Chalmers Publication Library

Optimal Design of Energy-Efficient HetNets: Joint Precoding and Load Balancing

IEEE International Conference on Communications, ICC 2015, London, UK

(article starts on next page)

Optimal Design of Energy-Efficient HetNets: Joint Precoding and Load Balancing

Jingya Li[†⋆], Emil Björnson[‡], Tommy Svensson[†], Thomas Eriksson[†], and Mérouane Debbah[⋆]

[†]Department of Signals and Systems, Chalmers University of Technology, Gothenburg, Sweden
[⋆]Ericsson Research, Ericsson AB, Gothenburg, Sweden
[‡]Department of Electrical Engineering (ISY), Linköping University, Linköping, Sweden
[⋆]Supélec, Gif-sur-Yvette, France
Email: jingya.li@ericsson.com, emil.bjornson@liu.se, tommy.svensson@chalmers.se,
thomase@chalmers.se, merouane.debbah@supelec.fr

Abstract—This paper considers the downlink of a heterogeneous network, where multiple base stations (BSs) can serve the users by non-coherent multiflow beamforming. We assume imperfect channel state information at both BSs and users. The objective is to jointly optimize the precoding, load balancing, and BS operation mode (active or sleep) for improving the energy efficiency of the network. The considered problem is to minimize the weighted total power consumption (both circuit power and dynamic transmit power), while satisfying per-user quality of service constraints and per-BS transmit power constraints. This problem is non-convex, but we prove that for each combination of BS modes, the considered problem has a hidden convexity structure. Thus, the global optimal solution is obtained by an exhaustive search over all possible BS mode combinations. Furthermore, by iterative convex approximations of the non-convex power consumption functions, a heuristic algorithm is proposed to obtain a local optimal solution with low complexity. Simulation results illustrate that our proposed algorithms significantly reduce the total power consumption, compared to the scheme where all BSs are continuously active. This implies that putting a BS into sleep mode by proper load balancing is an important solution for energy savings in heterogeneous networks.

I. INTRODUCTION

The concept of heterogeneous dense networks, which is based on dense deployment of low-cost and low-power base stations (BSs) coexisting with the macro BSs, has been considered as a key technique to improve the spectral efficiency and energy efficiency of future wireless communication systems [1]. By creating a large number of small cells, these low-power BSs have the potential to offload traffic from macro BSs, reduce the average distance between users and BSs, increase the system capacity and/or reduce the transmit power.

The total power consumption of a network can be modeled with a circuit part that depends on the transceiver hardware and a dynamic part which is a function of the transmitted signal power [2]–[4]. Adding more low-power BSs can reduce the dynamic power consumption due to the shorter propagation distances, but require more hardware; thus, it will increase the circuit power part. Note that the circuit power consumption

also depends on the operational mode of each BS, i.e., whether the BS is active or in the sleep mode. Therefore, to actually improve the overall energy efficiency of a heterogeneous network, the cooperation scheme, BS operational mode and load balancing must be properly and jointly optimized.

Joint load balancing and power control was studied in [5] and [6] for multi-cell single-antenna-BS heterogeneous networks, where different algorithms were proposed to maximize the minimum rate subject to per-BS power constraints. Considering multi-cell multi-antenna-transmitter heterogeneous networks, joint load balancing and precoding algorithms were designed in [7]–[9] to maximize the system utility. In [10], downlink linear precoding problems were studied jointly with active BS selection to either minimize the total transmit power or maximize the sum rate performance. With the objective of improving network energy efficiency, the power allocation, subcarrier allocation and the number of activated transmit antennas were jointly optimized in [11] for the downlink of an OFDM system. However, the work in [11] did not consider joint optimization of precoding vectors and the results are limited to a single cell scenario. In [12], [13], under the assumption of perfect CSI at both BSs and users, joint precoding and BS selection were studied for multi-cell coherent joint transmission systems to minimize the overall BS power consumption. In [14], using a stochastic geometry based model, the energy efficiency of both multi-cell homogeneous and heterogeneous networks was analyzed by considering active and sleep modes for macro BSs with fixed power control. Since both BSs and users are assumed to have a single antenna in [14], precoding design was not considered, either.

In this paper, we study joint precoding and load balancing optimization for energy-efficient heterogeneous networks, where all BSs are allowed to transmit to all users at the same time-frequency resource. The goal is to minimize the weighted total power consumption while satisfying per-user QoS constraints and per-BS transmit power constraints. We show that for each fixed combination of BS modes, the considered optimization problem has a hidden convexity structure. Thus, the optimal solution is obtained by an exhaustive search over all possible BS mode combinations. The obtained global optimal solution serves as an upper bound for any other suboptimal precoding and load balancing solutions. Moreover, we propose

an efficient iterative algorithm that resolves the non-convexity of the original problem by iterative convex approximations of the power consumption functions. Numerical results show that putting a BS into sleep mode by proper load balancing is an important solution for energy savings in heterogeneous networks. Moreover, the BS activation probability depends on the target QoS requirements, as well as the ratio between the circuit power consumed in the active mode and that consumed in the sleep mode.

