Interference Management and User Association for Nested Array-Based Massive MIMO HetNets

Mingjie Feng, *Student Member, IEEE*, and Shiwen Mao, *Senior Member, IEEE*

***Abstract*—The nested array, implemented by nonuniform antenna placement, is an effective approach to achieve $O(N^2)$ degrees of freedom (DOF) with an antenna array of N antennas. Such DOF refers to the number of directions of incoming signals that can be resolved. With the increased number of DOF, an important application of nested array is to nullify interference signals from multiple directions. In this paper, we apply nested array in a massive multiple input multiple output (MIMO) heterogeneous network (HetNet) for interference management. With a nested array, a base station (BS) can nullify a certain number of interference signals based on their directions. Then, a key design issue is how to select the set of interference sources to be nullified at each BS. As the DOF of each BS is used to resolve both desired signals and interference, the number of interference signals that can be nullified depends on the number of users served by the BS. Thus, user association is another factor that impacts the system performance and should be jointly considered with interference nulling. We formulate the joint interference nulling scheduling and user association problem as an integer programming problem, aiming to maximize the sum rate of all users subject to BS DOF constraints. We first investigate the case of interference nulling with a given user association, and propose a scheme to solve a relaxed problem as well as derive a performance upper bound. Then, we propose a distributed joint interference nulling and user association scheme based on a poly matching between users and BSs. Simulation results show that the proposed schemes effectively improve the sum rate and achieves a near optimal performance.**

***Index Terms*—5G wireless, heterogeneous networks (HetNet), interference nulling, massive MIMO, nested array.**

I. Introduction

MASSIVE *MIMO* (Multiple Input Multiple Output) and *small cell* are recognized as two key technologies for 5G wireless systems due to their great potential to enhance network capacity [1], [2]. In a massive MIMO system, the base station (BS) is equipped with more than 100 antennas and serves multiple users with the same spectrum band [3]. With aggressive spatial multiplexing, a massive MIMO can dramatically improve both energy and spectral efficiency compared to traditional wireless systems [4]–[6]. On the other hand, small cell deployment achieves high signal to noise ratio (SNR) and high spectrum spatial reuse due to the short transmission distance and small coverage area. As a result, a heterogeneous network (HetNet) with small cells can significantly boost network capacity compared to the traditional macrocell network.

Due to these benefits, massive MIMO HetNet, which integrates these two techniques, has drawn considerable attention [7]–[11], [13]–[16]. A massive MIMO HetNet consists of a macrocell BS (MBS) and multiple small cell BSs (SBS), the MBS is equipped with a large number of antennas. Due to spectrum scarcity in cellular networks, small cells are expected to share the same spectrum band with the macrocell, resulting in cross-tier interference. While interference management in a regular HetNet mainly focuses on resource allocation in the time-frequency domain [17], the spatial characteristics of massive MIMO can be exploited to mitigate interference in massive MIMO HetNets. In [18], a spatial blanking scheme was proposed in which the transmission energy of the MBS is focused on certain directions that do not cause interference to small cells. In [7], a reversed time division duplex (RTDD) architecture was introduced. Since the channels between the MBS and SBSs are quasi-static, the MBS can carry out zero-forcing beamforming based on the estimated channel covariance. In [9], coordinated transmissions are assumed between the BSs, so that each user receives signals from both the MBS and SBSs. Through coordinated beamforming, the mutual interference can be minimized.

Interference management in MIMO-based networks in most existing works are performed with baseband processing in the digital domain, which requires channel state information (CSI) between the interfering transceivers. However, in a massive MIMO HetNet with dense small cell deployment, acquiring the CSI of interfering links is difficult and causes a large overhead, due to the large number of antennas at the MBS and the large number of SBSs. To overcome this drawback, an efficient approach is to mitigate inter-cell interference with antenna array processing in the analog domain and deal with intra-cell interference with baseband processing in the digital domain. With antenna array processing, the directions of arrival (DoA) of interfering sources can be estimated. Then, a beamforming with respect to different directions can be applied to nullify the interference from certain directions. However, unlike typical application scenarios where DoA estimation is based on the line of sight (LOS) component, e.g., in a radar system, the DoA estimation in a wireless network is highly challenging due to

Manuscript received March 30, 2017; revised July 11, 2017; accepted August 13, 2017. Date of publication August 18, 2017; date of current version January 15, 2018. This work was supported in part by the U.S. National Science Foundation under Grant CNS-1320664, and in part by the Wireless Engineering Research and Engineering Center at Auburn University. This paper was presented in part at the IEEE Global Communications Conference, Washington, DC, USA, December 2016. The review of this paper was coordinated by Dr. Lian Zhao. *(Corresponding author: Shiwen Mao.)*

The authors are with the Department of Electrical and Computer Engineering, Auburn University, Auburn, AL 36849-5201 USA (e-mail: mzf0022@auburn.edu; smao@ieee.org).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TVT.2017.2741900

the rich scattering and multipath effect. As the signals of a user received by a BS come from multiple directions, the antenna array needs to resolve a large number of DoAs. However, with traditional antenna array configurations, such as uniform linear array, the number of directions that can be resolved is $O(N)$, which is insufficient for a massive MIMO HetNet. To this end, we employ a second order antenna array processing technique called *nested array* [19] for inter-cell interference management in massive MIMO HetNets. Based on the concept of *difference co-array*, a nested array is implemented by nonuniform antenna placement to achieve $O(N^2)$ degrees of freedom (DoF) with only N antennas. An advantage of nested array is that the DoA estimation is performed with a *passive sensing* pattern, i.e., the antenna array does not need to send out signals for detection. Due to the significantly increased DoF and its easy implementation, we apply nested array at the SBSs of a massive MIMO HetNet to identify a number of $O(N^2)$ directions of incoming signals, which include both desired signals and interference. Then, the signals from different directions can be filtered out, such that the desired signals remain while the interfering signals are nullified. Note that, the nested array is not applied at an MBS with massive MIMO, since the number of antennas is already sufficiently large.

Due to the multipath effect, the signals of a user received by an SBS come from multiple directions. Then, a certain number of DoFs are required to estimate the DoAs of these signals. Hence, both the channel gain and the number of multipaths of each user should be taken into consideration to enhance the system performance. Since both service provisioning and interference nulling require the use of DoF, there is a tradeoff between these two objectives. When an SBS serves more nearby users, the sum rate would first increase. However, the available DoF for interference nulling would be reduced, resulting in degraded signal to interference and noise ratio (SINR). Thus, given the DoFs of each SBS, a key design problem is to select the set of users to be served and the set of users for interference nulling, to optimize the system performance. In this paper, we consider user association and interference nulling scheduling in a nested array-based massive MIMO HetNet to fully harness the benefit of interference mitigation brought about by the nested array.

Specifically, we formulate the user association and interference nulling problem as an integer programming problem with the objective of maximizing the sum rate, subject to constraints on the DoF of each BS. We first consider interference nulling schedule with a given user association. The resulting integer programming problem has a nonlinear and nonconvex objective function. We propose a series of approximations that transform the original problem into an integer programming problem with a linear objective function. The optimal solution to the relaxed problem can be obtained with a *cutting plane* approach. Moreover, we find that when each BS only receives a fixed number of strongest signals from each user, the constraint matrix becomes *unimodular* and the integer programming problem will become equivalent to an linear programming (LP) problem obtained by relaxing the integer constraints. Thus an optimal solution can be obtained with an LP solver. We then consider *joint interference nulling schedule and user association*. Due to the highly complicated structure of the original problem, we propose a distributed scheme based on a poly matching between users and BSs. We show that the matching process converges and the outcome yields a stable matching that is optimal for each user and BS. The proposed schemes are compared with several benchmark schemes through simulations. The results show that near optimal performance can be achieved.

The remainder of this paper is organized as follows. The nested array based interference nulling method is introduced in Section II. The system model and problem formulation are presented in Section III. The solution for interference nulling with a given user association is presented in Section IV. The distributed algorithm for joint interference nulling scheduling and user association is presented in Section V. The simulation results are discussed in Section VI. We present related works in Section VII and conclude this paper in Section VIII.

II. Preliminaries

A. Signal Model of Difference Co-Array

Consider an antenna array with N antennas, the $N \times 1$ steering vector corresponding to direction θ is denoted as $\mathbf{a}(\theta)$. Let d_i be the position of the ith antenna and λ the carrier wavelength. The ith element of $\mathbf{a}(\theta)$ is $e^{j(2\pi/\lambda)d_i \sin\theta}$. Suppose D narrowband sources from directions $\{\theta_i, i = 1, 2, \ldots, D\}$ impinge upon the antenna array with powers $\{\sigma_i^2, i = 1, 2, \ldots, D\}$. The received signal is given by

$$\mathbf{r}[m] = \mathbf{F}\boldsymbol{\gamma}[m] + \mathbf{n}[m],\ m = 1, 2, \ldots, N, \tag{1}$$

where $\boldsymbol{\gamma}[m]_{D\times 1} = [\gamma_1[m], \gamma_2[m], \ldots, \gamma_D[m]]^T$ is the source signal vector, $\mathbf{F} = [\mathbf{f}(\theta_1), \mathbf{f}(\theta_1), \ldots, \mathbf{f}(\theta_{\mathbf{D}})]$ is the array manifold matrix, and $\mathbf{n}[m]$ is the white noise vector with power σ_0^2. Assuming the sources are temporally uncorrelated, hence the autocorrelation matrix of $\boldsymbol{\gamma}[m]$ is diagonal. The autocorrelation matrix of the received signal is given by [19]

$$\begin{aligned}\boldsymbol{\Theta}_{\mathbf{rr}} &= \mathbb{E}\left[\mathbf{r}\mathbf{r}^H\right] = \mathbf{F}\boldsymbol{\Theta}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}\mathbf{F}^H + \sigma_0^2\mathbf{I} \\ &= \mathbf{F}\begin{pmatrix}\sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \ddots & \\ & & & \sigma_D^2\end{pmatrix}\mathbf{F}^H + \sigma_0^2\mathbf{I}.\end{aligned} \tag{2}$$

We next vectorize $\boldsymbol{\Theta}_{\mathbf{rr}}$ and obtain the following vector [19].

$$\begin{aligned}\mathbf{z} &= \operatorname{vec}(\boldsymbol{\Theta}_{\mathbf{rr}}) = \operatorname{vec}\left[\sum_{i=1}^{D}\sigma_i^2\left(\mathbf{f}(\theta_i)\mathbf{f}^H(\theta_i)\right)\right] + \sigma_0^2\vec{\mathbf{1}} \\ &= (\mathbf{F}^* \odot \mathbf{F})\mathbf{p} + \sigma_0^2\vec{\mathbf{1}},\end{aligned} \tag{3}$$

where $\mathbf{p} = [\sigma_1^2, \sigma_2^2, \ldots, \sigma_D^2]^T$ is the power vector of the D sources, $\vec{\mathbf{1}} = [\mathbf{e}_1^T, \mathbf{e}_2^T, \ldots, \mathbf{e}_N^T]^T$, and $\mathbf{e}_i$ is a column vector with 1 at the ith position and 0 at all other positions. Comparing (3) with (1), we find that $\mathbf{z}$ can be regarded as a signal received at an array with a manifold matrix given as $\mathbf{F}^* \odot \mathbf{F}$, where $\odot$ denotes the Khatri-Rao (KR) product. The corresponding source signal is $\mathbf{p}$ and the noise vector is given as $\sigma_0^2\vec{\mathbf{1}}$. Analyzing the manifold matrix $\mathbf{F}^* \odot \mathbf{F}$, we find that the distinct rows of

$\mathbf{F}^* \odot \mathbf{F}$ behave like the manifold of an array with antenna positions given by distinct values in the set $\{\vec{\varepsilon_i} - \vec{\varepsilon_j}, 1 \le i, j \le N\}$, where $\vec{\varepsilon_i}$ is the position vector of the original array. The new array is the difference co-array of the original array [20].

