Performance Analysis of MIMO CR Multihop Relaying with Imperfect CSI in Short-Packet URLLCs

Ngo Hoang Tu[1], Hoang Trong Dai[2], and Kyungchun Lee[1,3]
[1]Department of Electrical and Information Engineering
[3]Research Center for Electrical and Information Technology
Seoul National University of Science and Technology, Seoul 01811, Republic of Korea
[2]School of Electrical and Data Engineering, University of Technology Sydney, Ultimo, NSW 2007, Australia
Email: ngohoangtu@seoultech.ac.kr, dai.t.hoang@student.uts.edu.au, kclee@seoultech.ac.kr

***Abstract*—This paper evaluates the multiple-input multiple-output underlay cognitive multihop relay networks with short-packet communications, where general and practical scenarios are considered with multiple primary users and imperfect channel state information of the interference channels. For performance evaluation, the closed-form expressions of the end-to-end (E2E) block error rate for the considered systems are derived under consideration of quasi-static Rayleigh fading channels and the finite-blocklength (FBL) regime, from which the E2E throughput, energy efficiency (EE), latency, and reliability are also studied. Based on the analytical results, we adopt a machine learning-aided estimator, i.e., extreme gradient boosting (XGB), to predict the E2E throughput, EE, latency, and reliability for real-time configurations. The XGB-based evaluation achieves equivalent performance while significantly reducing the execution time compared to conventional analytical and simulation methods, which makes XGB an efficient tool to estimate the system performance in real-time applications.**

***Index Terms*—Short-packet communication (SPC), multiple-input multiple-output (MIMO), underlay cognitive radio (CR), multihop relaying, and machine learning.**

I. Introduction

The spectral scarcity in fifth-generation (5G) and beyond networks has pushed the telecommunication systems to operate at a higher spectral efficiency (SE) [1]. To improve the SE, the underlay cognitive radio (CR) enables secondary users (SUs) to simultaneously leverage the licensed frequency band of the primary network without causing any harmful interference to the primary users (PUs) [2]. In the CR paradigm, the transmit power of the secondary transmitters must be limited, which leads to performance degradation of the secondary network. Herein, a multiple-input multiple-output (MIMO) multihop relay network along with CR has been proposed to combat the spectrum scarcity and limited transmit power problems, extend the network coverage, and improve reliability [3], [4]. However, these traditional MIMO CR multihop relay networks may not satisfy the rigorous conditions of ultra-reliable and low-latency communications (uRLLCs) in 5G and beyond networks. Because the uRLLCs' stringent requirements for 5G and beyond require not only a very low latency of around $1-10$ ms, but also ultra-reliable with a block error rate (BLER) of less than 10^{-5} [5], short-packet communication (SPC), also known as finite-blocklength (FBL) communication, has been considered as an efficient enabling technology to support the uRLLCs [6], [7].

It is conceivable that taking the closed-form expressions of the performance analysis will no longer be sufficient when the complexity and heterogeneity of future wireless networks grow enormously. Numerical integration and simulation approaches can be utilized as alternative methods to evaluate the system performance, but suffer from long run times to exactly achieve the converged value. Recently, machine learning (ML) has been realized as a powerful tool to accurately evaluate system performance while dramatically reducing the execution time compared to conventional simulation methods. Because ML-based applications are able to accurately estimate non-linear functions with low complexity, they enable a wide variety of networks for real-time applications [8], [9].

In this paper, we first study MIMO CR multihop relay networks with multiple PUs for SPCs to satisfy uRLLC requirements, where the comprehensive analysis is conducted in a general and practical scenario with imperfect channel-state information (CSI). Subsequently, a new ML evaluation approach is designed to effectively predict the system performance. The main contributions of this paper are summarized as follows:

- We propose a MIMO underlay CR multihop relay network with multiple PUs using SPCs in accordance with uRLLC requirements, where the closed-form expressions of the end-to-end (E2E) BLER, effective throughput (ETP), energy efficiency (EE), latency, and reliability are obtained to evaluate the system performance.
- To develop real-time system configurations for the proposed networks, we design a new ML approach, i.e., extreme gradient boosting (XGB), to estimate the system performance in terms of ETP, EE, latency, and reliability.
- The numerical results show that the strengths of the proposed framework in accordance with uRLLC requirements are confirmed. In addition, our proposed ML-aided estimator exhibits highly accurate prediction with short execution times, making XGB the efficient estimator for future practical real-time applications.

This work was supported in part by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education under Grant NRF-2019R1A6A1A03032119 and in part by the NRF Grant funded by the Korean Government (MSIT) under Grant NRF-2022R1A2C1006566.(*Corresponding author: Kyungchun Lee*.)

II. Networks and Channel Model

We consider a MIMO underlay CR multihop network for SPCs, where Rayleigh fading channels are assumed. In the primary network, there coexist N PUs that share the same licensed frequency band, whereas in the secondary network, the signal is sent from the secondary source SU_0 equipped with N_T transmit antennas to the secondary destination SU_{K+1} equipped with N_R receive antennas via K secondary relays, named $SU_1, SU_2, ..., SU_K$, that are equipped with N_T transmit and N_R receive antennas. We assume that the direct link between SU_0 and SU_{K+1} is not supported and all secondary devices' antennas are equipped with the half-duplex mode. Let $h_k^{(i,j)}$ denote the fading coefficient for the channel between transmit antenna i and receive antenna j at hop k, whereas $g_{k,n}^{(i)}$ denotes the channel coefficient between transmit antenna i of SU_{k-1} and PU_n. Throughout this paper, $i = \overline{1, N_T}$, $j = \overline{1, N_R}$, $k = \overline{1, K+1}$, and $n = \overline{1, N}$ are assumed. We note that only a single transmit antenna of node SU_{k-1} is activated for transmission by the transmit antenna selection (TAS) scheme, and $g_{k,n}$ represents the channel between the activated transmit antenna of SU_{k-1} and PU_n.

