

An Online Algorithm for Joint Long-Term Base Station Activation and Beamforming in Green Downlink MISO Networks

Jingran Lin, Qiang Li, and Hongwei Deng

School of Communication and Information Engineering

University of Electronic Science and Technology of China (UESTC), Chengdu China, 611731

Abstract—In modern wireless networks, optimizing the association between base stations (BSs) and users effectively improves network performance. On the other hand, a frequently changing BS-user association renders considerable operational burden for network management, e.g., it consumes extra power to awaken the deactivated BSs and to support users' switching among BSs. This motivates us to balance the flexibility and stability of BS-user association, leading to a long-term BS association problem. In this paper, we address this by considering the green communication problem in a MISO network where multiple BSs cooperatively forward information to multiple users. We minimize total power consumption (including not only the transmit power and device maintenance power, but also the switching power consumed as the BS-user association changes) by joint long-term BS activation and beamforming, i.e., we jointly optimize the active BSs, the transmit beamformer and the user switching frequency among BSs. Due to the inherent causality constraints on channel state information (CSI), we develop an online algorithm for the optimization problem, utilizing the sample average approximation (SAA) method. To improve the algorithm's efficiency, we further fit the problem into the framework of alternating direction method of multipliers (ADMM), and finally design a low-complexity distributed online algorithm for the green communication problem.

I. INTRODUCTION

As one key issue in green wireless communications, power minimization has drawn extensive attention recently [1], [2]. In addition to transmit power, the maintenance power consumed to keep the devices in good conditions becomes non-negligible in modern wireless networks, where a large number of base stations (BSs) are densely deployed for coverage enhancement [3]. Therefore, one popular strategy for green communication is joint BS activation and beamforming [4], which effectively balances the maintenance power and the transmit power.

On the other hand, most existing joint BS activation and beamforming algorithms optimize the active BSs according to *instantaneous* channel state information (CSI) [4]–[9]. Considering the time-varying characteristics of wireless channels, the BS's active status changes quickly. Consequently, BSs are activated and deactivated frequently and users switch among different BSs continually. However, this renders considerable operational burdens in practice. For instance, frequently shutting down and restarting BSs consumes more power than that

in normal maintenance. Some necessary switching (handover) operations need to be implemented as users switch to another BS, thus introducing extra power consumption [3]. Moreover, changing the BS-user association frequently may even cause service break.

This motivates us to consider the stability of BS-user association in joint BS activation and beamforming optimization, referred to as joint *long-term* BS activation and beamforming. Unfortunately, few works have emphasized the stability of BS-user association. In [10], a switching decision strategy considering transmission latency and handover signaling overhead is proposed utilizing Markov decision processes. A long-term transmit point (TP) association method is developed in [11] to maximize system throughput, where the stability of TP-user association is guaranteed by prohibiting users from switching among TPs in a relatively long time period.

In this paper, we consider the green communication problem in a downlink MISO network that consists of multiple BSs and multiple users. Under some quality-of-service (QoS) constraints, we minimize the total network power consumption, including not only the maintenance power and the transmit power, but also the switching power consumed to awaken the deactivated BSs and to support the users' switching among different BSs. To this end, we jointly optimize the active BSs, the BS transmit beamformer and the user switching frequency. We should mention that instead of forcing a static BS-user association in some given time period as in [11], we formulate the long-term BS activation problem from the perspective of sparse optimization, i.e., we require the variation of BS-user association to be a sparse vector.

Considering the inherent causality constraints on CSI [12], we develop an online algorithm to successively solve the joint BS activation and beamforming problems in different timeslots. Specifically, the active BSs in each timeslot are determined jointly with the consideration of the active BSs in the previous timeslot and the *expected* active BSs in the next timeslot. To avoid directly handling a complicated stochastic optimization problem, we utilize the sample average approximation (SAA) method to approximate the expectation by a sample average [13]. Moreover, to solve the approximated problem efficiently, we further fit the problem into the alternating direction method of multipliers (ADMM) framework [14], [15]. Finally, a low-

This work was supported by the National NSFC [61671120, 61401073, and 61471103], and by the Fundamental Research Funds for Central Universities [ZYGX2016J007, ZYGX2016J011].

complexity distributed online algorithm is developed.

II. SIGNAL MODEL AND PROBLEM STATEMENT

Consider the downlink transmission in a network where K N -antenna BSs forward information to M single-antenna users cooperatively. We employ block fading channels here, namely, the channels remain static in each fading block, while changing from one block to another according to the channel distribution [16]. We define each fading block as a timeslot, and consider the transmission in a time period of L timeslots.

Next, define $\mathbf{w}_m(l)$ and $\mathbf{h}_m(l) \in \mathbb{C}^{KN \times 1}$ as the multi-BS cooperative transmit beamformer and the associated channel for user m in timeslot l , where $\mathbf{w}_m^k(l)$ and $\mathbf{h}_m^k(l) \in \mathbb{C}^{N \times 1}$ are the k th row blocks of $\mathbf{w}_m(l)$ and $\mathbf{h}_m(l)$, respectively, denoting the beamformer of BS k to user m and the associated channel for $k = 1, \dots, K$, $m = 1, \dots, M$, $l = 1, \dots, L$. P is the power budget of the BSs, and σ^2 is the noise power. Define $a^k(l) \in \{0, 1\}$ as the active indicator of BS k ; that is, BS k is active in timeslot l if $a^k(l) = 1$, and inactive if $a^k(l) = 0$. Let $\mathbf{W}(l)$, $\mathbf{H}(l)$ and $\mathbf{a}(l)$ denote the collections of the cooperative beamformers, the associated channels and the BS active indicators in timeslot l , respectively, i.e., $\mathbf{W}(l) = [\mathbf{w}_1(l), \mathbf{w}_2(l), \dots, \mathbf{w}_M(l)]$, $\mathbf{H}(l) = [\mathbf{h}_1(l), \mathbf{h}_2(l), \dots, \mathbf{h}_M(l)]$, and $\mathbf{a}(l) = [a^1(l), a^2(l), \dots, a^K(l)]^T$.

