QoS-Aware Resource Allocation for Mobile Edge Networks: User Association, Precoding and Power Allocation

Guanchong Niu, Qi Cao, and Man-On Pun

Abstract—Mobile edge computing (MEC) can provide computing and storage services to user equipments (UEs) by utilizing edge nodes known as the small base stations (SBS’s) deployed at the edge of the network. The short-distance transmission nature between SBS’s and UEs makes the millimeter-wave (mmWave) communication empowered with multiple-input multiple-output (MIMO) hybrid precoding techniques particularly attractive for MEC. In this work, we consider the UE-SBS association, precoding design and power allocation for MEC networks endowed with mmWave MIMO. More specifically, the user association problem is first formulated as a max-k-cut (M & C) problem and then, solved by a distributed local-search algorithm. Next, the joint optimization of precoding and power allocation is cast into the difference of two convex functions (D.C.) programming framework before an iterative rank-constrained D.C. programming algorithm is developed to maximize the weighted sum-rate (WSR) of all UEs while taking into account the quality of service (QoS) requirement of each UE. Furthermore, the monotonic convergence of the proposed iterative algorithm is analytically proven. Finally, extensive computer simulation is conducted to demonstrate the effectiveness of the proposed iterative algorithm.

Index Terms—Mobile edge computing, user association, hybrid beamforming, D.C. programming, power allocation.

I. INTRODUCTION

The explosive growth in both mobile data traffic and the number of mobile terminals as smartphones and tablets has presented major challenges on the wireless network design [1], [2]. To meet these challenges, the future networks have to minimize the transmission latency to support real-time applications without being overloaded by data exchanges between UEs and base stations [3]. To circumvent these problems, mobile edge computing (MEC) has recently been proposed as one of the most promising techniques [4]–[8]. In MEC, the computing and storage resources are deployed from the central cloud server to edge nodes (SBS’s) of the network [9]. As a result, the computing task of stringent low-latency requirements can be processed by the proximate SBS’s without being sent to the cloud server. Furthermore, the close proximity between UEs and SBS’s enables the network to collect more recent UE information such as behaviors, locations and environment, which facilitates the network to provide agile services.

Meanwhile, the millimeter wave (mmWave) transmission endowed with massive antenna arrays has been adopted in the fifth generation (5G) wireless communication systems [10]. The synergy of mmWave transmission and MEC has been well regarded as a strong candidate for the future wireless networks. By exploiting the large available mmWave spectrum and the high antenna gain [11], the mmWave massive multiple-input multiple-output (MIMO) transmission enables the short-distance high-speed data communications between UEs and their associated SBS’s [12]. However, the implementation of fully digital mmWave massive MIMO transmission incurs prohibitively expensive hardware costs. To cope with this problem, the hybrid analog and digital beamforming has been proposed [13]–[15]. In hybrid beamforming, the precoding process is divided into two phases, namely the analog precoding with a large number of phase shifters, and the digital precoding with a smaller number of Radio Frequency (RF) chains. By carefully designing the analog and digital precoders, it has been shown that the low-cost hybrid beamforming can achieve comparable performance as compared to the fully digital beamforming.

In addition, user association has become an important problem in MEC as SBS’s are expected to be densely deployed with each SBS covering a much smaller area as compared to the conventional macro base stations. Thus, a carefully designed user association can match each UE with the best SBS by maximizing the transmission data rate while minimizing its incurred interference [16]. However, the user association problem is NP-hard due to its combinatorial nature. Some pioneering studies attempted to solve the problem by simplifying the deployment scenarios. For instance, in [17], the association problem was addressed based on the Euclidean distance without taking into account the

Manuscript received September 8, 2020; revised February 10, 2021; accepted April 25, 2021. Date of publication April 28, 2021; date of current version December 17, 2021. This work was supported in part by the Shenzhen Institute of Artificial Intelligence and Robotics for Society (AIRS) under Grant AC01202005001, in part by the Shenzhen Science and Technology Innovation Committee under Grants ZDSYS20170725140921348 and JCYJ2019081317083617, and in part by the National Natural Science Foundation of China under Grant 61731018. The review of this article was coordinated by Prof. G. Marchetto. (Corresponding author: Man-On Pun.)

Guanchong Niu is with the School of Science, and Engineering, The Chinese University of Hong Kong, Shenzhen 518172, China, and also with the Shenzhen Research Institute of Big Data, Shenzhen 518172, China (e-mail: 216019007@link.cuhk.edu.cn).

Qi Cao is with the School of Science and Engineering, The Chinese University of Hong Kong, Shenzhen 518172, China (e-mail: cnqcao@hotmail.com).

Man-On Pun is with the School of Science and Engineering, The Chinese University of Hong Kong, Shenzhen 518172, China, with the Shenzhen Research Institute of Big Data, Shenzhen 518172, China, and also with the Shenzhen Key Laboratory of IoT Intelligent System and Wireless Network Technology, Shenzhen 518172, China (e-mail: SimonPun@cuhk.edu.cn).

Digital Object Identifier 10.1109/TVT.2021.3076353
dynamic nature of the wireless channel model. Furthermore, [18] proposed a user association algorithm based on the physical resource block (PRB) table. In both [17], [18], user association was investigated without considering the potential impact caused by multiuser interference.

Finally, power allocation is considered as one of the most effective methods to achieve required quality of service (QoS) for each UE by better managing co-channel interference in multiuser networks [19], [20]. However, joint precoding and power allocation design is usually analytically intractable due to its non-convex nature [21]. To circumvent this obstacle, [22] proposed to optimize the power allocation for the sum-rate capacity maximization problem by utilizing the water-filling algorithm, assuming that the spatial channels of all UEs are perfectly orthogonal. Clearly, this assumption is an issue of concern in practice. In addition, the task of finding the optimal precoding and power allocation design that maximizes the spectral efficiency is very challenging due to the fact that all system parameters under consideration are highly coupled.

Motivated by the aforementioned challenges, this work studies the MEC system equipped with mmWave MIMO as shown in Fig. 1. As depicted in Fig. 2, the user association problem is first transformed into an intra-SBS interference minimization problem before being solved by a distributed max-$k$-cut (M$k$C)-based algorithm in Section IV. In contrast to the suboptimal resource allocation algorithm proposed in [23], we consider the user association problem for multiple SBS’s based on intra-SBS interference minimization. Finally, we cast the non-convex problem as a rank-constrained difference of two convex functions (D.C.) programming problem to solve the precoding and power allocation in Section V.

In summary, the main contributions of this work are summarized as follows:

- We first propose a novel distributed M$k$C-based algorithm for user association using practical wireless channel models. Furthermore, the performance bounds of the proposed algorithm are derived;
- Furthermore, joint precoding and power allocation is proposed using the rank-constrained D.C. programming technique. More specifically, the original non-convex optimization problem is formulated as a rank-constrained D.C. programming problem before an iterative algorithm is devised to maximize the weighted sum-rate (WSR) with QoS constraints for each UE. The convergence of the proposed iterative algorithm is proven.

The rest of this paper is organized as follows. First, an introduction to the D.C. programming techniques and the M$k$C problem is provided in Section II. Next, the MEC system model is presented in Section III before an M$k$C-based algorithm is developed for user association in Section IV. After that, an iterative joint precoding and power allocation algorithm using the rank-constrained D.C. programming techniques is established in Section V followed by the derivation of the performance bound of the proposed iterative algorithm in Section VI. Finally, extensive computer simulation results are presented in Section VII.

