

Multi-cell MIMO Downlink with Fairness Criteria: the Large System Limit

Hoon Huh*, Giuseppe Caire*, Sung-Hyun Moon†, and Inkyu Lee‡

*University of Southern California, Los Angeles, CA, USA, Email: hhuh, caire@usc.edu

†Korea University, Seoul, Korea, Email: shmoon, inkyu@korea.ac.kr

Abstract—We consider the downlink of a cellular network with multiple cells and multi-antenna base stations under arbitrary inter-cell cooperation, realistic distance-dependent pathloss, and general “fairness” requirements. Beyond Monte Carlo simulation, no efficient computation method to evaluate the ergodic throughput of such systems has been provided so far. We propose an analytic method based on the combination of the large random matrix theory with Lagrangian optimization. The proposed method is computationally much more efficient than Monte Carlo simulation and provides a very accurate approximation (almost indistinguishable) for the actual finite-dimensional case, even for a small number of users and base station antennas. Numerical examples include linear 2-cell and planar three-sectored 7-cell layouts, with no inter-cell cooperation, sector cooperation, and full cooperation.

I. INTRODUCTION

Multiuser MIMO (MU-MIMO) technology is expected to play a key role in future wireless cellular networks. The MIMO Gaussian broadcast channel serves as an information theoretic model for the MU-MIMO downlink and its capacity was studied and characterized in [1]–[3]. In a multi-cell scenario, depending on the level of base station (BS) cooperation, we are in the presence of a MIMO broadcast and interference channel problem, which is not yet fully understood in the information theoretic sense. In this work, we restrict our attention to the case where clusters of cooperating BSs act as a single distributed MIMO transmitter and where interference from other clusters of BSs is treated as noise. Our framework encompasses arbitrary cooperation clusters.

Multi-cell systems under full and partial cooperation have been widely studied under the so-called “Wyner model” [4]–[7] which assumes a very simplified and rather unrealistic model in regard to the pathloss and makes the system essentially identical with respect to any given user. As a matter of fact, users in different locations under the cellular coverage are subject to distance-dependent pathloss that may have more than 30 dB dynamic range [8]. It follows that in these realistic propagation conditions the channels from a BS to users are not identically distributed at all. Furthermore, studying the downlink sum-capacity (or achievable sum-throughput, under some suboptimal scheme) is not a very meaningful approach from a system performance viewpoint: rate and power optimization would lead to the solution of letting a BS transmit only to users very close to the BS, while leaving users at the cell edges to starve. In order to avoid this problem, *downlink scheduling* has been widely studied, which achieves a good

balance between network efficiency and “fairness” (see for example [9]–[11] and refs. therein). The goal of fairness scheduling is to make the system operate at a rate point in its ergodic achievable rate region such that a suitable *network utility function* is maximized [12]. Taking into account realistic pathloss models, multi-cell systems with arbitrary inter-cell cooperation, and fairness requirements makes system analysis extremely complicated. In fact, so far, the system performance of such systems has been evaluated only through extensive and computationally very intensive Monte Carlo simulation.

In this work, we consider a particular “large system limit” that is studied with known results from the large random matrix theory [13]–[15] and combine it with Lagrangian optimization in order to obtain a numerical “almost” closed-form analysis tool that incorporates all the above system aspects. The proposed method is much more efficient than Monte Carlo simulation and, somehow surprisingly, provides results that match very closely the performance of finite-dimensional systems, even for very small dimension.

II. SYSTEM SETUP

Consider M BSs with γN antennas each, and KN single-antenna user terminals distributed in the cellular coverage region. Users are divided into K co-located “user groups” of equal size N . Users in the same group are statistically equivalent: they experience the same pathloss from all BSs and their small-scale fading channel coefficients are iid.¹ The received signal vector $\mathbf{y}_k = [y_{k,1} \cdots y_{k,N}]^T \in \mathbb{C}^N$ for user group k is given by

$$\mathbf{y}_k = \sum_{m=1}^M \alpha_{m,k} \mathbf{H}_{m,k}^H \mathbf{x}_m + \mathbf{n}_k \quad (1)$$

where $\alpha_{m,k}$ and $\mathbf{H}_{m,k}$ denote the the distance dependent pathloss and $\gamma N \times N$ small-scale channel fading matrix from BS m to user group k , respectively, $\mathbf{x}_m = [x_{m,1} \cdots x_{m,\gamma N}]^T \in \mathbb{C}^{\gamma N}$ is the transmitted signal vector of BS m subject to the power constraint $\text{tr}(\text{Cov}(\mathbf{x}_m)) \leq P$, and $\mathbf{n}_k = [n_{k,1} \cdots n_{k,N}]^T \in \mathbb{C}^N$ denotes the AWGN at the user terminals. The elements of \mathbf{n}_k and of $\mathbf{H}_{m,k}$ are iid $\sim \mathcal{CN}(0, 1)$.

