Resource Allocation for MC-NOMA Systems with Cognitive Relaying

Yan Sun[*], Derrick Wing Kwan Ng[†], and Robert Schober[*]

[*]Institute for Digital Communications, Friedrich-Alexander-University Erlangen-Nürnberg (FAU), Germany
[†]School of Electrical Engineering and Telecommunications, The University of New South Wales, Australia

Abstract—In this paper, we investigate the resource allocation algorithm design for cooperative cognitive relaying multicarrier non-orthogonal multiple access (MC-NOMA) systems. In particular, the secondary base station serves multiple secondary users and simultaneously acts as a relay assisting the information transmission in the primary network. The resource allocation aims to maximize the weighted system throughput by jointly optimizing the power and subcarrier allocation for both the primary and the secondary networks while satisfying the quality-of-service requirements of the primary users. The algorithm design is formulated as a mixed combinatorial non-convex optimization problem. We apply monotonic optimization theory to solve the problem leading to an optimal resource allocation policy. Besides, we develop a low-complexity scheme to find a suboptimal solution. Our simulation results reveal that the performance of the proposed suboptimal algorithm closely approaches that of the optimal one. Besides, the combination of MC-NOMA and cognitive relaying improves the system throughput considerably compared to conventional multicarrier cognitive relaying systems.

I. Introduction

Spectrum has become a scarce resource due to the continuously growing demand for high-data rate communications. This has created a bottleneck for providing ubiquitous communication services. To handle this issue, cognitive radio (CR) has been proposed as a promising technique to improve spectrum utilization by enabling an unlicensed network (e.g. the secondary network) to dynamically access the licensed spectrum of the primary network [1]. However, the deployment of CR degrades the performance of the primary network due to the co-channel interference originating from the secondary network [2]. Recently, cooperative CR has attracted significant research interest since it can reduce the performance degradation to the primary network caused by CR deployment. In particular, in cooperative CR networks, the secondary base station (BS) acts as a relay to assist the signal transmission of the primary BS, while simultaneously utilizing the licensed spectrum to serve the secondary users (SUs). In [3], the authors studied the outage probability of cooperative CR relaying systems. The authors of [4] focused on precoder design for cooperative CR systems where a multiple-antenna secondary BS was deployed for suppressing the interference to primary users (PUs). However, since the primary network and the secondary network coexist in the same frequency band, there is a nontrivial tradeoff between the performance of the two networks due to the mutual interference.

Recently, non-orthogonal multiple access (NOMA) has been proposed to improve spectral efficiency by harnessing co-channel interference via superposition coding at the transmitter and successive interference cancellation (SIC) at the receiver [5]–[7]. Specifically, NOMA multiplexes the message of multiple users on the same time-frequency resource and

exploits the power domain for multiple access. Motivated by this, the application of NOMA in CR systems for improving the spectral efficiency was investigated in [8], where the secondary users were equipped with successive interference cancellers to cancel the interference generated by the primary network. Subsequently, the concept of NOMA was extended to cognitive relaying systems [9], [10]. For instance, in [9], the outage probability of a NOMA-based cognitive relaying system was investigated with the goal to satisfy the quality-of-service (QoS) requirements of the PU. The authors of [10] studied the performance of a cognitive NOMA system where a group of multicast SUs relayed information to a single PU. However, only fixed power allocation was adopted in [9], [10] to facilitate the performance analysis. However, power allocation is critical for improving the performance of CR systems and the optimal resource allocation design for NOMA with cognitive relaying is still unknown. Moreover, [9], [10] studied only single-carrier systems for serving a single PU, thus the optimal resource allocation design for multiuser cognitive relaying multicarrier NOMA (MC-NOMA) systems is still an open problem.

In this paper, we address the above issues. To this end, the resource allocation algorithm design for cognitive relaying MC-NOMA systems is formulated as a non-convex optimization problem for the maximization of the weighted system throughput. We solve the considered problem optimally via monotonic optimization theory [11], [12] and obtain the optimal power and subcarrier allocation policy. In addition, we also develop a low-complexity suboptimal scheme which achieves a close-to-optimal performance.

II. System Model

In this section, we present the adopted notation and the considered cognitive relaying MC-NOMA system model.

A. Notation

We use boldface capital and lower case letters to denote matrices and vectors, respectively. $\mathbf{a}^T$ denotes the transpose of vector $\mathbf{a}$; $\text{Tr}(\mathbf{A})$ denotes the trace of matrix $\mathbf{A}$; $\mathbb{C}$ denotes the set of complex numbers; $\mathbb{R}^+$ denotes the set of non-negative real numbers; $|\cdot|$ denotes the absolute value of a complex scalar; $\mathcal{E}\{\cdot\}$ denotes statistical expectation. The circularly symmetric complex Gaussian distribution with mean w and variance σ^2 is denoted by $\mathcal{CN}(w, \sigma^2)$; and $\sim$ stands for "distributed as". $\nabla_{\mathbf{x}} f(\mathbf{x})$ denotes the gradient vector of function $f(\mathbf{x})$ whose components are the partial derivatives of $f(\mathbf{x})$.

B. Cognitive Relaying MC-NOMA System Model

The considered cognitive relaying MC-NOMA system comprises one primary BS, K PUs, one secondary BS, and J
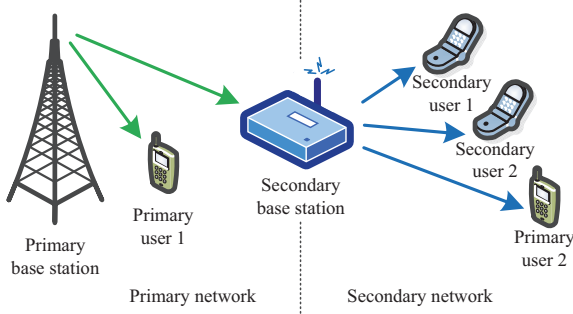
Fig. 1. A cognitive relaying system with a primary base station, $K = 2$ primary users, one secondary BS, and $J = 2$ secondary users.

SUs. All transceivers are single-antenna half-duplex (HD) devices, cf. Figure 1. The entire frequency band of W Hertz is partitioned into N_F orthogonal subcarriers. The duration of a time slot is T. The primary BS serves the PUs in the downlink (DL) in each scheduling slot. For the case that a PU is far away from the primary BS or has a poor channel to the primary BS, the secondary BS can act as a decode-and-forward (DF) relay [13] to assist the primary BS by forwarding the information received from the primary BS to the intended PU. For example, assuming that the secondary BS assists the transmission of PU k, the primary BS transmits the information of PU k to the secondary BS in the first half of a time slot on subcarrier i. Then, during the second half of the time slot, the secondary BS decodes the received signal on subcarrier i and forwards the recovered information to PU k on the same subcarrier, cf. Figure 2. Meanwhile, the secondary BS also serves the SUs via the subcarriers that are used for forwarding information to the PUs in the second half of the time slot. We assume that the secondary BS adopts NOMA to provide wireless service to the SUs by multiplexing the message of one secondary user (SU) and one PU on the same subcarrier. In addition, we assume that the SUs are equipped with successive interference cancellers for multiuser detection but the PUs are only equipped with linear receivers for single-user detection. Thus, if a PU is assisted by the secondary BS, it will be subject to the co-channel interference originating from the signal of the SU that is served on the same subcarrier.

Remark 1: We assume that the secondary BS can multiplex at most one SU and one PU on each subcarrier[1]. For the primary network, each subcarrier can only be allocated to at most one PU or to the secondary BS for information relaying[2].

C. Channel Model

For the primary network, the primary BS employs orthogonal multiple access which indicates that each subcarrier can either be allocated for transmitting the signal to one PU or to the secondary BS. Specifically, for a given time slot, if PU $k \in \{1, \ldots, K\}$ is scheduled on subcarrier $i \in \{1, \ldots, N_\mathrm{F}\}$, the received signal at PU k on subcarrier i is given by

$$x_k^i = \sqrt{q_k^i} f_k^i d_{\mathrm{PU}_k}^i + z_{\mathrm{PU}_k}^i, \qquad (1)$$

where $d_{\mathrm{PU}_k}^i \in \mathbb{C}$ denotes the information symbol intended for PU k on subcarrier i and we assume $\mathcal{E}\{|d_{\mathrm{PU}_k}^i|^2\} = 1$

[1]Multiplexing more than two users on a subcarrier leads to more severe co-channel interference and requires several stages of SIC at the users which increases the complexity and introduces additional delays.