II. SYSTEM AND SIGNAL MODEL

We consider a downlink heterogeneous network consisting of M base stations (BSs) and K single-antenna users. These M BSs are divided into different tiers or categories, and differ in terms of the number of transmit antennas, the power consumption model, the channel propagation model and CSI quality. BS v is assumed to have N_v antennas. The channels from BS v to user k is assumed to be flat-fading, and denoted by $\mathbf{h}_{k,v}^H \in \mathbb{C}^{1 \times N_v}$ for $v = 1, \dots, M$ and $k = 1, \dots, K$. In practice, these channels are imperfectly known to the user k and the BSs. This is modeled as $\mathbf{h}_{k,v} = \hat{\mathbf{h}}_{k,v} + \mathbf{e}_{k,v}$ with $v = 1, \dots, M$, where $\hat{\mathbf{h}}_{k,v}$ is the known channel estimate (at both the corresponding transmitter and receiver). The error vectors $\mathbf{e}_{k,v} \backsim \mathcal{CN}(\mathbf{0}, \mathbf{E}_{k,v})$ are assumed to be zero-mean with covariance matrix $\mathbf{E}_{k,v} \in \mathbb{C}^{N_v \times N_v}$. The errors can, for example, originate from channel estimation/prediction errors. The received signal at user k is

$$y_k = \sum_{v=1}^M \mathbf{h}_{k,v}^H \mathbf{x}_v + n_k \qquad (1)$$

where $\mathbf{x}_v \in \mathbb{C}^{N_v \times 1}$ is the transmitted signal from BS v, and $n_k \backsim \mathcal{CN}(0, \sigma_k^2)$ is the independent additive receiver noise at user k.

We assume that the BSs are connected via backhaul links, and all BSs are able to transmit to all users at the same time-frequency resource. However, motivated by the fact that tight phase synchronization between BSs is extremely difficult to achieve in practice, only linear non-coherent joint transmission is allowed; that is, each user can be served by a set of BSs using superposition coding. This scheme can be referred to as spatial multiflow transmission [15], which allows each user to receive different data streams from multiple BSs. Define $\mathcal{V} \triangleq \{1, 2, \dots, M\}$ as the set of all BSs in the network, and let $\mathcal{V}_k \subseteq \mathcal{V}$ denote the set of BSs that provide data transmission to user k. Then, the set of users assigned to BS v can be represented by $\mathcal{U}_v = \{k | v \in \mathcal{V}_k\}$. Let $s_{k,v} \backsim \mathcal{CN}(0, 1)$ be the coded information symbols for user k, transmitted from BS v. With non-coherent transmission, the symbols transmitted from different BSs to the same user is assumed to be independent. Then, the desired signals for user k transmitted from BS v is $\mathbf{w}_{k,v} s_{k,v}$, where the vector $\mathbf{w}_{k,v}$ is referred to as the precoding vector for user k from BS v. The aggregated transmitted signal from BS v is

$$\mathbf{x}_v = \sum_{k \in \mathcal{U}_v} \mathbf{w}_{k,v} s_{k,v}. \qquad (2)$$

A. Power Consumption Model

From (2), the expected transmit power from BS v can be calculated as

$$P_{\text{trans},v} = \sum_{k \in \mathcal{U}_v} \|\mathbf{w}_{k,v}\|^2 \, \mathbb{E}\left\{|s_{k,v}|^2\right\} = \sum_{k \in \mathcal{U}_v} \|\mathbf{w}_{k,v}\|^2 \qquad (3)$$

where the operator $\mathbb{E}\{\cdot\}$ stands for expectation, and $\|\cdot\|$ represents the Euclidean norm. In this paper, we adopt the linear approximated power consumption model proposed in [2, Eq. (4-3)] for 10MHz bandwidth, where the total consumed power of BS v, for $v \in \mathcal{V}$, is

$$P_{\mathbf{v}} = \begin{cases} N_v P_{\text{active},v} + \Delta_v P_{\text{trans},v}, & 0 < P_{\text{trans},v} \leq P_{v,\max} \\ N_v P_{\text{sleep},v}, & P_{\text{trans},v} = 0 \end{cases} \qquad (4)$$

where $P_{\text{active},v}$ is the hardware power consumption at BS v at the minimum non-zero transmit power, $P_{\text{sleep},v}$ denotes the sleep mode power consumption of BS v with $P_{\text{sleep},v} \leq P_{\text{active},v}$. Note that in the sleep mode, due to the power consumption for DC-DC power supply, mains supply, active cooling, maintaining backhaul connections, and enabling fast turn on control signaling, etc, $P_{\text{sleep},v}$ is typically non-zero [2]. Here, $P_{v,\max}$ is the peak transmit power constraint for BS v. The scaling factor, $\Delta_v \geq 1$, models the extra power consumed by the power amplifiers at BSs when the transmitted power is $P_{\text{trans},v}$. Some example values of $P_{\text{active},v}$, $P_{\text{sleep},v}$, $P_{v,\max}$ and Δ_v for different BS types can be found in [2, Table 8].