In a difference co-array with antenna positions given in the set $\{\vec{\varepsilon_i} - \vec{\varepsilon_j}\}$, for all $i, j = 1, 2, \ldots, N$, it is easy to see that the number of elements in this set is $N(N-1)+1$. Thus, given the original N-antenna array, the maximum DoFs of a difference co-array is

$$\mathrm{DOF}_{\max} = N(N-1). \tag{4}$$

We thus conclude that $O(N^2)$ DoFs can be achieved with N antennas by exploiting the second order statistics of the received signal [19].

Based on the difference co-array framework, the nested array was proposed in [19] as an effective solution to the problem of resolving more sources than antenna elements. Nested array is characterized by a *non-uniform antenna array placement* and second order statistic processing of the received signal. According to (3), the difference co-array of a nested array has $O(N^2)$ antenna elements, and thus a nested array can achieve $O(N^2)$ DoFs. Compared to existing methods on increasing DoFs, the nested array is easier to implement with reduced overhead and can be applied to more general scenarios. In addition, the nested array operates in a passive sensing pattern, which only needs to receive source signals. These favorable features make nested array suitable to applications in cellular networks. The implementation and setup process of a nested array are described in [19].

B. Interference Nulling with Nested Array

An important application of a nested array is interference nulling. Let $\mathbf{z} = (\mathbf{F}^* \odot \mathbf{F})\mathbf{p} + \sigma_0^2 \vec{\mathbf{1}}$ be the equivalent received signal at the difference co-array of the nested array of an SBS. Suppose a beamforming with weight vector $\mathbf{w}$. Then, the resulting signal is given by

$$r' = \mathbf{w}^H \mathbf{z} = \sum_{i=1}^{D} \mathbf{w}^H \left(\mathbf{f}^*(\theta_i) \otimes \mathbf{f}(\theta_i)\right) \sigma_i^2 + \sigma_0^2 \mathbf{w}^H \vec{\mathbf{1}}, \tag{5}$$

where $\otimes$ denotes the the Kronecker product. In (5), r' can be regarded as a weighted sum of $\sigma_i^2, i = 1, 2, \ldots, D$, and σ_0^2 with wights given as $\mathbf{w}^H(\mathbf{f}^*(\theta_i) \otimes \mathbf{f}(\theta_i))$ and $\mathbf{w}^H \vec{\mathbf{1}}$, respectively. Define the new beam pattern as

$$B(\theta_i) = \mathbf{w}^H \left(\mathbf{f}^*(\theta_i) \otimes \mathbf{f}(\theta_i)\right). \tag{6}$$

Thus, the powers of sources from different directions get spatially filtered by $B(\theta_i), i = 1, 2, \ldots, D$. One can adjust the new beam patterns so that the antenna array only receives desired signals, while nullifying noise and interference signals.

For an SBS with nested array, suppose the directions of its small cell user equipments (SUE) are $\{\delta_l, l = 1, 2, \ldots, L\}$, and the directions of interfering SUEs and macrocell user equipments (MUE) are $\{\eta_i, i = 1, 2, \ldots, I\}$. Then, the beam patterns of different directions are expected to be

$$\begin{cases} B(\delta_l) = 1, \; l = 1, 2, \ldots, L \\ B(\eta_i) = 0, \; i = 1, 2, \ldots, I. \end{cases} \tag{7}$$

According to (3), the beamforming weight vector satisfies

$$\begin{pmatrix} \left(\mathbf{f}^*(\delta_1) \otimes \mathbf{f}(\delta_1)\right)^H \\ \vdots \\ \left(\mathbf{f}^*(\delta_L) \otimes \mathbf{f}(\delta_L)\right)^H \\ \left(\mathbf{f}^*(\eta_1) \otimes \mathbf{f}(\eta_1)\right)^H \\ \vdots \\ \left(\mathbf{f}^*(\eta_I) \otimes \mathbf{f}(\eta_I)\right)^H \\ (\vec{\mathbf{1}})^T \end{pmatrix} \mathbf{w} = \begin{pmatrix} 1 \\ \vdots \\ 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}. \tag{8}$$

With the solution of $\mathbf{w}$, the weight vector of the original antenna array can be determined by the method presented in [19]. It can be observed from (8) that the number of DoFs to identify and manage the desired signals, noise, and interfering signals is $O(N^2)$. This enforces a constraint on the total number of desired and interfering sources that can be resolved. By employing spatial filtering on different directions, the nested array-based approach provides a new perspective to interference management. With the desirable features of nested array, it is highly promising to apply this technique to interference management in massive MIMO HetNets.

III. System Model and Problem Formulation

Consider a two-tier massive MIMO HetNet consists of an MBS with a massive MIMO (i.e., BS $j = 0$) and multiple SBSs, each with a regular MIMO (denoted as $j = 1, 2, \ldots, J$). There are K users (indexed by $k = 1, 2, \ldots, K$) to be served. Let $x_{k,j}$ be the *user association variable* defined as

$$x_{k,j} \doteq \begin{cases} 1, & \text{user } k \text{ is associated with BS } j \\ 0, & \text{otherwise,} \end{cases}$$
$$k = 1, 2, \ldots, K, \; j = 0, 1, \ldots, J, \tag{9}$$

The macrocell and small cells share the same spectrum band and both tiers adopt the time division duplex (TDD) mode in a synchronized way, i.e., the two tiers use the same time period for uplink or downlink transmissions. The SBSs use the nested array to perform interference nulling, so that the uplink interference from a certain number of users served by other BSs can be nulled with the beamforming process presented in (8). The MBS with a massive MIMO adopts the traditional linear array for DoA estimation and interference management due to two reasons: (i) the DoF of the MBS is sufficiently large; (ii) the nested array requires second order processing at all antennas, which may not be feasible at the MBS due to complexity concerns. With the DoAs of the interference signals, both the MBS and SBSs can optimize the direction of departure (DoD) of their transmissions with analog domain beamforming to avoid downlink interference to a certain number of users. This way, the mutual interference between the BSs and some users can be eliminated.

Define binary *interference nulling indicators* $n_{k,j}$ as,

$$n_{k,j} \doteq \begin{cases} 1, & \text{BS } j \text{ nulls the interference from user } k \\ 0, & \text{otherwise,} \end{cases}$$
$$k = 1, 2, \ldots, K,\ j = 0, 1, \ldots, J. \tag{10}$$

Due to multipath propagation, the signal of each user is received by a BS from multiple directions. Thus, the DoF assigned to a user is determined by the number of multipath components between the user and the BS. According to (8), $n_{k,j}$ should satisfy

$$\sum_{k=1}^{K} x_{k,j} q_{k,j} + \sum_{k=1}^{K} n_{k,j} q_{k,j} + 1 \le D_j,\ j = 0, 1, \ldots, J, \tag{11}$$

where $q_{k,j}$ is the number of multipaths from user k to BS j, and D_j is the DoF of BS j, which serves as the upper bound for the number of directions that can be resolved by BS j. Let M_j be the number of antennas at BS j, and assume the SBSs adopt the optimal N-level nested array. Then we have $D_j = M_j(M_j - 1) + 1,\ j = 1, \ldots, J$. For the MBS without nested array, we have $D_0 = M_0$. We also assume that noise is always nulled by the BSs using one DoF.

As in Fig. 1, the transmission/reception at each BS is a two-stage process, one is analog processing in the radio frequency (RF) domain and the other is digital processing at baseband. As presented in Section II, the nested array based interference nulling is performed in the RF domain of each SBS. By identifying the directions of desired signals and interference, the signals from different directions are spatially filtered with the second order processing given in (8). Then, the signals of users served by an SBS remain, while part of the interference from other users are nullified. After interference nulling using nested array, the intra-cell interference between users served by the same BS still exists and can only be mitigated with baseband processing in the digital domain. For each SBS, we assume that a matched filter is use for precoding and detection. Then, the sum of uplink and downlink data rates of user k connecting to SBS j can be approximated as in (12), and (13) shown at the bottom of this page, where p_k and p_j are the powers of user k and BS j, respectively; $h_{k,j}$ is the channel gain between user k and BS j; and $g_{k,j}$ is the large-scale channel gain between user k and BS j.

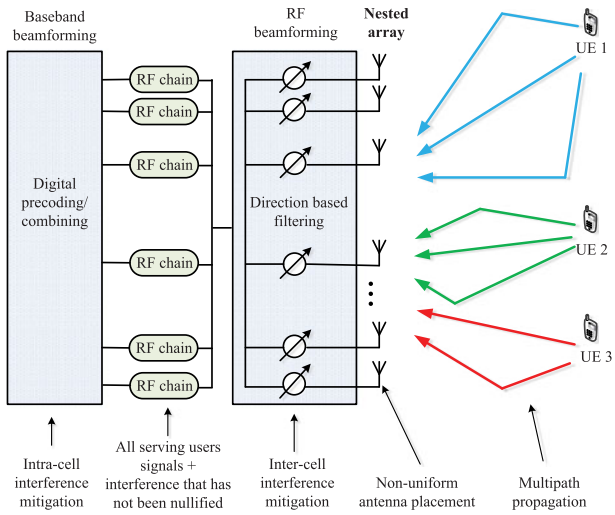


Fig. 1. System architecture and signal processing of nested array-based interference management.

For a macrocell user, due to the law of large numbers, the intra-cell interference can be averaged out in a massive MIMO system. Using the data rate model of massive MIMO HetNet in [10], the sum of uplink and downlink data rates of user k connecting to MBS is given by (13), where M_0 is the number of antennas of MBS, S_0 is the beamforming size of MBS, $\frac{M_0 - S_0 + 1}{S_0}$ is the antenna array gain of massive MIMO.

