IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received August 2, 2018, accepted September 2, 2018, date of publication September 6, 2018,
date of current version September 28, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2868947

Downlink Multi-User Scheduling With Zero-Forcing Precoding in Cognitive Hetnets: From Performance Tradeoff Perspective

ZICHEN CHEN, JIANDONG LI, (Senior Member, IEEE), AND JINJING HUANG
State Key Laboratory of Integrated Service Networks, Xidian University, Xi'an 710071, China

Corresponding author: Jiandong Li (jdli@ieee.org)

This work was supported by the Key Program of the National Natural Science Foundation of China under Grant 91638202 and Grant 61231008.

ABSTRACT In a cognitive heterogeneous network (HetNet), cognitive small cell base stations (SBSs) can perceive and utilize the frequency band belonging to massive MIMO enabled macrocell base stations opportunistically. There is a performance tradeoff between the macrocell-tier and the SC-tier. To restrict the cross-tier interference from SBSs to macrocell users (MUs), we adopt the concept of exclusion region (ER). A multi-user scheduling with zero-forcing precoding strategy is proposed to maximize the utility of the entire HetNet. The involved multi-user scheduling problem is formulated as the maximization of a general utility function. However, the problem is intractable for two reasons: First, it depends on ergodic rate bothering online scheduling. Second, ERs formed by different scheduled MUs may intersect with each other which makes the problem difficult to solve. Therefore, we convert the maximization problem into a simplified version and design an interfering SBS cluster splitting-based iterative user selection algorithm to figure out the scheduled MU set and its corresponding precoder. Simulation results reveal that the proposed scheme achieves significant gains compared with existing strategies and that the proposed strategy is efficient for dense HetNets.

INDEX TERMS Cognitive HetNets, exclusion region, interfering small cell base station cluster, massive MIMO, multi-user scheduling, performance tradeoff.

I. INTRODUCTION

In order to meet the ever growing demands of wireless data transmitting, standard bodies and academic communities have introduced cutting edge techniques to improve data rate and network capacity [1], such as heterogeneous networks (HetNets) [2], massive MIMO [3], cognitive small cell (SC) [4], secure transmissions [5], [6], etc. HetNets of macrocells underlaid with multiple SCs is an emerging technology to address the network performance challenges, in which macrocells can provide large scale coverage, SCs can provide higher spatial reuse and network capacity. With elaborate precoding schemes, a massive MIMO macrocell base station (MBS) can transmit signal concurrently to multiple intended macrocell users (MUs) over the same band with the benefit of space division multiple access (SDMA). Furthermore, it can create energy efficiency and spectrum efficiency downlink transmission [7]. The HetNet of a multi-user massive MIMO macrocell underlaid with dense SCs has been considered as one of the candidate techniques for 5G cellular networks [8]. Unfortunately, the cross-tier interference between the macrocell-tier and the SC-tier has become one of severest challenges for successful HetNets deployment [9]. One of typical approaches for mitigating the interference involves orthogonalizing the frequency allocated to the macrocell-tier and the SC-tier [10]. Nevertheless, this operation mode decreases frequency reuse efficiency [11]. In co-channel deployment scenarios, orthogonalization over time domain is proposed by introducing almost blank subframes (ABSs) where some subframes of the macrocell are reserved for SCs so as to reduce the cross-tier interference, which is also known as enhanced inter-cell interference coordination (eICIC) [12]. An alternative solution is to empower cognitive capabilities to SC base stations (SBSs) to form cognitive HetNets [4]. Cognitive SBSs can perceive channel occupation of the MBS and access the channel opportunistically [13]. When the interference generated by a cognitive SBS to a scheduled MU exceeds a certain threshold I_{th}, the SBS will be obliged to abandon transmitting [14].

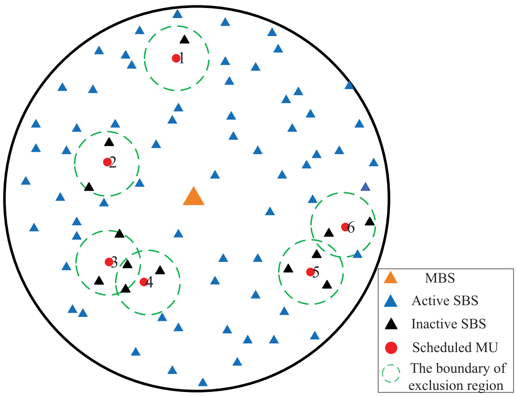


FIGURE 1. The cognitive HetNets model.

Thus, multiple exclusion regions (ERs) are formed around co-scheduled MUs [15].

In this paper, we consider the co-channel cognitive HetNet consisting of a multi-user massive MIMO macrocell underlaid with dense cognitive SCs. Thus, multiple co-scheduled MUs will form multiple ERs, as shown in Fig.1. SBSs within ERs which are called inactive SBSs have to abandon transmitting to protect scheduled MUs. Meanwhile, SBSs outside ERs which are called active SBSs can transmit insusceptibly. In this manner, the dominant cross-tier interference from SBSs to co-scheduled MUs is avoided by the cognitive capabilities of SBSs. However, the cross-tier interference avoidance is at the expense of sacrificing transmitting opportunities of SBSs in ERs. Scheduling MUs will increase network utility of the macrocell-tier, but will reduce network utility of the SC-tier. It can be observed that there is a performance tradeoff between the macrocell-tier and the SC-tier. Therefore, multi-user scheduling strategy of the MBS is crucial for the cognitive HetNets.

Existing researches on multi-user scheduling lie in single-tier network, in which multi-user scheduling strategies depend on the status of single-tier users, e.g., channel quality, relationships between channels, geographical locations, etc. In conventional multi-user MIMO networks, multi-user scheduling strategies have been extensively studied in [16]–[19] and references therein. Dimic and Sidiropoulos [16] consider greedy user selection with two types of precoding schemes, called zero-forcing dirty-paper (ZF-DP) precoding and ZF precoding, respectively. The work in [17] provides a suboptimal user selection with ZF beamforming, in which selected users are semi-orthogonal to each other. Summarizing user selection with ZF precoding schemes in [17] and [18] the work in [18] proposes a general mathematical framework of greedy user selection based on ZF precoding. Huang *et al.* [19] discover flaws of "redundant users" and "local optimum" in previous works, and propose a greedy user selection with swap (GUSS) algorithm. In multi-user massive MIMO networks, conventional multi-user scheduling strategies are inadvisable owing to the fact that huge number of antennas will heavily complicate multi-user scheduling algorithms. Therefore, multi-user scheduling strategies for massive MIMO networks mainly concentrate on designing low-complexity scheduling algorithms and on scheduling strategies with limited channel state information (CSI). The work in [20] proposes a two-stage precoder in FDD massive MIMO networks, in which scheduled MUs are selected from multiple groups based on their covariance matrix and the downlink precoder is decomposed into an outer precoder and an inner-precoder. Considering the randomness of both channel matrixes and locations of MUs, the work in [21] proposes two low-complexity user selection methods for downlink TDD massive MIMO systems. Zhang *et al.* [22] propose a joint beamforming and user scheduling approach based on statistical channel state information (SCI) of MUs. The work in [23] designs a novel two-phase-feedback based user scheduling and beamforming method, in which cones constructed by orthogonal reference beams are designed to select semi-orthogonal MUs.

Although multi-user scheduling with precoder design has been extensively studied, little attention is paid to multi-user scheduling strategies in HetNets. In the co-channel cognitive HetNet consisting of a multi-user massive MIMO macrocell underlaid with dense cognitive SCs, we propose a multi-user scheduling strategy considering not only the gains of the macrocell-tier, but also the loss of the SC-tier. Then, the multi-user scheduling strategy is formulated as the maximization of a general utility function which relies on ergodic rate of both the marocell-tier and the SC-tier. However, the problem is intractable for two reasons: Firstly, it depends on ergodic rate which is difficult to acquire an online solution. Secondly, ERs formed by co-scheduled MUs may intersect with each other which makes the problem intractable. So, we design an interfering SBS cluster (ISC) splitting based online solution to figure out the scheduled MU set and its corresponding precoder.

The key contributions of the paper can be summarized in the following:

- We consider the multi-user scheduling strategy in a co-channel cognitive HetNet, where the MBS transmits signal concurrently to multiple MUs over the same band. Multiple ERs are formed around co-scheduled MUs owing to the cognitive capabilities of SBSs.
- We propose a multi-user scheduling strategy from the perspective of performance tradeoff. The multi-user scheduling strategy considers both gains of the macrocell-tier and loss of the SC-tier.
- The multi-user scheduling strategy is formulated as the maximization of a general utility function, which can achieve sum rate maximization, proportional fairness and harmonic mean fairness according to different utility functions.
- Since the maximization problem is intractable, it is converted into a simplified version. Then, we proposed an ISC splitting based iterative user selection algorithm to figure out the scheduled MU set and its corresponding precoder.

The rest of the paper is organized as follows. Section II describes the network model and transmitting model. Section III gives the problem formulation of the multi-user scheduling strategy. Section IV presents the ISC splitting based multi-user scheduling algorithm. Numerical results are presented in section V. Finally, we conclude this work in section VI.