Taking the maintenance power, transmit power and switching power into account, we calculate the total power consumption in the length- L time period as

$$p = \underbrace{\sum_{l=1}^L \|\mathbf{W}(l)\|_F^2}_{\text{transmit power}} + \underbrace{\lambda_1 \sum_{l=1}^L \|\mathbf{a}(l)\|_0}_{\text{maintenance power}} + \underbrace{\lambda_2 \sum_{l=1}^{L-1} \|\mathbf{a}(l+1) - \mathbf{a}(l)\|_0}_{\text{switching power}} \quad (1)$$

where $\|\cdot\|_F$ is the Frobenius norm; $\lambda_1 > 0$ can be interpreted as the maintenance power to keep one BS active, and $\lambda_2 > 0$ as the average switching power consumed when the BS-user association changes.

Thus, the power minimization problem based on joint long-term BS activation and beamforming is expressed as

$$\min_{\{\mathbf{W}(l), \mathbf{a}(l)\}_{l=1}^L} p \quad (P1)$$

$$\text{s.t. SINR}_m(l) \triangleq \frac{|\mathbf{h}_m^\dagger(l) \mathbf{w}_m(l)|^2}{\sigma^2 + \sum_{n \neq m} |\mathbf{h}_m^\dagger(l) \mathbf{w}_n(l)|^2} \geq \gamma, \quad \forall m, l \quad (2)$$

$$\|\mathbf{W}^k(l)\|_F^2 \leq a^k(l)P, \quad a^k(l) \in \{0, 1\}, \quad \forall k, l \quad (3)$$

where (2) is the QoS constraint based on SINR; $(\cdot)^\dagger$ denotes Hermitian transpose; $\mathbf{W}^k(l) = [\mathbf{w}_1^k(l), \mathbf{w}_2^k(l), \dots, \mathbf{w}_M^k(l)] \in \mathbb{C}^{N \times M}$ denotes the k th row block of $\mathbf{W}(l)$, containing all the transmit beamformers of BS k at timeslot l .

III. OFFLINE OPTIMIZATION

Clearly, all the CSIs, i.e., $\mathbf{H}(l)$ for $l = 1, \dots, L$, are assumed to be available a priori in (P1). It means that (P1) is an offline

optimization problem. Considering that (P1) is a challenging nonlinear mixed-integer program, we try to find some efficient approximate solution to it.

To this end, we first approximate l_0 -norm by l_1 -norm and relax $a^k(l)$ as a continuous variable $0 \leq a^k(l) \leq 1$. Next, we perform the SOCP reformulation [17] to the QoS constraint (2), and then (P1) can be reformulated as

$$\min_{\{\mathbf{W}(l), \mathbf{a}(l)\}_{l=1}^L} \tilde{p} \quad (P2)$$

$$\text{s.t.} \begin{cases} \sqrt{1 + \gamma^{-1}} \mathbf{h}_m^\dagger(l) \mathbf{w}_m(l) \geq \|[\mathbf{h}_m^\dagger(l) \mathbf{W}(l), \sigma]\|_2, \\ \text{Im}\{\mathbf{h}_m^\dagger(l) \mathbf{w}_m(l)\} = 0, \quad \forall m, l \end{cases} \quad (4)$$

$$\|\mathbf{W}^k(l)\|_F^2 \leq a^k(l)P, \quad 0 \leq a^k(l) \leq 1, \quad \forall k, l \quad (5)$$

where \tilde{p} is defined as

$$\tilde{p} = \sum_{l=1}^L (\|\mathbf{W}(l)\|_F^2 + \lambda_1 \|\mathbf{a}(l)\|_1) + \lambda_2 \sum_{l=1}^{L-1} \|\mathbf{a}(l+1) - \mathbf{a}(l)\|_1 \quad (6)$$

Since (P2) is convex, it can be directly solved utilizing the standard soft-solver, e.g., CVX [18].

IV. ONLINE OPTIMIZATION

Considering the causality constraint on CSI, we design some online algorithm for (P2). In particular, we successively solve the joint BS activation and beamforming problems in different timeslots, using the current CSI and the statistics of future CSI.

A. Online Optimization Framework

From (6), $\mathbf{a}(l)$ is coupled with $\mathbf{a}(l-1)$ and $\mathbf{a}(l+1)$ in the switching power terms. Hence in timeslot l , we optimize $\mathbf{a}(l)$ jointly with the consideration of $\mathbf{a}(l-1)$ and $\mathbf{a}(l+1)$. However, $\mathbf{a}(l+1)$ is unknown in timeslot l , and should be determined based on the CSI $\mathbf{H}(l+1)$. Since only the statistics of $\mathbf{H}(l+1)$ are available, we integrate the expected power consumption in timeslot $(l+1)$ into the problem in timeslot l , and define the objective p_l as

$$p_l = \|\mathbf{W}(l)\|_F^2 + \lambda_1 \|\mathbf{a}(l)\|_1 + \lambda_2 \|\mathbf{a}(l) - \mathbf{a}(l-1)\|_1 + \mathbb{E} \left\{ \|\mathbf{W}(l+1)\|_F^2 + \lambda_1 \|\mathbf{a}(l+1)\|_1 + \lambda_2 \|\mathbf{a}(l+1) - \mathbf{a}(l)\|_1 \right\} \quad (7)$$

where $\mathbb{E}\{\cdot\}$ denotes expectation; $\{\mathbf{W}(l), \mathbf{a}(l)\}$ depends on the accurate value of $\mathbf{H}(l)$, while $\{\mathbf{W}(l+1), \mathbf{a}(l+1)\}$ depends on the statistics of $\mathbf{H}(l+1)$.