**Notation:** Uppercase boldface and lowercase boldface letters are used to denote matrices and vectors, respectively. $I_N$ represents the identity matrix with size $N \times N$. $A^T$ and $A^H$ are the transpose and conjugate transpose of $A$, respectively. $|a|$ denotes the $i$-th element of vector $a$. In addition, $\|A\|$ stands for the $\ell_2$ norm of $A$ while $|A|$ denotes the absolute value of $A$. $|A|$ is the cardinality of the enclosed set $A$. Furthermore, $\text{rank}(A)$ and $\text{trace}(A)$ represent the rank and trace of $A$, respectively. $A \succeq 0$
Table I: Summary of Key Notations

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_{k,n}$</td>
<td>The entry of $A$ representing the belonging of the $n$-th user</td>
</tr>
<tr>
<td>$V_i$</td>
<td>The $i$-th vertex</td>
</tr>
<tr>
<td>$K$</td>
<td>The number of SBS’s</td>
</tr>
<tr>
<td>$a(\phi, \theta)$</td>
<td>Array response vector with azimuth angle $\phi$ and elevation angle $\theta$</td>
</tr>
<tr>
<td>$N_R$</td>
<td>The number of antennas in the receiver</td>
</tr>
<tr>
<td>$N_T$</td>
<td>The number of antennas in the transmitter</td>
</tr>
<tr>
<td>$w_{k,u}$</td>
<td>Analog beamforming vector employed by the $u$-th UE of the $k$-th SBS</td>
</tr>
<tr>
<td>$H_{k,u}$</td>
<td>The $u$-th UE’s channel model under the $k$-th SBS</td>
</tr>
<tr>
<td>$f_{k,u}$</td>
<td>Digital beamforming vector for the $u$-th user of the $k$-th SBS</td>
</tr>
<tr>
<td>$P_k$</td>
<td>Digital precoding matrix for UEs under the $k$-th SBS</td>
</tr>
<tr>
<td>$v_{k,u}$</td>
<td>Analog beamforming vector for the $u$-th user of the $k$-th SBS</td>
</tr>
<tr>
<td>$V_k$</td>
<td>Analog precoding matrix for UEs under the $k$-th SBS</td>
</tr>
<tr>
<td>$p$</td>
<td>The set for all users’ power allocation vector</td>
</tr>
<tr>
<td>$S_k$</td>
<td>The maximal number of users for the $k$-th SBS</td>
</tr>
<tr>
<td>$\Omega_{k,u,j}$</td>
<td>The sum of the weights of the edges from a vertex $u$ in the $k$-th SBS</td>
</tr>
<tr>
<td>$\omega_{k,u,j}$</td>
<td>The weight between two vertices</td>
</tr>
<tr>
<td>$N_k$</td>
<td>The partition set for the $k$-th SBS with size $</td>
</tr>
<tr>
<td>$\bar{x}$</td>
<td>The reformulated digital precoder to release non-convex rank constraint</td>
</tr>
<tr>
<td>$F_i$</td>
<td>The set of digital precoder in the rank-constrained problem</td>
</tr>
<tr>
<td>$\bar{g}$</td>
<td>The effective array gain of the $i$-th UE of the $k$-th SBS</td>
</tr>
<tr>
<td>$x$</td>
<td>The weight of rate for the $u$-th UE of the $k$-th SBS</td>
</tr>
<tr>
<td>$P_k$</td>
<td>Transmit power constraint for the $k$-th SBS</td>
</tr>
<tr>
<td>$U^{(m)}$</td>
<td>Eigenvectors corresponding to the $N - 1$ smallest eigenvalues of $F_k^{(m)}$</td>
</tr>
</tbody>
</table>

Fig. 3. Illustration of an M\&C problem in a graph $|V| = 8, |E| = 10$ and $k = 3$. The maximal solution of this M\&C problem is $6 + 5 + 5 = 16$.

The CCP algorithm is formulated as
\[
\begin{align*}
\text{minimize} & \quad f_0(x) - g_0(x^{(m)}) - \langle \nabla g_0(x^{(m)}), x - x^{(m)} \rangle \\
\text{subject to} & \quad f_i(x) - g_i(x) \leq 0, \quad i = 1, \ldots, M.
\end{align*}
\] (2)
where $x^{(m)}$ represents the state at the $m$-th iteration.

Since Equation (2) is convex, it can be straightforwardly solved by most convex optimization software packages [25]. Furthermore, by taking advantage of the convex property, the $(m+1)$-th solution is always better than the $m$-th solution, which guarantees the convergence of the iterative algorithm stated in Equation (2). The detailed CCP algorithm can be found in [26].

B. Review of Capacitated M\&C Problem

For the M\&C problem, we are given a weighted graph $H = (V, E)$ consisting of a vertex set and an edge set denoted by $V$ and $E$, respectively. The goal is to partition $V$ into $k$ non-empty sets in such a way that the sum of the weights of edges $w_{ij}$ between different partitions is maximized as shown in Fig. 3. It is worth noting that the partition is not unique in general. Some M\&C-related problems are summarized as follows:

- MC: A special case of M\&C when $k = 2$ [27].
- Conventional M\&C: Just divide $V$ vertices into $k$ disjoint non-empty partitions without additional constraints [28].
- Max Steiner $k$-cut: Each partition must include at least one vertex from a given set $T = \{t_1, t_2, \ldots, t_l\} \subseteq V$, where $l \geq k$ [29].
- M\&C with given sizes: The partitions $V_1, V_2, \ldots, V_k$ must satisfy $|V_i| = s_i$ with a given set $\{s_1, s_2, \ldots, s_k\}$, for all $1 \leq i \leq k$ [30].
- Capacitated M\&C: Given a set $\{s_1, s_2, \ldots, s_k\}$, the partitions $V_1, V_2, \ldots, V_k$ must satisfy $|V_i| \leq s_i$ [31].

Since the inherent NP-hardness, the capacitated M\&C can only find a locally optimal solution in polynomial time. A binary edge formulation (i.e. $w_{ij} = 0, 1$) is proposed in [32], where the integer variable $x_{ij}$ for $i, j \in V$ is defined as:
\[
x_{ij} = \begin{cases} 
1 & \text{if } i \text{ and } j \text{ are in the same partition}, \\
0 & \text{otherwise}. 
\end{cases}
\] (3)

Thereby the M\&C problem is formulated as the following integer linear programming (ILP) formulation [33]:
\[
\begin{align*}
\text{maximize} & \quad \sum_{i, j \in V, i < j} \omega_{ij}(1 - x_{ij}) \\
\text{subject to} & \quad x_{ih} + x_{hj} - x_{ij} \leq 1, \quad \forall i, j, h \in V.
\end{align*}
\] (4b)

represents that $A$ is a positive semi-definite matrix. Finally, $\nabla f(A)$ represents the gradient of function $f(A)$.

Symbol: Table I summarizes the key symbols applied in this paper.
where constraints (4b) and (4c) are the triangle and clique inequalities respectively. Triangle inequality implies the logical conditions that if \( x_{ih} = 1 \) and \( x_{hj} = 1 \) hold, then by transitivity the value of \( x_{ij} \) should be 1 as well, meaning vertices \( i \) and \( j \) are in the same partition. Constraint (4c) imposes that at least two from every subset of \( k + 1 \) vertices have to be in the same partition. Constraints (4b) and (4c) imply that there are at most \( k \) partitions.

### III. SYSTEM MODEL AND PROBLEM FORMULATION

#### A. MEC System Model

We consider an MEC system of \( k \) SBS’s serving \( N \) UEs as shown in Fig. 1. We denote by \( N = \{1, 2, \ldots, N\} \) the UE index set while \( N_k = \{1, 2, \ldots, N_k\} \) stands for the UE index set for UEs served by the \( k \)-th SBS. Furthermore, we assume that each SBS can serve maximal \( S_k \) users, i.e., \( N_k \leq S_k \) for \( k = 1, 2, \ldots, K \). The set of \( k \) is denoted by \( K = \{1, 2, \ldots, K\} \).

Assuming that each UE can attach to only one SBS, we define \( A(n) = k \) to denote that the \( n \)-th UE is attached to the \( k \)-th SBS. Using \( A(n) \), we further define a UE-SBS association matrix denoted by \( A \) of dimension \( K \times N \). The entry of \( A \) is given by:

\[
A_{k,n} = \begin{cases} 
1 & \text{if } A(n) = k; \\
0 & \text{otherwise}, 
\end{cases}
\]

for \( k \in K \) and \( n \in N \).

#### B. Downlink Transmission Model

As shown in Fig. 4, the SBS’s are equipped with mmWave MU-MIMO systems, each with \( N_{RF} \) RF chains and \( N_T \) antennas aiming to transmit \( N_k \) data streams to \( N_k \) of \( N \) receivers. Furthermore, we assume that SBS’s operate in orthogonal channels. As a result, inter-SBS interference is not considered in the sequel.

