This article has been accepted for publication in IEEE Transactions on Wireless Communications. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TWC.2024.3396496

Extreme Value Theory-based Robust Minimum-Power Precoding for URLLC

Dian Echevarría Pérez, *Graduate Student Member, IEEE*, Onel L. Alcaraz López, *Member, IEEE*, Hirley Alves, *Member, IEEE*

***Abstract*—Channel state information (CSI) is crucial for achieving ultra-reliable low-latency communication (URLLC) in wireless networks. The main associated problems are the CSI acquisition time, which impacts the latency requirements of time-critical applications, and the estimation accuracy, which degrades the signal-to-interference-plus-noise ratio, thus, reducing communication reliability. In this work, we formulate and solve a minimum-power precoding design problem simultaneously serving multiple URLLC users in the downlink with imperfect CSI. Specifically, we develop an algorithm that exploits state-of-the-art precoding schemes such as maximal ratio transmission and zero-forcing, and adjust the power of the precoders to compensate for the channel estimation error uncertainty based on the extreme value theory framework. Finally, we evaluate the performance of our method and show its superiority with respect to a worst-case robust precoding benchmark.**

***Index Terms*—Extreme value theory, imperfect CSI, multi-antenna precoding, URLLC.**

I. Introduction

Ultra-reliable low-latency communication (URLLC), also known as critical machine-type communication, is an essential operation mode in current and future generations of wireless networks [1]. However, achieving reliability and low latency simultaneously, *e.g.*, for supporting applications like factory automation, vehicular communications, and telesurgery with latency-reliability pairs of (10 ms, $1-10^{-4}$), (1 ms, $1-10^{-5}$), and (1 ms, $1-10^{-9}$) [2], respectively, is extremely challenging. Indeed, efficiently supporting URLLC services requires an accurate statistical model characterization of the operational system, including channel conditions, interference statistics, user mobility, and the behavior of the communication protocols [3].

The use of multiple antennas at either one or both sides of a communication system, *e.g.,* single-input multiple-output, multiple-input single-output, and multiple-input multiple-output (MIMO), is a fundamental URLLC enabler [2]. Multiple antennas allow performing precoding/combining techniques to improve the signal-to-interference-plus-noise ratio (SINR) by boosting the received signal power, suppressing interference, or even both. This reduces the probability of error when decoding the signal, leading to higher reliability, and reduces latency since fewer packet re-transmissions are required.

The authors are with the Centre for Wireless Communications (CWC), University of Oulu, Finland. {dian.echevarriaperez, onel.alcarazlopez, hirley.alves}@oulu.fi

This research has been financially supported by Academy of Finland, 6G Flagship programme (Grant no. 346208), and the Finnish Foundation for Technology Promotion.

However, efficient precoding/combining methods are strictly tied to the availability of channel state information (CSI). Indeed, poor CSI estimations can cause a degradation in the quality of service (QoS) experienced by the user's equipment (UE) since the SINR may fall below a required threshold γ^{tar}. For URLLC, it is particularly important to keep the probability $\Pr\{\text{SINR} < \gamma^{tar}\}$ below a stringent permissible error target. Notably, when the reliability requirement is extremely tight, *i.e.,* $\Pr\{\text{SINR} < \gamma^{tar}\} \ll 1$, classical statistical methods derived from the central limit theorem (CLT) are not useful. This is because they rely on the behavior of averages over large sample sizes, which can mask the extreme values and tail behaviors, thus, fail to accurately capture the occurrence of rare error events [4]. Then, alternative approaches must be considered to overcome this issue, for instance, by exploiting the extreme value theory (EVT) framework. The latter focuses on characterizing/approximating distribution tails and is adept at predicting the probability of extreme events [4].

A. EVT for URLLC

EVT deals with the stochastic behavior of events that arise in the tail of probability distributions, thus, it is a handy tool for URLLC [3]. For instance, the authors in [5] presented a methodology to model extreme fade events on the channel. More specifically, they proposed techniques for fitting the tail distribution of the received power to the Generalized Pareto Distribution (GPD), determined the optimal threshold over which the statistics are derived, and calculated the optimal number of samples for fitting the model. The authors in [6] studied a power minimization problem with second-order statistical constraints on latency and reliability. They proposed semi-centralized and distributed queue-aware power allocation techniques for vehicle-to-vehicle communications taking advantage of EVT and Lyapunov stochastic optimization. In [7], the authors aimed to minimize the power consumption of vehicular UEs subject to reliability requirements in terms of probabilistic queuing delays. They presented an approach that combines federated learning and EVT to learn the network statistics and exploit them to reduce the occurrence of extreme events, *i.e.,* queue lengths growing beyond a predefined threshold. EVT and federated learning were also exploited in [8] to model extreme events in vehicular networks, *i.e.,* very large queues, and to learn the distribution of such events, respectively. The authors proposed a resource allocation strategy that enhances the worst-case system reliability by minimizing the maximum queue length. In [9], the authors

exploited EVT, federated learning, and large deviation theory to find the approximate distribution of the delay and the corresponding tail distribution in a scenario with a base station (BS) serving a large number of UEs in uplink (UL) and downlink (DL) fading channels. In [10], a vehicle's power minimization problem with strict URLLC demands regarding the probabilistic age of information was formulated. Also, a novel approach to map the distributions of queue length and age of information was presented, while EVT and the Lyapunov optimization framework were exploited to solve the problem. Meanwhile, the work in [11] presented an EVT-based rate selection approach for URLLC. Therein, the tail of the distribution of the received powers was fitted to the GPD, and the authors determined the maximum transmission rate by including the GPD in a proposed rate selection function.

B. Related works and Motivation

In recent years, several works have focused on the solution to minimum-power precoding design problems where the UEs have strict QoS requirements, *e.g.,* SINR, outage probability, data rate [12]–[19]. For instance, the work in [13] proposed a worst-case MIMO precoding design to guarantee an SINR performance of the UE for every channel realization and channel estimation error, thus, robust against channel/estimation uncertainties. Therein, the authors assumed the channel estimate as the center of an ellipsoid in a multidimensional complex space where the radius is determined by the norm of the CSI error vector, and any channel realization lies inside the ellipsoid. A similar problem was presented in [14] for the multi-user case. The original non-convex problem was re-formulated into a semi-definite programming (SDP) form via the S-procedure method and rank relaxations. The authors also presented an algorithm that extends the robust solution for the multi-user case with both perfect and imperfect CSI at the receiver side while guaranteeing that all the SINRs are above the required target. The authors in [15] addressed the minimum-power precoding design problem (hereinafter termed as transmit power minimization problem) with UE's outage constraints. They showed that the probabilistic approach can be converted into a deterministic one with SINR constraints and the same structure as the SDP problem in [14]. Moreover, the proposed approach allows controlling the radius of the ellipsoid according to the outage demands instead of fixing it to a pre-established value. The work in [16] also rearranged the outage constraints into SINR's for the frequency division duplex case. After applying the S-procedure and rank relaxation, the non-convex form was reformulated into a linear objective with linear-matrix inequalities (LMIs) constraints. The work in [17] also considered the outage constraints, but with beamforming directions being fixed beforehand. Therein, the authors took advantage of existing precoding methods such as maximal ratio transmission (MRT) or zero-forcing (ZF) and determined the power allocation for each UE. Again, relaxation of the constraints was needed to convert the original problem into an equivalent convex form. The authors in [18] also considered the transmit power minimization problem with outage constraints by establishing fixed beamforming directions. However, their proposal leads to many outage violations for moderate SINR targets, while the performance was evaluated for outage probabilities higher than 10^{-2}, which is still far from the most stringent requirements of URLLC. Finally, the work in [19] solved the transmit power minimization problem with per-user rate constraints in the finite block length regime. The minimum rates were set to meet specific block error rates in DL transmissions. The original problem was transformed into an SDP problem requiring rank relaxations.

Notice that for the aforementioned minimum-power precoding designs, the authors resorted to approximations or relaxation of constraints that do not fully guarantee to find optimal solutions to the original problem. Some of the approximations are conservative, meaning that the feasible set of precoders of the transformed problem may be smaller than the feasible set of the original formulation. In some cases, the procedures involve LMIs, *e.g.,* [13], [15], [16], [18], [19], which require high computational and processing costs. Also, some works, *e.g.,* [15]–[18], evaluate targets that are still far from those required in URLLC applications with strict QoS demands or evaluate the performance with parameters that may not be practical for real applications. The accuracy of the presented approaches for capturing critical events that arise far in the tail of the distributions may be questionable, being EVT a useful tool to overcome this issue.

C. Contributions

Our work focuses on a minimum-power precoding design to support URLLC in scenarios with imperfect CSI. Our contributions are four-fold:

- We formulate a precoding design optimization problem for transmit power minimization while ensuring URLLC demands at the UEs. We exploit EVT to impose the reliability requirements of the UEs based on the channel estimation and its related uncertainty. Specifically, we fit the data obtained from artificially-generated SINR values to the GPD to model the ultra-reliability region.
- We propose an algorithm that leverages state-of-the-art precoding methods to solve the problem. This promotes low complexity and thus low computational costs.
- We evaluate the performance of the proposed method using ZF and MRT precoding schemes. We use a worst-case robust precoding scheme as a benchmark to compare the results, showing the superiority of our proposed method.
- We analyze the impact on the performance of the number of estimation error samples, the confidence when fitting the obtained data to the GPD, the number of URLLC UEs, and the pilot length. We show that there is an optimal pilot length that minimizes the total transmit power, which also increases proportionally with the fitting confidence. Moreover, we show that the fitting confidence must be set larger as the reliability target gets stricter.

The work is structured as follows. Section II describes the system model and main assumptions, after which the optimization problem is formulated. In Section III, we present the

TABLE I: Main symbols used throughout the paper

Symbol	Definition
M	number of transmit antennas at the BS
K	total number of UEs
N	number of estimation error vectors available at the BS
$\mathbf{h}_k$	channel vector between the BS and UE k
$\hat{\mathbf{h}}_k$	estimate of $\mathbf{h}_k$
$\mathbf{e}_k$	estimation error of $\hat{\mathbf{h}}_k$
$\mathbf{s}_k^p$	pilot sequence transmitted by UE k
$\mathbf{s}_k^d$	data sequence transmitted to UE k
p_{ul}	uplink transmit power
τ_f	length of the data frame
τ_e	length of the pilot sequence
τ_{dl}	number of symbols for DL transmission
$\mathbf{w}_k$	precoder intended to UE k
$\mathbf{u}_k$	normalized precoder intended to UE k
γ_k	SINR at UE k
γ_k^{tar}	SINR target at UE k
γ_k°	SINR sample for UE k
σ_v^2	noise power
p_k	power allocated to UE k
p_{min}	minimum transmit power at the BS
p_{max}	maximum transmit power at the BS
NF	noise figure
BW	bandwidth
Γ	fitting confidence of the GPD
ρ	quantile value of the samples
ζ_k	outage probability target at UE k
β_k	average channel gain in the link between the BS and UE k
$\mathcal{O}_k$	outage probability of UE k
κ_k	Rician factor of the channel between the BS and UE k

EVT-based beamforming design and the proposed algorithm, and discuss MRT and ZF-based implementations. Section IV presents a benchmark approach to compare with our scheme. In Section V, we illustrate numerical results and validate the proposed algorithm. Finally, Section VI concludes the paper.