Let $\mathbf{x}$ and $\mathbf{n}$ be the matrices of $\{x_{k,j}\}$ and $\{n_{k,j}\}$, respectively. The sum rate maximization of a massive MIMO HetNet is formulated as follows.

$$\mathbf{P1}: \max_{\{\mathbf{x},\mathbf{n}\}} \left\{ \sum_{k=1}^{K} x_{k,0} R_{k,0} + \sum_{k=1}^{K} \sum_{j=1}^{J} x_{k,j} R_{k,j} \right\} \tag{14}$$

subject to:

$$\sum_{k=1}^{K} x_{k,j} q_{k,j} + \sum_{k=1}^{K} n_{k,j} q_{k,j} + 1 \le D_j,\ j = 0, 1, \ldots, J \tag{15}$$

$$R_{k,j} = \log\left(1 + \frac{p_k \left|h_{k,j}^H h_{k,j}\right|^2}{\sum_{k' \ne k} x_{k',j} p_{k'} \left|h_{k,j}^H h_{k',j}\right|^2 + \sum_{j' \ne j} \sum_{k' \ne k} x_{k',j'} p_{k'} g_{k',j} \left(1 - n_{k',j}\right)}\right) + \log\left(1 + \frac{p_j \left|h_{k,j}^H h_{k,j}\right|^2}{1 + \sum_{k' \ne k} x_{k',j} p_j \left|h_{k,j}^H h_{k',j}\right|^2 + v \sum_{j' \ne j} p_{j'} g_{k,j'} \left(1 - n_{k,j'}\right)}\right),$$
$$k = 1, 2, \ldots, K,\ j = 1, 2, \ldots, J. \tag{12}$$

$$R_{k,0} = \log\left(1 + \frac{M_0 - S_0 + 1}{S_0} \frac{p_k g_{k,0}}{\sum_{j=1}^{J} \sum_{k' \ne k} x_{k',j} p_{k'} g_{k',0} \left(1 - n_{k',0}\right)}\right) + \log\left(1 + \frac{M_0 - S_0 + 1}{S_0} \frac{p_0 g_{k,0}}{1 + \sum_{j=1}^{J} p_j g_{k,j} \left(1 - n_{k,j}\right)}\right),$$
$$k = 1, 2, \ldots, K. \tag{13}$$

$$n_{k,j} \leq 1 - x_{k,j},\ k = 1, 2, \ldots, K,\ j = 0, 1, \ldots, J \tag{16}$$

$$\sum_{j=0}^{J} x_{k,j} \leq 1,\ k = 1, 2, \ldots, K \tag{17}$$

$$x_{k,j}, n_{k,j} \in \{0, 1\},\ k = 1, 2, \ldots, K,\ j = 0, 1, \ldots, J. \tag{18}$$

Constraint (16) is due to the fact that when BS j serves user k, it does not need to null interference from user k. Constraint (17) indicates that each user can only be served by at most one BS.

IV. Interference Management with a Given User Association

In this section, we consider the case that user association is pre-determined (e.g., through a user association algorithm [11], [12]). Then, problem **P1** is reduced to

$$\mathbf{P2}: \max_{\{\mathbf{n}\}} \left\{ \sum_{k=1}^{K} x_{k,0} R_{k,0} + \sum_{k=1}^{K} \sum_{j=1}^{J} x_{k,j} R_{k,j} \right\} \tag{19}$$

subject to:

$$\sum_{k=1}^{K} x_{k,j} q_{k,j} + \sum_{k=1}^{K} n_{k,j} q_{k,j} + 1 \leq D_j,\ j = 0, 1, \ldots, J \tag{20}$$

$$n_{k,j} \leq 1 - x_{k,j},\ k = 1, 2, \ldots, K,\ j = 0, 1, \ldots, J \tag{21}$$

$$n_{k,j} \in \{0, 1\},\ k = 1, 2, \ldots, K,\ j = 0, 1, \ldots, J. \tag{22}$$

Problem **P2** is an integer programming program with a nonlinear and non-convex objective function, which is generally NP-hard. To make the problem tractable, we assume the system operate in the high SINR regime, so that $\log(1 + \mathrm{SINR}) \approx \log(\mathrm{SINR})$. The high SINR assumption is reasonable in a massive MIMO HetNet due to the large antenna array gain of massive MIMO and the short transmission distance of small cells. Applying this approximation to (12) and (13), the objective function of problem **P2** can be written as

$$\sum_{k=1}^{K} \sum_{j=0}^{J} x_{k,j} V_{k,j},\ k = 1, 2, \ldots, K,\ j = 0, 1, \ldots, J, \tag{23}$$

$V_{k,j}$ is given as

$$\begin{aligned} V_{k,0} = \ & \log\left(\frac{M_0 - S_0 + 1}{S_0} p_k g_{k,0}\right) \\ & + \log\left(\frac{M_0 - S_0 + 1}{S_0} p_0 g_{k,0}\right) \\ & - \log\left(\sum_{j=1}^{J} \sum_{k' \neq k} x_{k',j} p_{k'} g_{k',0} \left(1 - n_{k',0}\right)\right) \\ & - \log\left(1 + \sum_{j=1}^{J} p_j g_{k,j} \left(1 - n_{k,j}\right)\right) \end{aligned} \tag{24}$$

$$\begin{aligned} V_{k,j} = \ & \log\left(p_k g_{k,j}\right) + \log\left(p_j g_{k,j}\right) \\ & - \log\left(I_{k,j}^{U} + \sum_{j' \neq j} \sum_{k' \neq k} x_{k',j'} p_{k'} g_{k',j} \left(1 - n_{k',j}\right)\right) \\ & - \log\left(1 + I_{k,j}^{D} + \sum_{j' \neq j} p_{j'} g_{k,j'} \left(1 - n_{k,j'}\right)\right) \\ & j = 1, 2, \ldots, J, \end{aligned} \tag{25}$$

where $I_{k,j}^{U} = \sum_{k' \neq k} x_{k',j} p_{k'} |h_{k,j}^{H} h_{k',j}|^2$ and $I_{k,j}^{D} = \sum_{k' \neq k} x_{k',j} p_j |h_{k,j}^{H} h_{k',j}|^2$.

Let $\mathcal{U}_j$ be the set of users served by BS j, $\mathcal{U}_j = \{k | x_{k,j} = 1\}$. Then, the objective function of **P2** can be rewritten as $\sum_{j=0}^{J} \sum_{k \in \mathcal{U}_j} x_{k,j} V_{k,j}$. We remove the constants in (24) and (25) and apply the property $\sum_i \log x_i = \log(\prod_i x_i)$. Since $\log(\cdot)$ is a monotonic function, **P2** can be transformed into the following problem.

$$\mathbf{P3}: \max_{\{\mathbf{n}\}} \prod_{j=0}^{J} \prod_{k \in \mathcal{U}_j} W_{k,j} \tag{26}$$

$$\text{subject to: } (20), (21), \text{and } (22),$$

where

$$\begin{aligned} W_{k,0} = \ & \left[\sum_{j=1}^{J} \sum_{k' \neq k} x_{k',j} p_{k'} g_{k',0} \left(1 - n_{k',0}\right)\right] \\ & \times \left[1 + \sum_{j=1}^{J} p_j g_{k,j} \left(1 - n_{k,j}\right)\right], \end{aligned} \tag{27}$$

$$\begin{aligned} W_{k,j} = \ & \left[I_{k,j}^{U} + \sum_{j' \neq j} \sum_{k' \neq k} x_{k',j'} p_{k'} g_{k',j} \left(1 - n_{k',j}\right)\right] \\ & \times \left[I_{k,j}^{D} + 1 + \sum_{j' \neq j} p_{j'} g_{k,j'} \left(1 - n_{k,j'}\right)\right] \\ & j = 1, 2, ..., J, \end{aligned} \tag{28}$$

It can be seen that the objective function of **P3** is a product of linear expressions, which can be expressed as a polynomial on the set of variables $\{n_{k,j}\}$. Thus, **P3** is a nonlinear integer programming problem with a complicated form, which is hard to solve with general approaches. However, we can exploit a property of 0–1 problems to approximate problem **P3** with a linear integer programming problem.

A. Linear Approximation of *P3*

Consider the product of multiple i.i.d. 0–1 variables. When the number of variables is increased, the product becomes less likely to be 1 since it is less likely that all the variables are 1. As the objective function of **P2** is a weighted sum of products of 0–1 variables, the values of higher-order parts are more likely to be 0. Thus, the impact of the higher-order parts is limited.

Let P be the probability that an arbitrary $n_{k,j}$ equals to 1. In the objective function of **P3**, the probability for an M-th order product to be 1 is P^M.

Lemma 1: P can be approximated by $\frac{\bar{D}_j-1}{\bar{q}_{k,j}K}-\frac{1}{J}-\frac{1}{\bar{q}_{k,j}JK}$, where $\bar{z}$ is the mean of a variable z.

Proof: To maximize the sum rate, all the DoFs of each BS are expected to be used for data transmission and interference nulling. Thus all the constraints in (20) are close to equality. Summing up from $j=0$ to J, we have $\sum_{k=1}^{K}\sum_{j=0}^{J} n_{k,j}q_{k,j} = \sum_{j=0}^{J} D_j - \sum_{k=1}^{K}\sum_{j=0}^{J} x_{k,j}q_{k,j} - J - 1$. The probability that $n_{k,j}$ equals to 1 can be derived as

$$\begin{aligned} P &= \frac{\sum_{k=1}^{K}\sum_{j=0}^{J} n_{k,j}q_{k,j}}{\sum_{k=1}^{K}\sum_{j=0}^{J} q_{k,j}} \\ &= \frac{\sum_{j=0}^{J} D_j - \sum_{k=1}^{K}\sum_{j=0}^{J} x_{k,j}q_{k,j} - J - 1}{\bar{q}_{k,j}JK} \\ &= \frac{\sum_{j=0}^{J} D_j - \bar{q}_{k,j}K - J - 1}{\bar{q}_{k,j}JK} = \frac{\bar{D}_j - 1}{\bar{q}_{k,j}K} - \frac{1}{J} - \frac{1}{\bar{q}_{k,j}JK}. \end{aligned}$$

■

In a typical cellular network, the number of users in a macrocell can be more than 100, i.e., $K > 100$. The DoFs are $D_j = O(N^2)$ for SBSs and $D_0 = O(N)$ for the MBS. As the typical number of antennas for the MBS and an SBS are 100 and 10, respectively, we have $\bar{D}_j \approx 100$. Therefore, the value of P is expected to be small in a practical system, and the values of higher-order terms of P are close to 0. Due to this fact, we derive linear approximations for the higher-order parts in the polynomial of (26) and transform the objective function of **P3** to a linear function.