In the underlay CR systems, the transmit power at each SU must be lower than the maximum tolerable interference power to avoid severe interference at the PUs. Therefore, the transmit power of SU_{k-1} can be expressed as $P_{k-1} = \min\left(P_s, \frac{\mathcal{I}_p}{\max\limits_{n=1,\dots,N}|g_{k,n}|^2}\right)$, where P_s is the maximum allowable transmit power per transmit station in the secondary network and $\mathcal{I}_p$ is the maximum allowable interference power for a PU. However, perfect CSI of the interference channels is not available in practice because of complex channel intricacies, e.g., feedback delay, limited feedback, and channel estimation errors. Henceforth, the estimate of the channel coefficient $g_{k,n}$ can be expressed as $\hat{g}_{k,n} = \beta g_{k,n} + \sqrt{1-\beta^2} g_{k,n}^{\text{err}}$, where $g_{k,n}^{\text{err}}$ is a circular symmetric complex Gaussian random variable with zero-mean and the same variance of $g_{k,n}$, whereas $\beta \in [0,1]$ represents the correlation coefficient between the real and estimated channels, which is utilized to measure the CSI condition. As a result, the transmit power of SU_{k-1} with imperfect CSI is expressed as $\hat{P}_{k-1} = \min\left(P_s, \frac{\mathcal{I}_p}{\max\limits_{n=1,\dots,N}|\hat{g}_{k,n}|^2}\right)$. We note that a higher value of β implies a lower channel estimation error. Accordingly, the instantaneous received SNR at SU_k with imperfect CSI for the channel between the ith transmit antenna and the jth receive antenna is determined by

$$\gamma_k^{(i,j)} = \frac{\hat{P}_{k-1}}{\mathcal{N}_0}\left|h_k^{(i,j)}\right|^2 = \min\left(\bar{\gamma}_P, \frac{\bar{\gamma}_I}{\max\limits_{n=1,\dots,N}|\hat{g}_{k,n}|^2}\right)\left|h_k^{(i,j)}\right|^2, \quad (1)$$

where $\bar{\gamma}_P = P_s/\mathcal{N}_0$, $\bar{\gamma}_I = \mathcal{I}_p/\mathcal{N}_0$, and $\mathcal{N}_0$ is the noise variance. For the Rayleigh fading channel, $\left|h_k^{(i,j)}\right|^2$ and $|g_{k,n}|^2$ follow an exponential distribution with characteristic parameters $\lambda_k^{(i,j)}$ and $\lambda_{k,n}$. We assume the channel coefficients of each hop and interference link are independent and identically distributed, i.e., $\lambda_k^{(i,j)} = \lambda_k$, $\lambda_{k,n} = \lambda_{kp}$, $\forall i, j, n$.

For the transmission schemes, the TAS technique is employed at the transmit side to achieve transmit diversity, power consumption reduction, and hardware cost reduction [10]. For TAS, only a single optimal antenna out of N_T providing the highest received SNR is selected to transmit the signal. Meanwhile, either maximum ratio combining (MRC) or selection combining (SC) is utilized at the receive side to achieve receive diversity.

According to the MRC principle, the received signals from all branch channels are coherently combined [11]. Consequently, the instantaneous output SNR at hop k of the scheme TAS/MRC is determined as

$$\gamma_k^{\text{TAS/MRC}} = \max_{1\le i\le N_T}\sum\nolimits_{j=1}^{N_R}\gamma_k^{(i,j)}. \quad (2)$$

Proposition 1. *By denoting* $\mathcal{X}_k = \sum\limits_{j=1}^{N_R}\left|h_k^{(i,j)}\right|^2$ *and* $\mathcal{Y}_k = \max\limits_{1\le n\le N}|\hat{g}_{k,n}|^2$*, the cumulative distribution functions (CDFs) and probability density functions (PDFs) of* $\mathcal{X}_k$ *and* $\mathcal{Y}_k$ *are given, respectively, by*

$$F_{\mathcal{X}_k}(x) = 1 - \exp(-\lambda_k x)\sum_{n=0}^{N_R-1}\frac{1}{n!}(\lambda_k x)^n, \quad (3)$$

$$f_{\mathcal{X}_k}(x) = \frac{\lambda_k^{N_R} x^{N_R-1}}{(N_R-1)!}\exp(-\lambda_k x), \quad (4)$$

$$F_{\mathcal{Y}_k}(x) = 1 - \sum_{i=0}^{N-1}\frac{\phi(N,i)}{2}\exp\left(-\hat{\lambda}_{kp}^{(i)}x\right), \quad (5)$$

$$f_{\mathcal{Y}_k}(x) = \sum_{i=0}^{N-1}\frac{\phi(N,i)}{2}\hat{\lambda}_{kp}^{(i)}\exp\left(-\hat{\lambda}_{kp}^{(i)}x\right), \quad (6)$$

where $\phi(N,i) = \binom{N-1}{i}\frac{2N(-1)^i}{i+1}$ *and* $\hat{\lambda}_{kp}^{(i)} = \frac{(i+1)\lambda_{kp}}{i+1-i\beta^2}$*. By substituting* (1) *into* (2) *and utilizing* (3)–(6)*, the CDF of* $\gamma_k^{\text{TAS/MRC}}$ *is obtained as*