Following the same design commonly implemented in the literature [34], only one data stream is designated to each scheduled receiver. We use \( s_k \in \mathbb{C}^{N_k} \) to denote the data to be transmitted with \( \mathbb{E}[s_k s_k^H] = \frac{1}{N_k} \mathbb{I}_{N_k} \).

Without loss of generality, we focus our following discussions on the \( k \)-th SBS. The hybrid precoding system first multiplies \( s_k \) with the digital precoding matrix \( F_k = [f_{k,1}, f_{k,2}, \ldots, f_{k,N_k}] \) with \( f_{k,u} \in \mathbb{C}^{N_{RF} \times 1} \) being the digital beamforming vector for the \( u \)-th UE. After that, the output signal is multiplied by the analog precoding matrix \( V_k = [v_{k,1}, v_{k,2}, \ldots, v_{k,N_{RF}}] \) with \( v_{k,u} \in \mathbb{C}^{N_T \times N_{RF}} \). The resulting precoded signal \( x_k \) of length \( N_T \) can be expressed as

\[
x_k = V_k \cdot F_k \cdot s_k = V_k \sum_{n \in N} A_{k,n} f_{k,n} s_k,n.
\]

The precoded signal \( x_k \) is then broadcast to \( N_k \) UEs. The signal received by the \( n \)-th UE associated with the \( k \)-th SBS is given by [35]

\[
y_{k,u} = H_{k,u} x_k + n_{k,u} = H_{k,u} V_k f_{k,u} s_k,u + H_{k,u} V_k \sum_{n \in N_k \setminus u} A_{k,n} f_{k,n} s_k,n + n_{k,u},
\]
where $H_{k,u} \in \mathbb{C}^{N_R \times N_T}$ is the MIMO channel matrix between the $k$-th SBS and its $u$-th UE, and $n_{k,u}$ is the complex additive white Gaussian noise with zero mean and variance equal to $\sigma_u^2$.

Assuming that UEs under consideration are all low-cost terminals with analog-only decoding, the decoded signal according to $s_{k,u}$ is given by

$$\hat{s}_{k,u} = w_{k,u}H_{k,u}V_k f_{k,u}s_{k,u} + \tilde{w}_{k,u}, \quad (8)$$

where $w_{k,u}$ of length $N_R$ is the analog beamforming vector employed by the $u$-th UE of the $k$-th SBS with the power constraint of $\|w_{k,u}\|^2 = 1$ and

$$\tilde{w}_{k,u} = H_{k,u}V_k \sum_{n \in \mathbb{N}} \alpha_{k,u,n} f_{k,u}s_{k,u} + n_{k,u}, \quad (9)$$

Note that the first term in Equation (8) stands for the desired signal while the second term $\tilde{w}_{k,u}$ is the sum of the receiver noise and interference from other UEs in the same SBS network.

C. Channel Model

It has been shown that the mmWave wireless channel can be well modeled by the Saleh-Valenzuela model [15]. We assume that the channel state information (CSI) is perfectly known. Following the same technique proposed in [34], each scatter is assumed to contribute only one propagation path. As a result, the $u$-th UE’s channel model under the $k$-th SBS can be modeled as

$$H_{k,u} = \sqrt{N_TN_R} \sum_{l=1}^{L_{k,u}} \alpha_{k,u,l} \times a_R(\phi_{k,u,l}, \theta_{k,u,l}, \theta_{k,u,l}^e)^H \cdot a_H(\phi_{k,u,l}, \theta_{k,u,l}, \theta_{k,u,l}^e)^T, \quad (10)$$

where $L_{k,u}$ is the number of scatter of the $u$-th UE’s channel. Furthermore, $\alpha_{k,u,l}$ is the complex path gain with zero mean while $\{\theta_{k,u,l}, \phi_{k,u,l}^e\}$ and $\{\theta_{k,u,l}, \phi_{k,u,l}^e\}$ are the azimuth and elevation angles of arrival (AoA) and angles of departure (AoD), respectively. $a$ is the array response vector. For the uniform planar array (UPA) with size $W \times Q$ considered in this work, the array response vector $a$ is represented by

$$a(\phi, \theta) = \frac{1}{\sqrt{WQ}} [1, e^{j\kappa d \sin \phi \sin \theta + \cos \theta}, \ldots, e^{j\kappa d (P-1) \sin \phi \sin \theta + (Q-1) \cos \theta}]^T, \quad (11)$$

where $\kappa = \frac{2\pi}{\lambda}$ is the wavenumber and $d$ is the distance between two adjacent antennas.

D. Analog Precoding

We firstly design the analog precoding schemes on both UEs and SBS’s. The power constraints of analog precoders can be formulated by

$$\|w_{k,u}\|^2 = 1/N_R, \quad i = 1, 2, \ldots, N_R;$$

$$\|v_{k,u}\|^2 = 1/N_T, \quad i = 1, 2, \ldots, N_T. \quad (12)$$

It is well-known that distinct array response vectors are asymptotically orthogonal as the number of antennas in an antenna array goes to infinity [36], i.e.

$$\lim_{N \to \infty} a^H(\phi_{k,u}^1, \theta_{k,u}^1) \cdot a(\phi_{k,u}^2, \theta_{k,u}^2) = \delta(k-\ell)\delta(u-v). \quad (13)$$

However, since the infinite number of antennas is not practical, the residual interference must be taken into consideration for the analog precoding design. Recalling the channel model in Equation (10), we can asymptotically orthogonalize the transmitted signals by optimizing the design of $w_{k,u}$ and $v_{k,u}$ with given user association:

$$\{w_{k,u}, v_{k,u}\} = \arg \max_{\{w_{k,u}, v_{k,u}\}} \sum_{u=1}^{M_B} \log_2 (1 + \text{SINR}(\hat{w}_{k,u}, \hat{v}_{k,u}))$$

subject to $\hat{v}_{k,u} \in B_{k,u}^r$, $\hat{w}_{k,u} \in B_{k,u}^e$, \quad (14)

where $B_{k,u}^r$ and $B_{k,u}^e$ are the array response vectors for the SBS and UE denoted by

$$B_{k,u}^r = [a_R(\phi_{k,u,1}^r, \theta_{k,u,1}^r), \ldots, a_R(\phi_{k,u,n_u}^r, \theta_{k,u,n_u}^r)],$$

$$B_{k,u}^e = [a_T(\phi_{k,u,1}^e, \theta_{k,u,1}^e), \ldots, a_T(\phi_{k,u,n_u}^e, \theta_{k,u,n_u}^e)]. \quad (15)$$

The analog precoders are selected from the array response vectors by taking advantage of the perfect CSI.

Furthermore, SINR$(\hat{w}_{k,u}, \hat{v}_{k,u})$ is given by

$$\text{SINR}(\hat{w}_{k,u}, \hat{v}_{k,u}) = \frac{\|\hat{w}_{k,u}^H H_{k,u} \hat{v}_{k,u}\|^2}{\sum_{i \notin N_u, n \neq u} A_{k,u,n} \|H_{k,u} \hat{v}_{k,u}\|^2 + 1}, \quad (16)$$

where $\gamma = \frac{P}{N^2\sigma_u^2}$ is the SNR for each user. In deriving the results above, we have assumed $f_{k,u} f_{k,u}^H \approx I_k$. As a result, the optimal analog precoders at both transmitter and receiver can be straightforwardly found by exhaustively searching in the feasible sets of $B_{k,u}^r$ and $B_{k,u}^e$.

E. Problem Formulation

For notational simplicity, we denote by $g_{k,u}^H$ the effective array gain of the $u$-th UE with

$$g_{k,u}^H = w_{k,u}^H H_{k,u} V_k. \quad (17)$$

Then, the channel capacity of the $u$-th UE under the $k$-th SBS is given by

$$R_{k,u} = \log_2 \left( 1 + \frac{p_{k,u} \|g_{k,u}^H f_{k,u}\|^2}{\sum_{i \notin N_u} A_{k,n} p_{k,i} \|g_{k,u}^H f_{k,i}\|^2 + \sigma_u^2} \right). \quad (18)$$

Subsequently, the system WSR can be formulated as

$$R_{tot} = \sum_{k=1}^{K} \sum_{u=1}^{N_k} \tau_{k,u} R_{k,u}(p_k, F_k, A), \quad (19)$$

where $\tau_k = [\tau_{k,1}, \tau_{k,2}, \ldots, \tau_{k,N_k}]$ with $\sum_{u=1}^{N_k} \tau_{k,u} = N_k$ is the vector containing the corresponding weights for UEs while
$p_k = [p_{k,1}, p_{k,2}, \ldots, p_{k,N_k}]$ with $\sum_{u=1}^{N_k} p_{k,u} \leq P_k$ is the vector containing the transmit power of the $k$-th SBS. We further denote by $p = [p_1, \ldots, p_K]$ the vector containing all users’ power allocation.