Let $\tilde{\mathbf{n}}$ be the vector concatenating the columns of matrix $[n_{k,j}]_{K\times J}$. For a product with M elements in $\tilde{\mathbf{n}}$ given as $\tilde{\mathbf{n}}_{i_1}, \tilde{\mathbf{n}}_{i_2}, \cdots, \tilde{\mathbf{n}}_{i_M}$, we have the following approximation.

$$\tilde{\mathbf{n}}_{i_1}\tilde{\mathbf{n}}_{i_2}\cdots\tilde{\mathbf{n}}_{i_M} \approx \frac{P^{M-1}}{M}\left(\tilde{\mathbf{n}}_{i_1} + \tilde{\mathbf{n}}_{i_2} + \cdots + \tilde{\mathbf{n}}_{i_M}\right). \quad (29)$$

It can be easily verified that the expectations of both sides are equal to P^M, thus the long term performance of the approximation problem equals to that of the original problem. The expected value of the gap between the two sides of (29) is given as

$$\mathbb{E}\left\{\tilde{\mathbf{n}}_{i_1}\tilde{\mathbf{n}}_{i_2}\cdots\tilde{\mathbf{n}}_{i_M} - \frac{P^{M-1}}{M}\left(\tilde{\mathbf{n}}_{i_1} + \tilde{\mathbf{n}}_{i_2} + \cdots + \tilde{\mathbf{n}}_{i_M}\right)\right\} = (M-1)P^M, \quad (30)$$

which is a quite small value even when M is small. In addition, the product on the left hand side of (29) is approaching 0 as M increases. Thus, the approximation given by (29) is expected to be accurate. Note that, when M is larger than a certain value, e.g., $M > 5$, the left hand side of (29) approaches 0, and we can approximate the higher order parts with 0 to reduce the number of variables and computational complexity.

With the linear approximation of the polynomial objective function, problem **P3** is transformed to the following integer programming problem.

$$\mathbf{P4}: \max_{\{\tilde{\mathbf{n}}\}} \ \mathbf{c}\tilde{\mathbf{n}} \quad (31)$$

$$\text{subject to: } \mathbf{A}\tilde{\mathbf{n}} \le \mathbf{b}. \quad (32)$$

The vector $\mathbf{c}$ is determined by applying the linear transformation of (29) to (26). The constraint matrix $\mathbf{A}$ is given by

$$\mathbf{A}_{(K+1)(J+1)\times K(J+1)} \doteq \begin{pmatrix} \mathbf{Q} \\ \mathbf{I} \end{pmatrix}, \quad (33)$$

where $\mathbf{I}$ is a $K(J+1)\times K(J+1)$ identity matrix, and $\mathbf{Q}$ is given by

$$\mathbf{Q}_{(J+1)\times K(J+1)} = \begin{pmatrix} \mathbf{q}_0 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{q}_1 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{q}_J \end{pmatrix}, \quad (34)$$

where $\mathbf{q}_j = [q_{1,j}, q_{2,j}, \ldots, q_{K,j}]$, $j = 0, 1, \ldots, J$. The vector $\mathbf{b}$ in (32) is given by

$$\begin{aligned} \mathbf{b}_{(K+1)(J+1)\times 1} \doteq\ & [E_0, \ldots, E_J, 1-x_{1,0}, \ldots, 1-x_{K,0}, \\ & 1-x_{1,1}, \ldots, 1-x_{K,J}]^T, \end{aligned} \quad (35)$$

where

$$E_j = D_j - \sum_{k=1}^{K} x_{k,j}q_{k,j} - 1,\ j = 0, 1, \ldots, J. \quad (36)$$

In **P4**, matrix $\mathbf{A}$ characterizes the coefficients of the linear constraints (20) and (21) on $\{n_{k,j}\}$. In $\mathbf{A}$, the matrix $\mathbf{Q}$ corresponds to the information for the number of multipath between each user and each BS. The vector $\mathbf{b}$ consists of the values of the right-hand side of constraints (20) and (21). In particular, $E_0, \ldots, E_J$ are the upper bounds for the DoF that can be used by the BSs for interference nulling.

B. Performance Upper Bound

To verify the effectiveness of the approximation, we derive a performance upper bound for **P3** and compare it with the proposed scheme in our simulations (see Section VI). Applying the following linear approximation

$$\tilde{\mathbf{n}}_{i_1}\tilde{\mathbf{n}}_{i_2}\cdots\tilde{\mathbf{n}}_{i_M} \approx \frac{\tilde{\mathbf{n}}_{i_1} + \tilde{\mathbf{n}}_{i_2} + \cdots + \tilde{\mathbf{n}}_{i_M}}{M}, \quad (37)$$

we have the resulting integer programming problem as

$$\mathbf{P5}: \max_{\tilde{\mathbf{n}}} \ \mathbf{c}'\tilde{\mathbf{n}} \quad (38)$$

$$\text{subject to: } \mathbf{A}\tilde{\mathbf{n}} \le \mathbf{b}, \quad (39)$$

where $\mathbf{c}'$ is determined by (37).

Lemma 2: With the linear approximation described in (37), the objective function of problem **P5** is an upper bound for that of problem **P3**.

Proof: Due to the fact that the geometric mean is no greater than the arithmetic mean, we have $\sqrt[M]{\tilde{\mathbf{n}}_{i_1}\tilde{\mathbf{n}}_{i_2}\cdots\tilde{\mathbf{n}}_{i_M}} \le (\tilde{\mathbf{n}}_{i_1} + \tilde{\mathbf{n}}_{i_2} + \cdots + \tilde{\mathbf{n}}_{i_M})/M$. Since all elements of $\tilde{\mathbf{n}}$ are 0–1 variables,

it can be easily verified that $\sqrt[M]{\tilde{\mathbf{n}}_{i_1}\tilde{\mathbf{n}}_{i_2}\cdots\tilde{\mathbf{n}}_{i_M}} = \tilde{\mathbf{n}}_{i_1}\tilde{\mathbf{n}}_{i_2}\cdots\tilde{\mathbf{n}}_{i_M}$. Thus, we have

$$\tilde{\mathbf{n}}_{i_1}\tilde{\mathbf{n}}_{i_2}\cdots\tilde{\mathbf{n}}_{i_M} \leq \frac{\tilde{\mathbf{n}}_{i_1}+\tilde{\mathbf{n}}_{i_2}+\cdots+\tilde{\mathbf{n}}_{i_M}}{M},\ \forall M \geq 2. \quad (40)$$

Applying (40) to all the higher-order expressions in (26), $\mathbf{c}'\tilde{\mathbf{n}}$ is an upper bound to the objective function of **P3**

Based on Lemma 2, we further conclude that the optimal solution to problem **P5** provides an upper bound to the optimal solution of **P3**. ■

C. Optimal Solution to *P4* with the Cutting Plane Method

Since problem **P4** has a linear objective function, the cutting plane method [21] can be used to derive the optimal solution. The idea of cutting plane is to find a plane that separates the non-integer solution from the polyhedron that satisfies the constraints and contains all the integer, feasible solutions.

Consider the polyhedron defined by $\mathbf{A}\tilde{\mathbf{n}} \leq \mathbf{b}$, determined by a combination of $(K+1)(J+1)$ inequalities as

$$\mathbf{a}_i\tilde{\mathbf{n}} \leq b_i,\ i = 1, 2, \ldots, (K+1)(J+1), \quad (41)$$

where $\mathbf{a}_i$ is the ith row of $\mathbf{A}$. Let $y_1, y_2, \ldots, y_{(K+1)(J+1)} \geq 0$ and set

$$\mathbf{a}^* = \sum_{i=1}^{(K+1)(J+1)} y_i\mathbf{a}_i,\ \ b^* = \sum_{i=1}^{(K+1)(J+1)} y_i b_i. \quad (42)$$

Obviously, all the solutions in the polyhedron $\mathbf{A}\tilde{\mathbf{n}} \leq \mathbf{b}$ also satisfy $\mathbf{a}^*\tilde{\mathbf{n}} \leq b^*$. If $\mathbf{a}^*$ is integral, i.e., all elements of $\mathbf{a}^*$ are integers, then all the *integer* solutions should satisfy

$$\mathbf{a}^*\tilde{\mathbf{n}} \leq \lfloor b^* \rfloor, \quad (43)$$

where $\lfloor b^* \rfloor$ is the largest integer that is smaller than b^*. Then, (43) defines a cutting plane for **P4**.

With an additional constraint described by the cutting plane, all the integer solutions are still included while some non-integer solutions are removed. Due to this property, we can first relax the integer constraint in **P4** and solve the linear programming problem. If there are non-integer solutions, we add an additional constraint in the form of (43). Then, we solve the linear programming problem with the updated constraints. If there are still non-integer elements in the solution vector, we continue to add another constraint following (43). Such process terminates when all solution variables are integer. However, the effectiveness of this approach depends on the proper setting of parameters y_i. An efficient scheme to find the effective cutting plane was proposed in [21]; the details are omitted here due to lack of space.

D. A Special Case Without the Need for Cutting Plane

We consider a special case with an additional condition.

Assumption 1: Each BS j only receives a fixed amount of L_j strongest multipath signals from each user and neglect the other multipath components with weaker signal strengths, and L_j is set to a value such that E_j/L_j is an integer, where E_j is defined in (36).

With Assumption 1, we then have

$$q_{1,j} = q_{2,j} = \cdots = q_{K,j} = L_j,\ j = 0, 1, \ldots, J. \quad (44)$$

Note that, when L_j is sufficiently large, this special case can be regarded as the real case. Given (42), we divide both sides of (20) by L_j and $\mathbf{A}$ is updated by replacing $\mathbf{q}_j$ with

$$\mathbf{q}_j = [1, 1, \ldots, 1]^T. \quad (45)$$

The vector $\mathbf{b}$ is updated as

$$\mathbf{b}_{(K+1)(J+1)\times 1} \doteq \left[\frac{E_0}{L_0}, \ldots, \frac{E_J}{L_J}, 1-x_{1,0}, \ldots, 1-x_{K,0}, 1-x_{1,1}, \ldots, 1-x_{K,J}\right]^T. \quad (46)$$

There are exactly two 1's in each column of $\mathbf{A}$, with one from a column of $\mathbf{Q}$ and the other from a column of $\mathbf{I}$.

Definition 1: A matrix $\mathbf{A}$ is totally unimodular if the determinant of every square submatrix of $\mathbf{A}$ is either 0, +1 or −1 [22].

Lemma 3: Under Assumption 1, $\mathbf{A}$ is a totally unimodular matrix.