$$F_{\gamma_k}^{\text{TAS/MRC}}(\gamma) = \left\{F_{\mathcal{Y}_k}\left(\frac{\bar{\gamma}_I}{\bar{\gamma}_P}\right)F_{\mathcal{X}_k}\left(\frac{\gamma}{\bar{\gamma}_P}\right) + \sum_{i=0}^{N-1}\frac{\phi(N,i)}{2}\cdot\exp\left(-\hat{\lambda}_{kp}^{(i)}\frac{\bar{\gamma}_I}{\bar{\gamma}_P}\right) - \sum_{i=0}^{N-1}\frac{\phi(N,i)}{2}\hat{\lambda}_{kp}^{(i)}\sum_{n=0}^{N_R-1}\frac{1}{n!}\left(\frac{\lambda_k\gamma}{\bar{\gamma}_I}\right)^n\cdot\left(\hat{\lambda}_{kp}^{(i)} + \frac{\lambda_k\gamma}{\bar{\gamma}_I}\right)^{-n-1}\Gamma\left(n+1, \frac{\hat{\lambda}_{kp}^{(i)}\bar{\gamma}_I + \lambda_k\gamma}{\bar{\gamma}_P}\right)\right\}^{N_T}, \quad (7)$$

where $\Gamma(\alpha, x) = \int_x^\infty e^{-t}t^{\alpha-1}dt$ *denotes the upper incomplete gamma function [12, Eq. (8.350.2)].*

Proof. Please see Appendix A. □

According to the SC principle, only the link with the highest average received SNR is chosen to perform the signal detection [13]. When utilizing both the TAS and SC schemes, the instantaneous output SNR can be expressed as

$$\gamma_k^{\text{TAS/SC}} = \max_{1\le i\le N_T, 1\le j\le N_R}\gamma_k^{(i,j)}. \quad (8)$$

Proposition 2. *Based on* (8)*, the CDF of* $\gamma_k^{\mathrm{TAS/SC}}$ *is obtained as*

$$F_{\gamma_k}^{\mathrm{TAS/SC}}(\gamma) = \left\{ F_{\mathcal{Y}_k}\left(\frac{\bar{\gamma}_I}{\bar{\gamma}_P}\right) F_{\left|h_k^{(i,j)}\right|^2}\left(\frac{\gamma}{\bar{\gamma}_P}\right) + \sum_{i=0}^{N-1} \frac{\phi(N,i)}{2} \exp\left(-\hat{\lambda}_{kp}^{(i)} \frac{\bar{\gamma}_I}{\bar{\gamma}_P}\right) - \sum_{i=0}^{N-1} \frac{\phi(N,i)}{2} \cdot \frac{\hat{\lambda}_{kp}^{(i)} \bar{\gamma}_I}{\hat{\lambda}_{kp}^{(i)} \bar{\gamma}_I + \lambda_k \gamma} \exp\left(-\frac{\hat{\lambda}_{kp}^{(i)} \bar{\gamma}_I + \lambda_k \gamma}{\bar{\gamma}_P}\right) \right\}^{N_T N_R}, \quad (9)$$

where $F_{\left|h_k^{(i,j)}\right|^2}(x) = 1 - \exp(-\lambda_k x)$.

Proof. The proof of (9) follows the same steps as **Proposition 1**. □

III. Performance Analysis

A. E2E BLER

We assume that SU_0 sends $\mathcal{B}$ information bits to SU_{K+1} with the assistance of $SU_1, ..., SU_K$ via quasi-static fading channels, where $\mathcal{B}$ information bits are encoded into a block of τ channel uses (CUs). Thus, the channel coding rate is given by $\mathcal{R} = \mathcal{B}/\tau$. Under the FBL regime with an SPC of $\tau > 100$ CUs [14], the average BLER at link k is given by

$$\bar{\varepsilon}_k^{\mathcal{S}} = \mathbb{E}\left\{ Q\left(\frac{\log_2\left(1+\gamma_k^{\mathcal{S}}\right) - \mathcal{B}/\tau}{\sqrt{\mathcal{V}\left(\gamma_k^{\mathcal{S}}\right)/\tau}} \right) \right\}, \quad (10)$$

where $\mathcal{S}$ denotes one type of the transmission scheme, i.e., $\mathcal{S} \in \{\mathrm{TAS/MRC}, \mathrm{TAS/SC}\}$; $\mathbb{E}(\cdot)$ represents the expectation operator, $\mathcal{V}\left(\gamma_k^{\mathcal{S}}\right) \triangleq \left(1 - \frac{1}{\left(1+\gamma_k^{\mathcal{S}}\right)^2}\right)(\log_2 e)^2$ denotes the channel dispersion; and $Q(\cdot)$ is the Gaussian Q-function. It is challenging to directly derive the closed-form expression of $\bar{\varepsilon}_k^{\mathcal{S}}$ in (10) because of the complicated Q-function. To tackle this problem, we utilize a tight approximation of the Q-function, as discussed in [14], which yields

$$\bar{\varepsilon}_k^{\mathcal{S}} = \vartheta\sqrt{\tau} \int_{\psi_L}^{\psi_H} F_{\gamma_k}^{\mathcal{S}}(\gamma)\, d\gamma, \quad (11)$$

where $\vartheta = \left[2\pi\left(2^{2\mathcal{B}/\tau} - 1\right)\right]^{-1/2}$, $\psi_L = \theta - 1/(2\vartheta\sqrt{\tau})$, $\psi_H = \theta + 1/(2\vartheta\sqrt{\tau})$, and $\theta = 2^{\mathcal{B}/\tau} - 1$.

It is noted that the integral in (11) is very strenuous to calculate due to the complexity of the output SNR's CDF. Motivated by this issue, we utilize two approximations to obtain a tightly bounded closed-form expression for the average BLER at each hop. These approximated frameworks can achieve not only very high accuracy, but also low complexity, as presented in the following propositions.