B. Aggregated Received SINR

When each user performs *successive interference cancellation* on its own information symbols, and treats both channel errors and co-user interference as Gaussian noise [16], the total achievable spectral efficiency of user k is given as $R_k = \log_2(1 + \gamma_k)$ bit/s/Hz, where the *aggregated* SINR at user k is

$$\gamma_k = \frac{\sum_{v \in \mathcal{V}_k} \left| \hat{\mathbf{h}}_{k,v}^H \mathbf{w}_{k,v} \right|^2}{I_k + E_k + \sigma_k^2} \qquad (5)$$

where

$$I_k \triangleq \sum_{v \in \mathcal{V}} \sum_{\substack{l \in \mathcal{U}_v \\ l \neq k}} \mathbf{w}_{l,v}^H \left(\hat{\mathbf{h}}_{k,v} \hat{\mathbf{h}}_{k,v}^H + \mathbf{E}_{k,v} \right) \mathbf{w}_{l,v} \qquad (6)$$

denotes the co-user interference, and

$$E_k \triangleq \sum_{v \in \mathcal{V}_k} \mathbf{w}_{k,v}^H \mathbf{E}_{k,v} \mathbf{w}_{k,v} \qquad (7)$$

is the interference caused by the imperfect knowledge of the channels to user k.

C. Problem Formulation

The focus of this paper is on the joint design of load balancing ($\mathcal{U}_v$) and precoding vectors ($\mathbf{w}_{k,v}$) for $v = 1, \dots, M$ and $k = 1, \dots, K$ to minimize the weighted total power consumption of the system, while satisfying a set of SINR constraints (or, equivalently, spectral efficiency constraints) for each user and a set of transmit power constraints for each

BS. These constraints are referred to as the QoS constraints. With (3), (4) and (5) in hand, the optimization problem can be formulated as

$$\underset{\{\mathbf{w}_{k,v}, \mathcal{U}_v\}}{\text{minimize}} \sum_{v=1}^{M} a_v P_{\mathbf{v}}$$

$$\text{subject to } \gamma_k \geq \Gamma_k, \forall k \tag{8}$$

$$P_{\text{trans},v} \leq P_{v,\max}, \forall v$$

where $\Gamma_k > 0$ is the target SINR value for user k. The weights $a_v > 0$ are used to balance the power consumptions of different BSs. For the rest of the paper, we assume that the problem (8) has at least one feasible solution, which is reasonable in dense networks with an over-provisioning of access points. In practice, if no feasible solution exists, the SINR constraints have to be relaxed either by decreasing the target SINRs or by removing users [17].

III. Optimal Precoding and Load Balancing

From (8), we see that the load balancing and precoding design are inherently coupled with each other. In fact, by setting $\mathbf{w}_{k,v} = \mathbf{0}$ for $v \notin \mathcal{V}_k$ (or, equivalently, by setting $\mathbf{w}_{k,v} = \mathbf{0}$ for $k \notin \mathcal{U}_v$), the original problem (8) is equivalent to

$$\underset{\{\mathbf{w}_{k,v}\}}{\text{minimize}} \sum_{v=1}^{M} a_v P_{\mathbf{v}}$$

$$\text{subject to } \gamma_k \geq \Gamma_k, \forall k \tag{9}$$

$$\sum_{k=1}^{K} \|\mathbf{w}_{k,v}\|^2 \leq P_{v,\max}, \forall v$$

where $P_{\mathbf{v}}$ can be rewritten as

$$P_{\mathbf{v}} = \begin{cases} N_v P_{\text{active},v} + \Delta_v \sum_{k=1}^{K} \|\mathbf{w}_{k,v}\|^2, & 0 < P_{\text{trans},v} \leq P_{v,\max} \\ N_v P_{\text{sleep},v}, & P_{\text{trans},v} = 0 \end{cases} \tag{10}$$

and γ_k is reformulated as

$$\gamma_k = \frac{\sum_{v=1}^{M} \left| \hat{\mathbf{h}}_{k,v}^H \mathbf{w}_{k,v} \right|^2}{I_k + E_k + \sigma_k^2} \tag{11}$$

with I_k rewritten as

$$I_k \triangleq \sum_{v=1}^{M} \sum_{\substack{l=1 \\ l \neq k}}^{K} \mathbf{w}_{l,v}^H \left(\hat{\mathbf{h}}_{k,v} \hat{\mathbf{h}}_{k,v}^H + \mathbf{E}_{k,v} \right) \mathbf{w}_{l,v} \tag{12}$$

and E_k replaced by

$$E_k \triangleq \sum_{v=1}^{M} \mathbf{w}_{k,v}^H \mathbf{E}_{k,v} \mathbf{w}_{k,v}. \tag{13}$$

Hence, by solving the global precoding optimization problem (9), we can immediately obtain the optimal load balancing solution; that is, the set of users assigned to BS v is $\mathcal{U}_v = \{k | \mathbf{w}_{k,v} \neq \mathbf{0}, k \in \{1, \dots, K\}\}$, and the set of BSs that provide data transmission to user k is $\mathcal{V}_k = \{v | \mathbf{w}_{k,v} \neq \mathbf{0}, v \in \mathcal{V}\}$.

The optimization problem (9) is not convex. In particular, the power consumption function in (10), which is in the

form of a fixed transaction cost function, leads to a hard combinatorial problem [18]. Moreover, the SINR constraints of (9) do not have a standard convex form. In the following, we first show that, for each combination of BS modes, problem (9) can be reformulated into a convex problem. Then, the global optimum can be found by an exhaustive combinatorial search over these 2^M convex problems with different mode selections.