Notation Uppercase and lowercase boldface letters denote matrices and vectors, respectively. Superscript $(\cdot)^H$ depicts the Hermitian operator, $(\cdot)^T$ denotes the transpose operator, $(\cdot)^{-1}$ represents the matrix inverse operation, and $||\cdot||$ depicts the norm of a vector. Moreover, $\mathcal{CN}(\mathbf{v},\mathbf{R})$ denotes a complex Gaussian distribution with mean vector $\mathbf{v}$ and covariance matrix $\mathbf{R}$, and $\mathcal{U}(a,b)$ depicts a uniform distribution in the range $[a,b]$. $F_Q(\cdot)$ denotes the cumulative density function (CDF) of the random variable Q and $\mathbb{i} = \sqrt{-1}$ denotes the imaginary operator. Finally, $\mathcal{Q}(c,D)$ represents the $c\%$-quantile operator of the sample set D and $\mathcal{I}(\cdot)$ denotes the indicator operator. Table I summarizes the main symbols used throughout the paper.

II. System model

We consider a scenario where a BS equipped with M antennas serves $K \leq M$ single-antenna low-mobility URLLC UEs in the DL channel (see Fig. 1). The channels between the BS and the UEs remain constant within a time-frequency coherence block and change independently from block to block. Moreover, channel reciprocity is exploited for channel estimation. Before DL transmissions, the BS estimates the channel coefficients from the K pilot signals of length τ_e transmitted in the UL channel by the UEs. We assume that $\tau_e \geq K$ guarantees the generation of orthogonal sequences, thus avoiding pilot contamination. Let us denote τ_{dl} as the number of symbols dedicated for DL transmission, therefore

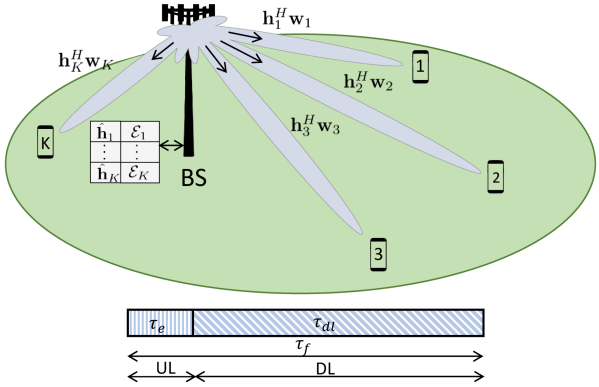

Fig. 1: System model and frame structure. The BS at the top of the figure serves a set of K single-antenna URLLC UEs in the DL. $\mathbf{h}_k$ and $\mathbf{w}_k$ represent the channel vector from the BS's antennas to UE k and the precoder vector intended for UE k, respectively. The beams from the BS to UE k are formed by precoding the signal through the communication channel, *i.e.,* $\mathbf{h}_k^H\mathbf{w}_k$. Prior to DL transmissions, pilot sequences are transmitted from the UEs to the BS for channel estimation. Thus, the BS stores the channel estimates $\hat{\mathbf{h}}_1 ... \hat{\mathbf{h}}_K$. From previous data, the BS also stores K sets of CSI estimation errors $\mathcal{E}_1 ... \mathcal{E}_K$. Moreover, the frame structure is displayed at the bottom of the figure with τ_f, τ_e, and τ_{dl} representing the length of the data frame, length of the pilot sequence, and the number of symbols for DL transmission, respectively.

$\tau_f = \tau_e + \tau_{dl}$ constitutes the frame duration, while $\mathbf{h}_k$, $\hat{\mathbf{h}}_k \in \mathbb{C}^{M\times 1}$ are respectively the true and estimated channel coefficients between the $k-$th UE and the BS's antennas. In practice, estimation errors arise due to noise and uncontrolled interfering signals and cannot be completely removed due to a finite τ_e. We also assume that the BS has previously stored N error samples for pilot sequences of length τ_e as detailed in Section II-B. Moreover, these samples are utilized together with the estimated channels $\hat{\mathbf{h}}_k$ for precoding design.

A. Signal model

In the UL, the UE k transmits a pilot sequence $\mathbf{s}_k^p \in \mathbb{C}^{\tau_e\times 1}$ with $||\mathbf{s}_k^p||^2 = \tau_e$ such that the signal $\mathbf{Y} \in \mathbb{C}^{M\times\tau_e}$ received at the BS is given by

$$\mathbf{Y} = \sum_{k=1}^{K} \sqrt{p_{ul}}\mathbf{h}_k(\mathbf{s}_k^p)^H + \mathbf{V}, \tag{1}$$

where p_{ul} is the average transmit power of the UEs and $\mathbf{V} \in \mathbb{C}^{M\times\tau_e}$ includes the influence not only of the additive white Gaussian noise (AWGN) but also potential interfering signals at the receiver[1]. We assume $\mathbb{E}\{\mathbf{V}\} = \mathbf{0}$, which holds in most practical setups as radio signals and AWGN have no direct current level. Herein, we adopt the least square (LS) channel estimate

$$\hat{\mathbf{h}}_k = \frac{1}{\sqrt{p_{ul}\tau_e}}\mathbf{Y}\mathbf{s}_k^p, \tag{2}$$

which exploits the fact that the pilot sequences corresponding to different UEs are orthogonal. Moreover,

$$\mathbf{h}_k = \hat{\mathbf{h}}_k + \mathbf{e}_k, \tag{3}$$

[1]Interference may arise from non-orthogonal (pilot) signals transmitted by UEs in neighboring cells.

with $\mathbb{E}\{\mathbf{e}_k\} = \mathbf{0}$ and $\sigma^2_{\hat{h}_k} = \sigma^2_{h_k} - \sigma^2_{e_k}$, where $\mathbf{e}_k$ is the CSI estimation error vector and $\sigma^2_{\hat{h}_k}$, $\sigma^2_{h_k}$, and $\sigma^2_{e_k}$ represent the variances of $\hat{\mathbf{h}}_k$, $\mathbf{h}_k$, and $\mathbf{e}_k$, respectively. Notice that $\sigma^2_{e_k}$ is inversely proportional to the UL SINR and the number of pilot symbols [20].

In the DL, the BS transmits the complex data signal $\mathbf{s}^d_k$ to UE k such that $\mathbb{E}\{||\mathbf{s}^d_k||^2\} = \tau_{dl}$ and $\mathbb{E}\{\mathbf{s}^d_k(\mathbf{s}^d_i)^H\} = \mathbf{0}\ \forall k \neq i$. Then, the signal $\mathbf{y}_k \in \mathbb{C}^{\tau_{dl}}$ received at UE k is given by

$$\mathbf{y}_k = \mathbf{h}_k^H \mathbf{w}_k \mathbf{s}^d_k + \sum_{i \neq k} \mathbf{h}_k^H \mathbf{w}_i \mathbf{s}^d_i + \mathbf{v}_k, \tag{4}$$

where $\mathbf{w}_k \in \mathbb{C}^{M \times 1}$ depicts the precoding vector intended to UE k. Moreover, similar to the UL signal, $\mathbf{v}_k \in \mathbb{C}^{\tau_{dl} \times 1}$, with $\mathbb{E}\{\mathbf{v}_k\} = \mathbf{0}$ and $\mathbb{E}\{||\mathbf{v}_k||^2\}/\tau_{dl} = \sigma_v^2$, comprises the contribution of interference signals and AWGN at the UE k. Finally, the SINR at UE k is given by

$$\gamma_k(\{\mathbf{w}_k\}, \{\mathbf{h}_k\}) = \frac{|\mathbf{h}_k^H \mathbf{w}_k|^2}{\sum_{i \neq k} |\mathbf{h}_k^H \mathbf{w}_i|^2 + \sigma_v^2}. \tag{5}$$

B. Acquisition and impact of the estimation error

As mentioned earlier, the BS stores N estimation error vectors from previous data, denoted as $\mathcal{E}_k = \{\mathbf{e}_{k,1}\ \mathbf{e}_{k,2}\ \ldots\ \mathbf{e}_{k,N}\}\ \forall k$. These samples are assumed to be independent and identically distributed (i.i.d.), which can be justified by spacing the sampling at time intervals larger than the coherence time. The error samples can be obtained in practice by performing multiple channel estimations within a coherence block. For instance, two pilot sequences of lengths τ_{e1} and τ_{e2} with $\tau_{e2} \gg \tau_{e1}$ may be used to obtain an error vector from the corresponding estimates $\hat{\mathbf{h}}_{k1}$ and $\hat{\mathbf{h}}_{k2}$ as $\mathbf{e}_k \approx \hat{\mathbf{h}}_{k2} - \hat{\mathbf{h}}_{k1}$. This exploits the fact that $\hat{\mathbf{h}}_{k2} \to \mathbf{h}_k$ as $\tau_{e2} \to \infty$ for efficient linear estimation methods. Although τ_e is finite in practice, a good estimate of the error is attainable if τ_e is set large. The acquisition of estimation errors must be performed with low periodicity, mainly when the UEs are inactive to reduce overhead and computational costs.

