Energy Efficient Resource Optimization in User-Centric UDNs with NOMA and Beamforming

Long Zhang, Guobin Zhang, Xiaofang Zhao, and Enchang Sun

***Abstract*—In this paper, we address the problem of energy efficient resource optimization for downlink transmission in user-centric ultra-dense networks enabled by wireless access via non-orthogonal multiple access and wireless backhaul via beamforming. Our objective is to maximize the system energy efficiency by optimizing user/access point scheduling, subchannel assignment, and power allocation jointly. The problem is formulated as a non-convex mixed-integer nonlinear programming problem which is NP-hard. We then transform it into a convex subproblem using the sum-of-ratios decoupling and the iterative successive convex approximation method. An overall algorithm is further developed to solve the subproblem iteratively. Simulation results show that the proposed algorithm has improved the system-wide energy efficiency significantly when compared to the benchmark scheme.**

I. INTRODUCTION

Recently, increased interest in emerging applications, e.g., extended reality, holographic display, tele-surgery, etc., has propelled the explosive growth in mobile data traffic. Such an 1000x traffic growth necessitates the configuration of ultra-dense networks (UDNs) to fulfill network capacity and spectral efficiency (SE) enhancement requirements for 5G and beyond [1]. Instead of relying on a macro base station (MBS) sending signals to users, UDNs deploy tens or hundreds more of small access points (APs) to provide wireless access service for users, which has potentials to enlarge cell coverage, improve spatial reuse of resources, enhance performance gains, etc.

Due to the overlapped coverage for users caused by dense deployment of APs, traditional cell-centric architecture poses extra challenges on network planning and design for UDNs. It is vital to transform network architecture from cell-centric to user-centric via the idea of "network serving user" and cell-free concept [2]. In user-centric UDNs, a user is simultaneously served by an AP group (APG) wherein the AP density is comparable to or even higher than the user density. Through the deconstruction of cellular structure, user-centric UDNs not only eliminate cell boundaries with entirely suppressed inter-cell interference, but also achieve dynamic APG configuration and flexible resource allocation in a user-centric manner.

Although user-centric UDNs bring about multi-Gigabit-per-second user experience and SE increases in access downlink, limited wireless resources lead to serious competitions among APs for massive access opportunities of users. Recently, non-orthogonal multiple access (NOMA) has been considered as an enabling technique due to its high SE, massive connectivity, high user fairness, and low latency [3]. Power-domain NOMA allows multiple signals multiplexed to transmit simultaneously on the same spectrum resource by differentiating the signals via power levels. User can use successive interference cancellation (SIC) to decode its own received signal and reduce the undesired interference effectively. On the other hand, for backhauling, it is uneconomical for every AP to be connected via fiber to core networks. An alternative is to use wireless backhauling that allows low-cost APs to employ wireless links to MBS for backhauling. Multiple-antenna technique has been recently proposed as a promising solution to obtain higher SE and powerful interference mitigation via beamforming. Given this scenario, integration of wireless access via NOMA and wireless backhaul via beamforming into user-centric UDNs is not only an extension of UDNs, but also a practical application incentive promoted to provide significant performance gains. However, such a coupling in user-centric UDNs raise important concerns about resource allocation and user scheduling, among which notably is energy efficiency (EE) balance.

Several recent works are devoted to energy efficient resource allocation in user-centric UDNs. In [4], Park *et al.* proposed a user-centric reverse association scheme for joint optimization of handover and power control to maximize the AP's EE. In [5], Zhang *et al.* developed a joint optimization framework of load-aware user association and power allocation in mmWave-based UDNs to maximize the system EE. Additionally, there are a few existing works that investigate resource optimization problem by incorporating either NOMA or beamforming into user-centric UDNs. Liu *et al.* [6] devised a resource optimization framework in NOMA-based user-centric UDNs with access and backhaul downlink to maximize the system EE. In [7], Qin *et al.* used matching theory to study the problem of resource allocation and user association under a unified NOMA framework in UDNs. In [8], Kwon and Park explored the joint problem of resource allocation, user association, and hybrid beamforming design in mmWave UDNs to maximize the weighted sum rate with limited feedback.

However, aforementioned research are mainly highlighted as (i) the impact of resource optimization on the EE balance for wireless access [4], [5], (ii) joint design of access and backhaul downlink using NOMA [6], (iii) uplink and downlink design for wireless access via NOMA [7], and (iv) design of both access and backhaul downlink through hybrid beamforming

---

L. Zhang is with the School of Information and Electrical Engineering, Hebei University of Engineering, Handan 056038, China. G. Zhang and X. Zhao are with the School of Electrical Engineering and Intelligentization, Dongguan University of Technology, Dongguan 523808, China. E. Sun is with the Beijing Advanced Innovation Center for Future Internet Technology, Beijing University of Technology, Beijing 100124, China, and also with the Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China.

This work was supported in part by NSFC (61901115), Hebei Program for Top-notch Young Talents (BJ2017037), and NSF of Hebei (F2019402206).

[8]. Few consider utilizing NOMA and beamforming simultaneously for resource allocation in user-centric UDNs. This research gap motivates us to pursue a solution for the problem of energy efficient resource optimization to maximize the system EE of downlink transmission integrating both access downlink via NOMA and backhaul downlink via beamforming. Main contributions of our work include:

- We develop a resource optimization framework in an energy-efficient manner for downlink user-centric UDNs with a close coupling of wireless access via NOMA and wireless backhaul via beamforming.
- We formulate the system EE maximization problem as an MINLP problem by jointly optimizing user/AP scheduling, subchannel assignment, and power allocation.
- We transform the problem into a standard convex problem via the relaxation of binary variables, the sum-of-ratios decoupling, and the successive convex approximation (SCA), and solve it via Lagrangian dual decomposition.

The rest of the paper is organized as follows. Section II introduces the system model and problem formulation. Section III proposes the problem transformation and algorithm design. Simulation results are presented in Section IV, followed by concluding remarks in Section V.

