the queue (see Fig. 3). We note that random variable J has a geometric distribution with $\Pr(J = j) = (1-\varrho_r)^{j-1}\varrho_r$, where $j \geq 1$ and $\varrho_r = 1-(1-\varrho)^{\tilde{R}N}$. Consequently, the number of coded blocks per data packet possesses the memoryless property, a highly desirable attribute for the purpose of analysis.

This communication system operates on top of the finite-state channel discussed earlier. The resulting framework allows us to rigorously characterize the queueing performance while varying the block length and the code rate. The scaling property of the Poisson packet arrivals with geometric packet lengths is crucial in enabling the fair comparison of systems with different code parameters. In particular, the arrival process is defined at the symbol level and we account for channel 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.

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. The code rate $\tilde{R}$ has a major impact on performance. Generally, a lower code rate will have a smaller probability of decoding failure. However, a lower rate also implies more segments 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

$$\begin{aligned}
&\Pr(U_{s+1} = (d, q_{s+1}) | U_s = (c, q_s)) \\
&= \sum_{n_1=0}^{N} P_{Q_{s+1}|N_1,Q_s}(q_{s+1}|n_1, q_s) \\
&\qquad \times P_{N_1,C_{(s+1)N+1}|C_{sN+1}}(n_1, d|c),
\end{aligned} \tag{63}$$

where the components of the summands are given by Lemma 3 and by

$$\begin{aligned}
&P_{Q_{s+1}|N_1,Q_s}\left(q_{s+1}|n_1, q_s\right) \\
&= \sum_{e_1=0}^{n_1}\sum_{e_2=0}^{n_2} P_{Q_{s+1},E_1,E_2|N_1,Q_s}(q_{s+1}, e_1, e_2|n_1, q_s) \\
&= \sum_{e_1=0}^{n_1}\sum_{e_2=0}^{n_2} P_{Q_{s+1}|E_1,E_2,N_1,Q_s}(q_{s+1}|e_1, e_2, n_1, q_s) \\
&\qquad \times P_{E_1,E_2|N_1,Q_s}(e_1, e_2|n_1, q_s) \\
&= \sum_{e_1=0}^{n_1}\sum_{e_2=0}^{n_2} \binom{n_1}{e_1}\binom{n_2}{e_2} \varepsilon_1^{e_1}(1-\varepsilon_1)^{n_1-e_1}\varepsilon_2^{e_2}(1-\varepsilon_2)^{n_2-e_2} \\
&\qquad \times P_{Q_{s+1}|E_1,E_2,N_1,Q_s}(q_{s+1}|e_1, e_2, n_1, q_s).
\end{aligned} \tag{64}$$

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

$$\begin{aligned}
&P_{Q_{s+1}|E_1,E_2,N_1,Q_s}(q_s+i|e_1, e_2, n_1, q_s) \\
&= a_i P_{\mathrm{e}|E_1,E_2,N_1}(e_1, e_2, n_1) \\
&+ a_{i+1}(1-P_{\mathrm{e}|E_1,E_2,N_1}(e_1, e_2, n_1))\varrho_r \\
&+ a_i(1-P_{\mathrm{e}|E_1,E_2,N_1}(e_1, e_2, n_1))(1-\varrho_r)
\end{aligned} \tag{65}$$

and the probability of the queue decreasing is

$$\begin{aligned}
&P_{Q_{s+1}|E_1,E_2,N_1,Q_s}(q_s-1|e_1, e_2, n_1, q_s) \\
&= a_0\left(1-P_{\mathrm{e}|E_1,E_2,N_1}(e_1, e_2, n_1)\right)\varrho_r.
\end{aligned} \tag{66}$$

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 (57). Variable 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, (65) applies for cases where $i \geq 1$. However, in this case, the queue length cannot decrease and the conditional transition probability for $i = 0$ amounts to

$$\begin{aligned}
&P_{Q_{s+1}|E_1,E_2,N_1,Q_s}(0|e_1, e_2, n_1, 0) \\
&= a_0 + a_1\left(1-P_{\mathrm{e}|E_1,E_2,N_1}(e_1, e_2, n_1)\right)\varrho_r.
\end{aligned} \tag{67}$$

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 [46], [43].