Define $\mathbf{w}_k \triangleq \left[\mathbf{w}_{k,1}^T, \mathbf{w}_{k,2}^T, \dots, \mathbf{w}_{k,M}^T \right]^T \in \mathbb{C}^{\left(\sum_{v=1}^{M} N_v \right) \times 1}$ as the aggregated precoding vector for user k from all BSs. To allow only non-coherent joint transmission from multiple BSs, the covariance matrix of the estimated channels from all BSs to user k is defined as [17]

$$\hat{\mathbf{R}}_k \triangleq \begin{bmatrix} \hat{\mathbf{R}}_{k,1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \hat{\mathbf{R}}_{k,2} & \mathbf{0} & \vdots \\ \vdots & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \vdots & \mathbf{0} & \hat{\mathbf{R}}_{k,M} \end{bmatrix}$$

where $\hat{\mathbf{R}}_{k,v} \triangleq \hat{\mathbf{h}}_{k,v} \hat{\mathbf{h}}_{k,v}^H \in \mathbb{C}^{N_v \times N_v}$ for $v = 1, \dots, M$. Similarly, we define the covariance matrix of the channel errors for user k as $\mathbf{E}_k \triangleq \text{diag}(\mathbf{E}_{k,1}, \mathbf{E}_{k,2}, \dots, \mathbf{E}_{k,M})$. Then, the received SINR, γ_k in (11), can be rewritten as

$$\gamma_k = \frac{\mathbf{w}_k^H \hat{\mathbf{R}}_k \mathbf{w}_k}{\sum_{\substack{l=1 \\ l \neq k}}^{K} \mathbf{w}_l^H \left(\hat{\mathbf{R}}_k + \mathbf{E}_k \right) \mathbf{w}_l + \mathbf{w}_k^H \mathbf{E}_k \mathbf{w}_k + \sigma_k^2}.$$

Using the semi-definite relaxation trick from [19], the SINR constraints can be rewritten in convex forms as a set of second order cone constraints. To do so, we define $\mathbf{W}_k \triangleq \mathbf{w}_k \mathbf{w}_k^H \succeq 0$, and

$$\mathbf{Q}_v \triangleq \text{diag}(\mathbf{Q}_{1,v}, \mathbf{Q}_{2,v}, \dots, \mathbf{Q}_{M,v}) \tag{14}$$

where

$$\mathbf{Q}_{i,v} = \begin{cases} \mathbf{I}_{N_v}, & \text{if } i = v \\ \mathbf{0}_{N_v \times N_v}, & \text{otherwise.} \end{cases} \tag{15}$$

By gathering the power weights in a diagonal matrix form as $\mathbf{A} \triangleq \text{diag}(a_1 \Delta_1 \mathbf{I}_{N_1}, a_2 \Delta_2 \mathbf{I}_{N_2}, \dots, a_M \Delta_M \mathbf{I}_{N_M})$ and noting that $\mathbf{w}_k^H \mathbf{Q} \mathbf{w}_k = \text{Tr}(\mathbf{Q} \mathbf{W}_k)$ for any matrix $\mathbf{Q}$, problem (9) can be reformulated as

$$\underset{\{\mathbf{W}_k \succeq 0\}}{\text{minimize}} \sum_{k=1}^{K} \text{Tr}(\mathbf{A} \mathbf{W}_k) + \mathcal{J}(\mathbf{z})$$

$$\text{subject to } \text{Tr}\left(\hat{\mathbf{R}}_k \mathbf{W}_k \right) - \Gamma_k \sum_{\substack{l=1 \\ l \neq k}}^{K} \text{Tr}\left((\hat{\mathbf{R}}_k + \mathbf{E}_k) \mathbf{W}_l \right)$$

$$- \Gamma_k \text{Tr}(\mathbf{E}_k \mathbf{W}_k) \geq \Gamma_k \sigma_k^2, \forall k \tag{16}$$

$$\sum_{k=1}^{K} \text{Tr}(\mathbf{Q}_v \mathbf{W}_k) \leq z_v P_{v,\max}, \forall v$$

$$z_v \in \{0, 1\}, \forall v$$

where

$$\mathcal{J}(\mathbf{z}) = \sum_v a_v N_v \left(P_{\text{active},v} z_v + P_{\text{sleep},v} (1 - z_v) \right) \tag{17}$$

and with the additional constraints $\text{rank}\,(\mathbf{W}_k) = 1, \forall k$. Here, z_v is the BS mode indicator for $v \in \mathcal{V}$: $z_v = 1$ if BS v is active, and $z_v = 0$ if BS v is in the sleep mode. $\mathbf{W}_k \succeq 0$ represents that the matrix $\mathbf{W}_k$ is positive semidefinite, and $\text{Tr}(\cdot)$ refers to the matrix trace.

Based on [20, Theorem 1], it can be shown that (16) always has a rank one solution, if the problem is feasible. Therefore, the rank-one constraints can be dropped without loss of optimality. For any combination of BS modes, the problem (16) is a convex semi-definite program. Thus, by using standard convex optimization software, such as CVX or YALMIP, we can efficiently obtain the optimal solution for each fixed $\mathbf{z}$. By searching over all 2^M mode combinations, the global optimum is found by choosing the one with the lowest weighted total power consumption. This global optimum can serve as an upper bound for any other suboptimal heuristic load-balancing and precoding algorithm.

IV. ITERATIVE HEURISTIC ALGORITHM DESIGN

In this section, we tackle the non-convex problem (16) by iterative convex approximations of the power consumption functions. In particular, each iteration solves a problem with a modified objective function, which is convex. This convex objective function is updated in each iteration such that most of the BSs with small transmit powers in the solution are driven to the sleep mode. The proposed algorithm will find a local optimum to the original problem in (16).

