IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received September 20, 2021, accepted October 7, 2021, date of publication October 11, 2021, date of current version October 29, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3119318

Joint Design of Beamforming and Antenna Selection in Short Blocklength Regime for URLLC in Cognitive Radio Networks

NGO TAN VU KHANH[1] AND TIEN-TUNG NGUYEN[2]

[1]Department of Ecommerce Technology, School of Business Information Technology, University of Economics Ho Chi Minh City, Ho Chi Minh City 70000, Vietnam

[2]Telecommunication Division, Industrial University of Ho Chi Minh City, Ho Chi Minh City 700000, Vietnam

Corresponding author: Ngo Tan Vu Khanh (khanhntv@ueh.edu.vn)

This work was supported by the University of Economics Ho Chi Minh City, Vietnam.

ABSTRACT This paper considers the short blocklength regime for cognitive radio networks (CRNs) to deliver ultra-reliable and low-latency communications (URLLCs) promised for beyond 5G networks. The secondary system consists of a secondary transmitter (ST) and multiple secondary users, which are allowed to access the same spectrum of licensed users (i.e., the primary system). Adopting linear beamforming at ST, we formulate the optimization problem of the energy-efficient maximization for the secondary system under the power constraint at ST and interference power constraints at primary receivers. In the short blocklength regime, the rate function is more complex and computationally intractable than the traditional Shannon rate function, which makes the formulated problem more difficult to solve. By leveraging techniques from the Dinkelbach method and the inner approximation method, we first devise newly approximated functions to convexify nonconvex constraints, and the iterative algorithm is then developed to obtain at least a locally optimal solution. To further enhance the energy efficiency of the secondary system, we consider a joint optimization of beamforming and antenna selection at ST, where binary variables are introduced to establish the operation modes of transmit antennas. To solve the mixed-integer nonconvex problem, we incorporate the penalty function into the objective function to dealing with the uncertainty of binary variables. Numerical results are provided to demonstrate the fast convergence and merits of the proposed algorithms, as well as to confirm the role of antenna selection in improving energy efficiency.

INDEX TERMS Antenna selection, beamforming, broadcast channel, cognitive radio, Dinkelbach method, energy efficiency, inner approximation, mixed-integer programming.

I. INTRODUCTION

A prominent solution to tackle the spectrum scarcity is based on cognitive radio (CR) techniques, where the secondary system (i.e., unlicensed users) is allowed to access the frequency spectrum of the primary system (i.e., licensed users). There are two well-known CR schemes, such as the opportunistic spectrum access model and spectrum sharing model [1]. In the former, the secondary transmitter (ST) acquirers the full information of the primary network to coexist with it to avoid interference [2], [3], and in the latter, the aggregated interference power at the primary receivers (PRs) caused by ST must be lower than a predefined threshold [4], [5]. Towards this end, spectral efficiency (SE) and energy efficiency (EE) have been be considered as the key performance metrics in designing an effective CR network (CRN).

On the other hand, ultra-reliable low-latency communication (URLLC) is considered as one of the key pillars of 5G New Radio to support emerging services and mission-critical applications such as Internet of Things (IoT), smart grids, remote surgery and intelligent transportation systems [6]. In 4G-LTE, the end-to-end (e2e) latency, including processing delay, wireless transmission delay and queuing delay etc., is in the 4-millisecond range, and targets even less than one millisecond in 3GPP Release 15 for 5G networks and beyond. In addition, URLLC requires 99.999 percent reliability and a very low block error rate, such as in the range of $10^{-9} - 10^{-5}$, depending on the specific use cases [7]. This makes a considerable challenge for system design and operations and will require fundamentally different approaches

The associate editor coordinating the review of this manuscript and approving it for publication was Mugen Peng.

due to a conflicting goal of low latency and ultrahigh reliability. In particular, low-latency requirements mandate the use of short packets to transmit a small amount of data which leads to a severe degradation in the reliability and channel coding gain, while ensuring ultrahigh reliability demands for more resources (e.g., parity, redundancy), resulting in higher latency [6], [8]. Therefore, more practical approaches with low computational complexity for efficient deployment of URLLC-enabled networks are of crucial importance.

A. RELATED WORKS

URLLC is able to unleash a plethora of emerging applications such as intelligent transportation system, tactile internet, remote surgery, autonomous driving, and factory automation, etc [8]. Recently, there has been several studies on resource allocation of URLLC-enabled networks. Elayoubi *et al.* [9] developed different resource allocation strategies for transmissions and re-transmission in the Industrial Internet of Things (IoT), depending on the requirements of different services and on the traffic characteristics. A joint design of subchannel assignment and power control to maximize the total network energy efficiency was studied in [10], where deep reinforcement learning (DRL) was developed for its solution. By characterizing the effects of system parameters, the authors in [11] developed an efficient model for URLLC packet transmissions and derived scaling results for the URLLC capacity.

In order to guarantee low-latency communication, short block-length communication must be used [12], which will pose significant challenges in terms of system design and performance optimization. Under the short block-length regime, the performance analysis in terms of throughput and decoding error probability is more complicated than in the long block-length regime using the standard Shannon capacity formula. For a given blocklength and error probability, Polyanskiy *et al.* [13] first derived the approximated achievable rate in short block-length regime as a function of channel capacity, channel dispersion and complementary Gaussian cumulative distribution function. This work was further extended in [14] which investigated the maximal achievable rate under quasi-static multiple-input multiple-output fading channels. Naturally, the resource allocation problems in URLLC are more difficult to solve, compared to the standard Shannon capacity formula, which requires new approaches with low computational complexity.

Recently, the literature on resource allocation in URLLC has recently received considerable attention. Specifically, the authors in [15] proposed a packet delivery mechanism to reduce the required bandwidth for URLLC while meeting queueing delay. References [16] developed a proactive packet dropping mechanism to optimize the packet dropping, power allocation and bandwidth allocation policies, which helps satisfy both transmission and queueing delays. The EE maximization problem was considered in [17] by jointly optimizing bandwidth, power control and antenna configuration, subject to latency and reliability constraints. The unsupervised deep learning framework was introduced in [18] to minimize bandwidth, where the QoS constraints of URLLC are taken into account. To deploy URLLC in real-time wireless control systems, a co-design problem of URLLC and control was introduced in [19] which developed a low-complexity iteration algorithm. It is noted that all these aforementioned works adopted lower-bound approximations of URLLC rate functions, aiming to simplify the optimization problems.

Despite its potential, there is only a few attempts on studying resource allocation of URLLC-enabled CRNs. By considering both massive machine-type communication (mMTC) and URLLC services for a cognitive unmanned aerial vehicle (UAV)-aided network, the authors in [20] first derived the analytical expressions of the throughput and then developed an efficient algorithm for EE maximization problem. Tackling the stringent requirement of latency and reliability in the URLLC and enhanced mobile broadband (eMBB) coexistence system was studied in [21], where the authors developed a dynamic spectrum allocation scheme to help alleviating the packet collisions of multiplexing URLLC and eMBB packets. However, these works only considered a single antenna at the secondary transmitter while in practice it is often equipped with multiple antennas to guarantee high throughput, which requires an efficient beamforming design.

B. MAIN CONTRIBUTIONS

Against the above background, we consider the URLLC in downlink cognitive radio networks with short block-length regime, where a multi-antenna ST transmits data to multiple secondary receivers (SRs). We study the optimization problem of maximizing energy efficiency of the secondary system subject to power constraint at ST and interference power constraint at PRs. We also propose a joint design of beamforming and antenna selection (JDBAS) at ST to select a subset of active antennas to transmit data in each transmission block. This design helps reduce the total power consumption while still maintaining a good SE, resulting in better EE performance. The optimization problem of interest is either nonconvex programming or mixed-integer non-convex programming, which is often very challenging and even impossible to obtain its globally optimal solution. We develop low-complexity iterative algorithms to address these problems efficiently. Our main contributions are summarized as follows.

- We formulate the EE maximization (EEmax) problem for an URLLC-aided CRN subject to the power constraint at ST and interference power constraints at PRs. Our consideration leads to a highly nonconvex optimization problem, which is hard to obtain a globally optimal solution. Towards an appealing application, we first apply Dinkelbach method [22] to make the original problem more computational tractable and then employ inner approximation (IA) framework [23] to approximate nonconvex parts. An iterative algorithm of low-computational complexity is developed for its

solution, which guarantees to achieve at least a local optimal solution.

- To further improve EE performance, we consider the JDBAS-EEmax problem by jointly designing beamforming and antenna selection at ST. The key idea is to find the optimal subset of active antennas by introducing binary variables to establish on/off operation modes of transmit antennas at ST, which are capable of reducing the power consumption of radio frequency (RF) chains, resulting in higher EE. The problem is formulated as a mixed-integer nonconvex problem, which is even more difficult to solve compared to the EEmax problem. To this end, we introduce a parameterized relaxed problem to tackle uncertainties of binary nature and develop an efficient iterative algorithm based on the approximate convex functions developed in solving the EEmax problem.
- Numerical results are provided to show fast convergence of the proposed algorithms. They also reveal the effectiveness of optimizing antenna selections in terms of the achievable EE performance.

C. PAPER STRUCTURE AND NOTATION

The rest of the paper is organized as follows. Section II presents the system model and problem formulation. We provide the iterative algorithms for solving EEmax and JDBAS-EEmax problems in Sections III and IV, respectively. Numerical results and conclusions are provided in Sections V and VI, respectively.

Notations: We summarize the main mathematical notations which are frequently used in the paper as follows. Bold lowercase letters and lowercase letters represent for vectors and scalars, respectively. $(\cdot)^H$, $(\cdot)^T$ and $(\cdot)^*$ denote the Hermitian transpose, normal transpose, and conjugate, respectively. $|\cdot|$ and $||\cdot||_2$ correspond to the cardinality, and $l2-$norm operators, respectively. $\mathbb{E}[\cdot]$ represents the expectation operation. $\mathcal{CN}(\mu, \sigma^2)$ is circularly symmetric complex Gaussian random variable with mean μ and variance σ^2.