[2]The primary BS is assumed to be a legacy infrastructure which adopts conventional orthogonal multiple access.
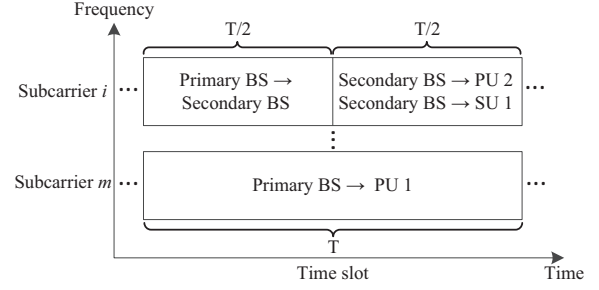

Fig. 2. Illustration of the subcarrier allocation in one time slot.

without loss of generality. $q_k^i \in \mathbb{R}^+$ is the transmit power for the signal transmitted directly to PU k. $f_k^i \in \mathbb{C}$ denotes the channel coefficient of the link between the primary BS and PU k on subcarrier i and captures the joint effect of pathloss, small scale fading, and shadowing. $z_{\mathrm{PU}_k}^i \sim \mathcal{CN}(0, \sigma_{\mathrm{PU}_k}^2)$ denotes the complex additive white Gaussian noise (AWGN) on subcarrier i at PU k. Besides, if subcarrier i is allocated for transmitting the information of PU k via the secondary BS, the received signal at the secondary BS on subcarrier i in the first half of the time slot is given by

$$x_{\mathrm{ST}}^i = \sqrt{q_{\mathrm{ST}}^i} f_{\mathrm{ST}}^i d_{\mathrm{PU}_k}^i + z_{\mathrm{ST}}^i, \qquad (2)$$

where $q_{\mathrm{ST}}^i \in \mathbb{R}^+$ is the transmit power for the signal transmitted to the assisting secondary BS. $f_{\mathrm{ST}}^i \in \mathbb{C}$ denotes the channel coefficient between the primary BS and the secondary BS on subcarrier i. $z_{\mathrm{ST}}^i \sim \mathcal{CN}(0, \sigma_{\mathrm{ST}}^2)$ denotes the complex AWGN on subcarrier i at the secondary BS.

On the other hand, in the second half of the time slot, the secondary BS can multiplex at most one PU and one SU on each subcarrier in the secondary network. Specifically, assuming that PU k and SU $j \in \{1, \ldots, J\}$ are multiplexed on subcarrier i, the received signals at PU k and SU j are given by

$$y_{\mathrm{PU}_k}^i = \sqrt{p_{\mathrm{PU}_k}^i} h_k^i d_{\mathrm{PU}_k}^i + \sqrt{p_{\mathrm{SU}_j}^i} h_k^i d_{\mathrm{SU}_j}^i + z_{\mathrm{PU}_k}^i \quad \text{and} \quad (3)$$

$$y_{\mathrm{SU}_j}^i = \sqrt{p_{\mathrm{SU}_j}^i} g_j^i d_{\mathrm{SU}_j}^i + \sqrt{p_{\mathrm{PU}_k}^i} g_j^i d_{\mathrm{PU}_k}^i + z_{\mathrm{SU}_j}^i, \qquad (4)$$

respectively, where $d_{\mathrm{SU}_j}^i \in \mathbb{C}$ denotes the information symbol intended for SU j on subcarrier i, and we assume $\mathcal{E}\{|d_{\mathrm{SU}_j}^i|^2\} = 1$ without loss of generality. $p_{\mathrm{PU}_k}^i \in \mathbb{R}^+$ and $p_{\mathrm{SU}_j}^i \in \mathbb{R}^+$ denote the transmit powers for the signals intended for PU k and SU j on subcarrier i at the secondary BS, respectively. $h_k^i \in \mathbb{C}$ and $g_j^i \in \mathbb{C}$ denote the coefficients for the secondary BS-to-PU k link and the secondary BS-to-SU j link on subcarrier i, respectively. We note that the joint effect of pathloss, small scale fading, and shadowing is captured by variables h_k^i and g_j^i. $z_{\mathrm{SU}_j}^i \sim \mathcal{CN}(0, \sigma_{\mathrm{SU}_j}^2)$ denotes the complex AWGN on subcarrier i at SU j. Besides, for the study of optimal resource allocation algorithm design, we assume that the global channel state information (CSI) of all links is perfectly known at the primary and secondary BSs.

III. PROBLEM FORMULATION

In this section, we first define the adopted performance measure for the considered cognitive relaying MC-NOMA system. Then, we formulate the power and subcarrier allocation design as an optimization problem.

A. Achievable Rate and Weighted System Throughput

In the primary network, if subcarrier i is allocated to PU k for direct transmission, the achievable data rate of PU k on subcarrier i is given by

$$C^i_{\mathrm{PU}_k} = \log_2\left(1 + q^i_k F^i_k\right), \tag{5}$$

where $F^i_k = |f^i_k|^2/\sigma^2_{\mathrm{PU}_k}$; otherwise if the secondary BS assists the information transmission to a PU on subcarrier i, the achievable data rate of the secondary BS on subcarrier i is given by

$$C^i_{\mathrm{ST}} = \frac{1}{2}\log_2\left(1 + q^i_{\mathrm{ST}} F^i_{\mathrm{ST}}\right), \tag{6}$$

where $F^i_{\mathrm{ST}} = |f^i_{\mathrm{ST}}|^2/\sigma^2_{\mathrm{ST}}$ and the pre-log factor $1/2$ appears in (6) since the primary BS communicates with the secondary BS only in the first half of the time slot.

On the other hand, the secondary network adopts NOMA for improving the system performance. Since only SUs are equipped with successive interference cancellers, we adopt a fixed SIC decoding order to enable NOMA. Specifically, assuming that PU k and SU j are multiplexed simultaneously on subcarrier i in the second half of the time slot, SU j performs SIC to decode and remove the signal of PU k before decoding its own signal. Then, the achievable data rates of PU k and SU j are given by

$$R^i_{\mathrm{PU}_{k,j}} = \frac{1}{2}\log_2\left(1 + \frac{p^i_{\mathrm{PU}_k} H^i_k}{p^i_{\mathrm{SU}_j} H^i_k + 1}\right) \quad \text{and} \tag{7}$$

$$R^i_{\mathrm{SU}_{k,j}} = \frac{1}{2}\log_2(1 + p^i_{\mathrm{SU}_j} G^i_j), \tag{8}$$

respectively, where $H^i_k = |h^i_k|^2/\sigma^2_{\mathrm{PU}_k}$ and $G^i_j = |g^i_j|^2/\sigma^2_{\mathrm{SU}_j}$. The achievable rates in (7) and (8) are multiplied with a factor of $1/2$ since the secondary BS utilizes only the second half of the time slot for transmitting signals to the SUs and PUs. Besides, the following inequality specifies the maximum achievable data rate for relaying the information of PU k on subcarrier i in the secondary network:

$$s^i_{k,j} R^i_{\mathrm{PU}_{k,j}} \leq c^i_{\mathrm{ST}} C^i_{\mathrm{ST}}, \forall i, k, j, \tag{9}$$

where $s^i_{k,j} \in \{1,0\}$ and $c^i_{\mathrm{ST}} \in \{1,0\}$ are the binary subcarrier allocation indicators. In particular, $s^i_{k,j} = 1$ indicates that PU k and SU j are multiplexed on subcarrier i in the secondary network, $c^i_{\mathrm{ST}} = 1$ indicates that subcarrier i is allocated for transmitting information from the primary BS to the secondary BS in the primary network, and $s^i_{k,j} = 0$ and $c^i_{\mathrm{ST}} = 0$ if subcarrier i is not used due to unsatisfactory channel conditions.

