# Network MIMO with Linear Zero-Forcing Beamforming: Large System Analysis, Impact of Channel Estimation and Reduced-Complexity Scheduling

Hoon Huh, *Student Member, IEEE*, Antonia M. Tulino, *Senior Member, IEEE*, and Giuseppe Caire, *Fellow, IEEE*

**Index Terms**

Large random matrices, linear zero-forcing beamforming, network MIMO, channel estimation, downlink scheduling

H. Huh and G. Caire are with the Department of Electrical Engineering, University of Southern California, Los Angeles, CA 90089 USA (e-mail: hhuh, caire@usc.edu).

A. M. Tulino is with the Department of Wireless Communications, Bell Laboratories, Alcatel-Lucent, Holmdel, NJ 07733 USA (e-mail: a.tulino@alcatel-lucent.com).

The material in this paper was presented in part at the 44th Annual Conference on Information Sciences and Systems (CISS), Princeton, NJ, March 2010.

**Abstract**

We consider the downlink of a multi-cell system with multi-antenna base stations and single-antenna user terminals, arbitrary base station cooperation clusters, distance-dependent propagation pathloss, and general "fairness" requirements. We focus on the *joint transmission* from the base stations in a cooperation cluster based on linear zero-forcing beamforming, subject to sum or per-base station power constraints. Analytic expressions for the system spectral efficiency are found in the large-system limit where both the numbers of users and antennas per base station tend to infinity with a given ratio. In particular, for the per-base station constraint, we find new results in random matrix theory, yielding the squared Frobenius norm of submatrices of the Moore-Penrose pseudo-inverse for the structured non-i.i.d. channel matrix resulting from the cooperation cluster, user distribution, and path-loss coefficients. The analysis is extended to the case of non-ideal channel state information obtained through explicit downlink channel training, and it is instrumental in providing insight in the joint operations of downlink multiuser beamforming, inter-cell cooperation, and opportunistic "fair" scheduling. Specifically, our results illuminate the trade-off between the benefit of a larger number of cooperating antennas and the cost of estimating higher-dimensional channel vectors. Furthermore, our analysis lead to a new simplified downlink scheduling scheme that pre-selects the users according to probabilities obtained from the large-system results. The proposed scheme performs close to the optimal (finite-dimensional) opportunistic user selection while requiring significantly less channel state feedback, since only a small fraction of pre-selected users must feed back their channel state information.

# I. Introduction

The next generation of wireless communication systems (e.g., 802.16m [1], LTE-Advanced [2]) consider multiuser MIMO (MU-MIMO) as one of the core technologies. A considerable research effort has been dedicated to the performance evaluation of MU-MIMO systems under the realistic cellular environments [3]–[5]. The downlink capacity in a single cell setting under perfect Channel State Information at the Transmitter (CSIT) was thoroughly characterized with the full information theoretic understanding of the underlying MIMO Gaussian broadcast channel [6]–[10]. However, in a multi-cell scenario, we are in the presence of a MIMO broadcast and interference channel which is not yet fully understood in an information theoretic sense.

A simple and practical approach consists of treating Inter-Cell Interference (ICI) as noise and, in this case, the system capacity may be significantly limited by the ICI. A variety of inter-cell cooperation schemes have been proposed to mitigate ICI, ranging from a fully cooperative network MIMO [11]–[14] to a partially coordinated beamforming [15]–[17]. In this work, we focus on the network MIMO approach, where clusters of cooperating base stations (BSs) act as a single distributed MIMO transmitter and interference from other clusters of BSs is treated as noise.

In a cellular environment, the received signal power is a polynomially decreasing function of the distance between the BS and the user terminals, with a dynamic range typically larger than 30 dB [18]. Thus, users close to the cell (or cluster of cells) boundary experience strong inter-cell interference, whereas the signal received from the intended BS is relatively weak. These "edge" users cannot be just *ignored* by the system. For example, maximizing the system achievable sum-throughput leads in general to a very unfair solution, where the system resources are concentrated on users close to a BS. In contrast, *fairness scheduling* has been proposed and widely studied in order to achieve a good balance between throughput efficiency and fairness (see for example [19]–[21] and references therein). Fairness scheduling can be systematically implemented under the framework of *stochastic network optimization* [20], which enforces that the system operates at a rate point of its *ergodic achievable rate region* that maximizes some suitable concave and increasing *network utility function* [22]. For a given network MIMO scheme (defined by the BS clusters and the joint MU-MIMO transmission scheme), the optimal operating point can be provably approached as close as desired by applying a dynamic scheduling policy at each time slot. While the analytical characterization of the optimal operating point is hopelessly complicated in any realistic scenario, the system performance has been evaluated so far through computationally very intensive Monte Carlo simulation [3]–[5], [13], [14], [23]–[26], where the actual scheduling algorithm

evolves in time and the ergodic rates are computed by time-averaging.

The capacity of a multi-cell network MIMO system under fairness criteria was evaluated in the large-system limit in [27], assuming ideal channel state knowledge and the Gaussian *Dirty Paper Coding* (DPC) transmission strategy. In [27], the asymptotic analysis method based on large random matrix results demonstrated its effectiveness by the comparison with finite-dimensional Monte Carlo simulation. In this work, we apply a similar large-system analysis to *Linear Zero-forcing Beamforming* (LZFB). It turns out that the analysis in this case is significantly more complicated, in particular, in order to take into account per-BS power constraint. In addition, we extend our analysis to the case where the channel state information is obtained through explicit downlink training and MMSE estimation. In these conditions, we obtain a lower bound on the achievable ergodic rates (referred to as "throughput" in the following), that takes into account the overhead due to training-based channel estimation. Several novel and important aspects are illuminated in this paper. Specifically: 1) As in [27], our analysis allows precise performance evaluation of systems for which brute-force Monte Carlo simulation would be very demanding. 2) By including the effect of downlink training, we can investigate the tradeoff between ICI reduction owing to BS cooperation and the cost of estimating higher and higher dimensional channels. Unlike previous results which assumed ideal CSIT at no cost [24], [25], [27], we observe that there exist an optimal cooperation cluster size that depends on the channel coherence time and bandwidth, beyond which cooperative network MIMO is not convenient, consistently with the finite-dimensional simulation findings of [26], [28]. 3) We provide novel results in random matrix theory, in particular, related to the evaluation of the coefficients appearing in the per-BS power constraint. 4) We make use of our asymptotic analysis in order to design a *random scheduling* algorithm that pre-selects the users with assigned probabilities obtained from the large-system results, and therefore requires much less CSIT feedback than the standard opportunistic scheduling scheme.

This last point deserves some remarks, since for the first time (to the authors' knowledge) asymptotic results are used not only for performance *analysis* but also for *system design* in network MIMO. [1] The standard approach to scheduling for downlink beamforming consists of having a large number of users feeding back their CSIT and selecting a subset of users with cardinality not larger than the number of jointly coordinated transmit antennas, such that the channel vectors of the selected users have both large

[1]In the unrelated context of multiuser detection, asymptotic large-system results were used to design low-complexity linear multiuser detectors based on the polynomial approximations, where the polynomial coefficients were obtained from large random matrix theory [29], [30].

norm and are mutually approximately orthogonal [31], [32]. This *multiuser diversity* selection, combined with LZFB precoding, is shown to attain the same performance as Gaussian DPC in the limit of a large number of users and fixed number of transmit antennas. However, in this limit, the throughput per user vanishes as $O(\frac{\log\log n}{n})$ where $n$ is the number of users. Therefore, a more meaningful regime is one in which the number of users is proportional to the number of antennas, yielding constant throughput per user. This is in fact the large-system regime investigated in this paper. Comparing the results of our asymptotic analysis with the Monte Carlo simulation of finite dimensional systems, including user selection as said before, we notice that multiuser diversity yields larger throughput per user for low-dimensional systems, but this gain reduces as the system dimension grows. This is a manifestation of the "channel hardening effect" noticed in [33], and agrees with the theoretical findings in [34] that show that the probability of finding a subset of approximately orthogonal users vanishes as the system dimension increases. It follows that for large systems there is diminishing return in selecting users from a large set. In contrast, the cost of CSIT feedback grows at least linearly with the number of users feeding back their estimated CSIT. Therefore, we advocate a probabilistic scheduling algorithm for which users are pre-selected at random using the probabilities derived from our large-system analysis, and only the selected users are required to feed back their CSIT. The performance of this scheme are shown to be close to the much more costly full user selection scheme, and become closer and closer as the system dimension increases (again, by the large-system limit and channel hardening effect).

In comparison with concurrent existing literature, we notice that the LZFB MU-MIMO performance analysis with non-ideal CSIT was extensively studied in the finite-dimensional case (see for example [35]–[37]) and in the large-system limit (see for example [38]–[40]). Unlike concurrent works, our paper focuses explicitly on the system optimization under the fairness criteria in the multi-cell downlink with inter-cell cooperation. This particular angle, allows us to illuminate aspects that are not present in other works, such as the distribution of the per-user throughput under fairness and, as a consequence, the design of the random scheduling scheme said before.

The remainder of this paper is organized as follows. In Section II, we describe the general finite-dimensional system model including the arbitrary clustering of cooperative BSs, formulate the system optimization problem, and provide its numerical solutions for a given channel realization. In Section III, we take the large system limit and present the large-system regime of the LZFB precoder and the optimization algorithm for user selection and power allocation. The opportunistic fairness scheduling scheme is also described in this section. The impact of non-perfect CSIT and training overhead is analyzed in Section IV. Numerical results and the low-complexity randomized scheduling scheme are presented

in Section V and some concluding remarks are given in Section VI. The most lengthy and technical derivations are relegated into the appendices.

## II. Finite Dimensional System

### A. System setup

Consider a cellular system formed by $M$ BSs, with $\gamma N$ antennas each, and $KN$ single-antenna user terminals spatially distributed in the coverage area. We assume that the users are divided into $K$ co-located "user groups" with $N$ users each. Users in the same group are statistically equivalent: they see the same pathloss coefficients from all BSs and their small-scale fading channel coefficients are i.i.d.. The received signal vector $\mathbf{y}_k = [y_{k,1} \cdots y_{k,N}]^{\mathsf{T}} \in \mathbb{C}^N$ for the $k$-th user group is given by

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

where $\alpha_{m,k}$ and $\mathbf{H}_{m,k}$ denote the distance dependent pathloss coefficient and $\gamma N \times N$ small-scale channel fading matrix from the $m$-th BS to the $k$-th user group, respectively, $\mathbf{x}_m = [x_{m,1} \cdots x_{m,\gamma N}]^{\mathsf{T}} \in \mathbb{C}^{\gamma N}$ is the transmitted signal vector of the $m$-th BS, subject to the power constraint $\mathrm{tr}\left(\mathrm{Cov}(\mathbf{x}_m)\right) \leq P_m$, and $\mathbf{n}_k = [n_{k,1} \cdots n_{k,N}]^{\mathsf{T}} \in \mathbb{C}^N$ denotes the additive white Gaussian noise (AWGN) at the user receivers. The elements of $\mathbf{n}_k$ and of $\mathbf{H}_{m,k}$ are i.i.d. $\mathcal{CN}(0,1)$.

A cooperative cell arrangement with $L$ cooperation clusters is defined by the BS partition $\{\mathcal{M}_1, \cdots, \mathcal{M}_L\}$ of the BS set $\{1, \cdots, M\}$ and the corresponding user group partition $\{\mathcal{K}_1, \cdots, \mathcal{K}_L\}$ of the user group set $\{1, \cdots, K\}$. We assume that the BSs in each cluster $\mathcal{M}_\ell$ act as a single distributed multi-antenna transmitter with $\gamma|\mathcal{M}_\ell|N$ antennas, perfectly coordinated by a central cluster controller, and serve users in groups $k \in \mathcal{K}_\ell$. The clusters do not cooperate and treat ICI from other clusters as noise. Assuming that each BS operates at its maximum individual transmit power, the ICI plus noise power at any user terminal in group $k \in \mathcal{K}_\ell$ is given by

$$\sigma_k^2 = 1 + \sum_{m \notin \mathcal{M}_\ell} \alpha_{m,k}^2 P_m. \tag{2}$$

Each cluster seeks to maximize its own objective function defined by the fairness scheduling. It is easy to show that, under the above system assumptions, the selfish optimal strategy that operates at maximum per-BS power is a Nash equilibrium of the system. At this Nash equilibrium, the clusters are effectively decoupled since the effect that other clusters have on each cluster $\ell$ is captured by the ICI terms in (2) that do not depend on the actual BS transmit covariances $\mathrm{Cov}(\mathbf{x}_m)$.

From the viewpoint of cluster $\ell$, the system is equivalent to a single-cell MIMO downlink channel with a modified channel matrix and noise levels and a per-BS power constraint. Therefore, from now on we focus on a given reference cluster $\ell = 1$ and, without loss of generality, we indicate the user groups in the reference cluster as $k = 1, \ldots, A$, with $A = |\mathcal{K}_1|$, and the BSs in $\mathcal{M}_1$ as $m = 1, \ldots, B$ with $B = |\mathcal{M}_1|$. After a convenient re-normalization of the channel coefficients, we arrive at the equivalent channel model for the reference cluster given by

$$\mathbf{y} = \mathbf{H}^{\mathsf{H}}\mathbf{x} + \mathbf{z} \tag{3}$$

with $\mathbf{y} \in \mathbb{C}^{AN}$, $\mathbf{x} \in \mathbb{C}^{\gamma BN}$, $\mathbf{z} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{AN})$ and the channel matrix $\mathbf{H} \in \mathbb{C}^{\gamma BN \times AN}$ is given by

$$\mathbf{H} = \begin{bmatrix} \beta_{1,1}\mathbf{H}_{1,1} & \cdots & \beta_{1,A}\mathbf{H}_{1,A} \\ \vdots & & \vdots \\ \beta_{B,1}\mathbf{H}_{B,1} & \cdots & \beta_{B,A}\mathbf{H}_{B,A} \end{bmatrix}, \tag{4}$$

where we define $\beta_{m,k} = \alpha_{m,k}/\sigma_k$. The pathloss coefficients are fixed constant, that depend only on the geometry of the system, and the small-scale fading coefficients are assumed to change independently from time slot to time slot according to a classical block-fading model. This is representative of a typical situation where the distance between BSs and users changes significantly over a time-scale of the order of the tens of seconds, while the small-scale fading decorrelates completely within a few milliseconds [41]. Here, a "time slot" indicates the number of channel uses over which the small-scale coefficients can be considered constant. This is approximately equal to the product of the channel coherence time and the channel coherence bandwidth [41].

### *B. Downlink scheduling optimization problem*

Each cluster controller operates according to a *downlink scheduling* scheme that allocates instantaneously the transmission resource (signal dimensions and transmit power) to the users. Following [22], the scheduling problem is formulated as the maximization of a suitable strictly increasing and concave *network utility function* $g(\cdot)$ over the region of achievable ergodic rates (throughput region). For users in group $k$, we denote the normalized sum of individual user throughputs by $\overline{R}_k = \frac{1}{N}\sum_{i=1}^{N}\overline{R}_k^{(i)}$. By symmetry, the users in the same group achieve the same throughput, therefore, the throughput vector with $\overline{R}_k^{(i)} = \overline{R}_k$ for all users $i$ in group $k$ is achievable. We assume that $g(\cdot)$ is symmetric in the throughput of users belonging to the same group. Therefore, letting $\overline{\mathbf{R}} = (\overline{R}_1, \cdots, \overline{R}_A)$, the fairness scheduling

problem is formulated as

$$\begin{aligned} &\text{maximize} \quad g(\overline{\mathbf{R}}) \\ &\text{subject to} \quad \overline{\mathbf{R}} \in \overline{\mathcal{R}} \end{aligned} \tag{5}$$

In this work, we consider LZFB downlink precoding, such that $\overline{\mathcal{R}}$ indicates the throughput region achievable by LZFB for the channel model (3), under the assumption of operating at the Nash equilibrium said above. Notice that $\overline{\mathcal{R}}$ includes all rates obtained by long-term (ergodic) average over the time-varying fading channels and by time-sharing of all possible LZFB policies (including user selection and power allocation). A scheduling policy achieving the optimum throughput point $\overline{\mathbf{R}}^{\star}$ solution of (5) consists of a rule that, at each scheduling time slot, maps the available channel information $\mathbf{H}$ into a set of users and transmit powers, such that the resulting rates, averaged over time, converge to $\overline{\mathbf{R}}^{\star}$.

As a first step towards the solution of (5), we focus on the weighted *instantaneous* sum-rate maximization problem:

$$\begin{aligned} &\text{maximize} \quad \sum_{k=1}^{A} \sum_{i=1}^{N} W_k^{(i)} R_k^{(i)} \\ &\text{subject to} \quad \mathbf{R} \in \mathcal{R}_{\text{lzfb}}(\mathbf{H}) \end{aligned} \tag{6}$$

where $W_k^{(i)}$ denotes the rate weight for user $i$ in group $k$, and $\mathcal{R}_{\text{lzfb}}(\mathbf{H})$ is the achievable *instantaneous* rate region of LZFB for given channel matrix $\mathbf{H}$. By "instantaneous", we mean that this rate region depends on the given channel realization $\mathbf{H}$, in contrast with the throughput region $\overline{\mathcal{R}}$, that depends on the statistics of $\mathbf{H}$. Realistically, we assume that $A \geq \gamma B$ (i.e., the number of users in the cluster is larger than or equal to the total number of BS antennas in the cluster) and that all coefficients $\beta_{m,k}$ are strictly positive. Therefore, $\text{rank}(\mathbf{H}) = \gamma BN$ is almost surely satisfied. In this case, LZFB cannot serve simultaneously all users in the cluster, and the scheduler must *select* a subset of users not larger than $\gamma BN$, to be served at each time slot.

The solution of (6) is generally difficult, since it requires a search over all user subsets of cardinality less or equal to $\gamma BN$. Well-known approaches (see [31], [32]) consider the selection of a user subset in some greedy fashion, by adding users to the active user set one by one, till the objective function in (6) cannot be improved further. Finally, even for a fixed set of active users, the problem of optimal LZFB precoding subject to a per-BS power constraint is non-trivial and has been recently addressed in [42]–[44] through fairly involved numerical algorithms. Because of these difficulties, problem (6) has so far escaped a clean analytical solution and most studies resort to extensive and costly Monte Carlo simulation.

In order to overcome the above difficulties, we make the following simplifying assumptions: 1) The scheduler picks a fraction $\mu_k$ of users in group $k$ by random selection inside the group. User selection is random and independent on each scheduling time slot; 2) The LZFB precoder is obtained by normalizing the columns of the Moore-Penrose pseudo-inverse of the channel matrix, although this choice is not necessarily optimal under the per-BS power constraint [42].

Under these assumptions, let $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_A)$ denote the fractions of active users in groups $1, \ldots, A$, respectively. For given $\boldsymbol{\mu}$, the corresponding effective channel matrix is given by

$$\mathbf{H}_{\boldsymbol{\mu}} = \begin{bmatrix} \beta_{1,1}\mathbf{H}_{1,1}(\mu_1) & \cdots & \beta_{1,A}\mathbf{H}_{1,A}(\mu_A) \\ \vdots & & \vdots \\ \beta_{B,1}\mathbf{H}_{B,1}(\mu_1) & \cdots & \beta_{B,A}\mathbf{H}_{B,A}(\mu_A) \end{bmatrix}, \tag{7}$$

where the blocks $\mathbf{H}_{m,k}(\mu_k)$ is a $\gamma N \times \mu_k N$ dimensional submatrix of $\mathbf{H}_{m,k}$. The user fractions must satisfy $\mu_k \in [0, 1]$ for each $k = 1, \ldots, A$ and $\mu \stackrel{\Delta}{=} \mu_{1:A} \leq \gamma B$ where we introduce the notation

$$\mu_{1:k} = \sum_{j=1}^{k} \mu_j. \tag{8}$$

Hence, $\mathrm{rank}(\mathbf{H}_{\boldsymbol{\mu}}) = \mu N$ is almost surely satisfied.

