BASE STATION ACTIVATION AND LINEAR TRANSCEIVER DESIGN FOR UTILITY MAXIMIZATION IN HETEROGENEOUS NETWORKS

Wei-Cheng Liao, Mingyi Hong, and Zhi-Quan Luo

Department of Electrical and Computer Engineering
University of Minnesota, Minneapolis, MN 55455, USA

ABSTRACT

In a densely deployed Heterogeneous network (HetNet), the number of pico/micro base stations (BS) can be comparable or more than the number of the users. To reduce the operational overhead of the HetNet, selection of serving BSs becomes an important design issue. In this work, we propose to jointly optimize the transceiver and active BSs to trade off the overall spectrum efficiency with the operational overhead. We formulate this problem as a regularized sum rate maximization problem and solve it using sparse optimization techniques. The proposed algorithm is guaranteed to converge to a local optimal solution. The efficiency and the efficacy of the algorithm are demonstrated via realistic numerical simulations.

Index Terms— Heterogeneous Networks, LASSO, BS selection/clustering

1. INTRODUCTION

A popular approach to cope with the explosive growth of mobile wireless data traffic is to deploy more transmitters to cellular networks, especially near cell edges or hot spots. As a result, the architecture of traditional cellular networks has undergone a fundamental shift in which a macro base station (BS) and various pico/micro BSs may coexist in a cell and collaborate to serve the users while sharing the same spectrum [1]. In LTE-A [2], two main modes of cooperation have been considered [3]: *(1)* Joint Processing (JP): several BSs jointly transmit to users by sharing transmitted data via high speed backhaul network; *(2)* Coordinated Beamforming (CB): BSs avoid interference by cooperatively design the transmit beamformers without sharing the users' data. Clearly these two approaches complement each other—JP achieves high spectrum efficiency, while the CB requires less backhaul capacity. Recently there has been many works that propose to strike a balance between these two approaches [4, 5, 6, 7, 8, 9]. The idea is to cluster the BSs together such that JP is used only within each BS cluster.

The main strength of the Heterogeneous network (HetNet) lies in its architecture: it can bring the transmitters and receivers close to each other, so that significantly less transmit power is needed to deliver the same signal quality. As a result, in the HetNet the low-power BSs are typically deployed densely within each cell. However, increasing the number of BSs also incurs substantial operational costs [3, 10]. These costs can take the form of power consumption, complexity for encoding/decoding, and overhead related to BS management or information exchanges among the BSs. To keep the operational cost under control, it is necessary to appropriately select a subset of active BSs while shutting down the rest. To the best of our knowledge, none of the existing works on BS clustering considers this factor in their formulations; see e.g., [4, 5, 6, 7, 8, 9]. As a result, the solutions computed by these algorithms typically require most BSs in the network to remain active.

In this work, we propose to design the downlink transmit strategies for a HetNet so as to achieve a high spectral efficiency with few active base stations. Mathematically, a small number of active BSs is equivalent to requiring that the most of the precoders of the BSs are zero. This leads to a sparsity requirement on the precoders. This observation motivates us to formulate the joint design problem as a *penalized* sum rate maximization problem, where the penalization is appropriately chosen to promote sparsity in the precoders. In the compressive sensing community [11, 12, 13], it is well-known that sparsity can be induced by using a so-called LASSO regularization term (e.g., [14]). Recently, this idea has been introduced to different applications in wireless communications, e.g., antenna selection in downlink transmit beamforming [15] and the joint precoder design with dynamical BS clustering [7, 8, 9]. Since, these works do not seek to reduce the number of active BSs in the network, none of them can be directly applied to our considered problem.

The main contribution of this paper is a novel single-stage formulation of the joint dynamic active BS selection and linear transceiver design problem. An efficient algorithm, inspired by the weighted minimum mean square error (WMMSE) algorithm [16], is then devised to compute a locally optimal solution for the problem. Extension to include dynamical BS clustering into the formulation is also considered.

2. SIGNAL MODEL AND PROBLEM STATEMENT

Consider a downlink multi-cell HetNet consisting of a set $\mathcal{K} \triangleq \{1, \ldots, K\}$ of cells. Within each cell k, there is a set $\mathcal{Q}_k = \{1, \ldots, Q_k\}$ distributed base stations (BSs) which provide service to users located in different areas of the cell. Assume that in each cell k, a central controller has the knowledge of all the users' data as well as their channel state information (CSI). Its objective is to determine the precoders for all BSs within the cell. Let $\mathcal{I}_k \triangleq \{1, \ldots, I_k\}$ denote the users located in cell k, and $\mathcal{I} \triangleq \bigcup_{k=1}^{K} \mathcal{I}_k$ is the set of all users. Each user $i_k \in \mathcal{I}$ is served jointly by a subset of BSs in $\mathcal{Q}_k$. For simplicity of notations, let us assume that each BS has M transmit antennas, and each user has N receive antennas.