This implies we need to solve a series of stochastic problems in the online optimization. To avoid handling this challenging problem directly, we employ the SAA method [13] to approximate (7) as an average of S samples, i.e.,

$$\hat{p}_l = \|\mathbf{W}(l)\|_F^2 + \lambda_1 \|\mathbf{a}(l)\|_1 + \lambda_2 \|\mathbf{a}(l) - \mathbf{a}(l-1)\|_1 + \frac{1}{S} \sum_{s=1}^S \left\{ \|\hat{\mathbf{W}}(l+1, s)\|_F^2 + \lambda_1 \|\hat{\mathbf{a}}(l+1, s)\|_1 + \lambda_2 \|\hat{\mathbf{a}}(l+1, s) - \mathbf{a}(l)\|_1 \right\} \quad (8)$$

where $\hat{\mathbf{W}}(l+1, s)$ and $\hat{\mathbf{a}}(l+1, s)$ are the transmit beamformers and BS active indicators in timeslot $(l+1)$, based on the s th sample of $\mathbf{H}(l+1)$, denoted by $\hat{\mathbf{H}}(l+1, s)$.

The stochastic problem in timeslot l can then be transformed to a deterministic convex problem,

$$\begin{aligned} \min_{\substack{\{\mathbf{W}(l), \mathbf{a}(l)\} \\ \{\hat{\mathbf{W}}(l+1, s), \hat{\mathbf{a}}(l+1, s)\}_{s=1}^S}} \quad & \hat{p}_l \quad (\text{P3}) \\ \text{s.t.} \quad & (4) \text{ and } (5) \text{ satisfied for } l, \\ & \begin{cases} \sqrt{1 + \gamma^{-1}} \hat{\mathbf{h}}_m^\dagger(l+1, s) \hat{\mathbf{w}}_m(l+1, s) \\ \geq \|[\hat{\mathbf{h}}_m^\dagger(l+1, s) \hat{\mathbf{W}}(l+1, s), \sigma]\|_2, \\ \text{Im}\{\hat{\mathbf{h}}_m^\dagger(l+1, s) \hat{\mathbf{w}}_m(l+1, s)\} = 0, \quad \forall m, s \\ \|\hat{\mathbf{W}}^k(l+1, s)\|_F^2 \leq \hat{a}^k(l+1, s)P, \quad \forall k, s \\ 0 \leq \hat{a}^k(l+1, s) \leq 1, \quad \forall k, s \end{cases} \end{aligned}$$

To summarize, we optimize the active BSs and the transmit beamformers timeslot by timeslot in the online optimization framework, i.e., we solve (P3) for $l = 1, 2, \dots, L$, successively. In timeslot l , we generate S samples of $\mathbf{H}(l+1)$, namely, $\{\hat{\mathbf{H}}(l+1, s)\}_{s=1}^S$, according to channel distribution, and then obtain $\{\mathbf{W}(l), \mathbf{a}(l)\}$ by solving (P3) based on the current CSI $\mathbf{H}(l)$, the previous active BSs $\mathbf{a}(l-1)$ and the estimated CSI samples $\{\hat{\mathbf{H}}(l+1, s)\}_{s=1}^S$.

However, the dimension and complexity are big issues in solving (P3), since the sample size S cannot be too small in order to approximate the expectation accurately. Then, to solve (P3) efficiently, we further fit (P3) into the ADMM framework to develop a low-complexity distributed algorithm.

B. ADMM-Based Algorithm for (P3)

To this end, we first define $\check{\mathbf{W}}(t)$, $\check{\mathbf{H}}(t)$, $\check{\mathbf{a}}(t)$, $\check{\lambda}_0(t)$, $\check{\lambda}_1(t)$, and $\check{\lambda}_2(t)$, $t = 1, 2, \dots, S+1$, to simplify the notations, where $\check{\mathbf{W}}(1) = \mathbf{W}(l)$, $\check{\mathbf{H}}(1) = \mathbf{H}(l)$, $\check{\mathbf{a}}(1) = \mathbf{a}(l)$, $\check{\lambda}_0(1) = 1$, $\check{\lambda}_1(1) = \lambda_1$; $\check{\mathbf{W}}(t) = \hat{\mathbf{W}}(l+1, t-1)$, $\check{\mathbf{H}}(t) = \hat{\mathbf{H}}(l+1, t-1)$, $\check{\mathbf{a}}(t) = \hat{\mathbf{a}}(l+1, t-1)$, $\check{\lambda}_0(t) = \frac{1}{S}$, $\check{\lambda}_1(t) = \frac{\lambda_1}{S}$ for $t = 2, \dots, S+1$; $\check{\mathbf{a}}(S+2) = \mathbf{a}(l-1)$; $\check{\lambda}_2(t) = \frac{\lambda_2}{S}$ for $t = 1, 2, \dots, S$, and $\check{\lambda}_2(S+1) = 1$. Thus, (P3) can be equivalently rewritten in a uniform form as

$$\begin{aligned} \min_{\{\check{\mathbf{W}}(t), \check{\mathbf{a}}(t)\}_{t=1}^{S+1}} \quad & \hat{p}_l \quad (\text{P4}) \\ \text{s.t.} \quad & \hat{p}_l = \sum_{t=1}^{S+1} \left\{ \check{\lambda}_0(t) \|\check{\mathbf{W}}(t)\|_F^2 + \check{\lambda}_1(t) \|\check{\mathbf{a}}(t)\|_1 \right. \\ & \left. + \check{\lambda}_2(t) \|\check{\mathbf{a}}(t+1) - \check{\mathbf{a}}(1)\|_1 \right\} \\ & \begin{cases} \sqrt{1 + \gamma^{-1}} \check{\mathbf{h}}_m^\dagger(t) \check{\mathbf{w}}_m(t) \geq \|[\check{\mathbf{h}}_m^\dagger(t) \check{\mathbf{W}}(t), \sigma]\|_2, \\ \text{Im}\{\check{\mathbf{h}}_m^\dagger(t) \check{\mathbf{w}}_m(t)\} = 0, \quad \forall m, t \\ \|\check{\mathbf{W}}^k(t)\|_F^2 \leq \check{a}^k(t)P, \quad 0 \leq \check{a}^k(t) \leq 1, \quad \forall k, t \end{cases} \end{aligned}$$