Finally, the optimal design of user scheduling and hybrid precoding matrices, as well as power allocation can be formulated as

$$
P_1 : \max_{p,F,A} \quad R_{tot}(p, F, A)
$$

subject to

$$
P_2 : \max_{A} \sum_{k=1}^{K} \sum_{u=1}^{N_k} \log_2 (1 + \text{SINR} (\tilde{w}_{k,u} \cdot \tilde{v}_{k,u}))
$$

subject to

$$
P_3 : \sum_{n=1}^{N} A_{k,n} \leq S_k, \quad \forall k \in K,
$$

$$
P_4 : \sum_{k=1}^{K} A_{k,n} = 1, \quad \forall n \in N.
$$

where $P_1$ ensures that each RF chain is of unit power. The total transmit power is limited by $P_k$ as shown in $P_2$ while $P_3$ ensures that the QoS for the $u$-th UE of the $k$-th SBS with a minimum data rate of $\lambda_{k,u}$. In addition, in constraints $P_3$ and $P_4$, the number of UEs served by one SBS is constrained while one UE can only be attached to one SBS for transmission.

Since the problem $P_1$ is highly non-convex, it is analytically intractable to derive a closed-form solution to $P_1$. In Section IV, we first formulate the UE association problem as a capacitated M$k$C problem, which can be solved by a distributed local-search algorithm. After that, a general rank-constrained D.C. programming technique is developed in Section V to solve the local optimal digital precoder and power allocation by approximating the problem in each iteration as a standard convex optimization problem.

IV. MAX-$k$-CUT PROBLEM FOR USER ASSOCIATION

In the sequel, the user allocation is designed to minimize the interference caused by the side lobe of beamformers based on the proposed capacitated M$k$C algorithm before the performance bounds are derived.

Since the antenna number is finite in practice, the residual interference must be considered in the analog precoding design. Recalling the channel model presented in Equation (10), we can asymptotically orthogonalize the transmitted signals by optimizing the design of $w_u$ and $v_u$:

$$
P_2 : \max_{A} \sum_{k=1}^{K} \sum_{u=1}^{N_k} \log_2 (1 + \text{SINR} (\tilde{w}_{k,u} \cdot \tilde{v}_{k,u}))
$$

subject to

$$
P_3 : \sum_{n=1}^{N} A_{k,n} \leq S_k, \quad \forall k \in K;
$$

$$
P_4 : \sum_{k=1}^{K} A_{k,n} = 1, \quad \forall n \in N.
$$

A. A Distributed Local-Search Algorithm

As exhibited in [37], the user association problem is NP-hard since the number of possible combinations between UEs and SBS’s is $\prod_{j=1}^{K} (N - \sum_{i=0}^{j-1} N_{i,j})$, where we set $N_0 = 0$. To cope with this obstacle, a heuristic algorithm is proposed by exploiting the interference minimization. The intra-SBS interference minimization problem is formulated as

$$
\min_{A} \sum_{k=1}^{K} \sum_{u=1}^{N_k} \sum_{n=1}^{N} A_{k,n} |w_{k,u}^H H_{k,u} v_{k,n}|^2.
$$

In the MEC system model, we assume that the channels operated in different SBS’s are orthogonal. As a result, the inter-SBS interference can be eliminated by the orthogonal channels. Inspired by the M$k$C technique in Section II-B, we can transform the intra-SBS minimization problem in Equation (22) to maximize the inter-SBS interference as follows:

$$
\max_{A} \quad I_{tot} = \sum_{k=1}^{K} \sum_{u=1}^{N_k} \sum_{n=1}^{N} A_{k,n} |w_{k,u}^H H_{k,u} v_{k,n}|^2.
$$

Thus, as shown in Fig. 5, the counterpart weight between two vertices (UEs) is given by

$$
\omega_{k,u,j,v} = \begin{cases} 0, & \text{if } j = k, \\ |w_{k,u}^H H_{k,u} v_{j,v}|^2, & \text{otherwise.} \end{cases}
$$

In this way, the problem in Equation (24) can be transformed as an M$k$C problem, which can be solved using a distributed local-search algorithm. We can regard an SBS network as one “cut”. To maximize the inter-SBS interference, we propose a local-search algorithm to solve the problem $P_2$. We first denote by $\Omega_{k,u,j}$ the sum of the weights of the edges from a vertex $u$ in Set $k$ to all the vertices in Set $j$. $\Omega_{k,u,j}$ takes the following form:

$$
\Omega_{k,u,j} = \sum_{q=1}^{N} \omega_{k,u,j,q}.
$$

A distributed local-search algorithm for the capacitated M$k$C problem [32] is given as follows:
1) **Initialization**: Partition the UEs into $K$ sets, $N_1, N_2, \ldots, N_K$ before randomly initializing $\sum_k A_{k,u} \leq S_k$, i.e., $N_k \leq S_k$ for $k = 1, 2, \ldots, K$.

2) **Iterative Step**: Determine if there is an association pair $q$ in Set $j$ and $u$ in Set $k$ for which

$$N_k \Omega_{k,u} + N_j \Omega_{j,q} > N_k \Omega_{j,k,u} + N_j \Omega_{k,j,q}.$$  

(26)

If such a pair exists, we reassign the $q$-th UE to the $k$-th SBS while the $u$-th UE to the $j$-th SBS.

3) **Termination**: Stop the iteration when all pairs violate Equation (26), i.e.,

$$N_k \Omega_{k,u} + N_j \Omega_{j,q} \leq N_k \Omega_{j,k,u} + N_j \Omega_{k,j,q}.$$  

(27)

for all $k, j \in K$ and $u \in N_k$, $q \in N_j$.

The computational complexity of the local-search algorithm can be analyzed as follows: In each iteration, one UE needs to examine at most $N - 1$ times to check if any pair satisfies Equation (26). Thus, the local-search algorithm has computational complexity of the order of $O(N^2)$.

Since no central coordinators are assumed in the search algorithm proposed above, the proposed iterative algorithm can be executed in a distributed manner. In the MEC system, we consider a fully coordinated scenario where the SBS’s collaborate to address the user association problem. SBS’s are able to share with each other the CSI of UEs. Since the inter-SBS interference can be eliminated by the orthogonal channels, we can optimize the user association using the proposed local-search algorithm by maximizing the inter-SBS interference.

**B. Worst-Case Analysis**

The theorem below characterizes the performance lower bound of the distributed local-search algorithm for all $K \geq 2$.

We define

$$I_{ls} = \sum_{k=1}^{K} \sum_{u=1}^{N_k} \Omega_{k,u}.$$  

(28)

where $I_{ls}$ is the solution given by the local-search algorithm.

**Theorem 1**: Assuming that the maximal value of a MkC problem is $I_{max}$. The solution obtained using the local-search algorithm has a value no smaller than $1 - \frac{1}{K}$ of the optimal solution value, i.e.,

$$\frac{I_{ls}}{I_{max}} \geq 1 - \frac{1}{K}.$$  

(29)

**Proof**: See Appendix A.

Since it is non-trivial to analytically evaluate Equation (28), we will use computer simulation to confirm the effectiveness of our proposed association algorithm in Section VII.

**V. PROPOSED ITERATIVE ALGORITHM FOR RANK-CONSTRAINED D.C. PROBLEM**

In this section, we will derive the optimal digital precoder and power allocation at each SBS for a given association matrix $A$ and a set of analog precoders. Without loss of generality, we focus the following discussion on the $k$-th group while omitting the subscript $k$ for presentational simplicity. In addition, we set the number of UEs attached to each SBS to $\bar{N}$ while the maximum total transmit power $P_0$.