Let us denote $\mathbf{v}_{i_k}^{q_k} \in \mathbb{C}^{M\times 1}$ as the transmit beamformer of BS q_k to user i_k, and $\mathbf{v} \triangleq \{\mathbf{v}_{i_k}^{q_k} | i_k \in \mathcal{I}, q_k \in \mathcal{Q}_k, k \in \mathcal{K}\}$ is the set of all transmit beamformers. Hence, the virtual precoder for user i_k from all BSs in cell k is $\mathbf{v}_{i_k} \triangleq [(\mathbf{v}_{i_k}^{1_k})^H, (\mathbf{v}_{i_k}^{2_k})^H, \ldots, (\mathbf{v}_{i_k}^{Q_k})^H]^H$, and $\mathbf{v}^{q_k} \triangleq [(\mathbf{v}_{1_k}^{q_k})^H, (\mathbf{v}_{2_k}^{q_k})^H, \ldots, (\mathbf{v}_{I_k}^{q_k})^H]^H$ is defined as the collection of all transmit beamformers of BS q_k. Let $s_{i_k} \in \mathbb{C}$ denote the unit variance transmitted data for user i_k, then the transmitted signal of BS q_k can be expressed as

This work is supported by the National Science Foundation, grant number CCF-1216858.

$$\mathbf{x}^{q_k} = \sum_{i_k \in \mathcal{I}_k} \mathbf{v}_{i_k}^{q_k} s_{i_k}. \tag{1}$$

The corresponding received signal of user i_k is expressed as

$$\mathbf{y}_{i_k} = \sum_{l \in \mathcal{K}} \mathbf{H}_{i_k}^{l} \mathbf{x}^{l} + \mathbf{z}_{i_k}, \tag{2}$$

where $\mathbf{H}_{i_k}^{l} \triangleq \{\mathbf{H}_{i_k}^{q_l}\} \in \mathbb{C}^{N \times MQ_l}$ and $\mathbf{H}_{i_k}^{q_l} \in \mathbb{C}^{N \times M}$ denotes the channel matrix between the BS q_l to user i_k, $\mathbf{x}^k \in \mathbb{C}^{MQ_k \times 1}$ is the stacked transmitted signal $[(\mathbf{x}^{1_k})^H, \ldots, (\mathbf{x}^{Q_k})^H]^H$ of all BSs in the kth cell, and $\mathbf{z}_{i_k} \in \mathbb{C}^{N \times 1} \sim CN(0, \sigma_{i_k}^2)$ is the additive white Gaussina noise (AWGN) at user i_k. We assume that each user treats the interference as noise, and thus the achievable rate for user i_k is

$$R_{i_k}(\mathbf{v}) = \log\det\Bigg(\mathbf{I} + \mathbf{H}_{i_k}^{k} \mathbf{v}_{i_k} \mathbf{v}_{i_k}^{H} (\mathbf{H}_{i_k}^{k})^H \times \Big(\sum_{(l,j)\neq(k,i)} \mathbf{H}_{i_k}^{l} \mathbf{v}_{jl} \mathbf{v}_{jl}^{H} (\mathbf{H}_{i_k}^{l})^H + \sigma_{i_k}^2 \mathbf{I}\Big)^{-1}\Bigg).$$

Our objective is to appropriately design the precoders so that the system sum rate is maximized while using a small number of BSs.

3. PROPOSED BASE STATIONS SELECTION SCHEME WITH LASSO REGULARIZER

The linear precoder design problem for sum rate maximization can be stated as follows

$$\begin{aligned} &\max_{\mathbf{v}} \sum_{k \in \mathcal{K}} \sum_{i_k \in \mathcal{I}_k} R_{i_k}(\mathbf{v}) \\ &\text{s.t. } (\mathbf{v}^{q_k})^H \mathbf{v}^{q_k} \le P_{q_k},\ \forall q_k \in \mathcal{Q}_k,\ k \in \mathcal{K}, \end{aligned} \tag{3}$$

where $P_{q_k} > 0$ is the power budget for BS q_k. To reduce the number of active BSs, let us split the transmit beamformer $\mathbf{v}_{i_k}^{q_k}$ by $\mathbf{v}_{i_k}^{q_k} = \alpha_{q_k} \bar{\mathbf{v}}_{i_k}^{q_k}$, with $\alpha_{q_k} \in [-1, 1]$ representing whether BS q_k is switched on. That is, when $\alpha_{q_k} = 0$, BS q_k is switched off, otherwise, BS q_k is turned on. Define $\boldsymbol{\alpha}_k \triangleq [\alpha_{1_k}, \alpha_{2_k}, \ldots, \alpha_{Q_k}]$. The requirement that only a small number of BSs is active is equivalent to having a few nonzero elements in the vector $\boldsymbol{\alpha}_k$, a property that can be promoted by penalizing $\boldsymbol{\alpha}_k$ using an ℓ_1 regularization term. As a result we propose to solve the following regularized sum-rate maximization problem instead

$$\begin{aligned} &\max_{\boldsymbol{\alpha}, \bar{\mathbf{v}}} \sum_{k \in \mathcal{K}} \sum_{i_k \in \mathcal{I}_k} \left(R_{i_k}(\mathbf{v}) - \mu_k \|\boldsymbol{\alpha}_k\|_1\right) \\ &\text{s.t. } \alpha_{q_k}^2 (\bar{\mathbf{v}}^{q_k})^H \bar{\mathbf{v}}^{q_k} \le P_{q_k},\ \forall q_k \in \mathcal{Q}_k,\ k \in \mathcal{K}, \end{aligned} \tag{4}$$

where $\mu_k \ge 0,\ \forall k \in \mathcal{K}$, is the parameter to control how many BSs will be active in cell k, and $\boldsymbol{\alpha} \triangleq \{\boldsymbol{\alpha}_k\}_{k \in \mathcal{K}}$.