Note that $0 \le P_{\text{trans},v} \le P_{v,\max}$ for each BS v, $v \in \mathcal{V}$. Thus, the total consumed power of BS v, P_v in (4), can be relaxed with its convex envelope, $P_\text{v}^{\text{c.e.}}$ over the interval $[0, P_{v,\max}]$, where

$$P_\text{v}^{\text{c.e.}} \left(P_{t,v}\right) \triangleq N_v P_{\text{sleep},v} + \Delta_v' P_{\text{trans},v} \tag{18}$$

with

$$\Delta_v' \triangleq \frac{N_v \left(P_{\text{active},v} - P_{\text{sleep},v}\right)}{P_{v,\max}} + \Delta_v \tag{19}$$

which is the largest convex function smaller or equal to P_v over the interval. Replacing P_v with $P_\text{v}^{\text{c.e.}}$, problem (9) is relaxed to a convex optimization problem

$$\begin{aligned}
\underset{\{\mathbf{W}_k \succeq 0\}}{\text{minimize}} \quad & \sum_{k=1}^{K} \text{Tr}\,(\mathbf{A}'\mathbf{W}_k) + \sum_{v=1}^{M} a_v N_v P_{\text{sleep},v} \\
\text{subject to} \quad & \text{Tr}\left(\hat{\mathbf{R}}_k \mathbf{W}_k\right) - \Gamma_k \sum_{\substack{l=1 \\ l \ne k}}^{K} \text{Tr}\left((\hat{\mathbf{R}}_k + \mathbf{E}_k)\mathbf{W}_l\right) \\
& - \Gamma_k \text{Tr}\,(\mathbf{E}_k \mathbf{W}_k) \ge \Gamma_k \sigma_k^2, \forall k \\
& \sum_{k=1}^{K} \text{Tr}\,(\boldsymbol{Q}_v \mathbf{W}_k) \le P_{v,\max}, \forall v
\end{aligned} \tag{20}$$

where $\mathbf{A}'$ is a modified block diagonal matrix of $\mathbf{A}$, with Δ_v replaced by Δ_v' for each block v. Note that based on [20, Theorem 1], the rank one constraints are dropped without loss of optimality. Compared to the original problem (9), the relaxed problem (20) has the same feasible set, but a modified objective function. The optimal value of (20) is a lower bound on the optimal value of the original problem (9).

The proposed iterative heuristic algorithm is as follows:

1) $i := 0$; Initialize $\mathbf{W}_k^{(0)}$ for $k = 1, \ldots, K$ by solving (20).
2) $i := i + 1$; Obtain the transmit power of each BS v as $P_{\text{trans},v}^{(i-1)} = \sum_{k=1}^{K} \text{Tr}\left(\boldsymbol{Q}_v \mathbf{W}_k^{(i-1)}\right)$. Define $\hat{P}_\text{v}^{(i)}\left(P_{\text{trans},v}\right) \triangleq N_v P_{\text{sleep},v} + \Delta_v^{(i)} P_{\text{trans},v}$, where

$$\Delta_v^{(i)} \triangleq \frac{N_v \left(P_{\text{active},v} - P_{\text{sleep},v}\right)}{P_{\text{trans},v}^{(i-1)} + \delta} + \Delta_v. \tag{21}$$

Solve the modified optimization problem

$$\begin{aligned}
\underset{\{\mathbf{W}_k \succeq 0\}}{\text{minimize}} \quad & \sum_{k=1}^{K} \text{Tr}\left(\mathbf{A}^{(i)}\mathbf{W}_k\right) + \sum_{v=1}^{M} a_v N_v P_{\text{sleep},v} \\
\text{subject to} \quad & \text{Tr}\left(\hat{\mathbf{R}}_k \mathbf{W}_k\right) - \Gamma_k \sum_{\substack{l=1 \\ l \ne k}}^{K} \text{Tr}\left((\hat{\mathbf{R}}_k + \mathbf{E}_k)\mathbf{W}_l\right) \\
& - \Gamma_k \text{Tr}\,(\mathbf{E}_k \mathbf{W}_k) \ge \Gamma_k \sigma_k^2, \forall k \\
& \sum_{k=1}^{K} \text{Tr}\,(\boldsymbol{Q}_v \mathbf{W}_k) \le P_{v,\max}, \forall v
\end{aligned}$$
$$\tag{22}$$

where $\mathbf{A}^{(i)}$ is the modified block diagonal matrix of $\mathbf{A}$, with Δ_v replaced by $\Delta_v^{(i)}$ for each block v.

3) Let $\mathbf{W}_k^{(i)}$ be the solution to this problem.
4) If $P_{\text{trans},v}^{(i-1)}$ and $P_{\text{trans},v}^{(i)}$ are approximately equal for each v, return $\mathbf{W}_k^* := \mathbf{W}_k^{(i)}$. Otherwise, go back to step 2).