¹In practice, co-located users are separated by a sufficient number of wavelength such that they undergo iid small-scale fading. However, the wavelength is sufficiently small so that they all have essentially the same distance-dependent pathloss.

We define the BS partition $\{\mathcal{M}_1, \dots, \mathcal{M}_L\}$ of set $\{1, \dots, M\}$ and the corresponding user group partition $\{\mathcal{K}_1, \dots, \mathcal{K}_L\}$ of set $\{1, \dots, K\}$, where L is the number of cooperation clusters. Cooperation cluster ℓ is formed by the BSs in \mathcal{M}_ℓ , acting effectively as a distributed network MIMO transmitter, and the user groups in \mathcal{K}_ℓ . We assume that the BSs in each cluster have perfect channel state information (CSI) for all users in the same cluster and *statistical information* (i.e. known distributions but not the instantaneous values) relative to channels from BSs in other clusters. The impact of non-perfect CSI at the transmitter and the required overhead for CSI estimation for inter-cell cooperation is addressed in [16]. We restrict to the symmetric cellular structure for clusters such that each cluster has the same number of BSs and the same distribution of user groups with respect to all the BSs in the cluster. Specific examples of such a structure will be given in Section V. Under these assumption, owing to the symmetric nature of clusters, the sum transmit power of each BS is equal to P when we consider the cluster-wise sum-power constraint $|\mathcal{M}_\ell|P$ and it is satisfied with an equality. Then, the interference plus noise variance at any user terminal in group $k \in \mathcal{K}_\ell$ is obtained as

$$\sigma_k^2 = 1 + \sum_{m \notin \mathcal{M}_\ell} \alpha_{m,k}^2 P \quad (2)$$

and, from the viewpoint of cluster ℓ , the system is equivalent to a single-cell MU-MIMO downlink given by

$$\mathbf{y}_\ell = \mathbf{H}_\ell^H \mathbf{x}_\ell + \mathbf{z}_\ell \quad (3)$$

where $\mathbf{y}_\ell \in \mathbb{C}^{|\mathcal{K}_\ell|N}$ is the received signal vector formed by $\{\mathbf{y}_k : k \in \mathcal{K}_\ell\}$, $\mathbf{z}_\ell \in \mathbb{C}^{|\mathcal{K}_\ell|N}$ is a Gaussian interference plus noise vector with independent elements and block-diagonal covariance matrix $\mathbb{E}[\mathbf{z}_\ell \mathbf{z}_\ell^H] = \text{diag}(\sigma_k^2 \mathbf{I}_N : k \in \mathcal{K}_\ell)$, and $\mathbf{H}_\ell \in \mathbb{C}^{\gamma|\mathcal{M}_\ell|N \times |\mathcal{K}_\ell|N}$ is formed as

$$\mathbf{H}_\ell = \begin{bmatrix} \alpha_{m_1,k_1} \mathbf{H}_{m_1,k_1} & \cdots & \alpha_{m_1,k_{|\mathcal{K}_\ell|}} \mathbf{H}_{m_1,k_{|\mathcal{K}_\ell|}} \\ \vdots & & \vdots \\ \alpha_{m_{|\mathcal{M}_\ell|},k_1} \mathbf{H}_{m_{|\mathcal{M}_\ell|},k_1} & \cdots & \alpha_{m_{|\mathcal{M}_\ell|},k_{|\mathcal{K}_\ell|}} \mathbf{H}_{m_{|\mathcal{M}_\ell|},k_{|\mathcal{K}_\ell|}} \end{bmatrix} \quad (4)$$

For the sake of system optimization, we introduce the “dual uplink” channel model [2], [17], [18] corresponding to (3) given by

$$\mathbf{r} = \tilde{\mathbf{H}} \mathbf{s} + \mathbf{w} \quad (5)$$

where we define the transmit signal $\mathbf{s} \in \mathbb{C}^{AN}$ with diagonal covariance matrix \mathbf{Q} subject to $\text{tr}(\mathbf{Q}) \leq Q$, the AWGN vector $\mathbf{w} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{BN})$, and the channel matrix $\tilde{\mathbf{H}} \in \mathbb{C}^{BN \times AN}$ formed by blocks $\beta_{m,k} \mathbf{H}_{m,k}$ in the same way as in (4), and where, for notation simplicity, we let $A = |\mathcal{K}_\ell|$, $B = \gamma|\mathcal{M}_\ell|$, $Q = |\mathcal{M}_\ell|P$, and $\beta_{m,k} = \frac{\alpha_{m,k}}{\sigma_k}$.