D. 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. One can employ such upper bounds to assess the queueing performance of the system through stochastic dominance.

The evolution of the queue length is governed by the Lindley equation [47, pp. 92–97],

$$Q_{s+1} = (Q_s + A_s - D_s)^+ = \max\{0, Q_s + A_s - D_s\} \quad (68)$$

where A_s is the number of arrivals that occurred during block 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 $\tilde{Q}_{s+1} = \left(\tilde{Q}_s + A_s - \tilde{D}_s\right)^+$, where $\tilde{D}_s$ is drawn according to the distribution implied by the upper bound. Defining Markov chain $\tilde{U}_s = (C_{sN+1}, \tilde{Q}_s)$ and paralleling our previous arguments for the evolution of the system, one can write

$$\begin{aligned} &\Pr\left(\tilde{U}_{s+1} = (d, q_s + i) | \tilde{U}_s = (c, q_s)\right) \\ &= a_i \tilde{P}_{\mathrm{e},S_N|S_0}(d|c) + a_i \left(1 - \tilde{P}_{\mathrm{e},S_N|S_0}(d|c)\right)(1 - \varrho_r) \\ &\quad + a_{i+1}\left(1 - \tilde{P}_{\mathrm{e},S_N|S_0}(d|c)\right)\varrho_r \end{aligned} \quad (69)$$

$$\begin{aligned} &\Pr\left(\tilde{U}_{s+1} = (d, q_s - 1) | \tilde{U}_s = (c, q_s)\right) \\ &= a_0 (1 - \tilde{P}_{\mathrm{e},S_N|S_0}(d|c))\varrho_r \end{aligned} \quad (70)$$

where $i \geq 0$ and $\tilde{P}_{\mathrm{e},S_N|S_0}(j|i)$ is given by (62).

We note that D_s and $\tilde{D}_s$ are Bernoulli random variables with $\Pr(D_s = 1) \geq \Pr\left(\tilde{D}_s = 1\right)$ at every time instant s. It follows that the process Q_s is stochastically dominated by $\tilde{Q}_s$, provided that the two queues are equal at the onset of the process [48], [49]. In addition, when the Markov chains $U_s = (C_{sN+1}, Q_s)$ and $\tilde{U}_s = (C_{sN+1}, \tilde{Q}_s)$ are positive recurrent, the corresponding queueing processes Q_s and $\tilde{Q}_s$ are stable. Then, for any integer q, we have $\Pr(Q_s > q) \leq \Pr\left(\tilde{Q}_s > q\right)$ and, in the limit, we obtain

$$\begin{aligned} \Pr(Q > q) &= \lim_{s\to\infty} \Pr(Q_s > q) \\ &\leq \lim_{s\to\infty} \Pr\left(\tilde{Q}_s > q\right) = \Pr\left(\tilde{Q} > q\right). \end{aligned} \quad (71)$$

That is, Q is stochastically dominated by $\tilde{Q}$, where Q and $\tilde{Q}$ denote stationary distributions. Intuitively, when comparing two queueing systems with a same arrival process, a same underlying channel, and a same code generator, increasing the probability of failure can only result in fewer departures and exacerbate the size of the queue. This observation holds in some generality and can be employed when the exact decoding error probabilities are not known or difficult to compute. Obtaining exact error probabilities can be computationally challenging for block lengths greater than 150, even for the two-state Gilbert-Elliott channel. Our approach enables one to provide performance guarantees for a queueing system using bounds on the probabilities of decoding failure.

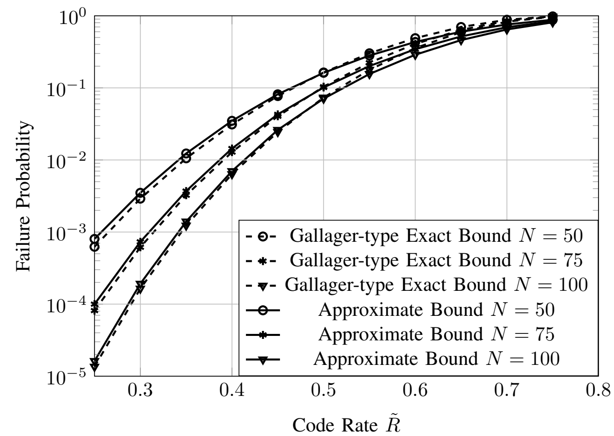

Fig. 4. Comparison of the approximate upper bound (37) with the exact bound (14) in the rare-transition regime with $[\mathbf{P}_N^N]_{12} \approx 4$ and $[\mathbf{P}_N^N]_{21} \approx 6$.