Next, we decouple the variables in (P4) by introducing

$$\mathbf{U}(t) = \check{\mathbf{W}}(t), \quad \mathbf{F}(t) = [\check{\mathbf{H}}^\dagger(t) \mathbf{U}(t), \sigma \mathbf{1}] \in \mathbb{C}^{M \times (M+1)} \quad (9)$$

$$\mathbf{y} = \mathbf{Q} \mathbf{x} \in \mathbb{R}^{(S+1)K \times 1}, \quad \mathbf{x} = [\mathbf{x}_a^T, \mathbf{a}^T(l-1)]^T \quad (10)$$

$$\mathbf{x}_a = \check{\mathbf{a}} \triangleq [\check{\mathbf{a}}^T(1), \dots, \check{\mathbf{a}}^T(S+1)]^T \in \mathbb{R}^{(S+1)K \times 1} \quad (11)$$

$$\mathbf{Q} = \begin{bmatrix} -\mathbf{I} & \mathbf{I} & \mathbf{0} & \cdots & \mathbf{0} \\ -\mathbf{I} & \mathbf{0} & \mathbf{I} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\mathbf{I} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I} \end{bmatrix} \in \mathbb{R}^{(S+1)K \times (S+2)K} \quad (12)$$

where $\mathbf{1}$ is the $M \times 1$ all-one vector; \mathbf{I} and $\mathbf{0}$ are the $K \times K$ identity and all-zero matrices, respectively. Hence, (P4) is equivalent to

$$\min_{\{\check{\mathbf{W}}, \check{\mathbf{a}}, \mathbf{F}\}} \sum_{t=1}^{S+1} \sum_{k=1}^K \left\{ \check{\lambda}_0(t) \|\check{\mathbf{W}}^k(t)\|_F^2 + \check{\lambda}_1(t) \check{a}^k(t) \right. \\ \left. + \check{\lambda}_2(t) |y^k(t)| \right\} \quad (\text{P5})$$

$$\text{s.t.} \quad \sqrt{\gamma^{-1}} f_m^m(t) \geq \|\mathbf{f}_{-m}^m(t)\|_2, \text{Im}\{f_m^m(t)\} = 0, \quad \forall m, t \quad (13)$$

$$\|\check{\mathbf{W}}^k(t)\|_F^2 \leq \check{a}^k(t)P, \quad 0 \leq \check{a}^k(t) \leq 1, \quad \forall k, t \quad (14)$$

(9) ~ (12) satisfied.

where $\check{a}^k(t)$ is the k th element of the $K \times 1$ vector $\check{\mathbf{a}}(t)$, which is the t th subvector of $\check{\mathbf{a}}$ for $t = 1, 2, \dots, (S+1)$, see (11); $y^k(t)$ is defined similarly. (13) is actually the QoS constraint, where $\mathbf{f}^m(t) = [\check{\mathbf{h}}_m^\dagger(t) \mathbf{U}(t), \sigma] \in \mathbb{C}^{1 \times (M+1)}$ is the m th row vector of $\mathbf{F}(t)$; $f_m^m(t) = \check{\mathbf{h}}_m^\dagger(t) \mathbf{u}_m(t)$ is the m th element of $\mathbf{f}^m(t)$ with $\mathbf{u}_m(t)$ being the m th column vector of $\mathbf{U}(t)$; $\mathbf{f}_{-m}^m(t)$ is the remaining subvector after removing $f_m^m(t)$, i.e., $\mathbf{f}_{-m}^m(t) = [f_1^m(t), \dots, f_{m-1}^m(t), f_{m+1}^m(t), \dots, f_M^m(t)] \in \mathbb{C}^{1 \times M}$.

The partial augmented Lagrangian function of (P5) is given by (15), where $c > 0$ is the penalty factor, and α, β, Φ and Ψ are the associated Lagrangian multipliers. Dividing the variables $\check{\mathbf{W}}, \check{\mathbf{a}}, \mathbf{U}, \mathbf{F}, \mathbf{x}_a$ and \mathbf{y} into two blocks of $\{\check{\mathbf{W}}, \check{\mathbf{a}}, \mathbf{F}, \mathbf{y}\}$ and $\{\mathbf{U}, \mathbf{x}_a\}$, we can employ the ADMM framework to solve (P5) iteratively, with global convergence guarantee [14]. The online ADMM algorithm is summarized in Table 1.

In the ADMM online algorithm, (P5) is solved by updating $\{\check{\mathbf{W}}, \check{\mathbf{a}}, \mathbf{F}, \mathbf{y}\}$ and $\{\mathbf{U}, \mathbf{x}_a\}$ alternatively in each timeslot. More interestingly, these two subproblems can be divided into smaller problems further. Specifically, the problem of $\{\check{\mathbf{W}}, \check{\mathbf{a}}, \mathbf{F}, \mathbf{y}\}$ can be divided into $(S+1)K$ problems of $\{\check{\mathbf{W}}^k(t), \check{a}^k(t)\}$, $(S+1)M$ problems of $\{\mathbf{f}^m(t)\}$, and $(S+1)K$ problems of $\{y^k(t)\}$; the problem of $\{\mathbf{U}, \mathbf{x}_a\}$ can be divided into $(S+1)M$ problems of $\{\mathbf{u}_m(t)\}$ and one problem of $\{\mathbf{x}_a\}$, $k = 1, \dots, K$, $t = 1, \dots, (S+1)$, and $m = 1, \dots, M$.