II. SYSTEM MODEL AND PROBLEM FORMULATION

A. System Overview

Consider the downlink of a user-centric UDN, where an MBS with an antenna array is located at the center with M APs, denoted by set $\mathcal{M}=\{1,2,\cdots,M\}$, densely deployed within the macrocell coverage of that MBS. The macrocell's radius is r. There also exist N users, denoted by set $\mathcal{N}=\{1,2,\cdots,N\}$, randomly distributed in the overlapping coverage area sharing the same spectrum resource with MBS and APs. The locations of APs follow independent homogeneous Poisson point processes (HPPPs) with density that is comparable to or even larger than user density. The bandwidth of spectrum resource is equally divided to K subchannels, denoted by set $\mathcal{K}=\{1,2,\cdots,K\}$. To avoid the interference between access and backhaul, subchannel set $\mathcal{K}$ is separated into $\mathcal{A}=\{1,2,\cdots,\delta\}$ for access and $\mathcal{B}=\{\delta+1,\delta+2,\cdots,K\}$ for backhaul. Moreover, densely distributed APs are grouped into F disjoint clusters based on spatial directions, denoted by set $\mathcal{F}=\{1,2,\cdots,F\}$. Thus, an AP can only provide wireless access exactly for one or more user(s) on a subset of $\mathcal{A}$ within the same cluster to avoid inter-cluster interference. In cluster f, user n can be simultaneously associated with at most M_f APs on one or more subchannel(s), for $M_f \ll M$ and $f \in \mathcal{F}$. As such, M_f APs in cluster f constitute a generalized APG, denoted by set $\mathcal{G}_f$, to serve user n by concurrently transmitting independent signals in a user-centric fashion, for $\mathcal{G}_f \subset \mathcal{M}$.

B. Communication Model

1) Access Downlink via NOMA: For access downlink, a user in each cluster can be simultaneously served by multiple APs via an assigned subchannel from $\mathcal{A}$ in a user-centric way. The power-domain NOMA is adopted for access downlink, which enables that multiple signals from APs in a cluster multiplex on the same subchannel at the same time. According to the NOMA principle, one user can receive from APs in the same cluster via multiple subchannels, and one subchannel can be assigned to multiple users. To represent the association status between user and AP, we introduce a binary variable a_{fmn}^k such that if user n on subchannel k associates with AP m in cluster f then $a_{fmn}^k=1$, otherwise $a_{fmn}^k=0$.

We assume that all the subchannels for access downlink follow a quasi-static block fading, where the channel gains remain to be constant within the time duration. As such, we denote the downlink channel coefficient from AP m in cluster f to user n on subchannel k as $h_{fmn}^k=g_{fmn}^k d_{fmn}^{-\vartheta_1}$, where g_{fmn}^k is the flat Rayleigh fading channel gain, d_{fmn} is the distance between AP m in cluster f and user n, and ϑ_1 is the path loss exponent. After receiving the superposed signals from M_f^k APs on subchannel k in $\mathcal{G}_f$, user n employs the SIC technique to decode its desired messages, for $0 \le M_f^k < M_f$. Let $H_{fmn}^k=\frac{\left|h_{fmn}^k\right|^2}{\sigma_{nk}^2}$ be the channel to noise ratio (CNR) of subchannel k from AP m in cluster f to user n, where σ_{nk}^2 is the noise variance at user n on subchannel k. Without loss of generality, the CNRs of the received signals at user n on subchannel k served by M_f^k APs on subchannel k in $\mathcal{G}_f$ are sorted as $H_{f1n}^k \le \cdots \le H_{fmn}^k \le \cdots \le H_{fM_f^k n}^k$. With the NOMA principle, the achievable rate (in bps/Hz) of user n on subchannel k served by AP m in $\mathcal{G}_f$ can be expressed as

$$R_{fmn}^k=\log_2\left(1+\frac{p_{fmn}^k H_{fmn}^k}{1+\sum_{j=m+1}^{M_f^k} p_{fjn}^k H_{fjn}^k}\right), \qquad (1)$$

where p_{fmn}^k is the transmit power of AP m in cluster f to user n on subchannel k.

2) Backhaul Downlink via Beamforming: For backhaul downlink, the MBS concurrently transmits independent signals to the APs in different clusters over the sharing subchannels. By exploiting multiple antennas at both the MBS and the APs, downlink beamforming is considered in wireless backhaul not only to increase the SE, but also to combat the inter-cluster and intra-cluster interference. To characterize the association status between MBS and AP, we introduce a binary variable b_{fm}^k such that if AP m in cluster f associates with the MBS using subchannel k then $b_{fm}^k=1$, otherwise $b_{fm}^k=0$.

Let Q be the number of transmit antennas in the MBS's antenna array. Denote ϕ_f^k as the number of APs on subchannel k in cluster f, for $0 \le \phi_f^k \ll M \le Q$. The downlink channel coefficient vector between MBS and AP m on subchannel k in cluster f is given by $\mathbf{h}_{fm}^k=\tilde{\mathbf{h}}_{fm}^k d_{fm}^{-\vartheta_2}$, where d_{fm} is the distance between MBS and AP m in cluster f, ϑ_2 is the path loss exponent, and $\tilde{\mathbf{h}}_{fm}^k$ is the small scale Rayleigh fading channel coefficient vector that is assumed to be complex Gaussian distributed with zero mean and unit variance matrix. For beamforming, let $\mathbf{w}_f^k=\left[\mathbf{w}_{f1}^k,\mathbf{w}_{f2}^k,\cdots,\mathbf{w}_{f\phi_f^k}^k\right]^{\mathrm{T}}$ be the beamforming vector for ϕ_f^k APs on subchannel k in cluster f. To simplify analysis, we consider that the number of transmit antennas for beamforming at MBS is equal to the number of APs on subchannel k in cluster f. As such, the received signal at AP m on subchannel k in cluster f is corrupted by three parts, i.e., intra-cluster interference, inter-cluster interference, and AWGN. For analytical simplicity, we employ the zero-forcing beamforming to eliminate the inter-cluster interference. Thus, the achievable rate (in bps/Hz) of AP m on subchannel k in cluster f can be obtained by

$$R_{fm}^{k} = \log_2\left(1+\frac{q_{fm}^{k}\left|\mathbf{h}_{fm}^{k}\mathbf{w}_{f}^{k}\right|^2}{\left|\mathbf{h}_{fm}^{k}\mathbf{w}_{f}^{k}\right|^2\sum_{j=1,j\neq m}^{\phi_f^k} q_{fj}^{k}+\sigma_{mk}^2}\right), \quad (2)$$

where q_{fm}^{k} is the transmit power of MBS to AP m on subchannel k in cluster f and σ_{mk}^2 is the noise variance at AP m on subchannel k.