Note that δ in (21) is a non-negative small value, which can be interpreted as a soft threshold for deciding when a BS is set to the sleep mode. Define $P_{\text{trans},v}^* \triangleq \sum_{k=1}^{K} \text{Tr}\,(\boldsymbol{Q}_v \mathbf{W}_k^*)$. Thus, for $P_{\text{trans},v}^* \gg \delta$, we have $\hat{P}_\text{v}\left(P_{t,v}^*\right) \triangleq N_v P_{\text{sleep},v} + \left(\frac{N_v(P_{\text{active},v} - P_{\text{sleep},v})}{P_{\text{trans},v}^* + \delta} + \Delta_v\right) P_{\text{trans},v}^* \approx N_v P_{\text{active},v} + \Delta_v P_{\text{trans},v}^* = P_\text{v}\left(P_{\text{trans},v}^*\right)$, and BS v is in the active mode; while for $P_{\text{trans},v}^* = 0$, $\hat{P}_\text{v}\left(P_{\text{trans},v}^*\right) \triangleq N_v P_{\text{sleep},v}$ and BS v is in the sleep mode. For each iteration as shown in step 2), when $P_{\text{trans},v}^{(i-1)}$ is small, the modified $\Delta_v^{(i)}$ in (21) becomes large, i.e., the derivative of the power consumption function $\hat{P}_\text{v}^{(i)}\left(P_{\text{trans},v}\right)$ increases. Therefore, the modified optimization problem (22) will push the smalls $P_{\text{trans},v}^{(i-1)}$ to zero; that is, the BSs with small transmit powers in the solution to the previous problem are driven to the sleep mode. This leads to sparse solutions of $\mathbf{W}_k^*$. Note that the modified objective function of problem (22) is on the form of the objective function in [18, Eq. (21)], which always converges to a local optimal point. A proof of convergence for this type of heuristic algorithms is given in [18, Appendix B].

V. NUMERICAL RESULTS

In this section, numerical results are presented to illustrate our analytical results and the proposed algorithms. The propagation environment is a simplified version of the dense urban information society model used in the METIS project [21],
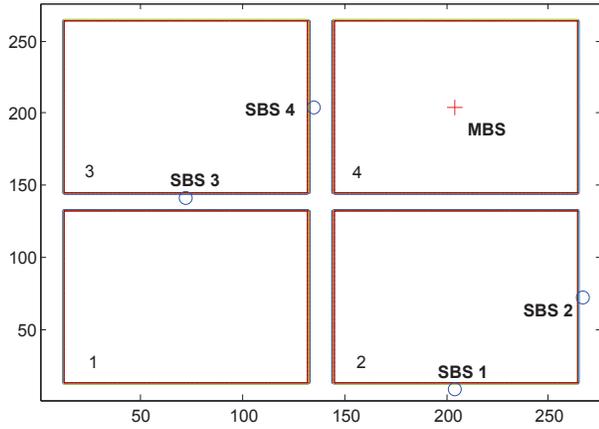
Fig. 1. The MBS (cross) and SBSs (circles) deployment considered in Section V.

TABLE I
POWER MODEL PARAMETERS FOR DIFFERENT BS TYPES.

BS type	N_v	$P_{v,\max}$ [W]	$P_{\text{sleep},v}$ [W]	$P_{\text{active},v}$ [W]	Δ_v
MBS	4	39.8	75.0	130.0	4.7
SBS	2	6.3	39.0	56.0	2.6

as illustrated in Fig. 1. The model consists of four square-shaped buildings of dimensions 120m×120m, each with 6 floors. A macro BS (MBS) is complemented with 4 small cell BSs (SBSs). The MBS has 4 transmit antennas, and the SBSs have 2 transmit antennas each. Five users are randomly and uniformly dropped in the network, whereof 4 users are indoors and 1 user is outdoors in every user drop. The system bandwidth is 10MHz. Here, we adopt the indoor and outdoor propagation models, PS#1 - PS#4, identified in METIS. More details regarding network deployment and propagation modes can be found in [21, Table 3.7 and Section 8.1]. We assume independent Rayleigh small-scale fading. The MMSE channel estimation errors are calculated based on [22] with the total pilot power $p = P_{v,\max}/2$. Table I shows the power model parameters and is based on [2, Table 6 and Table 8].

Three different schemes, namely "Optimal", "Heuristic" and "All Active", are compared. The "Optimal" scheme obtains the global optimal solution by an exhaustive search over all 2^5 possible BS mode combinations. The "Heuristic" scheme follows the algorithm proposed in Section IV, and the value of the soft threshold δ is set to 10^{-4}. The "All Active" scheme is used as our performance baseline, which solves the optimization problem (16) by assuming that all BSs are active, i.e., the BS mode indicator $z_v = 1$ for all BSs $v \in \mathcal{V}$. For each scheme, the performance is average over 1000 independent user drops that provide feasible solutions for our optimization problem (16). For each user drop, the algorithms are evaluated over 50 independent channel realizations. The weights a_v are set to 1 for all BSs.

