THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Fundamental Limits of Ultra-Reliable Low-Latency Communication

A. Oguz Kislal

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Electrical Engineering
Chalmers University of Technology
Gothenburg, Sweden, 2024

Fundamental Limits of Ultra-Reliable Low-Latency Communication

A. Oguz Kislal
ISBN 978-91-8103-091-4

Acknowledgements, dedications, and similar personal statements in this thesis, reflect the author's own views.

Doktorsavhandlingar vid Chalmers tekniska högskola
Ny serie nr 5549
ISSN 0346-718X
This thesis has been prepared using LaTeX, Tikz, and PgfPlots.

Department of Electrical Engineering
Chalmers University of Technology
SE-412 96 Göteborg, Sweden
Phone: +46 (0)31 772 1000
www.chalmers.se

Printed by Chalmers Reproservice
Göteborg, Sweden, September, 2024

Fundamental Limits of Ultra-Reliable Low-Latency Communication

A. Oguz Kislal
Department of Electrical Engineering
Chalmers University of Technology

Abstract

This thesis focuses on the fundamental performance of ultra-reliable low-latency communication (URLLC) systems, particularly in terms of achievable error probabilities. Since low latency necessitates the use of short data packets, understanding the trade-off between blocklength and reliability is crucial. This thesis provides design guidelines for URLLC systems, including those with multiple antennas, imperfect time synchronization, and error detection, based on three key papers.

In Paper A, we present a numerically efficient method for evaluating the random coding union bounds with parameter s for the error probability achievable by a pilot-assisted transmission method on a block-fading channel. Our approach, which uses the saddlepoint approximation with respect to the number of fading blocks, significantly reduces the number of Monte Carlo samples required to accurately estimate the achievability bound for the error probability, particularly in scenarios with multiple diversity branches.

In Paper B, we investigate the achievable error probability in the finite blocklength regime for a pilot-assisted transmission scheme operating over an imperfectly synchronized, memoryless block-fading waveform channel. Numerical experiments show that the number of pilot symbols needed to estimate the fading channel gains with the accuracy required for URLLC is also sufficient for synchronization when delays are modeled as fully dependent across fading blocks.

In Paper C, we explore a URLLC system with an erasure decoder and investigate the trade-off between the total error probability and the probability of undetected errors in the short blocklength regime by developing two achievability bounds. These bounds are tighter than Forney's error exponent bound for the additive white Gaussian noise channel and can be evaluated for practically relevant channels.

Keywords: Ultra-reliable low-latency, short packets, mismatched decoding, imperfect CSI, imperfect synchronization, multiuser massive MIMO, erasure decoding.

List of Publications

This thesis is based on the following publications:

[A] **A. Oguz Kislal**, Alejandro Lancho, Giuseppe Durisi, Erik G. Ström, "Efficient evaluation of the error probability for pilot-assisted URLLC with Massive MIMO". *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 7, pp. 1969–1981, May. 2023.

[B] **A. Oguz Kislal**, Madhavi Rajiv, Giuseppe Durisi, Erik G. Ström, Urbashi Mitra, "Is Synchronization a Bottleneck for Pilot-Assisted URLLC Links?". To appear in *IEEE Transactions on Wireless Communications*.

[C] Alexander Sauter, **A. Oguz Kislal**, Giuseppe Durisi, Gianluigi Liva, Balazs Matuz, Erik G. Ström, "Undetected Error Probability in the Short Blocklength Regime: Approaching the Bounds with Polar Codes". Submitted to *IEEE Transactions on Communications*, Jun. 2024.

Other publications by the author, not included in this thesis, are:

[D] **A. Oguz Kislal**, Madhavi Rajiv, Giuseppe Durisi, Erik G. Ström, Urbashi Mitra, "Pilot-Assisted URLLC Links: Impact of Synchronization Error," *To appear in Proc. IEEE Int. Conf. on Commun. (ICC),* Denver, CO, USA Jun. 2024.

[E] **A. Oguz Kislal**, Alejandro Lancho, Giuseppe Durisi, Erik G. Ström, "Efficient evaluation of the error probability for pilot-assisted finite-blocklength transmission," *in Proc. Asilomar Conf. Signals, Syst., Comput.,* Pacific Grove, CA, USA Nov. 2022.

[F] **A. Oguz Kislal**, Bayram C. Akdeniz, Changmin Lee, Ali E. Pusane, Tuna Tugcu, Chan-Byoung Chae, "ISI-mitigating channel codes for molecular communication via diffusion," *IEEE Access* vol. 8, pp. 24588-24599, Jan. 2020.

[G] **A. Oguz Kislal**, Bayram C. Akdeniz, Changmin Lee, Ali E. Pusane, Tuna Tugcu, Chan-Byoung Chae, "ISI-aware channel code design for molecular communication via diffusion," *IEEE Trans. on NanoBioscience* vol. 18, no. 2, pp. 205-213, Apr. 2019.

[H] Osman Ceylan, Alican Caglar, Halim Bahadir Tugrel, Hasan Onur Cakar, **A. Oguz Kislal**, Kaan Kula, Hasan Bulent Yagci, "Small satellites rock a software-defined radio modem and ground station design for cube satellite communication," *IEEE Microwave Magazine* vol. 17, no. 3, pp. 26-33, Feb. 2016.

Acknowledgments

My life as a doctoral student was like a roller coaster highs were very high and lows were very low. During this journey, my two supervisors Prof. Erik G. Ström and Prof. Giuseppe Durisi made sure that I did not jump off from my seat. Your attention to detail and your ability to express complex problems in a clean way amazes me to this day. I would like to thank you both for making this journey possible.

During my PhD, I had the opportunity to work with great researchers. I would like to thank Prof. Urbashi Mitra and Prof. Gianluigi Liva for the fruitful discussions and Madhavi Rajiv for meeting me in the middle of the night. I also would like to thank Alejandro Lancho for holding my hand as I tried to take my first steps into information theory. Alex, this work would not have been possible without you. I hope that one day I can inspire someone the way you did for me.

The inviting nature of the ComSys group helped me a lot in this journey. Thank you Andreas for keeping the time for the fruit baskets, Arni for playing good music in the office, Liqin for listening all my whining without a sign of boredom, Jose for being my tennis buddy, Furkan for being an amazing friend and my lunch buddy, and Prof. Fredrik Brännström for taking great care to create a friendly environment in our group.

Outside of Chalmers, there were many that supported me throughout this journey. I would like to thank Ulas, Berkin, Ayhan, Ekin, Ogo, and Bugra for carrying me game after game, Abdullah, Ramazan and Alican for teaching me F1, Can Sezgin, Necla, Sina, Sepi and Dilan for encouraging me to jump in a frozen lake and Can Tok for hosting me every time I go to Istanbul. This was a great journey thanks to all of you.

During this journey, there were times that everything looked dark and gloomy. I was very lucky to have my family reminding me that there is light at the end of the tunnel. Without the immense support and the unconditional love of my parents and my brother, I would never be able to see the end of this journey. I am deeply grateful for being the source of strength I needed to reach the end.

A. Oguz Kislal
Gothenburg, September, 2024

Financial support

This research work was funded by Swedish Research Council grants 2016-03293, 2018-04359 and 2021-04970. Moreover, the computation were enabled by resources provided by Chalmers e-Commons.

Acronyms

AR:	augmented reality
ARQ:	automatic repeat request
AWGN:	additive white Gaussian noise
biAWGN:	binary-input additive white Gaussian noise
BS:	base station
CGF:	cumulant generating function
CLT:	central limit theorem
CRC:	cyclic redundancy check
CSI:	channel state information
DMC:	discrete memoryless channel
DT:	dependence testing
eMBB:	enhanced mobile broadband
i.i.d.:	identically and independently distributed
IoT:	internet of Things
MGF:	moment generating function
MIMO:	multiple-input multiple-output
ML:	maximum likelihood
mMTC:	massive machine-type communication
OFDM:	orthogonal frequency-division multiplexing
PAT:	pilot-assisted transmission
pdf:	probability density function

RCU:	random coding Union
RCUs:	random coding Union with parameter s
r.h.s.:	right hand side
SISO:	single-input single-output
SNN:	scaled nearest-neighbor
TDD:	time division duplex
TEP:	total error probability
UEP:	undetected error probability
URLLC:	ultra-reliable low latency communication
WSS:	wide-sense stationary

Contents

Part I

Overview

CHAPTER 1

Introduction

Since the emergence of the initial iteration of analog wireless cellular systems, the past five decades have witnessed swift advancements in communication. With the introduction of next-generation wireless communications, new applications, that fall under the Internet of Things (IoT) paradigm, are enabled. In the context of IoT, it is anticipated that next-generation wireless communication systems will facilitate interconnection of a wide range of devices. These devices will span from vehicles or drones, functioning within dynamic and high-mobility environments, to autonomous machines or stationary sensors, operating within static or low-mobility settings [1]. To accommodate diverse performance needs and facilitate a range of services, 5G has identified three primary service categories: enhanced mobile broadband (eMBB), which aims to deliver high data rates and spectral efficiency; massive machine-type communication (mMTC), designed to enable connectivity for a large volume of machine-type devices; and ultra-reliable and low latency communication (URLLC), intended to support services where the data throughput is low (e.g., 50 Mbit/s) but the transmission needs to satisfy the stringent requirements in terms of reliability (e.g., 99.999%) and latency (e.g., 1 ms) [2], [3]. This thesis aims to provide a practically relevant framework for benchmarking

URLLC systems.

URLLC plays a crucial role in enabling mission-critical services and holds the potential to revolutionize various industries. Here are some major examples:

- Industrial automation [4], [5]: URLLC is anticipated to be implemented in factories to automate mission-critical processes such as robot motion control and tactile interaction. Utilizing this technology can potentially decrease the operational costs significantly.
- Intelligent transportation [6], [7]: URLLC facilitates the swift and reliable exchange of information among vehicles, infrastructure, and pedestrians, thereby contributing to improved road safety and enhanced traffic efficiency. Typical usage scenarios encompass automated overtaking, cooperative collision avoidance, and traffic management.
- Augmented reality (AR) [8]: AR is an approach to enhance the perception of the real-world environment by incorporating computer-generated information, including audio, video, and geographic data. This application imposes highly stringent demands on latency and reliability, where the use of URLLC could alleviate the discomfort caused by prolonged delays between actions and responses.

URLLC systems cover all these use cases with the key features like: short transmission time [9], error detection and feedback schemes, e.g. automatic repeat request (ARQ) [10], [11], the channel codes designed for short-packets [12], and exploitation of diversity [13]. Diversity here is especially important to satisfy the reliability constraints. The use of time diversity, which is popular in traditional communication systems, may not be practical in URLLC due to latency requirements [14, Ch. 1]. Therefore, in URLLC, different sources of diversity must be utilized, such as frequency diversity, which consists of transmission over a bandwidth spanning multiple coherent bandwidths, or spatial diversity [15].

Obtaining channel state information (CSI) is another important process for communication systems. Traditionally, the resources required to obtain CSI have rarely been accounted for. In fact, one can estimate the channel using pilot-assisted transmission (PAT) without significantly degrading the performance of the system for long packets. However, when communicating

with short packets, the impact of pilots can be significant and should be taken into account.

Traditional wireless communication systems were designed relying on the guidelines provided by information-theory analyses based on ergodic and outage capacity. Ergodic capacity is the maximum coding rate achievable with vanishing error probability when the blocklength tends to infinity, and the outage capacity, which is also called ϵ-capacity, is the maximum coding rate achievabile, given that the probability of a certain outage probability is smaller than ϵ. Due to the infinite blocklength assumption in their definitions, both of these capacities do not provide insightful benchmarks when the blocklength is short [1], [16]. In this case, a characterization of the maximum coding rate achievable as a function of the blocklength and the error probability is needed.

This thesis aims to analyze the fundamental performance, specifically focusing on the achievable error probability of communication systems operating over block-fading channels within the URLLC regime. In particular, we aimed to answer three research questions, all concerning URLLC systems, as follows:

1. The use of nonasymptotic achievability bounds within URLLC optimization routines, such as resource allocation and scheduling algorithms, is challenging due to their time-consuming nature. Therefore, how can we develop a numerically efficient method to evaluate these achievability bounds?

2. The timing synchronization is typically assumed to be perfect in the analyses available in the literature. This assumption is validated for traditional communication systems through the use of many pilot symbols. However, this argument does not hold for URLLC due to the use of short packets. So, what is the impact of timing synchronization error in URLLC systems?

3. The reliability of URLLC systems is often quantified by the block error probability at the decoder output. However, for certain mission-critical applications, this measure alone is insufficient. In such cases, it is important to distinguish between two kinds of decoding errors: detected and undetected errors. Given this distinction, how does a URLLC system utilizing such a decoder perform in terms of both undetected and detected errors?

1.1 Thesis Organization

This thesis is divided into two parts. In Part I, an overview of finite-blocklength information theory is presented. We delve into the critical differences between traditional communication systems and URLLC systems mentioned in the introduction and explore the basics of the finite-blocklength tools we used in our papers. Part I reviews the models and tools used in the appended papers in Part II.

In particular, Chapter 2, presents a review of the nonasymptotic achievability bounds on the error probability. Chapter 3 presents practically relevant system models describing the propagation of the channel and provide the background for imperfect CSI, mismatched decoding and massive multiple-input multiple-output (MIMO) communication systems. Chapter 4 reviews the asymptotic expansions of random coding union bound with parameter s. Chapter 5 reviews the erasure decoding and the fundamental trade-off between erasure probability and undetected error probability. Chapter 6 contains the concluding remarks for the first part of the thesis including an outline of the contributions and future work.

1.2 Notation

We denote random vectors and random scalars by upper-case boldface letters such as $\boldsymbol{X}$ and upper-case standard letters, such as X, respectively. Their realizations are indicated by lower-case letters of the same font. To avoid ambiguities, we use another font, such as R for rate, to denote constants that are typically capitalized in the literature. The circularly-symmetric Gaussian distribution is denoted by $\mathcal{CN}(0, \sigma^2)$, where σ^2 denotes the variance. The superscript $(\cdot)^H$ denotes Hermitian transposition. We write $\log(\cdot)$ to denote the natural logarithm, $\|\cdot\|$ stands for the ℓ^2-norm, $\mathbb{P}[\cdot]$ for the probability of an event, $\mathbb{E}[\cdot]$ for the expectation operator, $\mathbb{V}\mathrm{ar}[\cdot]$ for the variance of a random variable, $\mathbb{1}\{\cdot\}$ is the indicator function, and $Q(\cdot)$ for the Gaussian Q-function. Finally, $\mathbb{N}_0$ to denote the set of natural numbers including 0, $(a)^+$ stands for $\max(a)$, and for two functions $f(n)$ and $g(n)$, the notation $f(n) = o(g(n))$ means that $\lim_{n\to\infty} f(n)/g(n) = 0$ and the notation $f(n) = \mathcal{O}(g(n))$ means that $\limsup_{n\to\infty} |f(n)/g(n)| < \infty$.

CHAPTER 2

Finite-Blocklength Achievability Bounds

In this chapter, we review some of the main tools used in finite blocklength information theory analysis. The discussion focuses on the achievability bounds over the codes with a fixed blocklength, which relate to the results in Paper A, Paper B and Paper C.

2.1 Overview

In URLLC, devices communicate at low data rates, aiming for an error probability of 10^{-5} or less, while meeting strict latency requirements. This is generally achieved by transmitting short packets. Since capacity and outage capacity are traditionally defined under the assumption of infinite blocklength, it is necessary to provide a more precise characterization of the maximum coding rate versus blocklength.

We consider an arbitrary discrete-time memoryless communication channel that maps input symbols from the set $\mathcal{X}$ to output symbols from the set $\mathcal{Y}$. Specifically, let $\mathbf{x} = [x_1, \ldots, x_n] \in \mathcal{X}^n$ be the input vector and $\boldsymbol{Y} = [Y_1, \ldots, Y_n] \in \mathcal{Y}^n$ be the output vector of the channel. The channel is characterized by its transition probability $P_{\boldsymbol{Y}|\boldsymbol{X}}(\mathbf{y}|\mathbf{x})$. To simplify the no-

tation, we define the transition probability using the conditional probability mass function $P_{\boldsymbol{Y}|\boldsymbol{X}}(\mathbf{y}|\mathbf{x})$, which assumes that the set $\mathcal{Y}$ has finite cardinality. Nevertheless, most analysis here is general and applicable to channels with continuous input and output. In such cases, $P_{\boldsymbol{Y}|\boldsymbol{X}}(\mathbf{y}|\mathbf{x})$ should be replaced by the conditional probability density function. For memoryless channels, $P_{\boldsymbol{Y}|\boldsymbol{X}}(\mathbf{y}|\mathbf{x})$ factorizes as

$$P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{Y}|\boldsymbol{X}) = \prod_{i=1}^{n} P_{Y_i|X_i}(Y_i|X_i). \tag{2.1}$$

We next define the notion of a channel code.