The LZFB precoding scheme yields the transmitted signal for active users, $\mathbf{x}_{\boldsymbol{\mu}}$ in the form

$$\mathbf{x}_{\boldsymbol{\mu}} = \mathbf{V}_{\boldsymbol{\mu}}\mathbf{Q}^{1/2}\mathbf{u} \tag{9}$$

where $\mathbf{u}$ is the independently coded unit-power user symbol vector of length $\mu N$, $\mathbf{V}_{\boldsymbol{\mu}}$ is the precoding matrix with unit-norm columns and $\mathbf{Q}$ is the diagonal matrix which contains user powers on the diagonal. In particular, here we assume that $\mathbf{V}_{\boldsymbol{\mu}}$ is obtained from the Moore-Penrose pseudo-inverse as follows: define the pseudo-inverse of $\mathbf{H}_{\boldsymbol{\mu}}^{\mathsf{H}}$ as

$$\mathbf{H}_{\boldsymbol{\mu}}^{+} = \mathbf{H}_{\boldsymbol{\mu}}(\mathbf{H}_{\boldsymbol{\mu}}^{\mathsf{H}}\mathbf{H}_{\boldsymbol{\mu}})^{-1}, \tag{10}$$

and then we let $\mathbf{V}_{\boldsymbol{\mu}} = \mathbf{H}_{\boldsymbol{\mu}}^{+}\boldsymbol{\Lambda}_{\boldsymbol{\mu}}^{1/2}$ where the column-normalizing diagonal matrix $\boldsymbol{\Lambda}_{\boldsymbol{\mu}}$ contains the reciprocal of the squared norm of columns of $\mathbf{H}_{\boldsymbol{\mu}}^{+}$ on the diagonal. Letting $\Lambda_k^{(i)}(\boldsymbol{\mu})$ denote the diagonal element of $\boldsymbol{\Lambda}_{\boldsymbol{\mu}}$ in position $\mu_{1:k-1}N + i$, for $i = 1, \ldots, \mu_k N$, we have

$$\Lambda_k^{(i)}(\boldsymbol{\mu}) = \frac{1}{\left[\left(\mathbf{H}_{\boldsymbol{\mu}}^{\mathsf{H}}\mathbf{H}_{\boldsymbol{\mu}}\right)^{-1}\right]_k^{(i)}} \tag{11}$$

where $\left[\left(\mathbf{H}_{\boldsymbol{\mu}}^{\mathsf{H}}\mathbf{H}_{\boldsymbol{\mu}}\right)^{-1}\right]_k^{(i)}$ denotes the element in the corresponding position $\mu_{1:k-1}N + i$ of the main diagonal of the matrix $\left(\mathbf{H}_{\boldsymbol{\mu}}^{\mathsf{H}}\mathbf{H}_{\boldsymbol{\mu}}\right)^{-1}$. Rewriting (3) with (7) and (9) and noticing that $\mathbf{H}_{\boldsymbol{\mu}}^{\mathsf{H}}\mathbf{V}_{\boldsymbol{\mu}} = \boldsymbol{\Lambda}_{\boldsymbol{\mu}}^{1/2}$,

we arrive at the "parallel" channel model for all users selected simultaneously by the LZFB precoder in the form

$$\mathbf{y}_{\boldsymbol{\mu}} = \boldsymbol{\Lambda}_{\boldsymbol{\mu}}^{1/2}\mathbf{Q}^{1/2}\mathbf{u} + \mathbf{z}_{\boldsymbol{\mu}}. \tag{12}$$

The optimization of (6) for the channel model (12) is still involved, since the channel coefficients $\Lambda_k^{(i)}(\boldsymbol{\mu})$ depend on the active user fractions $\boldsymbol{\mu}$ in a complicated and non-convex way. Nevertheless, as an intermediate step, we consider here the solution of (6) for fixed user fractions $\boldsymbol{\mu}$.

### *C. Power allocation under sum-power or per-BS power constraints*

We divide all channel matrix coefficients by $\sqrt{N}$ and multiply the BS input power constraints $P_m$ by $N$, thus obtaining an equivalent system where the channel coefficients have variance that scales as $1/N$. This is useful when we consider the large-system limit for $N \to \infty$ in Section III.

Let $q_k^{(i)}$ denote the diagonal element in position $\mu_{1:k-1}N+i$ of $\mathbf{Q}$, corresponding to the power allocated to the $i$-th user of group $k$. The sum-power constraint is given by

$$\frac{1}{N}\text{tr}(\mathbf{Q}) = \frac{1}{N}\sum_{k=1}^{A}\sum_{i=1}^{\mu_k N} q_k^{(i)} \leq P_{\text{sum}} \tag{13}$$

where $P_{\text{sum}} = \sum_{m=1}^{B} P_m$. In order to express the per-BS power constraint, let $\boldsymbol{\Phi}_m$ denote a diagonal matrix with all zeros, but for $\gamma N$ consecutive ones, corresponding to positions from $(m-1)\gamma N + 1$ to $m\gamma N$ on the main diagonal. Then, the per-BS power constraint is expressed in terms of the partial trace of the transmitted signal covariance matrix as

$$\frac{1}{N}\text{tr}\left(\boldsymbol{\Phi}_m \mathbf{V}_{\boldsymbol{\mu}}\mathbf{Q}\mathbf{V}_{\boldsymbol{\mu}}^{\mathsf{H}}\right) \leq P_m, \quad m = 1, \ldots, B \tag{14}$$

More explicitly, (14) can be written in terms of the powers $q_k^{(i)}$ as

$$\sum_{k=1}^{A}\sum_{i=1}^{\mu_k N} q_k^{(i)}\theta_{m,k}^{(i)} \;\leq\; P_m, \quad m = 1, \ldots, B \tag{15}$$

where we define the coefficients

$$\theta_{m,k}^{(i)}(\boldsymbol{\mu}) = \frac{1}{N}\sum_{\ell=(m-1)\gamma N+1}^{m\gamma N} \left|\left[\mathbf{V}_{\boldsymbol{\mu}}\right]_{\ell,k}^{(i)}\right|^2 \tag{16}$$

and where $\left[\mathbf{V}_{\boldsymbol{\mu}}\right]_{\ell,k}^{(i)}$ denotes the element of $\mathbf{V}_{\boldsymbol{\mu}}$ corresponding to the $\ell$-th row and the $(\mu_{1:k-1}N + i)$-th column. Since $\mathbf{V}_{\boldsymbol{\mu}}$ has unit-norm columns, then $\sum_{m=1}^{B}\theta_{m,k}^{(i)} = 1/N$ for all $k, i$.

The power optimization problem that solves (6) for fixed user fractions is given by

$$\text{maximize} \quad \sum_{k=1}^{A}\sum_{i=1}^{\mu_k N} W_k^{(i)}\log(1 + \Lambda_k^{(i)}(\boldsymbol{\mu})q_k^{(i)}) \tag{17}$$

subject to (13) in the case of sum-power constraint, or to (15) for the case of per-BS power constraint.

The solution of (17) subject to the sum-power constraint is immediately given by the water-filling formula

$$q_k^{(i)} = \left[ \frac{W_k^{(i)}}{\lambda} - \frac{1}{\Lambda_k^{(i)}(\boldsymbol{\mu})} \right]_+ \tag{18}$$

where $\lambda \geq 0$ is the Lagrange multiplier corresponding to the sum-power constraint.

In the case of per-BS power constraint, we can use Lagrange duality and sub-gradient iteration method as given in the following. The Lagrangian for (17) is given by (dependency on $\boldsymbol{\mu}$ is dropped for notation simplicity)

$$\mathcal{L}(\mathbf{q}, \boldsymbol{\lambda}) = \sum_{k=1}^{A} \sum_{i=1}^{\mu_k N} W_k^{(i)} \log(1 + \Lambda_k^{(i)} q_k^{(i)}) - \boldsymbol{\lambda}^{\mathsf{T}} \left[ \boldsymbol{\Theta}\mathbf{q} - \mathbf{P} \right] \tag{19}$$

where $\boldsymbol{\lambda} \geq 0$ is a vector of dual variables corresponding to the $B$ BS power constraints, $\boldsymbol{\Theta}$ is the $B \times \mu N$ matrix containing the coefficients $\theta_{m,k}^{(i)}$ and $\mathbf{P} = (P_1, \ldots, P_B)^{\mathsf{T}}$. The KKT conditions are given by

$$\frac{\partial \mathcal{L}}{\partial q_k^{(i)}} = W_k^{(i)} \frac{\Lambda_k^{(i)}}{1 + \Lambda_k^{(i)} q_k^{(i)}} - \boldsymbol{\lambda}^{\mathsf{T}} \boldsymbol{\theta}_k^{(i)} \leq 0 \tag{20}$$

where $\boldsymbol{\theta}_k^{(i)}$ is the column of $\boldsymbol{\Theta}$ containing the coefficients $\theta_{m,k}^{(i)}$ for $m = 1, \ldots, B$. Solving for $q_k^{(i)}$, we find

$$q_k^{(i)}(\boldsymbol{\lambda}) = \left[ \frac{W_k^{(i)}}{\boldsymbol{\lambda}^{\mathsf{T}} \boldsymbol{\theta}_k^{(i)}} - \frac{1}{\Lambda_k^{(i)}} \right]_+ \tag{21}$$

Then, replacing this solution into $\mathcal{L}(\mathbf{q}, \boldsymbol{\lambda})$, we solve the dual problem by minimizing $\mathcal{L}(\mathbf{q}(\boldsymbol{\lambda}), \boldsymbol{\lambda})$ with respect to $\boldsymbol{\lambda} \geq 0$. It is immediate to check that for any $\boldsymbol{\lambda}' \geq 0$,

$$\begin{aligned} \mathcal{L}(\mathbf{q}(\boldsymbol{\lambda}'), \boldsymbol{\lambda}') &\geq \mathcal{L}(\mathbf{q}(\boldsymbol{\lambda}), \boldsymbol{\lambda}') \\ &= (\boldsymbol{\lambda}' - \boldsymbol{\lambda})^{\mathsf{T}} (\mathbf{P} - \boldsymbol{\Theta}\mathbf{q}(\boldsymbol{\lambda})) + \mathcal{L}(\mathbf{q}(\boldsymbol{\lambda}), \boldsymbol{\lambda}) \end{aligned} \tag{22}$$

Therefore, $(\mathbf{P} - \boldsymbol{\Theta}\mathbf{q}(\boldsymbol{\lambda}))$ is a subgradient for $\mathcal{L}(\mathbf{q}(\boldsymbol{\lambda}), \boldsymbol{\nu})$. It follows that the dual problem can be solved by a simple $B$-dimensional subgradient iteration over the vector of dual variables $\boldsymbol{\lambda}$.

## III. Large System Limit

In this section, we consider the limit of the above instantaneous rate maximization problems in the limit $N \to \infty$, when $\gamma, A, B$, and $\boldsymbol{\mu}$ are fixed. While the coefficients of the limiting optimization problem are obtained in general through the solution of fixed-point equations, under certain system symmetries, a closed-form explicit expression can be found. Then, we consider the simultaneous optimization of the user fractions and powers in the large system limit. Finally, we consider the solution of the downlink scheduling optimization problem (5) in the large system regime.

*A. Asymptotic analysis*

We start by finding the large system limit expression for the coefficients $\Lambda_k^{(i)}(\boldsymbol{\mu})$. This is provided by:

*Theorem 1:* For all $i = 1, \ldots, \mu_k N$, the following limit holds almost surely:

$$\lim_{N\to\infty} \Lambda_k^{(i)}(\boldsymbol{\mu}) = \Lambda_k(\boldsymbol{\mu}) = \gamma \sum_{m=1}^{B} \beta_{m,k}^2 \eta_m(\boldsymbol{\mu}) \tag{23}$$

where $(\eta_1(\boldsymbol{\mu}), \ldots, \eta_B(\boldsymbol{\mu}))$ is the unique solution in $[0,1]^B$ of the fixed point equations

$$\eta_m = 1 - \sum_{q=1}^{A} \mu_q \frac{\eta_m \beta_{m,q}^2}{\gamma \sum_{\ell=1}^{B} \eta_\ell \beta_{\ell,q}^2}, \quad m = 1, \ldots, B \tag{24}$$

with respect to the variables $\boldsymbol{\eta} = \{\eta_m\}$.

*Proof:* See Appendix A. ■

Notice that the limit (23) depends only on $k$ (user group index) and not on $i$ (user index within the group), consistently with the fact that, in our model, users in the same co-located group are statistically equivalent.

Next, we consider the per-BS power constraint given in (15). Since the users in group $k$ have identical $\Lambda_k(\boldsymbol{\mu})$, independent of $i$, and that we can assume without loss of generality that they are all given positive powers $q_k^{(i)} > 0$ (notice: if a user were given zero power we could decrease the corresponding fraction $\mu_k$ and take it out of the active user set), also in this case we arrive at the conclusions that in the large system limit the allocated powers to users in group $k$ must be identical, i.e., $q_k^{(i)} = q_k$, independent of $i$. Using this in the constraint (15), we obtain

$$\sum_{k=1}^{A} q_k \theta_{m,k}(\boldsymbol{\mu}) \leq P_m, \quad m = 1, \ldots, B, \tag{25}$$

where

$$\theta_{m,k}(\boldsymbol{\mu}) = \sum_{i=1}^{\mu_k N} \theta_{m,k}^{(i)}(\boldsymbol{\mu}) = \frac{1}{N} \sum_{i=1}^{\mu_k N} \sum_{\ell=1+(m-1)\gamma N}^{m\gamma N} \left| \left[\mathbf{V}_{\boldsymbol{\mu}}\right]_{\ell,k}^{(i)} \right|^2. \tag{26}$$

It is interesting to notice that $\theta_{m,k}(\boldsymbol{\mu})$ is the squared Frobenius norm (normalized by $N$) of the submatrix of $\mathbf{V}_{\boldsymbol{\mu}}$ corresponding to the users in group $k$ (columns from $\mu_{1:k-1}N + 1$ to $\mu_{1:k}N$) and the antennas of BS $m$ (rows from $(m-1)\gamma N + 1$ to $m\gamma N$). The next result yields an analytical expression for the large-system limits of the coefficients $\theta_{m,k}(\boldsymbol{\mu})$:

*Theorem 2:* For all $m, k$, the following limit holds almost surely:

$$\lim_{N\to\infty} \theta_{m,k}(\boldsymbol{\mu}) = \frac{\mu_k \eta_m^2(\boldsymbol{\mu}) \left(\beta_{m,k}^2 + \xi_{m,k}\right)}{\sum_{\ell=1}^{B} \eta_\ell(\boldsymbol{\mu}) \beta_{\ell,k}^2} \tag{27}$$

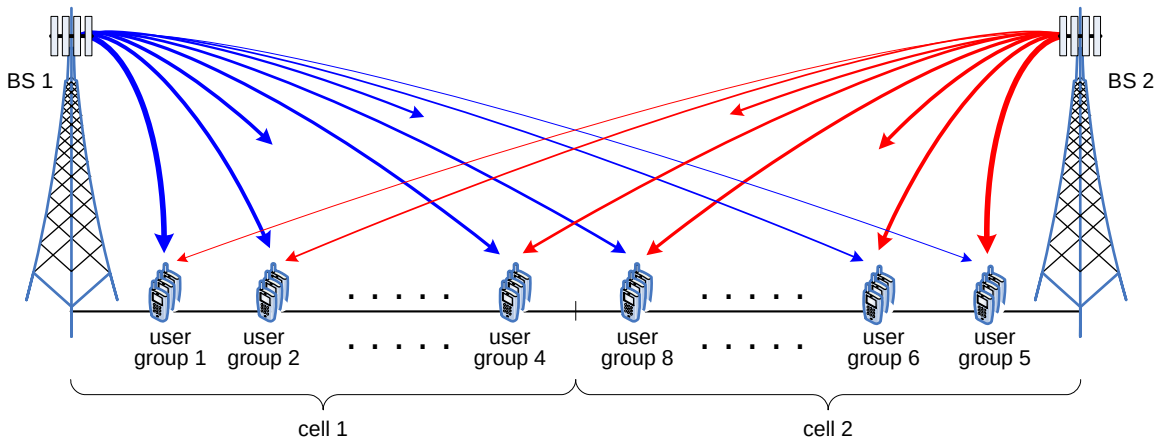



Fig. 1. A linear cellular layout with 2 BSs and 8 user groups

where $\boldsymbol{\xi}_m = (\xi_{m,1}, \ldots, \xi_{m,A})^{\mathsf{T}}$ is the solution to the linear system

$$[\mathbf{I} - \gamma \mathbf{M}]\, \boldsymbol{\xi}_m = \gamma \mathbf{M} \mathbf{b}_m \tag{28}$$

where $\mathbf{M}$ is the $A \times A$ matrix

$$\mathbf{M} = \left[ \sum_{\ell=1}^{B} \eta_\ell^2(\boldsymbol{\mu}) \mathbf{b}_\ell \mathbf{b}_\ell^{\mathsf{T}} \right] \operatorname{diag}\left( \frac{\mu_1}{\Lambda_1^2(\boldsymbol{\mu})}, \ldots, \frac{\mu_A}{\Lambda_A^2(\boldsymbol{\mu})} \right) \tag{29}$$

and $\mathbf{b}_\ell = (\beta_{\ell,1}^2, \ldots, \beta_{\ell,A}^2)^{\mathsf{T}}$, and the coefficients $\{\eta_m(\boldsymbol{\mu})\}$ and $\{\Lambda_k(\boldsymbol{\mu})\}$ are provided by Theorem 1.

*Proof:* See Appendix B. ■

Under some special symmetry conditions, the general problem simplifies significantly. In particular, assume that $B$ divides $A$, let $A' = A/B$, and assume that the $B \times A$ matrix of channel gains $\boldsymbol{\beta} = [\beta_{m,k}^2]$ can be partitioned into $A'$ submatrices of size $B \times B$ such that each submatrix has the property that all rows are permutations of the first row, and all columns are permutations of the first column. In analogy with "strongly symmetric" discrete memoryless channels, we shall refer to these submatrices as "strongly symmetric blocks." Also, we define a set of user groups whose corresponding columns in matrix $\boldsymbol{\beta}$ form a strongly symmetric block, as a user *equivalence class*. Then, we have $A'$ user equivalence classes, where each class corresponding to the user groups whose columns in the gain matrix $\boldsymbol{\beta}$ belong to the same strongly symmetric block. We re-index the user groups such that groups $\{(j-1)A' + i : j = 1, \ldots, B\}$ form the $i$-th equivalence class, for $i = 1, \ldots, A'$. To fix ideas, consider Figure 1 where a linear cellular layout with 2 cells and 8 user groups is considered, and assume that the gain coefficients depend on

distance and therefore are given as

$$\boldsymbol{\beta} = \begin{bmatrix} a & b & c & c & f & e & e & d \\ f & e & e & d & a & b & c & c \end{bmatrix} \tag{30}$$

for some positive numbers $a, b, c, d, e, f$. We notice that this matrix can be decomposed into the $A' = 4$ strongly symmetric blocks

$$\begin{bmatrix} a & f \\ f & a \end{bmatrix}, \quad \begin{bmatrix} b & e \\ e & b \end{bmatrix}, \quad \begin{bmatrix} c & e \\ e & c \end{bmatrix}, \quad \begin{bmatrix} c & d \\ d & c \end{bmatrix}$$

with the required property.

When this symmetry condition holds, under mild conditions on the concave network utility function, we have that the scheduler must set the fractions of all user groups in the same equivalence class to be equal. This is because such groups (e.g., user group pairs $(1, 5)$, $(2, 6)$, $(3, 7)$, and $(4, 8)$ in the example) are completely equivalent in terms of the set of channel gains seen from all BS in the cluster. We indicate the corresponding common fraction values as $\mu'_i$ for $i = 1, \ldots, A'$, such that $\mu_{(j-1)A'+i} = \mu'_i$ for all $j = 1, \ldots, B$. In this case, for any $m$, we have

$$\sum_{q=1}^{A} \mu_q \frac{\beta^2_{m,q}}{\gamma \sum_{\ell=1}^{B} \beta^2_{\ell,q}} = \frac{1}{\gamma} \sum_{i=1}^{A'} \mu'_i \sum_{j=1}^{B} \frac{\beta^2_{m,(j-1)A'+i}}{\sum_{\ell=1}^{B} \beta^2_{\ell,(j-1)A'+i}} = \frac{1}{\gamma} \sum_{i=1}^{A'} \mu'_i = \frac{\mu}{\gamma B}$$

where we have used the fact that, by the symmetry condition, $\sum_{j=1}^{B} \frac{\beta^2_{m,(j-1)A'+i}}{\sum_{\ell=1}^{B} \beta^2_{\ell,(j-1)A'+i}} = 1$ and $\sum_{i=1}^{A'} \mu'_i = \frac{1}{B} \sum_{q=1}^{A} \mu_q = \frac{\mu}{B}$. It follows that the solution of the fixed point equation (24) is given explicitly by

$$\eta_m(\boldsymbol{\mu}) = 1 - \frac{\mu}{\gamma B} \tag{31}$$

which is independent of $m$, and (23) yields

$$\Lambda_k(\boldsymbol{\mu}) = \gamma \left(1 - \frac{\mu}{\gamma B}\right) \sum_{m=1}^{B} \beta^2_{m,k}. \tag{32}$$

For all $k = (j-1)A' + i, \quad \forall j = 1, \ldots, B$ (indicating user groups in the same equivalence class), the sum $\sum_{m=1}^{B} \beta^2_{m,k}$ is a constant independent of $k$. Then, with some abuse of notation, we introduce the notation $\beta^2_i = \sum_{m=1}^{B} \beta^2_{m,k}$ for all groups $k$ in the $i$-th equivalence class. In addition, we have:

*Theorem 3:* For symmetric systems with $\mu_k = \mu'_i$ for all $k$ in equivalence class $i$, we have

$$\theta_{m,k}(\boldsymbol{\mu}) = \frac{\mu'_i}{B} \tag{33}$$

independently of $m$.

