Aalborg Universitet

AALBORG UNIVERSITY
DENMARK

A Statistical Learning Approach to Ultra-Reliable Low Latency Communication

Angjelichinoski, M.; Trillingsgaard, K. F.; Popovski, P.

Published in:
IEEE Transactions on Communications

DOI (link to publication from Publisher):
10.1109/TCOMM.2019.2907241

Publication date:
2019

Document Version
Accepted author manuscript, peer reviewed version

Link to publication from Aalborg University

Citation for published version (APA):
Angjelichinoski, M., Trillingsgaard, K. F., & Popovski, P. (2019). A Statistical Learning Approach to Ultra-Reliable Low Latency Communication. *IEEE Transactions on Communications*, *67*(7), 5153-5166. Article 8673783. https://doi.org/10.1109/TCOMM.2019.2907241

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

Take down policy
If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from vbn.aau.dk on: March 29, 2024

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2019.2907241, IEEE Transactions on Communications

A Statistical Learning Approach to Ultra-Reliable Low Latency Communication

Marko Angjelichinoski, *Student Member, IEEE,* Kasper Fløe Trillingsgaard, *Student Member, IEEE,* and Petar Popovski, *Fellow, IEEE*

***Abstract*—Mission-critical applications require Ultra-Reliable Low Latency (URLLC) wireless connections, where the packet error rate (PER) goes down to 10^{-9}. Fulfillment of the bold reliability figures becomes meaningful only if it can be related to a statistical model in which the URLLC system operates. However, this model is generally not known and needs to be learned by sampling the wireless environment. In this paper we treat this fundamental problem in the simplest possible communication-theoretic setting: selecting a transmission rate over a dynamic wireless channel in order to guarantee high transmission reliability. We introduce a novel statistical framework for design and assessment of URLLC systems, consisting of three key components: (i) channel model selection; (ii) learning the model using training; (iii) selecting the transmission rate to satisfy the required reliability. As it is insufficient to specify the URLLC requirements only through PER, two types of statistical constraints are introduced, Averaged Reliability (AR) and Probably Correct Reliability (PCR). The analysis and the evaluations show that adequate model selection and learning are indispensable for designing consistent physical layer that asymptotically behaves as if the channel was known perfectly, while maintaining the reliability requirements in URLLC systems.**

***Index Terms*—URLLC, channel uncertainty, parametric models, non-parametric models, learning, MLE, training.**

I. Introduction

ULTRA-reliable low latency communication (URLLC) is among the most exciting novelties in 5G networks [1]–[4]. The projected reliability guarantees of 10^{-5}, and even going down to 10^{-9}, represent the ultra-reliable (UR) regime of wireless operation. This regime is necessary to support a multitude of mission-critical applications, such as remote control of robots, autonomous coordination among vehicles as well as many yet-to-be-defined use cases. The strictness of the reliability requirements increases the focus on the performance of the *physical layer*, its main design assumptions [1], [5], the modeling of the inherently unreliable wireless channel [6]–[8] and the adequacy of existing transmission strategies in UR-relevant regime [8].

Fulfillment of the bold figures on reliability for URLLC becomes meaningful only if it can be related to a statistical model in which the URLLC system operates, as in that case one can calculate the probability of error or failure. However, the statistical model and the probability distribution of the parameters that affect the occurrence of errors are, generally, unknown and the URLLC system needs to invest resources to learn them. In general, the statistical model can encompass the interference from other devices, behavior of the protocols, etc. Here we consider the simplest possible communication-theoretic setting: selecting a transmission rate over a dynamic wireless channel, in absence of interference, in order to guarantee high transmission reliability. For example, even if it is known that a channel is subject to a Rayleigh fading, guaranteeing certain reliability means that the average gain of that channel is known perfectly. We take a closer look into the impact of channel uncertainty on the reliability that can be attained in wireless systems operating in UR-relevant regime. Specifically, we consider two fundamental questions:

1) *What level of knowledge is required about the wireless channel to be able to guarantee ultra-reliable support of a certain transmission rate?* To the best of our knowledge, the existing URLLC studies select the transmission parameters by ignoring the uncertainty introduced by the transmission environment. As a result, the assumption of perfect knowledge of the channel statistics needs to be revised, as without it, the extreme reliability guarantees become questionable. This assumption is fundamentally *unreliable*, and although it might be acceptable in existing wireless systems with reliability targets in the order of 10^{-3}, the impact of channel uncertainty will lead to severe performance degradation in URLLC systems.
2) *How to pose the requirements for ultra-reliability in a statistically correct way?* Stating only that we need a packet error rate of e.g. $\epsilon = 10^{-6}$ is insufficient. We define two different way to measure reliability: *Averaged Reliability (AR)*, suitable for dynamically changing environments, and *Probably Correct Reliability (PCR)*, where we assign a confidence that the reliability target will be met in a relatively static environment.

Addressing these questions calls naturally for the use of the statistical learning methodology, which has recently started to get a traction in the wireless communication community [9]. Our investigation, which to the best of our knowledge is the first of its kind, shows that, the knowledge about the true channel statistics affects significantly the amount of effort that has to be invested in guaranteeing high reliability. Furthermore, compared to the case in which the channel statistics is perfectly known, the transmitter needs to sacrifice the spectral efficiency in order to meet the reliability requirements. Hence, more

The authors are with the Department of Electronic Systems, Aalborg University, Denmark (e-mail: {maa,kft,petarp}@es.aau.dk). M. Angjelichinoski is now with the Department of Electrical and Computer Engineering, Duke University, USA (e-mail: marko.angjelichinoski@duke.edu).

This work has been in part supported by the European Research Council (ERC) under the European Union Horizon 2020 research and innovation program, through the ERC Consolidator Grant Nr. 648382 WILLOW.

0090-6778 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2019.2907241, IEEE Transactions on Communications

robust methods for design and assessment of URLLC systems are required. We illustrate this in the following example.

A. A Motivating Example

Alice is selling an ultra-reliable wireless communication system. In the advertisement she claims that the system can deliver information at a rate of up to R with the probability of failing being no larger than ϵ (e.g. 10^{-5}). Bob considers purchasing the system and wants the following two questions answered:

1) *How does Alice measure the reliability performance?*
2) *Under what conditions can the system offer the advertised performance?*

To investigate, Bob contacts Alice and gets the following response: *when established over a flat fading wireless channel with received power drawn from a cumulative distribution function F, the system can support one-way transmission at a maximal rate $R = R_\epsilon(F)$ while maintaining link outage probability equal to ϵ.* Alice also provides Bob a list of rates $R_\epsilon(F)$ for various values of F and ϵ.

Alice's answer might sound reassuring to some customers, but not to Bob. Instead, he finds the answer limiting and responds with several other questions:

- *What happens if the system has only limited/imperfect knowledge of F?*
- *What if the true channel differs from F?*
- *What if the system has no knowledge of F at all?*

Is Alice in a position to make an informed statement about the achievable rates and the corresponding reliability performance of the system under all of the above circumstances? Although she might be able to give some insights on part of them, Alice has a hard time in giving Bob technically precise or fully general answers to all questions. This is because, unless F is fully and perfectly known, it is unclear to Alice how the reliability should be assessed.

B. URLLC Revisited

This paper answers Bob's questions in a novel *statistical framework* for design and performance assessment. Being inspired by supervised learning, the framework consists of three key elements: *model selection*, *learning*, and *rate selection*.

Model Selection: As indicated in the above example, when transmitting at rate $R_\epsilon(F)$, the corresponding reliability, assessed in terms of e.g. link outage probability ϵ, can be guaranteed with *certainty* only in the case when the system has perfect knowledge of the true distribution F i.e. the exact channel statistics. Such knowledge is seldom available in practice and in all other cases the system has to learn, i.e., estimate F. In order to do so, the system first selects a model for F. The choice will, in general, depend on the state of knowledge of the system regarding the true distribution F. In some cases, a side information on the physical properties of the wireless channel might be available. For instance, the system might know that the channel exhibits a strong diffuse/specular component and adopt the Rayleigh/Rician fading model [10]. In other cases, physical knowledge will be unavailable; hence, no parametric model is suitable, and the system should resort to non-parametric models. Alternatively, given that in URLLC applications only the regions of very low outages are of practical interest, the system might resort to simple and general first order approximations of the lower tail of F [6].

Learning: After selecting an appropriate model, the system applies a learning procedure that generates an estimate of F, denoted by $\widehat{F}$, using a finite number of channel measurements. We refer to the measurements as the *training sample* and assume that they are collected in a dedicated *training phase* prior to transmission. For parametric models including the models based on tail approximations, we use *maximum likelihood estimation (MLE)*, as a standard learning tool in absence of informative priors.

Rate Selection: The estimated distribution $\widehat{F}$ is a function of the random training sample which makes it a random quantity itself and is, therefore, inherently uncertain. Choosing the transmission rate as $R = R_\epsilon(\widehat{F})$ randomizes the outage probability, i.e., different realizations of the training sample induce different outage probabilities; this is why Alice is unable to give Bob deterministic reliability guarantees. In other words, the commonly accepted methodology of specifying ultra-reliability through an outage probability value is essentially insufficient since, under limited channel knowledge and uncertainty, the outage probability is a random variable. In our framework, the system selects the transmission rate such that predefined statistical reliability constraints are satisfied; hence, the reliability now is guaranteed either on average or probabilistically. These reliability constraints rely on the statistical characterization of the outage probability as a random variable and impose limits on some specific probabilistic measures. Furthermore, due to the statistical interpretation, the reliability constraints naturally render themselves useful when the objective is the joint co-design of the communication and control planes. We consider two types of constraints. The first type of constraint, termed *Averaged Reliability (AR)*, controls the mean of the outage probability over all possible realizations of the training sample and is suitable for designing URLLC systems that perform as desired on the average, over all transmissions. The second constraint, termed *Probably Correct Reliability (PCR)*, controls the probability ξ that the outage probability violates ϵ for a given specific training sample. The latter constraint is more restrictive and suitable for applications that require tighter, per-transmission control of the performance of the system. In principle, the second constraint generalizes the first as it provides the system designer with the freedom to control the higher order moments of the distribution of the outage probability via an additional free parameter ξ. This type of statistical approach in characterizing the reliability performance probabilistically via two parameters is reminiscent to the framework of *probably approximately correct (PAC) learning* [11], [12]. The reader would recall that the goal in PAC learning, after acquiring the training samples, is to select a generalization function which, with high probability, has a low generalization error. Drawing the parallel, in our framework, when designing the URLLC system according to the second constraint, the goal of the designer, after training the channel, is to select a transmission rate which

0090-6778 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2019.2907241, IEEE Transactions on Communications

with high probability $1-\xi$ will have link outage probability equal to ϵ. Clearly, in both frameworks the performance is assessed probabilistically.

