Resource Allocation in Virtualized CoMP-NOMA HetNets: Multi-Connectivity for Joint Transmission

Sepehr Rezvani, Nader Mokari, Senior Member, IEEE, Mohammad R. Javan, Senior Member, IEEE, and Eduard A. Jorswieck, Fellow, IEEE

Abstract

In this work, we design a generalized joint transmission coordinated multi-point (JT-CoMP)-non-orthogonal multiple access (NOMA) model for a virtualized multi-infrastructure network. In this model, all users can benefit from multiple joint transmissions of CoMP thanks to the multi-connectivity opportunity provided by wireless network virtualization (WNV). We propose an unlimited NOMA clustering (UNC) scheme, where the order of NOMA clusters is the maximum possible value (called global NOMA cluster). We show that UNC provides the maximum overall spectral efficiency in CoMP-NOMA with maximum successful interference cancellation (SIC) complexity at users. To strike a balance between spectral efficiency and SIC complexity, we propose a limited NOMA clustering (LNC), where the SIC is performed to a subset of the global NOMA cluster sets. We formulate the problem of joint power allocation and user association in UNC and LNC such that CoMP scheduling and NOMA
clustering are determined by the user association policy. Then, one globally and one locally optimal solutions are proposed for each problem based on mixed-integer monotonic optimization and sequential programming, respectively. Numerical assessments reveal that WNV and LNC improves users sum-rate and reduces users SIC complexity up to 65% and 45% compared to non-virtualized CoMP-NOMA and UNC, respectively.

**Index Terms**

Coordinated multi-point, NOMA, wireless network virtualization, global programming, monotonic optimization, sequential programming, convex optimization.

**I. INTRODUCTION**

Among various existing coordinated multi-point (CoMP) techniques for mitigating inter-cell interference (ICI) in multi-cell wireless networks, joint transmission CoMP (JT-CoMP) has attracted significant attention. In JT-CoMP, multiple base stations (BSs) are allowed to schedule/transmit the same message to a user over the same frequency band which facilitates ICI management and empowers the received signals at users [1]–[3]. In this work, the term CoMP is referred to JT-CoMP. By introducing the orthogonal multiple access (OMA) techniques on CoMP, the overall interference caused by coordinated BSs (CoMP-BSs) could be eliminated at the CoMP-user. However, due to the orthogonality of OMA, this technique restricts the coordination opportunities in CoMP [4], [5]. In this line, power-domain non-orthogonal multiple access (NOMA) has been introduced on CoMP, called CoMP-NOMA, where the resource blocks are shared between users under the successive interference cancellation (SIC) technique which improves both the spectral effectiveness and users connectivity [2]–[6].

In multi-infrastructure wireless networks operating in different protected frequency bands, each user is restricted to be subscribed to only one infrastructure provider (InP) [7], [8]. This

---

1In this work, the term NOMA is referred to power-domain NOMA.
restriction degrades the users connectivity, specifically for those who are close to co-located BSs belonging to different InPs. There has been numerous studies to design efficient methods for sharing InPs resources by means of wireless network virtualization (WNV) [7]–[9]. In WNV, InPs lease their scheduled resources to a number of virtual networks, also called mobile virtual network operators (MVNOs). Each MVNO acts as a service provider for its subscribed users based on the service level agreements (SLAs) between MVNOs and end-users [7]–[10]. In virtualized multi-infrastructure networks, co-located BSs of different InPs form a virtual BS (VBS). Hence, users with strict SLA could be connected to the nearby VBS to benefit from the multi-connectivity opportunity[2] provided by WNV improving the spectral/energy efficiency [9]. Accordingly, WNV can be introduced on multi-infrastructure CoMP-NOMA systems, where cell-center users could be connected to the nearby VBS saving more physical resources for the cell-edge users. Besides, nearby VBSs can jointly transmit/schedule signals to the cell-edge users, specifically ICI-prone users suffering from poor channel qualities providing better user fairness and massive connectivity. However, virtualized CoMP-NOMA needs a centralized resource management to fully utilize the benefits of WNV on multi-infrastructure CoMP-NOMA and fulfill global constraints, e.g., user scheduling, which may be impractical due to the isolated resource management between InPs. Software-defined networking (SDN) is the promising solution for this issue enabling separation of the control plane from the data plane, centralized controlling by means of connected switches and routers to all the network elements which improves the network flexibility and scalability [9]. Despite the huge potential of software-defined virtualized CoMP-NOMA (SV-CoMP-NOMA), resource management is not straightforward in this system, due to the following challenges:

1) **SIC Ordering**: Generally, SIC ordering in CoMP-NOMA systems is not straightforward.

   In CoMP-NOMA, each CoMP-user receives multiple signal powers from CoMP-BSs. In

---

[2] The term 'multi-connectivity' refers to the association of a user to multiple InPs over orthogonal bands by applying WNV.
this line, CoMP-users with worse channel qualities may receive a higher power than the non-CoMP-users with better channel qualities \[2, 3\]. To this end, based on the principle of NOMA, non-CoMP-users may not guarantee to decode and cancel the desired signals of CoMP-users meaning that CoMP-users may be NOMA cluster-head, i.e., the user which is able to cancel all the signals of other users in a NOMA cluster. Therefore, the traditional optimal SIC ordering in NOMA systems which follows only the ascending order of NOMA users’ channel gain may not be a good criterion in CoMP-NOMA systems \[2\].

2) **Joint Power Allocation, CoMP Scheduling, and NOMA Clustering:** In this system with heterogeneous SLAs, adopting CoMP transmission to only cell-edge users may not be beneficial for the system \[11, 12\]. Actually, cell-edge users with low SLAs do not need to be scheduled in CoMP transmission while some cell-center users with strict SLA may need CoMP transmission. In NOMA, it is verified that the power allocation is a key factor to enable SIC at users and achieve the target spectral efficiency \[13–15\]. Furthermore, in CoMP-NOMA, SIC ordering for all of the CoMP-users should be the same in all NOMA clusters \[2, 3\]. Therefore, the traditional power allocation for spectral efficiency maximization in multi-cell NOMA systems may not be efficient for other cells calling the design of a generalized CoMP-NOMA model with efficient joint power allocation, CoMP scheduling, i.e., determining the set of CoMP-users and their coordinated BSs (CoMP-BSs), and NOMA clustering.

3) **SIC Complexity at CoMP-users:** In CoMP-NOMA, each CoMP-user forms multiple local NOMA clusters through the network (depending on the number of CoMP-BSs), and the SIC is performed based on the union of these local NOMA cluster sets \[2, 3\] called global NOMA cluster. This operation will be more crucial for the NOMA cluster-head CoMP-users with larger numbers of CoMP-BSs. Hence, the traditional approach which limits the number of multiplexed users over the shared frequency band in each cell \[13, 16\] has no refined control on the order of global NOMA clusters due to the joint transmission of
CoMP. This calls the design of a low-complexity NOMA clustering scheme, where each CoMP-user performs SIC to a subset of potential users in its global NOMA cluster.

Resource allocation in CoMP-NOMA consists of three parts: CoMP scheduling which determines the set of CoMP and non-CoMP users with their CoMP-BSs, NOMA clustering of all users, and power allocation among users. An opportunistic NOMA clustering scheme for a group of CoMP-users at cell-edge is designed in [5] by adopting an efficient power allocation algorithm. A novel multi-tier NOMA scheme is proposed in [4] to serve CoMP-users with poor channel qualities by relaying signals to them. A selective-transmission strategy for determining the set of CoMP-BSs for a fixed power allocation strategy in CoMP-NOMA is proposed in [17]. The authors in [3] first design a CoMP-NOMA model, where CoMP scheduling and NOMA clustering are heuristically determined. Then, two centralized and distributed power allocation algorithms per NOMA cluster are proposed to maximize users spectral efficiency in the NOMA cluster. In the mentioned works, the joint transmission of CoMP is considered for only cell-edge users. For instance, in [3], the set of CoMP-users are determined based on the received signal strength (RSS) at users. And, users with weak RSS, i.e., cell-edge users, are scheduled for joint transmission. A number of research efforts addressed the benefits of joint transmission of CoMP-NOMA for both the cell-edge and cell-center users in improving overall spectral efficiency [11], [18], and outage probability [12]. In [11], the CoMP scheduling and NOMA clustering are heuristically determined based on the quality of service (QoS) requirements of users. Then, a locally optimal joint beamforming and power allocation is designed. The fundamental limits of introducing the mutual SIC technique on CoMP-NOMA in 2-user and 3-user systems is investigated in [18], where the users simultaneously cancel their corresponding interfering signals. In [12], a generalized CoMP-NOMA system is proposed, where all the users are considered as potential CoMP-users. It is shown that generalizing CoMP to all users improves the overall spectral efficiency. However, generalized CoMP-NOMA inherently increases the NOMA clustering orders.
when the number of CoMP-BSs grows. To reduce the order of NOMA clusters, a heuristic low-complexity \(3^{rd}\) NOMA clustering strategy based on the channel qualities is devised, where the order of CoMP-BSs is reduced. After defining the NOMA clusters, an optimal power allocation strategy per NOMA cluster is proposed. Since CoMP scheduling and NOMA clustering affect the interference level at users, the joint power allocation, CoMP scheduling, and NOMA clustering would result in the maximum overall spectral efficiency \([3]\). However, the combinatorial nature of user scheduling makes the joint strategy complicated and even infeasible \([3]\). Hence, addressing these policies jointly is still an open problem. Besides, the power allocation strategies in \([3]\), \([12]\) are devised for each NOMA cluster independently. These strategies may not be efficient for the users forming multiple NOMA clusters since for such users, the allocated powers through all the NOMA clusters should be optimized jointly. Also, all the prior studies on CoMP-NOMA considered a single InP. Therefore, the impact of isolation among co-located BSs, and applications of WNV in providing the multi-connectivity opportunity for CoMP-NOMA are not yet addressed.

In the current study, we consider a multi-infrastructure heterogeneous network (HetNet) consisting of multiple isolated InPs and apply our proposed SV-CoMP-NOMA system to this network. The main contributions are summarized as follows:

- We propose a generalized CoMP-NOMA model, where both the cell-edge and cell-center users are considered as potential CoMP-users, and the set of CoMP-BSs for each user is chosen through resource allocation optimization to achieve maximum users sum-rate.
- In this system, we design a new resource sharing scheme, where co-located BSs of different InPs form a VBS, and nearby VBSs can jointly transmit signals to a CoMP-user providing multiple joint transmissions over orthogonal bands.
- To fully utilize the benefits of NOMA on CoMP, we design an unlimited NOMA clustering

\[3^{rd}\] In this context, the term complexity is referred to the complexity of SIC which is directly proportional to the order of NOMA clusters.
(UNC) scheme, where each CoMP-user forms a global NOMA cluster consisting of all
users connected to at least one of its CoMP-BSs on the assigned frequency band. We show
that this scheme provides maximum achievable sum-rate while maximum order of NOMA
clusters in orthogonal bands increasing the SIC complexity at users.

- A low-complexity NOMA clustering scheme, called limited NOMA clustering (LNC), is
  also designed to reduce the order of NOMA clusters of CoMP-users. In this scheme, the
  SIC at CoMP-users is limited to a subset of users in their global NOMA clusters.

- To address the problem of power allocation, CoMP scheduling, and NOMA clustering jointly
  in UNC and LNC, we first formulate joint power allocation and user association problems
  to maximize users sum-rate subject to their QoS requirements. In these formulations, CoMP
  scheduling and NOMA clustering are determined with the user association indicator.

- The problem formulations are nonconvex and NP-hard. To solve each problem, we propose
two globally and locally optimal solutions. The globally optimal solution is based on mixed-
integer monotonic optimization. The locally optimal solution is based on the successive
convex approximation (SCA) algorithm. To apply this method, we first propose a series of
transformations to simplify the main problem.

- Numerical results show that the joint optimization of power allocation, CoMP scheduling,
  and NOMA clustering outperforms the existing power allocation algorithms up to 20%.
  Moreover, LNC reduces the SIC complexity of users in UNC up to 45%. Furthermore,
  applying WNV to CoMP-NOMA systems, and CoMP to virtualized NOMA systems show
  performance gains nearly 65% and 37%, respectively.

The rest of this paper is organized as follows: Section II describes the SV-CoMP-NOMA
system, and NOMA clustering and SIC ordering models, and then formulates the UNC and
LNC optimization problems. These problems are solved in Section III. Section IV provides the
simulation results. Finally, our conclusions are presented in Section V.
Fig. 1. Exemplary illustration of the SV-CoMP-NOMA system in a 2-infrastructure HetNet, where each user is able to be associated to multiple nearby VBSs through orthogonal bands.

II. SYSTEM MODEL

A. Network Model & SV-CoMP-NOMA System

We consider the downlink transmission of a multi-user multi-infrastructure HetNet as shown in Fig. 1. This network consists of multiple InPs each of which includes a specific set of single-antenna BSs, and a dedicated licensed wireless band ($W_i$ Hz for InP $i$) that is orthogonal to other InPs [8]. The set of InPs and the set of BSs owned by InP $i$ are denoted by $I = \{1, \cdots, I\}$ and $B_i = \{0, \cdots, B_i\}$, respectively. The set of $K$ single-antenna users is indicated by $K = \{1, \cdots, K\}$. Moreover, a hypervisor located on the top of InPs is responsible for collecting users information and virtualizing InPs resources [10]. In addition, a SDN controller with a global view of the network is responsible for the centralized resource management [9].

Here, we first propose a generalized and flexible CoMP-NOMA model, where each CoMP-user can be associated to a specific set of CoMP-BSs which may differ from other CoMP-users. Then, the CoMP-BSs simultaneously transmit the same data to the CoMP-user over the same frequency band [2], [3]. In this model, each user forms a specific global NOMA cluster over the assigned frequency band which is described in the following: First, let us denote the user
association indicator by $\theta_{i,b,k} \in \{0,1\}$, where if user $k$ is associated to the $b$th BS of InP $i$ (on frequency band $W_i$), $\theta_{i,b,k} = 1$ and otherwise, $\theta_{i,b,k} = 0$. In addition, assume that $\lambda_{i,k}$ is the SIC ordering of user $k$ on frequency bandwidth $W_i$, and $\lambda_{i,k} > \lambda_{i,k'}$ indicates that user $k$ has a higher SIC ordering than user $k'$ on bandwidth $W_i$. In this system, user $k$ with $\theta_{i,b,k} = 1$ forms a local NOMA cluster set $\Phi_{\text{Cell}}_{i,b,k}$ with all the users $k' \in K$ with $\theta_{i,b,k'} = 1$ and $\lambda_{i,k} > \lambda_{i,k'}$. Actually, $\Phi_{\text{Cell}}_{i,b,k}$ is the set of users associated to the same BS $b_i$ over the same band $W_i$ with lower SIC ordering such that their desired signals could be removed by user $k$. In other words, $\Phi_{\text{Cell}}_{i,b,k}$ indicates the set of potential users associated to BS $b_i$ for performing SIC at user $k$. Therefore, each CoMP-user $k$ forms a global NOMA cluster on bandwidth $W_i$ obtained by $\Phi_i,k = \bigcup_{b \in B_i} \Phi_{\text{Cell}}_{i,b,k}$. To this end, user $k'$ belongs to the global NOMA cluster of user $k$ on bandwidth $W_i$, i.e., $k' \in \Phi_i,k$, if and only if $\sum_{b \in B_i} \theta_{i,b,k} \theta_{i,b,k'} \geq 1$ and $\lambda_{i,k} > \lambda_{i,k'}$. Hence, if user $k' \in \Phi_i,k$, user $k$ is able to decode and cancel all the desired signals of user $k'$ on bandwidth $W_i$ [3]. For the case that $\sum_{b \in B_i} \theta_{i,b,k} \theta_{i,b,k'} \geq 1$ and $\lambda_{i,k} < \lambda_{i,k'}$, all the desired signals of user $k'$ are treated as Intra-NOMA-interference (INI) at user $k$. Obviously, if $\sum_{b \in B_i} \theta_{i,b,k} \theta_{i,b,k'} = 0$, i.e., $k' \notin \Phi_i,k$, all the desired signals of user $k'$ are treated as ICI at user $k$.

Due to the isolation among InPs, when WNV is not applied to the proposed CoMP-NOMA model, each CoMP-user can be assigned to only one InP. In other words, for the case that WNV is not applied, the following isolation constraint should be satisfied:

$$\theta_{i,b,k} + \theta_{i',b',k'} \leq 1, \forall k \in K, i, i' \in I, i' \neq i, b \in B_i, b' \in B_{i'}.$$  \hspace{1cm} (1)

Here, we apply WNV to the generalized CoMP-NOMA model resulting in the SV-CoMP-NOMA system. In this line, we design a resource sharing scheme, where the InPs resources are shared among $V$ MVNOs with the set of $V = \{1, \cdots, V\}$. Each MVNO $v$ acts as a service provider for its subscribed users in $K_v$. Moreover, each user is owned by only one MVNO, i.e., $\bigcup_{v \in V} K_v \triangleq K$

\footnote{For the non-CoMP-user with $\theta_{i,b,k} = 1$, since $\sum_{b \in B_i} \theta_{i,b,k} = 1$, $\Phi_i,k \triangleq \Phi_{\text{Cell}}_{i,b,k}$ meaning that the local NOMA cluster of a non-CoMP-user refers to its global NOMA cluster.}
Fig. 2. A two-infrastructure CoMP system with/without WNV. In these figures, it is assumed that messages M1 and M2 are sent by the CoMP-BSs of InPs 1 and 2, respectively.

and $\mathcal{K}_v \cap \mathcal{K}_{v'} = \emptyset, \forall v, v' \in \mathcal{V}, v \neq v'$ [19]. To reduce conflicts between MVNOs, a specific minimum data rate $R_{v}^{\text{sv}}$ is contracted between each MVNO $v$ and users in $\mathcal{K}_v$ [19], [20]. In this system, WNV breaks the isolation among InPs and provides the multi-connectivity opportunity with/without CoMP for all the users over orthogonal bands. Hence, each user can be associated to multiple BSs owned by different InPs. Indeed, the isolation constraint (1) will be removed. An exemplary illustration of our proposed resource sharing scheme in a virtualized CoMP system and its comparison to a non-virtualized CoMP system is shown in Fig. 2.

B. NOMA-User Clustering for Performing SIC

In the following, we propose two NOMA clustering schemes for performing SIC in the SV-CoMP-NOMA system.

1) Unlimited NOMA Clustering: In UNC, each user performs SIC to all the users belonging to its global NOMA cluster $\Phi_{i,k}$ over frequency band $W_i$. Actually, UNC fully utilizes the SIC opportunity provided by NOMA, where each user decodes and cancels all the signals of all potential users belonging to its global NOMA clusters over the assigned frequency bands. In this scheme, since the same data is transmitted by the CoMP-BSs to a CoMP-user over the same bandwidth, the CoMP-user does not experience any ICI by its CoMP-BSs on that bandwidth.
Fig. 3. A UNC-based single-carrier CoMP-NOMA system consisting of one InP including 3 BSs with 3 subscribed users. The signals with the same colour refer to the desired signals of a user.

However, the CoMP-user may do experience INI incurred by its CoMP-BSs depending on the SIC ordering of users. Let us denote the channel power gain from BS $b \in B_i$ to user $k$ by $h_{i,b,k}$. Moreover, the transmit power of BS $b \in B_i$ to user $k$ is indicated by $p_{i,b,k}$. Fig. 3 shows an exemplary single-carrier UNC-based CoMP-NOMA system consisting of 3 BSs owned by a single InP, and 3 users with a unique SIC ordering $\lambda_{1,3} > \lambda_{1,2} > \lambda_{1,1}$. In this system, the non-CoMP-user 1 forms only one local NOMA cluster set $\Phi_{\text{Cell}}^{\text{Cell}}_{1,1,1} = \{\}$ which is equal to its global NOMA cluster set $\Phi_{1,1} = \{\}$ due to the lowest SIC ordering. This user receives both the INI and ICI in the network. The INI power of user 1 is $(p_{1,1,2}h_{1,1,1} + p_{1,2,2}h_{1,2,1} + p_{1,3,2}h_{1,3,1})$ since $\sum_{b} \theta_{1,b,1} \theta_{1,b,3} = 0$. Besides, CoMP-user 2 forms three local NOMA cluster sets $\Phi_{\text{Cell}}^{\text{Cell}}_{1,1,2} = \{1\}$ and $\Phi_{\text{Cell}}^{\text{Cell}}_{1,2,2} = \Phi_{\text{Cell}}^{\text{Cell}}_{1,3,2} = \{\}$, and subsequently a global NOMA cluster $\Phi_{1,2} = \{1\}$. Also, user 2 does not experience any ICI since $\sum_{b} \theta_{1,b,2} \theta_{1,b,1}$ and $\sum_{b} \theta_{1,b,2} \theta_{1,b,3}$ are nonzero. However, user 2 does experience INI power $(p_{1,2,3}h_{1,2,2} + p_{1,3,3}h_{1,3,2})$, due to $\sum_{b} \theta_{1,b,2} \theta_{1,b,3} \geq 1$ and $\lambda_{1,2} < \lambda_{1,3}$. Finally, the NOMA cluster-head user 3 forms a global NOMA cluster $\Phi_{1,3} = \{2\}$. Since $\lambda_{1,3} > \lambda_{1,2}$ and $\sum_{b} \theta_{1,b,1} \theta_{1,b,3} = 0$, this user does not experience any INI. However, user 3 receives ICI.
power \((p_{1,1,1} h_{1,1,3})\). Accordingly, the SINR of user \(k\) on bandwidth \(W_i\) in the UNC model is given by

\[
\gamma_{i,k}^{\text{UNC}} = \frac{s_{i,k}}{I_{i,k}^{\text{UNC,INI}} + I_{i,k}^{\text{UNC,ICI}} + N_0 W_i},
\]

where \(s_{i,k} = \sum_{b \in B_i} \theta_{i,b,k} p_{i,b,k} h_{i,b,k}\) is the total received desired signal power at user \(k\) on bandwidth \(W_i\), \(I_{i,k}^{\text{UNC,INI}} = \sum_{k' \in K, \lambda_{i,k'} > \lambda_{i,k}} \min \left\{ \sum_{b \in B_i} \theta_{i,b,k} \theta_{i,b,k'}, 1 \right\} \sum_{b \in B_i} \theta_{i,b,k'} p_{i,b,k'} h_{i,b,k}\) is the INI at user \(k\) on bandwidth \(W_i\), \(I_{i,k}^{\text{UNC,ICI}} = \sum_{k' \in K, k' \neq k} (1 - \min \left\{ \sum_{b \in B_i} \theta_{i,b,k} \theta_{i,b,k'}, 1 \right\}) \sum_{b \in B_i} \theta_{i,b,k'} p_{i,b,k'} h_{i,b,k}\) is the ICI at user \(k\) on bandwidth \(W_i\), and \(N_0\) is the power spectral density (PSD) of the additive white gaussian noise (AWGN). Therefore, the spectral efficiency of user \(k\) on bandwidth \(W_i\) is \(r_{i,k}^{\text{UNC,SE}} = \log_2 \left(1 + \gamma_{i,k}^{\text{UNC}}\right)\). In UNC, user \(k\) can successfully decode and cancel the signals of user \(k'\) when the SINR at user \(k\) according to the signals (and interferences) of user \(k'\) is larger than or equal to the SINR at user \(k'\) for its own signals \([21], [22]\). In our model, the following SIC constraint should be satisfied:

\[
\frac{s_{i',k',k}^{\text{VP}}}{I_{i,k',k}^{\text{UNCI,INI}} + I_{i,k',k}^{\text{UNCI,ICI}} + N_0 W_i} \geq \min \left\{ \sum_{b \in B_i} \theta_{i,b,k} \theta_{i,b,k'}, 1 \right\} \gamma_{i',k'}^{\text{UNC}}, \quad \forall i \in I, k, k' \in K, \lambda_{i,k} > \lambda_{i,k'},
\]

where \(s_{i',k',k}^{\text{VP}} = \sum_{b \in B_i} \theta_{i,b,k'} p_{i,b,k'} h_{i,b,k}\) is the desired signal power of user \(k'\) on bandwidth \(W_i\) received at user \(k\), \(I_{i,k',k}^{\text{UNCI,INI}} = \sum_{k'' \in K, k'' \neq k} \min \left\{ \sum_{b \in B_i} \theta_{i,b,k} \theta_{i,b,k''}, 1 \right\} \sum_{b \in B_i} \theta_{i,b,k''} p_{i,b,k''} h_{i,b,k}\) denotes the INI power of user \(k'\) on bandwidth \(W_i\) received at user \(k\), and \(I_{i,k',k}^{\text{UNCI,ICI}} = \sum_{k'' \in K, k'' \neq k'} (1 - \min \left\{ \sum_{b \in B_i} \theta_{i,b,k} \theta_{i,b,k''}, 1 \right\}) \sum_{b \in B_i} \theta_{i,b,k''} p_{i,b,k''} h_{i,b,k}\) indicates the ICI power of user \(k'\) on bandwidth \(W_i\) received at user \(k\).

2) Limited NOMA Clustering: Although UNC provides the maximum possible interference cancellation of NOMA in our CoMP-NOMA model achieving to the maximum spectral efficiency, this scheme inherently increases the order of global NOMA cluster \(|\Phi_{i,k}|\) of CoMP-users, specifically when the number of CoMP-BSs increases. To reduce the order of NOMA clusters for performing SIC, we propose a LNC scheme, where each CoMP-user is restricted to perform SIC to the users belonging to only one of its local NOMA clusters \(|\Phi_{i,b,k}^{\text{Cell}}|\). Therefore, user \(k\)
choosing BS $b \in \mathcal{B}_i$ on bandwidth $W_i$ (or choosing local NOMA cluster set $\Phi_{i,b,k}^{Cell}$) can decode and cancel the signals of user $k'$ with $\lambda_{i,k} > \lambda_{i,k'}$ if and only if $k' \in \Phi_{i,b,k}^{Cell}$ or $\theta_{i,b,k'} = 1$.

For the case that $\theta_{i,b,k'} = 0$, all the desired signals of user $k'$ are treated as ICI at user $k$ even $k'$ is a potential user belonging to $\Phi_{i,k}$, i.e., $\sum_{b' \in \mathcal{B}_i, b' \neq b} \theta_{i,b',k'} \theta_{i,b,k'} \geq 1$. For instance, in Fig. 3, with another predefined SIC ordering $\lambda_{1,2} > \lambda_{1,3} > \lambda_{1,1}$ for LNC, if user 2 selects $\Phi_{i,b,k}^{Cell} = \{3\}$, it decodes and cancels the signals of user 3 (because $\theta_{1,2,2} = 1$). However, user 2 receives ICI power $(p_{1,1,1} h_{1,1,2})$ since $\theta_{1,2,2} = 0$. Besides, if user 2 selects $\Phi_{i,b,k}^{Cell} = \{1\}$, it receives ICI power $(p_{1,2,3} h_{1,2,2} + p_{1,3,3} h_{1,3,2})$ since $\theta_{1,2,2} = 0$. In the sequel, we first introduce a new binary variable $x_{i,b,k}$ called local NOMA cluster selection indicator, where if user $k$ selects the local NOMA cluster of BS $b \in \mathcal{B}_i$ on bandwidth $W_i$, $x_{i,b,k} = 1$, and otherwise, $x_{i,b,k} = 0$. To ensure that each user selects at most one local NOMA cluster on each bandwidth, the following constraint should be satisfied:

$$\sum_{b \in \mathcal{B}_i} x_{i,b,k} \leq 1, \forall i \in \mathcal{I}, k \in \mathcal{K}. \tag{4}$$

Obviously, user $k$ can select the local NOMA cluster of BS $b \in \mathcal{B}_i$ on bandwidth $W_i$ if it is associated with that BS. Therefore, we have

$$x_{i,b,k} \leq \theta_{i,b,k}, \forall i \in \mathcal{I}, b \in \mathcal{B}_i, k \in \mathcal{K}. \tag{5}$$

The SINR of user $k$ on bandwidth $W_i$ is thus given by

$$\gamma_{i,k}^{LNC} = \frac{s_{i,k}}{I_{i,k}^{LNC,INI} + I_{i,k}^{LNC,ICI} + N_0 W_i}, \tag{6}$$

where $I_{i,k}^{LNC,INI} = \sum_{b \in \mathcal{B}_i} x_{i,b,k} \sum_{k' \in \mathcal{K}} \theta_{i,b,k'} \sum_{b' \in \mathcal{B}_i} \theta_{i,b',k'} p_{i,b',k'} h_{i,b',k}$ is the INI at user $k$ on bandwidth $W_i$, and $I_{i,k}^{LNC,ICI} = \sum_{b \in \mathcal{B}_i} x_{i,b,k} \sum_{k' \in \mathcal{K}} \theta_{i,b,k'} (1 - \theta_{i,b,k'}) \sum_{b' \in \mathcal{B}_i} \theta_{i,b',k'} p_{i,b',k'} h_{i,b',k}$ is the ICI at user $k$ on bandwidth $W_i$. According to (4)-(6), the spectral efficiency of user $k$ on bandwidth $W_i$ is given by $r_{i,k}^{LNC,SE} = \log_2 (1 + \gamma_{i,k}^{LNC})$. Similar to (3), the SIC constraint of LNC is formulated as

$$\frac{s_{i,k}^{VP}}{I_{i,k'}^{LNC,INI} + I_{i,k'}^{LNC,ICI} + N_0 W_i} \geq x_{i,b,k} \theta_{i,b,k'} \gamma_{i,k'}^{LNC}, \forall i \in \mathcal{I}, b \in \mathcal{B}_i, k' \in \mathcal{K}, \lambda_{i,k} > \lambda_{i,k'}, \tag{7}$$
where \( I_{i,k',k}^{\text{LNC,INI}} \) denotes the INI power of user \( k' \) on bandwidth \( W_i \) received at user \( k \), and \( I_{i,k',k}^{\text{LNC,ICI}} \) indicates the ICI power of user \( k' \) on bandwidth \( W_i \) received at user \( k \).

C. SIC Ordering in SV-CoMP-NOMA

In the SV-CoMP-NOMA system, SIC ordering should be the same for all NOMA clusters of a CoMP-user [2], [3]. Finding an optimal SIC ordering jointly with power allocation and user association is very challenging and is not yet investigated by all the prior works on CoMP-NOMA. Here, we propose an efficient approach to determine SIC ordering prior to resource allocation optimization. In this method, each BS \( b \in B_i \) broadcasts a specific reference signal on bandwidth \( W_i \) such that its signal power is \( P_{i,b}^{\text{max}}/K \), where \( P_{i,b}^{\text{max}} \) is the maximum available transmit power of BS \( b \in B_i \). On the other hand, these signals are summed up at users and the SIC ordering at each bandwidth will be defined based on the ascending order of total received powers at users. In this line, we set \( \lambda_{i,k'} > \lambda_{i,k} \) if \( \sum_{b \in B_i} P_{i,b}^{\text{max}} h_{i,b,k}/K > \sum_{b \in B_i} P_{i,b}^{\text{max}} h_{i,b,k'}/K \).

D. Problem Formulations

In this network, the frequency band \( W_i \) may be different for each InP. In this way, the overall spectral efficiency of a user connected to different InPs would not correspond to its overall data rate. Here, we assume that the revenue of each MVNO comes from providing data rates for its subscribed users [9], [23] such that \( \omega_v \) units/bps denotes the unit price of revenue of MVNO \( v \) due to providing data rates for users in \( K_v \). In the following, we formulate the problem of maximizing total revenue of MVNOs corresponding to the weighted sum-rate of users. The data rate of user \( k \) in the UNC and LNC schemes can be obtained by \( r_{i,k}^{\text{UNC}} = \sum_{i \in I} W_i r_{i,k}^{\text{UNC,SE}} \) and \( r_{i,k}^{\text{LNC}} = \sum_{i \in I} W_i r_{i,k}^{\text{LNC,SE}} \), respectively. The UNC problem is formulated as follows:

\[
\text{UNC: } \max_{p, \theta} \sum_{v \in V} \sum_{k \in K_v} \omega_v r_k^{\text{UNC}} \quad (8a)
\]
s.t. (3),

\[ r_{k}^{\text{UNC}} \geq R_{v}^{\text{rsv}}, \forall v \in \mathcal{V}, k \in \mathcal{K}_{v}, \quad (8b) \]

\[ \sum_{k \in \mathcal{K}} p_{i,b,k} \leq P_{i,b}^{\text{max}}, \forall i \in \mathcal{I}, b \in \mathcal{B}_{i}, \quad (8c) \]

\[ \sum_{b \in \mathcal{B}_{i}} \theta_{i,b,k} \leq \Psi_{i}^{\text{max}}, \forall i \in \mathcal{I}, k \in \mathcal{K}, \quad (8d) \]

\[ \theta_{i,b,k} \in \{0, 1\}, \forall i \in \mathcal{I}, b \in \mathcal{B}_{i}, k \in \mathcal{K}, \quad (8e) \]

\[ p_{i,b,k} \geq 0, \forall i \in \mathcal{I}, b \in \mathcal{B}_{i}, k \in \mathcal{K}, \quad (8f) \]

where \( p = [p_{i,b,k}] \) and \( \theta = [\theta_{i,b,k}] \). Furthermore, (8b) is the minimum required data rate constraint of user \( k \in \mathcal{K}_{v} \), and (8c) is the maximum available power constraint of each BS. Moreover, (8d) indicates that each user \( k \in \mathcal{K} \) can be associated to at most \( \Psi_{i}^{\text{max}} \) BSs in \( \text{InP} i \) [3]. The main advantages of restricting the order of CoMP-BSs for each CoMP-user are listed as follows:

1) alleviating the backhaul traffic, due to the joint transmission of CoMP [24]; 2) reducing the complexity of synchronization at CoMP-BSs; 3) decreasing the order of NOMA clusters which reduces: 1. SIC complexity at CoMP-users; 2. superposition coding at CoMP-BSs; 3. the negative side effect of SIC on users sum-rate [3], [12]. The LNC problem is formulated as

**LNC:**

\[
\begin{align*}
\text{max}_{p, \theta, x} & \sum_{v \in \mathcal{V}} \sum_{k \in \mathcal{K}_{v}} \omega_{v} r_{k}^{\text{LNC}} \\
\text{s.t.} & \quad (4), (5), (7), (8c)-(8f), \\
& \quad r_{k}^{\text{LNC}} \geq R_{v}^{\text{rsv}}, \forall v \in \mathcal{V}, k \in \mathcal{K}_{v}, \\
& \quad x_{i,b,k} \in \{0, 1\}, \forall i \in \mathcal{I}, b \in \mathcal{B}_{i}, k \in \mathcal{K}, \\
\end{align*}
\]

where \( x = [x_{i,b,k}], \forall i \in \mathcal{I}, b \in \mathcal{B}_{i}, k \in \mathcal{K} \). Compared to the UNC problem, LNC adds a new binary optimization variable \( x \) with two integer linear inequality constraints (4) and (5).
III. JOINT POWER ALLOCATION AND USER ASSIGNMENT ALGORITHMS

The problems (8) and (9) are classified as MINLP which are intractable and NP-hard [3], [25], [26]. Additionally, as mentioned before, the existing optimal solutions for single-cell or multi-cell NOMA systems cannot be directly applied to the CoMP-NOMA systems [3].

A. Solution Algorithms for the UNC Problem

1) Global Optimality: Mixed-Integer Monotonic Optimization: Here, we find a globally optimal solution for the problem (8) by proposing a mixed-integer monotonic program. The basic idea of this approach is reducing the exploration area for finding the global optimal solution of a monotonic problem to its outer boundary which reduces the computational complexity, and provides a guaranteed convergence. The poly block or branch-reduce-and-bound algorithms can solve these problems [21], [27]. The monotonic optimization can solve problems, where the objective function is monotone and constraints are the intersection of their normal and co-normal sets [21], [27]. However, (8) is not a monotonic problem in canonical form because of the following issues: 1) The user association variable $\theta$ in (8e) results in a non-continuous domain in (8); 2) The rate function in (8a) and (8b) cannot be directly transformed into the difference of two increasing functions, because of the non-increasing term $\left(1 - \min_{b \in B_i} \{ \theta_{i,b,k} \theta_{i,b,k} ', 1 \} \right)$ in $I_{i,k}^{\text{UNC,ICI}}$ in (2) with respect to $\theta$; 3) The objective function (8a) is not monotonic, since the SINR fraction in (2) is increasing neither in $p$ nor in $\theta$; 4) The constraint sets in (3) and (8b) are not guaranteed to be the intersection of normal and co-normal sets, since the difference of two increasing functions is in general nonincreasing. However, (8) shows a hidden monotonicity structure after issues 1 and 2 in above are solved. To tackle the combinatorial nature of (8), $\theta$ should be transformed into a continuous variable. Unfortunately, in contrast to prior works on OMA [28], [29], we cannot relax $\theta$ to a continuous variable between 0 and 1 by using the time sharing method. Actually, in downlink NOMA systems, we need to determine which part of a frame time is assigned to which user since the superposition coding and SIC (at BSs
and users, respectively) are performed according to the set of users receiving signals on the same frequency band at the same time. To overcome this challenge, we transform (8e) into the following equivalent constraint sets as

\[ \theta_{i,b,k} \leq \theta_{i,b,k}', \quad 0 \leq \theta_{i,b,k} \leq 1. \]  

(10)

Since the square of each variable in \((0, 1)\) is smaller than that variable, with (10), the variable \(\theta_{i,b,k}\) can only take zero or one while it has a continuous domain \([0, 1]\). By substituting (8e) with (10), the problem (8) is equivalently transformed into a problem with a continuous domain. In contrast to the prior works [21], [27], the data rate function \(r_{k}^{\text{UNC}}\) cannot be directly transformed into the difference of two increasing functions. To tackle this, we first substitute the term \(1 - \min\{\sum_{b \in B_{i}} \theta_{i,b,k} \theta_{i,b,k}', 1\}\) in (2) with a new auxiliary variable \(\alpha_{i,k,k'} \in [0, 1]\) by adding the following constraints

\[ \alpha_{i,k,k'} \leq 1 - \theta_{i,b,k} \theta_{i,b,k}', \quad 0 \leq \alpha_{i,k,k'} \leq 1. \]  

(11)

\[ \alpha_{i,k,k'} \geq 1 - \sum_{b \in B_{i}} \theta_{i,b,k} \theta_{i,b,k}'. \]  

(12)

According to (11) and (12), the SINR of user \(k\) on bandwidth \(W_{i}\) can be rewritten as

\[ \hat{\gamma}_{i,k}^{\text{UNC}} = \frac{s_{i,k}}{I_{i,k}^{\text{UNC,INI}} + I_{i,k}^{\text{UNC,ICI}} + N_{0}W_{i}}, \]  

(13)

where \(I_{i,k}^{\text{UNC,ICI}} = \sum_{k' \in K, k' \neq k} \alpha_{i,k,k'} \sum_{b \in B_{i}} \theta_{i,b,k} p_{i,b,k} h_{i,b,k}\). The spectral efficiency of user \(k\) on bandwidth \(W_{i}\) is rewritten as \(\hat{r}_{i,k}^{\text{UNC,SE}} = \log_{2} (1 + \hat{\gamma}_{i,k}^{\text{UNC}})\), and the data rate of user \(k\) is \(\hat{r}_{k}^{\text{UNC}} = \sum_{i \in I} W_{i} \hat{r}_{i,k}^{\text{UNC,SE}}\).

Accordingly, (8) is rewritten as

\[
\begin{align*}
\max_{\theta, p, \alpha} & \sum_{v \in V} \sum_{k \in K_{v}} \omega_{v} r_{k}^{\text{UNC}} \\
\text{s.t.} \quad & (8c), \quad (8d), \quad (8f), \quad (10)-(12), \\
& \frac{s_{i,k}^{\text{VP}}}{I_{i,k',k}^{\text{UNC,INI}} + I_{i,k',k}^{\text{UNC,ICI}} + N_{0}W_{i}} \geq \min \left\{ \sum_{b \in B_{i}} \theta_{i,b,k} \theta_{i,b,k}', 1 \right\} \hat{\gamma}_{i,k'}^{\text{UNC}}, \quad \forall i \in I, k, k' \in K, \lambda_{i,k} > \lambda_{i,k'},
\end{align*}
\]  

(14a)

(14b)
\[ \hat{r}_k^{\text{UNC}} \geq R_v^{\text{sv}}, \forall v \in \mathcal{V}, k \in \mathcal{K}_v, \]  

where \( \hat{r}_{i,k,k'}^{\text{UNC,ICI}} = \sum_{k'' \in \mathcal{K}_v, \ k'' \neq k'} \alpha_{i,k,k'} \sum_{b \in B_i} \theta_{i,b,k''} p_{i,b,k''} h_{i,b,k} \) and \( \alpha = [\alpha_{i,k,k'}] \). Problem (14) exhibits a hidden monotonicity structure as shown in the following theorem:

**Theorem 1.** Problem (14) can be expressed as a monotonic problem in canonical form.

**Proof:** Please see Appendix A.

2) **First-Order Optimality: Sequential Programming:** Despite the global optimality of the proposed mixed-integer monotonic program, the complexity of this method is still exponential in the number of optimization variables. Indeed, this algorithm can be considered as a benchmark for any low-complexity yet suboptimal method. Here, we apply the SCA algorithm which is a locally optimal solution with a polynomial time complexity [15], [23], [26], [30], [31]. Note that the SCA algorithm cannot be directly applied to (8), because of: 1) Combinatorial nature of (8), due to binary variable \( \theta \); 2) Multiplication of \( \theta \) and \( p \) in the objective function (8a), and constraints (3) and (8b) (please see (2)); 3) The term \((1 - \min\{\cdot\})\) in (2) with respect to \( \theta \); 4) The term \( \theta_{i,b,k} \theta_{i,b,k'} \) in \( \min \{ \sum_{b \in B_i} \theta_{i,b,k} \theta_{i,b,k'}, 1 \} \) and \((1 - \min \{ \sum_{b \in B_i} \theta_{i,b,k} \theta_{i,b,k'}, 1 \})\) (please see (2)); 5) Multiplications of \( \min \{ \sum_{b \in B_i} \theta_{i,b,k} \theta_{i,b,k'}, 1 \} \) and \((1 - \min \{ \sum_{b \in B_i} \theta_{i,b,k} \theta_{i,b,k'}, 1 \})\) with \( \theta_{i,b,k} p_{i,b,k'} \) in (2). However, (8) can be transformed into an equivalent form which can be solved by directly applying the SCA algorithm. The series of equivalent transformations of this MINLP problem is presented in Appendix B. After these transformations, we apply the SCA algorithm with the difference of convex (DC) approximation method to the transformed nonconvex problem as follows: We first initialize the approximation parameters. After that, the convex approximated problem is solved. These iterations are repeated until the convergence is achieved. The pseudo code of the SCA algorithm is presented in Alg. 1. The derivations of SCA for solving the transformed problem (41) is presented in Appendix C. Moreover, we analytically show that the proposed SCA algorithm generates a sequence of improved solutions in Appendix D. Hence, we
Algorithm 1 SCA with DC programming.

1: Initialize $\vartheta^{(0)}$, the maximum number of iterations $L_{\text{max}}$, and penalty factor $\eta \gg 1$.

   repeat

2: Find $\vartheta^{(l)}$ by solving the convex approximated form of (41) for a given $\vartheta^{\text{old}}$.
3: Set $\vartheta^{\text{old}} = \vartheta^{(l)}$, and store it.
4: Set $l = l + 1$

   Until Convergence of (41a) or $l = L_{\text{max}}$.
5: $\vartheta^* = \vartheta^{(l)}$ is the output of the algorithm.
6: $\theta^*$ and $p^*$ are adopted for the network.

prove that SCA converges to a stationary point which is a local maximum of (8).

B. Solution Algorithms for the LNC Problem

1) Global Optimality: Mixed-Integer Monotonic Optimization: The problems (8) and (9) have similar structure and nonconvexity challenges. Hence, to find a globally optimal solution for (9), we modify our proposed mixed-integer monotonic optimization for solving (8) to be applied to (9). In this line, we show how (9) can be equivalently transformed into a monotonic optimization problem in canonical form in Appendix E.

2) First-Order Optimality: Sequential Programming: Since (8) and (9) have the same structure, the proposed SCA algorithm for solving (9) is similar to that of proposed for solving (8). Due to the space limitation, the presentation of the proposed SCA algorithm with DC programming for solving (9) is not included here. It can be easily proved that this algorithm also converges to a locally optimal solution.
IV. Simulation Results

In this section, we investigate the performance of our proposed UNC and LNC schemes in the SV-CoMP-NOMA system with different resource allocation strategies.

A. Simulation Settings

Here, we consider 2 InPs each having one MBS and 4 femto BSs (FBSs). We assume that the BSs of different InPs are co-located. Actually, we have a virtual MBS (VMBS) and 4 virtual FBSs (VFBSs). The VMBS is positioned at the center of a circular area (macro-cell), and VFBSs are positioned in coordinates (femto-cells) $300 \angle 0^\circ$, $300 \angle 22.5^\circ$, $300 \angle 67.5^\circ$, and $300 \angle 90^\circ$ [30]. Assume that 6 users are uniformly (and independently) distributed in the area of each femto-cell with radii of 80 m [6]. Fig. 4 shows the network topology with an exemplary user placement.

Following the Third Generation Partnership Project (3GPP) Long Term Evolution-Advanced (LTE-A), the orthogonal wireless bandwidth of each InP is set to $W_i = 20$ MHz with a carrier frequency of 2 GHz [8]. The wireless fading channels include both the large-scale and small-scale fading. The large-scale fading is modeled as $128.1 + 37.6 \log_{10} d_{i,b,k}$ in dB, where $d_{i,b,k}$
is the distance from BS $b \in B_i$ to user $k$ in $K_m^3$. The small-scale fading is modeled as independent and identically distributed (i.i.d.) Rayleigh fading with zero mean and variance 1. The PSD of AWGN is -174 dBm/Hz $^{[15]}$. The transmit power of each MBS and each FBS are set to 46 dBm and 30 dBm, respectively $^3$. Without loss of generality, we assume that the minimum required data rate of users is 8 Mbps$^{[5]}$. Then, we set $\omega_v = 1$ unit/bps for each $v \in V$.

In the following, we investigate the performance of our proposed UNC and LNC schemes in different NOMA systems equipped with/without WNV and CoMP. To this end, we investigate the performance gains of WNV and CoMP by comparing our proposed SV-CoMP-NOMA system with the following systems:

- **Non-Virtualized CoMP (NoWNV-CoMP):** In this system, each user can be associated to only one InP, due to the isolation among InPs (For more details, please see Subsection II-A). Therefore, constraint (1) is added to the resource allocation problems.

- **Virtualized Non-CoMP (WNV-NoCoMP):** In this system, each user can be associated to only one BS at each InP. Hence, the joint transmission of CoMP is eliminated. However, each user could benefit from the multi-connectivity opportunity. In this way, the following constraint should be satisfied: $\sum_{b \in B_i} \theta_{i,b,k} \leq 1$, $\forall i \in I, k \in K$.

- **Non-Virtualized Non-CoMP (NoWNV-NoCoMP):** In this system, each user can be associated to only one BS through the network. Hence, the following constraint should be satisfied: $\sum_{i \in I} \sum_{b \in B_i} \theta_{i,b,k} \leq 1$, $\forall k \in K$.

To investigate the benefits of optimizing CoMP scheduling and NOMA clustering jointly with power allocation in the UNC and LNC schemes, we compare our proposed joint strategy with a power allocation approach in which the NOMA clustering and CoMP scheduling are predefined (actually $\theta$ is predefined) based on the RSS at users $^3$. In this line, for LNC, we

$^5$In our system, WNV breaks the isolation between InPs. Therefore, the number of MVNOs does not impact on the system performance when all the users have the same SLAs and rewards.

July 21, 2020 DRAFT
also need to apply a heuristic approach (as a benchmark) determining \(x\) before power allocation optimization. Since this scheme is not yet investigated in the literature, we propose a heuristic approach to determine \(x\) according to the determined \(\theta\) (based on the RSS at users \([3]\)) as follows: First, we note that \(x\) should satisfy constraint \([5]\). Furthermore, the choice of \(x\) affects the INI power at users which impacts on the ICI power. In this approach, each CoMP-user selects the local NOMA cluster (among associated BSs) which results in the lowest interference power at the user. According to \([6]\), for each user \(k\) over bandwidth \(W_i\), we set \(x_{i,b,k} = 1\) if

\[
\begin{aligned}
&b = \arg\min_{b \in B_i} \left\{ \sum_{b' \in B_i} \sum_{i' \in K, i' > i} \theta_{i',b',k'} p_{i',b',k'} h_{i',b',k'} + \sum_{b' \in B_i} \theta_{i',b'} \left( 1 - \theta_{i,b,k} \right) \theta_{i,b,k} p_{i,b,k} h_{i,b,k} \right\}
&\quad \text{and} \quad \theta_{i,b,k} = 1.
\end{aligned}
\]

The value of \(p\) is determined based on the equal power allocation strategy.

It is noteworthy that the heuristic approaches mentioned above are also used for initializing parameters in our proposed sequential programming algorithms.

We also investigate the impact of number of users, SLAs, and transmit power of FBSs on the performance of SV-CoMP-NOMA. Last but not least, we compare the performance of our proposed locally and globally optimal solutions for the UNC and LNC schemes.

### B. Convergence Speed

Fig. \([5]\) investigates the convergence speed of our proposed SCA algorithms for the UNC and LNC schemes. As shown, these iterative algorithms converge to stable values in maximum 6 iterations. It is noteworthy that after only 3 and 4 iterations, SCA achieves to over 90% and 96% of its upper-bound value (the dash-lines refer to the upper-bound solution of SCA), respectively in both the UNC and LNC schemes. Accordingly, our proposed algorithms with the fast convergence speed could be good candidate solutions for real-time SV-CoMP-NOMA systems. Also, we set the maximum SCA iteration number to 4 in the following numerical results. In Fig. \([5]\) it can be observed that UNC outperforms LNC in terms of users sum-rate. Here, at the converged points, UNC has a performance gain up to 17% compared to LNC. This is due to the fact that in LNC, each user \(k\) performs SIC (on each bandwidth \(W_i\)) to only the signals of users belonging to the
selected local NOMA cluster set $\Phi_{i,b,k}^{\text{Cell}}$. Since $\Phi_{i,b,k}^{\text{Cell}} \subseteq \Phi_{i,k}$, the signals of users belonging to $\Phi_{i,k}$ while do not belong to $\Phi_{i,b,k}^{\text{Cell}}$ are treated as ICI at user $k$. In UNC, all the signals of users belonging to its global NOMA cluster sets $\Phi_{i,k}$ will be decoded and removed by user $k$.

We also compare our proposed joint strategy with the power allocation strategy alone for a predefined user scheduling described in Subsection IV-A. Fig. 5 shows that the joint optimization of power allocation and user scheduling improves users sum-rate up to 20% compared to the power allocation optimization for predefined NOMA clustering and CoMP scheduling policies. It is noteworthy that the convergence speed of the power allocation optimization alone for a fixed $\theta$ and $\alpha$ would be faster than the joint optimization algorithm, which imposes additional auxiliary variables and constraints due to the series of transformations discussed in Subsection III-A2. However, the gaps between the convergence speed of our proposed algorithm and benchmark are pretty low and negligible compared to the performance gaps.
C. Impact of Number of Users and Service Level Agreements

Fig. 6 investigates the impact of number of users on the system performance for different SLAs in the UNC-based SV-CoMP-NOMA system. From Fig. 6(a), it can be observed that for smaller order of number of users, increasing the number of users improves the users sum-rate. This is due to the fact that when the number of users is small enough, the network can efficiently exploit the multi-user diversity. However, when the number of users keeps increasing, the system needs to allocate its resources to a larger number of users with poor channel conditions, due to the SLA constraints in (8b). The performance loss due to the restriction of the flexibility of resource allocation is dominant compared to the multi-user diversity gain, when the number of users is large enough. As a result, this increment leads to sum-rate degradation shown in 6(a) specifically when the number of users is large enough. Inversely, in Fig. 6(b) we show

---

6Our proposed LNC model is a special case of UNC. Therefore, SLAs have the same impact in both the UNC and LNC schemes. To avoid duplicated presentations, we present the impact of SLAs on only the UNC model.
Fig. 7. Sum-rate of users vs. number of users per femto-cell for different schemes with/without CoMP and WNV, when $R_v^{rv} = 8$ Mbps.

the impact of SLA on users sum-rate for different number of users. According to the above discussions, increasing $R_v^{rv}$ degrades the users sum-rate. However, when the number of users is small enough, SLA has not a significant impact on the users sum-rate. Actually, the impact of SLA would be more crucial for larger number of users.

D. Effect of WNV and CoMP on the UNC and LNC Schemes

Fig. 7 shows the impact of the number of users in different scenarios with/without WNV and CoMP in UNC and LNC. It can be observed that UNC always outperforms LNC in terms of sum-rate of users, due to the reasons presented in Subsection IV-B. In the non-CoMP systems, since each user is associated to only one cell, the global NOMA cluster of users is equal to their local NOMA clusters. Hence, the performance of UNC and LNC is the same.

From Fig. 7, it can be observed that applying WNV to multi-infrastructure CoMP-NOMA systems would result in a performance gain up to 65%. This significant performance gain is achieved due to the following two unique advantages of WNV: 1) Providing multi-connectivity opportunity for users who are not scheduled for joint transmission of CoMP (e.g., strong users);
2) Providing multiple joint transmissions of CoMP from nearby VBSs over orthogonal bands (e.g., weak users). The first advantage has more impact on maximizing users sum-rate, and the second one can improve fairness among users. Besides, applying CoMP to the virtualized multi-infrastructure NOMA systems would result in a performance gain near to 37%.

E. SIC Complexity of UNC and LNC Schemes

The SIC complexity at each user is directly proportional to the number of NOMA users decoded and canceled by that user (called order of NOMA cluster) over the shared wireless band. Different metrics could be considered as SIC complexity cost of users. Here, we consider the following two metrics: 1) SIC energy consumption; 2) Complexity of users hardware for decoding and canceling signals of multiple users in a NOMA cluster. The SIC energy consumption at each user corresponds to the total number of users decoded and canceled by that user. Therefore, we consider the total number of users decoded and canceled by each user as SIC energy consumption of that user. In this paper, we consider a simplified NOMA-layer metric as SIC complexity of users, where the SIC complexity is evaluated by the order of NOMA clusters [12], [13], [15]. The total number of users decoded and canceled by user $k$ in UNC and LNC can be obtained by $\sum_{i \in I} |\Phi_{i,k}|$, and $\sum_{i \in I} \sum_{b \in B_i} x_{i,b,k} |\Phi_{i,b,k}^{Cell}|$, respectively. The complexity of users hardware for performing SIC is directly proportional to the maximum number of users decoded and canceled by that user among orthogonal bands. Therefore, we consider this metric as complexity of users hardware for performing SIC. The maximum number of users decoded and canceled by user $k$ in UNC and LNC can be obtained by $\max_{i \in I} |\Phi_{i,k}|$, and $\max_{i \in I} \sum_{b \in B_i} x_{i,b,k} |\Phi_{i,b,k}^{Cell}|$, respectively.

It is noteworthy that for the non-virtualized CoMP-NOMA system, the total and maximum number of users is equal, since each user is assigned to only one InP. In other word, each user forms only one NOMA cluster, and subsequently we have $\sum_{i \in I} |\Phi_{i,k}| = \max_{i \in I} |\Phi_{i,k}|$ in UNC, and $\sum_{i \in I} \sum_{b \in B_i} x_{i,b,k} |\Phi_{i,b,k}^{Cell}| = \max_{i \in I} \sum_{b \in B_i} x_{i,b,k} |\Phi_{i,b,k}^{Cell}|$ in LNC.

Fig. 8 demonstrates the average complexity costs of performing SIC at each user in two
users energy consumption and users hardware metrics described in above. As shown, increasing the number of users inherently increases the size of NOMA clusters in single-carrier NOMA systems. In this way, the number of decoded and canceled users increases shown in Fig. 8. Despite the huge potential of WNV in improving users sum-rate in multi-infrastructure CoMP-NOMA systems (shown in Fig. 7), this technology inherently increases the number of NOMA clusters at each user. This is due to breaking of isolation among InPs by WNV resulting in multiple NOMA clusters for each user over orthogonal frequency bands. From Fig. 8 it can be observed that WNV inherently increases the SIC complexity of users nearly 168% and 307% in LNC and UNC schemes, respectively. As a result, SIC complexity is more challenging in the SV-CoMP-NOMA system compared to the non-virtualized one, specifically for the larger number of users. More importantly, it can be observed that our proposed LNC has a reduced SIC complexity up to 45% compared to UNC. However, the order of NOMA clusters are still large, due to the single-carrier nature of our system (at each InP). To overcome this issue, the multi-carrier technology can be introduced on SV-CoMP-NOMA [13], [15], where each sub-band is
shared among a limited number of users in a cell. By adopting this technology to LNC, the order of NOMA clusters will be reduced to the NOMA systems without CoMP. However, the multi-carrier systems inherently increase the computational complexity of the central controller on the order of number of sub-bands. The trade-off between the computational complexity of the central controller and the users SIC complexity could be considered as a future work.

F. Optimality Gap: Algorithm Performance vs. Computational Complexity

As mentioned before, the complexity of our globally optimal solution based on monotonic program is still exponential on the number of optimization variables. Here, we first discuss about the computational complexity of our proposed SCA algorithms. Suppose that we solve the convex approximated form of the UNC problem at SCA iteration $i$ by using the barrier method with inner Newton’s method to achieve an $\epsilon$-suboptimal solution. The number of barrier (outer) iterations required to achieve $\frac{m}{t} = \epsilon$-suboptimality is exactly $\Upsilon_i = \left\lceil \frac{\log(m/(\epsilon t^{(0)}))}{\log \mu} \right\rceil$, where $m$ is the total number of inequality constraints, $t^{(0)}$ is the initial accuracy parameter for approximating the functions in inequality constraints in standard form, and $\mu$ is the step size for updating the accuracy parameter $t$ [32]. The number of inner Newton’s iterations depends on $\mu$ and how good is the initial points at each barrier iteration. [32]. If the SCA algorithm converges to a locally optimal solution after $\kappa$ iterations, the total number of barrier iterations is $\sum_{i=1}^{\kappa} \Upsilon_i$.

Fig. 9 shows sum-rate of users versus maximum power of FBSs in UNC and LNC with our proposed globally and locally optimal solutions. In this simulation, due to the high computational complexity of the monotonic optimization, we consider a small scale network [27] with one InP including a single MBS and 2 FBSs in coordinates $0^\circ$, $300^\circ$, and $300^\circ 22.5^\circ$, respectively. In each femto-cell, we uniformly distribute 2 users [27] (the total number of users is 4) with the same simulation settings as Subsection IV-A. From Fig. 9 it is shown that the optimality gaps are less than 7.5% verifying the efficiency of our proposed locally optimal solutions.
Fig. 9. Sum-rate of users vs. transmit power of each femto-cell for globally and locally optimal algorithms in the UNC and LNC schemes.

V. CONCLUDING REMARKS

In this work, we designed a generalized CoMP-NOMA model, where all the cell-edge and cell-center users can benefit from joint transmission of CoMP with specific set of CoMP-BSs. In this model, we devised two NOMA clustering models as UNC and LNC, where UNC performs SIC to all the potential users with lower SIC orders. Besides, LNC performs SIC to only a subset of potential users which significantly reduces the SIC complexity at users. Here, we proposed two globally and locally optimal solutions for the problem of finding joint power allocation, CoMP scheduling, and NOMA clustering strategies. We also investigated the benefits and challenges of applying WNV in multi-infrastructure CoMP-NOMA systems. In simulation results, we observed that our proposed LNC scheme reduces the SIC complexity of users up to 45% compared to UNC. Moreover, it is shown that WNV significantly improves users sum-rate by breaking isolation among InPs while increasing the number of NOMA clusters at each user.
APPENDIX A

CANONICAL TRANSFORMATION OF (14)

Observe that (14a) can be equivalently rewritten as \( q^+(\theta, p, \alpha) - q^-(\theta, p, \alpha) \), wherein \( q^+(\theta, p, \alpha) \) and \( q^-(\theta, p, \alpha) \) are increasing in all optimization variables and given by

\[
q^+(\theta, p, \alpha) = \sum_{v \in V} \sum_{k \in K_v} \omega_v \sum_{i \in I} W_i \log_2 \left( I_{i,k}^{\text{UNCI,INI}} + I_{i,k}^{\text{UNCI,ICI}} + N_0 W_i + s_{i,k} \right),
\]

and

\[
q^-(\theta, p, \alpha) = \sum_{v \in V} \sum_{k \in K_v} \omega_v \sum_{i \in I} W_i \log_2 \left( I_{i,k}^{\text{UNCI,INI}} + I_{i,k}^{\text{UNCI,ICI}} + N_0 W_i \right).
\]

Then, we define \( p_{\text{max}} = \{ p_{\theta, b, k}^{\text{mask}} \}, \forall i, b, k, \theta_{\text{max}} = \{ \theta_{i, b, k}^{\text{mask}} \}, \forall i, b, k, \) and \( \alpha_{\text{max}} = \{ \alpha_{i, k, k'}^{\text{mask}} \}, \forall i, k, k' \neq k \), where \( p_{i, b, k}^{\theta}, \theta_{i, b, k}, \) and \( \alpha_{i, k, k'}^{\theta} \) are the maximum possible values that \( p_{i, b, k}, \theta_{i, b, k}, \) and \( \alpha_{i, k, k'} \) can take. Next, we define a new auxiliary variable \( s_0 = q^-(\theta_{\text{max}}, p_{\text{max}}, \alpha_{\text{max}}) - q^-(\theta, p, \alpha) \). Accordingly, (14) can be rewritten as

\[
\max_{\theta, p, \alpha, s_0} q^+(\theta, p, \alpha) + s_0
\]

s.t. (8c), (8d), (8f), (10)-(12), (14b), (14c),

\[
0 \leq s_0 + q^-(\theta, p, \alpha) \leq q^-(\theta_{\text{max}}, p_{\text{max}}, \alpha_{\text{max}}),
\]

\[
0 \leq s_0 \leq q^-(\theta_{\text{max}}, p_{\text{max}}, \alpha_{\text{max}}) - q^-(0_{i, b, k}, 0_{i, b, k}, 0_{i, k, k'}).\]

Problem (17) is not a monotonic problem, due to constraints (10), (14b), and (14c). Constraint (10) can be equivalently rewritten as the following single constraint \( \min_{k \in K} \left[ c_{i, b, k}^{1+}(\theta) - c_{i, b, k}^{1-}(\theta) \right] \geq 0 \), where \( c_{i, b, k}^{1+}(\theta) = \theta_{i, b, k}^{2} \) and \( c_{i, b, k}^{1-}(\theta) = \theta_{i, b, k} \). The latter constraint is equivalent to

\[
\min_{k \in K} \left[ c_{i, b, k}^{1+}(\theta) - \left( \sum_{i \in I} \sum_{b \in B_i} \sum_{k \in K} c_{i, b, k}^{1-}(\theta) - \sum_{i \in I} \sum_{b \in B_i} \sum_{k' \in K, k' \neq k} c_{i, b, k'}^{1-}(\theta) \right) \right] =
\]

\[
\min_{k \in K} \left[ c_{i, b, k}^{1+}(\theta) + \sum_{i \in I} \sum_{b \in B_i} \sum_{k' \in K, k' \neq k} c_{i, b, k'}^{1-}(\theta) - \sum_{i \in I} \sum_{b \in B_i} \sum_{k \in K} c_{i, b, k}^{1-}(\theta) \right] \geq 0, \quad (18)
\]
which is the difference of two increasing functions $c^{1+}(\theta)$ and $c^{1-}(\theta)$. Similarly, by introducing a new auxiliary variable $s_1$, (17) can be rewritten as

$$\max_{\theta, p, \alpha, s_0, s_1} q^+(\theta, p, \alpha) + s_0$$

s.t. (8c), (8d), (8f), (11), (12), (14b), (14c), (17b), (17c),

$$0 \leq s_1 + c^{1-}(\theta) \leq c^{1-}(\theta_{\text{max}}),$$

$$0 \leq s_1 + c^{1-}(\theta_{\text{max}}) - c^{1-}(0_{i,b,k}),$$

$$s_1 + c^{1+}(\theta) \geq c^{1-}(\theta_{\text{max}}).$$

Similar to (10), constraints (14b) and (14c) can be equivalently transformed into the difference of two increasing functions. After adopting this method to (14b) and (14c), the problem (19) can be reformulated as

$$\max_{\theta, p, \alpha, s_0, s_1, s_2, s_3} q^+(\theta, p, \alpha) + s_0$$

s.t. (8c), (8d), (8f), (11), (12), (17b), (17c), (19b)-(19d),

$$0 \leq s_2 + c^{2-}(\theta, p, \alpha) \leq c^{2-}(\theta_{\text{max}}, p_{\text{max}}, \alpha_{\text{max}}),$$

$$0 \leq s_2 \leq c^{2-}(\theta_{\text{max}}, p_{\text{max}}, \alpha_{\text{max}}) - c^{2-}(0_{i,b,k}, 0_{i,b,k}, 0_{i,k,k'}),$$

$$s_2 + c^{2+}(\theta, p, \alpha) \geq c^{2-}(\theta_{\text{max}}, p_{\text{max}}, \alpha_{\text{max}}),$$

$$0 \leq s_3 + c^{3-}(\theta, p, \alpha) \leq c^{3-}(\theta_{\text{max}}, p_{\text{max}}, \alpha_{\text{max}}),$$

$$0 \leq s_3 \leq c^{3-}(\theta_{\text{max}}, p_{\text{max}}, \alpha_{\text{max}}) - c^{3-}(0_{i,b,k}, 0_{i,b,k}, 0_{i,k,k'}),$$

$$s_3 + c^{3+}(\theta, p, \alpha) \geq c^{3-}(\theta_{\text{max}}, p_{\text{max}}, \alpha_{\text{max}}).$$
where $c^2+(\theta, p, \alpha) = \min_{i, i', k, k' \in K, \lambda_i, \lambda_{i'} \geq \lambda_i, \lambda_{i'}} \left[ s^\text{UNC}_{i,k',k}(i, i', k, \alpha) + s^\text{ICI}_{i,k',k}(i, i', k, \alpha) + N_0 W_i \right]$,

\[
\sum_{i \in I} \sum_{b \in B_i} \sum_{j \in K} \sum_{j' \in K, j' \neq k} s_{i,j'} \min \left\{ \sum_{b \in B_i} \theta_{i,b,j} \theta_{i,b,j'}, 1 \right\} \left( I^\text{UNC}_{i,b,j} + I^\text{ICI}_{i,b,j} + N_0 W_i \right),
\]

\[
\sum_{i \in I} \sum_{b \in B_i} \sum_{k \in K} \sum_{k' \in K} c^2_{i,b,k,k'}(\theta, p, \alpha),
\]

\[
c^2-(\theta, p, \alpha) = \sum_{i \in I} \sum_{b \in B_i} \sum_{k \in K} \sum_{k' \in K} c^2_{i,b,k,k'}(\theta, p, \alpha),
\]

\[
c^3+(\theta, p, \alpha) = \min_{k \in K} \left\{ \sum_{i \in I} W_i \log_2 \left( I^\text{UNC}_{i,k} + I^\text{ICI}_{i,k} + N_0 W_i + s_{i,k} \right) \right\},
\]

\[
\sum_{j \in K} \sum_{j' \in K} \sum_{i \in I} \sum_{j \neq k} \sum_{j' \neq k'} c^3_k(\theta, p, \alpha),
\]

\[
c^3-(\theta, p, \alpha) = \sum_{v \in V} \sum_{k \in K_v} c^3_k(\theta, p, \alpha) + R_v^\text{sv}.
\]

In the following, we prove that (20) is a monotonic optimization problem in canonical form. At first, observe that the objective function (20a) is monotonic in $(\theta, p, \alpha, s_0, s_1, s_2, s_3)$. Then, to show that the feasible set of (20) is an intersection of the normal and co-normal sets (according to Definitions 3-5 in [27]), for any $(\theta, p, \alpha)$ in the feasible set of (20), we have

\[
q^-(0_{i,b,k}, 0_{i,b,k}, 0_{i,k,k'}) \leq q^-(\theta, p, \alpha),
\]

\[
c^1-(0_{i,b,k}) \leq c^1-(\theta),
\]

\[
c^2-(0_{i,b,k}, 0_{i,b,k}, 0_{i,k,k'}) \leq c^2-(\theta, p, \alpha),
\]

and

\[
c^3-(0_{i,b,k}, 0_{i,b,k}, 0_{i,k,k'}) \leq c^3-(\theta, p, \alpha).
\]

To this end, the feasible set of (20) can be written as the intersection of the following two sets as

\[
S = \left\{ (\theta, p, \alpha, s_0, s_1, s_2, s_3) : \theta \leq \theta_{\text{max}}, p \leq p_{\text{max}}, \alpha \leq \alpha_{\text{max}}, (8c), (8d), (11), (17b), (17c), (19b), (19c), (20b), (20c), (20e), (20f) \right\},
\]

(25)
and
\[ S_C = \left\{ (\theta, p, \alpha, s_0, s_1, s_2, s_3) : \theta \geq 0, \ p \geq 0, \ \alpha \geq 0, \ (12), \ (19d), \ (20d), \ (20g) \right\}. \quad (26) \]

All the constraint sets in (25) and (26) are monotonic and continuous (because of employing again (21) and Proposition 2 in [27]), resulting \( S \) and \( S_C \) in (25) and (26) are normal and co-normal sets, respectively in the hyper-rectangle given by
\[
[0, \theta_{\text{max}}] \times [0, p_{\text{max}}] \times [0, \alpha_{\text{max}}] \times [0, c_1 - (\theta_{\text{max}}) - c_1 - (0_{i,b,k})] \times [0, c_2 - (\theta_{\text{max}}, p_{\text{max}}, \alpha_{\text{max}}) - c_2 - (0_{i,b,k}, 0_{i,k,k'})] \times [0, q_{\text{max}} - (\theta_{\text{max}}, p_{\text{max}}, \alpha_{\text{max}}) - q_{\text{max}} - (0_{i,b,k}, 0_{i,b,k}, 0_{i,k,k'})]. \quad (27)
\]

Thus, (20) fulfills Definition 5 in [27] and the proof is completed. The transformed monotonic optimization problem in canonical form can be easily solved by using the poly block and branch-reduce-and-bound algorithms [21], [27].

**APPENDIX B**

**EQUIVALENT TRANSFORMATION OF (8)**

To tackle the combinatorial nature of (8), we first relax (8e) by using (10) [21]. Then, we replace the term \( \min \left\{ \sum_{b \in B} \theta_{i,b,k} \theta_{i,b,k'} , 1 \right\} \) in (2) with a new auxiliary variable \( \alpha_{i,k,k'} \in [0, 1] \) by adding the following linear constraints:
\[
\beta_{i,b,k,k'} \leq \frac{\theta_{i,b,k} + \theta_{i,b,k'}}{2}, \quad 0 \leq \beta_{i,b,k,k'} \leq 1, \quad (28)
\]
\[
\alpha_{i,k,k'} \leq \sum_{b \in B} \beta_{i,b,k,k'}, \quad (29)
\]
\[
\alpha_{i,k,k'} \geq \max_{b \in B} \left\{ \theta_{i,b,k} + \theta_{i,b,k'} - 1 \right\}, \quad 0 \leq \alpha_{i,k,k'} \leq 1, \quad (30)
\]
in which the binary variable \( \beta_{i,b,k,k'} \) is added to tackle the binary bilinear product \( \theta_{i,b,k} \theta_{i,b,k'} \).

According to the above transformations, (8) can be rewritten as
\[
\max_{\theta, p, \alpha, \beta} \sum_{v \in V} \sum_{k \in K_v} \omega_{v,k}^{\text{UNC}} \tau_k \quad (31a)
\]
in which

\[ \hat{r}_k = \max_{v \in \mathcal{V}, k \in \mathcal{K}_v} R^v_s, \]

where \( \alpha = [\alpha_{i,k,k'}] \), \( \beta = [\beta_{i,b,k,k'}] \), and

\[ \gamma_{l,k}^{\text{UNC}} = \frac{s_{i,k}}{I_{i,k}^{\text{UNC},\text{ini}} + I_{i,k}^{\text{UNC},\text{ICI}} + N_0 W_i}, \]

in which

\[ I_{i,k}^{\text{UNC},\text{ini}} = \sum_{k' \in \mathcal{K}, k' > \lambda_{i,k}} \alpha_{i,k,k'} \frac{s_{i,k}}{I_{i,k}^{\text{UNC},\text{ini}} + I_{i,k}^{\text{UNC},\text{ICI}} + N_0 W_i}, \]

and

\[ \hat{r}_k = \max_{i \in \mathcal{I}} W_i \hat{r}_{i,k}^{\text{UNC},\text{sec}}, \]

in which \( r_{i,k}^{\text{UNC}} = \log_2 (1 + \gamma_{l,k}) \). Problem (31) cannot be directly solved by the SCA algorithm yet, due to the multiplications of \( \theta, p, \) and \( \alpha \) (please see (32)), and fractional constraint (31b). In this regard, we first substitute the product term \( \theta_{i,b,k} p_{i,b,k} \) with \( \tilde{p}_{i,b,k} \) by imposing the following constraints [21]:

\[ \tilde{p}_{i,b,k} \leq \theta_{i,b,k} P_{i,b,k}^{\text{max}}, \quad \tilde{p}_{i,b,k} \leq p_{i,b,k}, \quad \tilde{p}_{i,b,k} \geq p_{i,b,k} - (1 - \theta_{i,b,k}) P_{i,b,k}^{\text{max}}. \]

After these transformations, we substitute \( \alpha_{i,k,k'} \tilde{p}_{i,b,k'} \) and \( (1 - \alpha_{i,k,k'}) \tilde{p}_{i,b,k'} \) with \( q_{i,b,k',k} \) and \( \bar{q}_{i,b,k',k} \), respectively by adding the following constraints:

\[ q_{i,k,k',k} \leq \alpha_{i,k,k'} P_{i,b,k}^{\text{max}}, \quad q_{i,k,k',k} \leq \tilde{p}_{i,b,k'}, \quad q_{i,k,k',k} \geq \tilde{p}_{i,b,k'} - (1 - \alpha_{i,k,k'}) P_{i,b,k}^{\text{max}}, \]

\[ \bar{q}_{i,k,k',k} \leq (1 - \alpha_{i,k,k'}) P_{i,b,k}^{\text{max}}, \quad \bar{q}_{i,k,k',k} \leq \tilde{p}_{i,b,k'}, \quad \bar{q}_{i,k,k',k} \geq \tilde{p}_{i,b,k'} - (1 - \alpha_{i,k,k'}) P_{i,b,k}^{\text{max}}. \]

According to the above transformations, \( \gamma_{l,k}^{\text{UNC}} \) in (32) can be rewritten as

\[ \gamma_{l,k}^{\text{UNC}} = \frac{\sum_{b \in B_i} \tilde{p}_{i,b,k} h_{i,b,k}}{I_{i,k}^{\text{UNC},\text{ini}} + I_{i,k}^{\text{UNC},\text{ICI}} + N_0 W_i}, \]

in which \( I_{i,k}^{\text{UNC},\text{ini}} = \sum_{k' \in \mathcal{K}, k' > \lambda_{i,k}} q_{i,k,k',k} h_{i,b,k}, \)

\[ I_{i,k}^{\text{UNC},\text{ICI}} = \sum_{k' \in \mathcal{K}} \sum_{b \in B_i} \bar{q}_{i,b,k',k} h_{i,b,k}. \]

Moreover, (31b) can be rewritten as

\[ \sum_{b \in B_i} \tilde{p}_{i,b,k} h_{i,b,k} \geq \sum_{b \in B_i} q_{i,b,k',k} h_{i,b,k'}, \forall i \in \mathcal{I}, k, k' \in \mathcal{K}, \lambda_{i,k} > \lambda_{i,k'}. \]
where \( \tilde{I}_{UNC,INI, i,k} = \sum_{k'' \in K, \lambda_{i,k''} > \lambda_{i,k'}} \sum_{b \in B_i} q_{i,b,k'',k} h_{i,b,k} \), and \( \tilde{I}_{UNC,ICI, i,k} = \sum_{k'' \in K, k'' \neq k'} \sum_{b \in B_i} q_{i,b,k'',k} h_{i,b,k} \). Thus, (31) can be transformed into the following problem as

\[
\max_{\vartheta} \sum_{v \in V} \sum_{k \in K} \omega_v \tilde{T}_{k}^{UNC}
\]

\[
\text{s.t. (8c), (8d), (10), (28)-(30), (33)-(35), (37)}.
\]

(38a)

\[
\tilde{T}_{k}^{UNC} \geq R_{v}^{eq}, \forall v \in V, k \in K_v,
\]

(38b)

where \( \tilde{T}_{k}^{UNC} = \sum_{i \in I} W_i \tilde{T}_{i,k}^{UNC,SE} \) in which \( \tilde{T}_{i,k}^{UNC,SE} = \log_2 (1 + \gamma_{i,k}) \). Moreover, to ease of convenience, we denote \( q = [q_{i,b,k'}], \bar{q} = [\bar{q}_{i,b,k'}], \tilde{p} = [\tilde{p}_{i,b,k}], \) and \( \vartheta = [\theta, p, \alpha, \beta, q, \bar{q}, \tilde{p}] \).

Problem (38) is still nonconvex, due to the nonconcavity of the objective function in (38a), and nonconvexity of constraints (10), (37), and (38b). To handle (10), we use the penalty factor approach, where for a sufficiently large constant \( \eta \gg 1 \), (38) can be equivalently transformed into the following problem [21]:

\[
\max_{\vartheta} \sum_{v \in V} \sum_{k \in K_v} \omega_v \tilde{T}_{k}^{UNC} - \eta \left( \sum_{i \in I} \sum_{b \in B_i} \sum_{k \in K_v} \theta_{i,b,k} - \theta_{i,b,k}^2 \right)
\]

\[
\text{s.t. (8c), (8d), (8f), (28)-(30), (33)-(35), (37), (38b)}.
\]

(39a)

\[
0 \leq \theta_{i,b,k} \leq 1.
\]

(39b)

In fact, \( \eta \) acts as a penalty factor for the objective function to penalize the cost term \( \left( \theta_{i,b,k} - \theta_{i,b,k}^2 \right) \) in (39a). The proof of this theorem is similar to the proof presented in the appendix of [21].

The resulting problem (39) is still nonconvex, due to the nonconcavity of the SINR function in (37), (38b), and (39a), and also the term \( \theta_{i,b,k}^2 \) in (39a). In contrast to prior works [21], [23], [26], [30], we cannot directly apply the SCA algorithm with DC programming to solve (39), since (37) cannot be transformed into a linear constraint. This is because, in CoMP-NOMA systems, multiple signal powers are summed-up at a CoMP-user. This summation is appeared in the numerator of the SINR functions in the SIC constraint (37). To tackle this, we first transform (37) into the difference of two concave functions as
\[
\begin{align*}
\log_2 \left( \tilde{I}_{i,k}^{\text{UNC},\text{INI}} + \tilde{I}_{i,k}^{\text{UNC},\text{ICI}} + N_0 W_i + \sum_{b \in B_i} \tilde{p}_{i,b,k'} h_{i,b,k'} \right) + \log_2 \left( \tilde{I}_{i,k'}^{\text{UNC},\text{INI}} + \tilde{I}_{i,k'}^{\text{UNC},\text{ICI}} + N_0 W_i \right) \\
- \log_2 \left( \tilde{I}_{i,k'}^{\text{UNC},\text{INI}} + \tilde{I}_{i,k'}^{\text{UNC},\text{ICI}} + N_0 W_i + \sum_{b \in B_i} g_{i,b,k',k} h_{i,b,k} \right) - \log_2 \left( \tilde{I}_{i,k'}^{\text{UNC},\text{INI}} + \tilde{I}_{i,k'}^{\text{UNC},\text{ICI}} + N_0 W_i \right) \geq 0,
\end{align*}
\]

∀i ∈ I, k, k' ∈ K, λ_{i,k} > λ_{i,k'}.

(40)

Problem (39) can thus be rewritten as

\[
\max_{\bar{\vartheta}} \sum_{v \in V} \sum_{k \in K_v} \omega_v \tilde{r}_{i_k}^{\text{UNC}} - \eta \left( \sum_{i \in I} \sum_{b \in B_i} \sum_{k \in K} \theta_{i,b,k} - \theta_{i,b,k}^2 \right)
\]

(41a)

s.t. (8c), (8d), (8f), (28)-(30), (33)-(35), (37), (38b), (39b), (40).

Now, to find a locally optimal solution for the nonconvex problem (8), we can directly apply the SCA algorithm with DC programming to its equivalent form in (41).

**APPENDIX C**

**SCA WITH DC PROGRAMMING FOR SOLVING (41)**

To tackle the nonconcavity of the rate function in (41a) and (38b), we first define \( \tilde{r}_{i,k}^{\text{UNC,SE}}(q, \bar{q}, \tilde{p}) \) as the difference of two concave functions as

\[
\tilde{r}_{i,k}^{\text{UNC,SE}}(q, \bar{q}, \tilde{p}) = f_{i,k}^{\text{UNC}}(q, \bar{q}, \tilde{p}) - g_{i,k}^{\text{UNC}}(q, \bar{q}),
\]

where

\[
f_{i,k}^{\text{UNC}}(q, \bar{q}, \tilde{p}) = \log_2 \left( \tilde{I}_{i,k}^{\text{UNC},\text{INI}} + \tilde{I}_{i,k}^{\text{UNC},\text{ICI}} + N_0 W_i + \sum_{b \in B_i} \tilde{p}_{i,b,k} h_{i,b,k} \right),
\]

and

\[
g_{i,k}^{\text{UNC}}(q, \bar{q}) = \log_2 \left( \tilde{I}_{i,k}^{\text{UNC},\text{INI}} + \tilde{I}_{i,k}^{\text{UNC},\text{ICI}} + N_0 W_i \right).
\]

Note that \( \tilde{r}_{i,k}^{\text{UNC,SE}} \) is concave with respect to \( \tilde{p} \). Then, at each iteration \( l \), the term \( g_{i,k}^{\text{UNC}}(q^{(l-1)}, \bar{q}^{(l-1)}) \) is approximated to its first order Taylor series approximation around \((q^{(l-1)}, \bar{q}^{(l-1)})\) as

\[
g_{i,k}^{\text{UNC}}(q^{(l)}, \bar{q}^{(l)}) \approx g_{i,k}^{\text{UNC}}(q^{(l-1)}, \bar{q}^{(l-1)}) + \nabla_q g_{i,k}^{\text{UNC}}(q^{(l-1)}, \bar{q}^{(l-1)})(q^{(l)} - q^{(l-1)}) + \nabla_{\bar{q}} g_{i,k}^{\text{UNC}}(q^{(l-1)}, \bar{q}^{(l-1)})(\bar{q}^{(l)} - \bar{q}^{(l-1)}),
\]

(44)
where the gradient functions \( \nabla q g^\text{UNC}_{i,k}(\mathbf{q}, \bar{\mathbf{q}}) \) and \( \nabla q g^\text{UNC}_{i,k}(\mathbf{q}, \bar{\mathbf{q}}) \) are defined, respectively as follows:

\[
\nabla q g^\text{UNC}_{i,k}(\mathbf{q}, \bar{\mathbf{q}}) = \begin{cases} \frac{h_{i,b,k}}{(\ln 2)(I^\text{UNC,INI}_{i,k} + I^\text{UNC,ICI}_{i,k} + N_0 W_i)}, & \forall \lambda_{i,k'} > \lambda_{i,k}, \ b \in B_i; \\ 0, & \text{otherwise}, \end{cases}
\]

(45)

\[
\nabla q g^\text{UNC}_{i,k}(\mathbf{q}, \bar{\mathbf{q}}) = \begin{cases} \frac{h_{i,b,k}}{(\ln 2)(I^\text{UNC,INI}_{i,k} + I^\text{UNC,ICI}_{i,k} + N_0 W_i)}, & \forall k' \in K \setminus \{k\}, \ b \in B_i; \\ 0, & \text{otherwise}. \end{cases}
\]

(46)

Therefore, at each iteration \( l \), \( \hat{r}^\text{UNC,SE}_{i,k}(\mathbf{q}^{(l)}, \bar{\mathbf{q}}^{(l)}, \bar{\mathbf{p}}^{(l)}) \) is approximated by

\[
\hat{r}^\text{UNC,SE}_{i,k}(\mathbf{q}^{(l)}, \bar{\mathbf{q}}^{(l)}, \bar{\mathbf{p}}^{(l)}) \approx f^\text{UNC}_{i,k}(\mathbf{q}^{(l)}, \bar{\mathbf{q}}^{(l)}, \bar{\mathbf{p}}^{(l)}) - g^\text{UNC}_{i,k}(\mathbf{q}^{(l-1)}, \bar{\mathbf{q}}^{(l-1)}) - \nabla q g^\text{UNC}_{i,k}(\mathbf{q}^{(l-1)}, \bar{\mathbf{q}}^{(l-1)})) (\mathbf{q}^{(l)} - \mathbf{q}^{(l-1)}) - \nabla q g^\text{UNC}_{i,k}(\mathbf{q}^{(l-1)}, \bar{\mathbf{q}}^{(l-1)})) (\mathbf{q}^{(l)} - \bar{\mathbf{q}}^{(l-1)}).
\]

(47)

Similarly, to handle the nonconvexity of (40), at each iteration \( l \), we approximate \( T_{i,b,k',k}^1(\mathbf{q}, \bar{\mathbf{q}}) = \log_2 \left( I^\text{UNC,INI}_{i,k'}^1 + I^\text{UNC,ICI}_{i,k'}^1 + N_0 W_i \right) \), and \( T_{i,b,k',k}^2(\mathbf{q}, \bar{\mathbf{q}}) = \log_2 \left( I^\text{UNC,INI}_{i,k,k'} + I^\text{UNC,ICI}_{i,k,k'} + N_0 W_i \right) \) to affine functions as follows:

\[
\hat{T}_{i,b,k',k}^1(\mathbf{q}^{(l)}, \bar{\mathbf{q}}^{(l)}) \approx T_{i,b,k',k}^1(\mathbf{q}^{(l-1)}, \bar{\mathbf{q}}^{(l-1)}) + \nabla q T_{i,b,k',k}^1(\mathbf{q}^{(l-1)}, \bar{\mathbf{q}}^{(l-1)})) (\mathbf{q}^{(l)} - \mathbf{q}^{(l-1)}) + \nabla q T_{i,b,k',k}^1(\mathbf{q}^{(l-1)}, \bar{\mathbf{q}}^{(l-1)})) (\mathbf{q}^{(l)} - \bar{\mathbf{q}}^{(l-1)}),
\]

in which

\[
\nabla q T_{i,b,k',k}^1(\mathbf{q}, \bar{\mathbf{q}}) = \begin{cases} \frac{2h_{i,b,k'}}{(\ln 2)(2^\text{T}_{i,b,k',k}^1(\mathbf{q}, \bar{\mathbf{q}}))}, & \forall \lambda_{i,k''} > \lambda_{i,k'}, \ b \in B_i; \\ \frac{h_{i,b,k'}}{(\ln 2)(2^\text{T}_{i,b,k',k}^1(\mathbf{q}, \bar{\mathbf{q}}))}, & \forall \lambda_{i,k''} > \lambda_{i,k'}, \ b \in B_i; \\ 0, & \text{otherwise}, \end{cases}
\]

(49)

and

\[
\nabla q T_{i,b,k',k}^1(\mathbf{q}, \bar{\mathbf{q}}) = \begin{cases} \frac{h_{i,b,k'}}{(\ln 2)(2^\text{T}_{i,b,k',k}^1(\mathbf{q}, \bar{\mathbf{q}}))}, & \forall k'' \in K \setminus \{k'\}, \ b \in B_i; \\ 0, & \text{otherwise}. \end{cases}
\]

(50)

In addition,
\[
\hat{T}_{i,b,k',k}^2(q^{(l)}, \bar{q}^{(l)}) \approx T_{i,b,k',k}^2(q^{(l-1)}, \bar{q}^{(l-1)}) + \nabla_q T_{i,b,k',k}^2(q^{(l-1)}, \bar{q}^{(l-1)}) (q^{(l)} - q^{(l-1)}) + \\
\nabla_{\bar{q}} T_{i,b,k',k}^2(q^{(l-1)}, \bar{q}^{(l-1)}) (\bar{q}^{(l)} - \bar{q}^{(l-1)}),
\]

where

\[
\nabla_q T_{i,b,k',k}^2(q, \bar{q}) = \begin{cases} 
\frac{h_{i,b,k}}{\ln 2} \left( \frac{2}{h_{i,b,k'}} (q, \bar{q}) \right), & \forall \lambda_{i,k'} > \lambda_{i,k}, b \in B_i; \\
0, & \text{otherwise},
\end{cases}
\]

and

\[
\nabla_{\bar{q}} T_{i,b,k',k}^2(q, \bar{q}) = \begin{cases} 
\frac{h_{i,b,k}}{\ln 2} \left( \frac{2}{h_{i,b,k'}} (q, \bar{q}) \right), & \forall k' \in \mathcal{K} \setminus \{k\}, b \in B_i; \\
0, & \text{otherwise}.
\end{cases}
\]

Similar to (47), (48), and (51), at each iteration \( l \), the nonconcave term \( T_{i,b,k}^3 = \theta_{i,b,k}^2 \) in (41a) is approximated to its first order Taylor series as

\[
\hat{T}_{i,b,k}^3(\theta^{(l)}) \approx T_{i,b,k}^3(\theta^{(l-1)}) + \nabla_{\theta} T_{i,b,k}^3(\theta^{(l-1)}) (\theta^{(l)} - \theta^{(l-1)}),
\]

where \( \nabla_{\theta} T_{i,b,k}^3 = 2\theta_{i,b,k}^2 \). According to (47), (48), (51), and (54), the convex approximated problem of (41) is given by

\[
\max_{\tau} \sum_{v \in \mathcal{V}} \sum_{k \in \mathcal{K}_v} \omega_v \sum_{i \in \mathcal{I}} W_i \hat{r}_{i,k}^{unc,se} - \eta \left( \sum_{i \in \mathcal{I}} \sum_{b \in B_i} \sum_{k \in \mathcal{K}} \theta_{i,b,k} - \hat{T}_{i,b,k}^3 \right)
\]

s.t. (8c), (8d), (8f), (28)-(30), (33)-(35), (39b),

\[
\sum_{i \in \mathcal{I}} \hat{r}_{i,k}^{unc,se} \geq R_v^{\text{se}}, \forall v \in \mathcal{V}, k \in \mathcal{K}_v,
\]

\[
\log_2 \left( \hat{r}_{i,k',k}^{\text{unc.ini}} + \hat{r}_{i,k',k}^{\text{unc.ini}} + N_0 W_i + \sum_{b \in B_i} \tilde{p}_{i,b,k'} \hat{r}_{i,k,b} \right) + \log_2 \left( \hat{r}_{i,k'}^{\text{unc.ini}} + \hat{r}_{i,k'}^{\text{unc.ini}} + N_0 W_i \right)
\]

\[
- T_{i,b,k',k}^1 - T_{i,b,k',k}^2 \geq 0, \forall i \in \mathcal{I}, k, k' \in \mathcal{K}, \lambda_{i,k} > \lambda_{i,k'}.
\]

At each iteration \( l \), the convex optimization problem (55) can be solved by using the standard convex optimization solvers such as Lagrange dual method, interior point methods, or standard optimization software CVX [21], [23], [26], [30].
APPENDIX D

CONVERGENCE OF THE PROPOSED SCA ALGORITHM FOR SOLVING (41)

In order to prove that the proposed SCA algorithm converges to a locally optimal solution, we first note that the gradients of $g_{i,k}^{\text{UNC}}(q, \bar{q})$, $\hat{T}_{i,b,k'}^1(q^{(l)}, \bar{q}^{(l)})$, $\hat{T}_{i,b,k'}^2(q^{(l)}, \bar{q}^{(l)})$, and $\hat{T}_{i,b,k'}^3(\theta^{(l)})$ are indeed their supergradients meaning that at each iteration $l$, for any feasible point $\bar{q}^{(l)}$, we have:

$$
g_{i,k}^{\text{UNC}}(q^{(l)}, \bar{q}^{(l)}) \leq g_{i,k}^{\text{UNC}}(q^{(l-1)}, \bar{q}^{(l-1)}) + \nabla_q g_{i,k}^{\text{UNC}}(q^{(l-1)}, \bar{q}^{(l-1)}) (q^{(l)} - q^{(l-1)}) + \nabla_q g_{i,k}^{\text{UNC}}(q^{(l-1)}, \bar{q}^{(l-1)}) (\bar{q}^{(l)} - \bar{q}^{(l-1)}) , \quad (56)$$

$$
\hat{T}_{i,b,k'}^1(q^{(l)}, \bar{q}^{(l)}) \leq T_{i,b,k'}^1(q^{(l-1)}, \bar{q}^{(l-1)}) + \nabla_q T_{i,b,k'}^1(q^{(l-1)}, \bar{q}^{(l-1)}) (q^{(l)} - q^{(l-1)}) + \nabla_q T_{i,b,k'}^1(q^{(l-1)}, \bar{q}^{(l-1)}) (\bar{q}^{(l)} - \bar{q}^{(l-1)}) , \quad (57)$$

$$
\hat{T}_{i,b,k'}^2(q^{(l)}, \bar{q}^{(l)}) \leq T_{i,b,k'}^2(q^{(l-1)}, \bar{q}^{(l-1)}) + \nabla_q T_{i,b,k'}^2(q^{(l-1)}, \bar{q}^{(l-1)}) (q^{(l)} - q^{(l-1)}) + \nabla_q T_{i,b,k'}^2(q^{(l-1)}, \bar{q}^{(l-1)}) (\bar{q}^{(l)} - \bar{q}^{(l-1)}) , \quad (58)$$

$$
\hat{T}_{i,b,k'}^3(\theta^{(l)}) \leq T_{i,b,k'}^3(\theta^{(l-1)}) + \nabla_{\theta} T_{i,b,k'}^3(\theta^{(l-1)}) (\theta^{(l)} - \theta^{(l-1)}) . \quad (59)$$

According to (56)-(59), it can be easily shown that the optimal solution of (55) remains in the feasible region of (41) which is equivalent to (8). Moreover, according to (56) and (59), for any optimal solution $(\theta^{(l)}, q^{*}(l), \bar{q}^{*}(l), \bar{p}^{*}(l))$, it is concluded that at each iteration $l$, the following inequality holds:

$$
\sum_{v \in V} \sum_{k \in \mathcal{K}_v} \omega_v \sum_{i \in I} W_{i,v}^{\text{UNC,SE}}(q^{*}(l), \bar{q}^{*}(l), \bar{p}^{*}(l)) - \eta \left( \sum_{b \in B_i} \sum_{k \in \mathcal{K}} \theta_{i,b,k}^{*}(l) - T_{i,b,k}^3(\theta^{*}(l)) \right) \geq \sum_{v \in V} \sum_{k \in \mathcal{K}_v} \omega_v \sum_{i \in I} W_{i,v}^{\text{UNC,SE}}(q^{*}(l), \bar{q}^{*}(l), \bar{p}^{*}(l)) - \eta \left( \sum_{b \in B_i} \sum_{k \in \mathcal{K}} \theta_{i,b,k}^{*}(l) - \hat{T}_{i,b,k}^3(\theta^{*}(l)) \right) . \quad (60)
$$
According to the fact that at each iteration \( l \), \( \theta^{* (l)} \) is the globally optimal solution of the convex problem (55) which is in the feasible region of (41), it can be concluded that

\[
\sum_{v \in \mathcal{V}} \sum_{k \in \mathcal{K}_v} \omega_v \sum_{i \in \mathcal{I}} W_i^{\text{T}_{\text{SE}} \text{SE}} (q^{* (l)}, \bar{q}^{* (l)}, \bar{p}^{* (l)}) - \eta \left( \sum_{b \in \mathcal{B}_i} \sum_{k \in \mathcal{K}} \theta_{i,b,k}^{* (l)} - \hat{T}_{i,b,k}^{3} (\theta^{* (l)}) \right)
\]

\[
= \max_{\theta} \sum_{v \in \mathcal{V}} \sum_{k \in \mathcal{K}_v} \omega_v \sum_{i \in \mathcal{I}} W_i^{\text{T}_{\text{SE}} \text{SE}} (q^{(l)}, \bar{q}^{(l)}, \bar{p}^{(l)}) - \eta \left( \sum_{b \in \mathcal{B}_i} \sum_{k \in \mathcal{K}} \theta_{i,b,k}^{(l)} - \hat{T}_{i,b,k}^{3} (\theta^{(l)}) \right)
\]

\[
\geq \sum_{v \in \mathcal{V}} \sum_{k \in \mathcal{K}_v} \omega_v \sum_{i \in \mathcal{I}} W_i^{\text{T}_{\text{SE}} \text{SE}} (q^{(l-1)}, \bar{q}^{(l-1)}, \bar{p}^{(l-1)}) - \eta \left( \sum_{b \in \mathcal{B}_i} \sum_{k \in \mathcal{K}} \theta_{i,b,k}^{(l-1)} - \hat{T}_{i,b,k}^{3} (\theta^{(l-1)}) \right)
\]

\[
= \sum_{v \in \mathcal{V}} \sum_{k \in \mathcal{K}_v} \omega_v \sum_{i \in \mathcal{I}} W_i^{\text{T}_{\text{SE}} \text{SE}} (q^{(l-1)}, \bar{q}^{(l-1)}, \bar{p}^{(l-1)}) - \eta \left( \sum_{b \in \mathcal{B}_i} \sum_{k \in \mathcal{K}} \theta_{i,b,k}^{(l-1)} - \hat{T}_{i,b,k}^{3} (\theta^{(l-1)}) \right).
\]

According to (60) and (61), it can be derived that after each SCA iteration, the objective function (41a) is improved (increased) or remains constant. Therefore, the proposed SCA algorithm with DC programming converges to a locally optimal solution and the proof is completed.

**APPENDIX E**

**CANONICAL TRANSFORMATION OF (9)**

In order to transform (9) into a monotonic-based optimization problem in canonical form, we first relax \( \theta \) and \( x \) by using the relaxation approach in (10). Then, we substitute the term \((1 - \theta_{i,b,k})\) in \( r_{i,k}^{\text{LNC,ICI}} \) in (6) with \( \tilde{\theta}_{i,b,k} \) by adding the following constraint:

\[
\tilde{\theta}_{i,b,k} \geq 1 - \theta_{i,b,k}, \quad 0 \leq \tilde{\theta}_{i,b,k} \leq 1.
\]

In this regard, \( r_{i,k}^{\text{LNC,ICI}} \) and \( r_{i,k'}^{\text{LNC,ICI}} \) in (6) and (7) are substituted with \( \hat{r}_{i,k}^{\text{LNC,ICI}} = \sum_{b \in \mathcal{B}_i} \sum_{k' \in \mathcal{K}_v} \tilde{\theta}_{i,b,k'} \)

\[
\sum_{b' \in \mathcal{B}_i, b' \neq b} \sum_{k' \in \mathcal{K}_v} \tilde{\theta}_{i,b,k'} p_{i,b',k'} h_{i,b',k'}
\]

and \( \hat{r}_{i,k'}^{\text{LNC,ICI}} = \sum_{b \in \mathcal{B}_i} \sum_{k'' \in \mathcal{K}_v} \tilde{\theta}_{i,b,k''} \theta_{i,b,k''} p_{i,b',k''} h_{i,b',k'}, \)

respectively. Therefore, (9) can be rewritten as

\[
\begin{equation}
\max_{\theta,x,p,\tilde{\theta}} \sum_{v \in \mathcal{V}} \sum_{k \in \mathcal{K}_v} \omega_v \hat{r}_{k}^{\text{LNC}}
\end{equation}
\]
s.t. \((4), (5), (8c), (8d), (8f), (10)\),

\[
\text{s.t. } (4), (5), (8c), (8d), (8f), (10),
\]

\[
s_{i,k',k}^{\text{UNC}} \left( I_{i,k'}^{\text{LNC,INI}} + \bar{r}_{i,k'}^{\text{LNC,ICI}} + N_0 W_i \right) - \left( I_{i,k'}^{\text{LNC,INI}} + \bar{r}_{i,k'}^{\text{LNC,ICI}} + N_0 W_i \right) x_{i,b,k} \theta_{i,b,k'} s_{i,k'} \geq 0, \\
\forall i \in I, b \in B_i, k, k' \in K, \lambda_{i,k} > \lambda_{i,k'},
\]

\[
\hat{r}_{k}^{\text{LNC}} \geq \hat{R}_{sv}, \forall v \in V, k \in K_v,
\]

\[
x_{i,b,k} \leq x_{i,b,k}^2, \quad 0 \leq x_{i,b,k} \leq 1,
\]

where \( \hat{r}_{k}^{\text{LNC}} = \sum_{i \in I} W_i \log_2 \left( 1 + \frac{s_{i,k}^{\text{LNC,INI}} + \bar{s}_{i,k}^{\text{LNC,ICI}} + N_0 W_i}{I_{i,k}^{\text{LNC,INI}} + I_{i,k}^{\text{LNC,ICI}} + N_0 W_i} \right) \), and \( \bar{\theta} = [\bar{\theta}_{i,b,k}] \). Problem (63) is not yet a monotonic optimization problem in canonical form, because of the objective function (63a) and constraints (10), (63b)-(63d). To tackle the non-monotonicity of (10) and (63d), we apply a similar transformation method that is used in (18) and constraints (19b)-(19d). Moreover, for the non-monotonic constraints (63b) and (63c), we apply a similar method to the approach used in constraint sets (20b)-(20d) and (20e)-(20g), respectively. In addition, for the non-monotonic objective function (63a), we apply a similar method that is used in (15)-(17). After adopting the above steps, the resulting monotonic problem would be canonical which can be optimally solved by using the poly block or branch-reduce-and-bound algorithms [27]. In order to avoid duplicated discussions, the canonical form of (63) is not mathematically formulated here.

**REFERENCES**

[1] R. Irmer, H. Droste, P. Marsch, M. Griefer, G. Fettweis, S. Brueck, H. Mayer, L. Thiele, and V. Jungnickel, “Coordinated multipoint: Concepts, performance, and field trial results,” IEEE Communications Magazine, vol. 49, no. 2, pp. 102–111, Feb. 2011.

[2] M. S. Ali, E. Hossain, and D. I. Kim, “Coordinated multipoint transmission in downlink multi-cell NOMA systems: Models and spectral efficiency performance,” IEEE Wireless Communications, vol. 25, no. 2, pp. 24–31, Apr. 2018.

[3] M. S. Ali, E. Hossain, A. Al-Dweik, and D. I. Kim, “Downlink power allocation for CoMP-NOMA in multi-cell networks,” IEEE Transactions on Communications, vol. 66, no. 9, pp. 3982–3998, Sept. 2018.

[4] Y. Tian, A. Nix, and M. Beach, “On the performance of a multi-tier NOMA strategy in coordinated multi-point networks,” IEEE Communications Letters, vol. 21, no. 11, pp. 2448–2451, Nov. 2017.

[5] Y. Tian, A. R. Nix, and M. Beach, “On the performance of opportunistic NOMA in downlink CoMP networks,” IEEE Communications Letters, vol. 20, no. 5, pp. 998–1001, May 2016.

[6] Z. Liu, G. Kang, L. Lei, N. Zhang, and S. Zhang, “Power allocation for energy efficiency maximization in downlink CoMP systems with NOMA,” in Proc. IEEE Wireless Communications and Networking Conference (WCNC), Mar. 2017, pp. 1–6.

[7] I. Khan, F. Belqasmi, R. Glitho, N. Crespi, M. Morrow, and P. Polakos, “Wireless sensor network virtualization: A survey,” IEEE Communications Surveys Tutorials, vol. 18, no. 1, pp. 553–576, Mar. 2016.
[8] D. B. Rawat, “A novel approach for shared resource allocation with wireless network virtualization,” in Proc. IEEE International Conference on Communications Workshops (ICC Workshops), May 2017, pp. 665–669.

[9] E. J. Kitindi, S. Fu, Y. Jia, A. Kabir, and Y. Wang, “Wireless network virtualization with SDN and C-RAN for 5G networks: Requirements, opportunities, and challenges,” IEEE Access, vol. 5, pp. 19 099–19 115, Sept. 2017.

[10] L. Li, N. Deng, W. Ren, B. Kou, W. Zhou, and S. Yu, “Multi-service resource allocation in future network with wireless virtualization,” IEEE Access, vol. 6, pp. 53 854–53 868, Sept. 2018.

[11] X. Sun, N. Yang, S. Yan, Z. Ding, D. W. K. Ng, C. Shen, and Z. Zhong, “Joint beamforming and power allocation in downlink NOMA multiuser MIMO networks,” IEEE Transactions on Wireless Communications, vol. 17, no. 8, pp. 5367–5381, Aug. 2018.

[12] Y. Al-Eryani, E. Hossain, and D. I. Kim, “Generalized coordinated multipoint (GCoMP)-enabled NOMA: Outage, capacity, and power allocation,” IEEE Transactions on Communications, vol. 67, no. 11, pp. 7923–7936, Nov. 2019.

[13] L. Lei, D. Yuan, C. K. Ho, and S. Sun, “Power and channel allocation for non-orthogonal multiple access in 5G systems: Tractability and computation,” IEEE Transactions on Wireless Communications, vol. 15, no. 12, pp. 8580–8594, Dec. 2016.

[14] F. Fang, H. Zhang, J. Cheng, and V. C. M. Leung, “Energy-efficient resource allocation for downlink non-orthogonal multiple access network,” IEEE Transactions on Communications, vol. 64, no. 9, pp. 3722–3732, Sept. 2016.

[15] B. Di, L. Song, and Y. Li, “Sub-channel assignment, power allocation, and user scheduling for non-orthogonal multiple access networks,” IEEE Transactions on Wireless Communications, vol. 15, no. 11, pp. 7686–7698, Nov. 2016.

[16] F. Liu and M. Petrova, “Performance of dynamic power and channel allocation for downlink MC-NOMA systems,” IEEE Transactions on Wireless Communications, vol. 19, no. 3, pp. 1650–1662, Mar. 2020.

[17] C. Chung and C. Ku, “Selective transmission strategy for NOMA in downlink CoMP,” in Proc. IEEE 89th Vehicular Technology Conference (VTC2019-Spring), 2019, pp. 1–5.

[18] A. Kliži, J. Farah, C. Abdel Nour, and C. Douillard, “Mutual successive interference cancellation strategies in NOMA for enhancing the spectral efficiency of CoMP systems,” IEEE Transactions on Communications, vol. 68, no. 2, pp. 1213–1226, Feb. 2020.

[19] D. Tweed, S. Parsaeefard, M. Derakhshani, and T. Le-Ngoc, “Dynamic resource allocation for MC-NOMA VWNs with imperfect SIC,” in Proc. IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), Oct. 2017, pp. 1–5.

[20] T. M. Ho, N. H. Tran, S. M. A. Kazmi, Z. Han, and C. S. Hong, “Wireless network virtualization with non-orthogonal multiple access,” in Proc. NOMS 2018 - 2018 IEEE/IFIP Network Operations and Management Symposium, Apr. 2018, pp. 1–9.

[21] Y. Xu, 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 Transactions on Communications, vol. 65, no. 3, pp. 1077–1091, Mar. 2017.

[22] Y. Fu, L. Salan, C. W. Sung, and C. S. Chen, “Subcarrier and power allocation for the downlink of multicarrier NOMA systems,” IEEE Transactions on Vehicular Technology, vol. 67, no. 12, pp. 11 833–11 847, 2018.

[23] S. Rezvani, S. Parsaeefard, N. Mokari, M. R. Javan, and H. Yaniyomeroglu, “Cooperative multi-bitrate video caching and transcoding in multicarrier NOMA-assisted heterogeneous virtualized MEC networks,” IEEE Access, vol. 7, pp. 93 511–93 536, July 2019.

[24] S. Rezvani and E. A. Jorswieck, “Cache replacement in CoMP-based multi-cell networks with unknown content popularity,” in Proc. 24th International ITG Workshop on Smart Antennas (WSA), May 2020, pp. 1–6.

[25] Y. Xu, C. Chen, Z. Ding, X. Sun, S. Yan, G. Zhu, and Z. Zhong, “Joint beamforming and power-splitting control in downlink cooperative SWIPT NOMA systems,” IEEE Transactions on Signal Processing, vol. 65, no. 18, pp. 4874–4886, Sept. 2017.

[26] M. Moltafet, R. Joda, N. Mokari, M. R. Sabagh, and M. Zorzi, “Joint access and fronthaul radio resource allocation in PD-NOMA-based 5G networks enabling dual connectivity and CoMP,” IEEE Transactions on Communications, vol. 66, no. 12, pp. 6463–6477, 2018.

[27] A. Zappone, E. Bjrnson, L. Sanguinetti, and E. A. Jorswieck, “ Globally optimal energy-efficient power control and receiver design in wireless networks,” IEEE Transactions on Signal Processing, vol. 65, no. 11, p. 28442859, June 2017.

[28] C. Y. Wong, R. S. Cheng, K. B. Lataief, and R. D. Murch, “Multiuser OFDM with adaptive subcarrier, bit, and power allocation,” IEEE Journal on Selected Areas in Communications, vol. 17, no. 10, pp. 1747–1758, Oct. 1999.

[29] M. Tao, Y. C. Liang, and F. Zhang, “Resource allocation for delay differentiated traffic in multiuser OFDM systems,” IEEE Transactions on Wireless Communications, vol. 7, no. 6, pp. 2190–2201, June 2008.
[30] J. Zhao, Y. Liu, K. K. Chai, A. Nallanathan, Y. Chen, and Z. Han, “Spectrum allocation and power control for non-orthogonal multiple access in HetNets,” *IEEE Transactions on Wireless Communications*, vol. 16, no. 9, pp. 5825–5837, Sept. 2017.

[31] D. T. Ngo, S. Khakurel, and T. Le-Ngoc, “Joint subchannel assignment and power allocation for OFDMA femtocell networks,” *IEEE Transactions on Wireless Communications*, vol. 13, no. 1, pp. 342–355, Jan. 2014.

[32] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2009.