Definition 2.1. *An $(n, \mathrm{M}, \epsilon)$-code for the channel $P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{Y}|\boldsymbol{X})$ consists of:*

- *An encoder $f : \{1, \ldots, \mathrm{M}\} \to \mathcal{X}^n$ that maps a message W, which we assume to be uniformly distributed on $\{1, \ldots, \mathrm{M}\}$, to a codeword in the set $\mathcal{C} = \{\mathbf{x}_1, \ldots, \mathbf{x}_{\mathrm{M}}\} \subset \mathcal{X}^n$.*
- *A decoder $g : \mathcal{Y}^n \to \{1, \ldots, \mathrm{M}\}$, that maps the received vector to a message $\hat{W} = g(\boldsymbol{Y})$ The decoder satisfies the average error probability constraint*

$$\mathbb{P}\left[\hat{W} \neq W\right] \leq \epsilon \tag{2.2}$$

The performance of $(n, \mathrm{M}, \epsilon)$-code is typically assessed, for a fixed blocklength n and average error probability ϵ, via the maximum coding rate as

$$\mathsf{R}^*(n, \epsilon) = \frac{\log(\mathrm{M}^*(n, \epsilon))}{n} \tag{2.3}$$

where

$$\mathrm{M}^*(n, \epsilon) = \sup\{\mathrm{M} : \exists (n, \mathrm{M}, \epsilon)\text{-channel code}\}. \tag{2.4}$$

The supremum in (2.4) is over all encoder/decoder pairs that form a $(n, \mathrm{M}, \epsilon)$ channel code. Since the exact characterization of $\mathrm{R}^*(n, \epsilon)$ is generally NP-hard [17], non-asymptotic information-theoretic analyses typically focus on establishing bounds for $\mathrm{R}^*(n, \epsilon)$ and their asymptotic expansions. Note that we shall refer n as blocklength, R as rate in nats per channel use[1]. Then the size of the codebook is $\mathrm{M} = e^{n\mathrm{R}}$.

[1] With an abuse of notation, we will use R to denote also the rate measured in bits per channel use when presenting numerical results throughout the thesis.

For a fixed codebook $\mathcal{C}$, it is (in general) very hard to compute the error probability. To overcome this problem, we will use the argument of random coding. To do so, we consider the average error probability over an ensamble of codebooks (for instance the set of all Gaussian codebooks). The key idea here is to select the message W and the codebook ensemble uniformly so that each symbol is independent. Assuming that we draw each symbol of a codeword from a particular distribution P_X, each symbol is i.i.d. with P_X, which greatly simplifies our analysis. We also assume that we have a maximum likelihood (ML) decoder, i.e., that the receiver searches for the codeword in the codebook that is closest to the received signal.

Next, we will review commonly used non-asymptotic bounds in the literature, focusing on the random coding union (RCU) bound and the RCU bound with parameter s (RCUs). Then, starting from RCUs bound, we will also derive Gallager's Random Coding Bound (also known as Gallager Bound). The analysis in Paper A and Paper B are based on the RCUs bound, and Paper C uses both the random coding bound and the RCU bound.

2.2 Non-asymptotic Bounds

By determining an upper limit for the maximum coding rate, the achievable rate for a certain error probability and blocklength can be characterized. The dependence testing (DT) and the beta-beta bounds are two such non-asymptotic bounds which are given as:

- DT bound: A threshold decoder sequentially consider all codewords and, for each codeword, evaluates the information density $\imath(\mathbf{x};\mathbf{y})$, which is given as
$$\imath(\mathbf{x};\mathbf{y}) = \frac{P_{\boldsymbol{Y}|\boldsymbol{X}}(\mathbf{y}|\mathbf{x})}{P_{\boldsymbol{Y}}(\mathbf{y})} \tag{2.5}$$
where
$$P_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{X},\boldsymbol{Y}) = P_{\boldsymbol{X}}(\boldsymbol{X})P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{Y}|\boldsymbol{X}) \tag{2.6}$$
and $P_{\boldsymbol{Y}}$ is the resulting marginal distribution. Then, the decoder outputs the first codeword whose information density exceeds a certain threshold. We obtain DT bound after an analysis of the error probabil-

ity using this decoder as [16], [18]

$$\epsilon \leq \mathbb{E}\left[e^{\left(\imath(\boldsymbol{X},\boldsymbol{Y})-\log\frac{\mathrm{M}-1}{2}\right)^{+}}\right]. \tag{2.7}$$

- Beta-beta bound: In [19], the authors have extended the so-called golden formula, which expresses the mutual information of two random variables as the difference of two relative entropies involving an auxiliary distribution [20], to the finite-blocklength regime. To this end, using the random coding argument and a suboptimal decoder based on the Neyman-Pearson test between $P_{\boldsymbol{X},\boldsymbol{Y}}$ and $P_{\boldsymbol{X}}Q_{\boldsymbol{Y}}$, where $Q_{\boldsymbol{Y}}$ is the auxiliary output distribution, the beta-beta bound is given as [19]

$$\frac{\mathrm{M}}{2} \geq \sup_{0<\tau<\epsilon}\left(\frac{\beta_{\tau}(P_{\boldsymbol{Y}},Q_{\boldsymbol{Y}})}{\beta_{1-\epsilon+\tau}(P_{\boldsymbol{X},\boldsymbol{Y}},P_{\boldsymbol{X}}Q_{\boldsymbol{Y}})}\right) \tag{2.8}$$

  where $\beta_{\alpha}(P,Q)$ is the classic Neyman-Pearson function given as

$$\beta_{\alpha}(P,Q) = \min\sum_{x} P_{Z|X}(1|x)Q(x) \tag{2.9}$$

  where $Z \in \{0,1\}$ is a randomized test between P and Q with $Z=0$ indicates that the test chooses Q. In (2.9) minimum is over all tests satisfying

$$\sum_{x} P_{Z|X}(1|x)P(x) \geq \alpha. \tag{2.10}$$

The DT and beta-beta bounds are intractable for the practical system models we will consider in this thesis (see Chapter 3 for more details). The RCUs bound, which we review in detail next, is tractable even when applied to practical system models.

2.2.1 RCU and RCUs Bounds

As usual, we use a random coding argument and analyze the average error probability under ML decoding, averaged over all possible codewords generated from the distribution $P_{\boldsymbol{X}}$. Under ML decoding, an error occurs if the transmitted codeword has a lower posterior probability than one of the other codewords. Mathematically, the error probability for a given realization of

the random codebook can be upper bounded as

$$\epsilon(\mathbf{x}_1, \ldots, \mathbf{x}_{\mathrm{M}}) \leq \frac{1}{\mathrm{M}} \sum_{i=1}^{\mathrm{M}} \mathbb{P}\left[\bigcup_{j=1, j\neq i}^{\mathrm{M}} P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{Y}|\mathbf{x}_j) \geq P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{Y}|\mathbf{x}_i) \right] \tag{2.11}$$

where the inequality here arises from the assumption that all ties result in errors. Next, we evaluate the error probability averaged over all codebooks, and since all codewords are identically and independently distributed (i.i.d.), we assume, without loss of generality, that the first codeword $\boldsymbol{X}_1$ is transmitted. Then, the average error probability $\epsilon = \mathbb{E}[\epsilon(\mathbf{x}_1, \ldots, \mathbf{x}_{\mathrm{M}})]$ can be evaluated as

$$\epsilon \leq \mathbb{P}\left[\bigcup_{j=2}^{\mathrm{M}} P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{Y}|\boldsymbol{X}_j) \geq P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{Y}|\boldsymbol{X}_1) \right] \tag{2.12}$$

$$= \mathbb{E}\left[\mathbb{P}\left[\bigcup_{j=2}^{\mathrm{M}} P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{Y}|\boldsymbol{X}_j) \geq P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{Y}|\boldsymbol{X}_1) \;\middle|\; \boldsymbol{X}_1, \boldsymbol{Y} \right] \right] \tag{2.13}$$

$$\leq \mathbb{E}\left[\min\left\{ 1, \sum_{j=2}^{\mathrm{M}} \mathbb{P}\big[P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{Y}|\boldsymbol{X}_j) \geq P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{Y}|\boldsymbol{X}_1) \;\big|\; \boldsymbol{X}_1, \boldsymbol{Y} \big] \right\} \right] \tag{2.14}$$

$$= \mathbb{E}\big[\min\big\{ 1, (\mathrm{M}-1)\, \mathbb{P}\big[P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{Y}|\bar{\boldsymbol{X}}) \geq P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{Y}|\boldsymbol{X}) \;\big|\; \boldsymbol{X}, \boldsymbol{Y} \big] \big\} \big]\,. \tag{2.15}$$

where

$$P_{\boldsymbol{X},\boldsymbol{Y},\bar{\boldsymbol{X}}}(\boldsymbol{X}, \boldsymbol{Y}, \bar{\boldsymbol{X}}) = P_{\boldsymbol{X}}(\boldsymbol{X}) P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{Y}|\boldsymbol{X}) P_{\boldsymbol{X}}(\bar{\boldsymbol{X}}) \tag{2.16}$$

meaning that $\bar{\boldsymbol{X}}$ has same distribution as $\boldsymbol{X}$ but it is independent of $\boldsymbol{X}$ and $\boldsymbol{Y}$. Here, in (2.13), we condition on $\boldsymbol{X}_1$ and $\boldsymbol{Y}$ to have a tighter bound once we apply the union bound in (2.14), and (2.15) follows since all codewords are i.i.d., so the $(\mathrm{M}-1)$ probability terms are identical. Note that RCU bound can also equivalently be stated as

$$\epsilon \leq \mathbb{E}\big[\min\big\{ 1, (\mathrm{M}-1)\, \mathbb{P}\big[\imath(\bar{\boldsymbol{X}}; \boldsymbol{Y}) \geq \imath(\boldsymbol{X}; \boldsymbol{Y}) \;\big|\; \boldsymbol{X}, \boldsymbol{Y} \big] \big\} \big]\,. \tag{2.17}$$

The RCU bound is extremely difficult to compute for many practical channels. In general, there is no closed-form expression for the probability in (2.15), and this probability can be extremely small, since it is scaled by $(\mathrm{M}-1)$. For $n = 100$, $\mathrm{R} = 1/2$ we have $\mathrm{M} - 1 \approx 10^{15}$, which means that the probability

term in (2.15) must be evaluated with a resolution of 10^{-15}. The evaluation with a Monte Carlo simulation would require $\sim 10^{17}$ samples for the probability event. In addition, the probability term must be evaluated for different realizations of $\boldsymbol{X}$ and $\boldsymbol{Y}$. In [21], the authors evaluated both the probability term and the expectation in (2.15) using saddlepoint approximation for a binary-input additive white Gaussian noise (biAWGN) channel. We will also review saddlepoint approximation in detail in Chapter 4.

One may relax the RCU bound using a generalized Markov inequality which, for nonnegative random variable Z and $s > 0$, is given as

$$\mathbb{P}(Z > \gamma) \leq \frac{\mathbb{E}[Z^s]}{\gamma^s}. \tag{2.18}$$

After applying Markov inequality to the probability term in (2.15) we get

$$\epsilon \leq \mathbb{E}\left[\min\left\{1, (\mathrm{M}-1)\frac{\mathbb{E}_{\bar{\boldsymbol{X}}}\left[P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{Y}|\bar{\boldsymbol{X}})^s\right]}{P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{Y}|\boldsymbol{X})^s}\right\}\right] \tag{2.19}$$

$$= \mathbb{E}\left[e^{-(\imath_s(\boldsymbol{X};\boldsymbol{Y})-\log(\mathrm{M}-1))^+}\right] \tag{2.20}$$

where $\imath_s(\boldsymbol{X};\boldsymbol{Y})$ is the so-called generalized information density and defined as

$$\imath_s(\mathbf{x};\mathbf{y}) = \log\frac{P_{\boldsymbol{Y}|\boldsymbol{X}}(\mathbf{y}|\mathbf{x})^s}{\mathbb{E}\left[P_{\boldsymbol{Y}|\boldsymbol{X}}(\mathbf{y}|\bar{\boldsymbol{X}})^s\right]}. \tag{2.21}$$

Using the identity $\mathbb{E}[\min\{1, Z\}] = \mathbb{P}\{Z \geq U\}$ for a non-negative random variable Z and uniform random variable U in $(0, 1)$ [16], we may express the RCUs bound as a tail probability as

$$\epsilon \leq \mathbb{P}[\imath_s(\boldsymbol{X};\boldsymbol{Y}) + \log U \leq \log(\mathrm{M}-1)]\,. \tag{2.22}$$

2.2.2 Gallager's Random Coding Bound

Refined asymptotic expansions are derived by asymptotic analyses of the maximum coding rate or error probability, which become more and more precise with increasing blocklength. These expansions are usually expressed in closed form and illustrate how the maximum coding rate approaches capacity for a given coding rate or how the error probability decays with increasing blocklength. The Gallager's random coding error exponent, which we shall derive

next, is one of the classical examples of such an expression in the information-theory literature.

Let $\kappa_s(\mathbf{y})$ be the cumulant-generating function (CGF) of $\imath(\bar{\boldsymbol{X}}, \mathbf{y})$, which is the logarithm of moment generating function (MGF), as

$$\kappa(s, \mathbf{y}) = \log \mathbb{E}\left[e^{s\imath(\bar{\boldsymbol{X}}, \mathbf{y})}\right] \tag{2.23}$$

Then,

$$\begin{aligned}
\epsilon &\leq \inf_{s>0} \mathbb{E}\left[\min\left\{1, (\mathrm{M}-1)e^{-(s\imath(\boldsymbol{X};\boldsymbol{Y})-\kappa(s,\boldsymbol{Y}))}\right\}\right] && (2.24)\\
&\leq \min_{0\leq p\leq 1} \inf_{s>0} \mathbb{E}\left[(M-1)^p e^{-(ps\imath(\boldsymbol{X};\boldsymbol{Y})-p\kappa(s,\boldsymbol{Y}))}\right] && (2.25)\\
&\leq \min_{0\leq p\leq 1} \inf_{s>0} \mathbb{E}\left[e^{-\left(ps\imath(\bar{\boldsymbol{X}},\boldsymbol{Y})-p\kappa(s,\boldsymbol{Y})-\imath(\bar{\boldsymbol{X}};\boldsymbol{Y})-pn\mathrm{R}\right)}\right] && (2.26)\\
&= \min_{0\leq p\leq 1} \inf_{s>0} \mathbb{E}\left[e^{(1-sp)\imath(\bar{\boldsymbol{X}};\boldsymbol{Y})+p\kappa(s,\boldsymbol{Y})+pn\mathrm{R}}\right] && (2.27)\\
&= \min_{0\leq p\leq 1} \inf_{s>0} \mathbb{E}\left[e^{\kappa(1-sp,\boldsymbol{Y})+p\kappa(s,\boldsymbol{Y})+pn\mathrm{R}}\right] && (2.28)\\
&= \min_{0\leq p\leq 1} \mathbb{E}\left[e^{(1+p)\kappa\left(\frac{1}{1+p},\boldsymbol{Y}\right)+pn\mathrm{R}}\right] && (2.29)
\end{aligned}$$

where (2.24) is equivalent to (2.19) and infimum over s is included to obtain a tighter bound; (2.25) follows from $\min\{1, x\} \leq x^p$ for all $x \geq 0$ and $0 \leq p \leq 1$; (2.26) follows from the change of measure $\mathbb{E}[f(\boldsymbol{X}, \boldsymbol{Y})] = \mathbb{E}\left[f(\bar{\boldsymbol{X}}, \boldsymbol{Y}), e^{\imath(\bar{\boldsymbol{X}};\boldsymbol{Y})}\right]$ and $\mathrm{M}-1 \leq e^{n\mathrm{R}}$; (2.28) follows from (2.23), and $\bar{\boldsymbol{X}}$ is independent of $\boldsymbol{X}$ and $\boldsymbol{Y}$; (2.29) follows from setting $s = 1/(1+p)$, which can be shown to be optimal. Using Gallager's E_0-function, which is given as

$$E_0(p) = -\frac{1}{n}\log \mathbb{E}\left[e^{(1+p)\kappa\left(\frac{1}{1+p},\boldsymbol{Y}\right)}\right] \tag{2.30}$$

we obtain the Gallager's random coding bound as

$$\epsilon \leq \min_{0\leq p\leq 1} e^{-n(E_0(p)-p\mathrm{R})} \tag{2.31}$$

Gallager's random coding bound is tight when n is large but loose when n is small [16], which makes this bound a poor choice to benchmark URLLC systems.

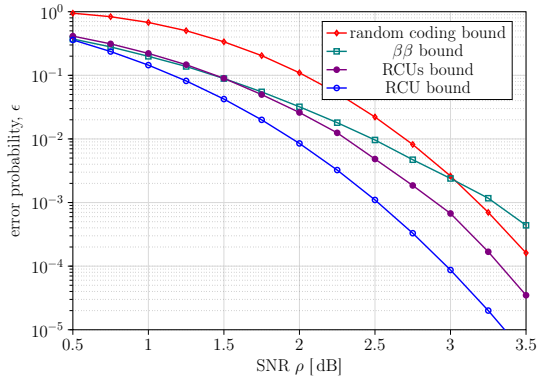

Figure 2.1: Error probability as a function of SNR. Here $n = 128$ and R = 0.5 BPCU; $Q_Y = P_Y$ for beta-beta bound and s is optimized for RCUs bound.

2.3 Results and Discussion

In Fig. 2.1 we have compared Gallager's random coding bound, the beta-beta bound, the RCU and the RCUs bounds for a biAWGN channel R = 0.5 bits per channel use (BPCU), and $n = 128$. We observe that the RCU bound is significantly tighter than all other bounds, but it is also the most complex in terms of computational complexity. The RCUs bound is tighter than the beta-beta bound and Gallager bound, and its computational complexity is significantly lower than that of the RCU bound.

CHAPTER 3

System Models

In this chapter, we review some of the practically relevant system models that we have used in this work. In particular, we introduce the block fading channel model, which serves as the basis for the remaining models and is used in Paper A, Paper B, Paper C and Chapter 4; block fading channel with imperfect synchronization, which models imperfect timing synchronization in addition to the block fading channel and is used in Paper B; massive MIMO network, which is used in Paper A.

3.1 Block-Fading Channel

3.1.1 Propagation Model

The propagation environment is crucial for the development of a wireless communication system. Since using an exact model of the environment is impractical, simplified channel models are used in the analysis. There is a trade-off between the accuracy and simplicity of such channel models. To avoid using a model that is too specific to a particular environment, the statistics of the propagation environment are usually described with several simplifying

assumptions. In the following, we present the simplifying assumptions that enables the use of block-fading channel model.

Electromagnetic waves get reflected, refracted, and diffracted in the environment. This interaction may separate the wave into so-called multipath components that arrive at the receiver with a potentially different delay, amplitude, phase, and angle. The impact of the channel is mathematically described as [22, Ch. 6.3.1]

$$y(t) = \int_{-\infty}^{\infty} x(t-\tau)h(t,\tau)\,\mathrm{d}\tau \tag{3.1}$$

where $x(t)$ and $y(t)$ are the complex signal transmitted and received at time t by the transmitter and receiver respectively, and $h(t,\tau)$ is the time-varying impulse response of the channel. Here, τ can be considered as the time delay between the transmission and reception of the signal. In fact, $h(t,\tau)$ depends on the multipath components and the radiation patterns of the transmit and receive antennas.

The delay, angle and number of multipath components may change over time due to the changing propagation environment. However, we assume that the change in these variables over time is slow enough in comparison to the packet duration such that they are considered time invariant. We model $h(t,\tau)$ as a realization of the random process $H(t,\tau)$. If the channel is wide-sense stationary (WSS) and scattering is uncorrelated, i.e., if we assume that $H(t,\tau)$ is WSS in t and uncorrelated in τ, then one can show that the time-frequency response $H(t,f)$, which is the Fourier transform of $H(t,\tau)$, is WSS in both t and f. Hence, the correlation function $R_H(t_1,t_2,f_1,f_2) = E[H^*(t_1,f_1)H(t_2,f_2)]$ depends only on $\Delta t = t_2 - t_1$ and $\Delta f = f_2 - f_1$ as we shall refer as $R_H(\Delta t, \Delta f)$.

Multipath components take different routes to the recipient, which is partly characterized by the delay τ. This leads to a time dispersion at the receiver due to multipath components with different delays. In the frequency domain, multipath components with different delays correspond to different phase shifts of the same signal, therefore, the combination of these components can lead to interference. If the interference is destructive, the output power of the channel can be significantly reduced. Since the phase shifts vary with frequency, the same multipath components can cause constructive interference at one frequency and destructive interference at another, typically known as

frequency selectivity. A common simplifying assumption is that dispersion is uncorrelated, meaning that multipath components with different delays do not correlate with each other.

A stochastic description of the time-varying impulse response requires the joint probability density function (pdf) of the complex amplitudes of the multipath components at all times, delays and angles. This is of course generally not available, and a common simplification is to use an autocorrelation function as an approximate description of the impulse response instead.

The coherence time T_c indicates the rate at which the channel changes over time and is defined as the duration during which the channel remains statistically correlated. It is given as [22, Ch.6.5.4]

$$\mathrm{T}_c = \max\left\{\Delta t : \frac{|R_H(\Delta t, 0)|}{R_H(0,0)} = \alpha\right\} \tag{3.2}$$

for some $\alpha \in [0, 1]$. Similarly, the coherence bandwidth B_c in the frequency domain indicates how fast the channel changes with frequency. It is defined as the maximum frequency spacing over which the channel remains correlated. It is given by

$$\mathrm{B}_c = \max\left\{\Delta f : \frac{|R_H(0, \Delta f)|}{R_H(0,0)} = \alpha\right\}. \tag{3.3}$$

where $\alpha \in [0, 1]$ as in (3.2). We assume that α is chosen large enough such that the channel remains the same for the time duration T_c and the frequency interval B_c. Such time-frequency blocks with the time duration T_c and the bandwidth B_c are usually referred to as coherence blocks, also known as fading blocks.

Assume that the signal bandwitdh B satisfy $\mathrm{B} \ll \mathrm{B}_c$ and we choose the duration of the transmitted signal T to be $\mathrm{T} \ll \mathrm{T}_c$. Then, we may exploit frequency diversity by transmitting the signal over different fading blocks with independent fading gains, and fading gains stay same during a transmission as $H = H(t, \tau)$.

3.1.2 Pilot-Assisted Transmission and Channel Estimation

We begin by examining a memoryless block fading channel within single-input single-output (SISO) setting. In this model, the channel remains unchanged during the transmission of a block containing n_c channel uses, but it varies

independently from one block to the next. Each transmitted packet spans n_b of these fading blocks. Consequently, a packet contains $n = n_\mathrm{b} n_\mathrm{c}$ symbols that are complex-valued. Within each block, the first n_p symbols are pilot symbols known to the receiver, while the remaining $n_\mathrm{d} = n_\mathrm{c} - n_\mathrm{p}$ symbols are data symbols.

The pilot symbols are used to estimate the fading channel within their respective blocks. The input-output relationship during the pilot transmission phase within the block $\ell = 1, \ldots, n_\mathrm{b}$ is represented as follows:

$$\boldsymbol{Y}_\ell^{(p)} = H_\ell \mathbf{x}_\ell^{(p)} + \boldsymbol{W}_\ell^{(p)}. \tag{3.4}$$

Here, $\mathbf{x}_\ell^{(p)}$ stands for the deterministic n_p-dimensional vector of pilot symbols, which we assume to obey the power constraint $\|\mathbf{x}_\ell^{(p)}\|^2 = \rho n_\mathrm{p}$, where ρ denotes the average transmission power per symbol. Moreover, H_ℓ denotes the scalar random complex fading channel gain and $\boldsymbol{W}_\ell^{(p)}$ denotes the n_p-dimensional additive noise vector. We assume that the entries of $\boldsymbol{W}_\ell^{(p)}$ are i.i.d. $\mathcal{CN}(0, \sigma_\ell^2)$.

The receiver uses the received vector $\boldsymbol{Y}_\ell^{(p)}$ and the pilot sequence $\mathbf{x}_\ell^{(p)}$ to estimate the channel H_ℓ, resulting in an estimate $\hat{H}_\ell$. It is important to note that we have not defined the fading distribution nor the specific algorithm that the receiver uses to estimate the fading channel. The error probability bounds presented in this section apply to any fading distribution and any channel estimation algorithm.

Within each block, the pilot-transmission phase is followed by a data-transmission phase, which involves the transmission of n_d symbols per block, i.e., a total of $n_\mathrm{b} n_\mathrm{d}$ symbols. Conditioned on the data, the input-output relation for the ℓth block in the data phase is given by

$$\boldsymbol{Y}_\ell = H_\ell \mathbf{x}_\ell + \boldsymbol{W}_\ell. \tag{3.5}$$

This setup is illustrated in Fig. 3.1 for an orthogonal frequency-division multiplexing (OFDM) scheme where a codeword with length $n = 108$ is transmitted over $n_\mathrm{b} = 3$ fading blocks, each block contains 3 resource blocks and each resource block contains one OFDM symbol spanning 12 subcarrier [15].

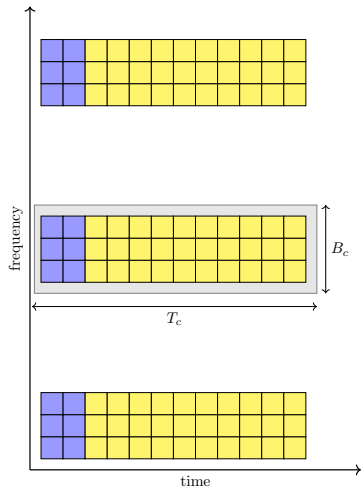

Figure 3.1: Pilot-assisted transmission over a block-fading channel for $n_\mathrm{b} = 3$, $n_\mathrm{c} = 36$, $n_\mathrm{p} = 6$. Here, blue squares corresponds to the pilot symbols and yellow squares corresponds to the data symbols.

3.1.3 Mismatched Decoding

To perform decoding, the receiver seeks the codeword in the codebook that is closest to the received signal, once each part of the codeword corresponding to a different fading block is scaled by the available channel estimate. Mathematically, given the received vector $[\mathbf{y}_1, \ldots, \mathbf{y}_{n_\mathrm{b}}]^T$ and the channel estimates $\{\hat{h}_1, \ldots, \hat{h}_{n_\mathrm{b}}\}$, the decoded codeword $\hat{\mathbf{x}} = [\hat{\mathbf{x}}_1, \ldots, \hat{\mathbf{x}}_{n_\mathrm{b}}]^T$ is determined as

$$\hat{\mathbf{x}} = \underset{\bar{\mathbf{x}}=[\bar{\mathbf{x}}_1,\ldots,\bar{\mathbf{x}}_{n_\mathrm{b}}]\in\mathcal{C}}{\arg\min} \sum_{\ell=1}^{n_\mathrm{b}} \|\mathbf{y}_\ell - \hat{h}_\ell \bar{\mathbf{x}}_\ell\|^2. \tag{3.6}$$

This decoder, which is known as mismatched scaled nearest-neighbor (SNN) decoder, coincides with the ML decoder only when the receiver has perfect channel-state information, i.e., $\hat{h}_\ell = h_\ell$ for $\ell = 1, \ldots, n_\mathrm{b}$. The attractive feature of this decoder is that information-theoretic bounds on its error probability can be approached in practice using good channel codes for the nonfading additive white Gaussian noise (AWGN) channel [23].

3.2 Imperfect Timing Synchronization

Timing synchronization refers to the process by which wireless communication systems align the timing of their operations to ensure accurate and coherent data exchange. This alignment is important to minimize interference, optimize spectral efficiency and ensure the integrity of transmitted data. In essence, timing synchronization is critical to the smooth operation of wireless networks and has a direct impact on their performance and reliability.

Many studies in the literature concerning timing synchronization consider clock synchronization for different system models (see [24], [25] for a review on this topic). However, we consider the problem of the receiver detecting the time of packet start. The imperfect time synchronization in this case can be caused by the communication channel or clock skew. At the waveform level, this problem is usually solved using pilot symbols, a solution that is at least 40 years old [26] and is still used today [27]. Mathematically, the continuous-time signal $x^{(p)}(t)$ for a fading block is defined as

$$x^{(p)}(t) = \sum_{k=1}^{n_{\mathrm{p}}} x_k^{(p)} s_{t_{\mathrm{p}}}(t - (k-1)t_{\mathrm{p}}) \tag{3.7}$$

where $s_{t_{\mathrm{p}}}(t)$ is a rectangular pulse given as

$$s_{t_{\mathrm{p}}}(t) = \begin{cases} \frac{1}{\sqrt{t_{\mathrm{p}}}}, & t \in [0, t_{\mathrm{p}}) \\ 0, & \text{otherwise .} \end{cases} \tag{3.8}$$

The data symbols in the ℓth block are sent after the pilot symbols via the continuous-time signal

$$x_\ell^{(d)}(t) = \sum_{k=1}^{n_s} x_{k,\ell}^{(d)} s_{t_{\mathrm{p}}}(t - (k-1)t_{\mathrm{p}} - n_p t_p). \tag{3.9}$$

The total continuous-time signal corresponding to the ℓth subpacket is put through a flat-fading channel to obtain the received continuous signal

$$Y_\ell(t) = H_\ell\Big(x^{(p)}(t - D_\ell) + x_\ell^{(d)}(t - D_\ell)\Big) + Z_\ell(t) \tag{3.10}$$

where H_ℓ denotes the scalar random fading complex channel gain for the ℓth

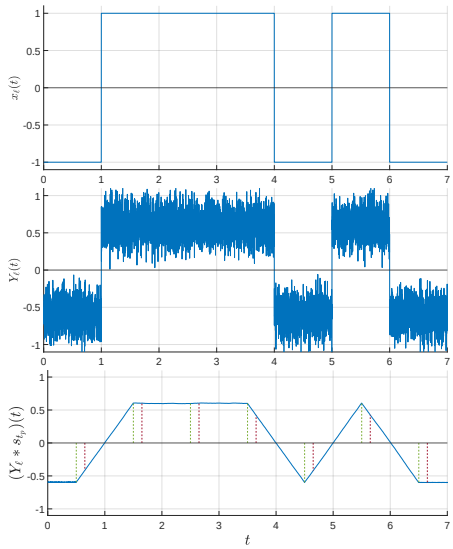

Figure 3.2: An example of transmission and detection. Here, we only consider real symbol transmission and to plot $Y_\ell(t)$, we oversampled the received signal after passed it through an anti-aliasing filter; $H_\ell = 0.6$, $\sigma_\ell^2 = 1$, $\rho = 15\,\text{dB}$; green and red dashed lines shows the sampling at the output of the matched filter when time synchronization is perfect and imperfect, respectively.

fading block, D_ℓ is the time delay for the ℓth fading block, which we assume to be uniform in $[0, d_{\max}]$, and $Z_1(t), \ldots, Z_{n_\text{b}}(t)$ are independent white complex Gaussian processes with power spectral density N_0.

Once a packet has been transmitted over a noisy channel and received by the receiver, it is passed through a matched filter. Due to imperfect time synchronization, the output of the matched filter cannot be sampled at the moment when a peak would occur if the signal were not distorted by the channel, as illustrated by a simple example in Fig. 3.2. Such a shift in sampling causes intersymbol interference, which should be taken into account when analyzing the performance of a communication system. In conventional systems, it can be assumed that the pilot sequences are long enough to eliminate this effect. With URLLC, however, the use of pilot symbols is heavily penalized due to the use of short packets. Therefore, it is crucial to evaluate the impact of imperfect time synchronization in addition to the impact of using pilot symbols. In Paper B, we analyzed the effects of imperfect timing synchronization

in detail and showed that for a block-fading channel with fading blocks that have the same delay, the pilot symbols used for channel estimation can be used for both coarse and fine timing synchronization.

3.3 Massive MIMO Network

A massive MIMO network consists of L cells that work with a synchronous time division duplex (TDD) protocol, in which an uplink transmission is usually followed by a downlink transmission [28, Ch. 2.1]. Each cell is served by a base station (BS) equipped with m antennas, where $\mathrm{m} \gg 1$ and each BS communicates with k single antenna users with $\mathrm{m/k} > 1$. Massive MIMO technology enables the design of a highly spectrally efficient cellular network [29], [30] and has been implemented in real-time massive MIMO testbeds [31].

Practical channels are usually spatially correlated because antennas have non-uniform radiation patterns, and certain spatial directions are more likely to carry strong signals from the transmitter to the receiver due to the physical propagation environment. This spatial channel correlation is particularly important for large arrays, as these have better spatial resolution in relation to the number of scattering clusters. Consequently, the correlated Rayleigh fading channel model is given as $\boldsymbol{H}^j_{\ell,i,k} \sim \mathcal{CN}(\mathbf{0}_{\mathrm{M}}, \mathsf{R}^j_{i,k})$ where $\boldsymbol{H}^j_{\ell,i,k} \in \mathbb{C}^{\mathrm{M}}$ the channel gain vector within the ℓth fading block between user k in cell i and the BS in cell j, and $\mathsf{R}^j_{i,k}$ is the spatial correlation matrix which shows the propagation effects like antenna gains and radiation patterns at the transmitted and receiver [28, Ch. 2.2]. The normalized trace $\beta^j_{i,k} = \mathrm{tr}(\mathsf{R}^j_{i,k})/\mathrm{M}$ determines the average channel gain between user k in cell i and the BS in cell j which can be modelled in dB scale as

$$\beta^j_{i,k} = \lambda - 10\alpha \log_{10}\left(d^j_{i,k}\right) \tag{3.11}$$

where λ the median channel gain at a reference distance of one kilometer, α is the pathloss exponent and $d^j_{i,k}$ is the distance between the BS in the cell j and the user k in cell i in kilometer. The parameters α and λ either needs to be measured or may be computed using the propagation model (see [32] for a survey on propagation models for mobile communication) and how the spatial correlation matrix can be evaluated is described in [28, Ch. 2.6].

Two important features that improve the spectral efficiency of massive

MIMO channels are channel hardening and favorable propagation.

Channel hardening describes the phenomenon that a fading channel behaves almost deterministically due to spatial diversity, which means the SNR is not affected by small-scale fading (the Rayleigh fading in this case). Mathematically, this is described as

$$\frac{\|\boldsymbol{H}^j_{\ell,j,k}\|^2}{\mathbb{E}\left[\|\boldsymbol{H}^j_{\ell,j,k}\|^2\right]} \to 1 \tag{3.12}$$

as $\mathrm{m} \to \infty$.

Favorable propagation stands for the phenomenon that the channel directions becoming orthogonal as $\mathrm{m} \to \infty$. It can be mathematically described as

$$\left\langle \frac{\boldsymbol{H}^j_{\ell,i,k_1}}{\sqrt{\mathbb{E}\left[\|\boldsymbol{H}^j_{\ell,i,k_1}\|^2\right]}}, \frac{\boldsymbol{H}^j_{\ell,j,k_2}}{\sqrt{\mathbb{E}\left[\|\boldsymbol{H}^j_{\ell,j,k_2}\|^2\right]}} \right\rangle \to 0 \tag{3.13}$$

as $\mathrm{m} \to \infty$. Favorable propagation makes it easier for the base station to combat interference between users and enables the use of linear combining and precoding.

The idea of block-fading, which is explained in the Sec. 3.1, can be used in massive MIMO transmission by using multi-carrier modulation methods such as conventional OFDM. In this case, the subcarriers are transmitted via coherence blocks, whereas the channel realizations between the blocks are independent. This allows the use of analysis tools developed for block-fading channels for massive MIMO settings, as done for example in [33] and Paper A.

CHAPTER 4

Asymptotic Expansions of RCUs Bound

In this chapter, we will present the saddlepoint expansion of the tail probability of the sum of i.i.d. random variables used in Paper A, Paper B and Paper C. We will then show that, through a first-order asymptotic expansion of RCUs bound, how one may capture both the outage and the ergodic capacity.

4.1 Asymptotic Expansions of a Tail Probability

RCU and RCUs bounds require the evaluation of tail probabilities, which are generally not available in closed form, and their direct evaluation (e.g. via Monte-Carlo simulations) has a high computational complexity. Tail probabilities instead can be approximated as $n \to \infty$ using probabilistic results. One of the most fundamental results of probability theory that is relevant here is the law of large numbers, which is given as follows [34, Ch. VII.8]: Let $Z_1, \ldots, Z_n$ be i.i.d. real-valued continuous random variables with finite mean $\mu = \mathbb{E}[Z_j]$, finite variance $\sigma^2 = \mathbb{V}\mathrm{ar}[Z_j]$ and finite third absolute central

moment $\theta = \mathbb{E}\left[|Z_j|^3\right]$ for $j = 1, \ldots, n$. Then,

$$\lim_{n\to\infty} \mathbb{P}\left[\frac{1}{n}\sum_{j=1}^{n} Z_j \geq \gamma\right] = \mathbb{1}\{\gamma \leq \mu\} \tag{4.1}$$

which shows that $\sum_{j=1}^{n} Z_j$ concentrates around its mean as n tends to infinity. The law of large numbers provides no information about the rate at which the sum of the random variables converges to its mean. We, however, would like to characterize how quickly the sum approaches its asymptotic limit. The behavior of the sum of random variables may be refined using central limit theorem (CLT). It provides a characterization of the error that arises when approximating the tail probability of an appropriate normalized sum of independent random variables with the tail probability of a standard normal random variable, which we shall refer as Gaussian approximation. The Berry-Esseen CLT for every $\lambda \in \mathbb{R}$ is given as [34, Chapter XVI]

$$\left|\mathbb{P}\left(\frac{\sum_{j=1}^{n}(Z_j - \mu)}{\sqrt{\sum_{j=1}^{n}\sigma^2}} \geq \lambda\right) - Q(\lambda)\right| \leq 6\frac{\sum_{j=1}^{n}\theta}{\left(\sum_{j=1}^{n}\sigma^2\right)^{3/2}}. \tag{4.2}$$

Note that the right hand side (r.h.s.) of (4.2) does not depend on λ, which indicates that Gaussian approximation is accurate when λ is close to μ, but, depending on n, may be inaccurate otherwise. To tackle this problem, we can utilize an exponentially tilted random variable and control its moments such that its mean is matched with λ, evaluate the desired tail probability via applying Gaussian approximation to the tilted random variable, and characterizes the relative error that results from approximating the exponentially tilted distribution with a normal distribution.. This method is called saddlepoint expansion which we shall review next.

The proof of the saddlepoint expansion is provided in detail in [35, Ch. 6], which is mostly based on [34, Ch. XVI.4, Th. 1] and [36, App. E] (see also [37, App. I.A]). For completeness, we will present the proof in [35, Ch. 6] here, using our notation with some minor changes for clarity. In the proof, for notational brevity, we use upper-case letters in a special font to denote the distribution of a continuous random variable (i.e., Y).

As before, let $Z_1, \ldots, Z_n$ be i.i.d. real-valued continuous random variables.

Additionally, let $m(\zeta)$ and $\kappa(\zeta)$ denote the MGF and CGF of Z_j, respectively, which are given as follows:

$$m(\zeta) = \mathbb{E}\left[e^{\zeta Z_j}\right] \tag{4.3}$$

$$\kappa(\zeta) = \log(m(\zeta)). \tag{4.4}$$

The saddlepoint expansion of the tail probability $\mathbb{P}[\sum_{j=1}^{n} Z_j > \lambda]$ for $\lambda \in \mathbb{R}$ is given as follows: Let ζ be the solution of $n\kappa(\zeta) = \lambda$ and we assume that

$$\sup_{\underline{\zeta}<\zeta<\overline{\zeta}} \left| \frac{\mathrm{d}^3}{\mathrm{d}\zeta^3} m(\zeta) \right| < \infty. \tag{4.5}$$

For $\zeta \in [0, \overline{\zeta})$

$$\mathbb{P}\left[\sum_{j=1}^{n} Z_j > \lambda\right] = e^{n\left(\kappa(\zeta)-\zeta\kappa'(\zeta)\right)} \left[\Psi_{n,\zeta}(\zeta) + o\left(\frac{1}{\sqrt{n}}\right)\right] \tag{4.6}$$

and for $\zeta \in (\underline{\zeta}, 0)$

$$\mathbb{P}\left[\sum_{j=1}^{n} Z_j > \lambda\right] = 1 - e^{n\left(\kappa(\zeta)-\zeta\kappa'(\zeta)\right)} \left[\Psi_{n,\zeta}(-\zeta) + o\left(\frac{1}{\sqrt{n}}\right)\right] \tag{4.7}$$

where $\kappa'(\zeta)$ and $\kappa''(\zeta)$ denote the first and second derivatives of $\kappa(\zeta)$, respectively, and

$$\Psi_{b,\zeta}(u) = e^{b\frac{u^2}{2}\kappa''(\zeta)} Q\left(u\sqrt{b\kappa''(\zeta)}\right). \tag{4.8}$$

We begin the proof with $\mathbb{P}\left[\sum_{j=1}^{n} Z_j > \lambda\right]$ for $\zeta \in [0, \overline{\zeta})$. Let $Y_j = Z_j - \tilde{\lambda}$, where $\tilde{\lambda} = \lambda/n$ and let Y denote the distribution of Y_j. Then, the CGF of Y_j is given by $\tilde{\gamma}(\zeta) = \kappa(\zeta) - \zeta\tilde{\lambda}$. Let the tilted random variable V_j have the

distribution V_ζ as[2]

$$\mathsf{V}_\zeta(x) = e^{-\tilde{\gamma}(\zeta)} \int_{-\infty}^{x} e^{\zeta t} \, \mathrm{d}\mathsf{Y}(t) \tag{4.9}$$

$$= e^{-\kappa(\zeta)+\zeta\tilde{\lambda}} \int_{-\infty}^{x} e^{\zeta t} \, \mathrm{d}\mathsf{Y}(t). \tag{4.10}$$

Also let $\psi_\zeta(\tau)$ denote the MGF of the tilted random variable V_j given as

$$\psi_\zeta(\tau) = \int_{-\infty}^{\infty} e^{\tau x} \, \mathrm{d}\mathsf{V}_\zeta(x) \tag{4.11}$$

$$= \int_{-\infty}^{\infty} e^{\tau x - \kappa(\zeta)+\zeta\tilde{\lambda}+\zeta x} \, \mathrm{d}\mathsf{Y}(x) \tag{4.12}$$

$$= e^{-\kappa(\zeta)+\zeta\tilde{\lambda}} \int_{-\infty}^{\infty} e^{(\tau+\zeta)x} \, \mathrm{d}\mathsf{Y}(x) \tag{4.13}$$

$$= e^{-\kappa(\zeta)+\zeta\tilde{\lambda}} \, \mathbb{E}\left[e^{(\tau+\zeta)(Z_j-\tilde{\lambda})}\right] \tag{4.14}$$

$$= e^{-\kappa(\zeta)} \, \mathbb{E}\left[e^{(\tau+\zeta)Z_j}\right] e^{-\tau\tilde{\lambda}} \tag{4.15}$$

$$= \frac{m(\tau+\zeta)}{m(\zeta)} e^{-\tau\tilde{\lambda}}. \tag{4.16}$$

Since $\mathbb{E}[V_j] = \psi'_\zeta(0)$ where the derivative is taken with respect to τ, it follows that

$$\mathbb{E}[V_j] = \psi'_\zeta(0) \tag{4.17}$$

$$= \left(\frac{m'(\tau+\zeta)}{m(\zeta)} e^{-\tau\tilde{\lambda}} - \tilde{\lambda}\frac{m(\tau+\zeta)}{m(\zeta)} e^{-\tau\tilde{\lambda}}\right)\Bigg|_{\tau=0} \tag{4.18}$$

$$= \frac{m'(\zeta)}{m(\zeta)} - \tilde{\lambda} \tag{4.19}$$

$$= \kappa'(\zeta) - \tilde{\lambda}. \tag{4.20}$$

Similarly, $\mathbb{V}\mathrm{ar}[V_j] = \mathbb{E}\left[V_j^2\right] - \mathbb{E}[V_j]^2 = \kappa''(\zeta)$.

Let Y^{*n} be the distribution of $\sum_{j=1}^{n} Y_j$ and V_ζ^{*n} be the distribution of

[2]Note that $\mathbb{P}[Y_j > \lambda] = \int_\lambda^\infty \mathrm{d}\mathsf{Y}(t)$

$\sum_{j=1}^{n} V_j$. Proceeding as in (4.9), we have

$$\mathsf{V}_{\zeta}^{*n}(x) = e^{-n\tilde{\gamma}(\zeta)} \int_{-\infty}^{x} e^{\zeta t} \, \mathrm{d}\mathsf{Y}^{*n}(t) \tag{4.21}$$

$$= e^{-n\kappa(\zeta)+\zeta\tilde{\lambda}} \int_{-\infty}^{x} e^{\zeta t} \, \mathrm{d}\mathsf{Y}^{*n}(t). \tag{4.22}$$

Since $\mathbb{P}[\sum_{j=1}^{n} Z_j > \lambda] = 1 - \mathsf{Y}^{*n}(\lambda)$, we require an expression for $1 - \mathsf{Y}^{*n}(\lambda)$ as a function of $\mathsf{V}_{\zeta}^{*n}(x)$. This can be obtained by inverting (4.22) as

$$\mathbb{P}\left[\sum_{j=1}^{n} Z_j > \lambda\right] = e^{n\kappa(\zeta)-\zeta\lambda} \int_{0}^{\infty} e^{-\zeta y} \, \mathrm{d}\mathsf{V}_{\zeta}^{*n}(y). \tag{4.23}$$

We next set ζ such that $n\kappa'(\zeta) = \lambda$, which ensures that the distribution V_{ζ}^{*n} has zero mean, and replace the distribution V_{ζ}^{*n} by the zero-mean normal distribution with variance $n\kappa''(\zeta)$, which we denote by Φ_{ζ}. To do so, let $a_{\zeta}(\lambda)$ be the expression after replacing V_{ζ}^{*n} by Φ_{ζ} in (4.23). Then,

$$a_{\zeta}(\lambda) = e^{n\kappa(\zeta)-\zeta\lambda} \int_{0}^{\infty} e^{-\zeta y} \, \mathrm{d}\Phi_{\zeta}(y) \tag{4.24}$$

$$= \frac{e^{n\left[\kappa(\zeta)-\zeta\kappa'(\zeta)\right]}}{\sqrt{2\pi n\kappa''(\zeta)}} \int_{0}^{\infty} e^{-\zeta y} e^{-\frac{y^2}{2n\kappa''(\zeta)}} \, \mathrm{d}y \tag{4.25}$$

$$= \frac{e^{n\left[\kappa(\zeta)-\zeta\kappa'(\zeta)\right]}}{\sqrt{2\pi}} \int_{0}^{\infty} e^{-\zeta t\sqrt{n\kappa''(\zeta)}} e^{-\frac{t^2}{2}} \, \mathrm{d}t \tag{4.26}$$

$$= \frac{e^{n\left[\kappa(\zeta)-\zeta\kappa'(\zeta)+\frac{\zeta^2}{2}\kappa''(\zeta)\right]}}{\sqrt{2\pi}} \int_{0}^{\infty} e^{-\frac{1}{2}\left(t+\zeta\sqrt{n\kappa''(\zeta)}\right)^2} \, \mathrm{d}t \tag{4.27}$$

$$= \frac{e^{n\left[\kappa(\zeta)-\zeta\kappa'(\zeta)+\frac{\zeta^2}{2}\kappa''(\zeta)\right]}}{\sqrt{2\pi}} \int_{\zeta\sqrt{n\kappa''(\zeta)}}^{\infty} e^{-\frac{x^2}{2}} \, \mathrm{d}x \tag{4.28}$$

$$= e^{n\left[\kappa(\zeta)-\zeta\kappa'(\zeta)+\frac{\zeta^2}{2}\kappa''(\zeta)\right]} Q\Big(\zeta\sqrt{n\kappa''(\zeta)}\Big) \tag{4.29}$$

$$= e^{n\left[\kappa(\zeta)-\zeta\kappa'(\zeta)\right]} \Psi_{n,\zeta}(\zeta) \tag{4.30}$$

where the (4.26) follows from the change of variable $t = y/\sqrt{n\kappa''(\zeta)}$, and

(4.28) follows from the change of variable $x = t + \zeta\sqrt{n\kappa''(\zeta)}$.

We next asses the error incurred by replacing V_ζ^{*n} with Φ_ζ in (4.23), which is given by

$$\begin{aligned}
&e^{n\kappa(\zeta)-\zeta\lambda}\int_0^\infty e^{-\zeta y}\,\mathrm{d}\mathsf{V}_\zeta^{*n}(y) - a_\zeta(\lambda) \\
&\quad= e^{n[\kappa(\zeta)-\zeta\kappa'(\zeta)]}\left[\frac{\kappa'''(\zeta)}{6\kappa''(\zeta)^{3/2}\sqrt{n}}\left(-\frac{1}{\sqrt{2\pi}}+\frac{\zeta^2 n\kappa''(\zeta)}{\sqrt{2\pi}}\right.\right. \\
&\qquad\left.\left.-\zeta^3\kappa''(\zeta)^{3/2}n^{3/2}\Psi_{n,\zeta}(\zeta)\right)+o\left(\frac{1}{\sqrt{n}}\right)\right] \qquad (4.31)\\
&\quad= e^{n[\kappa(\zeta)-\zeta\kappa'(\zeta)]}\left[\frac{K(\zeta,\zeta,n)}{\sqrt{n}}+o\left(\frac{1}{\sqrt{n}}\right)\right] \qquad (4.32)\\
&\quad= e^{n[\kappa(\zeta)-\zeta\kappa'(\zeta)]}o\left(\frac{1}{\sqrt{n}}\right) \qquad (4.33)
\end{aligned}$$

where

$$K(u,\zeta,n) = \frac{\kappa'''(\zeta)}{6\kappa''(\zeta)^{3/2}}\left(-\frac{1}{\sqrt{2\pi}}+\frac{u^2 n\kappa''(\zeta)}{\sqrt{2\pi}}-u^3(\kappa''(\zeta)n)^{3/2}\Psi_{n,\zeta}(u)\right). \qquad (4.34)$$

Here, (4.31) follows from integration by parts [34, Ch. V.6, Eq.(6.1)] and from [34, Ch. XVI.4, Th. 1], which requires the condition (4.5) to be met, and (4.33) follows since $K(u,\zeta,n) = O(1/n)$ [35, App. B.7] which indicates $K(u,\zeta,n)/\sqrt{n} = o(1/\sqrt{n})$. By combining (4.30) and (4.33), we establish (4.6).