Define the dynamic part of the total power consumption as the total RF power (i.e., $\sum_{v=1}^{M} a_v \Delta_v P_{\text{trans},v}$), and the remaining part of the total power consumption as the circuit power (i.e., $\sum_v a_v N_v \left(P_{\text{active},v} z_v + P_{\text{sleep},v} (1 - z_v) \right)$). Figs. 2 and 3 demonstrate the total RF power and the total power
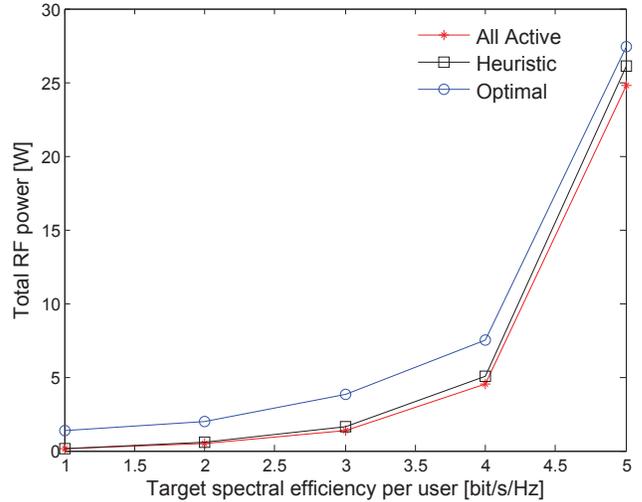
Fig. 2. Total RF power (the dynamic part $\sum_{v=1}^{M} a_v \Delta_v P_{\text{trans},v}$) vs. target spectral efficiency per user (R_k).

consumption as a function of target spectral efficiency per user, respectively. As expected, the total power consumption and the RF power increase as the target spectral efficiency increases. Fig. 2 shows that the RF power for the "All Active" scheme is less than that of the "Heuristic" and "Optimal" schemes. This is expected since all BSs are active in the "All Active" scheme, whileas for the "Heuristic" and "Optimal" schemes, some BSs are put into the sleep mode. With more BSs being active, the "All Active" scheme provides better energy-focusing and less propagation losses between the users and the transmitters, and will therefore reduce the total RF power. However, as can be seen from Fig. 3, compared to the "All Active" scheme, the "Heuristic" and "Optimal" schemes can substantially reduce the total power consumption, especially when the target QoS is small. This is because the circuit power consumption under the sleep mode is much lower compared to the one under the active mode, i.e., $P_{\text{sleep},v} \ll P_{\text{active},v}$. For the "All Active" scheme, the increase in the circuit part from the extra power consumed by activating BSs clearly outweighs the decrease in the dynamic part. This implies that putting a BS into sleep mode by proper load balancing is an important solution for energy savings in heterogeneous networks.

Fig. 4 demonstrates the BS activation probability versus the target spectral efficiency per user. Here, the activation probability of the SBS is averaged over the probabilities of the four SBSs depicted in Fig. 1. We see that for the "All Active" scheme, the activation probabilities of the MBS and SBS is always one, since all BSs are always active in this scheme. Moreover, as anticipated, for both the "Heuristic" and "Optimal" schemes, the BS activation probabilities of the MBS and SBS increase as the target spectral efficiency per user increases. This is because in order to satisfy the raised QoS expectations of all users, the probability that a BS becomes active should increase so as to provide better energy-focusing and less propagation losses. Over the considered range of target spectral efficiency per user, the "Optimal" scheme has lower activation probability for the MBS and
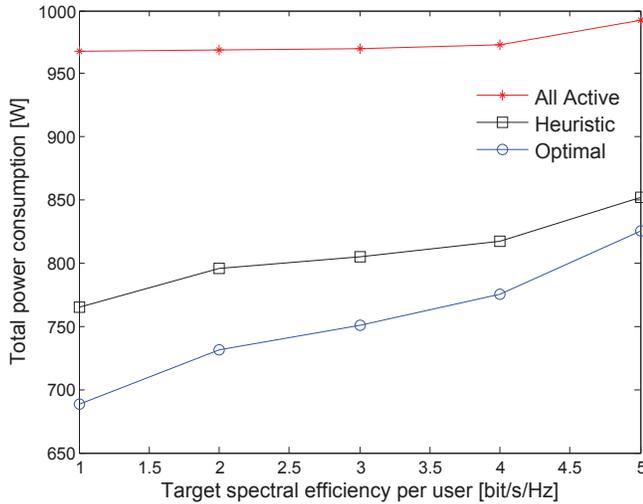
Fig. 3. Total power consumption ($\sum_{v=1}^{M} a_v P_v$) vs. target spectral efficiency per user (R_k).
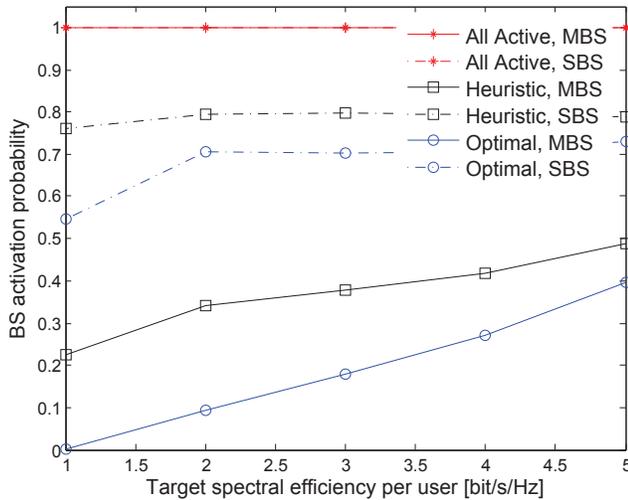
Fig. 4. BS activation probability vs. target spectral efficiency per user (R_k).

higher activation probability for the SBS as compared to the "Heuristic" scheme. Note that the circuit power consumed under the active mode $P_{\text{active},v}$ for the MBS is much higher than that of the SBSs. Thus, as shown in Fig. 3, the "Optimal" scheme results in better energy saving as compared to the "Heuristic" scheme.