Remark 1 *Instead of splitting $\mathbf{v}_{i_k}^{q_k}$ and penalizing $\|\boldsymbol{\alpha}_k\|_1$, another natural modification is to add a group LASSO regularization term for each BS's beamformer directly, i.e., use the regularization term $\|\mathbf{v}^{q_k}\|$ for BS q_k in the objective function of problem* (3)*. However, when the power used by BS q_k is large, the magnitude of penalization term can dominate that of the system sum rate. Thus solving such group-LASSO penalized problem would effectively force the BSs to use only a small portion of its power budget, which could lead to a dramatic reduction of the system sum rate. Therefore, in our formulation, a different regularization is needed.*

However, the penalized sum rate maximization problem (4) is shown to be NP-hard even for single antenna BSs and users for $\mu_k = 0,\ \forall k \in \mathcal{K}$ [17]. In what follows, we will propose an efficient algorithm to compute a locally optimal solution.

3.1. Active BS Selection via Sparse WMMSE Algorithm

By using a similar argument as in [9, Proposition 1], we can show that the penalized sum rate maximization problem (4) is equivalent to the following penalized weighted mean square error (MSE) minimization problem

$$\min_{\boldsymbol{\alpha}, \bar{\mathbf{v}}, \mathbf{u}, \mathbf{w}} f(\mathbf{v}, \mathbf{w}, \mathbf{u}) + \sum_{k \in \mathcal{K}} \mu_k \|\boldsymbol{\alpha}_k\|_1 \tag{5a}$$

$$\text{s.t. } f(\mathbf{v}, \mathbf{w}, \mathbf{u}) = \sum_{i,k} w_{i_k} e_{i_k}(\mathbf{u}_{i_k}, \mathbf{v}) - \log(w_{ik}) \tag{5b}$$

$$\alpha_{q_k}^2 (\bar{\mathbf{v}}^{q_k})^H \bar{\mathbf{v}}^{q_k} \le P_{q_k},\ \forall q_k \in \mathcal{Q}_k,\ k \in \mathcal{K}. \tag{5c}$$

In the above expression, $\mathbf{u} \triangleq \{\mathbf{u}_{i_k}\}_{i_k \in \mathcal{I}}$ is the set of all receive beamformers of the users; $\mathbf{w} \triangleq \{w_{i_k}\}_{i_k \in \mathcal{I}}$ is the set of non-negative weights; e_{i_k} is the MSE for estimating s_{i_k}:

$$\begin{aligned} e_{i_k}(\mathbf{u}_{i_k}, \mathbf{v}) \triangleq\ & (1 - \mathbf{u}_{i_k}^H \mathbf{H}_{i_k}^{k} \mathbf{v}_{i_k})(1 - \mathbf{u}_{i_k}^H \mathbf{H}_{i_k}^{k} \mathbf{v}_{i_k})^H \\ & + \sum_{(\ell,j)\neq(k,i)} \mathbf{u}_{i_k}^H \mathbf{H}_{i_k}^{\ell} \mathbf{v}_{j\ell} \mathbf{v}_{j\ell}^H (\mathbf{H}_{i_k}^{\ell})^H \mathbf{u}_{i_k} + \sigma_{i_k}^2 \mathbf{u}_{i_k}^H \mathbf{u}_{i_k}. \end{aligned}$$

To make sure the proposed algorithm converges (which will be explained later in the proof of Theorem 1), below we will replace the power constraint (5c) by a more conservative condition, namely $(\bar{\mathbf{v}}^{q_k})^H \bar{\mathbf{v}}^{q_k} \le P_{q_k},\ \alpha_{q_k}^2 \le 1$. Therefore, the modified penalized weighted MSE minimization problem for active BS selection is

$$\boxed{\begin{aligned} &\min_{\boldsymbol{\alpha}, \bar{\mathbf{v}}, \mathbf{u}, \mathbf{w}} f(\mathbf{v}, \mathbf{w}, \mathbf{u}) + \sum_{k \in \mathcal{K}} \mu_k \|\boldsymbol{\alpha}_k\|_1 \\ &\quad \text{s.t. } f(\mathbf{v}, \mathbf{w}, \mathbf{u}) = \sum_{i,k} w_{i_k} e_{i_k}(\mathbf{u}_{i_k}, \mathbf{v}) - \log(w_{ik}) \\ &\qquad (\bar{\mathbf{v}}^{q_k})^H \bar{\mathbf{v}}^{q_k} \le P_{q_k}, \\ &\qquad \alpha_{q_k}^2 \le 1,\ \forall q_k \in \mathcal{Q}_k,\ \forall k \in \mathcal{K}. \end{aligned}} \tag{6}$$

Although the modified power constraint will shrink the original feasible set whenever $\alpha_{q_k}^2 \neq 0$ or ± 1, thus may reduce the sum rate performance of the obtained transceiver, our numerical experiments (to be shown in Section 4) suggest that satisfactory sum rate performance can still achieved.