Next, we show the step by step computation of the ADMM algorithm, with each step being solved in closed form.

$$\begin{aligned} L_c(\check{\mathbf{W}}, \check{\mathbf{a}}, \mathbf{U}, \mathbf{F}, \mathbf{x}_a, \mathbf{y}, \alpha, \beta, \Phi, \Psi) = & \sum_{t=1}^{S+1} \sum_{k=1}^K \left[\check{\lambda}_0(t) \|\check{\mathbf{W}}^k(t)\|_F^2 + \check{\lambda}_1(t) \check{a}^k(t) + \check{\lambda}_2(t) |y^k(t)| \right] + \alpha^T (\mathbf{y} - \mathbf{Q} \mathbf{x}) \\ & + \frac{c}{2} \|\mathbf{y} - \mathbf{Q} \mathbf{x}\|_2^2 + \beta^T (\mathbf{x}_a - \check{\mathbf{a}}) + \frac{c}{2} \|\mathbf{x}_a - \check{\mathbf{a}}\|_2^2 + \sum_{t=1}^{S+1} \left\{ \text{Re} \left[\text{Tr} \left(\Phi^\dagger(t) [\mathbf{U}(t) - \check{\mathbf{W}}(t)] \right) \right] + \frac{c}{2} \|\mathbf{U}(t) - \check{\mathbf{W}}(t)\|_F^2 \right\} \\ & + \sum_{t=1}^{S+1} \left\{ \text{Re} \left[\text{Tr} \left(\Psi^\dagger(t) (\mathbf{F}(t) - [\check{\mathbf{H}}^\dagger(t) \mathbf{U}(t), \sigma \mathbf{1}]) \right) \right] + \frac{c}{2} \|\mathbf{F}(t) - [\check{\mathbf{H}}^\dagger(t) \mathbf{U}(t), \sigma \mathbf{1}]\|_F^2 \right\} \quad (15) \end{aligned}$$

Table 1: Online algorithm based on ADMM

(1)	For $l = 1, 2, \dots, L$, do
(2)	Generate S CSI samples according to channel distribution;
(3)	Solve (P5) utilizing ADMM: Repeat
(4)	Update $\{\check{\mathbf{W}}, \check{\mathbf{a}}, \mathbf{F}, \mathbf{y}\}$ with other variables fixed;
(5)	$\{\check{\mathbf{W}}, \check{\mathbf{a}}, \mathbf{F}, \mathbf{y}\} \leftarrow \operatorname{argmin}_{\{\check{\mathbf{W}}, \check{\mathbf{a}}, \mathbf{F}, \mathbf{y}\}} L_c(\cdot)$
(6)	s.t. (14)
(7)	Update $\{\mathbf{U}, \mathbf{x}_a\}$ with other variables fixed;
(8)	$\{\mathbf{U}, \mathbf{x}_a\} \leftarrow \operatorname{argmin}_{\{\mathbf{U}, \mathbf{x}_a\}} L_c(\cdot)$
(9)	s.t. (13)
(10)	Update Lagrangian variables α, β, Φ and Ψ ;
(11)	Until some stopping criterion is satisfied;
(12)	Output $\{\mathbf{W}(l), \mathbf{a}(l)\}$;
(13)	End

1) *Updating* $\{\check{\mathbf{W}}, \check{\mathbf{a}}, \mathbf{F}, \mathbf{y}\}$: The problem of $\{\check{\mathbf{W}}, \check{\mathbf{a}}, \mathbf{F}, \mathbf{y}\}$ is completely separable among $\{\check{\mathbf{W}}^k(t), \check{a}^k(t)\}$, $\{\mathbf{f}^m(t)\}$ and $\{y^k(t)\}$ for $\forall k, t, m$. We can update them in parallel.

• The problem of $\{\check{\mathbf{W}}^k(t), \check{a}^k(t)\}$ is a constrained quadratic problem (QP) and expressed as

$$\min_{\{\check{\mathbf{W}}^k(t), \check{a}^k(t)\}} \begin{cases} \check{\lambda}_0(t) \|\check{\mathbf{W}}^k(t)\|_F^2 - \operatorname{Re}\{\operatorname{Tr}[(\Phi^k(t))^\dagger \check{\mathbf{W}}^k(t)]\} \\ + \frac{c}{2} \|\check{\mathbf{W}}^k(t) - \mathbf{U}^k(t)\|_F^2 \\ + [\check{\lambda}_1(t) - \beta^k(t)] \check{a}^k(t) + \frac{c}{2} [\check{a}^k(t) - x^k(t)]^2 \end{cases}$$

s.t. $\|\check{\mathbf{W}}^k(t)\|_F^2 \leq \check{a}^k(t)P$, $0 \leq \check{a}^k(t) \leq 1$

where $\Phi^k(t)$ and $\mathbf{U}^k(t) \in \mathbb{C}^{N \times M}$ are defined similarly as $\check{\mathbf{W}}^k(t)$, denoting the k th row blocks of $\Phi(t)$ and $\mathbf{U}(t)$; $x^k(t)$ and $\beta^k(t)$ are similarly defined as $\check{a}^k(t)$, denoting the elements in \mathbf{x}_a and β corresponding to $\check{a}^k(t)$.

Employing the Lagrangian method, we solve it as

$$\begin{cases} \check{\mathbf{W}}^k(t) = \frac{c\mathbf{U}^k(t) + \Phi^k(t)}{c + 2\check{\lambda}_0(t) + 2\mu^k(t)} \\ \check{a}^k(t) = \left[\frac{c\check{\lambda}_1(t) + \mu^k(t)P + \beta^k(t) - \check{\lambda}_1(t)}{c} \right]_0^1 \end{cases} \quad (16)$$

where $\mu^k(t) \geq 0$ is the Lagrangian multiplier of $\|\check{\mathbf{W}}^k(t)\|_F^2 \leq \check{a}^k(t)P$ and should be chosen to satisfy the KKT conditions [17]; $[\cdot]_0^1$ denotes the projection onto the range of $[0, 1]$.