We assume that the pathloss coefficients are fixed while the small-scale fading coefficients change in time according to some ergodic process with the assigned first-order iid $\mathcal{CN}(0, 1)$ distribution. This is representative of a typical situation where the distance between BSs and users changes

significantly over a time-scale of the order of tens of seconds, whereas the small-scale fading decorrelates completely within a few milliseconds [19]. In this regime, the goal of downlink scheduling is to maximize a suitable strictly increasing and concave network utility function $g(\cdot)$ of ergodic user rates [12]. By symmetry, all users in the same group should achieve the same ergodic rate. However, users in different groups may operate at different rates, depending on the network utility function $g(\cdot)$. For $\mathbf{R} = [R_1, R_2, \dots, R_A]$ where R_k is the ergodic group rate given by the mean user rate of user group k , the fairness scheduling problem is formulated as

$$\begin{aligned} & \text{maximize} && g(\mathbf{R}) \\ & \text{subject to} && \mathbf{R} \in \mathcal{R}_{\text{erg}}(Q) \end{aligned} \quad (6)$$

where $\mathcal{R}_{\text{erg}}(Q)$ is the region of ergodic achievable group rates for system (5) (equivalently, for cluster ℓ with channel model defined in (3)) for given sum-power constraint Q .

III. WEIGHTED ERGODIC SUM-RATE MAXIMIZATION

In this section, we recall the computation algorithm of [15] to solve, in the limit for $N \rightarrow \infty$, the weighted ergodic sum-rate maximization problem:

$$\begin{aligned} & \text{maximize} && \sum_{k=1}^A W_k R_k \\ & \text{subject to} && \mathbf{R} \in \mathcal{R}_{\text{erg}}(Q). \end{aligned} \quad (7)$$

This is a fundamental building block for the computation of (6). The exposition in this section is necessarily brief and more details can be found in [15] and references therein.

Let π denote the permutation such that $W_{\pi_1} \leq \dots \leq W_{\pi_A}$, and define $\Theta_{k:A} = \tilde{\mathbf{H}}_{k:A} \mathbf{Q}_{k:A} \tilde{\mathbf{H}}_{k:A}^H$, where $\mathbf{H}_{k:A}$ and $\mathbf{Q}_{k:A}$ denote the channel and the input covariance sub-matrices restricted to user groups π_k, \dots, π_A . By symmetry, we have that all users in the same group k are allocated the same (uplink) power Q_k . After dividing the channel coefficients by \sqrt{N} , the power constraint for any N is given by $\sum_{k=1}^A Q_k \leq Q$. For the dual uplink channel (5), the maximum weighted sum-rate is achieved by a block successive interference cancellation decoding strategy that jointly decodes users from group π_1 to π_A , and, after decoding each group, subtracts it from the received signal. It follows that problem (7) reduces to the maximization with respect to \mathbf{Q} of the objective function

$$F_{\mathbf{W}}(\mathbf{Q}) = \sum_{k=1}^A \Delta_k \frac{1}{N} \mathbb{E} [\log |\mathbf{I} + \Theta_{k:A}|] \quad (8)$$

where $\Delta_k = W_{\pi_k} - W_{\pi_{k-1}}$ with $W_{\pi_0} \triangleq 0$. From the *Karush-Kuhn-Tucker* (KKT) conditions applied to the Lagrangian function $\mathcal{L}(\mathbf{Q}, \lambda) = F_{\mathbf{W}}(\mathbf{Q}) - \lambda(\text{tr}(\mathbf{Q}) - Q)$, we can eliminate the Lagrange multiplier λ by imposing that the power constraint holds with equality and then we obtain

$$Q_{\pi_k} = Q \frac{\sum_{i=1}^k \Delta_i \left(1 - \mathbb{E} [\text{mmse}_{i:A}^{(k)}]\right)}{\sum_{j=1}^A \sum_{i=1}^j \Delta_i \left(1 - \mathbb{E} [\text{mmse}_{i:A}^{(j)}]\right)} \quad (9)$$

where $\text{mmse}_{i:A}^{(k)} = \frac{1}{N} \text{tr}(\mathbb{E}[\mathbf{I} - Q_{\pi_k} \tilde{\mathbf{H}}_{\pi_k}^H (\mathbf{I} + \Theta_{j:A})^{-1} \tilde{\mathbf{H}}_{\pi_k}])$ for $k \geq i$ is the average *Minimum Mean Square Error* (MMSE) for the input symbols of group π_k when the symbols of groups π_1, \dots, π_{i-1} are cancelled from the observation \mathbf{r} in (5), and $\tilde{\mathbf{H}}_{\pi_k}$ denotes the $BN \times N$ submatrix of $\tilde{\mathbf{H}}$ corresponding to group π_k .