Due to the fact that problem (6) is convex in each block variables, global minimum can be obtained for each block variable when fixing the rest. Furthermore, the problem is strongly convex for block $\mathbf{u}$ and $\mathbf{w}$, respectively, and the unique optimal solution $\mathbf{u}_{i_k}^\star$ and $w_{i_k}^\star,\ \forall i_k \in \mathcal{I}$, can be obtained in closed form:

$$\begin{aligned} \mathbf{u}_{i_k}^\star(\mathbf{v}) &= \left(\sum_{(j,l)} \mathbf{H}_{i_k}^{l} \mathbf{v}_{jl} \mathbf{v}_{jl}^H (\mathbf{H}_{i_k}^{l})^H + \sigma_{i_k}^2 \mathbf{I}\right)^{-1} \mathbf{H}_{i_k}^{k} \mathbf{v}_{i_k}, \\ &\triangleq \mathbf{J}_{i_k}^{-1}(\mathbf{v}) \mathbf{H}_{i_k}^{k} \mathbf{v}_{i_k} \end{aligned} \tag{7}$$

$$w_{i_k}^\star(\mathbf{v}) = \left(1 - \mathbf{v}_{i_k}^H \left(\mathbf{H}_{i_k}^{k}\right)^H \mathbf{J}_{i_k}^{-1}(\mathbf{v}) \mathbf{H}_{i_k}^{k} \mathbf{v}_{i_k}\right)^{-1}. \tag{8}$$

On the other hand, problem (6) can also be rewritten as

$$\min_{\boldsymbol{\alpha}, \bar{\mathbf{v}}, \mathbf{u}, \mathbf{w}} f(\mathbf{v}, \mathbf{w}, \mathbf{u}) + \sum_{k \in \mathcal{K}} \mu_k \|\alpha^k\|_1 + I_1(\bar{\mathbf{v}}) + I_2(\boldsymbol{\alpha}) \tag{9}$$

where $I_1(\bar{\mathbf{v}})$ and $I_2(\boldsymbol{\alpha})$ are indicator functions for both constraints defined respectively as

$$I_1(\bar{\mathbf{v}}) = \begin{cases} 0, & \text{if } (\bar{\mathbf{v}}^{q_k})^H \bar{\mathbf{v}}^{q_k} \le P_{q_k},\ \forall q_k \in \mathcal{Q}_k,\ k \in \mathcal{K}, \\ \infty, & \text{otherwise} \end{cases},$$

$$I_2(\boldsymbol{\alpha}) = \begin{cases} 0, & \text{if } \alpha_{q_k}^2 \le 1,\ \forall q_k \in \mathcal{Q}_k,\ k \in \mathcal{K}. \\ \infty, & \text{otherwise} \end{cases}$$

Observe that when the problem is written in the form of (9), all its nonsmooth parts are *separable* across block variables $\boldsymbol{\alpha}$, $\bar{\mathbf{v}}$, $\mathbf{u}$, and $\mathbf{w}$. Such separability is guaranteed by our modified power constraints, and is referred to as the "regularity condition" for nonsmooth optimization; see [18] for details about this condition. Combining this property with the fact that at most two blocks, namely $\boldsymbol{\alpha}$ and $\bar{\mathbf{v}}$, may not have unique minimizer, a block coordinate descent (BCD) procedure [1] is guaranteed to converge to the stationary point of problem (6). This is proven by Lemma 3.1 and Theorem 4.1 of [18]. The following theorem summarizes the preceding discussion.

Theorem 1 *A BCD procedure that iteratively optimizes problem* (6) *for each block variables* $\boldsymbol{\alpha}$, $\bar{\mathbf{v}}$, $\mathbf{u}$, *and* $\mathbf{w}$, *can always converge to a stationary solution of problem* (6).

In the following, we discuss in detail how problem (6) can be solved for each block variables in an efficiently manner. For blocks $\mathbf{u}$ and $\mathbf{w}$, optimal solutions are shown in (7) and (8), respectively. Notice that when fixing $(\mathbf{u}, \mathbf{w}, \bar{\mathbf{v}})$, the objective of problem (6) is separable among the cells. Therefore K independent subproblems can be solved simultaneously, with k-th subproblem assumes the following form

$$\begin{aligned}\min_{\boldsymbol{\alpha}_k} \ & (\boldsymbol{\alpha}_k)^T \mathbf{A}_k \boldsymbol{\alpha}_k - 2Re(\mathbf{b}_k^H \boldsymbol{\alpha}_k) + \mu_k ||\boldsymbol{\alpha}_k||_1 \\ \text{s.t. } & \alpha_{q_k}^2 \le 1, \ \forall q_k \in \mathcal{Q}_k \end{aligned} \tag{10}$$

where

$$\mathbf{A}_k \triangleq \sum_{i_k \in \mathcal{I}_k} \text{diag}(\bar{\mathbf{v}}_{i_k})^H \left(\sum_{(l,j)} w_{j_l} (\mathbf{H}_{j_l}^k)^H \mathbf{u}_{j_l} \mathbf{u}_{j_l}^H \mathbf{H}_{j_l}^k \right) \text{diag}(\bar{\mathbf{v}}_{i_k})$$

$$\mathbf{b}_k \triangleq \sum_{i_k \in \mathcal{I}_k} w_{i_k} \text{diag}(\bar{\mathbf{v}}_{i_k})^H (\mathbf{H}_{i_k}^k)^H \mathbf{u}_{i_k}.$$