Proposition 3 (First-order Riemann integral approximation)**.** *The tightly bounded closed-form expression for the average BLER in the kth hop can be derived by using the first-order Riemann integral approximation [15]:* $\bar{\varepsilon}_k^{\mathcal{S},\mathrm{Rie}} = F_{\gamma_k}^{\mathcal{S}}(\theta)$.

Proof. When $\tau > 100$ CUs, it is observed that $\psi_H - \psi_L = \sqrt{2\pi\left(2^{2\mathcal{B}/\tau} - 1\right)/\tau}$ is very small. Therefore, it is valid to utilize the first-order Riemann integral approximation $\int_x^y f(z)\, dz \cong (y-x) f\left(\frac{x+y}{2}\right)$ for (11). As a result, (11) is approximated as $\bar{\varepsilon}_k^{\mathcal{S},\mathrm{Rie}} = \vartheta\sqrt{\tau}\left(\psi_H - \psi_L\right) F_{\gamma_k}^{\mathcal{S}}\left(\frac{\psi_H+\psi_L}{2}\right) \overset{(a)}{=} F_{\gamma_k}^{\mathcal{S}}(\theta)$, where step (a) is based on the observation that $\psi_H - \psi_L = 1/(\vartheta\sqrt{\tau})$ and $\psi_H + \psi_L = 2\theta$, which completes the proof. □

Proposition 4 (Gauss–Chebyshev quadrature)**.** *By using the Gauss–Chebyshev quadrature integral* $\int_x^y f(z)\, dz \cong \frac{(y-x)}{2} \sum_{u=1}^{T} \frac{\pi}{T}\sqrt{1-x_u^2} f\left(\frac{y-x}{2} x_u + \frac{y+x}{2}\right)$, *where* $x_u = \cos\left(\frac{(2u-1)\pi}{2T}\right)$ *and* T *denotes the number of tradeoff terms, i.e., the tradeoff parameter between the complexity and accuracy [16], the tightly bounded closed-form expression for the average BLER at link* k *can be obtained as* $\bar{\varepsilon}_k^{\mathcal{S},\mathrm{GCheb}} = \sum_{u=1}^{T} \frac{\pi}{2T}\sqrt{1-x_u^2} F_{\gamma_k}^{\mathcal{S}}\left(\frac{x_u}{2\vartheta\sqrt{\tau}} + \theta\right)$.

According to the selective decode-and-forward principle [17], the E2E BLER for the considered system is given by

$$\bar{\varepsilon}_{\mathrm{E2E}}^{\mathcal{S},\mathcal{Z}} = \bar{\varepsilon}_1^{\mathcal{S},\mathcal{Z}} + \sum_{k=1}^{K} \left\{ \bar{\varepsilon}_{k+1}^{\mathcal{S},\mathcal{Z}} \cdot \prod_{m=1}^{k} \left(1 - \bar{\varepsilon}_m^{\mathcal{S},\mathcal{Z}}\right) \right\}, \quad (12)$$

where $\mathcal{Z} \in \{\mathrm{Rie}, \mathrm{GCheb}\}$ indicates whether the Riemann approach in **Proposition 3** or the Gauss–Chebyshev approach in **Proposition 4** is applied.

B. ETP, EE, Latency, and Reliability

We consider the latency-limited transmission mode, where SU_0 transmits its data with a fixed transmission rate of $\mathcal{R} = \mathcal{B}/\tau$. The system ETP is defined as the achievable effective rate of the considered network, which is measured in bits per CU (BPCU). As a result, the ETP for scheme $\mathcal{S}$ is given by $\mathrm{ETP}_{\mathrm{E2E}}^{\mathcal{S}} = \left(1 - \bar{\varepsilon}_{\mathrm{E2E}}^{\mathcal{S}}\right)\mathcal{R}/(K+1)$. Furthermore, to obtain more insights into the tradeoff between the ETP and energy consumption, we determine the EE for scheme $\mathcal{S}$, which is measured in BPCU per watt (BPCU/W) as $\mathrm{EE}_{\mathrm{E2E}}^{\mathcal{S}} = \left(1 - \bar{\varepsilon}_{\mathrm{E2E}}^{\mathcal{S}}\right)\mathcal{R} \cdot \left\{ (K+1) \sum_{k=1}^{K+1} \hat{P}_{k-1} \right\}^{-1}$.

The efficient gains from SPCs over long-packet conventional communications are low latency and ultra-reliability, where the latency and reliability for scheme $\mathcal{S}$ are presented as [8] $\mathrm{Latency}_{\mathrm{E2E}}^{\mathcal{S}} = \frac{\mathcal{B}}{\mathrm{TP}_{\mathrm{E2E}}^{\mathcal{S}}} = \frac{\tau(K+1)}{1-\bar{\varepsilon}_{\mathrm{E2E}}^{\mathcal{S}}}$ and $\mathrm{Reliability}_{\mathrm{E2E}}^{\mathcal{S}} = \left(1 - \bar{\varepsilon}_{\mathrm{E2E}}^{\mathcal{S}}\right) \cdot 100\%$, respectively.

