# Semidynamic Green Resource Management in Downlink Heterogeneous Networks by Group Sparse Power Control

# Semi-dynamic Green Resource Management in Downlink Heterogeneous Networks by Group Sparse Power Control

Pan Cao, *Member, IEEE*, Wenjia Liu, John S. Thompson, *Fellow, IEEE*, Chenyang Yang, *Senior Member, IEEE* and Eduard A. Jorswieck, *Senior Member, IEEE*

***Abstract*—This paper addresses an energy-saving problem for the downlink of a cloud-assisted heterogeneous network (HetNet) using a time-division duplex (TDD) model, which aims to minimize the base stations (BSs) sum power consumption while meeting the rate requirement of each user equipment (UE). The basic idea of this work is to make use of the scalability of system configurations such that green resource management can be employed by flexibly switching off some unnecessary hardware components, especially for off-peak traffic scenarios. This motivates us to utilize a flexible BS power consumption formulation to jointly model its signal processing and circuit power, transmit power and backhaul transmission power. Instead of using the integer variables $\{1, 0\}$ to control the "on/off" two status of a BS in most previous work, we employ the group sparsity of a transmit power vector to denote the activity of each frequency carrier (FC) such that the signal processing and circuit power can be scaled with the effective bandwidth, thereby leading to multiple sleep modes for a BS in multi-FC systems. Based on this BS power model and the group sparsity concept, a simplified resource allocation scheme for joint BS-UE association, FC assignment, downlink power allocation and BS sleep modes determination is presented which is based on the average channel statistics computed over the coherence time of the large scale fading (LSF). This semi-dynamic green resource management mechanism can be formulated as a NP-hard optimization problem. In order to make it tractable, the successive convex approximation (SCA)-based algorithm is applied to efficiently find a stationary solution using a cloud-based centralized optimization. Simulation results also verify the effectiveness of the proposed mechanism under the developed BS power consumption model.**

***Index Terms*—Heterogeneous network, energy consumption minimization, group sparsity, power control, green scheduling, fractional frequency reuse, multiple base station sleeping modes, successive convex approximation**

## I. Introduction

The definition of the next generation (5G) networks gives the main focus on providing *ubiquitous* and high data rate services for massive devices [1]. Network densification and offloading, increased bandwidth (e.g., by spectrum sharing [2] and carrier aggregation [3]), and advanced multiple-input and multiple-output (MIMO) techniques (e.g., scaling up the number of antennas [4]) are recognized as the three key technologies for future 5G networks to increase the spectral efficiency [5]. By employing these concepts, future 5G networks are more likely to become increasingly dense, massive and heterogeneous in order to *target very high data rates everywhere*. However, like a double-edge sword, these *dense*, *massive* and *heterogeneous* advances in return may result in high energy consumption if proper green resource management is not adopted, since high data rates provide the possibility to transmit the same or even more information in a shorter time and thus cells may be lightly loaded for much of time (off-peak)[1] [6]. Therefore, if a heterogeneous network (HetNet)[2] is already planned or deployed in a typical area, a question arises:

*Q: How can we save the energy consumption of a HetNet by efficient resource management when rate demands in the network are low?*

This question on green resource management has attracted intensive research since last decade. According to the report from Nokia Networks [7], base stations (BSs) consume over 80 percent of a cellular network's energy consumption, and thus this work focuses on the problem of energy saving for BSs in the downlink of a HetNet. To reduce the energy consumption of BSs, there are three main methods *from the perspective of resource management*: 1) green scheduling (e.g., traffic-offloading and flexible frequency reuse), 2) transmit power allocation and 3) sleep mode for lightly loaded hardware components. Following these three aspects, a brief, comprehensive, yet non-exhaustive review of related work is given as follows.

### A. Related Work

The general BS and user equipment (UE) association is a popular way to improve the overall network performance by scheduling the connections between BSs and UEs such that the

The work of P. Cao and J. Thompson is supported by the UK EPSRC grant number EP/L026147/1; The work of W. Liu and C. Yang is supported by Natural Science Foundation of China (NSFC) under Grant 61120106002; The work of E. Jorswieck was partly funded by the German-Israel Foundation (GIF) for Scientific Research and Development under the Research Grant Number I-1243-406.10.

P. Cao and J. Thompson are with the Institute for Digital Communications (IDCoM), The University of Edinburgh, Edinburgh EH3 9JL, United Kingdom (email: {p.cao, john.thompson}@ed.ac.uk). W. Liu and C. Yang are with the School of Electronics and Information Engineering, Beihang University, Beijing 100191, China (email: {liuwenjia, cyyang}@buaa.edu.cn). E. Jorswieck is with the Chair of Communications Theory, Communications Laboratory, TU Dresden, Dresden 01062, Germany (email: eduard.jorswieck@tu-dresden.de).

[1]In this work, both "off-peak traffic" and "partially loaded scenarios" refer to the same status of a network whose throughput is smaller than its network capacity, e.g., with less active users or lower rate targets.

[2]Hereafter, we use the general "HetNet" to denote all the types of (single-tier or multi-tier) multi-cell environment, because our proposed mechanism is independent of the BSs' tiers/density, and the number of BSs' antennas.

inter-BS interference can be properly managed, see [8], [9] and the references therein for the HetNets. When the green communications is the goal, an adaptive BS-UE association can be used to reduce the network energy consumption by power control. In [10], both the power allocation and BS assignment in non-orthogonal downlink transmission code-division multiple-access (CDMA) communication systems are jointly studied, where each UE is allowed to connect to more than one BS. The authors in [11] propose a joint BS association and power control algorithm to simultaneously maximize the system revenue and minimize the total transmit power consumption such that each UE can be served by the right BS. Two types of BS-UE association problems are addressed in [12] by minimizing the total network power consumption (global throughtput) and minimizing each UE's power consumption (UE equilibrium), respectively. In [13], BS association and downlink beamforming are jointly optimized by minimizing the sum power consumption while guaranteeing a minimum signal-to-interference-and-noise-ratio (SINR) per UE. Instead of studying the BS-UE association under universal spectrum reuse, a joint design of flexible spectrum assignment and BS-UE association might further improve the network performance [14]. Another special case of spectrum reuse is orthogonal frequency division multiple access (OFDMA), which leads to a joint frequency subcarrier assignment and BS-UE association problem. Some recent works on energy efficiency maximization for the downlink multi-cell OFDMA system have been addressed in [15]–[17] and the references therein. In order to make flexible use of spectrum, fractional/partial spectrum reuse among BSs is considered to improve energy efficiency by flexibly improving the bandwidth and also avoiding some significant inter-BS interference in [18], [19].

In addition to the green scheduling and power allocation in the above works, another important way to save network energy consumption is to *completely or partially* turn off some "free" BSs with no/low load, e.g., [20]–[28] and the references therein. For instance, from the point of view of reducing energy consumption, the authors in [27], [28] provide some interesting performance analysis by optimizing the number and density of active BSs according to the varying traffic load in the network. In order to control the "on/off" status of a BS, the integer variables $\{1, 0\}$ are usually introduced and optimized, e.g., in [23]–[26] and in particular [25], [26] also with the consideration of scheduling and transmit power minimization. However, the "on/off" two-status decision for a whole BS might be crude and coarse, since this binary power model implies that all the "on" BSs consume the same constant circuit power in spite of their different traffic loads, which is not true in practical systems. In addition, it is not realistic to switch off and wake up a whole BS in a very short time (dynamically), and also a BS still needs to transmit and receive some basic signals for detection and control even when no UEs are connected. This motivates that hardware components of a network should be as flexible and reconfigurable as possible, since this hardware flexibility and scalability can be exploited to further improve energy efficient/saving performance, by reconfiguring the BS hardware components according to the effectively used resources [21], [29], [30]. In particular, the authors in [29] also suggest a modular hardware design approach based on a multi-core microprocessor in order to avoid the dependence of different hardware components on each other and enable flexible reconfiguration. Thus, it is possible to flexibly turn off or deactivate only unnecessary hardware components to reduce the signal processing and circuit power, e.g., the antenna muting/adaptation [31], [32]. In the time domain, the discontinuous transmission (DTX) [33] based on the varying channel quality is another example of hardware inactivity, which is extended in [34] by combining the scheduling and power control to minimize the BS energy consumption.

However, most previous work has not considered jointly solving green scheduling (BS-UE association and FC assignment), downlink power allocation and multiple BS sleep modes, and the system configurations are not as flexible and scalable as possible based on *some of* the following assumptions: *R1. both BSs and UEs are equipped with a single antenna; R2. each BS is allowed to serve one UE at a time on each FC; R3. each UE is allowed to be connected to only one BS at a time; R4. each UE is allowed to operate on only one FC at a time; R5. each FC is not allowed to be reused by two or multiple UEs at a time; R6. simple transmit power control for each UE on a FC is adopted, e.g., fixed power allocation or fractional power control; R7. the "on/off" two-status BS sleep mode is used.* In fact, these "restricted" system assumptions should be and can be relaxed due to recent hardware and signal processing capabilities in order to further improve the green performance.

### *B. System Assumptions and Explanations*

With the purpose of reducing BSs energy consumption, we desire to flexibly and jointly implement green scheduling, transmit power allocation and multiple sleep modes for BSs in a HetNet based on the following system assumptions

*A1*. *Multi-Antenna System:* Each BS is equipped with multiple or even a large scale antenna array. MIMO technology is maturing and is being incorporated into emerging wireless broadband standards like long-term evaluation (LTE) [35]. Furthermore, the recent massive MIMO (with large-scale antenna arrays) can increase the capacity 10 times or more and simultaneously improve the radiated energy efficiency on the order of 100 times, and is considered as an exciting 5G potential technology [36], [37];

*A2*. *Dual Multi-Connectivity/Access Enabled Operation:* Each BS can simultaneously serve more than one UEs on each individual FC, since a multi-antenna BS can transmit multiple data streams independently and simultaneously to multiple users using multiple degrees of freedom (i.e., multi-user transmission) [4]. Meanwhile, each UE can be simultaneously served by more than one BSs on each individual FC. One example is the coordinated multi-point (CoMP) transmission, which exploits the potential interference links for desired data transmission and plays an important role in interference-limited small cells to enhance the effective strength of signals [38], [39];

*A3. Dual Multi-Carrier Enabled Operation:* Each BS and each UE can operate simultaneously on one or more FCs. 3GPP Release 12 has already proposed the inter-site carrier aggregation in the HetNet, for example, a device can maintain parallel connections to a macro cell on some of the low frequency bands and to a small cell at higher frequency band [40];

*A4. Spectrum Reuse or Not:* Each FC is allowed to be reused by any BS set and UE set. By allowing spectrum reuse, a defined number of BSs or UEs are granted rights to use the same spectrum. The shared license model provides 5G systems and deployments with an important flexibility to use spectrum that is under-utilized by other services or fully utilized by other equipments which are located far away to provide additional capacity [41];

*A5. Frequency-Selective Fading Channel Model:* The same communication link on different FCs may experience different channel qualities. This is generally true in realistic wireless communication environment, since radio transmissions on different FCs usually have different wave propagation properties.

*A6. Flexible Transmit Power Allocation:* Flexible downlink transmit power is allocated subject to the per-BS transmit power budget. In this work, $N_k$ linear power ampliers are equipped at each BS $k$, since a linear amplifier is effectively transparent to the carriers modulation and the number of carriers and can linearly amplify all types of signals, e.g., a multi-carrier signal where each carrier has a constant, non-constant, or a mixture of both envelope [42].

In contrast to the assumptions *R1-R7*, these general system assumptions *A1-A6* allow us to formulate a series of the flexible scheduling and efficient resource management problems: such as *P1. BS-UE association problem (BS-selection and "many-to-many" assignment)*, *P2. BS/UE-FC assignment problem (FC-selection and "many-to-many" assignment)*, *P3. downlink transmit power allocation problem*, *P4. intra-carrier interference management problem (a side-product of P1-P3)*, and *P5. flexible BS power model (multiple sleeping modes enabled)*. In order to efficiently and jointly solve these resource management problems, we assume that all BSs in the HetNet are connected to and controlled by a central processor (CP)[3] via a backhaul network (in fact, this work requires only a low backhaul overhead) such that the high computation load of BSs can be transferred to the supercomputer at the CP, which avoids allocating an advanced processor to each BS (low cost) and reduce inter-BS information-exchange overhead for implementing an iterative coordinated algorithm (low overhead). In particular, [43], [44] provide some suggestions on architectures, flexible operation and centralized management for a cloud-assisted HetNet.

[3]The CP could be either the central data center in the Cloud radio access network (C-RAN) or a macro BS who has the capability to do central optimization for the entire network.

### C. Contributions

Consider a cloud-assisted HetNet with the assumptions *A1-A6*. all the BSs, FCs, time blocks, transmit power can be considered as the available radio frequency "resources", and can form a "pool" (i.e., through the supercomputer at the CP). The output of a pre-defined centralized optimization of green resource management will give the answer to Question *Q*. Therefore, this work is aimed to design a flexible and efficient green resource management mechanism. More precisely, the main contributions along with the organization of this paper are listed as follows.

- In Section II: We propose a semi-dynamic green resource management mechanism, which is implemented in two time scales: 1) The green scheduling, downlink transmit power allocation and BS sleep modes are jointly optimized and determined at the CP in a centralized fashion only based on the large scale fading (LSF) values, and thus these strategies are fixed while the LSF values stay constant; 2) The low-complexity maximum ratio transmission (MRT) beamforming is designed and employed locally at each BS based on the instantaneous small scale fading (SSF) coefficients. Compared with the previous dynamic and long-term resource management mechanisms, this semi-dynamic green resource management scheme has the following advantages: 1) It is semi-dynamic and also gains the benefit of varying LSF by dynamically employing the MRT beamforming; 2) It has a low computation and overhead demand for BSs in the dynamic transmission, and transfers the LSF values based optimization to the CP (the slowly-varying LSF values based optimization is not as delay-sensitive as dynamic transmission);
- In Section III: Since the BSs' signal processing circuit power is flexibly scaled by the effective bandwidth, some unnecessary hardware components of the unassigned FCs can be switched off to reduce the signal processing power consumption rather than the whole BS. This leads to multiple signal processing power levels that can be adapted flexibly to the varying traffic load. Inspired by [45], [46], where the $\ell_0$ norm of a beamforming vector is used to *dynamically* denote the integer variables $\{1, 0\}$, we employ group sparsity of *a transmit power vector* to *semi-dynamically* denote the activity of a FC, and then use a $\log$-based expression to better approximate the $\ell_0$ norm than the mixed $\ell_1/\ell_2$ norm approximation in [45], [46]. Based on this idea, a flexible and scalable BS downlink power consumption model is developed, which jointly contains signal processing and circuit power, downlink transmit power and backhaul transmission power. Furthermore, this BS power consumption formulation is a function of a single transmit power vector, and provides the potential to jointly solve the above problems *P1-P5*;
- In Section IV, we derive a closed-form expression to approximate the average achievable rate based on the channel estimation for the time-division duplex (TDD) model. Based on this average rate expression and the flexible power model, we formulate a *semi-dynamic* BSs

energy consumption minimization problem subject to UEs' rate constraints based on the slowly-varying LSF values. Solving this optimization problem provides solutions to the problems *P1-P5*. Since this big optimization problem is shown to be a NP-hard problem, we apply a successive convex approximation (SCA)-based algorithm in Section V to solve it efficiently, and its convergence to a stationary solution is proved.

*Notations:* $|\mathcal{X}|$ and $|\boldsymbol{x}|$ denote the number of the elements of a set $\mathcal{X}$ and a vector $\boldsymbol{x}$; $\mathcal{X}(i)$ denotes the $i$-th element in the set $\mathcal{X}$; $\mathcal{X}_1 \backslash \mathcal{X}_2$ denotes the set $\mathcal{X}_1$ but excluding all the elements in the set $\mathcal{X}_2$; $\mathrm{diag}[\boldsymbol{x}]$ denotes a diagonal matrix with the elements in $\boldsymbol{x}$ as its diagonal elements; $\binom{n}{L}$ denotes the number of $n$-combinations for a $L$-element set.