V. Numerical Results

In this section, we present numerical results for probabilities of decoding error and we compare 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.

A. Comparison of Exponential Upper Bounds

We consider a communication system that transmits data over a Gilbert-Elliott channel. We assume the setting of the SNR threshold for transitions between the first and second states are such that the cross-over probabilities of the Gilbert-Elliott channel are $\varepsilon_1 = 0.01$ and $\varepsilon_2 = 0.1$. Figure 4 shows the approximate upper bounds of (37) as functions of block length and code rate, and compares them to the standard Gallager-type bounds of (14). 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. The difference becomes negligible as N grows larger.

In Fig. 5, we plot the probabilities of decoding failure for the maximum-likelihood (ML) and minimum-distance (MD) decoders given by (53)–(56), against the bounds provided in (37). 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 occurs because, for rates close to capacity, the primary source of error in Gallager-type bounds is the loss associated with the prefactor [50], [51]. The figure only features short block lengths, as it is impractical to compute exact performance for long lengths.

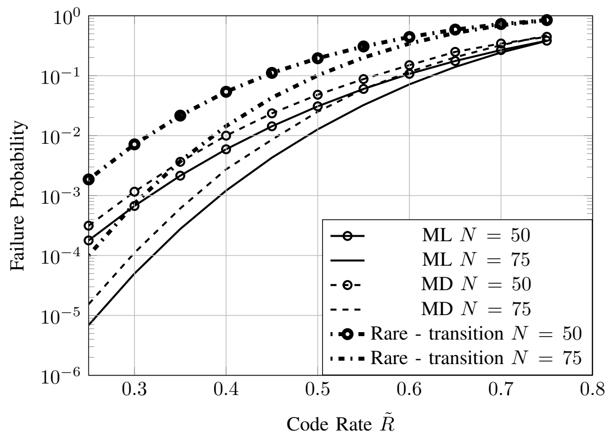

Fig. 5. Comparison of the approximate upper bound (37) with the exact probabilities of decoding failure under ML and MD decoding for $[\mathbf{P}_N]_{12} = 0.0533$ and $[\mathbf{P}_N]_{21} = 0.08$.

B. Evaluation of Queueing Performance

We evaluate the performance of our proposed methodology using traffic parameters roughly selected based on an EVDO system, a 3G component of CDMA2000 [52]. This system offers an uplink sector capacity of 500 Kb/s with 16 active users per sector [53]. For systems with more users and lower per-user rates, this is somewhat optimistic. Accordingly, 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 used by CDMA2000 systems, 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 [52]. 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 $1/\varrho = \sum_i h_i(l_i + \text{overhead}) = 88.55$ bits, where h_i is the relative frequency of state i and l_i denotes the frame size for the same state. Throughout the numerical evaluation, packets are assumed to arrive according to a Poisson process with $\lambda = 50$ packets per second (on average, packets are generated every 20 msec) and we receive an average of $50/R_{\mathrm{b}}$ packets/channel use. The state transition probabilities are considered to be $\alpha = 0.0533$ and $\beta = 0.08$, which roughly result in having an average of 4 and 6 transitions per block. Shannon capacity for this system when the channel state is known at the receiver is equal to 0.764 bits per channel use.

Increasing code rate $\tilde{R}$ 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 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. 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; this approach is standard when it comes to analysis. One can picture two systems operating side by side, one with the genie which informs the occurrence of undetected errors, and one without. The systems are performing the same, until there is an undetected error at which point, the former system reports it and immediately retransmits the corresponding information segment. The latter system does not report the occurrence of undetected error and continues until it gets into the end of the packet when the packet CRC does not check and it retransmits the entire packet. The difference between the two systems can be made arbitrarily small by reducing the probability of undetected error. We required the system to feature a very low probability of undetected error, less than 10^{-5} by a proper choice of the safety margin.