II. SYSTEM MODEL AND PROBLEM FORMULATION

A. SIGNAL MODEL

We consider a downlink CRN consisting of an N-antenna ST (or the secondary base station (SBS)) serving K single-antenna SRs,[1] which are allowed to share the same frequency spectrum licensed for M PRs, as shown Fig. 1. Let us denote by $\mathcal{N} \triangleq \{1, \cdots, N\}$, $\mathcal{K} \triangleq \{1, \cdots, K\}$ and $\mathcal{M} \triangleq \{1, \cdots, M\}$ the sets of transmit antennas at ST, K SRs and M PRs, respectively. The channel vectors from ST to SR $k \in \mathcal{K}$ and PR $m \in \mathcal{M}$ are denoted by $\mathbf{h}_k \in \mathbb{C}^{N\times 1}$ and $\mathbf{g}_m \in \mathbb{C}^{N\times 1}$, respectively, which account for the effects of both large-scale and small-scale fading. All the channels are assumed to be unchanged during a transmission block and change independently from one block to another.

[1]The case with multi-antenna SRs allows ST to transmit multiple independent data streams to each SR, leading to higher SE [24]. This is an interesting future work that is out of the scope of this paper.

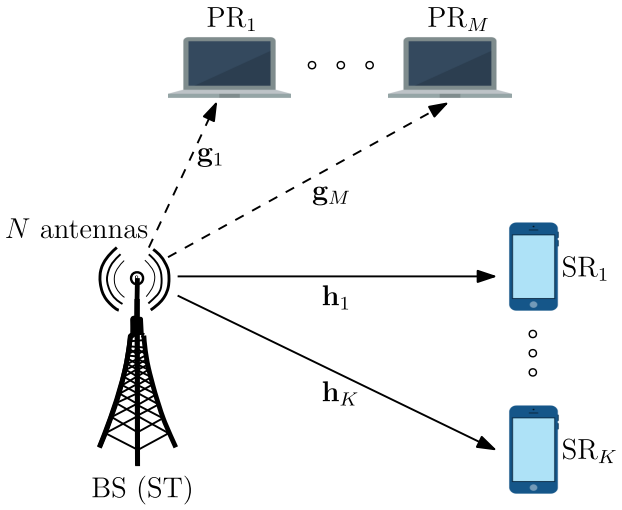

FIGURE 1. Illustration of a cognitive network model with K SRs and M PRs.

For the secondary system, we adopt linear beamforming at ST to convey independent symbols to SRs. Let $x_k \in \mathbb{C}$ with $\mathbb{E}\{|x_k|^2\} = 1$ and $\mathbf{w}_k \in \mathbb{C}^{N\times 1}$ denote the information symbol and corresponding beamforming, respectively, intended to SR k. The received signal at SR k can be expressed as

$$y_k = \mathbf{h}_k^H \mathbf{w}_k x_k + \sum_{j\in\mathcal{K}\setminus\{k\}} \mathbf{h}_k^H \mathbf{w}_j x_j + n_k \tag{1}$$

where $n_k \sim \mathcal{CN}(0, \sigma^2)$ is the additive white Gaussian noise (AWGN) with zero-mean and variance σ^2.

Channel State Information (CSI) Model: Unlike [24]–[26], which assumed the perfect CSI of all channels, we consider the worst-case design where the CSI of both $\mathbf{h}_k, \forall k$ and $\mathbf{g}_m, \forall m$ is imperfectly known at transceivers. Let $\hat{\mathbf{h}}_k$ and $\hat{\mathbf{g}}_m$ denote the channel estimates of $\mathbf{h}_k, \forall k$ and $\mathbf{g}_m, \forall m$. We note that the specific channel estimation scheme is out of the scope of this paper. Following [27], [28], the associated CSI errors are deterministic and bounded as

$$f(\mathbf{h}_k\mathbf{h}_k^H - \hat{\mathbf{h}}_k\hat{\mathbf{h}}_k^H) \le \xi_k, \quad \forall k, \tag{2}$$
$$f(\mathbf{g}_m\mathbf{g}_m^H - \hat{\mathbf{g}}_m\hat{\mathbf{g}}_m^H) \le \zeta_m, \quad \forall m \tag{3}$$

with $f(\mathbf{X}) = \max_i |\lambda_i(\mathbf{X})|$ where $\lambda_i(\mathbf{X})$ is the eigenvalues of $\mathbf{X}$; $\xi_k \in [0, 1)$ and $\zeta_m \in [0, 1)$ are the upper bounded CSI errors. To simplify the worst-case design, we further re-express ξ_k and ζ_m as

$$\xi_k = \delta_s \|\hat{\mathbf{h}}_k\|^2, \quad \forall k, \tag{4}$$
$$\zeta_m = \delta_p \|\hat{\mathbf{g}}_m\|^2, \quad \forall m \tag{5}$$

where δ_s and δ_p denote the normalized channel uncertainty levels of $\mathbf{h}_k$ and $\mathbf{g}_m$, respectively. In practice, the primary system operates independently of the secondary system, leading to $\delta_s < \delta_p$.

B. PROBLEM FORMULATION

Power Consumption Model: In this work, we consider the linear power model in [29] which has been widely adopted in the

literature, taking into account both the total circuit power and the power consumption for the transmitted data. In particular, the total power consumption at ST can be modeled as

$$P_{\text{tot}}(\mathbf{w}) = \frac{1}{\mu}\sum_{k\in\mathcal{K}} \|\mathbf{w}_k\|^2 + NP_{\text{dyn}} + P_{\text{sta}} \tag{6}$$

where $\mu \in (0, 1]$, P_{dyn} and P_{sta} are the power amplifier efficiency, dynamic power consumption (e.g., power radiation of RF chain) and the static power (e.g., cooling system), respectively.

By incorporating channel uncertainties (4) and (5) into (1) and treating multiuser interference as thermal noise [28], [30], the worst-case signal-to-interference-plus-noise ratio (SINR) of SR k can be expressed as

$$\gamma_k(\mathbf{w}) \triangleq \frac{|\hat{\mathbf{h}}_k^H \mathbf{w}_k|^2 - \xi_k\|\mathbf{w}_k\|^2}{\sum_{j\in\mathcal{K}\setminus\{k\}} |\hat{\mathbf{h}}_k^H \mathbf{w}_j|^2 + \sum_{j\in\mathcal{K}\setminus\{k\}} \xi_k\|\mathbf{w}_j\|^2 + \sigma^2} \tag{7}$$

where $\mathbf{w} \triangleq [\mathbf{w}_1^H, \cdots, \mathbf{w}_K^H]^H \in \mathbb{C}^{KN\times 1}$. The worst-case throughput of SR k in nats/s for URLLC can be approximated as [13]:

$$R_k(\mathbf{w}) = B\Big(\ln\big(1+\gamma_k(\mathbf{w})\big) - \sqrt{\frac{V(\gamma_k(\mathbf{w}))}{\tau B}}Q^{-1}(\epsilon_k)\Big) \tag{8}$$

where $V(\gamma_k(\mathbf{w})) = 1 - \big(1+\gamma_k(\mathbf{w})\big)^{-2}$ is the channel dispersion and $Q^{-1}(\cdot)$ is the inverse function with $Q(x) = \frac{1}{\sqrt{2\pi}}\int_x^\infty \exp(-t^2/2)dt$; B, τ and ϵ_k are the system bandwidth, transmission time interval (TTI) and decoding error probability, respectively.

The EE in nats/Joule can be expressed as

$$\mathsf{EE}(\mathbf{w}) \triangleq \frac{\sum_{k\in\mathcal{K}} R_k(\mathbf{w})}{P_{\text{tot}}(\mathbf{w})} \tag{9}$$

which is defined as the ratio between the sum throughput (nats/s) of all SRs and the total power consumption (Watt) at ST. Here the power consumed for the signal processing at SRs is ignored as it is dominated by $P_{\text{tot}}(\mathbf{w})$ at ST. The EE optimization problem of interest is mathematically stated as:

$$\begin{aligned}
(\text{EEmax}): \underset{\mathbf{w}}{\text{maximize}} \quad & \mathsf{EE}(\mathbf{w}) && (10a)\\
\text{subject to} \quad & \Psi_m(\mathbf{w}) \le \mathcal{I}_m, \quad \forall m && (10b)\\
& R_k(\mathbf{w}) \ge \mathrm{R}_k^{\min}, \quad \forall k && (10c)\\
& \sum_{k\in\mathcal{K}} \|\mathbf{w}_k\|^2 \le P_{\max} && (10d)
\end{aligned}$$

where $\Psi_m(\mathbf{w}) \triangleq \sum_{j\in\mathcal{K}} |\hat{\mathbf{g}}_m^H \mathbf{w}_j|^2 + \sum_{j\in\mathcal{K}} \zeta_m\|\mathbf{w}_j\|^2$ given in constraint (10b) is the total interference power at PR m caused by the downlink transmission of the secondary system, which is limited by $\mathcal{I}_m$. Constraint (10c) is to ensure the predetermined rate requirement $\mathrm{R}_k^{\min}$ for SR k, and constraint (10d) is the power constraint at ST with the power budget $P_{\max}$. It is easy to observe that the rate function $R_k(\mathbf{w})$ is either nonconvex or nonconcave, leading to the nonconvexity of constraint (10c) and nonconcavity of the objective (10a). Hence, problem (10) is nonconvex which is challenging to solve due to the complicated throughput function in (8).

III. PROPOSED ITERATIVE ALGORITHM FOR EEmax

In this section, we focus on designing an efficient algorithm to solve the EEmax problem (10). To doing so, we first convert (10) to an equivalent problem by the Dinkelbach's approach [22], and then employ IA framework to iteratively convexify nonconvex functions [23].

A. TRACTABLE FORMULATION

To start with we first equivalently rewrite (10) as

$$\begin{aligned}
\underset{\mathbf{w},\mathbf{r},\rho}{\text{maximize}} \quad & \frac{\sum_{k\in\mathcal{K}} r_k}{\rho} && (11a)\\
\text{subject to} \quad & (10b),\ (10d) && (11b)\\
& R_k(\mathbf{w}) \ge r_k, \quad \forall k && (11c)\\
& r_k \ge \mathrm{R}_k^{\min}, \quad \forall k && (11d)\\
& P_{\text{tot}}(\mathbf{w}) \le \rho && (11e)
\end{aligned}$$

where $\mathbf{r} \triangleq \{r_k\}_{\forall k}$ and ρ are newly introduced optimization variables, which can be considered as soft throughputs of SRs and soft power consumption, respectively. We note that the equivalence between (10) and (11) is ensured by the fact that constraints (11c) and (11e) must hold with equality at optimum. Constraint (11d) is derived from (10c) and (11c). Next, by the Dinkelbach transformation, we rewrite (11) as

$$\begin{aligned}
\underset{\mathbf{w},\mathbf{r},\rho}{\text{maximize}} \quad & \sum_{k\in\mathcal{K}} r_k - \psi\rho && (12a)\\
\text{subject to} \quad & (10b),\ (10d),\ (11c),\ (11d),\ (11e) && (12b)
\end{aligned}$$

where ψ is a given parameter, which will be iteratively updated at iteration i as $\psi^{(i)} = \sum_{k\in\mathcal{K}} r_k^{(i)}/\rho^{(i)}$. We can observe that the objective function (12a) is linear while constraints (10b), (10d), (11d), and (11e) are convex.

B. PROPOSED ITERATIVE ALGORITHM

To convexify (11c), we will apply IA method to approximate the nonconvex rate function. Let us rewrite (8) as

$$R_k(\mathbf{w}) = B\Big(F_k(\mathbf{w}) - G_k(\mathbf{w})\frac{Q^{-1}(\epsilon_k)}{\sqrt{\tau B}}\Big) \tag{13}$$

where $F_k(\mathbf{w}) \triangleq \ln\big(1+\gamma_k(\mathbf{w})\big)$ and $G_k(\mathbf{w}) \triangleq \sqrt{V(\gamma_k(\mathbf{w}))}$. We are now in position to find a concave lower bound of $F_k(\mathbf{w})$ and a convex upper bound of $G_k(\mathbf{w})$.

Concave Lower Bound of $F_k(\mathbf{w})$: We introduce new variables $\boldsymbol{\varphi} \triangleq \{\varphi_k\}_{\forall k}$ to express $F_k(\mathbf{w})$ as

$$F_k(\mathbf{w}) \ge F_k(\mathbf{w}, \varphi_k) \triangleq \ln\big(1+\gamma_k(\mathbf{w},\varphi_k)\big) \tag{14}$$

with the additional constraint

$$|\hat{\mathbf{h}}_k^H \mathbf{w}_k|^2 - \xi_k\|\mathbf{w}_k\|^2 \ge \varphi_k^2 \tag{15}$$

where

$$\gamma_k(\mathbf{w},\varphi_k) \triangleq \frac{\varphi_k^2}{\chi_k(\mathbf{w})} \tag{16}$$

and

$$\chi_k(\mathbf{w}) \triangleq \sum_{j\in\mathcal{K}\setminus\{k\}} |\hat{\mathbf{h}}_k^H \mathbf{w}_j|^2 + \sum_{j\in\mathcal{K}\setminus\{k\}} \xi_k\|\mathbf{w}_j\|^2 + \sigma^2. \tag{17}$$

By [31, Eq. (20)], a lower bound of $F_k(\mathbf{w}, \varphi_k)$ at a given feasible point $(\mathbf{w}^{(\kappa)}, \varphi_k^{(\kappa)})$ found at iteration κ of an iterative algorithm presented shortly is

$$\begin{aligned} F_k(\mathbf{w}, \varphi_k) \;\geq\; & F(\mathbf{w}^{(\kappa)}, \varphi_k^{(\kappa)}) - \gamma_k(\mathbf{w}^{(\kappa)}, \varphi_k^{(\kappa)}) + 2\frac{\varphi_k^{(\kappa)}\varphi_k}{\chi_k(\mathbf{w}^{(\kappa)})} \\ & - \frac{(\varphi_k^{(\kappa)})^2}{\chi_k(\mathbf{w}^{(\kappa)})(\chi_k(\mathbf{w}^{(\kappa)}) + (\varphi_k^{(\kappa)})^2)}\big(\chi_k(\mathbf{w}) + \varphi_k^2\big) \\ := \; & \mathcal{F}^{(\kappa)}(\mathbf{w}, \varphi_k). \end{aligned} \tag{18}$$

It is observed that the approximate function $\mathcal{F}^{(\kappa)}(\mathbf{w}, \varphi_k)$ is concave, satisfying

$$F_k(\mathbf{w}^{(\kappa)}, \varphi_k^{(\kappa)}) = \mathcal{F}^{(\kappa)}(\mathbf{w}^{(\kappa)}, \varphi_k^{(\kappa)}). \tag{19}$$

Next, we re-express (15) as

$$|\hat{\mathbf{h}}_k^H \mathbf{w}_k|^2 \geq \varphi_k^2 + \xi_k \|\mathbf{w}_k\|^2 \tag{20}$$

where $|\hat{\mathbf{h}}_k^H \mathbf{w}_k|^2$ is the quadratic function, which is convex and can be approximated by the first-order Taylor approximation. By [31, Eq. (21)], we innerly approximate (20) as

$$\Phi_k^{(\kappa)}(\mathbf{w}_k) \geq \varphi_k^2 + \xi_k \|\mathbf{w}_k\|^2 \tag{21}$$

where $\Phi_k^{(\kappa)}(\mathbf{w}_k) \triangleq 2\Re\{(\mathbf{w}_k^{(\kappa)})^H \hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^H \mathbf{w}_k\} - |\hat{\mathbf{h}}_k^H \mathbf{w}_k^{(\kappa)}|^2$ is the linear function and the first-order Taylor approximation of $|\hat{\mathbf{h}}_k^H \mathbf{w}_k|^2$.

Convex Upper Bound of $G_k(\mathbf{w})$: We first rewrite $G_k(\mathbf{w})$ as

$$G_k(\mathbf{w}) \geq G_k(\mathbf{w}, \phi_k) \triangleq \sqrt{1 - \big(1 + \gamma_k(\mathbf{w}, \phi_k)\big)^{-2}} \tag{22}$$

with the additional constraint

$$|\hat{\mathbf{h}}_k^H \mathbf{w}_k|^2 - \xi_k \|\mathbf{w}_k\|^2 \leq \phi_k \tag{23}$$

where

$$\gamma_k(\mathbf{w}, \phi_k) \triangleq \frac{\phi_k}{\chi_k(\mathbf{w})}. \tag{24}$$

We note that $\sqrt{x}$ with $x > 0$ is a concave function, where its convex approximation around the feasible point $x^{(\kappa)}$ is given as [32]:

$$\sqrt{x} \leq \frac{\sqrt{x^{(\kappa)}}}{2} + \frac{x}{2\sqrt{x^{(\kappa)}}}. \tag{25}$$

By (25), we have

$$\begin{aligned} G_k(\mathbf{w}, \phi_k) \;\leq\; & \frac{G_k(\mathbf{w}^{(\kappa)}, \phi_k^{(\kappa)})}{2} + \frac{G_k^2(\mathbf{w}, \phi_k)}{2G_k(\mathbf{w}^{(\kappa)}, \phi_k^{(\kappa)})} \\ = \; & \mathtt{a}_k - \mathtt{b}_k \frac{\chi_k^2(\mathbf{w})}{\big(\chi_k(\mathbf{w}) + \phi_k\big)^2} \end{aligned} \tag{26}$$

where

$$\mathtt{a}_k \triangleq \frac{G_k(\mathbf{w}^{(\kappa)}, \phi_k^{(\kappa)})}{2} + \frac{1}{2G_k(\mathbf{w}^{(\kappa)}, \phi_k^{(\kappa)})} \tag{27}$$

$$\mathtt{b}_k \triangleq \frac{1}{2G_k(\mathbf{w}^{(\kappa)}, \phi_k^{(\kappa)})}. \tag{28}$$

By the inequality

$$\frac{1}{x} \geq \frac{2}{x^{(\kappa)}} - \frac{x}{(x^{(\kappa)})^2} \tag{29}$$

it follows that [32]:

$$\begin{aligned} \frac{\chi_k^2(\mathbf{w})}{\big(\chi_k(\mathbf{w}) + \phi_k\big)^2} &= \frac{\chi_k^2(\mathbf{w})}{\chi_k(\mathbf{w}) + \phi_k} \frac{1}{\chi_k(\mathbf{w}) + \phi_k} \\ &\geq \frac{\chi_k^2(\mathbf{w})}{\chi_k(\mathbf{w}) + \phi_k}\Big(\frac{2}{\chi_k(\mathbf{w}^{(\kappa)}) + \phi_k^{(\kappa)}} - \frac{\chi_k(\mathbf{w}) + \phi_k}{(\chi_k(\mathbf{w}^{(\kappa)}) + \phi_k^{(\kappa)})^2}\Big) \\ &= \frac{2}{\chi_k(\mathbf{w}^{(\kappa)}) + \phi_k^{(\kappa)}} \frac{\chi_k^2(\mathbf{w})}{\chi_k(\mathbf{w}) + \phi_k} - \frac{\chi_k^2(\mathbf{w})}{(\chi_k(\mathbf{w}^{(\kappa)}) + \phi_k^{(\kappa)})^2} \end{aligned} \tag{30}$$

under the condition:

$$\chi_k(\mathbf{w}) + \phi_k \leq 2(\chi_k(\mathbf{w}^{(\kappa)}) + \phi_k^{(\kappa)}). \tag{31}$$

We now apply the inequality [31, Eq. (21)] to approximate $\chi_k^2(\mathbf{w})/(\chi_k(\mathbf{w}) + \phi_k)$ as

$$\begin{aligned} \frac{\chi_k^2(\mathbf{w})}{\chi_k(\mathbf{w}) + \phi_k} \;\geq\; & \frac{2\chi_k(\mathbf{w}^{(\kappa)})}{\chi_k(\mathbf{w}^{(\kappa)}) + \phi_k^{(\kappa)}} \chi_k(\mathbf{w}) \\ & - \frac{\chi_k^2(\mathbf{w}^{(\kappa)})}{(\chi_k(\mathbf{w}^{(\kappa)}) + \phi_k^{(\kappa)})^2}(\chi_k(\mathbf{w}) + \phi_k) \end{aligned} \tag{32}$$

$$\begin{aligned} \geq \; & \frac{2\chi_k(\mathbf{w}^{(\kappa)})}{\chi_k(\mathbf{w}^{(\kappa)}) + \phi_k^{(\kappa)}} \mathcal{L}_k^{(\kappa)}(\mathbf{w}) \\ & - \frac{\chi_k^2(\mathbf{w}^{(\kappa)})}{(\chi_k(\mathbf{w}^{(\kappa)}) + \phi_k^{(\kappa)})^2}(\chi_k(\mathbf{w}) + \phi_k) \end{aligned} \tag{33}$$

under the condition:

$$\chi_k(\mathbf{w}) + \phi_k \leq 2\chi_k(\mathbf{w}) \frac{\chi_k(\mathbf{w}^{(\kappa)}) + \phi_k^{(\kappa)}}{\chi_k(\mathbf{w}^{(\kappa)})} \tag{34}$$

where $\mathcal{L}_k^{(\kappa)}(\mathbf{w})$ is the first-order Taylor approximation of $\chi_k(\mathbf{w})$:

$$\begin{aligned} \chi_k(\mathbf{w}) \geq \mathcal{L}_k^{(\kappa)}(\mathbf{w}) := & \sum_{j \in \mathcal{K} \setminus \{k\}} 2\Re\{(\mathbf{w}_j^{(\kappa)})^H \hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^H \mathbf{w}_j\} \\ & + \sum_{j \in \mathcal{K} \setminus \{k\}} 2\xi_k \Re\{(\mathbf{w}_j^{(\kappa)})^H \mathbf{w}_j\} - \sum_{j \in \mathcal{K} \setminus \{k\}} |\hat{\mathbf{h}}_k^H \mathbf{w}_j^{(\kappa)}|^2 \\ & - \sum_{j \in \mathcal{K} \setminus \{k\}} \xi_k \|\mathbf{w}_j^{(\kappa)}\|^2 + \sigma^2. \end{aligned} \tag{35}$$

Substituting (30) and (33) into (26), we have

$$\begin{aligned} G_k(\mathbf{w}, \phi_k) \;\leq \; & \mathtt{a}_k - \mathtt{b}_k \Big(\frac{4\chi_k(\mathbf{w}^{(\kappa)})}{(\chi_k(\mathbf{w}^{(\kappa)}) + \phi_k^{(\kappa)})^2} \mathcal{L}_k^{(\kappa)}(\mathbf{w}) \\ & - \frac{2\chi_k^2(\mathbf{w}^{(\kappa)})}{(\chi_k(\mathbf{w}^{(\kappa)}) + \phi_k^{(\kappa)})^3}(\chi_k(\mathbf{w}) + \phi_k) \\ & - \frac{\chi_k^2(\mathbf{w})}{(\chi_k(\mathbf{w}^{(\kappa)}) + \phi_k^{(\kappa)})^2} \Big) \\ := \; & \mathcal{G}_k^{(\kappa)}(\mathbf{w}, \phi_k) \end{aligned} \tag{36}$$

where $\mathcal{G}_k^{(\kappa)}(\mathbf{w}, \phi_k)$ is a convex function which satisfies

$$G_k(\mathbf{w}^{(\kappa)}, \phi_k^{(\kappa)}) = \mathcal{G}_k^{(\kappa)}(\mathbf{w}^{(\kappa)}, \phi_k^{(\kappa)}). \tag{37}$$

As a result, the approximate concave of $R_k(\mathbf{w})$, denoted by $\mathcal{R}_k^{(\kappa)}(\mathbf{w}, \varphi_k, \phi_k)$, is given as

$$\mathcal{R}_k^{(\kappa)}(\mathbf{w}, \varphi_k, \phi_k) = B\Big(\mathcal{F}_k(\mathbf{w}) - \mathcal{G}_k(\mathbf{w})\frac{Q^{-1}(\epsilon_k)}{\sqrt{\tau B}}\Big). \tag{38}$$

We also iteratively replace (39) by the following convex constraint:

$$|\hat{\mathbf{h}}_k^H \mathbf{w}_k|^2 \le \phi_k + \xi_k\big(2\Re\{(\mathbf{w}_k^{(\kappa)})^H \mathbf{w}_k\} - \|\mathbf{w}_k^{(\kappa)}\|^2\big). \tag{39}$$

In summary, we solve the following approximate convex program of (10) at iteration $\kappa + 1$:

$$\begin{aligned}
\underset{\mathbf{w},\mathbf{r},\rho,\boldsymbol{\varphi},\boldsymbol{\phi}}{\text{maximize}} \quad & \sum_{k\in\mathcal{K}} r_k - \psi\rho && \text{(40a)}\\
\text{subject to} \quad & \text{(10b), (10d), (11d), (11e)} && \text{(40b)}\\
& \Phi_k^{(\kappa)}(\mathbf{w}_k) \ge \varphi_k^2 + \xi_k\|\mathbf{w}_k\|^2, \quad \forall k && \text{(40c)}\\
& \chi_k(\mathbf{w}) + \phi_k \le 2(\chi_k(\mathbf{w}^{(\kappa)}) + \phi_k^{(\kappa)}), \quad \forall k && \text{(40d)}\\
& \chi_k(\mathbf{w}) + \phi_k \le 2\chi_k(\mathbf{w})\frac{\chi_k(\mathbf{w}^{(\kappa)}) + \phi_k^{(\kappa)}}{\chi_k(\mathbf{w}^{(\kappa)})}, \quad \forall k && \text{(40e)}\\
& |\hat{\mathbf{h}}_k^H \mathbf{w}_k|^2 \le \phi_k + \xi_k\big(2\Re\{(\mathbf{w}_k^{(\kappa)})^H \mathbf{w}_k\} \\
& \quad - \|\mathbf{w}_k^{(\kappa)}\|^2\big), \quad \forall k && \text{(40f)}\\
& \mathcal{R}_k^{(\kappa)}(\mathbf{w}, \varphi_k, \phi_k) \ge r_k, \quad \forall k. && \text{(40g)}
\end{aligned}$$

The solution obtained by solving (40) is updated as the feasible point for next iteration. This procedure is repeated until convergence. The IA-based iterative algorithm often requires an initial feasible point to ensure that the convex optimization problem (40) is successfully solved at the first iterations. To doing so, we randomly generate $(\mathbf{w}^{(0)}, \boldsymbol{\varphi}^{(0)}, \boldsymbol{\phi}^{(0)})$ satisfying (10b) and (10d) and solve the following simple convex program:

$$\begin{aligned}
\underset{\mathbf{w},\mathbf{r},\rho,\boldsymbol{\varphi},\boldsymbol{\phi},\varrho}{\text{maximize}} \quad & \varrho && \text{(41a)}\\
\text{subject to} \quad & \text{(10b), (10d), (11e), (40c), (40d),} \\
& \text{(40e), (40f), (40g)} && \text{(41b)}\\
& r_k - \mathrm{R}_k^{\min} \ge \varrho, \quad \forall k && \text{(41c)}
\end{aligned}$$

where ϱ is a slack variable. The initial feasible point $(\mathbf{w}^{(0)}, \boldsymbol{\varphi}^{(0)}, \boldsymbol{\phi}^{(0)})$ is found by successively solving (41) until reaching $\varrho \ge 0$. The proposed iterative algorithm to solve the EEmax problem (10) is summarized in Algorithm 1.

C. CONVERGENCE AND COMPLEXITY ANALYSIS

The iterative Algorithm 1 is mainly based on the principle of IA framework, where its convergence to a local optimal solution was investigated in [33]. Let $\mathrm{EE}^{(\kappa)}$ denote the EE value obtained at iteration κ. Due to (18), (19), (36) and (37), we can show that Algorithm 1 produces a non-decreasing sequence of EE values, i.e., $\mathrm{EE}^{(\kappa+1)} \ge \mathrm{EE}^{(\kappa)}$, which is bounded above due to constraints (10b) and (10d). Moreover,

Algorithm 1 Proposed IA-Based Iterative Algorithm to Solve EEmax Problem (10)

Initialization: Set $\kappa := 0$ and generate an initial feasible point $(\mathbf{w}^{(0)}, \boldsymbol{\varphi}^{(0)}, \boldsymbol{\phi}^{(0)})$ by successively solving (41);
1: **repeat**
2: Solve (40) to obtain the solution $(\mathbf{w}^\star, \mathbf{r}^\star, \rho^\star, \boldsymbol{\varphi}^\star, \boldsymbol{\phi}^\star)$;
3: Update $(\mathbf{w}^{(\kappa+1)}, \boldsymbol{\varphi}^{(\kappa+1)}, \boldsymbol{\phi}^{(\kappa+1)}) := (\mathbf{w}^\star, \boldsymbol{\varphi}^\star, \boldsymbol{\phi}^\star)$;
4: Update $\psi^{(\kappa+1)} = \sum_{k\in\mathcal{K}} r_k^\star/\rho^\star$;
5: Set $\kappa := \kappa + 1$;
6: **until** Convergence
7: **Output:** the optimal solution $\mathbf{w}^\star$.

the feasible set of (40) is convex and connected. After a finite number of iterations, Algorithm 1 is ensured to achieve at least at a local optimal solution, which is shown to satisfy the Karush-Kuhn-Tucker conditions of (12) (and hence (10)) [23, Theorem 1].

The convex program (40) includes $KN + 3K + 1$ scalar optimization variables and $6K + M + 2$ linear and second-order cone (SOC) constraints. As a result, the worst-case of per-iteration complexity of Algorithm 1 using the interior-point method is estimated as $\mathcal{O}\big(\sqrt{6K + M}(KN + 3K)^3\big)$ [34, Chap. 6].

IV. PROPOSED ITERATIVE ALGORITHM FOR JDBAS-EEmax

In practice, ST can be equipped with N_{RF} radio frequency (RF) chains with $N_{\mathrm{RF}} \le N$, which aims at reducing hardware complexity. In other words, the number of active antennas is less than the number of RF chains. This design can potentially offer higher spatial diversity gains, minimize the interference power at PRs and reduce the power consumption at ST, resulting in better EE performance.

To find the optimal subset of active transmit antennas at ST, we introduce $\pi_n \in \{0, 1\}$ to indicate the operation mode (ON/OFF) of the n-th antenna, i.e.,

$$\pi_n = \begin{cases} 1, & \text{if the } n\text{-th antenna is activated,} \quad \text{(42a)}\\ 0, & \text{otherwise.} \quad \text{(42b)} \end{cases}$$

We define $\boldsymbol{\pi} \triangleq [\pi_1 \cdots, \pi_N]$ as the state vector of all transmit antennas. Similar to [35], let us denote the beamforming weights of all SRs involved with the n-th antenna at ST by $\bar{\mathbf{w}}_n \triangleq \big[[\mathbf{w}_1]_n, \cdots, [\mathbf{w}_K]_n\big]^T$, where $[\mathbf{w}_k]_n$ denotes the n-th element of $\mathbf{w}_k$. To facilitate the optimization design, we consider the following convex constraint:

$$\|\bar{\mathbf{w}}_n\|^2 \le \omega_n, \quad \forall n \tag{43}$$

where $\boldsymbol{\omega} \triangleq \{\omega_n\}_{\forall n}$ are newly introduced variables. The n-th antenna is inactive if $\pi_n = 0$ and $\bar{\mathbf{w}}_n = 0$, which implies that

$$\omega_n \le \pi_n P_{\max}, \quad \forall n. \tag{44}$$

The total power consumption $P_{\mathrm{tot}}(\mathbf{w})$ in (6) is rewritten as

$$P_{\mathrm{tot}}(\boldsymbol{\omega}, \boldsymbol{\pi}) = \frac{1}{\mu}\sum_{n\in\mathcal{N}} \omega_n + P_{\mathrm{dyn}}\sum_{n\in\mathcal{N}} \pi_n + P_{\mathrm{sta}}. \tag{45}$$

Based on (10) and the discussions above, the joint beamforming and transmit antenna selection for EE maximization problem (called JDBAS-EEmax for short) is stated as

$$
\begin{aligned}
&(\text{JDBAS} - \text{EEmax}): \\
&\underset{\mathbf{w},\boldsymbol{\omega},\boldsymbol{\pi}}{\text{maximize}} \quad \mathsf{EE}(\mathbf{w},\boldsymbol{\omega},\boldsymbol{\pi}) \triangleq \frac{\sum_{k\in\mathcal{K}} R_k(\mathbf{w})}{P_{\text{tot}}(\boldsymbol{\omega},\boldsymbol{\pi})} && (46a)\\
&\text{subject to} \quad (10b),\ (10c) && (46b)\\
&\qquad \|\bar{\mathbf{w}}_n\|^2 \le \omega_n, \quad \forall n && (46c)\\
&\qquad \omega_n \le \pi_n P_{\max}, \quad \forall n && (46d)\\
&\qquad \sum_{n\in\mathcal{N}} \omega_n \le P_{\max} && (46e)\\
&\qquad 1 \le \sum_{n\in\mathcal{N}} \pi_n \le N_{\text{RF}} && (46f)\\
&\qquad \pi_n \in \{0,1\}, \quad \forall n && (46g)
\end{aligned}
$$

where constraint (46f) is added to guarantee the minimum and maximum number of active antennas. It is clear that the JDBAS-EEmax problem (46) is a mixed-integer nonconvex problem, which is even more challenging to solve, compared to the EEmax problem (10).

A. BRUTE-FORCE SEARCH-BASED ITERATIVE ALGORITHM

A simple way to solve (46) is based on the Brute-Force Search (BFS) method, which is done by generating all possible subsets of $\boldsymbol{\pi}$ and strictly satisfying the conditions in (46f), and then reuse Algorithm 1 to solve each subproblem corresponding to each set of $\boldsymbol{\pi}$. The final optimal solution is the solution of subproblem having the highest EE. We note that this method requires to solve $\sum_{n=1}^{N_{\text{RF}}} \frac{N!}{n!(N-n)!}$ possible subproblems, resulting in the extremely high computational complexity. As a result, the EE performance obtained by the BFS-based method can be used as a benchmarking purpose (i.e., an upper bound) for other suboptimal schemes.

The JDBAS-EEmax subproblem of (46) for a given $\boldsymbol{\pi}$ is expressed as

$$
\begin{aligned}
&\underset{\mathbf{w},\boldsymbol{\omega}}{\text{maximize}} \quad \mathsf{EE}(\mathbf{w},\boldsymbol{\omega}|\boldsymbol{\pi}) \triangleq \frac{\sum_{k\in\mathcal{K}} R_k(\mathbf{w})}{P_{\text{tot}}(\boldsymbol{\omega}|\boldsymbol{\pi})} && (47a)\\
&\text{subject to} \quad (10b),\ (10c),\ (46c),\ (46d),\ (46e),\ (46f). && (47b)
\end{aligned}
$$

We can see that problem (47) has the similar structure with (10), while newly constraints (46c), (46d), (46e) and (46f) are linear and quadratic constraints. Therefore, we can directly reuse the approximate convex functions presented in Section III to convexify (47). In particular, we solve the following approximate convex program of (47) at iteration $\kappa+1$:

$$
\begin{aligned}
&\underset{\mathbf{w},\boldsymbol{\omega},\mathbf{r},\rho,\boldsymbol{\varphi},\boldsymbol{\phi}}{\text{maximize}} \quad \sum_{k\in\mathcal{K}} r_k - \psi\rho && (48a)\\
&\text{subject to} \quad (10b),\ (11d),\ (40c)-(40g),(46c)-\ (46f) && (48b)\\
&\qquad P_{\text{tot}}(\boldsymbol{\omega}|\boldsymbol{\pi}) \le \rho && (48c)
\end{aligned}
$$

which has $KN+3K+N+1$ scalar optimization variables and $6K+M+2N+2$ linear and second-order cone (SOC) constraints. We summarize the BFS-based iterative algorithm to solve the JDBAS-EEmax problem (46) in Algorithm 2. The worst-case of per-iteration complexity of Algorithm 2 is thus $\sum_{n=1}^{N_{\text{RF}}} n!\frac{N!}{(N-n)!} \times \mathcal{O}\big(\sqrt{6K+M+2N}(KN+3K+N)^3\big)$, which is seen much higher than that of Algorithm 1.

Algorithm 2 Proposed BFS-Based Iterative Algorithm to Solve JDBAS-EEmax Problem (46)

1: **Initialization:** Set $\mathsf{EE}^\star = 0$;
2: **for** each given set of $\boldsymbol{\pi}$ **do** {solving subproblem (47)}
3: **Initialization:** Set $\kappa := 0$ and generate an initial feasible point $(\mathbf{w}^{(0)}, \boldsymbol{\varphi}^{(0)}, \boldsymbol{\phi}^{(0)})$ for each subproblem;
4: **repeat**
5: Solve (48) to obtain the solution $(\mathbf{w}^\star, \boldsymbol{\omega}^\star, \mathbf{r}^\star, \rho^\star, \boldsymbol{\varphi}^\star, \boldsymbol{\phi}^\star)$;
6: Update $(\mathbf{w}^{(\kappa+1)}, \boldsymbol{\varphi}^{(\kappa+1)}, \boldsymbol{\phi}^{(\kappa+1)}) := (\mathbf{w}^\star, \boldsymbol{\varphi}^\star, \boldsymbol{\phi}^\star)$ and compute $\mathsf{EE}(\mathbf{w}^{(\kappa+1)}, \boldsymbol{\omega}^{(\kappa+1)}|\boldsymbol{\pi})$;
7: Update $\psi^{(\kappa+1)} = \sum_{k\in\mathcal{K}} r_k^\star/\rho^\star$;
8: Set $\kappa := \kappa+1$;
9: **until** Convergence
10: **if** $\mathsf{EE}(\mathbf{w}^{(\kappa)}, \boldsymbol{\omega}^{(\kappa)}|\boldsymbol{\pi}) > \mathsf{EE}^\star$ **then**
11: Set $\mathsf{EE}^\star := \mathsf{EE}(\mathbf{w}^{(\kappa)}, \boldsymbol{\omega}^{(\kappa)}|\boldsymbol{\pi})$ and $(\mathbf{w}^*, \boldsymbol{\pi}^*, \boldsymbol{\omega}^*) := (\mathbf{w}^{(\kappa)}, \boldsymbol{\pi}^{(\kappa)}, \boldsymbol{\omega}^{(\kappa)})$ as the newly updated optimal solution;
12: **end if**
13: **end for**
14: **Output:** The optimal solution $(\mathbf{w}^\star, \boldsymbol{\pi}^\star, \boldsymbol{\omega}^\star)$.

B. IA-CR BASED ITERATIVE ALGORITHM

We now focus on developing a suboptimal yet computationally efficient solution for (46), which aims at jointly optimizing both continuous and binary variables in a single layer. As a standard way to solve a mixed-integer nonconvex problem [35], [36], we relax binary variables to be continuous as $\pi_n \in [0,1], \forall n$. The continuous relaxation (CR) of the JDBAS-EEmax problem (46) is expressed as

$$
\begin{aligned}
&\underset{\mathbf{w},\boldsymbol{\omega},\boldsymbol{\pi}}{\text{maximize}} \quad \mathsf{EE}(\mathbf{w},\boldsymbol{\omega},\boldsymbol{\pi}) \triangleq \frac{\sum_{k\in\mathcal{K}} R_k(\mathbf{w})}{P_{\text{tot}}(\boldsymbol{\omega},\boldsymbol{\pi})} && (49a)\\
&\text{subject to} \quad (10b),\ (10c),\ (46c),\ (46d),\ (46e),\ (46f)(49b)\\
&\qquad 0 \le \pi_n \le 1, \quad \forall n. && (49c)
\end{aligned}
$$

From the relaxed constraint (49c), we can see that any feasible point of (49) is also feasible to (46), but not vice versa. Inspired by [36], we incorporate a penalty function into the objective function to obtain near exact binary values of (49) at optimum.

For any $\pi_n \in [0,1]$, we have $\pi_n \ge \pi_n^2$. The equality holds (i.e., $\pi_n = \pi_n^2$) only if $\pi_n = \{0,1\}$. Constraint (46g) can be equivalently expressed as

$$
(46g) \Leftrightarrow \pi_n = \{0,1\}\ \&\ \pi_n \le \pi_n^2, \quad \forall n. \tag{50}
$$

Therefore, to enforce $\pi_n = \pi_n^2$ at optimum, we consider the penalty function $\Upsilon(\boldsymbol{\pi}) = \sum_{n\in\mathcal{N}}(\pi_n^2 - \pi_n)$, which is always non-positive and can be used to guarantee the satisfaction in (50). The parameterized CR problem, which incorporates the penalty function $\Upsilon(\boldsymbol{\pi})$ into the objective function, is expressed as

$$
\underset{\mathbf{w},\boldsymbol{\omega},\boldsymbol{\pi}}{\text{maximize}} \quad \mathsf{EE}(\mathbf{w},\boldsymbol{\omega},\boldsymbol{\pi}) \triangleq \frac{\sum_{k\in\mathcal{K}} R_k(\mathbf{w})}{P_{\text{tot}}(\boldsymbol{\omega},\boldsymbol{\pi})} + \beta\Upsilon(\boldsymbol{\pi}) \tag{51a}
$$

Algorithm 3 Proposed IA-CR-Based Iterative Algorithm to Solve JDBAS-EEmax Problem (46)

Initialization: Set $\kappa := 0$ and generate an initial feasible point $(\mathbf{w}^{(0)}, \boldsymbol{\pi}^{(0)}, \boldsymbol{\varphi}^{(0)}, \boldsymbol{\phi}^{(0)})$;
1: **repeat**
2: Solve (53) to obtain the solution $(\mathbf{w}^\star, \boldsymbol{\pi}^\star, \boldsymbol{\omega}^\star, \mathbf{r}^\star, \rho^\star, \boldsymbol{\varphi}^\star, \boldsymbol{\phi}^\star)$;
3: Update $(\mathbf{w}^{(\kappa+1)}, \boldsymbol{\pi}^{(\kappa+1)}, \boldsymbol{\varphi}^{(\kappa+1)}, \boldsymbol{\phi}^{(\kappa+1)}) := (\mathbf{w}^\star, \boldsymbol{\pi}^\star, \boldsymbol{\varphi}^\star, \boldsymbol{\phi}^\star)$;
4: Update $\psi^{(\kappa+1)} = \sum_{k\in\mathcal{K}} r_k^\star/\rho^\star$;
5: Set $\kappa := \kappa + 1$;
6: **until** Convergence
7: Recover exact binary values as: $\pi_n^\star = \lfloor \pi_n^{(\kappa)} + 0.5 \rfloor, \forall n$;
8: Repeat Steps 1-6 with a given $\boldsymbol{\pi}^\star$ to find the exact solution of $(\mathbf{w}^\star, \boldsymbol{\omega}^\star)$;
9: **Output:** The optimal solution $(\mathbf{w}^\star, \boldsymbol{\pi}^\star, \boldsymbol{\omega}^\star)$.

subject to (10b), (10c), (46c), (46d), (46e), (46f), (49c) (51b)

where β is the constant penalty parameter. β should be a positive value and sufficiently large to guarantee the optimality of (51) and enforce $\Upsilon(\boldsymbol{\pi})$ close to zeros. The function $\Upsilon(\boldsymbol{\pi})$ is convex and can be directly approximated by IA method. Denoting by $\Upsilon^{(\kappa)}(\boldsymbol{\pi})$ the approximate function of $\Upsilon(\boldsymbol{\pi})$, we have

$$\begin{aligned}\Upsilon(\boldsymbol{\pi}) &\geq \Upsilon^{(\kappa)}(\boldsymbol{\pi}) \\ &:= \sum_{n\in\mathcal{N}} \Big(2\pi_n^{(\kappa)}\pi_n - (\pi_n^{(\kappa)})^2 - \pi_n\Big). \end{aligned} \tag{52}$$

The approximate convex program of (51) solved at iteration $\kappa + 1$ is given as

$$\underset{\mathbf{w},\boldsymbol{\omega},\boldsymbol{\pi},\mathbf{r},\rho,\boldsymbol{\varphi},\boldsymbol{\phi}}{\text{maximize}} \quad \sum_{k\in\mathcal{K}} r_k - \psi\rho + \beta\Upsilon^{(\kappa)}(\boldsymbol{\pi}) \tag{53a}$$

$$\text{subject to} \quad (10b),\ (11d),\ (40c) - (40g),\ (46c) - (46f),\ (49c) \tag{53b}$$

$$P_{\text{tot}}(\boldsymbol{\omega}, \boldsymbol{\pi}) \leq \rho \tag{53c}$$

where its per-iteration complexity is $\mathcal{O}\big(\sqrt{6K + M + N}(KN + 3K)^3\big)$, which is seen much lower than that of the BFS method presented in Algorithm 2. By numerical results, we have observed that there may exist some optimal solutions of $\boldsymbol{\pi}$ which are close but not exact binary values. To overcome this issue, we use the following floor function

$$\pi_n^\star = \lfloor \pi_n^{(\kappa)} + 0.5 \rfloor, \quad \forall n \tag{54}$$

and then re-optimize over other parameters to refine the optimal solution. The overall iterative algorithm to solve (46) based on IA-CR is summarized in Algorithm 3.

V. NUMERICAL RESULTS AND DISCUSSIONS

In this section, we use computer simulations to evaluate the performance of the proposed algorithms. All SRs and PRs are randomly placed within a ring with the outer radius of 100 meters and the inner radius of 10 meters, while ST is placed in the centered cell. The path-loss (PL) exponent of all the channel vectors is set to PL $= 3$, while the small-scale fading follows Rician fading with the Rician factor of 10 dB [28]. The decoding error probability and the duration for one TTI in URLLC are set to $\epsilon_k = 10^{-5}, \forall k$ and $\tau = 0.05$ ms [37], respectively. For simplicity, the maximum allowable interference power and minimum predetermined rate requirement are set to be identical to all PRs and SRs, respectively. Due to the non-cooperation between two systems, we set $\delta_p > \delta_s$. Unless stated otherwise, other system parameters are given in Table 1 for ease of cross-reference, which mainly follow studies in [28], [35].

TABLE 1. Simulation parameters.

Parameter	Value
System bandwidth, B	10 MHz
Noise power at SRs, σ^2	-70 dBm
Predetermined rate requirement, $\mathrm{R}^{\min} \equiv \mathrm{R}_k^{\min}, \forall k$	1 Mbits/s
Maximum allowable interference power at PRs, $\mathcal{I} \equiv \mathcal{I}_m, \forall m$	-20 dBm
Power budget at ST, $P_{\max}$	26 dBm
Number of transmit antennas at ST, N	8
Number of SRs, K	4
Number of PRs, M	2
Normalized channel uncertainties of SRs, δ_s	10^{-3}
Normalized channel uncertainties of PRs, δ_p	10^{-2}
Power amplifier efficiency, μ	0.69
Dynamic power consumption, P_{dyn}	42 dBm
Static power consumption, P_{sta}	33 dBm

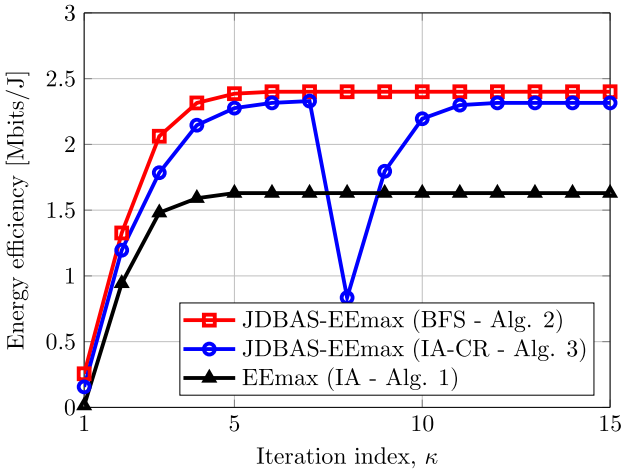

FIGURE 2. Convergence behavior of the proposed algorithms.

In the following, the EE results are obtained by averaging over 1000 simulation runs. The proposed algorithms are stopped if the error tolerance between any two consecutive iterations is smaller than 10^{-3}. The achieved EE performance given in nats/J is divided by ln(2) to have at unit of bits/J. The convex problems are solved by SeDuMi solver with the modelling toolbox YALMIP [38] in MATLAB environment.

In Fig. 2, we depict the convergence behavior of Algorithm 1 for the EEmax problem and Algorithms 2 and 3 for the JFBAS-EEmax problem over one random channel realization. We note that for Algorithm 2, Fig. 2 only shows the convergence for solving one subproblem (47) with the highest EE. As can be seen from Fig. 2 that all algorithms

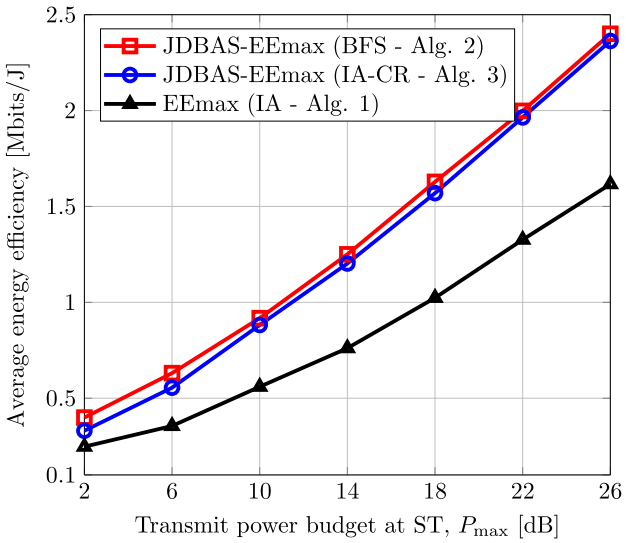

FIGURE 3. Average energy efficiency versus the transmit power budget at ST, $P_{\max} \in [0, 26]$ dBm.

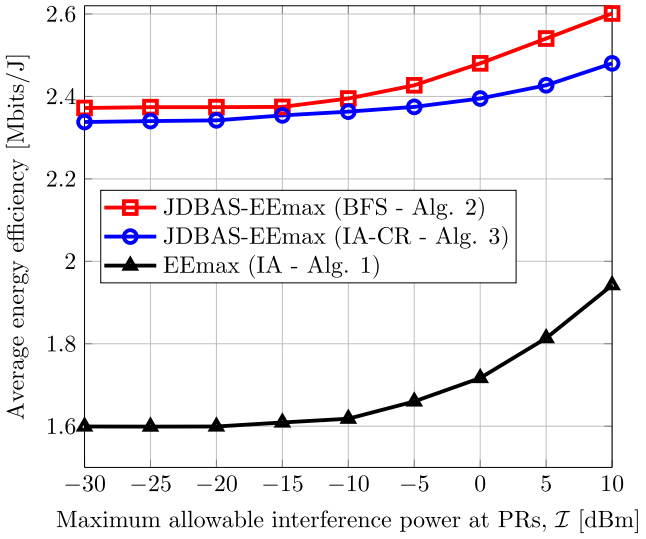

FIGURE 4. Average energy efficiency versus the maximum allowable interference power at PRs, $\mathcal{I} \in [-30, 10]$ dBm.

converge within a few iterations, and the EE is monotonically increased after every iteration. The EE of Algorithm 3 decreases at iteration 8 due to the binary recovery in Step 7, which is to guarantee the feasible solution to the JDBAS-EEmax problem. This also confirms the importance of the second phase (i.e., Step 8) to re-optimize other parameters $(\mathbf{w}, \boldsymbol{\omega})$. We remark that although Algorithm 2 with the BFS method provides the best EE, its performance only acts as an upper bound of the considered system due to the extremely high computational complexity required to solve a large number of subproblems.

Fig. 4 illustrates the average EE versus the transmit power budget at ST, $P_{\max}$, for three resource allocation schemes. We can see that the JDBAS-based schemes (Algorithms 2 and 3) provide significant performance gains over the traditional beamforming scheme (Algorithm 1) in terms of achievable EE, which clearly confirms the effectiveness of joint design of beamforming and antenna selection. In addition, we can see that the impact of antenna selection is even more significant when $P_{\max}$ is large. This is due to the fact that, for large $P_{\max}$, ST will allocate more power to the best antennas to reduce the total power consumption. Moreover, the loss in EE performance of the proposed IA-CR method compared to the optimal BFS is quite small, i.e., about 0.04 Mbits/J at the typical value of $P_{\max} = 26$ dBm.

Fig. 4 studies the impact of the maximum allowable interference power $\mathcal{I}$ on the achievable EE. As seen, the higher the maximum allowable interference power, the better the EE of all the considered schemes can be obtained. When $\mathcal{I}$ is extremely small (e.g., $\mathcal{I} < -15$ dBm), the EE is mostly unchanged. In this case, the ST must reduce its transmit power or design an effective beamforming to avoid or even cancel strong interference power at PRs.

Next, Fig. 5 plots the average EE of the secondary system as a function the number of PRs, M. As expected, the average

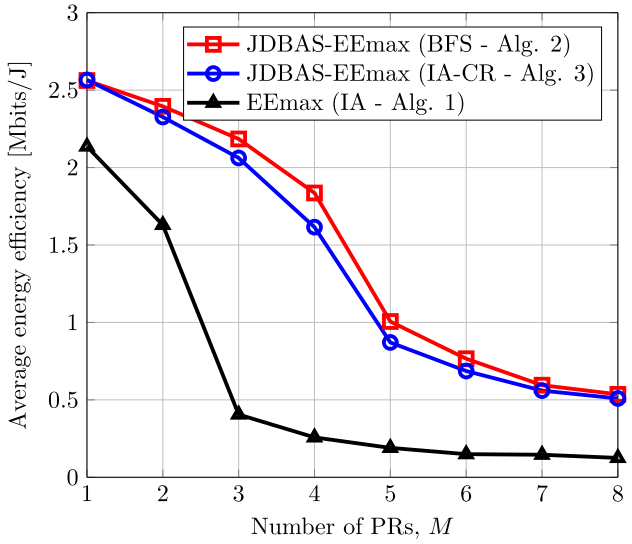

FIGURE 5. Average energy efficiency versus the number of PRs, M.

EE of the considered resource allocation schemes is degraded when M increases. The reason is that ST needs to scale down its transmission power to satisfy interference power constraints in (10b), resulting in lower EE. Another interesting observation is that the EE tends to saturate when M becomes large. In this case, ST is likely to avoid transmitting its downlink signals over the spatial space of PRs. We also see that the EE gain of joint design of beamforming and antenna selection over the traditional beamforming is smaller for a large number of PRs, as the possible optimal sets of active antennas are reduced significantly.

Finally, Fig. 6 shows the average EE versus the number of transmit antennas at ST, $N \in [4, 10]$. For small N, three resource allocation schemes provide a similar EE performance since all antennas should be active to achieve

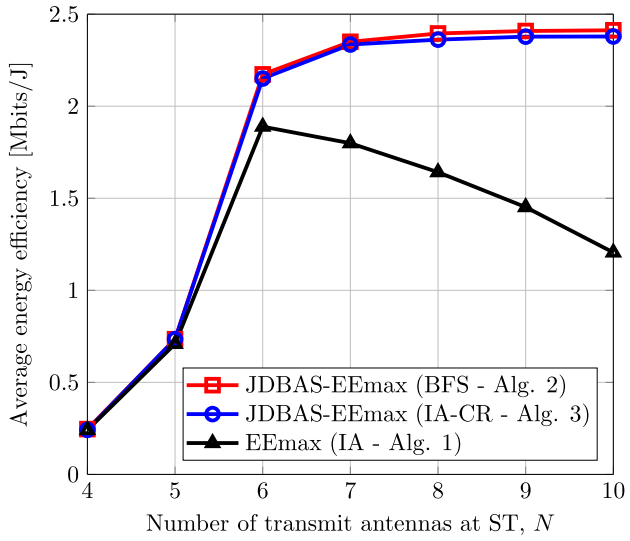

FIGURE 6. Average energy efficiency versus the number of transmit antennas at ST, N.

maximum EE. For large $N \geq 6$, joint beamforming and antenna selection offers a great EE improvement over the traditional beamforming scheme. In addition, the EE of the traditional beamforming scheme is dramatically degraded when N is larger than a certain threshold. The reason is that the total power consumption in (6) increases linearly with N, and the dynamic power consumption (i.e., NP_{dyn}) will dominate the power consumed for the transmitted data (i.e., $\frac{1}{\mu}\sum_{k\in\mathcal{K}}\|\mathbf{w}_k\|^2$). This again confirms the effectiveness of the proposed JDBAS design in (46).

VI. CONCLUSION

We have considered the short blocklength regime for URLLC in downlink CRNs. We have studied the optimization problem of energy efficiency maximization for the secondary system, where two different transmission designs have been proposed. In the first design, we have focused on developing an effective beamforming design based on the inner approximation framework and the Dinkelbach approach. The second design focused on joint beamforming and antenna selection to further improve the energy efficiency, which is formulated as the mixed-integer nonconvex problem. We have proposed two iterative algorithms, namely BFS and IA-CR, to successfully solve this challenging problem. The former aims to find the optimal set of transmit antennas among all possible sets while the latter developed an iterative algorithm to effectively solve the problem in a single layer. The EE performance of the proposed algorithms has been analyzed through numerical results, which have confirmed that joint beamforming and antenna selection provides significant energy efficiency improvement, especially when the large number of transmit antennas at ST becomes large.

REFERENCES

[1] G. Hattab and M. Ibnkahla, "Multiband spectrum access: Great promises for future cognitive radio networks," *Proc. IEEE*, vol. 102, no. 3, pp. 282–306, Mar. 2014.

[2] S. Haykin, "Cognitive radio: Brain-empowered wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 2, pp. 201–220, Feb. 2005.

[3] X. Gan and B. Chen, "A novel sensing scheme for dynamic multichannel access," *IEEE Trans. Veh. Technol.*, vol. 61, no. 1, pp. 208–221, Jan. 2012.

[4] J. T. Wang, "Maximum–minimum throughput for MIMO systems in cognitive radio networks," *IEEE Trans. Veh. Technol.*, vol. 63, no. 1, pp. 217–224, Jan. 2014.

[5] L. Zhang, Y.-C. Liang, and Y. Xin, "Joint beamforming and power allocation for multiple access channels in cognitive radio networks," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 1, pp. 38–51, Jan. 2008.

[6] 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, Aug. 2016.

[7] *Study on Scenarios and Requirements for Next Generation Access Technologies*, document 38.913, 3GPP, 2017, pp. 38–51, vol. 26, no. 1.

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

[9] S. E. Elayoubi, P. Brown, M. Deghel, and A. Galindo-Serrano, "Radio resource allocation and retransmission schemes for URLLC over 5G networks," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 4, pp. 896–904, Apr. 2019.

[10] H. Yang, Z. Xiong, J. Zhao, D. Niyato, C. Yuen, and R. Deng, "Deep reinforcement learning based massive access management for ultra-reliable low-latency communications," *IEEE Trans. Wireless Commun.*, vol. 20, no. 5, pp. 2977–2990, May 2021.

[11] A. Anand and G. de Veciana, "Resource allocation and HARQ optimization for URLLC traffic in 5G wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 11, pp. 2411–2421, Nov. 2018.

[12] *Study on New Radio (NR) Access Technologies, Technical Specification Group Radio Access Network*, document 38.802, 3GPP, Mar. 2017.

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

[14] W. Yang, G. Durisi, T. Koch, and Y. Polyanskiy, "Quasi-static multiple-antenna fading channels at finite blocklength," *IEEE Trans. Inf. Theory*, vol. 60, no. 7, pp. 4232–4265, Jun. 2014.

[15] C. She, C. Yang, and T. Q. S. Quek, "Joint uplink and downlink resource configuration for ultra-reliable and low-latency communications," *IEEE Trans. Commun.*, vol. 66, no. 5, pp. 2266–2280, May 2018.

[16] C. She, C. Yang, and T. Q. S. Quek, "Cross-layer optimization for ultra-reliable and low-latency radio access networks," *IEEE Trans. Wireless Commun.*, vol. 17, no. 1, pp. 127–141, Jan. 2018.

[17] C. Sun, C. She, C. Yang, T. Q. S. Quek, Y. Li, and B. Vucetic, "Optimizing resource allocation in the short blocklength regime for ultra-reliable and low-latency communications," *IEEE Trans. Wireless Commun.*, vol. 18, no. 1, pp. 402–415, Jan. 2019.

[18] C. Sun and C. Yang, "Learning to optimize with unsupervised learning: Training deep neural networks for URLLC," in *Proc. IEEE 30th Annu. Int. Symp. Pers., Indoor Mobile Radio Commun. (PIMRC)*, Sep. 2019, pp. 1–7.

[19] B. Chang, L. Zhang, L. Li, G. Zhao, and Z. Chen, "Optimizing resource allocation in URLLC for real-time wireless control systems," *IEEE Trans. Veh. Tech.*, vol. 68, no. 9, pp. 8916–8927, Jul. 2019.

[20] S. R. Sabuj, A. Ahmed, Y. Cho, K.-J. Lee, and H.-S. Jo, "Cognitive UAV-aided URLLC and mMTC services: Analyzing energy efficiency and latency," *IEEE Access*, vol. 9, pp. 5011–5027, 2021.

[21] S.-F. Cheng, L.-C. Wang, C.-H. Hwang, J.-Y. Chen, and L.-Y. Cheng, "On-device cognitive spectrum allocation for coexisting URLLC and eMBB users in 5G systems," *IEEE Trans. Cognit. Commun. Netw.*, vol. 7, no. 1, pp. 171–183, Mar. 2021.

[22] W. Dinkelbach, "On nonlinear fractional programming," *Manage. Sci.*, vol. 13, no. 7, pp. 492–498, Mar. 1967.

[23] B. R. Marks and G. P. Wright, "A general inner approximation algorithm for nonconvex mathematical programs," *Oper. Res.*, vol. 26, pp. 681–683, Jul. 1978.

[24] V.-D. Nguyen, L.-N. Tran, T. Q. Duong, O.-S. Shin, and R. Farrell, "An efficient precoder design for multiuser MIMO cognitive radio networks with interference constraints," *IEEE Trans. Veh. Technol.*, vol. 66, no. 5, pp. 3991–4004, May 2017.

[25] V.-D. Nguyen, C. T. Nguyen, H. V. Nguyen, and O.-S. Shin, "Joint beamforming and antenna selection for sum rate maximization in cognitive radio networks," *IEEE Commun. Lett.*, vol. 21, no. 6, pp. 1369–1372, Jun. 2017.

[26] L. Zhang, Y. Xin, and Y. C. Liang, "Weighted sum rate optimization for cognitive radio MIMO broadcast channels," *IEEE Trans. Wireless Commun.*, vol. 8, no. 6, pp. 2950–2959, Jun. 2009.
[27] J. Huang and A. L. Swindlehurst, "Robust secure transmission in MISO channels based on worst-case optimization," *IEEE Trans. Signal Process.*, vol. 60, no. 4, pp. 1696–1707, Apr. 2012.
[28] V.-D. Nguyen and O.-S. Shin, "Cooperative prediction-and-sensing-based spectrum sharing in cognitive radio networks," *IEEE Trans. Cogn. Commun. Netw.*, vol. 4, no. 1, pp. 108–120, Mar. 2018.
[29] O. Arnold, F. Richter, G. Fettweis, and O. Blume, "Power consumption modeling of different base station types in heterogeneous cellular networks," in *Proc. 19th Future Netw. Mobile Summit*, Florence, Italy, Jun. 2010, pp. 1–8.
[30] U. Rashid, H. D. Tuan, and H. H. Nguyen, "Relay beamforming designs in multi-user wireless relay networks based on throughput maximin optimization," *IEEE Trans. Commun.*, vol. 61, no. 5, pp. 1739–1749, May 2013.
[31] V.-D. Nguyen, T. Q. Duong, H. D. Tuan, O.-S. Shin, and H. V. Poor, "Spectral and energy efficiencies in full-duplex wireless information and power transfer," *IEEE Trans. Commun.*, vol. 65, no. 5, pp. 2220–2233, May 2017.
[32] A. A. Nasir, H. D. Tuan, H. H. Nguyen, M. Debbah, and H. V. Poor, "Resource allocation and beamforming design in the short blocklength regime for URLLC," *IEEE Trans. Wireless Commun.*, vol. 20, no. 2, pp. 1321–1335, Feb. 2021.
[33] A. Beck, A. Ben-Tal, and L. Tetruashvili, "A sequential parametric convex approximation method with applications to nonconvex truss topology design problems," *J. Global Optim.*, vol. 47, no. 1, pp. 29–51, May 2010.
[34] A. Ben-Tal and A. Nemirovski, *Lectures on Modern Convex Optimization* (MPS-SIAM Series on Optimization). Philadelphia, PA, USA: SIAM, 2001.
[35] O. Tervo, L.-N. Tran, and M. Juntti, "Optimal energy-efficient transmit beamforming for multi-user MISO downlink," *IEEE Trans. Signal Process.*, vol. 63, no. 20, pp. 5574–5588, Oct. 2015.
[36] H. V. Nguyen, V. Nguyen, O. A. Dobre, D. N. Nguyen, E. Dutkiewicz, and O. Shin, "Joint power control and user association for NOMA-based full-duplex systems," *IEEE Trans. Commun. Mag.*, vol. 67, no. 11, pp. 8037–8055, Aug. 2019.
[37] J. Sachs, G. Wikstrom, T. Dudda, R. Baldemair, and K. Kittichokechai, "5G radio network design for ultra-reliable low-latency communication," *IEEE Netw.*, vol. 32, no. 2, pp. 24–31, Mar. 2018.
[38] J. Lofberg, "YALMIP: A toolbox for modeling and optimization in MATLAB," in *Proc. IEEE Int. Symp. Comput. Aided Control Syst. Design*, Sep. 2004, pp. 284–289.

NGO TAN VU KHANH received the B.A. degree in information technology from the Ho Chi Minh City University of Science, Ho Chi Minh City, Vietnam, in 2004, the M.S. degree in computer information system technology from the Asia Institute of Technology, Bangkok, Thailand, in 2008, and the Ph.D. degree in management information system from Soongsil University, Seoul, South Korea, in 2014.

In 2009, he was the Vice Head of the Software Technology Department, Tiên Giang University. He is currently an Information Design Technology Master Program Director at the University of Economics Ho Chi Minh City–UEH. He is in charge of all operations within the Information Design and Technology Master Program (IDT). He is also the one of experts in the field of information technology in Vietnam secured many country prizes, with over 15 years of managing experience in the field of computer and information systems in the largest technology companies, like Microsoft, Oracle, DXC Technology, and Kaspersky. In addition, he also successfully conveys and shares his practical experience with those who love and are passionate about emerging technologies. With this desire, he has been the Co-Founder of the FinTech Institute under the Ho Chi Minh City Computer Association. His current research interests include network security, emerging applications and AI-based solutions for Fintech, and the Internet of Things.

TIEN-TUNG NGUYEN received the B.Sc. and M.Sc. degrees from the University of Science Ho Chi Minh City, Vietnam, in 2005 and 2010, respectively, and the Ph.D. degree in electronics from Myongji University, South Korea, in 2021. Since 2011, he has been an Assistant Professor with the Industrial University of Ho Chi Minh City. His research interests include emerging topics of wireless communication for 5G & 6G, including energy harvesting, physical layer security, cognitive radio, non-orthogonal multiple access (NOMA), short-packet communications, the Internet of Things (IoT), and applications of optimization and machine learning for wireless communications. He has served as a reviewer for many IEEE journals and conferences.