Then, the optimal design of the precoding matrices as well as power allocation can be formulated as

$$\mathcal{P}_3 : \text{maximize } R_{tot}(\mathbf{p}, \mathbf{F})$$  

subject to

$$C_1 : \| \mathbf{Vf}_u \|^2 = 1, \quad u = 0, 1, \ldots, \bar{N},$$  

(26)

$$C_2 : \sum_{u=1}^\bar{N} p_u \leq P_0,$$

$$C_3 : R_u \geq \lambda_u, \quad u = 0, 1, \ldots, \bar{N}.$$  

(27)

Since the tasks of precoding design and power allocation are highly coupled, we will propose a joint optimization algorithm to maximize the WSR in this section.

**A. Rank-Constrained D.C. Problem**

It can be easily seen that the challenge in solving $\mathcal{P}_3$ is the non-convexity of $R_{tot}(\mathbf{p}, \mathbf{F})$. To cope with this problem, we first express the digital precoder as

$$\mathbf{F}_u = p_u \mathbf{f}_u \mathbf{f}_u^H,$$  

(28)

with constraints $\text{rank}(\mathbf{F}_u) \leq 1$ and $\mathbf{F}_u \succeq \mathbf{0}$.

Furthermore, the constraints $C_1$ and $C_2$ in $\mathcal{P}_3$ can be transformed as

$$\sum_{u=1}^\bar{N} p_u \| \mathbf{Vf}_u \|^2 = \sum_{u=1}^\bar{N} \mathbf{Vf}_u \mathbf{V}^H \leq P_0.$$  

(29)

It is worth noting that the optimal digital precoder $\mathbf{f}_u^*$ is the eigenvector corresponding to the only non-zero eigenvalue $p_u^*$ of optimal $\mathbf{F}_u^*$, where $p_u^*$ is the optimal power allocation [38].

The variable $p_u$ is designed to ensure that $C_1$ in $\mathcal{P}_3$ is satisfied. Substituting Equation (31) into Equation (18), we have

$$R_u(\mathbf{F}_u) = \log_2 \left( 1 + \frac{g_u^H \mathbf{F}_u g_u}{\sum_{i \neq u} g_i^H \mathbf{F}_u g_i + \sigma_u^2} \right)$$

$$= \log_2 \left( \sum_{i=1}^\bar{N} g_i^H \mathbf{F}_i g_i + \sigma_u^2 \right)$$

$$- \log_2 \left( \sum_{i=1}^\bar{N} g_i^H \mathbf{F}_i g_i + \sigma_u^2 \right).$$  

(30)

Therefore, $\mathcal{P}_3$ can be reformulated as a D.C. problem with a rank constraint:

$$\mathcal{P}_4 : \text{maximize } f(\mathbf{F}) - g(\mathbf{F})$$  

subject to

$$C_1 : \sum_{u=1}^\bar{N} \mathbf{V}^H \mathbf{F}_u \mathbf{V} \leq P_0,$$

$$C_2 : R_u \geq \lambda_u, \quad u = 1, 2, \ldots, \bar{N}.$$  

(31)
where $\mathcal{F} = [\tilde{F}_1, \tilde{F}_2, \ldots, \tilde{F}_N]$ and
\begin{align}
f(\mathcal{F}) &= \sum_{u=1}^{\tilde{N}} \tau_u \log_2 \left( \sum_{i=1}^{\tilde{N}} g_u^H \tilde{F}_i g_u + \sigma_u^2 \right), \\
g(\mathcal{F}) &= \sum_{u=1}^{\tilde{N}} \tau_u \log_2 \left( \sum_{i=1}^{\tilde{N}} g_u^H \tilde{F}_i g_u + \sigma_u^2 \right).
\end{align}

Next, we propose to transform the QoS constraint $C_2$ into a convex constraint as follows:
\begin{equation}
g_u^H \tilde{F}_u g_u + (1 - 2^\lambda_u) \left( \sum_{i=1}^{\tilde{N}} g_u^H \tilde{F}_i g_u + \sigma_u^2 \right) \geq 0. \tag{37}
\end{equation}

In the following, we devise an iterative algorithm to solve $\mathcal{P}_4$ as a standard D.C. programming problem by first ignoring $C_4$. More specifically, the optimal $\mathcal{F}^*$ is obtained by iteratively solving $\hat{F}_u$ for $u = 1, 2, \ldots, \tilde{N}$. Similar to the procedures in [26], the iterative algorithm gives the optimal $\hat{F}^{(m+1)}_u$ at the $m$-th iteration by solving the following convex problem (See Appendix B):
\begin{align}
\text{maximize} & \quad f(\hat{F}_u) - g(\hat{F}^{(m)}_u) - \langle \hat{\nabla} g(\hat{F}^{(m)}_u), \hat{F}_u - \hat{F}^{(m)}_u \rangle \\
\text{subject to} & \quad C_1, C_2, C_3 \text{ in } \mathcal{P}_4,
\end{align}

where $\langle \cdot, \cdot \rangle$ denotes the inner product of two matrices, i.e., $\langle A, B \rangle = \text{trace}(A^H B)$. The gradient of $g(\hat{F}^{(m)}_u)$ can be computed as
\begin{equation}
\hat{\nabla} g(\hat{F}^{(m)}_u) = \sum_{j=1}^{\tilde{N}} \frac{w_{ij}}{\ln 2} \sum_{l=1, l \neq j}^{\tilde{N}} g_u^H \tilde{F}_i g_j + \sigma_j^2 g_j g_j^H. \tag{39}
\end{equation}

It is worth noting that $\langle \hat{\nabla} g(\hat{F}^{(m)}_u), \hat{F}_u - \hat{F}^{(m)}_u \rangle$ is a real value since $\hat{\nabla} g(\hat{F}^{(m)}_u)$ and $\hat{F}_u - \hat{F}^{(m)}_u$ are both Hermitian.

B. The General RCOP

To solve $\mathcal{P}_4$, we first introduce a rank constrained optimization problem (RCOP). Inspired by RCOP, we will then propose an iterative algorithm to solve $\mathcal{P}_4$ by formulating a convex optimization problem in each iteration.

We now consider the rank constraint $C_4$ in $\mathcal{P}_4$. A general RCOP to optimize a convex objective subject to a set of convex and rank constraints can be formulated as follows:
\begin{align}
\mathcal{P}_5 : \quad & \text{minimize} \quad f(X) \\
\text{subject to} & \quad X \succeq 0; \\
& \quad X \in \mathcal{C}; \\
& \quad \text{rank}(X) \leq r, \tag{40}
\end{align}

where $f(X)$ is a convex function, $\mathcal{C}$ is the set of given convex constraints and $X \in \mathbb{C}^{\tilde{N} \times \tilde{N}}$ is a general positive semidefinite matrix.

As shown in [39], RCOP can be solved by an iterative method to gradually approach the constrained rank. In the $m$-th iteration, we solve the following semidefinite programming (SDP) problem:
\begin{align}
\text{minimize} & \quad f(X^{(m+1)}) + w e^{(m+1)} \\
\text{subject to} & \quad X^{(m+1)} \succeq 0, \\
& \quad X^{(m+1)} \in \mathcal{C}, \\
& \quad e^{(m+1)} I_{\tilde{N} - r} - U^{(m)H} X^{(m+1)} U^{(m)} \succeq 0, \\
& \quad e^{(m+1)} \leq e^{(m)}, \tag{41}
\end{align}

where $w > 0$ is the weighting factor. Using the eigenvalue decomposition (EVD), $U^{(m)} \in \mathbb{C}^{\tilde{N} \times (\tilde{N} - r)}$ is the orthonormal eigenvectors corresponding to the $\tilde{N} - r$ smallest eigenvalues of $X^{(m)}$ solved in the previous iteration. In the first iteration where $m = 0$, $e^{(0)}$ is the $(\tilde{N} - r)$-th smallest eigenvalue of $X^{(0)}$ that can be obtained by
\begin{align}
\text{minimize} & \quad f(X^{(0)}) \\
\text{subject to} & \quad X^{(0)} \succeq 0; \\
& \quad X^{(0)} \in \mathcal{C}, \tag{42}
\end{align}

and $U^{(0)}$ is the eigenvectors corresponding to $\tilde{N} - r$ smallest eigenvalues of $X^{(0)}$.