*Proof:* See Appendix C. ■

As an immediate corollary, we have that if all the BSs in the cluster have the equal power constraint, i.e., $P_1 = \ldots = P_B = P$, then the per-BS power constraint (25) coincides with the sum power constraint with $P_{\text{sum}} = BP$.

### *B. Weighted sum-rate maximization*

Using the obtained asymptotic results, we consider the weighted sum-rate maximization problem in (17). First, we focus on the sum-power constraint (13). In the large system limit, we consider maximizing the weighted aggregate user group rates, normalized by $N$, in the case where the weights $W_k^{(i)}$ for all users in group $k$ are identical (i.e., independent of $i$), and denoted by $W_k$. We shall see later that the weights are calculated by the scheduler that solves the general network utility maximization problem (5) and, under our assumptions on the form of $g(\cdot)$, these weights must be identical for statistically equivalent users. Therefore, this assumption does not involve any loss of generality when the weighted-sum rate maximization problem is used as an intermediate step for the scheduling rule that addresses the throughput maximization problem (5). Furthermore, by the symmetry of the water-filling power allocation, it also follows that the power $q_k^{(i)}$ allocated to all active users in group $k$ must be identical (independent of $i$), and can be denoted by $q_k$. In these conditions, from (23) we have that the instantaneous group rate converges to $\frac{1}{N}\sum_{i=1}^{N} R_k^{(i)} \to \mu_k R_k$ with

$$R_k = \log\left(1 + \Lambda_k(\boldsymbol{\mu}) q_k\right) \tag{34}$$

Notice that $R_k$ is the limit instantaneous rate for any user in group $k$ that is effectively scheduled (the non-scheduled users have zero instantaneous rate). Because of the large-system limit, this rate is a deterministic quantity.

Using (24), we can write the large-system limit weighted sum-rate maximization problem subject to the sum-power constraint in the form:

$$\text{maximize} \quad \sum_{k=1}^{A} W_k \mu_k \log\left(1 + \gamma\left(\sum_{m=1}^{B} \beta_{m,k}^2 \eta_m\right) q_k\right) \tag{35a}$$

$$\text{subject to} \quad \sum_{k=1}^{A} \mu_k q_k \le P_{\text{sum}}, \quad \sum_{k=1}^{A} \mu_k \le \gamma B, \tag{35b}$$

$$\eta_m = 1 - \sum_{k=1}^{A} \mu_k \frac{\eta_m \beta_{m,k}^2}{\gamma \sum_{\ell=1}^{B} \eta_\ell \beta_{\ell,k}^2}, \quad m = 1, \ldots, B \tag{35c}$$

$$0 \le \eta_m \le 1, \quad m = 1, \ldots, B \tag{35d}$$

$$q_k \ge 0, \;\; 0 \le \mu_k \le 1, \quad k = 1, \ldots, A \tag{35e}$$

This problem generally non-convex in $\mathbf{q}, \boldsymbol{\mu}$ and $\boldsymbol{\eta}$. However, for fixed $\boldsymbol{\eta}$ and $\boldsymbol{\mu}$, it is convex in $\mathbf{q}$, and the solution is given by water-filling (see also (18)):

$$q_k = \left[ \frac{W_k}{\lambda} - \frac{1}{\gamma \left( \sum_{m=1}^{B} \beta_{m,k}^2 \eta_m \right)} \right]_+ \tag{36}$$

For fixed $\boldsymbol{\eta}$ and $\mathbf{q}$, we have a linear program with respect to $\boldsymbol{\mu}$. Finally, for fixed $\boldsymbol{\mu}$ and $\mathbf{q}$ the problem is degenerate with respect to $\boldsymbol{\eta}$ since the equality constraint (35c), that corresponds to the fixed-point equation (24) has a unique solution $\boldsymbol{\eta} \in [0,1]^B$ for all feasible $\boldsymbol{\mu}$.

In the symmetric system case with the conditions given in the previous section, we have that user groups in the same equivalence class are completely symmetric, since the limits $\Lambda_k(\boldsymbol{\mu})$ depend only on the equivalence class and not on the specific user group in the class. Assuming that the user groups in the same equivalence class also have the same rate weights, the optimization problem in the symmetric case reduces to optimizing the powers $q_i'$ and the fractions $\mu_i'$ for the equivalence classes. Letting $\mu' = \sum_{i=1}^{A'} \mu_i' = \mu/B$, we can state the optimization problem in the symmetric case as:

$$\text{maximize} \quad B \sum_{i=1}^{A'} W_i \mu_i' \log \left( 1 + \gamma \left( 1 - \frac{\mu'}{\gamma} \right) \beta_i^2 q_i' \right) \tag{37a}$$

$$\text{subject to} \quad B \sum_{i=1}^{A'} \mu_i' q_i' \leq P_{\text{sum}}, \tag{37b}$$

$$\sum_{i=1}^{A'} \mu_i' \leq \gamma, \tag{37c}$$

$$q_i' \geq 0, \quad 0 \leq \mu_i' \leq 1, \quad i = 1, \ldots, A' \tag{37d}$$

The net effect of the symmetry is a sort of "resource pooling": the system with a cluster of $B$ cooperating BSs reduces to an equivalent single-BS system with total transmit power $P_{\text{sum}}/B$, load $\mu' = \mu/B$, $A' = A/B$ user classes, and equivalent channel path gains $\beta_i^2 = \sum_{m=1}^{B} \beta_{m,k}^2$ given by the combination of the path gains from all BSs in the cluster to the user groups in the $i$-th equivalence class.

In the case of per-BS power constraint with the general non-symmetric system involved, the power constraint in problem (35) must be replaced by (25) where the coefficients $\{\theta_{m,k}(\boldsymbol{\mu})\}$ are provided by Theorem 2. Finally, in the symmetric case, using Theorem 3 and assuming $P_m = P$ for all $m$, the power constraint reduces to

$$\sum_{i=1}^{A'} q_i' \mu_i' \leq P, \quad m = 1, \ldots, B.$$

Summing over $m$ we obtain the equivalent constraint $B \sum_{i=1}^{A'} q_i' \mu_i' \leq BP = P_{\text{sum}}$ which coincides with the sum power constraint in (37), as anticipated before.

### C. Optimization of the user fractions and powers

While (35) is still a non-convex problem in $(\mathbf{q}, \boldsymbol{\mu})$, we can find near-optimal solutions by borrowing from the greedy user selection heuristic used in the finite-dimensional case (see [31], [32]). In particular, we consider the approach of incrementing user fractions $\mu$ sequentially in very small steps, $\Delta\mu \ll 1$, until the objective function value cannot be increased any longer. If we take the infinitesimal of $\Delta\mu$, this is equivalent to greedy user selection in the large system limit where $\Delta\mu$ denoting the fraction of one user to the total number of users goes to zero. We start from $\boldsymbol{\mu} = \mathbf{0}$ and at each step we find $k$ such that incrementing $\mu_k$ by $\Delta\mu$ yields the largest improvement and the resulting new $\boldsymbol{\mu}$ is feasible. For the tentative configuration of the fractions $\boldsymbol{\mu}$, the corresponding power allocation is obtained from the waterfilling solution. We stop when no further increment can improve the objective function value. The detailed description is given in the following:

1) Initialize variables such that $n = 0$, $R_{\text{wsr}}(0) = 0$, $\boldsymbol{\mu} = \mathbf{0}$, and $\mu = 0$.
2) Set $n \leftarrow n + 1$. For $\Delta\mu \ll 1$, set $\boldsymbol{\mu}^{(k)} = \boldsymbol{\mu} + \Delta\mu \mathbf{e}_k$ (note: $\mathbf{e}_k$ denotes a vector of length $A$ of all zeros with a single 1 in position $k$), for $k \in \mathcal{S} = \{j : \mu_j + \Delta\mu \leq 1, \forall j\}$. If $\mathcal{S}$ is empty or $\mu + \Delta\mu > \gamma$, then exit and keep the current $\boldsymbol{\mu}$ and the corresponding rates as the final values of the algorithm. Otherwise, compute the tentative weighted sum rate value $R_{\text{wsr}}^{(k)}$ for each $k$, by solving the optimization problem in (35) for fixed $\boldsymbol{\mu}^{(k)}$ with the waterfilling power allocation.
3) Let $\widehat{k} = \arg\max_{k \in \mathcal{S}} R_{\text{wsr}}^{(k)}$ and set $R_{\text{wsr}}(n) = R_{\text{wsr}}^{(\widehat{k})}$.
4) If $R_{\text{wsr}}(n) > R_{\text{wsr}}(n-1)$, then set $\boldsymbol{\mu} \leftarrow \boldsymbol{\mu}^{(\widehat{k})}$, $\mu \leftarrow \mu + \Delta\mu$ and go back to step 2.
5) Otherwise, if $R_{\text{wsr}}(n) \leq R_{\text{wsr}}(n-1)$, exit and take the current $\boldsymbol{\mu}$ and the corresponding rates as the final values of the algorithm.

Figure 2 shows the sum rate versus $\mu' = \mu/B$ in a symmetric setting, when the above user fraction and power allocation algorithm is applied and compares it with the globally optimal value obtained from the exhaustive search algorithm. Under the two-cell model described in Section III-A, we assume the two BSs are cooperating (B=2), the channel gain coefficients are given as (30) with $[a\ b\ c\ d\ e\ f] = [1.5\ 1.3\ 1.0\ 0.5\ 0.3\ 0.2]$, and the antenna ratio, transmit power, and user group weight are $\gamma = 4$, $P = 15$ dB, and $W_i = 1,\ \forall i$, respectively. The exhaustive algorithm searches for the optimal weighted sum rate in the $A'$ dimensional space of $\boldsymbol{\mu}'$ where each $\mu_i'$ is ranged from 0 to 1 with $\sum_{i=1}^{A'} \mu_i' < \gamma$. If we discretize this domain with $\Delta\mu$ step size for each dimension, the computational complexity of the exhaustive algorithm is roughly $O((1/\Delta\mu)^{A'})$, whereas the greedy algorithm has $O(A'\gamma/\Delta\mu)$ complexity. For the greedy algorithm curve, we removed the comparison between $R_{\text{wsr}}(n)$ and $R_{\text{wsr}}(n-1)$ in step 4 to see

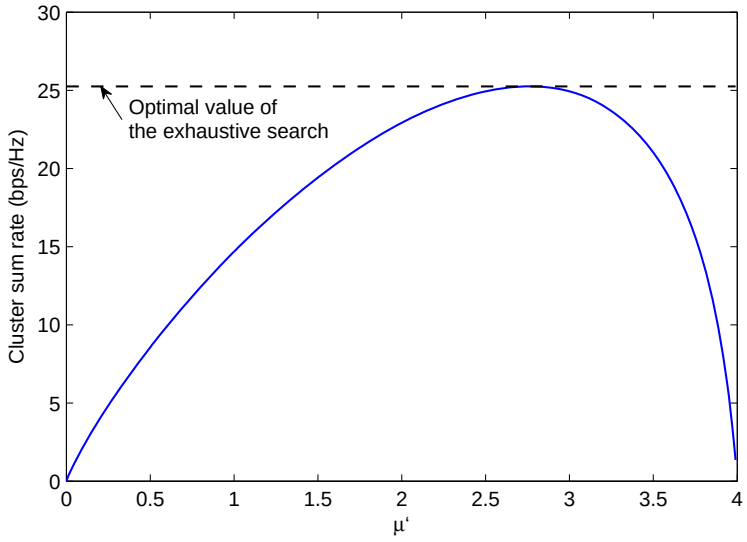



Fig. 2. Cluster sum rate ($B = 2$) of the proposed user fraction and group power optimization as $\mu'$ increases from 0 to $\gamma = 4$.

the objective function over the range from 0 to $\gamma$. When $\Delta\mu = 0.01$, the greedy algorithm achieves the same optimal value with the exhaustive algorithm at $\mu' = 2.76$.

### *D. Network utility function maximization*

In general, the solution of (6) (or (35) in the large system limit) for the case $A > \gamma B$ (more users than antennas) yields an unbalanced distribution of instantaneous rates, where some user classes are not served at all (we have $\mu_k = 0$ for some $k$). This shows that, for a general strictly concave network utility function $g(\cdot)$, the ergodic rate region $\overline{\mathcal{R}}$ requires time-sharing even in the asymptotic large-system case. Finding the solution of (5) is therefore extremely hard. Nevertheless, this solution can be computed to any level of accuracy by using a method inspired by the stochastic optimization approach of [20]. Interestingly, the algorithm can be used both as for the computation of the optimum throughput point in the large system limit, and as an actual downlink scheduling algorithm, that can be applied almost verbatim to the actual finite-dimensional system. In the former case, the algorithm is equivalent to Lagrangian iteration where the "virtual queues" (to be defined in the following) plays the role of Lagrange multipliers. In the latter case, when applied to the finite dimensional system, the algorithm performs a stochastic "Lyapunov drift" optimization (see [20]).

For each user group $k = 1, \ldots, A$, define a *virtual queue* that evolves according to

$$Q_k(t+1) = [Q_k(t) - r_k(t)]_+ + a_k(t) \tag{38}$$

where $r_k(t)$ denotes the virtual service rate and $a_k(t)$ the virtual arrival process. The queues are initialized by $Q_k(0) = r_k(0) = 0$. Then, at each iteration $t = 1, 2, \ldots$, the virtual arrival processes is given by $a_k(t) = a_k^\star$ where $\mathbf{a}^\star$ is the solution of the convex problem:

$$\begin{aligned} &\text{maximize} \quad Vg(\mathbf{a}) - \sum_{k=1}^{A} Q_k(t) a_k \\ &\text{subject to} \quad 0 \leq a_k \leq a_{\max}, \quad \forall\, k \end{aligned} \tag{39}$$

and where $V, a_{\max} > 0$ are some suitably chosen constants, that determine the convergence properties of the iterative algorithm. The service rates are given by

$$r_k(t) = \mu_k(t) \log \left( 1 + \gamma \left( \sum_{m=1}^{B} \beta_{m,k}^2 \eta_m(t) \right) q_k(t) \right)$$

where $(\boldsymbol{\mu}(t), \mathbf{q}(t), \boldsymbol{\eta}(t))$ is the solution of (35) for weights $W_k = Q_k(t)$. Then, the virtual queues are updated according to (38). The theory developed in [20] (see also [21]) ensures the following result. Let $\mathbf{r}(t)$ denote the vector of service rates generated by the above iterative algorithm. Then,

$$\liminf_{t \to \infty} g \left( \frac{1}{t} \sum_{\tau=0}^{t-1} B\mathbf{r}(\tau) \right) \geq g(\overline{\mathbf{R}}^\star) - \frac{\mathcal{K}}{V} \tag{40}$$

where $\overline{\mathbf{R}}^\star$ is the solution of (5) and $\mathcal{K}$ is a constant that depends on the system parameters and on $a_{\max}$. In particular, using the results in [21] we can prove the bound

$$\mathcal{K} \leq \frac{A}{2} \left( a_{\max}^2 + \log^2 \left( 1 + \gamma \max\{\beta_{m,k}^2 : \forall\, m, k\} P_{\text{sum}} \right) \right)$$

By choosing $V$ and $a_{\max}$ appropriately, we can ensure a desired tradeoff between the accuracy of the approximation of the optimum point and the convergence speed of the iterative algorithm.

It should be noticed that if we use the greedy optimization of the user fractions as described in Section III-C instead of the exact solution of (37), then the performance guarantee (40) is no longer valid. However, the algorithm ensures that the throughput point that maximizes $g(\cdot)$ over the ergodic rate region achievable with the (suboptimal) greedy optimization of the user fractions can be approached arbitrarily closely.

## IV. Channel Estimation and Non-Perfect CSIT

So far, we have assumed that the transmitter (cluster controller) has perfect CSIT. In this section we consider the case where the BSs in the cluster broadcast a set of downlink pilot signals in order to enable the users to measure their downlink channels and feed back channel state information in some form, in order to provide CSIT and enable the computation of the LZFB precoding matrix, user selection, and

power allocation. We seek the characterization of the non-trivial tradeoff between the advantage of having a large number of transmit antennas or cluster size (large $\gamma$ and/or large $B$) and the overhead required for estimating the channels. We assume that the channels are constant over time-frequency blocks of size $WT$ complex dimensions, where $W$ denotes to the system coherence bandwidth (in Hz) and $T$ denotes the system coherence time (in sec.). For each such block, $\gamma_p BN$ dimensions are dedicated to downlink training, in order to allow all users in the cluster to estimate the composite channel (i.e., the corresponding column of $\mathbf{H}$ in (4)) formed by $\gamma BN$ coefficients. Since the channel vectors are Gaussian, linear MMSE estimation is optimal with respect to the MSE criterion. A simple dimensionality argument shows that the MSE can be made arbitrarily small as $\sigma_k^2 \to 0$ (vanishing noise plus ICI) if and only if $\gamma_p \geq \gamma$. The ratio $\gamma_p/\gamma$ denotes the "pilot dimensionality overhead", relative to the minimum number of pilots dimensions that allow MMSE estimation, under the condition that the MMSE vanishes as $\sigma_k^2 \to 0$. In the following, we assume that this condition holds.

Focusing on the estimation of a generic column of $\mathbf{H}$ in (4) corresponding to some user $j$ in group $k$ of the reference cluster, the channel model of downlink channel estimation based on the common pilots is given by

$$\mathbf{y}_k^{(j)} = \mathbf{T}\underline{\mathbf{h}}_k^{(j)} + \mathbf{z}_k^{(j)} \tag{41}$$

where $\mathbf{T}$ is a $\gamma_p BN \times \gamma BN$ training matrix with equal-energy orthogonal columns, corresponding to the training sequences sent in parallel from the $\gamma BN$ antennas of the cluster joint transmitter (notice that the vertical dimension corresponds to channel uses, or "time", and the horizontal dimension corresponds to the antennas), the vector $\underline{\mathbf{h}}_k^{(j)}$ is the corresponding channel vector, obtained by the stacking of the channels (including their path coefficients) from the different base stations forming the cluster, and $\mathbf{z}_k^{(j)}$ is a vector of i.i.d. $\mathcal{CN}(0,1)$ normalized noise plus interference samples. For simplicity, we re-index the base stations forming the reference cluster as $m = 1, \ldots, B$ and the user groups as $k = 1, \ldots, A$. With this notation, from (4) we have

$$\underline{\mathbf{h}}_k^{(j)} = \begin{bmatrix} \beta_{1,k}\mathbf{h}_{1,k}^{(j)} \\ \vdots \\ \beta_{B,k}\mathbf{h}_{B,k}^{(j)} \end{bmatrix} \tag{42}$$

where $\mathbf{h}_{i,k}^{(j)}$ denotes the $j$-th column of the block $\mathbf{H}_{i,k}$, with i.i.d. $\mathcal{CN}(0,1)$ elements.

