Offloading Design for Energy and Spectral Efficiencies Tradeoff in Massive MIMO Enabled Heterogeneous Cellular Networks

TIANQING ZHOU1, NAN JIANG1, DONG QIN2, XUAN LI1, HENG ZHANG1, CHUNGUO LI3, LUXI YANG3
1School of Information Engineering, East China Jiaotong University, Nanchang 330013, China
2School of Information Engineering, Nanchang University, Nanchang 330031, China
3School of Information Science and Engineering, Southeast University, Nanjing 210096, China

Corresponding author: L. X. Yang (e-mail: lxyang@seu.edu.cn).

ABSTRACT
To fully utilize the system resources and enhance the user experience, two types of offloading mechanisms are designed for massive MIMO (multiple-input-multiple-out) enabled heterogeneous cellular networks (HCNs). Such mechanisms can achieve a tradeoff between energy efficiency (EE) and spectral efficiency (SE) experiences, but also have a difference on whether or not a power coordination measure is mentioned. At last, they are mathematically written as the network-wide utility maximization problems that are closely related to the SEs and EEs of associated users. In them, we introduce a crucial parameter $\alpha$ to adjust the EE-SE preference. It is noteworthy that the finally formulated problems are in some relatively complicated forms. To solve them, some necessary changes should be made at first, and then we can design some feasible algorithms. Specifically, we try to design a distributed algorithm for a mere offloading problem using dual decomposition (DD), and develop a two-layer iterative algorithm for the joint power coordination and offloading problem using DD and two-sided scalable (2.s.s.) function update. Regarding such these algorithms, we show the corresponding computation complexity and convergence analyses. In the simulation, we mainly investigate different network parameters on an EE-SE tradeoff for our advocated mechanisms and another existing mechanism.

INDEX TERMS
User association, offloading design, heterogeneous cellular networks, massive MIMO, energy efficiency, spectral efficiency.

I. INTRODUCTION
Massive MIMO (multiple-input-multiple-out) enabled heterogeneous cellular networks (HCNs) are envisioned as a crucial technology to improving the area spectrum efficiency (SE) for the 5th generation (5G) mobile communications [1], [2], which can approximately provide the 1000-fold wireless capacity and meet the ubiquitous access requirements for the future network development [3]–[6]. Massive MIMO enabled HCNs (MM-HCNs) integrate many kinds of low-power nodes (base stations, BSs) to enhance the network coverage [7]–[9], and equip large-scale antennas for the high-power nodes to achieve an unprecedented SE.

The offloading mechanism design is regarded as an important and indispensable part of radio resource management in wireless networks, and has attracted much and much attention. In fact, an offloading mechanism definitely assigns some users to one or several BSs, and thus it also refers to the user association (cell selection or BS assignment) [10], [11]. Most of the time, many designers often take account of an offloading operation to improve the (cell-edge) user...
experience by utilizing network resources economically. To this end, these designers often try to optimize the SEs or the utility of them.

As another important performance index, the energy efficiency (EE) has attracted increasing attention from both academia and industry, and it has been optimized to reduce operational cost and achieve a goal of green communications [12]. To reduce the energy consumption and thus improve EE, some energy-efficient offloading mechanisms are widely advocated, which may integrate a BS on/off operation, power coordination [13]–[15] and so on. On one hand, such these mechanisms can improve EE in the energy-efficient optimization, but they may result in some degraded SEs because of decreased transmit power or longer user-BS distance. On other hand, these mechanisms can enhance SEs by reducing network interference in the throughput optimization, but they often cannot achieve the best EEs. That is to say, there should exist a tradeoff between SE and EE.

In general, some offloading mechanisms designed for MM-HCNs may need to optimize some certain performance metric (e.g., SE, EE or EE-SE tradeoff) of interest. To this end, most of designers often consider some different optimization objectives, e.g., achievable rates, long-term rates, transmit power, geographical location, cell loads, etc. Next, we will give some detailed investigations on the offloading mechanism design for HCNs, especially for MM-HCNs. As we know, HCNs can be seen as a special case of MM-HCNs. In the existing efforts, there are many investigations on the offloading mechanism design for HCNs, but relatively few ones on it for MM-HCNs. In view of this, we may need to recall the relative work on both HCNs and MM-HCNs.

A. RELATED WORK
According to the difference of optimized performance metrics, most existing efforts on the offloading mechanism design can be roughly divided into three groups.

In the first group, some designers concentrated on the offloading mechanisms with SE optimization. So far, such a type has been thoroughly studied in the literatures. To balance the network loads, some designers in [16], [17] added a bias/offset to low-power nodes and thus theoretically enhanced the signal strength of users nearing to them. Through such a treatment, some users may be associated with some BSs with good biasing signal power. This method is often regarded as the cell range expansion, which can offload some cell-edge users at high-power nodes to some other low-power ones. However, it may be difficult for this approach to find a closed-form solution of optimal bias/offset, and thus it may not reach an optimal offloading performance. In another different light, some designers got involved with optimizing offloading indices in some formulated problems, e.g., network throughput maximization [17], logarithmic utility maximization [18]–[24] or $\alpha$-utility maximization [25], [26], where the mentioned utilities may refer to some functions with respect to users’ data (long-term) rates. In this type of offloading mechanisms, user’s long-term rate is an important parameter used for balancing the network loads and thus improving the experience of cell-edge users. Definitely, it is tightly coupling with both the load level of BS and the signal power of users associated with this BS. In order to achieve a high network experience, some users may have to give up overloaded BSs with good channel conditions, and access some underloaded BSs with relatively high channel qualities. Through this manner, the network resources can be fully utilized, and then users may get full and better services. In the first group, other designers also developed some different offloading mechanism to find a tradeoff between user experience/fairness and network throughput under the max-min optimization [27].

In the second group, some designers worked on the design of energy-efficient offloading mechanisms, which optimizes offloading indices in the formulated EE problems. To reduce the network energy consumption and reach a goal of green communications, some designers directly minimized total power consumption such as the sum power of BSs, the one of users, or the one of both users and BSs. To this end, many different energy-reducing mechanisms were introduced in the offloading design, which mainly included the power coordination/control, power allocation and beamforming. Certainly, the power consumption reduction should be performed under users’ QoS (quality-of-service) requirements. So far, the sum power minimization for offloading has been widely studied [28], [29]. In addition, there exist other energy-efficient offloading mechanisms that mentioned different optimization objectives, e.g., the sum energy efficiency/consumption [30]–[34], the sum utility of users’ EE [35], the overall (whole) EE [36], etc. All these mechanisms finally tried to optimize some performance metrics related to EEs under users’ QoS requirements, where the individual (user’s) EE represents a ratio of user’s data rate to the power consumed by it in uplink or its selected BS in downlink, and the overall EE denotes a ratio of sum rate to the sum power that is consumed by all users in uplink or all BSs in downlink. In fact, these two performance indices differ in perspectives. Specifically, an individual EE is introduced from user’s perspective, but an overall EE is mentioned from the system perspective.