Consistency and Reliability Assessment: As the size of the training sample grows, we intuitively expect that the transmission rate will converge to Alice's rate, i.e., the maximum achievable rate $R_\epsilon(F)$ for given ϵ when F is perfectly known; we refer to this desired property as *consistency*. We observe that ensuring consistency while meeting the statistical reliability requirements is crucially dependent on the relation between the assumed model and the true distribution. Summarizing the main findings, we conclude that the scheme is consistent when

- The assumed model belongs in the same parametric family as F, or
- The system uses a non-parametric model.

In the first case, the transmission rate converges rather quickly. However, the assumption of knowing which parametric family of models the channel belongs to is a strong one, difficult to fulfill in practice and prone to modeling mismatch which severely violates the reliability performance. Non-parametric models have generalization power and work for any channel. Nevertheless, the convergence is slow, requiring prohibitive training sample sizes. Finally, models based on first-order lower tail approximation offer "the best of the two worlds" [6]. Although such models do not maintain consistency due to inherent mismatch stemming from the approximation error, they still show superior performance compared to poor parametric modeling choices. In addition, they also require less channel training compared to non-parametric models.

C. Related Work and Organization of Contributions

There are several works that have focused on assessing the impact of channel uncertainty onto the realized reliability performance of URLLC systems. Preliminary insights are presented by Bennis et al. in the recent survey [5]. The work presented there gives an overview of key URLLC enablers with their corresponding performance metrics and identifies the challenges related to each of them; the authors also acknowledge the need for statistical characterization of the channel behavior in the regime of extremely rare events, i.e., the low tail of the channel distribution as well as the need to redefine the reliability metric when channel uncertainty and limited channel knowledge is being considered. However, apart from a discussion on potentially viable statistical tools for doing so (such as extreme value theory for instance), [5] does not provide any rigorous assessment of the performance of URLLC systems under channel uncertainty. Swamy et al. in [7] and [8] give more focused treatment of the problem of guaranteeing the reliability performance in block-fading URLLC systems under channel uncertainty. Nevertheless, the approach presented there relies heavily on consideration about the physical properties of the channel which makes it difficult to scale in practice; in addition, the analysis considers only the case of Rayleigh fading and provides little to no insight in general, non-Rayleighian environments. Within the context established in [5], an even earlier work by Eggers et al. [6] studies the tail behavior of the wireless channel using first-order expansion of the received power distribution; the main conclusion there is that wide variety of channel models in the UR-regime can be approximated via unified power law and represented only via two parameters. This result is further revisited in this paper in Section VI where it has been justified from extreme value theoretic point of view, which has not been done in [6]. Furthermore, the result has been used to show how to design channel-agnostic rate-selection function with satisfactory performance in terms of both reliability guarantees and spectral efficiency which is also missing in [6].

With respect to prior art, the main contribution of our work can be summarized as follows: we cast the problem of configuring the physical layer parameters of an URLLC system into a novel framework directly motivated and inspired from the theory of statistical learning. We apply the framework over a simple toy example of selecting the transmission rate in coherent, one-way communication system in order to illustrate how it can be used to deduce important insights and, in the same time, show how this well-established mathematical discipline should be applied for fundamental and rigorous assessment of URLLC communication systems. This type of treatment is missing in the literature, while its importance will only grow with the increasing importance of ultra-reliability in wireless systems. After introducing the system model in Section II, the specific contributions of the paper are distributed among the remaining sections in the following order.

- Motivated by the statistical interpretation of the outage probability under channel uncertainty, Section III introduces the two novel types of reliability constraints, already discussed in Section I-B, and also provides hints on the possible practical applications where they can be applied. In addition, Section II introduces the definition of the main problem considered in the rest of the paper: selection of transmission rate under limited channel knowledge while maintaining statistically-defined reliability constraints.
- Sections IV, V and VI focus on solving the probabilistically-constrained rate-selection problem using parametric, non-parametric and approximate, channel-agnostic models and identify several important trade-offs between consistency and training overhead, also briefly discussed in Section I-B.

To the best of the authors' knowledge and given the extent of the analysis, this is the first work of its kind.

II. Preliminaries

A. System model

We consider one-way communication link where the transmitter (Tx) sends a packet to a receiver (Rx) at rate R over a flat fading wireless communication channel. With F we denote the cumulative distribution function (CDF) of the received power $Y \geq 0$; we denote $Y \sim F$ and use the terms true channel statistics and true distribution interchangeably when

0090-6778 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2019.2907241, IEEE Transactions on Communications

referring to F.[1] We assume that F belongs to a class $\mathcal{F}$ of smooth distributions defined over the non-negative reals. We will consider two cases: 1) $\mathcal{F}$ is a *parametric*, and 2) $\mathcal{F}$ is an arbitrary non-parametric family of distributions. Throughout the rest of the paper, we assume that the F is a *stationary*, i.e., it does not change over time; this implies that the class $\mathcal{F}$ is fixed, which in the parametric case further implies that the parameters specifying the distribution class are not functions of time. Prior to transmission, in the training phase[2] the Tx collects n independent and identically distributed (i.i.d.) channel measurements from F; we refer to them as the *training sample* and denote them by $X^n = \{X_1, \ldots, X_n\}$.[3]

Parametric Channel Models: In this case, we assume that F belongs to a parametric family of distributions $\mathcal{F}_{\boldsymbol{\theta}} = \{F_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$ with $\boldsymbol{\Theta}$ denoting the parameter space. The parameter $\boldsymbol{\theta}$, which can be a vector or a scalar, usually has a specific interpretation stemming from the physical properties of the wireless channel. We review three common parametric models that are widely used in practice.

Rayleigh channel: Adopted in many wireless studies, the Rayleigh model represents a single scatterer (i.e. cluster) where the received signal is due to a diffuse component only [10], [13]. The received power Y follows an exponential distribution and the CDF is given by

$$F_\lambda(y) = 1 - e^{-\frac{y}{\lambda}}, \tag{1}$$

where $\lambda = \mathbb{E}[Y]$ denotes the average received power.

Rician channel: The Rician channel is an extension of the Rayleigh fading model, featuring an additional specular component [10], [13]. Let the power of the specular component be denoted by ρ, whereas, similarly to the Rayleigh case, the mean power of the diffuse component is denoted by λ. Let $k = \rho/\lambda \geq 0$ be the ratio between the average powers of the specular and diffuse components, also known as a Rician k-factor. The CDF of the received power under Rician fading is given as

$$F_{\lambda,k}(y) = 1 - Q_1\left(\sqrt{2k}, \sqrt{2\frac{y}{\lambda}}\right), \tag{2}$$

where $Q_1(\cdot,\cdot)$ is the 1st order Marcum Q-function. The average received power is $\mathbb{E}[Y] = \rho + \lambda$. The special case when $k = 0$, indicating no specular component, corresponds to Rayleigh-distributed channel as in (1).

Nakagami-m channel: The extension of the Rayleigh model in multi-cluster settings is the Nakagami-m model where the received envelope follows the Nakagami distribution with shape parameter m and scale parameter λ [14]. The Nakagami-m channel can be interpreted as the incoherent sum of m i.i.d. Rayleigh-type clusters, each with mean diffuse power λ. The CDF of the received power under Nakagami-m fading is

$$F_{\lambda,m}(y) = \frac{\gamma\left(m, \frac{y}{\lambda}\right)}{\Gamma(m)}, \tag{3}$$

where $\gamma(\cdot,\cdot)$ denotes the lower incomplete gamma function, whereas $\Gamma(\cdot)$ denotes the gamma function. Evidently, the average received power is $\mathbb{E}[Y] = m\lambda$. For generality, we assume that $m \geq 0.5$ [14]. The special case $m = 1$ indicates a single diffuse cluster and therefore corresponds to the Rayleigh channel (1).

B. Outage probability

To isolate and study the impact of channel uncertainty, we neglect the impact of noise and interference and consider errors due to *link outage* only. Link outages are defined by the following event:

$$R > \log_2(1 + Y). \tag{4}$$

Hence, the outage probability at transmission rate[4] R is defined as

$$p_F(R) = \mathbb{P}[R > \log_2(1 + Y)]. \tag{5}$$

This approach fits well our overall narrative; we study block-fading channel where we neglect other impairments such as receiver noise. In such case, the dominant source of error are the outages. In addition to this, deep fading is a wireless channel phenomenon; hence, using outage probability as a reliability metric extrapolates the impact of the wireless channel onto the performance of the communication system at the most fundamental, communication-theoretic level.

The goal of ultra-reliable communication is to choose the maximal rate that meets a predetermined reliability constraint, such as

$$p_F(R) \leq \epsilon. \tag{6}$$

However, designing the reliability criteria as well as determining the most favorable transmission rate is strongly linked to the amount of knowledge, i.e., state of knowledge the Tx has about the true distribution. As illustrated in the following section, specifying the reliability performance using only (6) when the Tx has limited knowledge of the channel is no longer sufficient.

III. Reliability Guarantees Under Limited Channel Knowledge

A. Perfect channel knowledge: ϵ-outage capacity

First, consider the benchmark case when the Tx *perfectly* knows F. In such circumstances, the Tx can easily determine

[1]Choosing the received signal power instead of the channel state information as a channel statistics implies that we assumed the transmit power to be fixed and known in advance. In such case, using the received signal power is equivalent to using the squared amplitude of the channel state information (more precisely, they are equivalent up to a scaling factor represented by the Signal-to-Noise Ratio).

[2]The reader will note that the definition of the training phase is vague, i.e., we intentionally do not specify the details on how the channel training is actually performed. In other words, our analysis is not bound to any specific standard or transmission format and is therefore valid for variety of channel training schemes. At one end, those are the conventional, dedicated pilot signals and training sequences which are commonly implemented in practical deployments. At the other end, the history of previous data transmissions and the associated feedback can be used as an input to the training process.

[3]Throughout the paper, we use small case letters to denote specific realizations of random variables.

[4]Note that in the definition (4) unit bandwidth is assumed; hence, strictly speaking R represents the transmission rate per unit bandwidth, i.e., spectral efficiency. However, we do stick to using "rate" throughout the paper for brevity.