• Let $\mathbf{Y}(t) \triangleq [\check{\mathbf{H}}^\dagger(t)\mathbf{U}(t), \sigma\mathbf{1}]$ and $\mathbf{Z}(t) \triangleq c\mathbf{Y}(t) - \Psi(t)$, with $\mathbf{y}^m(t)$ and $\mathbf{z}^m(t)$ being the m th row vectors of $\mathbf{Y}(t)$ and $\mathbf{Z}(t)$, respectively. The problem of $\mathbf{f}^m(t)$ is expressed as

$$\min_{\mathbf{f}^m(t)} \operatorname{Re}\{\operatorname{Tr}[(\psi^m(t))^\dagger \mathbf{f}^m(t)]\} + \frac{c}{2} \|\mathbf{f}^m(t) - \mathbf{y}^m(t)\|_2^2$$

s.t. $\sqrt{\gamma^{-1}} \mathbf{f}^m(t) \geq \|\mathbf{f}_{-m}^m(t)\|_2$

The first-order optimality conditions for $\mathbf{f}^m(t)$ are given by

$$\begin{cases} c\mathbf{f}_m^m(t) - \mathbf{z}_m^m(t) = \tau\sqrt{\gamma^{-1}} \\ \mathbf{z}_{-m}^m(t) - c\mathbf{f}_{-m}^m(t) \in \tau\partial\|\mathbf{f}_{-m}^m(t)\|_2 \end{cases} \quad (17)$$

where $\mathbf{z}_m^m(t)$ and $\psi_m^m(t)$ are the m th elements of $\mathbf{z}^m(t)$ and $\psi^m(t)$, respectively; $\mathbf{z}_{-m}^m(t)$ and $\psi_{-m}^m(t)$ are the remaining subvectors after removing $\mathbf{z}_m^m(t)$ and $\psi_m^m(t)$ from $\mathbf{z}^m(t)$ and $\psi^m(t)$, respectively; $\tau \geq 0$ is the Lagrangian multiplier with $\sqrt{\gamma^{-1}} \mathbf{f}_m^m(t) \geq \|\mathbf{f}_{-m}^m(t)\|_2$; $\partial\|\mathbf{f}_{-m}^m(t)\|_2$ is the subgradient of the nonsmooth function $\|\mathbf{f}_{-m}^m(t)\|_2$, i.e.,

$$\partial\|\mathbf{f}^m(t)\|_2 \triangleq \begin{cases} \frac{\mathbf{f}^m(t)}{\|\mathbf{f}^m(t)\|_2} \\ \{\mathbf{s} \mid \|\mathbf{s}\|_2 \leq 1, \mathbf{s} \in \mathbb{C}^{1 \times (M+1)}\} \end{cases} \quad (18)$$

Then, the optimal solution to $\mathbf{f}^m(t)$ is given by:

If $\|\mathbf{z}_{-m}^m(t)\|_2 \leq -\sqrt{\gamma}z_m^m(t)$, we have $\mathbf{f}^m(t) = \mathbf{0}$.

Else, we have

$$\begin{cases} \mathbf{f}_m^m(t) = \frac{\tau\sqrt{\gamma^{-1}} + z_m^m(t)}{c} \\ \mathbf{f}_{-m}^m(t) = \frac{\|\mathbf{z}_{-m}^m(t)\|_2 - \tau}{\|\mathbf{z}_{-m}^m(t)\|_2} \cdot \frac{\mathbf{z}_{-m}^m(t)}{\|\mathbf{z}_{-m}^m(t)\|_2} \end{cases} \quad (19)$$

where $\tau = \frac{[\gamma\|\mathbf{z}_{-m}^m(t)\|_2 - \sqrt{\gamma}z_m^m(t)]^+}{1 + \gamma}$ and $[\cdot]^+ = \max\{0, \cdot\}$.

• To solve $y^k(t)$, let us define $\mathbf{q}^k(t)$ as the $[(t-1)K + k]$ th row vector of \mathbf{Q} . Thus $\mathbf{q}^k(t)\mathbf{x}$ is the element in $\mathbf{Q}\mathbf{x}$ associated with $y^k(t)$. Similarly, let $\alpha^k(t)$ denote the $[(t-1)K + k]$ th element of α . Then the problem of $y^k(t)$ is

$$\min_{y^k(t)} \check{\lambda}_2(t)|y^k(t)| + \alpha^k(t)y^k(t) + \frac{c}{2}[y^k(t) - \mathbf{q}^k(t)\mathbf{x}]^2$$

Let $\xi = c\mathbf{q}^k(t)\mathbf{x} - \alpha^k(t)$, and then $y^k(t)$ can be solved as

$$y^k(t) = \begin{cases} 0, & |\xi| \leq \check{\lambda}_2(t) \\ \frac{|\xi| - \check{\lambda}_2(t)}{c} \cdot \frac{\xi}{|\xi|}, & |\xi| > \check{\lambda}_2(t) \end{cases} \quad (20)$$

2) *Updating* $\{\mathbf{U}, \mathbf{x}_a\}$: Again, the problem of $\{\mathbf{U}, \mathbf{x}_a\}$ can be solved by independently updating $\{\mathbf{u}_m(t)\}$ and $\{\mathbf{x}_a\}$.