The equal-energy and orthogonality condition on the columns of the training matrix $\mathbf{T}$ yield that the total transmit power (energy per channel use) in the training phase is given by

$$\frac{1}{\gamma_p BN}\mathrm{tr}\left(\mathbf{T}^{\mathsf{H}}\mathbf{T}\right) = \frac{\gamma}{\gamma_p}p \tag{43}$$

where we let $\mathbf{T}^{\mathsf{H}}\mathbf{T} = p\mathbf{I}$, and $p$ denotes the energy of the training sequences. Letting the total training power equal to the total cluster transmit power, we obtain

$$p = \frac{\gamma_p}{\gamma} \sum_{m=1}^{B} P_m$$

Noticing that

$$\mathrm{Cov}(\underline{\mathbf{h}}_k^{(j)}) = \mathrm{diag}(\beta_{1,k}^2\mathbf{I}, \ldots, \beta_{B,k}^2\mathbf{I}) \triangleq \mathbf{D}_k$$

has block-diagonal structure with diagonal blocks given by scaled $\gamma N \times \gamma N$ identity matrices, we immediately obtain the MMSE estimator of $\underline{\mathbf{h}}_k^{(j)}$ in the form

$$\widehat{\underline{\mathbf{h}}}_k^{(j)} = \mathbf{D}_k \left(\mathbf{I} + p\mathbf{D}_k\right)^{-1} \mathbf{T}^{\mathsf{H}} \mathbf{y}_k^{(j)} \tag{44}$$

with estimation error covariance given by

$$\mathbf{\Sigma}_k = \mathbf{D}_k - p\mathbf{D}_k \left(\mathbf{I} + p\mathbf{D}_k\right)^{-1} \mathbf{D}_k = \mathbf{D}_k \left(\mathbf{I} + p\mathbf{D}_k\right)^{-1} \tag{45}$$

The MMSE covariance matrix is also block diagonal, with scaled identities diagonal blocks, and it depends only on the user group index $k$ and not on the individual user in the group (this is obvious since the users in the same group are statistically equivalent).

From the well-known orthogonality condition of MMSE estimation and from joint Gaussianity, we have the canonical decomposition

$$\underline{\mathbf{h}}_k^{(j)} = \widehat{\underline{\mathbf{h}}}_k^{(j)} + \underline{\mathbf{e}}_k^{(j)} \tag{46}$$

where the estimator $\widehat{\underline{\mathbf{h}}}_k^{(j)}$ and the error $\underline{\mathbf{e}}_k^{(j)}$ are independent, and such that

$$\mathrm{Cov}(\widehat{\underline{\mathbf{h}}}_k^{(j)}) = \mathbf{D}_k - \mathbf{\Sigma}_k = p\mathbf{D}_k \left(\mathbf{I} + p\mathbf{D}_k\right)^{-1} \mathbf{D}_k \tag{47}$$

Putting everything together, we can write the channel matrix $\mathbf{H}$ in (4) in the form $\mathbf{H} = \widehat{\mathbf{H}} + \mathbf{E}$, where

$$\widehat{\mathbf{H}} = \begin{bmatrix} \widehat{\beta}_{1,1}\mathbf{H}_{1,1} & \cdots & \widehat{\beta}_{1,A}\mathbf{H}_{1,A} \\ \vdots & & \vdots \\ \widehat{\beta}_{B,1}\mathbf{H}_{B,1} & \cdots & \widehat{\beta}_{B,A}\mathbf{H}_{B,A} \end{bmatrix}, \tag{48}$$

with

$$\widehat{\beta}_{m,k} = \frac{\beta_{m,k}^2}{\sqrt{1/p + \beta_{m,k}^2}}, \tag{49}$$

and the blocks $\mathbf{H}_{m,k}$ are independent with i.i.d. $\mathcal{CN}(0,1)$ elements, and where $\mathbf{E}$ is independent of $\widehat{\mathbf{H}}$, and is given in the form

$$\mathbf{E} = \begin{bmatrix} \bar{\beta}_{1,1}\mathbf{E}_{1,1} & \cdots & \bar{\beta}_{1,A}\mathbf{E}_{1,A} \\ \vdots & & \vdots \\ \bar{\beta}_{B,1}\mathbf{E}_{B,1} & \cdots & \bar{\beta}_{B,A}\mathbf{E}_{B,A} \end{bmatrix}, \tag{50}$$

with

$$\bar{\beta}_{m,k} = \sqrt{\beta_{m,k}^2 - \widehat{\beta}_{m,k}^2} = \frac{\beta_{m,k}}{\sqrt{1+p\beta_{m,k}^2}}, \tag{51}$$

and the blocks $\mathbf{E}_{m,k}$ and independent with i.i.d. $\mathcal{CN}(0,1)$ elements.

In a practical FDD system, the users should feed back their estimated channel on each time-frequency block, i.e., for each new observation. Several schemes have been proposed for closed-loop CSIT feedback, including codebook-based vector quantization, scalar quantization of the channel coefficients, and unquantized "analog" feedback. Furthermore, the feedback takes place on the uplink, and can be performed by accessing the uplink channel in FDMA/TDMA, or exploiting the MIMO-MAC nature of the uplink in order to allow a number of users proportional to the number of receiving antennas to send their feedback signals simultaneously [37], [45], [46]. Analyzing the system in the presence of a specific feedback scheme is possible, although even more cumbersome [37]. However, from the results in the above mentioned papers we know that a well-designed digital feedback scheme can achieve a quantization error that is negligible with respect to the downlink training estimation error. Furthermore, this can be done with a moderate use of the uplink feedback total capacity, provided that the number of users feeding back their CSIT is not too large (see for example the optimization tradeoff in [47]. For the sake of simplicity, here we assume an ideal genie-aided CSIT feedback that provides $\widehat{\mathbf{H}}$ directly to the centralized cluster controller at no additional costs, either in terms of rate or in terms of CSIT distortion. This provides a "best case" for any CSIT scheme based on explicit downlink training and any form of feedback. Then, since an actual system implementation CSIT feedback has a cost that impacts on the uplink capacity, we shall propose a randomized scheduling scheme that pre-selects a subset of users and therefore limits the number of users actually feeding back their CSIT in the next section.

Under these assumptions, the cluster transmitter computes a mismatched LZFB precoding matrix from the estimated channel matrix $\widehat{\mathbf{H}}$ instead of $\mathbf{H}$. The following theorem yields an achievability lower bound on the large-system performance of the mismatched LZFB:

*Theorem 4:* Under the downlink training scheme described above and assuming genie-aided CSIT

feedback, the achievable rate of users in group $k$ is lower bounded by

$$R_k \geq \log\left(1 + \frac{\widehat{\Lambda}_k(\boldsymbol{\mu}) q_k}{1 + \sum_{m=1}^{B} \bar{\beta}_{m,k}^2}\right) \tag{52}$$

where

$$\widehat{\Lambda}_k(\boldsymbol{\mu}) = \gamma \sum_{m=1}^{B} \widehat{\beta}_{m,k}^2 \eta_m(\boldsymbol{\mu}) \tag{53}$$

where $(\eta_1(\boldsymbol{\mu}), \ldots, \eta_B(\boldsymbol{\mu}))$ is the unique solution with components in $[0,1]$ of the fixed point equation

$$\eta_m = 1 - \sum_{q=1}^{A} \mu_q \frac{\eta_m \widehat{\beta}_{m,q}^2}{\gamma \sum_{\ell=1}^{B} \eta_\ell \widehat{\beta}_{\ell,q}^2}, \quad m = 1, \ldots, B \tag{54}$$

with respect to the variables $\boldsymbol{\eta} = \{\eta_m\}$.

*Proof:* See Appendix D. ■

The conclusion of this section is that all the derivations and the optimization made before for the case of perfect CSIT, including the system symmetry conditions, can be applied straightforwardly to the case of non-ideal CSIT, provided that the per-user rates are replaced with the corresponding terms in (52). In particular, Theorem 2 is valid by replacing $\{\beta_{m,k}\}$ with $\{\widehat{\beta}_{m,k}\}$ given in (49).

Finally, the system spectral efficiency must be scaled by the factor $\left[1 - \frac{\gamma_p NB}{WT}\right]_+$, that takes into account the downlink training overhead, i.e., fraction of dimensions per block dedicated to training. In particular, letting $\tau = \frac{N}{WT}$ denote the ratio between the number of users per group, $N$, and the dimensions in a time-frequency slot, we can investigate the system spectral efficiency for fixed $\tau$, in the limit of $N \to \infty$. The ratio $\tau$ captures the "dimensional crowding" of the system. It is clear that a highly underspread system ($WT \gg 1$) can accommodate more users, and more jointly coordinated antennas at the transmitter. Vice versa, if $WT$ is not much larger than $N$, then the number of jointly coordinated transmit antennas (captured by the product $\gamma B$) is intrinsically limited by the channel time-frequency coherence.

## V. Numerical Results and Simplified Scheduling

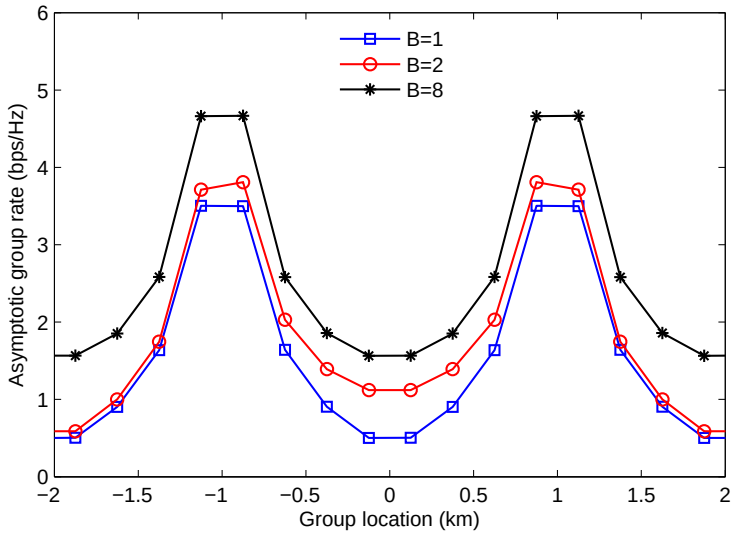
In this section, first we provide a comparison between the large-system limit results and the Monte Carlo simulation of the corresponding finite-dimensional systems with greedy user selection. Then, driven by the behavior of the finite-dimensional system, we propose a simplified scheduling algorithm that randomly pre-select users according to the probability obtained from the asymptotic analysis. As explained in Section I, this algorithm has the advantage that the feedback required for CSIT is greatly reduced, since only the users that are effectively served are going to feedback their CSIT. However, quantifying the feedback resource requirement in a precise manner is out of the scope of this paper. Finally, we consider

the case of non-perfect CSIT and investigate the tradeoff between increasing the number of jointly coordinated antennas and the dimensionality cost incurrent by downlink training for channel estimation.

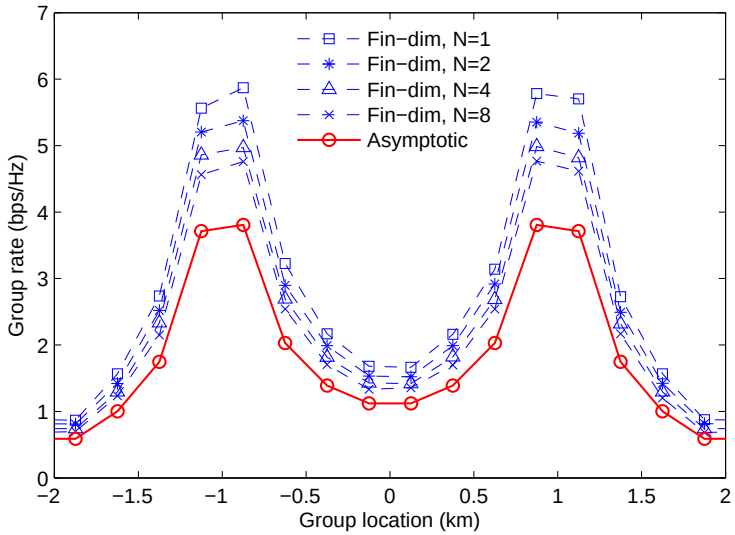
In the multi-cell system model, we consider a linear cellular arrangement where $M$ base stations are equally spaced on the segment $[-M, M]$ km, in positions $2m - M - 1$ for $m = 1, \ldots, M$ and $K$ user groups are also equally spaced on the same segment, with $K/M$ user groups uniformly spaced in each cell. The distance $d_{m,k}$ between BS $m$ and user group $k$ is defined modulo $[-M, M]$, i.e., we assume a wrap-around topology in order to eliminate boundary effects. We use a distance-dependent pathloss model given by $\alpha_{m,k}^2 = G_0/(1+(d_{m,k}/\delta)^\nu))$ and the pathloss parameters, $G_0, \nu$, and $\delta$ follow the mobile WiMAX system evaluation specifications [18], such that the 3dB break point is $\delta = 36$m (i.e., 3.6% of 1 km cell radius), the pathloss exponent is $\nu = 3.504$, the reference pathloss at $d_{m,k} = \delta$ is $G_0 = -91.64$ dB, and the per-BS transmit power normalized by the noise power at user terminals is $P = 154$ dB.

*a) Comparison with finite-dimensional systems:* Figure 3 shows the average user throughputs (bit/s/Hz) versus user locations for the first two cells near the origin (given the symmetry, this pattern repeats periodically), for the case of $M = 8$ cells, $K = 64$ user groups, cluster size $B = 1, 2$ and 8 and $\gamma = 4$. Notice that with 8 user groups per cell and $\gamma = 4$, we have twice as much users than antennas in each cell. The case $B = 8$ corresponds to the network-wide full cooperation. For the finite-dimensional Monte Carlo simulation, we applied the same stochastic optimization algorithm described in Section III-D, where now $t$ denotes a scheduling time slot index, and for each $t$ a new set of i.i.d. channel vectors is generated. In this case, the instantaneous weighted sum-rate is obtained via the user selection algorithm of [31], assuming that the CSIT for all users in the systems is available at the cluster controllers. As far as the network utility function $g(\cdot)$ is concerned, we consider the Proportional Fairness (PF) criterion, corresponding to $g(\overline{\mathbf{R}}) = \sum_k \log \overline{R}_k$. This PF criterion is applied to all the numerical results in this section.

From Figure 3(a), we notice that the advantage of full cooperation is significant, whereas the cluster of size $B = 2$ yields a significant improvement for the users in the center of the cluster, with respect to the basic cellular system with no cluster cooperation ($B = 1$). In Figure 3(b), we compare the asymptotic results with the finite-dimensional simulation results in the case of $B = 2$. The finite dimensional system yields better per-user throughput than the large-system limit, thanks to the ability of the user selection to exploit the randomness in the instantaneous channel realizations (multiuser diversity). However, as the number of users at each location, $N$ grows, the diversity gain continues to decline, for example, in this figure, the relative gain of the finite-dimensional rate to the asymptotic rate is about $55\%$ for $N = 1$ but $25\%$ for $N = 8$. In the case of $B = 1$ and $B = 8$ which are not shown here, the same trends are observed




(a) Asymptotic analysis




(b) Finite dimension simulation for $B = 2$

Fig. 3. User throughputs obtained from asymptotic analysis for cooperation clusters of size $B$ =1 (no cooperation), 2, and 8 (full cooperation) and from finite dimension simulation with greedy user selection for $B = 2$ and $N$ =1, 2, 4, and 8, with perfect CSIT. $M = 8$ cells and $K = 64$ user groups.

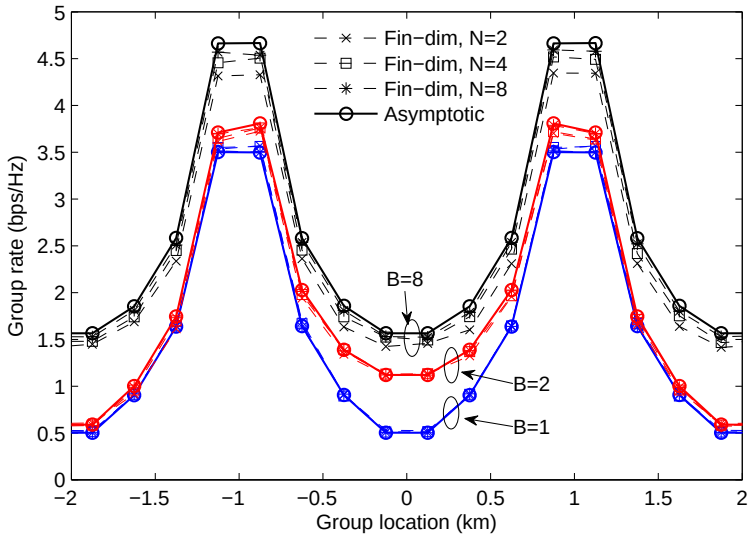



Fig. 4. User group rate in finite dimension ($N$ = 2, 4, and 8) with random user selection and power allocation aided by the asymptotic results for cooperation clusters of size $B$=1, 2, and 8, with perfect CSIT. $M = 8$ cells and $K = 64$ user groups.

even though the diversity gain is slightly larger ($B = 1$) or smaller ($B = 8$). It is well-known that for large systems (large $N$), this multiuser diversity effect disappears because of "channel hardening" [33], [34].

*b) Random user selection scheme for reduced CSIT feedback:* User selection requires a large amount of CSIT feedback since it needs CSIT from many users in order to select a good subset at each scheduling slot, even though no more users than the number of antennas can be served at a time. For systems with finite but large size, it is not wise to have many more users than transmit antennas to feedback their CSIT, since the multiuser diversity effect becomes marginal whereas the feedback resource grows at least linearly with the number of users feeding back their CSIT at each slot. In this regime, a better option consists of pre-select the users to be served in each slot, such that only these users feed back their CSIT. In this case, we have to design a user pre-selection scheme that approximately maximizes the desired network utility function. For example, a simple round-robin scheme may perform far from the desired PF optimal point.

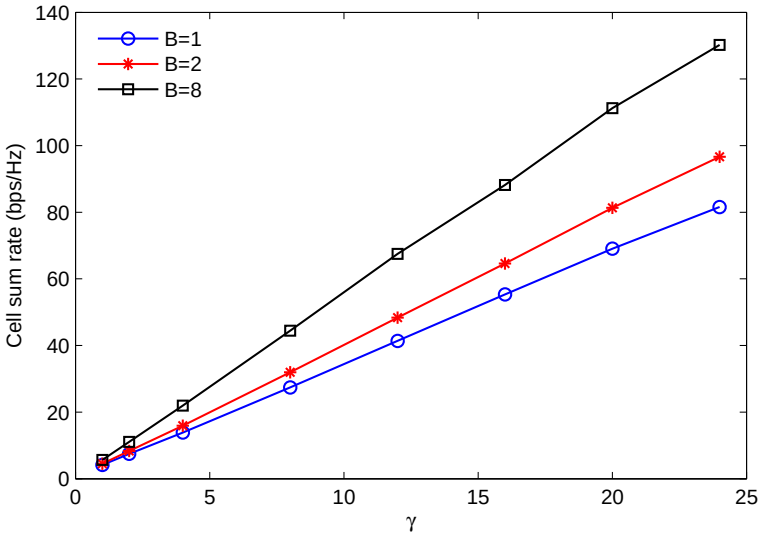
For this purpose, we consider a randomized scheduling scheme based on our asymptotic analysis that effectively provides such user pre-selection. We stress the fact that this may be useful for the sake of limiting the total CSIT feedback requirement, while still approximating the optimal proportional fairness throughput point. In the proposed scheme, the users to which CSIT feedback is requested are randomly

selected in each slot $t$ as follows: let $\{\mu_k\}$ be the user fractions per group of (approximately) co-located users, which is obtained from the asymptotic analysis. The cluster controller has a maximum of $\gamma BN$ independent data streams to transmit using LZFB (equal to the number of jointly coordinated transmit antennas). At each time $t$, the scheduler generates $\gamma BN$ i.i.d. random variables $S_1(t), \ldots, S_{\gamma BN}(t)$, taking values on the integers $\{0, 1, \ldots, A\}$ with probability $\mathbb{P}(S_i(t) = k) = \frac{\mu_k}{\gamma B}$ for $k \neq 0$ and $\mathbb{P}(S_i(t) = 0) = 1 - \sum_{k=1}^{A} \frac{\mu_k}{\gamma B}$. Then, user group $k$ is served by stream $i$ at time slot $t$ if $S_i(t) = k$. Notice that streams $i$'s for which $S_i(t) = 0$ are not used and that multiple streams may be associated to the same user group. Finally, for each stream a user in the associated group is selected at random, making sure that streams serve distinct users. Once the allocation of streams to users is determined, the selected users are requested to feedback their CSIT and the scheduler optimizes the transmit powers by solving the weighted sum rate maximization problem with weights $W_k = \partial g(\overline{\mathbf{R}})/\partial \overline{R}_k$, corresponding to the optimal asymptotic throughput point. In the special case of PF scheduling, this is given by $W_k = 1/\overline{R}_k$, [19].

