# Technical Report: Joint Power and Antenna Selection Optimization in Large Distributed MIMO Networks

An Liu, Vincent Lau, Fellow IEEE,

Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology

#### Abstract

Large MIMO network promises high energy efficiency by employing a large number of antennas. However, the overhead to obtain the full channel state information is very large. To reduce the overhead, we propose a downlink antenna selection scheme, which selects a subset of antennas based on the knowledge of large scale fading factors to serve a given set of users in large distributed MIMO networks employing regularized channel inversion. We study the joint optimization of antenna selection, regularization factor, and power allocation to maximize the average weighted sum-rate. The problem is a mixed combinatorial and non-convex problem whose objective and constraints have no closedform expressions. Random matrix theory is used to derive asymptotically accurate expressions for the objective and constraints. The joint optimization problem is decomposed into three subproblems, each of which is solved by an efficient algorithm. We derive structural solutions for some special cases and obtained the capacity scaling law under very large distributed MIMO networks. We also show that for sufficiently large number of distributed antennas, there is an asymptotic decoupling effect, which can be exploit to simplify algorithms and physical layer processing. Simulations illustrate that the proposed scheme achieves significant performance gain compared with various baselines.

#### **Index Terms**

Large MIMO network, Distributed antenna system, Antenna selection, Asymptotic Analysis

#### I. INTRODUCTION

Large MIMO network has been a hot research topic lately due to the potentially high energy efficiency [1]. Such a network is equipped with an order of magnitude more antennas than conventional systems. For a base station (BS) with  $M \gg 1$  antennas, the total transmit power can be made as O(1/M), and the

transmit power per antenna would be  $O(1/M^2)$  [2]. Furthermore, the gain in multiuser system is very impressive due to the increased degrees of freedom for large MIMO systems. There have been plenty of works on large MIMO networks, including various topics from information theoretical capacity [3], [4] to more practical issues such as transceiver design [5]-[7], channel state information (CSI) acquisition, and pilot contamination problem [8], [9]. Various downlink precoding schemes have been proposed and analyzed. Remarkably, the simple channel inversion (CI) precoding is shown to achieve most of the capacity of large MIMO downlink [2]. One of the main challenges of achieving the performance predicted by the idea analysis is how to obtain the CSI at the transmitter (CSIT) for a very large number of antennas with acceptable amount of overhead. In most of the existing works, Time-Division Duplex (TDD) is assumed and channel reciprocity can be exploited to obtain CSIT via uplink pilot training. However, there is still no efficient method for CSIT training and feedback in Frequency-Division Duplex (FDD) networks. Even if perfect channel reciprocity is assumed, the performance of multi-cell large MIMO system is still limited by the CSI error caused by pilot contamination [8], [9]. In [10], [11], random matrix theory has been used to analyze the asymptotic performance of CI and/or regularized channel inversion (RCI) [12]. However, they did not consider large scaling fading such as path loss and shadow fading, which has huge impact on the performance of large distributed MIMO networks.

In this paper, we focus on large distributed MIMO networks in which there are M distributed antennas (thin base stations) linked together by high speed fiber backhaul as illustrated in Fig. 1. In order to limit the pilot training overhead for downlink transmissions, a subset of S active antennas is selected to serve a given set of K users. The main gain of employing a large number of distributed antennas is due to the reduced transmission distances between each user and the antennas nearby. There are some works studying antenna selection problems in point-to-point MIMO links [13], [14]. Yet. these approaches are not applicable due to the heterogeneous path loss induced by the distributed topology. We study the joint optimization of antenna selection, regularization factor in RCI precoding, and power allocation, which is formulated as an average weighted sum-rate maximization problem under sum power and per antenna power constraints. The optimization requires the knowledge of large scale fading factors only but there are several first-order technical challenges that need to be addressed.

- Closed-Form Optimization Objective and Constraints: There is no simple closed-form expressions for the average weighted sum-rate and the average per-antenna transmission power due to the distributed topology. The asymptotic approach in [10], [11] cannot be applied here due to the heterogeneous path loss among antenna-user pairs.
- Combinatorial Optimization Problem: Even if there were expressions for the optimization



Figure 1. Illustration of a large distributed MIMO network

objectives, the antenna selection problem is combinatorial and brute force solution requires exhaustive search, which is highly undesirable. Furthermore, traditional scheme in which each user is associated with the closest antennas (thin base stations) is highly suboptimal due to the cooperative processing.

In this paper, we first outline the system model and the antenna selection formulation in Section II and III. Using random matrix theory, we derive an asymptotically accurate expressions for the optimization objectives and constraints in Section III. By exploiting the implicit structure in the objective and constraints functions, the joint optimization problem is decomposed into simpler subproblems, each of which is solved by an efficient algorithm in Section IV. In Section V, we focus on studying the structural properties of the solution for some interesting special cases. We show that for large M, there is an asymptotic decoupling effect in the distributed MIMO networks. Simulations in Section VI illustrate that the proposed solution achieve significant performance gain compared with various baselines.

#### II. SYSTEM MODEL

## A. Channel Model

Consider the downlink of a large distributed MIMO network with  $M \gg 1$  distributed transmit antennas and K single-antenna users as illustrated in Fig. 1. The M transmit antennas are distributed geographically and connected to a Cloud RAN [15] via high speed fiber backhaul. Denote  $h_{km}$  as the channel between the  $m^{\text{th}}$  transmit antenna and the  $k^{\text{th}}$  user. We consider composite fading channel, i.e.,  $h_{km} = \sigma_{km}W_{km}, \forall k, m$ , where  $\sigma_{km} > 0$  is the large scale fading factor caused by, e.g., path loss and/or shadow fading, and  $W_{km}$  is the small scale fading factor. Assumption 1 (Assumptions on the composite fading channel model): The small scale fading process  $W_{km}(t) \sim C\mathcal{N}(\mathbf{0}, 1)$  is quasi-static within a time slot but i.i.d. w.r.t. time slots and the spatial indices k, m. The large scale fading process  $\sigma_{km}(t)$  is assumed to be a slow ergodic random process according to a general distribution. It is also independent w.r.t. the spatial indices k, m.

## B. Physical Layer Processing

We assume  $M \gg K$  in the distributed MIMO network. While the M antennas are geographically distributed in the coverage area, the baseband processing is centralized at the cloud RAN. To limit the pilot training overhead to serve the K users, we consider antenna selection scheme where only a subset  $\mathcal{A}$ ,  $|\mathcal{A}| = S$  of the M distributed antennas are selected (activated) to serve the K users. For convenience, let  $\mathcal{A}_j$  denote the  $j^{\text{th}}$  element in  $\mathcal{A}$ . Let  $\mathbf{H}(\mathcal{A}) \in \mathbb{C}^{K \times S}$  denote the composite downlink channel matrix between the selected S antennas and the K users, and define  $\Sigma(\mathcal{A}) \in \mathbb{R}_{++}^{K \times S}$  as the corresponding large scale fading matrix, whose element at  $k^{\text{th}}$  row and the  $j^{\text{th}}$  column is  $\sigma_{k\mathcal{A}_j}$ . In the rest of the paper, we will use  $\mathbf{H}$  and  $\Sigma$  as the simplified notations for  $\mathbf{H}(\mathcal{A})$  and  $\Sigma(\mathcal{A})$  when there is no ambiguity. The cloud RAN is assumed to have knowledge of the  $K \times M$  large scale fading factors  $\sigma_{km}$ 's for antenna selection but only the  $K \times S$  small scale fading factors for linear precoding.

We consider linear precoding processing to support simultaneous downlink transmissions to the K users using the subset of active antennas A. The  $K \times 1$  composite receive signal vector for the K users can be expressed as:

$$y = HFs + n$$
,

where  $\mathbf{s} = [s_1, ..., s_K] \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_K)$  is the symbol vector,  $\mathbf{F} \in \mathbb{C}^{S \times K}$  is the pre-coding matrix and  $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_K)$  is the AWGN noise vector. We employ regularized channel inversion (RCI) pre-coding [12] which is given by

$$\mathbf{F} = \left(\mathbf{H}^{\dagger}\mathbf{H} + \alpha \mathbf{I}_{S}\right)^{-1}\mathbf{H}^{\dagger}\mathbf{P}^{1/2} = \mathbf{H}^{\dagger}\left(\mathbf{H}\mathbf{H}^{\dagger} + \alpha \mathbf{I}_{K}\right)^{-1}\mathbf{P}^{1/2},$$
(1)

where  $\alpha$  is the regularization factor and  $\mathbf{P} = \text{diag}(p_1, ..., p_K)$  is a power allocation matrix. Define power allocation vector as  $\mathbf{p} = [p_1, ..., p_K]^T$ . Note that the above RCI is a generalization of the conventional RCI in [12], where  $\mathbf{P}$  is fixed as  $c\mathbf{I}_K$  and c is chosen to satisfy sum power constraint.

For convenience, define the normalized channel matrix  $\hat{\mathbf{H}} = \mathbf{H}/\sqrt{S}$  and normalized regularization factor  $\rho = \alpha/S$ . Let  $\hat{\mathbf{h}}_k$  denote the  $k^{\text{th}}$  row of  $\hat{\mathbf{H}}$  and  $\hat{\mathbf{g}}_j$  denote the  $j^{\text{th}}$  column of  $\hat{\mathbf{H}}$ . Define  $\hat{\mathbf{H}}_k$  as the matrix  $\hat{\mathbf{H}}$  with the  $k^{\text{th}}$  row removed, and  $\mathbf{P}_k \triangleq \text{diag}(p_1, ..., p_{k-1}, p_{k+1}, ..., p_K)$ . Using matrix inversion lemma, it can be shown that the SINR of user k is given by [10]

$$\gamma_k \left( \mathbf{\Sigma}, \rho, \mathbf{p} \right) = \frac{p_k A_k^2}{B_k + \left( 1 + A_k \right)^2},\tag{2}$$

where

$$A_{k} = \hat{\mathbf{h}}_{k} \left( \hat{\mathbf{H}}_{k}^{\dagger} \hat{\mathbf{H}}_{k} + \rho \mathbf{I}_{S} \right)^{-1} \hat{\mathbf{h}}_{k}^{\dagger},$$
  

$$B_{k} = \hat{\mathbf{h}}_{k} \left( \hat{\mathbf{H}}_{k}^{\dagger} \hat{\mathbf{H}}_{k} + \rho \mathbf{I}_{S} \right)^{-1} \hat{\mathbf{H}}_{k}^{\dagger} \mathbf{P}_{k} \hat{\mathbf{H}}_{k} \left( \hat{\mathbf{H}}_{k}^{\dagger} \hat{\mathbf{H}}_{k} + \rho \mathbf{I}_{S} \right)^{-1} \hat{\mathbf{h}}_{k}^{\dagger}.$$

The transmit power of the  $j^{\text{th}}$  selected antenna in  $\mathcal{A}$  is given by

$$p_{\mathcal{A}_j}\left(\mathbf{\Sigma}, \rho, \mathbf{p}\right) = \hat{\mathbf{g}}_j^{\dagger} \left(\hat{\mathbf{H}}\hat{\mathbf{H}}^{\dagger} + \rho \mathbf{I}_K\right)^{-1} \mathbf{P} \left(\hat{\mathbf{H}}\hat{\mathbf{H}}^{\dagger} + \rho \mathbf{I}_K\right)^{-1} \hat{\mathbf{g}}_j / S.$$
(3)

## III. OPTIMIZATION FORMULATION FOR DYNAMIC ANTENNA SELECTION

## A. Optimization Variables, Objective and Constraints

In this paper, we consider the joint optimization of active antenna set  $\mathcal{A}$ , regularization factor  $\rho$  and the power allocation **p**. The optimization is performed over the time scale of large scale fading, i.e.,  $\mathcal{A}$ ,  $\rho$  and **p** are only adaptive to the large scale fading factors. The objective is to maximize the weighted sum-rate averaged over one large scale fading block. Given a realization of  $\sigma_{km}$ 's and an active antenna set  $\mathcal{A}$ , the large scale fading matrix  $\Sigma(\mathcal{A})$  is fixed within a large scale fading block, and the conditioned average weighted sum-rate is given by

$$\mathcal{I}(\mathcal{A},\rho,\mathbf{p}) = \mathbf{E}\left[\sum_{k=1}^{K} w_k \log\left(1+\gamma_k\left(\mathbf{\Sigma},\rho,\mathbf{p}\right)\right) |\mathbf{\Sigma}(\mathcal{A})\right],\tag{4}$$

where the conditioned expectation  $E[\cdot | \mathbf{\Sigma} (\mathcal{A})]$  is taken over the small scale fading factors, the weights satisfies  $w_k > 0$  and  $\sum_{k=1}^{K} w_k = 1$  without loss of generality. We consider both per antenna and sum power constraint, which are given by

$$\mathbb{E}\left[p_{m}\left(\boldsymbol{\Sigma},\rho,\mathbf{p}\right)|\boldsymbol{\Sigma}\left(\mathcal{A}\right)\right] \leq \bar{p}_{m}, \ m \in \mathcal{A},\tag{5}$$

$$\mathbb{E}\left[\sum_{m\in\mathcal{A}}p_{m}\left(\boldsymbol{\Sigma},\rho,\mathbf{p}\right)|\boldsymbol{\Sigma}\left(\mathcal{A}\right)\right] \leq P_{T}.$$
(6)

## B. Problem Formulation

The optimization problem is formulated as follows

$$\max_{\mathcal{A},\rho>0,\mathbf{p}\geq 0} \mathcal{I}\left(\mathcal{A},\rho,\mathbf{p}\right), \text{ s.t. } (5) \text{ and } (6) \text{ are satisfied}, |\mathcal{A}| = S.$$
(7)

There are several challenges to solve Problem (7). First, there is no analytical expression for the optimization objective in (4) and the constraints in (5) and (6). Second, determining the optimal  $\mathcal{A}$  in (7) requires an exhaustive search over the entire antenna set. Third, even for fixed  $\mathcal{A}$ , the problem is in general non-convex w.r.t.  $\rho$  and **p**, as will be shown later. In this paper, the first challenge is tackled in this section by using random matrix theory to derive asymptotically accurate expressions for the optimization objective and constraints. The last two challenges are tackled by decomposing the problem into simpler subproblems, each of which is solved by an efficient algorithm in Section IV.

