Multivariate Extreme Value Theory Based Rate Selection for Ultra-Reliable Communications

Niloofar Mehrnia, *Graduate Student Member, IEEE*, and Sinem Coleri, *Fellow, IEEE*

***Abstract*—Diversity schemes play a vital role in improving the performance of ultra-reliable communication (URC) systems by transmitting over two or more communication channels to combat fading and co-channel interference. Determining an appropriate transmission strategy that satisfies the ultra-reliability constraint necessitates the derivation of the statistics of the channel in the ultra-reliable region and, subsequently, integration of these statistics into the rate selection while incorporating a confidence interval to account for potential uncertainties that may arise during estimation. In this paper, we propose a novel framework for ultra-reliable real-time transmission considering both spatial diversities and ultra-reliable channel statistics based on multivariate extreme value theory (MEVT). First, the tail distribution of the joint received power sequences obtained from different receivers is modeled while incorporating the inter-relations of extreme events occurring rarely based on the Poisson point process approach in MEVT. The optimum transmission strategies are then developed by determining the optimum transmission rate based on the estimated joint tail distribution and incorporating confidence intervals (CIs) into the estimations to cope with the availability of limited data. Finally, the system reliability is assessed by utilizing the outage probability metric. Through analysis of the data obtained from the engine compartment of the Fiat Linea, our study showcases the effectiveness of the proposed methodology in surpassing traditional extrapolation-based approaches. This innovative method not only achieves a higher transmission rate, but also effectively addresses the stringent requirements of ultra-reliability. The findings indicate that the proposed rate selection framework offers a viable solution for achieving a desired target error probability by employing a higher transmission rate and reducing the amount of training data compared to the conventional rate selection methods.**

***Index Terms*—Multivariate extreme value theory (MEVT), URLLC, wireless channel modeling, spatial diversity, rate selection, confidence interval, ultra-reliable communication (URC), 6 G.**

I. Introduction

ULTRA-RELIABLE communication (URC) plays a critical role in fifth-generation (5 G) and beyond networks, enabling mission-critical applications in transportation, manufacturing, and healthcare. Examples of these applications include remote control of robots, remote surgery, wireless sensor networks within vehicles, autonomous vehicles, and vehicular teleoperation [1], [2], [4], [5]. In current 5 G standards, ultra-reliability and low latency are typically combined to form ultra-reliable low latency communication (URLLC). However, there are specific scenarios, such as health monitoring or disaster recovery, where the highest level of reliability is essential, but a slightly higher latency of more than the conventional 1 ms (ms) is acceptable [5], [6]. To achieve the ultra-reliability target of 10^{-9} to 10^{-5} packet error rate (PER), diversity schemes including time diversity [7], frequency diversity [8], spatial diversity [9], and interface diversity [10] have been proposed. These schemes aim to reduce the required signal-to-noise ratio (SNR) to meet the stringent reliability constraints. Spatial diversity, which involves using multiple transmit and/or receive antennas to achieve reliable communication over fading channels, is preferred over frequency and interface diversities due to its cost-effectiveness and minimal transmission delay compared to time diversity [11]. However, implementing ultra-reliable communication relies on accurately modeling extreme quantiles that occur infrequently. Therefore, significant advancements in statistical modeling of multiple channels and the development of corresponding transmission strategies are necessary for the successful utilization of spatial diversity in designing ultra-reliable communication systems.

In a URC system with spatial diversity, a proper channel model can be derived by i) estimating the channel statistic within a more nuanced notation of coherence distance for ultra-reliable communication [9], ii) using machine learning data-driven frameworks [12], or iii) incorporating novel techniques to the characteristics associated with extreme events across multiple variables, such as proposing the utilization of power law expression techniques by extrapolating from channel data [5] and multivariate extreme value theory (MEVT) [13]. In [9], adjustments to the coherence distance definitions are made to accurately capture the distance at which wireless channels can be reliably predicted. The authors employ a quasi-static Rayleigh fading model, assuming spatial independence, to study channel dynamics in the context of URC. Furthermore, [12] addresses the estimation of channel-blocking incidents with probabilities on the order of $10^{-9}-10^{-5}$ using data-driven methods, without prior knowledge of dynamic channel statistics. This allows for the assessment of wireless connectivity reliability in ultra-reliable communication systems. However, machine learning

Manuscript received 20 December 2023; accepted 14 May 2024. Date of publication 22 May 2024; date of current version 17 October 2024. The work of Sinem Coleri was supported in part by Ford Otosan, and in part by the Scientific and Technological Research Council of Turkey 2247-A National Leaders Research under Grant #121C314. The review of this article was coordinated by Dr. Cunhua Pan. *(Corresponding author: Niloofar Mehrnia.)*

Niloofar Mehrnia is with the Department of Electrical and Electronics Engineering, Koc University, 34450 Istanbul, Türkiye, and also with the Department of Electrical Engineering and Computer Sciences, KTH Royal Institute of Technology, SE-100 44 Stockholm, Sweden (e-mail: nmehrnia17@ku.edu.tr).

Sinem Coleri is with the Department of Electrical and Electronics Engineering, Koc University, 34450 Istanbul, Türkiye (e-mail: scoleri@ku.edu.tr).

Digital Object Identifier 10.1109/TVT.2024.3404111

© 2024 The Authors. This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

data-driven approaches require a significant amount of training samples, typically exceeding a factor of $(10 \times \epsilon^{-1})$ to achieve the desired error probability ϵ. Additionally, [5] proposes a power law framework that extends traditional channel models into the ultra-reliable region through extrapolation. This framework can be further enhanced to accommodate multiple receivers by employing a simplified expression of maximum ratio combining (MRC). Nevertheless, these initial studies still rely on average statistics channels, such as Gaussian, Rayleigh, or Rician, to characterize channel behavior in the ultra-reliable region. Only recently, a novel methodology utilizing MEVT has been introduced in [13] to model multiple channels in a multiple-input multiple-output (MIMO)-URC system by deriving the lower tail statistics of received power. MEVT is a statistical discipline specializing in formulating methods for modeling dependencies among infrequent events by leveraging multidimensional boundary relationships. Despite numerous studies in the field of channel modeling, none have proposed a transmission strategy that relies on real-time assessment of the lower tail statistics of received power across multiple diversities.

The design of communication systems for ultra-reliable communications has been addressed in the context of adaptive relay selection techniques [14], and rate selection frameworks [6], [15], [16]. A cooperative communication-based relaying scheme has been proposed in [14] to provide ultra-high reliability while meeting the latency requirements at a moderate signal-to-noise ratio (SNR) by modifying the definition of traditional coherence time and revising fading dynamics of wireless channels in the context of ultra-high reliability. On the other hand, [6] uses a rate selection framework based on the extrapolation of average statistic-based channel models to design an ultra-reliable system. Moreover, [15] develops a data-driven rate selection framework where federated learning (FL) training has been used to estimate the transmission rate and assess the reliability of the system with minimum training time in the absence of knowledge about the channel state information (CSI). However, all of these designs use the extrapolation of average-statistics channel models, which have been demonstrated not to fit the empirical data in the ultra-reliable region [13], [16], [17], [18]. We have recently proposed a novel extreme value theory (EVT)-based framework in [16] for estimating and validating the optimal transmission rate to address the constraints of ultra-reliable communications for a single transmitter-receiver pair. Nevertheless, none of these studies incorporate the statistics of multiple channels in the ultra-reliable region based on real data into the design of communication systems.

The estimation accuracy of the channel parameters and the transmission strategies need to be derived in the ultra-reliable regime due to the limited amount of training data by using confidence interval (CI) [12], [19], [20] and bootstrapping techniques [17]. In [12], the authors propose a multi-layer perceptron (MLP) neural network (NN)-based machine learning data-driven framework incorporating CI to the statistics of a non-blocking connectivity duration to maintain the URLLC with the penalty of large training samples to satisfy the target error probability ϵ. In [19], the authors evaluate the outage probability within a confidence level to address the reliability of the URC system by incorporating a non-parametric statistical learning algorithm into a rate selection framework assuming extrapolation of the average statistics fading channels. Only recently, in [20], we proposed a pioneering framework based on EVT that estimates the optimal transmission rate with a high degree of confidence using a reduced number of training samples. The proposed approach was further validated by evaluating the outage probability for ultra-reliable communications in a single transmitter-receiver scenario. On the other hand, in [17], we develop a bootstrapping-based algorithm for the determination of the stopping conditions in collecting sufficient samples to estimate the tail statistics based on the normality assumptions of the return levels. Accordingly, if the sample size is sufficiently large to estimate the channel tail distribution of values exceeding a predetermined threshold by using the generalized Pareto distribution (GPD), the corresponding return levels obtained in the bootstrap iterations are normally distributed. However, none of these studies incorporate ultra-reliable communication statistics of the diversities into the accuracy of simultaneously estimating the channel parameters and resulting transmission strategies.