IV. Machine-Learning Application

A. Dataset Generation

The data is generated for training based on **Proposition 3** and Section III-B. The related system parameters are chosen via an artificial stochastic process within their range. Specifically, $N \in [1,5]$, $K \in [0,10]$, $N_T \in [1,5]$, $N_R \in [1,5]$, $\overline{\gamma}_P \in [0,35]$ dB, $\overline{\gamma}_I \in [15,20]$ dB, and the correlation

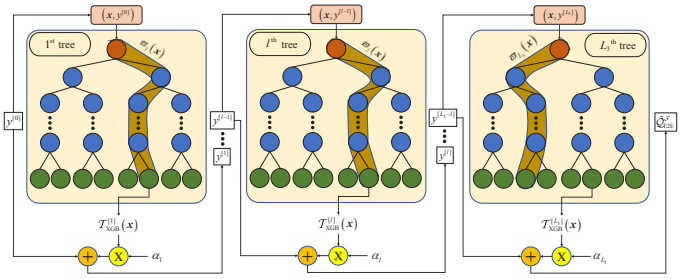

Fig. 1: Description of the decision trees in the XGB algorithm.

coefficient $\beta \in [0,1]$. Suppose that the information length of $\mathcal{B} \in [64, 8192]$ bits are encoded into the blocklength with the number of $\tau \in \left[10^2, 10^5\right]$ CUs. Then, the transmission rate $\mathcal{R} \in \left[6.4 \times 10^{-4}, 81.92\right]$ can be used as an input variable. In short, the input features for the ML-based estimation is vectorized as $\boldsymbol{x} = [\bar{\gamma}_P, \bar{\gamma}_I, \mathcal{R}, N_T, N_R, K, N, \beta]^T$. A dataset with one million samples is artificially generated based on these assumptions and 80% of the data points are used for training.

B. Description of XGB Model

The XGB model employs L_3 consecutive trees to predict $\mathcal{Q}_{\text{E2E}}^{\mathcal{S}} \in \left\{\text{ETP}_{\text{E2E}}^{\mathcal{S}}, \text{EE}_{\text{E2E}}^{\mathcal{S}}, \text{Latency}_{\text{E2E}}^{\mathcal{S}}, \text{Reliability}_{\text{E2E}}^{\mathcal{S}}\right\}$ by summing the output values of all estimators as $\widehat{\mathcal{Q}}_{\text{E2E}}^{\mathcal{S}} = \sum_{l=1}^{L_3} \alpha_l \mathcal{T}_{\text{XGB}}^{[l]}(\boldsymbol{x})$, where $\mathcal{T}_{\text{XGB}}^{[l]}(\boldsymbol{x})$ and α_l represent a nonlinear transformation and a learning rate chosen at the lth iteration, respectively. As shown in Fig. 1, the numbers of leaves are equally configured as ℓ for all trees. Each leaf of the lth regression tree is determined as a continuous value, denoted by $w_q^{[l]}$, for $q \in [1, \ell]$. For a given sample $\boldsymbol{x}$, a set of decision rules $\boldsymbol{\varpi}$ assigns an input toward a target leaf. In other words, we have $\mathcal{T}_{\text{XGB}}^{[l]}(\boldsymbol{x}) = w_q^{[l]}$ for $q = \boldsymbol{\varpi}_l(\boldsymbol{x})$. Furthermore, we assume that the predicted value at the lth tree's output is modeled as $y^{[l]} = y^{[l-1]} + \alpha_l \mathcal{T}_{\text{XGB}}^{[l]}(\boldsymbol{x})$, where $y^{[0]}$ equals the mean value of labels in the training dataset, which yields $y^{[L_3]} = \widehat{\mathcal{Q}}_{\text{E2E}}^{\mathcal{S}}$. According to the additive training, the regularized MSE loss function at the lth tree is given by [18]

$$\mathcal{J}\left(\left\{w_q^{[l]}\right\}_{q=1}^{\ell}\right) = \sum_{q=1}^{\ell}\Bigg\{\sum_{b\in I_q^{[l]}} \frac{2}{\Upsilon}\left[\mathcal{Q}_{\text{E2E},b}^{\mathcal{S}} - y_b^{[l-1]}\right] w_q^{[l]} + \left(\sum_{b\in I_q^{[l]}} \frac{1}{\Upsilon} + \chi\right)\left(w_q^{[l]}\right)^2\Bigg\}, \quad (13)$$

where χ represents a regularization parameter to reduce the overfitting problem during the training phase and $I_q^{[l]} = \left\{b | \pi^{[l]}(\boldsymbol{x}_b) = q\right\}$ contains the indices of the data points oriented to the qth leaf of the lth tree with input $\boldsymbol{x}_b$, whereas $y_b^{[l]}$ and $\mathcal{Q}_{\text{E2E},b}^{\mathcal{S}}$ are the estimated residual value at step l and the expected value of $\mathcal{Q}_{\text{E2E}}^{\mathcal{S}}$ corresponding to the bth sample point, respectively.

The number of FLOPs associated with employing a trained XGB system is mainly a comparison of the feature values in all nodes. In practical applications, the lth decision tree with a depth of $\mathcal{D}_l$ infers the result of $\mathcal{T}_{\text{XGB}}^{[l]}(\boldsymbol{x})$ after $(\mathcal{D}_l - 1)$ FLOPs. As a result, the required operations of the analysis based on the tree boosting algorithm are given by

$$\mathcal{O}_{\text{XGB}}\left(\{\mathcal{D}_l\}_{l=0}^{L_3}\right) = \sum_{l=1}^{L_3}(\mathcal{D}_l - 1) + 2L_3 = \sum_{l=1}^{L_3}\mathcal{D}_l + L_3. \quad (14)$$

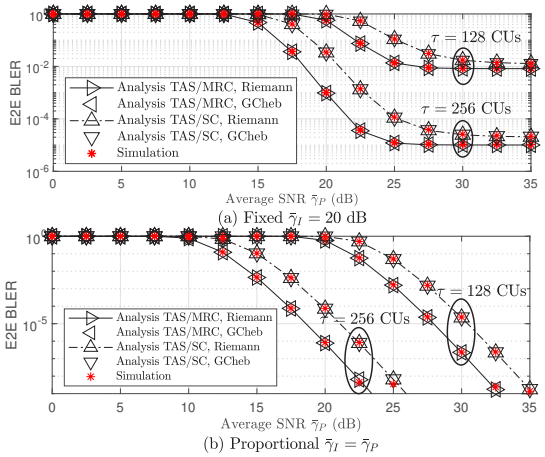

Fig. 2: E2E BLER of the MIMO underlay CR multihop relay network under SPCs in spectrum sharing.