0090-6778 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2019.2907241, IEEE Transactions on Communications

the maximum rate as a function of F at which the outage probability is no larger than ϵ, i.e., (6) can be guaranteed deterministically:

$$R_\epsilon(F) = \sup\{R \geq 0 : p_F(R) \leq \epsilon\} \tag{7}$$
$$= \log_2\left(1 + F^{-1}(\epsilon)\right). \tag{8}$$

The term $R_\epsilon(F)$ is also known as the ϵ-outage capacity, whereas $F^{-1}(\epsilon)$ is the ϵ-quantile of F. For parametric models, knowing the channel implies that $\boldsymbol{\theta}$ is known perfectly and we use the notation $R_\epsilon(\boldsymbol{\theta})$.

B. Limited channel knowledge: MLE of the ϵ-outage capacity

To illustrate the impact of channel uncertainty, consider the following: the Tx knows that $Y \sim F_{\boldsymbol{\theta}}$ but has no knowledge of $\boldsymbol{\theta}$. Having acquired a training sample x^n prior to transmission, the Tx can learn $\boldsymbol{\theta}$ via MLE as follows

$$\hat{\boldsymbol{\theta}}(x^n) = \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sum_{i=1}^{n} \log F'_{\boldsymbol{\theta}}(x)|_{x=x_i}. \tag{9}$$

Then, the "plug-in" estimator

$$R(x^n) = R_\epsilon\Big(\hat{\boldsymbol{\theta}}(x^n)\Big) \tag{10}$$

is MLE of the ϵ-outage capacity. Selecting the rate this way is naïve due to the uncertainty of $R(X^n)$, which is itself a random variable. It follows from (5) that different R yields different outage probability for fixed F; in other words, the random sequence X^n induces a distribution on the outage probability and the Tx can no longer guarantee with certainty that the outage probability for the transmission rate $R(X^n)$, denoted by $p_{\boldsymbol{\theta}}(R(X^n))$, will be less than or equal to ϵ. In fact, if the probability density function of the outage probability is symmetric (i.e., the mean and the median coincide), we have that $\mathbb{P}\left[p_{\boldsymbol{\theta}}(R(X^n)) > \epsilon\right] = 0.5$ which in some practical setups is clearly unacceptable, as discussed below.

C. Problem formulation

The discussion in the previous subsection shows that, when the Tx has a limited knowledge about the channel, it can only guarantee the reliability in a probabilistic manner. Formally stated, this is done by defining a *rate-selection function* $R(X^n)$ such that a predetermined statistical reliability constraint is satisfied.

1) Statistical reliability constraints: We consider two different approaches, resulting in two types of constraints, each of them suited to a specific set of use cases.

Averaged Reliability (AR): We consider the probability

$$\overline{p}_F = \mathbb{P}\left[R(X^n) > \log_2(1+Y)\right] \tag{11}$$

computed w.r.t. the joint distribution of X^n and Y. Conditioning on X^n and using the law of total expectation, (11) can be rewritten as

$$\overline{p}_F = \mathbb{E}[\mathbb{P}[R(X^n) > \log_2(1+Y)|X^n]] \tag{12}$$
$$= \mathbb{E}[p_F(R(X^n))], \tag{13}$$

where the outer expectation is taken over the distribution of the training sample X^n. It follows that $\overline{p}_F$ is the *mean* of the outage probability, as defined in (5), averaged over the training sample X^n. We consider the reliability constraint

$$\sup_{F\in\mathcal{F}} \overline{p}_F \leq \epsilon \tag{14}$$

which controls the *worst-case* mean outage probability over the whole class $\mathcal{F}$ and provides a firm reliability guarantee.

The mean outage probability (11) can be used when the Tx's objective is to optimize the transmission rate *jointly* over the training and the transmission, when training is performed *prior* to each transmission. This approach is suitable in a dynamic environment in which the channel changes frequently, requiring frequent channel training and estimation; an example of this is a vehicular communication scenario.

Probably Correct Reliability (PCR): Differently from the cases suited for AR, in many URLLC applications, such as monitoring and control in power grids or automated production in industrial complexes, the environment remains reasonably static for long periods of time. In such cases, it makes sense to train the channel infrequently implying that the most recent channel estimate will be used by the system over many future transmissions. Clearly, the Tx has to be more conservative here and choose the rate such that the outage probability in all the following transmissions is below ϵ with high probability.

This scenario is captured by PCR, which is more restrictive and effectively controls the higher order moments of the distribution of the outage probability. PCR is suitable when the Tx sets the transmission rate for all future transmissions *after* obtaining the training sample. We rely on the concept of meta-probability [15] and introduce

$$\widetilde{p}_F = \mathbb{P}\left[\mathbb{P}[R(X^n) > \log_2(1+Y)|X^n] > \epsilon\right] \tag{15}$$
$$= \mathbb{P}[p_F(R(X^n)) > \epsilon], \tag{16}$$

where the outer probability is computed w.r.t. the joint distribution of the training sample X^n. In PCR we define the statistical reliability constraint

$$\sup_{F\in\mathcal{F}} \widetilde{p}_F \leq \xi. \tag{17}$$

The probability ξ bounds the worst-case probability that the conditional outage probability (5) given X^n is larger than ϵ. Borrowing the terminology from PAC learning, ξ is the confidence parameter that indicates how likely it is to meet the reliability requirement [12].

So far, we have omitted to include the latency explicitly in the definition of the reliability metric (5). Nevertheless, the reader should note that the latency is included implicitly in the way the new reliability constraints are defined; in PCR for instance, the Tx does not have channel state information (CSI) before each transmission as there is no time to convey the CSI due to strict latency constraints.

2) Rate-selection function: There is a whole family of rate-selection functions $R(X^n)$ that satisfy (14) or (17). In order to find the most favorable, one should introduce an objective function of $R(X^n)$ and define an optimization problem that will give the *optimal* rate-selection function subject to (14)/(17). A possible objective function might be the ratio

$$\omega_\epsilon^n(F) = \frac{\mathbb{E}\left[R(X^n)1_{R(X^n)\leq\log_2(1+Y)}\right]}{R_\epsilon(F)(1-\epsilon)} \tag{18}$$

0090-6778 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2019.2907241, IEEE Transactions on Communications

between the throughput using $R(X^n)$ and the optimal throughput, given that F is known perfectly. Formulating such optimization problem using (18) is a non-trivial task and therefore out of the scope of the paper. One issue is the fact that (18) depends on a specific and fixed true distribution F while our aim is to design a robust rate-selection function that maximizes the transmission rate over potentially large class $\mathcal{F}$ of channel distributions.

In the rest of the paper, we will limit our discussion to the heuristic, yet intuitive choice, inspired by the naïve, MLE-based approach described in the previous subsection. Namely, the Tx uses

$$R(X^n) = \log_2\left(1 + \widehat{F}^{-1}(\varepsilon_n)\right), \tag{19}$$

where $\widehat{F}^{-1}(\varepsilon_n)$ is an estimate of the ε_n-quantile of the channel for some positive sequence $\varepsilon_n > 0$ and training sample X^n. The objective now is to find ε_n that *maximizes* $R(X^n)$ while meeting either (14) or (17). Note that if $\varepsilon_n = \epsilon$ for every n, we have the plug-in solution where $R(X^n)$ is just the MLE of the ϵ-outage capacity. Note that the rate-selection function (19) is still an estimate of the ϵ-outage capacity; however, by choosing ε_n such that (14)/(17) is satisfied, intuitively the Tx controls the uncertainty of the transmission rate introduced by limited channel knowledge.

We next introduce the notion of *consistency*. Namely, a rate-selection function is said to be consistent if $R(X^n)$ converges to the ϵ-outage capacity $R_\epsilon(F)$ as $n \to \infty$ with probability 1 for all $F \in \mathcal{F}$ while simultaneously satisfying either (14) or (17). In such case

$$\lim_{n\to\infty} \omega_\epsilon^n(F) = 1. \tag{20}$$

Clearly, consistent rate-selection functions are desirable, but, depending on the relation between the model and the true distribution, not always possible. Specifically, if there is a model mismatch, such that the model differs from the actual channel distribution, then the rate-selection function will not be consistent and $\lim_{n\to\infty} \omega_\epsilon^n(F) \neq 1$. This phenomenon, which can be also linked to the bias-variance trade-off [12], is discussed in more detail in the following section.

IV. Parametric Rate-selection Functions

We begin by considering parametric channel models. In this case, (19) can be rewritten as

$$R(X^n) = R_{\varepsilon_n}\left(\hat{\boldsymbol{\theta}}(X^n)\right) \tag{21}$$

for some $\varepsilon_n > 0$; here, $\hat{\boldsymbol{\theta}}$ is the MLE of the parameter $\boldsymbol{\theta}$ using the training sample X^n. We illustrate some of our main insights through several case studies for which we impose different assumptions regarding the true distribution and the state of knowledge at the Tx. In particular, we focus on the following cases:

1) The channel is Rayleigh-distributed and the Tx knows this; however, the Tx has no knowledge of λ.
2) The channel is not Rayleigh distributed and the Tx does not know this; nevertheless, the Tx still assumes that the channel is Rayleigh-distributed, with unknown λ.

For known λ, the ϵ-outage capacity is easily computed as

$$R_\epsilon(\lambda) = \log_2(1 - \lambda \log(1-\epsilon)). \tag{22}$$

Note that, even for the simple and intuitive choice (21), finding ε_n that maximizes $R(X^n)$ such that (14)/(17) is satisfied, is a non-trivial exercise for most of the remaining parametric channel models and one needs to resort to numerical methods to compute the transmission rate.

A. The true distribution is Rayleigh

The MLE of λ under Rayleigh fading is just the sample mean and the transmission rate for specific training sample x^n becomes

$$R(x^n) = \log_2\left(1 - \frac{\log(1-\varepsilon_n)}{n}\sum_{i=1}^n x_i\right). \tag{23}$$

We show how to find ε_n for any n such that (23) is maximized and either (14) or (17) is satisfied. Specifically, the AR as defined in (11) can be computed as

$$\overline{p}_\lambda = \mathbb{E}\left[F_\lambda\left(-\frac{\log(1-\varepsilon_n)}{n}\sum_{i=1}^n X_i\right)\right] \tag{24}$$

$$= 1 - \mathbb{E}\left[\exp\left\{\frac{\log(1-\varepsilon_n)}{n\lambda}\sum_{i=1}^n X_i\right\}\right] \tag{25}$$

$$= 1 - \left(1 - \frac{\log(1-\varepsilon_n)}{n}\right)^{-n}. \tag{26}$$

In (25) we have used (1) and in (26) the moment generating function (MGF) of an exponential random variable with mean λ given by $M_\lambda(t) = (1-t\lambda)^{-1}$. Now, ε_n can be computed from (14) by equating (26) with ϵ; we obtain

$$\varepsilon_n = 1 - e^{-n\left((1-\epsilon)^{-\frac{1}{n}} - 1\right)}. \tag{27}$$

Interestingly, for the specific case of Rayleigh channel, the mean outage probability in AR does not depend on λ and selecting ε_n according to (27) gives a transmission rate that satisfies (14) for all λ.

Similarly, the meta-probability (15) in PCR evaluates to

$$\widetilde{p}_\lambda = \mathbb{P}\left[F_\lambda\left(-\frac{\log(1-\varepsilon_n)}{n}\sum_{i=1}^n X_i\right) > \epsilon\right] \tag{28}$$

$$= \mathbb{P}\left[1 - \exp\left\{\frac{\log(1-\varepsilon_n)}{n\lambda}\sum_{i=1}^n X_i\right\} > \epsilon\right] \tag{29}$$

$$= \mathbb{P}\left[\sum_{i=1}^n X_i > n\lambda\frac{\log(1-\epsilon)}{\log(1-\varepsilon_n)}\right] \tag{30}$$

$$= 1 - \frac{\gamma\left(n, n\frac{\log(1-\epsilon)}{\log(1-\varepsilon_n)}\right)}{(n-1)!}. \tag{31}$$

In (31) we used the CDF of an Erlang-n random variable with shape parameters λ (obtained as a sum of n i.i.d. exponential random variables with mean λ), given by $\mathbb{P}\left[\sum_{i=1}^n X \leq x\right] = \gamma(n, x\lambda)/(n-1)!$. Observe again that, similarly as in AR, the

0090-6778 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2019.2907241, IEEE Transactions on Communications

meta-probability in PCR does not depend on λ. By choosing ε_n as the maximum value satisfying

$$1 - \frac{\gamma\left(n, n\frac{\log(1-\epsilon)}{\log(1-\varepsilon_n)}\right)}{(n-1)!} \leq \xi, \tag{32}$$

we obtain a rate-selection function that meets (17) for any λ.

B. Mismatch: The true distribution is not Rayleigh

Next, we study the impact of channel mismatch on the reliability performance of the system. We assume that the channel is no longer Rayleigh, i.e., $F_{\boldsymbol{\theta}}$ is different from (1); yet, the Tx maintains the assumption that the channel is Rayleigh, setting the rate as in (23) with ε_n computed via (27)/(32). Due to mismatch, the rate-selection function can no longer be guaranteed to be consistent, i.e., $R(X^n)$ does not converge to $R_\epsilon(\boldsymbol{\theta})$ computed w.r.t. the true distribution $F_{\boldsymbol{\theta}}$. As a result, one can no longer guarantee that the reliability constraints (14) or (17) will be satisfied.

Given that the Tx sets the transmission rate as in (23), the mean outage probability (11) and the meta-probability (15) can be written as

$$\overline{p}_{\boldsymbol{\theta}} = \mathbb{E}\left[F_{\boldsymbol{\theta}}\left(-\frac{\log(1-\varepsilon_n)}{n}\sum_{i=1}^{n} X_i\right)\right], \tag{33}$$

$$\tilde{p}_{\boldsymbol{\theta}} = \mathbb{P}\left[F_{\boldsymbol{\theta}}\left(-\frac{\log(1-\varepsilon_n)}{n}\sum_{i=1}^{n} X_i\right) > \epsilon\right]. \tag{34}$$

The outer expectations are taken w.r.t. the true distribution $F_{\boldsymbol{\theta}}$. The above quantities can be easily evaluated numerically for any $F_{\boldsymbol{\theta}}$. To gain more insight into the impact of channel mismatch, we derive simple approximations relaying on the assumption that the outage probability conditioned on x^n is small, i.e., in the order of ϵ. In such case, for wide variety of channels, $F_{\boldsymbol{\theta}}$ can be approximated via simple *power law* (see [6] for detailed derivations based on first-order approximations for different channels and Section VI for an argument from the extreme value theory)

$$F_{\boldsymbol{\theta}}(y) \approx \alpha_{\boldsymbol{\theta}} y^{1/\kappa_{\boldsymbol{\theta}}},\ y \to 0, \tag{35}$$

where $\alpha_{\boldsymbol{\theta}}$ and $\kappa_{\boldsymbol{\theta}}$ depend on the true distribution $F_{\boldsymbol{\theta}}$. Then, we obtain the simple approximations

$$\overline{p}_{\boldsymbol{\theta}} \approx \alpha_{\boldsymbol{\theta}}\left(-\log(1-\varepsilon_n)\mathbb{E}_{\boldsymbol{\theta}}[X]\right)^{1/\kappa_{\boldsymbol{\theta}}}\left(1 + \frac{1-\kappa_{\boldsymbol{\theta}}}{2n\kappa_{\boldsymbol{\theta}}^2}\frac{\mathrm{Var}_{\boldsymbol{\theta}}[X]}{\left(\mathbb{E}_{\boldsymbol{\theta}}[X]\right)^2}\right), \tag{36}$$

$$\widetilde{p}_{\boldsymbol{\theta}} \approx \exp\left\{\frac{t^* n}{\log(1-\varepsilon_n)}\left(\frac{\epsilon}{\alpha_{\boldsymbol{\theta}}}\right)^{\kappa_{\boldsymbol{\theta}}}\right\}(M_{\boldsymbol{\theta}}(t^*))^n. \tag{37}$$

In (36) we used a second-order Taylor expansion to approximate the expectation $\mathbb{E}\left[\left(\sum_i X_i\right)^{1/\kappa_{\boldsymbol{\theta}}}\right]$ and in (37) we used the Chernoff method to approximate $\mathbb{P}\left[\sum_i X_i > x\right]$; $M_{\boldsymbol{\theta}}(t^*)$ is MGF of X with t^* obtained as

$$t^* = \inf_{t>0} \exp\left\{\frac{tn}{\log(1-\varepsilon_n)}\left(\frac{\epsilon}{\alpha_{\boldsymbol{\theta}}}\right)^{\kappa_{\boldsymbol{\theta}}}\right\}(M_{\boldsymbol{\theta}}(t))^n. \tag{38}$$

Examples: Consider the Ricean channel (2) whose power law approximation has the form [6]

$$F_{k,\lambda}(y) \approx \frac{e^{-k}}{\lambda} y, \tag{39}$$

i.e., $\alpha_{k,\lambda} = e^{-k}/\lambda$ and $\kappa_{k,\lambda} = 1$. Note that the Rician channel asymptotically exhibits the same slope as the Rayleigh channel but with different scaling. The mean outage probability is

$$\overline{p}_{\lambda,k} \approx -\frac{k+1}{e^k}\log(1-\varepsilon_n). \tag{40}$$

Note that under Rayleigh fading, $k = 0$ and (26) can be approximated as $\overline{p}_\lambda \approx -\log(1-\varepsilon_n)$. Hence, we see from (40) that in case of Rician fading but with transmission rate dimensioned for the empirical channel mean, the mean outage probability is always smaller than ϵ, i.e., the transmission rate will always be pessimistic and lower than the maximum rate the Rician channel can support for outage ϵ.

Now, let us consider the Nakagami-m channel for $0.5 \leq m \leq 1$ (note that for $m = 0.5$, the distribution of received amplitude $\sqrt{Y}$ under Nakagami-m fading is exponential); the power law approximation obtains the form

$$F_{\lambda,m}(y) \approx \frac{1}{\lambda^m \Gamma(m+1)} y^m \tag{41}$$

and the mean outage probability can be approximated as

$$\overline{p}_{\lambda,m} \approx \frac{(-m)^m}{\Gamma(m+1)}(\log(1-\varepsilon_n))^m. \tag{42}$$

It is easy to check that for $0.5 \leq m \leq 1$, the above expression is always larger than ϵ, i.e., the transmission rate will be always optimistic and larger than the maximum rate the Nakagami-m channel offers for given ϵ. These results fit well the intuition: for equal average power λ of the diffuse component, the Rayleigh CDF (1) is always an upper/lower bound bound on the Rician/Nakagami-m (for $m < 1$) CDFs given in (2)/(3).

Another interesting observation that can be deduced from (26), (40) and (42) concerns the impact of the training sample size n on the convergence. Provided that ε_n is selected as in (27), we have that $-\log(1-\varepsilon_n) = n\left((1-\epsilon)^{-\frac{1}{n}} - 1\right)$. Given that $\epsilon/n \ll 1$ for any $n \geq 1$, and applying the binomial approximation, we obtain

$$\overline{p}_\lambda \approx \epsilon,\ \overline{p}_{\lambda,k} \approx \frac{k+1}{e^k}\epsilon,\ \overline{p}_{\lambda,m} \approx \frac{m^m}{\Gamma(m+1)}\epsilon^m. \tag{43}$$

We conclude that the mean outage probability depends very weakly on n. This implies that the transmission rate calculated under AR (14) converges quickly.

Now, we turn to PCR; in order to compute the meta-probability, we observe that the MGF of X under Rician fading is given by

$$M_{\lambda,k}(t) = \frac{\exp\left\{\frac{k\lambda t}{1-\lambda t}\right\}}{1-\lambda t} \tag{44}$$

for $t < 1/\lambda$. Hence, we obtain

$$\widetilde{p}_{\lambda,k} \approx \frac{\exp\left\{\left(\frac{k}{1-\lambda t^*} + \frac{\epsilon e^k}{\log(1-\varepsilon_n)}\right) n\lambda t^*\right\}}{(1-\lambda t^*)^n}, \tag{45}$$

0090-6778 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2019.2907241, IEEE Transactions on Communications

$$R(x^n) = \sup\Big\{R > 0 : \widehat{F}\left(2^R - 1\right) \le \epsilon\Big\} \tag{52}$$
$$= \sup\big\{R > 0 : \exists i \in \{1, \dots, n+1\}, x_{(i-1)} \le 2^R - 1 < x_{(i)}, i \le n\epsilon + 1\big\} \tag{53}$$
$$= \sup\big\{R > 0 : x_{(\lfloor n\epsilon+1\rfloor - 1)} \le 2^R - 1 < x_{(\lfloor n\epsilon+1\rfloor)}\big\} \tag{54}$$
$$= \sup\big\{R > 0 : \log_2(1 + x_{(\lfloor n\epsilon+1\rfloor-1)}) \le R < \log_2(1 + x_{(\lfloor n\epsilon+1\rfloor)})\big\} \tag{55}$$
$$= \log_2(1 + x_{(\lfloor n\epsilon+1\rfloor)}). \tag{56}$$

and t^* is the solution to the quadratic equation

$$n\lambda(1 - \lambda t + k) + \frac{\epsilon n\lambda e^k}{\log(1-\varepsilon_n)}(1 - \lambda t)^2 = 0 \tag{46}$$

that satisfies $t < 1/\lambda$. Similarly, the MGF of X under Nakagami-m fading is

$$M_{\lambda,m}(t) = \frac{1}{(1 - t\lambda)^m}, \tag{47}$$

for $t < 1/\lambda$; the meta-probability can be approximated as

$$\widetilde{p}_{\lambda,m} \approx \frac{\exp\left\{\frac{t^* n\lambda}{\log(1-\varepsilon_n)}(\epsilon\Gamma(m+1))^{\frac{1}{m}}\right\}}{(1 - t^*\lambda)^{mn}} \tag{48}$$

with

$$t^* = \frac{1}{\lambda}\left(1 + m\log(1 - \varepsilon_n)(\epsilon\Gamma(m+1))^{-1/m}\right). \tag{49}$$

The resulting approximations are less insightful than the approximations for the mean outage probability; in Section VII we show numerically that the mismatch impacts the meta-probability performance more severely.