Problem (10) is a quadratically constrained LASSO problem. It can be solved optimally by again applying a BCD procedure, with the block variables given by α_{q_k}, $\forall q_k \in \mathcal{Q}_k$ (e.g., [11]). For the q_k-th block, its optimal solution $\alpha_{q_k}^\star$ must satisfy the following first-order optimality condition

$$2(c_{q_k} - (\mathbf{A}_k[q,q] + \gamma_{q_k}^\star)\alpha_{q_k}^\star) \in \mu_k \partial |\alpha_{q_k}^\star|, \tag{11}$$

$$\gamma_{q_k}^\star \ge 0, \ (1 - (\alpha_{q_k}^\star)^2) \ge 0 \tag{12}$$

$$(1 - (\alpha_{q_k}^\star)^2)\gamma_{q_k}^\star = 0 \tag{13}$$

where $\gamma_{q_k}^\star$ is the optimal dual variable for the q_kth power constraint of problem (10), $\partial(\cdot)$ is the subgradient, and $c_{q_k} \triangleq Re(\mathbf{b}_k[q]) - \sum_{p \ne q} \mathbf{A}_k[p,q]\alpha_{p_k}$. Therefore, when $2\,|c_{q_k}| \le \mu_k$, we have $\alpha_{q_k}^\star = 0$. In the following, let us focus on the case where $2|c_{q_k}| > \mu_k$. In this case, from the expression of the subgradient (11), we have $\alpha_{q_k}^\star = \frac{-\mu_k \text{sign}(\alpha_{q_k}^\star) + 2c_{q_k}}{2(\mathbf{A}_k[q,q] + \gamma_{q_k}^\star)}$. Since $\gamma_{q_k}^\star \ge 0$, $\mathbf{A}_k[q,q] \ge 0$, and $2|c_{q_k}| > \mu_k$, we have $\text{sign}(\alpha_{q_k}^\star) = \text{sign}(c_{q_k})$. By plugging $\alpha_{q_k}^\star$ into the objective function of problem (10), it can be shown the objective value is an increasing function of $\gamma_{q_k}^\star$. Therefore, by the monotonicity of $\gamma_{q_k}^\star$, primal and dual constraints (12), and the complementarity condition (13), in the case of $2|c_{q_k}| > \mu_k$, $\alpha_{q_k}^\star$ has the following structure

$$\alpha_{q_k}^\star = \begin{cases} \frac{-\mu_k \text{sign}(c_{q_k}) + 2c_{q_k}}{2\mathbf{A}_k[q,q]}, & \text{if } \left| \frac{-\mu_k \text{sign}(c_{q_k}) + 2c_{q_k}}{2\mathbf{A}_k[q,q]} \right| < 1 \\ \text{sign}(c_{q_k}), & \text{otherwise} \end{cases} \tag{14}$$

[1] In our context, the BCD procedure refers to the computation strategy that cyclically updates the blocks $\boldsymbol{\alpha}$, $\bar{\mathbf{v}}$, $\mathbf{u}$, and $\mathbf{w}$ one at a time.

Similarly, when fixing $(\boldsymbol{\alpha}, \mathbf{w}, \mathbf{u})$, the solution for the beamformer $\bar{\mathbf{v}}$ can be obtained by iteratively updating its block components $\bar{\mathbf{v}}^{q_k}$, $\forall q_k \in \mathcal{Q}_k$. By checking the first order optimality condition, the q_k-th optimal block component, denoted as $\bar{\mathbf{v}}^{q_k\star}$, can be expressed as

$$\begin{aligned}\bar{\mathbf{v}}_{i_k}^{q_k\star}(\delta_{q_k}) = & \left(\mathbf{C}_k[q_k,q_k] + \delta_{q_k}^\star \mathbf{I}\right)^{-1} \\ & \times \left(\mathbf{D}_{i_k}[q_k] - \sum_{j_k \ne q_k} \mathbf{C}_k[q_k, j_k] \bar{\mathbf{v}}_{i_k}^{j_k\star} \right), \ \forall i_k \in \mathcal{I}_k. \end{aligned} \tag{15}$$