Given this framework, a primary goal is to minimize the tail probability of the queue over all admissible values of N, $\tilde{R}$, and τ or ν. To perform this task using the approximate exponential bound, we first evaluate the bound on undetected error probabilities for different rates and for $\tau = 0$ in (61). Then, for rates with high probability of undetected error, we increase τ so that the bound on probability of undetected error is reduced. Recall that this increases the probability of decoding failure, as seen in (62). Since 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 $\tilde{R}$ 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 (57) and (60).

Figure 6 shows the approximate bound on 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 (61)–(62) such that $\max_{i,j} \tilde{P}_{\mathrm{ue},S_N|S_0}(j|i)$ remains below 10^{-5}. Code rate varies 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. 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 $\tilde{R}$. We see in Fig. 6 that the optimum code

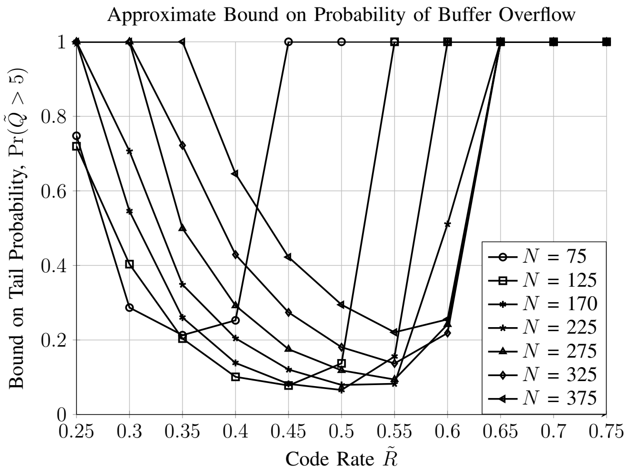

Fig. 6. Approximate bounds on probability of queue exceeding threshold as functions of block length N and code rate $\tilde{R}$. System parameters are subject to $\max_{i,j} \tilde{P}_{\mathrm{ue},S_N|S_0}(j|i) \leq 10^{-5}$.

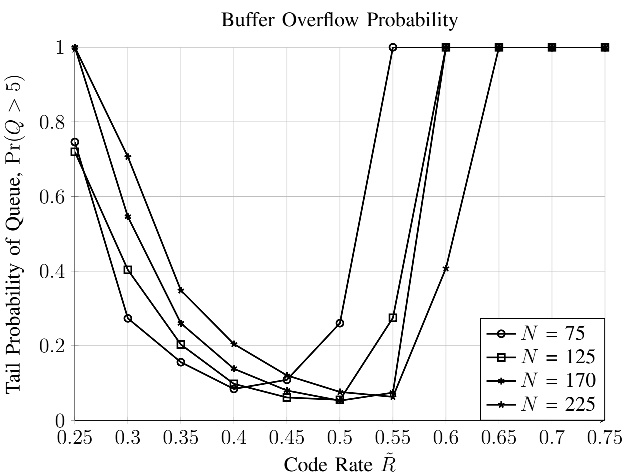

Fig. 7. Exact probability of queue exceeding a threshold as functions of block length N and code rate $\tilde{R}$. System parameters meet constraint $\max_{i,j} \bar{P}_{\mathrm{ue},S_N|S_0}(j|i) \leq 10^{-5}$.

parameters are close to $(N, \tilde{R}) = (170, 0.5)$. For this particular set of code parameters, we have $\tau = 0.048$.

Figure 7 offers similar plots using the exact failure probability. Again, the optimum code parameters are near $(N, \tilde{R}) = (170, 0.5)$. 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. In this scenario, 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. Since the approximate bounds are slightly pessimistic, they produce conservative estimates of overall performance. Empirically, the systems perform better than predicted by the approximate error bounds.

VI. Conclusion

The rare-transition regime is a 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. In this article, we derived an approximate upper bound specifically tailored to the rare-transition regime to estimate the probability of decoding error over finite-state channels. This methodology accounts for both dependence within and across codewords. Furthermore, the proposed bound is numerically efficient to compute. It can be employed for parameter selection and performance analysis in communication links subject to queueing constraints.