## II. System Model

Consider the downlink transmission in a cloud-assist HetNet, where $K$ BSs communicate with $L$ active single-antenna UEs employing $F$ orthogonal FCs, and all BSs are connected to the CP via a backhaul network. Let $\mathcal{K} \triangleq \{1, 2, \cdots, K\}$, $\mathcal{L} \triangleq \{1, 2, \cdots, L\}$ and $\mathcal{F} \triangleq \{1, 2, \cdots, F\}$ denote the index set of the BSs, UEs and FCs, respectively. This setup is denoted by $\mathcal{K} \times \mathcal{L} \times \mathcal{F}$. Based on the general system assumptions *A1-A6* in Section I-B, we let $N_k$ and $W_f$ Hz denote the number of antenna of BS $k \in \mathcal{K}$ and the bandwidth of FC $f \in \mathcal{F}$. Let $p_{k,\ell}^f \geq 0$ denote the downlink transmit power at BS $k$ allocated for the transmission to UE $\ell \in \mathcal{L}$ on FC $f \in \mathcal{F}$. The transmit power $\{p_{k,\ell}^f\}_{\ell \in \mathcal{L}, f \in \mathcal{F}}$ at each BS $k$ are allowed to be flexibly allocated to the $LF$ channels but subject to the per-BS transmit power budget $P_{BS,k}^{max}$, i.e., $\sum_{\ell=1}^{L} \sum_{f=1}^{F} p_{k,\ell}^f \leq P_{BS,k}^{max}$. To be clear, some abbreviations and variables used in this paper are listed in Table I.

TABLE I: Abbreviations and Variables

| | |
|---|---|
| LSF, SSF | Large scale fading, small scale fading |
| A-LSF | The age of large scale fading during which LSF values stay constant |
| $\tau_f$ | The length of pilot sequence on FC $f$ in channel training |
| $\beta_{1,f}, \beta_{2,f}$ | No. of symbols transmitted in an A-LSF, and in a coherence time |
| $\mathcal{K}, \mathcal{L}, \mathcal{F}$ | The set of $K$ BSs, the set of $L$ UEs, the set of $F$ FCs |
| $\mathcal{U}_k, \mathcal{B}_\ell$ | The set of UEs *initially* selected by BS $k$, the set of BSs *initially* serving UE $\ell$ |
| $p_{k,\ell}^f$ | Transmit power from BS $k$ to UE $\ell$ on FC $f$ |
| $\boldsymbol{p}_{BS,k}^f$ | Transmit power vector from BS $k$ to all the UEs in $\mathcal{U}_k$ on FC $f$ |
| $\boldsymbol{p}_{BS,k}$ | Transmit power vector from BS $k$ to all the UEs in $\mathcal{U}_k$ on all the FCs |
| $\boldsymbol{p}_{FC,f}$ | Transmit power vector from all the BSs in $\mathcal{K}$ to all the UEs in $\mathcal{L}$ on FC $f$ |
| $\boldsymbol{p}_{UE,\ell}$ | Transmit power vector from all the BSs in $\mathcal{B}_\ell$ to UE $\ell$ on all the FCs |
| $\boldsymbol{p}$ | Transmit power vector from all the BSs in $\mathcal{K}$ to their UEs in $\{\mathcal{U}_k\}_{k \in \mathcal{K}}$ on all the FCs |
| $P_{BS}$ | Sum BS power consumption in the downlink of a HetNet |
| $R_\ell^f$ | Average rate of UE $\ell$ on FC $f$ during an A-LSF [bits/Hz/second] |
| $R_\ell(\boldsymbol{p})$ | Average sum rate of UE $\ell$ on all the FCs during an A-LSF [bits/second] |

### A. Channel Model

We assume that the channel on each FC is quasi-static block-fading which is constant for a number of *symbol intervals*[4] [47]. Let $\boldsymbol{h}_{k,\ell}^f = \sqrt{\alpha_{k,\ell}^f} \tilde{\boldsymbol{h}}_{k,\ell}^f \in \mathbb{C}^{N_k \times 1}$ be the instantaneous channel state information (CSI) from BS $k \in \mathcal{K}$ to UE $\ell \in \mathcal{L}$ on FC $f \in \mathcal{F}$ in a certain time slot, where $\alpha_{k,\ell}^f$ denotes the LSF gain including path loss and shadowing, and $\tilde{\boldsymbol{h}}_{k,\ell}^f$ denotes the corresponding SSF vector where each entry is assumed to satisfy independent and identically distribution (i.i.d.) with zero mean and unit covariance [4], [48]. The *age of LSF (A-LSF)* is defined as the time duration over which the LSF of a communication link is considered to be not varying. The time duration over which the SSF stays constant is in fact the *coherence time*. In many mobile radio situations, the A-LSF is usually tens or hundreds of times longer than the coherence time of the SSF [47]. Without loss of generality, we assume $\beta_{1,f}$ and $\beta_{2,f}$ symbols can be transmitted during an A-LSF and a coherence time on FC $f$.

[4]The symbol interval denotes the time consumed for a transmission of one symbol.

### B. Green Resource Management Mechanism

In terms of resource management, dynamic design based on instantaneous CSI significantly benefits channel gains by adjusting strategies with the varying CSI but at the cost of high complexity. In most practical mobile communication scenarios, it is usually not allowed to design complicated instantaneous transmission strategies (e.g., by the high overhead required and high-complexity iterative algorithms) because of the limited coherence time. In contrast, the long-term fixed transmission strategies for a long time duration have a very low complexity but usually result in a very inefficient usage of the resources because of the mismatch between the fixed strategies and the varying CSI. This motivates us to design a *semi-dynamic* hybrid resource management mechanism:

*M1. MRT Beamforming:* During each coherence time, the low-overhead and low-complexity MRT downlink beamforming scheme is used. Each BS can design the MRT beamforming patterns for its serving UEs *locally* based on only the instantaneous CSI of the desired links, which has a low computation time (the remaining time can be left for uplink/downlink transmission) and only low backhaul overhead is needed by the coordinated BSs to

adjust phases if coherent CoMP transmission is desired.[5] One beamforming design is sufficient for each coherence time of the SSF;

*M2*. *Resource Management:* During each A-LSF, green resource management problem is optimized at the CP based on only the LSF values. The solution will suggest the strategies for scheduling, transmit power allocation and BS sleep modes, and *these strategies are fixed for a whole A-LSF*. Only one implementation is needed for each A-LSF.

In *M1*, no optimization but only the computation of the simple MRT beamforming pattern is required. Thus, our main focus will be on the optimization in *M2*, which only requires that LSF values are available at the CP. Therefore, the basic idea of this semi-dynamic green resource management mechanism is to design the low-complexity MRT beamforming dynamically but use the fixed scheduling, power allocation and BS sleep modes optimized in *M2* during an A-LSF computation hence we call it a *semi-dynamic* method. The main advantage of this mechanism is to reduce the computation and overhead requirement for BSs in the dynamic transmission (short-delay and low-cost), and to transfer the optimization based on the slowly-varying LSF values to the supercomputer at the CP and this optimization is not as delay-sensitive as dynamic transmission. The outline of how this mechanism is implemented will be shown in Section V-E.

### C. Channel Acquisition

In order to implement *M1* and *M2*, the acquisition of SSF and LSF are required, respectively. Some symbol intervals within each coherence time might be taken for channel training, e.g., by pilot sequence transmission, and the remainder is left for downlink data symbol transmission[6].

In this work, TDD operation model is employed, because the feedback phase under the frequency-division duplex (FDD) operation can be eliminated by using *channel reciprocity* and additionally the pilot overhead might be reduced for multi-antenna systems, especially for massive MIMO [50]. In the uplink channel training, all UEs transmit pilot sequences to their associated BSs on the assigned FCs. Let $\sqrt{\tau_f}\boldsymbol{\phi}_\ell^f$ with $||\boldsymbol{\phi}_\ell^f|| = 1$ be the training vector with the length $\tau_f$ transmitted from UE $\ell$ with the transmit power $p_{UE,\ell}^f$ to its associated BS $k$ on an assigned FC $f$. Let $\mathcal{U}_{FC,f} \subseteq \mathcal{L}$ denote the set of UEs who reuse FC $f$.[7] Then, by employing minimum mean square error (MMSE) estimation, the SSF from a typical UE $\ell \in \mathcal{U}_{FC,f}$ to its associated BS $k$ on FC $f$ can expressed as

**Lemma 1** *The MMSE estimate of the SSF $\tilde{\boldsymbol{h}}_{k,\ell}^f$ can be expressed as*

$$\tilde{\boldsymbol{h}}_{k,\ell}^f = \hat{\boldsymbol{h}}_{k,\ell}^f + \boldsymbol{e}_{k,\ell}^f, \tag{1}$$

*where $\hat{\boldsymbol{h}}_{k,\ell}^f \sim \mathcal{CN}(\boldsymbol{0}, \delta_{k,\ell}^f \boldsymbol{I})$ is independent of the estimation error $\boldsymbol{e}_{k,\ell}^f \sim \mathcal{CN}(\boldsymbol{0}, (1-\delta_{k,\ell}^f)\boldsymbol{I})$ with*

$$\delta_{k,\ell}^f \triangleq \frac{\tau_f p_{UE,\ell}^f \alpha_{k,\ell}^f}{\tau_f p_{UE,\ell}^f \alpha_{k,\ell}^f + \sum_{j \in \mathcal{U}_{FC,f}^\ell \setminus \{\ell\}} \tau_f \alpha_{k,j}^f p_{UE,j}^f + W_f \sigma^2}, \tag{2}$$

*where $W_f\sigma^2$ denotes the thermal noise power linearly with the operating bandwidth $W_f$.* □

*Proof:* See Apendix A. ■

**Remark 1** *When no pilot sequence is reused ($|\mathcal{U}_{FC,f}| = 1$), the channel estimation quality in* (2) *becomes $\delta_{k,\ell}^f = \frac{\tau_f p_{UE,\ell}^f \alpha_{k,\ell}^f}{\tau_f p_{UE,\ell}^f \alpha_{k,\ell}^f + W_f\sigma^2}$, and thus the channel estimation error $1-\delta_{k,\ell}^f$ becomes negligible as $\delta_{k,\ell}^f \to 1$ when $\tau_f p_{UE,\ell}^f \alpha_{k,\ell}^f$ is sufficiently large and $W_f$ is not very large. Interestingly,* (2) *also implies that pilot sequences can be reused on the same FC without significant performance loss by those UEs with small LSF gains or low uplink training power to the same BS.* □

In terms of the LSF values, they can be easily estimated at BSs and then reported to the CP via a backhaul network, and this procedure requires a very low overhead because LSF values are scalers. Since the LSF values depend on the specific locations of UEs in realistic communication environments, it is possible to employ a LSF map-based method to estimate LSF values at the CP directly instead of via backhaul transmission [51].

**Definition 1** *A LSF map is defined as a set of LSF values of dense sampling locations in a geographic area. A "point" on the LSF map contains $KF$-dimension LSF values of the downlink channels from $K$ BSs to the corresponding geo-locations on $F$ FCs, respectively.* □

A LSF map can be generated offline by measuring the LSF values of sampling locations in advance once the deployment is given [52], and thus it can be used as a *prior information* stored at the CP to implement the optimization in *M2*. For example, combining a LSF map and current UEs' locations (maybe provided by GPS), the LSF values in next A-LSF can be estimated based on UEs' mobility prediction [53].

[5]This is also the motivation for us not to use joint precessing, which requires the *dynamic* centralized beamforming design for the coordinated BSs and thus causes high overhead (transmit/receive channel vectors and beamforming vectors) and also high latency, but to use multiple signal enhancement in the CoMP that only requires the inter-BS phase adjustment [49]. In principle, some other beamforming schemes could be also employed here, such as zero-forcing (ZF) and minimum mean sum error (MMSE) beamforming, if each BS has an advanced processor to implement the $N_k$-dimension matrix inverse calculations required by ZF and MMSE beamforming design because each inverse calculation has a very high complexity of $\mathcal{O}(N_k^3)$ when $N_k$ becomes large.

[6]The uplink data transmission is not considered here in order to focus on the downlink transmission, since the total network energy is mainly consumed by BSs in the downlink transmission. Otherwise, it is equivalent to consider the "coherence time" used in this work to be a shorter one excluding the uplink transmission time.

[7]In fact, the channel estimation is implemented based on the optimized scheduling result in *M2*, i.e., the determined BS-UE association and FC assignment. Thus, each UEs' set $\mathcal{U}_{FC,f}, \forall f \in \mathcal{F}$ and their served BSs are already known before dynamical channel estimation. Without loss of generality, we assume $\mathcal{U}_{FC,f} \neq \emptyset$.

### D. Initial BS-UE Association

Let $\mathcal{U}_k^f \subseteq \mathcal{L}$ and $\mathcal{B}_\ell^f \subseteq \mathcal{K}$ denote the UEs set simultaneously served by BS $k \in \mathcal{K}$ and the BSs set simultaneously serving UE $\ell \in \mathcal{L}$, respectively, on FC $f \in \mathcal{F}$. Note that some UEs in $\mathcal{U}_k^f$ might not be in the "cell" of BS $k$ because of the CoMP transmission.

**Lemma 2** *For the setup $\mathcal{K} \times \mathcal{L} \times \mathcal{F}$, there exist at most $\sum_{k=1}^{K} \sum_{n=1}^{\min(FN_k, L)} \binom{n}{L}$ possible solutions to the BS-UE association problem in P1.* □

*Proof:* In principle, it is possible for each BS $k$ equipped with $N_k$ antennas to simultaneously and independently serve up to $N_k$ UEs on each FC $f$, and thus up to $\min(FN_k, L)$ UEs can be served by BS $k$ if it serves different UEs set on different FCs (i.e., $\mathcal{U}_k^f \cap \mathcal{U}_k^{\overline{f}} = \emptyset, \forall f \neq \overline{f}$). Then, the proposed result can be obtained by solving a combinatorial problem. ■

In order to remove unlikely solutions to reduce the complexity, we propose an initial BS-UE association to shrink the solutions set as follows. Each BS $k$ with $N_k$ antennas initially selects $N_k$ UEs with the strongest LSF gains on each FC to form its initial set of UEs. Without loss of generality, we assume $\mathcal{U}_k \triangleq \mathcal{U}_k^1 = \mathcal{U}_k^2 = \cdots \mathcal{U}_k^F$ and $|\mathcal{U}_k| \leq N_k$ (the inequality happens when $L < N_k$). After selecting UEs by all BSs, each UE $\ell \in \mathcal{L}$ might be simultaneously selected by multiple BSs for a potential CoMP transmission. We let $\mathcal{B}_\ell \triangleq \mathcal{B}_\ell^1 = \mathcal{B}_\ell^2 = \cdots \mathcal{B}_\ell^F$ denote the initial BSs set consisting of all the serving BSs who initially select UE $\ell$.

**Remark 2** *In general, it is reasonable to assume that each UE $\ell \in \mathcal{L}$ is initially selected by at least one BS, i.e., $|\mathcal{B}_\ell| \geq 1$. In fact, it is rare that a UE cannot be initially selected by any BS, since BSs are equipped with multiple antennas and the BSs deployment is in practice based on UEs' traffic load density. If it really happens, it means that there exist more UEs than the network capacity can support or the non-selected UEs suffer from very bad channel conditions, and thus they should be deactivated during the next A-LSF.* □

After the initial BS-UE association, the number of feasible solutions to Problem *P1* is reduced to $\Pi_{\ell=1}^{L}(|\mathcal{B}_\ell|!)$, thereby resulting in $\Pi_{\ell=1}^{L}(|\mathcal{B}_\ell|! \times F!)$ feasible solutions to the FC assignment problem *P2*. The power model in Section III will be used to the algorithm in Section V of the paper to find a good solution from these candidates.

## III. BSs Power Consumption Model

For the setup $\mathcal{K} \times \mathcal{L} \times \mathcal{F}$ after initial BS-UE association, the downlink transmit power $\{p_{k,\ell}^f\}_{k \in \mathcal{B}_\ell, \ell \in \mathcal{L}, f \in \mathcal{F}}$ forms an *irregular*[8] three-dimensional "tensor" with the size of $|\mathcal{B}_\ell| \times L \times F$. In particular, the status of a link from BS $k$ to UE $\ell$ on FC $f$ can be implied by $p_{k,\ell}^f$. More precisely, the link is *on* if $p_{k,\ell}^f > 0$. Otherwise, it is *off*. This motivates us to propose a general BSs downlink energy consumption model based on the transmit power control.

[8]The irregularity is because $|\mathcal{B}_\ell|$ might be different for each UE $\ell$.

