

Energy Efficient Hybrid Precoding in Heterogeneous Networks with Limited Wireless Backhaul Capacity

Zheng Chu¹, Wanming Hao¹, Pei Xiao¹, Fuhui Zhou², De Mi¹, Zhengyu Zhu³, and Victor C.M. Leung⁴

¹ Institute for Communication Systems (ICS), University of Surrey, Guildford, GU2 7XH, UK

² Department of Electrical and Computer Engineering, Utah State University, Logan, UT, USA

³ School of Information Engineering, Zhengzhou University, Zhengzhou, China

⁴ Department of Electrical and Computer Engineering, The University of British Columbia, Vancouver, Canada
(Email: zheng.chu@surrey.ac.uk)

Abstract—This paper investigates a two-tier heterogeneous networks (HetNets) with wireless backhaul, where millimeter wave (mmWave) frequency is adopted at the macro base station (MBS), and the cellular frequency is considered at small cell BS (SCS) with orthogonal frequency division multiple access (OFDMA). Subarray structure based hybrid analog/digital precoding scheme is investigated to reduce the hardware cost and energy consumption. Our goal is to maximize the energy efficiency (EE) of the HetNets with limited wireless backhaul capacity and all users' quality of service (QoS) constraints. The formulated problem is non-convex mixed integer nonlinear fraction programming (MINLFP), which is non-trivial to solve directly. In order to circumvent this issue, we propose a two-loop iterative resource allocation algorithm. Specifically, we transform the outer-loop problem into a difference of convex programming (DCP) by employing integer relaxation and Dinkelbach method. In addition, the first-order approximation is considered to linearize this inner-loop DCP problem into a convex optimization framework. Lagrange dual method is adapted to achieve the optimal closed-form power allocation. Furthermore, we analyze the convergence of the proposed iterative algorithm. Numerical results are presented to demonstrate our proposed schemes.

I. INTRODUCTION

With the explosive growth of the high-data-rate multimedia wireless services in fifth generation (5G) networks, the massive multiple-input-multi-output (mMIMO) heterogeneous network (HetNets) has been considered as a promising system architecture, especially in the case that ultra-dense small cells (SCs) overlay the macro cell (MC) [1]. The MC base station (MBS) is equipped with a large number of antennas to support high-mobility macro users (MUs) and manage resource allocation, whereas the SC BSs (SBSs) with a few number of antennas to serve the SC users (SUs) with low-mobility. Millimeter-Wave (mmWave) i.e., from 30 GHz to 300 GHz, has been regarded as a promising technique to exploit extra available spectrum to enhance the whole network throughput due to its bandwidth resources [2], [3].

Although the mmWave transmissions suffer from the severe propagation loss, the short wavelength allows more antenna elements to be packed into the same physical space. It compensates the poor propagation channel via narrow and high-gain beams provided by the large antennas arrays. Therefore, mmWave frequency has been widely applied at the large-scale antennas MBS, e.g., [2], [4]. However, it is not a trivial task to realize the large-scale antennas configuration at mmWave MBS. One of the challenging issues comes from that each antenna usually requires one dedicated radio-frequency (RF) chain (i.e., high-resolution digital-to-analog converter

(DAC), up-converter, etc). With the increase of the antennas at the MBS, it is more challenge to configure more and more RF chains in the limited physical space. Besides, it brings higher hardware cost and energy consumption [5]. To circumvent this issue, various low-complexity RF techniques in mmWave mMIMO system have been proposed (i.e., hybrid analog/digital precoding [6]). Hybrid analog/digital precoding structure consists of a low-dimension baseband digital precoder (requiring a small number of RF chains) and a high-dimension analog beamformer (including a large number of analog phase shifters) [6], [7]–[8]. Based on connectivity, it can be categorized into two classes, i.e., fully connected and subarray. The fully connected structure has been investigated in [6] and [7], where single user and multi-user scenario cases are exploited. For subarray structure, [9] proposed a successive interference cancellation (SIC)-based hybrid precoding and obtain the near-optimal performance. However, existing literature mainly focuses on system throughput maximization, and there is few work focusing on the energy efficiency (EE) in mmWave mMIMO HetNet. Although [8], [10] investigated the EE maximization problem, the EE in mmWave mMIMO HetNets has not yet been studied. The equal power allocation is assumed and the quality of service (QoS) requirements of users are also not imposed in [8].

Recently, backhaul solution has been considered as one of the main challenges in two-tier mmWave mMIMO HetNets. Although wired backhaul link can provide high data rates, it is impractical to have wired connections between all SBSs and MBS due to the high implementation cost [11]. Therefore, wireless backhaul link has been regarded as a suitable and cost-effective approach [12]–[15]. In [12], the downlink (DL) cell association and bandwidth allocation are jointly investigated to maximize the sum logarithmic user rate with limited wireless backhaul. In [13], energy harvesting SBS-based user association and power allocation with the self-backhaul is investigated to maximize the network EE. In [14], the mmWave frequency is used for wireless backhaul of SCs, forming a multi-hop backhaul network, where user association is investigated to jointly maximize the network throughput and EE. In [15], the energy consumption minimization was investigated with mmWave wireless backhaul.

Unlike other work in the literature, this paper investigates the EE maximization problem in a two-tier mmWave mMIMO HetNet with multiple frequency bands. Particularly, the MBS, with large-scale antenna arrays working in mmWave frequency, transmits downlink (DL) signal to MUs and backhaul

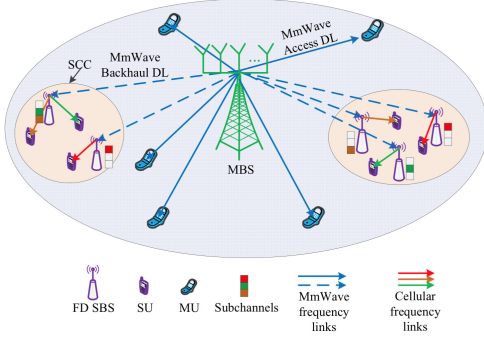


Fig. 1: The system model for the two-tier mmWave HetNets with wireless backhaul.

signal to SBSs simultaneously. The SC cluster (SCC)-based single-antenna SBSs operate in traditional cellular frequency with orthogonal frequency division multiple access (OFDMA) technology and provide the service for SUs. To the best of our knowledge, the EE resource allocation under the above system configuration has not been investigated in the existing works. *The contributions of this paper are summarized as follows:*

We consider subarray structure based hybrid precoding in the two-tier mmWave HetNets. First, the beamsteering-codebook searching algorithm is proposed for analog beamforming. Then, the zero-forcing (ZF) digital precoding is adopted to cancel interference between MUs and SBSs. We jointly design power and subchannel allocation (i.e., power allocation for MUs and SBSs at MBS, SUs at SBSs, and subchannel allocation for SUs in each SCC), to maximize the EE of the two-tier mmWave HetNets. Meanwhile, the power constraints for MBS and each SBS are considered, and the QoS requirements of each MU and SU are given. The formulated problem is a non-convex mixed-integer non-linear fractional programming (MINLFP) problem, which is challenging to solve directly. To circumvent this issue, we propose a two-loop iterative resource allocation algorithm. Particularly, the integer relaxation and Dinkelbach method are introduced to transform the original problem into a difference of convex programming (DCP), which is still intractable. To proceed, we propose a constrained concave convex procedure (CCCP)-based inner iterative algorithm to transform it into a convex optimization problem. On this basis, Dinkelbach method based outer iterative algorithm is applied to obtain the solution. Furthermore, the convergence of the proposed iterative algorithms is proved. Finally, numerical results highlight the advantage of the subarray structure based hybrid precoding in terms of EE performance with low energy consumption.

II. SYSTEM DESCRIPTION

In this section, we investigate a two-tier HetNet consisting of one MBS and L SCCs, where each SCC includes multiple SCs, as shown in Fig. 1.¹ The MBS is equipped with N_{TX} antennas to serve K MUs, and all SBSs receive the SUs' data (i.e., backhaul data) from MBS. Similar to [2], [9],

¹Each SCC can be regarded as a hot spot. Besides, a SCC can be formed by various neighbouring SCs to cancel the inter-cell interference. In this paper, we assume that multiple SCCs have been determined, where the detailed SC clustering scheme is discussed in [16].

mmWave frequency (i.e., 73 GHz) is used at MBS with W Hz bandwidth. It is assumed that the l -th SCC consists of M_l single-antenna SBSs and K_l single-antenna SUs. To circumvent the inter-tier interference, all SBSs encode their received data using OFDMA technology at cellular frequency (i.e., 2GHz) with total B Hz bandwidth, which is equally divided into N orthogonal subchannels. Thus, there is no intra-cluster interference because SUs are served by different subchannels. The inter-cluster interference is small enough which can be omitted due to the large distance among different SCCs and N subchannels are reused at all SCCs. Moreover, we employ full-duplex (FD) mode in backhaul links, where FD mode is equipped at SBSs. Specifically, SBSs can receive data from MBS in mmWave frequency and transmit data to SUs with cellular frequency simultaneously [2]. Let $\mathcal{L} = \{1, \dots, L\}$ and $\mathcal{K} = \{1, \dots, K\}$ denote the sets of SCCs and MUs, respectively, and $\mathcal{M}_l = \{1, \dots, M_l\}$ and $\mathcal{K}_l = \{1, \dots, K_l\}$ denote the sets of SBSs and SUs in the l -th SCC, respectively.