In the last group, there exist many designs that attempted to achieve a tradeoff between some special performance metrics and energy consumption. Specifically, such these metrics may refer to the long-term or achievable rates, delay, and network loads. In many existing efforts, the researchers often utilized one or more of them to design some energy-efficient offloading mechanisms, and finally reached a tradeoff between these metrics and power consumption. To date, such a type of mechanisms mainly focused on a tradeoff between SE and EE [37]–[40]. By optimizing the delay, rate or load distribution, the SE (experience) may be improved, and the EE experience can be further enhanced by properly adjusting the transmit power [41]. In addition, some QoS requirements may also need to be involved in this kind of offloading designs.

To improve the performance metrics or the user expe-
erience, many designers took account of some interference 
management measures in the aforementioned groups. Among
them, the resource (time/frequency/space/power) allocation 
and beamforming have been widely advocated for offloading 
designs, but the resource partitioning and power coordination 
(adjustment) are rarely utilized in downlink HCNs, especially 
downlink MM-HCNs. It is easy to find that the resource 
allocation may result in a waste of network resources if
some selected users cannot be served in a scheduling loop.
In addition, the beamforming and offloading often take place 
at small and large time scales respectively. That means they 
utilize fast-fading and slow-fading channels respectively.
In general, the resource allocation and beamforming may not
be well suitable for offloading designs. However, the resource
partitioning and power control (coordination) may be well
applied in them since they just refer to the resource adjust-
ment of nodes and don’t mention any specific distribution for
users.

In this paper, we design two types of offloading mecha-
nisms to achieve a tradeoff between EE and SE (experiences),
which differ from the most ones mentioned in the third group.
Although our work seems to be similar to the one of [37]
and [40], there exist some distinct differences. At first, the
optimization objective in this paper differs from the ones in
[37] and [40]. The mentioned one in this paper is tightly
related to SE and EE, which can be beneficial to the pure SE
optimization, pure EE optimization, and joint optimization of
SE and EE. Compared with the references [37] and [40], our
optimization objective may have a clearer insight. It is easy
to find that the optimization problems in [37] and [40] may have
no relation with association (offloading) variables if weight-
ing parameter is equal to zero. In other word, these problems
are not the offloading ones at this time. Secondly, the lower
and upper bounds of sum utility and total power consumption
may have a great impact on the performance of offloading
algorithms in [37] and [40], but it may not be the case in our
algorithms. Generally speaking, our designed algorithms are
independent on the lower and upper bounds of SEs and EEs.
That means the performance of these algorithms may have
no any direct relation with the mentioned bounds. Thirdly, the
optimization problem designed in this paper further considers
the long-term rate constraints of users. Evidently, such a con-
sideration should be necessary for guaranteeing users’ QoS
requirements when the energy-efficient offloading attracts
more and more attention. However, the references [37] and
[40] may not provide such a support for users. At last, we
try to find the optimal power of BSs using two-sided scalable
(2.s.s.) function update in this paper, but not the approaches
in [37] and [40]. Through a direct observation on [37] and
[40], we find that too small SINR may result in "log(0)",
which may let the designed algorithms cannot work well.

B. CONTRIBUTIONS AND ORGANIZATION

Generally, the existing efforts tried to get a tradeoff between
some certain performance metrics and power consumption,
but we consider the one between SE and EE (experiences)
under some necessary constraints, which is a completely
aspect for HCNs and MM-HCNs. The main contributions
in this paper can be listed as follows.

- Mere Offloading for EE-SE Tradeoff (MOET). We
just consider a mere offloading for an EE-SE tradeoff, which
is hardly involved in the existing literatures. Such a mech-
anism (MOET) integrates network loads and some other
parameters that can directly reflect SEs and EEs. Mathemati-
cally, it is formulated as a network-wide utility maximization
problem in a relatively complicated form, where a utility
function is used to further enhance the user fairness, and a
parameter $\alpha$ is used for adjusting the EE-SE preference.
To solve it, we may need to make some promising changes
and then develop a distributed algorithm using dual decomposi-
tion.

- Joint Power Coordination and Offloading for EE-
SE Tradeoff (PCOET). To mitigate the network interference
and reduce the power consumption under some strict users’
QoS constraints, a power coordination measure is integrated
into MOET. Such a mechanism (PCOET) takes account of
power coordination for downlink MM-HCNs, which may be
seldom mentioned in the past studies. Evidently, compared
to the problem formulated for MOET, the one for PCOET
should be more complicated since it owns some coupling
optimization parameters (i.e., offloading indices and transmit
power). To solve such a type of problem, there may not be
a good method but alternate optimization. Based on this, we
design a feasible algorithm consisting of two layer loops.
In the inner layer loops, the offloading indices and transmit
power are separately optimized; in the outer layer loop, the
operation in all inner layer loops is repeatedly carried out
until it converges or achieves the maximal allowed number
of iterations.

- Convergence and Complexity Analyses. As for the
algorithms designed for MOET and PCOET, we give some
investigations on the convergence and complexity analy-
ses. Particularly, we show some convergence proofs for
offloading and power coordination subalgorithms mentioned
in the inner layer loops. In addition, there also exist some
comprehensive discussions for the complexity reduction
and some specific analyses for all steps involved in the solving
processes of MOET and PCOET problems.

The remainder of this paper can be organized as follows.
The system model of MM-HCNs is given in Section II; two
types of offloading mechanisms (i.e., MOET and PCOET)
are designed, and the corresponding convergence and com-
plexity analyses are attached in Section III; the simulation
results are given and discussed in Section IV; some conclu-
sions are drawn in Section V.

II. SYSTEM MODEL

In this paper, we consider two-tier MM-HCNs consisting of
macro BSs (MBSs) and pico BSs (PBSs). In such networks,
the large-scale antennas are implemented at each MBS, but
single antenna is employed at each PBS [35]. The corre-
ponding deployment can be found in Fig. 1, where the

VOLUME 4, 2016
extended region is caused by an offloading mechanism that offloads some users to low-power nodes (PBSs) from high-power nodes (MBSs).

In MM-HCNs, there exist $S$ BSs consisting of PBSs and MBSs in the set $S$, and $U$ users in the set $U$. We assume that any user can just select (be associated with) only one BS at any time slot. That is to say, we take account of single association but not the multiple one. Since the latter lets one user be served by multiple BSs at the same, it often incurs a more higher implementation difficulty than the former.

A. KEY PARAMETERS FOR MASSIVE MIMO MBSS

In MM-HCNs, any MBS is equipped with $M$ antennas. Under the equal power allocation, any resource block (RB) can simultaneously serve at most $N$ downlink data streams at MBSs [35], where the condition $1 \ll N \ll M$ needs to be satisfied. Under the LZFMBF (linear zero-forcing beamforming) precoding, the data rates of users gradually concentrate on deterministic limits when $M$ approaches to infinity. That is to say, users’ data rates are independent on the fast fading (instantaneous realization) of communication channels, and just rely on the slow fading $g_{su}$ including pathloss and shadowing [42]. In the reality, the number of users associated with any MBS is always larger than $N$, which means the load constraints are unnecessary for all MBSs [35], [37]. Assume that all associated users of any MBS are served via resource sharing [35], [37]. Then, the downlink data rate $R_{su}$ of user $u$ associated with MBS $s$ is given by

$$\hspace{1cm} R_{su} = \frac{N}{\sum_{k \in \mathcal{U}} \rho_{sk}} \log_2 \left( 1 + \frac{N - 1}{N} \sinr_{su} \right),$$

where

$$\hspace{1cm} \sinr_{su} = \frac{p_s g_{su}}{\sum_{i \in S_s} p_i g_{isu} + \theta_s}.$$  

In (1) and (2), $p_s$ is the transmit power of BS $s$, $\Theta_s$ is the noise power of BS $s$, $S_s$ represents the BS set that $S$ doesn’t include BS $s$, $g_{su}$ denotes the channel gain between user $u$ and BS $s$, and $\rho_{su}$ denotes the offloading (link usage) index between user $u$ and BS $s$. In an offloading mechanism, $\rho_{su}$ should be 1 if user $u$ selects (is associated with) BS $s$, 0 otherwise.

In fact, the data rate model (1) implies that a Round Robin scheduler is employed by all BSs, and the time-frequency resources are equally allocated for the users associated with BSs.

As revealed in [35], the power consumption of massive MIMO MBS may not be proportional to the radiated transmit power. Based on this, we consider a more reasonable power consumption model for PBSs, which can clearly shows how the power consumption of massive MIMO MBSs scales with the number of antennas implemented at them. In such a model, total power $P_s$ is consumed by the transceiver chains, channel estimation and precoding, coding and decoding, and architectural cost [35]. Mathematically, the power consumption $P_s$ of MBS $s$ is given by

$$\hspace{1cm} P_s = \varepsilon_s p_s + \sum_{m=0}^{3} C_{m0} N^m + M \sum_{m=0}^{2} C_{m1} N^m,$$

where $\varepsilon_s$ represents the power amplifier coefficient of BS $s$; $C_{m0}$ and $C_{m1}$ are the power coefficients.

B. KEY PARAMETERS FOR PBSS

In MM-HCNs, any PBS may just be equipped with only one antenna. Under the equal resource sharing, the downlink data rate $R_{su}$ of user $u$ associated with PBS $s$ can be given by

$$\hspace{1cm} R_{su} = \frac{1}{\sum_{k \in \mathcal{U}} \rho_{sk}} \log_2 (1 + \sinr_{su}).$$

Unlike the massive MIMO MBSs, the conventional linear power consumption model should be feasible for PBSs and has been widely utilized in the existing efforts, which includes the static power and adaptive power generally. In such two types of power consumption, the former is often linear to the radiated power of PBS, and the latter has a tight relation with the power consumed by transceiver chains. Mathematically, the power consumption $P_s$ of PBS $s$ is given by

$$\hspace{1cm} P_s = \varepsilon_s p_s + \vartheta_s,$$

where $\vartheta_s$ represents the static power of PBS $s$.

C. AUXILIARY PARAMETERS

After providing the data rate and power consumption models, we may also need to introduce other auxiliary parameters. To this end, we give some definitions as follows.

Definition 1: The number of users associated with BS $s$ is definitely denoted as the load of this BS, which is given by $\varepsilon_s = \sum_{u \in \mathcal{U}} \rho_{su}$ in a mathematical manner.

Definition 2: The downlink EE $E_{su}$ of user $u$ associated with BS $s$ is denoted as a ratio of downlink data rate of user $u$ to total power consumption of BS $s$. Mathematically, it can be given by

$$\hspace{1cm} E_{su} = \frac{R_{su}}{P_s} = \frac{R_{su}}{\varepsilon_s p_s + \vartheta_s},$$

where

FIGURE 1. Massive MIMO enabled HCNs (MM-HCNs). Under the full frequency reuse, the users associated with some BS receive the co-channel interference from other BSs, especially from MBSs.
where
\[ R_{su} = \frac{K_s r_{su}}{\sum_{k \in \ell} \rho_{sk}} = \frac{K_s}{\sum_{k \in \ell} \rho_{sk}} \log_2 (1 + \kappa_s \text{SINR}_{su}), \quad (7) \]
\[ \omega_s = \left\{ \begin{array}{ll}
3 \left( C_{m0} N^m + M \sum_{m=0}^{2} C_{m1} N^m \right), & \forall s \in S_m, \\
\delta_s, & \forall s \in S_p,
\end{array} \right. \quad (8) \]
\[ \kappa_s = (M - N + 1)/N \text{ for any MBS } s, \kappa_s = 1 \text{ for any PBS } s, K_s = N \text{ for any MBS } s, K_s = 1 \text{ for any PBS } s, \quad \forall s \in S_m \text{ and } S_p \text{ represents the sets of MBSs and PBSs respectively.} \]

**Definition 3:** The supported ratio is denoted as a ratio of special users among users, where these special users refer to the ones whose downlink data rates are greater than or equal to a required threshold.

### III. OFFLOADING DESIGN

In order to achieve a tradeoff between EE and SE experiences, we will try to design two types of offloading mechanisms including MOET and PCOET.

#### A. DESIGN FOR MOET

At first, we just concentrate on the mere offloading for EE-SE tradeoff (MOET) in MM-HCNs, which jointly optimizes the EE and SE experiences under some downlink data rate constraints. In such a mechanism, a logarithmic utility function is introduced to enhance the user fairness and improve the user experience. Mathematically, the optimization problem for MOET is finally formulated as

\[ \max_{\rho} F(\rho) = \sum_{s \in S} \sum_{u \in \ell} \rho_{su} \ln \psi_{su} \]
\[ \text{s.t.} \quad \sum_{s \in S} \rho_{su} = 1, \forall u \in U, \]
\[ \sum_{s \in S} \rho_{su} R_{su} \geq \tau_u, \forall u \in U, \]
\[ \rho_{su} \in \{0, 1\}, \forall s \in S, \forall u \in U, \quad (9) \]

where \( \rho = \{\rho_{su}, \forall s \in S, \forall u \in U\} \); \( \ln x \) represents the logarithmic function with respect to \( x \) under the base \( e \approx 2.7183 \); the first constraint shows any user \( u \) can just select (be associated with) only one serving BS; the second constraint reveals that any user \( u \) needs to satisfy a minimal rate requirement \( \tau_u \); \( \psi_{su} \) is a utility of user \( u \) associated with BS \( s \) and given by

\[ \psi_{su} = R_{su}^{\alpha_u} E_{su}^{1 - \alpha_u} = R_{su}^{\alpha_u} \left[ R_{su}/P_{su} \right]^{1 - \alpha_u}. \quad (10) \]

In (10), \( \psi_{su} \) integrates two types of crucial performance indices, i.e., SE and EE. In addition, \( \alpha_u \in [0, 1] \) is a weighting parameter of user \( u \), and used for adjusting the SE-EE preference. Evidently, the solving process of (9) is to find an optimal SE experience with guaranteed data rates when \( \alpha_u = 1 \), but it tries to find an optimal EE experience under these constraints if \( \alpha_u = 0 \).

As we know, some traditional offloading mechanisms often let users select some BSs with good channel qualities, which may result in an extremely unbalanced load distribution for HCNs, especially for MM-HCNs. In other words, most users are attracted by some high-power BSs with good channel qualities, and very few users can be served by other low-power BSs. However, it is not the case for our design. To meet the data rate constraints involved in (9), some users don’t always are attracted by some BSs with the best channel qualities, and they may prefer some underloaded BSs with relatively good channel qualities. In this way, the network resources may be utilized fully and the experiences of cell-edge users should also be improved greatly.

Considering that the constraint \( \sum_{s \in S} \rho_{su} R_{su} \geq \tau_u \) is equivalent to the one \( R_{su} \geq \rho_{su} R_{su} \), and the definition of downlink data rate, we can obtain the following equivalent problem:

\[ \max_{\rho} F(\rho) = \sum_{s \in S} \sum_{u \in \ell} \rho_{su} \ln \psi_{su} \]
\[ \text{s.t.} \quad \sum_{s \in S} \rho_{su} = 1, \forall u \in U, \]
\[ r_{su} \geq \sigma_u \rho_{su} \sum_{k \in \ell} \rho_{sk}, \forall s \in S, \forall u \in U, \]
\[ \rho_{su} \in \{0, 1\}, \forall s \in S, \forall u \in U, \quad (11) \]

where \( \sigma_u = \tau_u/K_s \). Similar to the treatment on the second constraint of (11) mentioned in [43], we can rewrite the problem (11) as

\[ \max_{\rho} F(\rho) = \sum_{s \in S} \sum_{u \in \ell} \rho_{su} \ln \psi_{su} \]
\[ \text{s.t.} \quad \sum_{s \in S} \rho_{su} = 1, \forall u \in U, \]
\[ r_{su} \geq \sigma_u \sum_{k \in \ell} \rho_{sk} + \delta_u (\rho_{su} - 1), \forall s, \forall u, \]
\[ \rho_{su} \in \{0, 1\}, \forall s \in S, \forall u \in U, \quad (12) \]

where \( \delta_u = U \sigma_u \). Significantly, the second constraints of (11) and (12) should be the same if \( \rho_{su} = 1 \). In addition, these constraints are always satisfied (equivalent) if \( \rho_{su} = 0 \). In (12), the objective function \( F(\rho) \) can be expanded into

\[ F(\rho) = \sum_{s \in S} \sum_{u \in \ell} \rho_{su} \left( \chi_{su} - \ln \sum_{k \in \ell} \rho_{ku} \right), \quad (13) \]

where \( \chi_{su} = \ln (K_s r_{su}) - (1 - \alpha_u) \ln P_{su} \). By employing an auxiliary parameter \( z = \{z_s, \forall s \in S\} \), we can reformulate (12) into

\[ \max_{\rho, z} G(\rho, z) = \sum_{s \in S} \sum_{u \in \ell} \rho_{su} \chi_{su} - \sum_{u \in \ell} z_u \ln z_u \]
\[ \text{s.t.} \quad \sum_{s \in S} \rho_{su} = 1, \forall u \in U, \]
\[ r_{su} \geq \sigma_u \rho_{su} + \delta_u (\rho_{su} - 1), \forall s, \forall u, \]
\[ \sum_{u \in \ell} \rho_{su} = z_s, \forall s \in S, \quad (14) \]
\[ z_u \leq U, \forall s \in S, \forall u \in U, \]
\[ \rho_{su} \in \{0, 1\}, \forall s \in S, \forall u \in U, \]
where the fourth constraint introduced in (14) show that the load of some BS cannot go beyond the number of users distributed in networks.

When the association variable $\rho$ is relaxed to the continuous domain $[0,1]$ from discrete one $\{0,1\}$, (14) is a convex optimization problem because of its concave objective function and linear constraints. As revealed in [44], a central algorithm developed for a continuous domain can achieve almost the same performance with a distributed algorithm designed for discrete one. In view of this, we will mainly concentrate on the development of some algorithms under a discrete domain.

**Theorem 1**: By introducing the Lagrange multipliers $\nu = \{\nu_s, \forall s \in S\}$ and $\upsilon = \{\upsilon_{su}, \forall s \in S, \forall u \in U\}$ for the third and second constraints of (14) respectively, a decomposable dual form of (14) is given by

$$
\min_{\nu, \upsilon} \mathcal{H}(\nu, \upsilon) = \mathcal{I}(\nu, \upsilon) + \mathcal{J}(\nu, \upsilon),
$$

(15)

where

$$
\mathcal{I}(\nu, \upsilon) = \left\{ \begin{array}{l}
\max_{\rho} \mathcal{C}(\rho, \nu, \upsilon) \\
\text{s.t. } \sum_{s \in S} \rho_{su} = 1, \forall u \in U, \\
\rho_{su} \in \{0, 1\}, \forall s \in S, \forall u \in U,
\end{array} \right.
$$

(16)

and

$$
\mathcal{J}(\nu, \upsilon) = \left\{ \begin{array}{l}
\max_{z} \mathcal{D}(z, \nu, \upsilon) \\
\text{s.t. } z_s \leq U, \forall s \in S.
\end{array} \right.
$$

(17)

In (16) and (17), the mentioned function is listed as follows.

$$
\mathcal{C}(\rho, \nu, \upsilon) = \sum_{s \in S} \sum_{u \in U} \rho_{su} (\chi_{su} - \upsilon_{su} \delta_{su} - \nu_s),
$$

(18)

and

$$
\mathcal{D}(z, \nu, \upsilon) = \sum_{s \in S} z_s \left\{ \nu_s - \ln z_s - \sum_{u \in U} \sigma_{su} \upsilon_{su} \right\} \\
+ \sum_{s \in S} \sum_{u \in U} \upsilon_{su} (r_{su} + \delta_{su}).
$$

(19)

Proof: By introducing the Lagrange multiplier $\upsilon$ associated with auxiliary constraints and the one $\nu$ associated with downlink data rate constraints, the partial Lagrange function of (14) is given by

$$
\mathcal{L}(\rho, z, \nu, \upsilon) = \sum_{s \in S} \sum_{u \in U} \upsilon_{su} \left\{ r_{su} - \sigma_{su} z_s - \delta_{su} (\rho_{su} - 1) \right\} \\
+ \sum_{s \in S} \sum_{u \in U} \rho_{su} \chi_{su} - \sum_{s \in S} z_s \ln z_s \\
+ \sum_{s \in S} \upsilon_s \left( z_s - \sum_{u \in U} \rho_{su} \right) \\
= \mathcal{C}(\rho, \nu, \upsilon, \upsilon) + \mathcal{D}(z, \nu, \upsilon).
$$

(20)

Then, the corresponding dual function can be given by

$$
\mathcal{H}(\nu, \upsilon) = \left\{ \begin{array}{l}
\max_{\rho, z} \mathcal{L}(\rho, z, \nu, \upsilon) \\
\text{s.t. } \sum_{s \in S} \rho_{su} = 1, \forall u \in U, \\
\rho_{su} \in \{0, 1\}, \forall s \in S, \forall u \in U,
\end{array} \right.
$$

(21)

and the dual problem can be written as

$$
\min_{\nu, \upsilon} \mathcal{H}(\nu, \upsilon).
$$

(22)

Since the optimization of $\rho$ and $z$ in (21) is decoupling, they can be separately tackled. Based on this, the dual problem (22) can be decomposed into two subproblems, which optimize the Lagrange multipliers under optimal $\rho$ and $z$ respectively. That is to say, $\mathcal{H}(\nu, \upsilon)$ can be decomposed into two parts including $\mathcal{I}(\nu, \upsilon)$ and $\mathcal{J}(\nu, \upsilon)$. $\square$

It is evident that the primal problem with respect to $\rho$ in (16) can be equivalent to

$$
b = \arg \max_{\upsilon \in \mathcal{S}} \{ \chi_{su} - \upsilon_{su} \delta_{su} - \nu_s \}, \forall u \in U.
$$

(23)

That is to say, any user $u$ finds some BS $b$ that its utility $\{\chi_{bu} - \upsilon_{bu} \delta_{bu} - \upsilon_b\}$ is the highest among all possible BSs, and then accesses it.

To solve the primal problem with respect to $z$ in (17), we employ an extreme value principle $\partial \mathcal{D}/\partial z_s = 0$, and then have

$$
z_s^{t+1} = \min \{ \nu_s^{t+1} - 1 - \sum_{u \in U} \sigma_{su} z_s^{t+1}, U \}, \forall s \in S.
$$

(24)

where $t$ represents an iteration index.

By following a subgradient method [45], the multiplier $\nu_s$ for BS $s$ can be updated by

$$
\nu_s^{t+1} = \nu_s^t - \xi_1 \left( z_s^t - \sum_{u \in U} \rho_{su}^t \right),
$$

(25)

and the multiplier $\upsilon_{su}$ for BS $s$ and user $u$ can be updated by

$$
\upsilon_{su}^{t+1} = \left[ \upsilon_{su}^t - \xi_2 \left( r_{su} - \sigma_{su} z_s^t - \delta_{su} (\rho_{su} - 1) \right) \right]^+, \hspace{1cm}
$$

(26)

where $\xi_1$ and $\xi_2$ are the sufficiently small fixed step sizes; $[x]^+$ is equal to the maximum of $x$ and 0.

Now, we can give a whole insight on the solving process of offloading problem (14), which is summarized in Algorithm MOET. In it, 1 represents a vector or matrix whose elements are equal to 1.

**Algorithm 1: MOET**

1: Initialization: $t_1 = 0, \nu^{t_1} = 1$ and $\upsilon^{t_1} = 1$.
2: Repeat:
3: Initialize the offloading state: let all elements of $\rho$ be 0.
4: Perform the cell selection according to the rule (23).
5: Update the auxiliary factor $z_1^{t+1}$ using (24).
6: Update the multiplier $\upsilon_{su}^{t+1}$ using (25).
7: Update the iteration index: $t_1 = t_1 + 1$.
8: Until $G(p, z)$ converges or $t_1$ reaches $T_1$ iterations.

Next, we will concentrate on the convergence analysis for Algorithm MOET.

**Theorem 2**: After a few iterations, Algorithm MOET finally converges to the optimum of (15).

Proof: The first-order partial derivatives of $\mathcal{H}(\nu, \upsilon)$ for $s$ and $\upsilon$ are calculated by

$$
\frac{\partial \mathcal{H}(\nu, \upsilon)}{\partial \nu_s} = z_s (\upsilon_s, \upsilon_s, ...) - \sum_{u \in U} \rho_{su} (\nu_s, \upsilon_s),
$$

(27)
\[
\frac{\partial \mathcal{H}(\nu, \nu)}{\partial v_{su}} = r_{su} - \sigma_{su} z_{su}(\nu_s, \nu_{sc}) - \delta_{su} \rho_{su}(\nu_s, \nu_{sc}) + \delta_{su},
\]

where \( v_s \) denotes the \( s \)-th row of \( v \).

In a real network system, the number of users is always limited. In view of this, \( \sum_{u \in \ell} \rho_{su}(\nu_s, \nu_{sc}) \) should be bounded. In addition, we know \( z_s(\nu_s, \nu_{sc}) \) should also be bounded. At last, we can conclude that all subgradients of \( \mathcal{H}(\nu, \nu) \) are bounded.

\[
\sup_t \| \frac{\partial \mathcal{H}(\nu, \nu)}{\partial v_s} \| \leq \eta, \forall s \in S,
\]

\[
\sup_t \| \frac{\partial \mathcal{H}(\nu, \nu)}{\partial v_s} \| \leq \eta, \forall s \in S, \forall u \in \mathcal{U},
\]

where \( |X| \) denotes a 2-norm of \( X \); \( \sup \) represents the maximal estimate of \( x \) among all iterations; \( \eta \) is a constant. Evidently, we can prove Theorem 1 by employing the convergence proof in [45] since the dual problem (15) meets its necessary conditions.

**B. DESIGN FOR PCOET**

To reduce the power consumption under some strict data rate constraints, a power coordination measure is integrated into the problem (14). Thereby we have

\[
\max_{\rho, p} H(p, \rho) = \sum_{s \in S} \sum_{u \in \mathcal{U}} \rho_{su} \ln \psi_{su}
\]

\[
\text{s.t. } \rho_{su} = 1, \forall u \in \mathcal{U},
\]

\[
\sum_{s \in S} \rho_{su} R_{su} \geq \tau_u, \forall u \in \mathcal{U},
\]

\[
0 \leq p_s \leq \bar{p}_s, \forall s \in S,
\]

\[
\rho_{su} \in \{0, 1\}, \forall s \in S, \forall u \in \mathcal{U},
\]

where \( p = \{p_s, \forall s \in S\} \); \( \bar{p}_s \) represents the maximal allowed transmit power of BS \( s \).

We can easily find that the problem (31) is hard to tackle since it is in a coupling and mixed-integer form. As a common method, the alternative optimization is widely advocated to solve this type of problems. In view of this, we will utilize it to handle the formulated problem (31).

When the transmit power \( p \) is fixed, the problem (31) can be easily simplified into (14) that can be solved by Algorithm MOET. When the offloading index \( \rho \) is given, we can easily know that the problem (31) just focuses on the power coordination, and it can be further converted into

\[
\max_{\rho} I(p) = \sum_{s \in S} \sum_{u \in \mathcal{U}} \rho_{su} \ln \psi_{su}
\]

\[
\text{s.t. } R_{su}(p) \geq \rho_{su} R_{su}, \forall s \in S, \forall u \in \mathcal{U},
\]

\[
0 \leq p_s \leq \bar{p}_s, \forall s \in S.
\]

**Theorem 3:** Under the parameters \( h_{su} = \sum_{u \in \mathcal{U}} (1 - \alpha_u) \rho_{su} \) and \( \gamma_{su} = \kappa_s^{-1} (2^{\rho_{su} \sigma_{su} \sum_{k \in \ell} \rho_{sk}} - 1) \) for any \( s \) and \( u \), the upper bound of (32) is given by

\[
\max_{\rho} I(p) = \sum_{s \in S} \sum_{u \in \mathcal{U}} \rho_{su} \ln \psi_{su}
\]

\[
\text{s.t. } \text{SINR}_{su}(p) \geq \gamma_{su}, \forall s \in S, \forall u \in \mathcal{U},
\]

\[
0 \leq p_s \leq \bar{p}_s, \forall s \in S.
\]

Proof: At first, we deduce the upper bound of \( I(p) \) in (32). Specifically, we have

\[
J(p) = \sum_{s \in S} \sum_{u \in \mathcal{U}} \rho_{su} \ln \psi_{su} - \sum_{s \in S} h_s \ln p_s
\]

\[
\text{s.t. } \text{SINR}_{su}(p) \geq \gamma_{su}, \forall s \in S, \forall u \in \mathcal{U},
\]

\[
0 \leq p_s \leq \bar{p}_s, \forall s \in S.
\]

After some simple operations on the first constraint of (36), we can easily obtain the problem (33).
where
\[
\Gamma_{su}(q) = \frac{e^{\nu_s}g_{su}}{\sum_{i \in S_u} e^{\nu_i}g_{iu} + \Theta_s}, \quad \forall s \in S, \forall u \in U. \tag{39}
\]

After some necessary operations, we can easily find that (38) is a convex optimization problem because of concave objective function [46] and constraints, and linear constraints.

By introducing \( \lambda = \{\lambda_{su}, \forall s \in S, \forall u \in U\} \) for the first constraint in (38), we can attain the partial Lagrange function as follows.
\[
L(q, \lambda) = \sum_{s \in S} \sum_{u \in U} \lambda_{su}(q) - b_{su}q_s + \sum_{s \in S} \lambda_{su}(\ln \Gamma_{su}(q) - \ln \gamma_{su}). \tag{40}
\]

According to the extreme value principle \( \partial L(q, \lambda)/\partial q_s = 0 \), we can achieve
\[
e^{\nu_s} = \frac{\sum_{u \in U} (\lambda_{su} + b_{su}) - b_s}{\sum_{n \in S_u} \sum_{u \in U} A_{nu}g_{su}}, \forall s \in S, \tag{41}
\]

where
\[
A_{nu} = \frac{\lambda_{nu} + b_{nu}}{\sum_{i \in S_u} e^{\nu_i}g_{iu} + \Theta_n}, \forall n \in S, \forall u \in U. \tag{42}
\]

Under the box-constrained projection [46], we rewrite (41) in an equivalent form, i.e.,
\[
p_{s}^{t+1} = \frac{\sum_{u \in U} (\lambda_{su} + b_{su}) - b_s}{\sum_{n \in S_u} \sum_{u \in U} B_{nu}(q^t)g_{su}} = \Upsilon_s(q^t), \forall s. \tag{43}
\]

where
\[
B_{nu}(q^t) = \frac{\lambda_{nu} + b_{nu}}{\sum_{i \in S_u} p^t_{iu}g_{iu} + \Theta_n}, \forall n \in S, \forall u \in U. \tag{44}
\]

In general, the finally transmit power can be updated by
\[
p_{s}^{t+1} = \phi_s(q^t) = \left[ \Upsilon_s(q^t) \right]^{p_s}, \forall n \in S, \tag{45}
\]

where \( [x]^\Delta = \min \{\max \{x, \Delta\}, \nabla\} \) means \( x \) takes a value from the closed interval \([\Delta, \nabla]\); although \( \alpha \) should be 0, we let it be \( 10^{-30} \) to avoid “\( \ln(0) \)”. According to subgradient method, the multiplier \( \lambda_{su} \) for any \( s \) and \( u \) can be updated by
\[
\lambda_{su}^{t+1} = \lambda_{su}^t - \xi_3(\ln \Gamma_{su}(q^t) - \ln \gamma_{su})^+ \tag{46}
\]

where \( \xi_3 \) represents a sufficiently small fixed stepszise.

After a box-constrained projection, the update rule (46) can be rewritten as
\[
\lambda_{su}^{t+1} = \left[ \lambda_{su}^t - \xi_3(\ln \text{SINR}_{su}(q^t) - \ln \gamma_{su}) \right]^+. \tag{47}
\]

Now, we can give a detailed procedure to solve the problem (37), which is described in Algorithm PCOET. In such an algorithm, the outer layer alternately optimizes \( \rho \) and \( p \): the steps 4-9 find the optimal \( p \) and establish an inner (power coordination) loop; the step 3 achieves the optimal \( \rho \) and also establishes an inner (offloading) loop; the step 7 is used for guaranteeing the boundness of function (40), which is similar to the operation in [46].

**Algorithm 2: PCOET**

1. Initialization: \( t_2 = 0, t_3 = 0, \) and \( p(t_2) = \{p_\|, \forall s \in S\} \).
2. **Repeat (Outer Loop):**
   3. Perform the cell selection using Algorithm MOET.
   4. **Repeat (Inner Loop):**
      5. Update the transmit power \( p(t_2+1) \) using (45).
      6. Update the multiplier \( \lambda(t_2+1) \) using (47).
      7. Normalize the multiplier \( \lambda(t_2+1) \) so that \( 1 + \lambda(t_2+1)=1 \).
      8. Update the iteration index: \( t_2 = t_2 + 1 \).
   9. Until \( J(p) \) converges or \( t_2 \) reaches \( T_2 \) iterations.
10. Update the iteration index: \( t_3 = t_3 + 1 \).
11. Until \( H(p, p) \) converges or \( t_3 \) reaches \( T_3 \) iterations.

Similar to most efforts, the convergence for a whole procedure of Algorithm PCOET cannot be proven theoretically. However, we can give some convergence proofs for offloading and power coordination subprocedures. Considering that the former has been proven in Theorem 2, we just need to prove the latter. To this end, it is required to prove that \( \phi_s(p) \) is a two-sided scalable (2.s.s.) function with respect to \( p \) for any \( s \). Then, we can easily prove the convergence of power coordination loop using some results of 2.s.s. function used for updating the power in [15], [47].

**Theorem 4:** \( \phi_s(p) \) is a 2.s.s. function with respect to \( p \) for any \( s \).

Proof: Before providing some proofs for 2.s.s. \( \phi_s(p) \), we may need to prove that \( \Upsilon_s(p) \), its upper and lower bounds are 2.s.s. functions with respect to \( p \) for any \( s \). Assume that \( (1/c) p \leq q \leq cp \) for any \( c > 1 \), where \( x \leq y \) if \( x \leq y_0 \) for any \( s \). Then, we can easily deduce
\[
(1/c)B_{su}(q) \leq B_{su}(q) \leq cB_{su}(p), \forall s, \forall u, \tag{48}
\]
and thus achieve
\[
(1/c) \Upsilon_s(q) \leq \Upsilon_s(q) \leq c \Upsilon_s(p), \forall s. \tag{49}
\]

According to the definition of a 2.s.s. function in [47], we know that \( \Upsilon_s(p) \) is a 2.s.s. function with respect to \( p \) for any \( s \). Similarly, we can easily prove that the upper and lower bounds of transmit power in (45) are also 2.s.s.

In conclusion, \( \phi_s(p) \) is a 2.s.s. function with respect to \( p \) for any \( s \).

**Theorem 5:** The power coordination loop including steps 4-8 in Algorithm PCOET converges to a unique fixed point.

Proof: As revealed in Theorem 4, we know that \( \phi_s(p) \) is a 2.s.s. function with respect to \( p \) for any \( s \). According to the results of a 2.s.s. function used for updating the transmit power in [47], we can easily prove the convergence of power coordination loop in Algorithm PCOET.

**C. Complexity Analysis**

In this section, we concentrate on the computation complexities of proposed algorithms.

**0 The Complexity of Algorithm MOET.** In Algorithm MOET, the computation complexity may be mainly dependent on steps 3-7. It is evident that any of these steps has a complexity of \( O(U S) \). After \( T_1 \) iterations, Algorithm MOET should have a complexity of \( O(U S T_1) \).

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/.
TABLE 1. SIMULATION PARAMETERS

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>MBS-MBS distance</td>
<td>1000 m</td>
<td>M</td>
<td>10 [35]</td>
</tr>
<tr>
<td>System bandwidth</td>
<td>10 MHz</td>
<td>$p_u$ for MBS s</td>
<td>46 dBm [15]</td>
</tr>
<tr>
<td>$\varepsilon_s$ for BS s</td>
<td>10/3 [38]</td>
<td>$p_u$ for PBS s</td>
<td>30 dBm</td>
</tr>
</tbody>
</table>

IV. PERFORMANCE EVALUATION

In MM-HCNs, we consider the power coefficients $C_{00} = 4$, $C_{10} = 4.8$, $C_{20} = 0$, $C_{30} = 2.08 \times 10^{-8}$, $C_{01} = 1$, $C_{11} = 9.5 \times 10^{-8}$ and $C_{21} = 6.25 \times 10^{-8}$ [35]. Without loss of generality, we assume that $\tau_u = 1$ bit/s/Hz for any user $u$. Moreover, the detailed settings of other parameters can be found in TABLE 1, where $\ell_{su}$ represents the distance (in km) between BS s and user u.

To highlight the characteristics and effectiveness of designed offloading mechanisms, we introduce another existing one (mere offloading, MO) [20] for comparison. Without loss of generality, in the simulation, we take account of $\alpha_u = w$ for any user $u$ in the presented mechanisms. Considering our association rules mainly concentrate on a tradeoff between SE and EE experiences, we will investigate the impacts of different weighting parameters and numbers of antennas at MBSs on the cumulative distribution functions (CDFs) of data rates (SEs), the CDFs of EEs, the 5th percentile average throughput (SE) and the 5th percentile average EE. Significantly, the 5th percentile throughput represents the average of the lowest 5% data rates of associated users, and it can also be regarded as the average of data rates of cell-edge users; the 5th percentile EE represents the average of the lowest 5% EEs of associated users. At last, we investigate the convergence of our offloading mechanisms by numerical simulation.

Under $w = 0.5$, Fig. 2 shows the impacts of $M$ (the number of antennas at MBS) on the CDFs of data rates (SEs) for different offloading mechanisms. Through the power coordination, PCOET mitigates the network interference and thus has fewer low-rate (cell-edge) users than MOET. As we know, MO just concentrates on the enhancement of users’ SE experiences, but others try to find a tradeoff between SE and EE experiences. Therefore, MO may have fewer low-rate users than MOET. However, it may have more low-rate users than PCOET due to the interference mitigation caused by power coordination in the latter. In addition, we can also find that the rate (SE) experience of users can be improved if $M$ is increased. That’s because the increased number of antennas lets the users be associated with massive MIMO MBSs having higher data rates. In fact, we can easily conclude this point from the formula (1).

Under $w = 0.5$, Fig. 3 investigates the impacts of $M$ on the CDFs of EEs for different offloading mechanisms. Among all offloading mechanisms, MO may have the most low-EE users since it doesn’t pay any attention to users’ EE experiences but others do it. Under the power coordination, the network interference is mitigated and the system power consumption is reduced at the same time. Therefore, PCOET may have fewer low-EE users than MOET. According to the formula (8), we know that the increased number of antennas at MBS may result in the increased circuit power consumption. Thus, the users’ EE experiences may degrade with increased $M$.

Under $M = 100$, Fig. 4 shows the impacts of weighting...
parameter \( w \) on the CDFs of data rates for different offloading mechanisms. As revealed in Fig. 2, the power coordination mitigates network interference. Thus, PCOET may have fewer low-rate users than MOET in Fig. 4. According to the association rules, we can easily find that MOET can be converted into MO with rate constraints if \( w = 1 \). Under some rate constraints, MOET with \( w = 1 \) may achieve a relatively better rate experience than MO. As illustrated in Fig. 4, the rate experience of users may be gradually improved with increased \( w \). According to the formula (10), it is easy to know that both PCOET and MOET are increasingly keen to optimize the rate experience if \( w \) increases.

Under \( M = 100 \), Fig. 5 investigates the impacts of \( w \) on the CDFs of EEs for different offloading mechanisms. By employing a power coordination technique to reduce the network interference and power consumption, PCOET may have fewer low-EE users than MOET. Unlike the results in Fig. 4, since MO has a better rate experience than MOET, the former may achieve a relatively better EE experience than the latter with \( w = 1 \). As revealed in the formula (10), we know that both PCOET and MOET should be increasingly keen to optimize the EE experience if \( w \) decreases. Thus, the users’ EE experiences are gradually enhanced with decreased \( w \).

Fig. 6 investigates the impacts of \( w \) and \( M \) on the 5th percentile average throughput (SE) for different offloading mechanisms. Since PCOET has fewer low-rate users than MOET, the former may have higher 5th percentile average throughput than the latter. Similarly, MOET with \( w = 1 \) has higher 5th percentile average throughput than MO. According to our association rules, we know that both PCOET and MOET are weighted in favour of SE optimization when \( w \) increases. Thus, the 5th percentile average throughput gradually increases with increased \( w \) in PCOET and MOET. When \( w \) takes a relatively high value, the 5th percentile average throughput in our offloading mechanisms may increase with increased \( M \). The reason for this is that many users may select MBSs with high data rates if \( w \) is relatively high. When \( w \) utilizes a relatively low value, the 5th percentile average throughput in our offloading mechanisms may initially increase with increased \( M \), but then it may be stable or even decreasing. That' because a smaller \( w \) may let more users select PBSs with relatively low data rates.

Fig. 7 shows the impacts of \( w \) and \( M \) on the 5th percentile average EE for different offloading mechanisms. Under the power coordination, PCOET achieves a higher 5th percentile average EE than MOET. As we know, the designed offloading mechanisms should pay more attention to optimizing users’ EE experiences when \( w \) takes a smaller value. Therefore, the 5th percentile EE should increase with decreased \( w \). In PCOET, MOET and MO, more users are attracted by massive MIMO MBSs. An increased \( M \) may result in the increased power consumption and thus a decreased EE for users associated with MBSs. That means the 5th percentile EEs in all offloading mechanisms almost decrease with increased \( M \) in
shown in Fig. 8 (b) and Fig. 8 (d). In the power coordination also converge after a relatively few iterations, which are Fig. 8 (c). In addition, other loops in Algorithm PCOET exist.

We cannot prove the convergence of outer coordination loop in Algorithm PCOET. Similar to most existing works, we cannot prove the convergence of offloading loop, outer loop and power coordination loop in Algorithm MOET; Fig. 8 (b), Fig. 8 (c) and Fig. 8 (d) show the convergence of designed Algorithms.

Specifically, Fig. 8 (a) illustrates the convergence of Algorithm MOET; Fig. 8 (b), Fig. 8 (c) and Fig. 8 (d) show the convergence of offloading loop, outer loop and power coordination loop in Algorithm PCOET. Similar to most existing works, we cannot prove the convergence of outer loop in Algorithm PCOET, but we find that it converges after very few iterations in the simulation, which is illustrated in Fig. 8 (c). In addition, other loops in Algorithm PCOET also converge after a relatively few iterations, which are shown in Fig. 8 (b) and Fig. 8 (d). In the power coordination loop of Algorithm PCOET, it is easy to find that \( J(p) \) has no significant fluctuation at each iteration. The reason for this may be that the power coordination loop of Algorithm PCOET gradually converges after a few alternate iterations.

V. CONCLUSION

In this paper, we design two types of offloading mechanisms (including PCOET and MOET) to achieve a tradeoff between EE and SE experiences for massive MIMO enabled HCNs. Such two mechanisms have a key difference on whether or not a power coordination technique is involved, and they are finally formulated as network-wide utility maximization problems. As for these problems, we design a distributed algorithm and a two-layer iterative algorithm for mechanisms PCOET and MOET respectively. Then, we give some corresponding computation complexity and convergence analyses for the designed algorithms. The simulation results show that the designed mechanisms can achieve a tradeoff between EE and SE experiences by properly adjusting users’ weighting parameters. Future works can include the resource partitioning, ultra-dense networks and so on.

REFERENCES


---


---