V. Non-parametric Rate-selection Functions

Using non-parametric rate-selection approach is suitable when the Tx has no, or very limited knowledge of the channel distribution and is unwilling to impose assumptions that might lead to mismatch and compromise the reliability performance. Clearly, one major advantage of the non-parametric approach is its generality and versatility, i.e., it is applicable to wide variety of channels, subject only to some smoothness constraints, such as e.g. existence of the first-order derivative of F, see [16]. However, this comes at the expense of the duration of the channel training phase; in general, non-parametric approaches often require excessive training sample lengths for reasonably reliable system performance, as shown below.

A. Non-parametric estimation of the ϵ-outage capacity

Similarly as in the parametric case, we will first look into the plug-in estimate of the ϵ-outage capacity. We consider a general case: Y is drawn from an arbitrary distribution F, defined over the non-negative reals. Consider a specific training sample x^n; the empirical CDF

$$\widehat{F}(y) = \frac{1}{n}\sum_{i=1}^{n} 1_{x_i \le y} \tag{50}$$

for $y \ge 0$ serves as an estimate of F. Let $X_{(1)} \le \ldots \le X_{(n)}$ denote the order statistics formed from the elements of the X^n and define for convenience $X_{(0)} = 0$ and $X_{(n+1)} = \infty$. Then, for every $y \ge 0$, there exists an integer $i \in \{1, \dots, n+1\}$ such that

$$\widehat{F}(y) = \frac{i-1}{n} \quad \text{and} \quad x_{(i-1)} \le y < x_{(i)}. \tag{51}$$

Using (51), one can easily derive the plug-in estimate of the ϵ-outage capacity, see eq. (55)-(56) at the top of the page. From (55) we observe that the number of samples necessary to obtain reliability performance of ϵ should satisfy $\lfloor n\epsilon+1\rfloor > 0$ or, equivalently $n\epsilon > 1$; we see that the number of channel samples in purely non-parametric setup grows as

$$n \sim \frac{1}{\epsilon}. \tag{57}$$

In the context of URLLC with target ϵ in the order of 10^{-6} and below, the non-parametric approach requires excessive channel training.

B. Non-parametric rate-selection function

We see from (56) that in non-parametric setup the l-th order statistic $x_{(l)}$ is an estimate of the ϵ-quantile. Motivated by this, for a given training sample x^n, the Tx fixes the rate as

$$R(x^n) = \log_2(1 + x_{(l)}) \tag{58}$$

and chooses the largest l that meets the statistical reliability constraints. Thus, using (58), the outage probability (5) conditioned on X^n obtains the simple form

$$p_F(R(X^n)) = \mathbb{P}\left[R(X^n) \ge \log_2(1+Y)|X^n\right] \tag{59}$$
$$= F(X_{(l)}). \tag{60}$$

Now, let $U_1, \dots, U_n$ be independent random variables uniformly distributed $[0, 1]$. Then, $F(X_{(l)})$ has the same distribution as $U_{(l)}$ which is beta-distributed with shape parameters l and $n+1-l$. With this property, we easily evaluate the mean outage probability (11) as

$$\overline{p}_F = \mathbb{E}\left[p_F(R(X^n))\right] = \mathbb{E}\left[U_{(l)}\right] \tag{61}$$
$$= \frac{l}{n+1}. \tag{62}$$

Clearly, the mean outage does not depend on F and $\sup_{F\in\mathcal{F}} \overline{p}_F = l/(n+1)$. Hence, the constraint (14) gives

$$l \le \epsilon(n+1). \tag{63}$$

From the above rate-selection rule, it is evident that when the number of training samples $n < 1/\epsilon - 1$ which corresponds to $l < 1$, the transmission rate is $R(x^n) = 0$.

0090-6778 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2019.2907241, IEEE Transactions on Communications

Similarly, for the meta-probability (15), we obtain

$$\tilde{p}_F = \mathbb{P}\left[p_F(R(X^n)) > \epsilon\right] = \mathbb{P}\left[U_{(l)} > \epsilon\right] \quad (64)$$

$$= 1 - I_\epsilon(l, n+1-l), \quad (65)$$

where $I_x(a,b)$ is the regularized incomplete beta function. As in the mean outage approach, the meta-probability does not depend on F and the constraint (17) implies that l should be chosen as a solution to the equation:

$$1 - I_\epsilon(l, n+1-l) = \xi. \quad (66)$$

Since l is an integer, we choose l to be the largest integer satisfying $1 - I_\epsilon(l, n+1-l) \leq \xi$. Albeit the above implicit equation does not reveal immediate insights, the numerical evaluations in Section VII show that the meta-probability constraint requires even bigger training samples sizes.

VI. Power law approximation of the channel tail

In Section IV, we saw that the training sample size required to learn the channel for parametric channel models is relatively low; however, the model mismatch may have a significant impact on the realized reliability. The non-parametric method in Section V does not suffer from this drawback, but requires a training sample size of the order $1/\epsilon$ which is enormous for many practical applications. This section takes an approximate approach using power law approximations and asymptotic properties. To this end, we first argue why the power law approximation for lower tail of F, introduced in (35) is of interest.

The Pickands-Balkema-de Haan theorem theorem (see [17, Th. 2.1.1] or [18, Th. 21.17]) in extreme value theory states that, for a large class of distributions F, there exists a constant $\kappa > 0$ such that

$$\lim_{t\to 0} \frac{F(ty)}{F(t)} = y^{1/\kappa} \quad (67)$$

for every $y > 0$. Hence, justifying the use of the *power law* approximation

$$F(y) \approx \alpha y^{1/\kappa} \quad (68)$$

for small $y \geq 0$. Here, α and κ are parameters that depend on the true but unknown distribution F. For convenience, throughout the section, we consider the transformed variables $Z = \log(Y)$ and $Z^n = \{Z_i = \log(X_i)\}_{i=1}^n$. After this transformation, the power law approximation in (68) implies that

$$F_Z(z) \approx \alpha e^{z/\kappa}. \quad (69)$$

Under the additional assumption that F is smooth and that $F'(y) > 0$ for $y > 0$, the density of Z is also well-approximated as $f_Z(z) \approx \frac{\alpha}{\kappa} e^{z/\kappa}$.

The importance of the power law approximation is that it allows one to treat the tail of any distribution F, satisfying the conditions stated before, as a parametric distribution of only two parameters. We remark that there are three domains of attraction for the extreme value distribution and that (67) only captures one of these, see [6] for a list of common fading distributions satisfying the power law approximation and [17, Th. 2.1.2] sufficient conditions. As a result, we can take an approach similar to the one in Section IV for parametric distributions even though no full parameterization of F is given.

In order to estimate α and κ, we shall apply a variation of the ML estimator. In particular, given a training sample z^n and a small constant β, we use only the $l = \lceil \beta n \rceil$ smallest order statistics $z_{(1)}, \ldots, z_{(l)}$ and choose the parameters $\hat{\alpha}$ and $\hat{\kappa}$ such that the likelihood $f(z_{(1)}, \ldots, z_{(l)}; \hat{\alpha}, \hat{\kappa})$ is maximized. The intuition is that only the smallest observations contain information about the tail of F.

This approximate approach has some clear disadvantages; namely, the range of validity of the power law approximation depends crucially on the value of β and on the true distribution F. Hence, no strict statistical guarantees can be derived. It is, however, worth noting that in cases where no parameterization of F is available, only the non-parametric approach can provide such true statistical guarantees for reliability at the expense of a large required sample size, and approximate statistical guarantees may therefore be favored.

By using the power law approximation in (68) for approximating $f(z_{(1)}, \ldots, z_{(l)}; \hat{\alpha}, \hat{\kappa})$, we find that [19, Sec. 7.2.4]

$$f(z_{(1)}, \ldots, z_{(l)}; \alpha, \kappa) = \frac{n!}{(n-l)!} F_Z(z_{(l)})^{n-l} \prod_{i=1}^{l} f_Z(z_i) \quad (70)$$

$$\approx \left(\frac{\alpha}{\kappa}\right)^l \frac{n!}{(n-l)!} (1 - \alpha e^{z_{(l)}/\kappa})^{n-l} e^{\frac{1}{\kappa}\sum_{i=1}^{l} z_{(i)}}. \quad (71)$$

As a result, by differentiating the logarithm of the right-hand side of (71) with respect to α and κ, by equating with zero, and by solving for α and κ, we find the following simple expressions for the estimator $(\hat{\alpha}, \hat{\kappa})$

$$\hat{\kappa} = \frac{1}{l} \sum_{i=1}^{l} (Z_{(l)} - Z_{(i)}), \quad (72)$$

$$\hat{\alpha} = \frac{l}{n} e^{-Z_{(l)}/\hat{\kappa}}. \quad (73)$$

We let $\widehat{F}_Z(z)$ be the estimate of the tail distribution of Z given by $\hat{\alpha} e^{z/\hat{\kappa}}$. Then, we obtain an estimate of the lower ε_n-quantile of Z as follows:

$$\widehat{F}_Z^{-1}(\varepsilon_n) = \hat{\kappa} \log\Big(\frac{\varepsilon_n}{\hat{\alpha}}\Big) \quad (74)$$

$$= Z_{(l)} + \frac{1}{l} \log\Big(\frac{n\varepsilon_n}{l}\Big) \sum_{i=1}^{l} (Z_{(l)} - Z_{(i)}). \quad (75)$$

Under the power law approximation in (69), one can show that $Z_{(l)}$ and $\sum_{i=1}^{l}(Z_{(l)} - Z_{(i)})$ are independent random variables. In particular, $Z_{(l)}$ is asymptotically normal and $\sum_{i=1}^{l}(Z_{(l)} - Z_{(i)})$ is Erlang distributed with parameters $l-1$ and $1/\kappa$ and its distribution is not altered when conditioning on $Z_{(l)}$.