Moreover, the use of NOMA requires successful SIC at the SU for interference mitigation. In practice, SU j can successfully decode and remove the co-channel interference causing by PU k on subcarrier i by SIC only when the following inequality holds:

$$\log_2\left(1 + \frac{p^i_{\mathrm{PU}_k} H^i_k}{p^i_{\mathrm{SU}_j} H^i_k + 1}\right) \leq \log_2\left(1 + \frac{p^i_{\mathrm{PU}_k} G^i_j}{p^i_{\mathrm{SU}_j} G^i_j + 1}\right) \tag{10}$$

$$\Longleftrightarrow \quad H^i_k - G^i_j \leq 0. \tag{11}$$

We note that the inequality in (10) is due to the fixed decoding order in the considered system. Besides, the inequality in (11) indicates that for a given subcarrier i, an SU can only be paired with a PU that has a worse DL transmission channel with respect to the secondary BS than the SU. Therefore, we define $\mathcal{K}(i,j)$ as the set of PUs whose channel condition satisfies (11), i.e., $H^i_k - G^i_j \leq 0, \forall k \in \mathcal{K}(i,j)$.

Therefore, the weighted system throughput on subcarrier i is given by

$$U^i = \sum_{j=1}^{J} \sum_{k \in \mathcal{K}(i,j)} s^i_{k,j}\left[w R^i_{\mathrm{PU}_{k,j}} + \mu R^i_{\mathrm{SU}_{k,j}}\right] + \sum_{k=1}^{K} c^i_k w C^i_{\mathrm{PU}_k}. \tag{12}$$

Here, $c^i_k \in \{1,0\}$ and $c^i_k = 1$ indicates that PU k is scheduled on subcarrier i in the primary network, otherwise $c^i_k = 0$. The positive constants $0 \leq w \leq 1$ and $0 \leq \mu \leq 1$ reflect the priority of the PUs and SUs in resource allocation, respectively, and are specified in the media access control (MAC) layer to achieve certain fairness objectives in resource allocation.

B. Optimization Problem Formulation

The design objective is to maximize the weighted sum throughput of the two systems, while guaranteeing minimum data rates for the PUs. The optimal joint power and subcarrier allocation policy is obtained by solving the following optimization problem:

$$\underset{\substack{q^i_k, q^i_{\mathrm{ST}}, p^i_{\mathrm{PU}_k}, p^i_{\mathrm{SU}_j} \geq 0, \\ c^i_{\mathrm{ST}}, c^i_k, s^i_{k,j}}}{\text{maximize}} \quad \sum_{i=1}^{N_{\mathrm{F}}} U^i \tag{13}$$

s.t. C1: $s^i_{k,j} R^i_{\mathrm{PU}_{k,j}} - c^i_{\mathrm{ST}} C^i_{\mathrm{ST}} \leq 0, \forall i, k, j,$

C2: $\sum_{i=1}^{N_{\mathrm{F}}}\left(c^i_k C^i_{\mathrm{PU}_k} + \sum_{j=1}^{J} s^i_{k,j} R^i_{\mathrm{PU}_{k,j}}\right) \geq R^{\mathrm{req}}_{\mathrm{PU}_k}, \ \forall k,$

C3: $\sum_{i=1}^{N_{\mathrm{F}}}\sum_{k=1}^{K}\sum_{j=1}^{J} s^i_{k,j}(p^i_{\mathrm{PU}_k} + p^i_{\mathrm{SU}_j}) \leq P^{\mathrm{ST}}_{\max},$

C4: $\sum_{i=1}^{N_{\mathrm{F}}}\left(c^i_{\mathrm{ST}} q^i_{\mathrm{ST}} + \sum_{k=1}^{K} c^i_k q^i_k\right) \leq P^{\mathrm{PT}}_{\max},$

C5: $s^i_{k,j} \in \{0,1\}, \forall i, k, j,$ C6: $c^i_{\mathrm{ST}}, c^i_k \in \{0,1\}, \forall i, k,$

C7: $c^i_k + \sum_{j=1}^{J} s^i_{k,j} \leq 1, \forall i, k,$ C8: $c^i_{\mathrm{ST}} + \sum_{k=1}^{K} c^i_k \leq 1, \forall i.$

In problem (13), constraint C1 guarantees successful information decoding on subcarrier i at PU k if the secondary BS is selected for assisting information transmission. Constraint C2 imposes a minimum data rate requirement $R^{\mathrm{req}}_{\mathrm{PU}_k}$ for PU k. Constraints C3 and C4 are the power budget constraints for the secondary and primary BSs with the maximum transmit powers $P^{\mathrm{ST}}_{\max}$ and $P^{\mathrm{PT}}_{\max}$, respectively. Constraints C5, C6, and C7 are imposed to guarantee that each subcarrier can only be allocated to at most one SU and one PU in the secondary network. Constraints C6 and C8 guarantee that each subcarrier can only be allocated to at most one PU in the primary network or forward the information of one PU to the secondary BS.

The problem in (13) is a mixed-integer non-convex optimization problem which is known to be NP-hard. There exists no systematic and computationally efficient approach for obtaining a globally optimal solution for such problems. Nevertheless, by exploiting the special structure of the considered problem, we will develop efficient resource allocation algorithms for finding the optimal and suboptimal resource allocation policies for (13) in the next section.

IV. SOLUTIONS OF THE OPTIMIZATION PROBLEM

In this section, we first solve problem (13) optimally by applying monotonic optimization theory [11], [12]. Then, we propose a suboptimal algorithm with low computational complexity for obtaining a close-to-optimal solution of (13).

A. Optimal Resource Allocation Scheme

First, we can rewrite the weighted system throughput on subcarrier i in the following equivalent form:

$$U^i = \sum_{j=1}^{J} \sum_{k \in \mathcal{K}(i,j)} \frac{1}{2} \Big[w \log_2 \Big(1 + \frac{s_{k,j}^i p_{\mathrm{PU}_k}^i H_k^i}{s_{k,j}^i p_{\mathrm{SU}_j}^i H_k^i + I_{\mathrm{PU}_{k,j}}^i + 1} \Big) + \mu \log_2 \Big(1 + \frac{s_{k,j}^i p_{\mathrm{SU}_j}^i G_j^i}{I_{\mathrm{SU}_{k,j}}^i + 1} \Big) \Big] + \sum_{k=1}^{K} w \log_2 \Big(1 + \frac{c_k^i q_k^i F_k^i}{S_{\mathrm{PU}_k}^i + 1} \Big), (14)$$

where

$$I_{\mathrm{PU}_{k,j}}^i = \Big(c_k^i q_k^i + \sum_{n \neq j}^{J} \sum_{m \neq k}^{K} s_{m,n}^i (p_{\mathrm{PU}_m}^i + p_{\mathrm{SU}_n}^i) \Big) H_k^i, \quad (15)$$

$$I_{\mathrm{SU}_{k,j}}^i = \Big(c_k^i q_k^i + \sum_{n \neq j}^{J} \sum_{m \neq k}^{K} s_{m,n}^i (p_{\mathrm{PU}_m}^i + p_{\mathrm{SU}_n}^i) \Big) G_j^i, \quad (16)$$

$$S_{\mathrm{PU}_k}^i = \Big(c_{\mathrm{ST}}^i q_{\mathrm{ST}}^i + \sum_{m \neq k}^{K} c_m^i q_m^i \Big) F_k^i. \quad (17)$$

We note that (14) is equivalent to (12) due to constraints C5–C8. In particular, $I_{\mathrm{PU}_{k,j}}^i$, $I_{\mathrm{SU}_{k,j}}^i$, and $S_{\mathrm{PU}_k}^i$ represent the undesired interference at the receivers. Specifically, if a given subcarrier allocation policy satisfies constraints C5–C8, $I_{\mathrm{PU}_{k,j}}^i$, $I_{\mathrm{SU}_{k,j}}^i$, and $S_{\mathrm{PU}_k}^i$ do not have any impact on the system performance, otherwise they will severely degrade the system throughput. In other words, $I_{\mathrm{PU}_{k,j}}^i$, $I_{\mathrm{SU}_{k,j}}^i$, and $S_{\mathrm{PU}_k}^i$ act as penalty terms to penalize the objective function for any violation of constraints C5–C8.

