Outage Efficient Strategies for Network MIMO with Partial CSIT

Mari Kobayashi, Member, IEEE, Sheng Yang, Member, IEEE, Mérouane Debbah, Senior Member, IEEE, and Jean-Claude Belfiore Member, IEEE

Abstract

We consider a multi-cell MIMO downlink (network MIMO) where several base-stations (BSs) with \( M \) antennas connected to a central station (CS) simultaneously serve \( K \) single-antenna user terminals (UTs). Although many works have shown the potential benefits of network MIMO, the conclusion critically depends on the underlying assumptions on available channel state information at transmitters (CSIT). This motivates us to study the impact of partial CSIT on the performance gain in terms of reliability. Assuming that each UT has a target rate fixed by some delay-sensitive applications, we wish to construct a robust strategy exploiting only partial CSIT. Building on distributed zero-forcing beamforming, we propose power allocation algorithms minimizing the outage probability for the case \( M \geq K \) and a simple user scheduling maximizing the diversity gain offered by the dimension of network MIMO for the case \( M > K \). Our main finding is that even in a realistic scenario with partial CSIT, network MIMO can be beneficial by providing high data rate with a sufficient small outage probability to individual UTs. Such merit of network MIMO has been somehow overlooked in most of existing works assuming perfect CSIT.
I. INTRODUCTION

Recently, network MIMO schemes, where neighboring Base-Stations (BSs) are connected to form an antenna array, have been proposed as a means to drastically increase the downlink capacity and solve the interference management problem of cellular systems [1]. Inspired by this result, we consider the multi-cell MIMO downlink where $B$ BSs with $M$ antennas connected to a Central Station (CS) communicate simultaneously with $K$ User Terminals (UTs) with a single antenna each. Fig. 1 illustrates an example of a multi-cell downlink system for $B = K = 3$ and $M = 3$. The channel at hand is modeled by the Multi-Input Single-Output (MISO) interference channel, defined by

$$y_k[t] = \sum_{i=1}^{B} h_{ik}^T x_i[t] + n_k[t], \quad t = 1, \ldots, T$$

where $y_k[t]$ is the channel output at UT $k$, $h_{ik} \in \mathbb{C}^{M \times 1}$ denotes the channel vector from BS $i$ to UT $k$, $n_k[t] \sim \mathcal{N}(0, 1)$ is the Additive White Gaussian Noise (AWGN), and $x_i[t] \in \mathbb{C}^{M \times 1}$ denotes the input vector transmitted by BS $i$. By letting $S_i$ denote the covariance of $x_i[t]$, we consider that each BS has the power constraint $\text{tr}(S_i) \leq P_i$ for $i = 1, \ldots, B$.

If the CS or equivalently all the BSs have perfect Channel State Information at the Transmitter (CSIT) and share the messages of all UTs, the channel at hand falls down into a classical $BM \times K$ MIMO broadcast channel with per-BS power constraints. In this case, the optimal strategy to maximize the multi-cell throughput is joint dirty-paper coding [1], [2], [3]. In order to capture the essential features of the multi-cell systems while enabling the analysis tractable, the Wyner model [4] has been widely

Fig. 1. A multicell downlink with $B = 3$ BSs and $K = 3$ UTs.
considered in the literature. In [5], the authors provide a survey on the information theoretic results on the multi-cell systems under the Wyner model for the Gaussian and fading channels with a single-antenna BS ($M = 1$). These include the per-cell downlink capacity based on the circular Wyner model [3], [6], [7] and the corresponding analysis in the different asymptotic regime such as high SNR, a large number of BSs and UTs [3], [6]. If each BS is equipped with multiple antennas ($M > 1$), the multi-cell downlink capacity can be naturally enhanced by exploiting the spatial degrees of freedom (see for example [8], [9], [10]). For a small $B$, the MIMO multi-cell downlink channel is also referred to the MIMO interference channel or MIMO-X channel under various message sharing assumptions (see [8], [11] and references therein). In these contributions, the sum degrees of freedom has been extensively studied.

Unfortunately, the global joint processing at the CS is difficult (if not impossible) and the net benefits of network MIMO critically depend on how to deal with the following realistic limitations. First, obtaining accurate and full CSI at the CS is extremely challenging. Notice that perfect CSIT may not be available even in a classical cellular system due to time-varying nature of wireless channels, limited resource for channel estimation, and strictly causal estimation process. In a network MIMO where all UTs are served simultaneously by cooperative BSs, the number of channel coefficients increases drastically and acquisition of CSIT becomes much more critical. Moreover, if CS attempts to obtain CSI of all UTs over the fast fading channels, the related overhead becomes substantial as the number of cooperative BSs, antennas, UTs increases. This in turn leaves few resource for the useful data transmission and other signaling necessary for the BS cooperation. A non-trivial tradeoff between the benefits of network MIMO and the overhead related to channel estimation has been studied for the case of the uplink in [12]. Second, the backbone links between the BSs and the CS are typically imperfect. They might be the capacity-limited [13], [14], [15], [16] erroneous, or delayed [17]. The backhaul imperfectness will prevent the BSs from fully sharing the side information on the messages, codebooks, or CSI of UTs. The impact of finite-capacity backhaul link has been extensively studied (e.g. [13], [14], [15] and references therein). [13], [14] characterized the achievable multi-cell throughput while [15] considered different precoder designs and the related optimization problems. Finally, the difficulties related to connecting all BSs to a CS have motivated the concept of local connectivity, i.e. neighboring BSs are directly connected each other without relying on the CS and expensive backhaul links [18], [19], [20]. Based on this architecture, recent works (e.g. [19], [21]) considered distributed strategies which require the side information sharing between adjacent BSs.

Most of recent contributions have attempted to design practical network MIMO schemes to overcome one of the realistic limitations mentioned above (see the tutorial [22] for a complete list of references).
Our work is no exception. We aim to design a robust network MIMO scheme exploiting partial CSI at
the CS and BSs. To this end, we assume that each BS $i$ has local CSIT, i.e., knows $h_{i1}, \ldots, h_{iK}$ while the
CS has only statistical CSIT. Albeit idealized, this assumption, also considered in [23], captures some
relevant aspects. On one hand, it calls for distributed strategies both at BSs and CS. On the other hand, CS
typically needs to track the downlink channels at a rate much slower than their coherence time. We further
assume that each UT has an individual target rate to support some delay-sensitive application. Under this
setting, our goal is on minimizing the outage probability, defined as the probability that that target rate
set of all UTs cannot be supported simultaneously. This scenario, although simplified, is inspired by the
current/next wireless standards [24], [25] which aim to offer high data rate with a sufficient reliability
to individual UTs. It is worth noticing that the outage probability or the diversity gain, i.e. the decaying
rate of the outage probability at high SNR, has been identified as a key measure reflecting the reliability
in wireless communication systems. First, we address the case where the number of UTs is smaller than
the number of transmit antennas, i.e., $K \leq M$. In this case, we propose distributed zero-forcing (ZF)
beamforming applied at each BS, which creates $K$ parallel MISO channels. For the equivalent parallel
MISO channels, we propose the rate-balancing algorithm A1 for perfect CSIT and the iterative algorithm
A2 for statistical CSIT. Both algorithms aim to minimize the outage probability. Next, we consider a
more relevant case of a large number of UTs ($K > M$). In this case, a reasonable set of $\tilde{K} < M$ UTs
must be selected beforehand in order to apply distributed ZF beamforming. Since the minimization of
the outage probability is untractable, we focus on the diversity metric and propose a simple scheduling
scheme called distributed diversity scheduling (DDS). The main technical contributions are;

- for the case of perfect CSIT with $K \leq M$, we prove that Algorithm A1 minimizes the outage
  probability (Theorem 2).
- for the case of statistical CSIT with $K \leq M$, we derive Algorithm A2 which minimizes the
  approximated outage probability and prove that Algorithm A2 converges the optimal solution in
  most cases of interest (Theorem 3).
- for the case of a large number UTs $K > M$, we prove that DDS offers a diversity gain of $B\frac{K}{\tilde{K}}(M - \tilde{K} + 1)$ to each UT and that this gain scales optimally in $B, K,$ and $M$, respectively (Theorem 4).

Numerical results validate the analysis in terms of diversity gain and show that our proposed distributed ZF
beamforming significantly outperforms the non-cooperative scheme by improving the outage performance
of our proposed scheme with the number of cooperative BSs, transmit antennas, and UTs. It is also
shown in a simple one-dimensional topology that the performance gain can be emphasized as the path-
loss between neighboring cells decreases. The most striking conclusion of this paper is that even in a realistic scenario with partial CSIT, network MIMO can be beneficial by providing high data rate with a sufficient small outage probability to individual UTs. Such merit of network MIMO has been somehow overlooked in most of existing works assuming perfect CSIT.

Finally, we remark that some realistic constraints of the underlying network are ignored in this work in order to highlight the impact of partial CSIT. In particular, the backbone links are assumed to be perfect such that $B$ BSs can fully share the messages. As discussed previously, the full message sharing may not be possible in a network architecture where neighboring BSs talk each other without CS. However, it is beyond our scope to compare different options for message sharing depending on the configuration as the network architecture is still an open issue. We refer to [16] for the optimization of the amount of message sharing over the finite-capacity backhaul links. A more refined analysis which accounts for the backhaul overhead of different side information (UT messages, CSI, and resource allocation parameters) along the line of [26] remains as a future investigation.

The rest of the paper is organized as follows. Section II describes the system model. In Section III, we present the power allocation policies that minimize the outage probability both under perfect and statistical CSI at the CS. Then, we consider the relevant case of $K \geq M$ in Section IV where we propose a simple user scheduling algorithm. Section V shows some numerical results in different scenarios. Finally, Section VI concludes the paper.

**Notations:** Throughout the paper, we use boldface lower case letters to denote vectors, boldface capital letters to denote matrices. $(\cdot)^*$, $(\cdot)^T$, and $(\cdot)^H$ respectively denote the complex conjugation, matrix transposition, and Hermitian transposition operations. $I_n$ and $0_{n \times m}$ represent the $n \times n$ identity matrix and $n \times m$ zero matrix. The determinant, rank, trace, and Frobenius norm of a matrix $A$ are denoted by $|A|$, rank$(A)$, tr$(A)$, and $\|A\|_F^2$, respectively. The dot-equality stands for the equality on the near-zero decay rate, that is, $f(\epsilon) \overset{\epsilon n}{=} \epsilon^n$ means $\lim_{\epsilon \to 0} \frac{\log f(\epsilon)}{\log \epsilon} = n$ and $f(\epsilon) \overset{g(\epsilon)}{=} g(\epsilon)$ means $\lim_{\epsilon \to 0} \frac{\log f(\epsilon)}{\log \epsilon} = \lim_{\epsilon \to 0} \frac{\log g(\epsilon)}{\log \epsilon}$. We let $\chi^2_m$ denote the chi-square distribution with $m$ degrees of freedom.

**II. DISTRIBUTED ZERO-FORCING BEAMFORMING**

In order to split the processing at the CS and the BSs, it is reasonable to assume that the CS generates the messages and performs the resource allocation whereas each BS encodes and then transmits the symbols

---

1If some backbone links are in failure, the corresponding BSs fail to encode the messages to $K$ UTs. This will reduce the effective number of BSs.
in a distributed fashion. More precisely, we assume that the CS broadcasts $K$ messages intended to $K$ UTs which enable the BSs to cooperate in transmission. Then, each BS $i$ encodes these messages into $K T$ symbols $\{s_{ik}^t\}$ for $k = 1, \ldots, K$ and $t = 1, \ldots, T$ with some capacity-achieving code (e.g., Gaussian random coding with a sufficiently large $T$ channel uses). Under this setting, this section focuses on a distributed precoder design at each BS which requires only local CSIT $\{h_{ik}\}_{k=1}^K$ for any $i$.

In the multi-cell downlink channel (1), we model the vector $h_{ik}$ of channel coefficients from BS $i$ to UT $k$ Gaussian distributed $\sim \mathcal{N}(0, \sigma_{ik}^2 I_M)$ where the variance $\sigma_{ik}^2$ captures the path-loss of the corresponding link assuming that the UTs are arbitrary distributed. Furthermore, $\{h_{ik}\}$ are assumed to be independent over any pair $(i, k)$. We start with the definition of zero-forcing beamforming vectors as well as a useful lemma on the resulting channel statistics.

**Definition 1 (Zero-forcing beamforming vectors):** For a channel matrix $H \in \mathbb{C}^{K \times M}$ with $K \leq M$ linearly independent row vectors $h_1^T, \ldots, h_K^T$, there exists a zero-forcing beamforming matrix $G = \begin{bmatrix} g_1 & \cdots & g_K \end{bmatrix} \in \mathbb{C}^{M \times K}$ with

$$G = H^+ \text{diag} \{a_k\}$$

(2)

where $H^+ \overset{\Delta}{=} H^H \left( H H^H \right)^{-1}$ is the Moore-Penrose pseudoinverse of $H$ and diag $\{a_k\}$ is a diagonal matrix that normalizes the norm of the columns of $H^+$ such that $\|g_k\| = 1, \forall k$.

**Lemma 1:** If $H \in \mathbb{C}^{K \times M}$ has i.i.d. entries where each row $h_k^H \sim \mathcal{N}(0, \sigma_k^2 I)$, then $|a_k|^2$ as defined in (2) is $\chi^2_{2(M-K+1)}$ distributed with mean $\mathbb{E}[|a_k|^2] = (M - K + 1) \sigma_k^2, \forall k$.

We consider a simple ZF beamforming which enables each UT to achieve a multiplexing gain of one. At each channel use $t$, BS $i$ applies the ZF beamforming to transmit $K$ symbols in a distributed manner. Due to the symmetry over all channel uses, we ignore the index $t$ hereafter. We form the transmit vector of BS $i$ at any channel use as

$$x_i = G_i u_i = \sum_{k=1}^K \sqrt{p_{ik}} g_{ik} s_{ik}$$

(3)

where we let $G_i \overset{\Delta}{=} \begin{bmatrix} g_{i1} & \cdots & g_{iK} \end{bmatrix}$ be the ZF beamforming matrix corresponding to the channel matrix from BS $i$ formed by row vectors $h_{i1}^H, \ldots, h_{iK}^H$, $u_i \overset{\Delta}{=} \begin{bmatrix} \sqrt{p_{i1}} s_{i1} & \cdots & \sqrt{p_{iK}} s_{iK} \end{bmatrix}^T$ is the vector of symbols where $s_{ik} \sim \mathcal{N}(0, 1)$ denotes the symbol transmitted by BS $i$ to UT $k$ with power $p_{ik}$. With this beamforming, the received signal at UT $k$ is given by

$$y_k = \sum_{i=1}^B \sqrt{p_{ik}} a_{ik} s_{ik} + n_k$$

(4)
where $a_{ik} \triangleq h_{ik}g_{ik}$ denotes the overall channel from BS $i$ to UT $k$ and coincides with the definition in (2) for each $i$. From Lemma 1 and the independence of $\{a_{ik}\}$ over $i$, we remark that the original $B \times K$ MISO interference channel (1) is decoupled into $K$ parallel $B \times 1$ MISO channels.

**Lemma 2:** Let $y = Hx + z$ denote a $n_r \times n_t$ MIMO slow fading channel with standard complex Gaussian noise $z$. Then, the maximum diversity $d$ of the channel is

$$d = \lim_{\epsilon \to 0} \frac{\log \left( \Pr \left( \|H\|_F^2 < \epsilon \right) \right)}{\log \epsilon}.$$  

(5)

We can also use the short-hand notation $\Pr \left( \|H\|_F^2 < \epsilon \right) \doteq \epsilon^d$.

**Proof:** See Appendix A. 

**Theorem 1:** With distributed ZF beamforming, the diversity order of each UT is $B(M - K + 1)$.

**Proof:** See Appendix B.

Before proceeding further, we provide the following definitions that we use repeatedly in the rest of the paper. We define two $B \times K$ matrices of coefficients $A \triangleq \{a_{ik}\}$ and $P \triangleq \{p_{ik}\}$, $i = 1, \ldots, B$, $k = 1, \ldots, K$. Moreover, we define the following vectors $a_k \in \mathbb{C}^{B \times 1}$, $p_{-k} \in \mathbb{C}^{B \times 1}$, $p_i \in \mathbb{C}^{K \times 1}$

$$a_k \triangleq \begin{bmatrix} a_{1k} & \cdots & a_{Bk} \end{bmatrix}^T,$$

(6)

$$p_{-k} \triangleq \begin{bmatrix} p_{1k} & \cdots & p_{Bk} \end{bmatrix}^T,$$

(7)

$$p_i \triangleq \begin{bmatrix} p_{i1} & \cdots & p_{iK} \end{bmatrix}^T.$$  

(8)

Assuming that each UT $k$ perfectly knows the channel state $a_k$, it can perform the decoding and achieves the following rate

$$R_k = \log \left( 1 + \sum_{i=1}^B |a_{ik}|^2 p_{ik} \right).$$

(9)

The capacity region of the $K$ parallel MISO channels (4) for a fixed set of transmit power $P$ and channel state $A$ is given by

$$\mathcal{R}(A; P) = \left\{ (R_1, \ldots, R_K) \in \mathbb{R}^K_+ : R_k \leq \log \left( 1 + \sum_{i=1}^B |a_{ik}|^2 p_{ik} \right), \forall k \right\}.$$  

(10)

The above region is clearly convex (rectangular for $K = 2$). The capacity region of the $K$ parallel MISO channels (4) subject to the individual BS power constraints

$$\sum_{k=1}^K p_{ik} \leq P_i, \quad \forall i$$

(11)

for a fixed channel state $A$ is given by

$$\mathcal{C}(A; P_1, \ldots, P_B) = \bigcup_{P \in \mathcal{P}(A)} \mathcal{R}(A; P).$$  

(12)
where the matrix function $F$ denotes a power allocation policy $A \mapsto P$ that maps the channel state $A$ into the power coefficients $P$; $F$ denotes the feasibility set satisfying the power constraints (11) for any channel realization $A$. The capacity region $\mathcal{C}(A; P_1, \ldots, P_B)$ is convex and its boundary can be explicitly characterized by solving the weighted sum rate maximization as specified in subsection III-A.

III. POWER ALLOCATION MINIMIZING OUTAGE PROBABILITY

In the previous section, we considered the distributed processing at each BS assuming that the power allocation is already done at the CS and each BS is informed about the resulting power partition. Here, we address the power allocation problem solved at the CS. Assuming that some delay-sensitive application imposes a set of target rate to each UT, we wish to find the power allocation minimizing the outage probability simultaneously for all UTs. In order to formalize the problem, we let $\gamma_k$ denote the target rate of UT $k$ and form the target rate tuple $\gamma = (\gamma_1, \ldots, \gamma_K)$. For a given $\gamma$, we define the outage probability as the average probability that $\gamma$ is not fulfilled by all $K$ UTs simultaneously, namely

$$P_{\text{out}}(\gamma) \triangleq 1 - \mathbb{E}_A \left[ 1 \{ \gamma \in \mathcal{C}(A; P_1, \ldots, P_B) \} \right] = 1 - \Pr (\gamma \in \mathcal{C}(A; P_1, \ldots, P_B)).$$

This section provides the power allocation policies minimizing the outage probability defined above under perfect and statistical CSIT.

A. Perfect CSI at CS

We start with a special case where the CS has perfect CSIT.\footnote{It is worth noting that even with perfect CSI, the outage event can occur if the desired rate tuple is outside of the achievable rate region related to the instantaneous channel gain.} In this case, we are particularly interested in the power allocation policy that provides the rate tuple proportional to the target rate tuple (rate-balancing). As seen shortly, this policy equalizes the individual outage probability of all UTs and thus provides the strict fairness to UTs. Our objective is to find the set of $\{p_{ik}\}$ satisfying

$$\frac{R_k(P)}{R_1(P)} = \frac{\gamma_k}{\gamma_1} \triangleq \alpha_k, \quad k = 2, \ldots, K$$

where we defined $\frac{\gamma_k}{\gamma_1} = \alpha_k$ and $\alpha_1 = 1$. More precisely, the power allocation is a solution of [27]

$$\min_{\sum_k \theta_k = 1} \max_{(R_1, \ldots, R_K) \in \mathcal{C}} \sum_{k=1}^K \theta_k \frac{R_k}{\alpha_k}. \quad (15)$$

It is worth noting that even with perfect CSI, the outage event can occur if the desired rate tuple is outside of the achievable rate region related to the instantaneous channel gain.
Note that the inner problem for a fixed set \( \{\theta_k\} \) is convex since the objective function (weighted sum rate) is concave and the constraints are linear. Hence, it is necessary and sufficient to solve the KKT condition [28], given by

\[
\frac{\theta_k}{\alpha_k} \frac{|a_{ik}|^2}{1 + \sum_{j=1}^{B} |a_{jk}|^2 p_{jk}} = \frac{1}{\mu_i}, \quad k = 1, \ldots, K, \ i = 1, \ldots, B
\] (16)

where \( \mu_i \) denotes the Lagrangian variable to be determined such that \( \sum_k p_{ik} \leq P_i \). Although a closed-form solution does not exist, the iterative multiuser waterfilling algorithm for the MIMO multiple access channel [29] can be easily modified to solve the KKT conditions (16) iteratively. Namely, at each iteration, we find a new power vector \( \mathbf{p}_i \) related to BS \( i \) by treating the powers of the other BSs \( \{\mathbf{p}_l\}_{l \neq i} \) constant, by computing

\[
\mathbf{p}^{(n)}_{ik} = \left[ \frac{\theta_k \mu_i}{\alpha_k} - \frac{1}{|a_{ik}|^2} \frac{1 + \sum_{j \neq i} |a_{jk}|^2 p^{(n)}_{jk}}{\mu_i} \right]^{+}, \quad \forall k
\] (17)

The outer problem consists of minimizing the solution of the inner problem with respect to \( \theta_2, \ldots, \theta_K \). Since the problem at hand is convex, a \((K - 1)\)-dimensional subgradient method can be suitably applied [30]. We summarize the overall algorithm as follows.

**Algorithm A1: Rate balancing algorithm:**

1) Initialize \( \theta_k^{(0)} \in [0, 1] \) for \( k > 1 \).

2) At iteration \( n \), for the fixed set of weights \( \theta_k^{(n)} \) compute the optimal rates \( R_1^{(n)}, \ldots, R_K^{(n)} \), solution of the following maximization

\[
\max_{\mathbf{c}(\mathbf{A}; \mathbf{P}_1, \ldots, \mathbf{P}_B)} \sum_{k=1}^{K} \theta_k^{(n)} \frac{R_k}{\alpha_k}
\] (18)

via waterfilling approach (17). Compute a new subgradient.

\[
\Delta_k^{(n)} = \frac{R_k^{(n)}}{\alpha_k} - R_1^{(n)}, \quad \forall k
\]

3) Update the weights by subgradient:

\[
\theta_k^{(n+1)} = \theta_k^{(n)} - \xi_n \Delta_k^{(n)}, \quad \forall k
\]

where \( \xi_n \) denotes the step size at iteration \( n \) following a decreasing rule [30].

The overall algorithm implements the rate-balancing by allocating the rates proportional to the target rate tuple. Hereafter, we let \( \mathbf{P}^*(\mathbf{A}) \) and \( \{R_k^*(\mathbf{A})\} \) denote respectively the optimal power and rates found by the above algorithms as a function the channel state \( \mathbf{A} \) since the achievable rate (9) is a deterministic function of \( \mathbf{P}^*(\mathbf{A}) \). Then, we have the following result.
Theorem 2: The rate-balancing power allocation minimizes the outage probability. Moreover, it equalizes the individual outage probability of all $K$ UT by letting $\Pr(\gamma_1 < R_1^*(A)) = \cdots = \Pr(\gamma_k < R_k^*(A))$.

Proof: A sketch of proof is provided as follows. First, we remark that only the case when the target rate tuple $\gamma$ is inside the region $R(\gamma_1, \gamma_2)$ needs to be considered (e.g., case (b) in Fig.2). In this case, we define the subregion $R_s(\gamma_1, \gamma_2)$

$$
R_s(\gamma_1, \gamma_2) \triangleq \{(R_1, \ldots, R_K) | (R_1, \ldots, R_K) \in R(\gamma_1, \gamma_2), R_k \geq \gamma_k, k = 1, \ldots, K\}
$$

(19)
depicted in a shadow area in Fig. 2 (b). We define $F_s$ as a class of the power allocation policies that maps $a$ into the rate tuple $R$ inside $R_s(a)$ whenever we have $\gamma \in R(\gamma_1, \gamma_2)$. Note that any policy belonging to $F_s$ results in a successful transmission, and thus minimizes the outage probability. Since the proposed rate balancing scheme allocates the power $P^*$ so that the rate-tuple $(R_1^*, \ldots, R_K^*)$ is proportional to $\gamma$ on the boundary of $R(\gamma_1, \gamma_2)$ whenever $\gamma \in R(\gamma_1, \gamma_2)$, it belongs to the class $F_s$. This establishes the first part.

For second part, it is immediate to see that with the rate balancing scheme, we have for any $A$

$$
1(\gamma_k < R_k^*(A)) = 1(\alpha_k \gamma_1 < \alpha_k R_1^*(A)) = 1(\gamma_1 < R_1^*(A)), \quad k = 2, \ldots, K
$$

(20)

where the first equality follows from (14). It is readily shown that

$$
P_{\text{out}}^{\text{balance}}(\gamma) = 1 - \Pr(\cap_{k=1}^K \{\gamma_k < R_k^*(A)\})$$

$$
= 1 - \Pr(\gamma_k < R_k^*(A))
$$

where the last equality follows from (20).

B. Statistical CSI at CS

We consider a more realistic case when the CS has only statistical knowledge of the equivalent channels $A$. Formally, we define the power allocation policy $F^{\text{stat}} : \Sigma \mapsto P$ as a function mapping the variances...
of channel coefficients $\Sigma \triangleq \{ \sigma_{ik}^2 \}$ into the transmit power $P$. Since the equivalent channel coefficients in $A$ after ZF beamforming are correlated over different $k$ for each BS $i$, the individual outage events are dependent. Nevertheless, this dependency can be made arbitrarily small if a simple interleaving can be performed over frequency bands for example Sheng, can you explain this better? Under this assumption, we approximate the outage probability for a fixed power allocation $P$ as

$$P_{\text{app-out}}(\gamma, P) = 1 - \prod_{k=1}^{K} \Pr (\Delta_k(P_{-k}) > 2^{\gamma_k} - 1)$$

where we let $\Delta_k(P_{-k}) = \sum_{i=1}^{B} |a_{ik}|^2 p_{ik}$. First, we remark that for a fixed power $P_{-k}(P_{-k}) = (p_{1k}, \ldots, p_{Bk})$ of UT $k$, $\Delta_k(P_{-k})$ is a Hermitian quadratic form of complex Gaussian random variables given by

$$\Delta_k(P_{-k}) = (a_{1k}^* \cdots a_{Bk}^*) \begin{pmatrix} p_{1k} & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & p_{Bk} \end{pmatrix} \begin{pmatrix} a_{1k} \\ \vdots \\ a_{Bk} \end{pmatrix}$$

$$= (w_{1k}^H \cdots w_{Bk}^H) \begin{pmatrix} p_{1k}I & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & p_{Bk}I \end{pmatrix} \begin{pmatrix} w_{1k} \\ \vdots \\ w_{Bk} \end{pmatrix}$$

where the second equality follows by replacing a chi-square random variable $|a_{ik}|^2 \sim \chi^2_{2(M-K+1)}$ with $||w_{ik}||^2$ where $w_{ik} \sim \mathcal{C}(0, \sigma_{ik}^2 I_{M-K+1})$. The individual outage probability that UT $k$ cannot support $\gamma_k$ for a fixed $P_{-k}$ is [31]

$$\Pr (\Delta_k(P_{-k}) \leq c_k) = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} e^{s c_{k}} \Phi_{\Delta_k(P_{-k})}(s) \frac{ds}{s}$$

where we let $c_k = 2^{\gamma_k} - 1$ denote the target SNR of UT $k$, and the Laplace transform of $\Delta_k(P_{-k})$ is given by

$$\Phi_{\Delta_k(P_{-k})}(s) = \prod_{i=1}^{B} \frac{1}{(1 + sp_{ik} \sigma_{ik}^2)^{M-K+1}}.$$

It is not difficult to see from (23) that each UT achieves a diversity gain of $B(M-K+1)^3$, which agrees well with Theorem 1. By expressing extracting the term It can be derived from (23) each UT achieves a diversity gain of $B(M-K+1)$, which agrees well with Theorem 1. The widely used upper

\[ \text{March 1, 2011 DRAFT} \]
bound is the Chernoff bound for fixed powers $p_k$ given by

$$\Pr(\Delta_k(p_k) \leq c_k) \leq \min_{\lambda \geq 0} e^{\lambda c_k} \Phi_{\Delta_k(p_k)}(\lambda) \triangleq \overline{F}(c_k, p_k).$$

(24)

Using the last expression of the Chernoff upper bound for each $k$, the approximated outage probability for a fixed $P$ is upper-bounded by

$$P_{\text{app-out}}(e, P) \leq 1 - \prod_{k=1}^{K} (1 - \overline{F}(c_k, p_k))$$

(25)

Since the power optimization based on the exact outage probability is not amenable, we search the power allocation that minimizes the Chernoff upper bound. First we define for $k = 1, \ldots, K$

$$\rho_k(\lambda_k, p_k) \triangleq \frac{e^{\lambda_k c_k}}{\prod_{i=1}^{B} (1 + \sigma_{ik}^2 \lambda_k p_{ik})^{M-K+1}}.$$ 

(26)

We consider the following maximization problem.

$$\text{maximize} \quad f(\{\lambda_k\}, \{p_{ik}\}) \triangleq \prod_{k=1}^{K} (1 - \rho_k(\lambda_k, p_k))$$

(27)

subject to

$$\sum_{k=1}^{K} p_{ik} \leq P_i, \quad i = 1, \ldots, B$$

$$\lambda_k \geq 0, \quad k = 1, \ldots, K$$

$$p_{ik} \geq 0, \quad i = 1, \ldots, B; \quad k = 1, \ldots, K$$

(28)

It should be remarked that depending on the target rates and power constraints, the solution may not maximize the objective value. This will be the case if there exists $k$ such that $\rho_k(\lambda_k, p_k) \geq 1$. In order to upper bound the probability that $K$ UTs satisfy the target rate-tuple $\gamma$ by (27), we restrict ourselves to a particular set $\mathcal{F}$ of solutions $(\lambda, P)$ which satisfy the following:

$$\rho_k(\lambda_k, p_k) \in (0, 1), \quad k = 1, \ldots, K.$$ 

(29)

Hereafter, we call the solution feasible only if the solution to the maximization problem (27) satisfies (29) in addition to the constraints (28). Assuming that a feasible solution exists, we prove the concavity of the objective function (27) in Appendix C. In this case, since the objective is concave and the constraints (28) are linear, the maximization problem is convex. Although any convex tool can be applied to solve the problem, we propose an efficient numerical method to exploit the structure of the problem.

First, we remark that the optimal $\{\lambda_k\}$ is given as a solution to the polynomial of degree $B$ for a given set of $\{p_k\}$. As shown in Appendix C, the optimal $\lambda_k$ maximizing $f$ minimizes $\rho_k$ and moreover
ρ_k is a convex function of λ_k. It readily follows that the KKT condition [28], necessary and sufficient for the optimality, is given by

\[
\frac{c_k}{M - K + 1} = \prod_{i=1}^{B} \frac{\sigma_{ik}^2 p_{ik}}{1 + \lambda_k \sigma_{ik}^2 p_{ik}} \tag{30}
\]

for each k. For B = 2, the solution is given in a closed form.

\[
\lambda_k = \begin{cases} 
\frac{M-K+1}{c_k} - \frac{1}{\sigma_{ik}^2 p_{1k} + \sigma_{ik}^2 p_{2k}}, & \text{if } \prod_{j=1}^{2} \sigma_{jk}^2 p_{jk} = 0 \\
-\left(\frac{\sigma_{ik}^2 p_{1k} + \sigma_{ik}^2 p_{2k} - 2(M-K+1)\sigma_{ik}^2 p_{1k} \sigma_{ik}^2 p_{2k}}{c_k}\right) + \sqrt{\left(\sigma_{ik}^2 p_{1k} - \sigma_{ik}^2 p_{2k}\right)^2 + \left(\frac{2(M-K+1)\sigma_{ik}^2 p_{1k} \sigma_{ik}^2 p_{2k}}{c_k}\right)^2}}{2\sigma_{ik}^2 p_{1k} \sigma_{ik}^2 p_{2k}}, & \text{otherwise.} 
\end{cases} \tag{31}
\]

Since the optimal \{λ_k\} is uniquely determined as a function of \{p_{ik}\}, we can focus on the maximization of the objective function (27) with respect to the power variables \{p_{ik}\}.

Due to the concavity of f, we solve the KKT conditions on the powers. We form the Lagrangian function by introducing B Lagrangian multipliers \{µ_i\} each of which is associated to the power constraint of BS i. The KKT conditions are readily given by

\[
\sigma_{ik}^2 \lambda_k \left(1 + p_{ik} \sigma_{ik}^2 \lambda_k \right) \{e^{-\lambda_k c_k} \prod_{l=1}^{B} (1 + \lambda_k \sigma_{lk}^2 p_{lk})^{M-K+1} - 1\} = \mu_i, \quad k = 1, \ldots, K. \tag{32}
\]

for i = 1, \ldots, B. It should be remarked that when treating the powers \{p_{ik}\}_{i \neq i} of the other BSs j ≠ i fixed, the LHS of (32) is a strictly positive and monotonically decreasing function of p_{ik} for a given λ_k > 0. It remains to determine µ_i such that the power constraint of BS i is satisfied, i.e., p_{i1} + \cdots + p_{iK} = P_i. This can be done efficiently by a one-dimensional search of µ_i following a similar approach of [32, Appendix].

We summarize our proposed iterative algorithm to minimize the Chernoff upper bound, equivalently solve (27).

**Algorithm A2 : iterative algorithm for the Chernoff upper bound minimization:**

1) Initialize \( \mathbf{P}^{(0)} \) and \( \lambda^{(0)} \)

2) At iteration n

   For i = 1, \ldots, B

   - Find the new power vector \( \mathbf{p}_{i}^{(n)} \) of BS i by line search
   - Update \( \lambda^{(n)} \) by solving the polynomial (30)

   End

3) Continue until f converges
Theorem 3: Algorithm A2 converges to the optimal solution if feasible solution satisfying (29) exists.

Proof: Appendix D.

IV. EFFECT OF USER SCHEDULING

In this section we address the relevant case when the number of UTs is larger than the number of transmit antennas $K > M$. In order for each BS to apply the ZF beamforming in a distributed fashion, a set of $\tilde{K} \leq M$ UTs shall be selected beforehand. We assume that the user selection (scheduling) is handled by the CS together with the power allocation. In particular, we focus on a user selection method which achieves a high diversity order while limiting the amount of the side information necessary at the CS and the BSs. The rationale behind the adoption of diversity oriented scheduling is that an outage-optimal scheme is necessarily diversity-optimal in the high SNR regime. Moreover, deriving the outage-optimal scheduling is generally hard, if possible, due to the intractability of the outage probability. Fortunately, closed-form diversity analysis is feasible, thanks to the high SNR approximation. Furthermore, the diversity analysis can provide useful intuition and indications on system design. In the following, we present our proposed user selection scheme as well as the analysis on its achievable diversity gain.

A. Distributed Diversity Scheduling (DDS)

Let $\mathcal{S}, \mathcal{U}$ denote the set of all $K$ users UTs, the $\tilde{K}$ selected users, with $|\mathcal{S}| = K$, $|\mathcal{U}| = \tilde{K}$, respectively. Let us also define $\mathcal{Q}(\tilde{K})$ as the set of all possible user selections, i.e.,

$$\mathcal{Q}(\tilde{K}) = \left\{ \mathcal{U} \mid \mathcal{U} \subseteq \mathcal{S}, |\mathcal{U}| = \tilde{K} \right\}$$

for $\tilde{K} \leq M$. Then, the equivalent channel from the BSs to the selected users is

$$y_k = \mathbf{a}_k^T \mathbf{u}_k + z_k, \quad k \in \mathcal{U},$$

a MISO channel where we recall that $\mathbf{a}_k = [a_{1k} \cdots a_{Bk}]^T$ and $\mathbf{u}_k = [\sqrt{p_{1k}s_{1k}} \cdots \sqrt{p_{Bk}s_{Bk}}]^T$. For convenience, we only consider the diversity order of the worst user and refer it as the diversity of the system hereafter. Since the diversity order of a given channel depends solely on the Euclidean norm of the channel matrix, as shown in lemma 2, the following user selection scheme maximizes the diversity of the system

$$\mathcal{U}^* = \arg\max_{\mathcal{U}} \min_{k \in \mathcal{U}} ||\mathbf{a}_k||^2.$$

For convenience of notation, we will drop the argument $\tilde{K}$ whenever confusion is unlikely.
Unfortunately, this scheduling scheme has two major drawbacks: 1) perfect knowledge at the CS on \(\{\|a_k\|\}\), crucial for the scheduling, is hardly implementable as aforementioned, and 2) the maximization over all \(|Q(\tilde{K})| = (\frac{K}{\tilde{K}})\) possible sets \(\mathcal{U}\) grows in polynomial time with \(K\).

To overcome the first drawback, we use the following selection scheme

\[
\mathcal{U}_d = \arg \max_{\mathcal{U}} \max_{i = 1 \ldots B} \min_{\mathcal{U} \in \mathcal{P}_S} \min_{k \in \mathcal{U}} |a_{ik}|^2. \tag{35}
\]

That is, BS \(i\) finds out the set \(\mathcal{U}\) that maximizes \(\min_{k \in \mathcal{U}} |a_{ik}|^2\) and sends both the index of the set and the corresponding maximum value to the CS. Upon the reception of \(B\) values and the corresponding sets from the \(B\) BS, the CS makes a decision by selecting the largest one. Therefore, only partial channel state information is communicated in the BS-CS link. To address the second drawback, we narrow down the choices of \(\mathcal{U}\) to the following \(\kappa = \frac{K}{\tilde{K}}\) possibilities

\[
\mathcal{P}_S = \{\mathcal{U}_1, \mathcal{U}_2, \ldots, \mathcal{U}_\kappa\}, \quad \bigcup_i \mathcal{U}_i = \mathcal{S}, \quad \mathcal{U}_i \cap \mathcal{U}_j = \emptyset, \quad \forall i \neq j, \quad |\mathcal{U}_i| = \tilde{K}, \quad \forall i.
\]

In other words, \(\mathcal{P}_S\) is partition of the set of all users \(\mathcal{S}\). Furthermore, it is assumed that the partition \(\mathcal{P}_S\) is fixed by the CS and known to all BSs. Hence, the proposed scheduling scheme selects the following set of users

\[
\mathcal{U}_d = \arg \max_{i = 1 \ldots B} \max_{\mathcal{U} \in \mathcal{P}_S} \min_{k \in \mathcal{U}} |a_{ik}|^2. \tag{36}
\]

To summarize, the scheduling scheme works as follows

i) The CS fixes a partition \(\mathcal{P}_S\) and informs it to all BSs.

ii) BS \(i\) finds \(\max_{\mathcal{U} \in \mathcal{P}_S} \min_{k \in \mathcal{U}} |a_{ik}|^2\), and sends this value and the index of the maximizing set \(\mathcal{U}\) to the CS.

iii) The CS chooses the largest value and broadcasts the index of the winner set \(\mathcal{U}_d\) as defined in (36).

iv) All the BSs serve simultaneously the UTs in \(\mathcal{U}_d\).

An example of two BSs and six UTs is shown in Fig. 3. In this example, in order to serve two UTs simultaneously, a partition of three sets is fixed by the CS. With local CSI, each BS compares the coefficients \(\min_{k \in \mathcal{U}} |a_{ik}|^2\) for all three sets \(\mathcal{U}\), finds out the largest one, and sends the corresponding “index(value)” pair to the CS. The CSI compares the values and broadcasts the index of the winning set (set 1 in this example).

\(^5\text{Here, we assume that } K/\tilde{K} \text{ is integer for simplicity of demonstration. However, it will be shown that same conclusion holds otherwise.}\)
B. Diversity gain analysis

In this subsection, we analyze the diversity gain achieved by our proposed DDS and compare it with the upper bound. The result is summarized in the following Theorem.

Theorem 4: Let $K$, $B$, and $M$ denote the number of UTs, number of BSs, and number of antennas per BS. By serving $\tilde{K}$ UTs simultaneously, the following diversity gain is achievable with DDS

$$d_{UL}(\tilde{K}) \geq B \frac{K}{K} \left( M - \tilde{K} + 1 \right),$$

(37)

for $K = n\tilde{K}$ with an integer $n$. Furthermore, the optimal diversity gain achieved by (34) is upper-bounded by

$$d_{UL}(\tilde{K}) \leq B(K - \tilde{K} + 1) \left( M - \tilde{K} + 1 \right).$$

(38)

Proof: Appendix E.

Remark 4.1: For a fixed $\tilde{K}$, the diversity gain of the proposed scheduling scheme grows as $O(BKM)$, i.e., the optimal diversity scaling with $B$, $K$, and $M$. In this sense, the proposed scheme is order optimal in terms of diversity gain. Also note that the lower bound (37) and the upper bound (38) coincide in some specific settings. First, $d_{UL}(1) = d_{UL}(1) = BKM$, $\forall B, K, M$. That is, the proposed scheme is diversity optimal if only one user is served in the system. Then, for $K \leq M$, we have $d_{UL}(K) = d_{UL}(K) = B(M - K + 1)$. This corresponds to the case where all users in the system are served simultaneously.

Remark 4.2: Quite interestingly, the proposed DDS scheme is at least as good as the optimal
scheduling scheme defined in (34) if one replaces $Q$ with $P_S$ in (34). To see this, let us rewrite

$$\min_{k \in \mathcal{U}} \|a_k\|^2 = \max_{U \in P_S} \min_{k \in U} \|a_k\|^2$$

(39)

$$\leq \max_{U \in P_S} \|a_k\|^2, \quad \forall k \in \mathcal{U}$$

(40)

and that $\max_{U \in P_S} \|a_k\|^2$ is of diversity $B\tilde{K} \left(M + 1 - \tilde{K}\right)$. This observation implies that if the CS is somehow forced to optimize over a fixed partition instead of over all the possible user subsets (e.g., due to limitation of complexity, etc.), then perfect knowledge on $\{\|a_k\|\}$ does not help improve the diversity.

**Remark 4.3:** When $\tilde{K}$ does not divide $K$, we consider only $\left[\frac{K}{\tilde{K}}\right] \tilde{K}$ out of $K$ users. Since $\tilde{K}$ divides $\left[\frac{K}{\tilde{K}}\right] \tilde{K}$, the following diversity gain can be achieved

$$d_{U\cup}(\tilde{K}) = B \left[\frac{K}{\tilde{K}}\right] \left(M - \tilde{K} + 1\right)$$

(41)

with the proposed DDS.

V. NUMERICAL EXAMPLES

This section provides some numerical examples to verify the behavior of our proposed distributed ZF beamforming scheme in a simple network MIMO configuration with $B = 2$ cooperative BSs. We assume the same power constraint at both BSs $P_1 = P_2$ and let $P$ denote the SNR.

Fig. 5 shows the outage probability performance versus SNR for $K = 2$ and $M = 2, 4$. The target rate is fixed to $(\gamma_1, \gamma_2) = (3, 1)$ bit/channel use, and we let $\sigma^2_{i,k} = 1$ for all $i, k$. We compare the different power allocation strategies, algorithm A1 with perfect CSIT, algorithm A2 with statistical CSIT, and equal power allocation ($p_{i1} = p_{i2} = P/2$ for $i = 1, 2$). For the sake of comparison, we also consider the case without network MIMO (no message sharing) where each BS sends a message to its corresponding UT in a distributed fashion. In order to make the comparison fair in terms of complexity, we let each BS $i$ send the symbol $s_i$ by ZF beamforming, i.e. $x_i = \sqrt{P}g_is_i$ where $g_i$ is a unit-norm vector orthogonal to $h_{ik}$ for $k \neq i$. From Lemma 1, such a system offers a diversity order of $M - K + 1$ for each UT. As expected from Theorem 1, we observe that our BS cooperation schemes enables to achieve a diversity gain of $2(M - K + 1)$, i.e. 2, 6 with $M = 2, 4$, respectively. These gains are twice as large as the case without network MIMO. Moreover, the proposed algorithms provide a significant power gain compared to equal power allocation.

In Fig. 6, we plot the individual outage probability such that each UT $k$ cannot support its target rate $\gamma_k$ under the same setting as Fig. 5 for $M = 2$. With perfect CSIT, our proposed waterfilling allocation A1 guarantees the identical outage probability for both UTs by offering the strict fairness. This agrees
well with the second part of Theorem 2. Under statistical CSIT, algorithm A2 provides a better outage probability to UT 1 but keeps the gap between two UTs smaller than the equal power allocation.

In order to evaluate the impact of asymmetric path loss on the outage performance, we consider a simple 1-D configuration illustrated in Fig. 4 where UT 2 is located at \( x = 3 \) and UT 1 moves from \( x = 0 \) to \( x = 2 \). Assuming that BS 1, 2 is at \( x = 1, 3 \), respectively, we vary \( d_{11} = \sqrt{1 + (1-x)^2} \), \( d_{12} = \sqrt{1 + (3-x)^2} \) while we fix the position of UT2 by letting \( d_{12} = \sqrt{5}, d_{22} = 1 \). By taking into account the path loss \( \sigma_{ik}^2 = d_{ik}^{-3} \), we plot the outage probability as a function of the position \( x \) of UT 1 in Fig. 7.

We consider \( M = 4 \), SNR \( P = 10 \) dB and fix the target rate \( \gamma_1 = \gamma_2 = 1 \) bit/channel use. We observe that the proposed distributed ZF scheme provides a significant gain compared to the case without network MIMO especially as UT 1 gets closed to the cell boundary (\( x = 2 \)). This is because the performance without network MIMO only depends on \( d_{11} \) while the distributed ZF becomes more and more beneficial as \( d_{21} \) decreases.

Finally, Fig. 8 shows the outage probability versus SNR when we have more users than the number of served users, i.e. \( K \geq \tilde{K} = 2 \). Considering the same setting as Fig. 5 for \( M = 4 \), we apply distributed diversity scheme to select a set of two users among \( K = 2, 4, 6 \). Once the user selection is done, any power allocation studied in Section III can be applied. However, it is non-trivial (if not impossible) to characterize the statistics of the overall channel gains in the presence of any user scheduling. Hence, we illustrate here only the performance with equal power allocation under statistical CSIT. As a matter of fact, any smarter allocation shall perform between the waterfilling allocation and the equal power allocation. As expected from Theorem 4, the diversity gain increases significantly as the number \( K \) of users in the system gets large.
VI. CONCLUSIONS

We considered the multi-cell downlink system (network MIMO) where $B$ BSs, perfectly connected via the reliable backbone links to the CS, wish to communicate simultaneously with $K$ UTs. As one of the realistic limitations of network MIMO, we explicitly accounted for partial CSIT, i.e. local channel knowledge at each BS and statistical channel knowledge at the CS. Under this setting, we proposed an outage-efficient strategy which builds on distributed ZF beamforming to be performed at each BS and efficient power allocation algorithms at the CS. For the case of a small number of users $K \leq M$, the proposed scheme enables each UT to achieve a diversity gain of $B(M - K + 1)$. For the case of many users $K \geq M$, we proposed distributed diversity scheduling (DDS) which can be implemented in a distributed fashion at each BS and requires only limited amount of the backbone communications. We also proved that DDS can offer the diversity gain of $B\frac{K}{K}(M - \tilde{K} + 1)$ and this gain scales optimally with the number of cooperative BSs as well as the number of UTs. The main finding is that limited BS cooperation can still make network MIMO attractive in the sense that a well designed scheme can offer high data rates with sufficient reliability to individual UTs. The proposed scheme can be suitably applied to any other interference networks where the transmitters can perfectly share the messages to all UTs and a master transmitter can handle the resource allocation.

APPENDIX

A. Proof of Lemma 2

It is shown in [33] that the error probability of any coding scheme with rate $R$ bits per channel use is lower-bounded by the optimal outage probability of the channel with the same target rate in the high SNR regime. More specifically, let us set $\text{SNR} \triangleq \frac{E\{|x|^2\}}{\sigma^2}$ with $\sigma^2$ being the variance of the AWGN. Then, we have

$$\lim_{\text{SNR} \to \infty} \frac{\log (P_e(\text{SNR}))}{\log \text{SNR}} \geq \lim_{\text{SNR} \to \infty} \frac{\log (P_{\text{out}}(\text{SNR}))}{\log \text{SNR}}$$

for any coding scheme. Furthermore, the equality is attained with the Gaussian coding. Therefore, it is without loss of generality to use the outage probability to derive the maximum diversity. Following the footsteps in [33], the optimal outage probability has the same SNR exponent as the outage probability with isotropic Gaussian input giving

$$\Pr \left( \min\{n_t, n_r\} \sum_{i=1}^{\min\{n_t, n_r\}} \log \left( 1 + \frac{\text{SNR}}{n_t} \lambda_i \right) < R \right)$$

(43)
where \(\lambda_i\)'s are the non-zero eigenvalues of the matrix \(HH^H\). This probability is upper- and lower-bounded by
\[
\Pr \left( \log \left( 1 + \frac{\text{SNR}}{n_t} \lambda_{\text{max}} \right) < R \right) \quad \text{and} \quad \Pr \left( \log \left( 1 + \frac{\text{SNR}}{n_t} \lambda_{\text{max}} \right) < \frac{R}{\min \{n_t, n_r\}} \right).
\]
(44)

For a given \(R\) that does not scale with SNR, the above upper and lower bound coincides with the following probability in terms of SNR exponent in the high SNR regime
\[
\Pr \left( \lambda_{\text{max}} < \text{SNR}^{-1} \right)
\]
which is upper- and lower-bounded by
\[
\Pr \left( \sum_{i=1}^{\min \{n_t, n_r\}} \lambda_i < \min \{n_t, n_r\} \text{SNR}^{-1} \right) \quad \text{and} \quad \Pr \left( \sum_{i=1}^{\min \{n_t, n_r\}} \lambda_i < \text{SNR}^{-1} \right).
\]
(46)

Note that both bounds have the same SNR exponent at high SNR and that \(\sum \lambda_i = \|H\|_F^2\). Putting all the pieces together, we have
\[
d = \lim_{\text{SNR} \to \infty} \frac{\log(p_{\text{out}})}{\log(\text{SNR}^{-1})}
\]
\[
= \lim_{\text{SNR} \to \infty} \frac{\log(\Pr(\|H\|^2_F < \text{SNR}^{-1}))}{\log(\text{SNR}^{-1})}
\]
\[
= \lim_{\epsilon \to 0} \frac{\log(\Pr(\|H\|^2_F < \epsilon))}{\log(\epsilon)}
\]
(47)
(48)
(49)

B. Proof of Theorem 1

The channel (4) is a MISO channel defined by \(a_k\) and
\[
\Pr \{ \|a_k\|^2 < \epsilon \} \doteq \Pr \left\{ \max_i |a_{ik}|^2 < \epsilon \right\}
\]
\[
= \prod_i \Pr \{ |a_{ik}|^2 < \epsilon \}
\]
\[
= \left( \Pr \{ |a_{ik}|^2 < \epsilon \} \right)^B
\]
\[
\doteq \epsilon^{B(M-K+1)}
\]
(50)
(51)
(52)
(53)

From lemma 2, the maximum diversity is \(B(M-K+1)\).

C. Concavity of the objective function (27)

We prove the concavity of the objective function (27) by restricting ourselves to a feasible set \(\{p_{ik}\}\).

First we consider the second derivative with respect to \(\lambda_k\) given by
\[
\frac{\partial^2 f}{\partial \lambda_k^2} = \frac{\partial^2 p_k}{\partial \lambda_k^2} \prod_{j \neq k} (1 - p_j(\lambda_j, p_j))
\]
(54)
In order to prove the concavity of $f$ with respect to $\lambda_k$, it is sufficient to prove that $\frac{\partial^2 \rho_k}{\partial \lambda_k^2} < 0$ assuming that $\rho_j(\lambda_j, p_j) < 1$ for all $j \neq k$. We have the first and second derivatives of $\rho_k$ given by

$$\frac{\partial \rho_k}{\partial \lambda_k} = \frac{e^{\lambda_k c_k} \left[ c_k - (M - K + 1) \prod_p \frac{p_{ik} \sigma_{ik}^2}{1 + p_{ik} \sigma_{ik}^2 \lambda_k} \right]}{(M - K + 1) \prod_p \frac{p_{ik} \sigma_{ik}^2}{1 + p_{ik} \sigma_{ik}^2 \lambda_k}}$$

$$\frac{\partial^2 \rho_k}{\partial \lambda_k^2} = \frac{\partial \rho_k}{\partial \lambda_k} \left[ c_k - (M - K + 1) \prod_p \frac{p_{ik} \sigma_{ik}^2}{1 + p_{ik} \sigma_{ik}^2 \lambda_k} \right] + \rho_k (M - K + 1) \prod_p \frac{p_{ik} \sigma_{ik}^4}{(1 + p_{ik} \sigma_{ik}^2 \lambda_k)^2}$$

$$= \rho_k \left[ c_k - (M - K + 1) \prod_p \frac{p_{ik} \sigma_{ik}^2}{1 + p_{ik} \sigma_{ik}^2 \lambda_k} \right]^2 + \rho_k (M - K + 1) \prod_p \frac{p_{ik} \sigma_{ik}^4}{(1 + p_{ik} \sigma_{ik}^2 \lambda_k)^2} \geq 0$$  

(56)

where (56) follows by plugging (55). It readily follows that the two terms of (56) are non-negative for any non-negative set of $\{p_{ik}\}$, therefore the concavity of (27) with respect to $\lambda_k$ is verified.

Now, we check the concavity with respect to $p_{ik}$. It can be easily seen that the second derivative is non-positive. Namely we have the first and second derivatives given by

$$\frac{\partial f}{\partial p_{ik}} = (M - K + 1)e^{\lambda_k c_k} (\sigma_{ik}^2 \lambda_k (1 + \sigma_{ik}^2 \lambda_k p_{ik})^{-(M - K + 2)} \prod_{l \neq i} (1 + \sigma_{ik}^2 \lambda_k p_{ik})^{-(M - K + 1)} \prod_{j \neq k} (1 - \rho_j)$$

$$= -(M - K + 1)(M - K + 2)e^{\lambda_k c_k} (\sigma_{ik}^2 \lambda_k)^2$$

$$\times (1 + \sigma_{ik}^2 \lambda_k p_{ik})^{-(M - K + 3)} \prod_{l \neq i} (1 + \sigma_{ik}^2 \lambda_k p_{ik})^{-(M - K + 1)} \prod_{j \neq k} (1 - \rho_j) \leq 0.$$  

(58)

Hence, we have $\frac{\partial^2 f}{\partial p_{ik}^2} \leq 0$ for any $\lambda_k \geq 0$. This establishes the concavity of the objective function (27) for a feasible solution satisfying (29).

D. Convergence proof of Algorithm A2

The convergence proof follows in the same footsteps of the proof of convergence of [29, Theorem1, Theorem 2]. Let us first consider a special case with a single BS $B = 1$. Namely, consider the following problem:

$$\text{maximize} \quad f^{\text{single}}(\{\lambda_k\}, \{p_k\}) \triangleq \prod_{k=1}^K \left( 1 - \frac{e^{\lambda_k c_k}}{(1 + \lambda_k \sigma_{ik}^2 p_k)^{M - K + 1}} \right)$$

subject to \( \sum_{k=1}^K p_k \leq P, \)

\( \lambda_k \geq 0, \quad k = 1, \ldots, K \)

\( p_k \geq 0, \quad k = 1, \ldots, K \)

(59)
Since the problem is convex from Appendix C, we solve the KKT conditions, which are necessary and sufficient for the optimality. It is not difficult to show that the KKT conditions for a single BS system reduce to

\[
\lambda_k = \left[ \frac{M - K + 1}{c_k} - \frac{1}{\sigma^2_k p_k} \right]_+
\]

(61)

\[
\mu = \frac{\sigma^2_k \lambda_k}{(1 + \sigma^2_k \lambda_k p_k)} \left\{ e^{-\lambda_k c_k (1 + \sigma^2_k \lambda_k p_k)^{M - K + 1}} - 1 \right\}, \quad k = 1, \ldots, K
\]

(62)

where \(\mu \geq 0\) is a Lagrangian multiplier determined to satisfy the total power constraint. It follows that \(\lambda_k\) is given as a closed-form solution. Although \(p_k\) cannot be given in a closed-form solution, we can easily solve the polynomial function (62) simply by a linear search over \(\mu\) as discussed previously. Clearly, if a feasible solution exists such that \(\rho_{\text{single}}(p_k) = e^{\lambda_k c_k (1 + \sigma^2_k \lambda_k p_k)^{M - K + 1}} < 1\) is satisfied for all \(k\), the optimal solution is determined uniquely and the objective is maximized.

Now we consider the case of multiple BSs \(B > 1\). It is worth noticing that for each \(i = 1, \ldots, B\), the original objective function (27) coincides with \(f^{\text{single}}\) by considering

\[
\Theta_{ik} = \prod_{l \neq i} (1 + \sigma^2_{lk} \lambda_k p_{lk})^{-(M - K + 1)} \in [0, 1]
\]

constant. In other words, the KKT conditions of the multi-BS system (30) and (32) coincide with the KKT condition of the single-BS system when \(\Theta_{ik}\) are regarded constant for all \(i\) and \(k\). Following the same argument of [29, Theorem 1], it follows that \(p_i\) is an optimal solution to the original maximization problem if only if \(p_i\) is the solution to the single-BS problem (59) while treating \(\{p_l\}_{l \neq i}\) fixed. For each BS \(i\) at a given iteration \(n\), Algorithm A2 finds the power vector \(p_i\) by treating the powers of other BS \(\{p_l\}_{l \neq i}\), or equivalently \(\Theta_{i1}, \ldots, \Theta_{iK}\) fixed.

\[
\theta_{ik} = \prod_{l=1, l \neq i}^B (1 + \sigma^2_{lk} \lambda_k p_{lk})^{-(M - K + 1)} \in [0, 1]
\]

constant. In other words, the KKT conditions of the multi-BS system (30) and (32) coincide with the KKT condition of the single-BS system when \(\theta_{ik}\) are regarded constant for all \(i\) and \(k\). Following the same argument of [29, Theorem 1], it follows that \(p_i\) is an optimal solution to the original maximization problem if only if \(p_i\) is the solution to the single-BS problem (59) while treating \(\{p_l\}_{l \neq i}\) fixed. For each BS \(i\) at a given iteration \(n\), Algorithm A2 finds the power vector \(p_i\) by treating the powers of other BS \(\{p_l\}_{l \neq i}\), or equivalently \(\theta_{i1}, \ldots, \theta_{iK}\) fixed. Hence, the objective function is non-decreasing for each \(i\) and it converges to a limit (which is bounded). The convergence of \(\{P_{ik}\}\) also follows because each iteration of BS \(i\) finds the unique power vector \(p_i\) and thus strictly increases the objective or
keeps them unchanged. At the limit, all \( p_i \)s satisfy the KKT conditions (61) of the single-BS system simultaneously by incorporating the scaling factor \( \theta_{ik} \). Then, the set of \( p_1, \ldots, p_B \) must maximize the objective function (27).

E. Proof of Theorem 4

In order to examine the diversity gain of the proposed scheduling scheme, we first remark

\[
\min_{k \in \mathcal{U}_d} \|a_k\|^2 \geq \min_{k \in \mathcal{U}_d} \max_i |a_{ik}|^2 \geq \max_i \min_{k \in \mathcal{U}_d} |a_{ik}|^2 = \max \max_{i} \min_{k \in \mathcal{U}_d} |a_{ik}|^2
\]

where the (64) follows from the max-min inequality and the last equality holds since we can rewrite (36) as

\[
\mathcal{U}_d = \arg \max_{U} \max_{i} \min_{k \in \mathcal{U}} |a_{ik}|^2.
\]

by swapping the maximization over \( \mathcal{U} \) and that over \( i \). To find the diversity order of the scheme, we need to look at the following near-zero behavior of the channel coefficients

\[
\Pr \left( \min_{k \in \mathcal{U}_d} \|a_k\|^2 < \epsilon \right) \leq \Pr \left( \max_{i} \min_{k \in \mathcal{U}} |a_{ik}|^2 < \epsilon \right) \leq \left( \Pr \left( \min_{k \in \mathcal{U}} |a_{ik}|^2 < \epsilon \right) \right)^{B|\mathcal{P}_s|} \leq \left( \sum_{k \in \mathcal{U}} \Pr \left( |a_{ik}|^2 < \epsilon \right) \right)^{B|\mathcal{P}_s|} = \left( |\mathcal{U}| \Pr \left( |a_{ik}|^2 < \epsilon \right) \right)^{B|\mathcal{P}_s|} \leq \left( |\mathcal{U}| \epsilon^M + 1 - \tilde{K} \right)^{B|\mathcal{P}_s|} \leq \epsilon^{B \frac{K}{K}(M+1-\tilde{K})},
\]

where (68) follows from the fact that \( \mathcal{U} \)'s are disjoint in \( \mathcal{P}_s \) and that \( \min_{k \in \mathcal{U}} |a_{ik}|^2 \) are independent for different \( \mathcal{U} \) and \( i \); (69) is from the union bound. From lemma 2, (37) is straightforward. For the upper bound of the diversity gain of the scheduling scheme (34), let us first write

\[
\max \min_{\mathcal{U} \in \mathcal{Q}} \min_{k \in \mathcal{U}} \|a_k\|^2 \leq \max_{\mathcal{U} \in \mathcal{Q}} \|a_1\|^2 \leq B \max_{i} \max_{\mathcal{U} \in \mathcal{Q}} |a_{i1}|^2
\]

March 1, 2011
where the first inequality is from the fact that the worst user cannot be better than the first user; the second inequality is from \( \|a_1\|_2 \leq B \max_i |a_{i1}|^2 \). From [34, Theorem 1], we know that the diversity gain of \( |a_{i1}|^2 \) is \( (K - \bar{K} + 1)(M - \bar{K} + 1) \). Therefore, it readily follows that the diversity gain of \( \max_{U \in Q} \min_{k \in U} \|a_k\|^2 \) is upper-bounded by \( B(K - \bar{K} + 1)(M - \bar{K} + 1) \).

ACKNOWLEDGMENT

This work is partially supported by the European Commission through the FP7 project WiMAGIC (www.wimagic.eu) and by the French cluster System@tic through the project POSEIDON. The authors would like to thank Jakob Hoydis for his help in generating some numerical examples.

REFERENCES


Fig. 5. Outage probability vs. SNR with $B = K = 2$ and $M = 2, 4$. 
Fig. 6. Individual outage probabilities vs. SNR with $B = K = 2$ and $M = 2$. 
Fig. 7. Outage probability vs. location of UT1 with $B = K = 2$ and $M = 2, 4$. 

**Note:** The figure shows the outage probability for different channel conditions: 
- Equal power
- Statistical CSIT
- Perfect CSIT
- Without network MIMO

The outage probability is calculated with $\gamma = 1 \text{bit/ch.use}$ and $M = 4, 10 \text{dB}$. The x-axis represents the position $x$ of UT1.
Fig. 8. Outage probability vs. SNR for many users with $B = \tilde{K} = 2$ and $M = 4$. 