The main objective of this study is to propose a transmission strategy for a real-time URC system that employs spatial diversity by considering the MEVT-based statistics of multiple channels while incorporating confidence intervals into parameter estimation to cope with the limited availability of data. The spotlight is on two-dimensional or bi-variate instances, with the aim of illustrating key concepts and concerns related to MEVT while avoiding the added complexity associated with a comprehensive multivariate approach. However, this methodology can be easily generalized to multiple dimensions. According to the proposed methodology, the parameters of multiple channels are first estimated based on MEVT. Then, the optimal transmission rate is determined by incorporating the estimated multiple channel parameters while considering the ultra-reliability constraint subject to the target error probability ϵ. Finally, the reliability of a system operating at the determined transmission rate is assessed by employing the outage probability. The uncertainty of the transmission rate is also obtained in relation to the confidence intervals of the estimated channel parameters. The original contributions of the paper are listed as follows:

- We propose a novel framework for ultra-reliable real-time transmission considering both spatial diversities and ultra-reliable channel statistics based on MEVT, for the first time in the literature. The framework includes the modeling of the inter-relationship of the tail distributions of multiple channels by using the bi-variate GPD (BGPD) based on the Poisson point process approach, determination of the optimum transmission rate by using the estimated BGPD model, incorporating confidence intervals to the estimated transmission rate due to the restricted amount of data, and then assessment of the system reliability by utilizing the outage probability metric.
- We propose a novel algorithm for the rate selection in a multi-diversity system based on the Poisson point process approach, for the first time in the literature. The algorithm incorporates the parameters of the probability measure

function and BGPD into the optimum transmission rate to satisfy the ultra-reliability constraints.
- We propose an algorithm for the determination of the confidence interval of the transmission rate based on the confidence intervals of the estimated Uni-variate GPD (UGPD) parameters, for the first time in the literature. We determine the confidence intervals of the UGPD parameters by incorporating the bootstrapping-based bias-corrected accelerated (BCA) method that offers an order of magnitude improvement in accuracy for the parameter estimation using the maximum likelihood estimator (MLE).
- We utilize a combination of one omnidirectional transmitter and two directional receivers to capture the data within the engine compartment of a Fiat Linea vehicle under different driving scenarios and engine vibrations. Through our proposed methodology, we achieve superior modeling accuracy for URC compared to conventional extrapolation-based models that rely on average statistical channels for multi-channel modeling.

The rest of the paper is organized as follows: Section II describes the system model and assumptions considered throughout the paper. Section III presents the MEVT-based rate selection framework for determining the optimum transmission rate and its affiliated confidence interval for a system operating in URC. Section IV provides the channel measurement setup and the performance evaluation in determining the optimum transmission rate and outage probability, as well as the confidence interval for the estimated rate. Finally, concluding remarks and future works are given in Section V.

II. System Model

We consider a one-way communication in which a transmitter (Tx) sends data packets to two receivers, namely Rx1 and Rx2, at the total rate R, addressing the requirements of ultra-reliable communication. The framework can be easily extended for the scenario in which more than two receivers exist without loss of generality. Assuming a fixed and predetermined transmit power, estimating the received signal power is tantamount to determining the squared amplitude of the channel state information [6], [17].

At the training phase and before the transmission starts, the channel samples are collected from receivers Rx1 and Rx2 and converted into n independent and identically distributed (i.i.d.) sample sequences denoted by X^n and Y^n, respectively, by applying declustering method [17], [21]. The channel data in the training phase can be collected by using either the pilot signals or the records from the prior data transmissions. The recommended amount of required samples in the training phase is about $1/\epsilon$, where ϵ is the target error probability on the order of $10^{-9}-10^{-5}$ for the URC system. Suppose the channel is stationary according to the Augmented Dickey-Fuller (ADF) test results. In that case, the channel tail distribution is modeled by applying EVT to the received signal powers and estimating the parameters of the UGPD fitted to the channel tail distribution. Otherwise, external factors leading to variation over time in the parameters of the UGPD are identified and utilized to partition the sequence into stationary groups of size M. Then, EVT is applied to each stationary sequence for estimating the shape and scale parameters of UGPD as a change-point function of time, as explained in detail in [18]. The parameters of UGPDs obtained for the received powers of Rx1 and Rx2 are then mutually used to determine the tail distribution of the joint probability distribution of multiple channel sequences X^n and Y^n by using MEVT techniques to characterize the statistics of the inter-relationships of extreme events [13]. The transmitter assumes that the main source in the block fading channel is link outage [5], [6], [22].

Upon deriving MEVT-based channel model, transmission strategies are determined to estimate the optimum transmission rate with the goal of fulfilling a predetermined ϵ reliability constraint such that

$$p_F(R(X^n, Y^n)) \leq \epsilon, \tag{1}$$

where $R(X^n, Y^n)$ denotes the transmission rate estimated based on the n training received power samples of X^n and Y^n, F is the joint CDF of the training samples from channel data X^n and Y^n, and $p_F(R(X^n, Y^n))$ is the outage probability at transmission rate $R(X^n, Y^n)$ defined as

$$p_F(R(X^n, Y^n)) = P\left[R(X^n, Y^n) > \log_2(1+Z)\right], \tag{2}$$

where Z is any received power from test samples of receiver Rx1 or Rx2. The transmitter determines the optimum rate as a function of F as follows:

$$R(F) = sup\left\{R(X^n, Y^n) \geq 0 : p_F(R(X^n, Y^n)) \leq \epsilon\right\} \tag{3}$$

$$= \log_2\left(1 + F^{-1}(\varepsilon_n)\right), \tag{4}$$

where $F^{-1}(\varepsilon_n)$ is the ε_n-quantile of F, and ε_n is determined as a function of ϵ and the parameters of F with the goal of satisfying constraint (1).

Upon determining the optimal transmission rate using the proposed MEVT rate selection methodology at target error probability ϵ, the confidence intervals of the estimated UGPD parameters are computed, considering some uncertainty in the parameter estimation by incorporating the probability of wrong decision α. Subsequently, the statistics of transmission rate are estimated using MEVT, which incorporates the confidence intervals of the estimated UGPD parameters to satisfy the target error probability for a system equipped with spatial diversity operating in the ultra-reliable communication domain.

III. MEVT-Based Rate Selection Framework

The goal of the MEVT-based rate selection framework is to determine an optimum transmission rate along with the confidence interval based on the estimation of the lower tail statistics of multiple channels for real-time communication in the ultra-reliable regime. The proposed methodology involves a series of steps, including the conversion of the received power sample sequences into independent and identically distributed (i.i.d) samples. This is achieved through the application of the declustering method, which removes interdependency between the samples. Then, the inter-relationship of bi-variate extremes is modeled by utilizing BGPD based on the MEVT-Poisson

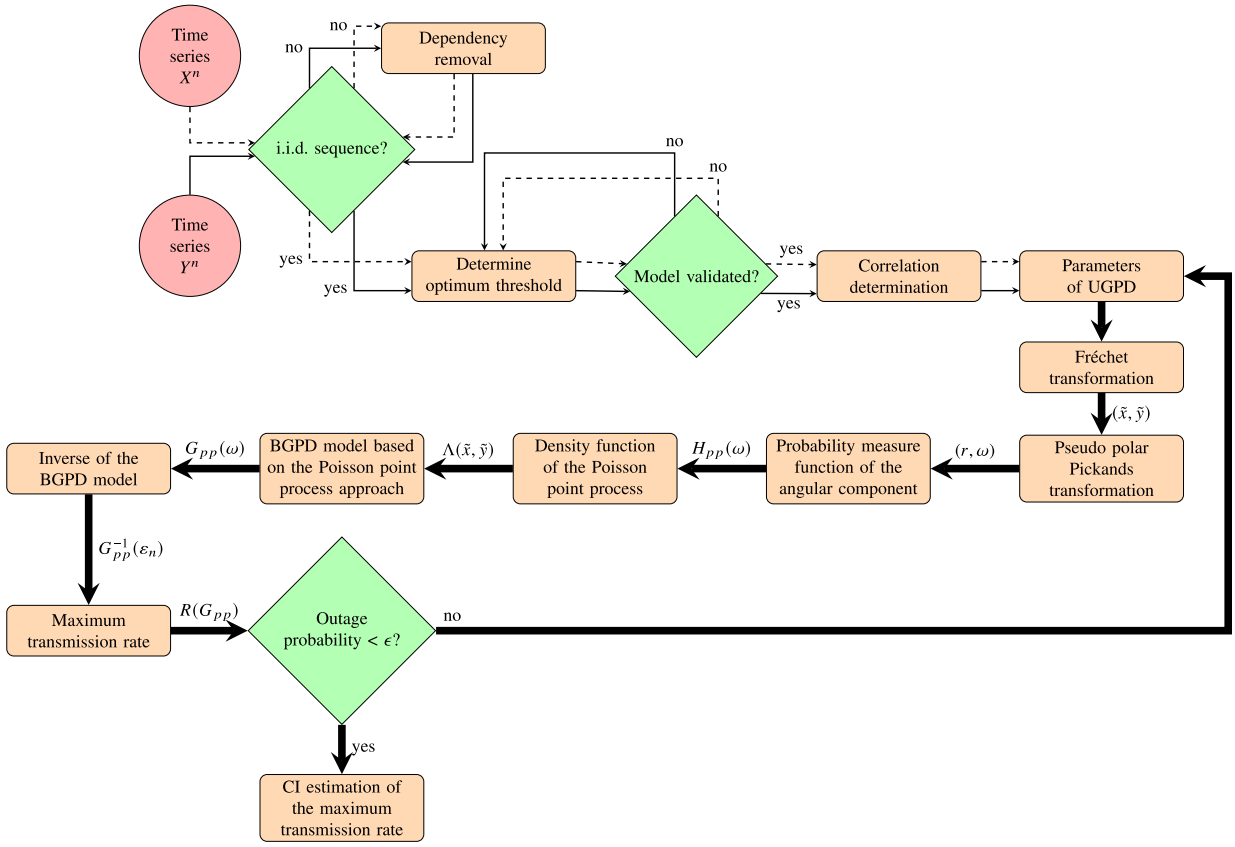


Fig. 1. Flow diagram of the proposed MEVT-based rate selection framework.

point process approach. Upon determining the optimum transmission rate based on the proposed MEVT-based rate selection framework, channel reliability is assessed based on the results of outage probability. Finally, the confidence interval is incorporated into the proposed rate selection framework by determining confidence intervals of the estimated parameters and the estimated optimum transmission rate for different values of training sample number n to specify the minimum required training samples for optimal estimation of the transmission rate. The applied algorithm for the proposed rate selection framework incorporating CI is depicted in Fig. 1 and explained in detail next.