In the above expression, $\delta_{q_k}^\star \ge 0$ is the optimal dual variable for the q_k-th power constraint; $\mathbf{C}_k[q_k,q_k] \in \mathbb{C}^{M \times M}$ and $\mathbf{D}_{i_k}[q_k] \in \mathbb{C}^{M \times 1}$ are, respectively, subblocks of matrices $\mathbf{C}_k$ and $\mathbf{D}_{i_k}$ given below

$$\mathbf{C}_k \triangleq \sum_{i_k \in \mathcal{I}_k} \hat{\boldsymbol{\alpha}}_k \left(\sum_{(l,j)} w_{j_l} (\mathbf{H}_{j_l}^k)^H \mathbf{u}_{j_l} \mathbf{u}_{j_l}^H \mathbf{H}_{j_l}^k \right) \hat{\boldsymbol{\alpha}}_k \in \mathbb{C}^{Q_k M \times Q_k M},$$

$$\mathbf{D}_{i_k} \triangleq \sum_{i_k \in \mathcal{I}_k} w_{i_k} \hat{\boldsymbol{\alpha}}_k (\mathbf{H}_{i_k}^k)^H \mathbf{u}_{i_k} \in \mathbb{C}^{Q_k M \times 1}, \ \forall i_k \in \mathcal{I}_k,$$

$$\hat{\boldsymbol{\alpha}}_k \triangleq \text{diag}(\boldsymbol{\alpha}_{1_k}\mathbf{I}, \ldots, \boldsymbol{\alpha}_{Q_k}\mathbf{I}) \in \mathbb{C}^{Q_k M \times Q_k M}.$$

By the complementary condition, $\delta_{q_k}^\star = 0$ if $(\bar{\mathbf{v}}^{q_k\star}(0))^H \bar{\mathbf{v}}^{q_k\star}(0) \le P_{q_k}$. Otherwise, it should satisfy $(\bar{\mathbf{v}}^{q_k\star}(\delta_{q_k}^\star))^H \bar{\mathbf{v}}^{q_k\star}(\delta_{q_k}^\star) = P_{q_k}$. For the latter case, $\delta_{q_k}^\star$ can be found by a simple bisection method.

In summary, our main algorithm can be summarized in the following table.

Spare WMMSE (S-WMMSE) algorithm:

1: **Initialization** Generate a feasible set of variables $\{\bar{\mathbf{v}}_{i_k}\}$, $i_k \in \mathcal{I}$, and let $\alpha_{q_k} = 1 \ \forall q_k \in \mathcal{Q}_k, k \in \mathcal{K}$.
2: **Repeat**
3: $\quad \mathbf{u}_{i_k} \leftarrow \mathbf{J}_{i_k}^{-1}(\mathbf{v}) \mathbf{H}_{i_k}^k \mathbf{v}_{i_k}, \ \forall i_k \in \mathcal{I}$
4: $\quad w_{i_k} \leftarrow (1 - \mathbf{v}_{i_k}^H \left(\mathbf{H}_{i_k}^k\right)^H \mathbf{J}_{i_k}^{-1}(\mathbf{v}) \mathbf{H}_{i_k}^k \mathbf{v}_{i_k})^{-1}, \ \forall i_k \in \mathcal{I}$
5: $\quad \alpha_{q_k}$ is iteratively updated by

$$\alpha_{q_k} = \begin{cases} 0, & \text{if } 2|c_{q_k}| \le \mu_k \\ (14), & \text{otherwise} \end{cases}, \ \forall q_k \in \mathcal{Q}_k, \ k \in \mathcal{K}$$

6: $\quad \bar{\mathbf{v}}^{q_k}$ is iteratively updated by (15), $\forall q_k \in \mathcal{Q}_k, \forall k \in \mathcal{K}$
7: **Until** Desired stopping criteria is met

After (6) is solved, an additional step of postprocessing can further improve the system sum rate performance. That is, with the given set of active BSs computed by the S-WMMSE algorithm, we can solve problem (6) again, this time *without* the sparse promoting terms. This can be done by making the following changes to the proposed algorithm: *1)* letting $\mu_k = 0$ for each $k \in \mathcal{K}$; *2)* skipping step 5; *3)* setting $\alpha_{q_k} = \text{sign}(\alpha_{q_k}^\star)$, $\forall q_k$. See reference [11] for further justification of using such debiasing technique in solving regularized optimization porblems.

3.2. Joint active BS selection and BS clustering

In addition to controlling the number of active BSs, we can further optimize the size of BS clusters by adding an additional penalization on the beamformers. Specifically, since $\mathbf{v}_{i_k}^{q_k}$ being zero means user i_k is not served by BS q_k, user i_k is served with a small BS cluster means $\|\mathbf{v}_{i_k}^{q_k}\|$ is nonzero for only a few q_ks. Thus, a set of group LASSO regularization terms, $\sum_{q_k \in \mathcal{Q}_k} \left\|\mathbf{v}_{i_k}^{q_k}\right\|$,$i_k \in \mathcal{I}$, can be added to the objective function of problem (3) to reduce the size of BS clusters; see [9] for details. Hence, to jointly control the size of BS

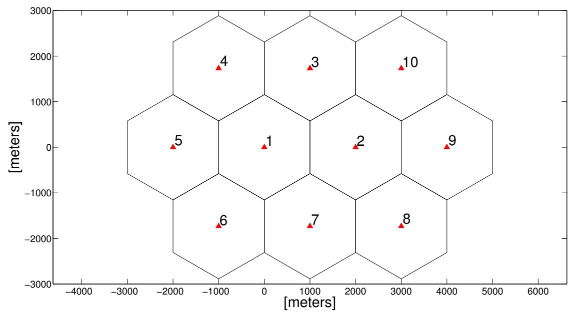

Fig. 1. Network configuration

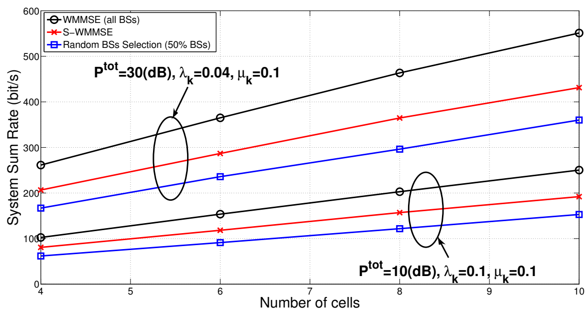

Fig. 2. The comparison on sum rate performance over different number of cells and total power budgets, P^{tot}, between proposed S-WMMSE algorithm, the performance upper bound, and a heuristic random selection.