### A. BSs Downlink Power Consumption Model

Before showing the BS power consumption model, we first give some definitions.

**Definition 2** *We let*

$$\boldsymbol{p}_{BS,k}^f \triangleq \left[p_{k,\mathcal{U}_k(1)}^f, p_{k,\mathcal{U}_k(2)}^f, \cdots, p_{k,\mathcal{U}_k(|\mathcal{U}_k|)}^f\right]^T \in \mathbb{R}_+^{|\mathcal{U}_k| \times 1},$$
$$\boldsymbol{p}_{BS,k} \triangleq \left[\boldsymbol{p}_{BS,k}^1, \boldsymbol{p}_{BS,k}^2, \cdots, \boldsymbol{p}_{BS,k}^F\right]^T \in \mathbb{R}_+^{|\mathcal{U}_k| F \times 1},$$
$$\boldsymbol{p} \triangleq [\boldsymbol{p}_{BS,1}, \boldsymbol{p}_{BS,2}, \cdots, \boldsymbol{p}_{BS,K}]^T \in \mathbb{R}_+^{F \sum_{k=1}^{K} |\mathcal{U}_k| \times 1}$$

*denote the transmit power of BS $k$ to all the UEs in $\mathcal{U}_k$ on FC $f$, the transmit power of BS $k$ to all the UEs in $\mathcal{U}_k$ on all the FCs, and the transmit power at all the $K$ BSs to their all initially selected UEs on all the FCs, respectively.* □

Let $\boldsymbol{T}_{BS,k}$ and $\boldsymbol{T}_{BS,k}^f$ denote $F|\mathcal{U}_k| \times F\sum_{k=1}^{K}|\mathcal{U}_k|$ and $|\mathcal{U}_k| \times F\sum_{k=1}^{K}|\mathcal{U}_k|$ selective matrices only consisting of $\{0,1\}$ such that $\boldsymbol{p}_{BS,k} = \boldsymbol{T}_{BS,k}\boldsymbol{p}$ and $\boldsymbol{p}_{BS,k}^f = \boldsymbol{T}_{BS,k}^f \boldsymbol{p}$, respectively.

In the initial BS-UE association, each BS $k$ is allowed to connect to $N_k$ UEs on all $F$ FCs. However, this initial *maximum-connectivity* rarely happens because it is usually inefficient and unnecessary for a HetNet to meet the UEs' transmission rate requirement, especially in off-peak traffic scenarios. Therefore, many elements of $\boldsymbol{p}_{BS,k}$ and $\boldsymbol{p}$ would be zeros, which implies that these transmit power vectors have the (group) sparse property.

**Definition 3** *A vector is group sparse if it has a grouping of its components and the components within each group are likely to be either all zeros or not. Let $\boldsymbol{x} \triangleq [\boldsymbol{x}_1^T, \boldsymbol{x}_2^T, \cdots, \boldsymbol{x}_G^T]^T$ be a $M \times 1$ vector with $G$ non-overlapping groups, where the vector $\boldsymbol{x}_g$ denotes the $g$-th group of the size $M_g \times 1$ satisfying $\sum_{g=1}^{G} M_g = M$. The weighted group sparsity of the vector $\boldsymbol{x}$ is defined by*

$$||\boldsymbol{x}||_{0,\boldsymbol{w}}^{G,M_g} \triangleq \sum_{g=1}^{G} w_g \cdot \text{sign}(||\boldsymbol{x}_g||_0), \tag{3}$$

*where $\boldsymbol{w} \triangleq [w_1, w_2, \cdots, w_G]$ with $w_g$ as the weight of the group $\boldsymbol{x}_g$ and*

$$\text{sign}(||\boldsymbol{x}_g||_0) = \begin{cases} 0 & \text{when } \boldsymbol{x}_g = \boldsymbol{0} \quad (4a) \\ 1 & \text{otherwise.} \quad (4b) \end{cases}$$

*When $\boldsymbol{w} = \boldsymbol{1}$, we use $||\boldsymbol{x}||_0^{G,M_g}$ to denote the standard unweighted group sparsity $\ell_0$ norm.* □

Inspired by this sparsity property, we propose to employ the *group sparsity* of the transmit power vectors to denote the activity of FCs. For example, $||\boldsymbol{p}||_0^{K,F|\mathcal{U}_k|}$ can be used to count the number of active BSs. Let $P_{BS}$ be the BSs sum power consumption in the downlink of a HetNet. Then, $P_{BS}$ can be modeled by transmit power vectors as follows.

**Proposition 1** *The BSs sum power consumption in the down-*

*link of a HetNet can be modeled as*

$$P_{BS} \triangleq \underbrace{\sum_{k=1}^{K} P_k^{sleep_0} + \sum_{k=1}^{K} ||\boldsymbol{p}_{BS,k}||_{0,\boldsymbol{\mu}_k}^{F,|\mathcal{U}_k|}}_{circuit\ \&\ signal\ processing\ power} + \underbrace{\sum_{f=1}^{F}\left(1-\frac{\tau_f}{\beta_{2,f}}\right)\sum_{k=1}^{K}\frac{1}{\eta_k}\mathbf{1}^T\boldsymbol{p}_{BS,k}^f}_{downlink\ transmit\ power} + \underbrace{P_{haul}\frac{R_{haul}}{C_{ref}}}_{backhaul\ power} \quad (5)$$

*where $P_k^{sleep_0}$ denotes the basic static power consumption to support the deep-sleep mode*[9]*, and $\boldsymbol{\mu}_k \triangleq [P_{sp,k}^1, P_{sp,k}^2, \cdots, P_{sp,k}^F]$ denotes the weights for the weighted group sparsity where $P_{sp,k}^f$ denotes the weight for the $f$-th group of $\boldsymbol{p}_{BS,k}$ and is expressed by [54]*

$$P_{sp,k}^f = N_k \frac{W_f}{10\ MHz}(P'_{BB} + P'_{RF}), \quad (6)$$

*where $P'_{BB}$ and $P'_{RF}$ are some reference baseband and RF related signal processing power consumption per 10 MHz bandwidth; and $\eta_k \in (0,1)$ denotes the downlink power amplifier (PA) efficiency ratio of BS $k$; and $P_{haul}$ is the reference backhaul power consumption for a backhaul collection of wireless links of a reference capacity $C_{ref}$ ($C_{ref} = 100Mbit/s$ in [55]) and $R_{haul}$ is the average total backhaul transmission rate.* □

### B. Explanation: Terms in Power Consumption Model

The proposed BS power consumption model in (5) is explained term by term as follows:

*1. Circuit & Signal Processing Power [54]:* 1) $\sum_{k=1}^{K} P_k^{sleep_0}$ denotes the very basic *static* power consumption of all the BSs to support their deep-sleep modes, where $P_k^{sleep_0}$ is power consumption when BS $k$ is in the deep-sleep mode, e.g., the power consumed by the DC-DC power supply, mains supply and active cooling system. This static power $P_k^{sleep_0}$ is usually different for different types of BSs. 2) $\sum_{k=1}^{K} ||\boldsymbol{p}_{BS,k}||_{0,\boldsymbol{\mu}_k}^{F,|\mathcal{U}_k|}$ denotes the power consumption by the baseband (BB) interface and the signaling of RF transceiver (RF-TRX) of all the BSs. The power consumption of the BB interface is mainly contributed by carrier aggregation, filtering, FFT/IFFT, modulation/demodulation, signal detection, channel coding/decoding, and the RF-TRX power consumption mainly depends on the bandwidth, the number of antennas and the resolution of the analogue-to-digital conversion.

**Remark 3** *We employ $||\boldsymbol{p}_{BS,k}||_{0,\boldsymbol{\mu}_k}^{F,|\mathcal{U}_k|}$ to count the number of effective FCs assigned to BS $k$, which allows that each BS to have maximum $(F!+1)$-level signal processing power by turning off partial hardware components according to different effective (assigned) bandwidth*[10]*. This term is load-dependent. For example, if a BS is required to support higher data rates of UEs, more FCs might be assigned at the cost of higher signal processing power. Otherwise, a BS could consume less power. Therefore, multi-level signal processing power enables multiple sleep modes for a BS, which can be determined by group sparsity power control based on UEs' rate requirements.* □

*2. Downlink Transmit Power:* A BS or UE can operate *simultaneously and in parallel* on different FCs (similar to the FDD mode). This parallel operation allows different length of pilot sequences for channel training on different FCs. The parameter $1-\frac{\tau_f}{\beta_{2,f}}$ denotes the ratio of downlink transmission time to the whole time period on a typical FC $f$. This term computes the total downlink transmit power consumption by all the BSs on all the FCs, while in fact, only the transmit power of the assigned FCs are counted because $\{p_{k,\ell}^f\}$ are zeros for un-assigned FCs.

*3. Backhaul Power:* This term is to measure the power consumption by the backhaul overhead, usually including the exchange of the CSI, transmission data and the signaling between coordinated BSs (e.g., in the iterative processing). The backhaul power consumption highly depends on the mechanism/algorithm itself. For instance, our proposed semi-dynamic resource management mechanism has no need for the backhaul communication during the channel training and only a very low backhaul overhead required in the MRT beamforming pattern design if the coherent CoMP transmission is employed. The main overhead is consumed by releasing the downlink data from the core network to the active BSs. Therefore, in our scenario the average total resulting backhaul rate for each UE is approximately its average downlink data rate[11], thereby

$$R_{haul} \approx \sum_{\ell=1}^{L} R_\ell(\boldsymbol{p}), \quad (7)$$

where $R_\ell(\boldsymbol{p})$ is defined in bits/s as the average downlink transmission rate for UE $\ell$ without consideration of the modulation.

The proposed BSs power consumption model in (5) is expressed as a function of transmit power vector $\boldsymbol{p}$. This implies that a series of resource management problems, such as the trade-offs between the BSs energy consumption and downlink transmission rate and the problems *P1-P4* in Section I-B, can be jointly solved by optimizing a single variable $\boldsymbol{p}$.

## IV. Downlink Transmission Rate and Problem Formulation

In this work, we desire to minimize BSs sum power consumption while each UE's required downlink rate is guaranteed. The downlink rate of an individual UE is first derived as follows.

### A. Downlink Transmission Rate

Given an initial BS-UE association, the average transmission rate of each UE $\ell \in \mathcal{L}$ during $T_{LSF}$ can be expressed

[9]The deep-sleep mode denotes the status of a BS without assigned FCs for downlink data transmission when $\boldsymbol{p}_{BS,k} = \mathbf{0}$. of BS $k$

[10]From (6), it implies that the signal processing power for each FC is different if all individual FCs have different bandwidth.

[11]In this setup, synchronization signaling, the inter-BS phase adjustment in MRT beamforming design, and the power allocation result announcement from the CP are also needed via backhaul links, which are not considered herein because of their very low overhead.

as

$$R_\ell(\boldsymbol{p}) = \sum_{f=1}^{F} \left(1 - \frac{\tau_f}{\beta_{2,f}}\right) W_f R_\ell^f \tag{8}$$

where $1 - \frac{\tau_f}{\beta_{2,f}}$ denotes the downlink data transmission time fraction in an A-LSF, and $R_\ell^f$ denotes the rate contribution from $\mathcal{B}_\ell$ to UE $\ell$ on FC $f$, i.e.[12],

$$R_\ell^f = \mathbb{E}_{\tilde{\boldsymbol{h}}}\left\{ \log_2 \left(1 + \frac{\sum_{k\in\mathcal{B}_\ell} |\boldsymbol{h}_{k,\ell}^{f,H} \boldsymbol{w}_{k,\ell}^f|^2}{W_f\sigma^2 + \mathrm{Inter}_{\mathrm{BS}\ell}^{\mathrm{f}} + \mathrm{Intra}_{\mathrm{BS}\ell}^{\mathrm{f}}}\right)\right\} \tag{9}$$

where

$$\mathrm{Inter}_{\mathrm{BS}\ell}^{\mathrm{f}} \triangleq \sum_{\overline{k}\in\mathcal{K}\setminus\mathcal{B}_\ell} \sum_{j\in\mathcal{U}_{\overline{k}}} |\boldsymbol{h}_{\overline{k},\ell}^{f,H} \boldsymbol{w}_{\overline{k},j}^f|^2 \tag{10}$$

$$\mathrm{Intra}_{\mathrm{BS}\ell}^{\mathrm{f}} \triangleq \sum_{k\in\mathcal{B}_\ell} \sum_{\overline{\ell}\in\mathcal{U}_k\setminus\{\ell\}} |\boldsymbol{h}_{k,\ell}^{f,H} \boldsymbol{w}_{k,f}^f|^2 \tag{11}$$

denote the inter-BS and the intra-BS interference to UE $\ell$ on FC $f$, respectively, and $\mathbb{E}_{\tilde{\boldsymbol{h}}}\{\}$ denotes the expectation only with respect to the SSF coefficients because the LSF values stay constant within an A-LSF, and $\boldsymbol{w}_{k,\ell}^f \in \mathbb{C}^{N_k\times 1}$ denotes the instantaneous downlink beamforming designed based on the estimated CSI at BS $k$ for UE $\ell$ on FC $f$.

**Lemma 3** *By using the MRT beamforming* $\boldsymbol{w}_{k,\ell}^f = \sqrt{p_{k,\ell}^f}\overrightarrow{\hat{\boldsymbol{h}}}_{k,\ell}^f$ *where* $p_{k,\ell}^f$ *is the fixed downlink transmit power within* $T_{LSF}$ *and* $\overrightarrow{\hat{\boldsymbol{h}}}_{k,\ell}^f \triangleq \frac{\hat{\boldsymbol{h}}_{k,\ell}^f}{||\hat{\boldsymbol{h}}_{k,\ell}^f||}$*, the average rate* $R_\ell^f$ *in* (9) *is approximately expressed as*

$$R_\ell^f \approx \log_2\left(1 + \frac{\sum_{k\in\mathcal{B}_\ell} p_{k,\ell}^f \alpha_{k,\ell}^f \left(\delta_{k,\ell}^f (N_k - 1) + 1\right)}{W_f\sigma^2 + \mathbb{E}_{\tilde{\boldsymbol{h}}}\{\mathrm{Inter}_{\mathrm{BS}\ell}^{\mathrm{f}}\} + \mathbb{E}_{\tilde{\boldsymbol{h}}}\{\mathrm{Intra}_{\mathrm{BS}\ell}^{\mathrm{f}}\}}\right), \tag{12}$$

*where*

$$\mathbb{E}_{\tilde{\boldsymbol{h}}}\{\mathrm{Inter}_{\mathrm{BS}\ell}^{\mathrm{f}}\} \triangleq \sum_{\overline{k}\in\mathcal{K}\setminus\mathcal{B}_\ell} \sum_{j\in\mathcal{U}_{\overline{k}}} p_{\overline{k},j}^f \alpha_{\overline{k},\ell}^f \tag{13}$$

$$\mathbb{E}_{\tilde{\boldsymbol{h}}}\{\mathrm{Intra}_{\mathrm{BS}\ell}^{\mathrm{f}}\} \triangleq \sum_{k\in\mathcal{B}_\ell} \sum_{\overline{\ell}\in\mathcal{U}_k\setminus\{\ell\}} p_{k,\overline{\ell}}^f \alpha_{k,\ell}^f, \tag{14}$$

*and* $\delta_{k,\ell}^f$ *is defined in* (2). □

*Proof:* See Appendix B. ■

**Remark 4** *The approximation is because* $\mathbb{E}_x\{\log_2(1 + \frac{f_1(x)}{f_2(x)})\} \approx \log_2(1 + \frac{\mathbb{E}_x\{f_1(x)\}}{\mathbb{E}_x\{f_2(x)\}})$ *is used, which is widely used and partially justified in the performance analysis for the multi-antenna systems (e.g., [56]). In particular, simulations in [57] imply this approximation has a high accuracy, especially for large scale antenna arrays.* □

[12]This rate expression is achieved by combining coherently all received desired signals at symbol level, which requires phase synchronization among the coordinated BSs.