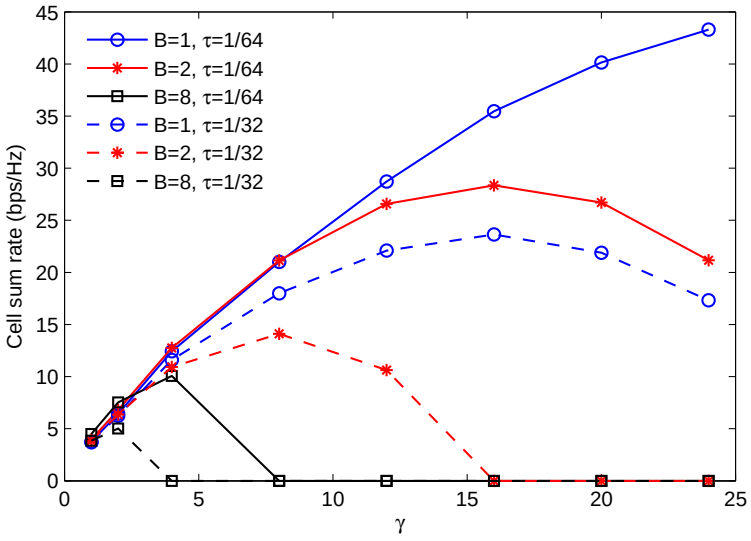
The finite-dimension simulation results under this random user selection scheme is compared with the asymptotic results in Figure 4 under the same system setting as in Figure 4. As $N$ increases, the finite-dimensional results converge to the infinite-dimensional limit and they are almost overlapped, especially when $B = 1$ or 2. Hence, the proposed scheme is effective for systems of finite but moderately large size.

*c) Non-perfect CSIT and coordination vs. estimation tradeoff:* Figure 5 shows the cell sum rate (cluster sum rate normalized by the number of cooperating cells in the reference cluster) versus values of $\gamma$ in the cases of (a) perfect CSIT and no consideration of training overhead, and (b) non-perfect CSIT and explicit downlink training with $\gamma_p = \gamma$. We consider a larger number of user groups, $K = 192$ in the $M = 8$ cells. As shown in Figure 5(a), under the assumption of perfect CSIT given at no cost, the cell sum rate grows almost linearly as $\gamma$ (the ratio of BS antennas over the users per group) increases, and grows also as $B$ (cluster size) increases, which shows the inter-cell cooperation and larger number of transmit antenna gain. However, when the CSIT estimation error and downlink training overhead are taken into account, there is a non-trivial tradeoff between the improvement owing to more and more jointly coordinated transmit antennas and the cost of estimating higher and higher dimensional channels.

Notice that this tradeoff is "fundamental", in the following sense: a trivial upper bound on the achievable sum capacity of the reference cluster is obtained by letting all users perfectly cooperate as a single multi-antenna receiver. The capacity of the resulting block-fading single-user MIMO channel with $\gamma BN$ transmit antennas and $AN$ receiving antennas and fading coherence block $WT = N/\tau$ was characterized in the high-SNR regime in [48], [49]. Using this result, in the case $\frac{1}{2\tau} \geq A \geq \gamma B$, the dimensionality




(a) Perfect CSIT and no training overhead




(b) Non-perfect CSIT and downlink training with $\gamma_p = \gamma$

Fig. 5. Cell sum rate versus the antenna ratio $\gamma$ for cooperation clusters of size $B$=1, 2, and 8. $M = 8$ cells and $K = 192$ user groups.

"pre-log" loss factor with respect to the case of ideal CSIT is given by $\left(1-\frac{\gamma BN}{WT}\right)$ that coincides with what is said at the end of Section IV with choice of $\gamma_p = \gamma$. In fact, the "pre-log" optimality of explicit training for single-user MIMO channels with block fading in the high-SNR regime is well-known [49], [50]. Also, the same result show that if $\frac{1}{2\tau} < \min\{A, \gamma B\}$, then there is no point in using more than $WT/2$ jointly coordinated antennas. Finally, notice that the recently proposed schemes for "blind" interference alignment [51], exploiting reconfigurable antennas at the user terminals, still require channel state information at the receiver (CSIR) for coherent detection at each user terminal. Since the resulting channel is MIMO point-to-point, the same downlink training said above appears. In other words, these "blind" interference alignment schemes avoid CSIT feedback, but still require downlink training in the same amount considered in this work. In conclusions, while we have analyzed a specific downlink training scheme, we have that, for a cluster in isolation, the sum capacity scaling in the high-rate regime (high-SNR) is indeed the correct one.

Figure 5 shows the cell sum rate with consideration of training overhead and estimation error for $\gamma_p = \gamma$. Inspired by practical system values, we chose $\tau = 1/64$ and $1/32$. In the finite-dimensional case, this corresponds to $WT = 640$ or 320 signal dimensions, respectively, with $N = 10$ users per user groups (total $KN/M = 240$ users per cell). We notice that as $\gamma$ increases, the sum rates in most cases grow at first, achieve some maximum point and decrease, due to the tradeoff between the benefit from a large number of antennas and the training overhead cost. For given $B$ and $\tau$, the maximum sum rate is achieved at $\gamma B = \frac{1}{2\tau}$, which is in line with the result of the high-SNR regime when $\gamma B \leq A$. For example, for $B = 2$ and $\tau = 1/64$, the sum rate is maximum at $\gamma = 16$ where $2\gamma = \frac{1}{2(1/64)}$. For $B = 1$ and $\tau = 1/64$, the optimal $\gamma$ is beyond the number of user groups per cell. We can also see that, when the number of antennas is large, the no cooperation case ($B = 1$) achieves the highest sum rate for both $\tau = 1/64$ and $1/32$, which suggests that no cooperation gain can be expected, because the improvement of multi-antenna gain does not compensate for the dimensional decrease (pre-log factor) due to the training overhead.

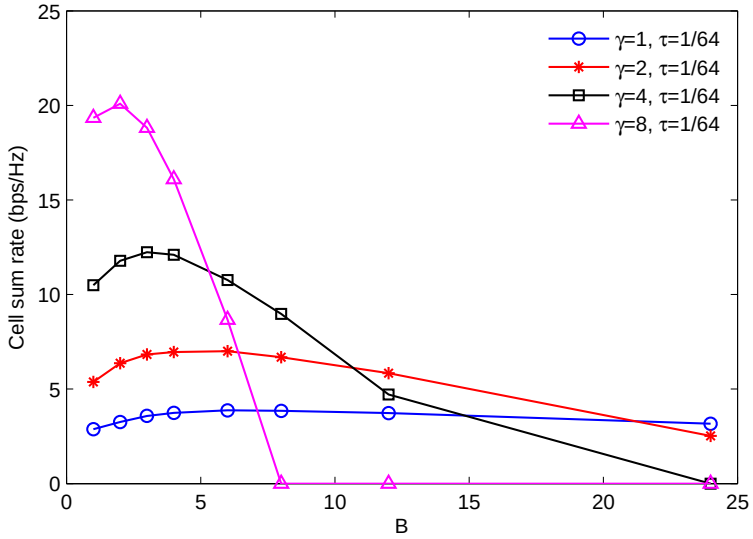
In order to see the best cluster size with downlink training and estimation, we consider a system with a large number of cells, $M = 24$. Figure 6 illustrates the cell sum rate versus the cluster size $B$ for $\gamma = 1, 2, 4$, and 8 and $\tau = 1/64$ and $1/32$ with $\gamma_p = \gamma$. In a linear cellular arrangement with $M = 24$, the clusters except for $B = 1, 2$, or 24, do not have the symmetric structure described in Section III-A. So for those clusters, we notice that the solution of problem (35) under the cluster sum-power constraint produces an upper-bound of the optimal value under per-BS power constraint. Even though not explicitly shown in the figure, we can confirm that the cluster sum rate ($B$ times the cell sum rate) is maximized

when $\gamma B = \frac{1}{2\tau}$ but as far as the cell sum rate is concerned, the optimal $\gamma$ for given $B$ and $\tau$ is smaller than the optimal one in terms of the cluster sum rate, i.e., $\frac{1}{2\tau\gamma}$. For example, in Figure 6(a), when $\gamma = 4$ and $\tau = 1/64$, the maximum cluster sum rate is achieved at $B = 8$, but the cell sum rate given as the cluster sum rate divided by $B$ is maximum at $B = 3$. When the channel is more time or frequency selective ($\tau = 1/32$), the optimum cluster size gets even smaller, as shown in Figure 6(b). Furthermore, the cell sum rate is more sensitive to the cluster size, when the number of antennas is larger.

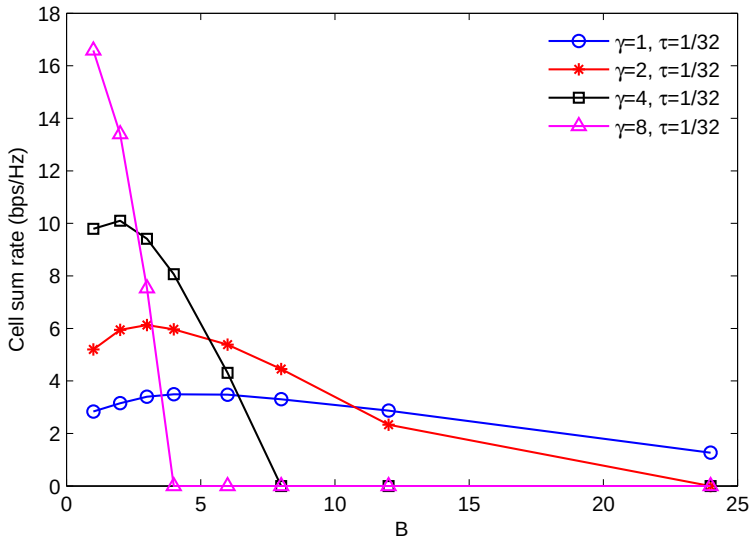
## VI. Conclusions

We considered a multi-cell "network MIMO" system in a realistic cellular scenario, with inter-cell cooperation and fairness criteria. Specifically, we focused on linear zero-forcing beamforming combined with user selection. We derived the asymptotic expression in the large system limit and proposed an algorithm that computes the throughput point under an arbitrary fairness criterion, expressed by the maximization of a suitable concave and componentwise increasing network utility function over the region of achievable ergodic user rates. The proposed method handles the per-cluster sum-power constraint. We showed that under certain system symmetries, this coincides with the more stringent per-BS power constraint. In particular, the system symmetries make the analysis much simpler, as it allows for a closed-form solution of a fixed-point equation that characterized the zero-forcing beamforming performance. The fairness scheduling was applied in the form of stochastic network optimization. The proposed asymptotic analysis is computationally much more efficient than the Monte Carlo simulation. It also provides a good approximation of finite-dimensional systems, when the users are randomly selected according to the asymptotic user fraction in the large system limit. In particular, we proposed a random user selection scheme that associates users with downlink data streams according to probabilities obtained from the asymptotic analysis, and provides a good approximation of the optimal throughput point while requiring much less CSIT feedback resource.

Our analytic tool can be extended to handle explicit channel state information estimation, obtained from downlink training. This allows the investigation of the tradeoff between the number of jointly coordinated antennas and the cost of estimating higher dimensional channels. This tradeoff yields the optimal "cooperation cluster size" that maximizes the system throughput subject to fairness, when the cost of channel estimation is also taken into account. Due to this training overhead, the increase in the cooperation cluster size does not necessarily correspond to a system throughput increase. As a matter of fact, our analysis shows that in most cases no cooperation among base stations (conventional cellular systems) with a significant number of antennas per base station (large $\gamma$) yields the best system




(a) $\tau = 1/64$




(b) $\tau = 1/32$

Fig. 6. Cell sum rate versus the cluster size $B$ for the antenna ratio $\gamma$=1, 2, 4, and 8 in the case of non-perfect CSIT and explicit downlink training with $\gamma_p = \gamma$. $M = 24$ cells and $K = 192$ user groups.

performance when the channel estimation cost is taken into account. This poses serious questions about whether "network MIMO" is a desirable solutions, also taking into account that base station cooperation yields a non-trivial complexity in system implementation, requiring some form of centralized processing of all the $B$ base stations in each cluster. It clearly appear that more effort should be devoted to using a larger number of antennas at each base station, since this yields larger system throughput and significantly less system complexity.

## APPENDIX A
## PROOF OF THEOREM 1

For the sake of clarity, we recall some definitions and facts about random matrices with independent non-identically distributed elements (see [52], [53]) of the type defined in (4), (7), that will be essential in the proof of Theorem 1.

*Definition 1:* Consider an $N_r \times N_c$ random matrix $\mathbf{H} = [\mathsf{H}_{i,j}]$, whose entries have variance

$$\mathrm{Var}[\mathsf{H}_{i,j}] = \frac{\mathsf{P}_{i,j}}{N_r} \tag{55}$$

such that $\mathbf{P} = [\mathsf{P}_{i,j}]$ is an $N_r \times N_c$ deterministic matrix with uniformly bounded entries. For given $N_r$, we define the *variance profile* of $\mathbf{H}$ as the function $v^{N_r} : [0,1) \times [0,1) \to \mathbb{R}$ such that

$$v^{N_r}(x,y) = \mathsf{P}_{i,j}, \qquad \tfrac{i-1}{N_r} \le x < \tfrac{i}{N_r}, \qquad \tfrac{j-1}{N_c} \le y < \tfrac{j}{N_c} \tag{56}$$

■

When we consider the limit for $N_r \to \infty$ with fixed ratio $\frac{N_c}{N_r} \to \nu$, we assume that $v^{N_r}(x,y)$ converges uniformly to a bounded measurable function $v(x,y)$, referred to as the *asymptotic variance profile* of $\mathbf{H}$. For random matrices distributed according to Definition 1, we have the following results.

*Theorem 5 ( [52, Theorem 2.52]):* Let $\mathbf{H}$ be an $N_r \times N_c$ random matrix whose entries are independent zero-mean complex circularly symmetric random variables satisfying the Lindeberg condition

$$\frac{1}{N_c} \sum_{i,j} \mathbb{E}\left[ |\mathsf{H}_{i,j}|^2 1\{|\mathsf{H}_{i,j}| \ge \delta\} \right] \to 0 \tag{57}$$

as $N_r, N_c \to \infty$ with $\frac{N_c}{N_r} \to \nu$, for all $\delta > 0$. Assume that the variances of the elements of $\mathbf{H}$ are given by Definition 1 and define the function

$$\mathcal{F}^{(N_r)}(y,s) = \mathbf{h}_j^{\mathsf{H}} \left( \mathbf{I} + s \sum_{\ell \ne j} \mathbf{h}_\ell \mathbf{h}_\ell^{\mathsf{H}} \right)^{-1} \mathbf{h}_j, \qquad \tfrac{j-1}{N_c} \le y < \tfrac{j}{N_c}. \tag{58}$$

As $N_r \to \infty$ with $\frac{N_c}{N_r} \to \nu$, $\mathcal{F}^{(N_r)}(y,s)$ converges almost surely to the limit $\mathcal{F}(y,s)$, given by the solution of the fixed-point equation

$$\mathcal{F}(y,s) = \mathbb{E}\left[\frac{v(\mathsf{X},y)}{1+s\,\nu\,\mathbb{E}\left[\frac{v(\mathsf{X},\mathsf{Y})}{1+s\,\mathcal{F}(\mathsf{Y},s)}\,|\mathsf{X}\right]}\right], \quad y \in [0,1]. \tag{59}$$

where $\mathsf{X}$ and $\mathsf{Y}$ are i.i.d. random variables uniformly distributed on $[0,1]$. ■

Defining the effective dimension ratio as

$$\nu' = \nu \frac{\mathbb{P}\left(\mathbb{E}[v(\mathsf{X},\mathsf{Y})|\mathsf{Y}] \neq 0\right)}{\mathbb{P}\left(\mathbb{E}[v(\mathsf{X},\mathsf{Y})|\mathsf{X}] \neq 0\right)},$$

the following high-SNR limit can be proved.

*Corollary 1 (see [52, Theorem 3.1]):* As $s$ goes to infinity, we have

$$\lim_{s\to\infty} \mathcal{F}(y,s) = \begin{cases} \Psi_\infty(y) & \text{if } \nu' < 1 \\ 0 & \text{if } \nu' \geq 1 \end{cases} \tag{60}$$

where, for $\nu' < 1$, $\Psi_\infty(y)$ is the positive solution to

$$\Psi_\infty(y) = \mathbb{E}\left[\frac{v(\mathsf{X},y)}{1+\nu\mathbb{E}\left[\frac{v(\mathsf{X},\mathsf{Y})}{\Psi_\infty(\mathsf{Y})}\middle|\mathsf{X}\right]}\right] \tag{61}$$

■

We now enter specifically the proof of Theorem 1. From the well-known formula for the inverse of a $2\times 2$ block matrix, we can write the $(j,j)$ diagonal element of the matrix $(\mathbf{I}+s\mathbf{H}^{\mathsf{H}}\mathbf{H})^{-1}$ as

$$\left[\left(\mathbf{I}+s\mathbf{H}^{\mathsf{H}}\mathbf{H}\right)^{-1}\right]_{j,j} = \frac{1}{1+s\mathbf{h}_j^{\mathsf{H}}\left(\mathbf{I}+s\sum_{\ell\neq j}\mathbf{h}_\ell\mathbf{h}_\ell^{\mathsf{H}}\right)^{-1}\mathbf{h}_j} \tag{62}$$

Furthermore, assuming that $\mathbf{H}$ has full rank, then

$$\begin{aligned}\left[\left(\mathbf{H}^{\mathsf{H}}\mathbf{H}\right)^{-1}\right]_{j,j} &= \lim_{s\to\infty} s\left[\left(\mathbf{I}+s\mathbf{H}^{\mathsf{H}}\mathbf{H}\right)^{-1}\right]_{j,j} \\ &= \lim_{s\to\infty}\frac{s}{1+s\mathbf{h}_j^{\mathsf{H}}\left(\mathbf{I}+s\sum_{\ell\neq j}\mathbf{h}_\ell\mathbf{h}_\ell^{\mathsf{H}}\right)^{-1}\mathbf{h}_j} \\ &= \frac{1}{\lim_{s\to\infty}\mathbf{h}_j^{\mathsf{H}}\left(\mathbf{I}+s\sum_{\ell\neq j}\mathbf{h}_\ell\mathbf{h}_\ell^{\mathsf{H}}\right)^{-1}\mathbf{h}_j}\end{aligned} \tag{63}$$

Comparing the definition of $\Lambda_k^{(i)}(\boldsymbol{\mu})$ in (11) with (63) and using Theorem 5 and Corollary 1, we have that the desired limiting value of $\Lambda_k^{(i)}(\boldsymbol{\mu})$ is given by $\Psi_\infty(y)$, evaluated at the corresponding value of $y$

such that $\frac{j-1}{N_c} \leq y < \frac{j}{N_c}$ for $j = \sum_{\ell=1}^{k-1} \mu_\ell N + i$, after replacing the general matrix $\mathbf{H}$ in Theorem 5 with $\mathbf{H}_{\boldsymbol{\mu}}$ given by our problem.