A. Hybrid Analog/Digital Precoding Model

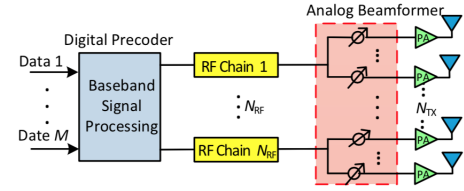


Fig. 2: Subarray structure diagram

In this paper, we assume that the MBS is equipped with N_{RF} ($N_{RF} \leq N_{TX}$) RF chains to reduce hardware cost and energy consumption, as shown in Fig. 2. Since the MBS simultaneously transmits MUs' data and SUs' data to MUs and SBSs, respectively, the number of RF chains should be greater than or equal to the total number of MUs and SUs, i.e., $N_{RF} \geq \sum_{l=1}^L M_l + K$. Herein, we consider the subarray structure, where each RF chain is only connected to a disjoint subset of antennas through phase shifters. In this paper, we assume that the number of antennas in all antenna subarrays is the same, denoted as $N_{SUB} = N_{TX}/N_{RF}$.² Accordingly, the received signal at the k -th MU can be expressed as:

$$\begin{aligned}
 y_{0,k} &= \sum_{i=1}^K \sqrt{P_{0,i}} \mathbf{h}_{0,k} \mathbf{F} \mathbf{v}_{0,i} x_{0,i} + \sum_{l=1}^L \sum_{j=1}^{M_l} \sqrt{P_{l,j}} \mathbf{h}_{0,k} \mathbf{F} \mathbf{v}_{l,j} x_{l,j} + n_{0,k} \\
 &= \underbrace{\sqrt{P_{0,k}} \mathbf{h}_{0,k} \mathbf{F} \mathbf{v}_{0,k} x_{0,k}}_{\text{Desired signal}} + \underbrace{\sum_{i \neq k} \sqrt{P_{0,i}} \mathbf{h}_{0,k} \mathbf{F} \mathbf{v}_{0,i} x_{0,i}}_{\text{Multi-MU interference}} + \underbrace{\sum_{l=1}^L \sum_{j=1}^{M_l} \sqrt{P_{l,j}} \mathbf{h}_{0,k} \mathbf{F} \mathbf{v}_{l,j} x_{l,j}}_{\text{SBSs' signals interference}} \\
 &\quad + \underbrace{n_{0,k}}_{\text{Noise}},
 \end{aligned}$$

where $P_{0,k}$ and $P_{l,j}$ denote the transmit power for the k -th MU and the j -th SBS in the l -th SCC, respectively; $\mathbf{h}_{0,k} \in \mathbb{C}^{1 \times N_{TX}}$ denotes the channel vector from MBS to the k -th MU; $x_{0,k}$ and $x_{l,j}$ are the transmitted symbols of the k -th MU and the i -th SBS in the l -th SCC, respectively (i.e., $\mathbb{E}\{|x_{0,k}|\} = 1$ and $\mathbb{E}\{|x_{l,j}|\} = 1$); $\mathbf{v}_{0,k}$ and $\mathbf{v}_{l,j}$ denote the digital precoders for the k -th MU and the j -th SBS in the l -th SCC, respectively; $n_{0,k}$ is an independent and identically distributed (i.i.d.) additive

²Note that N_{SUB} should be an integer. When N_{SUB} is not an integer, i.e. the number of antennas in different subarrays may be different, this situation also applies to our work.

white Gaussian noise (AWGN) $CN(0, N_0)$; $\mathbf{F} \in \mathbb{C}^{N_{\text{TX}} \times N_{\text{RF}}}$ is analog precoding performed by the equal power splitters and phase shifters. For the subarray structure, \mathbf{F} has the form of a block diagonal matrix as $\mathbf{F} = \text{Diag}[\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_{N_{\text{RF}}}]$, where $\mathbf{f}_k \in \mathbb{C}^{N_{\text{SUB}} \times 1}$ denotes the beamforming vector associated with the k -th RF chain with $|\mathbf{f}_k| = 1/\sqrt{N_{\text{SUB}}}$ ($i = 1, \dots, N_{\text{SUB}}$).

The received signal of the j -th SBS in the l -th SCC can be expressed as:

$$\begin{aligned} y_{l,j} &= \sum_{i=1}^L \sum_{k=1}^{M_l} \sqrt{P_{i,k}} \mathbf{h}_{l,j} \mathbf{F} \mathbf{v}_{i,k} x_{i,k} + \sum_{k=1}^K \sqrt{P_{0,k}} \mathbf{h}_{l,j} \mathbf{F} \mathbf{v}_{0,k} x_{0,k} + n_{l,j} \\ &= \underbrace{\sqrt{P_{l,j}} \mathbf{h}_{l,j} \mathbf{F} \mathbf{v}_{l,j} x_{l,j}}_{\text{Desired signal}} + \underbrace{\sum_{k=1}^K \sqrt{P_{0,k}} \mathbf{h}_{l,j} \mathbf{F} \mathbf{v}_{0,k} x_{0,k}}_{\text{MUs' signals interference}} \\ &\quad + \underbrace{\sum_{k \neq j}^{M_l} \sqrt{P_{l,k}} \mathbf{h}_{l,j} \mathbf{F} \mathbf{v}_{l,k} x_{l,k} + \sum_{i \neq l}^L \sum_{k=1}^{M_i} \sqrt{P_{i,k}} \mathbf{h}_{l,j} \mathbf{F} \mathbf{v}_{i,k} x_{i,k}}_{\text{Multi-SBS interference}} + \underbrace{n_{l,j}}_{\text{Noise}}, \end{aligned} \quad (1)$$

where $\mathbf{h}_{l,j} \in \mathbb{C}^{1 \times N_{\text{TX}}}$ denotes the channel vector from MBS to the j -th SBS in the l -th SCC, and $n_{l,j}$ is an i.i.d. AWGN $CN(0, N_0)$.

B. OFDMA of FD SBS in SCC

Let $\mathcal{N} = \{1, \dots, N\}$ denote the set of orthogonal subchannels. We employ $c_{k,n}^{l,j}$ to indicate whether the k -th user is assigned with the n -th subchannel at the j -th SBS in the l -th SCC, i.e.,

$$c_{k,n}^{l,j} = \begin{cases} 1, & \text{yes,} \\ 0, & \text{otherwise,} \end{cases} \quad (2)$$

where $k \in \mathcal{K}_l$, $l \in \mathcal{L}$, $j \in \mathcal{M}_l$ and $n \in \mathcal{N}$. In each SCC, we assume that different SBSs can transmit data to one SU with different subchannels, and each subchannel $n \in \mathcal{N}$ is assigned to at most one SU to avoid interference. Thus, we have $\sum_{k=1}^{K_l} \sum_{j=1}^{M_l} c_{k,n}^{l,j} \leq 1, n \in \mathcal{N}, l \in \mathcal{L}$. To proceed, the received signal from the j -th SBS to the k -th SU at the n -th subchannel in the l -th SCC can be expressed as $y_{k,n}^{l,j} = c_{k,n}^{l,j} \sqrt{P_{k,n}^{l,j}} h_{k,n}^{l,j} x_{k,n}^{l,j} + n_{k,n}^{l,j}$, where $P_{k,n}^{l,j}$, $h_{k,n}^{l,j}$ and $x_{k,n}^{l,j}$ denote the transmit power, channel coefficients and transmit signal of the k -th SU in the l -th SCC at the n -th subchannel through the j -th SBS, respectively. $n_{k,n}^{l,j}$ is an i.i.d. AWGN $CN(0, N_0)$.