### B. Problem Formulation

A semi-dynamic green resource management problem of BSs sum power minimization by group sparse power control is formulated as follows

$$\min_{\boldsymbol{p}\geq\boldsymbol{0}} \quad P_{BS} \tag{15a}$$

$$\text{s.t.} \quad \sum_{f=1}^{F} \left(1 - \frac{\tau_f}{\beta_{2,f}}\right) W_f R_\ell^f \geq \gamma_\ell, \ \forall \ell \in \mathcal{L} \tag{15b}$$

$$\mathbf{1}^T(\boldsymbol{T}_{BS,k}\boldsymbol{p}) \leq P_{BS,k}^{max}, \ \forall k \in \mathcal{K} \tag{15c}$$

where the objective function $P_{BS}$ is shown in (5), and $R_\ell^f$ in downlink transmission rate constraint (15b) is based on (12), and the constraint (15c) denotes per-BS transmit power constraint because of the hardware limits.

However, it is challenging to solve (15) directly. One reason is that it is a well-known NP hard problem to minimize the group sparsity ($\ell_0$ norm) in (3). Another reason is that the term $\sum_{f=1}^{F}\left(1 - \frac{\tau_f}{\beta_{2,f}}\right) W_f R_\ell^f$ with $R_\ell^f$ (12) in a coupled structure with the transmit power is like the sum rate expression of a single-input and single-output (SISO) interference network and also leads to a NP-hard problem in optimization. The goal of this work is to efficiently compute high-quality suboptimal solutions of Problem (15) by the centralized computation at the CP.

### C. Problem Reformulation

In order to make the problem (15) tractable, it is a common approach to relax a group sparsity $\ell_0$-norm to a mixed $\ell_2/\ell_1$ norm. The weighted group sparsity of a vector $\boldsymbol{x}$ in (3) is approximately expressed as $||\boldsymbol{x}||_{0,\boldsymbol{w}}^{G,|\boldsymbol{x}_g|} \approx \sum_{g=1}^{G} w_g ||\boldsymbol{x}_g||_2$, which is non-smooth but convex (its minimization is known as a group *Lasso* problem). However, [58] and [59] provided a comparison of serval non-convex approximations of $\ell_0$ norm and suggested that the following log-based approximation usually has a better sparse recovery performance

$$\begin{aligned} ||\boldsymbol{x}||_{0,\boldsymbol{w}}^{G,|\boldsymbol{x}_g|} &= \lim_{\epsilon\to 0} \sum_{g=1}^{G} w_g \frac{\log(1 + \epsilon^{-1}\mathbf{1}^T\boldsymbol{x}_g)}{\log(1+\epsilon^{-1})} \\ &\approx \sum_{g=1}^{G} w_g \frac{\log(1 + \epsilon^{-1}\mathbf{1}^T\boldsymbol{x}_g)}{\log(1+\epsilon^{-1})}, \end{aligned} \tag{16}$$

where $\epsilon$ in (16) is set to be a very small constant. The following simulations in this paper imply the choice of $\epsilon$ has a very low impact on the performance.

Based on (16) and (7), BSs sum power consumption in (5) approximately becomes

$$\begin{aligned} \widehat{P}_{BS} = &\sum_{k=1}^{K} P_k^{sleep_0} + \sum_{k=1}^{K}\sum_{f=1}^{F} P_{sp,k}^f \frac{\log(1 + \epsilon^{-1}\boldsymbol{t}_{k,f}^T\boldsymbol{p})}{\log(1+\epsilon^{-1})} \\ &+ \sum_{f=1}^{F}\left(1 - \frac{\tau_f}{\beta_{2,f}}\right)\sum_{k=1}^{K}\frac{1}{\eta_k}\boldsymbol{t}_{k,f}^T\boldsymbol{p} + P_{haul}\sum_{\ell=1}^{L}\frac{R_\ell(\boldsymbol{p})}{C_{ref}}, \end{aligned} \tag{17}$$

where $\boldsymbol{t}_k \triangleq \boldsymbol{T}_{BS,k}^T\mathbf{1}, \boldsymbol{t}_{k,f} \triangleq \boldsymbol{T}_{BS,k}^{f,T}\mathbf{1}$ and $R_\ell(\boldsymbol{p})$ in (12).

The average individual UE rate on FC $f$ in (12) can be rewritten in a vector-form as

$$R_\ell^f = \log_2\left(1+\frac{\boldsymbol{\alpha}_{\mathcal{B}_\ell,\ell}^{f,T}\boldsymbol{p}}{W_f\sigma^2+\boldsymbol{\alpha}_{\mathcal{K},\bar{\ell}}^{f,T}\boldsymbol{p}}\right) = \log_2\left(W_f\sigma^2+\boldsymbol{\alpha}_{\mathcal{K},\ell}^{f,T}\boldsymbol{p}\right)-\log_2\left(W_f\sigma^2+\boldsymbol{\alpha}_{\mathcal{K},\bar{\ell}}^{f,T}\boldsymbol{p}\right), \quad (18)$$

where $\boldsymbol{\alpha}_{\mathcal{B}_\ell,\ell}^f$ is a $LF|\mathcal{B}_\ell|\times 1$ all-zeros vector except for the corresponding positions of $\{\alpha_{k,\ell}^f(\delta_{k,\ell}^f(N_k-1)+1)\}_{k\in\mathcal{B}_\ell}$, and $\boldsymbol{\alpha}_{\mathcal{K},\bar{\ell}}^f$ is similarly defined. In (18), we define $\boldsymbol{\alpha}_{\mathcal{K},\ell}^f \triangleq \boldsymbol{\alpha}_{\mathcal{B}_\ell,\ell}^f + \boldsymbol{\alpha}_{\mathcal{K},\bar{\ell}}^f$. Observe that $R_\ell^f$ in (18) is a difference of two concave (DC) functions of $\boldsymbol{p}$.

Based on the reformulation in (17) and in (18) of the rate constraint and objective function, respectively, after moving the constant terms in the objective function Problem (15) becomes

$$\min_{\boldsymbol{p}\geq\boldsymbol{0}} \sum_{k=1}^K\sum_{f=1}^F\left(P_{sp,k}^f\frac{\log(\epsilon+\boldsymbol{t}_{k,f}^T\boldsymbol{p})}{\log(\epsilon+1)}+\left(1-\frac{\tau_f}{\beta_{2,f}}\right)\frac{\boldsymbol{t}_{k,f}^T\boldsymbol{p}}{\eta_k}\right) \quad (19a)$$

$$\text{s.t.} \sum_{f=1}^F\left(1-\frac{\tau_f}{\beta_{2,f}}\right)W_f\Big(\log_2\left(W_f\sigma^2+\boldsymbol{\alpha}_{\mathcal{K},\ell}^{f,T}\boldsymbol{p}\right) - \log_2\left(W_f\sigma^2+\boldsymbol{\alpha}_{\mathcal{K},\bar{\ell}}^{f,T}\boldsymbol{p}\right)\Big)\geq\gamma_\ell,\ \forall\ell\in\mathcal{L} \quad (19b)$$

$$\boldsymbol{t}_k^T\boldsymbol{p}\leq P_{BS,k}^{max},\ \forall k\in\mathcal{K}, \quad (19c)$$

where the total backhaul power consumption term is omitted in (19a), because the rate constraint (19b) will be optimally achieved with "equality", i.e., $R_\ell(\boldsymbol{p})=\gamma_\ell$ (constant term). However, Problem (19) is still difficult to solve, since it is a *concave-minimization* problem with *DC constraints*.

## V. SCA-based Algorithms and Solutions

In this section, the SCA-based algorithm is applied to compute the locally optimal solutions of the non-convex problem (19). The basic idea of the SCA-based algorithm (in spirit of [60], [61]) is to iteratively 1) construct a surrogate function as an upper bound for each objective/constraint function at the current solution and then 2) optimize the problem with surrogate functions which yields the next estimation of the variables.

### A. Technical Preliminaries

Consider the following non-convex optimization problem:

$$\min_{\boldsymbol{x}\in\mathbb{R}^M}\ y(\boldsymbol{x}) \quad (20a)$$

$$\text{s.t.}\ \ c_j(\boldsymbol{x})\leq 0,\ j=1,\cdots,J,\quad \boldsymbol{x}\in\Omega, \quad (20b)$$

where $y, c_j:\mathbb{R}^M\to\mathbb{R}$ are non-convex but smooth functions with the form of

$$y(\boldsymbol{x})\triangleq y^+(\boldsymbol{x})-y^-(\boldsymbol{x}),\ c_j(\boldsymbol{x})\triangleq c_j^+(\boldsymbol{x})-c_j^-(\boldsymbol{x}),\ \forall j \quad (21)$$

where $y^+, y^-, c_j^+, c_j^- : \mathbb{R}^M\to\mathbb{R}$ are continuous convex functions, and $\Omega$ is a convex set in $\mathbb{R}^M$. We define $\mathcal{X}\triangleq\{\boldsymbol{x}\in\Omega: c_j(\boldsymbol{x})\leq 0,\ j=1,\cdots,J\}$.

Problem (20) is a *DC program with DC constraints* (non-convex in general). By the SCA, a common scheme to generate a surrogate function is to *linearize* the non-convex functions by using a first-order Taylor series. For example, either the *completely linearized (CL)* function

$$y^{CL}(\boldsymbol{x},\boldsymbol{z}) = y(\boldsymbol{z})+(\nabla y(\boldsymbol{z}))^T(\boldsymbol{x}-\boldsymbol{z}) \quad (22)$$

or the *partially linearized (PL)* function

$$y^{PL}(\boldsymbol{x},\boldsymbol{z}) = y^+(\boldsymbol{x})-\left(y^-(\boldsymbol{z})+(\nabla y^-(\boldsymbol{z}))^T(\boldsymbol{x}-\boldsymbol{z})\right) \quad (23)$$

can be the surrogate function of $\boldsymbol{y}(\boldsymbol{x})$, which is tight at a feasible point $\boldsymbol{z}$, i.e.,

$$y^{CL}(\boldsymbol{x},\boldsymbol{z}),\ y^{PL}(\boldsymbol{x},\boldsymbol{z})\begin{cases}=y(\boldsymbol{x}) & \text{when } \boldsymbol{x}=\boldsymbol{z} \quad (24a)\\ \geq y(\boldsymbol{x}) & \text{otherwise.} \quad (24b)\end{cases}$$

Similarly, $c_j^{CL}(\boldsymbol{x})$ or $c_j^{PL}(\boldsymbol{x})$ is assumed to be a surrogate function of the DC constraint function $c_j(\boldsymbol{x}),\forall j$. Then, the DC program with DC constraints can be approximately formulated as a sequence of convex optimization problems (in multiple iterations), and each can be solved by using algorithms and toolbox from convex optimization theory. Therefore, Problem (20) can be suboptimally but efficiently solved by the following Algorithm 1 and its variants.

**Algorithm 1** SCA-based Algorithm to Solve DC Program (20)

Initialization: $i=0$, $\boldsymbol{x}^{(0)}\in\mathcal{X}$ and $\epsilon_{th}$.
**repeat**
Generate the surrogate functions $\boldsymbol{y}^{PL}(\boldsymbol{x},\boldsymbol{x}^{(i)})$ and $\boldsymbol{c}_j^{PL}(\boldsymbol{x},\boldsymbol{x}^{(i)})$ by following (22);
Solve the convex optimization problem

$$\boldsymbol{x}^{(i+1)} = \arg\min_{\substack{x\in\Omega,\\ c_j^{PL}(\boldsymbol{x},\boldsymbol{x}^{(i)})\leq 0,\ j=1,\cdots,J}} y^{PL}(\boldsymbol{x},\boldsymbol{x}^{(i)}); \quad (25)$$

$i\leftarrow i+1$.
**until** $||\boldsymbol{x}^{(i)}-\boldsymbol{x}^{(i-1)}||\leq\epsilon_{th}$;

**Remark 5** *In principle, both PL functions and the CL functions (if they are feasible) can be flexibly used as the surrogate functions of the non-convex objective and constraint functions, which might lead to some variants of Algorithm 1.* □

### B. Solutions of BS Energy Consumption Minimization

By the above SCA-based algorithm, Problem (19) as a DC program can be solved as follows.

At a feasible point $\boldsymbol{q}$, the surrogate functions of the concave objective function (19a) and the DC rate expression in (19b)

can be expressed by

$$\widehat{P}_{BS}^{S}(\boldsymbol{p},\boldsymbol{q}) \triangleq \sum_{k=1}^{K}\sum_{f=1}^{F}\frac{P_{sp,k}^{f}}{\log(\epsilon+1)}\frac{\boldsymbol{t}_{k,f}^{T}\boldsymbol{p}}{\epsilon+\boldsymbol{t}_{k,f}^{T}\boldsymbol{q}} + \sum_{f=1}^{F}\left(1-\frac{\tau_f}{\beta_{2,f}}\right)\sum_{k=1}^{K}\frac{1}{\eta_k}\boldsymbol{t}_{k,f}^{T}\boldsymbol{p}, \tag{26}$$

$$R_{\ell}^{S}(\boldsymbol{p},\boldsymbol{q}) \triangleq \sum_{f=1}^{F}(1-\frac{\tau_f}{\beta_{2,f}})W_f\left(\log_2\left(\frac{W_f\sigma^2+\boldsymbol{\alpha}_{\mathcal{K},\ell}^{f,T}\boldsymbol{p}}{W_f\sigma^2+\boldsymbol{\alpha}_{\mathcal{K},\bar{\ell}}^{f,T}\boldsymbol{q}}\right) - \frac{1}{\log(2)}\cdot\frac{\boldsymbol{\alpha}_{\mathcal{K},\bar{\ell}}^{f,T}(\boldsymbol{p}-\boldsymbol{q})}{W_f\sigma^2+\boldsymbol{\alpha}_{\mathcal{K},\bar{\ell}}^{f,T}\boldsymbol{q}}\right), \tag{27}$$

based on (22) and (23), respectively, and after omitting constant terms. In particular, the derivation of (27) from (19b) is also based on $\log(x_1)-\log(x_2)=\log(\frac{x_1}{x_2})$ and $\nabla_{\boldsymbol{p}}\log_2(W_f\sigma^2+\boldsymbol{\alpha}_{\mathcal{K},\bar{\ell}}^{f,T}\boldsymbol{p})=\frac{1}{\log(2)}\cdot\frac{\boldsymbol{\alpha}_{\mathcal{K},\bar{\ell}}^{f}}{W_f\sigma^2+\boldsymbol{\alpha}_{\mathcal{K},\bar{\ell}}^{f,T}\boldsymbol{q}}$.

After selecting a feasible initial point $\boldsymbol{p}^{(0)}$, Problem (19) can be sub-optimally solved by the following Algorithm 2.

**Algorithm 2** SCA-based Algorithm to Solve Problem (19)

Initialization: $i=0$, a feasible $\boldsymbol{p}^{(0)}$ and $\epsilon_{th}$.
**repeat**
  Solve the convex optimization problem

$$\boldsymbol{p}^{(i+1)}=\arg\min_{\substack{\boldsymbol{p}\geq\boldsymbol{0},\ \boldsymbol{t}_k^T\boldsymbol{p}\leq P_{BS,k}^{max},\ \forall k\in\mathcal{K}\\ R_{\ell}^{S}(\boldsymbol{p},\boldsymbol{p}^{(i)})\geq\gamma_{\ell},\ \forall\ell\in\mathcal{L}}}\widehat{P}_{BS}^{S}(\boldsymbol{p},\boldsymbol{p}^{(i)}); \tag{28}$$

  $i\leftarrow i+1$.
**until** $||\boldsymbol{p}^{(i)}-\boldsymbol{p}^{(i-1)}||_2\leq\epsilon_{th}$;

In Algorithm 2, (28) is a convex optimization problem with a linear objective function and convex constraints, which can be optimally solved by the convex optimization methods.

**Remark 6** *The surrogate function $R_{\ell}^{S}(\boldsymbol{p},\boldsymbol{p}^{(i)})$ in* (27) *is an upper bound of the real rate function $R_{\ell}(\boldsymbol{p})$, but in each iteration it is always achieved that $R_{\ell}^{S}(\boldsymbol{p}^{\star},\boldsymbol{p}^{(i)})=\gamma_{\ell},\forall\ell$ where $\boldsymbol{p}^{\star}$ is the optimal solution to* (28) *because of $R_{\ell}^{S}(\boldsymbol{p}^{\star},\boldsymbol{p}^{(i)})=R_{\ell}(\boldsymbol{p}^{\star},\boldsymbol{p}^{(i)})=\gamma_{\ell},\forall\ell$ (implied by* (24a)*). This makes that each UE rate requirement can be finally guaranteed.* □

**Proposition 2** *The SCA-based algorithm in Algorithm 2 always converges to a KKT stationary solution of Problem* (19). □

*Proof:* See Appendix C. ■

Therefore, a local-optimal solution $\overline{\boldsymbol{p}}$ to Problem (19) can be obtained by Algorithm 2, which is not guaranteed to be globally optimal. Then, this solution obtained at the CP determines the strategies for the problems *P1-P4* in Section I-B.