C. Iterative Algorithm for $\mathcal{P}_4$

It is straightforward to prove that Equation (38) is a concave optimization problem as shown in Section V-A. By combining $\mathcal{P}_4$ and $\mathcal{P}_5$, an iterative algorithm for the rank-constrained D.C. programming problem is derived. The optimal $g(\hat{F}^{(m+1)}_u)$ in the $m$-th iteration is given by solving the following convex problem:
\begin{align}
\mathcal{P}_6 : \quad & \text{minimize} \quad t(\hat{F}_u, e^{(m+1)}) \\
\text{subject to} & \quad C_1 : \sum_{u=1}^{\tilde{N}} V^H \hat{F}_u V \leq \tilde{N}, \\
& \quad C_2 : R_u \geq \lambda_u, u = 1, 2, \ldots, \tilde{N}, \\
& \quad C_3 : \hat{F}_u \succeq 0, \\
& \quad C_4 : e^{(m+1)} I_{\tilde{N} - 1} - U^{(m)H} \hat{F}_u U^{(m)} \succeq 0, \\
& \quad C_5 : e^{(m+1)} \leq e^{(m)}, \tag{43}
\end{align}

where $U^{(m)}$ is the eigenvectors corresponding to the $\tilde{N} - 1$ smallest eigenvalues of $\hat{F}^{(m)}_u$ and $t(\hat{F}_u, e^{(m+1)})$ has the following form:
\begin{align}
& t(\hat{F}_u, e^{(m+1)}) \\
& = g(\hat{F}^{(m)}_u) + \langle \hat{\nabla} g(\hat{F}^{(m)}_u), \hat{F}_u - \hat{F}^{(m)}_u \rangle - f(\hat{F}_u) + w e^{(m+1)} . \tag{44}
\end{align}
Algorithm 1: Proposed Iterative Algorithm for Rank-constrained D.C. Problem.

Input:
Effective channel: $g_1, g_2, \ldots, g_N$;
Random initialization of $F_u^{(0)}, u = 1, 2, \ldots, N$;
Initial information: $s, s', t(F_u^{(0)}, e(0))$;
Initialize $U^{(0)}, e(0)$ by solving Equation (42);
Stop criterion: $\epsilon_1, \epsilon_2$.

Procedures: Procedures:
1: while $|(s' - s)/s| \leq \epsilon_1$ do
2: Update $s': s' = (m + 1)\epsilon_2$
3: for $0 \leq u \leq N$ do
4: $m = 0$
5: while $|t(F_u^{(m+1)}, e(m+1)) - t(F_u^{(m)}, e(m))| \leq \epsilon_2$
6: Obtain the optimal value $t(F_u^{(m+1)}, e(m+1))$ of objective function and $F_u^{(m+1)}, e(m+1)$ in Equation (43);
7: Update $U^{(m)}$ from $F_u^{(m+1)}$ via EVD;
8: Update $m = m + 1$
9: end while
10: end for
11: Update $s': s' = t(m+1)$
12: end while
13: Outputs: $F^* = [F_1^{(m)}, F_2^{(m)}, \ldots, F_N^{(m)}]$. 

$\mathcal{P}_0$ is indeed a standard convex optimization problem that can be solved via available convex software packages, such as CVX [25]. $C_1$ ensures the total power constraint and $C_2$ guarantees the QoS constraint for each UE. Furthermore, $C_3, C_4$ and $C_5$ are the constraints followed from Equation (41) in which we set $r = 1$. The proposed iterative algorithm is summarized in Algorithm 1. In each iteration, we solve a D.C. programming problem and update the optimal $F_u^{(m)}$. If the WSR $t(F_u^{(m+1)})$ can no longer be improved, the digital precoder $F_u^{(m+1)}$ of the $(u + 1)$-th UE is then optimized successively in the same manner.

Finally, we analyze the computational complexity of the proposed rank-constrained D.C. programming algorithm. In each iteration, an SDP problem is derived as shown in $\mathcal{P}_0$. The existing SDP solver based on the interior point method has computational complexity of the order of $O(N^4_k)$ as $F_u \in C^N_k \times N_k$ [39]. Furthermore, the EVD conducted in each iteration has computational complexity of the order of $O(N^2_k)$. As a result, the complexity of the proposed algorithm is of the order of $O(N^3_k)$ in each iteration.

D. Convergence Analysis

In the following, we provide the convergence analysis of the proposed iterative algorithm for solving the rank-constrained D.C. programming problem.

As the function $g(F_u)$ is concave, its gradient $\nabla g(F_u)$ is super-gradient. Therefore, we have
\[
g(F_u) \leq g(F_u^{(m)}) + \langle \nabla g(F_u^{(m)}), F_u - F_u^{(m)} \rangle. \tag{45}
\]
Since $e^{(m+1)} \leq e^{(m)}$ and $F_u^{(m)}$ is feasible to Equation (43), it follows:
\[
g(F_u^{(m+1)}) - f(F_u^{(m+1)}) + e^{(m+1)} \leq g(F_u^{(m)}) + \langle \nabla g(F_u^{(m)}), F_u - F_u^{(m)} \rangle - f(F_u) + e^{(m)}, \tag{46}
\]

which shows that the solution $F_u^{(m+1)}$ is always better than or equal to the previous solution $F_u^{(m)}$. Thus, the algorithm must converge to a critical point.

VI. PERFORMANCE ANALYSIS

In this section, we will derive the performance upper and lower bounds of the proposed scheme. We will first begin with the analog-only scheme. The expected capacity is given by
\[
E[R_{k,u}] = E[\log_2 (1 + \rho)],
\]
\[
= E \left[ \log_2 \left( 1 + \frac{|w_{k,u}^H H_{k,u} v_{k,n}^2|}{1/\gamma + \mathcal{I}_{k,u}} \right) \right], \tag{47}
\]
where $\mathcal{I}_{k,u}$ is the received interference represented as
\[
\mathcal{I}_{k,u} = \sum_{n \in N, n \neq u} A_{k,n} |w_{k,u}^H H_{k,u} v_{k,n}^2|. \tag{48}
\]

From Equations (10) and (13), the optimal array response vectors in transmitter and receiver are denoted by $[15]$
\[
w_{k,u}^* = aR(\phi_{k,u}^*, \theta_{k,u}^*) \tag{49},
\]
\[
v_{k,n}^* = aT(\phi_{k,u}^*, \theta_{k,u}^*). \tag{50}
\]

**Proposition 1:** If the optimal analog beamformers are designed as $w_{k,u}^* = aR(\phi_{k,u}^*, \theta_{k,u}^*)$ and $v_{k,n}^* = aT(\phi_{k,u}^*, \theta_{k,u}^*)$, respectively, the following inequality holds:
\[
\sum_{n \in N, n \neq u} A_{k,n} |w_{k,u}^H H_{k,u} v_{k,n}^2| \leq Z(N_k - 1)\Xi. \tag{51}
\]
where $\Xi$ is a constant and $Z = D_{k,u}^2 \alpha_{k,u}^2$.

**Proof:** See Appendix C.

Recalling Equation (47), the expected downlink capacity for each UE can be rewritten as
\[
E[R_{k,u}] = E \left[ \log_2 \left( 1 + \frac{Z}{1/\gamma + Y} \right) \right], \tag{52}
\]
where
\[
Z = |w_{k,u}^H H_{k,u} v_{k,n}^2|, = D_{k,u}^2 \alpha_{k,u}^2. \tag{53}
\]

From Equation (49), we have
\[
Y = \sum_{n \in N, n \neq u} A_{k,n} |w_{k,u}^H H_{k,u} v_{k,n}^2|^2
\]
\[
= D_{k,u}^2 \alpha_{k,u}^2 \sum_{n \in N, n \neq u} A_{k,n} |v_{k,n}^H v_{k,n}^2|^2
\]
\[
= Z \sum_{n \in N, n \neq u} A_{k,n} |v_{k,n}^H v_{k,n}^2|. \tag{54}
\]
Proposition 2: The lower and upper bounds of the sum-rate capacity can be given by
\[
\int_0^\infty \log_2(1+x) dF_X(x) \leq \mathbb{E}[R_{k,u}] \leq \int_0^\infty \log_2(1+z) dF_Z(z),
\]
where \( F_X(x) \) and \( F_Z(z) \) are the CDF of SINR and \( Z \), respectively.