C. MmWave Channel Model

Due to the limited scattering in mmWave channel, we adopt a geometric channel model with G scatters, where each scatter is assumed to contribute a single propagation path between the MBS and MU (SBS). Thus, the channel vector $\mathbf{h}_{l,k}$ can be expressed as:

$$\mathbf{h}_{l,k} = \sqrt{\frac{N_{\text{TX}}}{G}} \sum_{g=1}^G \beta_{l,k}^g \mathbf{a}^H(\theta_{l,k}^g), \quad (3)$$

where $\beta_{l,k}^g$ denotes the complex gain of the g -th path between the MBS and the k -th MU. The path amplitudes are assumed to be Rayleigh distributed, i.e., $\beta_{l,k}^g \sim CN(0, \sigma_{l,k}^g)$, where $\sigma_{l,k}^g$ represents the average power of the g -th cluster for the k -th MU. $\theta_{l,k}^g \in [0, 2\pi]$ is the g -th path's azimuth angles of departure (AoDs) of the MBS for the k -th MU, and $\mathbf{a}(\theta_{l,k}^g)$ is the antenna array steering vector with respect to $\theta_{l,k}^g$. Here, we only consider the azimuth, but the extension to elevation and azimuth direction (i.e., 3-D beamforming) is possible. When $l \in \mathcal{L}$, $k \in \mathcal{K}_l$, $\mathbf{h}_{l,k}$ denotes the DL channel from MBS to the k -th SBS in the l -th SCC. For uniform linear

array (ULA) configuration in the subarray structure, RF chains are connected to different subsets of antennas, such that the antenna array steering vector $\mathbf{a}(\theta_{l,k}^g)$ consists multiple subarray steering vectors, which is given by

$$\mathbf{a}(\theta_{l,k}^g) = [\mathbf{a}_1(\theta_{l,k}^{g,1})^T, \dots, \mathbf{a}_{N_{\text{RF}}}(\theta_{l,k}^{g,N_{\text{RF}}})^T]^T, \quad (4)$$

where $\mathbf{a}_s(\theta_{l,k}^{g,s}) = \frac{1}{\sqrt{N_{\text{SUB}}}} [1, e^{j\frac{2\pi}{\lambda} d \sin(\theta_{l,k}^{g,s})}, \dots, e^{j\frac{2\pi}{\lambda} (N_{\text{SUB}}-1) d \sin(\theta_{l,k}^{g,s})}]^T$ ($s = \{1, \dots, N_{\text{RF}}\}$). Correspondingly, the channel $\mathbf{h}_{l,k}$ can be expressed as:

$$\mathbf{h}_{l,k} = [\mathbf{h}_{l,k}^1]^T, (\mathbf{h}_{l,k}^2)^T, \dots, (\mathbf{h}_{l,k}^{N_{\text{RF}}})^T]^T, \quad (5)$$

where $\mathbf{h}_{l,k}^s = \sqrt{\frac{N_{\text{SUB}}}{G}} \sum_{g=1}^G \beta_{l,k}^{g,s} \mathbf{a}_s^H(\theta_{l,k}^{g,s})$ ($s = \{1, \dots, N_{\text{RF}}\}$).

D. Power Consumption Model

The power consumption of the MBS consists of transmit power and circuit power consumption. The circuit power consumption mainly includes baseband, RF chains, phase shifters and power amplifiers (PAs) [5], [9]. For the subarray structure, there are N_{RF} RF chains, N_{TX} phases shifters and N_{TX} PAs, so the circuit power consumption can be written as

$$P_c^m = P_{\text{BB}} + N_{\text{RF}} P_{\text{RF}} + N_{\text{TX}} P_{\text{PS}} + N_{\text{TX}} P_{\text{PA}}. \quad (6)$$

Accordingly, the total power consumption at MBS is given by

$$P_{\text{MBS}} = P_c^m + \xi \left(\sum_{k=1}^K P_{0,k} + \sum_{l=1}^L \sum_{k=1}^{M_l} P_{l,k} \right), \quad (7)$$

where ξ is a constant which accounts for the inefficiency of the PA [17]. For the FD SBSs, the total power consumption is given by

$$P_{\text{SBS}} = \sum_{l=1}^L M_l P_c^s + \xi \left(\sum_{l=1}^L \sum_{j=1}^{M_l} \sum_{k=1}^{K_l} \sum_{n=1}^N c_{k,n}^{l,j} P_{k,n}^{l,j} \right), \quad (8)$$

where P_c^s is the circuit power consumption of SBS.

III. EE MAXIMIZATION

In this section, we design hybrid analog/digital precoding, optimizing power and subchannel allocations to maximize EE of the mmWave HetNets subject to the users' QoS and limited wireless backhaul capacity.

A. Hybrid Analog/Digital Precoding

In this subsection, we investigate the subarray structure based hybrid precoding, where RF chains are connected with different subsets of antennas. Also, the RF-chain based codebook can be searched based on subarray-based beamforming. We allocate the subarray antennas associated with one RF chain to one MU, until all RF chains are assigned. To guarantee fairness among all MUs and overall performance of system, we first obtain the maximal subarray channel gain of each MU from its N_{RF} subarray channel gains as $\{l, k, \|\mathbf{h}_{l,k}^s\|\} = \arg \max_{s=\{1, \dots, N_{\text{RF}}\}} \|\mathbf{h}_{l,k}^s\|$, $l \in \{0, \mathcal{L}\}$, $k \in \{\mathcal{K}, \mathcal{M}_l\}$. Next, we select one RF chain and MU pair that has the minimal subarray channel gain. Then, we remove the selected RF chain and MU pair and repeat the above procedure in remaining RF chains and MUs in each searching cycle.

To proceed, we denote the DL channel vector as $\tilde{\mathbf{h}}_{l,k} = \mathbf{h}_{l,k} \mathbf{F}$ ($l \in \{0, \mathcal{L}\}$, $k \in \{\mathcal{K}, \mathcal{M}_l\}$). For convenience, we define the whole DL channel matrix as $\tilde{\mathbf{H}} = [\tilde{\mathbf{h}}_{0,1}^T, \dots, \tilde{\mathbf{h}}_{0,K}^T, \tilde{\mathbf{h}}_{1,1}^T, \dots, \tilde{\mathbf{h}}_{1,M_1}^T, \dots, \tilde{\mathbf{h}}_{L,1}^T, \dots, \tilde{\mathbf{h}}_{L,M_L}^T]$. Then, we use the ZF method to eliminate the interference among MUs and SBSs as $\mathbf{V} = \tilde{\mathbf{H}}^H (\tilde{\mathbf{H}} \tilde{\mathbf{H}}^H)^{-1}$. As a result, the precoding vector of the k -th MU can be expressed as $\mathbf{v}_{0,k} = \frac{\mathbf{V}_k}{\|\mathbf{F} \mathbf{V}_k\|}$, $k \in \mathcal{K}$,

where \mathbf{V}_k denote the k -th column of \mathbf{V} . The precoding vector of the j -th SBS in the l -th SCC $\mathbf{v}_{l,j}$ can be expressed as $\mathbf{v}_{l,j} = \frac{\mathbf{V}_x}{\|\mathbf{FV}_x\|}$, $l \in \mathcal{L}$, $j \in \mathcal{M}_l$, where $x = K + j$ when $l = 1$, and $x = K + \sum_{l=1}^{l-1} M_l + j$ otherwise.

B. Problem Formulation

In this subsection, we aim to maximize the EE of the HetNets subject to the constraints of the QoSs for all users and the limited wireless backhaul capacity. By exploiting the hybrid precoding in Section III-A to eliminate the interference, the received signal of the k -th MU can be simplified as :

$$y_{0,k} = \sqrt{P_{0,k}} \mathbf{h}_{0,k} \mathbf{F} \mathbf{v}_{0,k} x_{0,k} + n_{0,k}, \quad (9)$$

and its achievable rate can be expressed as:

$$R_{0,k}^{\text{MU}}(\mathcal{P}_{\text{MU}}) = W \log_2 \left(1 + \frac{P_{0,k} |\mathbf{h}_{0,k} \mathbf{F} \mathbf{v}_{0,k}|^2}{WN_0} \right), \quad (10)$$

where \mathcal{P}_{MU} denotes the power allocation policy of MUs, i.e., $P_{0,k}$. Similarly, the received signal at the j -th SBS in the l -th SCC can be simplified as:

$$y_{l,j} = \sqrt{P_{l,j}} \mathbf{h}_{l,j} \mathbf{F} \mathbf{v}_{l,j} x_{l,j} + n_{l,j}, \quad (11)$$

and its achievable backhaul rate is given by

$$R_{l,j}^{\text{BH}}(\mathcal{P}_{\text{SBS}}) = W \log_2 \left(1 + \frac{P_{l,j} |\mathbf{h}_{l,j} \mathbf{F} \mathbf{v}_{l,j}|^2}{WN_0} \right), \quad (12)$$

where \mathcal{P}_{SBS} denotes the power allocation policy of SBSs, i.e., $P_{l,j}$. Since SUs are served by multiple SBSs with different subchannels on each SCC, the achievable rate at the k -th SU in the l -th SCC can be expressed as:

$$R_{l,k}^{\text{SU}}(\mathcal{P}_{\text{SU}}, C) = \sum_{j=1}^{M_l} \sum_{n=1}^N c_{k,n}^{l,j} B_0 \log_2 \left(1 + \frac{P_{k,n}^{l,j} |h_{k,n}^{l,j}|^2}{B_0 N_0} \right), \quad (13)$$

where $B_0 = B/N$ denotes the bandwidth of a subchannel. In addition, \mathcal{P}_{SU} and C denote power and subchannel allocation policies of SUs, i.e., $P_{k,n}^{l,j}$ and $c_{k,n}^{l,j}$, respectively. Similarly, the achievable rate provided by the j -th SBS in the l -th SCC can be formulated as:

$$R_{l,j}^{\text{SBS}}(\mathcal{P}_{\text{SU}}, C) = \sum_{k=1}^{K_l} \sum_{n=1}^N c_{k,n}^{l,j} B_0 \log_2 \left(1 + \frac{P_{k,n}^{l,j} |h_{k,n}^{l,j}|^2}{B_0 N_0} \right). \quad (14)$$