We provided supporting evidence for the accuracy of the bounding technique by deriving exact expressions for the Gilbert-Elliott channel. Both maximum-likelihood decoding and a minimum-distance decision rule are 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 with state information available at the destination. This methodology extends to performance criteria based on queueing behavior as well.

We described a practical method to choose 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 congestion control protocols. The queueing analysis based on upper bounds on error probability was shown to provide adequate estimates of system performance and optimum code parameters. Numerical studies suggest that, for fixed conditions, optimal system parameters are essentially unaffected by small variations in the buffer overflow threshold. The methodology and results can be generalized to more intricate channels with memory. In addition, the performance characterization of random codes over finite-state channels may extend to more practical schemes, such as iterative decoding of low density parity-check (LDPC) codes. Possible avenues of future research include the analysis of the rare-transition regime in the absence of side information. Our 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.

Appendix

A. *Proof of Lemma 2*

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 [41]. 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

$$\left[- \left(\mathbf{Q} - \begin{bmatrix} \theta & 0 \\ 0 & 0 \end{bmatrix} - \phi \mathbf{I} \right) \right]^{-1}. \tag{72}$$

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 [41] can be employed to get the desired format in terms of modified Bessel functions.

B. Proof of Lemma 3

Replicating existing results about state transitions in Markov models [41] and adapting them to the present scenario, we gather that

$$\begin{aligned} &P_{N_1|S_0}(m|1) = (1-\alpha)^{m-1}(1-\beta)^{N-m} \times \\ &\left[\sum_{k=0}^{\infty} \binom{m-1}{k} \binom{N-m-1}{k-1} \left(\frac{\alpha}{1-\beta} \right)^k \left(\frac{\beta}{1-\alpha} \right)^k \right. \\ &\left. + \sum_{k=0}^{\infty} \binom{m-1}{k} \binom{N-m-1}{k} \left(\frac{\alpha}{1-\beta} \right)^{k+1} \left(\frac{\beta}{1-\alpha} \right)^k \right]. \end{aligned} \tag{73}$$

Above, the summands are split into two sums according to whether the total number of transitions is even or odd. The upper and lower limits on k can be set to zero and infinity, respectively, because the binomial coefficients will be zero for all inadmissible terms. From the definition of the hypergeometric function ${}_2F_1(\cdot, \cdot; \cdot; \cdot)$, see e.g. [41, Lem. 1, p. 383], we gather that

$$\begin{aligned} &\sum_{k=0}^{\infty} \binom{m-1}{k} \binom{N-m-1}{k} \left(\frac{\alpha}{1-\beta} \right)^{k+1} \left(\frac{\beta}{1-\alpha} \right)^k \\ &\quad = \left(\frac{\alpha}{1-\beta} \right) {}_2F_1(-N+m+1, -m+1; 1; \psi) \end{aligned} \tag{74}$$

where $\psi = \frac{\alpha\beta}{(1-\beta)(1-\alpha)}$. These results lead to

$$\begin{aligned} P_{N_1,S_N|S_0}(m, 2|1) &= \alpha(1-\alpha)^{m-1}(1-\beta)^{N-m} \\ &\times {}_2F_1(-N+m+1, -m+1; 1; \psi) \end{aligned} \tag{75}$$

for $m = 1, \ldots, N-1$. When $m = 0$ and $m = N$, this conditional probability is equal to zero. Likewise, using the recursive formula for binomial coefficients $\binom{N-m-1}{k-1} = \binom{N-m}{k} - \binom{N-m-1}{k}$, we arrive at

$$\begin{aligned} P_{N_1,S_N|S_0}(m, 1|1) &= (1-\alpha)^m(1-\beta)^{N-m} \\ &\times \big({}_2F_1(-N+m, -m+1; 1; \psi) \\ &\quad - {}_2F_1(-N+m+1, -m+1; 1; \psi) \big) \end{aligned} \tag{76}$$

for $m = 1, \ldots, N-1$. 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 a similar manner.