• The problem of $\mathbf{u}_m(t)$ is an unconstrained QP, i.e.,

$$\min_{\mathbf{u}_m(t)} \begin{cases} \operatorname{Re}\{\operatorname{Tr}[\phi_m^\dagger(t)\mathbf{u}_m(t)]\} + \frac{c}{2}\|\mathbf{u}_m(t) - \check{\mathbf{w}}_m(t)\|_2^2 \\ - \operatorname{Re}\{\operatorname{Tr}[\psi_m^\dagger(t)\check{\mathbf{H}}(t)\mathbf{u}_m(t)]\} \\ + \frac{c}{2}\|\mathbf{f}_m(t) - \check{\mathbf{H}}(t)\mathbf{u}_m(t)\|_2^2 \end{cases}$$

and can be easily solved as

$$\mathbf{u}_m(t) = [\check{\mathbf{H}}(t)\check{\mathbf{H}}^\dagger(t) + \mathbf{I}]^{-1} \begin{bmatrix} \check{\mathbf{H}}(t)[\mathbf{f}_m(t) + c^{-1}\psi_m(t)] \\ + \check{\mathbf{w}}_m(t) - c^{-1}\phi_m(t) \end{bmatrix} \quad (21)$$

where $\mathbf{f}_m(t)$, $\psi_m(t)$, $\check{\mathbf{w}}_m(t)$ and $\phi_m(t)$ are defined similarly as $\mathbf{u}_m(t)$, denoting the m th column vectors of $\mathbf{F}(t)$, $\Psi(t)$, $\check{\mathbf{W}}(t)$ and $\Phi(t)$, respectively. Notice that $[\check{\mathbf{H}}(t)\check{\mathbf{H}}^\dagger(t) + \mathbf{I}]^{-1}$ can be computed in advance, thereby eliminating the complexity demanding matrix inverse operation in iterative process.

• Denoting $\mathbf{Q} = [\mathbf{Q}_a, \mathbf{Q}_{-a}]$ with $\mathbf{Q}_a \in \mathbb{R}^{(S+1)K \times (S+1)K}$ and $\mathbf{Q}_{-a} \in \mathbb{R}^{(S+1)K \times K}$, we express the problem of \mathbf{x}_a as,

$$\min_{\mathbf{x}_a} \begin{cases} -\alpha^T \mathbf{Q}_a \mathbf{x}_a + \frac{c}{2} \|\mathbf{y} - \mathbf{Q}_a \mathbf{x}_a - \mathbf{Q}_{-a} \mathbf{a}(l-1)\|_2^2 \\ + \beta^T \mathbf{x}_a + \frac{c}{2} \|\mathbf{x}_a - \check{\mathbf{a}}\|_2^2 \end{cases}$$

Then, \mathbf{x}_a can be solved as

$$\mathbf{x}_a = [\mathbf{Q}_a^T \mathbf{Q}_a + \mathbf{I}]^{-1} \begin{bmatrix} \mathbf{Q}_a^T [\mathbf{y} - \mathbf{Q}_{-a} \mathbf{a}(l-1) + c^{-1}\alpha] \\ + (\check{\mathbf{a}} - c^{-1}\beta) \end{bmatrix} \quad (22)$$

Again, since $[\mathbf{Q}_a^T \mathbf{Q}_a + \mathbf{I}]^{-1}$ can be calculated in advance, there is no matrix inverse operation in iterative process.

Remark 1 The online ADMM algorithm has very low complexity. In each timeslot, the complexity of solving (P4) directly, e.g., by the interior-point method, is about $O((KNM)^3(S+1)^3)$ per iteration, while the ADMM algorithm has a complexity of $O(KMN(M+KN)(S+1) + K^2(S+1)^2)$ only.

Remark 2 Due to the separable structure of the problem, the ADMM algorithm can be implemented distributively, then its efficiency will be further improved by exploiting the parallel computation.

V. SIMULATION AND CONCLUSION

We consider a MISO network with $K = 10$ BSs and $M = 10$ users, which are all randomly deployed in a hexagonal cell with the distance between the adjacent corners being $d = 1\text{km}$. Each BS is equipped with $N = 3$ antennas. The power budget for each BS is $P = 1000$, while the noise power is $\sigma^2 = 1$. The QoS requirement for the users is $\gamma = 4$.

We employ block fading channels. The channels in a fading block (or a timeslot) are assumed constant, while the channels in different blocks are generated according to Gaussian distribution. In particular, the elements of $\mathbf{h}_m^k(l)$ are generated according to the distribution $\mathcal{CN}(0, [\sigma_m^k(l)]^2)$, where the variance is given by $[\sigma_m^k(l)]^2 = \rho_m^k(l) \cdot [200/d_m^k(l)]^3$ with $d_m^k(l)$ the distance between BS k and user m at timeslot l , and $\rho_m^k(l)$ the shadowing effect satisfying $10\log_{10}[\rho_m^k(l)] \sim \mathcal{N}(0, 64)$.

We consider the data transmission within a time period of $L = 20$ timeslots, and compare the following three algorithms: 1) the joint BS activation and beamforming without stability control, which optimizes the active BSs and beamformers in different timeslots independently; 2) the offline joint long-term BS activation and beamforming algorithm, which solves (P2) directly by CVX; 3) the ADMM online algorithm, i.e., the proposed algorithm in Table 1.

The performance comparisons in terms of switching frequency and power consumption are displayed in Fig. 1, where the switching frequency is computed as the average switching times per timeslot. The curve labelled with “ $S = 0$ ” means we only consider $\mathbf{a}(l-1)$ and ignore $\mathbf{a}(l+1)$ when optimizing $\mathbf{a}(l)$. As expected, the offline algorithm performs best; the algorithm without stability control performs worst; the online algorithm has a performance in between. In general, switching frequency decreases with λ_2 . In the online algorithm, larger sample size S yields stabler BS-user association due to the more accurate average approximation. In the case of small λ_2 , switching power is negligible, thus the algorithms have similar power consumption. As λ_2 increases, switching power gradually dominates in the total power consumption. Then, obvious advantage of long-term BS activation in power consumption can be observed.