cluster and reducing the BS usage, the objective function of the penalized weighted MMSE minimization problem (6) is now modified as

$$f(\mathbf{v},\mathbf{w},\mathbf{u})+\sum_{k\in\mathcal{K}}\left(\sum_{i_k\in\mathcal{I}_k}\lambda_k\sum_{q_k\in\mathcal{Q}_k}\|\bar{\mathbf{v}}_{i_k}^{q_k}\|\right)+\mu_k\|\alpha^k\|_1, \quad (16)$$

where $\lambda_k \geq 0, \forall k \in \mathcal{K}$, is the parameter to control the size of BS cluster in cell k. For this modified problem, again a BCD procedure with block variables, α, $\bar{\mathbf{v}}$, $\mathbf{u}$, and $\mathbf{w}$, can be used to compute a locally optimal solution. The only difference from the algorithm proposed in the previous section is the computation of $\bar{\mathbf{v}}$. This can be carried out by solving a quadratically constrained group LASSO problem. See in [9, Table I] for details.

4. SIMULATION RESULTS AND CONCLUSIONS

In the following numerical experiments, we consider a HetNet with at most 10 cells while the distance between centers of adjacent cells is 2000 meters, and the network configuration is depicted in Fig. 1. For each cell, there are 10 users and 10 BSs, one of which is located in the center of the cell while the other BSs and the users are uniformly distributed in that cell. The channel model we use is Rayleigh channel with zero mean and variance $(200/d_{i_k}^{q_l})^3 L_{i_k}^{q_l}$, where $d_{i_k}^{q_l}$ is the distance between BS q_l and user i_k, and $10\log 10(L_{i_k}^{q_l}) \sim N(0,64)$. Let P^{tot} denote the sum of the power budget in each cell. We assume that the BSs located in the center of the cells are macro BSs, which have a power budget $P^{tot}/2$, and the rest of the BSs have equal power budgets. We set the number of antennas for the BSs and the users to be $M = 4$ and $N = 2$, respectively. All the simulation results are averaged over 100 channel realizations. After solving problem (16), we perform debiasing step to further improve the sum rate performance.

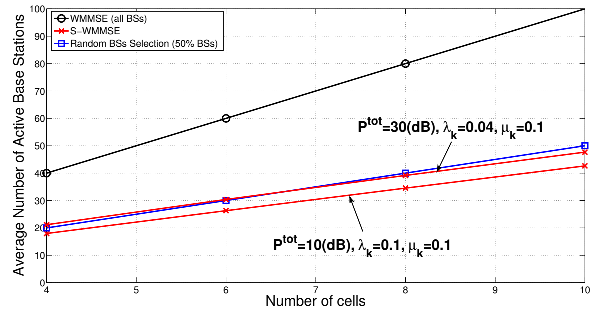

Fig. 3. The number of active BSs for each comparing scheme over different number of cells scenario.

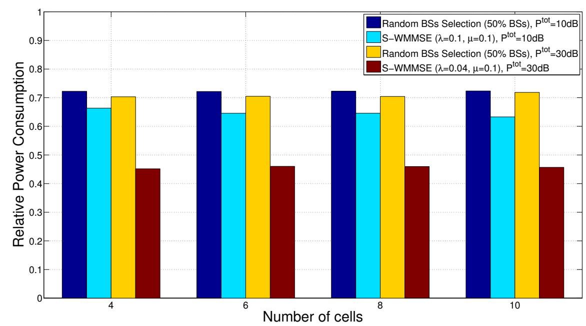

Fig. 4. Comparison of the power consumption for each comparing schemes with $P^{tot} = 10$dB or 30dB. The total power used for the all BSs being on case is normalized to 1.

The proposed algorithm is compared with the original WMMSE algorithm [16] applied to the following two scenarios: 1) all the BSs are turned on; 2) in each cell, the central BS and a randomly selected fixed number of the remaining BSs are turned on. Note that for both of these cases, full JP is used within each cell. Clearly, the first scenario can serve as the performance upper bound, and the latter can serve as a reasonable heuristic algorithm to select active BSs since BSs and users are uniformly distributed in each cell. In Fig. 2, the system sum rate performance for the proposed S-WMMSE algorithm is compared with the two scenarios for $P^{tot} = 10$dB and 30dB, respectively. To ensure a fair comparison, we choose λ_k and μ_k such that the number of active BSs for S-WMMSE is about the same as the random selection one (see Fig. 3). We can observe that S-WMMSE can achieve about 80% of the sum rate compared to the upper bound while activating around 50% BSs. Furthermore, the S-WMMSE can still achieve more than 20% and 25% improvement in sum rate performance for $P^{tot} = 10$dB and 30dB, respectively, compared to random selection. It is worth noting that compared with the random selection, the proposed algorithm can reduce the power consumption; see Fig. 4. It is because when we choose $\lambda_k > 0$, $\forall k \in \mathcal{K}$, S-WMMSE algorithm also dynamically optimizes the BS clustering. This can reduce the coverage of each BS, thus requiring less power consumption due to the decreased interference.