Note that the random fading effect cannot be completely removed due to imperfect CSI estimation. Thus, there is still some remaining randomness in the signal associated with using $\hat{\mathbf{h}}_k$. This may prevent the QoS demands from being met as the SINRs may fall below the target γ_k^{tar}. The impact of the estimation error becomes more severe when the UE transmit power is limited, the average channel gain is low, and/or there is pilot contamination. For example, consider that the BS in the system model serves one UE in the DL, and the minimum SINR to decode the signal with arbitrarily low error probability is $\gamma_1^{tar} = 10$ dB. Fig. 2 shows the empirical probability density function of the SINR achieved for a given channel realization $\mathbf{h}_k$ over 10^6 channel estimations $\hat{\mathbf{h}}_k$ with $\tau_e = \{1, 2\}$ and LS error estimation. As expected, the variance of the SINR decreases as τ_e increases. Notably, the probability of falling below the target γ_k^{tar} may be high if the precoding, and especially its power allocation (p_1), does not consider the estimation error $\big(3.29 \times 10^{-1}$ and 6.57×10^{-2} for $\tau_e = 1$ and $\tau_e = 2$, respectively, as shown in Fig. 2 (leftmost histograms)$\big)$. Also note that if the transmit power is increased by 3 dB

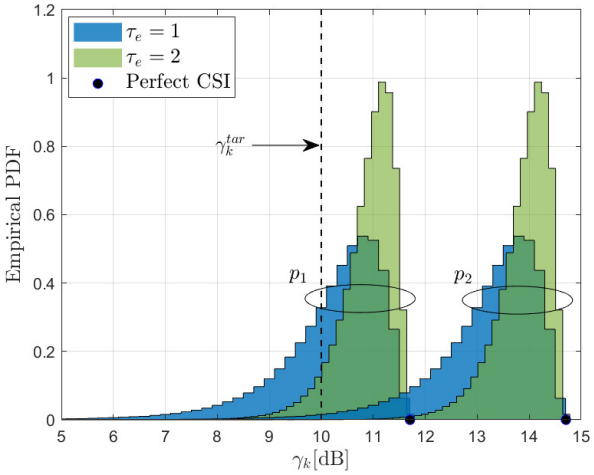

Fig. 2: Empirical distribution of γ_k [dB] for the channel realization $\mathbf{h}_k = \sqrt{10^{-13}}[0.118 + 0.501\mathbf{i},\ 0.145 + 0.058\mathbf{i},\ -0.051 + 0.022\mathbf{i},\ 0.087 - 0.176\mathbf{i}]^T$ in Rayleigh fading with $\tau_e = \{1, 2\}$ and DL transmit power $p_1 = 23$ dBm and $p_2 = 26$ dBm. MRT precoding is used with $M = 4$, UL transmit power of 20 dBm, and LS channel estimation.

(p_2), the SINR realizations are considerably moved to the right and the probability of not meeting the target γ_k^{tar} is highly reduced, *e.g.,* 1.83×10^{-2} and 2.87×10^{-4} for $\tau_e = 1$ and $\tau_e = 2$, respectively, as depicted in Fig. 2 (rightmost histograms). In general, reducing γ_k^{tar} and/or increasing the transmit power mitigates the impact of the estimation error on the performance. The reduction of γ_k^{tar} decreases the spectral efficiency, leading to a higher transmission latency over the same bandwidth. Therefore, increasing the transmit power seems more appealing if the power budget allows it. However, an arbitrarily high power allocation is not optimal from the energy efficiency point of view and might not guarantee the QoS requirements in the multi-UE case.

C. Problem formulation

As mentioned earlier, we focus on the precoding design to minimize the transmit power at the BS while ensuring URLLC constraints at each UE k. Specifically, we aim to solve the following optimization problem

$$\mathbf{P1}: \quad \underset{\{\mathbf{w}_k\}_{\forall k}}{\text{minimize}} \quad \sum_{k=1}^{K} ||\mathbf{w}_k||_2^2 \tag{6a}$$

$$\text{subject to} \quad \mathcal{O}_k \leq \zeta_k \quad \forall k, \tag{6b}$$

with

$$\mathcal{O}_k = \Pr\big\{\gamma_k\big(\{\mathbf{w}_k\}, \{\mathbf{h}_k\}\big) < \gamma_k^{tar}\big\}, \tag{7}$$

where γ_k^{tar} depicts the required SINR to achieve a successful transmission and ζ_k represents the target outage probability at UE k. Without loss of generality, we assume $\gamma_k^{tar} = 2^{r_k} - 1$, where $r_k = B/\tau_{dl}$ and B denotes the number of bits to be transmitted over $\tau_{dl} = \tau_f - \tau_e$ symbols. Notice that the constraint (6b) ensures that the outage probability of UE k is maintained below the target ζ_k. Interestingly, the objective function in (6a) is convex, but we cannot state the convexity of (6b) since the distribution of $\mathbf{h}_k$, and thus the distribution

of the SINR, is unknown. Even if the channel distribution is available, the accuracy of the obtained model for capturing events that arise in the tail of the distribution would be low. Therefore, we resort to EVT to reformulate constraint (6b) while proposing a framework that captures rare events and avoids using geometric-based approaches, *e.g., ellipse-based*, to model the uncertainties of the channel estimations $\hat{\mathbf{h}}_k$ around the actual channel $\mathbf{h}_k$.

III. EVT-based Optimization

A. EVT preliminaries

The main result we exploit from EVT is the following.

Theorem 1 (Theorem for Exceedances Over Thresholds [4])**.** *For an arbitrary random variable X from a non-degenerative distribution and for a large enough μ, the cumulative distribution function (CDF) of $Z = X - \mu$ conditioned on $X > \mu$ is given by*

$$F_Z(z) = 1 - \left[1 + \frac{\xi z}{\upsilon}\right]^{-\frac{1}{\xi}}, \tag{8}$$

defined on $\{z : z > 0$ and $1 + \xi z/\upsilon > 0\}$. The distribution in (8) is known as the GPD with shape and scale parameters ξ and υ, respectively.

The parameters of the GPD can be estimated from multiple observations of the random variable X. Log-likelihood and numerical methods relying on distribution fitting are commonly adopted for estimating ξ and υ. To obtain the threshold parameter μ, the *mean residual life method* may be used, which outputs the highest threshold below which the *mean residual life function* behaves linearly [4]. If the mean residual life function is almost linear independently of the threshold selection, a complementary approach termed as *parameter stability method* may be used. In this case, the optimal threshold is the highest one below which the estimated ξ and υ behave linearly concerning μ. Another common method is the so-called *fixed threshold approach* where the threshold is usually set before fitting. In this sense, simple quantile rules have been proposed, *e.g.*, the upper 10% rule of DuMouchel, which simply uses up to the upper 10% of the data to fit the GPD, *i.e.,* $\mu = \mathcal{Q}(u, X)$ with $u \geq 90\%$ [21], [22].

Notice that the selection of μ is a critical step in the accuracy of the GPD model. On the one hand, small values of μ may result in a large number of samples z (large bias), capturing not only events on the tail of the distribution but also values potentially close to the mean, thus affecting the fitting accuracy. On the other hand, large values of μ may result in a reduced data set (large variance), which would cause an inaccurate parameter estimation. Note that all the previously discussed methods for $\mu-$ estimation require subjective threshold selection to some extent. Considering this and from the point of view of computational cost and simplicity in the implementation, the *fixed threshold approach* is preferable. Thus, we adopt it from here onwards.

Fig. 3 shows the impact of the selection of μ on the accuracy of the fitting to the GPD fitting. Fig. 3 (a)−(c) show empirical distributions obtained from 10^6 samples drawn from a normal

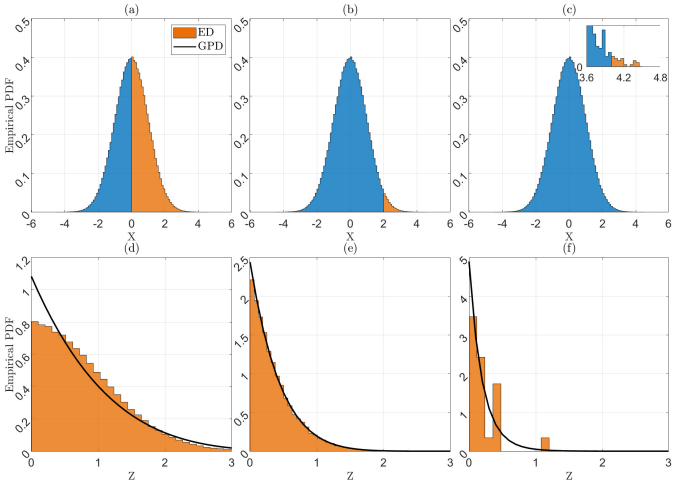

Fig. 3: Accuracy of the GPD fitting depending on the selection of μ. The data is drawn from a normal distribution, and the thresholds are $\mu = 0, \mu = 2$, and $\mu = 4$ in pairs (a)−(d), (b)−(e), and (c)−(f), respectively. Empirical distributions of the data are shown in (a)−(c) while histograms of the excesses are displayed in (d)−(f).

distribution. The portions of the distribution to the right of the vertical lines (orange color) depict the excess data, *i.e.*, the data exceeding μ, over different threshold values. Fig. 3 (d)−(f) represents the histograms of the excess data and the GPD fitting for the thresholds in the corresponding top plots. Notice that for $\mu = 0$ and $\mu = 4$, the data fitting to the GPD is not accurate. In the former, the excess data is large and captures events on the tail and around the mean. At the same time, in the latter, the excess data contains only a few samples, leading to an inaccurate fitting/parameter estimation. An accurate fitting is obtained in Fig. 3 (e) for $\mu = 2$ since the samples are located on the tail and their number is considerably larger than in Fig. 3 (f).

B. Problem reformulation

In this subsection, we exploit Theorem 1 from Section III-A and the sets $\mathcal{E}_k\ \forall k$ to rearrange the constraint (6b) and make it tractable. Let us consider an instantaneous channel estimation $\hat{\mathbf{h}}_k$ for UE k and the corresponding channel estimation error history $\mathcal{E}_k$. Because of the zero-mean properties of $\mathbf{e}_k$, the distributions of $\mathbf{h}_k$ and $\hat{\mathbf{h}}_k$ share the same mean but differ in their variances. Thus, real and imaginary components of $\mathbf{h}_k$ lie around the mean of real and imaginary components of $\hat{\mathbf{h}}_k$. Therefore, by adding up each entry of the error set $\mathcal{E}_k$ to $\hat{\mathbf{h}}_k$, we obtain the new set

$$\mathcal{H}_k = \mathcal{E}_k + \hat{\mathbf{h}}_k. \tag{9}$$

Notice that this is possible due to the independence between the error $\mathbf{e}_k$ and channel estimate $\hat{\mathbf{h}}_k$ enabled by the LS estimation method. Moreover, $\mathcal{H}_k = \{\tilde{\mathbf{h}}_{k,1}, \tilde{\mathbf{h}}_{k,2}, ..., \tilde{\mathbf{h}}_{k,N}\}$, with $\tilde{\mathbf{h}}_{k,n} = \mathbf{e}_{k,n} + \hat{\mathbf{h}}_k$, constitutes a group of possible channel realizations for the link between UE k and the BS that may have led to a channel estimation $\hat{\mathbf{h}}_k$. This implies that the larger the value of N, the smaller (probabilistically) the difference between at least one element $\tilde{\mathbf{h}}_{k,n}$ in $\mathcal{H}_k$ and

the actual $\mathbf{h}_k$. Additionally, the set $\mathcal{H}_k$ also contains entries that are farther from $\mathbf{h}_k$, which are beneficial for mimicking the poorest estimation cases (left tails in Fig. 2).