Then, we define $\tilde{p}_{\mathrm{PU}_{k,j}}^i = s_{k,j}^i p_{\mathrm{PU}_k}^i$, $\tilde{p}_{\mathrm{SU}_{k,j}}^i = s_{k,j}^i p_{\mathrm{SU}_j}^i$, $\tilde{q}_k^i = c_k^i q_k^i$, and $\tilde{q}_{\mathrm{ST}}^i = c_{\mathrm{ST}}^i q_{\mathrm{ST}}^i$. Therefore, the weighted system throughout in (14) can be rewritten in equivalent form as:

$$\tilde{U}^i = \sum_{j=1}^{J} \sum_{k \in \mathcal{K}(i,j)} \frac{1}{2} \Big[w \log_2 \Big(1 + \frac{\tilde{p}_{\mathrm{PU}_{k,j}}^i H_k^i}{\tilde{p}_{\mathrm{SU}_{k,j}}^i H_k^i + \tilde{I}_{\mathrm{PU}_{k,j}}^i + 1} \Big) + \mu \log_2 \Big(1 + \frac{\tilde{p}_{\mathrm{SU}_{k,j}}^i G_j^i}{\tilde{I}_{\mathrm{SU}_{k,j}}^i + 1} \Big) \Big] + \sum_{k=1}^{K} w \log_2 \Big(1 + \frac{\tilde{q}_k^i F_k^i}{\tilde{S}_{\mathrm{PU}_k}^i + 1} \Big), (18)$$

where $\tilde{I}_{\mathrm{PU}_{k,j}}^i = \Big(\tilde{q}_k^i + \sum_{n \neq j}^{J} \sum_{m \neq k}^{K} \tilde{p}_{\mathrm{PU}_{m,n}}^i + \tilde{p}_{\mathrm{SU}_{m,n}}^i \Big) H_k^i$, $\tilde{I}_{\mathrm{SU}_{k,j}}^i = \Big(\tilde{q}_k^i + \sum_{n \neq j}^{J} \sum_{m \neq k}^{K} \tilde{p}_{\mathrm{PU}_{m,n}}^i + \tilde{p}_{\mathrm{SU}_{m,n}}^i \Big) G_j^i$, and $\tilde{S}_{\mathrm{PU}_k}^i = \Big(\tilde{q}_{\mathrm{ST}}^i + \sum_{m \neq k}^{K} \tilde{q}_m^i \Big) F_k^i$.

Besides, we note that non-convex constraint C1 is the difference of two logarithmic functions and not monotonic. Hence, constraint C1 is not in the canonical form necessary for applying monotonic optimization. Thus, with the aforementioned definitions, we handle the non-convex constraint C1 by equivalently expressing it as

$$\text{C1a: } \log_2 \Big(1 + \frac{\tilde{p}_{\mathrm{PU}_{k,j}}^i H_k^i}{\tilde{p}_{\mathrm{SU}_{k,j}}^i H_k^i + \tilde{I}_{\mathrm{PU}_{k,j}}^i + 1} \Big) + \varsigma_{k,j}^i \leq \log_2 \Big(1 + P_{\max}^{\mathrm{ST}} H_k^i \Big), (19)$$

$$\text{C1b: } \log_2 \Big(1 + \tilde{q}_{\mathrm{ST}}^i F_{\mathrm{ST}}^i \Big) + \varsigma_{k,j}^i \geq \log_2 \Big(1 + P_{\max}^{\mathrm{ST}} H_k^i \Big), \quad (20)$$

where $\varsigma_{k,j}^i \geq 0$ is a slack scalar optimization variable. We note that constraints C1a and C1b are monotonically increasing functions.

Then, the original problem in (13) can be rewritten in the following equivalent form:

$$\underset{\tilde{q}_k^i, \tilde{q}_{\mathrm{ST}}^i, \tilde{p}_{\mathrm{PU}_{k,j}}^i, \tilde{p}_{\mathrm{SU}_{k,j}}^i}{\text{maximize}} \sum_{i=1}^{N_{\mathrm{F}}} \tilde{U}^i \quad (21)$$

s.t. C1a, C1b, C3: $\sum_{i=1}^{N_{\mathrm{F}}} \sum_{k=1}^{K} \sum_{j=1}^{J} \tilde{p}_{\mathrm{PU}_{k,j}}^i + \tilde{p}_{\mathrm{SU}_{k,j}}^i \leq P_{\max}^{\mathrm{ST}}$,

C2: $\sum_{i=1}^{N_{\mathrm{F}}} \Big(\log_2(1 + \tilde{q}_k^i F_k^i) + \sum_{j=1}^{J} \frac{1}{2} \log_2 \Big(1 + \frac{\tilde{p}_{\mathrm{PU}_{k,j}}^i H_k^i}{\tilde{p}_{\mathrm{SU}_{k,j}}^i H_k^i + 1} \Big) \Big)$

$\geq R_{\mathrm{PU}_k}^{\mathrm{req}}, \forall k,$ C4: $\sum_{i=1}^{N_{\mathrm{F}}} \Big(\tilde{q}_{\mathrm{ST}}^i + \sum_{k=1}^{K} \tilde{q}_k^i \Big) \leq P_{\max}^{\mathrm{PT}}$,

C7, C8, C9: $\tilde{q}_k^i, \tilde{q}_{\mathrm{ST}}^i, \tilde{p}_{\mathrm{PU}_{k,j}}^i, \tilde{p}_{\mathrm{SU}_{k,j}}^i \geq 0, \forall i, k, j.$

Here, we note that the binary constraints C5 and C6 are absorbed in the objective function via the penalty terms $\tilde{I}_{\mathrm{PU}_{k,j}}^i$, $\tilde{I}_{\mathrm{SU}_{k,j}}^i$, and $\tilde{S}_{\mathrm{PU}_k}^i$ in (18). Due to the monotonicity of the objective function in (21), for the optimal solution, each subcarrier will be allocated to a single receiver in the primary network or a single pair of PU and SU in the secondary network, i.e., undesired co-channel interference is completely avoided.

For facilitating monotonic optimization, we define three auxiliary variables $u_{k,j}^i$, $v_{k,j}^i$, and ξ_k^i, which satisfy the following constraints:

$$\text{M1: } 1 \leq u_{k,j}^i \leq 1 + \frac{\tilde{p}_{\mathrm{PU}_{k,j}}^i H_k^i}{\tilde{p}_{\mathrm{SU}_{k,j}}^i H_k^i + \tilde{I}_{\mathrm{PU}_{k,j}}^i + 1}, \forall i, j, \forall k \in \mathcal{K}(i,j), (22)$$

$$\text{M2: } 1 \leq v_{k,j}^i \leq 1 + \frac{\tilde{p}_{\mathrm{SU}_{k,j}}^i G_j^i}{\tilde{I}_{\mathrm{SU}_{k,j}}^i + 1}, \forall i, j, \forall k \in \mathcal{K}(i,j), \quad (23)$$

$$\text{M3: } 1 \leq \xi_k^i \leq 1 + \frac{\tilde{q}_k^i F_k^i}{\tilde{S}_{\mathrm{PU}_k}^i + 1}, \forall i, k. \quad (24)$$

In addition, we define

$$\underline{\tilde{U}}^i = \sum_{j=1}^{J} \sum_{k \in \mathcal{K}(i,j)} \frac{1}{2} [\log_2(u_{k,j}^i)^w + \log_2(v_{k,j}^i)^\mu] + \sum_{k=1}^{K} \log_2(\xi_k^i)^w. (25)$$