We now set the rate-selection function as

$$R(X^n) = \log_2\Big(1 + \exp\Big(\widehat{F}_Z^{-1}(\varepsilon_n)\Big)\Big), \quad (76)$$

where the exponential function is introduced to invert the log-transformation.

0090-6778 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2019.2907241, IEEE Transactions on Communications

We consider two different approaches for specifying ε_n. The first approach relies on asymptotic approximations of $\widehat{F}_Z^{-1}(\varepsilon_n)$ which holds in the limit $n \to \infty$. The second approach is simpler and does not rely on asymptotic approximations; however, it is computationally more tedious and does not provide simple analytical insights.

Asymptotic approach: Relying on the asymptotic normality of $\widehat{F}_Z^{-1}(\varepsilon_n)$, we choose ε_n according to the following simple formula:

$$\varepsilon_n = \epsilon \exp\left(-\sqrt{\frac{\overline{V}}{n}}Q^{-1}(\xi)\right), \tag{77}$$

where

$$\overline{V} = \frac{1}{\beta}\left(1-\beta+\log^2\frac{\epsilon}{\beta}\right). \tag{78}$$

Then, as shown in Appendix A, we obtain the following approximate meta-probability for PCR:

$$\lim_{n\to\infty}\widetilde{p}_F \approx \xi, \tag{79}$$

where the approximation stems from the power law approximation in (69). Hence, the approximation becomes increasingly more accurate as β is lowered towards zero. Similarly, it can be shown that $\varepsilon_n = \epsilon$ implies that

$$\lim_{n\to\infty}\overline{p}_F \approx \epsilon. \tag{80}$$

Non-asymptotic approach: While the asymptotic approach provides a simple closed-form expression for choosing the rate-selection function, asymptotic approximations are generally not accurate for small l. We can provide a simple, but less insightful, approximate bound for the meta-probability that does not exploit asymptotic normality as follows:

$$\widetilde{p}_F = \mathbb{P}\left[\widehat{F}_Z^{-1}(\varepsilon_n) > F_Z^{-1}(\epsilon)\right] \tag{81}$$

$$\leq \min_t\left\{\mathbb{P}\left[Z_{(l)} > t\right] + \mathbb{P}\left[\sum_{i=1}^{l}(Z_{(l)}-Z_{(i)}) > r\right]\right\} \tag{82}$$

$$\approx \min_{\tau\in[\epsilon,1]}\left\{2 - I_\tau(l,n+1-l) - \frac{\gamma\left(l-1,\frac{l\log\frac{\epsilon}{\tau}}{\log\left(\frac{n\varepsilon_n}{l}\right)}\right)}{(l-2)!}\right\}. \tag{83}$$

The bound (82) follows because, for any two random variables A and B, $\mathbb{P}[A+B>v] \leq \min_t\{\mathbb{P}[A>t]+\mathbb{P}[B>v-t]\}$; here, $r = l\left(\kappa\log\frac{\epsilon}{\alpha}-t\right)/\log\left(\frac{n\varepsilon_n}{l}\right)$. The approximation (83) follows from the substitution $\tau = \alpha e^{t/\kappa}$ and because, under the power law approximation, $\sum_{i=1}^{q}(Z_{(l)}-Z_{(i)})$ is Erlang distributed with parameters $1/\kappa$ and $l-1$ and $\mathbb{P}\left[Z_{(l)}\leq t\right] = \mathbb{P}\left[U_{(l)} \leq F_Z(t)\right] \approx I_\tau(l,n+1-l)$ where $U_{(l)}$ is the l-th order statistic of a sequence of n independent standard uniform random variables as in Section V.

By choosing ε_n such that right-hand side of (83) equals ξ, we find that

$$\widetilde{p}_F \lessapprox \xi. \tag{84}$$

The approximation in (84) again originates from the power law approximation in (69).

VII. Numerical evaluation

Before presenting the results, we advise the reader to refer to the caption of the individual figures for details regarding the notation of the curves.

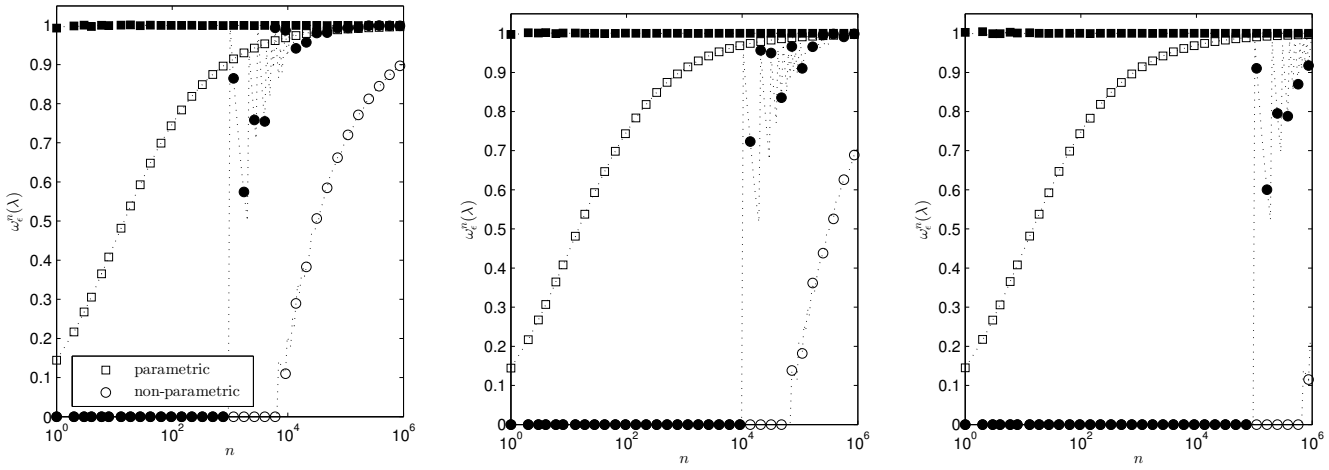
We begin by considering transmission over Rayleigh flat fading channel with average received power $\lambda = 1$. Fig. 1 shows the throughput ratio $\omega_\epsilon^n(\lambda)$ as defined in (18) for parametric (21) (square markers) and non-parametric (circle markers) rate-selection functions (58) for different training sample lengths n and $\epsilon \in \{10^{-3}, 10^{-4}, 10^{-5}\}$. We observe that the parametric and non-parametric rate-selection functions are consistent; the oscillations in the non-parametric case arise due to $(l) \in \mathbb{N}$. We also observe that the parametric rate-selection functions converge significantly faster to the ϵ-outage capacity. As expected, the non-parametric rate-selection functions require n to be of the order $1/\epsilon$ to produce non-zero throughput; the PCR constraint requires n to be even larger, almost an order of magnitude larger than the AR constraint.

We observe that $\omega_\epsilon^n(\lambda)$ approaches 1 faster when the rate-selection follows the AR constraint (14) (filled markers) as opposed to the PCR constraint (17) (empty markers); in fact, $\omega_\epsilon^n(\lambda) \approx 1$ and $R(x^n) \approx R_\epsilon(\lambda)$ even for $n < 10$. This is expected since the mean outage probability depends very weakly on n (see (43)). In contrast, the rate under meta-probability constraint converges slower which is also intuitively expected due to the strictness of the constraint. An interesting observation follows from Fig. 1: the convergence rate under the parametric rate-selection function appears to be (almost) independent from ϵ for fixed ξ, implying that the rate-selection function obtained for given ξ via (32) is valid for any ϵ. In Fig. 2 we depict the throughput ratio under PCR constraint for different values of the confidence parameter ξ (square/circle markers correspond to parametric/non-parametric rate selection). Note that lower values for ξ impose stricter requirements; this implies lower rate for fixed n and slower convergence. Interestingly, when $\xi \to 0.5$ we obtain similar behavior as in the case of AR-constrained rate-selection. In fact, when the distribution of the outage probability is symmetric, i.e., its mean coincides with the median, the PCR constraint for $\xi = 0.5$ is equivalent to the AR constraint. Hence, even though rigorously precise only in the case of symmetric outage probability distribution, the PCR (14) can, in general, be viewed as generalization of the AR constraint (17).

Fig. 3 evaluates the impact of mismatched model on the reliability performance: the Tx adopts Rayleigh model (filled markers), but the true distribution is different. We fix ϵ and ξ, we choose n to be large enough to ensure convergence and we plot the mean outage probability and the meta-probability as a function of the true channel distribution as specified by the corresponding parameters. In Fig. 3a, the channel follows Rice distribution (2) and we plot the reliability performance for a range of k-factors: note that, we plot both (33)/(34) with the corresponding approximations (36)/(37) (dotted lines), confirming that the latter approximate the former well for small ϵ and ξ. As already discussed in Section IV-B, assuming Rayleigh when the actual fading is Rician, always gives a

0090-6778 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2019.2907241, IEEE Transactions on Communications



(a) $\epsilon = 10^{-3}, \xi = 10^{-3}$ (b) $\epsilon = 10^{-4}, \xi = 10^{-3}$ (c) $\epsilon = 10^{-5}, \xi = 10^{-3}$

Fig. 1. Parametric vs non-parametric rate-selection under Rayleigh fading with $\lambda = 1$ (full/empty markers correspond to AR/PCR constraints (14)/(17)), respectively.

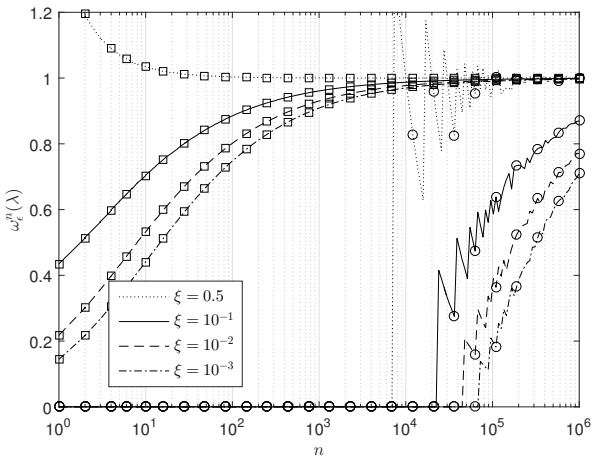


Fig. 2. Parametric rate-selection under Rayleigh fading and PCR constraint $\lambda = 1$, $\epsilon = 10^{-4}$ (square/circle markers correspond to parametric/non-parametric rate-selection functions, respectively).

pessimistic rate-selection function, i.e., the throughput ratio is strictly less than 1; in fact, as the specular component becomes stronger, the throughput becomes even lower. We conclude that by assuming Rayleigh, i.e., $k = 0$ when $k > 0$, the reliability constraints (14) and (17) will never be violated. However, this comes at the price of under-utilizing the degrees of freedom offered by the Rician channel; under PCR constraint, the underutilization is severe as seen in Fig. 3a, with the meta-probability quickly dropping and pushing the throughput towards 0. For Fig. 3b the true distribution is Nakagami-m (3) and it shows the reliability performance for different values of m. For $m > 1$, the behavior is similar to the Rician case. However, the numerical evaluations suggest that the under-utilization is more severe; already for $m = 2$, which corresponds to two Rayleigh-type of clusters, both the mean outage probability and the meta-probability are several order of magnitudes below the thresholds, implying that the throughput is very low. We observe opposite behavior for $m < 1$ (which we also predicted in Section IV-B). Here, the rate-selection function gives optimistic rates, larger than the corresponding ϵ-outage capacity; such rates violate the reliability constraints to produce throughput ratio larger than 1. Once again, the meta-probability shows severe under/over-utilization, quickly dropping/jumping towards $0/1$.