Notation: we use uppercase boldface letters for matrices and lowercase boldface letters for vectors. For a matrix $\boldsymbol{H}$, the symbols $\boldsymbol{H}^{\mathrm{T}}$, $\boldsymbol{H}^{\mathrm{H}}$, $\boldsymbol{H}^{-1}$, and $\boldsymbol{H}^{\dagger}$ denote the transpose, the conjugate transpose, the inverse, and the pseudo-inverse of matrix $\boldsymbol{H}$, respectively. $\boldsymbol{I}_N$ is $N \times N$ dimension identity matrix. We use $\mathcal{CN}(\boldsymbol{m}, \boldsymbol{N})$ to denote the circular symmetric complex Gaussian distribution with mean $\boldsymbol{m}$ and covariance matrix $\boldsymbol{N}$. The Euclidean norm of vector $\boldsymbol{x}$ is denoted by $\|\boldsymbol{x}\|$. $\mathcal{A}$ denotes a set and $|\mathcal{A}|$ denotes the cardinality of set $\mathcal{A}$. $\mathcal{A}\backslash\mathcal{B}$ denotes complementary set of $\mathcal{B}$ in $\mathcal{A}$. The subscript 0 is dedicated for the macrocell and subscripts of other letters are for SCs.

II. SYSTEM MODEL AND TRANSMITTING MODEL

A. NETWORK MODEL

We consider the downlink of a TDD co-channel cognitive HetNet, in which one macrocell is underlaid with dense cognitive SCs. The MBS is equipped with N antennas and serves a set of MUs in $\mathcal{U}$. Let $\mathcal{S} = \{1, 2, \cdots, |\mathcal{S}|\}$ denote the set of cognitive SBSs which are uniformly distributed under the coverage of the macrocell. Each SBS is equipped with M antennas and serves $|\mathcal{E}_j|$ small cell users (SUs), where $j \in \mathcal{S}$ is the index of the SBS, $\mathcal{E}_j$ denotes SU set served by SBS j $(1 \leq |\mathcal{E}_j| \leq M)$. Each user (MU or SU) is equipped with a single antenna and is only associated to one base station. In the scheduling slot t, the MBS selects a set of co-scheduled MUs $\mathcal{K}$ from the candidate MU pool $\mathcal{U}$, and transmits signal concurrently to selected MUs over the entire band. Since massive MIMO regime is applied, it is assumed that the number of co-scheduled MUs is significantly less than the antenna number of the MBS, i.e. $|\mathcal{K}| \ll N$. We assume that all base stations are perfectly synchronized and perfect CSI is available. It should be emphasized that the availability of perfect CSI is an idealistic assumption for massive MIMO networks. However, the proposed strategy here is still useful as a reference point.

B. CHANNEL MODEL

In this paper, TDD protocol is preferred for the reason that the downlink channel can be estimated with the uplink channel by channel reciprocity [24]. In the scheduling slot t (no more tautology later), the channel between the MBS and MU k is denoted by $\boldsymbol{g}_{0,k,t} \in \mathbb{C}^{N\times 1}$. We assume the channel $\boldsymbol{g}_{0,k,t}$ is equal to a complex fast fading factor times an amplitude factor that accounts for geometric attenuation and shadow fading, i.e.,

$$\boldsymbol{g}_{0,k,t} = \beta_{0,k,t}^{1/2}\boldsymbol{h}_{0,k,t}, \tag{1}$$

where $\beta_{0,k,t}^{1/2}$ is the large-scale fading which consists of path loss and shadow fading. $\beta_{0,k,t} = r_{0,k,t}^{-\alpha}\zeta_{0,k,t}$, where $r_{0,k,t}$ is the distance between the MBS and MU k, $\alpha \in [2, 6]$ is the path loss exponent, $\zeta_{0,k,t}$ is a shadow fading variable with the distribution of $10\log_{10}\zeta_{0,k,t} \sim \mathcal{N}(0, \sigma_i)$. The small-scale fading $\boldsymbol{h}_{0,k,t}$ is an independent and identical distribution (i.i.d.) random vector with the distribution of $\boldsymbol{h}_{0,k,t} \sim \mathcal{CN}(0, \boldsymbol{I}_N)$.

C. INTERFERING SBS CLUSTER MODEL

In the cognitive HetNet, cognitive SBSs can perceive the channel occupation of the MBS and access the channel opportunistically to avoid cross-tier interference. SBSs whose interference to surrounding MUs excesses the interference threshold I_{th}, will stop transmitting to guarantee their performance. Here, we assume that SBSs who interfere MU k as ISC $\mathcal{C}_k$, and

$$\mathcal{C}_k = \{j \in \mathcal{S} : \sum_{l\in\mathcal{E}_j} p_{j,l,t}\left\|\boldsymbol{g}_{j,k,t}^{\mathrm{H}}\boldsymbol{w}_{j,l,t}\right\|^2 \geq I_{th}\}, \tag{2}$$

where $p_{j,l,t}$ is the power allocated to SU $l \in \mathcal{E}_j$ by SBS j, $\boldsymbol{g}_{j,k,t}$ is the interference channel gain between SBS j and MU k, $\boldsymbol{w}_{j,l,t} \in \mathbb{C}^{M\times 1}$ is the unit-norm precoding vector of SBS j for SU l. We assume that each SBS develops its precoder and power allocation only depending on its own SUs regardless of scheduling status of the macrocell. For convenience, the union set of ISCs formed by co-scheduled MUs is denoted by $\mathcal{C}_{\text{sum}}$, where

$$\mathcal{C}_{\text{sum}} = \bigcup_{k\in\mathcal{K}} \mathcal{C}_k. \tag{3}$$

Similar to [25] and reference therein, we assume that interfering channels form cognitive SBSs to MUs are available. In the cognitive HetNet, each SBS can estimate channels from itself to surrounding MUs directly by pilot signal of MUs or indirectly through a band manager [26]. In addition, each SBS can sense pilot signal periodically and can estimate the interference to surrounding MUs. If possible interference from the SBS to an MU exceeds I_{th}, the SBS stops transmitting and feedbacks identifications (IDs) of interfered MUs to the MBS via X2\S1 interface. Otherwise, the SBS transmits signal insusceptibly. After that, the MBS can list all ISCs formed by MUs. For example, if MU k is interfered by SBS l, f, and n, the ISC formed by MU k is denoted by $\mathcal{C}_k = \{l, f, n\}$.

D. TRANSMITTING MODEL

For a scheduled MU $k \in \mathcal{K}$, the downlink received signal is given by

$$\begin{aligned} y_{0,k,t} &= \underbrace{\sqrt{p_{0,k,t}}\boldsymbol{g}_{0,k,t}^{\mathrm{H}}\boldsymbol{w}_{0,k,t}x_{0,k,t}}_{\text{intended signal}} + \underbrace{\sum_{i\in\mathcal{K},i\neq k}\sqrt{p_{0,i,t}}\boldsymbol{g}_{0,k,t}^{\mathrm{H}}\boldsymbol{w}_{0,i,t}x_{0,i,t}}_{\text{inter-user interference signal}} \\ &\quad + \underbrace{\sum_{j\in\mathcal{S}\backslash\mathcal{C}_{\text{sum}}}\sum_{l\in\mathcal{E}_j}\sqrt{p_{j,l,t}}\boldsymbol{g}_{j,k,t}^{\mathrm{H}}\boldsymbol{w}_{j,l,t}x_{j,l,t}}_{\text{cross-tier interference signal}} + \underbrace{n_t}_{\text{noise}}, \end{aligned} \tag{4}$$

where $p_{0,k,t}$ and $p_{j,l,t}$ are the power allocated to MU $k \in \mathcal{K}$ by the MBS and to SU $l \in \mathcal{E}_j$ by SBS j, respectively.

$\sum_{k\in\mathcal{U}} p_{0,k,t} \le P_0$ and $\sum_{l\in\mathcal{E}_j} p_{j,l,t} \le P_s$, where P_0 and P_s are the maximal transmitting power of the MBS and that of each SBS, respectively. $\boldsymbol{g}_{0,k,t} \in \mathbb{C}^{N\times 1}$ and $\boldsymbol{g}_{j,l,t} \in \mathbb{C}^{M\times 1}$ denote channel vectors from the MBS to scheduled MU k and from SBS j to SU $l \in \mathcal{E}_j$, respectively. $\boldsymbol{w}_{0,k,t} \in \mathbb{C}^{N\times 1}$ is the unit-norm precoding vector of the MBS for MU k. $x_{0,k,t} \sim \mathcal{CN}(0,1)$ and $x_{j,l,t} \sim \mathcal{CN}(0,1)$ denote the user data at the MBS towards MU k and at SBS j towards SU $l \in \mathcal{E}_j$, respectively. $n_t \sim \mathcal{CN}(0,\delta^2)$ is the additive white Gaussian noise (AWGN), where δ^2 is the noise power.

For simplicity, we assume that ZF precoding is applied by SBSs, the inter-user interference within each SC is eliminated. Then, the downlink received signal at SU $l \in \mathcal{E}_j$ is given by

$$\begin{aligned} y_{j,l,t} = {} & \underbrace{\sqrt{p_{j,l,t}}\boldsymbol{g}^{\mathrm{H}}_{j,l,t}\boldsymbol{w}_{j,l,t}x_{j,l,t}}_{\text{intended signal}} + \underbrace{\sum_{k\in\mathcal{K}}\sqrt{p_{0,k,t}}\boldsymbol{g}^{\mathrm{H}}_{0,l,t}\boldsymbol{w}_{0,k,t}x_{0,k,t}}_{\text{cross-tier interference signal}} \\ & + \underbrace{\sum_{f\in\mathcal{S}\setminus\mathcal{C}_{\mathrm{sum}}, f\neq j}\ \sum_{i\in\mathcal{E}_f}\sqrt{p_{f,i,t}}\boldsymbol{g}^{\mathrm{H}}_{f,l,t}\boldsymbol{w}_{f,i,t}x_{f,i,t}}_{\text{co-tier interference signal}} + \underbrace{n_t}_{\text{noise}}, \end{aligned} \tag{5}$$

where $\boldsymbol{g}_{0,l,t}$ is the interference channel gain from the MBS to SU $l \in \mathcal{E}_j$.

For MU $k \in \mathcal{K}$, the dominant cross-tier interference is avoided by cognitive capabilities of SBSs. The residual cross-tier interference generated by SBSs outside ERs is treated as noise. Then, the signal to interference plus noise ratio (SINR) of MU k is given by

$$\gamma_{0,k,t} = \frac{p_{0,k,t}\left\|\boldsymbol{g}^{\mathrm{H}}_{0,k,t}\boldsymbol{w}_{0,k,t}\right\|^2}{I_{0,k,t} + \hat{I}_{0,k,t} + \delta^2}, \tag{6}$$

where $I_{0,k,t} = \sum\limits_{i\in\mathcal{K}, i\neq k} p_{0,i,t}\left\|\boldsymbol{g}^{\mathrm{H}}_{0,k,t}\boldsymbol{w}_{0,i,t}\right\|^2$ and $\hat{I}_{0,k,t} = \sum\limits_{j\in\mathcal{S}\setminus\mathcal{C}_{\mathrm{sum}}}\sum\limits_{l\in\mathcal{E}_f} p_{j,l,t}\left\|\boldsymbol{g}^{\mathrm{H}}_{j,k,t}\boldsymbol{w}_{j,l,t}\right\|^2$ are the inter-user interference and the cross-tier interference suffered by MU k, respectively.

Analogously, the SINR of SU $l \in \mathcal{E}_j$ is given by

$$\gamma_{j,l,t} = \frac{p_{j,l,t}\left\|\boldsymbol{g}^{\mathrm{H}}_{j,l,t}\boldsymbol{w}_{j,l,t}\right\|^2}{I_{j,l,t} + \hat{I}_{j,l,t} + \delta^2}, \tag{7}$$

where $I_{j,l,t} = \sum\limits_{f\in\mathcal{S}\setminus\mathcal{C}_{\mathrm{sum}}, f\neq j}\ \sum\limits_{i\in\mathcal{E}_f} p_{f,i,t}\left\|\boldsymbol{g}^{\mathrm{H}}_{f,l,t}\boldsymbol{w}_{f,i,t}\right\|^2$ and $\hat{I}_j = \sum\limits_{k\in\mathcal{K}} p_{0,k,t}\left\|\boldsymbol{g}^{\mathrm{H}}_{0,l,t}\boldsymbol{w}_{0,k,t}\right\|^2$ are the co-tier interference and the cross-tier interference suffered by SU $l \in \mathcal{E}_j$, respectively.

Thus, the ergodic rate of MU $k \in \mathcal{K}$ and that of SU $l \in \mathcal{E}_j$ are given by

$$R_{0,k} = \mathbb{E}\{\log_2\left(1+\gamma_{0,k,t}\right)\}, \tag{8}$$

and

$$R_{j,l} = \mathbb{E}\{\log_2\left(1+\gamma_{j,l,t}\right)\}. \tag{9}$$

III. PROBLEM FORMULATION

To describe the multi-user scheduling problem, we define a scheduling indicator vector $\boldsymbol{\Gamma}_{0,t}$ to denote scheduling status of MUs, where $\boldsymbol{\Gamma}_{0,t} = [\Gamma_{0,1,t}, \cdots, \Gamma_{0,k,t}, \cdots, \Gamma_{0,|\mathcal{U}|,t}]$. When $\Gamma_{0,k,t} = 1$, it indicates that MU k is scheduled and SBSs in $\mathcal{C}_k$ are inactive. Otherwise, MU k is unscheduled and SBSs in $\mathcal{C}_k$ are active, i.e.,

$$\Gamma_{0,k,t} = \begin{cases} 1, & \text{MU } k \text{ is scheduled,} \\ 0, & \text{MU } k \text{ is unscheduled,} \end{cases} \quad k \in \mathcal{U}. \tag{10}$$

We define a utility function $u(z)$ to evaluate user performance, in which $u(z)$ is a concave and monotonically increasing function and $z \in [0,\infty)$ is the variable of the utility function. A typical utility function is $u(z) = z$, which maximizes sum rate. Other alternative choices are $u(z) = \log(z)$ for proportional fairness among users or $u(z) = -1/z$ for harmonic mean fairness. In this paper, we maximize the overall network utility which consists of three portions, the network utility of the macrocell-tier, that of SCs inside ISCs, and that of SCs outside ISCs.

Possible network utility of the macrocell-tier is given by

$$A = \sum_{k\in\mathcal{U}} \Gamma_{0,k,t}u(R_{0,k}). \tag{11}$$

For an ER formed by an MU (Supposing k), the scheduling status of MU k is repellent to active states of SCs in its formed ER. Since the scheduling indicator of MU k is denoted by $\Gamma_{0,k,t}$, active states of SCs in the ER are denoted by $1-\Gamma_{0,k,t}$. Therefore, possible network utility of SCs inside ERs is denoted by

$$B = \sum_{j\in\mathcal{C}_{\mathrm{sum}}}\sum_{l\in\mathcal{E}_j}\left(1-\Gamma_{0,k,t}\right)u(R_{j,l}). \tag{12}$$

When $\mathcal{U}$ and I_{th} are known in advance, the network utility of SCs outside ERs is irrelevant to $\boldsymbol{\Gamma}_{0,t}$. Therefore, the maximization of $A+B$ is selected as the objective function and the multi-user scheduling problem is formulated by

$$\begin{aligned} \mathrm{P1}: \quad & \max_{\Gamma_{0,k,t}} A + B \\ & \text{s.t. } \Gamma_{0,k,t} \in \{0,1\}, \quad \forall k \in \mathcal{U}. \end{aligned} \tag{13}$$

It can be observed that the multi-user scheduling problem relies on ergodic rate. However, practical scheduling strategies welcome online solutions. To seek a scheduling strategy working slot by slot, we denote the achievable rate of MU k and that of SU $l \in \mathcal{E}_j$ by

$$r_{0,k,t} = \log_2\left(1+\gamma_{0,k,t}\right), \tag{14}$$

and

$$r_{j,l,t} = \log_2\left(1+\gamma_{j,l,t}\right). \tag{15}$$

Meanwhile, the time average rate of MU k and that of SU $l \in \mathcal{E}_j$ in scheduling slot t are given by

$$R_{0,k,t} = \frac{1}{t-1}\sum_{s=1}^{t-1} r_{0,k,s}, \tag{16}$$

and

$$R_{j,l,t} = \frac{1}{t-1}\sum_{s=1}^{t-1} r_{j,l,s}. \tag{17}$$

By definition, ergodic rate can be approximately estimated by time average rate with sufficient duration, i.e.,

$$R_{0,k} \approx R_{0,k,t} = \frac{1}{t-1}\sum_{s=1}^{t-1} r_{0,k,s}, \quad t \gg 0, \tag{18}$$

$$R_{j,l} \approx R_{j,l,t} = \frac{1}{t-1}\sum_{s=1}^{t-1} r_{j,l,s}, \quad t \gg 0. \tag{19}$$

Without loss of generality, we take MU k as an example, SUs have similar forms. According to standard stochastic approximation recursions [27], the time average rate of MU k can be updated at every slot, i.e.,

$$R_{0,k,t+1} = R_{0,k,t} + \eta_{0,k,t}(r_{0,k,t} - R_{0,k,t}), \tag{20}$$

where $\eta_{0,k,t}$ is a sufficiently small step-size for MU k which can be either a constant or a known variable. Meanwhile, applying Taylor expansion, we have [28]:

$$u(R_{0,k,t+1}) \approx u(R_{0,k,t}) + \frac{\mathrm{d}u(R_{0,k,t})}{\mathrm{d}(R_{0,k,t})}(R_{0,k,t+1} - R_{0,k,t}), \tag{21}$$

where $\frac{\mathrm{d}u(R_{0,k,t})}{\mathrm{d}R_{0,k,t}}$ is the first-order derivative of the utility function on $R_{0,k,t}$. Substituting (20) into (21), we have

$$u(R_{0,k,t+1}) \approx u(R_{0,k,t}) + \frac{\mathrm{d}u(R_{0,k,t})}{\mathrm{d}(R_{0,k,t})}\eta_{0,k,t}(r_{0,k,t} - R_{0,k,t}). \tag{22}$$

When $t \gg 0$, we have

$$u(R_{0,k}) \approx u(R_{0,k,t}) + \frac{\mathrm{d}u(R_{0,k,t})}{\mathrm{d}(R_{0,k,t})}\eta_{0,k,t}(r_{0,k,t} - R_{0,k,t}). \tag{23}$$

Similarly, the network utility of SU $l \in \mathcal{E}_j$ is approximated by

$$u(R_{j,l}) \approx u(R_{j,l,t}) + \frac{\mathrm{d}u(R_{j,l,t})}{\mathrm{d}(R_{j,l,t})}\eta_{j,l,t}(r_{j,l,t} - R_{j,l,t}), \tag{24}$$

where $\eta_{j,l,t}$ is a sufficiently small step-size for SU $l \in \mathcal{E}_j$.

Since $R_{0,k,t}$, $u(R_{0,k,t})$, $\eta_{0,k,t}$, $\frac{\mathrm{d}u(R_{0,k,t})}{\mathrm{d}R_{0,k,t}}$, $R_{j,l,t}$, $u(R_{j,l,t})$, $\eta_{j,l,t}$, and $\frac{\mathrm{d}u(R_{j,l,t})}{\mathrm{d}R_{j,l,t}}$ are available at slot t, the maximization problem P1 can be decomposed into an optimization problem solved slot-by-slot. The online solution for scheduling slot t corresponds to a sum of weighted achievable rate as follows.

$$\begin{aligned} \text{P2}: \quad & \max_{\Gamma_{0,k,t}} \tilde{A} + \tilde{B} \\ & \text{s.t. } \Gamma_{0,k,t} \in \{0, 1\}, \quad \forall k \in \mathcal{U}, \end{aligned} \tag{25}$$

where

$$\begin{cases} \tilde{A} = \sum_{k\in\mathcal{U}} \Gamma_{0,k,t}\lambda_{0,k,t}r_{0,k,t}, \\ \tilde{B} = \sum_{j\in\mathcal{C}_{\text{sum}}}\sum_{l\in\mathcal{E}_j}\left(1 - \Gamma_{0,k,t}\right)\mu_{j,l,t}r_{j,l,t}, \end{cases} \tag{26}$$

$\lambda_{0,k,t} = \frac{\mathrm{d}u(R_{0,k,t})}{\mathrm{d}R_{0,k,t}}$ and $\mu_{j,l,t} = \frac{\mathrm{d}u(R_{j,l,t})}{\mathrm{d}R_{j,l,t}}$ depend on the derivative of the utility function $u(z)$. In case of $u(z) = z$, $\frac{\mathrm{d}u(R_{0,k,t})}{\mathrm{d}R_{0,k,t}} = 1$ represents sum rate maximization. In case of $u(z) = \log(z)$, $\frac{\mathrm{d}u(R_{0,k,t})}{\mathrm{d}R_{0,k,t}} = \frac{1}{R_{0,k,t}}$ represents proportional fairness among users.

Rearranging the objective function of P2, we have

$$\begin{aligned} \tilde{A} + \tilde{B} = & \sum_{k\in\mathcal{U}} \Gamma_{0,k,t}\lambda_{0,k,t}r_{0,k,t} - \sum_{j\in\mathcal{C}_{\text{sum}}}\sum_{l\in\mathcal{E}_j}\Gamma_{0,k,t}\mu_{j,l,t}r_{j,l,t} \\ & + \sum_{j\in\mathcal{C}_{\text{sum}}}\sum_{l\in\mathcal{E}_j}\mu_{j,l,t}r_{j,l,t}. \end{aligned} \tag{27}$$

Since the third term on the right side of (27) is positive and has nothing to do with $\Gamma_{0,k,t}$, the optimization problem P2 is equivalent to

$$\begin{aligned} \text{P3}: \quad & \max_{\Gamma_{0,k,t}} \sum_{k\in\mathcal{U}} \Gamma_{0,k,t}\lambda_{0,k,t}r_{0,k,t} - \sum_{j\in\mathcal{C}_{\text{sum}}}\sum_{l\in\mathcal{E}_j}\Gamma_{0,k,t}\mu_{j,l,t}r_{j,l,t} \\ & \text{s.t. } \Gamma_{0,k,t} \in \{0, 1\}, \quad \forall k \in \mathcal{U}, \end{aligned} \tag{28}$$

where the first term of the objective function of P3 can be considered as possible gains of the macrocell-tier, and the second term can be considered as the cost paid by the SC-tier. The difference of these two terms can be interpreted as the net gains of the network. It should be noticed that we do not consider the effect of scheduling overhead on network utility. In this paper, the scheduling overhead of an MU and that of a SBS are equal which will not affect scheduling decisions of the MBS.

The solution of the optimization problem P3 is intractable for two reasons. Firstly, optimal variables are discrete. Secondly, any one of optimal variables may be related to multiple different ISCs, which is called multi-couplings. Although the optimal solution of P3 can be acquired by exhaustively searching over all possible MU combinations, the size of searching space increases exponentially with $|\mathcal{U}|$, which is unaffordable for practical systems. Therefore, we propose a low-complexity algorithm to solve P3 in the next section.

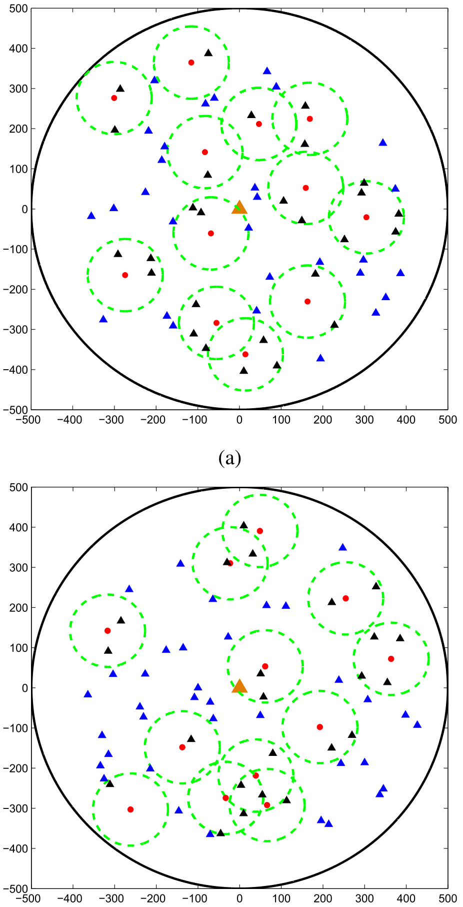
IV. ISC SPLITTING BASED MULTI-USER SCHEDULING STRATEGY

The intractability of the maximization problem P3 mainly lies in multi-couplings. Therefore, we firstly consider a type of special cases in which any one of optimization variables is only related to its corresponding ISC, which is called single-coupling (For the optimal variable $\Gamma_{0,k,t}$, its corresponding ISC is $\mathcal{C}_k$). Then, we propose an iterative user selection algorithm to obtain the scheduled MU set and its corresponding precoder. Secondly, we consider general cases with multi-couplings and propose an ISC splitting scheme to transform general cases into special versions.

A. A TYPE OF SPECIAL CASES

The type of special cases is described as follows. $\forall i, k \in \mathcal{U}$, ISCs formed by them are disjoint, i.e.,

$$\mathcal{C}_i \cap \mathcal{C}_k = \emptyset, \quad \forall i, k \in \mathcal{U}. \tag{29}$$



(a)

(b)

FIGURE 2. The comparison of cognitive HetNets layouts without scheduling and that with the proposed scheduling. The yellow triangle, blue triangles, black triangles, red dots, green dotted line, and black solid line represent the MBS, active SBSs, inactive SBSs, scheduled MUs, boundary of ERs, and boundary of the macrocell, respectively. (a) A random layout of special cases. (b) A random layout of general cases.

A random layout of special cases is shown in Fig. 2 (a). For a special case, the second term of the objective function of P3 can be disassembled into summations for each ER, i.e, the objective function of P3 can be arranged as

$$\begin{aligned}&\sum_{k\in\mathcal{U}}\Gamma_{0,k,t}\lambda_{0,k,t}r_{0,k,t}-\sum_{j\in\mathcal{C}_{\text{sum}}}\sum_{l\in\mathcal{E}_j}\Gamma_{0,k,t}\mu_{j,l,t}r_{j,l,t}\\&\quad=\sum_{k\in\mathcal{U}}\Gamma_{0,k,t}\lambda_{0,k,t}r_{0,k,t}-\sum_{k\in\mathcal{U}}\sum_{j\in\mathcal{C}_k}\sum_{l\in\mathcal{E}_j}\Gamma_{0,k,t}\mu_{j,l,t}r_{j,l,t}\\&\quad=\sum_{k\in\mathcal{U}}\Gamma_{0,k,t}(\lambda_{0,k,t}r_{0,k,t}-\sum_{j\in\mathcal{C}_k}\sum_{l\in\mathcal{E}_j}\mu_{j,l,t}r_{j,l,t}),\end{aligned}\tag{30}$$

where the weighted sum rate of SCs in all ERs is converted into summations of weighted sum rate of SCs in each ER. From (30), the optimal problem P3 can be decomposed into $|\mathcal{U}|$ subproblems. For an MU $k\in\mathcal{U}$, its corresponding subproblem is given by

$$\begin{aligned}\text{P4}:\quad&\max_{\Gamma_{0,k,t}}\Gamma_{0,k,t}(\lambda_{0,k,t}r_{0,k,t}-\sum_{j\in\mathcal{C}_k}\sum_{l\in\mathcal{E}_j}\mu_{j,l,t}r_{j,l,t})\\&\text{s.t. }\Gamma_{0,k,t}\in\{0,1\}.\end{aligned}\tag{31}$$

When $\lambda_{0,k,t}$, $\mu_{j,l,t}$, $r_{0,k,t}$, $r_{j,l,t}$ are available, the solution of P4 is easy to obtain by relaxing $\Gamma_{0,k,t}$ [29]. Let $0\le\Gamma_{0,k,t}\le1$, subproblem P4 is relaxed as a linear programming, i.e.,

$$f(\Gamma_{0,k,t})=\Gamma_{0,k,t}(\lambda_{0,k,t}r_{0,k,t}-\sum_{j\in\mathcal{C}_k}\sum_{l\in\mathcal{E}_j}\mu_{j,l,t}r_{j,l,t}),\tag{32}$$

where $f(\Gamma_{0,k,t})$ is the objective function of P4 after relaxing $\Gamma_{0,k,t}$. The solution of (32) is acquired by letting its derivative be zero. The derivative of (32) is given by

$$\frac{df(\Gamma_{0,k,t})}{d\Gamma_{0,k,t}}=\lambda_{0,k,t}r_{0,k,t}-\sum_{j\in\mathcal{C}_k}\sum_{l\in\mathcal{E}_j}\mu_{j,l,t}r_{j,l,t}.\tag{33}$$

Letting $\frac{df(\Gamma_{0,k,t})}{d\Gamma_{0,k,t}}=0$, the optimal solution of (32) is acquired. Then, the solution of P4 is described as follows. When $(\lambda_{0,k,t}r_{0,k,t}-\sum_{j\in\mathcal{C}_k}\sum_{l\in\mathcal{E}_j}\mu_{j,l,t}r_{j,l,t})\ge0$, $\Gamma_{0,k,t}=1$. Otherwise, $\Gamma_{0,k,t}=0$.

From the perspective of network performance, $\lambda_{0,k,t}r_{0,k,t}$ is the gains of scheduling MU k, $\sum_{j\in\mathcal{C}_k}\sum_{l\in\mathcal{E}_j}\mu_{j,l,t}r_{j,l,t}$ is the cost of scheduling MU k. $(\lambda_{0,k,t}r_{0,k,t}-\sum_{j\in\mathcal{C}_k}\sum_{l\in\mathcal{E}_j}\mu_{j,l,t}r_{j,l,t})$ is the net gains of scheduling MU k. When the net gains is positive, MU k should be scheduled. Otherwise, MU k should be unscheduled. This is consistent with the mathematical framework.

In the above analysis, the rates of MUs are assumed to be known in advance. However, they are difficult to acquire if all scheduled MU are not predetermined. Therefore, we design an iterative searching algorithm to figure out the scheduled MU set and its corresponding precoder.

To describe the iterative searching algorithm, we make the following assumptions. In the ith iteration, the candidate MU set is denoted by $\mathcal{M}_C^{(i)}$. The ith selected MU is denoted by $\mathcal{K}(i)$. The set of all currently selected MUs is denoted by $\mathcal{K}^{(i)}=\{\mathcal{K}(1),\mathcal{K}(2),\cdots,\mathcal{K}(i-1)\}$. The MU scheduling procedures are described as follows.

Initially, the candidate MU set is $\mathcal{M}_C^{(0)}=\mathcal{U}$, and the scheduled MU set is $\mathcal{K}^{(0)}=\emptyset$.

Step 1: Update the candidate MU set $\mathcal{M}_C^{(i)}$.

$$\mathcal{M}_C^{(i)}=\mathcal{U}\backslash\mathcal{K}^{(i)}.\tag{34}$$

Step 2: $\forall k\in\mathcal{M}_C^{(i)}$, possible net gains of scheduling MU k is given by

$$\Delta^{(k)}=\lambda_{0,k,t}r_{0,k,t}-\sum_{j\in\mathcal{C}_k}\sum_{l\in\mathcal{E}_l}\mu_{j,l,t}r_{j,l,t}.\tag{35}$$

Step 3: Select the MU with the maximum net gains in step 2.

$$\mathcal{K}(i)=\arg\max_{k\in\mathcal{M}_C^{(i)}}\Delta^{(k)}.\tag{36}$$

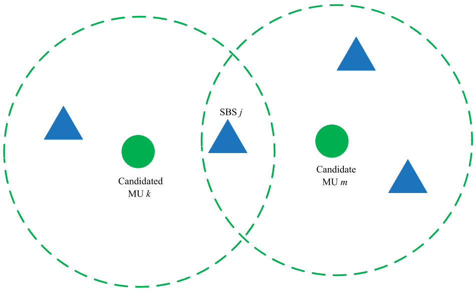


FIGURE 3. An example of two intersecting ISCs. ISC $\mathcal{C}_k$ and $\mathcal{C}_m$ are formed by candidate MU k and by candidate MU m. They are intersecting with each other and SBS j is the shared element.

In step 3, we denote the maximal objective function of (36) by $\Delta_{\max}$. The iterative algorithm is terminated when $\Delta_{\max} < 0$ or entire MUs in $\mathcal{U}$ are selected. $\Delta_{\max} < 0$ implies that scheduling remaining MUs will not increase overall network gains.

B. ISC SPLITTING BASED MULTI-USER SCHEDULING ALGORITHM

In this subsection, we extend network topology to general cases, i.e., there are multi-couplings in the network. Graphically, two or more ISCs may be intersecting. Fig. 2 (b) shows an example of general cases. Due to multi-couplings, we design an ISC splitting scheme to transform multi-couplings into single-coupling. Here, we illustrate the principle of ISC splitting scheme by an example of two intersecting ISCs shown in Fig. 3. Supposing ISC $\mathcal{C}_k$ and $\mathcal{C}_m$ are intersecting and SBS j is the shared element. Possible net gains of candidate MU m and that of candidate MU k are given by

$$\Delta^{(m)} = \lambda_{0,m,t} r_{0,m,t} - \sum_{f \in \mathcal{C}_m} \sum_{i \in \mathcal{E}_f} \mu_{f,i,t} r_{f,i,t}, \tag{37}$$

and

$$\Delta^{(k)} = \lambda_{0,k,t} r_{0,k,t} - \sum_{j \in \mathcal{C}_k} \sum_{l \in \mathcal{E}_j} \mu_{j,l,t} r_{j,l,t}. \tag{38}$$

When $\Delta^{(m)} > 0$ and $\Delta^{(m)} > \Delta^{(k)}$, MU m is the prior scheduled MU. When $\Delta^{(k)} > 0$ and $\Delta^{(k)} > \Delta^{(m)}$, MU k is the prior scheduled MU. If we assume that MU m is the prior scheduled MU and MU k is the candidate MU, possible net gains of MU k is given by

$$\Delta^{(k)} = \lambda_{0,k,t} r_{0,k,t} - \sum_{j \in \overline{\mathcal{C}}_k} \sum_{l \in \mathcal{E}_l} \mu_{j,l,t} r_{j,l,t}, \tag{39}$$

where $\overline{\mathcal{C}}_k$ denotes interfering SBS set excepted SBS j. Therefore, the proprietary interfering SBS set of candidate MU k is denoted by $\overline{\mathcal{C}}_k = \mathcal{C}_k \backslash \{j\}$. In summary, we have

$$\mathcal{C}_m \cup \mathcal{C}_k = \mathcal{C}_m \cup \overline{\mathcal{C}}_k = \overline{\mathcal{C}}_m \cup \mathcal{C}_k, \tag{40}$$

where $\overline{\mathcal{C}}_m = \mathcal{C}_m \backslash \{j\}$. Meanwhile, $\mathcal{C}_m \cap \overline{\mathcal{C}}_k = \emptyset$ and $\overline{\mathcal{C}}_m \cap \mathcal{C}_k = \emptyset$. Obviously, the union set of all ISCs is split into a group of newly disjoint ISCs in a certain order.

Next, we apply the ISC splitting scheme into generalized network cases. In the ith iteration, we assume that the interfering SBS union set formed by all currently selected MUs is denoted by $\mathcal{L}(\mathcal{K}^{(i)}) = \bigcup_{k \in \mathcal{K}^{(i)}} \mathcal{C}_k$. Then, the generalized MU scheduling procedures are described as follows.

Initially, the candidate MU set is $\mathcal{M}_C^{(0)} = \mathcal{U}$, the selected MU set is $\mathcal{K}^{(0)} = \emptyset$, and the interfering SBS union set is $\mathcal{L}(\mathcal{K}^{(0)}) = \emptyset$.

Step 1: Update the candidate MU set $\mathcal{M}_C^{(i)}$ and the interfering SBS union set.

$$\mathcal{M}_C^{(i)} = \mathcal{U} \backslash \mathcal{K}^{(i)}, \quad \mathcal{L}(\mathcal{K}^{(i)}) = \bigcup_{k \in \mathcal{K}^{(i)}} \mathcal{C}_k. \tag{41}$$

Step 2: $\forall k \in \mathcal{M}_C^{(i)}$, possible net gains of scheduling MU k is then calculated by

$$\Delta^{(k)} = \lambda_{0,k,t} r_{0,k,t} - \sum_{j \in \overline{\mathcal{C}}_k} \sum_{l \in \mathcal{E}_l} \mu_{j,l,l} r_{j,l,t}. \tag{42}$$

where $\overline{\mathcal{C}}_k = \mathcal{C}_k \backslash [\mathcal{C}_k \cap \mathcal{L}(\mathcal{K}^{(i)})]$ is the interfering SBS set of MU k excepted the intersection set of $\mathcal{C}_k$ and $\mathcal{L}(\mathcal{K}^{(i)})$. From the perspective of network architecture, the newly interfering SBS set of MU k should exclude SBSs whose active state has been determined by selected MUs.