In fact, $\underline{\tilde{U}}^i$ serves as a lower bound of the original objective function U^i in (12). However, due to the monotonicity of the original problem in (13), we can obtain the optimal solution of (13) by maximizing $\underline{\tilde{U}}^i$. In particular, the original problem in (13) can be rewritten as a monotonic optimization problem in canonical form as follows:

$$\underset{\mathbf{u}, \mathbf{v}, \boldsymbol{\xi}, \boldsymbol{\varsigma}}{\text{maximize}} \sum_{i=1}^{N_{\mathrm{F}}} \underline{\tilde{U}}^i \quad \text{s.t.} (\mathbf{u}, \mathbf{v}, \boldsymbol{\xi}, \boldsymbol{\varsigma}) \in \mathcal{G} \cap \mathcal{H}, \quad (26)$$

where vectors $\mathbf{u}$, $\mathbf{v}$, $\boldsymbol{\xi}$, and $\boldsymbol{\varsigma}$ contain all $u_{k,j}^i$, $v_{k,j}^i$, ξ_k^i, and $\varsigma_{k,j}^i$, respectively. In (25), $\mathcal{G}$ denotes a normal set [12] which accounts for the limitation of the available radio resources. In particular, $\mathcal{G}$ is spanned by constraints C1a, C3, C4, C9, M1, M2, and M3. Besides, $\mathcal{H}$ is a conormal set which is spanned by constraints C1b and C2. We note that the objective function and all constraint functions in (26) are monotonically increasing functions. Hence, from the theory of monotonic optimization, we know that the globally optimal solution of optimization problem (26) can be obtained by employing the outer polyblock approximation approach. In particular, a

globally optimal solution of (26) can be obtained via an outer polyblock approximation based algorithm as developed in our previous works [7], [14], and the details are omitted here due to page limitation.

We note that the monotonic optimization-based resource allocation algorithm entails a high computational complexity as the search space of the optimal solution grows exponentially with the number of vertexes involved in the problem, e.g. $2N_\mathrm{F}(K + KJ)$ for the considered system. Nevertheless, the performance of the optimal resource allocation policy serves as a benchmark for any suboptimal algorithm. In the next section, we develop a suboptimal scheme, which finds a locally optimal solution for problem (13), but requires only polynomial time computational complexity.

B. Suboptimal Solution

To facilitate the design of the suboptimal algorithm, we rewrite the original objective function in (12) as

$$\overline{U}^i = \sum_{j=1}^J \sum_{k \in \mathcal{K}(i,j)} \frac{1}{2}\Big[w \log_2\Big(1 + \frac{\tilde{p}^i_{\mathrm{PU}_{k,j}} H^i_k}{\tilde{p}^i_{\mathrm{SU}_{k,j}} H^i_k + 1}\Big) + \mu \log_2(1 + \tilde{p}^i_{\mathrm{SU}_{k,j}} G^i_j)\Big] + \sum_{k=1}^K w \log_2\Big(1 + \tilde{q}^i_k F^i_k\Big). \quad (27)$$

We note that the multiplicative terms in (27) involving binary variables and power allocation variables, i.e., $\tilde{p}^i_{\mathrm{PU}_{k,j}} = s^i_{k,j} p^i_{\mathrm{PU}_k}$, $\tilde{p}^i_{\mathrm{SU}_{k,j}} = s^i_{k,j} p^i_{\mathrm{SU}_j}$, $\tilde{q}^i_k = c^i_k q^i_k$, and $\tilde{q}^i_{\mathrm{ST}} = c^i_{\mathrm{ST}} q^i_{\mathrm{ST}}$, are obstacles for the design of computationally efficient resource allocation algorithms. To handle this issue, we adopt the big-M method [15] to decompose the multiplicative terms. In particular, we impose the following additional constraints:

C10: $\tilde{q}^i_k \leq P^\mathrm{PT}_\mathrm{max} c^i_k$,   C11: $\tilde{q}^i_\mathrm{ST} \leq P^\mathrm{PT}_\mathrm{max} c^i_\mathrm{ST}$,   (28)

C12: $\tilde{p}^i_{\mathrm{PU}_{k,j}} \leq P^\mathrm{ST}_\mathrm{max} s^i_{k,j}$,   C13: $\tilde{p}^i_{\mathrm{PU}_{k,j}} \leq P^\mathrm{ST}_\mathrm{max} s^i_{k,j}$, (29)

C14: $\tilde{q}^i_k \geq q^i_k - (1 - c^i_k) P^\mathrm{PT}_\mathrm{max}$,   C15: $\tilde{q}^i_k \leq q^i_k$,   (30)

C16: $\tilde{q}^i_\mathrm{ST} \geq q^i_\mathrm{ST} - (1 - c^i_\mathrm{ST}) P^\mathrm{PT}_\mathrm{max}$,   C17: $\tilde{q}^i_\mathrm{ST} \leq q^i_\mathrm{ST}$,   (31)

C18: $\tilde{p}^i_{\mathrm{PU}_{k,j}} \geq p^i_{\mathrm{PU}_k} - (1 - s^i_{k,j}) P^\mathrm{ST}_\mathrm{max}$, C19: $\tilde{p}^i_{\mathrm{PU}_{k,j}} \leq p^i_{\mathrm{PU}_k}$, (32)

C20: $\tilde{p}^i_{\mathrm{SU}_{k,j}} \geq p^i_{\mathrm{SU}_j} - (1 - s^i_{k,j}) P^\mathrm{ST}_\mathrm{max}$, C21: $\tilde{p}^i_{\mathrm{SU}_{k,j}} \leq p^i_{\mathrm{SU}_j}$. (33)

Besides, in order to handle the binary constraints C5 and C6 in problem (13), we replace constraints C5 and C6 with their equivalent constraints {C5a, C5b} and {C6a, C6b, C6c}, respectively, which are given by

C5a: $\sum_{i=1}^{N_\mathrm{F}} \sum_{k=1}^K \sum_{j=1}^J s^i_{k,j} - (s^i_{k,j})^2 \leq 0$,   C5b: $0 \leq s^i_{k,j} \leq 1$, (34)

C6a: $\sum_{i=1}^{N_\mathrm{F}}\Big(c^i_\mathrm{ST} - (c^i_\mathrm{ST})^2 + \sum_{k=1}^K c^i_k - (c^i_k)^2\Big) \leq 0$,   (35)

C6b: $0 \leq c^i_\mathrm{ST} \leq 1$,   and   C6c: $0 \leq c^i_k \leq 1$.   (36)

We note that in (34) and (36), $s^i_{k,j}$, c^i_ST, and c^i_k are continuous in the interval between zero and one. However, constraints C5a and C6a are reverse convex functions [16] which makes problem (13) still non-convex. To resolve this issue, we reformulate the problem in (13) as

$$\underset{\tilde{\mathbf{p}},\tilde{\mathbf{q}},\mathbf{c},\mathbf{s}}{\text{minimize}} \sum_{i=1}^{N_\mathrm{F}} \Big[-\overline{U}^i + \rho\Big(\sum_{k=1}^K \sum_{j=1}^J s^i_{k,j} - (s^i_{k,j})^2 + \sum_{k=1}^K c^i_k - (c^i_k)^2 + c^i_\mathrm{ST} - (c^i_\mathrm{ST})^2\Big)\Big]$$

s.t.   C1–C4, C5b, C6b, C6c, C7–C21,   (37)

Algorithm 1 Successive Convex Approximation

1: Initialize the penalty factor $\rho \gg 1$, iteration index $r = 1$, and initial point $\tilde{\mathbf{p}}^{(1)}$, $\tilde{\mathbf{q}}^{(1)}$, $\mathbf{c}^{(1)}$, and $\mathbf{s}^{(1)}$
2: **repeat**
3:    Solve (44) for given $\tilde{\mathbf{p}}^{(r)}$, $\tilde{\mathbf{q}}^{(r)}$, $\mathbf{c}^{(r)}$, and $\mathbf{s}^{(r)}$ and store the intermediate resource allocation policy $\{\tilde{\mathbf{p}}, \tilde{\mathbf{q}}, \mathbf{c}, \mathbf{s}\}$
4:    Set $r = r + 1$ and $\tilde{\mathbf{p}}^{(r)} = \tilde{\mathbf{p}}$, $\tilde{\mathbf{q}}^{(r)} = \tilde{\mathbf{q}}$, $\mathbf{c}^{(r)} = \mathbf{c}$, and $\mathbf{s}^{(r)} = \mathbf{s}$
5: **until** convergence
6: Obtain final resource allocation policy $\tilde{\mathbf{p}}^* = \tilde{\mathbf{p}}^{(r)}$, $\tilde{\mathbf{q}}^* = \tilde{\mathbf{q}}^{(r)}$, $\mathbf{c}^* = \mathbf{c}^{(r)}$, and $\mathbf{s}^* = \mathbf{s}^{(r)}$