Proof: We divide the constraint matrix $\mathbf{A}$ into blocks as

$$\mathbf{A} = \begin{pmatrix} \mathbf{Q}_1 & \mathbf{Q}_2 & \cdots & \mathbf{Q}_J \\ \mathbf{I}_1 & \mathbf{I}_2 & \cdots & \mathbf{I}_J \end{pmatrix},$$

where each $\mathbf{Q}_j$, $j = 1, 2, \ldots, J$, is a $J \times K$ matrix; the jth row of $\mathbf{Q}_j$ is all 1, while all the other rows are all 0; and each $\mathbf{I}_j$, $j = 1, 2, \ldots, J$, is a $K \times K$ identity matrix.

Denote G_n as an arbitrary $n \times n$ square submatrix of matrix $\mathbf{A}$. Obviously, the determinant of G_n is either 0 or 1 when $n = 1$. To analyze the determinant of G_n for $n \geq 2$, the following two cases need to be considered.

Case 1: G_n is a submatrix of $\mathbf{Q}_j$ or $\mathbf{I}_j$, $j = 1, 2, \ldots, J$. If G_n is a submatrix of $\mathbf{Q}_j$, we have $\det(G_n) = 0$, since at least one row would be all 0. If G_n is a submatrix of $\mathbf{I}_j$, $\det(G_n)$ would be either 0 or +1, since $\mathbf{I}_j$ is an identity matrix.

Case 2: The entries of G_n are from more than one $\mathbf{Q}_j$ or $\mathbf{I}_j$. We apply an induction method to analyze the determinant. For $n = 2$, $\det(G_n)$ can only be 0, +1, or −1, since the four entries are either 0 or 1 with at least one 0. Suppose $\det(G_{n-1})$ can only be 0, +1, or −1, we need to verify whether the same result hold for $\det(G_n)$. Denote $G_n(u, v)$ as the entry of G_n at row u, column v. Let $v^* = \arg\min_v\{\sum_u G_n(u, v)\}$. Then column v^* is the one with the minimum number of 1s in G_n. Let φ_{v^*} be the number of 1s in column v^*, which can be 0, 1, or 2 according to the structure of $\mathbf{A}$.

If $\varphi_{v^*} = 0$, column v^* of G_n is all 0 and $\det(G_n) = 0$.

If $\varphi_{v^*} = 1$, we calculate $\det(G_n)$ through column v^* and have $\det(G_n) = \pm\det(G_{n-1})$. According to the induction hypothesis, $\det(G_{n-1})$ can only be 0, −1, or 1. Therefore, $\det(G_n)$ can only be 0, −1, or 1.

If $\varphi_{v^*} = 2$, each column of G_n has exactly two 1s, with one in $\mathbf{Q}_j$ and the other in $\mathbf{I}_j$. Due to the equal number of 1s in $\mathbf{Q}_j$ and $\mathbf{I}_j$, we can obtain an all-zero row in G_n through some elementary transformations, which yields $\det(G_n) = 0$.

Consequently, the determinant of any square submatrix of $\mathbf{A}$ can only be either 0, -1, or 1. According to Definition 1, we conclude that $\mathbf{A}$ is totally unimodular. ■

For a linear programming problem with a unimodular constraint matrix $\mathbf{A}$ and integral right hand side vector $\mathbf{b}$, all decision variables to the optimal solution are integers [22]. Thus, the optimal solution of **P4** can be obtained by relaxing the integer constraints and solving the resulting linear programming problem.

V. Distributed Algorithm for Joint Interference Nulling Schedule and User Association

As the DoF of each BS is shared by interference nulling and data transmission, joint optimization of interference nulling schedule and user association, which corresponds to solving problem **P1**, could further optimize the system performance. However, problem **P1** is a highly complicated integer programming problem with two sets of variables, which cannot be solved with computational efficient techniques. In this section, we propose a distributed solution algorithm based on a *poly matching* between users and BSs, in which each user and BS makes its own decision to optimize its performance.

A. Poly Matching Between Users and BSs

We assume that each user has a *preference list* over the BSs. When a user is not served by any BS, the preference list is for user association, which is determined by the achievable rate of connecting to different BSs. For instance, if $j^* = \arg\max_j \{R_{k,j}\}$, BS j^* is on top of user k's preference list. When, a user is served by a BS, the preference list is for interference nulling, which is determined by the level of interference received from other BSs.

On the other hand, each BS has a preference list over users, which is determined by its performance gain achieved by serving a user or nullifying the interference of a user. Each BS also has a waiting list indicating the set of users that are currently held by the BS. With the objective of maximizing the sum rate of the users that it serves under the constraint $\sum_{k=1}^{K} x_{k,j} q_{k,j} + \sum_{k=1}^{K} n_{k,j} q_{k,j} + 1 \leq D_j$, the distributed user association and interference nulling strategy for BS j is presented in Algorithm 1.

In Algorithm 1, $\Delta_{k,j}$ is defined as the performance gain of BS j by serving user k or nullifying the interference from user k. Suppose BS j put user k^* into its waiting list, then $\Delta_{k^*,j}$ is given as

$$\Delta_{k^*,j} = \sum_{k=1}^{K} x^*_{k,j} R^*_{k,j} - \sum_{k=1}^{K} x_{k,j} R_{k,j}, \quad k = 1, 2, \ldots, K,\ j = 0, 1, \ldots, J, \tag{47}$$

where $R^*_{k,j}$ and $x^*_{k,j}$ are the data rate and user association indicator of user k after serving or nulling the interference of user k^*, respectively. Similarly, $\Delta_{\{1,\ldots,i\},j}$ is defined as the performance gain of BS j by adding the set of users $\{1, \ldots, i\}$ into its waiting list. The initial values of $\Delta_{k,j}$ are set to be $R_{k,j}$, and $\Delta_{k,j}$ is updated whenever user k proposes to a BS or BS j accepts the proposal of a user. In Algorithm 1, the information needed to calculate $\Delta_{k,j}$, e.g., channel gain and traffic load, is collected by each BS using the uplink signals from users, and each BS distributively updates its decision variables based on $\Delta_{k,j}$.

Algorithm 1: Distributed User Association and Interference Nulling Strategy of BS j

```
while (covergence not achieved) do
    For the users propose to BS j and the users that are not
    served but can be detected by BS j, put them into multiple
    sets according to the number of multipaths to BS j, given
    as Ω_q = {k | q_{k,j} = q} ;
    for q = q_min : q_max do
        Assign indices to users in Ω_q, k* = 1, 2, ..., according
        to the descending order of Δ_{k,j} ;
        while (user k* has not been rejected) do
            if Σ_{k=1}^{K} x_{k,j} q_{k,j} + Σ_{k=1}^{K} n_{k,j} q_{k,j} + 1 > D_j after
            accepting user k* then
                For users already in the waiting list, sort
                {Δ_{k,j}} is ascending order as {Δ_{i,j}} ;
                ρ_j = 0 ;
                i = 1 ;
                while ρ_j < q do
                    ρ_j = ρ_j + Δ_{i,j} ;
                    i = i + 1 ;
                end
                if Δ_{k*,j} > Δ_{{1,...,i},j} then
                    Add user k* into the waiting list ;
                    Reject user(s) {1, ..., i} ;
                    k* = k* + 1 ;
                else
                    Reject user k* ;
                end
            else
                Add user k* in the waiting list ;
                k* = k* + 1 ;
            end
        end
        Reject all users ranked after user k* in Ω_q ;
    end
end
```

The poly matching between users and BSs has three stages. In the *first* stage, each user proposes to the top BS in its preference list. In particular, if a user has not been served by any BS, the user proposes to be served; for users that are currently served by a BS, they propose to other BSs for interference nulling.

In the *second* stage, the BSs decide whether to accept the proposals of users according to Algorithm 1 and feedback the decision to users. Specifically, the evaluation of each BS begins from the users with the least number of multipaths, to the users with larger numbers of multipaths. For users in each Ω_q, where $\Omega_q = \{k | q_{k,j} = q\}$, the evaluation is performed in descending order of $\Delta_{k,j}$. When evaluating a user, if the DoF constraint is still satisfied after serving the user or nullifying the interference of the user, the user is directly put into the waiting list of the BS. If the DoF constraint is violated after adding the user into the waiting list, a BS first selects the user(s) that use a DoF of no less than the required DoFs of the requesting user and with the least performance gain. Then, the BS compares the performance gain of the selected user(s) with that of the new user, and the

one with a larger performance gain will be added to or kept in the waiting list.

In the *third* stage, a user that has been rejected first deletes the BS that rejected it from its preference list. Then, the user proposes to the most desirable BS among the remaining ones. After receiving a proposal, a BS compares it with users in its waiting list by evaluating the number of multipath and the achievable rate of the user according to Algorithm 1, and then decides whether to serve the user. If a user is rejected again, it continues to propose to other BSs following the preference order, and the BSs make decisions and feedback to users, and so forth. Such a matching process between users and BSs is continued until convergence, i.e., the users in the waiting list of each BS do not change anymore.

The complexity of the poly matching is upper bounded by $\mathcal{O}(JK)$, which corresponds to the case that every user has proposed to every BS. Since $\Delta_{k,j}$ is updated whenever user k proposes to a BS or BS j accepts the proposal of a user, the complexity of calculating $\Delta_{k,j}$ is also upper bounded by $\mathcal{O}(JK)$. In Algorithm 1, users with the same number of multipaths are put into set Ω_q. In each set, users are sorted following the descending order of $\Delta_{k,j}$, and then users are evaluated in such order until a user is rejected. When a user in a set is rejected after evaluated by a BS, all subsequent users in the same set will also be rejected since their performance gains are smaller while they have the same number of multipaths. The BS does not need to evaluate the remaining users and it can directly switch to users in another set. This way the complexity is significantly reduced.

B. Convergence Analysis

We next prove that the poly matching scheme converges and a stable matching can be achieved.

Definition 2: In a stable matching, there is no such pair of people who are not matched as partners, while both of them prefer each other to their current partners. In other words, there is no such a pair of people that both of them have a better choice than their current partners [23].

Lemma 4: With Algorithm 1, a user entering the waiting list of a BS has a larger performance gain compared to user(s) in the waiting list. In other words, the sum rate of each BS is always increased after adding a user to its waiting list.

Proof: In Algorithm 1, if the inequality $\sum_{k=1}^{K} x_{k,j} q_{k,j} + \sum_{k=1}^{K} n_{k,j} q_{k,j} + 1 > D_j$ holds after accepting a new user with q multipaths, a comparison would be performed between the incoming user with the user(s) already in the list. The one with a larger performance gain would win the competition. Thus, a newly accepted user has a larger performance gain compared to user(s) in the waiting list. We further conclude that the sum rate of users served by a BS is always increased after a user is added to the waiting list of the BS. ■

Lemma 5: The BSs proposed by a user is non-increasing in the user's preference list.