Step 3: Select the MU with the maximum net gains in step 2.

$$\mathcal{K}\,(i) = \arg \max_{k \in \mathcal{M}_C^{(i)}} \Delta^{(k)}. \tag{43}$$

Since $r_{0,k,t}$ is crucial to determine the scheduled MU $\mathcal{K}(i)$, it is necessary to calculate the allocated power and the precoding vector of MU k. In the ith iteration, ZF precoder of possible scheduled MU set $(\mathcal{K}^{(i)} \cup k)$ is given by

$$\boldsymbol{W}(\mathcal{K}^{(i)} \cup k) = \boldsymbol{G}(\mathcal{K}^{(i)} \cup k)^{\dagger}, \tag{44}$$

where $\boldsymbol{W}(\mathcal{K}^{(i)} \cup k) = [\boldsymbol{w}_{0,\mathcal{K}(1),t}, \cdots, \boldsymbol{w}_{0,\mathcal{K}(i-1),t}, \boldsymbol{w}_{0,k,t}]^{\mathrm{T}}$. $\boldsymbol{G}(\mathcal{K}^{(i)} \cup k) = [\boldsymbol{g}_{0,\mathcal{K}(1),t}, \cdots, \boldsymbol{g}_{0,\mathcal{K}(i-1),t}, \boldsymbol{g}_{0,k,t}]$ is the channel matrix of possible scheduled MU set $(\mathcal{K}^{(i)} \cup k)$ in the ith iteration.

As the dominant cross-tier interference has been avoided by ISCs, and the residual is treated as noise, the allocated power of MU k can be obtained directly by using water-filling, i.e.,

$$p_{0,k,t} = \left(\frac{\tau}{\|\boldsymbol{w}_{0,k,t}\|^2} - 1 \right)^+, \tag{45}$$

where $(\cdot)^+$ denotes $\max\{\cdot, 0\}$, and τ is the water level satisfying

$$\sum_{k \in \mathcal{K}^{(i)} \cup k} (\tau - \|\boldsymbol{w}_{0,k,t}\|^2)^+ = P_0. \tag{46}$$

Thus, we can calculate $r_{0,k,t}$ by the allocated power $p_{0,k,t}$ and the precoding vector $\boldsymbol{w}_{0,k,t}$. This strategy is called multi-user scheduling with direct ZF precoding. The procedures of the strategy are summarized in Algorithm 1.

Algorithm 1 The Iterative Multi-User Scheduling With Direct ZF Precoding

```
Require: $\mathcal{U}$.
Ensure: $\mathcal{K}$, $\boldsymbol{W}_{0,t}$, and allocated power $\boldsymbol{P}_{0,t}$.
  Initialize: $\mathcal{M}_C^{(0)} = \mathcal{U}$, $\mathcal{K}^{(0)} = \emptyset$, $\mathcal{L}(\mathcal{K}^{(0)}) = \emptyset$.
  repeat
    Update the candidate MU set $\mathcal{M}_C^{(i)}$.
    Update the union set of ISCs $\mathcal{L}(\mathcal{K}^{(i)})$.
    for $k \in \mathcal{M}_C^{(i)}$ do
      Calculate the precoding matrix by ZF and allocated power by water-filling.
      Calculate possible net gains of scheduling MU $k$ by (42).
    end for
    Select the MU with maximal net gains;
  until $\Delta_{\max} < 0$ or $\mathcal{K}^{(i)} \bigcap \mathcal{U} = \mathcal{U}$.
  $\mathcal{K} = \mathcal{K}^{(i)}$, precoder and allocated power are values in the $i$th iteration.
```

Algorithm 2 The Iterative Multi-User Scheduling With Indirect ZF Precoding

```
Require: $\mathcal{U}$.
Ensure: $\mathcal{K}$, $\boldsymbol{W}_{0,t}$, and allocated power $\boldsymbol{P}_{0,t}$.
  Initialize: $\mathcal{M}_C^{(0)} = \mathcal{U}$, $\mathcal{K}^{(0)} = \emptyset$, $\mathcal{L}(\mathcal{K}^{(0)}) = \emptyset$, and estimate the precoder $\hat{\boldsymbol{W}}_{0,t} = [\hat{\boldsymbol{w}}_{0,1,t}, \cdots, \hat{\boldsymbol{w}}_{0,|\mathcal{U}|,t}]$ by (47).
  repeat
    Update the candidate MU set $\mathcal{M}_C^{(i)}$.
    Update the union set of ISCs $\mathcal{L}(\mathcal{K}^{(i)})$.
    for $k \in \mathcal{M}_C^{(i)}$ do
      Calculate the allocated power by water-filling.
      Calculate possible net gains scheduling MU $k$ by (42).
    end for
    Select the MU with maximal net gains.
  until $\Delta_{\max} < 0$ or $\mathcal{K}^{(i)} \bigcap \mathcal{U} = \mathcal{U}$.
  $\mathcal{K} = \mathcal{K}^{(i)}$, precoder and allocated power are calculated by ZF precoding and by water-filling with determined $\mathcal{K}$.
```

In addition, we propose a low-complexity strategy called multi-user scheduling with indirect ZF precoding which is implemented in two steps. Firstly, the scheduled MU set is determined by estimated MU rate, and then water-filling power allocation and ZF precoder are calculated based on the scheduled MU set. To estimate MU rate in each iteration, we adopt the maximum ratio transmitting (MRT) precoding, which is proximate to ZF precoding in performance and is simply the conjugate transpose of the downlink channel, i.e.,

$$\hat{\boldsymbol{w}}_{0,k,t} = \frac{\boldsymbol{g}_{0,k,t}^{\mathrm{H}}}{\|\boldsymbol{g}_{0,k,t}\|^2}, \tag{47}$$

where $\hat{\boldsymbol{w}}_{0,k,t}$ is the estimated precoding vector. In this manner, the precoding vector is calculated once in initializing step. After scheduled MU set is determined, ZF precoding is applied to eliminate inter-user interference. The procedures of the multi-user scheduling with indirect ZF precoding algorithm are summarized in Algorithm 2.

C. COMPLEXITY ANALYSIS

The computation complexity is evaluated in terms of the number of complex multiplications. The optimal scheduled MU set (upper bound) can be obtained by exhaustively searching over all possible scheduled MU combinations. Approximately $\sum_{k=1}^{N} \binom{|\mathcal{U}|}{k} k^5 N$ complex multiplication operations are required to complete one selection [21], which is unaffordable in practice. The complexity of the multi-user scheduling with direct ZF precoding strategy is $\mathcal{O}(N^3|\mathcal{U}|)$, which is equivalent to the strategy in [19]. For the multi-user scheduling with indirect ZF precoding strategy, the evaluation of $\hat{\boldsymbol{w}}_{0,m,t}$ involves a multiplication of two $(1 \times N)$ vectors. This operation is replicated over $|\mathcal{U}|$ MUs. Therefore, the complexity of the multi-user scheduling with indirect ZF precoding strategy is roughly $\mathcal{O}(N|\mathcal{U}|)$. Although the complexities of strategies in [21] are $\mathcal{O}(1)$, these strategies are under the premise that the optimal number of scheduled MUs ($K^\star$) is predetermined which is a complicated issue.

V. PERFORMANCE EVALUATION

We assume that the macrocell has a geographical circular coverage with the MBS located at the center. The MBS is equipped with large scale antennas. Each SBS is deployed with 4 antennas and serves 2 SUs. The distance between adjacent antenna elements is half of the wavelength. ZF precoding and water-filling power allocation are applied by SBSs. MUs and SBSs are uniformly scattered within the coverage of the macrocell. The channel model in [30] is followed by SCs. Extra simulation parameters are summarized in Table 1.

TABLE 1. Simulation parameters.

Parameters	Macrocell	Small cell
The radius of cell coverage	500m	40
Number of antennas	128	4
Maximal transmitting power	43dBm	20dBm
Path loss exponent α	3.8	-
Antenna gain	5dBi	5dBi
Standard deviation of shadowing σ_i	8dB	6dB
The power of AWGN δ^2	-96dBm	-96dBm
Noise figure	6dB	6dB
Interference threshold I_{th}	-	-84dBm

We employ the widely-used utility function of sum rate maximization, i.e., $u(z) = z$. Monte Carlo simulation with 10000 times channel realizations is utilized to obtain meaningful simulation results. In addition, the average sum rate of the entire network is chosen as the performance metric, which is averaged over 2000 scheduling slots. We compare the proposed algorithm with several existing strategies as follows.

(1) Conventional multi-user scheduling with ZF precoding strategy [19].

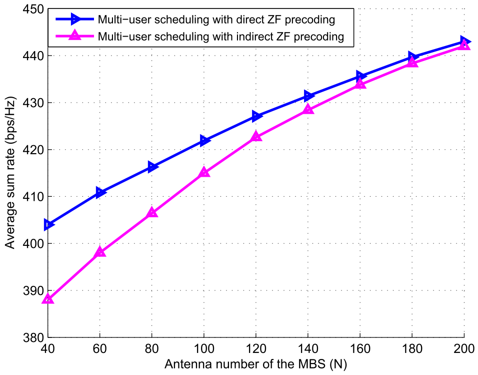


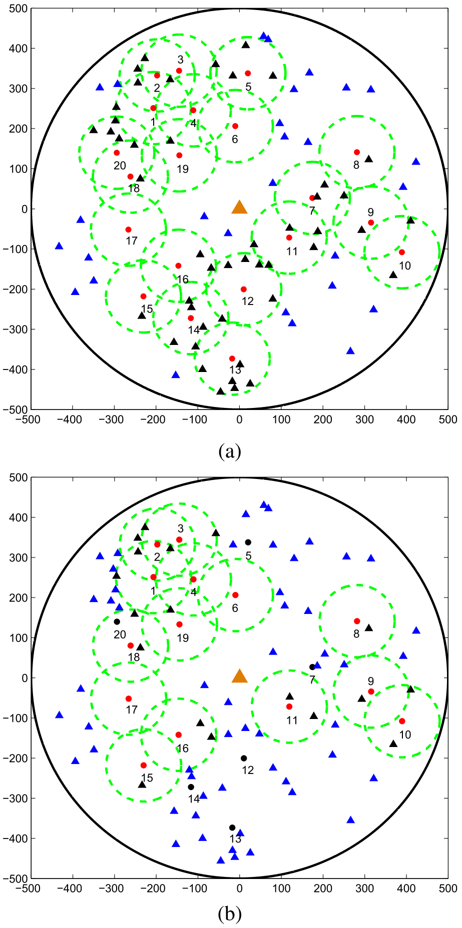
FIGURE 4. **Average sum rate vs. antenna number of the MBS.**

(2) Cognitive HetNets without scheduling [13], where the macrocell is equipped with massive MIMO and all of MUs are scheduled concurrently over the same band.

(3) The low-complexity multi-user scheduling strategy called $\text{K}^\star$-RUS [21].

(4) The optimal solution is acquired by exhaustively searching as the upper bound.

Fig.4 shows the average sum rate versus antenna number of the MBS (N) for the multi-user scheduling with direct ZF precoding strategy and for the multi-user scheduling with indirect ZF precoding strategy, respectively. We assume that $|\mathcal{U}| = 20$ and $|\mathcal{S}| = 60$. It is observed that there is a gap between two strategies. This is because that the scheduled MU set of the multi-user scheduling with indirect ZF precoding strategy is solved by estimated precoder. Moreover, the gap becomes narrower with N increasing. The reason behind this is that the performance of MRT precoding approaches that of ZF precoding with large scale antennas. The simulation result can provide guides in designing networks. When the antenna number of the MBS is small, multi-user scheduling with direct ZF precoding should be employed because of its higher performance and affordable complexity. Otherwise, multi-user scheduling with indirect ZF precoding strategy should be employed. Since the multi-user scheduling with direct ZF precoding strategy always outperforms the multi-user scheduling with indirect ZF precoding strategy, we will only compare the latter (referred to as the proposed strategy hereafter) with other strategies.



(a)

(b)

FIGURE 5. **The comparison of cognitive HetNets layouts without scheduling and that with the proposed scheduling. The yellow triangle, blue triangles, black triangles, red dots, green dotted line, and black solid line represent the MBS, active SBSs, inactive SBSs, scheduled MUs, unscheduled MUs, boundary of ERs, and boundary of the macrocell, respectively. (a) The network snapshot with the proposed strategy. (b) The network snapshot with the proposed strategy.**

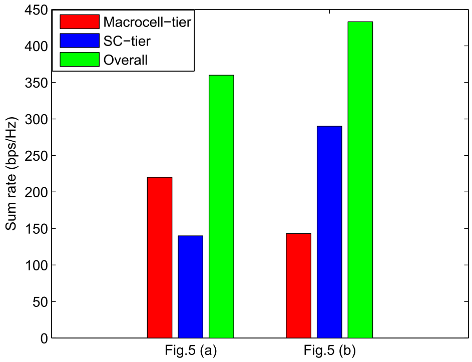


FIGURE 6. **The sum rate before and after the proposed strategy applied in Fig.5.**

In Fig.5, we illustrate the principle of the proposed strategy in a graphical way. We assume that SBSs have the same rate of $r_{j,t}$, that MUs have the same rate of $r_{0,k,t}$, and that $2r_{j,t} < r_{0,k,t} < 3r_{j,t}$, where $r_{j,t}$ is the sum rate of SBS j. For $\forall k \in \mathcal{U}$, we have $\sum_{j\in\mathcal{C}_k} r_{j,t} = |\overline{\mathcal{C}}_k| r_{j,t}$. Therefore, the sum rate of SBSs in $\mathcal{C}_k$ can be represented by the number of SBSs. We further assume that the MU is scheduled if the number of its proprietary interfering SBSs is less than 3, i.e., when $|\overline{\mathcal{C}}_k| < 3$, MU k is scheduled. Otherwise $|\overline{\mathcal{C}}_k| \geq 3$, the MU is unscheduled. Fig.5 (a) shows a random network snapshot of cognitive HetNets without scheduling [13]. Fig.5 (b) shows the network snapshot with the proposed strategy employed. It is observed that the number of dedicated interfering SBSs for MU 5,7,12,13,14, and 20 is greater than or equal to 3. Then, they are unscheduled. The number of dedicated interfering SBSs for other MUs is smaller than 3. Then, they are scheduled.

The sum rate of Fig.5 (a) and that of Fig.5 (b) are illustrated in Fig.6. The left group of histograms represents sum rate of the macrocell-tier, that of the SC-tier, and that of the overall network in Fig.5 (a). Meanwhile, the right group of histograms represents corresponding sum rate in Fig.5 (b). By comparison, it is found that the proposed strategy acquires more sum rate of the SC-tier by sacrificing less sum rate of the macrocell-tier. In this manner, the overall sum rate is improved.

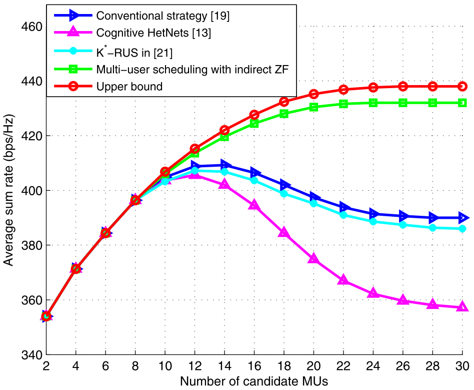


FIGURE 7. Average sum rate vs. number of candidate MUs.

Fig.7 illustrates average sum rate versus the number of MUs ($|\mathcal{U}|$) for the case of 70 SCs. It is observed that significant gains is obtained by the proposed scheme with large $|\mathcal{U}|$. For large number of candidate MUs ($|\mathcal{U}| > 26$), the proposed strategy achieves approximate gains of 40 bps/Hz, 43 bps/Hz, and 71 bps/Hz compared with conventional strategy, K*-RUS, and cognitive HetNets strategy. We discuss performance of these various strategies in three segments. The first segment is $0 < |\mathcal{U}| \leq 10$, the average sum rate of all strategies increases with $|\mathcal{U}|$ growing. This is because that these five strategies will schedule all of candidate MUs in this segment, their performance are similar. With $|\mathcal{U}|$ increasing to the second segment $10 < |\mathcal{U}| \leq 26$, performance of three compared strategies first increase, and at a turning point, they start to decrease, but the speeds diminish with the number of MUs growing. The reason behind the turning point is that the obtained average sum rate of the macrocell-tier is firstly greater than and is then less than the loss average sum rate of the SC-tier. The speed diminishment is due to the fact that the gap between the obtained average sum rate and the lost average sum rate declines with the number of candidate MUs growing. For the proposed strategy, it schedules MUs who can maximize net sum rate of the overall network. When the number of candidate MUs continues to increase into the third segment $|\mathcal{U}| > 26$, the performance of the proposed strategy, that of conventional strategy, and that of K*-RUS tend to be stable. For cognitive HetNets without scheduling strategy in the third segment, although the number of ISCs will increase with the growth of scheduled MUs, newly added ISCs will intersect with existing ISCs with higher probability, the increasing speed of inactive SBSs will decline. Therefore, the performance degradation is slower.

Fig.8 shows the average sum rate versus the number of SCs for the case of $|\mathcal{U}| = 20$. As the number of SCs increasing from 20 to 100, the average sum rate of these strategies increase, but increasing speeds decline. The decline is attributable to the fact that growing of SCs' population will intensify co-tier interference among SCs. When the number of SC is 100, increments of approximate 50 bps/Hz, 53 bps/Hz, and 80 bps/Hz are obtained by the proposed scheme compared to conventional strategy, K*-RUS, and cognitive HetNets, respectively. Furthermore, Fig. 8 reflects that the proposed strategy is more efficient than other strategies for dense HetNets.

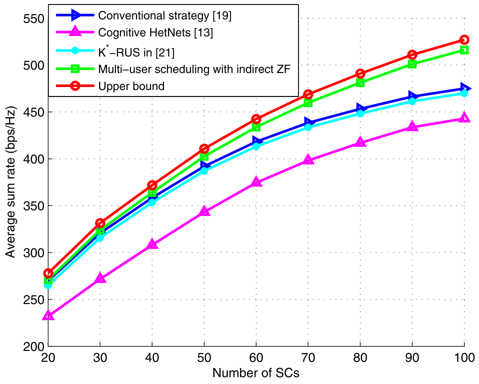


FIGURE 8. Average sum rate vs. number of SCs.

VI. CONCLUSION

In this paper, we consider a cognitive HetNet of a multi-user massive MIMO macrocell underlaid with dense cognitive SCs, in which multiple ERs are formed around co-scheduled MUs due to cognitive capabilities of SBSs. Then, the cross-tier interference from SBSs to MUs are avoided at the expense of sacrificing transmitting opportunities of cognitive SBSs. In such a cognitive HetNet, multi-user scheduling of the MBS depends not only on the macrocell-tier, but also on the SC-tier. Therefore, we propose a multi-user scheduling strategy from the performance tradeoff perspective. The proposed strategy considers both the gains of the macrocell-tier and the loss of the SC-tier. We formulate the multi-user scheduling problem as the maximization of a general utility function which relies on ergodic rate. Next, the maximization problem is converted into an online version which depends on achievable rate. Because of multi-couplings in the network, we design an ISC splitting based multi-user scheduling strategy to figure out the scheduled MU set and its corresponding precoder. Simulation results confirm that our proposed strategy is more favourable for improving network performance in HetNets and is more efficient for dense network.

For future work, we will extend the current research from two directions. The first is to evaluate the performance of the proposed strategy in the cognitive HetNets with limited CSI. The second is to analyze the performance of the proposed strategy under limited cognitive capabilities of SBSs.

REFERENCES

[1] A. Gupta and E. R. K. Jha, "A survey of 5G network: Architecture and emerging technologies," *IEEE Access*, vol. 3, pp. 1206–1232, Jul. 2015.

[2] A. Damnjanovic *et al.*, "A survey on 3GPP heterogeneous networks," *IEEE Wireless Commun.*, vol. 18, no. 3, pp. 10–21, Jun. 2011.

[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] H. ElSawy, E. Hossain, and D. I. Kim, "Hetnets with cognitive small cells: User offloading and distributed channel access techniques," *IEEE Commun. Mag.*, vol. 51, no. 6, pp. 28–36, Jun. 2013.

[5] D. Wang, P. Ren, Q. Du, L. Sun, and Y. Wang, "Security provisioning for MISO vehicular relay networks via cooperative jamming and signal superposition," *IEEE Trans. Veh. Technol.*, vol. 66, no. 12, pp. 10732–10747, Dec. 2017.

[6] D. Wang, P. Ren, J. Cheng, and Y. Wang, "Achieving full secrecy rate with energy-efficient transmission control," *IEEE Trans. Commun.*, vol. 65, no. 12, pp. 5386–5400, Dec. 2017.

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

[8] 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.

[9] E. Hossain, M. Rasti, H. Tabassum, and A. Abdelnasser, "Evolution toward 5G multi-tier cellular wireless networks: An interference management perspective," *IEEE Wireless Commun.*, vol. 21, no. 3, pp. 118–127, Jun. 2014.

[10] A. S. Hamza, S. S. Khalifa, H. S. Hamza, and K. Elsayed, "A survey on inter-cell interference coordination techniques in OFDMA-based cellular networks," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 4, pp. 1642–1670, 4th Quart., 2013.

[11] R. Madan, J. Borran, A. Sampath, N. Bhushan, A. Khandekar, and T. Ji, "Cell association and interference coordination in heterogeneous LTE-A cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1479–1489, Dec. 2010.

[12] K. I. Pedersen, Y. Wang, S. Strzyz, and F. Frederiksen, "Enhanced intercell interference coordination in co-channel multi-layer LTE-advanced networks," *IEEE Wireless Commun.*, vol. 20, no. 3, pp. 120–127, Jun. 2013.

[13] Z. Yan, W. Zhou, S. Chen, and H. Liu, "Modeling and analysis of two-tier HetNets with cognitive small cells," *IEEE Access*, vol. 5, pp. 2904–2912, 2017.

[14] H. ElSawy and E. Hossain, "Two-tier HetNets with cognitive femtocells: Downlink performance modeling and analysis in a multichannel environment," *IEEE Trans. Mobile Comput.*, vol. 13, no. 3, pp. 649–663, Mar. 2014.

[15] U. Tefek and T. J. Lim, "Interference management through exclusion zones in two-tier cognitive networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 3, pp. 2292–2302, Mar. 2016.

[16] G. Dimić and N. D. Sidiropoulos, "On downlink beamforming with greedy user selection: Performance analysis and a simple new algorithm," *IEEE Trans. Signal Process.*, vol. 53, no. 10, pp. 3857–3868, Oct. 2005.

[17] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE J. Sel. Areas Commun*, vol. 24, no. 3, pp. 528–541, Mar. 2006.

[18] J. Wang, D. J. Love, and M. D. Zoltowski, "User selection with zero-forcing beamforming achieves the asymptotically optimal sum rate," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3713–3726, Aug. 2008.

[19] S. Huang, H. Yin, J. Wu, and V. C. M. Leung, "User selection for multiuser MIMO downlink with zero-forcing beamforming," *IEEE Trans. Veh. Technol.*, vol. 62, no. 7, pp. 3084–3097, Sep. 2013.

[20] J. Nam, A. Adhikary, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing: Opportunistic beamforming, user grouping and simplified downlink scheduling," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 876–890, Oct. 2014.

[21] H. Liu, H. Gao, S. Yang, and T. Lv, "Low-complexity downlink user selection for massive MIMO systems," *IEEE Syst. J.*, vol. 11, no. 2, pp. 1072–1083, Jun. 2017.

[22] C. Zhang, Y. Huang, Y. Jing, S. Jin, and L. Yang, "Sum-rate analysis for massive MIMO downlink with joint statistical beamforming and user scheduling," *IEEE Trans. Wireless Commun.*, vol. 16, no. 4, pp. 2181–2194, Apr. 2017.

[23] G. Lee and Y. Sung, "A new approach to user scheduling in massive multiuser MIMO broadcast channels," *IEEE Trans. Commun.*, vol. 66, no. 4, pp. 1481–1495, Apr. 2018.

[24] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?" *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, Feb. 2013.

[25] P. V. Tuan and I. Koo, "Optimal multiuser MISO beamforming for power-splitting SWIPT cognitive radio networks," *IEEE Access*, vol. 5, pp. 14141–14153, 2017.

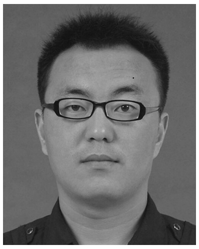
[26] G. Zheng, Z. Ho, E. A. Jorswieck, and B. Ottersten, "Information and energy cooperation in cognitive radio networks," *IEEE Trans. Signal Process.*, vol. 62, no. 9, pp. 2290–2303, May 2014.

[27] H. J. Kushner and G. G. Yin, *Stochastic Approximation Algorithms and Applications*, 2nd ed. Berlin, Germany: Springer-Verlag, 2003.

[28] X. Wang, G. B. Giannakis, and A. G. Marques, "A unified approach to QoS-guaranteed scheduling for channel-adaptive wireless networks," *Proc. IEEE*, vol. 95, no. 12, pp. 2410–2431, Dec. 2007.

[29] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[30] *Evolved universal terrestrial radio access (E-UTRA); Further advancements for E-UTRA physical layer aspects*, document TR 36.814-900, 3rd Generation Partnership Project, Mar. 2010.



ZICHEN CHEN received the B.E. degree in telecommunication engineering from Xidian University, Xi'an, China, in 2009, where he is currently pursuing the Ph.D. degree with the School of Telecommunications Engineering. His main research interests include heterogeneous network, massive MIMO, and distributed algorithms.

JIANDONG LI (M'01–SM'05) received the B.E., M.S., and Ph.D. degrees in communications engineering from Xidian University, Xi'an, China, in 1982, 1985, and 1991, respectively. He was a Visiting Professor with the Department of Electrical and Computer Engineering, Cornell University, Ithaca, NY, USA, from 2002 to 2003. He has been a Faculty Member with the School of Telecommunications Engineering, Xidian University, since 1985, where he is currently a Professor and the Vice Director of the Academic Committee, State Key Laboratory of Integrated Service Networks. His research interests include wireless communication theory, cognitive radio, and signal processing. He was a recipient of the Distinguished Young Researcher from NSFC and the Changjiang Scholar Award from the Ministry of Education, China. He served as the General Vice Chair of ChinaCom 2009 and the TPC Chair of the IEEE ICCC 2013.

JINJING HUANG received the B.S. degree in telecommunications engineering from the Shandong University of Science and Technology, Qingdao, China, in 2010. He is currently pursuing the Ph.D. degree with Xidian University, under the supervision of Prof. J.-D. Li. His research interests focus on inter-cell interference coordination in LTE\LTE-A networks and interference management in heterogeneous cellular networks.