Proof: From Proposition 1, \( Y \) can be upper bounded by
\[
Y \leq Z(N_k - 1)\Xi,
\]
where \( 0 \leq \Xi \leq 1 \) with \( \Xi \) being the expected residual interference power between distinct beams. In our proposed system, the beams will be selected and grouped to reduce the residual interference. Clearly, \( \Xi = 0 \) if the number of antennas goes to infinity or the steering vectors of different users are strictly orthogonal. In contrast, \( \Xi = 1 \) if different UEs have same AoDs. In general, the value of \( \Xi \) can be numerically derived between 0 and 1.

Capitalizing on the Extreme Value Theory [40], we can derive the cumulative distribution function (CDF) of \( Z \) as
\[
F_Z(z) = 1 - e^{-\frac{z}{C}},
\]
where \( C = 2N_TN_R/L_{k,u} \). By setting \( \Xi = 0 \), i.e., the interference is totally removed, we have
\[
\mathbb{E}[R_{k,u}] \leq \int_0^\infty \log_2(1+z) dF_Z(z).
\]
Finally, the CDF of the SINR lower bound can be given by
\[
F_X(x) = 1 - e^{-\frac{C}{\gamma(1-(N_k - 1)\Xi)}}.
\]
The detailed derivation can be found in Appendix D.

Using the CDF above, the lower bound of the sum-rate capacity can be derived as
\[
\mathbb{E}[R_{k,u}] \geq \int_0^\infty \log_2(1+x) dF_X(x).
\]

It is analytically intractable to obtain a closed-form solution to Equation (58). We will show the numerical results in simulation section.

VII. SIMULATION RESULTS

In this section, we will use computer simulations to confirm the performance of the proposed UE-SBS association and rank-constrained D.C. programming algorithms. In the simulations, we consider a system of multiple SBS’s each equipped with a \( 4 \times 4 \) UPA (i.e., \( N_T = 16 \)) and \( N \) ground UEs each equipped with a \( 2 \times 2 \) UPA (i.e. \( N_R = 4 \)). The number of paths in the wireless channel model is set to \( L_u = 4 \). We model the azimuth AoAs/AoDs uniformly distributed over \([0, 2\pi]\) while the elevation AoAs/AoDs uniformly distributed over \([-\pi/2, \pi/2]\), respectively. For each computer experiment, we compute the average over 100 realizations.

In Fig. 6, we evaluate the total amount of residual inter-user interference that all UE’s suffer as a function of the number of groups. Specifically, we compare the performance of the distributed local-search algorithm proposed in Section IV against the greedy selection algorithm proposed in [41]. Inspection of Fig. 6 suggests that the intra-SBS interference can be suppressed by maximizing the residual inter-SBS interference.

Fig. 7 shows the total amount of residual interference as a function of the total number of users for two SBS’s. The upper and lower bounds derived in Section VI are also shown in the figure. As derived in Theorem 1, the gap between
the performance upper bound and the solution obtained by the proposed algorithm is on the order of $1 - \frac{1}{K} = 0.5$, where $K = 2$ is the number of SBS’s in this simulation. The upper bound $I_{\text{max}}$ is obtained by exhaustive search.

Fig. 8 shows the WSR performance comparison as a function of SNR. In this experiment, we simulate 12 UEs served by one SBS. The curve labeled as “CCP-SVD” corresponds to the algorithm reported in [38] while the curves labeled as “ZF-GloPowAll” and “ZF-UnifPowAll” represent the zero-forcing scheme with global power allocation proposed in [26] and with uniform power allocation scheme in [15], respectively. Inspection of Fig. 8 reveals that the proposed algorithm has the best performance as compared to the conventional algorithms. More specifically, the proposed algorithm leverages a rank-constrained D.C. algorithm to jointly optimize precoding and power allocation. In contrast to the implementation of SVD in “CCP-SVD,” an iterative algorithm is proposed to satisfy the rank constraint for digital precoding design while optimizing the power allocation. The dotted and dashed curves are the upper bound and lower bound derived in Section VI, respectively.

Next, we examine the QoS for each UE by setting $\lambda_u = 1$ bps for $u = 1, \ldots, 12$. Fig. 9 depicts the CDF of the average UE data rate. It is evident that all UEs served by the proposed QoS-aware power allocation algorithms satisfy the minimum QoS requirement (i.e. 1 bps/Hz). Defining the outage as the average UE data rate being below the minimum required data rate, “Uniform Power HB” and “CCP-SVD” suffer from an outage rate of about 20% and 58%, respectively.

Fig. 10 shows the convergence behaviors of the proposed iterative algorithm at different SNR values ranging from $-10$ dB to 15 dB. Fig. 10 suggests that the proposed algorithm converges monotonically within 80 iterations at all SNR values.

Finally, the comparison of MkC and greedy selection for user association is presented in Fig. 11 using $N = 24$ UEs served by $K = 2, 3$ SBS’s. Fig. 11 shows that the proposed MkC scheme outperforms the greedy selection for both $K = 2, 3$. Furthermore, the WSR performance improves with the increment of $K$ since more frequency bands are introduced.
Fig. 12. The value of $\Xi$ for different number of transmitter antennas.

VIII. CONCLUSION

In this paper, we have investigated an mmWave MEC system in which each user has a QoS requirement. Taking into account the hardware cost, we have proposed hybrid precoding techniques to reduce the number of RF chains as compared to the conventional fully digital precoding approaches. To suppress the intra-SBS interference, an M$k$C approach has been developed to solve the user association problem. Moreover, we have proposed a D.C. programming-based iterative algorithm to jointly optimize the digital precoder and power allocation for QoS-aware WSR maximization for hybrid precoding systems. To solve this coupled non-convex problem, we have proposed to cast the optimization problem as a rank-constrained D.C. programming problem before an iterative algorithm is devised by combining the conventional D.C. programming problem with RCOP. Simulation results have shown that the proposed schemes can effectively improve WSR while satisfying the QoS of each UE.

APPENDIX A

PROOF FOR THEOREM 1

Proof: Assuming the maximal value of M$k$C is $I_{\text{max}}$. Let $N_1, N_2,\ldots,N_K$ be an arbitrary solution obtained while the value is $I_{ls}$ using the local-search algorithm. We can show straightforwardly

$$I_{ls} = \sum_{k=1}^{K} \sum_{u=1}^{N_k} \Omega_{k_u,k}.$$  \hfill (61)

where $I_{ls}$ is the solution given by the local-search algorithm.

By summing both sides of the inequalities (27) over all $u \in N_k, q \in N_j$, we have

$$N_k N_j \sum_{u=1}^{N_k} \Omega_{k_u,k} + N_k N_j \sum_{q=1}^{N_j} \Omega_{j_q,j} \leq N_k N_j \sum_{q=1}^{N_j} \Omega_{j_q,j}.$$ \hfill (62)

Setting

$$\sum_{u=1}^{N_k} \Omega_{k_u,k} = \Gamma_{kk}, \quad \sum_{q=1}^{N_j} \Omega_{j_q,j} = \Gamma_{jj},$$

$$\sum_{q=1}^{N_j} \Omega_{j_q,j} = \Gamma_{kq}, \quad \sum_{u=1}^{N_k} \Omega_{k_u,j} = \Gamma_{ju},$$

we have

$$\Gamma_{kk} + \Gamma_{jj} \leq \Gamma_{kq} + \Gamma_{ju}. \quad \hfill (64)$$

Summing both sides of Equation (64) for over all $k, j = 1, 2, \ldots, K$, for $k \neq j$, we have

$$\sum_{k=1}^{K} \sum_{j=1, j \neq k}^{K} (\Gamma_{kk} + \Gamma_{jj}) \leq \sum_{k=1}^{K} \sum_{j=1, j \neq k}^{K} (\Gamma_{kq} + \Gamma_{ju}). \quad \hfill (65)$$

The cut of this solution is then given by:

$$\sum_{k=1}^{K} \sum_{j=1, j \neq k}^{K} (\Gamma_{kq} + \Gamma_{ju}) = 2I_{ls}. \quad \hfill (66)$$

The above inequality can be rewritten as

$$\sum_{k=1}^{K} \sum_{j=1, j \neq k}^{K} (\Gamma_{kk} + \Gamma_{jj}) \leq \sum_{k=1}^{K} \sum_{j=1, j \neq k}^{K} \Gamma_{jj} \quad (K - 1) \sum_{k=1}^{K} \Gamma_{kk} + \Gamma_{ls} \quad (K - 1) \sum_{k=1}^{K} \Gamma_{jj}$$

$$= (K - 1) \sum_{k=1}^{K} \Gamma_{kk}.$$ \hfill (67)

Recalling Equation (66), we have

$$\sum_{k=1}^{K} \Gamma_{kk} \leq I_{ls}. \quad \hfill (68)$$

Since the optimal solution can contain all edges, we obtain

$$I_{\text{max}} \leq (K - 1) \sum_{k=1}^{K} \Gamma_{kk} + I_{ls} \leq I_{ls} + \frac{I_{ls}}{K - 1}. \quad \hfill (69)$$

or equivalently

$$\frac{I_{ls}}{I_{\text{max}}} \geq 1 - \frac{1}{K}. \quad \hfill (70)$$

Thus, the theorem is proved.

APPENDIX B

PROOF FOR EQUATION (38)

Proof: Suppose $f(p)$ is a concave function on a convex neighborhood $C$ and differentiable at $p$. Then, for every $y \in C$,
we have the following inequality based on the definition of concavity:
\[ f((1 - \lambda)p + \lambda y), \]
\[ = f(p + \lambda(y - p)), \]
\[ \geq f(p) + \lambda(f(y) - f(p)), \quad \text{(71)} \]
where \(0 < \lambda \leq 1\).

Rearranging the terms and dividing both sides by \(\lambda\), we have
\[ \frac{f(p + \lambda(y - p)) - f(p)}{\lambda} \geq f(y) - f(p). \quad \text{(72)} \]

Letting \(\lambda \to 0\), it can be shown that the left hand side of Inequality (72) converges to \(f'(p) \cdot (y - p)\). Finally, we have Equation (38) as follows:
\[ f(p) + f'(p) \cdot (y - p) \geq f(y). \quad \text{(73)} \]

**APPENDIX C**

**PROOF FOR PROPOSITION 1**

*Proof:* Recalling Equation (52), the SINR can be represented as
\[ X = \frac{Z}{\gamma + Y}. \quad \text{(74)} \]

To derive the CDF of SINR, we have to first calculate the CDF of \(Y\), which can be represented as
\[ Y = Z \mathbb{E} \left[ \sum_{i=1, i \neq u}^{N_k} |\mathbf{v}_k^H \mathbf{v}_{k,i}^*|^2 \right], \]
\[ \approx Z(N_k - 1) \mathbb{E} \left[ |\mathbf{v}_k^H \mathbf{v}_{k,i}^*|^2 \right], \quad \text{(75a)} \]
\[ \leq Z(N_k - 1) \Xi \quad \text{(75b)} \]

where we have assumed that the variance of \(|\mathbf{v}_k^H \mathbf{v}_{k,i}^*|^2\) for \(\forall i \in N_k\) is small by recalling the Equation (13) in Equation (75a). Equation (75b) is the average interference over all i.i.d UEs. Finally, the inequality of Equation (75b) comes from the effectiveness of association algorithm.

As the AoDs and AoAs are i.i.d for UPA antennas, the expectation can be calculated as
\[ \mathbb{E} \left[ |\mathbf{v}_k^H \mathbf{v}_{k,i}^*|^2 \right] \triangleq \Xi \]
\[ = \frac{1}{4\pi^2 N_T^2} \int_0^{2\pi} \int_0^{2\pi} |B(\theta_{k,u}, \theta_{k,i}, \phi_{k,u}, \phi_{k,i})|^2 d\theta_{k,u} d\phi_{k,i}, \quad \text{(76)} \]

where \(B(\cdot)\) is given by
\[ B(\theta_{k,u}, \theta_{k,i}, \phi_{k,u}, \phi_{k,i}) \]
\[ = \sum_{q=0}^{Q-1} \sum_{p=0}^{P-1} \exp \left[ jkd \left[ p (\sin \phi_{k,u} \sin \theta_{k,u} - \sin \phi_{k,i} \sin \theta_{k,i}) \right] + q (\cos \theta_{k,u} - \cos \theta_{k,i}) \right] , \]
\[ = \frac{(1 - e^{jkd}(\sin \phi_{k,u} \sin \theta_{k,u} - \sin \phi_{k,i} \sin \theta_{k,i}))^P}{1 - e^{jkd}(\sin \phi_{k,u} \sin \theta_{k,u} - \sin \phi_{k,i} \sin \theta_{k,i})} \times \frac{(1 - e^{jkd}(\cos \theta_{k,u} - \cos \theta_{k,i}))^Q}{1 - e^{jkd}(\cos \theta_{k,u} - \cos \theta_{k,i})}. \quad \text{(77)} \]

As shown in Fig. 12, the value of \(\Xi\) can be numerically estimated. As the number of antenna increases, the value of \(\Xi\) decreases gradually, which confirms Equation (13).

**APPENDIX D**

**DERIVATION FOR EQUATION (59)**

Since \(\alpha_{k,u,l}\) has the distribution of \(\chi^2(2)\), the CDF of \(Z\) can be derived in a straightforward manner [42]:
\[ F_Z(z) = 1 - e^{-\frac{z}{\gamma}}. \quad \text{(78)} \]

For a given CDF of \(Z\) in Equation (57), the CDF of SINR can be computed as
\[ F_X(x) = P(X \leq x) \]
\[ = P \left( \frac{Z}{1/\gamma + Z(N - 1)\Xi} \leq x \right), \]
\[ = P \left( Z \leq \frac{x}{\gamma(1 - x(N - 1)\Xi)} \right), \]
\[ = 1 - e^{-\frac{x}{(1-x)(\gamma(N-1)\Xi)}}. \quad \text{(79)} \]

where we have \(\gamma > 0\) in the derivation above.

**REFERENCES**


Guanchong Niu received the bachelor’s degree in physics from Jilin University, Changchun, China, in 2014. He is currently working toward the Ph.D. degree with the Chinese University of Hong Kong, Shenzhen, China. Between January and March 2020, he was an Intern with the Bell Labs, France, developing indoor localization systems. His current research interests include the millimeter wave and multiple-input multiple-output systems for the next-generation wireless communications and robotics.

Qi Cao received the bachelor’s degree in electronic communication engineering from the University of Liverpool, Liverpool, U.K., in 2013, the M.S. degree in communications and signal processing from Imperial College London, London, U.K., in 2014, and the Ph.D. degree from the School of Information and Control Engineering, China University of Mining and Technology, Xuzhou, China, in 2018. In 2019, he joined the Chinese University of Hong Kong, Shenzhen, China, as Postdoctoral Research Associate. His current research interests include MIMO wireless communications, machine learning, and artificial intelligent-driven network optimization.

Man-On Pun received the B.Eng. degree in electronic engineering from the Chinese University of Hong Kong (CUHK), Shenzhen, China, in 1996, the M.Eng. degree in computer science from the University of Tsukuba, Tsukuba, Japan, in 1999, and the Ph.D. degree in electrical engineering from the University of Southern California, Los Angeles, Los Angeles, CA, U.S.A in 2006. From 2006 to 2008, he was a Postdoctoral Research Associate with Princeton University, Princeton, NJ, USA.

He is currently an Associate Professor with the School of Science and Engineering, CUHKSZ. Prior to joining CUHKSZ in 2015, he held research positions with Huawei, USA, Mitsubishi Electric Research Labs, Boston, MA, USA and Sony, Tokyo, Japan. His research interests include AI Internet of Things and applications of machine learning in communications and satellite remote sensing.

Prof. Pun was the recipient of the Best Paper awards from IEEE VTC’06 Fall, IEEE ICC’08, and IEEE Infocom’09. From 2010 to 2014, he was an Associate Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS. He is the founding Chair of the IEEE Joint SPS-ComSoc Chapter, Shenzhen, China.