C. Details on Section IV-B

1) The Exact Approach: First, we revisit the ML decoding rule and error probability]for the Gilbert-Elliott channel when channel state information is available at the receiver [43]. Given state occupation $T = (n_1, N - n_1)$, we have

$$\begin{aligned} P_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) &= P_{\mathbf{Y}_1|\mathbf{X}_1}(\mathbf{y}_1|\mathbf{x}_1) P_{\mathbf{Y}_2|\mathbf{X}_2}(\mathbf{y}_2|\mathbf{x}_2) \\ &= \varepsilon_1^{e_1}(1-\varepsilon_1)^{n_1-e_1} \varepsilon_2^{e_2}(1-\varepsilon_2)^{n_2-e_2}. \end{aligned} \tag{77}$$

Upon receiving word $\mathbf{y}$, the ML decoder returns the codeword,

$$\arg\max_{\mathbf{x}\in\mathcal{C}} \ln P_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) = \arg\min_{\mathbf{x}\in\mathcal{C}} [\gamma e_1(\mathbf{x}) + e_2(\mathbf{x})], \tag{78}$$

where $\gamma = \frac{\ln\epsilon_1 - \ln(1-\epsilon_1)}{\ln\epsilon_2 - \ln(1-\epsilon_2)}$, $e_1(\mathbf{x}) = d_{\mathrm{H}}(\mathbf{x}_1, \mathbf{y}_1)$, and $e_2(\mathbf{x}) = d_{\mathrm{H}}(\mathbf{x}_2, \mathbf{y}_2)$. For a uniform random binary code with M codewords, the probability of error for ML decoding is upper bounded by $1 - (1 - V(n_1, n_2, e_1, e_2)/2^N)^{M-1}$, where

$$V(n_1, n_2, e_1, e_2) = \sum_{(\tilde{e}_1, \tilde{e}_2)\in\mathcal{M}(\gamma e_1 + e_2)} \binom{n_1}{\tilde{e}_1} \binom{n_2}{\tilde{e}_2}. \tag{79}$$

Under the 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 (78). To reduce the probability of undetected error, we adapt the technique established in [44] and introduce a safety margin ν. This scheme and its ramifications are easiest to explain for the binary symmetric channel. 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.

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 within the ball of radius e, then detected and undetected failures can occur, although for well-designed systems such events are very rare. 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.

The intuition developed above for binary symmetric channel also applies to the Gilbert-Elliott model except that weighted distance is used. See [43] for additional details.

2) Exponential Bounds: In [44], Forney shows that one can trade undetected errors for detected failures by using a set of disjoint non-exhaustive decision regions. Similar to the previous section, the idea is that the decoder outputs a codeword only if its posterior probably is a factor $e^{N\tau}$, for $\tau \geq 0$, larger than the total probability of all other codewords. If no codeword satisfies this condition, then the decoder declares failure and outputs an erasure. By adjusting τ, one can optimally tradeoff the probability of erasure with the probability of undetected error. Using this modification, one can generalize Gallager's derivation of the error exponent for random codes to show that the probability of undetected error is upper bounded by $\exp(-NE_1(R,\tau))$ and the probability of decoder failure (i.e., detected or undetected error) is upper bounded by $\exp(-NE_2(R,\tau))$, where

$$E_1(R,\tau) = \max_{0\leq v\leq\rho\leq 1} E_0(v,\rho,\mathbf{Q}) - \rho R - v\tau \tag{80}$$

$$E_2(R,\tau) = E_1(R,\tau) + \tau \tag{81}$$

$$\begin{aligned} E_0(v,\rho,\mathbf{Q}) = -\ln \sum_y \Bigg[& \left(\sum_x Q(x) P_{Y|X}(y|x)^{1-v} \right) \\ & \times \left(\sum_{x'} Q(x') P_{Y|X}(y|x')^{v/\rho} \right)^{\rho} \Bigg]. \end{aligned} \tag{82}$$

Forney provides lower bounds, which are tight for small τ, whose expressions in terms of Gallager's function are

$$E_1(R,\tau) \geq \max_{0\leq\rho\leq 1} E_0\left(\frac{\rho}{1+\rho}, \rho, \mathbf{Q}\right) - \frac{\rho}{1+\rho}\tau \tag{83}$$

$$E_2(R,\tau) \geq \max_{0\leq\rho\leq 1} E_0\left(\frac{\rho}{1+\rho}, \rho, \mathbf{Q}\right) + \frac{1}{1+\rho}\tau. \tag{84}$$

Adapting these to our framework gives rise to (61) and (62).

References

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

[2] A. Feinstein, "A new basic theorem of information theory," *IRE Trans. Inf. Theory*, vol. 4, no. 4, pp. 2–22, 1954.

[3] C. E. Shannon, "Certain results in coding theory for noisy channels," *Inf. Contr.*, vol. 1, no. 1, pp. 6–25, 1957.

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

[5] H. Yudkin, "On the exponential error bound and capacity for finite state channels," in *Proc. IEEE Int. Symp. Inform. Theory*, San Remo, Italy, Sept. 1967.

[6] M. Mushkin and I. Bar-David, "Capacity and coding for Gilbert-Elliott channels," *IEEE Trans. Inform. Theory*, vol. 35, no. 6, pp. 1277–1290, Nov. 1989.

[7] A. J. Goldsmith and P. P. Varaiya, "Capacity, mutual information, and coding for finite-state Markov channels," *IEEE Trans. Inform. Theory*, vol. 42, no. 3, pp. 868–886, May 1996.

[8] D. Arnold, H. A. Loeliger, P. O. Vontobel, A. Kavčić, and W. Zeng, "Simulation-based computation of information rates for channels with memory," *IEEE Trans. Inform. Theory*, vol. 52, no. 8, pp. 3498–3508, Aug. 2006.

[9] H. D. Pfister, J. B. Soriaga, and P. H. Siegel, "On the achievable information rates of finite state ISI channels," in *Proc. IEEE Global Telecom. Conf.*, San Antonio, TX, USA, Nov. 2001, pp. 2992–2996.

[10] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Dispersion of the Gilbert-Elliott channel," *IEEE Trans. Inform. Theory*, vol. 57, no. 4, pp. 1829–1848, 2011.

[11] G. Como, S. Yuksel, and S. Tatikonda, "On the Burnashev exponent for markov channels," in *Proc. IEEE Int. Symp. Inform. Theory*, 2007, pp. 1871–1875.

[12] N. Merhav, "On zero-rate error exponents of finite-state channels with input-dependent states," *IEEE Trans. Inform. Theory*, vol. 61, no. 2, pp. 741–750, Feb. 2015.

[13] 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.

[14] 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.

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

[16] Y. Altug and A. B. Wagner, "Moderate deviations in channel coding," *IEEE Trans. Inform. Theory*, vol. 60, no. 8, pp. 4417–4426, 2014.

[17] P. Wu and N. Jindal, "Performance of hybrid-ARQ in block-fading channels: a fixed outage probability analysis," *IEEE Trans. Commun.*, vol. 58, no. 4, pp. 1129–1141, 2010.

[18] T. A. Courtade and R. D. Wesel, "Optimal allocation of redundancy between packet-level erasure coding and physical-layer channel coding in fading channels," *IEEE Trans. Commun.*, vol. 59, no. 8, pp. 2101–2109, 2011.

[19] P. Wu and N. Jindal, "Coding versus ARQ in fading channels: how reliable should the PHY be?" *IEEE Trans. Commun.*, vol. 59, no. 12, pp. 3363–3374, 2011.

[20] H. V. Poor, "Information and inference in the wireless physical layer," *IEEE Trans. Wireless Commun.*, vol. 19, no. 1, pp. 40–47, 2012.

[21] B. Li, H. Shen, and D. Tse, "An adaptive successive cancellation list decoder for polar codes with cyclic redundancy check," *IEEE Commun. Letters*, vol. 16, no. 12, pp. 2044–2047, 2012.

[22] N. Gunaseelan, L. Liu, J.-F. Chamberland, and G. H. Huff, "Performance analysis of wireless hybrid-ARQ systems with delay-sensitive traffic," *IEEE Trans. Commun.*, vol. 58, no. 4, pp. 1262–1272, 2010.

[23] M. Zorzi, R. R. Rao, and L. B. Milstein, "ARQ error control for fading mobile radio channels," *IEEE Trans. Vehicular Technology*, vol. 46, no. 2, pp. 445–455, 1997.