A. Bi-Variate Channel Estimation

At the training phase, the channel samples are collected from receivers Rx1 and Rx2 and converted into i.i.d. samples through the declustering approach. In this approach, the samples are folded into multiple clusters, and then the minimum of each cluster is extracted; the minima are considered i.i.d. samples [17], [21]. Upon applying EVT to the resulting sequence of i.i.d. samples from individual channel sequences, the optimum thresholds are determined separately by applying two complementary methods based on EVT: mean residual life (MRL) and parameter stability methods [17], [21]. According to the MRL method, the optimum threshold is the highest threshold below which the mean of values exceeding a given threshold is linear with respect to the threshold. On the other hand, the optimum threshold based on the parameter stability method is the highest threshold below which the estimated shape and modified scale parameters are linear against the threshold [17], [21]. The MRL method can be applied prior to the estimation of the tail distribution by using UGPD, while the parameter stability method should be applied after deriving the tail statistics by using UGPD. Since the bi-variate analysis are limited to time periods where both thresholds are exceeded, in this study, only the exceedances that occur simultaneously are considered, while any exceedances that occur at different time intervals are excluded. Following this, the parameters of the UGPD are estimated for the exceedances of each channel data, using the maximum likelihood (ML) estimator with the corresponding optimal thresholds. The UGPD fitted to the tail of sample sequence X^n is formulated as

$$G_x(\tilde{\sigma}_x, \xi_x) = 1 - \left[1 + \frac{\xi_x(u_x - x)}{\tilde{\sigma}_x}\right]^{-1/\xi_x}, \tag{5}$$

where $(u_x - x)$ denotes the non-negative exceedance for observation $x \in X^n$ below optimum threshold u; and ξ_x and $\tilde{\sigma}_x$ are the estimated shape and scale parameters of UGPD, respectively. The validity of the fitted UGPD model to the tail distributions is assessed by using the probability plots, including the probability-probability (PP) plot and the quantile-quantile (QQ) plot. The PP and QQ plots are graphical techniques that are employed to compare the empirical values with the corresponding modeled results. In the PP plot, the empirical CDF of the

occurrence for each extreme value is plotted against the corresponding CDF obtained by the UGPD, while in the QQ plot, the empirical extreme quantile is plotted against the corresponding quantile obtained by the inverse of UGPD [17]. If the UGPD properly characterizes the extreme values exceeding threshold u, then both PP and QQ plots conform to the unit diagonal line, i.e., a straight line with the horizontal angle 45°. Thereafter, the inter-relationship of bi-variate extremes is modeled by applying the Poisson point process approach.

In bi-variate extreme analysis, the first step involves examining the presence of dependency within the tail samples and the total samples. The feasibility of spatial diversity is evaluated using the correlation between received powers in the total samples. Spatial diversity is considered reasonable if the correlation coefficient is within the range of 0.1 and 0.5 among the total samples, since otherwise fading occurs simultaneously at different links [23]. On the other hand, if the extremes of the two sample sequences are independent, i.e., the correlation coefficient approaches 0, there is no need for bi-variate modeling. However, if there is correlation between the tail samples, it is necessary to investigate the inter-relationship of the extreme values of receivers.

In the Poisson point process approach, the channel data is transformed in two steps: i) Fréchet transformation and ii) Pickands coordinates. The Fréchet transformation is applied on the obtained tail sequences of X^n and Y^n as follows:

$$\tilde{x} = -\left(\log\left\{1-\zeta_x\left[1+\frac{\xi_x(u_x-x)}{\tilde{\sigma}_x}\right]^{-1/\xi_x}\right\}\right)^{-1}, \quad x<u_x, \tag{6}$$

and

$$\tilde{y} = -\left(\log\left\{1-\zeta_y\left[1+\frac{\xi_y(u_y-y)}{\tilde{\sigma}_y}\right]^{-1/\xi_y}\right\}\right)^{-1}, \quad y<u_y, \tag{7}$$

where $\zeta_x = Pr(x<u_x)$ and $\zeta_y = Pr(y<u_y)$ for optimum thresholds u_x and u_y, respectively. Accordingly, the marginal distribution of variables $\tilde{x}$ and $\tilde{y}$ approximately have Fréchet distribution for $x<u_x$, $x\in X^n$, and $y<u_y$, $y\in Y^n$. After applying the Fréchet transformation, the pseudo polar Pickands transformation is applied by introducing pseudo polar radial r and angular ω components as

$$r = \frac{-\tilde{x}}{n}+\frac{-\tilde{y}}{n},$$
$$\omega = \frac{-\tilde{x}/n}{-\tilde{x}/n-\tilde{y}/n}, \tag{8}$$

where $n\in N$ is the length of vector $\tilde{X}^n$ and $\tilde{Y}^n$, and $\tilde{X}^n$ and $\tilde{Y}^n$ are the Fréchet transformation of X^n and Y^n, respectively. Upon determining the probability measure function of the Pickands angular coordinate $H_{pp}(\omega)$, satisfying the mean constraint $\int_{\omega=0}^{1}\omega\, H_{pp}(d\omega)=0.5$, the density function of the Poisson point process $\Lambda(\tilde{x},\tilde{y})$ is computed as

$$\Lambda(\tilde{x},\tilde{y}) = \int_{r=-\infty}^{0}\int_{\omega=0}^{1} 2\frac{dr}{r^2}dH(\omega)$$
$$= -2\int_{\omega=0}^{1}\max\left(\frac{\omega}{-\tilde{x}/n},\frac{1-\omega}{-\tilde{y}/n}\right)H_{pp}(d\omega), \tag{9}$$

where $\Lambda(\tilde{x},\tilde{y})$ is defined on space $\{(0,\infty)\times(0,\infty)\backslash(0,\tilde{x})\times(0,\tilde{y})\}$, denoting space $\{(0,\infty)\times(0,\infty)\}$ excluding sub-space $\{(0,\tilde{x})\times(0,\tilde{y})\}$. Finally, the BGPD model based on the Poisson point process approach is determined as [13]

$$G_{pp}(\tilde{x},\tilde{y}) = \exp(-\Lambda(\tilde{x},\tilde{y})). \tag{10}$$

B. Bi-Variate Rate Selection

The goal of this section is to choose the maximal transmission rate that achieves the target reliability ϵ by satisfying constraint (1). The maximum rate is determined as a function of G_{pp} as

$$R(G_{pp}) = sup\left\{R(X^n+Y^n)\geq 0 : p_F(R(X^n+Y^n))\leq\epsilon\right\}$$
$$= \log_2\left(1+G_{pp}^{-1}(\varepsilon_n)\right), \tag{11}$$

where $G_{pp}^{-1}(\varepsilon_n)$ is an estimate of a positive ε_n-quantile of the bi-variate channel G_{pp}. Here, the objectives are i) to determine $G_{pp}^{-1}(\varepsilon_n)$ and ii) to find ε_n that maximizes the transmission rate $R(G_{pp})$ while satisfying the target reliability ϵ based on constraint (1).

In order to determine $G_{pp}^{-1}(\varepsilon_n)$, we first obtain the inverse joint CDF function of the training samples from channel data X^n and Y^n.

Theorem 1: Let $G_{pp}(\hat{\tilde{x}},\hat{\tilde{y}})\approx G_{pp}(\hat{\omega})$ be the estimated BGPD model fitted to the training samples $X^n=\{x_1,x_2,\ldots,x_n\}$ and $Y^n=\{y_1,y_2,\ldots,y_n\}$. Then, the inverse joint CDF function $F^{-1}\sim G_{pp}^{-1}(\varepsilon_n)$ can be obtained as

$$G_{pp}^{-1}(\varepsilon_n) = H_{\hat{\omega}}^{-1}\left\{\frac{1}{2}\max(-\hat{\tilde{x}}/n-\hat{\tilde{y}}/n)\ln\varepsilon_n\right\}, \tag{12}$$

where $\hat{\tilde{x}}$ and $\hat{\tilde{y}}$ are the estimated Fréchet transformation of $G_x(\hat{\tilde{\sigma}}_x,\hat{\xi}_x)$ and $G_y(\hat{\tilde{\sigma}}_y,\hat{\xi}_y)$ fitted to X^n and Y^n training sequences, respectively, $\ln(.)$ function refers to the natural logarithm, $H_{\hat{\omega}}^{-1}(.)$ is the CDF-inverse of the β-distribution fitted to the pseudo-polar angular component $\hat{\omega}=\frac{\hat{\tilde{x}}/n}{\hat{\tilde{x}}/n+\hat{\tilde{y}}/n}$.