In this case, the number of rows in the matrix is given by $N_r = \gamma B N$ and the number of columns is given by $N_c = \mu N$. With the normalization by $1/\sqrt{N}$ of all the channel coefficients, the matrix $\mathbf{H}_{\boldsymbol{\mu}}$ defined in (7) is formed by independent blocks $\mathbf{H}_{m,k}(\mu_k)$ of dimension $\gamma N \times \mu_k N$, such that each block has i.i.d. $\mathcal{CN}(0, \beta_{m,k}^2/N)$ elements. As $N \to \infty$, we have that $N_c, N_r \to \infty$ with ratio $\nu = \frac{\mu}{\gamma B}$. By imposing the appropriate normalization, the asymptotic variance profile of $\mathbf{H}_{\boldsymbol{\mu}}$ is given by the piece-wise constant function

$$v(x,y) = \gamma B \beta_{m,k}^2 \;\text{ for }\; (x,y) \in \left[\frac{m-1}{B}, \frac{m}{B}\right) \times \left[\frac{\sum_{j=1}^{k-1} \mu_j}{\mu}, \frac{\sum_{j=1}^{k} \mu_j}{\mu}\right) \tag{64}$$

with $m = 1, \ldots, B$ and $k = 1, \ldots, A$. Also, we find explicitly

$$\nu' = \nu \, \frac{\sum_{k=1}^{A} \frac{\mu_k}{\mu} 1\left\{\frac{1}{B}\sum_{m=1}^{B} \beta_{m,k} \neq 0\right\}}{\frac{1}{B}\sum_{m=1}^{B} 1\left\{\frac{1}{\mu}\sum_{k=1}^{A} \mu_k \beta_{m,k} \neq 0\right\}} \tag{65}$$

and notice that the case $\nu' < 1$ in (60) always holds since, by construction, $\text{rank}(\mathbf{H}_{\boldsymbol{\mu}}) = \mu N$. Hence, the limit for $\Lambda_k^{(i)}(\boldsymbol{\mu})$ is obtained as the solution of the fixed point equation (61), for any $y \in \left[\frac{\sum_{j=1}^{k-1} \mu_j}{\mu}, \frac{\sum_{j=1}^{k} \mu_j}{\mu}\right)$. In fact, the piece-wise constant form of $v(x,y)$ yields that $\Lambda_k^{(i)}(\boldsymbol{\mu})$ converges to a limit that depends only on $k$ (the user group) and not on $i$ (the specific user in the group).

With some abuse of notation, we let $\Lambda_k(\boldsymbol{\mu}) = \Psi_\infty(y)$ for all $y \in \left[\frac{\sum_{j=1}^{k-1} \mu_j}{\mu}, \frac{\sum_{j=1}^{k} \mu_j}{\mu}\right)$, in order to denote this limit. Particularizing (61) to this case, we obtain

$$\Lambda_k(\boldsymbol{\mu}) = \gamma \sum_{m=1}^{B} \frac{\beta_{m,k}^2}{1 + \sum_{q=1}^{A} \mu_q \, \frac{\beta_{m,q}^2}{\Lambda_q(\boldsymbol{\mu})}}, \qquad k = 1, \ldots, A \tag{66}$$

It follows that the asymptotic limit of $\boldsymbol{\Lambda}_{\boldsymbol{\mu}}$ is block-diagonal, with scaled-identity diagonal blocks, where the $k$-th block is given by $\Lambda_k(\boldsymbol{\mu}) \mathbf{I}_{\mu_k N}$.

In order to obtain the more convenient expression (23), we introduce the variables $\eta_m \in [0,1]$, for $m = 1, \ldots, B$, and replace $\Lambda_k(\boldsymbol{\mu}) = \gamma \sum_{m=1}^{B} \beta_{m,k}^2 \eta_m$ into (66). Since $\eta_m$ takes values in $[0,1]$, we can write $\eta_m = 1/(1+z_m)$ for $z_m \geq 0$, and solving for $z_m$, we obtain $z_m = \sum_{q=1}^{A} \mu_q \frac{\beta_{m,q}^2}{\Lambda_q(\boldsymbol{\mu})}$. Eliminating the variables $z_m$ from the latter equation, we arrive at the desired fixed point equation (24), as given in Theorem 1.

As a final remark, notice that (24) has some significant advantages with respect to (66). In particular, the variables $\eta_m$ take values in $[0,1]$ (by construction), and typically we have $B < A$ (less BSs in a cluster than user groups). Therefore, (24) can be always initialized by letting $\eta_m = 1$, and the fixed point

equation iterative solution involves only $B$, rather than $A$, variables. Also, it is immediately evident by inspection that the solution of (24) for $\eta_m \in [0,1]$ always exists and it is unique.

## APPENDIX B
## PROOF OF THEOREM 2

We start with the following auxiliary result:

*Lemma 1:* Let $\mathbf{x}$ be a $n$-dimensional vector with i.i.d. entries with variance $\frac{1}{n}$. Let $\mathbf{A}$ and $\mathbf{C}$ be $n \times n$ Hermitian symmetric matrices independent on $\mathbf{x}$. Finally let $\mathbf{D}$ be a $n \times n$ diagonal matrix independent on $\mathbf{x}$. Then:

$$\mathbf{x}^{\mathsf{H}}\mathbf{D}^{\mathsf{H}}(\mathbf{D}\mathbf{x}\mathbf{x}^{\mathsf{H}}\mathbf{D}^{\mathsf{H}}+\mathbf{A})^{-1}\mathbf{C}(\mathbf{D}\mathbf{x}\mathbf{x}^{\mathsf{H}}\mathbf{D}^{\mathsf{H}}+\mathbf{A})^{-1}\mathbf{D}\mathbf{x} \rightarrow \frac{\phi(\mathbf{D}^{\mathsf{H}}\mathbf{A}^{-1}\mathbf{C}\mathbf{A}^{-1}\mathbf{D})}{(1+\phi(\mathbf{D}^{\mathsf{H}}\mathbf{A}^{-1}\mathbf{D}))^2}$$

where $\phi(\cdot) = \lim_{n\rightarrow\infty}\frac{1}{n}\mathrm{tr}(\cdot)$ and the convergence is almost surely.

*Proof:* Let

$$Q = \mathbf{x}^{\mathsf{H}}\mathbf{D}^{\mathsf{H}}(\mathbf{D}\mathbf{x}\mathbf{x}^{\mathsf{H}}\mathbf{D}^{\mathsf{H}}+\mathbf{A})^{-1}\mathbf{C}(\mathbf{D}\mathbf{x}\mathbf{x}^{\mathsf{H}}\mathbf{D}^{\mathsf{H}}+\mathbf{A})^{-1}\mathbf{D}\mathbf{x}$$

From the inversion lemma we have that

$$\left(\mathbf{D}\mathbf{x}\mathbf{x}^{\mathsf{H}}\mathbf{D}^{\mathsf{H}}+\mathbf{A}\right)^{-1} = \mathbf{A}^{-1} - \frac{1}{1+\mathbf{x}^{\mathsf{H}}\mathbf{D}^{\mathsf{H}}\mathbf{A}^{-1}\mathbf{D}\mathbf{x}}\mathbf{A}^{-1}\mathbf{D}\mathbf{x}\mathbf{x}^{\mathsf{H}}\mathbf{D}^{\mathsf{H}}\mathbf{A}^{-1} \tag{67}$$

Hence,

$$\begin{aligned} Q &= \mathbf{x}^{\mathsf{H}}\mathbf{D}^{\mathsf{H}}\mathbf{A}^{-1}\mathbf{C}(\mathbf{D}\mathbf{x}\mathbf{x}^{\mathsf{H}}\mathbf{D}^{\mathsf{H}}+\mathbf{A})^{-1}\mathbf{D}\mathbf{x} \\ &\quad -\frac{1}{1+\mathbf{x}^{\mathsf{H}}\mathbf{D}^{\mathsf{H}}\mathbf{A}^{-1}\mathbf{D}\mathbf{x}}\mathbf{x}^{\mathsf{H}}\mathbf{D}^{\mathsf{H}}\mathbf{A}^{-1}\mathbf{D}\mathbf{x}\mathbf{x}^{\mathsf{H}}\mathbf{D}^{\mathsf{H}}\mathbf{A}^{-1}\mathbf{C}(\mathbf{D}\mathbf{x}\mathbf{x}^{\dagger}\mathbf{D}^{\dagger}+\mathbf{A})^{-1}\mathbf{D}\mathbf{x} \\ &= \frac{\mathsf{a}}{1+\mathbf{x}^{\mathsf{H}}\mathbf{D}^{\mathsf{H}}\mathbf{A}^{-1}\mathbf{D}\mathbf{x}} \end{aligned} \tag{68}$$

where

$$\mathsf{a} = \mathbf{x}^{\mathsf{H}}\mathbf{D}^{\mathsf{H}}\mathbf{A}^{-1}\mathbf{C}(\mathbf{D}\mathbf{x}\mathbf{x}^{\mathsf{H}}\mathbf{D}^{\mathsf{H}}+\mathbf{A})^{-1}\mathbf{D}\mathbf{x}$$

Applying again the inversion lemma we have:

$$\begin{aligned} \mathsf{a} &= \mathbf{x}^{\mathsf{H}}\mathbf{D}^{\mathsf{H}}\mathbf{A}^{-1}\mathbf{C}\mathbf{A}^{-1}\mathbf{D}\mathbf{x} \\ &\quad -\frac{1}{1+\mathbf{x}^{\mathsf{H}}\mathbf{D}^{\mathsf{H}}\mathbf{A}^{-1}\mathbf{D}\mathbf{x}}\mathbf{x}^{\mathsf{H}}\mathbf{D}^{\mathsf{H}}\mathbf{A}^{-1}\mathbf{C}\mathbf{A}^{-1}\mathbf{D}\mathbf{x}\mathbf{x}^{\mathsf{H}}\mathbf{D}^{\mathsf{H}}\mathbf{A}^{-1}\mathbf{D}\mathbf{x} \\ &= \frac{\mathsf{b}}{1+\mathbf{x}^{\mathsf{H}}\mathbf{D}^{\mathsf{H}}\mathbf{A}^{-1}\mathbf{D}\mathbf{x}} \end{aligned} \tag{69}$$

where

$$\mathsf{b} = \mathbf{x}^{\mathsf{H}}\mathbf{D}^{\mathsf{H}}\mathbf{A}^{-1}\mathbf{C}\mathbf{A}^{-1}\mathbf{D}\mathbf{x}$$

From (68) and (69) we obtain

$$Q = \frac{\mathbf{x}^{\mathsf{H}}\mathbf{D}^{\mathsf{H}}\mathbf{A}^{-1}\mathbf{C}\mathbf{A}^{-1}\mathbf{D}\mathbf{x}}{\left(1+\mathbf{x}^{\mathsf{H}}\mathbf{D}^{\mathsf{H}}\mathbf{A}^{-1}\mathbf{D}\mathbf{x}\right)^2} \tag{70}$$

Finally, we arrive at the desired result by using the well-known result in random matrix according to which $\lim_{n\to\infty}\mathbf{x}^{\mathsf{H}}\mathbf{M}\mathbf{x}=\phi(\mathbf{M})$ provided that $\mathbf{x},\mathbf{M}$ are independent, that $\mathbf{M}$ has a well-defined limiting eigenvalue distribution and that $\mathbf{x}$ has i.i.d. elements with mean zero and variance $1/n$. ■

Using Lemma 1, we can proceed with the proof. From the expression of $\theta_{m,k}(\boldsymbol{\mu})$, it follows that

$$\begin{aligned}\theta_{m,k}(\boldsymbol{\mu}) &= \frac{1}{N}\sum_{i=1}^{\mu_k N}\sum_{\ell=1+(m-1)\gamma N}^{m\gamma N}\left|\left[\mathbf{V}_{\boldsymbol{\mu}}\right]_{\ell,k}^{(i)}\right|^2 \\ &= \frac{1}{N}\mathrm{tr}\left(\boldsymbol{\Phi}_m\mathbf{V}_{\boldsymbol{\mu}}\boldsymbol{\Theta}_k\mathbf{V}_{\boldsymbol{\mu}}^{\mathsf{H}}\right) \\ &= \frac{1}{N}\mathrm{tr}\left(\boldsymbol{\Phi}_m\mathbf{H}_{\boldsymbol{\mu}}(\mathbf{H}_{\boldsymbol{\mu}}^{\mathsf{H}}\mathbf{H}_{\boldsymbol{\mu}})^{-1}\boldsymbol{\Lambda}_{\boldsymbol{\mu}}^{1/2}\boldsymbol{\Theta}_k\boldsymbol{\Lambda}_{\boldsymbol{\mu}}^{1/2}(\mathbf{H}_{\boldsymbol{\mu}}^{\mathsf{H}}\mathbf{H}_{\boldsymbol{\mu}})^{-1}\mathbf{H}_{\boldsymbol{\mu}}^{\mathsf{H}}\boldsymbol{\Phi}_m\right)\end{aligned} \tag{71}$$

where $\boldsymbol{\Phi}_m$ is a diagonal matrix with all zeros, but for $\gamma N$ consecutive ones, corresponding to positions from $(m-1)\gamma N+1$ to $m\gamma N$ on the main diagonal, and where $\boldsymbol{\Theta}_k$ denotes the $\mu N$-dimensional diagonal matrix with all zeros, but for $\mu_k N$ consecutive ones, corresponding to positions from $\mu_{1:k-1}N+1$ to $\mu_{1:k}N$ on the main diagonal (recall that we define the partial sum $\mu_{1:k}=\sum_{j=1}^{k}\mu_j$).

The submatrix of $\boldsymbol{\Phi}_m\mathbf{H}_{\boldsymbol{\mu}}$ corresponding to the non-zero rows, i.e., including rows from $(m-1)\gamma N+1$ to $m\gamma N$, can be written as

$$[\beta_{m,1}\mathbf{H}_{m,1}(\mu_1),\cdots,\beta_{m,A}\mathbf{H}_{m,A}(\mu_A)]=\mathbf{W}_m\mathbf{B}_m$$

where $\mathbf{W}_m$ is a $\gamma N\times\mu N$ rectangular matrix with i.i.d. entries, with mean 0 and variance $1/N$, and

$$\mathbf{B}_m=\mathrm{diag}\left(\underbrace{\beta_{m,1},\ldots,\beta_{m,1}}_{\mu_1 N},\ldots,\underbrace{\beta_{m,k},\ldots,\beta_{m,k}}_{\mu_k N},\ldots,\underbrace{\beta_{m,A},\ldots,\beta_{m,A}}_{\mu_A N}\right) \tag{72}$$

Also, we let

$$\begin{aligned}\mathbf{C}_k &= \boldsymbol{\Lambda}_{\boldsymbol{\mu}}^{1/2}\boldsymbol{\Theta}_k\boldsymbol{\Lambda}_{\boldsymbol{\mu}}^{1/2} \\ &= \mathrm{diag}\left(\underbrace{0,\ldots,0}_{\mu_{1:k-1}N},\underbrace{\Lambda_k^{(1)}(\boldsymbol{\mu}),\ldots,\Lambda_k^{(\mu_k N)}(\boldsymbol{\mu})}_{\mu_k N},\underbrace{0,\ldots,0}_{(\mu-\mu_{1:k})N}\right)\end{aligned} \tag{73}$$

Notice that $\mathbf{B}_m$ and $\mathbf{C}_k$ have both dimension $\mu N\times\mu N$.

After simple algebraic manipulation, letting the $\ell$-th row of $\mathbf{W}_m$ be denoted by $\mathbf{w}_{m,\ell}^{\mathsf{H}}$ we can write

$$\begin{aligned}\mathbf{H}_{\boldsymbol{\mu}}^{\mathsf{H}}\mathbf{H}_{\boldsymbol{\mu}} &= \sum_{m=1}^{B}\mathbf{B}_m\mathbf{W}_m^{\mathsf{H}}\mathbf{W}_m\mathbf{B}_m \\ &= \mathbf{B}_m\mathbf{w}_{m,\ell}\mathbf{w}_{m,\ell}^{\mathsf{H}}\mathbf{B}_m + \sum_{j\neq\ell}\mathbf{B}_m\mathbf{w}_{m,j}\mathbf{w}_{m,j}^{\mathsf{H}}\mathbf{B}_m \\ &\quad + \sum_{q\neq m}\mathbf{B}_q\mathbf{W}_q^{\mathsf{H}}\mathbf{W}_q\mathbf{B}_q \end{aligned} \tag{74}$$

In order to be able to apply Lemma (1), we need that the variance of the elements of the i.i.d. vector $\mathbf{w}_{m,\ell}$ (playing the role of $\mathbf{x}$ in the Lemma), is equal to the inverse of the vector length. Therefore, dividing by $\mu$, we define

$$\mathbf{A} = \frac{1}{\mu}\sum_{q=1}^{B}\mathbf{B}_q\mathbf{W}_q^{\mathsf{H}}\mathbf{W}_q\mathbf{B}_q \tag{75}$$

and

$$\mathbf{A}_{m,\ell} = \mathbf{A} - \frac{1}{\mu}\mathbf{B}_m\mathbf{w}_{m,\ell}\mathbf{w}_{m,\ell}^{\mathsf{H}}\mathbf{B}_m \tag{76}$$

Using (72), (73), (75) and (76) in (71) we arrive at

$$\begin{aligned}\theta_{m,k}(\boldsymbol{\mu}) &= \frac{1}{N\mu}\mathrm{tr}\left(\frac{1}{\sqrt{\mu}}\mathbf{W}_m\mathbf{B}_m\mathbf{A}^{-1}\mathbf{C}_k\mathbf{A}^{-1}\mathbf{B}_m\mathbf{W}_m^{\mathsf{H}}\frac{1}{\sqrt{\mu}}\right) \\ &= \frac{1}{N\mu}\sum_{\ell=1}^{\gamma N}\frac{1}{\sqrt{\mu}}\mathbf{w}_{m,\ell}^{\mathsf{H}}\mathbf{B}_m\left(\frac{1}{\mu}\mathbf{B}_m\mathbf{w}_{m,\ell}\mathbf{w}_{m,\ell}^{\mathsf{H}}\mathbf{B}_m + \mathbf{A}_{m,\ell}\right)^{-1}\mathbf{C}_k \\ &\qquad \cdot\left(\frac{1}{\mu}\mathbf{B}_m\mathbf{w}_{m,\ell}\mathbf{w}_{m,\ell}^{\mathsf{H}}\mathbf{B}_m + \mathbf{A}_{m,\ell}\right)^{-1}\mathbf{B}_m\mathbf{w}_{m,\ell}\frac{1}{\sqrt{\mu}} \qquad (77) \\ &\to \frac{\gamma}{\mu}\frac{\phi\left(\mathbf{B}_m\mathbf{A}^{-1}\mathbf{C}_k\mathbf{A}^{-1}\mathbf{B}_m\right)}{\left(1+\phi\left(\mathbf{B}_m\mathbf{A}^{-1}\mathbf{B}_m\right)\right)^2} \qquad (78)\end{aligned}$$

where the last line follows by applying Lemma 1, and by noticing that the terms for different $\ell$ in the sum in (77) converge to the same limit (by symmetry), that can be obtained by using $\mathbf{A}$ *in lieu* of $\mathbf{A}_{m,\ell}$, while assuming $\mathbf{A}$ and $\mathbf{w}_{m,\ell}$ to be independent, for all $\ell = 1,\ldots,\gamma N$ and $m = 1,\ldots,B$.

At this point, our goal is to evaluate the two limit normalized traces in (78). We start by the term in the denominator , which is considerably simpler. We have

$$\begin{aligned}
\phi\left(\mathbf{B}_m\mathbf{A}^{-1}\mathbf{B}_m\right) &= \lim_{N\to\infty}\frac{1}{\mu N}\text{tr}\left(\mathbf{B}_m\mathbf{A}^{-1}\mathbf{B}_m\right) \\
&= \lim_{N\to\infty}\frac{1}{\mu N}\text{tr}\left(\left(\frac{1}{\mu}\mathbf{H}_{\boldsymbol{\mu}}^{\mathsf{H}}\mathbf{H}_{\boldsymbol{\mu}}\right)^{-1}\mathbf{B}_m^2\right) \\
&= \lim_{N\to\infty}\frac{1}{N}\text{tr}\left(\left(\mathbf{H}_{\boldsymbol{\mu}}^{\mathsf{H}}\mathbf{H}_{\boldsymbol{\mu}}\right)^{-1}\mathbf{B}_m^2\right) \\
&= \lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^{A}\sum_{i=1}^{\mu_k N}\frac{\beta_{m,k}^2}{\Lambda_k^{(i)}(\boldsymbol{\mu})} \\
&= \sum_{k=1}^{A}\frac{\mu_k\beta_{m,k}^2}{\Lambda_k(\boldsymbol{\mu})}
\end{aligned} \tag{79}$$

where we used the fact that, by definition,

$$\left[\left(\mathbf{H}_{\boldsymbol{\mu}}^{\mathsf{H}}\mathbf{H}_{\boldsymbol{\mu}}\right)^{-1}\right]_k^{(i)} = \frac{1}{\Lambda_k^{(i)}(\boldsymbol{\mu})}$$

for the diagonal elements of $\left(\mathbf{H}_{\boldsymbol{\mu}}^{\mathsf{H}}\mathbf{H}_{\boldsymbol{\mu}}\right)^{-1}$ in position $\mu_{1:k-1}N+i$ for $i=1,\ldots\mu_k N$, and the convergence result of Theorem 1. Also, comparing (79) with the expression of $z_m$ in the proof of Theorem 1 (see Appendix A, eq. (66) and below), we have that

$$z_m = \sum_{k=1}^{A}\frac{\mu_k\beta_{m,k}^2}{\Lambda_k(\boldsymbol{\mu})} \tag{80}$$

Since $\eta_m(\boldsymbol{\mu}) = 1/(1+z_m)$, where $\{\eta_m(\boldsymbol{\mu}) : m=1,\ldots,B\}$ are the auxiliary variables defined in Theorem 1 as the solutions of the fixed-point equation (24), we have that the denominator of (78) is given by

$$\left(1+\phi\left(\mathbf{B}_m\mathbf{A}^{-1}\mathbf{B}_m\right)\right)^2 = \eta_m^{-2}(\boldsymbol{\mu}) \tag{81}$$

Next, we consider the numerator of (78). For this purpose, let $\rho$ be a dummy non-negative real variable and consider the identity:

$$\frac{-d}{d\rho}\text{tr}\left(\left(\rho\mathbf{B}_m^2+\mathbf{A}\right)^{-1}\mathbf{C}_k\right) = \text{tr}\left(\mathbf{B}_m(\rho\mathbf{B}_m^2+\mathbf{A})^{-1}\mathbf{C}_k(\rho\mathbf{B}_m^2+\mathbf{A})^{-1}\mathbf{B}_m\right) \tag{82}$$

By almost-sure continuity of the trace in the left-hand side of (82) with respect to $\rho\geq 0$, it follows that the desired expression for the numerator of (78) can be calculated as

$$\phi\left(\mathbf{B}_m\mathbf{A}^{-1}\mathbf{C}_k\mathbf{A}^{-1}\mathbf{B}_m\right) = \lim_{\rho\downarrow 0}\frac{-d}{d\rho}\phi\left(\left(\rho\mathbf{B}_m^2+\mathbf{A}\right)^{-1}\mathbf{C}_k\right) \tag{83}$$

In order to compute the asymptotic normalized trace in (83), we use [52, Lemma 2.51], reported here for completeness.