### *C. Two-stage SCA-based Algorithm and its Complexity Analysis*

In order to analyze the complexity of Algorithm 2, we need to analyze the complexity of the optimization in (28). Problem (28) is a convex optimization problem with log-based functions in the constraints because of $R_{\ell}^{S}(\boldsymbol{p}^{\star},\boldsymbol{p}^{(i)})$ – it can be directly solved by the recent CVX toolbox [62]. However, since a log function cannot be simply supported by the symmetric primal/dual solvers within CVX, the principle of the recent CVX solver is to construct a successive approximation heuristic that allows the symmetric primal/dual solvers to support log functions [63]. This motivates us to apply the SCA-based algorithm to solve Problem (28) as follows.

In the $i$-th iteration of Algorithm 2, the surrogate function of the function (27) in Problem (28) at a point $\boldsymbol{s}$ can be generated as

$$R_{\ell}^{SS}(\boldsymbol{p},\boldsymbol{p}^{(i)},\boldsymbol{s}) \triangleq (\boldsymbol{g}_{\ell}(\boldsymbol{p}^{(i)},\boldsymbol{s}))^{T}\boldsymbol{p}+v_{\ell,1}(\boldsymbol{p}^{(i)},\boldsymbol{s}) + v_{\ell,2}(\boldsymbol{p}^{(i)},\boldsymbol{s}),\forall\ell\in\mathcal{L} \tag{29}$$

where $v_{\ell,1}(\boldsymbol{p}^{(i)},\boldsymbol{s}), v_{\ell,2}(\boldsymbol{p}^{(i)},\boldsymbol{s})$ and $\boldsymbol{g}_{\ell}(\boldsymbol{p}^{(i)},\boldsymbol{s})$ are defined as

$$v_{\ell,1}(\boldsymbol{p}^{(i)},\boldsymbol{s}) \triangleq \sum_{f=1}^{F}\left(1-\frac{\tau_f}{\beta_{2,f}}\right)W_f\times \log_2\left(\frac{W_f\sigma^2+\boldsymbol{\alpha}_{\mathcal{K},\ell}^{f,T}\boldsymbol{s}}{W_f\sigma^2+\boldsymbol{\alpha}_{\mathcal{K},\bar{\ell}}^{f,T}\boldsymbol{p}^{(i)}}\right), \tag{30}$$

$$v_{\ell,2}(\boldsymbol{p}^{(i)},\boldsymbol{s}) \triangleq \sum_{f=1}^{F}\left(1-\frac{\tau_f}{\beta_{2,f}}\right)\frac{W_f}{\log(2)}\times \left(\frac{\boldsymbol{\alpha}_{\mathcal{K},\bar{\ell}}^{f,T}\boldsymbol{p}^{(i)}}{W_f\sigma^2+\boldsymbol{\alpha}_{\mathcal{K},\bar{\ell}}^{f,T}\boldsymbol{p}^{(i)}}-\frac{\boldsymbol{\alpha}_{\mathcal{K},\ell}^{f,T}\boldsymbol{s}}{W_f\sigma^2+\boldsymbol{\alpha}_{\mathcal{K},\ell}^{f,T}\boldsymbol{s}}\right), \tag{31}$$

$$\boldsymbol{g}_{\ell}(\boldsymbol{p}^{(i)},\boldsymbol{s}) \triangleq \sum_{f=1}^{F}\left(1-\frac{\tau_f}{\beta_{2,f}}\right)\frac{W_f}{\log(2)}\times \left(\frac{\boldsymbol{\alpha}_{\mathcal{K},\bar{\ell}}^{f}}{W_f\sigma^2+\boldsymbol{\alpha}_{\mathcal{K},\bar{\ell}}^{f,T}\boldsymbol{p}^{(i)}}-\frac{\boldsymbol{\alpha}_{\mathcal{K},\ell}^{f}}{W_f\sigma^2+\boldsymbol{\alpha}_{\mathcal{K},\ell}^{f,T}\boldsymbol{s}}\right). \tag{32}$$

This derivation is based on (22) only for the log-term, and in (27) $\log_2(W_f\sigma^2+\boldsymbol{\alpha}_{\mathcal{K},\ell}^{f,T}\boldsymbol{p})$ is upper bounded by $\log_2(W_f\sigma^2+\boldsymbol{\alpha}_{\mathcal{K},\ell}^{f,T}\boldsymbol{s})+\frac{1}{\log(2)}\frac{\boldsymbol{\alpha}_{\mathcal{K},\ell}^{f,T}(\boldsymbol{p}-\boldsymbol{s})}{W_f\sigma^2+\boldsymbol{\alpha}_{\mathcal{K},\ell}^{f,T}\boldsymbol{s}}$.

When a SCA-based algorithm is employed to solve Problem (28), in each iteration the following linear optimization is required to be solved

$$\min_{\boldsymbol{p}\geq\boldsymbol{0}}\ \boldsymbol{r}(\boldsymbol{p}^{(i)})^{T}\boldsymbol{p} \tag{33a}$$

$$\text{s.t. } \boldsymbol{g}_{\ell}(\boldsymbol{p}^{(i)},\boldsymbol{s})^{T}\boldsymbol{p}+v_{\ell,1}(\boldsymbol{p}^{(i)},\boldsymbol{s})+v_{\ell,2}(\boldsymbol{p}^{(i)},\boldsymbol{s})\geq\gamma_{\ell},\ \forall\ell\in\mathcal{L} \tag{33b}$$

$$\boldsymbol{t}_k^T\boldsymbol{p}\leq P_{BS,k}^{max},\ \forall k\in\mathcal{K}, \tag{33c}$$

where $\boldsymbol{r}(\boldsymbol{p}^{(i)}) \triangleq \sum_{k=1}^{K}\sum_{f=1}^{F}\frac{P_{sp,k}^{f}}{\log(\epsilon+1)}\frac{\boldsymbol{t}_{k,f}}{\epsilon+\boldsymbol{t}_{k,f}^{T}\boldsymbol{p}^{(i)}} + \sum_{f=1}^{F}\left(1-\frac{\tau_f}{\beta_{2,f}}\right)\sum_{k=1}^{K}\frac{1}{\eta_k}\boldsymbol{t}_{k,f}$. Problem (33) can be further formulated as a standard linear program as

$$\min_{\boldsymbol{p}\geq\boldsymbol{0}}\ (\boldsymbol{r}(\boldsymbol{p}^{(i)}))^{T}\boldsymbol{p} \tag{34a}$$

$$\text{s.t. } \boldsymbol{R}(\boldsymbol{p}^{(i)},\boldsymbol{s})\boldsymbol{p}\leq\boldsymbol{b}(\boldsymbol{p}^{(i)},\boldsymbol{s}) \tag{34b}$$

$$\boldsymbol{p}\geq\boldsymbol{0} \tag{34c}$$

where $\boldsymbol{R}(\boldsymbol{p}^{(i)},\boldsymbol{s})$ and $\boldsymbol{b}(\boldsymbol{p}^{(i)},\boldsymbol{s})$ are defined as

$$\boldsymbol{R}(\boldsymbol{p}^{(i)},\boldsymbol{s}) \triangleq \big[-(\boldsymbol{g}_1(\boldsymbol{p}^{(i)},\boldsymbol{s}))^T;\cdots;-(\boldsymbol{g}_L(\boldsymbol{p}^{(i)},\boldsymbol{s}))^T; \boldsymbol{t}_1^T;\cdots;\boldsymbol{t}_K^T\big] \quad (35)$$

$$\boldsymbol{b}(\boldsymbol{p}^{(i)},\boldsymbol{s}) \triangleq \big[v_{1,1}(\boldsymbol{p}^{(i)},\boldsymbol{s})+v_{1,2}(\boldsymbol{p}^{(i)},\boldsymbol{s})-\gamma_1,\cdots, v_{L,1}(\boldsymbol{p}^{(i)},\boldsymbol{s})+v_{L,2}(\boldsymbol{p}^{(i)},\boldsymbol{s})-\gamma_L, P_{BS,1}^{max},\cdots,P_{BS,K}^{max}\big]^T. \quad (36)$$

Then, Problem (19) can be solved by the following two-stage SCA-based algorithm

**Algorithm 3** Two-stage SCA-based Algorithm to Solve Problem (19)

Initialization: $i=0$, a feasible $\boldsymbol{p}^{(0,0)}$ and $\epsilon_{th}$.
**repeat**
  Initialization: $j=0$, $\boldsymbol{p}^{(i,0)}$ and $\epsilon'_{th}$.
  **repeat**
    $$\boldsymbol{p}^{(i,j)} = \arg \min_{\substack{\boldsymbol{p}\geq\boldsymbol{0},\\ \boldsymbol{R}(\boldsymbol{p}^{(i,0)},\boldsymbol{p}^{(i,j)})\boldsymbol{p}\leq\boldsymbol{b}(\boldsymbol{p}^{(i,0)},\boldsymbol{p}^{(i,j)})}} (\boldsymbol{r}(\boldsymbol{q}^{(i,0)}))^T\boldsymbol{p}; \quad (37)$$
    $j \leftarrow j+1$.
  **until** $||\boldsymbol{p}^{(i,j)}-\boldsymbol{p}^{(i,j-1)}||_2 \leq \epsilon'_{th}$;
  $i \leftarrow i+1$;
  $\boldsymbol{p}^{i,0} \leftarrow \boldsymbol{p}^{(i-1,j)}$.
**until** $||\boldsymbol{p}^{(i,j)}-\boldsymbol{p}^{(i-1,j)}||_2 \leq \epsilon_{th}$;

**Lemma 4** *The two-stage SCA-based Algorithm 3 achieves the same solution to the single-stage SCA-based Algorithm 2.* □

*Proof:* Based on Proposition 2, it can be similarly proved that the inner SCA-based algorithm in Algorithm 3 can achieve a KKT stationary solution of Problem (28). Since Problem (28) is a strictly convex optimization, its has a unique KKT stationary solution (optimal solution), which can be achieved by the inner SCA-based algorithm. Therefore, this lemma holds. ■

In the following, the complexity of Algorithm 3 (equivalent to that of Algorithm 2) is derived.

**Proposition 3** *The number of operations to implement two-stage SCA-based algorithm 3 is of order*

$$N_{iter}^{out}N_{iter}^{in}\mathcal{O}\Bigg(\bigg(F\sum_{k=1}^{K}|\mathcal{U}_k|\bigg)^{3.5}\bigg((L+K)F\sum_{k=1}^{K}|\mathcal{U}_k| + F\sum_{k=1}^{K}|\mathcal{U}_k|+(L+K)\bigg)\zeta\Bigg) \quad (38)$$

*where $N_{iter}^{out}$ and $N_{iter}^{in}$ denote the average number of iterations of the outer SCA-based algorithm and the inner SCA-based algorithm in Algorithm [64], respectively, and $\zeta$ denotes the number of bits used to represent each real value of $\boldsymbol{R}(\boldsymbol{p}^{(i,0)},\boldsymbol{p}^{(i,j)})$, $\boldsymbol{b}(\boldsymbol{p}^{(i,0)},\boldsymbol{p}^{(i,j)})$ and $\boldsymbol{r}(\boldsymbol{q})$.* □

*Proof:* By using the two-stage SCA-based algorithm, the implementation of Algorithm 3 becomes an iterative optimization of standard linear programs. Based on Khachiyan's worst-case polynomial bound for the complexity of a standard linear programming [64], the complexity of (38) can be derived. ■

**Remark 7** *From* (38)*, the complexity of Algorithm 3 roughly scales as $(L+K)\left(F\sum_{k=1}^{K}|\mathcal{U}_k|\right)^{4.5}$ which is upper bounded by $(L+K)\left(F\sum_{k=1}^{K}N_k\right)^{4.5}$ because of $|\mathcal{U}_k|\leq N_k$. Thus, the number of FCs and the total number of BS antennas have a significant impact on the complexity. We stress that one implementation of Algorithm 3 at the CP is sufficient for a whole A-LSF time period.* □

### *D. Performance Analysis*

We compare our proposed algorithm based on the flexible assumptions *A2-A4* in Section I-B with some baselines that study the same BSs power minimization problem with the proposed BS power model but based on the assumptions *R2-R5* in Section I-A in a theoretical way.

**Proposition 4** *Based on the flexible system assumptions A2-A4 in Section I-B, our proposed green resource management mechanism always outperforms those baselines which are based on the assumptions R2-R5 in Section I-A.* □

*Proof:* Similar to Definition 2, we let $\boldsymbol{p}_{UE,\ell}\in\mathbb{R}^{F|\mathcal{B}_\ell|\times 1}$, $\boldsymbol{p}_{UE,\ell}^{f}\in\mathbb{R}^{|\mathcal{B}_\ell|\times 1}$, and $\boldsymbol{p}_{FC,f}\in\mathbb{R}^{L|\mathcal{B}_\ell|\times 1}$ denote the power of the BSs set $\mathcal{B}_\ell$ to UE $\ell$ on all FCs, the power of the BSs set $\mathcal{B}_\ell$ to UE $\ell$ on FC $f$, and the power of all the BSs to all the UEs on FC $f$, respectively. The "restricted" assumptions *R2-R5* can be equivalently formulated to the following theoretical constraints

$$\text{Assumption } R2 \Leftrightarrow ||\boldsymbol{p}_{BS,k}^{f}||_0\leq 1, \forall k\in\mathcal{K},\forall f\in\mathcal{F}, \quad (39)$$

$$\text{Assumption } R3 \Leftrightarrow ||\boldsymbol{p}_{UE,\ell}||_0^{|\mathcal{B}_\ell|,F}=1, \quad \forall \ell\in\mathcal{L}, \quad (40)$$

$$\text{Assumption } R4 \Leftrightarrow ||\boldsymbol{p}_{UE,\ell}||_0^{F,|\mathcal{B}_\ell|}=1, \quad \forall k\in\mathcal{K}, \quad (41)$$

$$\text{Assumption } R5 \Leftrightarrow ||\boldsymbol{p}_{FC,f}||_0^{L,|\mathcal{B}_\ell|}\leq 1, \quad \forall f\in\mathcal{F}, \quad (42)$$

respectively. Therefore, for example, one baseline with assumption *R2* can be formulated to the optimization problem (15) but with an extra constraint (39). In optimization, more constraints used for the same objective optimization problem will degrade the performance (or have the same performance when this extra constraint is inactive), since the feasible solution set is shrunk. In this work, these constraints (39)-(42) have been, in fact, relaxed by the general assumptions *A2-A4* as shown in Problem (15) , and thus its outperformance is verified. ■

### *E. Implementation*

The implementation of the proposed semi-dynamic green resource management mechanism during each A-LSF in a could-assisted HetNet is summarized as follows.

- **Step 1 (LSF Acquisition)**: At the beginning of an A-LSF, the CP collects the predicted LSF values of the network;
- **Step 2 (Green Resource Management)**: Based on the LSF values, the CP solves Problem (19) by Algorithm 2. According to the group sparse vector $\overline{\boldsymbol{p}}$ that is obtained,

the BS-UE association, FC assignment, downlink transmit power allocation and Bss sleep modes can be jointly determined, *and these strategies are fixed during the whole A-LSF*;

- **Step 3a (CSI Estimation)**: At the beginning of each coherence time, each UE transmits the uplink training sequences to its associated BSs on the assigned FCs under TDD model, based on which each BS estimates its local CSI of its serving UEs;
- **Step 3b (MRT Beamforming Design)**: Each BS locally designs the MRT beamforming vectors for its serving UEs on the assigned FC based on the estimated CSI in **Step 3a** and the transmit power vector $\overline{\boldsymbol{p}}$ in **Step 2**;
- **Step 3c (Downlink Transmission)**: Each BS transmits the desired data symbols to its serving UEs by the same MRT beamforming vectors and the fixed power allocation determined in **Step 2** until the end of the coherence time;
- **Step 4**: Repeat **Step 3a** to **Step 3c** until the end of the A-LSF.