For the $n-$th entry $\tilde{\mathbf{h}}_{k,n}$ of the set $\mathcal{H}_k$, we can generate a sample of the SINR of UE k using (5) as follows

$$\gamma^{\circ}_{k,n}(\{\mathbf{w}_k\},\{\tilde{\mathbf{h}}_{k,n}\})=\frac{|\tilde{\mathbf{h}}^H_{k,n}\mathbf{w}_k|^2}{\sum_{i\neq k}|\tilde{\mathbf{h}}^H_{k,n}\mathbf{w}_i|^2+\sigma_v^2}. \quad (10)$$

To meet the reliability demands, we must ensure that most samples satisfy $\gamma^{\circ}_{k,n}(\{\mathbf{w}_k\},\{\tilde{\mathbf{h}}_{k,n}\}) > \gamma_k^{tar}$ for a channel estimation $\hat{\mathbf{h}}_k$. However, in this case, the data of interest is located on the left tail of the distribution (samples that do not meet the SINR requirements). Still, to apply EVT, specifically the Theorem 1, we must have the data on the right tail. Thus, we may proceed with a simple transformation of (6b) as follows

$$\begin{aligned}&\Pr\{\gamma^{\circ}_{k,n}(\{\mathbf{w}_k\},\{\tilde{\mathbf{h}}_{k,n}\}) < \gamma_k^{tar}\}\\ &\qquad=\Pr\left\{\frac{1}{\gamma^{\circ}_{k,n}(\{\mathbf{w}_k\},\{\tilde{\mathbf{h}}_{k,n}\})} > \frac{1}{\gamma_k^{tar}}\right\}. \quad (11)\end{aligned}$$

Some samples of the random variable $1/\gamma^{\circ}_{k,n}(\{\mathbf{w}_k\},\{\tilde{\mathbf{h}}_{k,n}\})$ might be significantly dispersed from the rest which may affect the fitting to the excess data that we will perform in the next steps. To mitigate this issue, we introduce a concave transformation $f(\cdot)$ to all samples as

$$\Pr\left\{f\left(\frac{1}{\gamma^{\circ}_{k,n}(\{\mathbf{w}_k\},\{\tilde{\mathbf{h}}_{k,n}\})}\right) > f\left(\frac{1}{\gamma_k^{tar}}\right)\right\}. \quad (12)$$

Let us now define

$$\psi_k(\{\mathbf{w}_k\},\{\tilde{\mathbf{h}}_{k,n}\}) \triangleq f\big(1/\gamma^{\circ}_{k,n}(\{\mathbf{w}_k\},\{\tilde{\mathbf{h}}_{k,n}\})\big), \quad (13)$$

$$\phi_k \triangleq f\big(1/\gamma_k^{tar}\big) \quad (14)$$

for ease of notation. We can now set a threshold μ_k and apply the definition of conditional probability as

$$\begin{aligned}&\Pr\Big\{\psi_k(\{\mathbf{w}_k\},\{\tilde{\mathbf{h}}_{k,n}\}) > \phi_k\Big\}\\ &\quad=\Pr\Big\{\psi_k(\{\mathbf{w}_k\},\{\tilde{\mathbf{h}}_{k,n}\}) > \mu_k\Big\}\Pr\Big\{\psi_k(\{\mathbf{w}_k\},\{\tilde{\mathbf{h}}_{k,n}\}) - \mu_k\\ &\qquad> \phi_k-\mu_k\big|\psi_k(\{\mathbf{w}_k\},\{\tilde{\mathbf{h}}_{k,n}\}) > \mu_k\Big\}. \quad (15)\end{aligned}$$

Moreover, according to DuMouchel's rule, we can set the threshold $\mu_k = \mathcal{Q}(\rho\times 100, \psi_k)$, thus as a function of $\{\mathbf{w}_k\}$ and $\{\tilde{\mathbf{h}}_{k,n}\}$ such that

$$\begin{aligned}&\Pr\Big\{\psi_k(\{\mathbf{w}_k\},\{\tilde{\mathbf{h}}_{k,n}\}) > \mu_k(\{\mathbf{w}_k\},\{\tilde{\mathbf{h}}_{k,n}\})\Big\}\\ &\qquad\approx\frac{1}{N}\sum_{n=1}^{N}\mathcal{I}\Big[\psi_k(\{\mathbf{w}_k\},\{\tilde{\mathbf{h}}_{k,n}\}) > \mu_k(\{\mathbf{w}_k\},\{\tilde{\mathbf{h}}_{k,n}\})\Big] \quad (16)\end{aligned}$$

is equal to $1-\rho$. Thus, we have that

$$\begin{aligned}&\Pr\Big\{\gamma_k(\{\mathbf{w}_k\}) < \gamma_k^{tar}\Big\}\\ &\qquad=(1-\rho)\Big(1-F_{Q_k}\big(\phi_k-\mu_k(\{\mathbf{w}_k\},\{\tilde{\mathbf{h}}_{k,n}\})\big)\Big), \quad (17)\end{aligned}$$

where $\phi_k > \mu_k(\{\mathbf{w}_k\},\{\tilde{\mathbf{h}}_{k,n}\})$ and

$$\begin{aligned}Q_k \triangleq \Big(&\psi_k(\{\mathbf{w}_k\},\{\tilde{\mathbf{h}}_{k,n}\}) - \mu_k(\{\mathbf{w}_k\},\{\tilde{\mathbf{h}}_{k,n}\})\\ &\Big|\psi_k(\{\mathbf{w}_k\},\{\tilde{\mathbf{h}}_{k,n}\}) > \mu_k(\{\mathbf{w}_k\},\{\tilde{\mathbf{h}}_{k,n}\})\Big). \quad (18)\end{aligned}$$

Next, we proceed to fit all data samples Q_k to the GPD in (8) to obtain the estimates $\hat{\upsilon}_k(\{\mathbf{w}_k\},\{\tilde{\mathbf{h}}_{k,n}\}) = \hat{\upsilon}_k$ and $\hat{\xi}_k(\{\mathbf{w}_k\},\{\tilde{\mathbf{h}}_{k,n}\}) = \hat{\xi}_k$ of the parameters υ and ξ with $z = \phi_k - \mu_k\big(\{\mathbf{w}_k\},\{\tilde{\mathbf{h}}_{k,n}\}\big)$. With the estimates, (7) can be re-written as

$$\mathcal{O}_k = (1-\rho)\Big(1+\frac{\hat{\xi}_k}{\hat{\upsilon}_k}\big(\phi_k-\mu_k(\{\mathbf{w}_k\},\{\tilde{\mathbf{h}}_{k,n}\})\big)\Big)^{-1/\hat{\xi}_k}. \quad (19)$$

Notice that after the transformation of $\mathcal{O}_k$, **P1** remains as a non-convex problem which is yet challenging to solve. We address this issue in the next subsection.

C. Proposed algorithm

Note that common non-convex optimization solvers such as those based on genetic and particle swarm algorithms might not often provide feasible solutions to **P1** because of the high non-linearity of the constraints, the difficulty to properly configure the optimization hyperparameters, and the large amount of required computational resources. Similar to prior work, *e.g.,* [12], [17], [18], that fixed the precoding directions for reducing complexity, we propose an algorithm that exploits state-of-the-art linear precoding schemes, *e.g.,* ZF and MRT, for transmit power minimization.

First, we depart from the channel estimations $\hat{\mathbf{h}}_k$ to compute the precoders as $\mathbf{w}_k = \sqrt{p_k}\mathbf{u}_k$ with p_k as the power allocated to UE k which is initially set to a minimum value p_{min} to all UEs. The normalized precoding directions are given by

$$\mathbf{u}_k = \frac{\mathbf{z}_k^*}{||\mathbf{z}_k||}, \quad (20)$$

where $\mathbf{z}_k = \hat{\mathbf{h}}_k$ for MRT precoding, and $[\mathbf{z}_1,\mathbf{z}_2,...,\mathbf{z}_k] = \hat{\mathbf{H}}(\hat{\mathbf{H}}^H\hat{\mathbf{H}})^{-1}$ for ZF precoding with $\hat{\mathbf{H}} = [\hat{\mathbf{h}}_1,\hat{\mathbf{h}}_2,...,\hat{\mathbf{h}}_k]$. Then, compute the sets $\mathcal{H}_k$ according (9), and for every UE k, compute (10), (13) and (14). Next, determine the value of μ_k as the $\rho-$quantile (%) of ψ_k such that (16) is equal to $1-\rho$. Then, compute the excesses Q_k in (18) and estimate the parameters of the GPD with confidence Γ via log-likelihood estimation[2] to obtain bounds as $[\hat{\upsilon}_{k,LB},\hat{\xi}_{k,LB}]$ and $[\hat{\upsilon}_{k,UB},\hat{\xi}_{k,UB}]$. With the estimates, proceed to obtain outage probability bounds $\mathcal{O}_{k,LB}$ and $\mathcal{O}_{k,UB}$ by evaluating the parameters in (19). Then, if the bound $\mathcal{O}_{k,UB}$ is above the target ζ_k, the power p_k is increased in a small value Δp, and the process starts again from the computation in (10). Nevertheless, if the outage bound is below the target ζ_k, the real outage probability will also be below ζ_k if the parameters ρ and Γ are properly configured, and the number of samples N is sufficiently large. In such a case, a similar analysis must be done with the remaining UEs until the outage bounds for

[2]Log-likelihood methods are suitable for i.i.d. data samples. If non-i.i.d. sequences are available, the method of moments or probability-weighted moments could be used for parameter estimation. Also, the sequence may be transformed to i.i.d. using declustering methods [4].

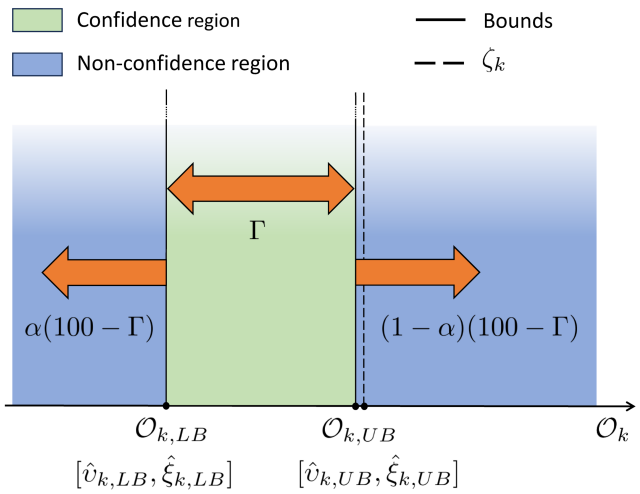

Fig. 4: Bound region for the EVT approach. The figure shows the confidence area in green color.

all UEs are below their respective targets ζ_k simultaneously or until the total power constraint p_{max} is violated, and there is no feasible solution. Notice that the selection of Δp significantly impacts the performance of the proposed algorithm. While large values may cause the algorithms not to find solutions to the problem, small values will make the processing time larger. Nevertheless, it is recommended to select a small value that ensures finding the solutions. Finally, the precoders $\mathbf{w}_k = \sqrt{p_k}\mathbf{u}_k \ \forall k$ constitute the solution to **P1**.

Algorithm 1 summarizes the steps discussed earlier. It also comprises the initialization of the transmit powers p_k to a minimum power p_{min} and the upper bounds in the outage probability $\mathcal{O}_{k,UB} = 1$ in lines 1 and 4, respectively. The value of p_{min} is recommended to be small, and p_{max} can be chosen according to hardware constraints, *e.g.,* $46-47$ dBm, as in typical BSs [23].

The computational complexity of the algorithm is mainly dominated by the total number of data samples N and the amount of excess data when fitting to the GPD, *i.e.*, $\mathcal{O}\big((1-\rho)N^2\big)$. According to [24], if we consider log-likelihood methods for parameter estimation, the variances of the estimates $\hat{\xi}_k$ and $\hat{\upsilon}_k$ are respectively given by $\frac{1}{(1-\rho)NJ(\xi_k)}$ and $\frac{1}{(1-\rho)NJ(\upsilon_k)}$. Herein, $J(\xi_k)$ and $J(\upsilon_k)$ represent the Fisher information[3], which grows with the sample size. Notice that the estimates' variance goes to zero as the number of samples grows. Thus, $\hat{\xi}_k \to \xi_k$, $\hat{\upsilon}_k \to \upsilon_k$ and the estimated distribution converges to the true GPD distribution, which represents the main strength of **Theorem 1**. This ensures enough tail information and, therefore, an accurate resource allocation. On the other hand, when $N \to 0$, the variance of the estimates grows unbounded, leading to poor parameter estimation and failure of the resource allocation method.

Notice that the use of confidence bounds of the parameters, *i.e.,* $[\hat{\upsilon}_{k,LB}, \hat{\xi}_{k,LB}]$ and $[\hat{\upsilon}_{k,UB}, \hat{\xi}_{k,UB}]$, and their associated outage probability bounds $\mathcal{O}_{k,LB}$ and $\mathcal{O}_{k,UB}$, instead of the actual estimates $[\hat{\upsilon}_k, \hat{\xi}_k]$ as in most EVT-related works, allows to cope with the uncertainties associated with the use of a limited number of samples. To compensate for inaccurate parameter estimation, the algorithm exploits this feature to allocate more resources, *i.e.,* transmit power. However, if N is very small, a significant bias may be introduced to the estimates, the bounds become very loose, and their use may not ensure the performance as we discuss later. It is important to notice that using the principle behind **Theorem 1** to characterize the distribution tail along with the use of confidence bounds of the parameters introduces robustness to our approach. This means that the QoS demands are guaranteed even in the presence of critical estimation errors in adverse network conditions. Fig. 4 sketches the behavior of the confidence region between the obtained bounds. Note that the actual outage probability will be within the confidence region (green color) with Γ confidence but will also be below the target ζ_k with $\Gamma+\alpha(100-\Gamma)$ confidence, where $\alpha \in (0,1)$ is unknown.

[3]For a full definition of the Fisher information, please refer to [25].

Algorithm 1 Robust Minimum-Power Precoding for URLLC.

Inputs: $\rho,\ \Gamma,\ \{\hat{\mathbf{h}}_k\},\ \{\mathcal{E}_k\},\ p_{min},\ p_{max},\ \Delta p$
Outputs: $\{\mathbf{w}_k\}$

1: Initialize $p_k \leftarrow p_{min}\ \forall k$
2: Compute $\mathcal{H}_k\ \forall k$ according to (9)
3: Compute $\mathbf{u}_k$ and $\phi_k\ \forall k$ according to (20) and (14), respectively
4: Initialize the outage bound $\mathcal{O}_{k,UB} = 1\ \forall k$
5: **while** $\mathcal{O}_{k,UB} > \zeta_k\ \forall k$ **and** $\sum_{k=1}^{K} p_k \leq p_{max}$ **do**
6: For UE k compute $\gamma_{k,N}$ and ψ_k according to (10) and (13), respectively
7: Find $\mu_k\big(\{\mathbf{w}_k\}, \{\tilde{\mathbf{h}}_{k,n}\}\big)$ as the $\rho-$quantile of ψ_k
8: Compute the data Q_k in (18)
9: Fit the GPD to Q_k with confidence Γ to obtain $[\hat{\upsilon}_{k,LB}, \hat{\xi}_{k,LB}], [\hat{\upsilon}_{k,UB}, \hat{\xi}_{k,UB}]$
10: Evaluate $[\hat{\upsilon}_{k,UB}, \hat{\xi}_{k,UB}]$ in (19) to obtain $\mathcal{O}_{k,UB}$
11: **if** $\mathcal{O}_{k,UB} < \zeta_k$ **then**
12: $\mathbf{w}_k = \sqrt{p_k}\mathbf{u}_k$
13: pick another UE k
14: **else**
15: $p_k = p_k + \Delta p$
16: **end if**
17: **end while**

IV. Benchmark Approach

As a benchmark, we consider the work in [13], where the authors solved the transmit power minimization problem with SINR constraint for a single UE-MIMO system given by

$$\mathbf{P2a}: \underset{\mathbf{W}_k \succeq 0}{\text{minimize}} \quad \text{Tr}(\mathbf{W}_k) \tag{21a}$$

$$\text{subject to } \gamma_k(\mathbf{W}_k, \hat{\mathbf{h}}_k + \mathbf{e}_k) < \gamma_k^{tar} \ \forall \mathbf{e}_k : ||\mathbf{e}_k|| \leq \epsilon, \tag{21b}$$

where $\mathbf{W}_k = \mathbf{w}_k\mathbf{w}_k^H$ and $||\mathbf{e}_k|| \leq \epsilon$ ensures that all possible channels $\mathbf{h}_k$ in $\mathbb{C}^M$ lie inside an ellipsoid centered at the estimated channel $\hat{\mathbf{h}}_k$ with radius ϵ. To guarantee a certain level of reliability ζ_k, it is enough to control the radius ϵ of the ellipsoid such that $100 \times (1-\zeta_k)\%$ of the channels $\mathbf{h}_k$

lie inside the boundaries. This can be achieved by defining $\epsilon = \mathcal{Q}\big(100 \times (1-\zeta_k), ||\mathcal{E}_k||\big)$, which also imposes a minimum number of required samples $N = 1/\zeta_k$ in the set $\mathcal{E}_k$ to effectively determine the quantile. **P2a** is not convex in its current form. Therefore, it is transformed into the equivalent SDP, thus convex problem [13]

$$\mathbf{P2b}: \underset{\mathbf{W}_k,\mathbf{Z},\Omega}{\text{minimize}} \quad \text{Tr}(\mathbf{W}_k) \tag{22a}$$

$$\text{subject to } \text{Tr}\big[(\mathbf{Z}-\mathbf{W}_k)\hat{\mathbf{h}}_k^H\hat{\mathbf{h}}_k\big]+\epsilon^2\Omega+\gamma_k^{tar} \leq 0, \tag{22b}$$

$$\begin{bmatrix} \mathbf{Z} & \mathbf{W}_k \\ \mathbf{W}_k & \mathbf{W}_k+\Omega\mathbf{T} \end{bmatrix} \succeq 0, \tag{22c}$$

$$\mathbf{W}_k \succeq 0, \tag{22d}$$

$$\Omega \geq 0, \tag{22e}$$

where $\mathbf{Z}$ and Ω depict auxiliary variables, and $\mathbf{T} = \Upsilon\mathbf{I}$ with $\mathbf{I}$ as the identity matrix. Υ determines the shape of the ellipsoid being a sphere for the case $\Upsilon = 1$. The complexity of this problem grows with the number of variables $5M^2+2$ in polynomial time [26], specifically as $\mathcal{O}\big((5M^2+2)^3\big)$, and the solution can be found using common solvers/algorithms such as CVX or interior point methods (IPM).

The main disadvantages of this approach are related to the use of LMIs and the computation complexity for finding the solution. Moreover, the need for at least $1/\zeta_k$ samples to efficiently compute ϵ represents another key drawback.

V. Numerical Results

In this section, we evaluate the performance of the proposed algorithm. For this, we consider that the BS is equipped with a uniform linear array and assume the spatially-correlated Rayleigh fading model for the channels where $\mathbf{h}_k \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}_k)$ such that $\mathbf{R}_k$ represents the spatial correlation matrix. For $\mathbf{R}_k$, we adopt the local scattering spatial correlation model and its approximation for Gaussian angular distribution with half-wavelength antenna separation [27]

$$[\mathbf{R}_k]_{t,m} = \frac{\beta_k}{L}\sum_{l=1}^{L}\exp(\pi\mathbb{i}(t-m)\sin\varphi_{k,l}) \times \exp(-\frac{1}{2}\sigma_{\varphi_k}^2\pi(t-m)\cos\varphi_{k,l}), \tag{23}$$

where β_k denotes the average channel gain (accounting only for path-loss) of UE k, L denotes the number of multi-path clusters, $\varphi_{k,l} \sim \mathcal{U}\big(\bar{\varphi}_k - \frac{2\pi}{9},\ \bar{\varphi}_k + \frac{2\pi}{9}\big)$ is the nominal angle of arrival of cluster l for UE k, which is uniformly distributed around the azimuth angle of the UEs relative to the boresight of the BS antenna array $\bar{\varphi}_k$. Moreover, σ_{φ_k} depicts the angular standard deviation of the paths within a multi-path cluster. The estimation error is assumed to be distributed as $\mathbf{e}_k \sim \mathcal{CN}(\mathbf{0}, \frac{\sigma_n^2}{p_{ul}\tau_e}\mathbf{I})$, which corresponds to the scenario without pilot contamination [20]. All UEs are assumed to have the same average channel gain β_k and UL transmit power p_{ul} for simplicity in the modeling, but $\bar{\varphi}_k \sim \mathcal{U}(0, 2\pi)$. The noise power is given by $\sigma_n^2 = -173.8 + 10\log_{10} BW + NF$ dBm where BW and NF represent the communication bandwidth and noise figure, respectively. Moreover, we use $f = 10\log_{10}(\cdot)$ as the concave function in the EVT-based approach. All simulation parameters are displayed in Table II and are based on [13], [22], [23], [27], [28]. **P2b** for the benchmark approach is solved using the CVX tool. Finally, the performance of the algorithm in terms of outage probability is evaluated in Matlab using Monte Carlo such that 3×10^7 error vectors are generated for each channel coefficient.

TABLE II: Simulation parameters

Parameter	Value
M	4, 8
K	$1-8$
p_{min}	-30 dBm
p_{max}	47 dBm [23]
N	10^4
NF	7 dB [27]
BW	60 kHz [28]
τ_f	42 (OFDM symbols)-0.75 ms [28]
τ_e	$1-8$ [28]
B	256 [28]
ρ	0.95 [22]
Γ	$10-90\%$
β_k	-115 dB [27]
ζ_k	$10^{-6}-10^{-1}$
σ_{φ_k}	$\frac{\pi}{36}$ [27]
L	10
p_{ul}	20 dBm
Υ	1 [13]

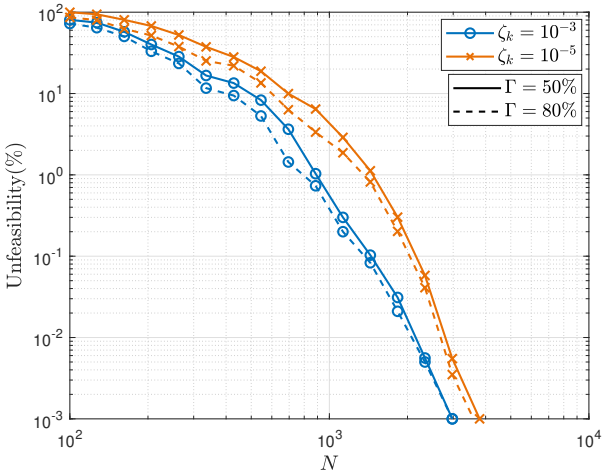

Fig. 5: Unfeasibility of the algorithm as a function of the number of samples N with $M = 8$ and $\tau_e = 1$.

A. Performance evaluation for single URLLC UE

In the single-UE case, we exploit MRT precoding, which allows maximizing the received power in the spatial direction of the UE. Fig. 5 shows the unfeasibility rate of the proposed algorithm as a function of the number of samples N for $\zeta_k = 10^{-3}$ and $\zeta_k = 10^{-5}$. Notice that as N increases, the algorithm will likely find a solution for any channel realization. However, for smaller N, the variance of the estimates increases, more bias is introduced, and the bounds become too loose. Therefore, the EVT approach is more prone to failures, *e.g.,* outage events, which impacts negatively the reliability of the system as displayed in Fig. 6.

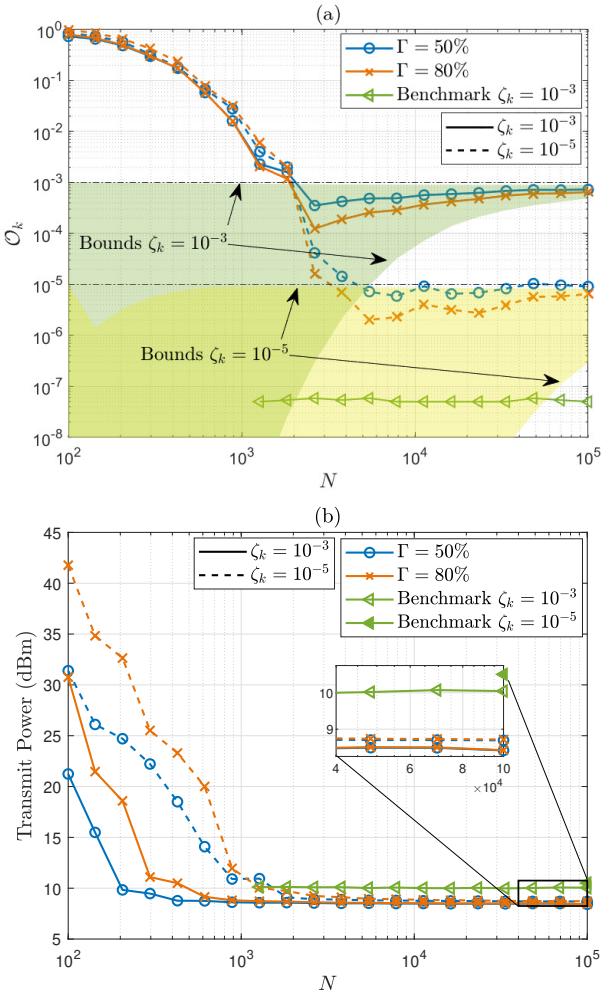

Fig. 6: Outage probability (a) and transmit power (b) as a function of the number of estimation error samples. We set $M = 8$ and $\tau_e = 1$, and employ MRT for the EVT approach. Green and yellow regions represent the outage probability regions between $\mathcal{O}_{k,LB}$ and $\mathcal{O}_{k,UB}$ for $\xi_k = 10^{-3}$ and $\xi_k = 10^{-5}$ with $\Gamma = 80\%$, respectively. The bound regions are below ζ_k for every found solution, even for small N.

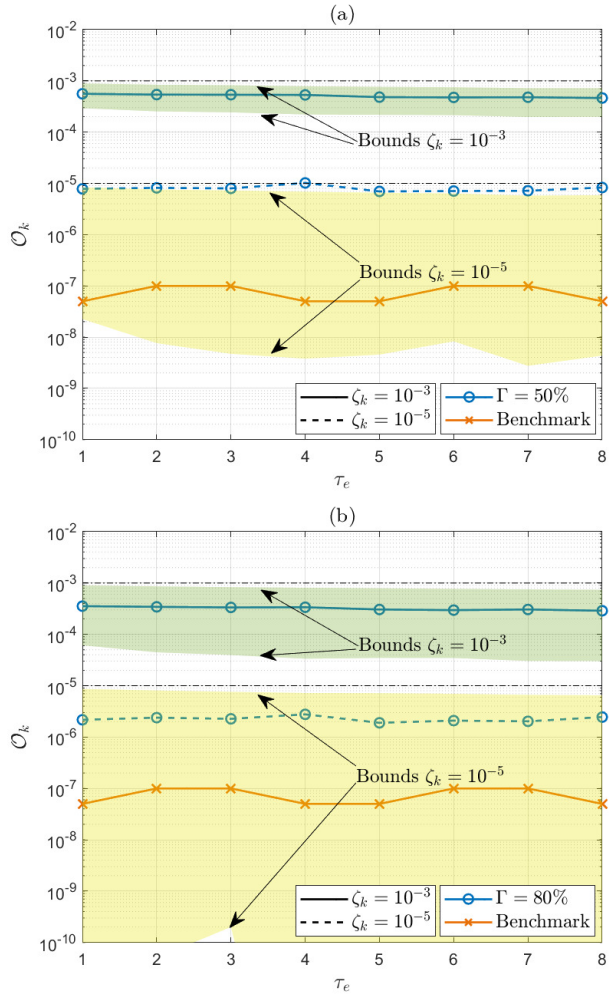

Fig. 7: Outage probability and confidence bounds for $\Gamma = 50\%$ (a) and $\Gamma = 80\%$ (b) as a function of the pilot lengths τ_e. We set $M = 8$ and $N = 10^4$, and employ MRT for the EVT approach. Green and yellow regions represent the outage probabilities between $\mathcal{O}_{k,LB}$ and $\mathcal{O}_{k,UB}$.

Fig. 6 (a) shows the achievable outage probabilities and confidence regions as functions of the total number of samples N with minimum pilot length, *i.e.*, $\tau_e = 1$. Note that the outage probabilities attained by our approach are above the target ζ_k and out of the confidence region when exploiting a small number of samples, *e.g.*, $N \leq 2000$ for $\zeta_k = 10^{-3}$. This is because the use of small sets Q_k introduces significant bias to the estimates leading to a higher failure rate of the algorithm as depicted in Fig. 5. Observe that the upper bound is always below ζ_k whenever a solution is found, even for the range of small N. Also note that as N increases for a given Γ, the confidence region becomes narrower, *i.e.,* less uncertainty, and the outage probabilities go below the targets ζ_k due to more reliable parameter estimation. Also, for larger N and high-reliability targets, *e.g.,* $\zeta = 10^{-5}$, the actual outage probabilities are below ζ_k especially if the fitting confidence is sufficiently high, *e.g.,* $\Gamma = 80\%$. Under these assumptions, the algorithm can ensure the outage requirements even in cases when $N < 1/\zeta_k$ by allocating extra transmit power, which compensates for the uncertainty that arises due to fitting with a limited number of samples. The benchmark scheme meets the outage requirements amply for both targets when $N \geq 1/\zeta_k$ since that represents the minimum number of required samples for finding the quantiles as discussed in Section IV, thus, evincing the advantages of EVT-based approaches over schemes derived from the CLT. The outage probability values in the case of $\zeta_k = 10^{-5}$ are not displayed as they are smaller than 10^{-8} and are thus difficult to estimate due to computational resource limitations. Notice that the benchmark approach achieves lower outage probabilities than ours but at the cost of higher power consumption, as we will discuss next. Fig. 6 (b) focuses the analysis on the required transmit power as a function of N. Given a relatively small N, the EVT-based scheme is somewhat unstable and incurs high power consumption due to inaccuracies in the parameter estimation of the GPD. In contrast, the transmit power tends to stably reduce with N, converging as $N \to \infty$. The power requirements of the benchmark approach increase slightly with N and are more than 1 dB above our approach's requirements in the stability region. Notice that for $\zeta_k = 10^{-5}$, there

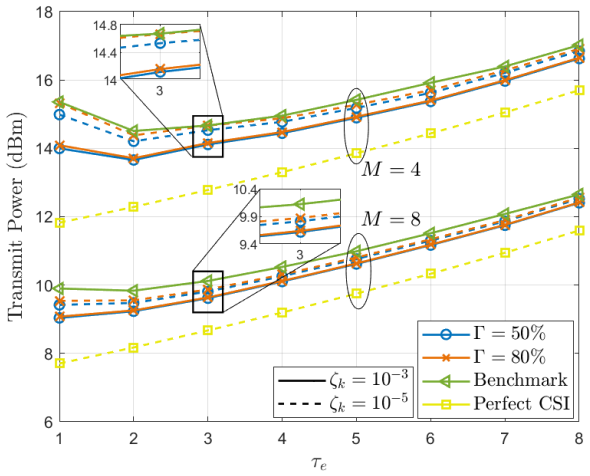

Fig. 8: Transmit power as a function of the pilot lengths τ_e. We set $M \in \{4, 8\}$ and $N = 10^4$, and employ MRT for the EVT approach.

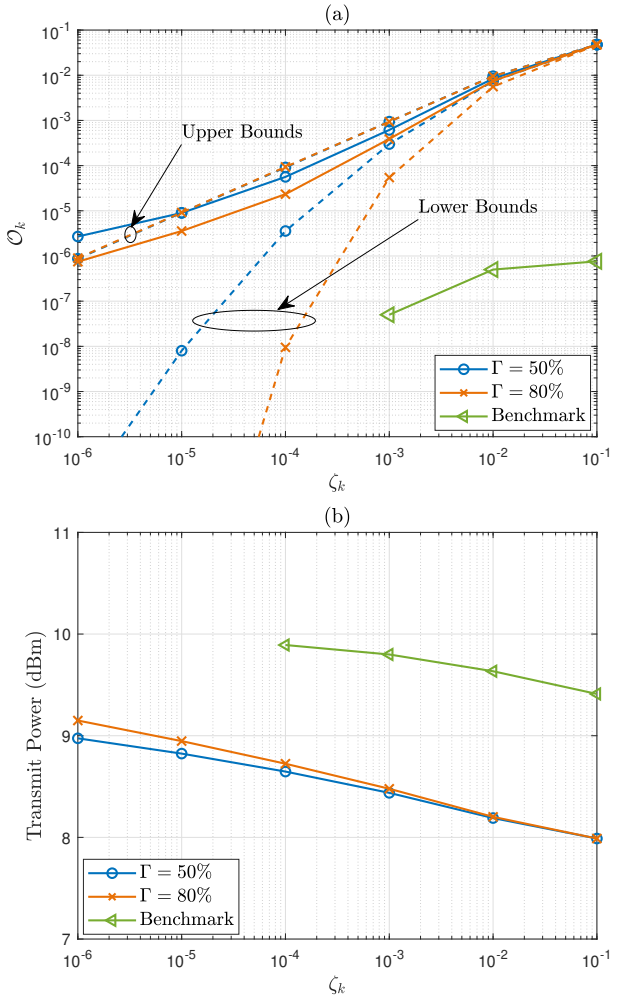

Fig. 9: Outage probability bounds (a) and transmit power (b) as a function of the outage targets ζ_k. We set $M = 8$, $\tau_e = 1$, and $N = 10^4$, and employ MRT for the EVT approach.

is only one feasible point at $N = 10^5$ since that is the minimum number to compute the quantile. The figure shows that our method's main advantage is reducing the transmit power while making the outage probabilities approach closely the targets, which is not achieved by the benchmark under the same requirements. Note that relying on historical data allows adaptation over time to the evolving nature of the system. We assume $N = 10^4$ for the remaining simulations.

Fig. 7 (a) and Fig. (b) show the outage probability and bound regions versus τ_e for $\Gamma = 50\%$ and $\Gamma = 80\%$, respectively. The figures also show the outage probabilities achieved by the benchmark approach. Notice that the bound regions get wider as the fitting confidence increases and that for low fitting confidence, *e.g.,* $\Gamma \leq 50\%$, the outage constraint is more likely to be violated, especially for smaller ζ_k as it is the case of $\zeta_k = 10^{-5}$ in Fig. 7 (a). Thus, a higher Γ may be required to meet the reliability requirements in practice. However, in such a case, we can notice a gap between the target ζ_k and the true outage that may lead to a slight extra power consumption. This gap cannot be safely reduced in practice since we cannot predict where the actual outage probability will be within the bound regions.

Fig. 8 shows the transmit power required for achieving $\zeta_k = 10^{-3}$ and $\zeta_k = 10^{-5}$ as a function of the pilot lengths given $M \in \{4, 8\}$ and $\Gamma \in \{50\%, 80\%\}$. Notice that the gap between the transmit power for $\Gamma = 50\%$ and $\Gamma = 80\%$ increases with the reliability level, being larger for $\zeta_k = 10^{-5}$. Interestingly, there is a pilot length that minimizes the transmit power depending on the number of antennas M. The estimation error may be significant given a relatively small τ_e, leading to higher power requirements to achieve a certain SINR. On the other hand, a relatively large τ_e implies better channel estimation and a smaller τ_{dl}, consequently higher SINR requirements and thus transmit power. For the specific results illustrated in Fig. 8, when the number of antennas is $M = 4$, the diversity and degrees of freedom gains of the system are low, which implies that a better channel estimation is required to achieve the requirements, *i.e.,* $\tau_e > 1$.

For $M = 8$, the system takes advantage of extra diversity and degrees of freedom gains offered by the additional four antennas, thus optimally meeting the requirements with a single-symbol pilot, *i.e.*, $\tau_e = 1$. Also note the need for only ~ 0.5 dB of power to go from $\zeta_k = 10^{-3}$ to $\zeta_k = 10^{-5}$ for the EVT approach at the optimal solution, which is significantly smaller than the gap in multi-UE scenarios due to interference as discussed later. Furthermore, the figure shows the required power when the BS has perfect CSI knowledge, *i.e.,* $\mathbf{e}_k = \mathbf{0}$, and the required power for the benchmark approach, which exceeds in ~ 1 dB our proposed method at the optimal solution. In the following, we adopt $\tau_e = 1$ for the results corresponding to a single URLLC UE.

Fig. 9 (a) shows the outage bounds as a function of the outage targets. Interestingly, the obtained upper bound obeys $\mathcal{O}_{k,UB} \approx \zeta_k$. Notice that for the most strict target, *i.e.,* $\zeta_k = 10^{-6}$, the outage probability violates the constraint (or is about to), which suggests increasing the number of samples N and/or increasing Γ. Fig. 9 (b) displays the transmit power required to maintain the actual outage probability inside the region delimited by the upper and lower bounds in Fig. 9 (a).

This article has been accepted for publication in IEEE Transactions on Wireless Communications. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TWC.2024.3396496

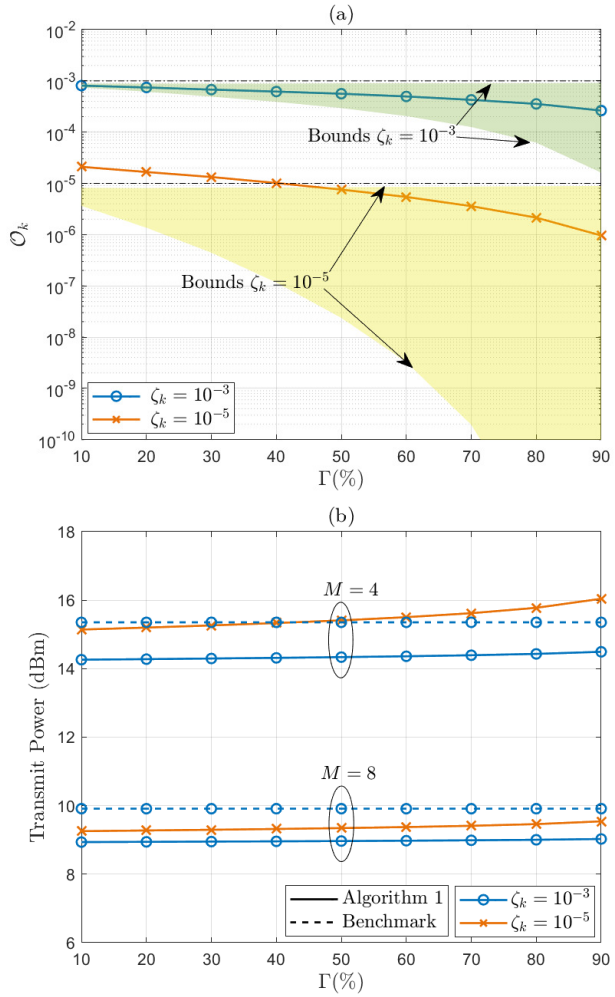

Fig. 10: Outage probability (a) and transmit power (b) as a function of the confidence level Γ. We set $M \in \{4, 8\}$, $\tau_e = 1$, $N = 10^4$, and employ MRT for the EVT approach. Green and yellow regions represent the outage probabilities between $\mathcal{O}_{k,LB}$ and $\mathcal{O}_{k,UB}$. A plot of transmit power for the benchmark with $\zeta_k = 10^{-5}$ is not displayed since the number of used samples is smaller than 10^5.

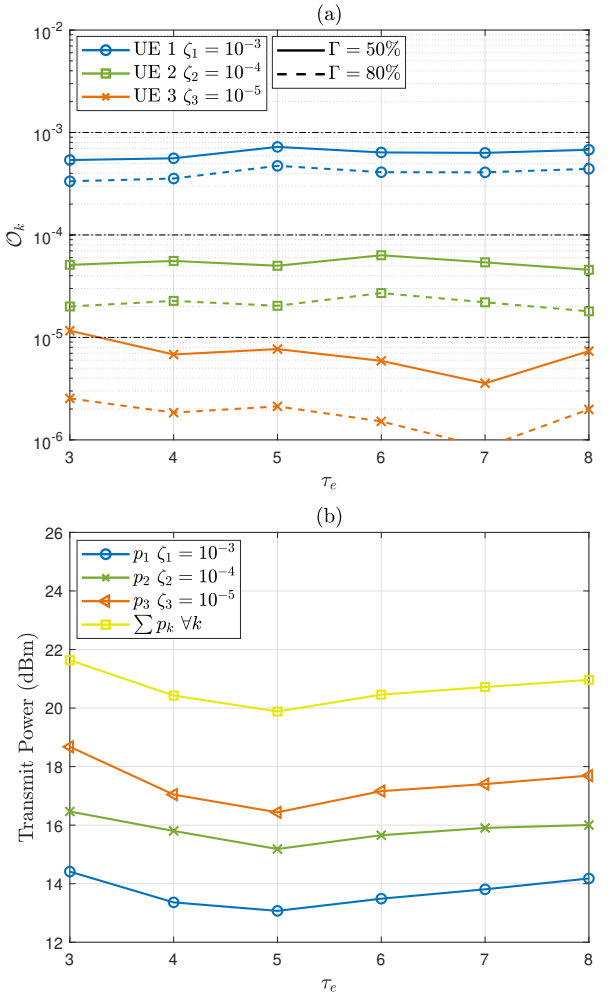

Fig. 11: Outage probability (a) and transmit power (b) as a function of the pilot lengths τ_e. In (b), we set $\Gamma = 80\%$, $M = 8$, $K = 3$, and $N = 10^4$, and we employ ZF precoding.

The required power increases as ζ_k decreases, which is crucial for achieving ultra-reliability. Also, notice that the gap between the transmit powers for $\Gamma = 50\%$ and $\Gamma = 80\%$ increases as the reliability target becomes more stringent. It is important to highlight that the energy efficiency gains with respect to the benchmark slightly increase when relaxing the outage requirements, *i.e.,* increasing ζ_k, since the transmit power decreases faster for the EVT scheme.

Fig. 10 (a) shows the actual outage probabilities and the outage bounds for a range of confidence levels on the GPD fitting. Notice that the target $\zeta_k = 10^{-5}$ is violated whenever $\Gamma < 40\%$ which does not occur for $\zeta_k = 10^{-3}$. This suggests using a larger Γ as the reliability requirement gets stricter. On the other hand, Fig. 10 (b) depicts the performance concerning required transmit power for $M = \{4, 8\}$. Here, it is shown that the power gap between configurations with different confidence levels Γ increases as M decreases and the target ζ_k becomes stricter. For instance, moving from $\Gamma = 10\%$ to $\Gamma = 90\%$ with $M = 8$ requires an increment of 0.09 dB and 0.286 dB for $\zeta_k = 10^{-3}$ and $\zeta_k = 10^{-5}$, respectively. Furthermore, with $M = 4$ the increments are 0.23 dB and 0.9 dB for $\zeta_k = 10^{-3}$ and $\zeta_k = 10^{-5}$, respectively. This means that the fitting confidence becomes less expensive regarding power consumption as the number of antennas increases.

B. Performance evaluation for multiple URLLC UEs

Fig. 11 (a) and Fig. (b) show the performance of the proposed algorithm for multiple UEs in terms of outage probability and power consumption, respectively, and as a function of the pilot lengths τ_e for $\zeta_1 = 10^{-3}$, $\zeta_2 = 10^{-4}$ and $\zeta_3 = 10^{-5}$ with ZF precoding. Note that ZF is an efficient multi-user precoding approach, especially in interference-limited networks, as it allows suppressing the interference in the spatial direction of the targeted UE. Notice that similar to the single-UE case, the constraint $\zeta_3 = 10^{-5}$ may be violated for the case $\Gamma = 50\%$, *e.g.,* for $\tau_e = 3$, but all targets are guaranteed for $\Gamma = 80\%$. Interestingly, the pilot length that minimizes the total transmit power in the multi-UE case is $\tau_e = 5$. Notably, the total power is minimized with $\tau_e > K$ driven by the imperfect interference cancellation. It is worth

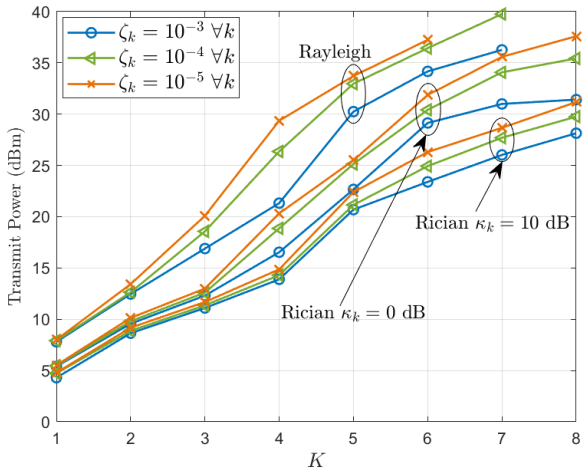

Fig. 12: Transmit power as a function of the number of UEs K. The figure shows the performance for Rayleigh fading and Rician fading with $\kappa_k = \{0, 10\}$ dB $\forall k$, $\Gamma = 80\%$, $M = 8$, and $N = 10^4$ while employing ZF precoding.

highlighting the requirement of around 2 dB of extra power to go up or down one order of magnitude in the reliability at the optimal solution.

Fig. 12 shows the transmit powers for different numbers of UEs in Rayleigh fading but also in Rician fading, *i.e.,* $\mathbf{h}_k = \sqrt{\frac{\kappa_k}{\kappa_k+1}}\mathbf{h}_{k,LOS} + \sqrt{\frac{1}{\kappa_k+1}}\mathbf{h}_{k,NLOS}$ where the first and second component represent the line-of-sight and scattering (non-line-of-sight) propagation components, respectively. Moreover, κ_k depicts the Rician factor, $\mathbf{h}_{k,NLOS} \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}_k)$, $\mathbf{h}_{k,LOS} = [1, e^{\mathbb{i}\theta_1}, ..., e^{\mathbb{i}\theta_{M-1}}]$ where θ_m represents the phase shift of the signal with respect of the first antenna element and κ_k depicts the line-of-sight factor of UE k. Notice that the gap in the transmit power for different outage targets increases with the number of UEs K. This is because the interference grows as the reliability target increases due to the increment in the required transmit power. Also, note that as the number of UEs increases, the power difference between different channel models increases for any outage target. For instance, for $\zeta_k = 10^{-5}$, the power difference is around 1 dB in a single-UE scenario, while the difference is around 7 dB in a network with eight UEs when comparing Rician fading with $\kappa_k = 0$ dB and $\kappa_k = 10$ dB.

VI. Conclusions

This work considered a minimum-power precoding design problem for serving multiple UEs in the DL with imperfect CSI and URLLC constraints. We proposed a solving algorithm that exploits CSI estimation error information and state-of-the-art precoding schemes such as MRT and ZF precoding. Moreover, we used the EVT framework to capture outage events that arise with low probability. Precisely, we fit data obtained from artificially generated SINR values to the GPD with different confidence levels to model rare error events. We evaluated the performance of the presented approach through simulations and compared it with a worst-case robust precoding method in the literature. We showed that the proposed method outperforms the benchmark approach and that there is an optimal pilot length that minimizes the transmit power. The confidence level influences the latter when fitting the data to the GPD.

References

[1] N. Mahmood, O. López, O. Park, I. Moerman, K. Mikhaylov, E. Mercier, A. Munari, F. Clazzer, S. Böcker, and H. Bartz *(Eds.)*, "White paper on critical and massive machine type communication towards 6G [white paper]," *6G Research Visions*, vol. 11, 2020. [Online]. Available: http://urn.fi/urn:isbn:9789526226781

[2] P. Popovski, C. Stefanović, J. J. Nielsen, E. de Carvalho, M. Angjelichinoski, K. F. Trillingsgaard, and A.-S. Bana, "Wireless access in ultra-reliable low-latency communication (URLLC)," *IEEE Transactions on Communications*, vol. 67, no. 8, pp. 5783–5801, Aug. 2019.

[3] O. L. A. López, N. H. Mahmood, M. Shehab, H. Alves, O. M. Rosabal, L. Marata, and M. Latva-Aho, "Statistical tools and methodologies for ultrareliable low-latency communication—a tutorial," *Proceedings of the IEEE*, vol. 111, no. 11, pp. 1502–1543, 2023.

[4] S. Coles, J. Bawa, L. Trenner, and P. Dorazio, *An introduction to statistical modeling of extreme values*. Springer, 2001, vol. 208.

[5] N. Mehrnia and S. Coleri, "Wireless channel modeling based on extreme value theory for ultra-reliable communications," *IEEE Transactions on Wireless Communications*, vol. 21, no. 2, pp. 1064–1076, 2021.

[6] C.-F. Liu and M. Bennis, "Ultra-reliable and low-latency vehicular transmission: An extreme value theory approach," *IEEE Communications Letters*, vol. 22, no. 6, pp. 1292–1295, 2018.

[7] S. Samarakoon, M. Bennis, W. Saad, and M. Debbah, "Distributed federated learning for ultra-reliable low-latency vehicular communications," *IEEE Transactions on Communications*, vol. 68, no. 2, pp. 1146–1159, 2019.

[8] P. Zheng, Y. Zhu, Y. Hu, and A. Schmeink, "Data-driven extreme events modeling for vehicle networks by personalized federated learning," in *International Symposium on Wireless Communication Systems (ISWCS)*. IEEE, 2022, pp. 1–6.

[9] L. Li, L. Yang, X. Guo, Y. Shi, H. Wang, W. Chen, and K. B. Letaief, "Delay analysis of wireless federated learning based on saddle point approximation and large deviation theory," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 12, pp. 3772–3789, 2021.

[10] M. K. Abdel-Aziz, C.-F. Liu, S. Samarakoon, M. Bennis, and W. Saad, "Ultra-reliable low-latency vehicular networks: Taming the age of information tail," in *IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2018, pp. 1–7.

[11] N. Mehrnia and S. Coleri, "Extreme value theory based rate selection for ultra-reliable communications," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 6, pp. 6727–6731, 2022.

[12] D. E. Pérez, O. L. A. López, and H. Alves, "Robust downlink multi-antenna beamforming with heterogenous CSI: Enabling eMBB and URLLC Coexistence," *IEEE Transactions on Wireless Communications*, vol. 22, no. 6, pp. 4146–4157, 2022.

[13] J. Wang and D. P. Palomar, "Worst-case robust MIMO transmission with imperfect channel knowledge," *IEEE Transactions on Signal Processing*, vol. 57, no. 8, pp. 3086–3100, 2009.

[14] G. Zheng, K.-K. Wong, and T.-S. Ng, "Robust linear MIMO in the downlink: A worst-case optimization with ellipsoidal uncertainty regions," *EURASIP Journal on Advances in Signal Processing*, vol. 2008, pp. 1–15, 2008.

[15] K.-Y. Wang, A. M.-C. So, T.-H. Chang, W.-K. Ma, and C.-Y. Chi, "Outage constrained robust transmit optimization for multiuser MISO downlinks: Tractable approximations by conic optimization," *IEEE Transactions on Signal Processing*, vol. 62, no. 21, pp. 5690–5705, 2014.

[16] M. Medra, Y. Huang, W.-K. Ma, and T. N. Davidson, "Low-complexity robust MISO downlink precoder design under imperfect CSI," *IEEE Transactions on Signal Processing*, vol. 64, no. 12, pp. 3237–3249, 2016.

[17] M. Medra and T. N. Davidson, "Robust MISO downlink: An efficient algorithm for improved beamforming directions," in *IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*. IEEE, 2016, pp. 1–5.

[18] F. Sohrabi and T. N. Davidson, "Coordinate update algorithms for robust power loading for the MU-MISO downlink with outage constraints," *IEEE Transactions on Signal Processing*, vol. 64, no. 11, pp. 2761–2773, 2016.

[19] T. Li, H. Zhang, J. Qiao, and J. Tian, "Robust beamforming design with finite blocklength for URLLC," *IEEE Transactions on Vehicular Technology*, 2022.

[20] S. Roy and P. Fortier, "Maximal-ratio combining architectures and performance with channel estimation based on a training sequence," *IEEE Transactions on Wireless Communications*, vol. 3, no. 4, pp. 1154–1164, 2004.

[21] C. Scarrott and A. MacDonald, "A review of extreme value threshold estimation and uncertainty quantification," *REVSTAT-Statistical journal*, vol. 10, no. 1, pp. 33–60, 2012.

[22] W. H. DuMouchel, "Estimating the stable index α in order to measure tail thickness: A critique," *the Annals of Statistics*, vol. 11, no. 4, pp. 1019–1031, 1983.

[23] S. Ahmadi, *5G NR: Architecture, technology, implementation, and operation of 3GPP new radio standards*. Academic Press, 2019.

[24] J. A. Rice, *Mathematical statistics and data analysis*. Cengage Learning, 2006.

[25] A. Ly, M. Marsman, J. Verhagen, R. P. Grasman, and E.-J. Wagenmakers, "A tutorial on Fisher information," *Journal of Mathematical Psychology*, vol. 80, pp. 40–55, 2017.

[26] R. D. Monteiro and T. Tsuchiya, "Polynomial convergence of primal-dual algorithms for the second-order cone program based on the MZ-family of directions," *Mathematical programming*, vol. 88, no. 1, pp. 61–83, 2000.

[27] 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. [Online]. Available: http://dx.doi.org/10.1561/2000000093

[28] A.-S. Bana, G. Xu, E. De Carvalho, and P. Popovski, "Ultra reliable low latency communications in massive multi-antenna systems," in *52nd Asilomar Conference on Signals, Systems, and Computers*. IEEE, 2018, pp. 188–192.