In summary, we develop a low-complexity online algorithm for joint long-term BS activation and beamforming. It achieves green communication by balancing the maintenance power, the transmit power and the switching power effectively. Numerical results have been provided to demonstrate its efficacy.

REFERENCES

- [1] J. Zander, “Performance of optimum transmitter power control in cellular radio systems,” *IEEE Trans. Vehicular Technology*, vol. 41, no. 3, pp. 57–62, Feb. 1992.
- [2] F. Rashid-Farrokhi, L. Tassiulas, and K.J.R. Liu, “Joint optimal power control and beamforming in wireless networks using antenna arrays,” *IEEE Trans. Communications*, vol. 46, no. 10, pp. 1313–1324, Oct. 1998.
- [3] O. Arnold, F. Richter, G. Fettweis, and O. Blume, “Power consumption modeling of different base station types in heterogeneous networks,” in *Proc. Future Network and Mobile Summit*, pp. 1–8, Florence, June 2010.

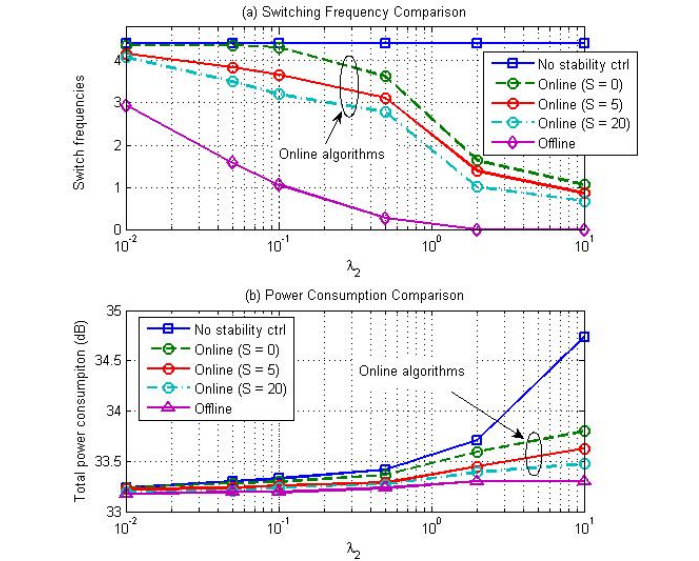


Fig. 1. Performance comparison at different λ_2 for $K = 10$, $M = 10$, $N = 2$, $\gamma = 4$, $\lambda_1 = 10$ and $L = 20$.

- [4] 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,” *IEEE Trans. Signal Processing*, vol. 62, no. 15, pp. 3939–3952, Aug. 2014.
- [5] B. Dai, and W. Yu, “Sparse beamforming for limited-backhaul network MIMO system via reweighted power minimization,” in *Proc. IEEE Globecom*, pp. 1962–1967, Atlanta, Dec. 2013.
- [6] O. Mehanna, N.D. Sidiropoulos, and G.B. Giannakis, “Joint multicast beamforming and antenna selection,” *IEEE Trans. Signal Processing*, vol. 61, no. 10, pp. 2660–2674, May 2013.
- [7] S. Luo, R. Zhang, and T.J. Lim, “Downlink and uplink energy minimization through user association and beamforming in cloud RAN,” *IEEE Trans. Wireless Communications*, vol. 14, no. 1, pp. 494–508, Jan. 2015.
- [8] Y. Shi, J. Zhang, and K.B. Letaief, “Group sparse beamforming for green Cloud-RAN,” *IEEE Trans. Wireless Communications*, vol. 13, no. 5, pp. 2809–2823, May 2014.
- [9] Y. Cheng, M. Pesavento, and A. Philipp, “Joint network optimization and downlink beamforming for CoMP transmissions using mixed integer conic programming,” *IEEE Trans. Signal Processing*, vol. 61, no. 16, pp. 3972–3987, Aug. 2013.
- [10] J. Pan and W. Zhang, “An MDP-based handover decision algorithm in hierarchical LTE networks,” in *Proc. IEEE VTC-Fall*, pp. 1–5, Quebec City, Sept. 2012.
- [11] R. Sun, H. Baligh and Z.-Q. Luo, “Long-term transmit point association for coordinated multipoint transmission by stochastic optimization,” in *Proc. IEEE SPAWC*, pp. 330–334, Darmstadt, June 2013.
- [12] W.-C Li, H.-S. Hsieh, and C.-Y. Chi, “An online algorithm for throughput maximization of wireless powered communication networks,” in *Proc. IEEE ICASSP*, pp. 3731–3735, Shanghai, Mar. 2016.
- [13] A. Shapiro, “On complexity of multistage stochastic programs,” *IEEE Trans. Signal Processing*, vol. 34, no. 1, pp. 1–8, Jan. 2006.
- [14] M. Fukushima, “Application of the alternating direction method of multipliers to separable convex programming problems,” *Computational Optimization and Applications*, vol. 1, no. 1, pp. 93–111, Oct. 1992.
- [15] J. Lin, Q. Li, C. Jiang, and H. Shao, “Joint multi-relay selection, power allocation, and beamformer design for multi-user decode-and-forward relay networks,” *IEEE Trans. Vehicular Technology*, vol. 65, no. 7, pp. 5073–5087, July 2016.
- [16] E. Biglieri, J. Proakis, and S. Shamai, “Fading channels: information theoretic and communication aspects,” *IEEE Trans. Information Theory*, vol. 44, no. 6, pp. 2619–2692, Oct. 1998.
- [17] S. Boyd, and L. Vandenberghe, *Convex Optimization*, Cambridge, UK: Cambridge University Press, 2004.
- [18] M. Grant, S. Boyd, “The CVX users’ guide,” *CVX Research, Inc.*, Oct. 24, 2014 [On-line]. <http://cvxr.com/cvx/>.