C. Power Consumption Model

For access downlink, power consumption depends on the power consumed at users in receiving mode and at APs in transmission mode, respectively. Power consumption for user n in cluster f is written as $P_{fn} = P_{fn}^{\mathrm{R}} + \psi_{\mathrm{A}} P_{fn}^{\mathrm{D}}$, where P_{fn}^{R} is the constant circuit power consumption for received signal processing, P_{fn}^{D} is the dynamic circuit power consumption for signal decoding, and ψ_{A} is correlated with the number of APs in every APG on each subchannel. Besides, power consumption for AP m in cluster f sending signal to user n on subchannel k is determined by transmitter circuit power consumption P_{m}^{C} and transmit power p_{fmn}^{k}, i.e., $P_m = P_{m}^{\mathrm{C}} + p_{fmn}^{k}$. Let N_f be the number of users that are associated with APs in cluster f, for $0 \leq N_f \ll N$. Then the sum power consumption for access downlink is equal to

$$P_{\mathrm{A}} = \sum_{f=1}^{F}\sum_{m=1}^{M_f}\sum_{n=1}^{N_f}\sum_{k=1}^{\delta} a_{fmn}^{k}\left(P_{fn} + P_{m}^{\mathrm{C}} + p_{fmn}^{k}\right). \quad (3)$$

For backhaul downlink, power consumption is aimed at the power consumed at APs in receiving mode and at MBS in transmission mode. Power consumption for AP m in cluster f can be modeled as $P_{fm} = P_{fm}^{\mathrm{R}} + \psi_{\mathrm{B}} P_{fm}^{\mathrm{D}}$, where P_{fm}^{R} is the constant circuit power consumption for received signal processing, P_{fm}^{D} is the dynamic circuit power consumption for signal decoding, and ψ_{B} is correlated with the number of APs in every cluster on each subchannel. Moreover, power consumption of MBS mainly depends on transmit power q_{fm}^{k} to AP m on subchannel k in cluster f. Thus, the sum power consumption for backhaul downlink is expressed by

$$P_{\mathrm{B}} = \sum_{f=1}^{F}\sum_{m=1}^{M_f}\sum_{k=\delta+1}^{K} b_{fm}^{k}\left(P_{fm} + q_{fm}^{k}\right). \quad (4)$$

D. Problem Formulation

The energy efficient resource optimization problem for the downlink is to maximize the system EE metric via jointly optimizing user/AP scheduling, subchannel assignment, and power allocation. Combining the access downlink via NOMA and the backhaul downlink via beamforming, the actual overall achievable rate of system can be obtained as

$$R_{\mathrm{S}} = \sum_{f=1}^{F}\sum_{m=1}^{M_f}\sum_{n=1}^{N_f}\sum_{k=1}^{\delta} a_{fmn}^{k} R_{fmn}^{k}. \quad (5)$$

Hence, the system EE for downlink transmission is defined by $\xi_{\mathrm{EE}} = \frac{R_{\mathrm{S}}}{P_{\mathrm{A}}+P_{\mathrm{B}}}$ (in bit/Hz/Joule). Let $\mathbf{A} = \{a_{fmn}^{k}\}_{m,n,k=1}^{M_f,N_f,\delta}$, $\mathbf{B} = \{b_{fm}^{k}\}_{m=1,k=\delta+1}^{M_f,K}$, $\mathbf{P} = \{p_{fmn}^{k}\}_{m,n,k=1}^{M_f,N_f,\delta}$, and $\mathbf{Q} = \{q_{fm}^{k}\}_{m=1,k=\delta+1}^{M_f,K}$, for $f \in \mathcal{F}$. The optimization problem is then formulated as

$$\max_{\mathbf{A},\mathbf{B},\mathbf{P},\mathbf{Q}} \frac{R_{\mathrm{S}}}{P_{\mathrm{A}} + P_{\mathrm{B}}} \quad (6a)$$

$$\text{s.t.} \sum_{f=1}^{F}\sum_{m=1}^{M_f}\sum_{k=1}^{\delta} a_{fmn}^{k} R_{fmn}^{k} \geq R_{n}^{\min}, \forall n, \quad (6b)$$

$$\sum_{f=1}^{F}\sum_{k=\delta+1}^{K} b_{fm}^{k} R_{fm}^{k} \geq \sum_{f=1}^{F}\sum_{n=1}^{N_f}\sum_{k=1}^{\delta} a_{fmn}^{k} R_{fmn}^{k}, \forall m, \quad (6c)$$

$$\sum_{f=1}^{F}\sum_{m=1}^{M_f}\sum_{k=\delta+1}^{K} b_{fm}^{k} q_{fm}^{k} \leq P_{\max}, \forall f, \forall m, \forall k, \quad (6d)$$

$$\sum_{n=1}^{N_f}\sum_{k=1}^{\delta} a_{fmn}^{k} p_{fmn}^{k} \leq P_{m}^{\max}, \forall f, \forall m, \quad (6e)$$

$$a_{fmn}^{k}, b_{fm}^{k} \in \{0,1\}, \forall f, \forall m, \forall n, \forall k, \quad (6f)$$

where $R_{n}^{\min}$ is the minimum data rate of user n, $P_{\max}$ is the MBS's maximum power, and $P_{m}^{\max}$ is the maximum power of AP m. The user's minimum rate constraint is shown in (6b). (6c) dictates the achievable rate of backhaul should be larger than that of access. (6d) is the MBS's maximum power constraint. (6e) denotes that AP's power is restricted by its maximum limit. Lastly, (6f) is the binary constraints to imply the user/AP scheduling relations. Due to the existence of interference terms in (6a), nonlinear rate constraints in (6b) and (6c), and binary variables in (6f), problem (6) is a non-convex MINLP problem. Such kind of problem is NP-hard and is very difficult to solve directly for the UDN scenario with larger numbers of densely distributed users and APs.

III. PROPOSED APPROACH

A. Problem Transformation

Considering that binary variables can be interpreted as user association-dependent indicators for assigning subchannels, we relax binary variables a_{fmn}^{k} and b_{fm}^{k} to be continuous real variables within the range of $[0,1]$ based on the time-sharing relaxation idea. As such, the actual power of AP m in cluster f to user n on subchannel k is represented as $\widetilde{p}_{fmn}^{k} = a_{fmn}^{k} p_{fmn}^{k}$, the actual power of MBS to AP m on subchannel k in cluster f is given by $\widetilde{q}_{fm}^{k} = b_{fm}^{k} q_{fm}^{k}$. Thus, we have $\widetilde{\mathbf{P}} = \{\widetilde{p}_{fmn}^{k}\}_{m,n,k=1}^{M_f,N_f,\delta}$, and $\widetilde{\mathbf{Q}} = \{\widetilde{q}_{fm}^{k}\}_{m=1,k=\delta+1}^{M_f,K}$, for $f \in \mathcal{F}$. Then problem (6) can be reformulated as

$$\max_{\mathbf{A},\mathbf{B},\widetilde{\mathbf{P}},\widetilde{\mathbf{Q}}} \frac{\widetilde{R}_{\mathrm{S}}}{\widetilde{P}_{\mathrm{A}} + \widetilde{P}_{\mathrm{B}}} \quad (7a)$$

$$\text{s.t.} \sum_{f=1}^{F}\sum_{m=1}^{M_f}\sum_{k=1}^{\delta} a_{fmn}^{k} \widetilde{R}_{fmn}^{k} \geq R_{n}^{\min}, \forall n, \quad (7b)$$

$$\sum_{f=1}^{F}\sum_{k=\delta+1}^{K} b^k_{fm}\widetilde{R}^k_{fm} \geq \sum_{f=1}^{F}\sum_{n=1}^{N_f}\sum_{k=1}^{\delta} a^k_{fmn}\widetilde{R}^k_{fmn}, \forall m, \quad \text{(7c)}$$

$$\sum_{f=1}^{F}\sum_{m=1}^{M_f}\sum_{k=\delta+1}^{K} \widetilde{q}^k_{fm} \leq P_{\max}, \forall f, \forall m, \forall k, \quad \text{(7d)}$$

$$\sum_{n=1}^{N_f}\sum_{k=1}^{\delta} \widetilde{p}^k_{fmn} \leq P^{\max}_m, \forall f, \forall m, \quad \text{(7e)}$$

$$a^k_{fmn}, b^k_{fm} \in [0,1], \forall f, \forall m, \forall n, \forall k. \quad \text{(7f)}$$

B. Sum-of-Ratios Decoupling

After the relaxation of binary variables, we can find that the reformulated problem (7) is still not a convex problem. To make this problem tractable, we recheck the structure of objective function in (7a)., and observe that objective function in (7a) holds the structure of a nonlinear sum of fractional functions. To maximize a sum of fractional functions subject to the non-convex constraints is a sum-of-ratios fractional programming problem, which is difficult to solve by conventional optimization methods [9]. To solve this problem, we use the sum-of-ratios algorithm by decoupling the numerators and denominators of objective function in (7a) into an equivalent parametric subtractive structure, which can be represented by

$$\widetilde{\xi}_{\text{EE}} = \widetilde{R}_{\text{S}} - \mu\left(\widetilde{P}_{\text{A}} + \widetilde{P}_{\text{B}}\right), \quad (8)$$

where μ is an auxiliary parameter. Note that objective function in (8) is still non-concave due to the interference terms in non-concave sum rate function $\widetilde{R}_{\text{S}}$. To obtain the convex structure of objective function, through the feature of logarithmic structure, we rewrite $\widetilde{R}_{\text{S}}$ as the difference of convex structure

$$\begin{aligned}\widetilde{R}_{\text{S}} =& \sum_{f=1}^{F}\sum_{m=1}^{M_f}\sum_{n=1}^{N_f}\sum_{k=1}^{\delta} a^k_{fmn} \\ &\times \log_2\left(1+\widetilde{p}^k_{fmn}H^k_{fmn}+\sum_{j=m+1}^{M^k_f}\widetilde{p}^k_{fjn}H^k_{fjn}\right) \\ &-\sum_{f=1}^{F}\sum_{m=1}^{M_f}\sum_{n=1}^{N_f}\sum_{k=1}^{\delta} a^k_{fmn}\log_2\left(1+\sum_{j=m+1}^{M^k_f}\widetilde{p}^k_{fjn}H^k_{fjn}\right). \quad (9)\end{aligned}$$

Based on the subtractive structure in (8) and the logarithmic operation in (9), problem (7) can be further rewritten by

$$\begin{aligned}\max_{\mathbf{A},\mathbf{B},\widetilde{\mathbf{P}},\widetilde{\mathbf{Q}}} & \sum_{f=1}^{F}\sum_{m=1}^{M_f}\sum_{n=1}^{N_f}\sum_{k=1}^{\delta} a^k_{fmn} \\ &\times \log_2\left(1+\widetilde{p}^k_{fmn}H^k_{fmn}+\sum_{j=m+1}^{M^k_f}\widetilde{p}^k_{fjn}H^k_{fjn}\right) \\ &-\sum_{f=1}^{F}\sum_{m=1}^{M_f}\sum_{n=1}^{N_f}\sum_{k=1}^{\delta} a^k_{fmn}\log_2\left(1+\sum_{j=m+1}^{M^k_f}\widetilde{p}^k_{fjn}H^k_{fjn}\right) \\ &-\mu\sum_{f=1}^{F}\sum_{m=1}^{M_f}\sum_{n=1}^{N_f}\sum_{k=1}^{\delta}\left(P_{fn}+P^{\text{C}}_m+\widetilde{p}^k_{fmn}\right) \\ &-\mu\sum_{f=1}^{F}\sum_{m=1}^{M_f}\sum_{k=\delta+1}^{K}\left(P_{fm}+\widetilde{q}^k_{fm}\right) \quad (10) \\ \text{s.t.} \quad & \text{(7b)}, \text{(7c)}, \text{(7d)}, \text{(7e)}, \text{(7f)}.\end{aligned}$$

C. Successive Convex Approximation

Due to the non-convexity of problem (10) caused by constraints in (7b) and (7c), we resort to the iterative SCA method for solving it, where, in each iteration, the original non-convex problem is approximately converted into a convex problem. For notational simplicity, let $\gamma^{k,1}_{fmn} = \frac{\widetilde{p}^k_{fmn}H^k_{fmn}}{1+\sum_{j=m+1}^{M^k_f}\widetilde{p}^k_{fjn}H^k_{fjn}}$. As in [10], a lower bound of $\widetilde{R}_{\text{S}}$ is determined by

$$\widetilde{R}_{\text{S}} \geq \sum_{f=1}^{F}\sum_{m=1}^{M_f}\sum_{n=1}^{N_f}\sum_{k=1}^{\delta} a^k_{fmn}\left(\alpha^k_{fmn}\log_2\left(\gamma^{k,1}_{fmn}\right)+\beta^k_{fmn}\right), \quad (11)$$

where α^k_{fmn} and β^k_{fmn} are the auxiliary approximation variables, respectively, which can be calculated as follows to tighten the lower bound in (11), i.e.,

$$\alpha^k_{fmn} = \frac{\gamma^{k,1}_{fmn}}{1+\gamma^{k,1}_{fmn}}, \quad (12)$$

$$\beta^k_{fmn} = \log_2\left(1+\gamma^{k,1}_{fmn}\right) - \frac{\gamma^{k,1}_{fmn}}{1+\gamma^{k,1}_{fmn}}\log_2\left(\gamma^{k,1}_{fmn}\right). \quad (13)$$

By letting $\gamma^{k,2}_{fm} = \frac{\widetilde{q}^k_{fm}\left|\mathbf{h}^k_{fm}\mathbf{w}^k_f\right|^2}{\left|\mathbf{h}^k_{fm}\mathbf{w}^k_f\right|^2\sum_{j=1,j\neq m}^{\phi^k_f}\widetilde{q}^k_{fj}+\sigma^2_{mk}}$, we can also obtain a lower bound of $\widetilde{R}^k_{fm}$, which can be given by

$$\widetilde{R}^k_{fm} \geq \varLambda^k_{fm}\log_2\left(\gamma^{k,2}_{fm}\right) + \varXi^k_{fm}, \quad (14)$$

where $\varLambda^k_{fm}$ and $\varXi^k_{fm}$ are the auxiliary approximation variables, respectively, which can be expressed as follows to tighten the lower bound in (14), i.e.,

$$\varLambda^k_{fm} = \frac{\gamma^{k,2}_{fm}}{1+\gamma^{k,2}_{fm}}, \quad (15)$$

$$\varXi^k_{fm} = \log_2\left(1+\gamma^{k,2}_{fm}\right) - \frac{\gamma^{k,2}_{fm}}{1+\gamma^{k,2}_{fm}}\log_2\left(\gamma^{k,2}_{fm}\right). \quad (16)$$

Define $\widehat{p}^k_{fmn} = \log_2\left(\widetilde{p}^k_{fmn}\right)$ and $\widehat{q}^k_{fm} = \log_2\left(\widetilde{q}^k_{fm}\right)$. Let $\widehat{\mathbf{P}} = \{\widehat{p}^k_{fmn}\}^{M_f,N_f,\delta}_{m,n,k=1}$ and $\widehat{\mathbf{Q}} = \{\widehat{q}^k_{fm}\}^{M_f,K}_{m=1,k=\delta+1}$, for $f \in \mathcal{F}$. By applying the lower bounds in (11) and (14) as well as the logarithmic change of variables into a logarithmic transformation of objective and constraint functions in problem (10), we arrive at the following approximate parametric subproblem

$$\begin{aligned}\max_{\mathbf{A},\mathbf{B},\widehat{\mathbf{P}},\widehat{\mathbf{Q}}} & \sum_{f=1}^{F}\sum_{m=1}^{M_f}\sum_{n=1}^{N_f}\sum_{k=1}^{\delta} a^k_{fmn}\left(\alpha^k_{fmn}\log_2\left(\gamma^{k,1}_{fmn}\right)+\beta^k_{fmn}\right) \\ &-\mu\sum_{f=1}^{F}\sum_{m=1}^{M_f}\sum_{n=1}^{N_f}\sum_{k=1}^{\delta}\left(P_{fn}+P^{\text{C}}_m+\exp\left(\widehat{p}^k_{fmn}\right)\right) \\ &-\mu\sum_{f=1}^{F}\sum_{m=1}^{M_f}\sum_{k=\delta+1}^{K}\left(P_{fm}+\exp\left(\widehat{q}^k_{fm}\right)\right) \triangleq \aleph\left(\mathbf{A},\mathbf{B},\widehat{\mathbf{P}},\widehat{\mathbf{Q}}\right) \quad (17)\end{aligned}$$

$$\begin{aligned}\text{s.t. } & \sum_{f=1}^{F}\sum_{m=1}^{M_f}\sum_{k=1}^{\delta} a_{fmn}^{k}\left(\alpha_{fmn}^{k}\log_2\left(\gamma_{fmn}^{k,1}\right)+\beta_{fmn}^{k}\right) \geq R_n^{\min},\ \forall n,\\ & \sum_{f=1}^{F}\sum_{k=\delta+1}^{K} b_{fm}^{k}\left(\mathit{\Lambda}_{fm}^{k}\log_2\left(\gamma_{fm}^{k,2}\right)+\mathit{\Xi}_{fm}^{k}\right)\\ & \geq \sum_{f=1}^{F}\sum_{n=1}^{N_f}\sum_{k=1}^{\delta} a_{fmn}^{k}\left(\alpha_{fmn}^{k}\log_2\left(\gamma_{fmn}^{k,1}\right)+\beta_{fmn}^{k}\right),\ \forall m,\\ & \sum_{f=1}^{F}\sum_{m=1}^{M_f}\sum_{k=\delta+1}^{K} \exp\left(\widehat{q}_{fm}^{k}\right) \leq P_{\max},\ \forall f, \forall m, \forall k,\\ & \sum_{n=1}^{N_f}\sum_{k=1}^{\delta} \exp\left(\widehat{p}_{fmn}^{k}\right) \leq P_m^{\max},\ \forall f, \forall m,\\ & a_{fmn}^{k}, b_{fm}^{k} \in [0,1],\ \forall f, \forall m, \forall n, \forall k.\end{aligned}$$

Note that subproblem (17) follows the log-sum-exp function structure. Thus, we finally convert the original problem (6) into a standard convex problem with logarithmic change variables. In fact, we only maximize a lower bound of objective function in (17). To effectively solve subproblem (17), with the SCA method, we need to tighten the bound in (11) by iteratively updating α_{fmn}^{k} and β_{fmn}^{k}, and also tighten the bound in (14) by iteratively updating $\mathit{\Lambda}_{fm}^{k}$ and $\mathit{\Xi}_{fm}^{k}$. Due to the space limitation, detailed procedure of the iterative algorithm via the SCA method to tighten the bounds in (11) and (14) is omitted here, and readers can refer to [10] for detailed description.

D. Lagrangian Dual Decomposition

In this subsection, the standard convex problem in (17) is solved by using the Lagrangian dual decomposition method. The Lagrangian function is given by

$$\begin{aligned} & L\left(\mathbf{A},\mathbf{B},\widehat{\mathbf{P}},\widehat{\mathbf{Q}},\boldsymbol{\lambda},\boldsymbol{\varphi},\eta,\boldsymbol{\chi}\right) = \aleph\left(\mathbf{A},\mathbf{B},\widehat{\mathbf{P}},\widehat{\mathbf{Q}}\right)\\ & +\sum_{n=1}^{N_f}\lambda_n\left(\sum_{f=1}^{F}\sum_{m=1}^{M_f}\sum_{k=1}^{\delta} a_{fmn}^{k}\left(\alpha_{fmn}^{k}\log_2\left(\gamma_{fmn}^{k,1}\right)+\beta_{fmn}^{k}\right)-R_n^{\min}\right)\\ & +\sum_{m=1}^{M_f}\varphi_m\left(\sum_{f=1}^{F}\sum_{k=\delta+1}^{K} b_{fm}^{k}\left(\mathit{\Lambda}_{fm}^{k}\log_2\left(\gamma_{fm}^{k,2}\right)+\mathit{\Xi}_{fm}^{k}\right)\right.\\ & \qquad\left.-\sum_{f=1}^{F}\sum_{n=1}^{N_f}\sum_{k=1}^{\delta} a_{fmn}^{k}\left(\alpha_{fmn}^{k}\log_2\left(\gamma_{fmn}^{k,1}\right)+\beta_{fmn}^{k}\right)\right)\\ & +\eta\left(P_{\max}-\sum_{f=1}^{F}\sum_{m=1}^{M_f}\sum_{k=\delta+1}^{K}\exp\left(\widehat{q}_{fm}^{k}\right)\right)\\ & +\sum_{f=1}^{F}\sum_{m=1}^{M_f}\chi_{fm}\left(P_m^{\max}-\sum_{n=1}^{N_f}\sum_{k=1}^{\delta}\exp\left(\widehat{p}_{fmn}^{k}\right)\right), \end{aligned} \tag{18}$$

where $\boldsymbol{\lambda}$, $\boldsymbol{\varphi}$, η, and $\boldsymbol{\chi}$ are the Lagrange multiplier vectors for the constraints except for binary constraints in (17). The boundary constraints for binary constraints will be absorbed in the KKT conditions. The Lagrange dual function is given as $g\left(\boldsymbol{\lambda},\boldsymbol{\varphi},\eta,\boldsymbol{\chi}\right)=\max\limits_{\mathbf{A},\mathbf{B},\widehat{\mathbf{P}},\widehat{\mathbf{Q}}} L\left(\mathbf{A},\mathbf{B},\widehat{\mathbf{P}},\widehat{\mathbf{Q}},\boldsymbol{\lambda},\boldsymbol{\varphi},\eta,\boldsymbol{\chi}\right)$. Then the Lagrangian dual problem is formulated by

$$\min_{\boldsymbol{\lambda},\boldsymbol{\varphi},\eta,\boldsymbol{\tau}\geq 0} g\left(\boldsymbol{\lambda},\boldsymbol{\varphi},\eta,\boldsymbol{\chi}\right). \tag{19}$$

The dual variables are optimized by subgradient method based on the KKT conditions, which are specified as in [11]. According to the KKT conditions, the optimal solutions of the subproblem (17), denoted by $\{\widehat{p}_{fmn}^{k,*}\}$, $\{\widehat{q}_{fm}^{k,*}\}$, $\{a_{fmn}^{k,*}\}$, and $\{b_{fm}^{k,*}\}$, can be respectively obtained as

$$\frac{\partial L\left(\cdots\right)}{\partial \widehat{p}_{fmn}^{k,*}}=0 \quad\text{and}\quad \frac{\partial L\left(\cdots\right)}{\partial \widehat{q}_{fm}^{k,*}}=0, \tag{20}$$

$$\frac{\partial L\left(\cdots\right)}{\partial a_{fmn}^{k,*}} = \begin{cases} <0 & a_{fmn}^{k,*}=0,\\ =0 & 0<a_{fmn}^{k,*}<1\ ,\\ >0 & a_{fmn}^{k,*}=1, \end{cases} \tag{21}$$

$$\frac{\partial L\left(\cdots\right)}{\partial b_{fm}^{k,*}} = \begin{cases} <0 & b_{fm}^{k,*}=0,\\ =0 & 0<b_{fm}^{k,*}<1,\\ >0 & b_{fm}^{k,*}=1. \end{cases} \tag{22}$$

The optimal power of AP m in cluster f to user n on sub-channel k and the optimal power of MBS to AP m on sub-channel k in cluster f is obtained by solving the equations

$$\begin{aligned}\widehat{p}_{fmn}^{k,*} &= \ln\left(\frac{a_{fmn}^{k,*}\alpha_{fmn}^{k}\left(\varphi_m-\lambda_n-1\right)}{\mu+\chi_{f,m}}\right.\\ & \left.\times\left(1-\frac{\sum_{j=m+1}^{M_f^k}\exp\left(\widehat{p}_{fmn}^{k,*}\right)H_{fjn}^{k}}{\left(1+\sum_{j=m+1}^{M_f^k}\exp\left(\widehat{p}_{fmn}^{k,*}\right)H_{fjn}^{k}\right)\ln 2}\right)\right),\end{aligned} \tag{23}$$

$$\begin{aligned}\widehat{q}_{fm}^{k,*} &= \ln\left(\frac{\varphi_m b_{fm}^{k}\mathit{\Lambda}_{fm}^{k}}{\mu+\eta}\right.\\ & \left.\times\left(1-\frac{\left|\mathbf{h}_{fm}^{k}\mathbf{w}_{f}^{k}\right|^2\sum_{j=1,j\neq m}^{\phi_f^k}\exp\left(\widehat{q}_{fm}^{k,*}\right)}{\left(\left|\mathbf{h}_{fm}^{k}\mathbf{w}_{f}^{k}\right|^2\sum_{j=1,j\neq m}^{\phi_f^k}\exp\left(\widehat{q}_{fm}^{k,*}\right)+\sigma_{mk}^2\right)\ln 2}\right)\right).\end{aligned} \tag{24}$$

Through the partial derivative of the Lagrangian, sub-channel k^* is assigned to user n by AP m in cluster f such that $a_{fmn}^{k^*,*}=1$, and sub-channel k^* is assigned to AP m in cluster f by MBS such that $b_{fm}^{k^*,*}=1$, which can be given as

$$a_{fmn}^{k^*,*}\Big|_{k^*=\arg\max\limits_{k}\frac{\partial L(\cdots)}{\partial a_{fmn}^{k,*}}}=1, \tag{25}$$

$$b_{fm}^{k^*,*}\Big|_{k^*=\arg\max\limits_{k}\frac{\partial L(\cdots)}{\partial b_{fm}^{k,*}}}=1. \tag{26}$$

We use the subgradient method and update the dual variables by setting the step sizes for each iteration. Due to the space limitation, the specific updated process for dual variables is omitted here, and readers can refer to [11] for detailed description. The overall algorithm to iteratively realize the joint optimization of user/AP scheduling, subchannel assignment, and power allocation is sketched in Algorithm 1.

IV. SIMULATION RESULTS

In this section, we present simulations results to verify the performance of our proposed algorithm as compared to the equal-power based allocation scheme as a benchmark. We consider a macrocell area with radius $r=200$m centered at MBS, wherein the locations of users and APs are randomly generated with equal possibility and deployed subject to the independent

Algorithm 1 Proposed Resource Allocation Algorithm

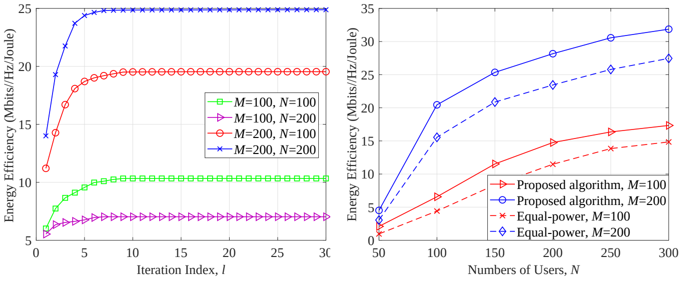
```
Initialize maximum number of iterations $L_{\max}$ and Lagrange multipliers $\boldsymbol{\lambda}$, $\boldsymbol{\varphi}$, $\eta$, $\boldsymbol{\chi}$, set iteration index $l = 1$.
Obtain updated variables $\alpha_{fmn}^k$, $\beta_{fmn}^k$, $\Lambda_{fm}^k$, $\Xi_{fm}^k$ via SCA in [10].
repeat
    for $f = 1$ to $F$ do
        for $m = 1$ to $M$ do
            Calculate sub-channel $k^*$ using (26) and update $b_{fm}^{k^*,*}$.
            Solve (24) to update $\widehat{q}_{fm}^{k^*,*}$.
            for $n = 1$ to $N$ do
                Calculate sub-channel $k^*$ using (25) and update $a_{fmn}^{k^*,*}$.
                Solve (23) to update $\widehat{p}_{fmn}^{k^*,*}$.
            end for
        end for
    end for
    Update Lagrange multipliers $\boldsymbol{\lambda}$, $\boldsymbol{\varphi}$, $\eta$, $\boldsymbol{\chi}$ by [11], and set $l = l + 1$.
until onvergence or $l = L_{\max}$.
```

TABLE I
SIMULATION PARAMETERS.

Parameter	Value
Maximum associated number of APs for user in f, M_f	16
Maximum associated number of APs for MBS in f, ϕ_f^k	10
Maximum number of APs on subchannel k in $\mathcal{G}_f$, M_f^k	12
Path loss exponent for access downlink, ϑ_1	2
Flat Rayleigh fading channel gain, g_{fmn}^k	$\mathcal{CN}(0,1)$
Noise variance at user n on subchannel k, σ_{nk}^2	-174dBm/Hz
Path loss exponent for backhaul downlink, ϑ_2	2
Small scale Rayleigh fading channel vector, $\tilde{\mathbf{h}}_{fm}^k$	$\mathcal{CN}(0, \mathbf{I}_Q)$
Noise variance at AP m on subchannel k, σ_{mk}^2	-174dBm/Hz

HPPPs, respectively. The minimum distance between APs is set to 3m, and the minimum distance between users is set to 1.2m. For downlink, the power consumption parameters are set as $P_{fn}^{\mathrm{R}} = 5$mW, $P_{fn}^{\mathrm{D}} = 10$mW, $P_m^{\mathrm{C}} = 50$mW, $P_{fm}^{\mathrm{R}} = 15$mW, and $P_{fm}^{\mathrm{D}} = 30$mW. All the APs are separated into $F = 8$ disjoint clusters based on their spatial direction to MBS, i.e., 45° direction angle interval. As in [12], beamforming vector for AP on each sub-channel in a cluster is generated through the channel coefficient vector $\tilde{\mathbf{h}}_{fm}^k$ between MBS and that AP. We assume that the number of transmit antennas for beamforming in antenna array of MBS is equal to the number of APs on each sub-channel in a cluster for simplicity of simulations. The other simulation parameters are summarized in Table I.

Fig. 1(a) shows the convergence of the proposed algorithm for four cases with different combinations with the numbers of APs and users. It can be seen that the proposed algorithm converges rapidly in less than 10 iterations to reach the optimal points. With different combinations of APs and users, better system EE performance is obtained when $M = N = 200$. From Fig. 1(a), we can conclude that the proposed algorithm has good convergence performance. In addition, Fig. 1(b) shows the performance comparison in terms of system EE versus the number of users between the proposed algorithm and the benchmark scheme. We can find that the system EE greatly increases with the continuous evolution of the number of users. It is further observed that the proposed algorithm significantly outperforms the benchmark scheme in terms of the system EE. Such an insight, to some extent, is aligned with the fact that the proposed algorithm fully achieves the joint optimization of resource allocation, and thereby achieves good performance.



(a) (b)

Fig. 1. Simulation results achieved when $R_n^{\min} = 100$bps/Hz, $P_{\max} = 46$dBm, and $P_m^{\max} = 32$dBm: (a) Convergence behavior of the proposed algorithm. (b) System EE versus the number of users.

V. CONCLUSION

This paper studied the energy efficient resource optimization problem for the downlink user-centric UDNs integrating NOMA and beamforming. The system EE maximization problem was formulated as a non-convex MINLP problem. By using the sum-of-ratios decoupling and the iterative SCA method, this problem was transformed into a convex parametric subproblem. Then the overall algorithm was devised to obtain the joint optimization of user/AP scheduling, subchannel assignment, and power allocation. The simulation results after comparison with the benchmark scheme revealed that our proposed algorithm accomplishes significant enhancement in the system EE.

REFERENCES

[1] X. Ge, S. Tu, G. Mao, C.-X. Wang, and T. Han, "5G ultra-dense cellular networks," *IEEE Wireless Commun.*, vol. 23, no. 1, pp. 72–79, Feb. 2016.
[2] P.-H. Kuo and A. Mourad, "User-centric multi-RATs coordination for 5G heterogeneous ultra-dense networks," *IEEE Wireless Commun.*, vol. 25, no. 1, pp. 6–8, Feb. 2018.
[3] Y. Liu, Z. Qin, M. Elkashlan, Z. Ding, A. Nallanathan, and L. Hanzo, "Nonorthogonal multiple access for 5G and beyond," *Proc. IEEE*, vol. 105, no. 12, pp. 2347–2381, Dec. 2017.
[4] J. Park, S. Y. Jung, S.-L. Kim, M. Bennis, and M. Debbah, "User-centric mobility management in ultra-dense cellular networks under spatio-temporal dynamics," in *Proc. IEEE GLOBECOM*, Washington, DC, USA, Dec. 2016.
[5] H. Zhang, S. Huang, C. Jiang, K. Long, V. C. M. Leung, and H. V. Poor, "Energy efficient user association and power allocation in millimeter-wave-based ultra dense networks with energy harvesting base stations," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1936–1947, Sep. 2017.
[6] Y. Liu, X. Li, F. R. Yu, H. Ji, H. Zhang, and V. C. M. Leung, "Grouping and cooperating among access points in user-centric ultra-dense networks with non-orthogonal multiple access," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2295–2311, Oct. 2017.
[7] Z. Qin, X. Yue, Y. Liu, Z. Ding, and A. Nallanathan, "User association and resource allocation in unified NOMA enabled heterogeneous ultra dense networks," *IEEE Commun. Mag.*, vol. 56, no. 6, pp. 86–92, Jun. 2018.
[8] G. Kwon and H. Park, "Joint user association and beamforming design for millimeter wave UDN with wireless backhaul," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 12, pp. 2653–2668, Dec. 2019.
[9] S. Schaible and J. Shi, "Fractional programming: The sum-of-ratios case," *Optim. Meth. and Softw.*, vol. 18, no. 2, pp. 219–229, Apr. 2003.
[10] J. Papandriopoulos and J. S. Evans, "SCALE: A low-complexity distributed protocol for spectrum balancing in multiuser DSL networks," *IEEE Trans. Inf. Theory*, vol. 55, no. 8, pp. 3711–3724, Aug. 2009.
[11] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge University Press, Mar. 2004.
[12] B. Kimy, S. Lim, H. Kim, S. Suh, J. Kwun, S. Choi, C. Lee, S. Lee, and D. Hong, "Non-orthogonal multiple access in a downlink multiuser beamforming system," in *Proc. IEEE MILCOM*, San Diego, CA, USA, Nov. 2013, pp. 1278–1283.