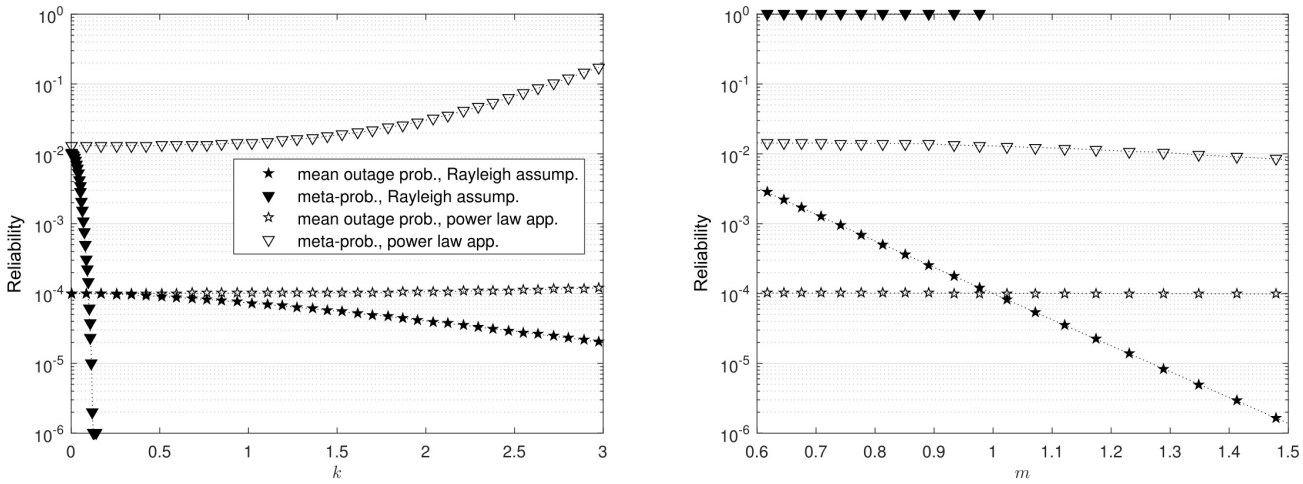
The performance of the parametric and non-parametric rate-selection functions can be also linked to the bias-variance trade-off. Namely, learning parametric models via MLE produces estimates with small variance which explains the relatively fast convergence of the rate, but they can be heavily biased in case of model mismatch. On the other hand, in the non-parametric approach the ε_n-quantile is estimated with a small bias but a large variance; hence, it works for any distribution F but requires a large number of samples.

The asymptotic rate-selection functions based on power law tail approximation address the trade-off among consistency, training sample length, and mismatch. Fig. 4 compares the performance of the power law tail approach against parametric/non-parametric rate-selection functions with Rayleigh true distribution; in such case $\kappa = 1$, $\alpha = 1/\lambda$. Clearly, the throughput ratio converges faster than the non-parametric case, which is particularly evident for the meta-probability constraint. Note that the approach requires $l \geq 2$, i.e., $n \geq 2/\beta$ samples. Also, for small l (less than 10 in this case study), the estimators (72) and (73) overestimate κ and α. For the meta-probability constraint, this effect is (partially) alleviated by using ε_n chosen according to the upper bound (81) at the expense of slower convergence. The rate-selection function (76) also suffers from mismatch due to approximation error; therefore, the approach does not guarantee consistency.

To investigate the effect, Fig. 3 shows the reliability per-

0090-6778 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2019.2907241, IEEE Transactions on Communications



(a) True distribution: Rician fading ($\lambda = 1$)

(b) True distribution: Nakagami-m fading ($\lambda = 1$)

Fig. 3. Impact of channel mismatch onto AR- and PCR-constrained rate-selection with $\epsilon = 10^{-4}, \xi = 10^{-2}, n = 10^6, \beta = 0.01$ (full/empty markers correspond to Rayleigh fading assumption/power law tail approximation, respectively).

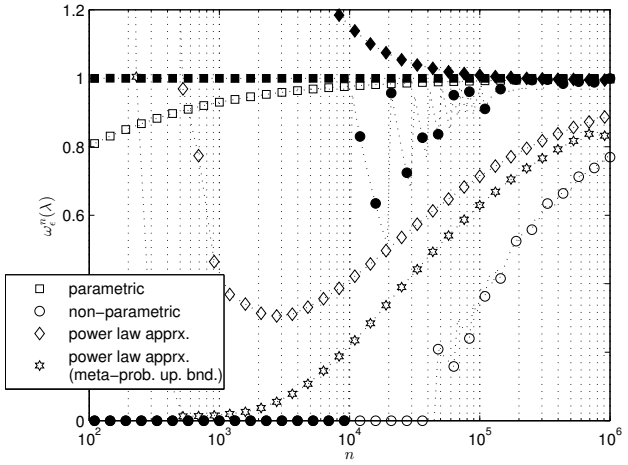


Fig. 4. Parametric, non-parametric and power law-based rate-selection under Rayleigh fading (full/empty markers correspond to AR/PCR constraints, $\epsilon = 10^{-4}, \xi = 10^{-2}, \beta = 0.01, \lambda = 1$).

formance of the approach (empty markers), compared against the case of incorrect model assumption. In both Figs. 3a and 3b, we see that even in Rayleigh case ($k = 0$ or $m = 1$, recall the special cases of the Rician or Nakagami-m channel in Section II) there is slight bias in the reliability performance which is more evident for the PCR constraint; hence, the rate is slightly inconsistent due to the approximation error. When the true distribution is Ricean, we see that the power law tail approach violates both constraints for large k, producing optimistic rates; similarly, in the case of Nakagami-m and $m < 1$, the approach gives optimistic rates. This is due to the fact that the power law tail approximation is a lower bound of the respective cdfs (see [6] for more details). Oppositely, for large enough $m > 1$ in the Nakagami-m channel, the approach leads to pessimistic rates since the power law approximation becomes an upper bound [6].

VIII. Conclusions and Discussion

The strict requirements of URLLC systems demand from us to reconsider the usual ways the physical layer is designed and its performance assessed. Our study shows that the phenomena of channel uncertainty strongly impact the performance in UR-relevant regime of operation, in the sense that the targeted reliability can no longer be guaranteed with certainty. Motivated by this, we introduced novel statistical framework in which the transmitter, given its best knowledge of the true channel distribution, determines the physical layer parameters, namely the transmission rate such that the reliability is guaranteed probabilistically. We showed that the approach based on parametric channel models requires the least amount of channel training but is highly susceptible to modeling mismatch. Non-parametric approaches do not suffer from such mismatch since they do not impose assumptions; however, the training sample sizes tend to be prohibitively large. As a consensus between the parametric and non-parametric approaches, the design approach based on the power law tail approximations offers versatile performance that reduces both, the impact of modeling mismatch and the training sample sizes.

Finally, some remarks on the perspectives and future work. Despite the fact that our work is limited to the simple case of one-way transmission, we note that the statistical treatment we introduce here opens up a methodology that can be readily extended to more advanced and complex transmission and networking scenarios that also introduce other sources of uncertainties such as ARQ, interference, etc. Most importantly, the framework can be extended to apply to higher layers as well. However, such an endeavor would require adequate redefinition of the reliability metric. For instance, when considering ARQ mechanisms, a possible reliability metric can be the average number of retransmissions until the packet has been delivered. When going to the higher layers, such as the

0090-6778 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2019.2907241, IEEE Transactions on Communications

application layer which employs TCP-like transport protocol, the reliability can be defined integrally as the probability of successful end-to-end packet delivery within given latency budget. Note that, in all cases above, the reliability is a probabilistic quantity, so one can apply the same statistical reliability constraints introduced in Section III for selecting the relevant communication parameters. In addition to the relatively straightforward appeal of our framework to higher layers, the analysis presented in this paper can also serve as a building block for more exhaustive characterization of the physical layer and analysis of advanced physical layer architectures, such as multiple antenna techniques. For instance, one can use the proposed framework to study the phenomenon of channel hardening in massive MIMO systems from a statistical learning perspective and quantify the impact of number of antennas, channel model, etc. on how fast the channel hardens; such investigations are part of our on-going research [3].

APPENDIX A
PROOF OF (79)

We are going to show that, under the power law approximation,

$$\sqrt{n}\left(\widehat{F}_Z^{-1}(\varepsilon_n) - \kappa \log\frac{\epsilon}{\alpha} + \sqrt{\frac{\kappa^2\overline{V}}{n}}Q^{-1}(\xi)\right) \tag{85}$$

is asymptotically normal with asymptotic mean and asymptotic variance given by 0 and $\kappa^2\overline{V}$, respectively. Let W be a standard normal random variable. Then, we obtain (79) through the following steps

$$\begin{aligned}
&\lim_{n\to\infty} \mathbb{P}\left[\mathbb{P}\left[R(X^n) \geq \log(1+Y)|X^n\right] \geq \epsilon\right] \\
&= \lim_{n\to\infty} \mathbb{P}\left[\mathbb{P}\left[\widehat{F}_Z^{-1}(\varepsilon_n) > \log(Y)|X^n\right] \geq \epsilon\right] && (86)\\
&= \lim_{n\to\infty} \mathbb{P}\left[\widehat{F}_Z^{-1}(\varepsilon_n) > F_Z^{-1}(\epsilon)\right] && (87)\\
&\approx \mathbb{P}\left[W > \sqrt{\frac{n}{\overline{V}}}\left(\frac{F_Z^{-1}(\epsilon)}{\kappa} - \log\frac{\epsilon}{\alpha} + \sqrt{\frac{\overline{V}}{n}}Q^{-1}(\xi)\right)\right] && (88)\\
&\approx \mathbb{P}\left[W > Q^{-1}(\xi)\right] && (89)\\
&= \xi. && (90)
\end{aligned}$$

Here, (87) follows from the log-transformation which implies that $\mathbb{P}\left[\cdot \geq \log(Y)\right] = F_Z(\cdot)$, (88) follows from the asymptotic normality of (85), and (89) follows because $F_Z^{-1}(\epsilon) \approx \kappa \log\frac{\epsilon}{\alpha}$.