We first have the following assumption on the large scale fading matrix  $\Sigma$ .

Assumption 2 (Finite Dynamic Range on Large Scale Fading Matrix): There exist finite positive constant  $\sigma_{max}$  and  $\sigma_{min}$  such that

$$\sigma_{\min} \leq |\sigma_{km}| \leq \sigma_{\max}, \ \forall 1 \leq k \leq K, \ m \in \mathcal{A}.$$

This assumption is a mild assumption and can easily be satisfied in practice.

In the following, we derive asymptotic expressions for SINR and per-antenna power constraints under Assumption 2 in large system limit when  $K, S \to \infty$  with the ratio  $\beta = K/S$  fixed.

Lemma 1 (Asymptotic SINR): For a given  $(\mathcal{A}, \rho, \mathbf{p})$ , the following are true:

1) For any  $k \in \{1, ..., K\}$ , the system of K equations:

$$\xi_{kl} = \frac{1}{S} \sum_{m \in \mathcal{A}} \left[ \sigma_{lm}^2 / f_m\left(\vec{\xi}_k\right) \right], \ l = 1, \dots, K, \tag{8}$$

admits a unique solution  $\vec{\xi}_k = [\xi_{k1}, ..., \xi_{kK}]^T$  in  $\mathbb{R}_{++}^K$ , where  $f_m\left(\vec{\xi}_k\right) \triangleq \rho + \frac{1}{S} \sum_{i \neq k}^K \frac{\sigma_{im}^2}{1 + \xi_{km}}$ .

2) For any  $k \in \{1, ..., K\}$ , define matrix  $\mathbf{D}_k \in \mathbb{R}^{K \times K}$  whose element at the *l*<sup>th</sup> row and *n*<sup>th</sup> column is given by

$$D_{k,ln} = \frac{1}{S} \sum_{m \in \mathcal{A}} \left[ \frac{1}{S} \sigma_{lm}^2 \sigma_{nm}^2 / \left( (1 + \xi_{kn})^2 f_m^2 \left( \vec{\xi}_k \right) \right) \right].$$

Define vector  $\mathbf{b}_k = [b_{k1}, ..., b_{kK}]^T$  whose  $l^{\text{th}}$  element is given by

$$b_{kl} = -\frac{1}{S} \sum_{m \in \mathcal{A}} \left[ \sigma_{lm}^2 / f_m^2 \left( \vec{\xi}_k \right) \right].$$

Define vector  $\mathbf{d}_k = [d_{k1}, ..., d_{kK}]^T$  whose  $l^{\text{th}}$  element is given by

$$d_{kl} = -\frac{1}{S} \sum_{m \in \mathcal{A}} \left[ \sigma_{km}^2 \sigma_{lm}^2 / f_m^2 \left( \vec{\xi}_k \right) \right]$$

Then  $I_K - D_k$  is invertible and thus we can define the following vectors

$$\vec{\phi}_k \triangleq (\mathbf{I}_K - \mathbf{D}_k)^{-1} \mathbf{b}_k, \tag{9}$$

$$\vec{\theta}_k \triangleq (\mathbf{I}_K - \mathbf{D}_k)^{-1} \mathbf{d}_k, \ k = 1, ..., K,$$
(10)

where the  $l^{\text{th}}$  element of  $\vec{\phi}_k$ , denoted by  $\phi_{kl}$ , is the partial derivative of  $\xi_{kl}$  over  $\rho$ .

3) As  $K, S \to \infty$  with the ratio  $\beta = K/S$  fixed,  $\gamma_k(\Sigma, \rho, \mathbf{p})$  in (2) converges almost surely to the following deterministic value

$$\bar{\gamma}_k\left(\mathbf{\Sigma}, \rho, \mathbf{P}\right) = \frac{p_k \xi_{kk}^2}{-\frac{1}{S} \sum_{l \neq k}^K \left[ p_l \theta_{kl} / \left(1 + \xi_{kl}\right)^2 \right] + \left(1 + \xi_{kk}\right)^2},\tag{11}$$

where  $\theta_{kl} < 0$  is the  $l^{\text{th}}$  element of  $\vec{\theta_k}$ .

Lemma 2 (Asymptotic per-antenna power): For a given  $(\mathcal{A}, \rho, \mathbf{p})$ , the following are true:

1) The system of K equations:

$$v_{l} = \frac{1}{\rho + \frac{1}{S} \sum_{m \in \mathcal{A}} \left[ \sigma_{lm}^{2} / h_{m} \left( \mathbf{v} \right) \right]}, \ l = 1, ..., K$$
(12)

admits a unique solution  $\mathbf{v} = [v_1, ..., v_K]^T$  in  $\mathbb{R}_{++}^K$ , where  $h_m(\mathbf{v}) \triangleq 1 + \frac{1}{S} \sum_{i=1}^K \sigma_{im}^2 v_i$ .

2) Define a matrix  $\mathbf{C} \in \mathbb{R}^{K \times K}$  whose element at the  $l^{\text{th}}$  row and  $n^{\text{th}}$  column is given by

$$C_{ln} = \frac{1}{S} \sum_{m \in \mathcal{A}} \left[ \frac{1}{S} \sigma_{lm}^2 \sigma_{nm}^2 v_l / h_m^2 \left( \mathbf{v} \right) \right].$$
(13)

Define a diagonal matrix  $\Delta$  with the  $l^{\text{th}}$  diagonal element given by

$$\Delta_{l} = \frac{1}{S} \sum_{m \in \mathcal{A}} \left[ \sigma_{lm}^{2} / h_{m} \left( \mathbf{v} \right) \right].$$
(14)

Define a vector  $\mathbf{c} = [c_1, ..., c_K]^T$  with the  $l^{\text{th}}$  element given by

$$c_{l} = \frac{1}{S} \sum_{m \in \mathcal{A}} \left[ \sigma_{lm}^{2} v_{l} \left( p_{l} + \frac{1}{S} \sum_{i=1}^{K} \sigma_{im}^{2} v_{i} \left( p_{l} - p_{i} \right) \right) / h_{m}^{2} \left( \mathbf{v} \right) \right]$$

Then  $\rho \mathbf{I}_K + \mathbf{\Delta} - \mathbf{C}$  is invertible and thus we can define the following vector

$$\vec{\varphi} \triangleq (\rho \mathbf{I}_K + \mathbf{\Delta} - \mathbf{C})^{-1} \mathbf{c}.$$
 (15)

As K, S → ∞ with the ratio β = K/S fixed, p<sub>m</sub> (Σ, ρ, p) in (3) converges almost surely to the following deterministic value

$$\bar{p}_m\left(\mathbf{\Sigma}, \rho, \mathbf{p}\right) = \frac{\rho^{-1}}{S^2} h_m^{-2}\left(\mathbf{v}\right) \sum_{i=1}^K \sigma_{im}^2\left(p_i v_i - \varphi_i\right), \ \forall m \in \mathcal{A},\tag{16}$$

where  $\varphi_i$  is the *i*<sup>th</sup> element of  $\vec{\varphi}$  in (15).

Please refer to Appendix A for the proof of the above two lemmas.

For convenience, define  $\bar{p}_{\mathcal{A}}(\mathbf{\Sigma}, \rho, \mathbf{p}) \triangleq \sum_{m \in \mathcal{A}} \bar{p}_m(\mathbf{\Sigma}, \rho, \mathbf{p})$ . Using the above two lemmas, we can prove the following theorem which gives an asymptotic equivalence of Problem (7).

Theorem 1 (Asymptotic equivalence of Problem (7)): Let  $\mathcal{A}^*, \rho^*, \mathbf{p}^*$  denote an optimal solution of the following optimization problem

$$\bar{\mathcal{I}}^* \triangleq \max_{\mathcal{A}, \rho > 0, \mathbf{p} \ge 0} \quad \bar{\mathcal{I}}\left(\mathcal{A}, \rho, \mathbf{p}\right) \triangleq \sum_{k=1}^{K} w_k \log\left(1 + \bar{\gamma}_k\left(\mathbf{\Sigma}, \rho, \mathbf{p}\right)\right) \tag{17}$$

s.t. 
$$\bar{p}_m(\mathbf{\Sigma}, \rho, \mathbf{p}) \le \bar{p}_m, \ \forall m \in \mathcal{A}, \ \bar{p}_{\mathcal{A}}(\mathbf{\Sigma}, \rho, \mathbf{p}) \le P_T,$$
 (18)

where  $\bar{\gamma}_k(\mathbf{\Sigma}, \rho, \mathbf{p})$  and  $\bar{p}_m(\mathbf{\Sigma}, \rho, \mathbf{p})$  are defined in (11) and (16) respectively. As  $K, S \to \infty$  with the ratio  $\beta = K/S$  fixed,  $\mathcal{A}^*, \rho^*, \mathbf{p}^*$  converges almost surely to the optimal solution of Problem (7), and  $\bar{\mathcal{I}}^*$  converges to the optimal value of Problem (7).

The above theorem implies that the solution of the complicated problem in (7) can be approximated by the solution of Problem (17), and the approximation is asymptotically accurate as  $K, S \rightarrow \infty$ . Fig. 6 shows that the approximation is very good even for a finite number of transmit antennas and users.

## IV. OPTIMIZATION SOLUTION FOR PROBLEM (17)

In this section, we shall tackle the remaining challenges of solving the optimization problem in (17). We first decompose the complex problems into simpler subproblems, and then propose efficient algorithms for solving the subproblems.

### A. Problem Decomposition

Using primal decomposition, problem (16) can be decomposed into the following 3 subproblems: **Subproblem 1:** Optimization of **p** under fixed A and  $\rho$ , which can be formulated as

$$\mathcal{P}_{1}\left(\mathcal{A},\rho\right): \max_{\mathbf{p}\geq0} \bar{\mathcal{I}}\left(\mathcal{A},\rho,\mathbf{p}\right), \text{ s.t. } (18) \text{ is satisfied.}$$
(19)

**Subproblem 2:** Optimization of  $\rho$  under fixed A, which can be formulated as

$$\mathcal{P}_{2}\left(\mathcal{A}\right): \max_{\rho>0} \bar{\mathcal{I}}\left(\mathcal{A}, \rho, \mathbf{p}^{*}\left(\mathcal{A}, \rho\right)\right), \text{ s.t. } (18) \text{ is satisfied},$$
(20)

where  $\mathbf{p}^{*}(\mathcal{A}, \rho)$  is the optimal solution of (19).

Subproblem 3: Optimization of A.

$$\mathcal{P}_{3}: \max_{\mathcal{A}} \bar{\mathcal{I}}\left(\mathcal{A}, \rho^{*}\left(\mathcal{A}\right), \mathbf{p}^{*}\left(\mathcal{A}, \rho^{*}\left(\mathcal{A}\right)\right)\right), \text{ s.t. } \mathcal{A} \subseteq \left\{1, ..., M\right\}, \text{ and } |\mathcal{A}| = S,$$
(21)

where  $\rho^*(\mathcal{A})$  is the optimal solution of subproblem 2.

Subproblem 1 and 2 are in general non-convex and it is difficult to obtain the optimal solution. In Section IV-B, we propose *Algorithm S1* which converges to a stationary point for Subproblem 1. In Section IV-C, a bisection method is used to solve Subproblem 2. In Section IV-D, we propose an efficient algorithm for Subproblem 3. For some special cases discussed in Section V, the proposed algorithms are asymptotically optimal.

## B. Algorithm S1 for Solving Subproblem 1

Subproblem 1 in (19) can be rewritten as a weighted sum-rate maximization problem under linear constraints for *K*-user interference channel as follows. First, rewrite the objective  $\overline{I}(\mathcal{A}, \rho, \mathbf{p})$  as

$$\bar{\mathcal{I}}\left(\mathcal{A},\rho,\mathbf{p}\right) = \sum_{k=1}^{K} w_k \log\left(1 + g_{kk}p_k / \left(1 + \sum_{l\neq k}^{K} g_{kl}p_l\right)\right),$$
$$g_{kk} \triangleq \frac{\xi_{kk}^2}{\left(1 + \xi_{kk}\right)^2}, \,\forall k, \, g_{kl} \triangleq \frac{-\theta_{kl}}{S\left(1 + \xi_{kl}\right)^2\left(1 + \xi_{kk}\right)^2}, \,\forall k \neq l.$$

Recall the definitions of  $\mathbf{v}$ ,  $h_m(\mathbf{v})$ ,  $m \in \mathcal{A}$ ,  $\boldsymbol{\Delta}$  and  $\mathbf{C}$  in Lemma 2. Define a  $K \times S$  matrix  $\hat{\mathbf{R}}$  with the element at the  $k^{\text{th}}$  row and the  $j^{\text{th}}$  column given by

$$\hat{R}_{kj} = \frac{1}{S^2} \rho^{-1} \sigma_{k\mathcal{A}_j}^2 h_{\mathcal{A}_j}^{-2} \left( \mathbf{v} \right),$$

and define  $\bar{\mathbf{R}}$  as a  $K \times K$  matrix with each element given by

$$\bar{R}_{kk} = \frac{1}{S} \sum_{m \in \mathcal{A}} \left[ \left( 1 + \frac{1}{S} \sum_{i \neq l}^{K} \sigma_{im}^2 v_i \right) \sigma_{lm}^2 v_l / h_m^2 \left( \mathbf{v} \right) \right], \forall k = \bar{R}_{kl} = -\frac{1}{S} \sum_{m \in \mathcal{A}} \left[ \frac{1}{S} \sigma_{lm}^2 v_l \sigma_{km}^2 v_k / h_m^2 \left( \mathbf{v} \right) \right], \forall k \neq l.$$

Let  $\mathbf{V} = \operatorname{diag}(v_1, ..., v_K)$ . Then the per antenna power constraint in (18) can be rewritten as  $\tilde{\mathbf{R}}\mathbf{p} \leq [\bar{p}_{\mathcal{A}_1}, ..., \bar{p}_{\mathcal{A}_S}]^T$ , where  $\tilde{\mathbf{R}} \triangleq \hat{\mathbf{R}}^T \left[ \mathbf{V} - (\rho \mathbf{I}_K + \boldsymbol{\Delta} - \mathbf{C})^{-1} \bar{\mathbf{R}} \right] \in \mathbb{R}^{S \times K}$ . The sum power constraint in (18) can be rewritten as  $\mathbf{1}^T \tilde{\mathbf{R}} \mathbf{p} \leq \bar{p}_{\mathcal{A}}$ . Finally, the overall constraint in (18) can be expressed in a compact form as  $\mathbf{R}\mathbf{p} \leq \bar{\mathbf{p}}$ ,  $\mathbf{p} \geq 0$ , where  $\mathbf{R} \triangleq \begin{bmatrix} \mathbf{I}_S, \mathbf{1} \end{bmatrix}^T \tilde{\mathbf{R}} \in \mathbb{R}^{(S+1) \times K}$ , and  $\bar{\mathbf{p}} = [\bar{p}_{\mathcal{A}_1}, ..., \bar{p}_{\mathcal{A}_S}, \bar{p}_{\mathcal{A}}]^T$ .

The Lagrange function of Subproblem 1 is given by

$$L\left(\vec{\lambda},\mathbf{p}\right) = \bar{\mathcal{I}}\left(\mathcal{A},\rho,\mathbf{p}\right) + \vec{\lambda}^{T}\left(\bar{\mathbf{p}} - \mathbf{R}\mathbf{p}\right), \ \mathbf{p} \ge 0,$$

where  $\vec{\lambda} \in \mathbb{R}^{(S+1)\times 1}_+$  is the Lagrange multipliers. It is well known that Problem (19) is usually a nonconvex problem with non-zero duality gap. Hence, the standard Lagrange dual method (LDM) [16] cannot be used to solve this problem. We proposed a *local LDM* to solve Problem (19) as follows.

Algorithm S1 (for solving Subproblem 1):

**Initialization**: Choose proper initial  $\vec{\lambda}$ , **p**.

**Step 1**: For fixed  $\vec{\lambda}$ , search for a stationary point  $\tilde{\mathbf{p}}\left(\vec{\lambda}\right)$  of

$$\max_{\mathbf{p}} L\left(\vec{\lambda}, \mathbf{p}\right), \text{ s.t. } \mathbf{p} \ge 0, \tag{22}$$

using Algorithm I described later starting from the latest p.

**Step 2**: Update  $\vec{\lambda}$  as

$$\vec{\lambda} = \left[\vec{\lambda} - \tau \left(\bar{\mathbf{p}} - \mathbf{R}\tilde{\mathbf{p}}\left(\vec{\lambda}\right)\right)\right]^+,\tag{23}$$

where  $\tau > 0$  is a small step size.

#### Return to Step 1 until convergence.

In the following, we will propose an efficient inner-loop algorithm to solve Problem (22) based on the interference pricing method [17], which strikes a balance between maximizing each user's own objective and minimizing interference by introducing interference prices in each user's objective function. Specifically, the interference price for user k is given by [17]

$$\pi_k = \sum_{l \neq k}^{K} w_l g_{lk} \frac{g_{ll} p_l}{\Omega_l \left(\Omega_l + g_{ll} p_l\right)},\tag{24}$$

where

$$\Omega_l = 1 + \sum_{i \neq l}^K g_{li} p_i, \tag{25}$$

is the interference-plus-noise power of user l. Given fixed interference prices and powers for the other users,  $p_k$  is updated by maximizing the following objective function

$$\max_{\mathbf{p}} w_k \log\left(1 + \frac{g_{kk} p_k}{1 + \sum_{l \neq k}^K g_{kl} p_l}\right) - \vec{\lambda}^T \mathbf{r}_k p_k - \pi_k p_k, \text{ s.t. } \mathbf{p} \ge 0,$$
(26)

where  $\mathbf{r}_k$  is the  $k^{\text{th}}$  column of **R**. Here, we present an algorithm with sequential power updates to solve the inner loop problem in (22).

Algorithm I (for solving inner-loop Problem (22)):

Initialization: Choose proper initial p.

While not converge do

For k = 1 to K

Calculate  $\pi_k$  in (24) and  $\Omega_k$  in (25) using current power allocation **p**.

Given  $\pi_k$  and current power allocation **p**, update  $p_k$  by solving problem (26) as

$$p_k = \left[\frac{w_k}{\pi_k + \vec{\lambda}^T \mathbf{r}_k} - \frac{\Omega_k}{g_{kk}}\right]^+.$$

End

End

Theorem 2 (Convergence of Alg. I): Algorithm I converges to a stationary point of the inner-loop Problem in (22).

The proof can be established using the same approach as in Appendix A of [17].

Theorem 3 (Convergence of Alg. S1): Let  $\tilde{\mathbf{p}}(\vec{\lambda})$  denote the stationary point of Problem (22) found by Algorithm I for fixed  $\vec{\lambda}$ . Assume that  $\tilde{\mathbf{p}}(\vec{\lambda})$  satisfies the following condition over the trajectory of  $\vec{\lambda}$  generated by Algorithm S1:

$$\lim_{\|\Delta\vec{\lambda}\|\to 0, \, \vec{\lambda}+\Delta\vec{\lambda}\ge 0} \frac{\left\|\tilde{\mathbf{p}}\left(\vec{\lambda}+\Delta\vec{\lambda}\right)-\tilde{\mathbf{p}}\left(\vec{\lambda}\right)\right\|}{\left\|\Delta\vec{\lambda}\right\|} \le B,\tag{27}$$

where B > 0 is some constant. Then Algorithm S1 converges to a stationary point of  $\mathcal{P}_1(\mathcal{A}, \rho)$ . The proof is given in Appendix B.

## C. Bisection Algorithm for Subproblem 2

One of the main challenges for solving Subproblem 2 is that the calculation of the objective function  $\overline{\mathcal{I}}(\mathcal{A},\rho,\mathbf{p}^*(\mathcal{A},\rho))$  requires solving the optimal solution  $\mathbf{p}^*(\mathcal{A},\rho)$  of Subproblem 1 which is a nonconvex problem. In this section, we propose a bisection algorithm with Algorithm S1 as a subroutine to find a good solution for Subproblem 2. This algorithm is also shown to be asymptotically optimal at high SNR under some specific topology in Section V. The algorithm relies on the following theorem.

Theorem 4 (Stationary point of an equivalent problem of  $\mathcal{P}_2(\mathcal{A})$ ): Consider the following joint optimization problem under fixed  $\mathcal{A}$ :

$$\mathcal{P}_{2a}\left(\mathcal{A}\right): \max_{\rho>0,\mathbf{p}\geq 0} \bar{\mathcal{I}}\left(\mathcal{A},\rho,\mathbf{p}\right), \text{ s.t. } (18) \text{ is satisfied.}$$
(28)

Them the followings are true:

- Let ρ\*, p\* denote the optimal solution of P<sub>2a</sub> (A). Then ρ\* must be an optimal solution of P<sub>2</sub> (A), and p\* must be an optimal solution of P<sub>1</sub> (A, ρ\*).
- 2) Let  $\tilde{\mathbf{p}}(\mathcal{A}, \rho)$  denote the stationary point of  $\mathcal{P}_1(\mathcal{A}, \rho)$  found by Algorithm S1. Define a function

$$\hat{\mathcal{I}}(\mathcal{A},\rho) \triangleq \bar{\mathcal{I}}(\mathcal{A},\rho,\tilde{\mathbf{p}}(\mathcal{A},\rho)).$$
<sup>(29)</sup>

Assume that  $\hat{\mathcal{I}}(\mathcal{A},\rho)$  is differentiable over  $\rho$  and let  $\tilde{\rho}(\mathcal{A})$  denote a solution of

$$\frac{\partial \hat{\mathcal{I}}\left(\mathcal{A},\rho\right)}{\partial\rho} = 0. \tag{30}$$

Then  $\tilde{\rho}(\mathcal{A}), \tilde{\mathbf{p}}(\mathcal{A}, \tilde{\rho}(\mathcal{A}))$  must be a stationary point of  $\mathcal{P}_{2a}(\mathcal{A})$ .

*Proof:* The first result is obvious. The second result can be proved using the facts that  $\tilde{\mathbf{p}}(\mathcal{A}, \tilde{\rho}(\mathcal{A}))$  satisfies the KKT condition of  $\mathcal{P}_1(\mathcal{A}, \tilde{\rho}(\mathcal{A}))$  and  $\tilde{\rho}(\mathcal{A})$  is a solution of (30). Details are omitted due to page limit.

The above theorem implies that we can find a good solution for Subproblem 2 by solving Equation (30) using the following bisection algorithm.

Algorithm S2 (Bisection search for solving Subproblem 2):

**Initialization**: Choose proper  $\rho_a$ ,  $\rho_b$  such that  $0 < \rho_a < \rho_b$  and  $\frac{\partial \hat{I}(\mathcal{A},\rho)}{\partial \rho_a} > 0$ ,  $\frac{\partial \hat{I}(\mathcal{A},\rho)}{\partial \rho_b} < 0$ . **Step 1**: Let  $\rho = (\rho_a + \rho_b)/2$ . If  $\frac{\partial \hat{I}(\mathcal{A},\rho)}{\partial \rho} \leq 0$ , let  $\rho_b = \rho$ . Otherwise, let  $\rho_a = \rho$ . **Return to Step 1 until**  $\rho_b - \rho_a$  is small enough.

The main challenge in implementing Algorithm S2 is to calculate the derivative  $\frac{\partial \hat{I}(\mathcal{A},\rho)}{\partial \rho}$  without an analytical expression for  $\hat{\mathcal{I}}(\mathcal{A},\rho)$ . In the following, we show how to calculate  $\frac{\partial \hat{I}(\mathcal{A},\rho)}{\partial \rho}$  from the output of Algorithm S1:  $\tilde{\mathbf{p}}(\mathcal{A},\rho)$  and  $\tilde{\lambda}$ , where  $\tilde{\lambda} \in \mathbb{R}^{(S+1)\times 1}_+$  is the Lagrange multipliers corresponding to  $\tilde{\mathbf{p}}(\mathcal{A},\rho)$ .

By Theorem 3,  $\tilde{\mathbf{p}}(\mathcal{A}, \rho)$  satisfies the KKT conditions of  $\mathcal{P}_1(\mathcal{A}, \rho)$ , which can be expressed as

$$\nabla_{\mathbf{p}} \bar{\mathcal{I}} \left( \mathcal{A}, \rho, \tilde{\mathbf{p}} \left( \mathcal{A}, \rho \right) \right) - \mathbf{R}^{T} \tilde{\lambda} + \tilde{\nu} = 0;$$

$$\operatorname{diag} \left( \bar{\mathbf{p}} \right) \tilde{\lambda} - \operatorname{diag} \left( \tilde{\lambda} \right) \mathbf{R} \tilde{\mathbf{p}} \left( \mathcal{A}, \rho \right) = 0;$$

$$\operatorname{diag} \left( \tilde{\nu} \right) \tilde{\mathbf{p}} \left( \mathcal{A}, \rho \right) = 0;$$

$$\operatorname{diag} \left( \tilde{\nu} \right) \tilde{\mathbf{p}} \left( \mathcal{A}, \rho \right) = 0;$$

$$\operatorname{diag} \left( \tilde{\nu} \right) \tilde{\mathbf{p}} \left( \mathcal{A}, \rho \right) = 0;$$

$$\operatorname{diag} \left( \tilde{\nu} \right) \tilde{\mathbf{p}} \left( \mathcal{A}, \rho \right) = 0;$$

$$\operatorname{diag} \left( \tilde{\nu} \right) \tilde{\mathbf{p}} \left( \mathcal{A}, \rho \right) = 0;$$

$$\operatorname{diag} \left( \tilde{\nu} \right) \tilde{\mathbf{p}} \left( \mathcal{A}, \rho \right) = 0;$$

$$\operatorname{diag} \left( \tilde{\nu} \right) \tilde{\mathbf{p}} \left( \mathcal{A}, \rho \right) = 0;$$

$$\operatorname{diag} \left( \tilde{\nu} \right) \tilde{\mathbf{p}} \left( \mathcal{A}, \rho \right) = 0;$$

where  $\tilde{\nu} \in \mathbb{R}^{K}_{+}$  is the Lagrange multipliers associated with the positive constraint  $\mathbf{p} \geq 0$ . Note that  $\tilde{\lambda}$  can be obtained by Algorithm S1 and  $\tilde{\nu} = \mathbf{R}^{T}\tilde{\lambda} - \nabla_{\mathbf{p}}\bar{\mathcal{I}}(\mathcal{A},\rho,\tilde{\mathbf{p}}(\mathcal{A},\rho))$ , where  $\nabla_{\mathbf{p}}\bar{\mathcal{I}}(\mathcal{A},\rho,\tilde{\mathbf{p}}(\mathcal{A},\rho)) = \begin{bmatrix} \frac{\partial \bar{\mathcal{I}}}{\partial p_{1}}, ..., \frac{\partial \bar{\mathcal{I}}}{\partial p_{K}} \end{bmatrix}$  and  $\frac{\partial \bar{\mathcal{I}}}{\partial p_{K}} = \mathbf{P}(\tilde{\mathcal{A}},\rho,\tilde{\mathbf{p}}(\mathcal{A},\rho))$ 

$$\frac{\partial \mathcal{I}}{\partial p_{k}} = w_{k}g_{kk} / \left(\tilde{\Omega}_{k} + g_{kk}\tilde{p}_{k}\left(\mathcal{A},\rho\right)\right) - \pi_{k},$$

where  $\tilde{\Omega}_k$  is the interference-plus-noise power in (25) calculated from  $\tilde{\mathbf{p}}(\mathcal{A},\rho)$ . Assuming that  $\frac{\partial \tilde{\mathbf{p}}(\mathcal{A},\rho)}{\partial \rho}$ ,  $\frac{\partial \tilde{\lambda}}{\partial \rho}$  and  $\frac{\partial \tilde{\nu}}{\partial \rho}$  exits and taking partial derivative of the equations in (31) with respect to  $\rho$ , we obtain a linear equation with  $\frac{\partial \tilde{\mathbf{p}}(\mathcal{A},\rho)}{\partial \rho}$ ,  $\frac{\partial \tilde{\lambda}}{\partial \rho}$  and  $\frac{\partial \tilde{\nu}}{\partial \rho}$  as the variables. Then we can calculate  $\frac{\partial \tilde{\mathbf{p}}(\mathcal{A},\rho)}{\partial \rho}$  by solving this linear equation. Finally, the derivative  $\frac{\partial \hat{\mathcal{I}}(\mathcal{A},\rho)}{\partial \rho}$  can be calculated as

$$\frac{\partial \hat{\mathcal{I}}\left(\mathcal{A},\rho\right)}{\partial \rho} = \frac{\sum_{l=1}^{K} \left( \tilde{p}_{l}\left(\mathcal{A},\rho\right) \frac{\partial g_{kl}}{\partial \rho} + g_{kl} \frac{\partial \tilde{p}_{l}\left(\mathcal{A},\rho\right)}{\partial \rho} \right)}{g_{kk} \tilde{p}_{k}\left(\mathcal{A},\rho\right) + \tilde{\Omega}_{k}} - \frac{\sum_{l\neq k}^{K} \left( \tilde{p}_{l}\left(\mathcal{A},\rho\right) \frac{\partial g_{kl}}{\partial \rho} + g_{kl} \frac{\partial \tilde{p}_{l}\left(\mathcal{A},\rho\right)}{\partial \rho} \right)}{\tilde{\Omega}_{k}}.$$
 (32)

The detailed calculations and expressions for  $\frac{\partial \tilde{\mathbf{p}}(\mathcal{A},\rho)}{\partial \rho}$ ,  $\frac{\partial g_{kl}}{\partial \rho}$ 's and  $\frac{\partial \hat{\mathcal{I}}(\mathcal{A},\rho)}{\partial \rho}$  can be found in Appendix C.

### D. Algorithm S3 for Solving Subproblem 3

Subproblem 3 is a combinatorial problem and optimal solution requires complex brute-force exhaustive search, which is undesirable. In this section, we shall propose a low complexity algorithm for  $\mathcal{P}_3$ . The proposed solution is also asymptotically optimal for large M as shown in Corollary 1.

A simple traditional antenna selection baseline algorithm is for each user to associate with the strongest base stations (antennas) and this simple algorithm has been adopted in 3G and LTE systems.



Figure 2. An example that strong cross link causes large interference



Figure 3. An example that strong cross links provide cooperative gain

We first illustrate the deficiency of this baseline algorithm using two simple numerical examples. In both examples, we assume S = 2 distributed antennas are selected to serve K = 2 users.

*Example 1 (Strong cross-link causes low SINR)):* Fig. 2 illustrates the path loss configuration of a distributed MIMO network. According to the baseline algorithm, the selected antennas will be  $\mathcal{A} = \{2,3\}$ . Yet, this is obviously not a good choice because antenna 2 causes strong interference to user 2 before pre-coding. Although the interference can be suppressed using RCI pre-coding, the overall SINR is still lower because the cross link from A3 to U1 is week and the joint transmission gain using RCI is very limited. A better choice will be  $\mathcal{A} = \{1,3\}$ .

*Example 2 (Strong cross-link provides cooperative gain):* Fig. 3 illustrates the path loss configuration. According to the baseline algorithm, the selected antennas will be  $\mathcal{A} = \{1, 3\}$ . However, a better performance can be achieved by letting  $\mathcal{A} = \{1, 2\}$  due to cooperative transmission.

As illustrated from Example 1 and 2, it is important to incorporate both the cross link and direct link in computing the antenna selection metric for distributed MIMO networks. If there is an antenna with strong cross links / direct links with several users, then it may or may not be good antenna because it can contribute to both cooperative gain or interference. There is no simple rule to identify good antennas but this insight is helpful to design a reasonable antenna selection algorithm.

Based on the above insight, we propose an efficient algorithm S3 for  $\mathcal{P}_3$ . It contains 4 steps. In step 1, the algorithm selects the antennas which has a direct link with a single user. Such antenna provides a direct link for a single user without causing strong interference to others<sup>1</sup>. In step 2, the algorithm selects the antennas which has strong cross link / direct link with several users. Such antennas have the potential to provide large cooperative gain. In this step, the "bad" antennas which cause strong interference may also be selected. However, they will be identified and deleted in step 4. In step 3, the algorithm selects the antennas which has a strong cross link with a single user (the weight of the user is also considered in the selection). Such antenna provides an (additional) strong channel for a single user without causing strong interference to others. In step 4, a greedy search is performed to replace the "bad" antennas with the "good" antennas chosen from a candidate antenna set  $\Gamma_j$ . In the  $j^{\text{th}}$  search, we switch the  $j^{\text{th}}$  selected antenna  $\mathcal{A}_j$  with each antenna  $m \in \Gamma_j$  and calculate the weighted sum-rate. If the weighted sum-rate is increased, we update  $\mathcal{A}$  as  $\mathcal{A}_j = m$ . The candidate antenna set  $\Gamma_j$  is carefully chosen to reduce the number of weighted sum-rate calculations as well as maintain a good performance. Fig. 4 illustrates the detailed steps of Algorithm S3.

## V. STRUCTURAL SOLUTION FOR SOME SPECIAL CASES

In this section, we focus on deriving structural properties of the optimal solutions under several special cases so as to obtain some design insights.

#### A. Large MIMO Network with Collocated Antennas

We first study the case when the antennas are collocated at the base station. In this case, it is reasonable to assume that all antennas experience the same large scale fading and thus  $\sigma_{km}^2 = \sigma_{k1}^2$ , k = 1, ..., K, m = 1, ..., M. Under this assumption, any subset A of S antennas is optimal for the antenna selection problem since all antennas are statistically equivalent. Hence, we focus on the structural properties of  $\mathbf{p}^*$  and  $\rho^*$ .

We first obtain simpler expressions for asymptotic SINR and transmit power for  $\mathcal{P}_1(\mathcal{A}, \rho)$  and  $\mathcal{P}_2(\mathcal{A})$ .

Theorem 5 (Asymptotic SINR and transmit power for collocated antennas): For a given  $(\mathcal{A}, \rho, \mathbf{p})$ , if  $\sigma_{km}^2 = \sigma_{k1}^2, \ k = 1, ..., K, \ m = 1, ..., M$ , then the followings are true:

<sup>&</sup>lt;sup>1</sup>In the rest of the paper, the phrase "an antenna causes interference to a user" refers to the case when an antenna causes strong interference to a user before pre-coding and the joint transmission using RCI does not provide much gain due to some other week cross links as shown in Example 1.

**Initialization**: Let  $\mathcal{A} = \Phi$ , where  $\Phi$  denote the void set.

Step 1 (Select antennas with a direct link and no cross link):

For k = 1 to K, if  $|\mathcal{K}_{\tilde{m}_k}| = 1$ , let  $\mathcal{A} = \mathcal{A} \cup \tilde{m}_k$ . If  $|\mathcal{A}| = S$ , goto step 4.

Step 2 (Select antennas with multiple strong links):

Let  $\bar{G}_m = |\mathcal{K}_m| B + \sum_{k=1}^K \sigma_{km}^2$ , where  $B > \max_{1 \le m \le M} \sum_{k=1}^K \sigma_{km}^2$  is a large enough constant. Let  $m^* = \underset{m \in \bar{\mathcal{A}}}{\operatorname{argmax}} \bar{G}_m$ While  $|\mathcal{K}_{m^*}| \ge 2$  and  $|\mathcal{A}| < S$ Let  $\mathcal{A} = \mathcal{A} \cup m^*$  and  $m^* = \underset{m \in \bar{\mathcal{A}}}{\operatorname{argmax}} \bar{G}_m$ .

End

If  $|\mathcal{A}| = S$ , goto step 4.

Step 3 (Select antennas with a single strong link):

Let 
$$\tilde{k}_m = \underset{k}{\operatorname{argmax}} \sigma_{km}^2$$
 and  $I_m = w_{\tilde{k}_m} \log \left(1 + \sigma_{\tilde{k}_m m}^2\right)$ .  
While  $|\mathcal{A}| < S$   
Let  $m^* = \underset{m \in \bar{\mathcal{A}}}{\operatorname{argmax}} I_m$  and  $\mathcal{A} = \mathcal{A} \cup m^*$ .  
End

Step 4 (Greedy search for replacing "bad" antennas with "good" ones):

For 
$$j = 1$$
 to  $S$   
Let  $\mathcal{A}_{-j} = \mathcal{A}/\mathcal{A}_j$ . Let  $n^* = \underset{m \in \bar{\mathcal{A}} \cap \{m: |\mathcal{K}_m| = 1\}}{\operatorname{argmax}} I_m$  and  $\Gamma_j^a = \bar{\mathcal{A}} \cap \{m: |\mathcal{K}_m| \ge 2\}$ .  
If  $I_{n^*} \ge I_{\mathcal{A}_j}$  or  $|\mathcal{K}_{\mathcal{A}_j}| > 1$ , let  $\Gamma_j = \Gamma_j^a \cup n^*$ ; otherwise, let  $\Gamma_j = \Gamma_j^a$ .  
Let  $m^* = \underset{m \in \Gamma_j}{\operatorname{argmax}} \tilde{\mathcal{I}}_{\mathcal{A}_{-j} \cup m}$ .  
If  $\tilde{\mathcal{I}}_{\mathcal{A}_{-j} \cup m^*} > \tilde{\mathcal{I}}_{\mathcal{A}}$ , let  $\mathcal{A}_j = m^*$ .  
End

Figure 4. Algorithm S3 for solving Subproblem 3. The following is a list of the notations used in the algorithm. Let  $\tilde{m}_k = \operatorname{argmax}_m \sigma_{km}^2$ , k = 1, ..., K. Define  $\bar{g}_k^d = \sigma_{kn\tilde{n}_k}^2$ . For k = 1, ..., K, and m = 1, ..., M, let  $G_{km} = 1$ , if  $\sigma_{km}^2 \ge \bar{g}_k^d/4$ , and otherwise, let  $G_{km} = 0$ . Define  $\mathcal{K}_m = \{k : G_{km} > 1\}$ , m = 1, ..., M. Let  $\tilde{\mathcal{I}}_{\mathcal{A}} \triangleq \bar{\mathcal{I}}(\mathcal{A}, \tilde{\rho}(\mathcal{A}), \tilde{\mathbf{p}}(\mathcal{A}, \tilde{\rho}(\mathcal{A})))$  denote the weighted sum-rate under  $\mathcal{A}$ . For any set of antennas  $\mathcal{A} \subseteq \{1, ..., M\}$ , let  $\bar{\mathcal{A}}$  denote the relative complement of  $\mathcal{A}$ .

1) The system of equation:

$$u = \frac{1}{\rho + \frac{1}{S} \sum_{i=1}^{K} \frac{\sigma_{i1}^2}{1 + \sigma_{i1}^2 u}},$$
(33)

admits a unique solution u in  $\mathbb{R}_{++}$ .

2) Define

$$F_{12} = \frac{1}{S} \sum_{i=1}^{K} \frac{\sigma_{i1}^2}{\left(1 + \sigma_{i1}^2 u\right)^2}, \ \bar{F}_{12}\left(\mathbf{p}\right) = \frac{1}{S} \sum_{i=1}^{K} \frac{p_i \sigma_{i1}^2}{\left(1 + \sigma_{i1}^2 u\right)^2}.$$

As  $K, S \rightarrow \infty$  with the ratio  $\beta = K/S$  fixed,  $\gamma_k(\mathbf{\Sigma}, \rho, \mathbf{p})$  in (2) and  $p_m(\mathbf{\Sigma}, \rho, \mathbf{p})$  in (3) respectively converge to the following deterministic values

$$\bar{\gamma}_{k}\left(\boldsymbol{\Sigma}, \boldsymbol{\rho}, \mathbf{P}\right) = \frac{p_{k}\sigma_{k1}^{4}u^{2}\left(\boldsymbol{\rho} + F_{12}\right)}{\bar{F}_{12}\left(\mathbf{p}\right)\sigma_{k1}^{2}u + \left(1 + \sigma_{k1}^{2}u\right)^{2}\left(\boldsymbol{\rho} + F_{12}\right)},$$
  
$$\bar{p}_{m}\left(\boldsymbol{\Sigma}, \boldsymbol{\rho}, \mathbf{p}\right) = \frac{\bar{F}_{12}\left(\mathbf{p}\right)u}{S\left(\boldsymbol{\rho} + F_{12}\right)}, \forall m \in \mathcal{A}.$$
(34)

The proof is similar to the proof in Appendix A.

Using Theorem 5,  $\mathcal{P}_1(\mathcal{A}, \rho)$  can be reformulated into a simpler form as follows. First, according to (34), all antennas always have the same transmit power. Furthermore, it can be verified that either the per antenna power constraint or the sum power constraint must be achieved with equality at the optimal solution. Combining these facts and the asymptotic expressions in Theorem 5, subproblem  $\mathcal{P}_1(\mathcal{A}, \rho)$ can be equivalent to the following optimization problem

$$\max_{\mathbf{p}\geq 0} \sum_{k=1}^{K} w_k \log\left(1 + \frac{p_k \sigma_{k1}^4 u^2}{\sigma_{k1}^2 P_T' + \left(1 + \sigma_{k1}^2 u\right)^2}\right), \text{ s.t.} \frac{\bar{F}_{12}\left(\mathbf{p}\right) u}{\left(\rho + F_{12}\right)} \leq P_T',$$
(35)

where  $P'_T = \min\left(P_T, \min_{m \in \mathcal{A}} S\bar{p}_m\right)$ . 1) Water-filling Structure of the Optimal Power Allocation: For fixed  $\mathcal{A}, \rho$ , the optimal power allocation  $\mathbf{p}^{*}(\mathcal{A}, \rho) = [p_{1}^{*}(\mathcal{A}, \rho), ..., p_{K}^{*}(\mathcal{A}, \rho)]$  is given by:

$$p_{k}^{*}(\mathcal{A},\rho) = \left(\frac{w_{k}S\left(1+\sigma_{k1}^{2}u\right)^{2}(\rho+F_{12})}{\lambda\sigma_{k1}^{2}u} - \frac{\sigma_{k1}^{2}P_{T}^{'}+\left(1+\sigma_{k1}^{2}u\right)^{2}}{\sigma_{k1}^{4}u^{2}}\right)^{+},$$
(36)

where  $\lambda$  is chosen such that  $\bar{F}_{12}\left(\mathbf{p}^{*}\left(\mathcal{A},\rho\right)\right)u/\left(\rho+F_{12}\right)=P_{T}^{'}$ .

2) Properties of the optimal  $\rho$  in High SNR Regime: The following theorem summarizes the structural properties of the optimal solution  $\rho^*$  for  $\mathcal{P}_2(\mathcal{A})$ .

Theorem 6 (Properties of  $\rho^*$  at high SNR): For fixed K and S, the following are true:

- 1)  $\rho^* = O\left(\frac{1}{P_T'}\right)$  for large  $P_T'$ .
- 2) There exists a small enough constant  $\rho'_{\max} > 0$  such that the objective of  $\mathcal{P}_{2}(\mathcal{A}): \overline{\mathcal{I}}(\mathcal{A}, \rho, \mathbf{p}^{*}(\mathcal{A}, \rho)),$ is a concave function of  $\rho$  for all  $\rho < \rho'_{\text{max}}$ .

The proof is given in Appendix D. Theorem 6 implies that with sufficiently small initial  $\rho_b > \rho_a > 0$ , Algorithm S2 will converge to the optimal  $\rho^*$  at high SNR.

## B. Large MIMO Network with Collocated Users

In this case, all users experience the same large scale fading:  $\sigma_{km}^2 = \sigma_{1m}^2$ , k = 1, ..., K, m = 1, ..., M. As a result, the antenna selection problem has a trivial solution:  $\mathcal{A}^* = \{m : \sigma_{1m}^2 \ge \sigma_{1S^{\text{max}}}^2\}$ , where  $\sigma_{1S^{\text{max}}}^2$  is the S<sup>th</sup> largest  $\sigma_{1m}^2$ 's and we shall focus on  $\mathbf{p}^*$  and  $\rho^*$  for subproblems  $\mathcal{P}_1(\mathcal{A}, \rho)$  and  $\mathcal{P}_2(\mathcal{A})$ . The following theorem summarizes the asymptotic SINR and sum power in this case.

Theorem 7 (Asymptotic SINR and sum power for collocated users): For a given  $(\mathcal{A}, \rho, \mathbf{p})$ , if  $\sigma_{km}^2 = \sigma_{1m}^2$ , k = 1, ..., K, m = 1, ..., M, then the followings are true:

1) The system of equation:

$$\xi = \frac{1}{S} \sum_{m \in \mathcal{A}} \frac{(1+\xi) \,\sigma_{1m}^2}{\rho + \rho \xi + \beta \sigma_{1m}^2},$$

admits a unique solution  $\xi$  in  $\mathbb{R}_{++}$ .

2) Define

$$E_{12} = \frac{1}{S} \sum_{m \in \mathcal{A}} \frac{\sigma_{1m}^2}{\left(\rho + \rho\xi + \beta \sigma_{1m}^2\right)^2}, \ E_{22} = \frac{1}{S} \sum_{m \in \mathcal{A}} \frac{\sigma_{1m}^4}{\left(\rho + \rho\xi + \beta \sigma_{1m}^2\right)^2}.$$

As  $K, S \to \infty$  with the ratio  $\beta = K/S$  fixed,  $\gamma_k(\Sigma, \rho, \mathbf{p})$  in (2) and  $\sum_{m=1}^{S} p_m(\Sigma, \rho, \mathbf{p})$  in (6) respectively converge to the following deterministic values

$$\bar{\gamma}_k(\mathbf{\Sigma}, \rho, \mathbf{P}) = \frac{p_k \xi^2}{\frac{\beta E_{22}}{1 - \beta E_{22}} \frac{1}{K} \sum_{l=1}^K p_l + (1 + \xi)^2},$$
(37)

$$\bar{p}_{\mathcal{A}}(\boldsymbol{\Sigma}, \boldsymbol{\rho}, \mathbf{P}) = \frac{\beta E_{12}}{1 - \beta E_{22}} \frac{1}{K} \sum_{l=1}^{K} p_l, \qquad (38)$$

*Remark 1:* Similarly, we can obtain simpler expression for per antenna power  $p_m(\Sigma, \rho, \mathbf{p})$  in (5). Using these simpler asymptotic expressions,  $\mathcal{P}_1(\mathcal{A}, \rho)$  can be reformulated into a simpler form, and the optimal power allocation is also given by a simple *water-filling solution*. The details are omitted due to limited space.

Similarly, using Theorem 7 and assuming we have sum power constraint  $\bar{p}_{\mathcal{A}}(\Sigma, \rho, \mathbf{p}) \leq P_T$  only, subproblem  $\mathcal{P}_1(\mathcal{A}, \rho)$  is equivalent to the following optimization problem:

$$\max_{\mathbf{p} \ge 0} \sum_{k=1}^{K} w_k \log\left(1 + \frac{\xi^2 p_k}{P_T E_{22}/E_{12} + (1+\xi)^2}\right), \text{ s.t. } \frac{\beta E_{12}}{1 - \beta E_{22}} \frac{1}{K} \sum_{l=1}^{K} p_l \le P_T$$

For fixed  $\mathcal{A}, \rho$ , the optimal power allocation  $\mathbf{p}^*(\mathcal{A}, \rho)$  is given by water-filling solution as

$$p_k^*(\mathcal{A},\rho) = \left(\frac{w_k S \left(1 - \beta E_{22}\right)}{\lambda E_{12}} - \frac{P_T E_{22} / E_{12} + \left(1 + \xi\right)^2}{\xi^2}\right)^+,$$

where  $\lambda$  is chosen such that  $\frac{\beta E_{12}}{1-\beta E_{22}} \frac{1}{K} \sum_{l=1}^{K} p_l^*(\mathcal{A}, \rho) = P_T$ . On the other hand, the optimal  $\rho^*$  for  $\mathcal{P}_2(\mathcal{A})$  is given by:

$$\rho^* = \frac{\beta}{P_T},$$



Figure 5. Illustration of asymptotic decoupling effect in a very large distributed MIMO network

where  $\beta = K/S$ . The proof can be extended from the proof in Appendix B of [10] and is omitted due to page limit.

#### C. Very Large Distributed MIMO Network

In this section, we analyze the asymptotic performance for very large distributed MIMO networks (i.e., M is very large). To simplify the analysis, we make the following assumptions.

Assumption 3 (Assumptions on the Very Large Distributed MIMO Network): The coverage area is a square with side length  $R_c$ . There are  $M = N^2$  distributed antennas evenly distributed in the square grid for some integer N. The locations of the K users are randomly generated from a uniformly distribution within the square. Assume that the large scale fading is purely caused by path loss. The pass loss model is given by  $\sigma_{km}^2 = G_0 r_{km}^{-\zeta}$ , where  $G_0 > 0$  is a constant,  $r_{km}$  is the distance between the  $m^{\text{th}}$  antenna and  $k^{\text{th}}$  user,  $\zeta$  is the pass loss factor.

Let  $\tilde{m}_k = \underset{m}{\operatorname{argmax}} \sigma_{km}^2$ . For each user, define  $\bar{g}_k^d = \sigma_{k\bar{m}_k}^2$  as the direct-link gain, and  $\bar{g}_k^c = \underset{l \neq k}{\max} \sigma_{k\bar{m}_l}^2$  as the maximum cross-link gain. Define the ratio  $\eta = \underset{k}{\min} \bar{g}_k^d / \bar{g}_k^c$ , which measures the coupling between the distributed antennas and the users. We have the following theorem

Theorem 8 (Asymptotic Decoupling and Capacity Scaling): For any  $\eta_0 > 0$ , we have

$$\Pr\left(\eta > \eta_0\right) \ge 1 - K\left[1 - \left(1 - \pi\left(\eta_0^{1/\zeta} + 1\right)^2 / (2M)\right)^{K-1}\right].$$
(39)

Furthermore, for any  $\epsilon > 0$ , the maximum achievable weighted sum-rate  $C_w$  almost surely satisfies  $O\left(K\left(\frac{\zeta}{2}-\epsilon\right)\log M\right) \le C_w \le O\left(K\left(\frac{\zeta}{2}+\epsilon\right)\log M\right)$  as  $M \to \infty$  with K, S fixed.

Please refer to Appendix E for the proof.

Remark 2 (Interpretation of Asymptotic Decoupling): For fixed  $\beta = K/S$  and reasonably large M/K, there is a high probability that  $\eta$  is large. This means that there is a large chance that the topology of the selected active antennas and the K users have strong direct-link and weak cross-links as illustrated in Fig. 5. Intuitively, this means that for large M, there is a high chance that each of the K users can find a set of nearby transmit antennas which are relatively far from other users. Due to this decoupling effect in large distributed MIMO system, simplified physical layer processing (such as Matched-Filter pre-coder [2] for each user using the selected antennas nearby) can also achieve good performance.

Corollary 1 (Asymptotic Optimality of Algorithm S3): For reasonably large M/K, Algorithm S3 is asymptotically optimal.

The proof is given in Appendix F.

## VI. NUMERICAL RESULTS

In this section, simulations are used to verify the accuracy of the asymptotic expressions in the paper, and the performance of the proposed algorithms. Consider a Cloud RAN serving K = 8 users lying inside a square with an area of 2Km × 2Km. Assume that the antennas are evenly distributed in the square. Assume the same pass loss model as in Assumption 3, where the pass loss factor is set as  $\zeta = 2.5$ . In the simulation figures, we plot the average weighted sum-rate<sup>2</sup> versus sum power constraint  $P_T$  in (5) for various cases. In all simulations, the per antenna power constraint is set as  $\bar{p}_m = 5$ dB, m = 1, ..., K. From the first user to the eighth user, the weights increase linearly from 0.0625 to 0.1875. The locations of users are randomly generated from a uniform distribution except for Fig. 7.

## A. Asymptotic Expressions and Power Allocation Gain

In Fig. 6, we verify the accuracy of the asymptotic expressions and the performance gain due to power allocation. We plot both the asymptotic and simulated weighted sum-rates averaged over a single

<sup>&</sup>lt;sup>2</sup>The weighted sum-rate shown in the simulations are enlarged by K times since the weights are normalized.



Figure 6. Comparison of asymptotic and simulated weighted sum-rates with and without power allocation

large fading block. The number of antennas S is fixed as 16. The performance without power allocation, i.e.,  $\mathbf{P} = c\mathbf{I}$  in (1), is also given for comparison. It can be seen that the asymptotic approximation is close to the simulated value. When power allocation is considered and optimized, a higher weighted sum-rate can be achieved compared to the case when only the regularization factor  $\rho$  is optimized. Note that as  $P_T$  increases, the weighted sum-rate get saturated due to per antenna power constraint.

#### B. Capacity Gain of the Proposed Scheme w.r.t. Baseline

In Fig. 7 and 8, we compare the performance of the proposed algorithm with the traditional antenna selection baseline algorithm described in Section IV-D. The power allocation and  $\rho$  are optimized in all cases. There are a total number of M = 25 antennas. We plot the weighted sum-rates averaged over different realizations of user locations for S = 8 and 16 respectively<sup>3</sup>. In Fig. 7, we consider strong cross link case, where the user locations are randomly generated but with the restriction that the distance between each user and the nearest antenna must be larger than a threshold. In this case, it can be seen that the proposed scheme achieves significant performance gain compared with the baseline. In Fig. 8, we consider normal case where the users are uniformly distributed. Similar results as in Fig. 7 can be observed, although the performance gain is smaller compared to the strong cross link case.

<sup>&</sup>lt;sup>3</sup>The baseline algorithm can be extended to select  $S \le n \times K$  antennas for some integer n by applying the baseline algorithm for several rounds and deleting the selected antennas from the entire antenna set after each round.



Figure 7. Comparison of proposed antenna selection scheme and baseline for strong cross-link case



Figure 8. Comparison of proposed antenna selection scheme and baseline for uniformly distributed users



Figure 9. Comparison of asymptotic and simulated weighted sum-rates for different choices of transmit antennas

## C. Comparison with Performance Upper and Lower Bounds

In Fig. 9, we compare the proposed antenna selection scheme with some upper and lower bounds. There are a total number of M = 49 antennas and S = 16 of them are selected for transmission. The performance of the following cases are compared. Case 1: All the 49 distributed antennas are used for transmission, which provides an performance upper bound. Case 2: All the 49 antennas are collocated at the BS and are used for transmission. Case 3: There are a total number of 16 antennas evenly distributed in the square and all the 16 antennas are used for transmission, which provides a performance lower bound. We plot the weighted sum-rates averaged over different realizations of user locations. The following advantages of the proposed antenna selection scheme can be observed. 1) It achieves a weighted sum-rate close to the upper bound in Case 1, and higher than Case 2, while the pilot training overhead is lower. 2) The performance is much better than Case 3 due to large antenna gain. Note that as  $P_T$  increases, the weighted sum-rate of the proposed scheme and Case 3 get saturated earlier because in this two cases, the number of active antennas is smaller and the actual total transmit power is smaller when  $P_T$  is large.

## VII. CONCLUSION

We consider downlink antenna selection problem in a large distributed MIMO network with  $M \gg 1$  geographically distributed antennas. Assume that RCI pre-coding is employed. The objective is to max-

imize the average weighted sum-rate under per antenna and sum power constraint by joint optimization of antenna selection, regularization factor, and power allocation based on the knowledge of large scale fading factors. The problem is a mixed combinatorial and non-convex problem. The objective and constraints have no closed-form expressions. We first derive asymptotically accurate expressions for average weighted sum-rate and transmit power. Then the joint optimization problem is decomposed into simpler subproblems and efficient algorithms are proposed to solve them. For the special cases of collocated antennas or collocated users, we obtain structural solution. We also show that the capacity of a very large distributed MIMO network scales according to  $O\left(K\frac{\zeta}{2}\log M\right)$ , where  $\zeta$  is the pass loss factor. Simulations show that the proposed antenna selection scheme provides a very good trade-off between performance and CSI acquisition overhead.

#### APPENDIX

## A. Proof of Lemma 1 and Lemma 2

a) Almost sure convergence of  $\bar{\gamma}_k(\Sigma, \rho, \mathbf{P})$ : We first prove the almost sure convergence of  $A_k$ and  $B_k$ , then the almost sure convergence of  $\bar{\gamma}_k(\Sigma, \rho, \mathbf{P})$  follows immediately.

 $A_k$  can be rewritten as  $A_k = \sigma_k^2 \tilde{\mathbf{h}}_k \left( \hat{\mathbf{H}}_k^{\dagger} \hat{\mathbf{H}}_k + \rho \mathbf{I}_S \right)^{-1} \tilde{\mathbf{h}}_k^{\dagger}$ , where  $\sigma_k^2 = \text{diag} \left( \sigma_{kA_1}^2, ..., \sigma_{kA_S}^2 \right)$ , and the elements of  $\tilde{\mathbf{h}}_k$  are i.i.d. complex random variables with zero mean, unit variance. Then it follows from [18, Corrolary 1] that

$$A_k - \frac{1}{S} \operatorname{Tr} \left( \boldsymbol{\sigma}_k^2 \left( \hat{\mathbf{H}}_k^{\dagger} \hat{\mathbf{H}}_k + \rho \mathbf{I}_S \right)^{-1} \right) \xrightarrow{a.s} 0.$$
(40)

Applying [19, Lemma 2.4] to the trace term in (40), we have  $A_k \xrightarrow{a.s} \xi_{kk}$ , where  $\xi_{kk}$  is defined in (8). Following a similar analysis as the above and by denoting  $\mathbf{Q}_k = \hat{\mathbf{H}}_k^{\dagger} \hat{\mathbf{H}}_k + \rho \mathbf{I}_S$ , we have

$$B_k - \frac{1}{S} \operatorname{Tr} \left( \boldsymbol{\sigma}_k^2 \mathbf{Q}_k^{-1} \hat{\mathbf{H}}_k^{\dagger} \mathbf{P}_k \hat{\mathbf{H}}_k \mathbf{Q}_k^{-1} \right) \xrightarrow{a.s} 0.$$
(41)

Using matrix inverse lemma, the trace term in (41) can be rewritten as

$$\frac{1}{S}\sum_{l\neq k}^{K}\frac{p_{l}}{\left(1+\hat{\mathbf{h}}_{l}\mathbf{Q}_{kl}^{-1}\hat{\mathbf{h}}_{l}^{\dagger}\right)^{2}}\hat{\mathbf{h}}_{l}\mathbf{Q}_{kl}^{-1}\boldsymbol{\sigma}_{k}^{2}\mathbf{Q}_{kl}^{-1}\hat{\mathbf{h}}_{l}^{\dagger},$$

where  $\mathbf{Q}_{kl} = \hat{\mathbf{H}}_{kl}^{\dagger} \hat{\mathbf{H}}_{kl} + \rho \mathbf{I}_S$  and  $\hat{\mathbf{H}}_{kl}$  is the matrix of  $\hat{\mathbf{H}}$  where the  $k^{\text{th}}$  and  $l^{\text{th}}$  rows are removed. Applying [18, Corrolary 1], [20, Lemma 2.1] and [19, Lemma 2.4] one by one, we have

$$\hat{\mathbf{h}}_{l} \mathbf{Q}_{kl}^{-1} \hat{\mathbf{h}}_{l}^{\dagger} \xrightarrow{a.s.} \frac{1}{S} \operatorname{Tr} \left( \boldsymbol{\sigma}_{l}^{2} \mathbf{Q}_{kl}^{-1} \right) \longrightarrow \frac{1}{S} \operatorname{Tr} \left( \boldsymbol{\sigma}_{l}^{2} \mathbf{Q}_{k}^{-1} \right) \xrightarrow{a.s.} \xi_{kl}.$$

$$(42)$$

By [18, Corrolary 1], we have

$$\hat{\mathbf{h}}_{l} \mathbf{Q}_{kl}^{-1} \boldsymbol{\sigma}_{k}^{2} \mathbf{Q}_{kl}^{-1} \hat{\mathbf{h}}_{l}^{\dagger} \xrightarrow{a.s} \frac{1}{S} \operatorname{Tr} \left( \boldsymbol{\sigma}_{l}^{2} \mathbf{Q}_{kl}^{-1} \boldsymbol{\sigma}_{k}^{2} \mathbf{Q}_{kl}^{-1} \right)$$

$$= -\frac{1}{S} \frac{\partial}{\partial z} \operatorname{Tr} \left( \boldsymbol{\sigma}_{l}^{2} \left( \hat{\mathbf{H}}_{kl}^{\dagger} \hat{\mathbf{H}}_{kl} + \rho \mathbf{I}_{S} + z \boldsymbol{\sigma}_{k}^{2} \right)^{-1} \right) \Big|_{z=0}$$

$$\longrightarrow -\frac{1}{S} \frac{\partial}{\partial z} \operatorname{Tr} \left( \boldsymbol{\sigma}_{l}^{2} \left( \hat{\mathbf{H}}_{k}^{\dagger} \hat{\mathbf{H}}_{k} + \rho \mathbf{I}_{S} + z \boldsymbol{\sigma}_{k}^{2} \right)^{-1} \right) \Big|_{z=0}.$$

$$(43)$$

where the last line follows from [20, Lemma 2.1]. Applying [19, Lemma 2.4] to the last trace term in (43), and calculate the partial derivative over z, it can be shown that

$$\hat{\mathbf{h}}_{l} \mathbf{Q}_{kl}^{-1} \boldsymbol{\sigma}_{k}^{2} \mathbf{Q}_{kl}^{-1} \hat{\mathbf{h}}_{l}^{\dagger} \xrightarrow{a.s} \theta_{kl}, \tag{44}$$

where  $\theta_{kl}$  is defined in (10). Combining the above results, we have  $B_k \xrightarrow{a.s} -\frac{1}{S} \sum_{l \neq k}^{K} \left[ p_l \theta_{kl} / (1 + \xi_{kl})^2 \right]$ .

b) Almost sure convergence of  $p_m(\mathbf{\Sigma}, \rho, \mathbf{p})$ : Without loss of generality, we assume  $m = \mathcal{A}_j$ . By denoting  $\tilde{\mathbf{Q}}_j = \left(\hat{\mathbf{H}}_j^c \hat{\mathbf{H}}_j^{c\dagger} + \rho \mathbf{I}_K\right)^{-1}$ , where  $\hat{\mathbf{H}}_j^c$  is the matrix of  $\hat{\mathbf{H}}$  where the  $j^{\text{th}}$  column is removed, and applying matrix inverse lemma to (3), we have

$$p_{\mathcal{A}_j}\left(\mathbf{\Sigma}, \boldsymbol{\rho}, \mathbf{p}\right) = \frac{1}{S} \hat{\mathbf{g}}_j^{\dagger} \tilde{\mathbf{Q}}_j^{-1} \mathbf{P} \tilde{\mathbf{Q}}_j^{-1} \hat{\mathbf{g}}_j \left(1 + \hat{\mathbf{g}}_j^{\dagger} \tilde{\mathbf{Q}}_j^{-1} \hat{\mathbf{g}}_j\right)^{-2}$$

Following similar analysis for the term  $\hat{\mathbf{h}}_l \mathbf{Q}_{kl}^{-1} \boldsymbol{\sigma}_k^2 \mathbf{Q}_{kl}^{-1} \hat{\mathbf{h}}_l^{\dagger}$  in (41), we have

$$\hat{\mathbf{g}}_{j}^{\dagger} \tilde{\mathbf{Q}}_{j}^{-1} \mathbf{P} \tilde{\mathbf{Q}}_{j}^{-1} \hat{\mathbf{g}}_{j} \xrightarrow{a.s} \frac{\rho^{-1}}{K} \sum_{i=1}^{K} \sigma_{i\mathcal{A}_{j}}^{2} \left( p_{i} v_{i} - \varphi_{i} \right),$$

where  $v_i$  is defined in (12) and  $\varphi_i$  is defined in (15). Following similar analysis for the term  $\hat{\mathbf{h}}_l \mathbf{Q}_{kl}^{-1} \hat{\mathbf{h}}_l^{\dagger}$  in (42), we have

$$\hat{\mathbf{g}}_{j}^{\dagger} \tilde{\mathbf{Q}}_{j}^{-1} \hat{\mathbf{g}}_{j} \xrightarrow{a.s} \frac{1}{S} \sum_{i=1}^{K} \sigma_{i\mathcal{A}_{j}}^{2} v_{i}$$

Combining the above results, we show that  $p_m(\mathbf{\Sigma}, \rho, \mathbf{p}) \xrightarrow{a.s} \bar{p}_m(\mathbf{\Sigma}, \rho, \mathbf{p})$  in (16).

Finally,  $\mathbf{I}_K - \mathbf{D}_k$  is invertible because the spectral radius of  $\mathbf{D}_k$  is less than 1 as  $K, S \to \infty$  with the ratio  $\beta = K/S$  fixed. And  $(\rho \mathbf{I}_K + \mathbf{\Delta} - \mathbf{C})^{-1}$  is invertible because it is a diagonally dominant matrix. This completes the proof of Lemma 1 and Lemma 2.

# B. Proof of Theorem 3

Under condition (27), it can be shown that  $\nabla_{\vec{\lambda}} L\left(\vec{\lambda}, \tilde{\mathbf{p}}\left(\vec{\lambda}\right)\right)$  exists and is given by  $\nabla_{\vec{\lambda}} L\left(\vec{\lambda}, \tilde{\mathbf{p}}\left(\vec{\lambda}\right)\right) = \bar{\mathbf{p}} - \mathbf{R}\tilde{\mathbf{p}}$ . Then (23) is a gradient projection update for solving the problem  $\min_{\vec{\lambda}} L\left(\vec{\lambda}, \tilde{\mathbf{p}}\left(\vec{\lambda}\right)\right)$ , s.t.  $\vec{\lambda} \ge 0$ , and thus Algorithm S1 must converge to a stationary point  $\tilde{\lambda}$  of this problem [21]. Finally, it can be verified that  $\tilde{\mathbf{p}}\left(\tilde{\lambda}\right)$  is a stationary point of  $\mathcal{P}_1(\mathcal{A}, \rho)$ . The details are skipped due to page limit.

C. Calculation of the Derivative  $\frac{\partial \hat{I}(\mathcal{A},\rho)}{\partial \rho}$ 

For convenience, define two (S + K + 1)-dimensional vectors

$$\bar{\mathbf{p}}_{\text{ext}} = \begin{bmatrix} \bar{\mathbf{p}} \\ \mathbf{0} \end{bmatrix}, \ \tilde{\lambda}_{\text{ext}} = \begin{bmatrix} \tilde{\lambda} \\ -\tilde{\nu} \end{bmatrix}$$

Define a  $(S + K + 1) \times K$  matrix  $\mathbf{R}_{ext} \triangleq \begin{bmatrix} \mathbf{R}^T, -\mathbf{I}_K \end{bmatrix}^T$ . Define a vector  $\mathbf{e} \in \mathbb{R}^K$  whose  $k^{\text{th}}$  element is

$$e_{k} = \sum_{l=1}^{K} \left( \frac{w_{l}g_{lk} \sum_{i=1}^{K} \tilde{p}_{i}\left(\mathcal{A},\rho\right) \frac{\partial g_{li}}{\partial \rho}}{\left(g_{ll}\tilde{p}_{l}\left(\mathcal{A},\rho\right) + \tilde{\Omega}_{l}\right)^{2}} - \frac{w_{l} \frac{\partial g_{lk}}{\partial \rho}}{g_{ll}\tilde{p}_{l}\left(\mathcal{A},\rho\right) + \tilde{\Omega}_{l}} \right) + \sum_{l \neq k}^{K} \left( \frac{w_{l} \frac{\partial g_{lk}}{\partial \rho}}{\tilde{\Omega}_{l}} - \frac{w_{l}g_{lk} \sum_{i \neq l}^{K} \tilde{p}_{i}\left(\mathcal{A},\rho\right) \frac{\partial g_{li}}{\partial \rho}}{\tilde{\Omega}_{l}^{2}} \right).$$

Define a  $K \times K$  matrix G whose element at the  $k^{\text{th}}$  row and  $l^{\text{th}}$  column is

$$G_{kl} = \sum_{l=1}^{K} \frac{-w_l g_{lk} g_{ln}}{\left(1 + \sum_{i=1}^{K} g_{li} \tilde{p}_i (\mathcal{A}, \rho)\right)^2} + \sum_{l \neq k, n} \frac{w_l g_{lk} g_{ln}}{\left(1 + \sum_{i \neq l}^{K} g_{li} \tilde{p}_i (\mathcal{A}, \rho)\right)^2}$$

Finally, define a  $(2K + S + 1) \times (2K + S + 1)$  matrix

$$\mathbf{G}_{\text{ext}} = \begin{bmatrix} \mathbf{G}; & -\mathbf{R}_{\text{ext}}^T \\ \text{diag}\left(\tilde{\lambda}_{\text{ext}}\right) \mathbf{R}_{\text{ext}}; & \text{diag}\left(\mathbf{R}_{\text{ext}}\tilde{\mathbf{p}}\left(\mathcal{A}, \rho\right) - \bar{\mathbf{p}}_{\text{ext}}\right) \end{bmatrix}.$$

Taking partial derivative of the equations in (31) with respect to  $\rho$ , we obtain the following linear equations

$$\mathbf{G}_{\text{ext}} \begin{bmatrix} \frac{\partial \tilde{\mathbf{p}}(\mathcal{A}, \rho)}{\partial \rho} \\ \frac{\partial \tilde{\lambda}_{\text{ext}}}{\partial \rho} \end{bmatrix} = \begin{bmatrix} \left( \begin{pmatrix} \partial \mathbf{R} \\ \partial \rho \end{pmatrix}^T \tilde{\lambda} + \mathbf{e} \\ -\text{diag} \left( \tilde{\lambda}_{\text{ext}} \right) \left( \frac{\partial \mathbf{R}_{\text{ext}}}{\partial \rho} \right)^T \tilde{\mathbf{p}} \left( \mathcal{A}, \rho \right) \end{bmatrix}.$$

Then we can obtain  $\frac{\partial \tilde{\mathbf{p}}(\mathcal{A}, \rho)}{\partial \rho}$  by solving the above linear equations.

Define  $\mathcal{J}_1 = \{j : \bar{p}_{\mathcal{A}_j} (\Sigma, \rho, \tilde{\mathbf{p}}(\mathcal{A}, \rho)) < \bar{p}_{\mathcal{A}_j}\}$  and  $\mathcal{K} = \{k : \tilde{p}_k (\mathcal{A}, \rho) > 0\}$ . If  $\bar{p}_{\mathcal{A}} (\Sigma, \rho, \tilde{\mathbf{p}}(\mathcal{A}, \rho)) < P_T$ , let  $\mathcal{J} = \mathcal{J}_1 \cup \{S+1\}$ , otherwise, let  $\mathcal{J} = \mathcal{J}_1$ . Note that we have  $\tilde{\lambda}_j = 0, \forall j \in \mathcal{J}$  and  $\tilde{\nu}_k = 0, \forall k \in \mathcal{K}$  according to the KKT conditions. It can be verified that  $\frac{\partial \tilde{\lambda}_j}{\partial \rho} = 0, \forall j \in \mathcal{J}$  and  $\frac{\partial \tilde{\nu}_k}{\partial \rho} = 0, \forall k \in \mathcal{K}$ . Therefore, we can delete these  $|\mathcal{J}| + |\mathcal{K}|$  variables and the corresponding linear equations whose index *i* satisfies  $i - K \in \mathcal{J}$  or  $i - S - K - 1 \in \mathcal{K}$ . The remaining  $2K + S + 1 - |\mathcal{J}| - |\mathcal{K}|$  variables can be determined by the remaining linear equations. After obtaining  $\frac{\partial \tilde{p}(\mathcal{A}, \rho)}{\partial \rho}$ , the derivative  $\frac{\partial \hat{I}(\mathcal{A}, \rho)}{\partial \rho}$  can be calculated using (32).

To complete the calculation of  $\frac{\partial \hat{I}(\mathcal{A},\rho)}{\partial \rho}$ , we still need to obtain  $\frac{\partial g_{kl}}{\partial \rho}$ ,  $\forall k, l$ , and  $\frac{\partial \mathbf{R}}{\partial \rho}$ . The following Lemma are useful and can be proved by a direct calculation.

Lemma 3 (Derivatives of the intermediate variables): For the intermediate variables  $\vec{\theta}_k$ , v,  $\Delta$  and C defined in Lemma 1 and Lemma 2, the partial derivatives of them with respective to  $\rho$  are given below.

$$\frac{\partial \vec{\theta}_k}{\partial \rho} = (\mathbf{I}_K - \mathbf{D}_k)^{-1} \left( \frac{\partial \mathbf{d}_k}{\partial \rho} + \frac{\partial \mathbf{D}_k}{\partial \rho} \vec{\theta} \right), \, \forall k,$$
(45)

where  $\frac{\partial \mathbf{d}_k}{\partial \rho}$  is given by

$$\frac{\partial d_{kl}}{\partial \rho} = \frac{2}{S} \sum_{m \in \mathcal{A}} \left[ \sigma_{km}^2 \sigma_{lm}^2 \left( 1 - \frac{1}{S} \sum_{i \neq k}^K \frac{\sigma_{im}^2 \phi_{ki}}{(1 + \xi_{ki})^2} \right) / f_m^3 \left( \vec{\xi}_k \right) \right], \,\forall l$$

and  $\frac{\partial \mathbf{D}_k}{\partial \rho}$  is given by

$$\frac{\partial D_{k,ln}}{\partial \rho} = \frac{1}{S^2} \sum_{m \in \mathcal{A}} \left[ \frac{2\sigma_{lm}^2 \sigma_{nm}^2}{(1+\xi_{kn})^2} \left[ \frac{1}{S} \sum_{i \neq k}^K \frac{\sigma_{im}^2}{1+\xi_{ki}} \left( \frac{\phi_{ki}}{1+\xi_{ki}} - \frac{\phi_{kn}}{1+\xi_{kn}} \right) - 1 - \frac{\rho \phi_{kn}}{1+\xi_{kn}} \right] / f_m^3 \left( \vec{\xi}_k \right) \right].$$

$$\frac{\partial C_{ln}}{\partial \rho} = \frac{1}{S} \sum_{m \in \mathcal{A}} \left[ \frac{1}{S} \sigma_{lm}^2 \sigma_{nm}^2 \frac{\partial v_l}{\partial \rho} / h_m^2 \left( \mathbf{v} \right) - \frac{2}{S} \sigma_{lm}^2 \sigma_{nm}^2 v_l \frac{1}{S} \sum_{i=1}^K \sigma_{im}^2 \frac{\partial v_i}{\partial \rho} / h_m^3 \left( \mathbf{v} \right) \right], \ \forall l, n,$$
(46)

$$\frac{\partial \Delta_l}{\partial \rho} = -\frac{1}{S} \sum_{m \in \mathcal{A}} \left[ \sigma_{lm}^2 \frac{1}{S} \sum_{i=1}^K \sigma_{im}^2 \frac{\partial v_i}{\partial \rho} / h_m^2 \left( \mathbf{v} \right) \right], \, \forall l,$$
(47)

$$\frac{\partial v_l}{\partial \rho} = -\left(\rho \mathbf{I}_K + \boldsymbol{\Delta} - \mathbf{C}\right)^{-1} \mathbf{v}, \,\forall l.$$
(48)

Using Lemma 3,  $\frac{\partial g_{kl}}{\partial \rho}$ ,  $\forall k, l$ , and  $\frac{\partial \mathbf{R}}{\partial \rho}$  can be obtained by a direct calculation as follows.

$$\frac{\partial g_{kk}}{\partial \rho} = \frac{2\xi_{kk}\frac{\partial\xi_{kk}}{\partial \rho}}{(1+\xi_{kk})^3}, \forall k,$$

$$\frac{\partial g_{kl}}{\partial \rho} = -\frac{\frac{\partial\theta_{kl}}{\partial \rho}}{S\left(1+\xi_{kl}\right)^2\left(1+\xi_{kk}\right)^2} + \frac{2\theta_{kl}\left[\frac{\partial\xi_{kl}}{\partial \rho}\left(1+\xi_{kk}\right) + \frac{\partial\xi_{kk}}{\partial \rho}\left(1+\xi_{kl}\right)\right]}{S\left(1+\xi_{kl}\right)^3\left(1+\xi_{kk}\right)^3}, \forall k \neq l,$$

where  $\frac{\partial \xi_{kk}}{\partial \rho} = \phi_{kk}$ ,  $\forall k$  is defined in (9),  $\vec{\theta}_k = [\theta_{k1}, ..., \theta_{kK}]^T$ ,  $\forall k$  is defined in (10) and  $\frac{\partial \vec{\theta}_k}{\partial \rho}$  is given in (45). To calculate  $\frac{\partial \mathbf{R}}{\partial \rho}$ , we first obtain  $\frac{\partial \hat{\mathbf{R}}}{\partial \rho}$  as

$$\frac{\partial \hat{R}_{kj}}{\partial \rho} = -\frac{\sigma_{k\mathcal{A}_j}^2 \rho^{-1}}{S^2} \left[ \rho^{-1} / h_{\mathcal{A}_j}^2 \left( \mathbf{v} \right) + \frac{2}{S} \sum_{i=1}^K \sigma_{i\mathcal{A}_j}^2 \frac{\partial v_i}{\partial \rho} / h_{\mathcal{A}_j}^3 \left( \mathbf{v} \right) \right], \ \forall k, j.$$

where  $\frac{\partial v_i}{\partial \rho}$ ,  $\forall i$  is given in (48). Then we obtain  $\frac{\partial \bar{\mathbf{R}}}{\partial \rho}$  as

$$\begin{split} \frac{\partial \bar{R}_{kk}}{\partial \rho} &= \frac{1}{S} \sum_{m \in \mathcal{A}} \left[ \left( \sigma_{km}^2 \frac{\partial v_k}{\partial \rho} + \sigma_{km}^2 \frac{1}{S} \sum_{i \neq k}^K \sigma_{im}^2 \left( v_k \frac{\partial v_i}{\partial \rho} + v_i \frac{\partial v_k}{\partial \rho} \right) \right) / h_m^2 \left( \mathbf{v} \right) \\ &- \left( 2\sigma_{km}^2 v_k \left( 1 + \frac{1}{S} \sum_{i \neq k}^K \sigma_{im}^2 v_i \right) \frac{1}{S} \sum_{i=1}^K \sigma_{im}^2 \frac{\partial v_i}{\partial \rho} \right) / h_m^3 \left( \mathbf{v} \right) \right], \forall k, \\ \frac{\partial \bar{R}_{kl}}{\partial \rho} &= -\frac{1}{S} \sum_{m \in \mathcal{A}} \left[ \sigma_{km}^2 \sigma_{lm}^2 \frac{1}{S} \left( v_l \frac{\partial v_k}{\partial \rho} + v_k \frac{\partial v_l}{\partial \rho} \right) / h_m^2 \left( \mathbf{v} \right) - \frac{2}{S} \sigma_{km}^2 \sigma_{lm}^2 v_k v_l \frac{1}{S} \sum_{i=1}^K \sigma_{im}^2 \frac{\partial v_i}{\partial \rho} / h_m^3 \left( \mathbf{v} \right) \right], \forall k \neq l. \end{split}$$

Finally,  $\frac{\partial \mathbf{R}}{\partial \rho} = \begin{bmatrix} \mathbf{I}_S, & \mathbf{1} \end{bmatrix}^T \frac{\partial \tilde{\mathbf{R}}}{\partial \rho}$  and  $\frac{\partial \tilde{\mathbf{R}}}{\partial \rho}$  is given by  $\frac{\partial \tilde{\mathbf{R}}}{\partial \rho} = \left(\frac{\partial \hat{\mathbf{R}}}{\partial \rho}\right)^T \left(\mathbf{V} - (\rho \mathbf{I}_K + \mathbf{\Delta} - \mathbf{C})^{-1} \bar{\mathbf{R}}\right) + \hat{\mathbf{R}}^T \left(\frac{\partial \mathbf{V}}{\partial \rho} - (\rho \mathbf{I}_K + \mathbf{\Delta} - \mathbf{C})^{-1} \frac{\partial \bar{\mathbf{R}}}{\partial \rho}\right) \\
+ \hat{\mathbf{R}}^T \left(\mathbf{I}_K + \frac{\partial \mathbf{\Delta}}{\partial \rho} - \frac{\partial \mathbf{C}}{\partial \rho}\right) (\rho \mathbf{I}_K + \mathbf{\Delta} - \mathbf{C})^{-2} \bar{\mathbf{R}},$ 

where  $\frac{\partial \mathbf{C}}{\partial \rho}$  and  $\frac{\partial \mathbf{\Delta}}{\partial \rho}$  are given in (46) and (47) respectively, and  $\frac{\partial \mathbf{V}}{\partial \rho} = \text{diag} \left[ \frac{\partial v_1}{\partial \rho}, ..., \frac{\partial v_K}{\partial \rho} \right]^T$ .

# D. Proof of Theorem 6

When  $P'_T$  is large enough, all users will be allocated with positive power. In this case, the SINR of user k under power allocation in (36) is given by

$$\hat{\gamma}_{k}\left(\rho, \mathbf{p}^{*}\left(\mathcal{A}, \rho\right)\right) = \frac{Sw_{k}\left(1 + \sigma_{k1}^{2}u\right)^{2}\sigma_{k1}^{2}\left(2\rho u P_{T}^{\prime} + \left(\beta - 1\right)P_{T}^{\prime} + \frac{1}{S}\sum_{l=1}^{K}\frac{1}{\sigma_{l1}^{2}}\right)}{\sigma_{k1}^{2}P_{T}^{\prime} + \left(1 + \sigma_{k1}^{2}u\right)^{2}} - 1, \qquad (49)$$

and the objective of  $\mathcal{P}_2(\mathcal{A})$  is given by  $\overline{\mathcal{I}}(\mathcal{A}, \rho, \mathbf{p}^*(\mathcal{A}, \rho)) = \sum_{k=1}^K w_k \log(1 + \hat{\gamma}_k(\rho, \mathbf{p}^*(\mathcal{A}, \rho))))$ . For any k, it can be shown that the solution  $\tilde{\rho}_k$  of  $\frac{\partial}{\partial \rho} \hat{\gamma}_k(\rho, \mathbf{p}^*(\mathcal{A}, \rho)) = 0$  must satisfy  $\tilde{\rho}_k = O\left(\frac{1}{P_T'}\right)$ . Since the optimal regularization factor  $\rho^*$  must satisfy  $\min_k \tilde{\rho}_k \leq \rho^* \leq \max_k \tilde{\rho}_k$ , we have  $\rho^* = O\left(\frac{1}{P_T'}\right)$ . To prove the second result, it can be verified that  $\frac{\partial^2}{\partial^2 \rho} \hat{\gamma}_k(\rho, \mathbf{p}^*(\mathcal{A}, \rho)) < 0$  and thus  $\hat{\gamma}_k(\rho, \mathbf{p}^*(\mathcal{A}, \rho))$  is concave when  $\rho$  is small enough. Since  $\overline{\mathcal{I}}(\mathcal{A}, \rho, \mathbf{p}^*(\mathcal{A}, \rho))$  is a concave increasing function of  $\hat{\gamma}_k(\rho, \mathbf{p}^*(\mathcal{A}, \rho))$ ,  $\overline{\mathcal{I}}(\mathcal{A}, \rho, \mathbf{p}^*(\mathcal{A}, \rho))$  must be a concave function of  $\rho$  [16].

## E. Proof of Theorem 8

We first derive the lower bound for the probability that the minimum distance  $r_{\min}$  between any two users is larger than a certain value  $r_0$ :  $\Pr(r_{\min} \ge r_0)$ . Let  $d_{kl}^u$  denote the distance between user k and user l. We have

$$\begin{aligned} \Pr\left(r_{\min} \ge r_{0}\right) &= 1 - \Pr\left(\min_{l \ne k} d_{kl}^{u} \le r_{0}, \ \exists k \in \{1, ..., K\}\right) \\ &\ge 1 - \sum_{k=1}^{K} \Pr\left(\min_{l \ne k} d_{kl}^{u} \le r_{0}\right) \\ &= 1 - K \Pr\left(\min_{l \ne 1} d_{1l}^{u} \le r_{0}\right) \\ &= 1 - K \left[1 - (\Pr\left(d_{12}^{u} \ge r_{0}\right))^{K-1}\right] \\ &\ge 1 - K \left[1 - \left(1 - \frac{\pi r_{0}^{2}}{R_{c}^{2}}\right)^{K-1}\right], \end{aligned}$$

where the second inequality follows from the union bound and the last inequality holds because  $\Pr(d_{12}^u \ge r_0) \le 1 - \frac{\pi r_0^2}{R_c^2}.$ 

Then we use the path loss model to transfer the probability  $\Pr(r_{\min} \ge r_0)$  to the probability  $\Pr(\eta > \eta_0)$ in (39). Note that for any k, we have  $\min_m r_{km} \le \frac{\sqrt{2}R_c}{2\sqrt{M}}$ , and  $\max_l r_{k\tilde{m}_l} \ge r_{\min} - \frac{\sqrt{2}R_c}{2\sqrt{M}}$ . Hence  $\eta \ge \left(r_{\min} / \left(\frac{\sqrt{2}R_c}{2\sqrt{M}}\right) - 1\right)^{\zeta}$  and

$$\begin{aligned} \Pr\left(\eta > \eta_{0}\right) &\geq & \Pr\left(\left(r_{\min} / \left(\frac{\sqrt{2}R_{c}}{2\sqrt{M}}\right) - 1\right)^{\zeta} > \eta_{0}\right) \\ &= & \Pr\left(r_{\min} > \frac{\sqrt{2}R_{c}}{2\sqrt{M}} \left(\eta_{0}^{1/\zeta} + 1\right)\right) \\ &\geq & 1 - K\left[1 - \left(1 - \pi \left(\eta_{0}^{1/\zeta} + 1\right)^{2} / (2M)\right)^{K-1}\right] \end{aligned}$$

Finally, we prove the capacity scaling law by deriving an upper and a lower bound for the achievable weightd sum-rate. The following lemma is useful for deriving the upper bound.

Lemma 4: For any  $\epsilon>0,$  as  $M\to\infty$  with K,S fixed, we have

$$\Pr\left(\min_{k,m} r_{km} \le M^{-\frac{1}{2}-\epsilon}\right) = \frac{\pi M^{-2\epsilon}}{R_c^2} \to 0,$$

and thus

$$\Pr\left(\bar{g}_k^d > G_0 M^{\frac{\zeta}{2} + \epsilon}\right) \to 0.$$

Let  $P_T^U = \max\left(P_T, \max_{m \in \mathcal{A}} S\bar{p}_m\right)$  and let  $X_S$  denote a random variable with  $\chi^2(2S)$  distribution. Assuming that each user is severed by S antennas without interference from other users, we obtain an upper bound for average weightd sum-rate as follows:

$$C_{w} \leq K \mathbb{E} \left[ \log \left( 1 + P_{T}^{U} \bar{g}_{k}^{d} X_{S} \right) \right]$$
  
$$\leq K \log \left( 1 + P_{T}^{U} \bar{g}_{k}^{d} \mathbb{E} \left[ X_{S} \right] \right).$$
(50)

Combining (50) with Lemma 4, we prove that  $C_w \leq O\left(K\left(\frac{\zeta}{2} + \epsilon\right)\log M\right)$  holds almost surely as  $M \to \infty$  with K, S fixed.

Furthermore, it follows from the lower bound provided in Appendix F that  $C_w \ge O\left(K\left(\frac{\zeta}{2} - \epsilon\right)\log M\right)$ . This completes the proof of Theorem 8.

# F. Proof of Corollary 1

Due to (39) in Theorem 8, the step 1 in Algorithm S3 will almost surely select a set of antennas  $\mathcal{A}$ ,  $|\mathcal{A}| = K$  such that each user has strong direct-link with one of the selected K antennas and weak

cross-links with other selected antennas for large M/K. Assume that each selected antenna only severs the nearest user, and assume equal power allocation for each user, i.e.,  $p_k = \min\left(P_T/K, \min_{m \in \mathcal{A}} \bar{p}_m\right), k =$ 1, ..., K. Let  $X_m$  denote a random variable with  $\chi^2(2m)$  distribution. Let  $\eta_0 = M^{\frac{\zeta}{2}-\epsilon_1}$  in (39). Then using (39) and the fact that  $\bar{g}_k^d \ge G_0\left(\frac{\sqrt{2}R_c}{2\sqrt{M}}\right)^{-\zeta}$ , we can show that as  $M \to \infty$  with K, S fixed, the average weighted sum-rate  $\mathcal{I}_{\mathcal{A}}$  is almost surely lower bounded by

$$\mathcal{I}_{\mathcal{A}} \stackrel{a.s}{\geq} KE\left[\log\left(1 + \frac{p_1 G_0 \left(\frac{\sqrt{2}R_c}{2\sqrt{M}}\right)^{-\zeta} X_1}{1 + p_1 M^{-\frac{\zeta}{2} + \epsilon_1} G_0 \left(\frac{\sqrt{2}R_c}{2\sqrt{M}}\right)^{-\zeta} X_{K-1}}\right)\right],\tag{51}$$

where  $X_1$  and  $X_{K-1}$  are independent. Choose  $B_1 > 0$  and  $B_2 > 0$  such that  $\Pr(X_1 \ge B_1) \Pr(X_{K-1} \le B_2) \ge 1 - \epsilon_2$ . Then as  $M \to \infty$  with K, S fixed, it follows from (51) that

$$\begin{aligned} \mathcal{I}_{\mathcal{A}} & \stackrel{a.s}{\geq} \quad K\left(1-\epsilon_{2}\right)\log\left(1+\frac{p_{1}G_{0}\left(\frac{\sqrt{2}R_{c}}{2\sqrt{M}}\right)^{-\zeta}B_{1}}{1+p_{1}M^{-\frac{\zeta}{2}+\epsilon_{1}}G_{0}\left(\frac{\sqrt{2}R_{c}}{2\sqrt{M}}\right)^{-\zeta}B_{2}}\right) \\ &= O\left(K\left(1-\epsilon_{2}\right)\left(\frac{\zeta}{2}-\epsilon_{1}\right)\log M\right). \end{aligned}$$

Choose  $\epsilon_1, \epsilon_2$  such that  $\epsilon_1 + \frac{\zeta}{2}\epsilon_2 - \epsilon_1\epsilon_2 = \epsilon$ . Then we have  $\mathcal{I}_{\mathcal{A}} \stackrel{a.s}{\geq} O\left(K\left(\frac{\zeta}{2} - \epsilon\right)\log M\right)$  as  $M \to \infty$  with K, S fixed. The rest steps in Algorithm S3 only increase the weighted sum-rate by a constant gain. This completes the proof.

#### REFERENCES

- [1] T. Marzetta, "How much training is required for multiuser MIMO?" Fortieth Asilomar Conf. on Signals, Systems, and Computers, Pacific Grove, CA, pp. 359 363, Oct. 2006.
- [2] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Scaling up MIMO: Opportunities and challenges with very large arrays," *to appear in IEEE Signal Processing Magazine*, 2012. [Online]. Available: http://arxiv.org/abs/1201.3210
- [3] A. Moustakas, S. Simon, and A. Sengupta, "Mimo capacity through correlated channels in the presence of correlated interferers and noise: a (not so) large n analysis," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2545 2561, oct. 2003.
- [4] A. Tulino, A. Lozano, and S. Verdu, "Impact of antenna correlation on the capacity of multiantenna channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 7, pp. 2491 2509, july 2005.
- [5] B. Hochwald and S. Vishwanath, "Space-time multiple access: Linear growth in the sum rate," Proc. Allerton Conf. on Commun., Control, and Computing, vol. 40, no. 1, pp. 387 – 396, 2002.
- [6] Y.-C. Liang, S. Sun, and C. K. Ho, "Block-iterative generalized decision feedback equalizers for large MIMO systems: algorithm design and asymptotic performance analysis," *IEEE Trans. Signal Processing*, vol. 54, no. 6, pp. 2035 – 2048, june 2006.
- [7] K. Vishnu Vardhan, S. Mohammed, A. Chockalingam, and B. Sundar Rajan, "A low-complexity detector for large mimo systems and multicarrier cdma systems," *IEEE J. Select. Areas Commun.*, vol. 26, no. 3, pp. 473 – 485, april 2008.

- [8] T. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590 3600, Nov. 2010.
- [9] J. Jose, A. Ashikhmin, T. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2640 – 2651, August 2011.
- [10] R. Muharar and J. Evans, "Downlink beamforming with transmit-side channel correlation: A large system analysis," *ICC 2011*, pp. 1 5, june 2011.
- [11] S. Wagner, R. Couillez, M. Debbah, and D. T. M. Slock, "Large system analysis of linear precoding in MISO broadcast channels with limited feedback," *submitted to IEEE Trans. Inf. Theory*. [Online]. Available: http://arxiv.org/abs/0906.3682
- [12] C. Peel, B. Hochwald, and A. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication-part i: channel inversion and regularization," *IEEE Trans. Commun.*, vol. 53, no. 1, pp. 195 – 202, Jan. 2005.
- [13] A. Molisch, "Mimo systems with antenna selection an overview," Proc. RAWCON'03, pp. 167 170, Aug. 2003.
- [14] I. Berenguer, X. Wang, and V. Krishnamurthy, "Adaptive MIMO antenna selection via discrete stochastic optimization," *IEEE Trans. Signal Processing*, vol. 53, no. 11, pp. 4315 – 4329, nov. 2005.
- [15] Huawei, "Cloud RAN introduction," The 4th CJK International Workshop, Sep. 2011.
- [16] S. Boyd and L. Vandenberghe, Convex Optimization. Cambridge University Press, 2004.
- [17] C. Shi, R. A. Berry, and M. L. Honig, "Monotonic convergence of distributed interference pricing in wireless networks," *in Proc. IEEE ISIT, Seoul, Korea*, June 2009.
- [18] J. Evans and D. Tse, "Large system performance of linear multiuser receivers in multipath fading channels," *IEEE Trans. Inf. Theory*, vol. 46, no. 6, pp. 2059 –2078, sep 2000.
- [19] W. Hachem, P. Loubaton, and J. Najim, "A clt for information-theoretic statistics of gram random matrices with a given variance profile," *Ann. Appl. Probab.*, vol. 18, no. 6, pp. 2071 – 2130, 2008.
- [20] Z. Bai and J. Silverstein, "On the signal-to-interference ratio of cdma systems in wireless communications," Ann. Appl. Probab., vol. 17, no. 1, p. 81, 2007.
- [21] D. P. Bertsekas, "Nonlinear programming," Belmont, MA: Athena Scientific, 1995.