## VI. Numerical Results

In this section, the performance of the proposed algorithm is evaluated on a 3-macro cell two-tier HetNet. Each macro cell is a regular hexagon with a radius of 250 meters and a single macro BS located at the center, where the same number of pico BSs and UEs are randomly deployed within each macro cell with the simulation parameters in Table II.

As shown in Section V-D, we have already proved that our proposed algorithm always outperforms the baselines based on the restricted BS-UE association and BS/UE-FC assignment assumption *R2-R5* in Section I-A, and thus the focus herein is on three other baselines:

- $L_{2,1}$ Approx: It denotes the performance of the same optimization by Algorithm 2 but using the $\ell_1/\ell_2$ mixed norm to approximate the $\ell_0$ norm instead of (16). This baseline is to show the impact of the $\ell_0$ norm approximation;
- Min. T-Power: This baseline is determined by minimizing only the sum downlink transmit power of BSs and no BS sleep modes are adopted. BSs are always *on* with full signal processing and circuit power, since no hardwares is switched off. The basic BS power and backhaul power are also considered in the computation of BSs sum power consumption according to (5);
- On/off only BS: This baseline is determined by minimizing only the number of active BSs, where each BS has binary choices: deep sleep or with full signal processing and circuit power. The basic BS circuit power and backhaul power are also considered in the computation of BSs sum power consumption according to (5).

### A. Deterministic Numerical Examples

We first evaluate the performance of Algorithm 2 within an A-LSF time period, where the UEs' locations can be considered to be fixed because the LSF is not varying during each LSF time period. We assume 5 pico BSs per macro cell. The partially loaded scenario is considered, where 6 UEs are located within each macro cell and each UE has a 12 Mbits/s data rate requirement. As shown in Table II, a total 40 MHz spectrum of $\{f_1, f_2\}$ is available.

TABLE II: HetNet system parameters

| | | | |
|---|---|---|---|
| Pilot length | Total No. of UEs | FCs | $f_1 = 783-803$ MHz, $f_2 = 1900-1920$ MHz |
| $P^f_{sp,k}$ | Reference [54] | $P_k^{sleep_0}$ | $75 \times N_k$ Watt (macro), $4.3 \times N_k$ Watt (pico) |
| $\sigma^2$ | -174 dBm/Hz | $P^{max}_{BS,k}$ | 40 Watt (macro), 1 Watt (pico) |
| Path Loss | Reference [65] | $\eta_k$ | 35% (macro), 25% (pico) |

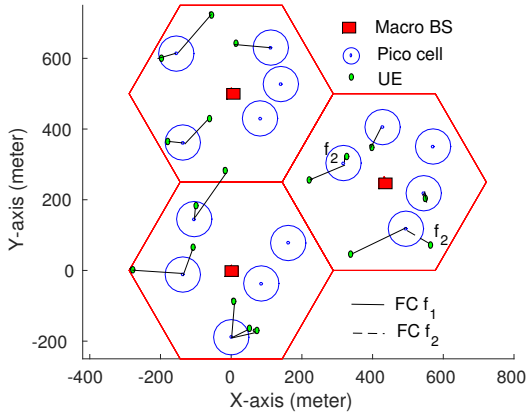



Fig. 1: A numerical example of Algorithm 2 with a per UE rate requirement 12 Mbits/s:. Each macro-BS and pico BS possess 16 and 4 antennas, respectively.

A result example for Algorithm 2 is shown in Fig. 1, where the BS-UE association, FC assignment and BSs status are clearly illustrated. We observe that all macro BSs are in deep-sleep mode as well as some pico BSs because of the off-peak traffic load. Another interesting observation is that all UEs except for only two UEs prefer to reuse the FC $f_1 = 783-803$ MHz which has lower path loss, where the assignment of $f_1$ and $f_2$ are denoted by the "solid lines" and "dashed lines", respectively.

In Fig. 2 the convergence behavior of Algorithm 2 is shown, where we set the parameter $\epsilon$ for the $\ell_0$ norm approximation in (16) as $\epsilon \in \{10^{-1}, 10^{-3}, 10^{-5}, 10^{-7}\}$, where for each $\epsilon$, 10 random initializations are used. It is shown in Fig. 2 that the used $\ell_0$ norm approximation in (16) is robust to the choice of $\epsilon$ and different initializations might lead to different KKT stationary solutions with similar convergence rate.

### B. Average Performance Evaluation

The average performance of the proposed algorithm is evaluated by 100 Monte Carlo simulations, where the locations of the UEs are randomly generated within each macro cell.

The average energy consumption for 5 pico cells in each macro cell with respect to the per-UE rate requirement is shown in Fig. 3. Observe that the energy consumption is increasing with the UE's rate requirement and our algorithm

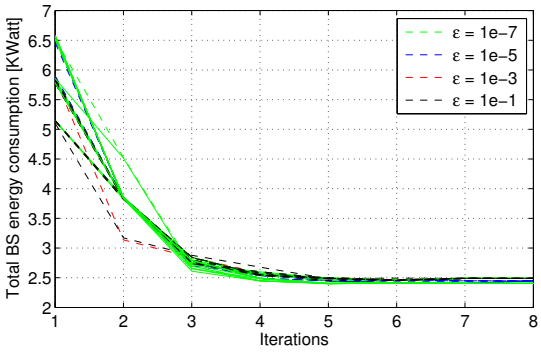



Fig. 2: Convergence performance of Algorithm 2 with a per UE rate requirement 2 Mbits/s: Each macro BS and pico BS possess 8 and 4 antennas, respectively.

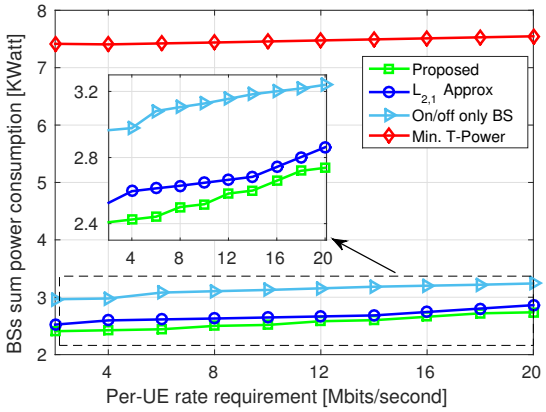



Fig. 3: Average total BS power consumption performance vs. UE rate: 5 pico cells and 6 UEs in each macro cell and each macro-BS and pico BS possess 8 and 4 antennas, respectively.

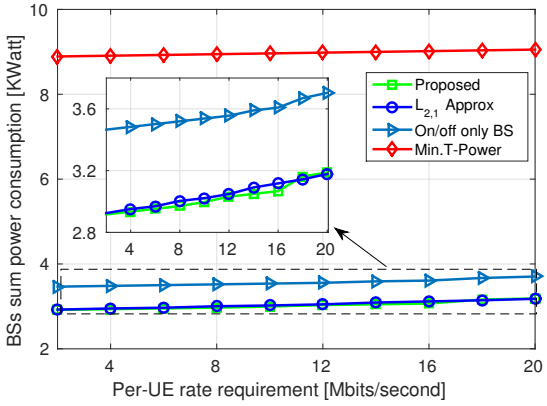



Fig. 4: Average total BS power consumption performance vs. UE rate: 10 pico cells and 6 UEs in each macro cell and each macro-BS and pico BS possess 8 and 4 antennas, respectively.

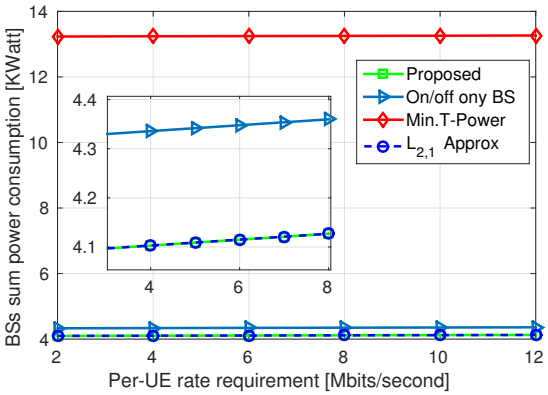



Fig. 5: Average total BS power consumption performance vs. UE rate: 5 pico cells and 2 UEs in each macro cell and each macro-BS and pico BS possess 16 and 4 antennas, respectively.

can achieve a more than 60% and 10% energy reduction compared with the "Min. T-Power" and the "On/off only BS", respectively, since the "Min. T-Power" does not adopt the BS sleep modes and "On/off only BS" cannot flexibly switch off hardware of the assigned FCs. This implies that our proposed flexible BS power model provides more degrees of freedom for increased energy saving. In addition, the log-based approximation slightly outperforms the $\ell_1/\ell_2$ mixed norm based approximation. Another energy consumption comparison for 10 pico cells in each macro cell with respect to the per-UE rate requirement is shown in Fig. 4. A more than 60% and 10% energy reduction compared with the "Min. T-Power" and the "On/off only BS" can be still achieved in the denser networks, respectively, while the performance gap between the proposed and $L_{2,1}$ Approx becomes small as the number of pico BSs increases.

In order to provide sufficient evaluations of the proposed algorithm, different system scenarios are simulated. In Fig. 5, the average BSs sum power consumption is compared in a very low traffic load secario, where only two UEs are located within a macro cell. Observe that the power consumption by "Min. T-Power" is about three times larger than the proposed. The gap between the proposed and the other two baselines becomes smaller in the very low traffic load scenario. In particular, the proposed and "$L_{2,1}$ Approx" achieve the same performance. Both Fig. 4 and Fig. 5 imply that the performance by "$L_{2,1}$ Approx" becomes approaching to the proposed when a network has a *relatively* small traffic load compared with its own capacity.

In Fig. 6, we illustrate the average total energy consumption versus with the number of pico-BS antennas for the per-UE rate requirement of 12 Mbits/s. The power consumption is increasing with the number of pico-BS antennas, since both the basic and signal processing circuit power are linearly increased with the number of BS antennas. This result still verifies the effectiveness of the proposed with different number of pico-BS antennas.

## VII. Conclusions

In this paper, motivated by the high demand for energy saving in a cloud-assisted HetNet with off-peak traffic loads, we propose a semi-dynamic green resource management mechanism to minimize BSs energy consumption and also to satisfy each UE's rate requirement. This mechanism fits well with

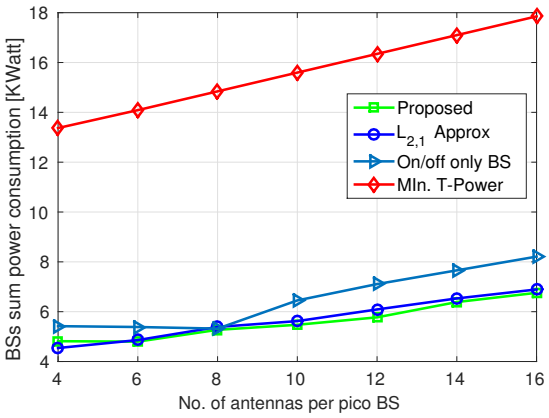



Fig. 6: Average total BS power consumption performance vs. No. of pico-BS antennas: Per-UE rate requirement 12 Mbits/s, 5 pico cells and 6 UEs in each macro cell and each macro-BS possess 16 antennas, respectively.

the architecture of the cloud-assisted HetNet, since BSs have a low requirement for computation and signalling transmission by locally employing the low-complexity MRT beamforming in dynamic downlink transmission. The computationally demanding optimization will be performed on a slower time scale relating to changes in large scale fading coefficients. In this approach, in order to benefit from the reconfiguration of a system, a flexible BS power consumption model is developed to support scalability, i.e., some unnecessary hardware components could be switched off to reduce the energy consumption. Furthermore, this power model is formulated as a function of a transmit power vector and reflects the power consumption of signal processing and circuits, downlink transmission and backhaul transmission. Based on this power model, a large scale fading based optimization problem is formulated and solved by the CP in a centralized fashion. The solution is used to determine the energy-saving strategies for scheduling, transmit power allocation and BSs sleep modes, which are fixed for the coherence time of the large scale fading. In addition, the green resource management mechanism proposed in this work serves as a general framework for BSs energy minimization, and much previous related work can be considered as special cases. Simulation results indicate that the proposed algorithm is capable of reducing BSs power consumption by more than 60% compared with some previous approaches.

## ACKNOWLEDGEMENT

The authors would like to thank the editors and anonymous reviewers for their helpful comments. Matlab codes for numerical simulations are available at http://www.homepages.ed.ac.uk/jst/.

## APPENDIX A
## PROOF OF LEMMA 1

*Proof:* For the UE set $\mathcal{U}_{FC,f}$, a $\tau_f \times |\mathcal{U}_{FC,f}|$ pilot sequence matrix is needed for channel training from $\mathcal{U}_{FC,f}$ to their associated BSs

$$\boldsymbol{\Phi}^f = [\boldsymbol{\phi}^f_{\mathcal{U}_{FC,f}(1)}; \cdots; \boldsymbol{\phi}^f_{\mathcal{U}_{FC,f}(|\mathcal{U}_{FC,f}|)}]. \tag{43}$$

If $\tau_f \geq |\mathcal{U}_{FC,f}|$, we can generate the pairwise orthogonal pilot sequences $\{\boldsymbol{\phi}^f_\ell\}_{\ell \in \mathcal{U}_{FC,f}}$. Otherwise, pilot reuse among the UEs in $\mathcal{U}_{FC,f}$ is needed and pilot contamination exists. To consider both cases, we generally denote by $\mathcal{U}^m_{FC,f} \subset \mathcal{U}_{FC,f}$ the set of UEs who use the same pilot sequence $\boldsymbol{\phi}^f_{\mathcal{U}_{FC,f}(m)}$ in $\boldsymbol{\Phi}^f$. If $|\mathcal{U}^m_{FC,f}| = 1$, it means no reuse of $\boldsymbol{\phi}^f_{\mathcal{U}_{FC,f}(m)}$. Otherwise, $|\mathcal{U}^m_{FC,f}|$ UEs reuse the same pilot sequence $\boldsymbol{\phi}^f_{\mathcal{U}_{FC,f}(m)}$.

By transmitting $\sqrt{\tau_f}\boldsymbol{\psi}^f_\ell$ from UE $\ell \in \mathcal{U}_{FC,f}$ with the uplink power $\sqrt{p^f_{UE,\mathcal{U}_{FC,f}(\ell)}}$, the $\tau_f$ length column vector received at the $m$-th antenna at BS $k$ on FC $f$ is

$$\boldsymbol{y}^f_{km} = \sqrt{\tau_f} \sum_{\ell \in \mathcal{U}_{FC,f}} \sqrt{p^f_{UE,\mathcal{U}_{FC,f}(\ell)}} h^f_{km,\ell} \boldsymbol{\psi}^f_\ell + \boldsymbol{z}^f_{k,m} \tag{44}$$

where $h^f_{km,\ell} \in \mathbb{C}$ denotes the channel coefficient from UE $\ell$ to the $m$-th antenna of BS $k$ on FC $f$. Then, the signal received at the BS $k$ can be expressed as

$$\begin{aligned} \boldsymbol{Y}^f_k &= [\boldsymbol{y}^f_{k1}, \boldsymbol{y}^f_{k2}, \cdots, \boldsymbol{y}^f_{kN_k}] \in \mathbb{C}^{\tau_f \times N_k} && \text{(45a)} \\ &= \sqrt{\tau_f} \boldsymbol{P}^{\frac{1}{2}}_{\mathcal{U}_{FC,f}} \boldsymbol{\Phi}^f \boldsymbol{H}^f_k + \boldsymbol{Z}^f_k && \text{(45b)} \\ &= \sqrt{\tau_f} \boldsymbol{P}^{\frac{1}{2}}_{\mathcal{U}_{FC,f}} \boldsymbol{\Phi}^f \boldsymbol{A}^f_k \tilde{\boldsymbol{H}}^f_k + \boldsymbol{Z}^f_k && \text{(45c)} \end{aligned}$$

where

$$\begin{aligned} \boldsymbol{P}_{\mathcal{U}_{FC,f}} &= \text{diag}\left[p^f_{UE,\mathcal{U}_{FC,f}(1)}, \cdots, p^f_{UE,\mathcal{U}_{FC,f}(|\mathcal{U}_{FC,f}|)}\right] \\ \tilde{\boldsymbol{H}}^f_k &= [\tilde{\boldsymbol{h}}^{f,T}_{k,\mathcal{U}_{FC,f(1)}}; \cdots; \tilde{\boldsymbol{h}}^{f,T}_{k,\mathcal{U}_{FC,f}(|\mathcal{U}_{FC,f}|)}] \\ \boldsymbol{A}^f_k &= \text{diag}\left[\sqrt{\alpha^f_{k,\mathcal{U}_{FC,f}(1)}}, \cdots, \sqrt{\alpha^f_{k,\mathcal{U}_{FC,f}(|\mathcal{U}_{FC,f}|)}}\right] \\ \boldsymbol{Z}^f_k &= [\boldsymbol{z}^f_{k,1}, \boldsymbol{z}^f_{k,2}, \cdots, \boldsymbol{z}^f_{k,N_k}] \in \mathbb{C}^{\tau \times N_k}, \end{aligned}$$

where $\boldsymbol{z}^f_{k,n} \in \mathbb{C}^{\tau_f \times 1}, \forall n \in \{1, \cdots, N_k\}$ denotes the noise vector at $n$-th antenna of BS $k$ in uplink training phase on FC $f$. We assume that $\boldsymbol{z}^f_{k,n} \sim \mathcal{CN}(\boldsymbol{0}, W_f \sigma^2 \boldsymbol{I}), \forall n$.