For finite N , the amount of calculation in order to evaluate the solution of (9) is tremendous because $\mathbb{E}[\text{mmse}_{i:A}^{(k)}]$ must be computed by Monte Carlo simulation. In the following, we consider the large system regime where we let $N \rightarrow \infty$ and, by using the asymptotic random matrix theory of [13], we arrive at a computationally efficient algorithm. Let $\Upsilon_{i:A}^{(k)} = \lim_{N \rightarrow \infty} \text{mmse}_{i:A}^{(k)}$ and define $\Gamma_{i:A}^{(k)} = 1/\Upsilon_{i:A}^{(k)} - 1$ and then a direct application of [13, Lemma 1] yields $\Gamma_{i:A}^{(k)}$ as the solution of the fixed-point equation

$$\Gamma_{i:A}^{(k)} = \gamma Q_{\pi_k} \sum_{m=1}^{B/\gamma} \frac{\beta_{m,\pi_k}^2}{1 + \sum_{j=i}^A \frac{\beta_{m,\pi_j}^2 Q_{\pi_j}}{1 + \Gamma_{i:A}^{(k)}}}. \quad (10)$$

Combining (10) with the iterative algorithm [20, Algorithm 1] that converges to the solution of (9), we obtain Algorithm 1 below. (for notation simplicity, $\pi_k = k$ is assumed for all $k = 1, \dots, A$).

Algorithm 1 Algorithm for weighted sum-rate maximization

- 1) Initialize $\mathbf{Q}(0) = \frac{Q}{A} \mathbf{I}_A$.
- 2) For $\ell = 0, 1, 2, \dots$, iterate until the following solution settles:

$$Q_k(\ell + 1) = Q \frac{\sum_{i=1}^k \Delta_i (1 - \Upsilon_{i:A}^{(k)}(\ell))}{\sum_{j=1}^A \sum_{i=1}^j \Delta_i (1 - \Upsilon_{i:A}^{(j)}(\ell))} \quad (11)$$

for $k = 1, \dots, A$, where $\Upsilon_{i:A}^{(k)}(\ell) = 1/(1 + \Gamma_{i:A}^{(k)}(\ell))$ and $\Gamma_{i:A}^{(k)}(\ell)$ is obtained as the solution (also obtained by iteration) of (10) for powers $Q_k = Q_k(\ell)$.

- 3) Denote by $\Gamma_{i:A}^{(k)}(\infty)$, $\Upsilon_{i:A}^{(k)}(\infty)$, and $Q_k(\infty)$ the fixed points reached by the iterations in step 2). If the condition

$$Q \sum_{i=1}^k \Delta_i \Gamma_{i:A}^{(k)}(\infty) \leq \sum_{j=1}^A \sum_{i=1}^j \Delta_i (1 - \Upsilon_{i:A}^{(j)}(\infty))$$

is satisfied for all k such that $Q_k(\infty) = 0$, then stop. Otherwise, set $Q_k = 0$ for k corresponding to the lowest value of $\sum_{i=1}^k \Delta_i \Gamma_{i:A}^{(k)}(\infty)$ and repeat steps 2) and 3).

After group powers $Q_k^* = Q_k(\infty)$ have been obtained from Algorithm 1, it remains to compute the ergodic group rates. We have

$$R_{\pi_k} = \frac{1}{N} \mathbb{E}[\log |\mathbf{I} + \Theta_{k:A}|] - \frac{1}{N} \mathbb{E}[\log |\mathbf{I} + \Theta_{k+1:A}|]. \quad (12)$$

In the limit for $N \rightarrow \infty$, we can use the asymptotic analytical expression for the mutual information given in [14]. Adapting [14, Result 1] to our case, we obtain $\lim_{N \rightarrow \infty} \frac{1}{N} \mathbb{E}[\log |\mathbf{I} + \Theta_{k:A}|]$ from expression (21) in [15].