VI. CONCLUSIONS

This paper analyzed the energy efficiency in the downlink of heterogeneous networks, where each user can be served by non-coherent joint transmission from multiple transmitters. More specifically, assuming imperfect CSI at both BSs and users, the optimal precoding vectors, load balancing, and BS operational modes were obtained to minimize the weighted total power consumption. Moreover, an iterative heuristic algorithm was proposed to find a local optimum with relatively low complexity. Numerical results showed that the total power consumption can be greatly reduced by putting a BS into sleep mode by using proper load balancing.

REFERENCES

[1] A. Ghosh, N. Mangalvedhe, R. Ratasuk, B. Mondal, M. Cudak, E. Visotsky, T. Thomas, J. Andrews, P. Xia, H. Jo, H. Dhillon, and T. Novlan, "Heterogeneous cellular networks: From theory to practice," *IEEE Commun. Mag.*, vol. 50, no. 6, pp. 54–64, June 2012.

[2] EU FP7 INFSO-ICT-247733 EARTH, D2.3, "Energy efficiency analysis of the reference systems, areas of improvements and target breakdown," Jan. 2012.

[3] E. Björnson, M. Kountouris, and M. Debbah, "Massive MIMO and small cells: Improving energy efficiency by optimal soft-cell coordination," in *Proc. IEEE Int. Conf. Telecommun.(ICT)*, 2013.

[4] J. Li, M. M. Matthaiou, S. Jin, and T. Svensson, "Energy efficiency analysis of rank-1 Ricean fading MIMO channels," in *Proc. IEEE Int. Workshop on Signal Processing Advances in Wireless Commun. (SPAWC)*, June 2014.

[5] R. Sun and Z.-Q. Luo, "Globally optimal joint uplink base station association and power control for max-min fairness," in *Proc. IEEE Int. Conf. Acoust., Speech and Signal Process. (ICASSP)*, May 2014, pp. 454–458.

[6] R. Sun, M. Hong, and Z.-Q. Luo, "Joint downlink base station association and power control for max-min fairness: Computation and complexity," 2014. Avaliable: http://arxiv.org/abs/1407.2791.

[7] 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. Select. Areas Commun.*, vol. 31, no. 2, pp. 226–240, Feb. 2013.

[8] M. Hong and Z.-Q. Luo, "Distributed linear precoder optimization and base station selection for an uplink heterogeneous network," *IEEE Trans. Signal Process.*, vol. 61, no. 12, pp. 3214–3228, June 2013.

[9] M. Sanjabi, M. Razaviyayn, and Z.-Q. Luo, "Optimal joint base station assignment and beamforming for heterogeneous networks," *IEEE Trans. Signal Process.*, vol. 62, no. 8, pp. 1950–1961, Apr. 2014.

[10] W.-C. Liao, M. Hong, Y.-F. Liu, and Z.-Q. Luo, "Base station activation and linear transceiver design for optimal resource management in heterogeneous networks," 2014. Avaliable: http://arxiv.org/abs/1309.4138.

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

[12] Y. Cheng, M. Pesavento, and A. Philipp, "Joint network optimization and downlink beamforming for CoMP transmissions using mixed integer conic programming," *IEEE Trans. Signal Process.*, vol. 61, no. 16, pp. 3972–3987, Aug 2013.

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

[14] Y. S. Soh, T. Quek, M. Kountouris, and H. Shin, "Energy efficient heterogeneous cellular networks," *IEEE J. Select. Areas Commun.*, vol. 31, no. 5, pp. 840–850, May 2013.

[15] H. Holma and A. Toskala, *LTE Advanced: 3GPP Solution for IMT-Advanced*, 1st ed. Wiley, 2012.

[16] M. Medard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Trans. Inform. Theory*, vol. 46, no. 3, pp. 933–9446, 2000.

[17] R. Stridh, M. Bengtsson, and B. Ottersten, "System evaluation of optimal downlink beamforming with congestion control in wireless communication," *IEEE Trans. Wireless Commun.*, vol. 5, no. 4, pp. 743–751, Apr. 2006.

[18] M. S. Lobo, M. Fazel, and S. Boyd, "Portfolio optimization with linear and fixed transaction costs," *Annals of Operations Research*, vol. 152, no. 1, pp. 341–365, July 2007.

[19] M. Bengtsson and B. Ottersten, *Optimal and suboptimal transmit beamforming.* in Handbook of Antennas in Wireless Communications, (L. C. Godara, ed.), CRC Press, 2001.

[20] E. Björnson, N. Jalden, M. Bengtsson, and B. Ottersten, "Optimality properties, distributed strategies, and measurement-based evaluation of coordinated multicell OFDMA transmission," *IEEE Trans. Signal Process.*, vol. 59, no. 12, pp. 6086–6101, Dec 2011.

[21] EU FP7 INFSO-ICT-317669 METIS, D6.1, "Simulation guidelines," Oct. 2013.

[22] T. Wild, "Comparing downlink coordinated multi-point schemes with imperfect channel knowledge," in *Proc. IEEE Veh. Technol. Conf. (VTC)*, Sept 2011, pp. 1–5.