Following the standard MMSE estimation in [66, Chapter 15.8], the MMSE estimate of the channel from a typical UE $\ell \in \mathcal{U}_{FC,f}$ to its associated BS $k$ on FC $f$ is $\sqrt{\alpha^f_{k,\ell}}\hat{\boldsymbol{h}}^f_{k,\ell}$, where

$$\hat{\boldsymbol{h}}^f_{k,\ell} = \frac{\sqrt{\tau_f \alpha^f_{k,\ell} p^f_{UE,\ell}} \boldsymbol{\phi}^{f,H}_\ell \boldsymbol{Y}^f_k}{\tau_f \alpha^f_{k,\ell} p^f_{UE,\ell} + \sum_{j \in \mathcal{U}^\ell_{FC,f} \setminus \{\ell\}} \tau_f \alpha^f_{k,j} p^f_{UE,j} + W_f \sigma^2}. \tag{46}$$

Then, the result in Lemma 1 is concluded. ■

## APPENDIX B
## PROOF OF LEMMA 3

*Proof:* With the MRT beamforming $\boldsymbol{w}^f_{k,\ell} = \sqrt{p^f_{k,\ell}} \overrightarrow{\hat{\boldsymbol{h}}}^f_{k,\ell}, \forall k, \ell, f$, (9) becomes

$$R_\ell^f = \mathbb{E}_{\tilde{\boldsymbol{h}}}\left\{\log_2\left(1+\frac{\sum_{k\in\mathcal{B}_\ell} p_{k,\ell}^f \alpha_{k,\ell}^f |\tilde{\boldsymbol{h}}_{k,\ell}^{f,H}\overrightarrow{\hat{\boldsymbol{h}}}_{k,\ell}^f|^2}{W_f\sigma^2+\sum\limits_{\overline{k}\in\mathcal{K}\setminus\mathcal{B}_\ell}\sum\limits_{j\in\mathcal{U}_{\overline{k}}} p_{\overline{k},j}^f \alpha_{\overline{k},\ell}^f |\tilde{\boldsymbol{h}}_{\overline{k},\ell}^{f,H}\overrightarrow{\hat{\boldsymbol{h}}}_{\overline{k},j}^f|^2 + \sum\limits_{k\in\mathcal{B}_\ell}\sum\limits_{\overline{\ell}\in\mathcal{U}_k\setminus\{\ell\}} p_{k,\overline{\ell}}^f \alpha_{k,\ell}^f |\tilde{\boldsymbol{h}}_{k,\ell}^{f,H}(t)\hat{\boldsymbol{h}}_{k,\overline{\ell}}^f(t)|^2}\right)\right\} \tag{47}$$

$$\approx \log_2\left(1+\frac{\sum_{k\in\mathcal{B}_\ell} p_{k,\ell}^f \alpha_{k,\ell}^f \mathbb{E}_{\tilde{\boldsymbol{h}}}\{|\tilde{\boldsymbol{h}}_{k,\ell}^{f,H}\overrightarrow{\hat{\boldsymbol{h}}}_{k,\ell}^f|^2\}}{W_f\sigma^2+\sum\limits_{\overline{k}\in\mathcal{K}\setminus\mathcal{B}_\ell}\sum\limits_{j\in\mathcal{U}_{\overline{k}}} p_{\overline{k},j}^f \alpha_{\overline{k},\ell}^f \mathbb{E}_{\tilde{\boldsymbol{h}}}\{|\tilde{\boldsymbol{h}}_{\overline{k},\ell}^{f,H}\overrightarrow{\hat{\boldsymbol{h}}}_{\overline{k},j}^f|^2\} + \sum\limits_{k\in\mathcal{B}_\ell}\sum\limits_{\overline{\ell}\in\mathcal{U}_k\setminus\{\ell\}} p_{k,\overline{\ell}}^f \alpha_{k,\ell}^f \mathbb{E}_{\tilde{\boldsymbol{h}}}\{|\tilde{\boldsymbol{h}}_{k,\ell}^{f,H}(t)\hat{\boldsymbol{h}}_{k,\overline{\ell}}^f(t)|^2\}}\right) \tag{48}$$

where (48) is derived based on the approximation $\mathbb{E}_x\{\log_2(1+\frac{f_1(x)}{f_2(x)})\} \approx \log_2(1+\frac{\mathbb{E}_x\{f_1(x)\}}{\mathbb{E}_x\{f_2(x)\}})$. Based on (48), Lemma 3 is derived according to the following results:

$$\mathbb{E}_{\tilde{\boldsymbol{h}}}\{|\tilde{\boldsymbol{h}}_{k,\ell}^f\overrightarrow{\hat{\boldsymbol{h}}}_{k,\ell}^f|^2\} = \mathbb{E}\{|\left(\hat{\boldsymbol{h}}_{k,\ell}^f+\boldsymbol{e}_{k,\ell}^f\right)^H\overrightarrow{\hat{\boldsymbol{h}}}_{k,\ell}^f|^2\} \tag{49}$$

$$= \mathbb{E}\{|\hat{\boldsymbol{h}}_{k,\ell}^{f,H}\overrightarrow{\hat{\boldsymbol{h}}}_{k,\ell}^f|^2\} + \mathbb{E}\{|\boldsymbol{e}_{k,\ell}^{f,H}\overrightarrow{\hat{\boldsymbol{h}}}_{k,\ell}^f|^2\} \tag{50}$$

$$= \mathbb{E}\{||\hat{\boldsymbol{h}}_{k,\ell}^f||^2\} + \overrightarrow{\hat{\boldsymbol{h}}}_{k,\ell}^{f,H}\mathbb{E}\{\boldsymbol{e}_{k,\ell}^f\boldsymbol{e}_{k,\ell}^{f,H}\}\overrightarrow{\hat{\boldsymbol{h}}}_{k,\ell}^f \tag{51}$$

$$= \delta_{k,\ell}^f N_k + (1-\delta_{k,\ell}^f), \quad \forall k\in\mathcal{B}_\ell \tag{52}$$

where (49) is based on the estimated channel model in (1), and (50) is based on the fact $\mathbb{E}\{\hat{\boldsymbol{h}}_{k,\ell}^{f,H}\overrightarrow{\hat{\boldsymbol{h}}}_{k,\ell}^f\overrightarrow{\hat{\boldsymbol{h}}}_{k,\ell}^{f,H}\boldsymbol{e}_{k,\ell}^f\} = 0$ because $\boldsymbol{e}_{k,\ell}^f$ is zero-mean Gaussian and is independent of $\hat{\boldsymbol{h}}_{k,\ell}^f$, and (52) is based on the derived result in (1).

The average inter-BS interference terms in the denominator of (48) are derived to

$$\mathbb{E}_{\tilde{\boldsymbol{h}}}\left\{|\tilde{\boldsymbol{h}}_{\overline{k},\ell}^{f,H}\overrightarrow{\hat{\boldsymbol{h}}}_{\overline{k},j}^f|^2\right\} = \overrightarrow{\hat{\boldsymbol{h}}}_{\overline{k},j}^{f,H}\mathbb{E}_{\tilde{\boldsymbol{h}}}\{\tilde{\boldsymbol{h}}_{\overline{k},\ell}^f\tilde{\boldsymbol{h}}_{\overline{k},\ell}^{f,H}\}\overrightarrow{\hat{\boldsymbol{h}}}_{\overline{k},j}^f = 1, \quad \forall j\in\mathcal{U}_{\overline{k}},\ \overline{k}\in\mathcal{K}\setminus\{\mathcal{B}_\ell\} \tag{53}$$

$$\mathbb{E}_{\tilde{\boldsymbol{h}}}\left\{|\tilde{\boldsymbol{h}}_{k,\ell}^{m,H}\overrightarrow{\hat{\boldsymbol{h}}}_{k,\overline{\ell}}^m|^2\right\} = 1, \quad \forall\overline{\ell}\in\mathcal{U}_k\setminus\{\ell\},\ \overline{k}\in\mathcal{B}_\ell \tag{54}$$

where (53) is because $\tilde{\boldsymbol{h}}_{\overline{k},\ell}^f$ is unit-variance Gaussian and is independent of $\overrightarrow{\hat{\boldsymbol{h}}}_{\overline{k},\overline{\ell}}^f, \forall\overline{\ell}\neq\ell$. ■

## APPENDIX C
## PROOF OF PROPOSITION 2

*Proof:* The proof of Proposition 2 has two aspects: 1) the convergence of Algorithm 2 and 2) the solutions of Algorithm 2.

*1) Convergence:* The convergence of the algorithm is implied by the fact that the iterative sequence $\{\widehat{P}_{BS}(\boldsymbol{p}^{(i)})\}_{i=1}^{+\infty}$ is monotonically decreasing. At the $i$-th iteration, we have

$$\widehat{P}_{BS}(\boldsymbol{p}^{(i+1)}) \overset{(a)}{=} \widehat{P}_{BS}(\boldsymbol{p}^{(i+1)},\boldsymbol{p}^{(i+1)}) \overset{(b)}{\leq} \widehat{P}_{BS}(\boldsymbol{p}^{(i+1)},\boldsymbol{p}^{(i)}) \tag{55}$$

$$\overset{(c)}{\leq} \widehat{P}_{BS}(\boldsymbol{p}^{(i)},\boldsymbol{p}^{(i)}) \overset{(d)}{=} \widehat{P}_{BS}(\boldsymbol{p}^{(i)}), \tag{56}$$

where both the equalities (a) and (d) are based on (24a), and the inequalities (b) and (c) are based on (24b) and the convex optimization of (28) (optimal updating). Considering that the constraints form a closed set, there exists a cluster point of the sequence $\{\widehat{P}_{BS}(\boldsymbol{p}^{(i)})\}_{i=1}^{+\infty}$. Let $\overline{\boldsymbol{p}} \triangleq \lim_{i\to+\infty}\boldsymbol{p}^{(i)}$ be the cluster point solution returned by Algorithm 2 with a sufficiently small $\epsilon_{th}$.

*2) KKT Solutions:* We will show the cluster point solution $\overline{\boldsymbol{p}}$ is a KKT stationary point of the original problem (19). Considering the properties of the cluster point, we have $\boldsymbol{p}^{(i)} = \boldsymbol{p}^{(i+1)} = \overline{\boldsymbol{p}}$ with $i\to+\infty$ for the optimization of (28). Therefore, given $\boldsymbol{p}^{(i)} = \overline{\boldsymbol{p}}$, the optimal solution $\boldsymbol{p}^{(i+1)} = \overline{\boldsymbol{p}}$ of (28) should satisfy the following KKT conditions

$$\sum_{k=1}^K\sum_{f=1}^F P_{sp,k}^f\frac{\boldsymbol{t}_{k,f}^T}{\epsilon+\boldsymbol{t}_{k,f}^T\overline{\boldsymbol{p}}} + \sum_{k=1}^K\frac{\theta_k\boldsymbol{t}_{BS,k}^T}{\eta_k} + \sum_{\ell=1}^L\zeta_\ell\sum_{f=1}^F\left(1-\frac{\tau_f}{\beta_{2,f}}\right)\frac{W_f}{\log(2)}\times \left(\frac{\boldsymbol{\alpha}_{\mathcal{K},\ell}^{f,T}}{W_f\sigma^2+\boldsymbol{\alpha}_{\mathcal{K},\ell}^{f,T}\overline{\boldsymbol{p}}} - \frac{\boldsymbol{\alpha}_{\mathcal{K},\overline{\ell}}^{f,T}}{W_f\sigma^2+\boldsymbol{\alpha}_{\mathcal{K},\overline{\ell}}^{f,T}\overline{\boldsymbol{p}}}\right) = 0 \tag{57a}$$

$$0 \leq \zeta_\ell \perp \left(\sum_{f=1}^F\left(1-\frac{\tau_f}{\beta_{2,f}}\right)\frac{W_f}{\log(2)}\times \left(\frac{\boldsymbol{\alpha}_{\mathcal{K},\ell}^{f,T}}{W_f\sigma^2+\boldsymbol{\alpha}_{\mathcal{K},\ell}^{f,T}\overline{\boldsymbol{p}}} - \frac{\boldsymbol{\alpha}_{\mathcal{K},\overline{\ell}}^{f,T}}{W_f\sigma^2+\boldsymbol{\alpha}_{\mathcal{K},\overline{\ell}}^{f,T}\overline{\boldsymbol{p}}}\right) - R_\ell\right) \geq 0, \forall\ell \tag{57b}$$

$$0 \leq \theta_k \perp \left(P_{BS,k}^{max} - \boldsymbol{t}_{BS,k}^T\overline{\boldsymbol{p}}\right) \geq 0, \forall k \tag{57c}$$

$$\overline{\boldsymbol{p}} \geq \boldsymbol{0} \tag{57d}$$

where $\zeta_\ell, \forall\ell\in\mathcal{L}$ and $\theta_k, \forall k\in\mathcal{K}$ are the Lagrangian multipliers. Observe that the KKT conditions (57a)-(57d) are exactly same as the KKT conditions of Problem (19). Therefore, it implies that $\overline{\boldsymbol{p}}$ with the associated Lagrangian multipliers $\{\zeta_\ell,\theta_k\}$ is a KKT stationary solution to the original problem (19). ■

## REFERENCES


[1] NGMN Alliance, "Next generation mobile networks (NGMN) 5G white paper," Tech. Rep., Feb. 2015.

[2] E.A Jorswieck, L. Badia, T. Fahldieck, E. Karipidis, and Jian Luo, "Spectrum sharing improves the network efficiency for cellular operators," *IEEE Commun. Mag.*, vol. 52, no. 3, pp. 129–136, Mar. 2014.

[3] 4G Americas, "LTE carrier aggregation technology development and deployment worldwide," Tech. Rep., Oct. 2014.

[4] F. Rusek, D. Persson, Buon Kiong Lau, E.G. Larsson, T.L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[5] J.G. Andrews, S. Buzzi, Wan Choi, S.V. Hanly, A Lozano, AC.K. Soong, and J.C. Zhang, "What will 5G be?," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

[6] Ericsson, "5G energy performance–key technologies and design principles," Tech. Rep., Ericsson White Paper, http://www.ericsson.com/res/docs/whitepapers/wp-5g-energy-performance.pdf, Apr. 2015.

[7] Nokia, "Technology vision 2020 flatten network energy consumption," Tech. Rep., Nokia Networks White Paper, http://networks.nokia.com/innovation/technology-vision/flatten-total-energy-consumption, 2014.

[8] 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.

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

[10] L. Smolyar, I. Bergel, and H. Messer, "Unified approach to joint power allocation and base assignment in nonorthogonal networks," *IEEE Trans. Veh. Technol.*, vol. 58, no. 8, pp. 4576–4586, Oct. 2009.

[11] L. Qian, Y. Zhang, Y. Wu, and J. Chen, "Joint base station association and power control via Benders' decomposition," *IEEE Trans. Wirel. Commun.*, vol. 12, no. 4, pp. 1651–1665, Apr. 2013.

[12] A. Silva, H. Tembine, E. Altman, and M. Debbah, "Optimum and Equilibrium in assignment problems with congestion: Mobile terminals association to base stations," *IEEE Trans. Autom. Control*, vol. 58, no. 8, pp. 2018–2031, Aug. 2013.