We next consider the case $\zeta \in (\underline{\zeta}, 0)$. Note that $\mathbb{P}\left[\sum_{j=1}^n Z_j < \lambda\right]$ is the tail of the distribution. To utilize this, we note that

$$\mathbb{P}\left[\sum_{j=1}^n Z_j > \lambda\right] = 1 - \mathbb{P}\left[\sum_{j=1}^n Z_j < \lambda\right] \qquad (4.35)$$

and

$$\mathbb{P}\left[\sum_{j=1}^n Z_j < \lambda\right] = e^{n\kappa(\zeta)-\zeta\lambda}\int_{-\infty}^0 e^{-\zeta y}\,\mathrm{d}\mathsf{V}_\zeta^{*n}(y). \qquad (4.36)$$

We again choose ζ such that $n\kappa'(\zeta) = \lambda$, and let $\tilde{a}_\zeta(\lambda)$ be the expression after

replacing V_ζ^{*n} by Φ_ζ in (4.36). Then,

$$\tilde{a}_\zeta(\lambda) = e^{n\kappa(\zeta)-\zeta\lambda} \int_{-\infty}^{0} e^{-\zeta y}\, \mathrm{d}\Phi_\zeta(y) \tag{4.37}$$

$$= \frac{e^{n[\kappa(\zeta)-\zeta\kappa'(\zeta)]}}{\sqrt{2\pi n\kappa''(\zeta)}} \int_{-\infty}^{0} e^{-\zeta y} e^{-\frac{y^2}{2n\kappa''(\zeta)}}\, \mathrm{d}y \tag{4.38}$$

$$= \frac{e^{n[\kappa(\zeta)-\zeta\kappa'(\zeta)]}}{\sqrt{2\pi}} \int_{-\infty}^{0} e^{-\zeta t\sqrt{n\kappa''(\zeta)}} e^{-t^2/2}\, \mathrm{d}t \tag{4.39}$$

$$= \frac{e^{n\left[\kappa(\zeta)-\zeta\kappa'(\zeta)+\frac{\zeta^2}{2}\kappa''(\zeta)\right]}}{\sqrt{2\pi}} \int_{-\infty}^{0} e^{-\frac{1}{2}\left(t+\zeta\sqrt{n\kappa''(\zeta)}\right)^2}\, \mathrm{d}t \tag{4.40}$$

$$= \frac{e^{n\left[\kappa(\zeta)-\zeta\kappa'(\zeta)+\frac{\zeta^2}{2}\kappa''(\zeta)\right]}}{\sqrt{2\pi}} \int_{-\infty}^{\zeta\sqrt{n\kappa''(\zeta)}} e^{-\frac{x^2}{2}}\, \mathrm{d}x \tag{4.41}$$

$$= \frac{e^{n\left[\kappa(\zeta)-\zeta\kappa'(\zeta)+\frac{\zeta^2}{2}\kappa''(\zeta)\right]}}{\sqrt{2\pi}} \int_{-\zeta\sqrt{n\kappa''(\zeta)}}^{\infty} e^{-\frac{x^2}{2}}\, \mathrm{d}x \tag{4.42}$$

$$= e^{n\left[\kappa(\zeta)-\zeta\kappa'(\zeta)+\frac{\zeta^2}{2}\kappa''(\zeta)\right]} Q\Big(-\zeta\sqrt{n\kappa''(\zeta)}\Big) \tag{4.43}$$

$$= e^{n\left[\kappa(\zeta)-\zeta\kappa'(\zeta)\right]} \Psi_{n,\zeta}(-\zeta) \tag{4.44}$$

where (4.39) follows from the change of variable $t = y/\sqrt{n\kappa''(\zeta)}$, and (4.41) follows from the change of variable $x = t+\zeta\sqrt{n\kappa''(\zeta)}$. The error incurred due to substituting V_ζ^{*n} by Φ_ζ can be obtained by following the steps leading to

(4.33) as

$$
\begin{aligned}
& e^{n\kappa(\zeta)-\zeta\lambda}\int_{-\infty}^{0} e^{-\zeta y}\,\mathrm{d}\mathsf{V}_{\zeta}^{*n}(y)-\tilde{a}_{\zeta}(\lambda) \\
&\quad = e^{n[\kappa(\zeta)-\zeta\kappa'(\zeta)]}\left[\frac{1}{\sqrt{2\pi}}\frac{\kappa'''(\zeta)}{6\kappa''(\zeta)^{3/2}\sqrt{n}}\left(1+\int_{-\infty}^{0}\zeta\sqrt{n\kappa''(\zeta)}(1-z^2)\right.\right. \\
&\qquad \left.\left.\times\, e^{-\zeta\sqrt{\kappa''(\zeta)n}z-\frac{z^2}{2}}\,\mathrm{d}z\right)+o\left(\frac{1}{\sqrt{n}}\right)\right] && (4.45)\\
&\quad = e^{n[\kappa(\zeta)-\zeta\kappa'(\zeta)]}\left[\frac{\kappa'''(\zeta)}{6\kappa''(\zeta)^{3/2}\sqrt{n}}\left(\frac{1}{\sqrt{2\pi}}-\frac{\zeta^2\kappa''(\zeta)n}{\sqrt{2\pi}}\right.\right. \\
&\qquad \left.\left.-\,\zeta^3(\kappa''(\zeta)n)^{3/2}\Psi_{n,\zeta}(-\zeta)\right)+o\left(\frac{1}{\sqrt{n}}\right)\right] && (4.46)\\
&\quad = e^{n[\kappa(\zeta)-\zeta\kappa'(\zeta)]}\left(-\frac{K(-\zeta,\zeta,n)}{\sqrt{n}}+o\left(\frac{1}{\sqrt{n}}\right)\right) && (4.47)\\
&\quad = e^{n[\kappa(\zeta)-\zeta\kappa'(\zeta)]}o\left(\frac{1}{\sqrt{n}}\right). && (4.48)
\end{aligned}
$$

Combining (4.44) and (4.48) for $n\kappa'(\zeta)=\lambda$, we establish (4.7).

The tail probability that appears in the RCUs bound is in the form of $\mathbb{P}\Big[\sum_{j=1}^{n} Z_j \geq \lambda+\log U\Big]$. Different from (4.6) and (4.7), we now have the term $\log U$. Specifically, we again let ζ be the solution of $\lambda=n\kappa'(\zeta)$ and if $\zeta\in[0,1]$, then

$$
\mathbb{P}\left[\sum_{j=1}^{n} Z_j \geq \lambda+\log U\right] = e^{n\left[\kappa(\zeta)-\zeta\kappa'(\zeta)\right]}\left[\Psi_{n,\zeta}(\zeta)+\Psi_{n,\zeta}(1-\zeta)+o\left(\frac{1}{\sqrt{n}}\right)\right] \tag{4.49}
$$

If $\zeta\in(1,\overline{\zeta})$, then

$$
\mathbb{P}\left[\sum_{j=1}^{n} Z_j \geq \lambda+\log U\right] = e^{n\kappa(1)-\lambda}\left[\tilde{\Psi}_n(1,1)+\tilde{\Psi}_n(0,-1)+\mathcal{O}\left(\frac{1}{\sqrt{n}}\right)\right] \tag{4.50}
$$

where

$$\tilde{\Psi}_b(a_1, a_2) = e^{ba_1\left[\frac{\lambda}{b} - \kappa'(1) + \frac{\kappa''(1)}{2}\right]} Q\left(a_1\sqrt{b\kappa''(1)} - a_2\frac{b(\kappa'(1)+\lambda)}{\sqrt{b\kappa''(1)}}\right). \quad (4.51)$$

If $\zeta \in (\underline{\zeta}, 0)$, then

$$\mathbb{P}\left[\sum_{j=1}^{n} Z_j \geq \lambda + \log U\right] = 1 - e^{n\left[\kappa(\zeta) - \zeta\kappa'(\zeta)\right]} \times \left[\Psi_{n,\zeta}(-\zeta) - \Psi_{n,\zeta}(1-\zeta) + o\left(\frac{1}{\sqrt{n}}\right)\right]. \quad (4.52)$$

We start with the case $\zeta \in [0, 1]$ and expand the tail probability as

$$\begin{aligned}
&\mathbb{P}\left[\sum_{j=1}^{n} Z_j \geq \lambda + \log U\right] \\
&\quad = e^{n\kappa(\zeta)-\zeta\lambda} \int_0^1 \int_{\log u}^{\infty} e^{-\zeta y}\, \mathrm{d}\mathsf{V}_\zeta^{*n}(y)\, \mathrm{d}u && (4.53)\\
&\quad = e^{n\kappa(\zeta)-\zeta\lambda} \int_{-\infty}^{\infty} \int_0^{\min(1,e^y)} e^{-\zeta y}\, \mathrm{d}u\, \mathrm{d}\mathsf{V}_\zeta^{*n}(y) && (4.54)\\
&\quad = e^{n\kappa(\zeta)-\zeta\lambda} \int_0^{\infty} e^{-\zeta y}\, \mathrm{d}\mathsf{V}_\zeta^{*n}(y) + e^{n\kappa(\zeta)-\zeta\lambda} \int_{-\infty}^{0} e^{(1-\zeta)y}\, \mathrm{d}G_\zeta^{*n}(y). && (4.55)
\end{aligned}$$

The first term in (4.55) coincides with the $a_\zeta(\lambda)$ given in (4.30). To analyze the second term, we let $b_\zeta(\lambda)$ be the expression after V_ζ^{*n} is replaced by Φ_ζ of

the second term in (4.55). Then,

$$b_\zeta(\lambda) = e^{n\kappa(\zeta)-\zeta\lambda}\int_{-\infty}^{0} e^{(1-\zeta)y}\,\mathrm{d}\Phi_\zeta(y) \tag{4.56}$$

$$= \frac{e^{n\left[\kappa(\zeta)-\zeta\kappa'(\zeta)\right]}}{\sqrt{2\pi n\kappa''(\zeta)}}\int_{-\infty}^{0} e^{(1-\zeta)y}e^{-\frac{y^2}{2n\kappa''(\zeta)}}\,\mathrm{d}y \tag{4.57}$$

$$= \frac{e^{n\left[\kappa(\zeta)-\zeta\kappa'(\zeta)\right]}}{\sqrt{2\pi}}\int_{-\infty}^{0} e^{(1-\zeta)t\sqrt{n\kappa''(\zeta)}}e^{-\frac{t^2}{2}}\,\mathrm{d}t \tag{4.58}$$

$$= \frac{e^{n\left[\kappa(\zeta)-\zeta\kappa'(\zeta)+\frac{(1-\zeta)^2}{2}\kappa''(\zeta)\right]}}{\sqrt{2\pi}}\int_{\zeta}^{\infty} e^{-\frac{1}{2}\left(t-(1-\zeta)\sqrt{n\kappa''(\zeta)}\right)^2}\,\mathrm{d}t \tag{4.59}$$

$$= \frac{e^{n\left[\kappa(\zeta)-\zeta\kappa'(\zeta)+\frac{(1-\zeta)^2}{2}\kappa''(\zeta)\right]}}{\sqrt{2\pi}}\int_{-\infty}^{-(1-\zeta)\sqrt{n\kappa''(\zeta)}} e^{-\frac{x^2}{2}}\,\mathrm{d}x \tag{4.60}$$

$$= \frac{e^{n\left[\kappa(\zeta)-\zeta\kappa'(\zeta)+\frac{(1-\zeta)^2}{2}\kappa''(\zeta)\right]}}{\sqrt{2\pi}}\int_{(1-\zeta)\sqrt{n\kappa''(\zeta)}}^{\infty} e^{-\frac{x^2}{2}}\,\mathrm{d}x \tag{4.61}$$

$$= e^{n\left[\kappa(\zeta)-\zeta\kappa'(\zeta)+\frac{(1-\zeta)^2}{2}\kappa''(\zeta)\right]}Q\Big((1-\zeta)\sqrt{n\kappa''(\zeta)}\Big) \tag{4.62}$$

$$= e^{n\left[\kappa(\zeta)-\zeta\kappa'(\zeta)\right]}\Psi_{n,\zeta}(1-\zeta) \tag{4.63}$$

where (4.58) follows from the change of variable $t = y/\sqrt{n\kappa''(\zeta)}$ and the (4.60) follows from the change of variable $x = t-(1-\zeta)\sqrt{n\kappa''(\zeta)}$. We next evaluate

the error incurred by substituting V_ζ^{*n} with Φ_ζ as

$$
\begin{aligned}
&e^{n\kappa(\zeta)-\zeta\lambda}\int_{-\infty}^{0} e^{(1-\zeta)y}\,\mathrm{d}\mathsf{V}_\zeta^{*n}(y) - b_\zeta(\lambda) \\
&\quad= e^{n(\kappa(\zeta)-\zeta\kappa'(\zeta))}\left[\frac{1}{\sqrt{2\pi}}\frac{\kappa'''(\zeta)}{6\kappa''(\zeta)^{3/2}\sqrt{n}}\right. \\
&\qquad\left.\times\left(1-\int_{-\infty}^{0}(1-\zeta)\sqrt{n\kappa''(\zeta)}(1-z^2)e^{(1-\zeta)\sqrt{n\kappa''(\zeta)}z-\frac{z^2}{2}}\,\mathrm{d}z\right)\right] \qquad (4.64)\\
&\quad= e^{n(\kappa(\zeta)-\zeta\kappa'(\zeta))}\left[\frac{\kappa'''(\zeta)}{6\kappa''(\zeta)^{3/2}\sqrt{n}}\left(\frac{1}{\sqrt{2\pi}}-\frac{(1-\tau)^2 n\kappa''(\zeta)}{\sqrt{2\pi}}\right.\right. \\
&\qquad\left.\left.+(1-\zeta)^3(n\kappa''(\zeta))^{3/2}\Psi_{n,\zeta}(1-\zeta)\right)+o\left(\frac{1}{\sqrt{n}}\right)\right] \qquad (4.65)
\end{aligned}
$$

where (4.64) follows from integration by parts and from [34, Ch. XVI.4, Th. 1]. By combining (4.65), (4.63), (4.30) and (4.33), we get (4.49).

When $\zeta \in (1, \overline{\zeta})$, the term $e^{(1-\zeta)y}$ diverges as $y \to -\infty$. Then, we cannot follow the steps (4.64)-(4.65) to get the error incurred by substituting V_ζ^{*n} with Φ_ζ for $\zeta \in (1, \overline{\zeta})$. To analyse this case, we instead set $\zeta = 1$ and as a consequence, we now need to analyze the error incurred by replacing V_ζ^{*n} with the normal distribution that has the mean $n\kappa'(\zeta) - \lambda$ and the variance $n\kappa''(\zeta)$. After following the steps leading to (4.63) and (4.33) for the two terms in (4.55) for $\zeta = 1$, we get (4.50).

It only remains to show the saddlepoint expansion of the tail probability $P[\sum_{j=1}^n Z_j \geq \lambda + \log U] = 1 - P[\sum_{j=1}^n Z_j < \lambda + \log U]$ when $\zeta \in (\underline{\zeta}, 0)$. In this case, it follows that

$$
\begin{aligned}
&\mathbb{P}\left[\sum_{j=1}^{n} Z_j < \lambda + \log U\right] && (4.66)\\
&= e^{n\kappa(\zeta)-\zeta\lambda}\int_0^1\int_{-\infty}^{\log u} e^{-\zeta y}\,\mathrm{d}\mathsf{V}_\zeta^{*n}(y)\,\mathrm{d}u && (4.67)\\
&= e^{n\kappa(\zeta)-\zeta\lambda}\int_{-\infty}^{0}\int_{e^y}^{1} e^{-\zeta y}\,\mathrm{d}u\,\mathrm{d}\mathsf{V}_\zeta^{*n}(y) && (4.68)\\
&= e^{n\kappa(\zeta)-\zeta\lambda}\left(\int_{-\infty}^{0} e^{-\zeta y}\,\mathrm{d}\mathsf{V}_\zeta^{*n}(y) + \int_{-\infty}^{0} e^{(1-\zeta)y}\,\mathrm{d}\mathsf{V}_\zeta^{*n}(y)\right). && (4.69)
\end{aligned}
$$

Here, the first term coincides with $\tilde{a}_\zeta(\lambda)$ and the second term coincides with $b_\zeta(\lambda)$. Thus, combining (4.44), (4.48), (4.63) and (4.65) we get (4.52).

Refined approximations based on saddlepoint approximation of the RCU and RCUs bounds are presented in Paper A, Paper B and Paper C for system models introduced in Chapter 3. The error probability expansions derived through saddlepoint expansion have a term that decays exponentially with the number of data symbols and thus captures the behavior of the error probability in the large-deviation regime. In addition, a pre-exponential term, obtained by applying a refined normal approximation, ensures the accuracy of the approximation in the short-packet regime. This approach results in an approximation that is accurate for a wide range of target error probabilities and rates, including those relevant to URLLC scenarios.

4.2 First-order Asymptotic Expansion of RCUs bound

In this section, we let the transmission takes place in a block-fading channel as described in Chapter 3.1. The RCUs bound for the block-fading channel is given as

$$\epsilon \leq \mathbb{P}\left[\sum_{\ell=1}^{n_\mathrm{b}} \imath_s(\boldsymbol{X}_\ell; \boldsymbol{Y}_\ell, \hat{H}_\ell) + \log U \leq \log(\mathrm{M}-1)\right] \tag{4.70}$$

$$= \mathbb{E}\left[\mathbb{P}\left[\sum_{\ell=1}^{n_\mathrm{b}} \imath_s(\boldsymbol{X}_\ell; \boldsymbol{Y}_\ell, \hat{H}_\ell) + \log U \leq \log(\mathrm{M}-1) \,\middle|\, \boldsymbol{H}, \hat{\boldsymbol{H}}\right]\right] \tag{4.71}$$

$$= \mathbb{E}\left[\mathbb{P}\left[\sum_{\ell=1}^{n_\mathrm{b}} \sum_{k=1}^{n_\mathrm{c}} \imath_s(X_{k,\ell}; Y_{k,\ell}, \hat{H}_\ell) + \log U \leq \log(\mathrm{M}-1) \,\middle|\, \boldsymbol{H}, \hat{\boldsymbol{H}}\right]\right] \tag{4.72}$$

where $X_{k,\ell}$ and $Y_{k,\ell}$ are the kth element of $\boldsymbol{X}_\ell$ and $\boldsymbol{Y}_\ell$ respectively.

There are two ways to perform asymptotic blocklength over block-fading channels. In particular, the blocklength n can grow infinitely as the size of the fading blocks n_c grows to infinity, or as the number of fading blocks n_b grows to infinity. As illustrated in Fig. 4.1, this results in two different asymptotic regimes: quasistatic setting, in which the asymptotic analysis is performed over n_c, and the ergodic setting, in which the asymptotic analysis

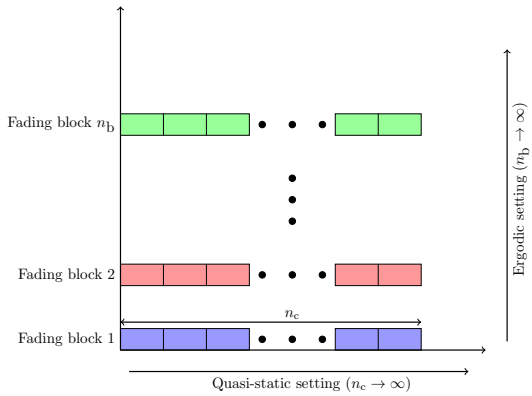

Figure 4.1: Ergodic and quasistatic regimes.

is performed over n_b.

4.2.1 Quasistatic Setting

For a quasistatic block-fading channel spanning n_b coherence blocks, the outage probability as a function of the rate R is defined by

$$P_{\text{out}}(R) = \mathbb{P}\left[\frac{1}{n_b}\sum_{\ell=1}^{n_b}\log\left(1+\rho\left|H_\ell\right|^2\right) \leq \mathrm{R}\right] \tag{4.73}$$

For a given $\epsilon > 0$, the outage capacity C_{out} is defined as the supremum of all rates R that satisfy $\sup\{\mathrm{R} : P_{\text{out}}(\mathrm{R}) < \epsilon\}$. Communication with arbitrarily small error probability is possible if $\mathsf{R} < \frac{1}{n_b}\sum_{\ell=1}^{n_b}\log\left(1+\rho\left|H_\ell\right|^2\right)$. Next, we will show that we can prove the achievability of P_{out} by expanding the RCUs bound.

Note that, when $n_c \to \infty$, the channel h_ℓ can be estimated perfectly with no rate penalty by simply using a sublinear number of channel uses for this task [38, p. 2632]. Thus, we may safely replace n_d by n_c and assume perfect CSI is available at the receiver, namely $\hat{H}_\ell = H_\ell$ for $\ell = 1, \ldots, n_b$.

The tail probability in (4.72) may be expanded as

$$\epsilon \leq \mathbb{E}\left[\mathbb{P}\left[\frac{1}{n_{\mathrm{c}}}\sum_{\ell=1}^{n_{\mathrm{b}}}\sum_{k=1}^{n_{\mathrm{c}}} \imath_s\Big(X_{k,\ell};Y_{k,\ell},\hat{H}_\ell\Big)+\frac{\log U}{n_{\mathrm{c}}} \leq n_{\mathrm{b}}\mathsf{R} \;\middle|\; \boldsymbol{H}\right]\right] \tag{4.74}$$

$$= \mathbb{E}\left[Q\left(\frac{\sum_{\ell=1}^{n_{\mathrm{b}}} I_s(H_\ell,H_\ell)-n_{\mathrm{b}}\mathsf{R}}{\sqrt{\sum_{\ell=1}^{n_{\mathrm{b}}} V_s(H_\ell,H_\ell)/n_{\mathrm{c}}}}\right)+o\Big(\frac{1}{\sqrt{n_{\mathrm{c}}}}\Big)\right]. \tag{4.75}$$

where

$$I_s\Big(H_\ell,\hat{H}_\ell\Big) = -s\left(\left|H_\ell-\hat{H}_\ell\right|^2\rho+1\right)+s\frac{|H_\ell|^2\rho+1}{1+s\rho\left|\hat{H}_\ell\right|^2}+\log\left(1+s\rho\left|\hat{H}_\ell\right|^2\right) \tag{4.76}$$

$$V_s\Big(H_\ell,\hat{H}_\ell\Big) = \mathbb{E}\left[\Big(\imath_s\Big(X_{k,\ell};Y_{k,\ell},\hat{H}_\ell\Big)-I_s\Big(H_\ell,\hat{H}_\ell\Big)\Big)^2 \;\middle|\; H_\ell,\hat{H}_\ell\right] \tag{4.77}$$

(4.74) follows from $\log(\mathrm{M}-1)\leq\log(\mathrm{M})=n\mathrm{R}$, and (4.75) follows from applying the Berry-Esseen CLT, performing multiple Taylor series expansions and gathering the terms vanishing faster than $1/\sqrt{n_{\mathrm{c}}}$ into $o(1/\sqrt{n_{\mathrm{c}}})$. Recall that $P_{\mathrm{out}}(R)$ is the error probability as $n_{\mathrm{c}}\to\infty$, then

$$P_{\mathrm{out}}(R)\leq\mathbb{P}\left\{\frac{1}{n_{\mathrm{b}}}\sum_{l=1}^{n_{\mathrm{b}}} I_s(H_\ell,H_\ell)\leq\mathsf{R}\right\}. \tag{4.78}$$

Here, (4.78) follows from using the identity

$$\lim_{n_{\mathrm{c}}\to\infty} Q\left(\frac{\sum_{\ell=1}^{n_{\mathrm{b}}} I_s(H_\ell,H_\ell)-n_{\mathrm{b}}\mathsf{R}}{\sqrt{\sum_{\ell=1}^{n_{\mathrm{b}}} V_s(H_\ell,H_\ell)/n_{\mathrm{c}}}}\right) = \mathbb{1}\left\{\sum_{\ell=1}^{n_{\mathrm{b}}} I_s(H_\ell,H_\ell)-n_{\mathrm{b}}\mathrm{R}\leq 0\right\}. \tag{4.79}$$

We next set $s=1$, as it maximizes $I_s(H_\ell,H_\ell)$, and get

$$I_1(H_\ell,H_\ell)=\log(1+\rho\,|H_\ell|^2) \tag{4.80}$$

Then, by substituting (4.80) into (4.78), we get the outage capacity.

4.2.2 Ergodic Setting

The well-known capacity expression for the block-fading channel with perfect CSI, i.e. $\hat{\boldsymbol{H}} = \boldsymbol{H}$ may be obtained by following similar steps to obtain outage capacity: We first apply the Berry-Esseen CLT to (4.70) as

$$\epsilon \leq Q\left(\frac{n_{\mathrm{b}} I_s - \log(\mathrm{M}-1)}{\sqrt{n_{\mathrm{b}} V_s}}\right) + o\left(\frac{1}{\sqrt{n_{\mathrm{b}}}}\right) \tag{4.81}$$

where $I_s = \mathbb{E}\Big[\imath_s(\boldsymbol{X};\boldsymbol{Y},\hat{H})\Big]$ is the generalized mutual information, $V_s = \mathbb{V}\mathrm{ar}\Big[\imath_s(\boldsymbol{X};\boldsymbol{Y},\hat{H})\Big]$ is the channel dispersion and $\boldsymbol{X}$, $\boldsymbol{Y}$ and H are distributed as one of the $\boldsymbol{X}_\ell$, $\boldsymbol{Y}_\ell$ and H_ℓ, respectively. It follows that

$$\lim_{n_{\mathrm{b}}\to\infty} \epsilon \leq \mathbb{1}\left\{\frac{I_s}{n_{\mathrm{c}}} < \mathrm{R}\right\}. \tag{4.82}$$

After picking the s that is maximizing I_s, we have $I_s = n_{\mathrm{c}}\,\mathbb{E}\Big[\log(1+\rho\,|H|^2)\Big]$ [39] which then can be substituted into (4.82) as

$$\lim_{n_{\mathrm{b}}\to\infty} \epsilon \leq \mathbb{P}\Big[\mathbb{E}\Big[\log(1+\rho\,|H|^2)\Big] < \mathrm{R}\Big]. \tag{4.83}$$

This expression indicates that a reliable communication can only be established when $\mathrm{R} < \mathbb{E}\Big[\log(1+\rho\,|H|^2)\Big]$ which is exactly the ergodic (or channel) capacity.

4.2.3 Results and Discussion

We next compare the accuracy of the saddlepoint approximation, outage probability and a simple yet theoretically ungrounded approximation that is sometimes used by researchers. Specifically, in [40], the ergodic capacity for block-fading channel under ML estimation of the channel is used to find the outage probability. This outage expression is given as

$$\epsilon \approx \mathbb{P}\left[\left(\frac{n_{\mathrm{c}}-n_{\mathrm{p}}}{n_{\mathrm{c}}}\right)\frac{1}{n_{\mathrm{b}}}\sum_{\ell=1}^{n_{\mathrm{b}}}\mathbb{E}\Big[\log(1+\hat{\rho}_{\mathrm{ML}}(\hat{H}))\Big] < \mathrm{R}\right] \tag{4.84}$$

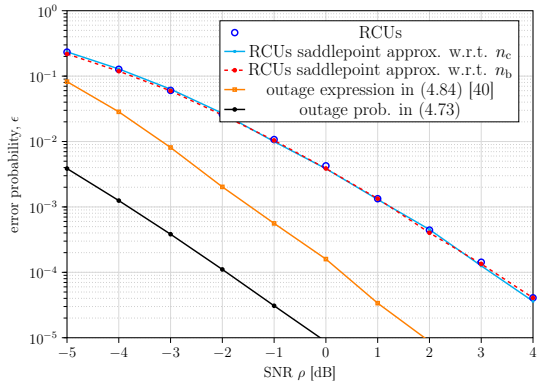

Figure 4.2: Error probability as a function of SNR. Here $n = 192$, $n_\mathrm{b} = 6$, R = 0.104 BPCU; n_p and s are optimized

where

$$\hat{\rho}_{\mathrm{ML}}\left(\hat{H}\right) = \rho \left|\hat{H}\right|^2 \left(\frac{n_\mathrm{p}\rho}{n_\mathrm{p}\rho + 1}\right)^2 \left(\frac{\rho + n_\mathrm{p}\rho + 1}{n_\mathrm{p}\rho + 1}\right)^{-1}. \tag{4.85}$$

Although this expression is not theoretically well-founded, it can nevertheless be regarded as an approximation.

To compare the tightness of the aforementioned expressions, we consider a Rayleigh fading scenario in which $\{H_\ell\}_{\ell=1}^{n_\mathrm{b}}$ are generated independently from a $\mathcal{CN}(0,1)$ distribution. We also consider i.i.d. $\mathcal{CN}(0,\rho)$ Gaussian codebooks and use an ML estimate of the channel. In Fig. 4.2 we report the error probability ϵ as a function of SNR ρ for $n = 192$, $n_\mathrm{b} = 6$, R = 0.104 BPCU of RCUs bound, its saddlepoint approximations w.r.t. n_b and n_c, the outage capacity, and the outage expression in (4.84). We observe that both saddlepoint approximations are very accurate for the considered parameters and that the outage capacity and the outage expression in (4.84) are not a good benchmark for the short-blocklength regime.

CHAPTER 5

Erasure Decoding in URLLC

In this chapter, we present an information theoretic analysis of erasure decoding, in which the decoder detects errors through the capability of the channel codes and in this case declares an erasure. This is fundamentally different than having a complete decoder, which always outputs a codeword as defined in Chapter 2.

5.1 Overview

In channel coding, the goal is to develop a code that can accurately decode a message transmitted over a noisy channel. However, when the noise is too severe for the decoding system to confidently determine the correct message, it is often better to declare an erasure. This approach prevents the system from making the costly mistake of passing an incorrect message to the upper layers, and for instance, allows the use of an automatic repeat request protocol to retransmit the intended message.

The reliability of URLLC systems are typically measured by the block error probability at the output of the decoder. However, for certain mission-critical applications, this metric alone is insufficient, and it is important to distin-

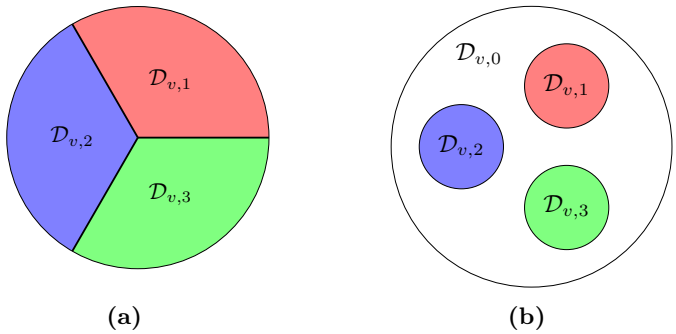

Figure 5.1: Representation of decision regions: (a) classical decoder, (b) erasure decoder

guish between two types of decoding errors: detected and undetected errors. A detected error occurs when the channel decoder declares an erasure. In contrast, an undetected error occurs when the decoder outputs an incorrect message, which is then passed on to the upper layers without being flagged as erroneous.

5.2 System Model

In classical decoding, the decision regions are mutually exclusive and collectively exhaustive, which means that every received vector is assigned to a decision region, and the decoder always outputs a codeword. In contrast, an erasure decoder operates differently: while the decision regions remain disjoint, they do not fully cover the output space. Consequently, certain received vectors fall outside any decision region and cause the decoder to declare an erasure. This is depicted in Fig. 5.1.

To analyse decoding with erasure option, we consider an arbitrary discrete-time communication channel that maps input symbols from the set $\mathcal{X}$ to output symbols from the set $\mathcal{Y}$, as we introduced in Chapter 2. However, our goal is to characterize the performance of the channel codes in terms of both undetected error probability (UEP), and total error probability (TEP), which is the combined probability of undetected and detected errors. To this end, we need to update the notion of channel code introduced in Chapter 2, which we define next.

Definition 5.1. *An $(n, \mathrm{M}, \epsilon_{\mathsf{U}}, \epsilon_{\mathsf{T}})$-code for the channel $P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{Y}|\boldsymbol{X})$ consists of*

- *A random variable V with distribution P_V defined on a set $\mathcal{V}$ with $|\mathcal{V}| \leq 2$ that is revealed to both the transmitter and the receiver before the start of transmission. This allows the transmitter and the receiver to time-share between deterministic $(n, \mathrm{M}, \epsilon_{\mathsf{U}}, \epsilon_{\mathsf{T}})$-codes.*
- *An encoder $f : \mathcal{V} \times \{1, \ldots, \mathrm{M}\} \to \mathcal{X}^n$ that maps a message W, which we assume to be uniformly distributed on $\{1, \ldots, \mathrm{M}\}$, to a codeword in the set $\mathcal{C}_V = \{\mathbf{x}_1, \ldots, \mathbf{x}_{\mathrm{M}}\} \subset \mathcal{X}^n$.*
- *An erasure decoder $g : \mathcal{V} \times \mathcal{Y}^n \to \{0, 1, \ldots, \mathrm{M}\}$, that maps the received vector to one of the messages in $\{1, \ldots, \mathrm{M}\}$, or declares an erasure, which we indicate by the message 0. Let $\mathcal{D}_{v,\hat{w}} = g^{-1}(v, \hat{w}) \subset \mathcal{Y}^n$ denote the decoding region associated to each $\hat{w} \in \{0, 1, \ldots, \mathrm{M}\}$ and v, and assume that they are disjoint. We require that the TEP and UEP do not exceed ϵ_{T} and ϵ_{U}, respectively:*

$$\frac{1}{\mathrm{M}} \sum_{v \in \mathcal{V}} \sum_{m=1}^{\mathrm{M}} \sum_{\substack{m'=0 \\ m' \neq m}}^{\mathrm{M}} \mathbb{P}[\boldsymbol{Y} \in \mathcal{D}_{v,m'} | W = m, V = v] \, P_V(v) \leq \epsilon_{\mathsf{T}} \tag{5.1}$$

$$\frac{1}{\mathrm{M}} \sum_{v \in \mathcal{V}} \sum_{m=1}^{\mathrm{M}} \sum_{\substack{m'=1 \\ m' \neq m}}^{\mathrm{M}} \mathbb{P}[\boldsymbol{Y} \in \mathcal{D}_{v,m'} | W = m, V = v] \, P_V(v) \leq \epsilon_{\mathsf{U}}. \tag{5.2}$$

The performance of $(n, \mathrm{M}, \epsilon_{\mathsf{U}}, \epsilon_{\mathsf{T}})$-code may also be assessed, for a fixed blocklength n and average error probabilities ϵ_{U} and ϵ_{T}, via the maximum coding rate as

$$\mathrm{R}^*(n, \epsilon_{\mathsf{U}}, \epsilon_{\mathsf{T}}) = \frac{\log_2(\mathrm{M}^*(n, \epsilon_{\mathsf{U}}, \epsilon_{\mathsf{T}}))}{n} \tag{5.3}$$

where

$$\mathrm{M}^*(n, \epsilon_{\mathsf{U}}, \epsilon_{\mathsf{T}}) = \sup\{\mathrm{M} : \exists (n, \mathrm{M}, \epsilon_{\mathsf{U}}, \epsilon_{\mathsf{T}})\text{-channel code}\}. \tag{5.4}$$

The supremum in (5.4) is formed over all encoder/decoder pairs that form a $(n, \mathrm{M}, \epsilon_{\mathsf{U}}, \epsilon_{\mathsf{T}})$ channel code.

The original argument of random coding states that for certain encoding and decoding rules, an upper bound on the average error probability, averaged over

an ensemble of randomly generated codebooks, indicates that there is at least one codebook within the ensemble that satisfies this bound. If two quantities need to be bounded, in this case TEP and UEP, this argument needs to be updated. In particular, an upper bound on the average UEP and the average TEP, both averaged over a codebook ensemble, does not guarantee that a single codebook exists that satisfies both bounds simultaneously. One can use a randomized coding strategy, i.e., a new code is drawn from the ensemble and used for each transmission to overcome this problem. It turns out that for this case, randomization between two deterministic codebooks is sufficient [11, App. A] and the random variable V in Definition 5.1 enables the use of the randomized coding strategy.

Next, we present three methods for erasure decoding that are commonly used in the literature: the optimal erasure decoder, the use of the CRC outer code for error detection, and thresholding (generalized) information density.

5.2.1 Optimal Erasure Decoder

An optimal incomplete decoding algorithm was presented and analyzed by Forney in [41]. This method can be seen as applying complete maximum likelihood decoding followed by a threshold test after decoding to decide whether to accept or reject the ML decoder's decision. Forney has shown that the decoder that optimally weighs between TEP and UEP follows this rule: It outputs the message whose likelihood is at least $2^{n\mathrm{T}}$ times greater than the sum of the likelihoods of all other messages. If no message fulfills this criterion, an erasure is declared. Mathematically, for a given observation $\boldsymbol{Y} = \mathbf{y}$, the decoder outputs the coderword $\mathbf{x}$ that fulfills the following conditions

$$\Lambda(\mathbf{x}, \mathbf{y}) > 2^{n\mathrm{T}} \tag{5.5}$$

where $\mathrm{T} \geq 0$ is a parameter that controls the tradeoff between ϵ_{U} and ϵ_{T}, and

$$\Lambda(\mathbf{x}, \mathbf{y}) = \frac{P_{\boldsymbol{Y}|\boldsymbol{X}}\left(\mathbf{y}|\mathbf{x}\right)}{\sum_{\mathbf{x}' \in \mathcal{C} \setminus \{\mathbf{x}\}} P_{\boldsymbol{Y}|\boldsymbol{X}}\left(\mathbf{y}|\mathbf{x}'\right)}. \tag{5.6}$$

If no codeword in the codebook satisfies (5.5), the decoder declares an erasure.

The main disadvantage of this decoder is that it is difficult to analyze and difficult to implement, which limits its use for our purposes.

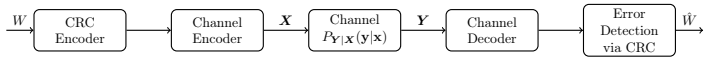

Figure 5.2: Reference model describing erasure decoding via CRC outer code.

5.2.2 Error Detection via CRC Outer Code

A CRC code can be appended to a data packet to enable the error detection after channel decoding as shown in Fig. 5.2 Although suboptimal, using CRC outer code for error detection is a very common choice due to their low encoding complexity and strong error detection performance [42]–[44]. With this scheme, the relation between TEP and UEP can be approximated as $\epsilon_{\mathrm{U}} \approx \epsilon_{\mathrm{T}} 2^{-k_{\mathrm{crc}}}$ where k_{crc} is the number of parity bits added by the CRC [45]. Note that when the blocklength is large, a strong error detection capability can be ensured by using CRC codes as the rate loss due to the addition of these parity bits may be negligible. In the short-blocklength regime however, the rate loss cannot be omitted anymore and should be characterized to have an accurate benchmark for the system.

5.2.3 Suboptimal Threshold Decoder

Another scheme used for error detection, although not as common as CRC codes, is thresholding the generalized information density (defined in (2.21)) [46], [47].[3] Specifically, the decoder selects the codeword with the highest likelihood and outputs the corresponding message if the generalized information density of this codeword exceeds a preset threshold $n\delta$. If this condition is not met, the decoder declares an erasure.

5.3 Finite Blocklength Achievability Bounds for Erasure Decoding

Forney in his seminal paper [41] showed that over a discrete memoryless channel (DMC) for an arbitrary input distribution P_X, and for $\mathcal{X} = \{x_1, \ldots, x_{m_i}\}$ and $\mathcal{Y} = \{y_1, \ldots, y_{m_o}\}$, there exists a block code of length n and rate R that

[3] In [46], [47] the authors threshold the information density in their erasure decoders. By setting $s = 1$ in generalized information density (2.21), we can get the information density, thus thresholding generalized information density is a more general method.

simultaneously satisfies

$$\epsilon_{\mathsf{T}} < 2^{-nE_1(\mathrm{R},\mathrm{T},P_X)} \tag{5.7}$$

$$\epsilon_{\mathsf{U}} < 2^{-nE_2(\mathrm{R},\mathrm{T},P_X)} \tag{5.8}$$

where

$$E_1(\mathrm{R},\mathrm{T},P_X) = \max_{0\le s\le p\le 1} \left[E_0(s,p,P_X) - p\mathrm{R} - sT\right] \tag{5.9}$$

$$E_2(\mathrm{R},\mathrm{T},P_X) = E_1(\mathrm{R},\mathrm{T},P_X) + \mathrm{T} \tag{5.10}$$

with

$$E_0(s,p,P_X) = -\log_2 \sum_{j=1}^{m_o} \left(\sum_{i=1}^{m_i} P_X(x_i) q_{ji}^{1-s}\right) \left(\sum_{i'=1}^{m_i} P_X(x_{i'}) q_{ji'}^{s/p}\right)^p \tag{5.11}$$

and with $q_{ji} = P_{Y|X}(y_j|x_i)$ being the transition probability of the DMC. In the literature, using the optimal erasure decoder and the Gallager's error exponent framework, the bounds are generalized for constant composition codes [48], [49] and improved for some linear block code ensembles [50]. The main drawback with the Forney's bounds (and their generalization) is that they are based on the Gallager's error exponent framework which typically yields loose bounds for short blocklengths [37].

One drawback of Forney's bound is that it is only valid for DMCs. To have achievability bounds on TEP and UEP for block-fading channel, we can use two bounds we introduced in Paper C. Both of these bounds are based on RCU bound, thus, they are computationally demanding, as we discussed in Chapter 2. To address this, we next introduce the relaxed versions of these bounds, relaxed via Markov inequality in (2.18), similar to how we relaxed RCU bound and obtained RCUs bound.

The finite-blocklength achievability bound using a CRC-outer code scheme is given as follows.

Theorem 5.1. *For an arbitrary input distribution $P_{\boldsymbol{X}}$, for all $s > 0$ and for every $k_{\mathrm{crc}} \in \mathbb{N}_0$, there exists an $(n, \mathrm{M}, \epsilon_{\mathsf{T}}, \epsilon_{\mathsf{U}})$-code for the channel*

$P_{\mathbf{y}|\mathbf{x}}(\boldsymbol{Y}|\boldsymbol{X})$ *simultaneously satisfying*

$$\epsilon_{\mathsf{T}} \leq \mathbb{E}\left[e^{-\left(\imath_s(\boldsymbol{X};\boldsymbol{Y})-\log(2^{k+k_{\mathrm{crc}}}-1)\right)^+}\right] \tag{5.12}$$

$$\epsilon_{\mathsf{U}} \leq \mathbb{E}\left[e^{-\left(\imath_s(\boldsymbol{X};\boldsymbol{Y})-\log(2^{k+k_{\mathrm{crc}}}-1)\right)^+}\right] 2^{-k_{\mathrm{crc}}} \tag{5.13}$$

where $k = n\mathrm{R}$.

Proof: See Paper C for the proof of RCU-based bounds for ϵ_{T} and ϵ_{U} for this setup. (5.12) and (5.13) can be obtained by relaxing these bounds using Markov inequality given in (2.18). ■

Here, we observe that when a CRC-outer code is used, the ϵ_{T} can be bounded by RCUs bound with a rate $\mathrm{R} = (k+k_{\mathrm{crc}})/n$ and the error detection capability is directly related with the number of CRC-bits k_{crc}.

The finite-blocklength achievability bound using the suboptimal threshold decoder is given as follows.

Theorem 5.2. *For an arbitrary input distribution* $P_{\boldsymbol{X}}$*, for all* $s > 0$*,* $\bar{s} > 0$ *and* $\delta \in \mathbb{R}$*, there exists an* $(n, \mathrm{M}, \epsilon_{\mathsf{T}}, \epsilon_{\mathsf{U}})$*-code for the channel* $P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{Y}|\boldsymbol{X})$ *satisfying*

$$\epsilon_{\mathsf{T}} \leq \mathbb{E}\left[\mathbb{1}\left\{\imath_{\bar{s}}(\boldsymbol{X},\boldsymbol{Y}) \geq n\delta\right\} e^{-(\imath_s(\boldsymbol{X};\boldsymbol{Y})-\log(\mathrm{M}-1))^+}\right] + \mathbb{P}[\imath_{\bar{s}}(\boldsymbol{X},\boldsymbol{Y}) < n\delta] \tag{5.14}$$

$$\epsilon_{\mathsf{U}} \leq \mathbb{E}\left[e^{-(\tilde{\imath}_{s,\bar{s}}(\boldsymbol{X};\boldsymbol{Y})-\log(\mathrm{M}-1))^+}\right] \tag{5.15}$$

where

$$\tilde{\imath}_{s,\bar{s}}(\boldsymbol{X};\boldsymbol{Y}) = \log \frac{\max\left\{P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{Y}|\boldsymbol{X})^s, \widetilde{\delta}_{\boldsymbol{Y},\bar{s}}\right\}}{\mathbb{E}_{\bar{\boldsymbol{X}}}\left[P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{Y}|\bar{\boldsymbol{X}})^s\right]} \tag{5.16}$$

$$\widetilde{\delta}_{\mathbf{y},\bar{s}} = \left(e^{n\delta}\,\mathbb{E}_{\bar{\boldsymbol{X}}}\left[P_{\boldsymbol{Y}|\boldsymbol{X}}(\mathbf{y}|\bar{\boldsymbol{X}})^{\bar{s}}\right]\right)^{1/\bar{s}} \tag{5.17}$$

and

$$P_{\boldsymbol{X},\boldsymbol{Y},\bar{\boldsymbol{X}}}(\boldsymbol{X},\boldsymbol{Y},\bar{\boldsymbol{X}}) = P_{\boldsymbol{X}}(\boldsymbol{X})P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{Y}|\boldsymbol{X})P_{\boldsymbol{X}}(\bar{\boldsymbol{X}}). \tag{5.18}$$

Proof: Similar to the Theorem 5.1, the RCU-based bounds for this setup on ϵ_{T} and ϵ_{U} are available in Paper C. (5.14) and (5.15) can be obtained by relaxing these bounds using Markov inequality given in (2.18). ■

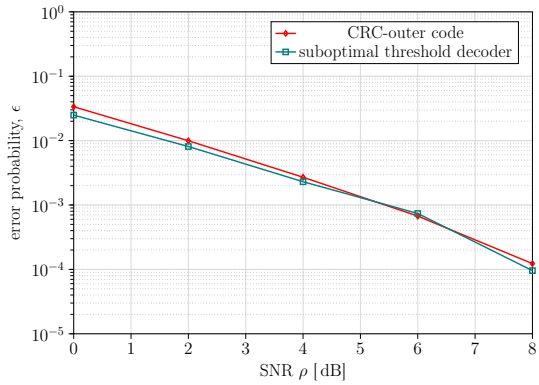

Figure 5.3: TEP as a function of SNR. Here, $n = 288$, $n_{\mathrm{b}} = 3$, R = 0.104 BPCU, $\epsilon_{\mathrm{U}} = 10^{-5}$; s and $\bar{s}$ are optimized.

5.4 Results and Discussion

Next, we compare the performances of the two decoders in a Rayleigh fading scenario. We use i.i.d. $\mathcal{CN}(0, \rho)$ Gaussian codebooks and an ML estimate of the channel at the receiver. In Fig. 5.3 we report TEP as a function of SNR ρ for $n = 288$, $n_{\mathrm{b}} = 3$, $n_{\mathrm{p}} = 10$, R = 0.104 BPCU and $\epsilon_{\mathrm{U}} = 10^{-5}$. Here we see that these two methods perform similarly. This indicates that in the short-blocklength regime, using the CRC outer code for error detection is as good as using the suboptimal threshold decoder. This is important because implementing a CRC outer code for error detection is much easier than implementing the suboptimal threshold decoder.

CHAPTER 6

Summary

This chapter concludes Part I of the thesis by summarizing the most important results and conclusions. In addition, we discuss the limitations of our results and provide directions for future research

6.1 Contributions

In this thesis, we study the performance of wireless communication systems operating in the URLLC regime. Specifically, this thesis contains nonasymptotic information-theoretic studies on:

- Pilot-assisted transmissions over SISO block-fading channels with mismatched SNN decoding
- Multicell multiuser massive MIMO communications over spatially correlated Rayleigh-fading channels
- Pilot-assisted transmissions over block-fading channels with imperfect timing synchronization and mismatched decoder

- The trade-off between the total and the undetected error probabilities over biAWGN and block-fading channels.
- How the saddlepoint approximation can be used to efficiently evaluate the nonasymptotic results

Our contributions are presented in Part II of the thesis in the form of three attached papers that are summarized below.

Paper A:"Efficient Evaluation of the Error Probability for Pilot-assisted URLLC with Massive MIMO"

In Paper A, we propose an efficient method for evaluating the random-coding union bound on the error probability in finite-blocklength regime for pilot-assisted transmission over memoryless block-fading channels using Gaussian codebooks. Our method uses the saddlepoint approximation with respect to the number of fading blocks (diversity branches) per codeword, thus avoiding the costly numerical averaging over fading process realizations and their pilot-assisted estimates at the receiver. This approach significantly reduces the number of channel realizations required to accurately estimate the error probability. Numerical experiments on single-antenna and massive MIMO systems show that, compared to traditional saddlepoint approximations with respect to the number of channel realizations per block, our method requires about two orders of magnitude fewer Monte-Carlo samples and achieves accurate error probability estimation when there are two or more fading blocks.

Paper B: "Is Synchronization a Bottleneck for Pilot-assisted URLLC Links?"

In Paper B, we propose a framework to evaluate the RCUs bound on the achievable error probability in the finite-blocklength regime for pilot-assisted transmission over imperfectly synchronized memoryless block-fading waveform channels. In contrast to previous methods, our system accounts for imperfect synchronization by using pilots for both synchronization and channel estimation. Using this framework, we introduce an RCUs bound for a receiver that treats these estimates as accurate and employ the saddlepoint approximation for numerical efficiency. Numerical experiments confirm the accuracy of the approximation. The results show that for fully dependent delays

across fading blocks, the number of pilots required for good enough channel estimation when time delays are known (i.e., when synchronization is perfect) is also sufficient when times delays are unknown and needs to be estimated in URLLC. However, with independent delays between blocks, synchronization limits system performance.

Paper C: "Undetected Error Probability in the Short Blocklength Regime: Approaching Finite-Blocklength Bounds with Polar Codes"

In paper C, we analyze the tradeoff between the total and the undetected error probability in short blocklength regimes. We introduce two achievability bounds and evaluate the performance of polar codes with outer CRC codes known as CRC-assisted polar codes. The first bound uses an outer detection code, and the second employs a threshold test for generalized information density, both of which are tighter than Forney's error exponent bound for the binary-input additive white Gaussian noise channel. Inspired by these bounds, CRC-assisted polar codes are compared under two error detection techniques: a threshold test that approximates Forney's optimal rule at the output of the successive cancellation list decoder, and an algorithm that splits CRC parity bits into subsets for error detection. The results show that the threshold test outperforms CRC-based error detection, although the advantage diminishes with increasing blocklength. In scenarios with noisy channel state information leading to mismatched decoding, outer CRC codes ensure robustness, while threshold-based techniques relying on mismatched likelihood exhibit significant performance degradation.

6.2 Conclusions

In this thesis, we provide nonasymptotic tools to benchmark wireless communication systems. The conclusions we draw from this work can be summarized as follows:

1. The saddlepoint approximation provides accurate results for the nonasymptotic bounds in short-blocklength regime. When possible, it should be utilized to reduce the numerical complexity evaluations.

2. Timing synchronization in short-packet transmission can be established via pilots. When delays are fully dependent across fading blocks, the

number of pilots required for sufficiently good channel estimation when delays are known is also sufficient to establish synchronization. Hence, in this sense, there is no penalty for not knowing the channel delays.

3. The optimal erasure decoder is both difficult to analyze and implement. In fact, the available achievability bound on the error probability for the optimal decoder in the short blocklength regime is loose. This bound can be significantly improved by analyzing suboptimal decoders that utilize error detection schemes, such as employing a CRC outer code or thresholding the information density.

6.3 Future Work

The results in all our studies are based on RCUs bound and saddlepoint and normal approximations. An alternative approach to obtain nonasymptotic result is studied in [51], where the authors considered the so-called moderate deviation regime and derived refined asymptotic expansions of the upper and lower bounds of the error probability based on another refined asymptotic expansion for independent random variables [52, Ch. 8, Th. 4]. In this regime, it is assumed that the error probability ϵ decay subexponentially with the blocklength n. Thus, the goal is to close the gap between the approximations based on the central limit theorem, which yield asymptotic expansions of the rate with increasing n for fixed ϵ, and the approximations based on large deviations such as the error exponent approximation, which yield asymptotic expansions of ϵ with fixed rate. It would be interesting to see how this approximation behaves, for example, in the context of for example, massive MIMO for URLLC.

Another interesting extension of all our studies would be to include the effects of practical components. In industrial environments, for example, electrical and other types of machinery are recognised sources of significant interference. In addition to the usual white Gaussian thermal noise, this interference can severely affect communication in radio frequency bands [53]. The block-fading channel that we have considered in this thesis cannot model the effects of such behaviour. For this, one should consider channel models that take impulsive noise into account [54].

A missing component in this study is the lack of nonasymptotic converse bounds for various system models considered in Paper A, Paper B and Paper

C. Unfortunately, it is a challenge to obtain a general nonasymptotic converse bound for mismatched decoding since there is no converse bound for mismatched capacity.

In Paper A, we claim that an advantage of saddlepoint approximation with respect to number of fading blocks is that it requires a small number of Monte Carlo samples. An interesting extension would be to derive an explicit expression for how many samples are required to achieve a certain accuracy with saddlepoint approximation. Such an expression would allow the extension of the results in Paper A to different system models.

In Paper B, timing synchronization was established only with the help of pilot symbols. A valuable extension would be to allow the system to use the data symbols for synchronization and then examine the trade-off between the error probability and the number of pilot symbols for a fixed blocklength and rate. It would also be interesting here to see how our bounds compare to an actual code for URLLC.

Finally, although we have some understanding of how the suboptimal threshold decoder used in Paper C relates to Forney's optimal erasure decoder, it would be interesting to analyse their relation rigorously. Indeed, such an analysis could also pave the wave for the discovery of better performing suboptimal decoders.

References

[1] G. Durisi, T. Koch, and P. Popovski, "Towards massive, ultra-reliable, and low-latency wireless communication with short packets", *Proc. IEEE*, vol. 104, no. 9, pp. 1711–1726, Sep. 2016.

[2] G. J. Sutton, J. Zeng, R. P. Liu, *et al.*, "Enabling technologies for ultra-reliable and low latency communications: From PHY and MAC layer perspectives", *IEEE Commun. Surveys & Tutorials*, vol. 21, no. 3, pp. 2488–2524, Feb. 2019.

[3] P. Popovski, K. F. Trillingsgaard, O. Simeone, and G. Durisi, "5G wireless network slicing for eMBB, URLLC, and mMTC: A communication-theoretic view", *IEEE Access*, vol. 6, pp. 55 765–55 779, Sep. 2018.

[4] G. Berardinelli, N. H. Mahmood, I. Rodriguez, and P. Mogensen, "Beyond 5G wireless IRT for industry 4.0: Design principles and spectrum aspects", in *2018 IEEE Globecom Workshops (GC Wkshps)*, Feb. 2018, pp. 1–6.

[5] G. Hampel, C. Li, and J. Li, "5G ultra-reliable low-latency communications in factory automation leveraging licensed and unlicensed bands", *IEEE Communications Magazine*, vol. 57, no. 5, pp. 117–123, May 2019.

[6] P. Schulz, M. Matthe, H. Klessig, *et al.*, "Latency critical IoT applications in 5G: Perspective on the design of radio interface and network architecture", *IEEE Communications Magazine*, vol. 55, no. 2, pp. 70–78, 2017.

[7] B. S. Khan, S. Jangsher, A. Ahmed, and A. Al-Dweik, "URLLC and eMBB in 5G industrial IoT: A survey", *IEEE Open J. Commun. Soc.*, vol. 3, pp. 1134–1163, Jul. 2022.

[8] S. Sukhmani, M. Sadeghi, M. Erol-Kantarci, and A. El Saddik, "Edge caching and computing in 5G for mobile AR/VR and tactile internet", *IEEE MultiMedia*, vol. 26, no. 1, pp. 21–30, Jan. 2019.

[9] J. Li, H. Sahlin, and G. Wikström, "Uplink PHY design with shortened tti for latency reduction", in *2017 IEEE Wireless Communications and Networking Conference (WCNC)*, Mar. 2017, pp. 1–5.

[10] A. Sauter, B. Matuz, and G. Liva, "Error detection strategies for CRC-concatenated polar codes under successive cancellation list decoding", in *2023 57th Annual Conference on Information Sciences and Systems (CISS)*, Apr. 2023, pp. 1–6.

[11] J. Östman, R. Devassy, G. Durisi, and E. G. Ström, "Short-packet transmission via variable-length codes in the presence of noisy stop feedback", *IEEE Transactions on Wireless Communications*, vol. 20, no. 1, pp. 214–227, Sep. 2021.

[12] M. C. Coşkun, G. Durisi, T. Jerkovits, *et al.*, "Efficient error-correcting codes in the short blocklength regime", *Physical Communication*, vol. 34, pp. 66–79, Jun. 2019, ISSN: 1874-4907.

[13] G. Durisi, G. Liva, and Y. Polyanskiy, "Short-packet transmission", in *Information Theoretic Perspectives on 5G Systems and Beyond*, I. Maric, S. Shamai, and O. Simeone, Eds., New York, NY, USA: Cambridge Univ. Press, 2021.

[14] J. Östman, "Ultra-reliable short-packet communications: Fundamental limits and enabling technologies", PhD thesis, Chalmers University of Technology, Gothenburg, Sweden, Oct. 2020.

[15] G. C. Ferrante, J. Östman, G. Durisi, and K. Kittichokechai, "Pilot-assisted short-packet transmission over multiantenna fading channels: A 5G case study", in *Proc. Conf. on Inf. Sci. and Sys. (CISS)*, Princeton, NJ, U.S.A., Mar. 2018, pp. 1–6.

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

[17] R. A. Costa, M. Langberg, and J. Barros, "One-shot capacity of discrete channels", in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Austin, TX, Jun. 2010, pp. 211–215.

[18] A. Martinez and A. Guillén i Fàbregas, "Saddlepoint approximation of random-coding bounds", in *Proc. Inf. Theory Appl. Workshop*, San Diego, CA, USA, Feb. 2011, pp. 257–262.

[19] W. Yang, A. Collins, G. Durisi, Y. Polyanskiy, and H. V. Poor, "Beta–Beta bounds: Finite-blocklength analog of the golden formula", *IEEE Trans. Inf. Theory*, vol. 64, no. 9, pp. 6236–6256, Sep. 2018.

[20] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed. Cambridge University Press, 2011.

[21] J. Font-Segura, G. Vazquez-Vilar, A. Martinez, A. Guillén i Fàbregas, and A. Lancho, "Saddlepoint approximations of lower and upper bounds to the error probability in channel coding", in *Proc. Conf. Inf. Sci. Sys. (CISS)*, Princeton, NJ, Mar. 2018.

[22] A. Molisch, *Wireless Communications*, 2nd ed. New York, NY: John Wiley & Sons, 2011.

[23] J. Östman, G. Durisi, E. G. Ström, M. C. Coşkun, and G. Liva, "Short packets over block-memoryless fading channels: Pilot-assisted or noncoherent transmission?", *IEEE Trans. Commun.*, vol. 67, no. 2, pp. 1521–1536, Feb. 2019.

[24] H. Puttnies, P. Danielis, A. R. Sharif, and D. Timmermann, "Estimators for time synchronization—Survey, analysis, and outlook", *IoT*, vol. 1, no. 2, pp. 398–435, 2020, ISSN: 2624-831X.

[25] A. Mahmood, R. Exel, H. Trsek, and T. Sauter, "Clock synchronization over IEEE 802.11—a survey of methodologies and protocols", *IEEE Transactions on Industrial Informatics*, vol. 13, no. 2, pp. 907–922, 2017.

[26] R. Scholtz, "Frame synchronization techniques", *IEEE Transactions on Communications*, vol. 28, no. 8, pp. 1204–1213, 1980.

[27] X. Hu, X. Cui, Z. Sha, and M. Lu, "A novel start-of-frame delimiter detection algorithm for IEEE 802.15.4a ultra-wideband", *IEEE Communications Letters*, vol. 28, no. 6, pp. 1387–1391, Apr. 2024.

[28] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO networks: Spectral, energy, and hardware efficiency", *Foundations and Trends® in Signal Processing*, vol. 11, no. 3-4, pp. 154–655, 2017, ISSN: 1932-8346.

[29] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas", *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[30] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO has unlimited capacity", *IEEE Trans. Wireless Commun.*, vol. 17, no. 1, pp. 574–590, Jan. 2018.

[31] J. Vieira, S. Malkowsky, K. Nieman, *et al.*, "A flexible 100-antenna testbed for massive MIMO", in *2014 IEEE Globecom Workshops (GC Wkshps)*, Dec. 2014, pp. 287–293.

[32] T. Sarkar, Z. Ji, K. Kim, A. Medouri, and M. Salazar-Palma, "A survey of various propagation models for mobile communication", *IEEE Antennas and Propagation Magazine*, vol. 45, no. 3, pp. 51–82, 2003.

[33] J. Östman, A. Lancho, G. Durisi, and L. Sanguinetti, "URLLC with massive MIMO: Analysis and design at finite blocklength", *IEEE Trans. Wireless Commun.*, vol. 20, no. 10, Oct. 2021.

[34] W. Feller, *An Introduction to Probability Theory and Its Applications*. New York, NY, USA: Wiley, 1971, vol. 2.

[35] A. L. Serrano, "Fundamental limits of short-packet wireless communications", PhD thesis, Universidad Carlos III de Madrid, Spain, Jun. 2019.

[36] J. Scarlett, A. Martinez, and A. Guillén i Fàbregas, "Mismatched decoding: Error exponents, second-order rates and saddlepoint approximations", *IEEE Trans. Inf. Theory*, vol. 60, no. 5, pp. 2647–2666, May 2014.

[37] A. Lancho, J. Östman, G. Durisi, T. Koch, and G. Vazquez-Vilar, "Saddlepoint approximations for short-packet wireless communications", *IEEE Trans. Wireless Commun.*, vol. 19, no. 7, pp. 4831–4846, Jul. 2020.

[38] E. Biglieri, J. Proakis, and S. Shamai, "Fading channels: Information-theoretic and communications aspects", *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2619–2692, Oct. 1998.

[39] A. Lapidoth and S. Shamai (Shitz), "Fading channels: How perfect need "perfect side information" be?", *IEEE Trans. Inf. Theory*, vol. 48, no. 5, pp. 1118–1134, May 2002.

[40] M. Karlsson, E. Björnson, and E. G. Larsson, "Performance of in-band transmission of system information in massive MIMO systems", *IEEE Trans. Wireless Commun.*, vol. 17, no. 3, pp. 1700–1712, Mar. 2018.

[41] G. Forney Jr., "Exponential error bounds for erasure, list, and decision feedback schemes", *IEEE Trans. Inf. Theory*, vol. 14, no. 2, pp. 206–220, Mar. 1968.

[42] W. W. Peterson and D. T. Brown, "Cyclic codes for error detection", *Proceedings of the IRE*, vol. 49, no. 1, pp. 228–235, Jan. 1961.

[43] M. El-Khamy, J. Lee, and I. Kang, "Detection analysis of CRC-assisted decoding", *IEEE Communications Letters*, vol. 19, no. 3, pp. 483–486, Jan. 2015.

[44] T. Tonnellier, C. Leroux, B. Le Gal, B. Gadat, C. Jego, and N. Van Wambeke, "Lowering the error floor of turbo codes with CRC verification", *IEEE Wireless Communications Letters*, vol. 5, no. 4, pp. 404–407, 2016.

[45] 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, Nov. 2011.

[46] M. Hayashi and V. Y. F. Tan, "Asymmetric evaluations of erasure and undetected error probabilities", *IEEE Trans. Inf. Theory*, vol. 61, no. 12, pp. 6560–6577, Dec. 2015.

[47] V. Y. F. Tan and P. Moulin, "Second-order capacities of erasure and list decoding", in *Proc. Int. Symp. Inf. Theory*, Honolulu, HI, USA, Jun. 2014, pp. 1887–1891.

[48] I. E. Telatar and R. G. Gallager, "New exponential upper bounds to error and erasure probabilities", in *Proc. Int. Symp. Inf. Theory*, Trondheim, Norway, Jun. 1994, p. 379.

[49] I. E. Telatar, "Multi-access communications with decision feedback decoding", PhD thesis, Mass. Inst. Technol., Cambridge, MA, May 1992.

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

[51] R. C. Yavas, V. Kostina, and M. Effros, "Third-order analysis of channel coding in the moderate deviations regime", in *2022 IEEE International Symposium on Information Theory (ISIT)*, Jun. 2022, pp. 2309–2314.

[52] W. Petrov, *Sums of Independent Random Variables*. Springer, 1975.

[53] M. G. Sanchez, I. Cuinas, and A. V. Alejos, "Interference and impairments in radio communication systems due to industrial shot noise", in *2007 IEEE International Symposium on Industrial Electronics*, Jun. 2007, pp. 1849–1854.

[54] G. Casciano, P. Baracca, and S. Buzzi, "Enabling ultra reliable wireless communications for factory automation with distributed MIMO", in *2019 IEEE 90th Vehicular Technology Conference (VTC2019-Fall)*, Nov. 2019, pp. 1–7.