[24] ——, "Error statistics in data transmission over fading channels," *IEEE Trans. Commun.*, vol. 46, no. 11, pp. 1468–1477, 1998.

[25] M. S. Pinsker, V. Prelov, and E. C. van der Meulen, "Transmission of a slowly varying Markov signal over memoryless channels," in *Proc. IEEE Int. Symp. Inform. Theory*, 2000, p. 488.

[26] 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*, 2005, pp. 1838–1842.

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

[28] M. Asadi, R. P. Torghabeh, and N. P. Santhanam, "Estimation in slow mixing, long memory channels," in *Proc. IEEE Int. Symp. Inform. Theory*, 2013, pp. 2104–2108.

[29] E. N. Gilbert, "Capacity of a burst-noise channel," *The Bell Syst. Techn. J.*, vol. 39, pp. 1253–1265, Sept. 1960.

[30] E. O. Elliott, "Estimates of error rates for codes on burst-noise channels," *The Bell Syst. Techn. J.*, vol. 42, pp. 1977–1997, Sept. 1963.

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

[32] V. G. Kulkarni, *Modeling and Analysis of Stochastic Systems*. Chapman and Hall, 1995.

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

[34] 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.

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

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

[37] P. Billingsley, *Convergence of probability measures*. New York, New York, USA: John Wiley & Sons, 1968.

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

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

[40] 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.

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

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

[43] F. Hamidi-Sepehr, H. Pfister, and J.-F. Chamberland, "Delay-sensitive communication over fading channel: Queueing behavior and code parameter selection," *IEEE Trans. Vehicular Technology*, 2015.

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

[45] 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.

[46] 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.*, 2002, pp. 86–96.

[47] S. Asmussen, *Applied Probability and Queues*. Springer Science & Business Media, 2003, vol. 51.

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

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

[50] Y. Altug and A. B. Wagner, "Refinement of the random coding bound," *IEEE Trans. Inform. Theory*, vol. 60, no. 10, pp. 6005–6023, 2014.

[51] J. Scarlett, A. Martinez, and A. Guillen i Fabregas, "The saddlepoint approximation: Unified random coding asymptotics for fixed and varying rates," in *Proc. IEEE Int. Symp. Inform. Theory*, 2014, pp. 1892–1896.

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

[53] 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.

Fatemeh Hamidi-Sepehr (S'09-14) received the B.S. degree from University of Tehran, Tehran, Iran, in 2006; the M.S. degree from Sharif University of Technology, Tehran, Iran, 2008; and the Ph.D. degree from Texas A&M University, College Station, TX, in 2014, all in electrical engineering. Her research interests include interplay between queueing theory and error-control coding with applications in delay-sensitive communications.

She has been with Qualcomm Inc. in Santa Clara, CA, since 2014, working in the area of wireless communications and design of cellular modems under 3G and LTE technologies.

Jean-Francois Chamberland (S'98–M'04–SM'09) received the Ph.D. degree in 2004 from the University of Illinois at Urbana-Champaign, the M.S. degree in 2000 from Cornell University, Ithaca, NY, and the B.Eng. degree in 1998 from McGill University, Montreal, Canada, all in electrical engineering. He joined Texas A&M University in 2004, where he is currently an associate professor in the Department of Electrical and Computer Engineering. His research interests include communication systems, queueing theory, detection and estimation, and statistical information processing. In 2006, he was the recipient of a Young Author Best Paper Award from the IEEE Signal Processing Society. He received a CAREER Award from the National Science Foundation in 2008.

Henry D. Pfister (S'99-M'03-SM'09) received his Ph.D. in electrical engineering from UCSD in 2003 and is currently an associate professor in the electrical and computer engineering department of Duke University. Prior to that, he was a faculty member at Texas A&M University (2006-2014), a post-doc at EPFL (2005-2006), and a senior engineer at Qualcomm Corporate R&D in San Diego (2003-2004).

He received the NSF Career Award in 2008, the Texas A&M ECE Department Outstanding Professor Award in 2010, and was a coauthor of the 2007 IEEE COMSOC best paper in Signal Processing and Coding for Data Storage. He is currently an associate editor in coding theory for the IEEE Transactions on Information Theory.