In summary, from the simulation results, the proposed S-WMMSE algorithm can effectively reduce the BS usage and the size of BS cluster simultaneously. Some interesting future directions of this work are under investigation, e.g., how to adaptively choose the sparse parameters, i.e., μ_k and λ_k [12, 15], and how to devise the S-WMMSE algorithm when only long-term channel statistics are available.

5. REFERENCES

[1] A. Damnjanovic, J. Montojo, Y. Wei, T. Ji, T. Luo, M. Vajapeyam, T. Yoo, O. Song, and D. Malladi, "A survey on 3GPP heterogeneous networks," *IEEE Trans. Wireless Communications*, vol. 18, no. 3, pp. 10–21, Jun. 2011.

[2] B. Clerckx, A. Lozano, S. Sesia, C. van Rensburg, and C. B. Papadias, "3GPP LTE and LTE Advanced," *EURASIP Journal on Wireless Communications and Networking*, 2009.

[3] D. Gesbert, S. Hanly, H. Huang, S. Shamai (Shitz), O. Simeone, and W. Yu, "Multi-cell MIMO cooperative networks: A new look at interference," *IEEE J. Sel. Areas in Communications*, vol. 28, no. 9, pp. 1380–1408, Dec. 2010.

[4] J. Zhang, R. Chen, J.G. Andrews, A. Ghosh, and R.W. Heath, "Networked mimo with clustered linear precoding," *IEEE Trans. Wireless Communications*, vol. 4, no. 8, pp. 1910–921, Jun. 2009.

[5] A. Papadogiannis, D. Gesbert, and E. Hardouin, "A dynamic clustering approach in wireless networks with multi-cell cooperative processing," in *Proc. of IEEE International Conference on Communications*, May 2008, pp. 4033–4037.

[6] A. Papadogiannis and G.C. Alexandropoulos, "The value of dynamic clustering of base stations for future wireless networks," in *IEEE International Conference on Fuzzy Systems (FUZZ)*, Jul. 2010, pp. 1–6.

[7] Y. Zeng, E. Gunawan, Y.-L. Guan, and J. Liu, "Joint base station selection and linear precoding for cellular networks with multi-cell processing," in *TENCON*, Nov. 2010, pp. 1976–1981.

[8] S.-J. Kim, S. Jain, and G. B. Giannakis, "Backhaul-constrained multi-cell cooperation using compressive sensing and spectral clustering," in *Proc. of IEEE 13th International Workshop on Signal Processing Advances for Wireless Communications*, Cesme, Jun. 17-20 2012, pp. 65–69.

[9] M. Hong, R. Sun, and Z.-Q. Luo, "Joint base station clustering and beamformer design for partial coordinated transmission in heterogenous networks," *to appear in IEEE Journal on Selected Areas in Communications*, 2012.

[10] O. Arnold, F. Richter, G. Fettweis, and O. Blume, "Power consumption modeling of different base station types in heterogeneous cellular networks," in *Proc. of Future Network and Mobile Summit*, Jun. 2010, pp. 1–8.

[11] S. J. Wright, R. D. Nowak, and M. A. T. Figueiredo, "Sparse reconstruction by separable approximation," *IEEE Trans. Signal Process.*, vol. 57, no. 7, pp. 2479–2493, Jul. 2009.

[12] E. Candes, M.Wakin, and S. Boyd, "Enhancing sparsity by reweighted l_1 minimization," *J. Fourier Analysis and Applications*, vol. 14, no. 5, pp. 877–905, Dec. 2008.

[13] E.J. Candes and M.B. Wakin, "An introduction to compressive sampling," *IEEE Trans. Signal Process.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.

[14] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *Journal of the Royal Statistical Society: Series B*, vol. 68, no. 1, pp. 49–67, Feb. 2006.

[15] O. Mehanna, N. D. Sidiropoulos, and G. B. Giannakis, "Multicast beamforming with antenna selection," in *Proc. of IEEE 13th International Workshop on Signal Processing Advances for Wireless Communications*, Cesme, Jun. 17-20 2012, pp. 70–74.

[16] Q. Shi, M. Razaviyayn, Z.-Q. Luo, and C. He, "An iteratively weighted mmse approach to distributed sum-utility maximization for a MIMO interfering broadcast channel," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4331–4340, Sep. 2011.

[17] Z.-Q. Luo and S. Zhang, "Dynamic spectrum management: Complexity and duality," *IEEE J. Sel. Topics Signal Process.*, vol. 2, no. 1, pp. 5773, Feb. 2008.

[18] P. Tseng, "Convergence of a block coordinate descent method for nondifferentiable minimization," *Journal of Optimization Theory and Applications*, vol. 109, no. 3, pp. 475–494, Jun. 2001.