*Lemma 2:* Let $\mathbf{H}$ be $N_r \times N_c$ of the type given in Definition 1, satisfying the same assumptions of Theorem 5. For any $a, b \in [0,1]$ with $a < b$,

$$\frac{1}{N_r} \sum_{i=\lfloor aN_r \rfloor}^{\lfloor bN_r \rfloor} \left[ \left( s\mathbf{H}\mathbf{H}^{\mathsf{H}} + \mathbf{I} \right)^{-1} \right]_{i,i} \to \int_a^b \Gamma_{\mathbf{H}\mathbf{H}^{\mathsf{H}}}(x,s)\, dx \tag{84}$$

where $N_c/N_r \to \nu$ and where $\Gamma_{\mathbf{H}\mathbf{H}^{\mathsf{H}}}(x,s)$ and $\Upsilon_{\mathbf{H}\mathbf{H}^{\mathsf{H}}}(y,s)$ are functions defined implicitly by the fixed-point equation

$$\begin{aligned} \Gamma_{\mathbf{H}\mathbf{H}^{\mathsf{H}}}(x,s) &= \frac{1}{1+\nu s \mathbb{E}\left[ v(x,\mathsf{Y}), \Upsilon_{\mathbf{H}\mathbf{H}^{\mathsf{H}}}(\mathsf{Y},s) \right]} \\ \Upsilon_{\mathbf{H}\mathbf{H}^{\mathsf{H}}}(y,s) &= \frac{1}{1+s \mathbb{E}\left[ v(\mathsf{X},y), \Gamma_{\mathbf{H}\mathbf{H}^{\mathsf{H}}}(\mathsf{X},s) \right]} \end{aligned} \tag{85}$$

for $(x,y) \in [0,1] \times [0,1]$, where $\mathsf{X}$ and $\mathsf{Y}$ are i.i.d. uniform-$[0,1]$ RVs and where the variance profile function $v(x,y)$ was introduced in Definition 1. ■

In order to use Lemma 2 we write

$$\begin{aligned} \mathrm{tr}\left( \left( \rho \mathbf{B}_m^2 + \mathbf{A} \right)^{-1} \mathbf{C}_k \right) &= \mathrm{tr}\left( \left( \rho \mathbf{I} + \mathbf{B}_m^{-1} \mathbf{A} \mathbf{B}_m^{-1} \right)^{-1} \mathbf{B}_m^{-1} \mathbf{C}_k \mathbf{B}_m^{-1} \right) \\ &= \frac{1}{\rho} \mathrm{tr}\left( \left( \mathbf{I} + \frac{1}{\rho} \mathbf{B}_m^{-1} \mathbf{A} \mathbf{B}_m^{-1} \right)^{-1} \mathbf{B}_m^{-1} \mathbf{C}_k \mathbf{B}_m^{-1} \right) \end{aligned} \tag{86}$$

Noticing that, by definition, $\mathbf{A} = \frac{1}{\mu} \mathbf{H}_{\boldsymbol{\mu}}^{\mathsf{H}} \mathbf{H}_{\boldsymbol{\mu}}$, we can identify the matrix $\frac{1}{\sqrt{\mu}} \mathbf{B}_m^{-1} \mathbf{H}_{\boldsymbol{\mu}}^{\mathsf{H}}$ with the matrix $\mathbf{H}$ of Lemma 2. In this case, $N_r = \mu N$ and $N_c = \gamma B N$. Using $\{\mathbf{B}_m\}$ and $\{\mathbf{W}_m\}$ defined before, we can write the block-matrix form

$$\mathbf{H}_{\mu}^{\mathsf{H}} = \left[ \mathbf{B}_1 \mathbf{W}_1^{\mathsf{H}}, \mathbf{B}_2 \mathbf{W}_2^{\mathsf{H}}, \ldots, \mathbf{B}_B \mathbf{W}_B^{\mathsf{H}} \right]$$

so that

$$\mathbf{B}_m^{-1} \mathbf{H}_{\mu}^{\mathsf{H}} = \left[ \mathbf{B}_m^{-1} \mathbf{B}_1 \mathbf{W}_1^{\mathsf{H}}, \mathbf{B}_m^{-1} \mathbf{B}_2 \mathbf{W}_2^{\mathsf{H}}, \ldots, \mathbf{B}_m^{-1} \mathbf{B}_B \mathbf{W}_B^{\mathsf{H}} \right]$$

It follows that the variance profile function of $\frac{1}{\sqrt{\mu}} \mathbf{B}_m^{-1} \mathbf{H}_{\boldsymbol{\mu}}^{\mathsf{H}}$ is given by

$$v_m(x,y) = \frac{\beta_{\ell,k}^2}{\beta_{m,k}^2}, \quad \text{for } (x,y) \in \left[ \frac{\mu_{1:k-1}}{\mu}, \frac{\mu_{1:k}}{\mu} \right) \times \left[ \frac{\ell-1}{B}, \frac{\ell}{B} \right) \tag{87}$$

Using this in Lemma 2 and letting $1/\rho = s$, we find

$$\frac{1}{\mu N} \sum_{i=\mu_{1:k-1}N+1}^{\mu_{1:k}N} \left[ \left( \mathbf{I} + s \mathbf{B}_m^{-1} \mathbf{A} \mathbf{B}_m^{-1} \right)^{-1} \right]_{i,i} \to \int_{\mu_{1:k-1}/\mu}^{\mu_{1:k}/\mu} \Gamma_m(x,s)\, dx \tag{88}$$

where $\Gamma_m(x,s)$ and $\Upsilon_m(y,s)$ are defined by

$$\begin{aligned}\Gamma_m(x,s) &= \frac{1}{1+\frac{\gamma B s}{\mu}\mathbb{E}\left[v_m(x,\mathsf{Y}),\Upsilon_m(\mathsf{Y},s)\right]}\\ \Upsilon_m(y,s) &= \frac{1}{1+s\mathbb{E}\left[v_m(\mathsf{X},y),\Gamma_m(\mathsf{X},s)\right]}\end{aligned} \tag{89}$$

Noticing that $v_m(x,y)$ is piecewise constant (see (87), we have that also the functions $\Gamma_m(x,s)$ and $\Upsilon_m(y,s)$ are piecewise constant. With some abuse of notation, we denote the values of these functions as $\{\Gamma_{m,q}(s), q=1,\ldots,A\}$ and $\{\Upsilon_{m,\ell}(s), \ell=1,\ldots,B\}$, respectively, we find that (89) can be re-written directly in terms of these values as

$$\begin{aligned}\Gamma_{m,q}(s) &= \frac{1}{1+\frac{s}{\mu}\sum_{\ell=1}^{B}\frac{\gamma\beta_{\ell,q}^2}{\beta_{m,q}^2}\Upsilon_{m,\ell}(s)}, \text{ for } q=1,\ldots,A\\ \Upsilon_{m,\ell}(s) &= \frac{1}{1+\frac{s}{\mu}\sum_{q=1}^{A}\frac{\mu_q\beta_{\ell,q}^2}{\beta_{m,q}^2}\Gamma_{m,q}(s)} \text{ for } \ell=1,\ldots,B\end{aligned} \tag{90}$$

Finally, using (88) and (86) and noticing that the non-zero diagonal elements of $\mathbf{B}_m^{-1}\mathbf{C}_k\mathbf{B}_m^{-1}$ converge to the constant $\Lambda_k(\boldsymbol{\mu})\beta_{m,k}^{-2}$, we arrive at:

$$\phi\left(\left(\rho\mathbf{B}_m^2+\mathbf{A}\right)^{-1}\mathbf{C}_k\right) = \frac{\mu_k}{\rho\mu}\Gamma_{m,k}(1/\rho)\Lambda_k(\boldsymbol{\mu})\beta_{m,k}^{-2} \tag{91}$$

It turns out that it is convenient to define the new variables

$$S_{m,q}(\rho) = \frac{1}{\rho\beta_{m,q}^2}\Gamma_{m,q}(1/\rho), \quad \text{and} \quad G_{m,\ell}(\rho) = \Upsilon_{m,\ell}(1/\rho)$$

Therefore, we can rewrite (90) and (91) as

$$S_{m,q}(\rho) = \frac{1}{\rho\beta_{m,q}^2+\frac{\gamma}{\mu}\sum_{\ell=1}^{B}\beta_{\ell,q}^2 G_{m,\ell}(\rho)}, \text{ for } q=1,\ldots,A \tag{92}$$

$$G_{m,\ell}(\rho) = \frac{1}{1+\frac{1}{\mu}\sum_{q=1}^{A}\mu_q\beta_{\ell,q}^2 S_{m,q}(\rho)}, \text{ for } \ell=1,\ldots,B \tag{93}$$

$$\phi\left(\left(\rho\mathbf{B}_m^2+\mathbf{A}\right)^{-1}\mathbf{C}_k\right) = \frac{\mu_k}{\mu}\Lambda_k(\boldsymbol{\mu})S_{m,k}(\rho) \tag{94}$$

Taking the derivative in (94), we have that the numerator of (78) can be obtained as:

$$\begin{aligned}\lim_{\rho\downarrow 0}\frac{-d}{d\rho}\phi\left(\left(\rho\mathbf{B}_m^2+\mathbf{A}\right)^{-1}\mathbf{C}_k\right) &= \frac{\mu_k}{\mu}\Lambda_k(\boldsymbol{\mu})\lim_{\rho\downarrow 0}\frac{-d}{d\rho}S_{m,k}(\rho)\\ &= \frac{\mu_k}{\mu}\Lambda_k(\boldsymbol{\mu})\dot{S}_{m,k}(0)\end{aligned} \tag{95}$$

where we define $\dot{S}_{m,k}(0) = \frac{-d}{d\rho}S_{m,k}(\rho)|_{\rho=0}$ and, for later use, $\dot{G}_{m,\ell}(0) = \frac{d}{d\rho}G_{m,\ell}(\rho)|_{\rho=0}$.

Next, we wish to find a fixed-point equation that yields directly $\dot{S}_{m,k}(0)$. By continuity, we can replace directly $\rho = 0$ into the fixed point equations after taking the derivatives. By doing so, from (92) and (93), we obtain:

$$\dot{S}_{m,q}(0) = \frac{\beta_{m,q}^2 + \frac{\gamma}{\mu}\sum_{\ell=1}^{B}\beta_{\ell,q}^2\dot{G}_{m,\ell}(0)}{\left(\frac{\gamma}{\mu}\sum_{\ell=1}^{B}\beta_{\ell,q}^2 G_{m,\ell}(0)\right)^2}, \text{ for } q = 1,\ldots,A \tag{96}$$

$$\dot{G}_{m,\ell}(0) = \frac{\frac{1}{\mu}\sum_{q=1}^{A}\mu_q\beta_{\ell,q}^2\dot{S}_{m,q}(0)}{\left(1+\frac{1}{\mu}\sum_{q=1}^{A}\mu_q\beta_{\ell,q}^2 S_{m,q}(0)\right)^2}, \text{ for } \ell = 1,\ldots,B \tag{97}$$

Also, the equations for $S_{m,q}(0)$ and $G_{m,\ell}(0)$, obtained by replacing $\rho = 0$ in (92), (93), read:

$$S_{m,q}(0) = \frac{1}{\frac{\gamma}{\mu}\sum_{\ell=1}^{B}\beta_{\ell,q}^2 G_{m,\ell}(0)}, \text{ for } q = 1,\ldots,A \tag{98}$$

$$G_{m,\ell}(0) = \frac{1}{1+\frac{1}{\mu}\sum_{q=1}^{A}\mu_q\beta_{\ell,q}^2 S_{m,q}(0)}, \text{ for } \ell = 1,\ldots,B \tag{99}$$

Replacing (98) into (99), we obtain, for all $\ell = 1,\ldots,B$,

$$G_{m,\ell}(0) = \frac{1}{1+\sum_{q'=1}^{A}\frac{\mu_{q'}\beta_{\ell,q'}^2}{\gamma\sum_{\ell'=1}^{B}\beta_{\ell',q'}^2 G_{m,\ell'}(0)}}. \tag{100}$$

By multiplying both sides by $\gamma\beta_{\ell,q}^2$ and summing over $\ell$, we find

$$U_{m,q} = \gamma\sum_{\ell=1}^{B}\frac{\beta_{\ell,q}^2}{1+\sum_{q'=1}^{A}\frac{\mu_{q'}\beta_{\ell,q'}^2}{U_{m,q'}}}, \tag{101}$$

where we define $U_{m,q} = \gamma\sum_{\ell=1}^{B}\beta_{\ell,q}^2 G_{m,\ell}(0)$. Comparing the fixed point equation (101) with (66), we discover that $U_{m,q} = \Lambda_q(\boldsymbol{\mu})$, independent of $m$. Using this result in (98), we obtain

$$S_{m,q}(0) = \frac{\mu}{\Lambda_q(\boldsymbol{\mu})} \tag{102}$$

Using the definition of $U_{m,q}$, (96) can be written as,

$$\dot{S}_{m,q}(0) = \frac{\mu^2\beta_{m,q}^2 + \mu\dot{U}_{m,q}}{\Lambda_q^2(\boldsymbol{\mu})}, \tag{103}$$

where, with some abuse of notation, we define $\dot{U}_{m,q} = \gamma\sum_{\ell=1}^{B}\beta_{\ell,q}^2\dot{G}_{m,\ell}(0)$.

Multiplying both sides of (97) by $\gamma\beta_{\ell,q}^2$, using (103) and (102) and summing over $\ell$, we obtain

$$\begin{aligned}
\dot{U}_{m,q} &= \gamma \sum_{\ell=1}^{B} \beta_{\ell,q}^2 \frac{\frac{1}{\mu}\sum_{q'=1}^{A} \mu_{q'} \beta_{\ell,q'}^2 \dot{S}_{m,q'}(0)}{\left(1 + \frac{1}{\mu}\sum_{q'=1}^{A} \mu_{q'} \beta_{\ell,q'}^2 S_{m,q'}(0)\right)^2} \\
&= \gamma \sum_{\ell=1}^{B} \beta_{\ell,q}^2 \frac{\frac{1}{\mu}\sum_{q'=1}^{A} \mu_{q'} \beta_{\ell,q'}^2 \frac{\mu^2 \beta_{m,q'}^2 + \mu \dot{U}_{m,q'}}{\Lambda_{q'}^2(\boldsymbol{\mu})}}{\left(1 + \sum_{q'=1}^{A} \frac{\mu_{q'} \beta_{\ell,q'}^2}{\Lambda_{q'}(\boldsymbol{\mu})}\right)^2} \qquad (104) \\
&= \gamma\mu \sum_{q'=1}^{A} \left[\sum_{\ell=1}^{B} \eta_\ell^2(\boldsymbol{\mu}) \beta_{\ell,q}^2 \beta_{\ell,q'}^2\right] \frac{\mu_{q'}}{\Lambda_{q'}^2(\boldsymbol{\mu})} \left(\beta_{m,q'}^2 + \frac{1}{\mu}\dot{U}_{m,q'}(\boldsymbol{\mu})\right) \qquad (105)
\end{aligned}$$

where we have used again the identity (80) in the denominator of (104). Somehow surprisingly, we notice that (105) is a system of $A$ linear equations in the $A$ unknown $\{\dot{U}_{m,q} : q = 1, \ldots, A\}$. Therefore, this can be solved explicitly (although not in closed form in general). In particular, we define the $A \times A$ matrix

$$\mathbf{M} = \left[\sum_{\ell=1}^{B} \eta_\ell^2(\boldsymbol{\mu}) \mathbf{b}_\ell \mathbf{b}_\ell^{\mathsf{T}}\right] \text{diag}\left(\frac{\mu_1}{\Lambda_1^2(\boldsymbol{\mu})}, \ldots, \frac{\mu_A}{\Lambda_A^2(\boldsymbol{\mu})}\right) \qquad (106)$$

where $\mathbf{b}_\ell = (\beta_{\ell,1}^2, \ldots, \beta_{\ell,A}^2)^{\mathsf{T}}$, and the vector of unknowns $\dot{\mathbf{U}}_m$, then, we the linear system corresponding to (105) is given by

$$[\mathbf{I} - \gamma\mathbf{M}]\, \dot{\mathbf{U}}_m = \gamma\mu \mathbf{M}\mathbf{b}_m \qquad (107)$$

Solving the system (107) and using (103) in (95), we obtain the sought numerator of (78) in the form

$$\frac{\mu_k}{\mu}\Lambda_k(\boldsymbol{\mu})\dot{S}_{m,k}(0) = \mu_k \frac{\mu\beta_{m,k}^2 + \dot{U}_{m,k}}{\Lambda_k(\boldsymbol{\mu})}. \qquad (108)$$

Finally, putting together (78), (81) and (108), we obtain our final result:

$$\begin{aligned}
\theta_{m,k}(\boldsymbol{\mu}) &= \frac{\gamma}{\mu} \frac{\phi\left(\mathbf{B}_m \mathbf{A}^{-1} \mathbf{C}_k \mathbf{A}^{-1} \mathbf{B}_m\right)}{\left(1 + \phi\left(\mathbf{B}_m \mathbf{A}^{-1} \mathbf{B}_m\right)\right)^2} \\
&= \frac{\gamma}{\mu} \frac{\mu_k(\mu\beta_{m,k}^2 + \dot{U}_{m,k})}{\Lambda_k(\boldsymbol{\mu})} \eta_m^2(\boldsymbol{\mu}) \\
&= \frac{\mu_k \eta_m^2(\boldsymbol{\mu}) \left(\beta_{m,k}^2 + \dot{U}_{m,k}/\mu\right)}{\sum_{\ell=1}^{B} \eta_\ell(\boldsymbol{\mu}) \beta_{\ell,k}^2} \qquad (109)
\end{aligned}$$

where in the last line we used Theorem 1. Comparing (27) and (109), we see that the two expression coincide by letting $\xi_m = \dot{U}_m/\mu$. Therefore, Theorem 2 is proved.

Fig. 7 shows finite dimensional samples of $\theta_{m,k}(\boldsymbol{\mu})$ for randomly generated channels and their asymptotic values obtained from (109) under the two cell setting used in the example of Fig. 2 with $\boldsymbol{\mu} = [0.5\ 0.5\ 0.75\ 1\ 1\ 0.75\ 0.5\ 0.5]$. As $N$ increases, the finite dimensional samples (dots) converge the asymptotic values (lines) and this example shows the validness of the asymptotic analysis.