[13] H. Pennanen, A. Tolli, and M. Latva-aho, "Decentralized base station assignment in combination with downlink beamforming," in *IEEE SPAWC*, Jun. 2010, pp. 1–5.

[14] Y. Lin, W. Bao, W. Yu, and B. Liang, "Optimizing user association and spectrum allocation in HetNets: A utility perspective," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 6, pp. 1025–1039, Jun. 2015.

[15] D.W.K. Ng, E.S. Lo, and R. Schober, "Energy-efficient resource allocation in OFDMA systems with large numbers of base station antennas," *IEEE Trans. Wirel. Commun.*, vol. 11, no. 9, pp. 3292–3304, Sep. 2012.

[16] L. Venturino, A. Zappone, C. Risi, and S. Buzzi, "Energy-efficient scheduling and power allocation in downlink OFDMA networks with base station coordination," *IEEE Trans. Wirel. Commun.*, vol. 14, no. 1, pp. 1–14, Jan. 2015.

[17] X. Wang, F. Zheng, P. Zhu, and X. You, "Energy-efficient resource allocation in coordinated downlink multi-cell OFDMA systems," *IEEE Trans. Veh. Technol.*, vol. 65, no. 3, pp. 1395–1408, Mar. 2016.

[18] D. Cao, S. Zhou, and Z. Niu, "Improving the energy efficiency of two-tier heterogeneous cellular networks through partial spectrum reuse," *IEEE Trans. Wirel. Commun.*, vol. 12, no. 8, pp. 4129–4141, Aug. 2013.

[19] K. Davaslioglu, C.C. Coskun, and E. Ayanoglu, "Energy-efficient resource allocation for fractional frequency reuse in heterogeneous networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 10, pp. 5484–5497, Oct. 2015.

[20] Z. Yang Z. Niu S. Zhou, J. Gong and P. Yang, "Green mobile access network with dynamic base station energy saving," *ACM MobiCom*, vol. 9, no. 262, pp. 10–12, 2009.

[21] R. Wang, J.S. Thompson, H. Haas, and P.M. Grant, "Sleep mode design for green base stations," *IET Commun.*, vol. 5, no. 18, pp. 2606–2616, Dec. 2011.

[22] J. Wu, Y. Zhang, M. Z., and E.K.-N. Yung, "Energy-efficient base-stations sleep-mode techniques in green cellular networks: A survey," *IEEE Communications Surveys Tutorials*, vol. 17, no. 2, pp. 803–826, Q2 2015.

[23] E. Pollakis, R.L.G. Cavalcante, and S. Stanczak, "Base station selection for energy efficient network operation with the majorization-minimization algorithm," in *IEEE SPAWC*, Jun. 2012, pp. 219–223.

[24] R. L. G. Cavalcante, S. Stańczak, M. Schubert, A. Eisenblätter, and U. Türke, "Toward energy-efficient 5G wireless communications technologies," *IEEE Signal Process. Mag.*, vol. 31, no. 6, pp. 24–34, Nov. 2014.

[25] L. Su, C. Yang, Z. Xu, and A.F. Molisch, "Energy-efficient downlink transmission with base station closing in small cell networks," in *IEEE ICASSP*, May 2013, pp. 4784–4788.

[26] S. Han, C. Yang, and A.F. Molisch, "Spectrum and energy efficient cooperative base station doze," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 2, pp. 285–296, Feb. 2014.

[27] S. Zhang, J. Gong, S. Zhou, and Z. Niu, "How many small cells can be turned off via vertical offloading under a separation architecture?," *IEEE Trans. Wirel. Commun.*, vol. 14, no. 10, pp. 5440–5453, Oct. 2015.

[28] S.-R. Cho and W. Choi, "Energy-efficient repulsive cell activation for heterogeneous cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 5, pp. 870–882, May 2013.

[29] H. Claussen, I. Ashraf, and Lester T. W. Ho, "Dynamic idle mode procedures for femtocells," *Bell Labs Technical Journal*, vol. 15, no. 2, pp. 95–116, 2010.

[30] B. Debaillie, C. Desset, and F. Louagie, "A flexible and future-proof power model for cellular base stations," in *IEEE VTC Spring*, May 2015.

[31] P. Skillermark and P. Frenger, "Enhancing energy efficiency in LTE with antenna muting," in *IEEE VTC Spring*, May 2012, pp. 1–5.

[32] M. Hedayati, M. Amirijoo, P. Frenger, and J. Moe, "Reducing energy consumption through adaptation of number of active radio units," in *IEEE VTC Spring*, May 2011, pp. 1–5.

[33] P. Frenger, P. Moberg, J. Malmodin, Y. Jading, and I. Godor, "Reducing energy consumption in LTE with cell DTX," in *IEEE VTC Spring*, May 2011, pp. 1–5.

[34] H. Holtkamp, G. Auer, S. Bazzi, and H. Hass, "Minimizing base station power consumption," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 2, pp. Feb., 2014.

[35] J. Sköld E. Dahlman, S. Parkvall and P. Beming, *3G Evolution HSPA and LTE for mobile broadband*, New York: Academic, 2008.

[36] E. Larsson, O. Edfors, F. Tufvesson, and T. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, February 2014.

[37] Erik Luther, "5G massive MIMO testbed: From theory to reality," Tech. Rep., National Instruments White Paper, http://www.ni.com/white-paper/52382/en/, Sep. 2015.

[38] V. Jungnickel, K. Manolakis, W. Zirwas, B. Panzner, V. Braun, M. Lossow, M. Sternad, R. Apelfröjd, and T. Svensson, "The role of small cells, coordinated multipoint, and massive MIMO in 5G," *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 44–51, May 2014.

[39] R. Irmer, H. Droste, P. Marsch, M. Grieger, G. Fettweis, S. Brueck, H.-P. Mayer, L. Thiele, and V. Jungnickel, "Coordinated multipoint: Concepts, performance, and field trial results," *IEEE Commun. Mag.*, vol. 49, no. 2, pp. 102–111, Feb. 2011.

[40] 3GPP, "3GPP Release 12," Tech. Rep., http://www.3gpp.org/specifications/releases/68-release-12, 2014.

[41] 4G Americas, "5G spectrum recommendations," Tech. Rep., The voice for 5G in the Americas, Aug. 2015.

[42] N. Pothecary, *Feedforward Linear Power Amplifiers*, chapter Power amplifiers and system design, p. 98, Artech House, 1999.

[43] P. Rost, C.J. Bernardos, A.D. Domenico, M.D. Girolamo, M. Lalam, A. Maeder, D. Sabella, and D. Wübben, "Cloud technologies for flexible 5G radio access networks," *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 68–76, May 2014.

[44] N. Zhang, N. Cheng, A.T. Gamage, K. Zhang, J.W. Mark, and X. Shen, "Cloud assisted HetNets toward 5G wireless networks," *IEEE Commun. Mag.*, vol. 53, no. 6, pp. 59–65, Jun. 2015.

[45] Y. Shi, J. Zhang, and K.B. Letaief, "Group sparse beamforming for green Cloud-RAN," *IEEE Trans. Wirel. Commun.*, vol. 13, no. 5, pp. 2809–2823, May 2014.

[46] M. Hong, R. Sun, H. Baligh, and Z.-Q. Luo, "Joint base station clustering and beamformer design for partial coordinated transmission in heterogeneous networks," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 226–240, Feb. 2013.

[47] J.S.Chitode, *Digital Communications*, Technical Publications, 2009.

[48] K. Zheng, S. Qu, and X. Yin, "Massive MIMO channel models: A survey," *International Journal of Antennas and Propagation*, pp. 1–10, 2014.

[49] NGMN Alliance, "RAN evolution project: COMP evaluation and enhancement," Tech. Rep. 2, Version 2.0, Mar. 2015.

[50] J. Jose, A. Ashikhmin, T.L. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems," *IEEE Trans. Wirel. Commun.*, vol. 10, no. 8, pp. 2640–2651, Aug. 2011.

[51] H. Li and G. Ascheid, "Long-term window scheduling in multiuser OFDM systems based on large scale fading maps," in *IEEE SPAWC*, Jun. 2012, pp. 324–328.

[52] J. Yu, X. Yin, J. Chen, N. Zhang, Z Zhong, W. Duan, and S. R. Boque, "Channel maps and stochastic models in elevation based on measurements in operating networks," in *WCSP*, Oct 2013, pp. 1–6.

[53] H. Li and G. Ascheid, "Mobility prediction based on graphical model learning," in *IEEE VTC Fall*, Sept 2012, pp. 1–5.

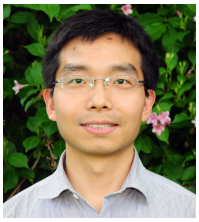
[54] H. Holtkamp, G. Auer, V. Giannini, and H. Haas, "A parameterized base station power model," *IEEE Commun. Lett.*, vol. 17, no. 11, pp. 2033–2035, Nov. 2013.

[55] A.J. Fehske, P. Marsch, and G.P. Fettweis, "Bit per Joule efficiency of cooperating base stations in cellular networks," in *IEEE GLOBECOM Workshops*, Dec 2010, pp. 1406–1411.

[56] B. Khoshnevis, W. Yu, and Y. Lostanlen, "Two-stage channel quantization for scheduling and beamforming in network MIMO systems: Feedback design and scaling laws," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 10, pp. 2028–2042, Oct. 2013.

[57] 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.

[58] J. Fan and R. Li, "Variable selection via nonconcave penalized likelihood and its oracle properties," *J. Amer. Statist. Assoc.*, vol. 96, no. 456, pp. 1348–1360, 2001.
[59] G. Gasso, A. Rakotomamonjy, and S. Canu, "Recovering sparse signals with a certain family of nonconvex penalties and DC programming," *IEEE Trans. Signal Process.*, vol. 57, no. 12, pp. 4686–4698, Dec 2009.
[60] A. Ben-Tal A. Beck and L. Tetruashvili, "A sequential parametric convex approximation method with applications to nonconvex truss topology design problems," *J. of Global Optimization*, vol. 47, no. 1, pp. 29–51, 2010.
[61] T. D. Quoc and M. Diehl, "Sequential convex programming methods for solving nonlinear optimization problems with DC constraints," http://arxiv.org/abs/1107.5841, Jul. 2011.
[62] Michael Grant and Stephen Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1 beta," Jun. 2015.
[63] M. Grant and S. Boyd, "The CVX users' guide: Release 2.1," Tech. Rep., CVX Research, Inc., Jun. 2015.
[64] Gilbert Strang, "Karmarkar's algorithm and its place in applied mathematics," *The mathematical Intelligencer*, vol. 9, no. 2, pp. 4–10, Jun. 1987.
[65] M. Riback, J. Medbo, J.-E. Berg, F. Harrysson, and H. Asplund, "Carrier frequency effects on path loss," in *IEEE VTC Spring*, May 2006, pp. 2717–2721.
[66] S. M. Kay, *Fundamentals of Statistical Signal Processsing: Estimation Theory*, Prentice Hall, 1993.

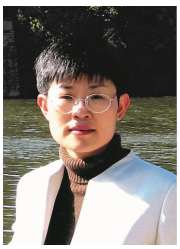


**Pan Cao** (S'12 – M'15) received the B.S. degree in Mechano-Electronic Engineering and the M.S. degree in Information and Signal Processing from Xidian University, Xi'an, P. R. China in 2008 and 2011, respectively, and the Doktor-Ingenieur (Ph.D.) degree in Electrical Engineering from the Technische Universität Dresden (TU Dresden), Germany in 2015. Since March 2015, he works as a Postdoctoral Research Associate in the Institutes for Digital Communications at The University of Edinburgh, UK, supported by the EPSRC project SERAN (Seamless and Efficient Wireless Access for Future Radio Networks).

His current research focuses on designing the novel architectures and algorithms for future wireless communication networks, e.g., green communications, dense networks, large scale antenna array and millimeter-wave systems, by optimization techniques, game theory and stochastic geometry. He received the Best Student Paper Award of the 13th IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Cesme, Turkey in 2012, and the Qualcomm Innovation Fellowship (QInF) Award (one of three winners in the continent of Europe) in 2013.

**Wenjia Liu** received her B. Eng. degree in electronics engineering in 2011 and now is pursuing her Ph.D. degree in signal and information processing, both in the School of Electronics and Information Engineering, Beihang University (BUAA), Beijing, China. From January 2015 to July 2015, she worked as a visiting student at Technische Universität Dresden (TUD), Germany. Her research interests include energy efficient large-scale antenna system, small cell network, and heterogeneous network.

**John S. Thompson** (F'16) is currently a Professor in Signal Processing and Communications at the School of Engineering in the University of Edinburgh. He specializes in antenna array processing, cooperative communications systems and energy efficient wireless communications. He has published in excess of three hundred papers on these topics, including one hundred journal paper publications. He is currently the project coodinator for UK EPSRC SERAN project on 5G technologies and the EU Marie Curie International Training Network project ADVANTAGE, which studies how communications and power engineering can provide future "smart grid" systems). He was an elected Member-at-Large for the Board of Governors of the IEEE Communications Society from 2012-2014, the second largest IEEE Society. He was also a distinguished lecturer on the topic of energy efficient communications and smart grid for the IEEE Communications Society during 2014-2015. He is an editor for the Green Communications and Computing Series that appears regularly in IEEE Communications Magazine. In January 2016, he was elevated to Fellow of the IEEE for contributions to to multiple antenna and multi-hop wireless communications.



**Chenyang Yang** received her Ph.D. degrees in Electrical Engineering from Beihang University (formerly Beijing University of Aeronautics and Astronautics, BUAA), China, in 1997. She has been a full professor with the School of Electronics and Information Engineering, BUAA since 1999. She has published over 200 papers and filed over 60 patents in the fields of energy efficient transmission, CoMP, interference management, cognitive radio, and relay, etc. She was nominated as an Outstanding Young Professor of Beijing in 1995 and was supported by the 1st Teaching and Research Award Program for Outstanding Young Teachers of Higher Education Institutions by Ministry of Education during 1999-2004. She was the chair of Beijing chapter of IEEE Communications Society during 2008-2012, and the MDC chair of APB of IEEE Communications Society during 2011-2013. She has served as TPC Member for a numerous IEEE conferences. She has ever served as an associate editor for IEEE Trans. on Wireless Communications during 2009-2014, guest editor for IEEE Journal of Selected Topics in Signal Processing and IEEE Journal of Selected Areas in Communications, an associate editor-in-chief of Chinese Journal of Communications and Chinese Journal of Signal Processing. Her recent research interests lie in emerging technologies for 5G networks.

**Eduard Jorswieck** (S'01 – M'03 – SM'08) received the Diplom-Ingenieur (M.S.) degree and Doktor-Ingenieur (Ph.D.) degree, both in electrical engineering and computer science, from the Technische Universitt Berlin, Germany, in 2000 and 2004, respectively. He was with the Broadband Mobile Communication Networks Department, Fraunhofer Institute for Telecommunications, Heinrich-Hertz-Institut, Berlin, from 2000 to 2008. From 2005 to 2008, he was a Lecturer with the Technische Universitt Berlin. From 2006 to 2008, he was with the Department of Signals, Sensors and Systems, Royal Institute of Technology, as a Post-Doctoral Researcher and an Assistant Professor. Since 2008, he has been the Head of the Chair of Communications Theory and a Full Professor with the Technische Universitt Dresden, Germany. He is principal investigator in the excellence cluster center for Advancing Electronics Dresden (cfAED) and founding member of the 5G lab Germany (5Glab.de).

His main research interests are in the area of signal processing for communications and networks, applied information theory, and communications theory. He has authored over 85 journal papers, 11 book chapters, some 225 conference papers and 3 monographs on these research topics. Eduard was a co-recipient of the IEEE Signal Processing Society Best Paper Award in 2006 and co-authored papers that won the Best Paper or Best Student Paper Awards at IEEE WPMC 2002, Chinacom 2010, IEEE CAMSAP 2011, IEEE SPAWC 2012, and IEEE WCSP 2012.

Dr. Jorswieck was a member of the IEEE SPCOM Technical Committee (2008–2013), and has been a member of the IEEE SAM Technical Committee since 2015. Since 2011, continuing until 2015, he has been an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING. Since 2008, continuing until 2011, he has served as an Associate Editor of the IEEE SIGNAL PROCESSING LETTERS, and until 2013, as a Senior Associate Editor. Since 2013, he has served as an Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS.