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 zero-forcing precoding. 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 closed-form 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, Cloud Radio Access Networks (C-RAN), 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 $O\left(1/M^2\right)$ [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 zero-forcing 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. In [10], [11], random matrix theory has been used to analyze the asymptotic performance of zero-forcing (ZF) and/or regularized ZF (RZF) [12], with focus on the case when all the antennas are collocated at a BS. In [13], the authors provided a unified large system analysis of the performance of linear detectors/precoders in multicell multiuser TDD systems under a general channel model.

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. This distributed network is also called the Cloud-RAN [14]. In such a scenario, it is likely that only a few nearby antennas (thin base stations) can contribute significantly to a user's communication due to path loss. To avoid expensive CSI acquisition and signal processing overheads for antennas with huge path loss to the users, a subset of S active antennas is selected to serve a given set of K users using RZF precoding, where $M \gg S \geq K$. RZF precoding is considered because it is asymptotically optimal for $S, K \to \infty$ [15], and it is more tractable than the more complicated non-linear precoding schemes.

There are some works studying antenna selection problems in multi-user MIMO systems [16], [17]. However, these approaches require the global knowledge of the instantaneous CSI in the network which causes unacceptable signaling overhead for large M. This problem could be avoided using a simple antenna selection algorithm which associates each user with the "strongest" antennas/BSs (i.e., the antennas/BSs who have the largest path gain with the user). This simple baseline algorithm has been adopted in conventional cellular networks such as 3G and LTE systems. However, it is inefficient when the antennas/BSs are allowed to perform cooperative MIMO (CoMP) [18] as illustrated in the following two examples. In both examples, we assume $S = 2$ distributed antennas are selected to serve $K = 2$

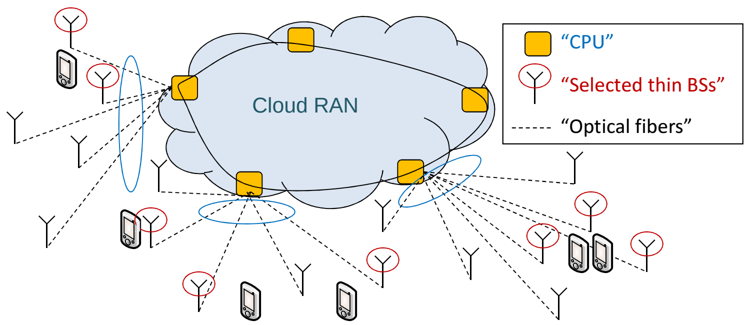

Figure 1. Illustration of a large distributed MIMO network, which consists of M thin base stations (distributed antennas) connected to a cloud RAN via high speed optical fiber.

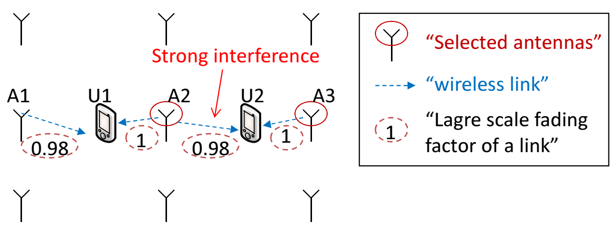

Figure 2. An example that strong cross link causes large interference

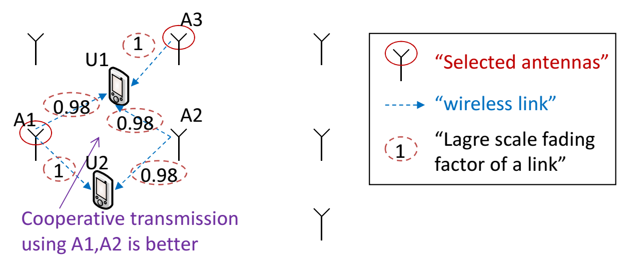

Figure 3. An example that strong cross links provide cooperative gain

users.

Example 1 (Strong cross link causing 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 precoding. Although the interference can be suppressed using RZF precoding, the overall SINR is still low because the cross link from A3 to U1 is weak and the joint transmission gain is very limited. A better choice will be $\mathcal{A} = \{1, 3\}$.

Example 2 (Strong cross link providing 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.

Hence, a more efficient antenna selection design is needed to better exploit the multi-user spatial multiplexing gain provided by a large number of distributed antennas with cooperation in the C-RAN. We study the joint optimization of antenna selection, regularization factor in RZF precoding, and power allocation, to maximize the average weighted sum-rate under sum power and per antenna power constraints. The optimization requires the knowledge of large scale fading factors only and the overhead for CSI acquisition is greatly reduced as discussed in Remark 1. The following are two first-order technical challenges that need to be addressed.

- **Combinatorial Optimization Problem**: As motivated by the examples, traditional antenna selection scheme in which each user is associated with the closest antennas (thin base stations) is highly sub-optimal in the CoMP configuration. In fact, the antenna selection problem with CoMP processing in the C-RAN is combinatorial with exponential complexity w.r.t. the total number of antennas M. Brute force solution requires exhaustive search which is highly undesirable.
- **Asymptotic Performance Analysis**: It is also important to derive closed-form performance expressions for the C-RAN with dynamic antenna selection supporting CoMP in order to obtain design insights on how the network performance scale with important system parameters. Yet, the performance analysis is highly non-trivial due to the heterogeneous path loss between antenna-user pair as well as the lack of closed form antenna selection solution.

In this paper, we first outline the system model and the antenna selection formulation in Section II and III. Using the random matrix theory, we extend the results in [11] to obtain asymptotically accurate expressions for the weighted sum-rate objective and the per-antenna transmit power constraints under a given antenna selection and power allocation. 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 and the asymptotic performance analysis for very large distributed MIMO networks. For large M, we show that there is an asymptotic decoupling effect in the distributed MIMO networks and the capacity grows logarithmically with M even when the number of active antennas S is fixed. Simulations in Section VI illustrate that the proposed solution achieves significant performance gain compared with various baselines.

II. System Model

A. Channel Model

Consider the downlink of a large distributed MIMO network with M distributed transmit antennas and K single-antenna users as illustrated in Fig. 1. The $M \gg K$ transmit antennas are distributed geographically and connected to a Cloud RAN [14] via high speed fiber backhaul. Denote h_{km} as the channel between the m^{th} transmit antenna and the k^{th} user. We consider composite fading channel, i.e.,

$h_{km} = \sigma_{km} W_{km}$, $\forall k, m$, where $\sigma_{km} \geq 0$ 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 (Composite fading channel model): The small scale fading process $W_{km}(t) \sim \mathcal{CN}(\mathbf{0}, 1)$ 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 and signal processing overheads to serve the K users, we consider antenna selection scheme where only a subset $\mathcal{A}$, $|\mathcal{A}| = S \geq K$ of the M distributed antennas are selected (activated) to serve the K users. For convenience, let $\mathcal{A}_j$ denote the j^{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 $\mathbf{\Sigma}(\mathcal{A}) \in \mathbb{R}_{++}^{K \times S}$ as the corresponding large scale fading matrix, whose element at k^{th} row and the j^{th} column is $\sigma_{k\mathcal{A}_j}$. In the rest of the paper, we will use $\mathbf{H}$ and $\mathbf{\Sigma}$ as the simplified notations for $\mathbf{H}(\mathcal{A})$ and $\mathbf{\Sigma}(\mathcal{A})$ when there is no ambiguity. The cloud RAN is assumed to have knowledge of the $K \times M$ large scale fading factors σ_{km}'s for antenna selection but only the $K \times S$ channel matrix $\mathbf{H}(\mathcal{A})$ for cooperative MIMO processing.

Remark 1 (Long-term and short-term CSI Acquisition): In TDD systems, channel reciprocity can be exploited and $\mathbf{H}(\mathcal{A})$ can be estimated at the cloud RAN by transmitting reference signals in the uplink. In FDD systems, the short-term channel matrix estimation $\mathbf{H}(\mathcal{A})$ can be obtained via explicit channel feedback. In both cases, the amount of training for the short-term channel matrix $\mathbf{H}(\mathcal{A})$ is limited by the coherence time of the channel. Since the channel coherent time is $O(1)$, estimated CSI quality $\mathbf{H}(\mathcal{A})$ at the C-RAN will be very poor if all the antennas in the network are active. Using antenna selection, the number of active antennas at any time slot is limited. Hence, the amount of instantaneous CSI feedback or the uplink overheads of channel sounding is greatly reduced by antenna selection $\mathcal{A}$. On the other hand, estimating the large scale fading matrix $\mathbf{\Sigma}$ is not limited by the channel coherent time because it is a long-term statistics. In both FDD and TDD systems, the large scale fading matrix $\mathbf{\Sigma}$ can be estimated at the C-RAN from the uplink reference signals [19], [20] due to the reciprocity of large scale path loss. ■

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

$$\mathbf{y} = \mathbf{HFs} + \mathbf{n},$$

where $\mathbf{s} = [s_1, ..., s_K] \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_K)$ is the symbol vector, $\mathbf{F} = [\mathbf{f}_1, ..., \mathbf{f}_K] \in \mathbb{C}^{S \times K}$ is the precoding matrix and $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_K)$ is the AWGN noise vector. We employ regularized zero-forcing (RZF) precoding [12] which is given by

$$\mathbf{F} = \left(\mathbf{H}^{\dagger}\mathbf{H} + \alpha\mathbf{I}_S\right)^{-1}\mathbf{H}^{\dagger}\mathbf{P}^{1/2} = \mathbf{H}^{\dagger}\left(\mathbf{H}\mathbf{H}^{\dagger} + \alpha\mathbf{I}_K\right)^{-1}\mathbf{P}^{1/2}, \tag{1}$$

where α 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 RZF is a generalization of the conventional RZF 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 $\hat{\mathbf{h}}_k^{\dagger}$ denote the k^{th} row of $\hat{\mathbf{H}}$ and $\hat{\mathbf{H}}\mathbf{1}_j$ denote the j^{th} column of $\hat{\mathbf{H}}$, where $\mathbf{1}_j$ denotes a $K \times 1$ vector of which the j^{th} element is 1 and all other elements are zeros. Define $\hat{\mathbf{H}}_k$ as the matrix $\hat{\mathbf{H}}$ with the k^{th} row removed, and $\mathbf{P}_k \triangleq \text{diag}(p_1, ..., p_{k-1}, p_{k+1}, ..., p_K)$. We assume that user k has perfect knowledge of the effective channel $\mathbf{h}_k^{\dagger}\mathbf{f}_k$ and the interference-plus-noise power, which can be estimated through dedicated downlink training. Under this assumption and using matrix inversion lemma, it can be shown that the SINR of user k is given by [10]

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

where

$$A_k = \hat{\mathbf{h}}_k^{\dagger}\left(\hat{\mathbf{H}}_k^{\dagger}\hat{\mathbf{H}}_k + \rho\mathbf{I}_S\right)^{-1}\hat{\mathbf{h}}_k,$$
$$B_k = \hat{\mathbf{h}}_k^{\dagger}\left(\hat{\mathbf{H}}_k^{\dagger}\hat{\mathbf{H}}_k + \rho\mathbf{I}_S\right)^{-1}\hat{\mathbf{H}}_k^{\dagger}\mathbf{P}_k\hat{\mathbf{H}}_k\left(\hat{\mathbf{H}}_k^{\dagger}\hat{\mathbf{H}}_k + \rho\mathbf{I}_S\right)^{-1}\hat{\mathbf{h}}_k.$$

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

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

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 ρ and the power allocation $\mathbf{p}$. The optimization is performed over the time scale of large scale fading, i.e., $\mathcal{A}$, ρ and $\mathbf{p}$ 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 σ_{km}'s and an active antenna set $\mathcal{A}$, the large scale fading matrix $\mathbf{\Sigma}(\mathcal{A})$ is fixed within a large scale fading block, and the conditioned average weighted sum-rate is given by

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

where the conditioned expectation $\mathrm{E}\left[\cdot\,|\mathbf{\Sigma}(\mathcal{A})\right]$ is taken over the small scale fading factors, and $w_k\geq 0$ is the weight for user k. We consider both per antenna and sum power constraint, which are given by

$$\mathrm{E}\left[p_m\left(\mathbf{\Sigma}(\mathcal{A}),\rho,\mathbf{p}\right)|\mathbf{\Sigma}(\mathcal{A})\right]\leq\bar{p}_m,\ m\in\mathcal{A},\tag{5}$$

$$\mathrm{E}\left[\sum_{m\in\mathcal{A}}p_m\left(\mathbf{\Sigma}(\mathcal{A}),\rho,\mathbf{p}\right)|\mathbf{\Sigma}(\mathcal{A})\right]\leq P_T.\tag{6}$$

B. Problem Formulation

The optimization problem is formulated as follows

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

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. ρ and $\mathbf{p}$, as will be shown later. In this paper, the first challenge is tackled in this section by using the random matrix theory in [11] 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.

To derive asymptotically accurate deterministic expressions, we require the following assumptions. Throughout the paper, the notation $S\overset{K=S\beta}{\longrightarrow}\infty$ refers to $K,S\rightarrow\infty$ such that $\lim_{S\rightarrow\infty}K/S=\beta\in(0,\infty)$.

Assumption 2 (Boundedness of Large Scale Fading): The elements of the large scale fading matrix $\mathbf{\Sigma}(\mathcal{A})$ are uniformly bounded on S, i.e.,

$$\limsup_{S\overset{K=S\beta}{\longrightarrow}\infty}\ \sup_{1\leq k\leq K,\,m\in\mathcal{A}}|\sigma_{km}|\ <\ \infty.$$

Moreover, $\frac{1}{S}\sum_{k=1}^{K}\sigma_{km}^2$, $\forall m\in\mathcal{A}$ is uniformly lower bounded by some nonnegative quantity, i.e.,

$$\liminf_{S\overset{K=S\beta}{\longrightarrow}\infty}\ \inf_{m\in\mathcal{A}}\frac{1}{S}\sum_{k=1}^{K}\sigma_{km}^2\ >\ 0.$$

The normalized channel matrix $\hat{\mathbf{H}}(\mathcal{A})$ has uniformly bounded spectral norm on S with probability 1, i.e.,

$$\limsup_{S \overset{K=S\beta}{\longrightarrow} \infty} \left\| \hat{\mathbf{H}}^{\dagger}(\mathcal{A}) \hat{\mathbf{H}}(\mathcal{A}) \right\| \quad < \quad \infty,$$

with probability one, where $\|\cdot\|$ denotes the spectral norm of a matrix.

The power allocation $\mathbf{p}$ satisfies $\max(p_1, ..., p_K) = O(1/K)$.

■

In the following, we derive asymptotic expressions for the 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): Based on Assumption 2, the following statements are true for any given $(\mathcal{A}, \rho > 0, \mathbf{p})$.

1) The system of K equations:

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

admits a unique solution $\vec{\xi} = [\xi_1, ..., \xi_K]^T$ in $\mathbb{R}_+^K$, where $f_m\left(\vec{\xi}\right) \triangleq \rho + \frac{1}{S}\sum_{i=1}^K \frac{\sigma_{im}^2}{1+\xi_i}$.

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

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

Define vector $\mathbf{b} = [b_1, ..., b_K]^T$ whose l^{th} element is given by

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

For any $k \in \{1, ..., K\}$, define vector $\mathbf{d}_k = [d_{k1}, ..., d_{kK}]^T$ whose l^{th} element is given by

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

Then $\mathbf{I}_K - \mathbf{D}$ is invertible and thus we can define the following vectors

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

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

where the l^{th} element of $\vec{\phi}$, denoted by ϕ_l, is the partial derivative of ξ_l over ρ.

3) Let $\gamma_k(\mathbf{\Sigma}, \rho, \mathbf{p})$ be the SINR of user k defined in (2). As $K, S \to \infty$ with the ratio $\beta = K/S$ fixed, we have $\gamma_k(\mathbf{\Sigma}, \rho, \mathbf{p}) - \bar{\gamma}_k(\mathbf{\Sigma}, \rho, \mathbf{P}) \overset{a.s}{\to} 0$, where

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

where θ_{kl} is the l^{th} element of $\vec{\theta}_k$.

Lemma 2 (Asymptotic per-antenna power): Based on Assumption 2, the following statements are true for any given $(\mathcal{A}, \rho > 0, \mathbf{p})$.

1) The system of K equations:

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

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

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

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

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

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

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

$$c_l = \frac{1}{S}\sum_{m\in\mathcal{A}}\left[\sigma_{lm}^2 v_l\left(p_l + \frac{1}{S}\sum_{i=1}^K \sigma_{im}^2 v_i(p_l - p_i)\right)/\psi_m^2(\mathbf{v})\right].$$

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

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

3) Let $p_m(\mathbf{\Sigma}, \rho, \mathbf{p})$ be the per-antenna power defined in (2). As $K, S \to \infty$ with the ratio $\beta = K/S$ fixed, we have $p_m(\mathbf{\Sigma}, \rho, \mathbf{p}) - \bar{p}_m(\mathbf{\Sigma}, \rho, \mathbf{p}) \overset{a.s}{\to} 0$, $\forall m \in \mathcal{A}$, where

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

where φ_i is the i^{th} element of $\vec{\varphi}$ in (15).

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

Remark 2: The equations in (8) and (12) can be solved using the fixed point iterations in Proposition 1 of [11]. ■

Remark 3: Another deterministic approximation of the SINR has been provided in Theorem 2 of [11]. The differences between the deterministic approximation in (11) and the one in [11] are as follows. 1) In [11], a more general channel model which includes per-user channel transmit correlation and channel estimation error is considered. Our channel model is a special case of the channel model in [11] with the channel correlation matrices given by $\mathbf{\Theta}_k = \text{diag}\left(\sigma_{km}^2,\, m \in \mathcal{A}\right)$, $\forall k$ and zero channel estimation error, i.e., $\tau_k = 0$ in [11]. Since we assume geographically distributed antennas, the correlations between the antennas are indeed negligible. Moreover, the algorithms and most results in this paper can be directly generalized to case with transmit correlation and channel estimation error with little modification. 2) We consider both sum power constraint and per-antenna power constraint, while [11] focused on sum power constraint only. Consequently, in the deterministic approximation in [11], the precoding matrix $\mathbf{F}$ is scaled to satisfy the sum power constraint. However, it is impossible to satisfy the per-antenna power constraint by simply scaling $\mathbf{F}$. Hence, the deterministic approximation in (11) is for a given power allocation $\mathbf{p}$ without scaling of $\mathbf{F}$. Moreover, we need to derive the deterministic approximation for per-antenna power constraint in (2), which is not given in [11]. Then the per-antenna power constraint is guaranteed by the optimization algorithm. ■

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 and following similar proof as that of [21, Prop. 1], we can prove the following theorem which gives an asymptotic equivalence of Problem (7).

Theorem 1 (Asymptotic equivalence of Problem (7)): Assume that Assumption 2 is satisfied. Let $\mathcal{A}^*$, $\rho^*, \mathbf{p}^*$ denote an optimal solution of the following optimization problem

$$\begin{aligned} \mathcal{P}_E: \quad \max \quad & \bar{\mathcal{I}}(\mathcal{A},\rho,\mathbf{p}) \triangleq \sum_{k=1}^{K} w_k \log\left(1+\bar{\gamma}_k(\mathbf{\Sigma},\rho,\mathbf{p})\right) \\ \text{s.t.} \quad & |\mathcal{A}| = S,\ \rho > 0,\ \mathbf{p} \geq 0, \\ \text{and} \quad & \bar{p}_m(\mathbf{\Sigma},\rho,\mathbf{p}) \leq \bar{p}_m,\ \forall m \in \mathcal{A},\ \bar{p}_{\mathcal{A}}(\mathbf{\Sigma},\rho,\mathbf{p}) \leq P_T, \end{aligned} \tag{17}$$

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. Let $\mathcal{I}^*$ denote the optimal value of Problem (7). As $K, S \rightarrow \infty$ with the ratio $\beta = K/S$ fixed, we have

$$\bar{\mathcal{I}}(\mathcal{A}^*,\rho^*,\mathbf{p}^*) \xrightarrow{a.s} \mathcal{I}(\mathcal{A}^*,\rho^*,\mathbf{p}^*) \xrightarrow{a.s} \mathcal{I}^*.$$

Remark 4: Theorem 1 implies that the solution of the complicated problem in (7) can be approximated by the solution of $\mathcal{P}_E$, and the approximation is asymptotically accurate as $S \overset{K=S\beta}{\longrightarrow} \infty$. Simulations in Section VI show that the deterministic approximation is still good even for a small number of transmit antennas and users. ■

IV. Optimization Solution for $\mathcal{P}_E$

In this section, we shall tackle the remaining challenges of solving $\mathcal{P}_E$. We first decompose the complex problems into simpler subproblems, and then propose efficient algorithms for solving the subproblems.

A. Problem Decomposition

Using primal decomposition, $\mathcal{P}_E$ can be decomposed into the following 3 subproblems:

Subproblem 1: Optimization of $\mathbf{p}$ under fixed $\mathcal{A}$ and ρ, which can be formulated as

$$\mathcal{P}_1(\mathcal{A},\rho):\ \max_{\mathbf{p}\geq 0}\bar{\mathcal{I}}(\mathcal{A},\rho,\mathbf{p}),\ \text{s.t. (17) is satisfied.} \tag{18}$$

Subproblem 2: Optimization of ρ under fixed $\mathcal{A}$, which can be formulated as

$$\mathcal{P}_2(\mathcal{A}):\ \max_{\rho>0}\bar{\mathcal{I}}(\mathcal{A},\rho,\mathbf{p}^*(\mathcal{A},\rho)),\ \text{s.t. (17) is satisfied,} \tag{19}$$

where $\mathbf{p}^*(\mathcal{A},\rho)$ is the optimal solution of (18).

Subproblem 3: Optimization of $\mathcal{A}$.

$$\mathcal{P}_3:\ \max_{\mathcal{A}}\bar{\mathcal{I}}(\mathcal{A},\rho^*(\mathcal{A}),\mathbf{p}^*(\mathcal{A},\rho^*(\mathcal{A}))),\ \text{s.t. } \mathcal{A}\subseteq\{1,...,M\},\ \text{and } |\mathcal{A}|=S, \tag{20}$$

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

$$\bar{\mathcal{I}}(\mathcal{A},\rho,\mathbf{p})=\sum_{k=1}^{K}w_k\log\left(1+g_{kk}p_k/\left(1+\sum_{l\neq k}^{K}g_{kl}p_l\right)\right),$$

$$g_{kk}\triangleq\frac{\xi_k^2}{(1+\xi_k)^2},\ \forall k,\ g_{kl}\triangleq\frac{\theta_{kl}}{S(1+\xi_l)^2(1+\xi_k)^2},\ \forall k\neq l.$$

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

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

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

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

Let $\mathbf{V} = \text{diag}(v_1, ..., v_K)$. Then the per antenna power constraint in (17) can be rewritten as $\tilde{\mathbf{R}}\mathbf{p} \leq [\bar{p}_{\mathcal{A}_1}, ..., \bar{p}_{\mathcal{A}_S}]^T$, where $\tilde{\mathbf{R}} \triangleq \hat{\mathbf{R}}^T\left[\mathbf{V} - (\rho\mathbf{I}_K + \mathbf{\Delta} - \mathbf{C})^{-1}\bar{\mathbf{R}}\right] \in \mathbb{R}^{S\times K}$. The sum power constraint in (17) can be rewritten as $\mathbf{1}^T\tilde{\mathbf{R}}\mathbf{p} \leq \bar{p}_{\mathcal{A}}$. Finally, the overall constraint in (17) 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

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

where $\vec{\lambda} \in \mathbb{R}_+^{(S+1)\times 1}$ is the Lagrange multipliers. It is well known that Problem (18) is usually a non-convex problem with non-zero duality gap. Hence, the standard Lagrange dual method (LDM) [22] cannot be used to solve this problem. We apply the *local LDM* in [23] to solve Problem (18) as follows.

Algorithm S1 (for solving Subproblem 1):

Initialization: Choose initial $\vec{\lambda}^{(0)} > \mathbf{0}$. Let $i = 1$.

Step 1 (Primal update in the inner loop): For fixed $\vec{\lambda} = \vec{\lambda}^{(i-1)}$, find a stationary point $\tilde{\mathbf{p}}\left(\vec{\lambda}^{(i-1)}\right)$ of

$$\max_{\mathbf{p}} L\left(\vec{\lambda}, \mathbf{p}\right), \text{ s.t. } \mathbf{p} \geq 0, \tag{21}$$

using *Algorithm I* described later with initial point[1] $\tilde{\mathbf{p}}\left(\vec{\lambda}^{(i-2)}\right)$.

Step 2 (Dual update in the outer loop): Update $\vec{\lambda}$ as

$$\vec{\lambda}^{(i)} = \vec{\lambda}^{(i-1)} + t^{(i)}\triangle\vec{\lambda}^{(i)}, \tag{22}$$

where $t^{(i)}$ denotes the step size, and the update direction $\triangle\vec{\lambda}^{(i)}$ is the optimal solution of the following quadratic programming problem:

$$\min_{\triangle\vec{\lambda}} \triangle\vec{\lambda}^T\mathbf{b}^{(i)} + \frac{1}{2}\triangle\vec{\lambda}^T\mathbf{B}^{(i)}\triangle\vec{\lambda}, \text{ s.t. } \vec{\lambda}^{(i-1)} + \triangle\vec{\lambda} \geq 0, \tag{23}$$

where $\mathbf{b}^{(i)} = \bar{\mathbf{p}} - \mathbf{R}\tilde{\mathbf{p}}\left(\vec{\lambda}^{(i-1)}\right)$, and $\mathbf{B}^{(i)} \in \mathbb{R}^{N\times N}$ is positive definite.

Return to Step 1 until convergence.

[1]Note that $\tilde{\mathbf{p}}\left(\vec{\lambda}^{(-1)}\right)$ is randomly generated.

Choice of the Matrix $\mathbf{B}^{(i)}$*:* $\mathbf{B}^{(i)}$ is obtained by the well known BFGS update as follows [24]

$$\mathbf{B}^{(i+1)} = \begin{cases} \mathbf{B}^{(i)} + \frac{\mathbf{q}_i \mathbf{q}_i^T}{\mathbf{q}_i^T \mathbf{z}_i} - \frac{\mathbf{B}^{(i)} \mathbf{z}_i \mathbf{z}_i^T \mathbf{B}^{(i)}}{\mathbf{z}_i^T \mathbf{B}^{(i)} \mathbf{z}_i}, & \mathbf{q}_i^T \mathbf{z}_i > 0, \\ \mathbf{B}^{(i)}, & \text{otherwise}, \end{cases}$$

$$\mathbf{z}_i = \vec{\lambda}^{(i)} - \vec{\lambda}^{(i-1)}, \ \mathbf{q}_i = \mathbf{b}^{(i+1)} - \mathbf{b}^{(i)}.$$

The initial $\mathbf{B}^{(1)} = 2^{m_b} \mathbf{I}_{S+1}$, where m_b is the smallest integer such that $\max_n \left| \triangle \tilde{\lambda}_n \right| < 0.5$, and $\triangle \tilde{\vec{\lambda}} = \left[\triangle \tilde{\lambda}_1, ..., \triangle \tilde{\lambda}_{S+1} \right]$ is the optimal solution of the problem $\max_{\triangle \vec{\lambda}} \triangle \vec{\lambda}^T \mathbf{b}^{(1)} + \frac{1}{2} \triangle \vec{\lambda}^T \mathbf{B}^{(1)} \triangle \vec{\lambda}$.

Choice of the Step Size $t^{(i)}$*:* Set $t^{(i)} = \tau^{(i)} 2^{-m_t}$, where m_t is an integer which is initialized as 0 and is incremented until one of the following conditions is satisfied: 1) $L\left(\vec{\lambda}^{(i)}, \tilde{\mathbf{p}}\left(\vec{\lambda}^{(i)}\right)\right) \leq L\left(\vec{\lambda}^{(i-1)}, \tilde{\mathbf{p}}\left(\vec{\lambda}^{(i-1)}\right)\right)$. 2) $m_t = m_t^0$, where m_t^0 is a small positive integer, e.g., we let $m_t^0 = 2$ in the simulations. 3) $re_i \leq re_{i-1}$, where

$$re_i = \max_n \left| \lambda_n^{(i)} b_n^{(i+1)} \right| + \left(\max_n b_n^{(i+1)} \right)^+, \tag{24}$$

is defined as the *residual error* after the i^{th} iteration, and $b_n^{(i)}$, $\lambda_n^{(i)}$ are the n^{th} element of $\mathbf{b}^{(i)}$, $\vec{\lambda}^{(i)}$ respectively. Finally, $\tau^{(i)}$ is given by

$$\tau^{(i+1)} = (1 - \varepsilon) \tau^{(i)} + \varepsilon t^{(i)},$$

where $0 < \tau^{(1)} \leq 0.5$, $0 < \varepsilon < 1$ and are set as $\tau^{(1)} = 0.25, \varepsilon = 0.2$ in the simulations.

In the following, we will propose an efficient primal algorithm to solve the inner loop problem (21) based on the interference pricing method [25], 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 [25]

$$\pi_k = \sum_{l \neq k}^{K} w_l g_{lk} \frac{g_{ll} p_l}{\Omega_l \left(\Omega_l + g_{ll} p_l \right)}, \tag{25}$$

where

$$\Omega_l = 1 + \sum_{i \neq l}^{K} g_{li} p_i, \tag{26}$$

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

$$\max_{\mathbf{p}} w_k \log \left(1 + \frac{g_{kk} p_k}{1 + \sum_{l \neq k}^{K} g_{kl} p_l} \right) - \vec{\lambda}^T \mathbf{r}_k p_k - \pi_k p_k, \text{ s.t. } \mathbf{p} \geq 0, \tag{27}$$

where $\mathbf{r}_k$ is the k^{th} column of $\mathbf{R}$. Here, we present a primal algorithm with sequential power updates to solve the inner loop problem (21).

Algorithm I (Primal algorithm for solving inner-loop Problem (21)):

Initialization: Choose proper initial $\mathbf{p}$.

While not converge **do**

For $k=1$ to K

Calculate π_k in (25) and Ω_k in (26) using current power allocation $\mathbf{p}$.

Given π_k and current power allocation $\mathbf{p}$, update p_k by solving problem (27) 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 Primal Alg. I): Algorithm I converges monotonically to a stationary point of the inner-loop Problem in (21).

The proof can be established using a similar approach as in Appendix A of [25].

Using the monotonic convergence of the primal Algorithm I in Theorem 2, it follows from Theorem 1 and Theorem 2 in [23] that the overall Algorithm S1 (with Algorithm I as primal algorithm) converges to a stationary point of $\mathcal{P}_1(\mathcal{A},\rho)$ under the mild conditions summarized in [23].

C. Bisection Algorithm for Subproblem 2

One of the main challenges for solving Subproblem 2 is that the calculation of the objective function $\bar{\mathcal{I}}(\mathcal{A},\rho,\mathbf{p}^*(\mathcal{A},\rho))$ requires solving the optimal solution $\mathbf{p}^*(\mathcal{A},\rho)$ of Subproblem 1 which is a non-convex 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 3 (Equivalent problem of $\mathcal{P}_2(\mathcal{A})$): Consider the following joint optimization problem under fixed $\mathcal{A}$:

$$\mathcal{P}_{2a}(\mathcal{A}):\ \max_{\rho>0,\mathbf{p}\geq 0} \bar{\mathcal{I}}(\mathcal{A},\rho,\mathbf{p}),\ \text{s.t. (17) is satisfied.} \tag{28}$$

Then the followings are true:

1) Let $\rho^*,\mathbf{p}^*$ denote the optimal solution of $\mathcal{P}_{2a}(\mathcal{A})$. Then ρ^* must be an optimal solution of $\mathcal{P}_2(\mathcal{A})$, and $\mathbf{p}^*$ must be an optimal solution of $\mathcal{P}_1(\mathcal{A},\rho^*)$.
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

$$\hat{\mathcal{I}}(\mathcal{A},\rho) \triangleq \bar{\mathcal{I}}(\mathcal{A},\rho,\tilde{\mathbf{p}}(\mathcal{A},\rho)). \tag{29}$$

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

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

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 ρ_a, ρ_b such that $0 < \rho_a < \rho_b$ and $\frac{\partial \hat{\mathcal{I}}(\mathcal{A},\rho_a)}{\partial \rho} > 0$, $\frac{\partial \hat{\mathcal{I}}(\mathcal{A},\rho_b)}{\partial \rho} < 0$.

Step 1: Let $\rho = (\rho_a + \rho_b)/2$. If $\frac{\partial \hat{\mathcal{I}}(\mathcal{A},\rho)}{\partial \rho} \leq 0$, 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 $\frac{\partial \hat{\mathcal{I}}(\mathcal{A},\rho)}{\partial \rho}$ without an analytical expression for $\hat{\mathcal{I}}(\mathcal{A}, \rho)$. In the following, we show how to calculate $\frac{\partial \hat{\mathcal{I}}(\mathcal{A},\rho)}{\partial \rho}$ from the output of Algorithm S1: $\tilde{\mathbf{p}}(\mathcal{A}, \rho)$ and $\tilde{\vec{\lambda}}$, where $\tilde{\vec{\lambda}} \in \mathbb{R}_+^{(S+1)\times 1}$ is the Lagrange multipliers corresponding to $\tilde{\mathbf{p}}(\mathcal{A}, \rho)$.

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

$$\begin{aligned} \nabla_{\mathbf{p}} \bar{\mathcal{I}}(\mathcal{A}, \rho, \tilde{\mathbf{p}}(\mathcal{A}, \rho)) - \mathbf{R}^T \tilde{\vec{\lambda}} + \tilde{\vec{\nu}} &= 0; \\ \text{diag}(\bar{\mathbf{p}}) \tilde{\vec{\lambda}} - \text{diag}\left(\tilde{\vec{\lambda}}\right) \mathbf{R} \tilde{\mathbf{p}}(\mathcal{A}, \rho) &= 0; \\ \text{diag}\left(\tilde{\vec{\nu}}\right) \tilde{\mathbf{p}}(\mathcal{A}, \rho) &= 0; \end{aligned} \tag{31}$$

where $\tilde{\vec{\nu}} \in \mathbb{R}_+^K$ is the Lagrange multipliers associated with the positive constraint $\mathbf{p} \geq 0$. By (31), $\tilde{\vec{\nu}} = \mathbf{R}^T \tilde{\vec{\lambda}} - \nabla_{\mathbf{p}} \bar{\mathcal{I}}(\mathcal{A}, \rho, \tilde{\mathbf{p}}(\mathcal{A}, \rho))$, where $\nabla_{\mathbf{p}} \bar{\mathcal{I}}(\mathcal{A}, \rho, \tilde{\mathbf{p}}(\mathcal{A}, \rho)) = \left[\frac{\partial \bar{\mathcal{I}}}{\partial p_1}, ..., \frac{\partial \bar{\mathcal{I}}}{\partial p_K}\right]$ and

$$\frac{\partial \bar{\mathcal{I}}}{\partial p_k} = w_k g_{kk} / \left(\tilde{\Omega}_k + g_{kk} \tilde{p}_k(\mathcal{A}, \rho)\right) - \pi_k,$$

where $\tilde{\Omega}_k$ is the interference-plus-noise power in (26) calculated from $\tilde{\mathbf{p}}(\mathcal{A}, \rho)$. Assuming that $\frac{\partial \tilde{\mathbf{p}}(\mathcal{A},\rho)}{\partial \rho}$, $\frac{\partial \tilde{\vec{\lambda}}}{\partial \rho}$ and $\frac{\partial \tilde{\vec{\nu}}}{\partial \rho}$ exist and taking partial derivative of the equations in (31) with respect to ρ, we obtain a linear equation with $\frac{\partial \tilde{\mathbf{p}}(\mathcal{A},\rho)}{\partial \rho}$, $\frac{\partial \tilde{\vec{\lambda}}}{\partial \rho}$ and $\frac{\partial \tilde{\vec{\nu}}}{\partial \rho}$ 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

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

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 B.

D. Algorithm S3 for Solving Subproblem 3

Subproblem 3 is a combinatorial problem and the 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.

As illustrated from Example 1 and 2 in the introduction, 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 good 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 have a direct link with a single user. Such antenna provides a direct link for a single user without causing strong interference to others[2]. In step 2, the algorithm selects the antennas which have 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 Γ_j. In the j^{th} search, we switch the j^{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 Γ_j is carefully chosen to reduce the number of weighted sum-rate calculations as well as maintain a good performance.

We first define some notations and then give the detailed steps of Algorithm S3. Let $\tilde{m}_k = \text{argmax}_m \sigma_{km}^2$, $k = 1, ..., K$. Define $\bar{g}_k^d = \sigma_{k\tilde{m}_k}^2$. For $k = 1, ..., K$, and $m = 1, ..., M$, let $G_{km} = 1$, if $\sigma_{km}^2 \geq \kappa \bar{g}_k^d$, and otherwise, let $G_{km} = 0$, where $\kappa \in (0, 1)$. Roughly speaking, G_{km} is an indication of whether antenna m contribute significantly to the communication of user k. Simulations show that the performance of Algorithm S3 is not sensitive to the choice of κ for $\kappa \in \left[\frac{1}{4}, \frac{1}{2}\right]$ and we set $\kappa = \frac{1}{4}$ for the simulations in Section VI. Define $\mathcal{K}_m = \{k : G_{km} > 1\}$, $m = 1, ..., M$. Let $\tilde{\mathcal{I}}_{\mathcal{A}} \triangleq \bar{\mathcal{I}}\left(\mathcal{A}, \tilde{\rho}\left(\mathcal{A}\right), \tilde{\mathbf{p}}\left(\mathcal{A}, \tilde{\rho}\left(\mathcal{A}\right)\right)\right)$ denote

[2]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 precoding and the joint transmission using RZF does not provide much gain due to some other weak cross links as shown in Example 1.

the optimized 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 $\mathcal{A}$.

Algorithm S3 (for solving Subproblem 3):

Initialization: Let $\mathcal{A} = \Phi$, where Φ 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 can be any constant larger than $\max\limits_{1 \leq m \leq M} \sum_{k=1}^{K} \sigma_{km}^2$.

Let $m^* = \underset{m \in \bar{\mathcal{A}}}{\text{argmax}}\, \bar{G}_m$

While $|\mathcal{K}_{m^*}| \geq 2$ and $|\mathcal{A}| < S$

Let $\mathcal{A} = \mathcal{A} \cup m^*$ and $m^* = \underset{m \in \bar{\mathcal{A}}}{\text{argmax}}\, \bar{G}_m$.

End

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

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

Let $\tilde{k}_m = \underset{k}{\text{argmax}}\, \sigma_{km}^2$ and $I_m = w_{\tilde{k}_m} \log\left(1 + \sigma_{\tilde{k}_m m}^2\right)$.

While $|\mathcal{A}| < S$

Let $m^* = \underset{m \in \bar{\mathcal{A}}}{\text{argmax}}\, I_m$ 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 $n^* = \underset{m \in \bar{\mathcal{A}} \cap \{m: |\mathcal{K}_m| = 1\}}{\text{argmax}} I_m$ and $\Gamma_j^a = \bar{\mathcal{A}} \cap \{m: \ |\mathcal{K}_m| \geq 2\}$.

If $I_{n^*} \geq I_{\mathcal{A}_j}$ or $\left|\mathcal{K}_{\mathcal{A}_j}\right| > 1$, let $\Gamma_j = \Gamma_j^a \cup n^*$; otherwise, let $\Gamma_j = \Gamma_j^a$.

Let $m^* = \underset{m \in \Gamma_j}{\text{argmax}}\, \tilde{\mathcal{I}}_{\mathcal{A}_{-j} \cup m}$.

If $\tilde{\mathcal{I}}_{\mathcal{A}_{-j} \cup m^*} > \tilde{\mathcal{I}}_{\mathcal{A}}$, let $\mathcal{A}_j = m^*$.

End

Finally, we give more elaboration on the choice of the candidate antenna set Γ_j in step 4. The I_m defined in step 3 roughly reflects the contribution of antenna m to the weighted rate of a single user. Then n^* in step 4 is the unselected antenna which is likely to contribute the most to the weighted rate of a single user without causing strong interference to other users. If $I_{n^*} \geq I_{\mathcal{A}_j}$, n^* is added in Γ_j. Even if $I_{n^*} < I_{\mathcal{A}_j}$, we still add n^* in Γ_j if $\mathcal{A}_j$ has the potential to cause large interference, i.e., $\left|\mathcal{K}_{\mathcal{A}_j}\right| > 1$. Γ_j^a are the set of unselected antennas which have the potential to provide cooperative gain, and are also added in Γ_j.

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 $\mathcal{A}$ 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 ρ^*.

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 4 (Asymptotic expressions for collocated antennas): For a given $(\mathcal{A}, \rho > 0, \mathbf{p})$, if $\sigma_{km}^2 = \sigma_{k1}^2$, $k = 1, ..., K$, $m = 1, ..., M$, then the followings are true:

1) The system of equation:
$$u = \frac{1}{\rho + \frac{1}{S}\sum_{i=1}^{K}\frac{\sigma_{i1}^2}{1+\sigma_{i1}^2 u}}, \tag{33}$$
admits a unique solution u in $\mathbb{R}_{++}$.

2) Define
$$F_{12} = \frac{1}{S}\sum_{i=1}^{K}\frac{\sigma_{i1}^2}{\left(1+\sigma_{i1}^2 u\right)^2},\ \bar{F}_{12}(\mathbf{p}) = \frac{1}{S}\sum_{i=1}^{K}\frac{p_i\sigma_{i1}^2}{\left(1+\sigma_{i1}^2 u\right)^2}.$$
As $K, S \to \infty$ 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 almost surely
$$\begin{aligned}\bar{\gamma}_k(\mathbf{\Sigma}, \rho, \mathbf{P}) &= \frac{p_k\sigma_{k1}^4 u^2(\rho + F_{12})}{\bar{F}_{12}(\mathbf{p})\sigma_{k1}^2 u + \left(1+\sigma_{k1}^2 u\right)^2(\rho + F_{12})},\\ \bar{p}_m(\mathbf{\Sigma}, \rho, \mathbf{p}) &= \frac{\bar{F}_{12}(\mathbf{p})u}{S(\rho + F_{12})},\ \forall m \in \mathcal{A}.\end{aligned} \tag{34}$$

The proof is similar to the proof in Appendix A.

Using Theorem 4, $\mathcal{P}_1(\mathcal{A}, \rho)$ 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 4, subproblem $\mathcal{P}_1(\mathcal{A}, \rho)$ can be equivalent to the following optimization problem

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

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:

$$p_k^*(\mathcal{A},\rho) = \left(\frac{w_k S\left(1+\sigma_{k1}^2 u\right)^2(\rho+F_{12})}{\lambda\sigma_{k1}^2 u} - \frac{\sigma_{k1}^2 P_T^{'} + \left(1+\sigma_{k1}^2 u\right)^2}{\sigma_{k1}^4 u^2}\right)^+, \tag{36}$$

where λ is chosen such that $\bar{F}_{12}(\mathbf{p}^*(\mathcal{A},\rho))\, u/(\rho+F_{12}) = P_T^{'}$.

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

Theorem 5 (Properties of ρ^* 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 $\mathcal{P}_2(\mathcal{A})$: $\bar{\mathcal{I}}(\mathcal{A},\rho,\mathbf{p}^*(\mathcal{A},\rho))$, is a concave function of ρ for all $\rho < \rho_{\max}^{'}$.

The proof is given in Appendix C. Theorem 5 implies that with sufficiently small initial $\rho_b > \rho_a > 0$, Algorithm S2 will converge to the optimal ρ^* at high enough 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: $\mathcal{A}^* = \left\{m:\ \sigma_{1m}^2 \geq \sigma_{1S^{\max}}^2\right\}$, where $\sigma_{1S^{\max}}^2$ is the S^{th} largest σ_{1m}^2's and we shall focus on $\mathbf{p}^*$ and ρ^* 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 6 (Asymptotic expressions for collocated users): For a given $(\mathcal{A}, \rho > 0, \mathbf{p})$, 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 ξ 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, $\gamma_k(\mathbf{\Sigma},\rho,\mathbf{p})$ in (2) and $\sum_{m\in\mathcal{A}}^S p_m(\mathbf{\Sigma},\rho,\mathbf{p})$ in (6)

respectively converge to the following deterministic values almost surely

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

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

Remark 5: Similarly, we can obtain simpler expression for per antenna power $p_m(\mathbf{\Sigma},\rho,\mathbf{p})$ 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 6 and assuming we have sum power constraint $\bar{p}_{\mathcal{A}}(\mathbf{\Sigma},\rho,\mathbf{p}) \leq P_T$ only, subproblem $\mathcal{P}_1(\mathcal{A},\rho)$ is equivalent to the following optimization problem:

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

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

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

where λ is chosen such that $\frac{\beta E_{12}}{1-\beta E_{22}}\frac{1}{K}\sum_{l=1}^{K}p_l^*(\mathcal{A},\rho)=P_T$. For the special case of sum rate maximization where $w_k=1,\ \forall k$, the above water-filling solution gives equal power allocation, which coinsides with the result in Section IV of [11]. On the other hand, the optimal ρ^* for $\mathcal{P}_2(\mathcal{A})$ is given by:

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

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 (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 path 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^{th} antenna and k^{th} user, ζ is the path loss factor. ■

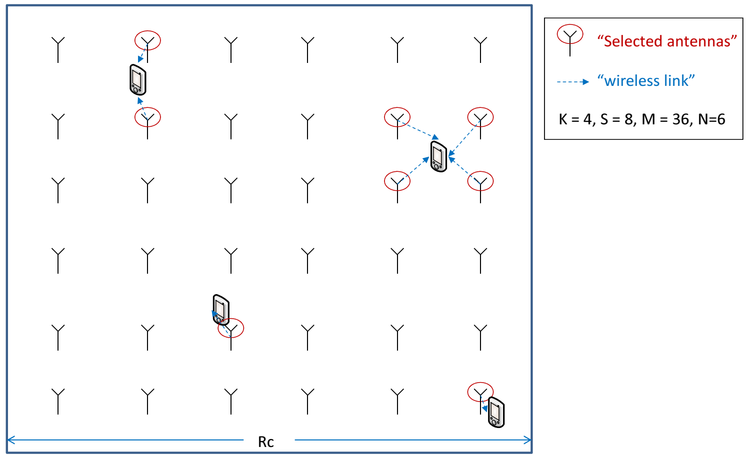

Figure 4. Illustration of asymptotic decoupling effect in a very large distributed MIMO network

Let $\tilde{m}_k = \underset{m}{\text{argmax}}\, \sigma^2_{km}$. For each user, define $\bar{g}^d_k = \sigma^2_{k\tilde{m}_k}$ as the direct-link gain, and $\bar{g}^c_k = \max_{l \neq k} \sigma^2_{k\tilde{m}_l}$ as the maximum cross-link gain. Define the ratio $\eta = \min_k \bar{g}^d_k / \bar{g}^c_k$, which measures the decoupling between the distributed antennas and the users. We have the following theorem

Theorem 7 (Asymptotic Decoupling and Capacity Scaling): For any $\eta_0 > 0$, we have

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

Furthermore, for any $\epsilon > 0$, the maximum achievable sum-rate C_s almost surely satisfies

$$O\left(K \left(\frac{\zeta}{2} - \epsilon \right) \log M \right) \leq C_s \leq O\left(K \left(\frac{\zeta}{2} + \epsilon \right) \log M \right),$$

as $M \to \infty$ with K, S fixed. ■

Please refer to Appendix D for the proof.

Remark 6 (Interpretation of Asymptotic Decoupling): For fixed $\beta = K/S$ and reasonably large M/K, there is a high probability that η 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. 4. 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 precoder [2] for each user using the selected antennas nearby) can also achieve good performance. ■

Corollary 1 (Asymptotic Optimality of Algorithm S3): Algorithm S3 is asymptotically optimal, i.e., for any $\epsilon > 0$, the achieved sum-rate $\mathcal{I}_{\mathcal{A}}$ satisfies $\mathcal{I}_{\mathcal{A}} \overset{a.s}{\geq} O\left(K \left(\frac{\zeta}{2} - \epsilon \right) \log M \right)$ as $M \to \infty$ with K, S fixed.

The proof is given in Appendix E.

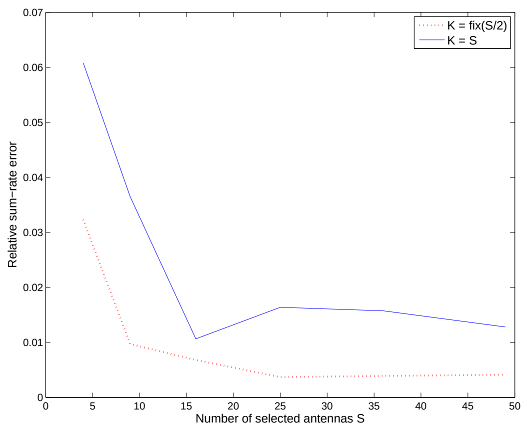

Figure 5. Relative sum-rate error versus S under sum power constraint $P_T = 10$

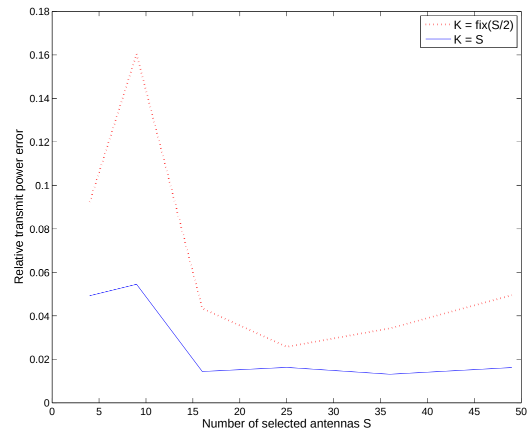

Figure 6. Relative transmit power error versus S under sum power constraint $P_T = 10$

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 users lying inside a square with an area of 2Km $\times$ 2Km. Assume that the antennas are evenly distributed in the square. Assume the same path loss model as in Assumption 3, where the path loss factor is set as $\zeta = 2.5$. The locations of users are randomly generated from a uniform distribution except for Fig. 8.

A. Verify the accuracy of the Asymptotic Expressions

We verify the accuracy of the asymptotic sum-rate (i.e., $w_k = 1, \forall k$) $\bar{\mathcal{I}}(\mathcal{A}, \rho, \mathbf{p})$ and transmit power $\bar{p}_{\mathcal{A}}(\mathbf{\Sigma}, \rho, \mathbf{p})$ by comparing them to those obtained by Monte-Carlo simulations: $\mathcal{I}^{\text{sim}}$ and $p_{\mathcal{A}}^{\text{sim}}$. In this simulation, all antennas are selected for transmission, i.e., $S = M$ and $\mathcal{A} = \{1, ..., M\}$. Assume sum power constraint $P_T = 10$ only and equal power allocation, i.e., $\mathbf{P} = c\mathbf{I}$ in (1), where c is chosen such that $\bar{p}_{\mathcal{A}}(\mathbf{\Sigma}, \rho, \mathbf{p}) = P_T$. The regularization factor is fixed as $\rho = \beta / P_T$, where $\beta = K/S$. In Fig. 5, we plot $\left(\bar{\mathcal{I}}(\mathcal{A}, \rho, \mathbf{p}) - \mathcal{I}^{\text{sim}}\right) / \mathcal{I}^{\text{sim}}$, the relative error of the asymptotic sum-rate compared to the simulated one, versus S. In Fig. 6, we plot $\left(\bar{p}_{\mathcal{A}}(\mathbf{\Sigma}, \rho, \mathbf{p}) - p_{\mathcal{A}}^{\text{sim}}\right) / p_{\mathcal{A}}^{\text{sim}}$, the relative error of the asymptotic sum transmit power compared to the simulated one, versus S. In each simulation, the large scale fading matrix $\mathbf{\Sigma}$ is generated according to a random realization of user locations and the simulated sum-rate / sum transmit power are averaged over a single large fading block. It can be seen that the asymptotic approximation is quite accurate. Note that the approximation error does not always decrease as M increases because in

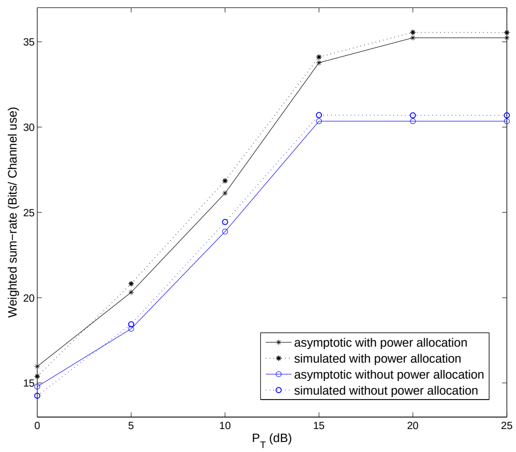

Figure 7. Comparison of asymptotic and simulated weighted sum-rates with and without power allocation

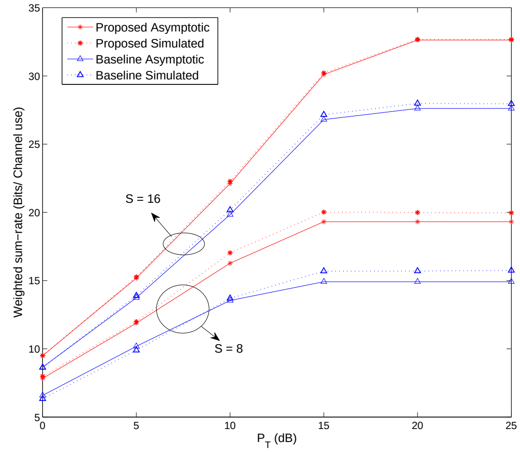

Figure 8. Comparison of proposed antenna selection scheme and baseline for strong cross-link case

each simulation, $\mathbf{\Sigma}$ is totally changed. But roughly speaking, the approximation error tends to decrease as M increases.

In the rest of the simulations, the per antenna power constraint is set as $\bar{p}_m = 5\text{dB}$, $m = 1, ..., M$. The number of users K is fixed as 8. From the first user to the eighth user, the weights increase linearly from 0.5 to 1.5.

B. Evaluate the Gain due to Power Allocation

In Fig. 7, we verify the performance gain due to power allocation using Algorithm S1. We plot both the asymptotic and simulated weighted sum-rates averaged over a single large fading block versus sum power constraint P_T. The number of antennas S is fixed as $S = M = 16$. The performance without power allocation, i.e., $\mathbf{P} = c\mathbf{I}$ in (1), is also given for comparison. When power allocation is considered and optimized, a higher weighted sum-rate can be achieved compared to the case when only the regularization factor ρ is optimized. Note that as P_T increases, the weighted sum-rate get saturated due to per antenna power constraint.

C. Performance Gain of the Proposed Scheme w.r.t. Baseline

In Fig. 8 and 9, we compare the performance of the proposed algorithm with the traditional antenna selection baseline algorithm described in Section IV-D. The power allocation and ρ are optimized using

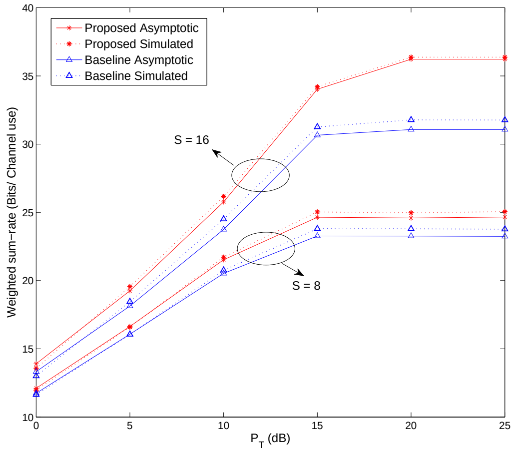

Figure 9. Comparison of proposed antenna selection scheme and baseline for uniformly distributed users

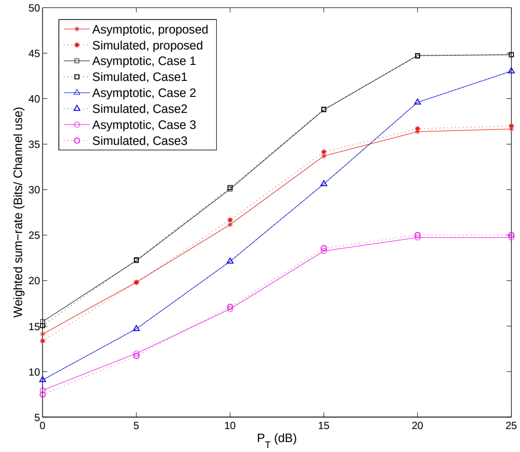

Figure 10. Comparison of asymptotic and simulated weighted sum-rates for different choices of transmit antennas

Algorithm S1 and S2 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[3]. In Fig. 8, 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. 9, we consider normal case where the users are uniformly distributed. Similar results as in Fig. 8 can be observed, although the performance gain is smaller compared to the strong cross link case.

D. Comparison with Performance Upper and Lower Bounds

In Fig. 10, 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.

[3]The baseline algorithm can be extended to select $S \leq n \times K$ 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.

We plot the weighted sum-rates averaged over different realizations of user locations versus sum power constraint P_T. 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 serving K users using RZF precoding. The objective is to maximize 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 ζ is the path 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) Deterministic equivalent for $\gamma_k\left(\mathbf{\Sigma},\rho,\mathbf{p}\right)$: We first derive the deterministic equivalent for A_k and B_k, then the deterministic equivalent for $\gamma_k\left(\mathbf{\Sigma},\rho,\mathbf{p}\right)$ follows immediately.

A_k can be rewritten as $A_k = \boldsymbol{\sigma}_k^2\tilde{\mathbf{h}}_k^\dagger\left(\hat{\mathbf{H}}_k^\dagger\hat{\mathbf{H}}_k+\rho\mathbf{I}_S\right)^{-1}\tilde{\mathbf{h}}_k$, where $\boldsymbol{\sigma}_k^2 = \mathrm{diag}\left(\boldsymbol{\sigma}_{k\mathcal{A}_1}^2,...,\boldsymbol{\sigma}_{k\mathcal{A}_S}^2\right)$, and the elements of $\tilde{\mathbf{h}}_k$ are i.i.d. complex random variables with zero mean, variance $1/S$. Applying [26, Corrolary 1] and [27, Lemma 2.1] one by one, we have

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

as $S \stackrel{K=S\beta}{\longrightarrow} \infty$. Applying [11, Theorem 1] to the trace term in (40), we have $A_k \stackrel{a.s}{\longrightarrow} \xi_k$ as $S \stackrel{K=S\beta}{\longrightarrow} \infty$, where ξ_k 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

$$B_k - \frac{1}{S}\mathrm{Tr}\left(\boldsymbol{\sigma}_k^2 \mathbf{Q}_k^{-1} \hat{\mathbf{H}}_k^\dagger \mathbf{P}_k \hat{\mathbf{H}}_k \mathbf{Q}_k^{-1}\right) \stackrel{a.s}{\longrightarrow} 0, \text{ as } S \stackrel{K=S\beta}{\longrightarrow} \infty. \tag{41}$$

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

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

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^{th} and l^{th} rows are removed. Applying [26, Corrolary 1], [27, Lemma 2.1] and [11, Theorem 1] one by one, we have

$$\hat{\mathbf{h}}_l^\dagger \mathbf{Q}_{kl}^{-1}\hat{\mathbf{h}}_l \stackrel{a.s}{\longrightarrow} \frac{1}{S}\mathrm{Tr}\left(\boldsymbol{\sigma}_l^2 \mathbf{Q}_{kl}^{-1}\right) \longrightarrow \frac{1}{S}\mathrm{Tr}\left(\boldsymbol{\sigma}_l^2 \mathbf{Q}^{-1}\right) \stackrel{a.s}{\longrightarrow} \xi_l, \tag{42}$$

as $S \stackrel{K=S\beta}{\longrightarrow} \infty$. By [26, Corrolary 1], as $S \stackrel{K=S\beta}{\longrightarrow} \infty$, we have

$$\begin{aligned}
\hat{\mathbf{h}}_l^\dagger \mathbf{Q}_{kl}^{-1} \boldsymbol{\sigma}_k^2 \mathbf{Q}_{kl}^{-1} \hat{\mathbf{h}}_l \quad & \stackrel{a.s}{\longrightarrow} \quad \frac{1}{S}\mathrm{Tr}\left(\boldsymbol{\sigma}_l^2 \mathbf{Q}_{kl}^{-1} \boldsymbol{\sigma}_k^2 \mathbf{Q}_{kl}^{-1}\right) \\
& = \quad -\frac{1}{S}\frac{\partial}{\partial z}\mathrm{Tr}\left(\boldsymbol{\sigma}_l^2\left(\hat{\mathbf{H}}_{kl}^\dagger \hat{\mathbf{H}}_{kl} + \rho \mathbf{I}_S + z\boldsymbol{\sigma}_k^2\right)^{-1}\right)\Bigg|_{z=0} \\
& \longrightarrow \quad -\frac{1}{S}\frac{\partial}{\partial z}\mathrm{Tr}\left(\boldsymbol{\sigma}_l^2\left(\hat{\mathbf{H}}^\dagger \hat{\mathbf{H}} + \rho \mathbf{I}_S + z\boldsymbol{\sigma}_k^2\right)^{-1}\right)\Bigg|_{z=0}.
\end{aligned} \tag{43}$$

where the last line follows from [27, Lemma 2.1]. Applying [11, Theorem 1] to the last trace term in (43), and calculate the partial derivative over z, it can be shown that

$$\hat{\mathbf{h}}_l^\dagger \mathbf{Q}_{kl}^{-1} \boldsymbol{\sigma}_k^2 \mathbf{Q}_{kl}^{-1} \hat{\mathbf{h}}_l \stackrel{a.s}{\longrightarrow} \theta_{kl}, \text{ as } S \stackrel{K=S\beta}{\longrightarrow} \infty, \tag{44}$$

where θ_{kl} is defined in (10) and the fact that $\mathbf{I}_K - \mathbf{D}$ is invertible has been proved in [11]. Combining the above results, we have $B_k \stackrel{a.s}{\longrightarrow} \frac{1}{S}\sum_{l\neq k}^{K}\left[p_l\theta_{kl}/\left(1+\xi_l\right)^2\right]$ as $S \stackrel{K=S\beta}{\longrightarrow} \infty$.

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

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

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

$$\mathbf{1}_j^T \hat{\mathbf{H}}^\dagger \tilde{\mathbf{Q}}_j^{-1} \mathbf{P} \tilde{\mathbf{Q}}_j^{-1} \hat{\mathbf{H}} \mathbf{1}_j \stackrel{a.s}{\longrightarrow} \frac{\rho^{-1}}{K}\sum_{i=1}^{K} \sigma_{i\mathcal{A}_j}^2 \left(p_i v_i - \varphi_i\right), \text{ as } S \stackrel{K=S\beta}{\longrightarrow} \infty,$$

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

$$\mathbf{1}_j^T \hat{\mathbf{H}}^\dagger \tilde{\mathbf{Q}}_j^{-1} \hat{\mathbf{H}} \mathbf{1}_j \xrightarrow{a.s} \frac{1}{S} \sum_{i=1}^{K} \sigma_{i\mathcal{A}_j}^2 v_i, \text{ as } S \overset{K=S\beta}{\longrightarrow} \infty.$$

Combining the above results, we show that as $S \overset{K=S\beta}{\longrightarrow} \infty$, $p_m(\mathbf{\Sigma}, \rho, \mathbf{p}) \xrightarrow{a.s} \bar{p}_m(\mathbf{\Sigma}, \rho, \mathbf{p})$ in (16).

Finally, $(\rho \mathbf{I}_K + \mathbf{\Delta} - \mathbf{C})^{-1}$ is invertible because it is a diagonally dominant matrix. This completes the proof of Lemma 1 and Lemma 2.

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

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

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

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

$$\begin{aligned} e_k &= \sum_{l=1}^{K} \left(\frac{w_l g_{lk} \sum_{i=1}^{K} \tilde{p}_i(\mathcal{A},\rho) \frac{\partial g_{li}}{\partial \rho}}{\left(g_{ll}\tilde{p}_l(\mathcal{A},\rho) + \tilde{\Omega}_l\right)^2} - \frac{w_l \frac{\partial g_{lk}}{\partial \rho}}{g_{ll}\tilde{p}_l(\mathcal{A},\rho) + \tilde{\Omega}_l} \right) \\ &\quad + \sum_{l \neq k}^{K} \left(\frac{w_l \frac{\partial g_{lk}}{\partial \rho}}{\tilde{\Omega}_l} - \frac{w_l g_{lk} \sum_{i \neq l}^{K} \tilde{p}_i(\mathcal{A},\rho) \frac{\partial g_{li}}{\partial \rho}}{\tilde{\Omega}_l^2} \right). \end{aligned}$$

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

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

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

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

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

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

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

Define $\mathcal{J}_1 = \left\{j : \bar{p}_{\mathcal{A}_j}(\mathbf{\Sigma}, \rho, \tilde{\mathbf{p}}(\mathcal{A},\rho)) < \bar{p}_{\mathcal{A}_j}\right\}$ and $\mathcal{K} = \{k : \tilde{p}_k(\mathcal{A},\rho) > 0\}$. If $\bar{p}_{\mathcal{A}}(\mathbf{\Sigma}, \rho, \tilde{\mathbf{p}}(\mathcal{A},\rho)) < P_T$, 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 $\frac{\partial \tilde{\mathbf{p}}(\mathcal{A},\rho)}{\partial \rho}$, the derivative $\frac{\partial \hat{\mathcal{I}}(\mathcal{A},\rho)}{\partial \rho}$ can be calculated using (32).

To complete the calculation of $\frac{\partial \hat{\mathcal{I}}(\mathcal{A},\rho)}{\partial \rho}$, 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 $\vec{\theta}_k$, $\mathbf{v}$, $\mathbf{\Delta}$ and $\mathbf{C}$ defined in Lemma 1 and Lemma 2, the partial derivatives of them with respective to ρ are given below.

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

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

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

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

$$\frac{\partial D_{ln}}{\partial \rho} = \frac{1}{S^2} \sum_{m \in \mathcal{A}} \left[\frac{2\sigma_{lm}^2 \sigma_{nm}^2}{(1+\xi_n)^2} \left[\frac{1}{S} \sum_{i=1}^{K} \frac{\sigma_{im}^2}{1+\xi_i} \left(\frac{\phi_i}{1+\xi_i} - \frac{\phi_n}{1+\xi_n} \right) - 1 - \frac{\rho \phi_n}{1+\xi_n} \right] / f_m^3 \left(\vec{\xi} \right) \right].$$

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

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

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

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.

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

where $\frac{\partial \xi_k}{\partial \rho} = \phi_k$, $\forall k$ 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

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

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

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

Finally, $\frac{\partial \mathbf{R}}{\partial \rho}=\left[\begin{array}{cc}\mathbf{I}_S, & \mathbf{1}\end{array}\right]^T\frac{\partial \tilde{\mathbf{R}}}{\partial \rho}$ and $\frac{\partial \tilde{\mathbf{R}}}{\partial \rho}$ is given by

$$\begin{aligned}\frac{\partial \tilde{\mathbf{R}}}{\partial \rho} &= \left(\frac{\partial \hat{\mathbf{R}}}{\partial \rho}\right)^T\left(\mathbf{V}-\left(\rho\mathbf{I}_K+\boldsymbol{\Delta}-\mathbf{C}\right)^{-1}\bar{\mathbf{R}}\right)+\hat{\mathbf{R}}^T\left(\frac{\partial \mathbf{V}}{\partial \rho}-\left(\rho\mathbf{I}_K+\boldsymbol{\Delta}-\mathbf{C}\right)^{-1}\frac{\partial \bar{\mathbf{R}}}{\partial \rho}\right)\\ &\quad+\hat{\mathbf{R}}^T\left(\mathbf{I}_K+\frac{\partial \boldsymbol{\Delta}}{\partial \rho}-\frac{\partial \mathbf{C}}{\partial \rho}\right)\left(\rho\mathbf{I}_K+\boldsymbol{\Delta}-\mathbf{C}\right)^{-2}\bar{\mathbf{R}},\end{aligned}$$

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

C. Proof of Theorem 5

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

$$\hat{\gamma}_k\left(\rho,\mathbf{p}^*\left(\mathcal{A},\rho\right)\right)=\frac{Sw_k\left(1+\sigma_{k1}^2u\right)^2\sigma_{k1}^2\left(2\rho uP_T^{'}+\left(\beta-1\right)P_T^{'}+\frac{1}{S}\sum_{l=1}^{K}\frac{1}{\sigma_{l1}^2}\right)}{\left(\sum_{l=1}^{K}w_l\right)\left(\sigma_{k1}^2P_T^{'}+\left(1+\sigma_{k1}^2u\right)^2\right)}-1, \tag{49}$$

and the objective of $\mathcal{P}_2(\mathcal{A})$ is given by $\bar{\mathcal{I}}\left(\mathcal{A},\rho,\mathbf{p}^*\left(\mathcal{A},\rho\right)\right)=\sum_{k=1}^{K}w_k\log\left(1+\hat{\gamma}_k\left(\rho,\mathbf{p}^*\left(\mathcal{A},\rho\right)\right)\right)$. For any k, it can be shown that the solution $\tilde{\rho}_k$ of $\frac{\partial}{\partial \rho}\hat{\gamma}_k\left(\rho,\mathbf{p}^*\left(\mathcal{A},\rho\right)\right)=0$ must satisfy $\tilde{\rho}_k=O\left(\frac{1}{P_T^{'}}\right)$. Since the optimal regularization factor ρ^* 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 $\frac{\partial^2}{\partial^2\rho}\hat{\gamma}_k\left(\rho,\mathbf{p}^*\left(\mathcal{A},\rho\right)\right)<0$ and thus $\hat{\gamma}_k\left(\rho,\mathbf{p}^*\left(\mathcal{A},\rho\right)\right)$ is concave when ρ is small enough. Since $\bar{\mathcal{I}}\left(\mathcal{A},\rho,\mathbf{p}^*\left(\mathcal{A},\rho\right)\right)$ is a concave increasing function of $\hat{\gamma}_k\left(\rho,\mathbf{p}^*\left(\mathcal{A},\rho\right)\right)$, $\bar{\mathcal{I}}\left(\mathcal{A},\rho,\mathbf{p}^*\left(\mathcal{A},\rho\right)\right)$ must be a concave function of ρ [22].

D. Proof of Theorem 7

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

$$\begin{aligned}\Pr\left(r_{\min}\geq r_0\right) &= 1-\Pr\left(\min_{l\neq k} d_{kl}^u \leq r_0,\ \exists k\in\{1,...,K\}\right)\\ &\geq 1-\sum_{k=1}^{K}\Pr\left(\min_{l\neq k} d_{kl}^u\leq r_0\right)\\ &= 1-K\Pr\left(\min_{l\neq 1} d_{1l}^u\leq r_0\right)\\ &= 1-K\left[1-\left(\Pr\left(d_{12}^u\geq r_0\right)\right)^{K-1}\right]\\ &\geq 1-K\left[1-\left(1-\frac{\pi r_0^2}{R_c^2}\right)^{K-1}\right],\end{aligned}$$

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

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

$$\begin{aligned}\Pr\left(\eta>\eta_0\right) &\geq \Pr\left(\left(r_{\min}/\left(\frac{\sqrt{2}R_c}{2\sqrt{M}}\right)-1\right)^{\zeta}>\eta_0\right)\\ &= \Pr\left(r_{\min}>\frac{\sqrt{2}R_c}{2\sqrt{M}}\left(\eta_0^{1/\zeta}+1\right)\right)\\ &\geq 1-K\left[1-\left(1-\pi\left(\eta_0^{1/\zeta}+1\right)^2/\left(2M\right)\right)^{K-1}\right].\end{aligned}$$

Finally, we prove the capacity scaling by deriving an upper and a lower bound for the sum-rate. The following lemma follows directly from Assumption 3 and is useful for deriving the upper bound.

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

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

and thus

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

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

$$\begin{aligned}C_s &\leq K\mathrm{E}\left[\log\left(1+P_T^U\bar{g}_k^d X_S\right)\right]\\ &\leq K\log\left(1+P_T^U\bar{g}_k^d\mathrm{E}\left[X_S\right]\right).\end{aligned}\tag{50}$$

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

Furthermore, it follows from the lower bound provided in Appendix E that $C_s \geq O\left(K\left(\frac{\zeta}{2}-\epsilon\right)\log M\right)$. This completes the proof of Theorem 7.

E. Proof of Corollary 1

Due to (39) in Theorem 7, 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 $\bar{g}_k^d \geq G_0\left(\frac{\sqrt{2}R_c}{2\sqrt{M}}\right)^{-\zeta}$, we can show that as $M \rightarrow \infty$ with K, S fixed, the average sum-rate $\mathcal{I}_{\mathcal{A}}$ is almost surely lower bounded by

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

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

$$\begin{aligned}\mathcal{I}_{\mathcal{A}} &\overset{a.s}{\geq} K\left(1-\epsilon_2\right)\log\left(1+\frac{p_1 G_0\left(\frac{\sqrt{2}R_c}{2\sqrt{M}}\right)^{-\zeta} B_1}{1+p_1 M^{-\frac{\zeta}{2}+\epsilon_1} G_0\left(\frac{\sqrt{2}R_c}{2\sqrt{M}}\right)^{-\zeta} B_2}\right)\\ &= O\left(K\left(1-\epsilon_2\right)\left(\frac{\zeta}{2}-\epsilon_1\right)\log M\right).\end{aligned}$$

Choose ϵ_1, ϵ_2 such that $\epsilon_1 + \frac{\zeta}{2}\epsilon_2 - \epsilon_1\epsilon_2 = \epsilon$. Then we have $\mathcal{I}_{\mathcal{A}} \overset{a.s}{\geq} O\left(K\left(\frac{\zeta}{2}-\epsilon\right)\log M\right)$ as $M \rightarrow \infty$ with K, S fixed. The rest steps in S3 only increase the sum-rate by a constant. 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, Jul. 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, Jun. 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,” *in Proc. IEEE ICC 2011*, pp. 1 – 5, Jun 2011.

[11] S. Wagner, R. Couillet, M. Debbah, and D. T. M. Slock, “Large system analysis of linear precoding in correlated MISO broadcast channels under 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] J. Hoydis, S. ten Brink, and M. Debbah, “Massive MIMO in UL/DL cellular systems: How many antennas do we need?” *submitted to IEEE J. Select. Areas Commun., Special Issue on "Large-Scale Multiple Antenna Wireless Systems"*, 2011.

[14] Huawei, “Cloud RAN introduction,” *The 4th CJK International Workshop*, Sep. 2011.

[15] R. Zakhour and S. V. Hanly, “Base station cooperation on the downlink: Large system analysis,” *submitted to IEEE Trans. Info. Theory*, Jun. 2010. [Online]. Available: http://arxiv.org/abs/1006.3360

[16] R. Chen, J. Andrews, and R. Heath, “Efficient transmit antenna selection for multiuser MIMO systems with block diagonalization,” *in Proc. IEEE GLOBECOM 2007*, pp. 3499 –3503, Nov. 2007.

[17] M. Mohaisen and K. Chang, “On transmit antenna selection for multiuser MIMO systems with dirty paper coding,” *in Proc. IEEE PIMRC 2009*, pp. 3074 –3078, Sep. 2009.

[18] R. Irmer, H. Droste, P. Marsch, M. Grieger, G. Fettweis, S. Brueck, H.-P. Mayer, L. Thiele, and V. Jungnickel, “Coordinated multipoint: Concepts, performance, and field trial results,” *IEEE Communications Magazine*, vol. 49, no. 2, pp. 102 –111, february 2011.

[19] B. Chalise, L. Haering, and A. Czylwik, “Robust uplink to downlink spatial covariance matrix transformation for downlink beamforming,” *in Proc. IEEE ICC 2004*, pp. 3010 – 3014, Jun. 2004.

[20] B. Hochwald and T. Maretta, “Adapting a downlink array from uplink measurements,” *IEEE Trans. Signal Processing*, vol. 49, no. 3, pp. 642 –653, Mar 2001.

[21] J. Hoydis, R. Couillet, and M. Debbah, “Random beamforming over correlated fading channels,” *submitted to IEEE Trans. Info. Theory*, 2011. [Online]. Available: http://arxiv.org/abs/1105.0569

[22] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[23] A. Liu, V. K. N. Lau, and Y. Liu, “Local dual method for optimization of parallel MIMO B-MAC interference networks under multiple linear constraints,” *submitted to IEEE Trans. Signal Processing*, Sep. 2011; revised Apr. 2012. [Online]. Available: http://www.ee.ust.hk/~eeknlau/HKUST-Office-HomePage/Publications.html

[24] D. G. Luenberger and Y. Ye, *Linear and Nonlinear Programming*, 3rd ed. New York: Springer, 2008.

[25] 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.

[26] 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.

[27] 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.