Proof: Assume that $G_{pp}(\hat{\tilde{x}},\hat{\tilde{y}}) = exp(-\Lambda(\tilde{x},\tilde{y}))$ models the bi-variate GPD fitted to the training samples X^n and Y^n. Then,

$$\Lambda(\tilde{x},\tilde{y}) = \int_{\hat{\omega}=0}^{1}\int_{\hat{r}=-\infty}^{0} 2\frac{d\hat{r}}{\hat{r}^2}dH(\hat{\omega})$$
$$= \int_{\hat{\omega}=0}^{1}\int_{\frac{-\hat{\tilde{x}}/n}{\min(-\hat{\tilde{x}}/n-\hat{\tilde{y}}/n)}}^{\frac{-\hat{\tilde{x}}/n}{\max(-\hat{\tilde{x}}/n-\hat{\tilde{y}}/n)}}\frac{-2}{\hat{\tilde{x}}/n}d\hat{\omega}\,dH(\hat{\omega}).$$

In the above equation, $\frac{d\hat{r}}{\hat{r}^2}$ has been replaced by $\frac{-2}{\hat{\tilde{x}}/n}d\hat{\omega}$ since $\hat{\omega}=\frac{-\hat{\tilde{x}}/n}{-\hat{\tilde{x}}/n-\hat{\tilde{y}}/n}=\frac{-\hat{\tilde{x}}/n}{\hat{r}}$, and therefore $\hat{r}$ can be written as $\hat{r}=\frac{-\hat{\tilde{x}}/n}{\hat{\omega}}$. Additionally, the boundary of $\hat{r}$ has been computed within the range $[\frac{-\hat{\tilde{x}}/n}{\min(-\hat{\tilde{x}}/n-\hat{\tilde{y}}/n)},\frac{-\hat{\tilde{x}}/n}{\max(-\hat{\tilde{x}}/n-\hat{\tilde{y}}/n)}]$ since

$$\min(-\hat{\tilde{x}}/n-\hat{\tilde{y}}/n)<\hat{r}<\max(-\hat{\tilde{x}}/n-\hat{\tilde{y}}/n)$$
$$\Rightarrow \frac{-\hat{\tilde{x}}/n}{\min(-\hat{\tilde{x}}/n-\hat{\tilde{y}}/n)}<\frac{-\hat{\tilde{x}}/n}{\hat{r}}=\hat{\omega}<\frac{-\hat{\tilde{x}}/n}{\max(-\hat{\tilde{x}}/n-\hat{\tilde{y}}/n)}$$

Moreover, $min(-\hat{\tilde{x}}/n - \hat{\tilde{y}}/n) \to -\infty$ and hence, $\Lambda(\tilde{x}, \tilde{y})$ is further simplified as

$$\Lambda(\tilde{x}, \tilde{y}) = -2\left(\frac{1}{\max(-\hat{\tilde{x}}/n - \hat{\tilde{y}}/n)}\right)\int_{\hat{\omega}=0}^{1} dH(\hat{\omega}) = \frac{-2}{\max(-\hat{\tilde{x}}/n - \hat{\tilde{y}}/n)}H(\hat{\omega}), \tag{13}$$

On the other hand, since $G_{pp}(\hat{\tilde{x}}, \hat{\tilde{y}}) = exp(-\Lambda(A))$, $\Lambda(A) = -\ln G_{pp}(\hat{\tilde{x}}, \hat{\tilde{y}})$, and therefore,

$$-\ln G_{pp}(\hat{\tilde{x}}, \hat{\tilde{y}}) = \frac{-2}{\max(-\hat{\tilde{x}}/n - \hat{\tilde{y}}/n)}H(\hat{\omega}), \tag{14}$$

$$\Rightarrow \hat{\omega} = H_{\hat{\omega}}^{-1}\left\{\frac{\max(-\hat{\tilde{x}}/n - \hat{\tilde{y}}/n)}{2}\ln G_{pp}(\hat{\tilde{x}}, \hat{\tilde{y}})\right\} \tag{15}$$

where $\hat{\omega} = \frac{-\hat{\tilde{x}}/n}{-\hat{\tilde{x}}/n - \hat{\tilde{y}}/n}$, the angular component, is a rational function defined only on [0, 1], and is fitted to the β-distribution [24] with parameters $\hat{p}$ and $\hat{q}$, i.e., $\hat{\omega} \sim \beta(\hat{\omega}; \hat{p}, \hat{q})$, and hence, $H_{\hat{\omega}}^{-1}(.)$ is the CDF-inverse of the β-distribution fitted to the angular component $\hat{\omega}$. ■

Since determining the inverse of β function is a non-trivial task, the Newton–Raphson method is utilized to determine the inverse of $\beta(\hat{\omega}; \hat{p}, \hat{q})$ [25]. Newton–Raphson method is a recursive approximation technique for finding the root and inverse of a differentiable function.

The optimum transmission rate based on the proposed rate selection framework is then determined by substituting the results of Theorem 1 into (11) as

$$R(G_{PP}) = \log_2\left(1 + H_{\hat{\omega}}^{-1}\left\{\frac{1}{2}\max(-\hat{\tilde{x}}/n - \hat{\tilde{y}}/n)\ln\varepsilon_n\right\}\right), \tag{16}$$

for the maximum allowed error probability ε_n, $\varepsilon_n \leq \epsilon$, where $H_{\hat{\omega}}^{-1}(.)$ denotes the CDF-inverse of β-distribution fitted to the angular component $\hat{\omega}$ with shape parameters $\hat{p}$ and $\hat{q}$, denoted by $\beta(\hat{\omega}; \hat{p}, \hat{q})$.

C. Outage Probability of the Transmission Rate

Suppose $R(G_{pp})$ is a valid transmission rate. In that case, the corresponding outage probability $p_F(R(G_{pp}))$ is expected not to exceed the target reliability ϵ, according to constraint (1). The average outage probability for the transmission rate obtained in the previous section is obtained by substituting (16) into the outage probability equation expressed in (2) as

$$\begin{aligned} p_F(R(G_{pp})) &= P\left[\log_2\left(1 + H_{\hat{\omega}}^{-1}\left(\frac{1}{2}\max(\hat{r})\ln\varepsilon_n\right)\right)\right. \\ &\quad \left.> \log_2(1+Z)\right] \\ &= H_Z\left(H_{\hat{\omega}}^{-1}\left(\frac{1}{2}\max(\hat{r})\ln\varepsilon_n\right)\right), \end{aligned} \tag{17}$$

where $\hat{r} = \frac{-\hat{\tilde{x}}}{n} + \frac{-\hat{\tilde{y}}}{n}$, $Z = \frac{-\tilde{x}}{N} + \frac{-\tilde{y}}{N}$, N is the total number of samples including training and test samples, and $H_Z(.)$ is the CDF of β-distribution with parameters p and q fitted to Z, i.e., $\beta(Z; p, q)$, assuming the channel is perfectly known. In order to determine the maximum allowed error probability ε_n, the constraint (1) should be satisfied as

$$\varepsilon_n = exp\left\{\frac{2}{\max(\hat{r})}H_{H_\epsilon^{-1}(p,q)}(\hat{p}, \hat{q})\right\}, \tag{18}$$

where $H_\epsilon^{-1}(p, q)$ denotes the ϵ-quantile of the β-distribution with parameters p and q fitted to Z, and $H_{H_\epsilon^{-1}(p,q)}(\hat{p}, \hat{q})$ is CDF of β-distribution with parameters $\hat{p}$ and $\hat{q}$ fitted to $\hat{\omega}$ and computed at point $H_\epsilon^{-1}(p, q)$.

D. Confidence Interval

The confidence interval for the transmission rate can be calculated by deriving confidence intervals for the estimated UGPD parameters. Although various methods can be used for computing confidence intervals, bootstrapping methods are mostly preferred due to their higher accuracy and avoidance of assumptions on the properties of the distribution of the original samples. Bootstrapping methods generate confidence intervals for statistical parameters by drawing pseudo-samples either from the original sample sequence or a model fitted to the original sample. In the realm of statistics, conventional bootstrapping-based methods are highly dependent on the asymptotic normality of parameter estimate to construct a standard confidence interval for estimation of the parameter $\hat{\theta}$. However, a standard confidence interval might not be a proper way of considering the estimation uncertainty for many applications, especially when the distribution of the parameter estimation, i.e., $\hat{\theta}$, is not normal. Alternatively, the bootstrap-based bias-corrected and accelerated (BCA) method has been demonstrated to provide higher accuracy with narrower CIs correcting (i) the non-normality of $\hat{\theta}$ by utilizing the bootstrap distribution where a parametric model assumption is made for $\hat{\theta}$, (ii) the bias of $\hat{\theta}$ by using the bias correction parameter, and (iii) the non-constant standard error of $\hat{\theta}$ by introducing the acceleration parameter [26], [27].

1) Fundamentals of Bootstrap-Based BCA: B bootstrap samples are generated at each bootstrapping round with replacement, and then an estimate of the parameter is obtained as $\hat{\theta}_b$ at the b-th round for $b \in \{1, 2, \ldots, B\}$. The BCA method is based on the assumption that there is a monotone transformation $\hat{\phi} = h(\hat{\theta})$ such that $\hat{\phi} \sim N(\phi - z_0^{\hat{\theta}}\tau_\phi, \tau_\phi^2)$, where the transformed variable $\hat{\phi}$ follows a normal distribution with mean $\phi - z_0^{\hat{\theta}}\tau_\phi$ and variance τ_ϕ^2, ϕ is the median of the distribution of the bootstrap estimates, $\tau_\phi = 1 + a^{\hat{\theta}}\phi$, and bias correction $z_0^{\hat{\theta}}$ and acceleration $a^{\hat{\theta}}$ factors are constant and estimated based on the CDF of normal distribution and jackknife re-sampling, respectively.

The bias correction factor $z_0^{\hat{\theta}}$ is estimated as

$$z_0^{\hat{\theta}} = \Phi^{-1}(\hat{F}_e(\hat{\theta})), \tag{19}$$

where Φ is the standard normal CDF and $\hat{F}_e(.)$ is the empirical probability of the estimated parameter $\hat{\theta}$ and calculated as

$$\hat{F}_e = P(\hat{\theta}_b < \hat{\theta}) = \frac{m}{B}, \quad b = 1, \ldots, B, \tag{20}$$

in which m is the number of estimated parameters of bootstrapping $\hat{\theta}_b$ smaller than the estimated parameter based on the original samples $\hat{\theta}$. Please note that $\hat{\theta}$ is the estimated parameter from the original samples while $\hat{\theta}_b$ is the estimated parameter from bootstrapping the original samples.

The acceleration factor $a^{\hat{\theta}}$ can be obtained based on jackknife re-sampling (or leave-one-out procedure) [27], [28], [29] by incorporating n replicates of the original samples, where n is the total number of the original samples. Jackknife resampling involves generating a series of jackknife replicates by repeatedly leaving out one observation from the original sample. Specifically, the first replicate is generated by excluding the first observation, the second replicate is generated by excluding the second observation, and so on until n replicates of size $n-1$ are obtained. For each sample number j, $j \in \{1, 2, \ldots, n\}$, the estimated parameter $\hat{\theta}^j$ is obtained. Then, the average of these estimations is calculated as $\bar{\theta} = \frac{1}{n}\sum_{j=1}^{n} \hat{\theta}^j$. The acceleration factor is computed as

$$a^{\hat{\theta}} = \frac{\sum_{j=1}^{n}[\hat{\theta}^j - \bar{\theta}]^3}{6\{\sum_{j=1}^{n}[\hat{\theta}^j - \bar{\theta}]^2\}^{\frac{3}{2}}}. \tag{21}$$

Finally, the confidence interval is constructed as

$$\hat{\theta} = [\theta_L, \theta_U], \tag{22}$$

where the estimated $\hat{\theta}_b$'s are ordered from smallest to the largest, $L = a_1^{\hat{\theta}} \times (B+1)$, and $U = a_2^{\hat{\theta}} \times (B+1)$ for $a_1^{\hat{\theta}}$ and $a_2^{\hat{\theta}}$ defined as

$$a_1^{\hat{\theta}} = \Phi\left\{ z_0^{\hat{\theta}} + \frac{z_0^{\hat{\theta}} - z_{1-\frac{\alpha}{2}}^{\hat{\theta}}}{1 - a^{\hat{\theta}}(z_0^{\hat{\theta}} - z_{1-\frac{\alpha}{2}}^{\hat{\theta}})} \right\}, \tag{23}$$

and

$$a_2^{\hat{\theta}} = \Phi\left\{ z_0^{\hat{\theta}} + \frac{z_0^{\hat{\theta}} + z_{1-\frac{\alpha}{2}}^{\hat{\theta}}}{1 - a^{\hat{\theta}}\left(z_0^{\hat{\theta}} + z_{1-\frac{\alpha}{2}}^{\hat{\theta}}\right)} \right\}, \tag{24}$$

where $z_{1-\frac{\alpha}{2}}^{\hat{\theta}}$ is the $100(1-\frac{\alpha}{2})$ percentile of the standard normal distribution for significance level (or probability of wrong decision) α [29].

2) BCA in CI Determination of UGPD Parameters: The estimated scale and shape parameters of UGPD are not exactly $\hat{\tilde{\sigma}}$, $\hat{\xi}$ and typically take values within the confidence intervals $[\hat{\tilde{\sigma}}_L, \hat{\tilde{\sigma}}_U]$ and $[\hat{\xi}_L, \hat{\xi}_U]$, respectively. The BCA-based algorithm that is utilized to determine the upper and lower bounds of the CI of the shape and scale parameters are given in Algorithm 1 and is described in detail as follows. The inputs of the algorithm are the observation sample set X^n; the significance level α, depending on the application and how much error is allowed in the parameter estimation; and the number of newly generated data sets in the bootstrapping process denoted by B. The algorithm starts by bootstrapping the original data set for B times and storing the new data sets in the vectors denoted by X_b^n, $b \in \{1, 2, \ldots B\}$ (Lines $1-2$). The UGPD is fitted to the tail of data set X_b^n, and the parameters of UGPD are estimated as $\hat{\tilde{\sigma}}_b$ and $\hat{\xi}_b$ for $b \in \{1, 2, \ldots B\}$ (Line 3). Upon estimating the UGPD parameters for each data set X_b^n, the bias correction factor is obtained

Algorithm 1: BCA-Based CI Determination Algorithm.

Input: $X^n = [x_1, x_2, \ldots, x_n]$, α, and B;
Output: $\tilde{\sigma}_L$, $\tilde{\sigma}_U$, ξ_L, and ξ_U;
1: **for** b = 1:B **do**
2: obtain bootstrap samples with replacement of X^n, as X_b^n;
3: estimate $\hat{\tilde{\sigma}}_b$ and $\hat{\xi}_b$, by fitting the UGPD to the tail samples of X_b^n;
4: **end for**
5: determine the bias correction factors $z_0^{\hat{\tilde{\sigma}}}$ and $z_0^{\hat{\xi}}$
6: **for** $j = 1 : n-1$ **do**
7: leave out the first j samples of X^n, as X^{n-j};
8: estimate $\hat{\tilde{\sigma}}^j$ and $\hat{\xi}^j$, by fitting UGPD to X^{n-j};
9: **end for**
10: determine acceleration factors $a^{\hat{\tilde{\sigma}}}$ and $a^{\hat{\xi}}$;
11: determine coefficients $a_1^{\hat{\tilde{\sigma}}}$, $a_2^{\hat{\tilde{\sigma}}}$, $a_1^{\hat{\xi}}$, and $a_2^{\hat{\xi}}$;
12: determine the CIs of the $(1-\alpha)$ BCA-CI, $[\tilde{\sigma}_L, \tilde{\sigma}_U]$ and $[\xi_L, \xi_U]$, for the UGPD parameters.
13: **return** $\tilde{\sigma}_L$, $\tilde{\sigma}_U$, ξ_L, and ξ_U;

based on (19) for the estimated scale and shape parameters of UGPD as $z_0^{\hat{\tilde{\sigma}}}$ and $z_0^{\hat{\xi}}$, respectively (Line 5). The corresponding acceleration factors of the scale and shape parameters (Line 10) based on (21) are determined by first applying the jackknife method and obtaining $n-1$ replica of X^n by leaving out the first j samples of X^n, denoted by X^{n-j} (Line $6-7$) and then, estimating the scale and shape parameters of fitted UGPD to each data set X^{n-j} as $\hat{\tilde{\sigma}}^j$ and $\hat{\xi}^j$, respectively (Line 8). Next, the $\{(1-\alpha)\times 100\}\%$ confidence intervals of the scale and shape parameters, denoted by $[\tilde{\sigma}_L, \tilde{\sigma}_U]$ and $[\xi_L, \xi_U]$, respectively, are estimated for the significance level α based on (22) (Line 12), through the determination of coefficients $a_1^{\hat{\tilde{\sigma}}}$, $a_2^{\hat{\tilde{\sigma}}}$, $a_1^{\hat{\xi}}$, and $a_2^{\hat{\xi}}$ according to (23) and (24) for the scale and shape parameters (Line 11). The same procedure is applied to the sample set Y^n to compute the corresponding CI of the estimated parameters of UGPD fitted to Y^n.

3) BCA in CI Determination of the Transmission Rate: Upon determining the upper and lower bounds of the estimated CI corresponding to the scale and shape parameters of X^n and Y^n, the $\tilde{x}$, $\tilde{y}$, $R(G_{pp})$, are estimated within the intervals $[\tilde{x}_L, \tilde{x}_U]$, $[\tilde{y}_L, \tilde{y}_U]$, $[R_L, R_U]$, respectively. The estimated transmission rate, $R(G_{pp})$, is a function of the pseudo-polar angular component $\hat{\omega}$, or equivalently Fréchet transformations $\hat{\tilde{x}}$ and $\hat{\tilde{y}}$. The CI of the angular component and the Fréchet transformed variables are functions of the confidence intervals of the scale and shape parameters. Therefore, the transmission rate $R(G_{pp})$ is estimated to be within the interval $[R_L, R_U]$, considering the confidence intervals of the estimated Pareto parameters.

Referring to (6) and (7), the upper (lower) bound of the scale parameters $\hat{\sigma}_x$ and $\hat{\sigma}_y$ along with the lower (upper) bound of the shape parameter $\hat{\xi}_x$ and $\hat{\xi}_y$ provide the lower (upper) bound of the Fréchet transformation parameters $\hat{\tilde{x}}$ and $\hat{\tilde{y}}$, respectively. Since the transmission rate obtained in Section III-B is a function of the CDF of angular component

$\hat{\omega}$, and

$$\hat{\omega} = \frac{-\hat{\tilde{x}}/n}{-\hat{\tilde{x}}/n - \hat{\tilde{y}}/n},$$

the upper (lower) bound of the $\hat{\tilde{y}}$ in conjunction with the lower (upper) bound of $\hat{\tilde{x}}$, results in the lower (upper) bound of $\hat{\omega}$. Moreover, since the CDF and CDF-inverse of $\hat{\omega}$ are monotonically increasing functions, the lower (upper) bound of $\hat{\omega}$ affects the lower (upper) bound of the estimated transmission rate $R(G_{pp})$ as follows:

$$R(G_{PP})_{L/U} = \log_2\left(1 + H^{-1}_{\hat{\omega}_{L/U}}\left\{\frac{1}{2}\max(-\hat{\tilde{x}}/n - \hat{\tilde{y}}/n)\ln\varepsilon_n\right\}\right), \tag{25}$$

where $R(G_{PP})_{L/U}$ and $H^{-1}_{\hat{\omega}_{L/U}}$ denotes the lower(L)/upper(U) bound of the estimated transmission rate $R(G_{PP})$ and CDF-inverse of the estimated Pickands angular component $H^{-1}_{\hat{\omega}}$, respectively.

It is worth noting that although $R(G_{pp})$ is a function of $\max(-\hat{\tilde{x}}/n - \hat{\tilde{y}}/n)$ and ε_n, these parameters do not affect the CI of the transmission rate for the following reasons: For the first parameter, the max function included in the ε_n formula in (18) cancels the effect of $\max(-\hat{\tilde{x}}/n - \hat{\tilde{y}}/n)$; for the second parameter, ε_n approaches a fixed value equal to the target error probability ϵ for large sample numbers.

The complexity of the proposed MEVT-based rate selection framework is $O(n\,N_{Tx}\,N_{Rx})$, where N_{Tx} and N_{Rx} are the number of transmitters and receivers, respectively, and n is the number of the training samples for individual channel sequences.

IV. Numerical Results

This section aims to evaluate the proposed rate selection framework compared to the traditional extrapolation-based approaches for a MIMO-URLLC system. In the following, we first provide the benchmark algorithm based on the traditional extrapolation-based approaches and simulation platform in Sections IV-A and IV-B, respectively. Then, in Section IV-C, we evaluate the performance of the proposed MEVT-based rate selection framework in determining the maximum transmission rate for URLLC in spatial diversity, evaluating the system reliability by using the outage probability metric and comparing it to the traditional extrapolation-based method. Finally, in Section IV-D, the estimation of the rate confidence interval is obtained by computing the confidence interval for the parameters of BGPD fitted to the tail distribution of the channel data, iterating over multiple sample sizes and values of the significance level α.

A. Benchmark Extrapolation-Based Algorithm

In the conventional extrapolation-based method, the distribution of the current channel data is estimated for the reliability order ranging from 10^{-3} to 10^{0} PER, based on individual channel data sequences [30], [31]. Subsequently, the tail distributions of these sequences are estimated by extrapolating the obtained results towards the ultra-reliable region of 10^{-9} to 10^{-5} PER [5]. After fitting different distributions to the samples in each observation sequence, the optimal distribution identified in our study is the Gaussian distribution, based on the Akaike information criterion/Bayesian information criterion (AIC/BIC) metric [13]. In order to estimate the tail distribution of joint probability and establish a fair comparison between our proposed approach with the extrapolation-based method, we apply the Fréchet transformation, and then the Pseudo-polar Pickands transformation to the extrapolated results [32]. Accordingly, the probability measure function of the corresponding angular component of Pickands transformation is modeled using the Beta distribution. Finally, the optimum transmission rate associated with the targeted error probability ϵ is modeled as

$$R(F_{ep}) = \log_2\left(1 + H^{-1}_{ep}\left\{\frac{1}{2}\max(-\hat{\tilde{x}}_{ep}/n - \hat{\tilde{y}}_{ep}/n)\ln\varepsilon^{\dagger}_n\right\}\right), \tag{26}$$

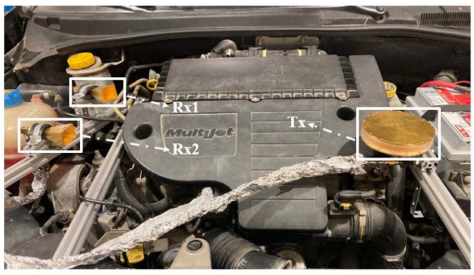
where F_{ep} is the extrapolated CDF, $\hat{\tilde{x}}_{ep}, \hat{\tilde{y}}_{ep}$ are the corresponding Fréchet variables of extrapolated results, H_{ep} is the probability measure function of the angular component of extrapolated results, and $\varepsilon^{\dagger}_n$ is the maximum allowed error probability for the extrapolation-based method calculated as

$$\varepsilon^{\dagger}_n = exp\left\{\frac{2}{\max(\hat{r}_{ep})}\, H^{ep}_{H^{-1}_{\epsilon,ep}(p_{ep},q_{ep})}(\hat{p}_{ep},\hat{q}_{ep})\right\}, \tag{27}$$

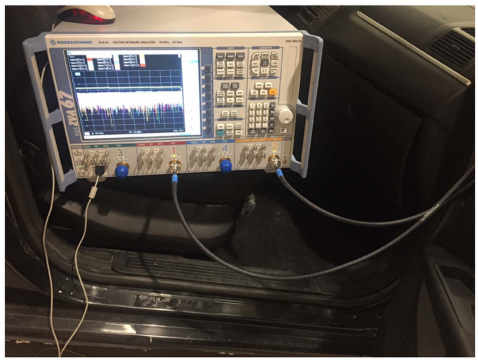
where $\hat{r}_{ep} = \hat{\tilde{x}}_{ep} + \hat{\tilde{y}}_{ep}$, $H^{-1}_{\epsilon,ep}(p_{ep},q_{ep})$ denotes the ϵ-quantile of the β distribution with parameters p_{ep} and q_{ep} fitted to the probability measure function of the angular component for the extrapolated results, and $H^{ep}_{H^{-1}_{\epsilon,ep}(p_{ep},q_{ep})}(\hat{p}_{ep},\hat{q}_{ep})$ is CDF of β distribution with parameters $\hat{p}_{ep}$ and $\hat{q}_{ep}$ computed at point $H^{-1}_{\epsilon,ep}(p_{ep},q_{ep})$.

B. Measurement Setup and Simulation Platform

The proposed methodology is implemented on the dataset obtained from the engine compartment of a Fiat Linea vehicle, as depicted in Fig. 2(a), utilizing a Vector Network Analyzer (VNA) (R & S ZVA67), as shown in Fig. 2(b). The VNA is connected to the transmitter and receivers, Rx1 and Rx2, via the R & S ZV-Z196 and PE361 port cables, respectively, which have a length of 610 mm. The transmitter is an omnidirectional antenna operating in the frequency range of 58 GHz to 63 GHz, with a nominal gain of 0 dBi. The receivers are horn antennas operating in the frequency range of 50–75 GHz, with a nominal gain of 24 dBi, and horizontal and vertical half power beamwidths of 11° and 9.5°, respectively. The locations of the transmitter and receiver antennas are selected out of the possible locations for the wireless sensors located within the engine compartment, namely locations 13 and 2&4 in [33]-Fig. 1, respectively, and shown in Fig. 2(a). The antennas are interconnected with the coaxial cables via a waveguide that operates within the frequency range of 50–65 GHz. The waveguide exhibits an insertion loss of 0.5 dB and impedance of 50Ω. It is worth noting that an omnidirectional antenna at the transmitter along with the directional antenna at two receivers are utilized to consider the usage of diversities in an ultra-reliable communication system and study



(a)



(b)

Fig. 2. Measurement setup with the transmitter (TX) and receiver (RX) antennas located in the engine compartment of Fiat Linea: (a) Engine compartment, and (b) VNA setup.

the inter-relationships of extreme events for a MIMO system under an intra-vehicular communication. MATLAB is used to implement the proposed framework on the collected data.

We categorize the samples from receivers Rx1 and Rx2 into two stationary groups: The first group, Gr_1, corresponds to driving on the smooth road, and the second group, Gr_2, corresponds to driving on the ramp. The optimum thresholds for the received power samples from receivers Rx1 and Rx2 are determined for each group of stationary data sequences based on the optimum threshold determination process in Section III-A. Accordingly, for the stationary group 1, the optimum thresholds of Rx1 and Rx2 are -7 dBm and -6.8 dBm, respectively, and for the stationary group 2, the optimum thresholds of Rx1 and Rx2 are -8 dBm and -27 dBm, respectively. In the following, we only present the results of Gr_2 due to the limited number of pages. However, a similar observation has been made in the sample data of Gr_1.

C. Transmission Rate

Fig. 3 shows the transmission rate of BGPD fitted to the filtered i.i.d. samples of the group 1 at different sample numbers, and error probabilities $\epsilon \in \{10^{-3}, 10^{-4}, 10^{-5}\}$ compared to the extrapolation-based results. To address the same target error probability ϵ, a higher value of optimum transmission rate is obtained based on the proposed MEVT-based framework, up to 10^3 more than the extrapolation-based benchmark algorithm. The gap between their rates increases as the error probability increases. Additionally, according to the results obtained based on the extrapolation method at $\epsilon = 10^{-3}$, the transmission rate

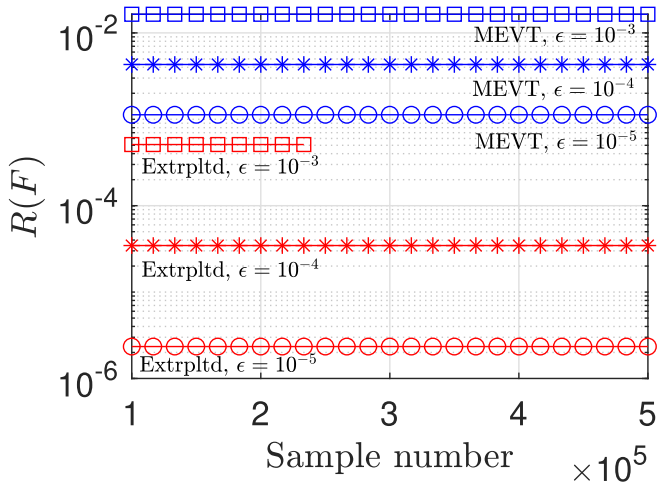


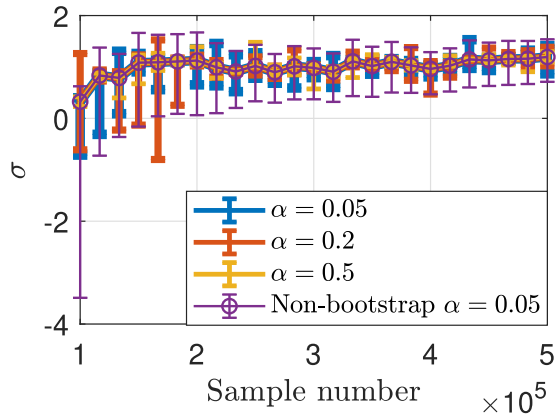
Fig. 3. The transmission rate of BGPD fitted to the filtered i.i.d. samples of the group 1 at different sample numbers, and targeted PER $\epsilon \in \{10^{-5}, 10^{-4}, 10^{-3}\}$.

drops significantly after some sample number since $\varepsilon_n^\dagger$ approaches 0, meaning that the maximum allowed error probability $\varepsilon_n^\dagger$ is estimated 0 instead of $\sim 10^{-3}$ which is the target error probability. Besides, the estimated transmission rate is not a function of the sample number for a large number of samples and varies only as a function of the target error probability ϵ. These results are consistent with the results of the EVT-based rate selection framework for single-input single-output (SISO)-URC presented in [16] and the extrapolation-based rate selection framework in [6].

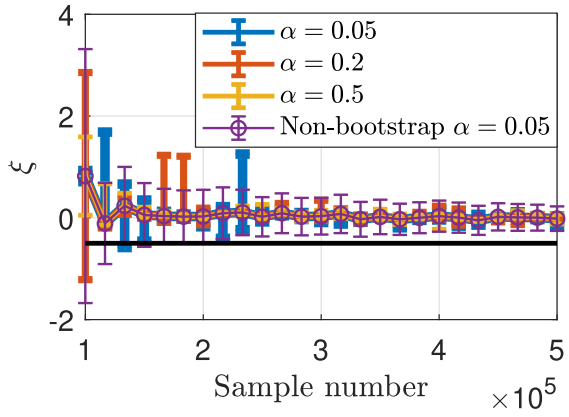
The outage probability of the estimated transmission rate for $R(G_{pp})$ at different target error probability ϵ has also been computed. The outage probability results are the same for the proposed framework and the traditional extrapolation-based method. However, to experience the same outage probability, the traditional method requires a lower transmission rate which is neither power nor time-efficient. Therefore, by applying the proposed MEVT-based rate selection framework, we achieve a higher transmission rate while addressing the ultra-reliable constraint in which the outage probability does not exceed the target error probability.

D. Confidence Interval

Fig. 4 shows the estimated UGPD parameters along with the lower and upper bounds of CIs for $\alpha = \{0.05, 0.2, 0.5\}$ and different sample numbers in both receivers Rx1 and Rx2 for stationary group 1. Compared to the standard CI results presented in [20], the confidence interval obtained by the non-bootstrap method is wider in general above a certain number of samples, and the estimated GPD parameters are almost constant. Moreover, it is observed that the MLEs of the scale (σ) and shape (ξ) parameters are subject to significant uncertainty when the sample size is limited. Nonetheless, the level of uncertainty reduces progressively with an increase in the sample size. This is consistent with the findings from the standard CI results as reported in [20]. This underscores the trade-off between the



(a)



(b)

Fig. 4. The estimated Pareto parameters along with their CI considering $\alpha = 0.05, 0.2, 0.5$ for stationary group 1, receivers Rx1 and Rx2, at different sample numbers: (a) Scale parameter and the corresponding CI for Rx1, and (b) Shape parameter and the corresponding CI for Rx1; the horizontal black line refer to the -0.5 minimum acceptable value for the shape parameter of UGPD.

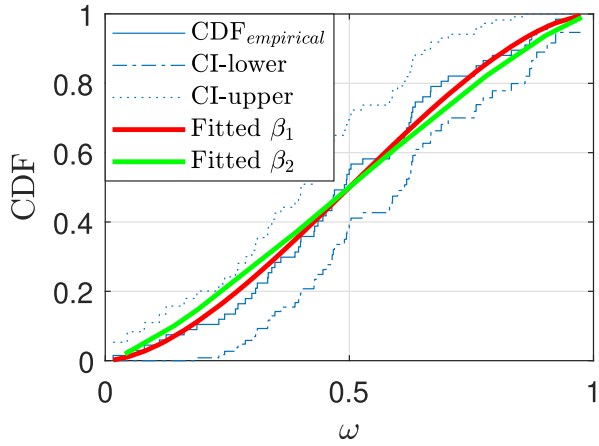


Fig. 5. The CDF of pseudo-polar Pickands angular component based on empirical results along with 0.01 confidence interval, the Beta distribution β_1 fitted to the sample numbers $\approx 1.3 \times 10^{-5}$, and the Beta distribution β_2 fitted to the sample number $\approx 5 \times 10^{-5}$.

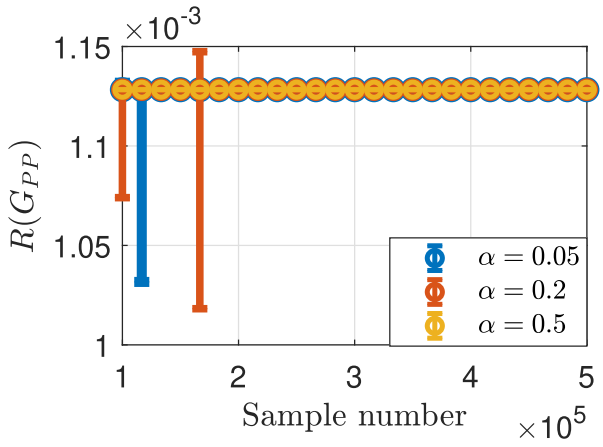


Fig. 6. Estimated transmission rate with the corresponding CI at $\alpha = 0.05, 0.2, 0.5$ at different sample numbers for targeted error probability of 10^{-5}.

uncertainty of the GPD parameters and the expenses involved in gathering data. In contrast to the conventional non-parametric technique for rate selection, which mandates noticeably more training on the channel, as outlined in [6], our framework based on MEVT appropriately infers the rate with a certain level of assurance, i.e., CI, using a smaller sample size of around $1/\epsilon$.

Fig. 5 illustrates the empirical CDF of the angular component of Pickands transformation along with the 0.01 confidence interval (the dotted plots), compared to the estimated Beta distribution fitted to the probability measure function of the Pickands' angular coordinate for two different sample numbers. β_1 and β_2 correspond to the fitted β-distribution for the sample numbers $n_1 \approx 1.3 \times 10^5$ and $n_2 \approx 5 \times 10^5$, respectively, where the estimation of UGPD parameters and the corresponding angular component ω become stable. It is observed that both β_1 and β_2 distributions fit the empirical results within the $\alpha = 0.01$ confidence interval. Consequently, utilizing MEVT for modeling the multivariate channel and determining the associated transmission rate can significantly reduce the required number of samples by a minimum of 10-fold compared to traditional extrapolation and data-driven models.

Fig. 6 depicts the estimated transmission rate based on the proposed MEVT-base rate selection framework for different sample numbers with the corresponding CI for $\alpha = \{0.05, 0.2, 0.5\}$ and target error probabilities $\epsilon \in \{10^{-5}, 10^{-4}, 10^{-3}\}$. The estimated transmission rate is constant with respect to the sample size, the same as what we have observed in Fig. 3. The confidence interval is also constant with respect to the sample size except for low sample numbers, where the estimation of UGPD parameters suffers from uncertainty and varies within a wider confidence interval. However, above a certain sample size, the estimation of CIs becomes stable and converges. These observations are in good agreement with what we observed for the SISO system illustrated in [20].

V. Conclusion

In this paper, we propose a novel framework based on the multivariate extreme value theory with the goal of determining an optimum transmission rate for a system operating in MIMO-URC, and incorporating a confidence interval for the estimated BGPD parameters and the corresponding optimum

transmission rate to enable the usage of MEVT in real-time communication. The framework includes the modeling of the inter-relationship of the tail distributions of the multiple channels by using the BGPD based on the Poisson point process approach, determination of the optimum transmission rate by using the estimated BGPD model, incorporation of the confidence intervals for the estimated transmission rate due to the availability of the restricted amount of data, and then assessment of the system reliability by utilizing the outage probability metric. The proposed MEVT-based rate selection framework achieves a higher transmission rate, up to about 10^3, than the traditional methods based on the extrapolation of average statistic channel models. Incorporating CI into the rate selection estimate can lower the required number of samples, at least by 10-fold, during the training phase necessary to achieve a specific level of reliability. This reduction in required samples subsequently decreases the sample complexity and contributes to a more efficient rate estimation. It has been demonstrated that calculating GPD parameters using limited samples leads to a wider confidence interval in rate estimation, implying a greater degree of uncertainty. Nevertheless, as additional data become accessible, the confidence limits become narrower because of the diminishing level of uncertainty. Moreover, based on the proposed methodology, we demonstrate the trade-off between the uncertainty in the rate selection and the cost of data collection. Since a large amount of data is required for deriving the statistics, our future endeavors encompass the expansion of our research through the implementation of transfer learning techniques, specifically employing knowledge-assisted training. This approach entails the development of a digital twin model that accurately represents a real-time network, incorporating essential elements such as network topology, channel characteristics, and queueing models for offline training purposes. Subsequently, this model is refined and optimized using a smaller dataset in the live operating environment.

REFERENCES

[1] P. Popovski et al., "Wireless access for ultra-reliable low-latency communication: Principles and building blocks," *IEEE Netw.*, vol. 32, no. 2, pp. 16–23, Mar./Apr. 2018.

[2] C. Gustafson, K. Haneda, S. Wyne, and F. Tufvesson, "On mm-wave multipath clustering and channel modeling," *IEEE Trans. Antennas Propag.*, vol. 62, no. 3, pp. 1445–1455, Mar. 2014.

[3] P. Popovski et al., "Wireless access in ultra-reliable low-latency communication (URLLC)," *IEEE Trans. Commun.*, vol. 67, no. 8, pp. 5783–5801, Aug. 2019.

[4] M. Bennis, M. Debbah, and H. V. Poor, "Ultra reliable and low-latency wireless communication: Tail, risk, and scale," *Proc. IEEE*, vol. 106, no. 10, pp. 1834–1853, Oct. 2018.

[5] P. C. F. Eggers, M. Angjelichinoski, and P. Popovski, "Wireless channel modeling perspectives for ultra-reliable communications," *IEEE Trans. Wireless Commun.*, vol. 18, no. 4, pp. 2229–2243, Apr. 2019.

[6] M. Angjelichinoski, K. F. Trillingsgaard, and P. Popovski, "A statistical learning approach to ultra-reliable low latency communication," *IEEE Trans. Commun.*, vol. 67, no. 7, pp. 5153–5166, Jul. 2019.

[7] M. Serror, C. Dombrowski, K. Wehrle, and J. Gross, "Channel coding versus cooperative ARQ: Reducing outage probability in ultra-low latency wireless communications," in *Proc. IEEE Globecom Workshops*, 2015, pp. 1–6.

[8] G. Pocovi, B. Soret, M. Lauridsen, K. I. Pedersen, and P. Mogensen, "Signal quality outage analysis for ultra-reliable communications in cellular networks," in *Proc. IEEE Globecom Workshops*, 2015, pp. 1–6.

[9] V. N. Swamy, P. Rigge, G. Ranade, B. Nikolić, and A. Sahai, "Wireless channel dynamics and robustness for ultra-reliable low-latency communications," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 4, pp. 705–720, Apr. 2019.

[10] J. J. Nielsen, R. Liu, and P. Popovski, "Ultra-reliable low latency communication using interface diversity," *IEEE Trans. Commun.*, vol. 66, no. 3, pp. 1322–1334, Mar. 2018.

[11] S. R. Khosravirad, H. Viswanathan, and W. Yu, "Exploiting diversity for ultra-reliable and low-latency wireless control," *IEEE Trans. Wireless Commun.*, vol. 20, no. 1, pp. 316–331, Jan. 2021.

[12] S. Samarakoon, M. Bennis, W. Saad, and M. Debbah, "Predictive ultra-reliable communication: A survival analysis perspective," *IEEE Commun. Lett.*, vol. 25, no. 4, pp. 1221–1225, Apr. 2021.

[13] N. Mehrnia and S. Coleri, "Multivariate extreme value theory based channel modeling for ultra-reliable communications," *IEEE Trans. Wireless Commun.*, vol. 23, no. 5, pp. 4964–4975, May 2024.

[14] V. N. Swamy, P. Rigge, G. Ranade, B. Nikolic, and A. Sahai, "Predicting wireless channels for ultra-reliable low-latency communications," in *Proc. IEEE Int. Symp. Inf. Theory*, 2018, pp. 2609–2613.

[15] F. Pase, M. Giordani, and M. Zorzi, "On the convergence time of federated learning over wireless networks under imperfect CSI," in *Proc. Int. Conf. Commun. Workshops*, 2021, pp. 1–7, doi: 10.1109/ICCWorkshops50388.2021.9473615.

[16] N. Mehrnia and S. Coleri, "Extreme value theory based rate selection for ultra-reliable communications," *IEEE Trans. Veh. Technol.*, vol. 71, no. 6, pp. 6727–6731, Jun. 2022.

[17] N. Mehrnia and S. Coleri, "Wireless channel modeling based on extreme value theory for ultra-reliable communications," *IEEE Trans. Wireless Commun.*, vol. 21, no. 2, pp. 1064–1076, Feb. 2022.

[18] N. Mehrnia and S. Coleri, "Non-stationary wireless channel modeling approach based on extreme value theory for ultra-reliable communications," *IEEE Trans. Veh. Technol.*, vol. 70, no. 8, pp. 8264–8268, Aug. 2021.

[19] W. Zhang, M. Derakhshani, and S. Lambotharan, "Non-parametric statistical learning for URLLC transmission rate control," in *Proc. IEEE Int. Conf. Commun.*, 2021, pp. 1–6.

[20] N. Mehrnia and S. Coleri, "Incorporation of confidence interval into rate selection based on the extreme value theory for ultra-reliable communications," in *Proc. IEEE Eur. Conf. Netw. Commun. (EuCNC) 6G Summit*, 2022, pp. 118–123.

[21] S. Coles, J. Bawa, L. Trenner, and P. Dorazio, *An Introduction to Statistical Modeling of Extreme Values*, vol. 208. Berlin, Germany: Springer, 2001.

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

[23] J. Shi, D. Anzai, and J. Wang, "Channel modeling and performance analysis of diversity reception for implant UWB wireless link," *IEICE Trans. Commun.*, vol. 95, no. 10, pp. 3197–3205, Oct. 2012.

[24] D. Cooley, R. A. Davis, and P. Naveau, "The pairwise Beta distribution: A flexible parametric multivariate model for extremes," *J. Multivariate Anal.*, vol. 101, no. 9, pp. 2103–2117, 2010.

[25] A. Gil, J. Segura, and N. M. Temme, *Numerical Methods for Special Functions*. Philadelphia, PA, USA: SIAM, 2007.

[26] M.-T. Puth, M. Neuhäuser, and G. D. Ruxton, "On the variety of methods for calculating confidence intervals by bootstrapping," *J. Animal Ecol.*, vol. 84, no. 4, pp. 892–897, 2015.

[27] B. Grün and T. Miljkovic, "The automated bias-corrected and accelerated bootstrap confidence intervals for risk measures," *North Amer. Actuarial J.*, vol. 27, no. 4, pp. 731–750, 2023. [Online]. Available: https://doi.org/10.1080/10920277.2022.2141781

[28] M. H. Quenouille, "Approximate tests of correlation in time-series 3," in *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 45, Cambridge, MA, USA: Cambridge Univ. Press, 1949, pp. 483–484.

[29] R. J. Tibshirani and B. Efron, "An introduction to the bootstrap," *Monographs Statist. Appl. Probability*, vol. 57, pp. 1–436, 1993.

[30] C. U. Bas and S. Coleri Ergen, "Ultra-wideband channel model for intra-vehicular wireless sensor networks beneath the chassis: From statistical model to simulations," *IEEE Trans. Veh. Technol.*, vol. 62, no. 1, pp. 14–25, Jan. 2012.

[31] U. Demir, C. U. Bas, and S. Coleri Ergen, "Engine compartment UWB channel model for intra-vehicular wireless sensor networks," *IEEE Trans. Veh. Technol.*, vol. 63, no. 6, pp. 2497–2505, Jun. 2013.

[32] P. Bortot, S. Coles, and J. Tawn, "The multivariate gaussian tail model: An application to oceanographic data," *J. Roy. Stat. Society, Ser. C. (Appl. Statistics)*, vol. 49, no. 1, pp. 31–049, 2000.

[33] U. Demir, C. U. Bas, and S. C. Ergen, "Engine compartment UWB channel model for intravehicular wireless sensor networks," *IEEE Trans. Veh. Technol.*, vol. 63, no. 6, pp. 2497–2505, Jun. 2013.



Niloofar Mehrnia (Graduate Student Member, IEEE) received the B.S. degree in biomedical engineering from Isfahan University, Isfahan, Iran, in 2014, the M.S. degree in electrical and electronics engineering from Istanbul Sehir University, Istanbul, Türkiye, in 2017, and the Ph.D. degree in electrical and electronics engineering from Koc University, Istanbul, in 2022. From 2019 to 2023, she was a Researcher with Koc University Ford Otosan Automotive Technologies Laboratory. Since 2024, she has been a Postdoctoral Researcher with KTH Royal Institute of Technology, Stockholm, Sweden. Her research interests includes wireless channel modeling, ultra-reliable and low-latency communications (URLLC), extreme value theory (EVT)-based channel tail modeling for URLLC, and vehicular communications.



Sinem Coleri (Fellow, IEEE) received the B.S. degree in electrical and electronics engineering from Bilkent University, Ankara, Türkiye, in 2000, and the M.S. and Ph.D. degrees in electrical engineering and computer sciences from the University of California Berkeley, Berkeley, CA, USA, in 2002 and 2005, respectively. She was a Research Scientist with Wireless Sensor Networks Berkeley Lab under sponsorship of Pirelli and Telecom Italia from 2006 to 2009. Since 2009, she has been a Faculty Member with the Department of Electrical and Electronics Engineering, Koc University, Istanbul, Türkiye. She is currently a Professor and the Chair with the Department of Electrical and Electronics Engineering, Koc University. She is also the founding Director of Wireless Networks Laboratory (WNL) and Director of Ford Otosan Automotive Technologies Laboratory. Her research interests include 6G wireless communications and networking, machine learning for wireless networks, machine-to-machine communications, wireless networked control systems, and vehicular networks. She was the recipient of numerous awards and recognitions, such as the N2Women: Stars in Computer Networking and Communications, TUBITAK (The Scientific and Technological Research Council of Turkey) Incentive Award and IEEE Vehicular Technology Society Neal Shepherd Memorial Best Propagation Paper Award. She is the Editor-in-Chief of IEEE OPEN JOURNAL OF THE COMMUNICATIONS SOCIETY. She is an IEEE ComSoc Distinguished Lecturer.