IV. INTRODUCING FAIRNESS

In a finite dimensional system, a dynamic scheduling policy allocates powers and rates in order to obtain the system ergodic rate point (i.e. the time-average user rates) as close to the solution of (6) as possible. This can be systematically obtained by the stochastic optimization approach of [10], [11], based on the idea of “virtual queues”. For a deterministic network, we can exploit Lagrangian duality with outer subgradient iteration, where the Lagrange dual variables play the roles of the virtual queue backlogs of the dynamic scheduling. In the large system limit, the channel uncertainty disappears and the multi-cell MU-MIMO system becomes indeed deterministic. Hence, problem (6) for large N can be addressed directly, using Lagrangian duality.

We rewrite (6) using auxiliary variables $\mathbf{r} = [r_1, \dots, r_A]$ as follows:

$$\begin{aligned} \max_{\mathbf{r}, \mathbf{Q}, \pi} \quad & g(\mathbf{r}) \\ \text{s.t.} \quad & r_{\pi_k} \leq \frac{1}{N} \mathbb{E} \left[\log \frac{|\mathbf{I} + \Theta_{k:A}|}{|\mathbf{I} + \Theta_{k+1:A}|} \right], \quad \forall k, \\ & \text{tr}(\mathbf{Q}) \leq Q. \end{aligned} \quad (13)$$

The Lagrangian function of problem (13) is given by

$$\begin{aligned} \mathcal{L}(\mathbf{r}, \mathbf{Q}, \pi, \boldsymbol{\mu}) \\ = g(\mathbf{r}) - \underbrace{\sum_{k=1}^A r_{\pi_k} \mu_{\pi_k}}_{f_{\boldsymbol{\mu}}(\mathbf{r})} + \underbrace{\sum_{k=1}^A \mu_{\pi_k} \frac{1}{N} \mathbb{E} \left[\log \frac{|\mathbf{I} + \Theta_{k:A}|}{|\mathbf{I} + \Theta_{k+1:A}|} \right]}_{h_{\boldsymbol{\mu}}^N(\mathbf{Q}, \pi)} \end{aligned} \quad (14)$$

where $\boldsymbol{\mu}$ denotes the dual variables for the rate constraints. The Lagrange dual function for (14) is given by

$$\begin{aligned} \mathcal{G}(\boldsymbol{\mu}) &= \max_{\mathbf{r}, \mathbf{Q}, \pi} \mathcal{L}(\mathbf{r}, \mathbf{Q}, \pi, \boldsymbol{\mu}) \\ &= \underbrace{\max_{\mathbf{r}} f_{\boldsymbol{\mu}}(\mathbf{r})}_{(a)} + \underbrace{\max_{\mathbf{Q}, \pi} h_{\boldsymbol{\mu}}^N(\mathbf{Q}, \pi)}_{(b)} \end{aligned} \quad (15)$$

and it is obtained by the decoupled maximization in (a) (with respect to \mathbf{r}) and in (b) (with respect to \mathbf{Q}, π) in (15). Finally, we solve the dual problem defined as

$$\min_{\boldsymbol{\mu} \geq 0} \mathcal{G}(\boldsymbol{\mu}). \quad (16)$$

Since the maximization of (b) is a weighted ergodic sum-rate maximization for weights $\mathbf{W} = \boldsymbol{\mu}$, it follows that the optimal π^* is the permutation that sorts $\boldsymbol{\mu}$ in non-decreasing order (see Section III). For $\pi = \pi^*$, the Lagrangian function is concave in \mathbf{r} and \mathbf{Q} and convex in $\boldsymbol{\mu}$ and the solution (saddle point of the min-max problem) can be found via inner-outer iterations.

Inner Problem: For given $\boldsymbol{\mu}$, we solve the maximization problem in (15) with respect to \mathbf{r}, \mathbf{Q} and π .

• Subproblem (a): Since $f_{\boldsymbol{\mu}}(\mathbf{r})$ is concave in $\mathbf{r} \geq 0$, the maximum of $f_{\boldsymbol{\mu}}(\mathbf{r})$ is achieved by imposing the KKT conditions:

$$\frac{\partial g(\mathbf{r})}{\partial r_k} - \mu_k \leq 0, \quad \forall k \quad (17)$$

where the equality must hold for all k such that the solution is positive, i.e., $r_k^* > 0$.

• Subproblem (b): As said, this problem is a weighed ergodic sum-rate maximization and can be solved for $N \rightarrow \infty$ by Algorithm 1.