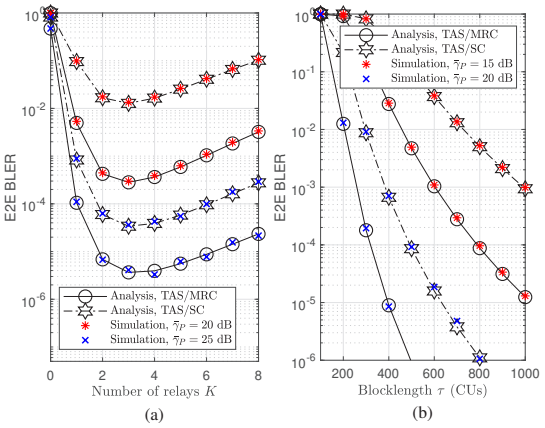

Fig. 3: E2E BLER versus (a) the number of relays K and (b) blocklength τ.

V. Numerical Results and Discussion

In this section, Monte Carlo simulations have been performed to verify our theoretical analysis. For channel settings, we consider a two-dimensional plane where PU_n, SU_0, SU_{K+1}, and SU_k are located at coordinates $(0.5, 0.5)$, $(0, 0)$, $(1, 0)$, and $\left(\frac{k}{K+1}, 0\right)$, respectively. Unless otherwise stated, we assume equal PA, $D = 1$, $\eta = 3$, $T = 50$, $\beta = 0.8$, $K = 3$, $N = 2$, $N_T = N_R = 3$, $\bar{\gamma}_P = \bar{\gamma}_I = 20$ dB, $\mathcal{B} = 1280$ bits, and $\tau = 128$ CUs. In the XBG model, the learning factor is kept constant as $\alpha_l = 0.3$ in all trees, and we assume that all trees have the same depth as $\mathcal{D}_l = \mathcal{D}$ for $l \in [1, L_3]$. The configurations of the XGB models for the TAS/MRC and TAS/SC schemes are set as $(\mathcal{D}, L_3) = (50, 95)$ and $(\mathcal{D}, L_3) = (65, 65)$, respectively.

Fig. 2 compares the E2E BLER of the proposed TAS/SC and TAS/MRC schemes for the underlay cognitive multihop relay SPC system, where two practical approaches, namely, the fixed $\bar{\gamma}_I = 20$ dB and proportional $\bar{\gamma}_I = \bar{\gamma}_P$, are shown in Fig. 2(a) and Fig. 2(b), respectively. First, in both Fig. 2(a) and Fig. 2(b), it is readily observed that the TAS/MRC scheme achieves better performance than the TAS/SC scheme over the entire SNR range. Second, in Fig. 2(a), the E2E BLER becomes saturated in the high-SNR regime as a result of the fixed tolerable interference power constraint, whereas in Fig. 2(b), the E2E performance is continuously improved with increasing average SNR $\bar{\gamma}_P$. We note that the E2E BLER floor in the fixed $\bar{\gamma}_I$ scenario can be improved significantly by adopting a longer blocklength. Third, because the Riemann integral analysis perfectly matches the Gauss–Chebyshev method, we only implement the Riemann approach in the subsequent figures for simplification. More importantly, the Monte Carlo simulation results agree well with all analytical results, which confirms of our analysis.

Fig. 3 presents the influence of the number of relays and blocklength on the system performance. In Fig. 3(a), it is clear that the E2E BLER as a function of K has convex formality, which yields the existence of an optimal K. In particular, $K = 3$ is the optimal choice in this simulation environment, which minimizes the E2E BLER. This behavior can be explained by increasing K leading to a smaller d_k, which reduces the BLER at each hop. As a result, the E2E BLER drops dramatically in the low-K range. Nevertheless, when K becomes too large, the number of time slots increases proportionally. In such a scenario, under the total transmit power budget P_{tot}, P_k becomes lower, which can result in performance degradation. In Fig. 3(b), the longer blocklength offers higher E2E performance. However, a short blocklength is expected in SPC systems, as it is selected to achieve low latency. Motivated by the tradeoff between the ultra-reliability and low-latency requirements, the value of τ should be chosen carefully.

The benefits of SPCs compared to long-packet communications (LPCs) are ultra-reliability and low latency, as shown in Fig. 4(a) and Fig. 4(b), respectively. We can observe that short-message scenarios (e.g., 128 bytes) offer higher reliability and lower latency than long-message scenarios (e.g., 512 and 1024 bytes). Encapsulating long messages into a long blocklength to guarantee ultra-reliability also causes higher latency. For example, 512 information bytes are encapsulated into packets with over 5500 CUs by using channel coding techniques (e.g., low-density parity check, polar codes, and Turbo codes) to satisfy the ultra-reliability requirement, as shown in Fig. 4(a). However, for a CU duration of 3 μs [5], Fig. 4(b) shows that the E2E latency is 22000 CUs = 66 ms, which is too high to serve uRLLC applications that require lower latency of 10 ms [5]. Meanwhile, 128 information bytes encapsulated into packets with 1000 CUs provide not only ultra-reliability, but also low latency.