To establish asymptotic normality of (85) under the power law approximation, we first substitute the expressions for $\widehat{F}_Z^{-1}(\varepsilon_n)$ and ε_n (see (75) and (77)) and rewrite (85) as follows

$$\begin{aligned}
\sqrt{n}\Bigg(& Z_{(l)} - \kappa\log\frac{\beta}{\alpha} \\
&+ \frac{1}{l}\log\Big(\frac{n\epsilon}{l}\Big)\sum_{i=1}^{l}(Z_{(l)} - Z_{(i)}) - \kappa\log\frac{\epsilon}{\beta} \\
&- \sqrt{\frac{\overline{V}}{n}}Q^{-1}(\xi)\left(\frac{1}{l}\sum_{i=1}^{l}(Z_{(l)} - Z_{(i)}) - \kappa\right)\Bigg). \qquad (91)
\end{aligned}$$

We now consider each line of (91) separately. First, it follows from [18, Th. 21.7] that

$$\sqrt{n}\Big(Z_{(l)} - \kappa\log\frac{\beta}{\alpha}\Big) \xrightarrow{\mathrm{d}} \mathcal{N}\Big(0, \widetilde{V}\Big) \tag{92}$$

as $n \to \infty$, where

$$\widetilde{V} = \frac{\beta(1-\beta)}{f_Z^2(F_Z^{-1}(\beta))} \approx \frac{\kappa^2(1-\beta)}{\beta}. \tag{93}$$

Next, it follows from the standard central limit theorem [18, Th. 2.17] that

$$\sqrt{n}\left(\frac{1}{l-1}\log\left(\frac{\epsilon}{\beta}\right)\sum_{i=1}^{l}(Z_{(l)} - Z_{(i)}) - \kappa\log\frac{\epsilon}{\beta}\right) \tag{94}$$

is asymptotically normal with mean 0 and variance $\frac{\kappa^2}{\beta}\log^2\frac{\epsilon}{\beta}$ (recall that the mean and variance of an Erlang distribution with parameters $1/\kappa$ and $l-1$ is $\kappa(l-1)$ and $\kappa^2(l-1)$, respectively. An application of Slutsky's theorem [18, Th. 2.8] shows that the second line in (91) (multiplied by $\sqrt{n}$) has the same asymptotic distribution as (94). We finally note that

$$\frac{1}{l}\sum_{i=1}^{l}(Z_{(l)} - Z_{(i)}) \tag{95}$$

converges to κ almost surely as $n \to \infty$.

These three properties imply that (85) is asymptotically normal distributed with asymptotic mean and asymptotic variance given by 0 and $\frac{\kappa^2}{\beta}(1-\beta+\log^2\frac{\epsilon}{\beta}) = \kappa^2\overline{V}$, respectively, as desired.[5]

REFERENCES

[1] P. Popovski, "Ultra-reliable communication in 5G wireless systems," in *IEEE International Conference on 5G for Ubiquitous Connectivity*. IEEE, Jul. 2013, pp. 146–151.

[2] P. Popovski, J. J. Nielsen, C. Stefanovic, E. d. Carvalho, E. Strom, K. F. Trillingsgaard, A. Bana, D. M. Kim, R. Kotaba, J. Park, and R. B. Sorensen, "Wireless access for ultra-reliable low-latency communication: Principles and building blocks," *IEEE Network*, vol. 32, no. 2, pp. 16–23, March 2018.

[3] P. Popovski, C. Stefanovic, J. J. Nielsen, E. de Carvalho, M. Angjelichinoski, K. F. Trillingsgaard, and A.-S. Bana, "Wireless access in ultra-reliable low-latency communication (urllc)," *submitted to IEEE Transactions on Communications*, Oct. 2018, arXiv:1810.06938v1 [cs.IT].

[4] P. Schulz, M. Matthe, H. Klessig, M. Simsek, G. Fettweis, J. Ansari, S. A. Ashraf, B. Almeroth, J. Voigt, I. Riedel, A. Puschmann, A. Mitschele-Thiel, M. Muller, T. Elste, and M. Windisch, "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, February 2017.

[5] M. Bennis, M. Debbah, and H. V. Poor, "Ultra-reliable and low-latency wireless communication: Tail, risk and scale," *submitted to Proceedings of IEEE*, May 2018, arXiv:1801.01270 [cs.IT].

[6] P. C. F. Eggers, M. Angjelichinoski, and P. Popovski, "Wireless channel modeling perspectives for ultra-reliable communications," *IEEE Transactions on Wireless Communications*, pp. 1–1, 2019.

[7] V. N. Swamy, P. Rigge, G. Ranade, B. Nikolic, and A. Sahai, "Predicting wireless channels for ultra-reliable low-latency communications," in *2018 IEEE International Symposium on Information Theory (ISIT)*, June 2018, pp. 2609–2613.

[5]Note that if $A_n \xrightarrow{\mathrm{d}} \mathcal{N}(0,\sigma_1^2)$ and $B_n \xrightarrow{\mathrm{d}} \mathcal{N}(0,\sigma_2^2)$ as $n \to \infty$, then it follows from Lévy's continuity theorem [18, Th. 2.13] that $A_n + B_n \to \mathcal{N}(0, \sigma_1^2 + \sigma_2^2)$.

0090-6778 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2019.2907241, IEEE Transactions on Communications

[8] ——, "Wireless channel dynamics and robustness for ultra-reliable low-latency communications," *submitted to IEEE Journal on Selected Areas in Communications*, 2018, arXiv:1806.08777 [cs.IT].
[9] O. Simeone, "A very brief introduction to machine learning with applications to communication systems," *arXiv preprint arXiv:1808.02342*, 2018.
[10] R. Vaughan, J. Andersen, J. Andersen, and I. of Electrical Engineers, *Channels, Propagation and Antennas for Mobile Communications*, ser. Electromagnetics and Radar Series. Institution of Engineering and Technology, 2003. [Online]. Available: https://books.google.com/books?id=9GO0ZKgKoIYC
[11] L. G. Valiant, "A theory of the learnable," *Commun. ACM*, vol. 27, no. 11, pp. 1134–1142, Nov. 1984. [Online]. Available: http://doi.acm.org/10.1145/1968.1972
[12] S. Shalev-Shwartz and S. Ben-David, *Understanding Machine Learning: From Theory to Algorithms*, ser. Understanding Machine Learning: From Theory to Algorithms. Cambridge University Press, 2014. [Online]. Available: https://books.google.com/books?id=ttJkAwAAQBAJ
[13] G. D. Durgin, T. S. Rappaport, and D. A. de Wolf, "New analytical models and probability density functions for fading in wireless communications," *IEEE Transactions on Communications*, vol. 50, no. 6, pp. 1005–1015, June 2002.
[14] M. Nakagami, "The m-distributiona general formula of intensity distribution of rapid fading," in *Statistical Methods in Radio Wave Propagation*, W. Hoffman, Ed. Pergamon, 1960, pp. 3 – 36. [Online]. Available: http://www.sciencedirect.com/science/article/pii/B9780080093062500054
[15] M. Haenggi, "The meta distribution of the sir in poisson bipolar and cellular networks," *IEEE Transactions on Wireless Communications*, vol. 15, no. 4, pp. 2577–2589, April 2016.
[16] A. B. Tsybakov, *Introduction to Nonparametric Estimation*, 1st ed. Springer Publishing Company, Incorporated, 2008.
[17] M. Falk, J. Hüsler, and R.-D. Reiss, *Laws of Small Numbers: Extremes and Rare Events, 3rd. ed.* Birkhäuser, 2010.
[18] A. W. van der Vaart, *Asymptotic Statistics*. Cambridge University Press, 1998.
[19] E. Castillo, A. S. H. N. Balakrishnan, and J. M. Sarabia, *Extreme Value and Related Models with Applications in Engineering and Science*. Wiley, 2005.



Marko Angjelichinoski (S'15) obtained the Dipl. Ing. and M.Sc. degrees in Telecommunications from the Ss. Cyril and Methodius University in Skopje in 2011 and 2014, respectively. He obtained the Ph.D. degree from Aalborg University in 2017. In spring 2016 he was a visiting researcher at the Arizona State University, AZ, USA. Presently, he is a Postdoctoral Associate at Duke University. His research interests lie in the broad areas of statistical signal processing, machine learning and data mining and their applications in various fields such as communications, power systems and brain-computer interfaces.



Kasper Fløe Trillingsgaard received his B.Sc. degree in electrical engineering, his M.Sc. degree in wireless communications, and his Ph.D. degree in electrical engineering from Aalborg University, Denmark, in 2011, 2013, and 2017, respectively. From 2017 to 2018, he was a postdoctoral researcher at the same institution. He was a visiting researcher at New Jersey Institute of Technology, NJ, USA, in 2012, at Chalmers University of Technology, Sweden, in 2014, and at Massachusetts Institute of Technology, MA, USA, in November 2017. His research interests are in the areas of information and communication theory. In September 2018, he joined In-Commodities as a Data Scientist.



Petar Popovski (S'97–A'98–M'04–SM'10–F'16) is a Professor of Wireless Communications with Aalborg University. He received his Dipl. Ing and Magister Ing. degrees in communication engineering from the University of Sts. Cyril and Methodius in Skopje and the Ph.D. degree from Aalborg University in 2005. He has over 300 publications in journals, conference proceedings, and edited books and he is featured in the list of Highly Cited Researchers 2018, compiled by Web of Science. He holds over 30 patents and patent applications. He received an ERC Consolidator Grant (2015), the Danish Elite Researcher award (2016), IEEE Fred W. Ellersick prize (2016) and IEEE Stephen O. Rice prize (2018). He is currently a Steering Committee Member of IEEE SmartGridComm and IEEE Transactions on Green Communications and Networking. He previously served as a Steering Committee Member of the IEEE INTERNET OF THINGS JOURNAL. He is currently an Area Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS. Prof. Popovski is the General Chair for IEEE SmartGridComm 2018 and IEEE Communication Theory Workshop 2019. His research interests are in the area of wireless communication and communication theory.

0090-6778 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.