Outer Problem: Once the optimal \mathbf{r}^* , \mathbf{Q}^* and π^* are obtained for given μ , the minimization of $\mathcal{G}(\mu)$ with respect to $\mu \geq \mathbf{0}$ can be performed by a subgradient-based method. For any fixed μ' and μ , we have

$$\begin{aligned} \mathcal{G}(\mu') &= \max_{\mathbf{r}} f_{\mu'}(\mathbf{r}) + \max_{\mathbf{Q}} h_{\mu'}^{\infty}(\mathbf{Q}, \pi^*) \\ &\geq f_{\mu'}(\mathbf{r}^*) + h_{\mu'}^N(\mathbf{Q}^*, \pi^*) \\ &= \mathcal{G}(\mu) + \sum_{k=1}^A (\mu'_{\pi_k^*} - \mu_{\pi_k^*}) (R_{\pi_k^*} - r_{\pi_k^*}) \end{aligned} \quad (18)$$

where $R_{\pi_k^*}$ is the rate of user group π_k^* obtained from Algorithm 1 with weights μ . Then, the vector with components $R_{\pi_k^*} - r_{\pi_k^*}$ is the subgradient for $\mathcal{G}(\mu)$ and dual variable μ_{π_k} is updated at each outer iteration n as

$$\mu_{\pi_k^*}(n+1) = \mu_{\pi_k^*}(n) - s(n) (R_{\pi_k^*}(n) - r_{\pi_k^*}(n)), \quad \forall k \quad (19)$$

for some step size $s(n) > 0$ which can be determined efficiently by a back-tracking line search method [21]. It should be noticed that by setting $s(n) = 1$ this subgradient update plays the role of the virtual queue update in the dynamic scheduling policy of [10], [11].

As an application example of the above general optimization, we focus on the two special cases, *proportional fairness scheduling* (PFS) and *hard fairness scheduling* (HFS), also known as max-min fairness scheduling. PFS corresponds to the network utility function $g(\mathbf{r}) = \sum_{k=1}^K C \log(r_k)$ for some constant $C > 0$ and the optimality condition (17) yields $r_k(n) = C/\mu_k(n)$ at outer iteration n .

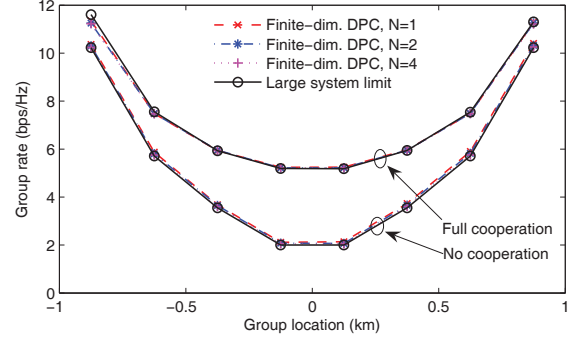
In the case of HFS, the network utility function is given as $g(\mathbf{r}) = \min_{k=1, \dots, A} C r_k$ and the corresponding optimization problem can be rewritten as

$$\begin{aligned} \max_{\mathbf{r}, \mathbf{Q}, \pi} \quad & Cr \\ \text{s.t.} \quad & r \leq \frac{1}{N} \mathbb{E} \left[\log \frac{|\mathbf{I} + \Theta_{k:A}|}{|\mathbf{I} + \Theta_{k+1:A}|} \right], \quad \forall k \\ & \text{tr}(\mathbf{Q}) \leq Q. \end{aligned} \quad (20)$$

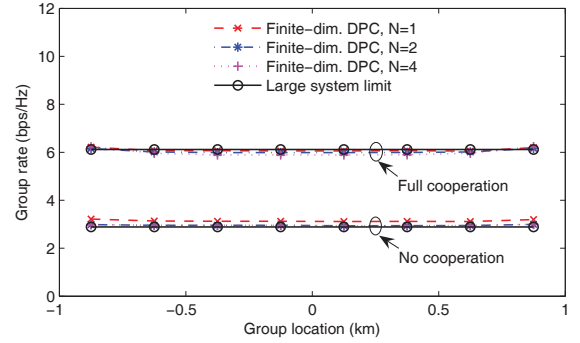
Replicating the former approach for this problem, with the corresponding changes, we find $r(n) = \frac{1}{A} \sum_{k=1}^A R_{\pi_k}(n)$ at outer iteration n .