Proof: Suppose user k is rejected by BS j. Following Algorithm 1, there must be a user k' with $\Delta_{k',j} > \Delta_{k,j}$ and $q_{k',j} \leq q_{k,j}$, and having the smallest value of performance gain among users with a number of multipaths no greater than $q_{k,j}$.

For the case when $q_{k',j} < q_{k,j}$, there is no user in the waiting list of BS j who has a larger or equal number of multipath and a smaller performance gain compared to user k. Hence, BS j would reject user k again, i.e., user k can never enter the waiting list of BS j again. For the case when $q_{k',j} = q_{k,j}$, BS j would make a comparison between user k' and user k. There are two subcases: (i) user k' is served by BS j, (ii) BS j nulls the interference of user k'. If user k' is served by BS j, then $\Delta_{k',j} > \Delta_{k,j}$ would always hold regardless of the interference pattern, since the SINR with user k' is always higher than that with user k. If user k' is selected by BS j for interference nulling, it is obvious that BS j would reject user k as long as user k' is still in the waiting list. Note that, user k' may be replaced by another user k'' in future rounds. With Lemma 4, we have $\Delta_{k'',j} > \Delta_{k',j} > \Delta_{k,j}$. If user k'' is selected for service provision, we have $\Delta_{k'',j} > \Delta_{k,j}$, user k'' would always bring a higher performance gain than user k regardless of the interference pattern. Thus, user k cannot be accepted by BS j again. If user k'' is selected for interference nulling, it must be the case that user k'' causes stronger interference to other users compared to user k'. Similarly, BS j would reject user k as long as user k'' is still in the waiting list. Thus, it is impossible for user k to be accepted by BS j again. We conclude that the sequence of BSs proposed by a user is non-increasing in its preference list. ■

With Lemma 5, a user deletes a BS from its preference list once it is rejected by the BS.

Theorem 1: The poly matching converges to a stable matching.

Proof: We provide a proof by contradiction. Suppose the matching is not stable. By definition, there must be a user k and a BS j such that: (i) user k prefers BS j to its current connecting BS j', (ii) BS j prefers user k to user k', who is currently in the waiting list of BS j and $q_{k',j} \geq q_{k,j}$. Note that, we use the case of one user as an example. The proof can be easily extend to the case of multiple users.

Since user k prefers BS j to BS j', user k will propose to BS j. With Algorithm 1, BS j would accept the proposal of user k and replace user k' since it prefers user k over user k'. By Lemma 4, the users being put into the waiting list has incremental performance gains. As user k contributes larger performance gain than user k' while user k is not in the waiting list, it must be the case that user k has never proposed to BS j. As user k prefers BS j over BS j', it must have proposed to BS j before BS j'. Then, the only explanation is that user k has never proposed to BS j'. However, user k is currently in the waiting list of BS j'. Hence, user k must have proposed to BS j' before, which is a contradiction. We conclude that the poly matching process converges and the outcome is a stable matching. ■

VI. Simulation Study

We validate the performance of the proposed scheme with Matlab simulations. Consider a macrocell overlaid with multiple small cells. The radii of the macrocell and a small cell are 1000 m and 50 m, respectively. The macrocell and small cells share a total bandwidth of 4 MHz. The transmit power of the MBS is set to 40 dBm, while the transmit power of the MUEs

has five levels ranging from 10 dBm to 30 dBm according to the distance between the MUE and MBS. The transmit power of an SBS and an SUE are set to 25 dBm and 15 dBm, respectively. We employ the ITU path loss model [24], the path loss from the MBS to a user and from an SBS to a user are $15.3 + 37.6\log_{10}d$ and $37 + 30\log_{10}d$, respectively. The ratio $\frac{M_0 - S_0 + 1}{S_0}$ is set to 100 for the MBS.

We consider four schemes in our simulations. The first one is interference nulling with a given user association (termed *IN Only*) described in Section IV, in which we assume that each user is connected to the BS with the strongest received signal strength. The second one is the proposed distributed joint interference nulling and user association scheme (termed *DJINUA*) presented in Section V. We also consider a heuristic scheme for comparison purpose (termed *Heuristic*). In the heuristic scheme, each user is served by the BS with the strongest signal strength, then each BS chooses the user that causes the strongest interference and nullifies the signals of this user. This process is continued until the DoF constraint at a BS is violated. We also consider the case where no interference nulling is performed as a baseline (termed *No Nulling*).

Two kinds of user distribution patterns are considered, i.e., the uniform distribution and the non-uniform distribution. Suppose the total number of users is U. In the uniform case, users are randomly distributed in the macrocell area; In the non-uniform case, the numbers of users within the coverage of SBSs are random numbers generated with the following approach. We first generate J random integers in $[0, U]$. Then, we sort the J integers in ascending order, given as $\kappa_1 \leq \kappa_2 \leq, \ldots, \leq \kappa_J$. Let $\pi_j = \kappa_j - \kappa_{j-1}$, for $j = 2, 3, \ldots, J$, and $\pi_1 = \kappa_1, \pi_{J+1} = U - \kappa_J$. Then, the sequence $\{\pi_1, \pi_2, \ldots, \pi_{J+1}\}$ includes $J + 1$ random integers in $[0, U]$ and the sum of these integers is U. The number of users in the coverage area of SBS j is set to $\pi_j, j = 1, 2, \ldots, J$, while the number of users that is not in the coverage area of any SBS is π_{J+1}.

The sum rates of different schemes versus the number of SBSs under uniform and non-uniform user distribution are presented in Figs. 2 and 3, respectively. It can be seen that without interference management, the sum rate first decreases as the SBSs are deployed, since the SINRs of MUEs are significantly reduced. As the number of SBSs continues to grow, the sum rate increases since more users can be served by nearby SBSs. Compared to the No Nulling scheme, a significant performance gain can be achieved by interference nulling, as a result of enhanced SINRs of both MUEs and SUEs. The sum rates with the DJINUA and IN only schemes are higher than the heuristic scheme since both schemes optimize the performance from the perspective of the entire network. As expected, the DJINUA scheme outperforms the IN only scheme since the user association is jointly considered with interference nulling. We can also observe that the performance of the IN only scheme is close to its upper bound, indicating that the solution with linear approximation is near optimal. The performance gaps between the proposed schemes (DJINUA and IN only) and heuristic scheme become larger as the number of SBSs increases, since the heuristic scheme only achieves a local optimal solution for each BS. The resulting performance loss increases as the network gets larger. Compared

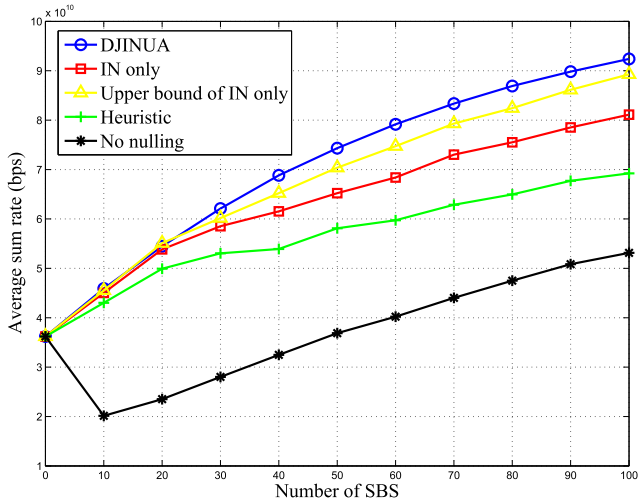


Fig. 2. Average sum rate versus number of SBSs. Uniform user distribution, 500 users, $\overline{q} = 6$, 10 antennas at each SBS.

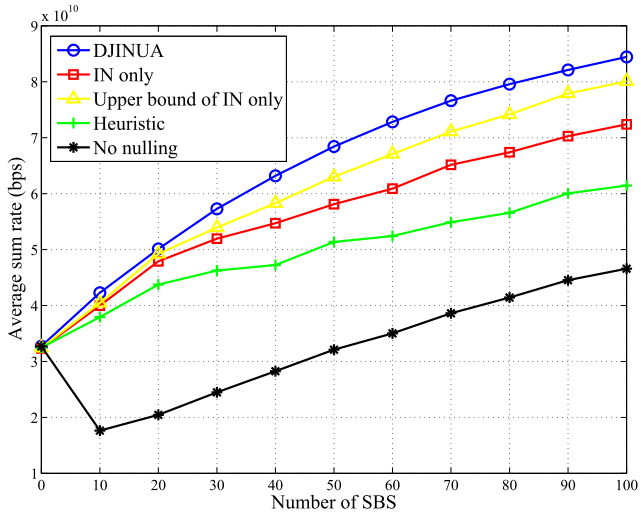


Fig. 3. Average sum rate versus number of SBSs. Non-uniform user distribution, 500 users, $\overline{q} = 6$, 10 antennas at each SBS.

to the case of uniform user distribution, the performance gain brought by interference nulling is decreased. This is because the number of users in each SBS varies from 0 to U, thus DoF of each BS is more likely to be under-utilized or insufficient. The performance gap between the DJINUA and IN only is increased compared to the uniform user distribution case, since the DJINUA is adaptive to the change of traffic with efficient use of DoF. Besides, each user and each BS can achieve higher data rate via the propose, compare, and reject operations in the poly matching.

In Figs. 4 and 5, we compare the sum rates under different numbers of users under uniform and non-uniform user distribution. Due to the same reasons, similar trends are observed for the different schemes. It can be seen when the number of users is sufficiently large, the performances of schemes with interference nulling are increasingly affected by interference, due to the fact that all the DoFs are used. The performance gap between the DJINUA and IN only becomes significant when the

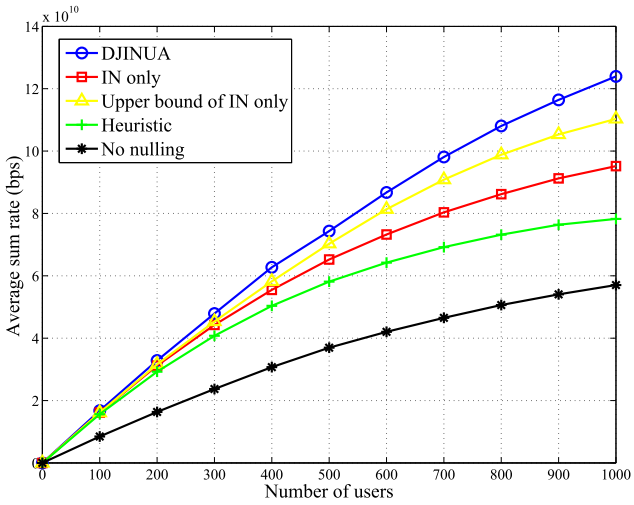


Fig. 4. Average sum rate versus average number of users. Uniform user distribution, 50 SBSs, $\overline{q} = 6$, 10 antennas at each SBS.

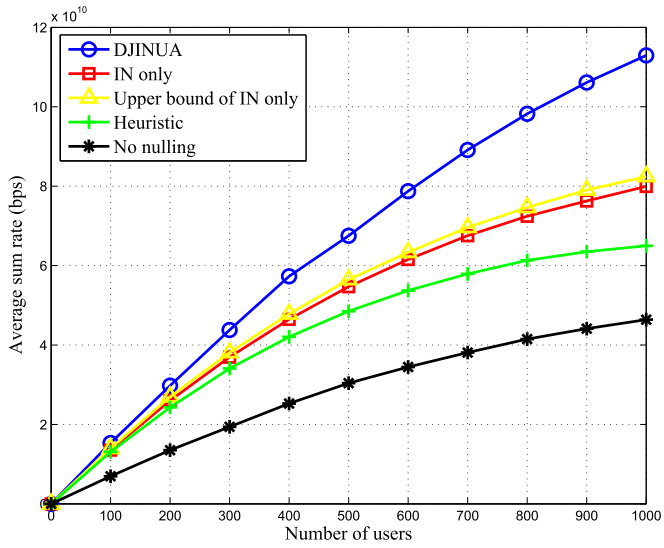


Fig. 5. Average sum rate versus average number of users. Non-uniform user distribution, 50 SBSs, $\overline{q} = 6$, 10 antennas at each SBS.

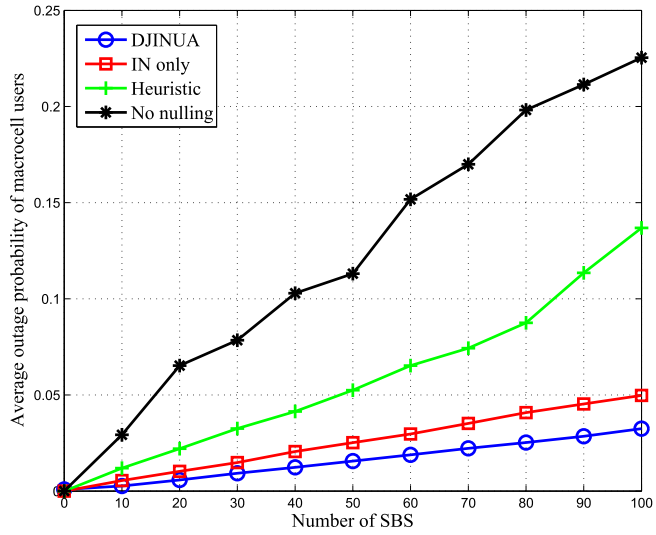


Fig. 6. Average outage probability of MUs versus number of SBSs. Uniform user distribution, 500 users, $\overline{q} = 6$, 10 antennas at each SBS.

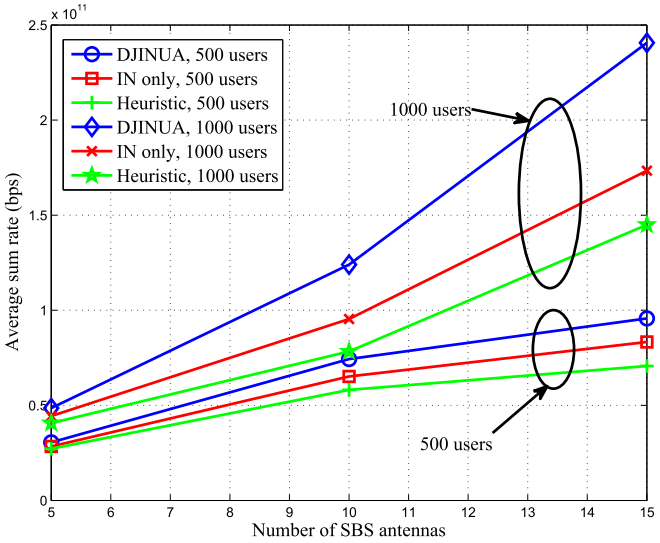


Fig. 7. Average sum rate versus number of SBS antennas. Uniform user distribution, 50 SBSs, 500 users, $\overline{q} = 6$.

number of users is large, showing the importance of DoF-aware joint schedule for user association and interference nulling in case of heavy traffic. The performance gap between the IN only scheme and its upper bound is increased when the number of users becomes large. This is because P becomes smaller as K is increased. Then, the higher-order products are more likely to be 0, and the linear approximation used to derive upper bound becomes less accurate. Similar to Figs. 2 and 3, the DJINUA scheme is more robust to the variations of traffic compared to other schemes, which also shows the benefits of DoF-aware joint schedule for user association and interference nulling, as well as the benefits of the operations in poly matching.

Fig. 6 shows the outage performance of macrocell users (MU). We chose to evaluate the MUs since their average SINRs are lower, and hence they are more vulnerable to interference compared to small cell users. Due to the aggregated interference caused by SBSs to MUE, the average outage probability of MUs increases as the number of SBSs grows. It can be seen that with interference nulling performed by SBSs, the average outage probability of MUs is significantly reduced. When the number of SBSs gets large, the outage probabilities of all schemes are increased, since part of the DoFs are used to deal with interference between an SBS and SUEs served by other SBSs. The DJINUA scheme achieves the best performance since the users with stronger interference and small number of multipath are more likely to be hold by the BSs for interference nulling. Thus, the DoF of each BS is efficiently used, resulting in most mitigated interference.

The impact of SBS antenna number is evaluated in Fig. 7. With increased antenna number at SBSs, more DoF is available for service provision and interference nulling, resulting in improved performance. When the number of users is large, the potential of increased DoF can be fully harnessed, and a

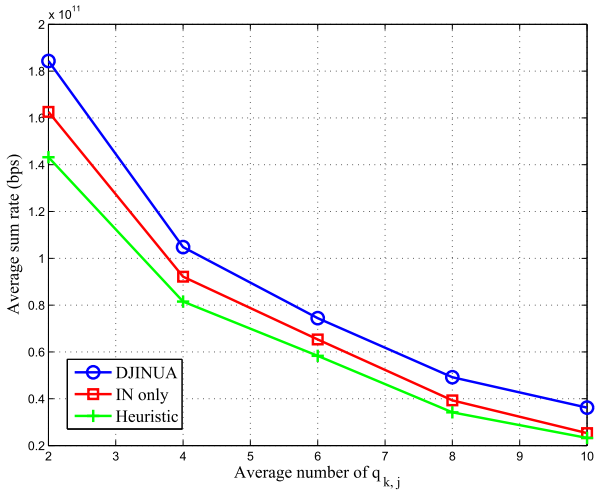


Fig. 8. Average sum rate versus average number of multipath, $\overline{q}$. Uniform user distribution, 50 SBSs, 500 users, 10 antennas at each SBS.

near-quadratic performance gain brought by $O(N^2)$ DoF can be achieved. However, we can also see from the Fig. 7 that when the number of users is relatively small, it is unnecessary to increase the number of SBS antennas as no significant performance gain can be achieved. Thus, the SBS antenna configuration should be based on the traffic pattern.

The impact of average number of multipath, $\overline{q}$, is shown in Fig. 8. When users have a larger average value of $q_{k,j}$, less users can be put into its waiting list of each BS, resulting in degraded system performance. In a practical system design, if we only consider a certain number of strongest multipath components and neglect the rest, which corresponds to a small value of $\overline{q}$, the evaluation of each users is less accurate. However, more users can be included by each BS for service provision or interference nulling. On the other hand, if we select a large value of $\overline{q}$ and take more multipath components into consideration, a more accurate information of each user can be obtained. However, fewer users can be associated by each BS for service provision or interference nulling. Such a tradeoff should be considered in the system design.

VII. Related Work

Massive MIMO is envisioned to be a potentially disruptive technology for future wireless communication systems [25], and has been extensively studied. The fundamental PHY layer techniques of massive MIMO were introduced in [3], [6]. Apart from the PHY, a wireless network with massive MIMO are also optimized from the perspective of upper layers [4], [11], [15], [27]. In [27], a time-shift frame structure was proposed to mitigate the effect of pilot contamination in a multi-cell massive MIMO system. With MAC layer design, the pilots of neighboring cells are transmitted at different time instants, the inter-cell interference can be well mitigated. In [15], the pilot reuse factor, BS deployment density, BS power, number of BS antennas, and number of users are optimized to maximize the energy efficiency of a massive MIMO HetNet.

DoA information has been considered in recent works to improve the performance of massive MIMO systems [28]–[31]. An ESPRIT-based DoA estimation scheme was proposed in [28] for 2D massive MIMO systems, and the mean square estimation error was derived. In [29], a multipath channel model was considered, where the channel gain is determined by the steering vector and the attenuation on each path. To reduce the channel estimation complexity and combat the effect of angular spread in DoA-based model, the low-rank property was employed in [30] with a spatial basis expansion model to represent the UL/DL channels. Using the spatial information and CSI of users, pilot contamination can be mitigated, and the system performance can be enhanced with user scheduling during the data transmission period. In a massive MIMO system with two-stage precoding, the angular spread of different user clusters may overlap, resulting in interference. In [31], a graph theory based pattern division scheme was proposed by assigning orthogonal subchannels to overlapping clusters.

QoS provisioning is a major design objective of wireless networks. Thus, interference management is a fundamental challenge, especially in a HetNet where both inter-tier and intra-tier interference need to be addressed. The major approaches include power control [33], spectrum allocation [34], access control [35], beamforming [36], and cognitive radio based interference avoidance [37]. Compared to these methods, the interference nulling considered in this paper is from the perspective of antenna processing, and interference is managed based on the directions of the sources.

User association in HetNet has been widely investigated, such as [12], [38], [39]. The objective in [12] is to minimize the maximum load among all BSs, several approximation algorithms were proposed with analysis on complexity and approximation ratio. In [38], user association and resource allocation were jointly optimized with the objective of sum utility maximization. Using a dual decomposition, each user can distributively update its lagrangian multiplier, and the solution is shown to be near-optimal. User association in massive MIMO HetNet was recently investigated in [39]. Based on the analysis for SINR expression under different precoding schemes, the network utility maximization problem was formulated and solved with optimization techniques.

VIII. Conclusion

In this paper, we applied the nested array technique in a massive MIMO HetNet and addressed the problem of joint user association and interference nulling scheduling to maximize the sum rate of MUEs and SUEs. We first considered the case with a given user association, and formulated the interference nulling scheduling problem as an integer programming problem. An approximation solution algorithm as well as a performance upper bound were derived to obtain the near optimal solution. We then considered joint user association and interference nulling and proposed a distributed scheme based on a poly matching between users and BSs. The simulation results validated the superior performance of the proposed schemes.