---

where the $\tilde{\mathbf{p}}, \tilde{\mathbf{q}}, \mathbf{c}, \mathbf{s}$ contain all $\{\tilde{p}^i_{\mathrm{PU}_{k,j}}, \tilde{p}^i_{\mathrm{SU}_{k,j}}\}$, $\{\tilde{q}^i_k, \tilde{q}^i_\mathrm{ST}\}$, $\{c^i_\mathrm{ST}, c^i_k\}$, and $s^i_{k,j}$, respectively. Constant $\rho \gg 1$ is the penalty factor. In fact, $\rho\Big(\sum_{k=1}^K \sum_{j=1}^J s^i_{k,j} - (s^i_{k,j})^2 + \sum_{k=1}^K c^i_k - (c^i_k)^2 + c^i_\mathrm{ST} - (c^i_\mathrm{ST})^2\Big)$ in (37) is the penalty term that penalizes the violation of constraints C5a and C6a, which forces the optimization variables $s^i_{k,j}$, c^i_k, and c^i_ST to be zero or one. It is shown in [7], [16] that (37) and (13) are equivalent for $\rho \gg 1$.

Now, the remaining non-convexity of problem (37) is due to constraints C1, C2, and the objective function. However, (37) can be rewritten in form of a standard difference of convex (d.c.) programming problem [16] as:

$$\underset{\tilde{\mathbf{p}},\tilde{\mathbf{q}},\mathbf{c},\mathbf{s}}{\text{minimize}} \quad \boldsymbol{F}(\tilde{\mathbf{p}}, \tilde{\mathbf{q}}) - \boldsymbol{G}(\tilde{\mathbf{p}}, \tilde{\mathbf{q}}) + \rho(\boldsymbol{H}(\mathbf{c}, \mathbf{s}) - \boldsymbol{M}(\mathbf{c}, \mathbf{s}))$$

s.t. $\widetilde{\text{C1}}$: $\boldsymbol{B}^i_{k,j}(\tilde{\mathbf{p}}, \tilde{\mathbf{q}}) - \boldsymbol{D}^i_{k,j}(\tilde{\mathbf{p}}, \tilde{\mathbf{q}}) \leq 0$,
      $\widetilde{\text{C2}}$: $\boldsymbol{R}^i_{k,j}(\tilde{\mathbf{p}}, \tilde{\mathbf{q}}) - \boldsymbol{T}^i_{k,j}(\tilde{\mathbf{p}}, \tilde{\mathbf{q}}) \leq 0$,
      C3, C4, C5b, C6b, C6c, C7–C21,   (38)

where

$$\boldsymbol{F}(\tilde{\mathbf{p}}, \tilde{\mathbf{q}}) = \sum_{i=1}^{N_\mathrm{F}} \Big(\sum_{j=1}^J \sum_{k \in \mathcal{K}(i,j)} \frac{1}{2}\big[w \log_2(1 + (\tilde{p}^i_{\mathrm{PU}_{k,j}} + \tilde{p}^i_{\mathrm{SU}_{k,j}}) H^i_k) + \mu \log_2(1 + \tilde{p}^i_{\mathrm{SU}_{k,j}} G^i_j)\big] + \sum_{k=1}^K w \log_2(1 + \tilde{q}^i_k F^i_k)\Big), \quad (39)$$

$$\boldsymbol{G}(\tilde{\mathbf{p}}, \tilde{\mathbf{q}}) = \sum_{i=1}^{N_\mathrm{F}} \sum_{j=1}^J \sum_{k \in \mathcal{K}(i,j)} \frac{1}{2} w \log_2(1 + \tilde{p}^i_{\mathrm{SU}_{k,j}} H^i_k), \quad (40)$$

$$\boldsymbol{H}(\mathbf{c}, \mathbf{s}) = \sum_{i=1}^{N_\mathrm{F}} \Big(c^i_\mathrm{ST} + \sum_{k=1}^K c^i_k + \sum_{k=1}^K \sum_{j=1}^J s^i_{k,j}\Big), \quad \text{and} \quad (41)$$

$$\boldsymbol{M}(\mathbf{c}, \mathbf{s}) = \sum_{i=1}^{N_\mathrm{F}} \Big((c^i_\mathrm{ST})^2 + \sum_{k=1}^K (c^i_k)^2 + \sum_{k=1}^K \sum_{j=1}^J (s^i_{k,j})^2\Big). \quad (42)$$

The definitions of $\boldsymbol{B}^i_{k,j}(\tilde{\mathbf{p}}, \tilde{\mathbf{q}})$ and $\boldsymbol{D}^i_{k,j}(\tilde{\mathbf{p}}, \tilde{\mathbf{q}})$ in $\widetilde{\text{C1}}$, and the definitions of $\boldsymbol{R}^i_{k,j}(\tilde{\mathbf{p}}, \tilde{\mathbf{q}})$ and $\boldsymbol{T}^i_{k,j}(\tilde{\mathbf{p}}, \tilde{\mathbf{q}})$ in $\widetilde{\text{C2}}$ are similar to those of $\boldsymbol{F}(\tilde{\mathbf{p}}, \tilde{\mathbf{q}})$ and $\boldsymbol{G}(\tilde{\mathbf{p}}, \tilde{\mathbf{q}})$. In particular, constraints $\widetilde{\text{C1}}$ and $\widetilde{\text{C2}}$ are written as a difference of logarithmic functions. We note that the problems in (21) and (38) are equivalent in the sense that they have the same optimal solution. Thus, we can obtain a locally optimal solution of (38) by applying successive convex approximation [17]. In particular, for any feasible point $\tilde{\mathbf{p}}^{(r)}$, $\tilde{\mathbf{q}}^{(r)}$, $\mathbf{c}^{(r)}$, and $\mathbf{s}^{(r)}$, we have the following inequalities:

$$\boldsymbol{G}(\tilde{\mathbf{p}}, \tilde{\mathbf{q}}) \geq \boldsymbol{G}(\tilde{\mathbf{p}}^{(r)}, \tilde{\mathbf{q}}^{(r)}) + \mathrm{Tr}(\nabla_{\tilde{\mathbf{p}}} \boldsymbol{G}(\tilde{\mathbf{p}}, \tilde{\mathbf{q}})(\tilde{\mathbf{p}} - \tilde{\mathbf{p}}^{(r)})) + \mathrm{Tr}(\nabla_{\tilde{\mathbf{q}}} \boldsymbol{G}(\tilde{\mathbf{p}}, \tilde{\mathbf{q}})(\tilde{\mathbf{q}} - \tilde{\mathbf{q}}^{(r)})) \triangleq \overline{G}(\tilde{\mathbf{p}}, \tilde{\mathbf{q}}, \tilde{\mathbf{p}}^{(r)}, \tilde{\mathbf{q}}^{(r)}), \quad (43)$$

Carrier center frequency and bandwidth	2 GHz and 2.5 MHz
Number of subcarriers, N_F	32
Bandwidth of subcarrier	78 kHz
Primary and secondary BS antenna gain	10 dBi and 5 dBi
Path loss exponent and reference distance, D_{ref}	3.6 and 10 meters
Receiver noise power, $\sigma_{ST}^2, \sigma_{PU_k}^2, \sigma_{SU_j}^2$	-110 dBm
Maximum power at primary BS, P_{max}^{PT}	40 dBm
Maximum power at secondary BS, P_{max}^{ST}	40 dBm
Minimum required rate for PUs, $R_{PU_k}^{req} = R_{req}$	1 bits/s/Hz
Penalty factor ρ for **Algorithm 1**	$10\log_2(1+P_{max}/\sigma_{ST}^2)$

where the right hand side of (43) is an affine function and represents a global underestimation of $G(\tilde{\mathbf{p}}, \tilde{\mathbf{q}})$. Similarly, we denote $\overline{M}(\mathbf{c}, \mathbf{s}, \mathbf{c}^{(r)}, \mathbf{s}^{(r)})$, $\overline{D}_{k,j}^i(\tilde{\mathbf{p}}, \tilde{\mathbf{q}}, \tilde{\mathbf{p}}^{(r)}, \tilde{\mathbf{q}}^{(r)})$, and $\overline{T}_{k,j}^i(\tilde{\mathbf{p}}, \tilde{\mathbf{q}}, \tilde{\mathbf{p}}^{(r)}, \tilde{\mathbf{q}}^{(r)})$ as the global underestimations of $M(\mathbf{c}, \mathbf{s})$, $D_{k,j}^i(\tilde{\mathbf{p}}, \tilde{\mathbf{q}})$, and $T_{k,j}^i(\tilde{\mathbf{p}}, \tilde{\mathbf{q}})$, respectively.

Therefore, for any given $\tilde{\mathbf{p}}^{(r)}$, $\tilde{\mathbf{q}}^{(r)}$, $\mathbf{c}^{(r)}$, and $\mathbf{s}^{(r)}$, we can obtain a lower bound of (38) by solving the following optimization problem:

$$\underset{\tilde{\mathbf{p}}, \tilde{\mathbf{q}}, \mathbf{c}, \mathbf{s}}{\text{minimize}} \quad F(\tilde{\mathbf{p}}, \tilde{\mathbf{q}}) - \overline{G}(\tilde{\mathbf{p}}, \tilde{\mathbf{q}}, \tilde{\mathbf{p}}^{(r)}, \tilde{\mathbf{q}}^{(r)})$$
$$+ \rho(H(\mathbf{c}, \mathbf{s}) - \overline{M}(\mathbf{c}, \mathbf{s}, \mathbf{c}^{(r)}, \mathbf{s}^{(r)}))$$
$$\text{s.t.} \quad \widetilde{C1}: B_{k,j}^i(\tilde{\mathbf{p}}, \tilde{\mathbf{q}}) - \overline{D}_{k,j}^i(\tilde{\mathbf{p}}, \tilde{\mathbf{q}}, \tilde{\mathbf{p}}^{(r)}, \tilde{\mathbf{q}}^{(r)}) \leq 0,$$
$$\widetilde{C2}: R_{k,j}^i(\tilde{\mathbf{p}}, \tilde{\mathbf{q}}) - \overline{T}_{k,j}^i(\tilde{\mathbf{p}}, \tilde{\mathbf{q}}, \tilde{\mathbf{p}}^{(r)}, \tilde{\mathbf{q}}^{(r)}) \leq 0,$$
$$\text{C3, C4, C5b, C6b, C6c, C7–C21.} \quad (44)$$

We successively tighten the obtained lower bound by applying the iterative algorithm summarized in **Algorithm 1**. The proposed suboptimal iterative algorithm converges to a locally optimal solution of (38) with polynomial time computational complexity [17]. We note that the convex problem in (44) can be solved efficiently by standard convex program solvers such as CVX [18].

V. SIMULATION RESULTS

In this section, we evaluate the system performance of the proposed scheme via simulations. The adopted simulation parameters are given in Table I, unless specified otherwise. We assume that the primary BS is L meters away from the secondary BS. There are K PUs and J SUs which are randomly and uniformly distributed between the reference distance and the maximum service distances of $D_{PT} = 500$ meters and $D_{ST} = 150$ meters for the primary BS and the secondary BS, respectively. The small-scale fading of the primary BS-to-PU channels, the secondary BS-to-PU channels, the secondary BS-to-SU channels, and the link between the primary BS and the secondary BS is modeled as independent and identically Rayleigh distributed. The weights of the PUs and SUs are set as $w = 2$ and $\mu = 1$, respectively, to provide higher priority for maximizing the throughput of the PUs. The results shown in this section are averaged over different realizations of pathloss and multipath fading.

For comparison, we also consider the performance of two baseline schemes. For baseline scheme 1, we consider a traditional multicarrier cognitive relaying system where the SUs cannot perform SIC for cancelling the interference from PUs in the secondary network. For baseline scheme 2, the user pair on each subcarrier in the secondary network is selected randomly and we optimize the corresponding transmit powers for the PUs and SUs.
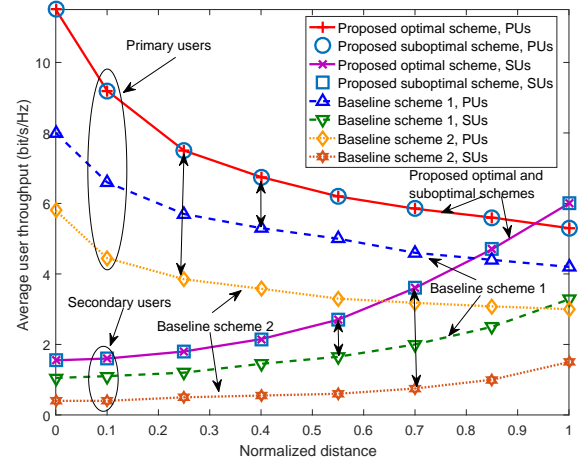

Fig. 3. Average user data rate (bit/s/Hz) versus the normalized distance between the primary and secondary BSs for different resource allocation schemes. The double-sided arrows indicate the performance gains of the proposed optimal scheme compared to the baseline schemes.

A. Average User Throughput vs. Normalized Distance

In Figure 3, we investigate the average user throughput of all PUs and all SUs versus (vs.) the normalized distance between the primary BS and the secondary BS, for $K = 3$ PUs and $J = 3$ SUs. In particular, the normalized distance between the primary BS and the secondary BS is given by $\frac{L - D_{ref}}{D_{PT} - D_{ref}}$. As can be observed from Figure 3, for the proposed optimal and suboptimal schemes, the average user throughput of the PUs decreases monotonically with the normalized distance. In particular, as the normalized distance increases, the quality of the primary BS-to-secondary BS link deteriorates and becomes the bottleneck for the throughput of the assisted PUs. Thus, the primary BS is more reluctant to let the secondary BS assist in the information transmission since forwarding the information over a weak channel requires high transmit power. As a result, the potential throughput gain of the PUs introduced by the secondary BS serving as a relay diminishes with increasing distance between the primary and secondary BSs. On the other hand, it can be observed that the average user throughput of the SUs increases with the normalized distance. In particular, as the normalized distance increases, the secondary BS assists a decreasing number of PUs since the quality of the information forwarding channel between the primary BS and the secondary BS becomes worse. Thus, less power is used at the secondary BS for satisfying the QoS requirements of the assisted PUs and the newly available power can be reallocated to SUs for improving the system throughput. We also note that the proposed suboptimal scheme closely approaches the performance of the proposed optimal resource allocation scheme. On the other hand, Figure 3 shows that both baseline schemes achieve a substantially lower average throughput for the PUs and the SUs compared to the proposed schemes. In particular, since baseline scheme 1 does not utilize interference cancellation at the receivers, the co-channel interference between the PU and the SU in the secondary network degrades the average user throughput of both the PUs and the SUs. Besides, for baseline scheme 2, the non-optimality of the subcarrier allocation leads to a substantial reduction in the average user throughputs. In particular, for a normalized distance of 0.55, the proposed
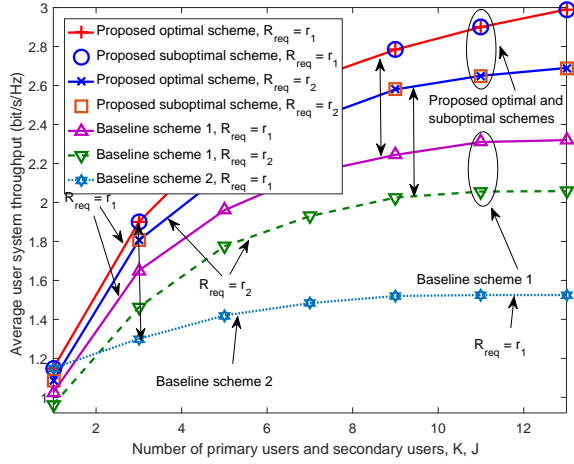
Fig. 4. Average user system throughput (bit/s/Hz) versus the number of primary and secondary users, K, J, for different resource allocation schemes. The double-sided arrows indicate the performance gains of the proposed optimal scheme compared to the baseline schemes.

schemes can achieve roughly a 24% and 87% higher average user throughput for the PUs and a 64% and 280% higher average user throughput for the SUs than baseline schemes 1 and 2, respectively.

B. Average User System Throughput vs. Number of Users

In Figure 4, we investigate the average user system throughput vs. the number of PUs and SUs for a normalized distance between the primary and secondary BSs of 0.6 and different minimum QoS requirements for the PUs, i.e., $r_1 = 1$ bits/s/Hz and $r_2 = 2$ bits/s/Hz. We assume that the numbers of PUs and SUs are identical, i.e., $K = J$. The average user system throughput is calculated as $\frac{\sum_{i=1}^{N_F} \sum_{k=1}^{K} (c_k^i C_{\mathrm{PU}_k}^i + \sum_{j=1}^{J} s_{k,j}^i (R_{\mathrm{PU}_{k,j}}^i + R_{\mathrm{SU}_{k,j}}^i))}{K+J}$. As can be seen from Figure 4, the average user system throughput of both proposed schemes and the baseline schemes increase monotonically with the number of users due to the ability of these schemes to exploit multiuser diversity. Nevertheless, it can be observed that the average user system throughput of the proposed schemes grows faster with increasing number of users than that of the baseline schemes. In fact, compared to baseline scheme 1, since the proposed schemes utilize SIC for multiuser detection at the SUs, the co-channel interference is significantly reduced in the secondary network. Besides, the proposed schemes exploit the power domain to facilitate multiuser access which offers additional degrees of freedom for user scheduling and power allocation. Moreover, since baseline scheme 2 adopts random user allocation in the secondary network, it can only exploit the multiuser diversity of the PUs to improve the average user throughput, which leads to a marginal performance gain. On the other hand, both the proposed schemes and the baseline schemes achieve a lower average user system throughput for larger minimum data rate requirements. In fact, both the primary BS and the secondary BS have to allocate more power and frequency resources to the PUs to meet more stringent QoS requirements even if the conditions of the corresponding channels are poor. We note that the proposed suboptimal scheme achieves a similar performance as the proposed optimal scheme in all considered scenarios but entails a lower computational complexity.

VI. CONCLUSION

In this paper, we studied the power and subcarrier allocation design for cooperative cognitive relaying MC-NOMA systems. The resource allocation design was formulated as a mixed-integer non-convex optimization problem for the maximization of the weighted system throughout. The optimal resource allocation policy was obtained by optimally solving the formulated problem via monotonic optimization. Besides, we developed a low-complexity suboptimal scheme for finding a locally optimal solution in order to achieve a balance between optimality and computational complexity. Our simulation results unveiled that the proposed cognitive relaying MC-NOMA system achieves a significantly higher system throughput compared to traditional multicarrier cognitive relaying systems.

REFERENCES

[1] X. Chen, H. H. Chen, and W. Meng, "Cooperative Communications for Cognitive Radio Networks: From Theory to Applications," *IEEE Commun. Surveys & Tutorials*, vol. 16, no. 3, pp. 1180–1192, Third Quarter 2014.

[2] Y. Sun, D. W. K. Ng, N. Zlatanov, and R. Schober, "Robust Resource Allocation for Full-Duplex Cognitive Radio Systems," in *Proc. European Signal Process. Conf.*, Aug. 2016, pp. 773–777.

[3] Y. Han, A. Pandharipande, and S. H. Ting, "Cooperative Decode-and-Forward Relaying for Secondary Spectrum Access," *IEEE Trans. Wireless Commun.*, vol. 8, no. 10, pp. 4945–4950, Oct. 2009.

[4] R. Manna, R. H. Y. Louie, Y. Li, and B. Vucetic, "Cooperative Spectrum Sharing in Cognitive Radio Networks with Multiple Antennas," *IEEE Trans. Signal Process.*, vol. 59, no. 11, pp. 5509–5522, Nov. 2011.

[5] V. W. S. Wong, R. Schober, D. W. K. Ng, and L.-C. Wang, *Key Technologies for 5G Wireless Systems*, 1st ed. Cambridge University Press, 2017.

[6] Z. Wei, D. W. K. Ng, J. Yuan, and H. M. Wang, "Optimal Resource Allocation for Power-Efficient MC-NOMA with Imperfect Channel State Information," *IEEE Trans. Commun.*, vol. PP, no. 99, pp. 1–1, May 2017.

[7] Y. Sun, D. W. K. Ng, Z. Ding, and R. Schober, "Optimal Joint Power and Subcarrier Allocation for Full-Duplex Multicarrier Non-Orthogonal Multiple Access Systems," *IEEE Trans. Commun.*, vol. 65, no. 3, pp. 1077–1091, Mar. 2017.

[8] Z. Ding, P. Fan, and V. Poor, "Impact of User Pairing on 5G Nonorthogonal Multiple-Access Downlink Transmissions," *IEEE Trans. Veh. Technol.*, vol. 65, no. 8, pp. 6010–6023, Aug. 2016.

[9] L. Lv, Q. Ni, Z. Ding, and J. Chen, "Application of Non-Orthogonal Multiple Access in Cooperative Spectrum-Sharing Networks over Nakagami-m Fading Channels," *IEEE Trans. Veh. Technol.*, vol. 66, no. 6, pp. 5506–5511, Jun. 2017.

[10] L. Lv, J. Chen, Q. Ni, and Z. Ding, "Design of Cooperative Non-Orthogonal Multicast Cognitive Multiple Access for 5G Systems: User Scheduling and Performance Analysis," *IEEE Trans. Commun.*, vol. 65, no. 6, pp. 2641–2656, Jun. 2017.

[11] H. Tuy, "Monotonic Optimization: Problems and Solution Approaches," *SIAM J. Optim.*, vol. 11, no. 2, pp. 464–494, 2000.

[12] Y. J. A. Zhang, L. Qian, and J. Huang, "Monotonic Optimization in Communication and Networking Systems," *Found. Trends in Netw.*, vol. 7, no. 1, pp. 1–75, Oct. 2013.

[13] J. Zhang, L. Dai, R. Jiao, X. Li, and Y. Liu, "Performance Analysis of Relay Assisted Cooperative Non-Orthogonal Multiple Access Systems," *submitted to IEEE Wireless Commun. Lett.*, 2017.

[14] Y. Sun, D. W. K. Ng, and R. Schober, "Optimal Resource Allocation for Multicarrier MISO-NOMA Systems," in *Proc. IEEE Intern. Commun. Conf.*, May 2017, pp. 1–7.

[15] J. Lee and S. Leyffer, *Mixed Integer Nonlinear Programming*. Springer Science & Business Media, 2011.

[16] D. W. K. Ng, Y. Wu, and R. Schober, "Power Efficient Resource Allocation for Full-Duplex Radio Distributed Antenna Networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 4, pp. 2896–2911, Apr. 2016.

[17] Q. T. Dinh and M. Diehl, "Local Convergence of Sequential Convex Programming for Nonconvex Optimization," in *Recent Advances in Optimization and its Applications in Engineering*. Springer, 2010, pp. 93–102.

[18] M. Grant and S. Boyd, "CVX: Matlab Software for Disciplined Convex Programming, version 2.1," [Online] http://cvxr.com/cvx, Mar. 2014.