Fig. 4(c) and Fig. 4(d) plot the ETP and EE versus $\bar{\gamma}_P$, whereas Fig. 4(e) and Fig. 4(f) display the ETP and EE versus K. In Fig. 4(c), we observe that the E2E throughput significantly improves when $\bar{\gamma}_P$ is increased. Apparently, in the region of high $\bar{\gamma}_P$, the ETP converges to the coding rate. In Fig. 4(d), it can be seen that there exists an optimal value of $\bar{\gamma}_P$ to achieve the highest EE. We also note that EE represents the tradeoff between the ETP and energy consumption. In the low-SNR region, EE monotonically increases as $\bar{\gamma}_P$ increases. Meanwhile, in the high-SNR region, EE is reduced. This is because, at a sufficiently high SNR value, the ETP achieves a target rate and remains saturated, whereas $\bar{\gamma}_P$ still increases, leading to EE degradation. Furthermore, it can be further observed that higher numbers of transmit and receive antennas provide better performance in terms of ETP and EE. Fig. 4(e) and Fig. 4(f) reveal that there exists an optimal K to maximize the ETP and EE. Notably, choosing the value of K is important in multihop networks to balance the tradeoffs among the implementation cost, transmit power, and system performance (e.g., reliability, latency, ETP, and EE). Furthermore, we note that even though the case $N_T = N_R = 6$ falls outside the range of the XGB input features, the XGB prediction results still closely align with the theoretical results.

TABLE I: Average run time (ms) versus (K, N) of the ML-based and analytical computations.

(K, N)	XGB		Riemann		GCheb	
	TAS/MRC	TAS/SC	TAS/MRC	TAS/SC	TAS/MRC	TAS/SC
$(2, 4)$	0.30	0.29	0.80	0.66	32.10	22.16
$(4, 6)$	0.31	0.36	2.12	1.21	90.12	53.51
$(6, 8)$	0.29	0.31	3.50	2.41	131.22	97.53
$(8, 10)$	0.34	0.34	6.12	3.58	180.46	114.88

Most importantly, Fig. 4 shows that the analysis, simulation, and XGB prediction results are in excellent agreement, which confirms our designed ML application and theoretical analysis. Table I shows the execution time of the XGB estimator and the computation based on the closed-form expressions. As can be observed, the amount of time spent on computing based on the analytical derivations significantly increases as K and N increase, whereas the XGB approach remains fixed regardless of execution time. In other words, the numbers of elements of the primary and secondary networks in the underlay CR multihop relay systems have no impact on the performance prediction. In Table I, the XGB run time remains stable at around 0.3 ms. In contrast, for the TAS/MRC scheme, as K and N increase to $(K, N) = (8, 10)$, the Riemann and Gauss–Chebyshev approximations gradually increase to 6.12 ms and 180.46 ms, which are approximately 20 and 600 times higher than the XGB prediction, respectively.

VI. Conclusions

This work studies SPCs for the MIMO underlay CR multihop relay network with multiple PUs, where transceivers transmit and receive short packets to provide uRLLCs. The approximated E2E BLER closed-form expressions of both the TAS/MRC and TAS/SC schemes are derived in a practical scenario under imperfect CSI of the interference channels, from which their ETP, EE, latency, and reliability are analyzed. The strengths of the MIMO implementation, SPC, and multihop relay networks in accordance with uRLLC requirements are

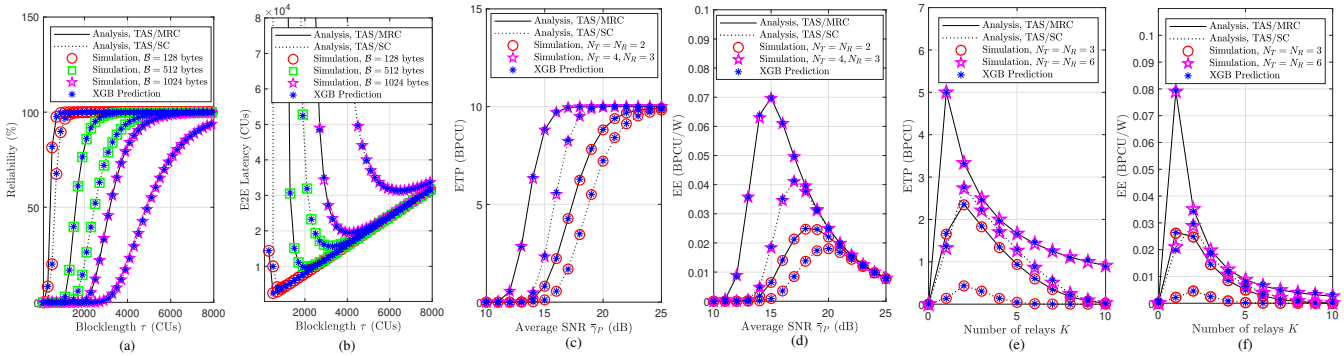

Fig. 4: The influence of blocklength τ and message length $\mathcal{B}$ on the performance in terms of (a) reliability and (b) E2E latency; (c) ETP and (d) EE versus the average SNR $\bar{\gamma}_P$; (e) ETP and (f) EE versus the number of relays K.

confirmed via the numerical results, and the impacts of the average SNR and information length of bits on the system performance are also shown. Furthermore, the employed XGB estimator provides both highly accurate predictions and a very short execution time, which demonstrates the possibility of realizing real-time configurations for wireless systems.