REFERENCES

[1] M. Feng and S. Mao, "Interference management in massive MIMO HetNets: A nested array approach," in *Proc. IEEE GLOBECOM'16*, Washington, DC, USA, Dec. 2016, pp. 1–6.

[2] J. Andrews *et al.*, "What will 5G be?" *IEEE J. Sel. Areas Commun.,* vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

[3] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.,* vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[4] M. Feng and S. Mao, "Harvest the potential of massive MIMO with multilayer techniques," *IEEE Netw.*, vol. 30, no. 5, pp. 40–45, Sept./Oct. 2016.

[5] Y. Xu, G. Yue, and S. Mao, "User grouping for massive MIMO in FDD systems: New design methods and analysis," *IEEE Access J.*, vol. 2, no. 1, pp. 947–959, Sep. 2014.

[6] F. Rusek *et al.*, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Sig. Proc.,* vol. 30, no. 1, pp. 40–60, Jan. 2013.

[7] K. Hosseini, J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO and small cells: How to densify heterogeneous networks," in *Proc. IEEE Int. Conf. Commun.,* Budapest, Hungary, Jun. 2013, pp. 5442–5447.

[8] M. Feng and S. Mao, "Adaptive pilot design for massive MIMO HetNets with wireless backhaul," in *Proc. IEEE Int. Conf. Sens., Commun., Netw.,* San Diego, CA, USA, Jun. 2017, pp. 1–9.

[9] E. Björnson, M. Kountouris, and M. Debbah, "Massive MIMO and small cells: Improving energy efficiency by optimal soft-cell coordination," in *Proc. Int. Conf. Telecommun.,* Casablanca, Morocco, May 2013, pp. 1–5.

[10] D. Bethanabhotla, O. Y. Bursalioglu, H. C. Papadopoulos, and G. Caire, "User association and load balancing for cellular massive MIMO," in *Proc. IEEE Inf. Theory Appl. Workshop,* San Diego, CA, USA, Feb. 2014, pp. 1–10.

[11] Y. Xu and S. Mao, "User association in massive MIMO HetNets," *IEEE Syst. J.*, vol. 11, no. 1, pp. 7–19, Mar. 2017.

[12] H. Zhou, S. Mao, and P. Agrawal, "Approximation algorithms for cell association and scheduling in femtocell networks," *IEEE Trans. Emerg. Topics Comput.,* vol. 3, no. 3, pp. 432–443, Sep. 2015.

[13] M. Feng, S. Mao, and T. Jiang, "BOOST: Base station on-off switching strategy for energy efficient massive MIMO HetNets," in *Proc. IEEE Int. Conf. Comput. Commun.,* San Francisco, CA, USA, Apr. 2016, pp. 1395–1403.

[14] M. Feng, S. Mao, and T. Jiang, "BOOST: Base station on-off switching strategy for green massive MIMO HetNets," *IEEE Trans. Wireless Commun.*, doi: 10.1109/TWC.2017.2746689.

[15] E. Björnson, L. Sanguinetti, and M. Kountouris, "Deploying dense networks for maximal energy efficiency: Small cells meet massive MIMO," *IEEE J. Sel. Areas Commun.,* vol. 34, no. 4, pp. 832–847, Apr. 2016.

[16] Q. Ye, O. Y. Bursalioglu, H. C. Papadopoulos, C. Caramanis, and J. G. Andrews, "User association and interference management in massive MIMO HetNets," *IEEE Trans. Wireless Commun.,* vol. 64, no. 5, pp. 2049–2065, May 2016.

[17] M. Feng, T. Jiang, D. Chen, and S. Mao, "Cooperative small cell networks: High capacity for hotspots with interference mitigation," *IEEE Wireless Commun.,* vol. 21, no. 6, pp. 108–116, Dec. 2014.

[18] A. Adhikary, H. S. Dhillon, and G. Caire, "Massive-MIMO meets HetNet: Interference coordination through spatial blanking," *IEEE J. Sel. Areas Commun.,* vol. 33, no. 6, pp. 1171–1186, Jun. 2015.

[19] P. Pal and P. P. Vaidyanathan, "Nested arrays: A novel approach to array processing with enhanced degrees of freedom," *IEEE Trans. Sig. Process.,* vol. 58, no. 8, pp. 4167–4180, Aug. 2010.

[20] R. T. Hoctor and S. A. Kassam, "The unifying role of the coarray in aperture synthesis for coherent and incoherent imaging," *Proc. IEEE,* vol. 78, no. 4, pp. 735–752, Apr. 1990.

[21] R. Gomory, "Outline of an algorithm for integer solutions to linear programs," *Bull. Amer. Math. Soc.,* vol. 64, no. 5, pp. 275–278, Sep. 1958.

[22] A. Schrijver, *Theory of Linear and Integer Programming.* Hoboken, NJ, USA: Wiley, 1998.

[23] R. W. Irving, "An efficient algorithm for the 'Stable Roommates' problem," *J. Algebra ,* vol. 6, no. 6, pp. 577–595, Dec. 1985.

[24] International Telecommunication Union, *Guidelines for Evaluation of Radio Transmission Technologies for IMT-2000*, Recommendation ITU-R M.1225, 1997.

[25] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Commun.*, vol. 52, no. 2, pp. 74–80, Feb. 2014.

[26] H. Xie, F. Gao, and S. Jin, "An overview of low-rank channel estimation for massive MIMO systems," *IEEE Access,* vol. 4, pp. 7313–7321, Nov. 2016.

[27] F. Fernandes, A. Ashikhmin, and T. L. Marzetta, "Inter-cell interference in noncooperative TDD large scale antenna systems," *IEEE J. Sel. Areas Commun.,* vol. 31, no. 2, pp. 192–201, Feb. 2013.

[28] Y. Zhu, L. Liu, A. Wang, K. Sayana, and J. Zhang, "DoA estimation and capacity analysis for 2D active massive MIMO systems," in *Proc. IEEE Int. Conf. Commun.,* Budapest, Hungary, Jun. 2013, pp. 4630–4634.

[29] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *IEEE J. Sel. Areas Commun.,* vol. 31, no. 2, pp. 264–273, Feb. 2013.

[30] H. Xie, F. Gao, S. Zhang, and S. Jin, "A unified transmission strategy for TDD/FDD massive MIMO systems with spatial basis expansion model," *IEEE Trans. Veh. Technol.,* vol. 66, no. 4, pp. 3170–3184, Apr. 2017.

[31] J. Ma, S. Zhang, H. Li, N. Zhao, and A. Nallanathan, "Pattern division for massive MIMO networks with two-stage precoding," *IEEE Commun. Lett.,* vol. 21, no. 7, pp. 1665–1668, Jul. 2017.

[32] S. Mao and S. S. Panwar, "A survey of envelope processes and their applications in quality of service provisioning," *IEEE Commun. Surveys Tut.,* vol. 8, no. 3, pp. 2–20, Jul.–Sep. 2006.

[33] V. Chandrasekhar, J. G. Andrews, T. Muharemovic, Z. Shen, and A. Gatherer, "Power control in two-tier femtocell networks," *IEEE Trans. Wireless Commun.,* vol. 8, no. 8, pp. 4316–4328, Aug. 2009.

[34] V. Chandrasekhar and J. G. Andrews, "Spectrum allocation in tiered cellular networks," *IEEE Trans. Commun.,* vol. 57, no. 10, pp. 3059–3068, Oct. 2009.

[35] M. Feng, D. Chen, Z. Wang, and T. Jiang, "Throughput improvement for OFDMA femtocell networks through spectrum allocation and access control strategy," in *Proc. IEEE Comput., Commun. Appl. Conf.,* Hong Kong, China, Jan. 2012, pp. 387–391.

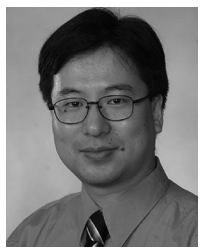
[36] D.-C. Oh, H.-C. Lee, and Y.-H. Lee, "Power control and beamforming for femtocells in the presence of channel uncertainty," *IEEE Trans. Veh. Technol.,* vol. 60, no. 6, pp. 2545–2554, Jul. 2011.

[37] M. Feng, S. Mao, and T. Jiang, "Joint duplex mode selection, channel allocation, and power control for full-duplex cognitive femtocell networks," *Elsevier Digital Commun. Netw.*, vol. 1, no. 1, pp. 30–44, Feb. 2015.

[38] Q. Ye, B. Rong, Y. Chen, M. A.-Shalash, C. Caramanis, and J. G. Andrews, "User association for load balancing in heterogeneous cellular networks," *IEEE Trans. Wireless Commun.,* vol. 12, no. 6, pp. 2706–2716, Jun. 2013.

[39] D. Bethanabhotla, O. Y. Bursalioglu, H. C. Papadopoulos, and G. Caire, "Optimal user-cell association for massive MIMO wireless networks," *IEEE Trans. Wireless Commun.,* vol. 15, no. 3, pp. 1835–1850, Mar. 2016.

Mingjie Feng (S'15) received the B.E. and M.E. degrees from Huazhong University of Science and Technology in 2010 and 2013, respectively, both in electrical engineering. He is currently working toward the Ph.D. degree in the Department of Electrical and Computer Engineering, Auburn University, Auburn, AL, USA. He was a Visiting Student in the Department of Computer Science, Hong Kong University of Science and Technology, Hong Kong, in 2013. His research interests include cognitive radio networks, heterogeneous networks, massive MIMO, mmWave networks, and full-duplex communications. He received a Woltosz Fellowship at Auburn University.



Shiwen Mao (S'99–M'04–SM'09) received the Ph.D. degree in electrical and computer engineering from Polytechnic University, Brooklyn, NY, USA, in 2004. He is the Samuel Ginn Distinguished Professor and the Director in the Wireless Engineering Research and Education Center, Auburn University, Auburn, AL, USA. He is a Distinguished Lecturer of the IEEE Vehicular Technology Society. He is on the Editorial Board of the IEEE TRANSACTIONS ON MULTIMEDIA, the IEEE INTERNET OF THINGS JOURNAL, the IEEE Multimedia, ACM GetMobile, among others. His research interests include wireless networks, multimedia communications, and smart grid. He received the 2015 IEEE ComSoc TC-CSR Distinguished Service Award, the 2013 IEEE ComSoc MMTC Outstanding Leadership Award, and the NSF CAREER Award in 2010. He is a co-recipient of the Best Demo Award from the IEEE SECON 2017, the Best Paper Awards from the IEEE GLOBECOM 2016 and 2015, the IEEE WCNC 2015, and the IEEE ICC 2013, and the 2004 IEEE Communications Society Leonard G. Abraham Prize in the Field of Communications Systems.