Then, the achievable EE of the system can be written as:

$$\eta_{EE}(\mathcal{P}_{\text{MU}}, \mathcal{P}_{\text{SBS}}, \mathcal{P}_{\text{SU}}, C) = \frac{R_{\text{M}}(\mathcal{P}_{\text{MU}}) + R_{\text{S}}(\mathcal{P}_{\text{SU}}, C)}{P_{\text{C}} + P_{\text{M}}(\mathcal{P}_{\text{MU}}) + P_{\text{B}}(\mathcal{P}_{\text{SBS}}) + P_{\text{S}}(\mathcal{P}_{\text{SU}}, C)},$$

where $R_{\text{M}}(\mathcal{P}_{\text{MU}}) = \sum_{k=1}^K R_{0,k}^{\text{MU}}(\mathcal{P}_{\text{MU}})$, $R_{\text{S}}(\mathcal{P}_{\text{SU}}, C) = \sum_{l=1}^L \sum_{k=1}^{K_l} R_{l,k}^{\text{SU}}(\mathcal{P}_{\text{SU}}, C)$, $P_{\text{C}} = P_{\text{c}}^m + \sum_{l=1}^L M_l P_{\text{c}}^s$, $P_{\text{M}}(\mathcal{P}_{\text{MU}}) = \xi \sum_{k=1}^K P_{0,k}$, $P_{\text{B}}(\mathcal{P}_{\text{SBS}}) = \xi \sum_{l=1}^L \sum_{j=1}^{M_l} P_{l,j}$, $P_{\text{S}}(\mathcal{P}_{\text{SU}}, C) = \xi \sum_{l=1}^L \sum_{j=1}^{M_l} \sum_{k=1}^{K_l} \sum_{n=1}^N c_{k,n}^{l,j} P_{k,n}^{l,j}$. Thus, the EE maximization problem can be formulated as

$$\max_{\{\mathcal{P}_{\text{MU}}, \mathcal{P}_{\text{SBS}}, \mathcal{P}_{\text{SU}}, C\}} \eta_{EE}(\mathcal{P}_{\text{MU}}, \mathcal{P}_{\text{SBS}}, \mathcal{P}_{\text{SU}}, C) \quad (15a)$$

$$\text{s.t. } R_{0,k}^{\text{MU}}(\mathcal{P}_{\text{MU}}) \geq R_{\min}, k \in \mathcal{K}, \quad (15b)$$

$$R_{l,k}^{\text{SU}}(\mathcal{P}_{\text{SU}}, C) \geq R_{\min}, l \in \mathcal{L}, k \in \mathcal{K}_l, \quad (15c)$$

$$R_{l,j}^{\text{BH}}(\mathcal{P}_{\text{SBS}}) \geq \tilde{R}_{l,j}^{\text{SBS}}(\mathcal{P}_{\text{SU}}, C), l \in \mathcal{L}, j \in \mathcal{M}_l, \quad (15d)$$

$$\sum_{k=0}^K P_{0,k} + \sum_{l=1}^L \sum_{j=1}^{M_l} P_{l,j} \leq P_{\max}^m, \quad (15e)$$

$$\sum_{k=1}^{K_l} \sum_{n=1}^N c_{k,n}^{l,j} P_{k,n}^{l,j} \leq P_{\max}^s, l \in \mathcal{L}, j \in \mathcal{M}_l, \quad (15f)$$

$$c_{k,n}^{l,j} \in \{0, 1\}, n \in \mathcal{N}, l \in \mathcal{L}, j \in \mathcal{M}_l, k \in \mathcal{K}_l, \quad (15g)$$

$$\sum_{j=1}^{M_l} \sum_{k=1}^{K_l} c_{k,n}^{l,j} \leq 1, n \in \mathcal{N}, l \in \mathcal{L}, \quad (15h)$$

where (15b) and (15c) impose the QoS requirements for MUs and SUs, respectively. (15d) ensures that the received backhaul rate of the SBS is no less than its achievable rate. (15e) and (15f) represent the maximum transmit power constraints for MBS and SBSs, respectively. In the above problem, we assume the same QoS requirements for all MUs and SUs, i.e., R_{\min} , which can be easily extended to the case of different QoS requirements.

C. Solution to Problem (15)

Problem (15) is a MINLFP but nonconvex, due to the fractional objective function (15a), binary subchannel indicator variable $c_{k,n}^{l,j}$ and non-convex constraint (15d). To circumvent this issue, we first transform the non-convex MINLFP problem into a DCP problem, and then further simplify the formulated DCP problem into a convex optimization problem by first-order approximation. A two-loop iterative algorithm is proposed to obtain the optimal power allocation of problem (15).

1) *Relaxation of Binary Variable*: First, we relax the binary variable $c_{k,n}^{l,j}$ to be a continuous value in the interval $[0, 1]$. It is assumed that one subchannel is only assigned to one SU-SBS pair, but the relaxation of the binary constraints means a time-sharing subchannel allocation among SUs and SBSs. Naturally, the original problem is not actually solved after relaxation. However, it has been verified when the number of available subchannels goes to infinity, the solution of the relaxed problem approaches the solution of the original problem [18]. We redefine the transmit power for the SU as $\tilde{P}_{k,n}^{l,j} = c_{k,n}^{l,j} P_{k,n}^{l,j}$, such that (13) and (14) are given by, respectively.

$$\tilde{R}_{l,k}^{\text{SU}}(\tilde{\mathcal{P}}_{\text{SU}}, C) = \sum_{j=1}^{M_l} \sum_{n=1}^N c_{k,n}^{l,j} B_0 \log_2 \left(1 + \frac{\tilde{P}_{k,n}^{l,j} |h_{k,n}^{l,j}|^2}{c_{k,n}^{l,j} B_0 N_0} \right), \quad (16a)$$

$$\tilde{R}_{l,j}^{\text{SBS}}(\tilde{\mathcal{P}}_{\text{SU}}, C) = \sum_{k=1}^{K_l} \sum_{n=1}^N c_{k,n}^{l,j} B_0 \log_2 \left(1 + \frac{\tilde{P}_{k,n}^{l,j} |h_{k,n}^{l,j}|^2}{c_{k,n}^{l,j} B_0 N_0} \right). \quad (16b)$$

Theorem 1: $\tilde{R}_{l,k}^{\text{SU}}(\tilde{\mathcal{P}}_{\text{SU}}, C)$ is jointly concave w.r.t. $\tilde{\mathcal{P}}_{\text{SU}}$ and C , while $\tilde{R}_{l,j}^{\text{SBS}}(\tilde{\mathcal{P}}_{\text{SU}}, C)$ is jointly concave w.r.t. $\tilde{\mathcal{P}}_{\text{SU}}$ and C .

Proof: Please refer to [19]. ■

Thus, problem (15) is reformulated as

$$\max_{\{\mathcal{P}_{\text{MU}}, \mathcal{P}_{\text{SBS}}, \tilde{\mathcal{P}}_{\text{SU}}, C\}} \frac{R_{\text{M}}(\mathcal{P}_{\text{MU}}) + \tilde{R}_{\text{S}}(\tilde{\mathcal{P}}_{\text{SU}}, C)}{P_{\text{C}} + P_{\text{M}}(\mathcal{P}_{\text{MU}}) + P_{\text{B}}(\mathcal{P}_{\text{SBS}}) + \tilde{P}_{\text{S}}(\tilde{\mathcal{P}}_{\text{SU}})} \quad (17a)$$

$$\text{s.t. } (15b), (15e), (15h), \quad (17b)$$

$$\tilde{R}_{l,k}^{\text{SU}}(\tilde{\mathcal{P}}_{\text{SU}}, C) \geq R_{\min}, l \in \mathcal{L}, k \in \mathcal{K}_l, \quad (17c)$$

$$R_{l,j}^{\text{BH}}(\mathcal{P}_{\text{SBS}}) \geq \tilde{R}_{l,j}^{\text{SBS}}(\tilde{\mathcal{P}}_{\text{SU}}, C), l \in \mathcal{L}, j \in \mathcal{M}_l, \quad (17d)$$

$$\sum_{k=1}^{K_l} \sum_{n=1}^N \tilde{P}_{k,n}^{l,j} \leq P_{\max}^s, l \in \mathcal{L}, j \in \mathcal{M}_l, \quad (17e)$$

$$c_{k,n}^{l,j} \in \{0, 1\}, n \in \mathcal{N}, l \in \mathcal{L}, j \in \mathcal{M}_l, k \in \mathcal{K}_l, \quad (17f)$$

where $\tilde{R}_{\text{S}}(\tilde{\mathcal{P}}_{\text{SU}}, C) = \sum_{l=1}^L \sum_{k=1}^{K_l} \tilde{R}_{l,k}^{\text{SU}}(\tilde{\mathcal{P}}_{\text{SU}}, C)$, $\tilde{P}_{\text{S}}(\tilde{\mathcal{P}}_{\text{SU}}) = \xi \sum_{l=1}^L \sum_{j=1}^{M_l} \sum_{k=1}^{K_l} \sum_{n=1}^N \tilde{P}_{k,n}^{l,j}$.

2) *Transformation to Objective Function* (17a): Next, we define the maximum EE q^* of the problem (17) as:

$$q^* = \frac{R_{\text{M}}(\mathcal{P}_{\text{MU}}^*) + \tilde{R}_{\text{S}}(\tilde{\mathcal{P}}_{\text{SU}}^*, C^*)}{P_{\text{C}} + P_{\text{M}}(\mathcal{P}_{\text{MU}}^*) + P_{\text{B}}(\mathcal{P}_{\text{SBS}}^*) + \tilde{P}_{\text{S}}(\tilde{\mathcal{P}}_{\text{SU}}^*)} \quad (18)$$

$$= \max_{\{\mathcal{P}_{\text{MU}}, \mathcal{P}_{\text{SBS}}, \tilde{\mathcal{P}}_{\text{SU}}, C\}} \frac{R_{\text{M}}(\mathcal{P}_{\text{MU}}) + \tilde{R}_{\text{S}}(\tilde{\mathcal{P}}_{\text{SU}}, C)}{P_{\text{C}} + P_{\text{M}}(\mathcal{P}_{\text{MU}}) + P_{\text{B}}(\mathcal{P}_{\text{SBS}}) + \tilde{P}_{\text{S}}(\tilde{\mathcal{P}}_{\text{SU}})},$$

where $\{\mathcal{P}_{\text{MU}}^*, \mathcal{P}_{\text{SBS}}^*, \tilde{\mathcal{P}}_{\text{SU}}^*, C^*\}$ is defined as the set of the optimal solution to problem (17).

Theorem 2: The maximum EE q^* is achieved if and only if

$$\max_{\{\mathcal{P}_{\text{MU}}, \mathcal{P}_{\text{SBS}}, \tilde{\mathcal{P}}_{\text{SU}}, C\}} R_{\text{M}}(\mathcal{P}_{\text{MU}}) + \tilde{R}_{\text{S}}(\tilde{\mathcal{P}}_{\text{SU}}, C) - q^* (P_{\text{C}} + P_{\text{M}}(\mathcal{P}_{\text{MU}}) + P_{\text{B}}(\mathcal{P}_{\text{SBS}}) + \tilde{P}_{\text{S}}(\tilde{\mathcal{P}}_{\text{SU}})) = R_{\text{M}}(\mathcal{P}_{\text{MU}}^*) + \tilde{R}_{\text{S}}(\tilde{\mathcal{P}}_{\text{SU}}^*, C^*)$$

$$- q^* (P_{\text{C}} + P_{\text{M}}(\mathcal{P}_{\text{MU}}^*) + P_{\text{B}}(\mathcal{P}_{\text{SBS}}^*) + \tilde{P}_{\text{S}}(\tilde{\mathcal{P}}_{\text{SU}}^*)) = 0.$$

Proof: Refer to [17], [20]. ■

To proceed, we propose a two-loop iterative algorithm to solve problem (17). Specifically, Dinkelbach method is employed to tackle the outer loop, and its convergence has been proved in [20]. While, with each iteration, the inner loop problem can be reformulated as

$$\max_{\{\Omega\}} R_{\text{M}}(\mathcal{P}_{\text{MU}}) + \tilde{R}_{\text{S}}(\tilde{\mathcal{P}}_{\text{SU}}, C) - q (P_{\text{C}} + P_{\text{M}}(\mathcal{P}_{\text{MU}}) + P_{\text{B}}(\mathcal{P}_{\text{SBS}}) + \tilde{P}_{\text{S}}(\tilde{\mathcal{P}}_{\text{SU}})) \quad (17b), (17c), (17e), (17f), \Omega = \{\mathcal{P}_{\text{MU}}, \mathcal{P}_{\text{SBS}}, \tilde{\mathcal{P}}_{\text{SU}}, C\}, \quad (20a)$$

$$R_{l,j}^{\text{BH}}(\mathcal{P}_{\text{SBS}}) - \tilde{R}_{l,j}^{\text{SBS}}(\tilde{\mathcal{P}}_{\text{SU}}, C) \geq 0, l \in \mathcal{L}, j \in \mathcal{M}_l, \quad (20b)$$

where (17d) is equivalently written as (20b). It is observed that the objective function in (20) is concave and all constraints in (20a) are convex. It is observed that (20b) is the DC constraint. Thus, (20) is a DCP problem, where the CCCP will be employed to tackle this problem.

3) *Reformulation to (20b):* In this subsection, we aim to transform (20b) into convex constraint via the first-order Taylor approximation. We denote the current point of $\tilde{R}_{l,j}^{\text{SBS}}(\tilde{\mathcal{P}}_{\text{SU}}, C)$ in (20b) as $[\tilde{\mathcal{P}}_{\text{SU}}^*, C^*]^{(t)}$ at the t -th iteration. Thus, (20b) can be approximated as

$$R_{l,j}^{\text{BH}}(\mathcal{P}_{\text{SBS}}) - \tilde{R}_{l,j}^{\text{SBS}}([\tilde{\mathcal{P}}_{\text{SU}}^*, C^*]^{(t)}, [\tilde{\mathcal{P}}_{\text{SU}}, C]) = R_{l,j}^{\text{BH}}(\mathcal{P}_{\text{SBS}}) - \tilde{R}_{l,j}^{\text{SBS}}([\tilde{\mathcal{P}}_{\text{SU}}^*, C^*]^{(t)}) - \left(\frac{(\tilde{\mathcal{P}}_{\text{SU}} - \tilde{\mathcal{P}}_{\text{SU}}^*)}{\partial \tilde{\mathcal{P}}_{\text{SU}}} + \frac{(C - C^*)}{\partial C} \right) \tilde{R}_{l,j}^{\text{SBS}}([\tilde{\mathcal{P}}_{\text{SU}}^*, C^*]^{(t)}) \geq 0, l \in \mathcal{L}, j \in \mathcal{M}_l. \quad (21)$$

Consequently, (20) is transformed into the following convex optimization problem at the t -th iteration:

$$\max_{\{\mathcal{P}_{\text{MU}}, \mathcal{P}_{\text{SBS}}, \tilde{\mathcal{P}}_{\text{SU}}, C\}} R_{\text{M}}(\mathcal{P}_{\text{MU}}) + \tilde{R}_{\text{S}}(\tilde{\mathcal{P}}_{\text{SU}}, C) - q (P_{\text{C}} + P_{\text{M}}(\mathcal{P}_{\text{MU}}) + P_{\text{B}}(\mathcal{P}_{\text{SBS}}) + \tilde{P}_{\text{S}}(\tilde{\mathcal{P}}_{\text{SU}})) \quad (22a)$$

$$\text{s.t. (17b), (17c), (17e), (17f), (21).} \quad (22b)$$

Since (22) is a convex optimization problem, the duality gap is zero and solving its dual problem is equivalent to solving the original problem [21]. To this end, we first write the Lagrangian dual problem to (22) as

$$\min_{\lambda, \mu, \nu, \beta, \alpha, \delta \geq 0} \mathcal{G}(\lambda, \mu, \nu, \beta, \alpha, \delta) = \min_{\lambda, \mu, \nu, \beta, \alpha, \delta \geq 0} \max_{\mathcal{P}_{\text{MU}}, \mathcal{P}_{\text{SBS}}, \tilde{\mathcal{P}}_{\text{SU}}, C} \mathcal{L}_a, \quad (23)$$

where

$$\mathcal{L}_a(\lambda, \mu, \nu, \beta, \alpha, \delta, \mathcal{P}_{\text{MU}}, \mathcal{P}_{\text{SBS}}, \tilde{\mathcal{P}}_{\text{SU}}, C) = R_{\text{M}}(\mathcal{P}_{\text{MU}}) + \tilde{R}_{\text{S}}(\tilde{\mathcal{P}}_{\text{SU}}, C)$$

$$- q (P_{\text{C}} + P_{\text{M}}(\mathcal{P}_{\text{MU}}) + P_{\text{B}}(\mathcal{P}_{\text{SBS}}) + \tilde{P}_{\text{S}}(\tilde{\mathcal{P}}_{\text{SU}})) + \sum_{k=1}^K \lambda_k (R_{0,k}^{\text{MU}}(\mathcal{P}_{\text{MU}}) - R_{\min}) + \sum_{l=1}^L \sum_{k=1}^{K_l} \mu_{l,k} (\tilde{R}_{l,k}^{\text{SU}}(\tilde{\mathcal{P}}_{\text{SU}}, C) - R_{\min}) + \sum_{l=1}^L \sum_{j=1}^{M_l} \nu_{l,j} (R_{l,j}^{\text{BH}}(\mathcal{P}_{\text{SBS}}) - \tilde{R}_{l,j}^{\text{SBS}}([\tilde{\mathcal{P}}_{\text{SU}}^*, C^*]^{(t)}) - \left(\frac{(\tilde{\mathcal{P}}_{\text{SU}} - \tilde{\mathcal{P}}_{\text{SU}}^*)}{\partial \tilde{\mathcal{P}}_{\text{SU}}} + \frac{(C - C^*)}{\partial C} \right) \tilde{R}_{l,j}^{\text{SBS}}([\tilde{\mathcal{P}}_{\text{SU}}^*, C^*]^{(t)})) + \sum_{l=1}^L \sum_{j=1}^{M_l} \beta_{l,j} \left(P_{\max}^s - \sum_{k=1}^{K_l} \sum_{n=1}^N \tilde{P}_{k,n}^{l,j} \right) + \sum_{l=1}^L \sum_{n=1}^N \alpha_{l,n} \left(1 - \sum_{j=1}^{M_l} \sum_{k=1}^{K_l} c_{k,n}^{l,j} \right) + \delta \left(P_{\max}^m - \sum_{k=0}^K P_{0,k} - \sum_{l=1}^L \sum_{j=1}^{M_l} P_{l,j} \right), \quad (24)$$

$\{\lambda, \mu, \nu, \beta, \alpha\}$ with elements $\{\lambda_k \ (k \in \mathcal{K}), \mu_{l,k} \ (l \in \mathcal{L}, k \in \mathcal{K}_l), \nu_{l,j}, \beta_{l,j} \ (l \in \mathcal{L}, j \in \mathcal{K}_l), \alpha_{l,n} \ (l \in \mathcal{L}, n \in \mathcal{N})\}$ are the nonnegative dual multiplier vectors associated with the constraints (15b), (17c), (21), (17e), and (15h), respectively. Also, δ is the nonnegative dual multiplier associated with the

Algorithm 1: CCCP-Based Iterative Algorithm For Solving Problem (20)

- 1 **Initialize** the maximum number of iterations L_{\max}^1 , the maximum tolerate ε_1 , feasible point $[\tilde{\mathcal{P}}_{\text{SU}}^*, C^*]^{(0)}$, and the iteration index $t = 0$.
- 2 **repeat**{Inner Loop}
- 3 Compute $\tilde{R}_{l,j}^{\text{SBS}}([\tilde{\mathcal{P}}_{\text{SU}}^*, C^*]^{(t)}, [\tilde{\mathcal{P}}_{\text{SU}}, C])$.
- 4 Initialize Lagrange multipliers $\lambda, \mu, \nu, \beta, \alpha, \delta$.
- 5 **repeat**{Solve problem (36)}
- 6 Obtain $P_{0,k}^*, P_{l,j}^*$ with (26), $\tilde{P}_{k,n}^{l,j*}$ and $c_{k,n}^{l,j*}$ with steepest descent method [21].
- 7 Update $\lambda, \mu, \nu, \beta, \alpha, \delta$ via the subgradient method.
- 8 **until** Lagrange multipliers converge;
- 9 Compute $\text{Obj}^{(t)} = R_{\text{M}}(\mathcal{P}_{\text{MU}}^*) + \tilde{R}_{\text{S}}(\tilde{\mathcal{P}}_{\text{SU}}^*, C^*) - q (P_{\text{C}} + P_{\text{M}}(\mathcal{P}_{\text{MU}}^*) + P_{\text{B}}(\mathcal{P}_{\text{SBS}}^*) + \tilde{P}_{\text{S}}(\tilde{\mathcal{P}}_{\text{SU}}^*))$.
- 10 Update $t = t + 1$, $[\tilde{\mathcal{P}}_{\text{SU}}^*, C^*]^{(t)} = \{\tilde{P}_{k,n}^{l,j*}, c_{k,n}^{l,j*}\} (n \in \mathcal{N}, l \in \mathcal{L}, j \in \mathcal{M}_l, k \in \mathcal{K}_l)$.
- 11 **until** $|\text{Obj}^{(t+1)} - \text{Obj}^{(t)}| \leq \varepsilon_1$ or $t \geq L_{\max}^1$;

constraint (15e). To solve (23) to obtain the optimal power and subchannel allocations, we first fix the Lagrange multipliers and q , and then consider the following Karush-Kuhn-Tucker (KKT) conditions

$$\frac{\partial \mathcal{L}_a}{\partial P_{0,k}} = 0, \text{ and } \frac{\partial \mathcal{L}_a}{\partial P_{l,j}} = 0. \quad (25)$$

After a few of mathematical manipulations, the optimal power allocations for MUs and SBSs are given by

$$P_{0,k}^* = \left[\frac{(1 + \lambda_k)W}{(q\xi + \delta) \ln 2} - \frac{1}{\gamma_{0,k}} \right]^+, P_{l,j}^* = \left[\frac{\nu_{l,j}W}{(q\xi + \delta) \ln 2} - \frac{1}{\gamma_{l,j}} \right]^+, \quad (26)$$

where $\gamma_{0,k} = |\mathbf{h}_{0,k} \mathbf{F} \mathbf{v}_{0,k}|^2 / W N_0$ and $\gamma_{l,j} = |\mathbf{h}_{l,j} \mathbf{F} \mathbf{v}_{l,j}|^2 / W N_0$.

Then, the power allocation $\tilde{P}_{k,n}^{l,j*}$ for SUs and the subchannel allocation $c_{k,n}^{l,j*}$ can be obtained by steepest descent method [21]. To proceed, the subgradient method can be employed to iteratively update the dual variables at each iteration. Thus, the CCCP-based iterative algorithm (i.e., inner loop) is used to find the solution of problem (20), which is summarized in Algorithm 1. Meanwhile, its convergence is analyzed via the following theorem:

Theorem 3: Algorithm 1 converges to a stationary and local optimal point after finite iterations.

Proof: Please refer to Appendix. ■

Since the binary subchannel allocation indicator $c_{k,n}^{l,j}$ is relaxed with the interval between $[0, 1]$ in Algorithm 1, we need to recover it to a Boolean. We first compute the marginal benefit for each $c_{k,n}^{l,j}$ as, $Q_{k,n}^{l,j} = \partial \mathcal{L}_a / \partial c_{k,n}^{l,j} |_{\tilde{P}_{k,n}^{l,j*}, c_{k,n}^{l,j} = \tilde{P}_{k,n}^{l,j*}}$. Then, the indicator $c_{k,n}^{l,j}$ can be recovered to zero or one by:

$$c_{k^*, n}^{l,j*} = \begin{cases} 1, & \text{if } \{k^*, j^*\} = \arg \max_{k \in \mathcal{K}_l, j \in \mathcal{M}_l} Q_{k,n}^{l,j} \text{ and } Q_{k,n}^{l,j} \geq 0, \\ 0, & \text{otherwise.} \end{cases} \quad (27)$$

Finally, we resolve the problem (22) and obtain the power allocation according to the recovered $c_{k,n}^{l,j*}$.

IV. SIMULATION RESULTS

In this section, numerical results are provided to demonstrate the performance of our proposed schemes. We consider the two-tier HetNets model shown Fig. 1, where the radius

of the MC is set to 500 meter and the MBS is located at the center of the MC which is equipped with 300 transmit antennas (i.e., $N_{TX} = 300$). The MUs and SCCs are randomly distributed within the MC, while M_l SBSs and K_l SUs are randomly distributed at the l -th SCC within a radius of 150 meter. It is assumed that $M_l = K_l = 3$, and $K = 6$. Also, the minimum number of RF chains is assumed to be $L = 3$. The mmWave channel is centered at 73 GHz with a bandwidth of 200 MHz. The path loss is modelled as $69.7 + 24 \log_{10}(d_m)$ dB [22], where d_m denotes the distance (meter). We assume that there are $G = 8$ clusters in the mmWave channel, and the azimuth AoA is uniformly distributed over $[0, 2\pi]$ and $\sigma_{l,k}^g = 1$. In addition, the maximum transmit power of the MBS is 46 dBm. Following the 3 GPP LTE-A standard [23], the cellular frequency is 2 GHz with a bandwidth of 20 MHz, which is divided into $N = 128$ subchannels. Additionally, the path loss is given as $38 + 30 \log_{10}(d_m)$ dB [24], whereas the multiple channels for the cellular frequency are considered with an exponential delay profile with $N/4$ taps. The QoS requirements of MUs and SUs are 10 Mbits/s. The maximum transmit power of each SBS is set to 23 dBm, and the noise power spectral density is -174 dBm/Hz. In addition, we set $P_{BB} = 200$ mW, $P_{RF} = 300$ mW, $P_{PA} = 40$ mW, $P_{PS} = 20$ mW, and $P_c^s = 100$ mW, while inefficiency of the power amplifier ξ is set as $1/0.38$. In our simulations, we compare our proposed subarray structure based hybrid precoding scheme with the full connected structure based hybrid precoding scheme in [8] and the digital precoding.

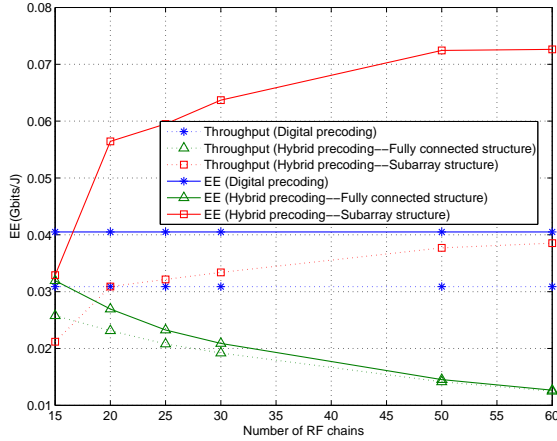


Fig. 3: EE versus number of RF chains.

First, we evaluate the the EE performance versus number of RF chains. In Fig. 3, “Throughput” denotes that the system EE performance is plotted with maximizing the throughput (i.e., $q=0$), whereas “EE” denotes the system EE is evaluated by solving the EE maximization problem. From this result, one can observe that the EE of the subarray structure based hybrid precoding increases with the number of RF chains, whereas the full connected based hybrid precoding scheme decreases as the number of RF chains increases. Also, the digital precoding scheme is not affected by the number of RF chains, as expected. In addition, the subarray structure based hybrid precoding scheme outperforms the digital precoding and full connected based hybrid precoding schemes in terms of EE and throughput performances, especially with a large number of RF chains. This is owing to a fact that a larger

number of RF chains leads to more energy consumption. Figs.

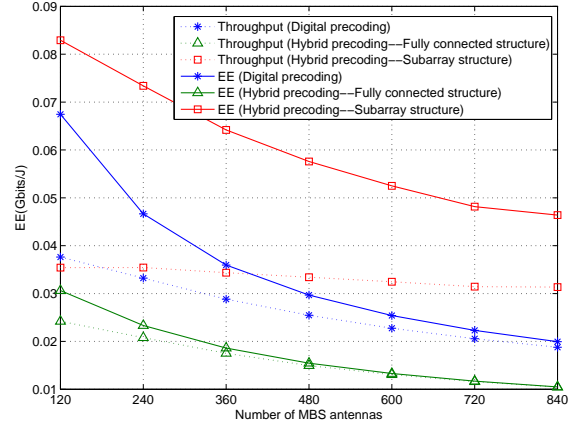


Fig. 4: EE versus number of MBS antennas.

4 shows that the EE performance versus the number of MBS antennas in the case that the number of RF chains is equal to 30. From this result, it is observed that the throughput and EE performances increase with the number of MBS antenna N_{TX} . This implies that more energy may be consumed with a larger number of antennas at the MBS. In addition, the subarray structure based hybrid scheme outperforms the digital precoding and full connected structure based schemes in terms of throughput and EE performances.

V. CONCLUSION

In this paper, we first designed hybrid precoding for MUs and SBSs at mmWave MBS with different structures. Then, the joint power and subchannel allocation problem was formulated to maximize the EE of the two-tier HetNets with limited wireless backhaul link. Due to non-convexity of the formulated MINLFP problem, we reformulated it into a DCP. In addition, a two-loop iterative algorithm was designed to obtain the power and subchannel allocation. Simulation results confirmed that the subarray hybrid precoding structure achieves the higher EE performance than the digital precoding and the fully connected based hybrid precoding schemes.

APPENDIX A THE PROOF OF Theorem 3

First, we analyze the convergence of Algorithm 1. For given initial feasible points $[\tilde{\mathcal{P}}_{SU}^*, C^*]^{(0)}$, according to Algorithm 1, we can obtain the feasible points $\{\mathcal{P}_{MU}, \mathcal{P}_{SBS}, \tilde{\mathcal{P}}_{SU}, C\}^{(t)}$ of convex optimization problem (22) at the t -th iteration. For convenience, we define the concave objective function in (22) as $\mathcal{U}(\mathcal{P}_{MU}, \mathcal{P}_{SBS}, \tilde{\mathcal{P}}_{SU}, C) = R_M(\mathcal{P}_{MU}) + \tilde{R}_S(\tilde{\mathcal{P}}_{SU}, C) - q(P_C + P_M(\mathcal{P}_{MU}) + P_B(\mathcal{P}_{SBS}) + \tilde{P}_S(\tilde{\mathcal{P}}_{SU}))$. Thus, the sequence $\{\mathcal{U}(\{\mathcal{P}_{MU}, \mathcal{P}_{SBS}, \tilde{\mathcal{P}}_{SU}, C\}^{(t)})\}$ monotonically increases as iteration number t grows. Due to the limited transmit power, sequence $\{\mathcal{U}(\{\mathcal{P}_{MU}, \mathcal{P}_{SBS}, \tilde{\mathcal{P}}_{SU}, C\}^{(t)})\}$ is upper bounded and convergent. Since objective function is strictly concave, the upper-bounded point of sequence $\{\mathcal{U}(\{\mathcal{P}_{MU}, \mathcal{P}_{SBS}, \tilde{\mathcal{P}}_{SU}, C\}^{(t)})\}$ is unique.

Next, we analyze the stability of Algorithm 1. We assume that $\{\mathcal{P}_{MU}, \mathcal{P}_{SBS}, \tilde{\mathcal{P}}_{SU}, C\}^\ddagger$ is the limit point of sequence

$\{\mathcal{P}_{\text{MU}}, \mathcal{P}_{\text{SBS}}, \tilde{\mathcal{P}}_{\text{SU}}, C\}^{(t)}$, when the iteration number t goes to infinity, we have the following definition:

$$\{\mathcal{P}_{\text{MU}}, \mathcal{P}_{\text{SBS}}, \tilde{\mathcal{P}}_{\text{SU}}, C\}^{\ddagger} \triangleq \lim_{t \rightarrow \infty} \{\mathcal{P}_{\text{MU}}, \mathcal{P}_{\text{SBS}}, \tilde{\mathcal{P}}_{\text{SU}}, C\}^{(t)}. \quad (28)$$

According to the above definition, we know that the limit point $\{\mathcal{P}_{\text{MU}}, \mathcal{P}_{\text{SBS}}, \tilde{\mathcal{P}}_{\text{SU}}, C\}^{\ddagger}$ is the solution of the convex optimization problem (22). Therefore, (22) can be also written as follows:

$$\max_{\{\mathcal{P}_{\text{MU}}, \mathcal{P}_{\text{SBS}}, \tilde{\mathcal{P}}_{\text{SU}}, C\}} \mathcal{U}(\mathcal{P}_{\text{MU}}, \mathcal{P}_{\text{SBS}}, \tilde{\mathcal{P}}_{\text{SU}}, C) \quad (29a)$$

$$\text{s.t. } R_{l,j}^{\text{BH}}(\mathcal{P}_{\text{SBS}}) - \tilde{R}_{l,j}^{\text{SBS}}([\tilde{\mathcal{P}}_{\text{SU}}, C]^{\ddagger}, [\tilde{\mathcal{P}}_{\text{SU}}, C]) \geq 0, l \in \mathcal{L}, j \in \mathcal{M}_l, \quad (29b)$$

$$\{\mathcal{P}_{\text{MU}}, \mathcal{P}_{\text{SBS}}, \tilde{\mathcal{P}}_{\text{SU}}, C\} \in \mathcal{F}, \quad (29c)$$

where \mathcal{F} is the feasible set of constraints (17b), (17c), and (17e)-(17f). Then, we have the following equalities:

$$R_{l,j}^{\text{BH}}(\mathcal{P}_{\text{SBS}}^{\ddagger}) - \tilde{R}_{l,j}^{\text{SBS}}([\tilde{\mathcal{P}}_{\text{SU}}, C]^{\ddagger}) = R_{l,j}^{\text{BH}}(\mathcal{P}_{\text{SBS}}^{\ddagger}) - \tilde{R}_{l,j}^{\text{SBS}}([\tilde{\mathcal{P}}_{\text{SU}}, C]^{\ddagger}, [\tilde{\mathcal{P}}_{\text{SU}}, C]^{\ddagger}) = 0, l \in \mathcal{L}, j \in \mathcal{M}_l, \quad (30)$$

where (30) means that (29b) are active for limit point $\{\mathcal{P}_{\text{MU}}, \mathcal{P}_{\text{SBS}}, \tilde{\mathcal{P}}_{\text{SU}}, C\}^{\ddagger}$. Next, we prove (30) by contradiction. We assume $R_{l,j}^{\text{BH}}(\mathcal{P}_{\text{SBS}}^{\ddagger}) > \tilde{R}_{l,j}^{\text{SBS}}([\tilde{\mathcal{P}}_{\text{SU}}, C]^{\ddagger}, [\tilde{\mathcal{P}}_{\text{SU}}, C]^{\ddagger})$, it means that the MBS can reduce the transmit power for the j -th SBS in the l -th SCC while the DL sum rate $R_{\text{M}}(\mathcal{P}_{\text{MU}}) + \tilde{R}_{\text{S}}(\tilde{\mathcal{P}}_{\text{SU}}, C)$ does not change. In this case, the MBS can transmit the remaining power for other MUs or SBSs, which improves the DL sum rate. Since the term $q(P_{\text{C}} + P_{\text{M}}(\mathcal{P}_{\text{MU}}) + P_{\text{B}}(\mathcal{P}_{\text{SBS}}) + \tilde{P}_{\text{S}}(\tilde{\mathcal{P}}_{\text{SU}}))$ does not change and the term $R_{\text{M}}(\mathcal{P}_{\text{MU}}) + \tilde{R}_{\text{S}}(\tilde{\mathcal{P}}_{\text{SU}}, C)$ increases, $\mathcal{U}(\mathcal{P}_{\text{MU}}, \mathcal{P}_{\text{SBS}}, \tilde{\mathcal{P}}_{\text{SU}}, C)$ will increase. Consequently, limit point $\{\mathcal{P}_{\text{MU}}, \mathcal{P}_{\text{SBS}}, \tilde{\mathcal{P}}_{\text{SU}}, C\}^{\ddagger}$ does not achieve the optimal solution of problem (29), which is contradictory with the original assumption. Therefore, we can obtain (30) at the limit point $\{\mathcal{P}_{\text{MU}}, \mathcal{P}_{\text{SBS}}, \tilde{\mathcal{P}}_{\text{SU}}, C\}^{\ddagger}$.

According to the above analysis, no matter how to choose the initial point $[\tilde{\mathcal{P}}_{\text{SU}}, C^*]^{(0)}$, only if it is feasible, the final convergence point, i.e., limit point $\{\mathcal{P}_{\text{MU}}, \mathcal{P}_{\text{SBS}}, \tilde{\mathcal{P}}_{\text{SU}}, C\}^{\ddagger}$, can be obtained by solving problem (29). In other words, the limit point is a stationary point of problem (20).

Finally, we prove that the stationary point is local optimal. It is well known that a stationary point may be a saddle point, a local minimum or a local maximum. Therefore, we need to prove that all stationary points are local maximum for the DCP (20). Next, we prove it by contraction, namely we assume that the limit point \mathbb{P}^{\ddagger} ($\mathbb{P}^{\ddagger} = \{\mathcal{P}_{\text{MU}}, \mathcal{P}_{\text{SBS}}, \tilde{\mathcal{P}}_{\text{SU}}, C\}^{\ddagger}$) is a local minimum, and there will exist a constant $\varsigma > 0$ that satisfies $\|\mathbb{P}^{\ddagger} - \mathbb{P}\| \leq \varsigma$ and $\mathcal{U}(\mathbb{P}) \geq \mathcal{U}(\mathbb{P}^{\ddagger})$. Next, we define: $\vartheta \triangleq \frac{\varsigma}{\|\mathbb{P}\|} > 0$, and $\mathbb{P}^{\ddagger*} \triangleq (1-\vartheta)\mathbb{P}^{\ddagger}$. Then, $\mathbb{P}^{\ddagger*}$ is also a feasible point of the DCP (20), and $\|\mathbb{P}^{\ddagger*} - \mathbb{P}^{\ddagger}\| \leq \varsigma$. Following this, we have

$$\mathcal{U}(\mathbb{P}^{\ddagger*}) \geq \mathcal{U}(\mathbb{P}^{\ddagger}). \quad (31)$$

On the other hand, from the objective function in (20), we have $\mathcal{U}(\mathbb{P}^{\ddagger*}) \leq \mathcal{U}(\mathbb{P}^{\ddagger})$, which contradicts with (31). Consequently, our original assumption is invalid, and the limit point \mathbb{P}^{\ddagger} is not a local minimum. Therefore, all stationary points should be local maximum for the DCP (20), and the limit point \mathbb{P}^{\ddagger} is local optimal.

REFERENCES

- [1] J. G. Andrews, H. Claussen, M. Dohler, S. Rangan, and M. C. Reed, "Femtocells: Past, present, and future," *IEEE J. Sel. Areas Commun.*, vol. 30, pp. 497–508, April Apr. 2012.
- [2] Z. Gao, L. Dai, D. Mi, Z. Wang, M. A. Imran, and M. Z. Shaker, "MmWave massive-MIMO-based wireless backhaul for the 5G ultra-dense network," *IEEE Wireless Commun.*, vol. 22, no. 5, pp. 13–21, Oct. 2015.
- [3] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter-wave cellular wireless networks: Potentials and challenges," *Proceedings of the IEEE*, vol. 102, no. 3, pp. 366–385, Mar. 2014.
- [4] D. Zhang, Z. Zhou, C. Xu, Y. Zhang, J. Rodriguez, and T. Sato, "Capacity analysis of noma with mmwave massive mimo systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 7, pp. 1606–1618, Jul. 2017.
- [5] R. W. Heath, N. Gonzalez-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 436–453, Apr. 2016.
- [6] M. M. Mulu, P. Xiao, M. Khalily, K. Cumanan, L. Zhang, and R. Tafazolli, "Low-complexity and robust hybrid beamforming design for multi-antenna communication systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 3, pp. 1445–1459, Mar. 2018.
- [7] A. Alkhatieb, G. Leus, and R. W. Heath, "Limited feedback hybrid precoding for multi-user millimeter wave systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 11, pp. 6481–6494, Nov. 2015.
- [8] C. Lin and G. Y. Li, "Energy-efficient design of indoor mmwave and Sub-THz systems with antenna arrays," *IEEE Trans. Wireless Commun.*, vol. 15, no. 7, pp. 4660–4672, Jul. 2016.
- [9] X. Gao, L. Dai, S. Han, C. L. I, and R. W. Heath, "Energy-efficient hybrid analog and digital precoding for MmWave MIMO systems with large antenna arrays," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 998–1009, Apr. 2016.
- [10] W. Hao, M. Zeng, Z. Chu, and S. Yang, "Energy-efficient power allocation in millimeter wave massive MIMO with non-orthogonal multiple access," *IEEE Wireless Commun. Lett.*, vol. 6, no. 6, pp. 782–785, Dec. 2017.
- [11] U. Siddique, H. Tabassum, E. Hossain, and D. I. Kim, "Wireless backhauling of 5G small cells: challenges and solution approaches," *IEEE Wireless Commun.*, vol. 22, no. 5, pp. 22–31, Oct. 2015.
- [12] N. Wang, E. Hossain, and V. K. Bhargava, "Joint downlink cell association and bandwidth allocation for wireless backhauling in two-tier HetNets with large-scale antenna arrays," *IEEE Trans. Wireless Commun.*, vol. 15, no. 5, pp. 3251–3268, May 2016.
- [13] L. Chen, F. R. Yu, H. Ji, B. Rong, X. Li, and V. C. M. Leung, "Green full-duplex self-backhaul and energy harvesting small cell networks with massive MIMO," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 12, pp. 3709–3724, Dec. 2016.
- [14] A. Mesodiakaki, F. Adelantado, L. Alonso, M. D. Renzo, and C. Verikoukis, "Energy- and spectrum-efficient user association in millimeter-wave backhaul small-cell networks," *IEEE Trans. Veh. Technol.*, vol. 66, no. 2, pp. 1810–1821, Feb. 2017.
- [15] Y. Niu, C. Gao, Y. Li, L. Su, D. Jin, Y. Zhu, and D. O. Wu, "Energy-efficient scheduling for mmwave backhauling of small cells in heterogeneous cellular networks," *IEEE Trans. Veh. Technol.*, vol. 66, no. 3, pp. 2674–2687, Mar. 2017.
- [16] W. Hao, O. Muta, H. Gacanin, and H. Furukawa, "Dynamic small cell clustering and non-cooperative game-based precoding design for two-tier heterogeneous networks with massive MIMO," *IEEE Trans. Commun.*, vol. 66, no. 2, pp. 675–687, Feb. 2018.
- [17] D. W. K. Ng, E. S. Lo, and R. Schober, "Energy-efficient resource allocation in OFDMA systems with large numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 11, no. 9, pp. 3292–3304, Sept. 2012.
- [18] W. Yu and R. Lui, "Dual methods for nonconvex spectrum optimization of multicarrier systems," *IEEE Trans. Commun.*, vol. 54, pp. 1310–1322, July 2006.
- [19] Z. Chu, F. Zhou, Z. Zhu, M. Sun, and N. Al-Dhahir, "Energy beamforming design and user cooperation for wireless powered communication networks," *IEEE Wireless Commun. Lett.*, vol. 6, no. 6, pp. 750–753, Dec. 2017.
- [20] W. Dinkelbach, "On nonlinear fractional programming," *Management Science*, vol. 13, no. 7, pp. 492–498, 1967.
- [21] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge University Press, 2004.
- [22] T. S. Rappaport, G. R. MacCartney, M. K. Samimi, and S. Sun, "Wide-band millimeter-wave propagation measurements and channel models for future wireless communication system design," *IEEE Transactions on Communications*, vol. 63, pp. 3029–3056, Sept. 2015.
- [23] "Evolved universal terrestrial radio access (e-utra); physical channels and modulation (release 12), 3gpp std. 36.211," 2014.
- [24] "Evolved universal terrestrial radio access (e-utra); radio frequency (rf) requirements for lte pico node b (release 12), 3gpp std. 36.931," 2014.