V. NUMERICAL RESULTS AND DISCUSSION

We present numerical examples for a one-dimensional 2-cell model ($M = 2$) and two-dimensional three-sectored 7-cell model ($M = 21$). In both models, the system parameters and pathloss model are based on the mobile WiMAX system evaluation specification [8], except for cell radius 1.0 km and no shadow fading assumption. In the 2-cell model, we assume two one-sided cells with BSs located at -1 km and 1 km with $\gamma = 4$, and $K = 8$ user groups equally spaced between the



(a) Proportional fairness scheduling



(b) Hard fairness scheduling

Fig. 1. Individual group rates under (a) PFS and (b) HFS for $\gamma = 4$ and $K = 8$ in the 2-cell model

two BSs. We consider the case of full BS cooperation and no cooperation with a symmetric partition of user groups, namely: $L = 2$, $\mathcal{K}_1 = \{1, 2, 3, 4\}$ and $\mathcal{K}_2 = \{5, 6, 7, 8\}$. Fig. 1 illustrates the individual group rates (i.e. mean user rates of each group) in the large system limit as a function of group locations and compares them with the achievable rates obtained by Monte Carlo simulation in finite dimensions with $N = 1, 2$, or 4 . In the simulation with finite N , the BSs are equipped with γN antennas and N users are located at each of K locations. Channel vectors are randomly generated and the dynamic scheduling [11] and DPC precoding with water-filling algorithm [22] are applied to each realization of channel vectors. In plots (a) and (b), the PFS and HFS are considered, respectively. Remarkably, the rates obtained by the large system analysis almost overlap with those obtained by the finite-dimensional simulations, even for $N = 1$. Notice that the dynamic scheduling policy should provide multiuser diversity gain and achieve in general higher rates than the large system limit which is not able to exploit the dynamic fluctuations of the small-scale fading due to “channel hardening” effect. However, it appears that in the regime where the pathloss is dominant over the randomness of multi-antenna channels and the number of users is not much larger than the number of BS antennas, the multiuser diversity gain is negligible and the asymptotic analysis produces rate points very close to the outputs of simulations with dynamic scheduling. Notice that in

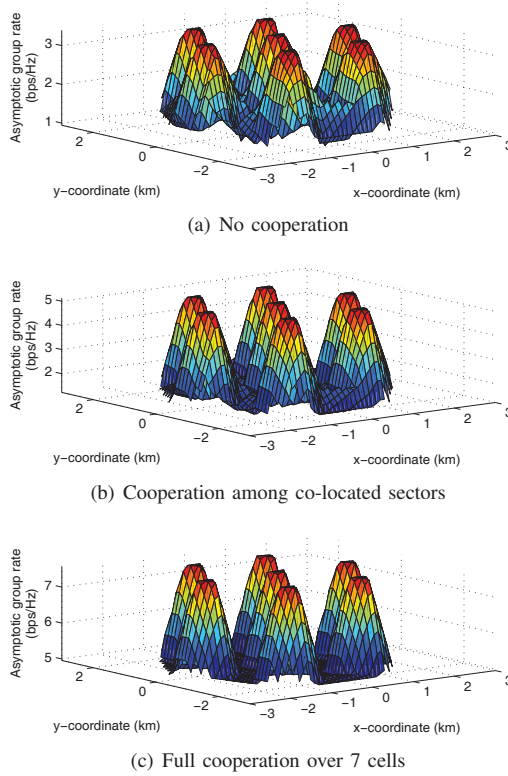


Fig. 2. Ergodic group rate distribution under PFS in the 7-cell model

HFS case all the users achieve the same individual rate which is slightly higher than the smallest rate under the PFS.

Using the proposed asymptotic analysis, validated with the simple 2-cell model, we can obtain ergodic rate distributions for much larger systems, for which a full-scale simulation would be very demanding. We consider a two-dimensional cell layout where 7 hexagonal cells form a network and each cell consists of three 120-degree sectors. Three BSs are co-located at the center of each cell such that each BS handles one sector in no cooperation case. Each sector is split into 4 diamond-shaped equal-area grids and one user group is placed at the center of each grid. Therefore there are total $M = 21$ BSs and $K = 84$ user groups in the network. In addition, we assume the 7 cells are wrap-around in a torus topology, such that each cell is virtually surrounded by the other 6 cells and all the cells have the symmetric inter-cell interference distribution. The antenna orientation and pattern follows [23] and the non-ideal antenna pattern generates inter-cell interference even between co-located sectors with no cooperation. This model indeed conjectures very large scale of cellular networks in that the dominant inter-cell interference and effective inter-cell cooperation gain comes from directly neighboring cells. Fig. 2 shows the user rate distribution under three levels of cooperation, (a) no cooperation ($L = 21$), (b) cooperation among co-located 3 sectors ($L = 7$), and (c) full cooperation over 7-cell network ($L = 1$). The simulation is very intricate especially in case (c). From the asymptotic rate results, it is

shown that in case (b), the cooperation gain over case (a) is primarily obtained for users around cell centers, whereas the gain is achieved over all the locations in case (c).

REFERENCES

- [1] G. Caire and S. Shamai (Shitz), "On the achievable throughput of a multiantenna Gaussian broadcast channel," *IEEE Trans. on Inform. Theory*, vol. 49, pp. 1691–1706, July 2003.
- [2] A. J. Goldsmith, S. A. Jafar, N. J. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE J. Select. Areas Commun.*, vol. 21, pp. 684–702, June 2003.
- [3] H. Weingarten, Y. Steinberg, and S. Shamai (Shitz), "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Trans. on Inform. Theory*, vol. 52, pp. 3936–3964, Sept. 2006.
- [4] A. D. Wyner, "Shannon-theoretic approach to a Gaussian cellular multiple access channel," *IEEE Trans. on Inform. Theory*, vol. 40, pp. 1713–1727, Nov. 1994.
- [5] S. Shamai (Shitz) and A. D. Wyner, "Information-theoretic considerations for symmetric, cellular, multiple-access fading channels – Part I & II," *IEEE Trans. on Inform. Theory*, vol. 43, pp. 1877–1894, Nov. 1997.
- [6] O. Somekh and S. Shamai (Shitz), "Shannon-theoretic approach to a Gaussian cellular multiple-access channel with fading," *IEEE Trans. on Inform. Theory*, vol. 46, pp. 1401–1425, July 2000.
- [7] A. Sanderovich, O. Somekh, H. Poor, and S. Shamai (Shitz), "Uplink macro diversity of limited backhaul cellular network," *IEEE Trans. on Inform. Theory*, vol. 55, pp. 3457–3478, Aug. 2009.
- [8] WiMAX Forum, "Mobile WiMAX – Part I: A technical overview and performance evaluation," Tech. Rep., Aug. 2006.
- [9] P. Viswanath, D. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Trans. on Inform. Theory*, vol. 48, pp. 1277–1294, June 2002.
- [10] L. Georgiadis, M. Neely, and L. Tassiulas, *Resource Allocation and Cross-Layer Control in Wireless Networks*. Foundations and Trends in Networking, 2006, vol. 1, no. 1.
- [11] H. Shirani-Mehr, G. Caire, and M. J. Neely, "MIMO downlink scheduling with non-perfect channel state knowledge," *submitted to IEEE Trans. on Commun.* (posted on arXiv:0904.1409 [cs.IT]), 2009.
- [12] J. Mo and J. Walrand, "Fair end-to-end window-based congestion control," *IEEE/ACM Trans. on Networking*, vol. 8, pp. 556–567, Oct. 2000.
- [13] A. M. Tulino, A. Lozano, and S. Verdu, "Impact of antenna correlation on the capacity of multiantenna channels," *IEEE Trans. on Inform. Theory*, vol. 7, pp. 2491–2509, July 2005.
- [14] D. Aktas, M. N. Bacha, J. S. Evans, and S. V. Hanly, "Scaling results on the sum capacity of cellular networks with MIMO links," *IEEE Trans. on Inform. Theory*, vol. 52, pp. 3264–3274, July 2006.
- [15] S.-H. Moon, H. Huh, Y.-T. Kim, I. Lee, and G. Caire, "Weighted sum-rate of multi-cell MIMO downlink channels in the large system limit," in *Proc. IEEE Int. Conf. on Commun. (ICC)*, Cape Town, South Africa, May 2010.
- [16] H. Huh, A. M. Tulino, and G. Caire, "Network MIMO large-system analysis and the impact of CSIT estimation," in *Proc. Conf. on Inform. Sciences and Systems (CISS)*, Princeton, NJ, March 2010.
- [17] N. Jindal and A. J. Goldsmith, "Capacity and optimal power allocation for fading broadcast channels with minimum rates," *IEEE Trans. on Inform. Theory*, vol. 49, pp. 2895–2909, November 2003.
- [18] N. Jindal, W. Rhee, S. Vishwanath, S. A. Jafar, and A. J. Goldsmith, "Sum power iterative water-filling for multi-antenna Gaussian broadcast channels," *IEEE Trans. on Inform. Theory*, vol. 51, pp. 1570–1580, April 2005.
- [19] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.
- [20] A. M. Tulino, A. Lozano, and S. Verdu, "Capacity-achieving input covariance for single-user multi-antenna channels," *IEEE Trans. on Wireless Commun.*, vol. 5, pp. 662–671, March 2006.
- [21] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
- [22] W. Yu, "Sum-capacity computation for the Gaussian vector broadcast channel via dual decomposition," *IEEE Trans. on Inform. Theory*, vol. 52, pp. 754–759, Feb. 2006.
- [23] IEEE 802.16 broadband wireless access working group, "IEEE 802.16m evaluation methodology document (EMD)," Tech. Rep., Jan. 2009.