Appendix A
Proof of Proposition 1

Based on (2), the CDF of $\gamma_k^{\mathrm{TAS/MRC}}$ is calculated as

$$\begin{aligned} &F_{\gamma_k}^{\mathrm{TAS/MRC}}(\gamma) \\ &= \prod_{i=1}^{N_T} \Pr\left\{ \min\left(\bar{\gamma}_P, \frac{\bar{\gamma}_I}{\max\limits_{1\le n\le N} |\hat{g}_{k,n}|^2} \right) \sum_{j=1}^{N_R} \left| h_k^{(i,j)} \right|^2 \le \gamma \right\} \\ &= \left\{ \Pr\left(\bar{\gamma}_P \le \frac{\bar{\gamma}_I}{\mathcal{Y}_k}, \bar{\gamma}_P \mathcal{X}_k \le \gamma \right) \right. \\ &\qquad \left. + \Pr\left(\bar{\gamma}_P > \frac{\bar{\gamma}_I}{\mathcal{Y}_k}, \frac{\bar{\gamma}_I}{\mathcal{Y}_k} \mathcal{X}_k \le \gamma \right) \right\}^{N_T}. \end{aligned} \tag{A.1}$$

It is noted that, in (A.1), the first term is the probability of two independent events, whereas the second term is the probability of two dependent events. Therefore, (A.1) can be rewritten as

$$\begin{aligned} F_{\gamma_k}^{\mathrm{TAS/MRC}}(\gamma) = \Bigg\{ & F_{\mathcal{Y}_k}\left(\frac{\bar{\gamma}_I}{\bar{\gamma}_P} \right) F_{\mathcal{X}_k}\left(\frac{\gamma}{\bar{\gamma}_P} \right) \\ & + \int_{\bar{\gamma}_I/\bar{\gamma}_P}^{\infty} f_{\mathcal{Y}_k}(y) F_{\mathcal{X}_k}\left(\frac{\gamma}{\bar{\gamma}_I} y \right) dy \Bigg\}^{N_T}. \end{aligned} \tag{A.2}$$

By substituting the CDFs and PDFs of $\mathcal{X}_k$ and $\mathcal{Y}_k$ in (3)–(6) into (A.2), after some mathematical manipulations, we derive (7). The proof is completed.

References

[1] N. N. Dao, Q. V. Pham, N. H. Tu, T. T. Thanh, V. N. Q. Bao, D. S. Lakew, and S. Cho, "Survey on aerial radio access networks: Toward a comprehensive 6G access infrastructure," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 2, pp. 1193–1225, Secondquarter 2021.

[2] A. Sharmila and P. Dananjayan, "Spectrum sharing techniques in cognitive radio networks–a survey," in *2019 IEEE Int. Conf. Syst. Comput. Automat. Netw. (ICSCAN)*, Pondicherry, India, Mar. 2019, pp. 1–4.

[3] W. Xu, W. Yuan, Q. Shi, X. Wang, and Y. Zhang, "Distributed energy-efficient cross-layer optimization for multihop MIMO cognitive radio networks with primary user rate protection," *IEEE Trans. Veh. Technol.*, vol. 66, no. 1, pp. 785–797, Jan. 2017.

[4] N. B. Halima and H. Boujemâa, "Energy harvesting with adaptive transmit power for multi-antenna multihop cognitive radio networks," *Sustainable Computing: Informatics and Systems*, p. 100567, Sep. 2021.

[5] O. L. A. López, H. Alves, R. D. Souza, and E. M. G. Fernández, "Ultrareliable short-packet communications with wireless energy transfer," *IEEE Signal Process. Lett.*, vol. 24, no. 4, pp. 387–391, Apr. 2017.

[6] N. T. Y. Linh, N. H. Tu, P. N. Son, and V. N. Q. Bao, "Dual-hop relaying networks for short-packet URLLCs: Performance analysis and optimization," *J. Commun. Netw.*, vol. 24, no. 4, pp. 408–418, 2022.

[7] N. H. Tu and K. Lee, "Performance analysis and optimization of multihop MIMO relay networks in short-packet communications," *IEEE Trans. Wireless Commun.*, vol. 21, no. 6, pp. 4549–4562, Jun. 2022.

[8] C. D. Ho, T.-V. Nguyen, T. Huynh-The, T.-T. Nguyen, D. B. da Costa, and B. An, "Short-packet communications in wireless-powered cognitive IoT networks: Performance analysis and deep learning evaluation," *IEEE Trans. Veh. Technol.*, vol. 70, no. 3, pp. 2894–2899, Mar. 2021.

[9] T.-H. Vu, T.-V. Nguyen, and S. Kim, "Wireless powered cognitive NOMA-based IoT relay networks: Performance analysis and deep learning evaluation," *IEEE Internet Things J.*, vol. 9, no. 5, pp. 3913–3929, Mar. 2022.

[10] L. Li, S. A. Vorobyov, and A. B. Gershman, "Transmit antenna selection based strategies in MISO communication systems with low-rate channel state feedback," *IEEE Trans. Wireless Commun.*, vol. 8, no. 4, pp. 1660–1666, Apr. 2009.

[11] Z. Chen, J. Yuan, and B. Vucetic, "Analysis of transmit antenna selection/maximal-ratio combining in rayleigh fading channels," *IEEE Trans. Veh. Technol.*, vol. 54, no. 4, pp. 1312–1321, Jul. 2005.

[12] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*. Academic Press, 2014.

[13] A. F. Coskun and O. Kucur, "Performance of joint transmit and receive antenna selection in nakagami-m fading channels," in *2010 European Wireless*, Lucca, Italy, Apr. 2010.

[14] B. Makki, T. Svensson, and M. Zorzi, "Finite block-length analysis of the incremental redundancy HARQ," *IEEE Wireless Commun. Lett.*, vol. 3, no. 5, pp. 529–532, Aug. 2014.

[15] R. M. McLeod, *The generalized Riemann integral*. Amer. Math. Soc., 1980, vol. 20.

[16] C. K. Chui, "Concerning Gaussian–Chebyshev quadrature errors," *SIAM J. Numer. Anal.*, vol. 9, no. 2, pp. 237–240, 1972.

[17] M. Khafagy, A. Ismail, M.-S. Alouini, and S. Aissa, "On the outage performance of full-duplex selective decode-and-forward relaying," *IEEE Commun. Lett.*, vol. 17, no. 6, pp. 1180–1183, Apr. 2013.

[18] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, San Francisco, USA, August 2016, pp. 785–794.