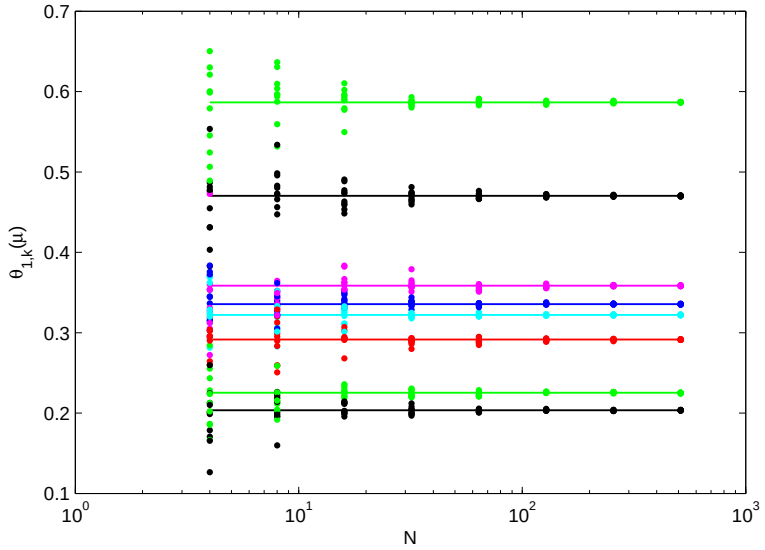



Fig. 7. Finite dimensional samples of $\theta_{m,k}(\boldsymbol{\mu})$ with $N = [4\ 8\ 16\ 32\ 64\ 128\ 256]$ (dots) and asymptotic values in the large system limit (lines) for $m = 1$, $k = 1, \ldots, 8$, and $\boldsymbol{\mu} = [0.5\ 0.5\ 0.75\ 1\ 1\ 0.75\ 0.5\ 0.5]$ under the same setting as in the example of Fig. 2

## APPENDIX C
## PROOF OF THEOREM 3

We wish to show that, under the symmetric system conditions, $\lim_{N\to\infty} \theta_{m,k}(\boldsymbol{\mu}) = \frac{\mu_i'}{B}$ for all user groups $k$ in equivalence class $i$. Again, we begin by recalling some facts from [52], [53].

*Definition 2:* An $N_r \times N_c$ matrix $\mathbf{P}$ with elements $\mathsf{P}_{i,j}$ is *asymptotically row-regular* if

$$\lim_{N_c\to\infty} \frac{1}{N_c} \sum_{j=1}^{N_c} 1\{\mathsf{P}_{i,j} \leq \alpha\}$$

is independent of $i$ for all $\alpha \in \mathbb{R}$, as the aspect ratio $\frac{N_c}{N_r}$ converges to a constant. ■

*Theorem 6 ( [52], [53]):* Define an $N_r \times N_c$ complex random matrix $\mathbf{H}$ whose entries are independent zero-mean complex circularly symmetric random variables (arbitrarily distributed) satisfying the Lindeberg condition (57) with variance given by (55) for an $N_r \times N_c$ deterministic asymptotically *row-regular matrix* $\mathbf{P}$ with uniformly bounded entries, for any $N_r$. Then, the asymptotic empirical eigenvalue distribution of $\mathbf{H}\mathbf{H}^{\mathsf{H}}$ converges almost surely to the asymptotic distribution of the random matrix $\mathbf{H}_w\mathbf{T}\mathbf{H}_w^{\mathsf{H}}$, where $\mathbf{H}_w$ is a matrix whose entries are i.i.d. complex circularly-symmetric with zero-mean and variance $1/N_r$ and where $\mathbf{T}$ is a diagonal matrix whose asymptotic empirical distribution is

given by

$$\mathsf{f}_{\mathbf{T}}(\alpha) = \lim_{N_c \to \infty} \frac{1}{N_c} \sum_{j=1}^{N_c} 1\{\mathsf{P}_{i,j} \leq \alpha\}$$

■

It is easy to check that when the system symmetry condition defined in Section III holds, then the variance matrix $\mathbf{P}$ of the random matrix $\mathbf{H}_{\boldsymbol{\mu}}$ is asymptotically row-regular as in Definition 2. Using Theorem 6, it follows that the asymptotic empirical spectral distribution of $\mathbf{H}_{\mu}\mathbf{H}_{\mu}^{\mathsf{H}}$ converges to that of $\mathbf{H}_w\mathbf{T}\mathbf{H}_w^{\mathsf{H}}$ where $\mathbf{T}$ is diagonal and invertible, with asymptotic spectral distribution $\mathsf{f}_{\mathbf{T}}(\alpha)$ given by the probability mass function with masses $\{\mu_k/\mu : k = 1, \ldots, A\}$ at the points $\{\beta_{m,k}^2 : k = 1, \ldots, A\}$ (notice that under the symmetry conditions the set of such points does not depend on $m$).

Under asymptotic row-regularity, $\mathbf{H}_{\boldsymbol{\mu}}$ is statistically equivalent to $\mathbf{W}\mathbf{T}^{1/2}$ with $\mathbf{W}$ Gaussian complex circularly symmetric, of dimension $\gamma NB \times \mu N$ and i.i.d. elements. In other words, we can replace $\mathbf{H}_{\boldsymbol{\mu}}$ with $\mathbf{H}_w\mathbf{T}^{1/2}$ in all normalized traces expressions, letting $N \to \infty$, and obtain the same result. We have

$$\begin{aligned}
\theta_{m,k}(\boldsymbol{\mu}) &= \frac{1}{N}\mathrm{tr}\left(\boldsymbol{\Phi}_m\mathbf{H}_{\boldsymbol{\mu}}(\mathbf{H}_{\boldsymbol{\mu}}^{\mathsf{H}}\mathbf{H}_{\boldsymbol{\mu}})^{-1}\boldsymbol{\Lambda}_{\boldsymbol{\mu}}^{1/2}\boldsymbol{\Theta}_k\boldsymbol{\Lambda}_{\boldsymbol{\mu}}^{1/2}(\mathbf{H}_{\boldsymbol{\mu}}^{\mathsf{H}}\mathbf{H}_{\boldsymbol{\mu}})^{-1}\mathbf{H}_{\boldsymbol{\mu}}^{\mathsf{H}}\boldsymbol{\Phi}_m\right) \\
&\to \frac{1}{N}\mathrm{tr}\left(\mathbf{T}^{1/2}\mathbf{W}^{\mathsf{H}}\boldsymbol{\Phi}_m\mathbf{W}\mathbf{T}^{1/2}(\mathbf{T}^{1/2}\mathbf{W}^{\mathsf{H}}\mathbf{W}^{\mathsf{H}}\mathbf{T}^{1/2})^{-1}\boldsymbol{\Lambda}_{\boldsymbol{\mu}}^{1/2}\boldsymbol{\Theta}_k\boldsymbol{\Lambda}_{\boldsymbol{\mu}}^{1/2}(\mathbf{T}^{1/2}\mathbf{W}^{\mathsf{H}}\mathbf{W}\mathbf{T}^{1/2})^{-1}\right) \\
&= \frac{1}{N}\mathrm{tr}\left((\mathbf{W}^{\mathsf{H}}\mathbf{W})^{-1}\mathbf{W}^{\mathsf{H}}\boldsymbol{\Phi}_m\mathbf{W}(\mathbf{W}^{\mathsf{H}}\mathbf{W})^{-1}\mathbf{T}^{-1/2}\boldsymbol{\Lambda}_{\boldsymbol{\mu}}^{1/2}\boldsymbol{\Theta}_k\boldsymbol{\Lambda}_{\boldsymbol{\mu}}^{1/2}\mathbf{T}^{-1/2}\right) \qquad (110)
\end{aligned}$$

where $\boldsymbol{\Phi}_m$ and $\boldsymbol{\Theta}_k$ were defined after eq. (71). From Theorem 2 we know that $\theta_{m,k}(\boldsymbol{\mu})$ converges almost surely. Furthermore, using (110) we see that the limit, in the symmetric case, is independent of $m$. In fact, because of the isotropic nature of $\mathbf{W}$, it is clear that we can replace $\mathbf{W}^{\mathsf{H}}\boldsymbol{\Phi}_m\mathbf{W}$ with $\mathbf{W}\boldsymbol{\Phi}_\ell\mathbf{W}$, for $\ell \neq m$, and the limit does not change.

Using these facts, in the limit for large $N$ we can write

$$\begin{aligned}
\lim_{N\to\infty}\theta_{m,k}(\boldsymbol{\mu}) &= \frac{1}{B}\sum_{\ell=1}^{B}\lim_{N\to\infty}\theta_{\ell,k}(\boldsymbol{\mu}) \\
&= \lim_{N\to\infty}\sum_{\ell=1}^{B}\frac{1}{NB}\mathrm{tr}\left(\boldsymbol{\Phi}_\ell\mathbf{H}_{\boldsymbol{\mu}}(\mathbf{H}_{\boldsymbol{\mu}}^{\mathsf{H}}\mathbf{H}_{\boldsymbol{\mu}})^{-1}\boldsymbol{\Lambda}_{\boldsymbol{\mu}}^{1/2}\boldsymbol{\Theta}_k\boldsymbol{\Lambda}_{\boldsymbol{\mu}}^{1/2}(\mathbf{H}_{\boldsymbol{\mu}}^{\mathsf{H}}\mathbf{H}_{\boldsymbol{\mu}})^{-1}\mathbf{H}_{\boldsymbol{\mu}}^{\mathsf{H}}\boldsymbol{\Phi}_\ell\right) \\
&= \lim_{N\to\infty}\frac{1}{NB}\mathrm{tr}\left((\mathbf{H}_{\boldsymbol{\mu}}^{\mathsf{H}}\mathbf{H}_{\boldsymbol{\mu}})^{-1}\boldsymbol{\Lambda}_{\boldsymbol{\mu}}^{1/2}\boldsymbol{\Theta}_k\boldsymbol{\Lambda}_{\boldsymbol{\mu}}^{1/2}(\mathbf{H}_{\boldsymbol{\mu}}^{\mathsf{H}}\mathbf{H}_{\boldsymbol{\mu}})^{-1}\mathbf{H}_{\boldsymbol{\mu}}^{\mathsf{H}}\mathbf{H}_{\boldsymbol{\mu}}\right) \\
&= \lim_{N\to\infty}\frac{1}{NB}\mathrm{tr}\left(\boldsymbol{\Lambda}_{\boldsymbol{\mu}}^{1/2}(\mathbf{H}_{\boldsymbol{\mu}}^{\mathsf{H}}\mathbf{H}_{\boldsymbol{\mu}})^{-1}\boldsymbol{\Lambda}_{\boldsymbol{\mu}}^{1/2}\boldsymbol{\Theta}_k\right) \\
&= \lim_{N\to\infty}\frac{1}{NB}\mathrm{tr}\left(\boldsymbol{\Theta}_k\right) \\
&= \frac{\mu_k}{B}. \qquad (111)
\end{aligned}$$

Finally, for a symmetric system we have that $\mu_k = \mu'_i$ for all user groups $k$ in equivalence class $i$. This concludes the proof.

## APPENDIX D
## PROOF OF THEOREM 4

Let $\widehat{\mathbf{V}}_{\boldsymbol{\mu}}$ denote the beamforming matrix for given user fractions $\boldsymbol{\mu}$, defined as in Section II-B after replacing $\mathbf{H}_{\boldsymbol{\mu}}$ with $\widehat{\mathbf{H}}_{\boldsymbol{\mu}}$, defined as in (7) with the change $\beta_{m,k} \to \widehat{\beta}_{m,k}$.

Let's focus on a generic user $j$ in group $k$. From (3) and (9) the received signal is given by

$$\begin{aligned} y_k^{(j)} &= \left(\underline{\mathbf{h}}_k^{(j)}\right)^{\mathsf{H}} \widehat{\mathbf{V}}_{\boldsymbol{\mu}} \mathbf{Q}^{1/2} \mathbf{u} + z_k^{(j)} \\ &= \left(\widehat{\underline{\mathbf{h}}}_k^{(j)}\right)^{\mathsf{H}} \widehat{\mathbf{v}}_k^{(j)} \sqrt{q_k^{(j)}} u_k^{(j)} + \left(\underline{\mathbf{e}}_k^{(j)}\right)^{\mathsf{H}} \widehat{\mathbf{V}}_{\boldsymbol{\mu}} \mathbf{Q}^{1/2} \mathbf{u} + z_k^{(j)} \end{aligned} \tag{112}$$

where we used the fact that $\widehat{\mathbf{v}}_k^{(j)}$ is orthogonal to all measured channel vectors $\widehat{\underline{\mathbf{h}}}_\ell^{(i)}$, for all other scheduled users, and we used the decomposition (46). The useful signal coefficient $\left(\widehat{\underline{\mathbf{h}}}_k^{(j)}\right)^{\mathsf{H}} \widehat{\mathbf{v}}_k^{(j)}$ is, by construction, equal to the diagonal element corresponding to user $j$ in group $k$ of the matrix $\widehat{\boldsymbol{\Lambda}}_{\boldsymbol{\mu}}^{1/2}$, calculated from $\widehat{\mathbf{H}}_{\boldsymbol{\mu}}$ as in (11). The additional interference term $\left(\underline{\mathbf{e}}_k^{(j)}\right)^{\mathsf{H}} \widehat{\mathbf{V}}_{\boldsymbol{\mu}} \mathbf{u}$ is the intra-cluster multiuser interference due to the fact that CSIT is not perfect.

A standard technique to lower bound the mutual information $I(u_k^{(j)}; y_k^{(j)} | \widehat{\mathbf{H}})$ is as follows:

$$\begin{aligned} I(u_k^{(j)}; y_k^{(j)} | \widehat{\mathbf{H}}) &= h(u_k^{(j)}) - h(u_k^{(j)} | y_k^{(j)}, \widehat{\mathbf{H}}) \\ &= \log \pi e q_k - h(u_k^{(j)} - a y_k^{(j)} | y_k^{(j)}, \widehat{\mathbf{H}}) \\ &\geq \log \pi e q_k - h(u_k^{(j)} - a y_k^{(j)} | \widehat{\mathbf{H}}) \\ &\geq \log \pi e q_k - \mathbb{E}\left[\log \pi e \mathrm{Var}(u_k^{(j)} - a y_k^{(j)} | \widehat{\mathbf{H}})\right] \end{aligned} \tag{113}$$

where we assumed that $u_k^{(j)}$ is Gaussian with variance 1 (denoting, as before, the transmit power to users in group $k$). The bound holds for any coefficient $a$. In particular, we wish to use the coefficient that minimizes the conditional variance $\mathrm{Var}(u_k^{(j)} - a y_k^{(j)} | \widehat{\mathbf{H}})$, given by the linear MMSE estimation of $u_k^{(j)}$ from $y_k^{(j)}$ for given $\widehat{\mathbf{H}}$. After standard algebra, omitted here for the sake of brevity, we obtain the variance (conditional MMSE estimation error)

$$\mathrm{Var}(u_k^{(j)} - a y_k^{(j)} | \widehat{\mathbf{H}}) = \frac{q_k \left[\mathbb{E}\left[\left(\underline{\mathbf{e}}_k^{(j)}\right)^{\mathsf{H}} \widehat{\mathbf{V}}_{\boldsymbol{\mu}} \mathbf{Q} \widehat{\mathbf{V}}_{\boldsymbol{\mu}}^{\mathsf{H}} \underline{\mathbf{e}}_k^{(j)} \middle| \widehat{\mathbf{H}}\right] + 1\right]}{\left|\left(\widehat{\underline{\mathbf{h}}}_k^{(j)}\right)^{\mathsf{H}} \mathbf{v}_k^{(j)}\right|^2 q_k + \mathbb{E}\left[\left(\underline{\mathbf{e}}_k^{(j)}\right)^{\mathsf{H}} \widehat{\mathbf{V}}_{\boldsymbol{\mu}} \mathbf{Q} \widehat{\mathbf{V}}_{\boldsymbol{\mu}}^{\mathsf{H}} \underline{\mathbf{e}}_k^{(j)} \middle| \widehat{\mathbf{H}}\right] + 1} \tag{114}$$

Replacing this into (113), we obtain the desired lower bound in the form

$$I(u_k^{(j)};y_k^{(j)}|\widehat{\mathbf{H}}) \geq \mathbb{E}\left[\log\left(1+\frac{\left|\left(\widehat{\underline{\mathbf{h}}}_k^{(j)}\right)^{\mathsf{H}}\mathbf{v}_k^{(j)}\right|^2 q_k}{1+\mathbb{E}\left[\left(\underline{\mathbf{e}}_k^{(j)}\right)^{\mathsf{H}}\widehat{\mathbf{V}}_{\boldsymbol{\mu}}\mathbf{Q}\widehat{\mathbf{V}}_{\boldsymbol{\mu}}^{\mathsf{H}}\underline{\mathbf{e}}_k^{(j)}\middle|\widehat{\mathbf{H}}\right]}\right)\right] \tag{115}$$

Let's examine the terms in (115) separately. As already said before, the coefficient in the numerator of the SINR term inside the logarithm, in the large system limit, is given by $\left|\left(\widehat{\underline{\mathbf{h}}}_k^{(j)}\right)^{\mathsf{H}}\mathbf{v}_k^{(j)}\right|^2 \to \widehat{\Lambda}_k(\boldsymbol{\mu})$, where the latter is obtained via Theorem 1 replacing the coefficients $\beta_{m,k}$ with the new coefficients $\widehat{\beta}_{m,k}$ defined in (49), thus obtaining (53) and (54).

The intra-cluster interference term in the denominator can be evaluated as follows. First, notice that because of the properties of the MMSE estimator, the channel error vector is independent of the estimator $\widehat{\mathbf{H}}$. Therefore, the conditioning with respect to $\widehat{\mathbf{H}}$ makes $\widehat{\mathbf{V}}_{\boldsymbol{\mu}}$ and the diagonal matrix of transmitted powers $\mathbf{Q}$ act as constant matrices with respect to the conditional expectation, since they are both functions of the CSIT $\widehat{\mathbf{H}}$. We have

$$\begin{aligned}
\mathbb{E}\left[\left(\underline{\mathbf{e}}_k^{(j)}\right)^{\mathsf{H}}\widehat{\mathbf{V}}_{\boldsymbol{\mu}}\mathbf{Q}\widehat{\mathbf{V}}_{\boldsymbol{\mu}}^{\mathsf{H}}\underline{\mathbf{e}}_k^{(j)}\middle|\widehat{\mathbf{H}}\right] &= \mathrm{tr}\left(\mathbf{Q}\widehat{\mathbf{V}}_{\boldsymbol{\mu}}^{\mathsf{H}}\mathrm{Cov}(\underline{\mathbf{e}}_k^{(j)})\widehat{\mathbf{V}}_{\boldsymbol{\mu}}\right)\\
&= \mathrm{tr}\left(\mathbf{Q}\widehat{\mathbf{V}}_{\boldsymbol{\mu}}^{\mathsf{H}}\boldsymbol{\Sigma}_k\widehat{\mathbf{V}}_{\boldsymbol{\mu}}\right)\\
&= \mathrm{tr}\left(\boldsymbol{\Sigma}_k\widehat{\mathbf{V}}_{\boldsymbol{\mu}}\mathbf{Q}\widehat{\mathbf{V}}_{\boldsymbol{\mu}}^{\mathsf{H}}\right)\\
&= \sum_{m=1}^{B}\bar{\beta}_{m,k}^2
\end{aligned} \tag{116}$$

where the last line follows from the definition of $\boldsymbol{\Sigma}_k$ in (45), which is block-diagonal with $B$ diagonal blocks of dimension $\gamma N \times \gamma N$, and the $m$-th diagonal block is given by $\bar{\beta}_{m,k}^2\mathbf{I}$ where $\bar{\beta}_{m,k}$ is defined in (51), and by noticing that $\widehat{\mathbf{V}}_{\boldsymbol{\mu}}\mathbf{Q}\widehat{\mathbf{V}}_{\boldsymbol{\mu}}^{\mathsf{H}}$ is the covariance matrix of the signal transmitted from all the base stations forming the cluster. Under a per-BS power constraint, the partial trace of this matrix on any diagonal segment corresponding to one base station (diagonal segments of length $\gamma N$) is equal to 1. Therefore, the simple form of (116) follows. This concludes the proof.

## REFERENCES


[1] IEEE 802.16 broadband wireless access working group, "IEEE 802.16m system requirements," IEEE 802.16m-07/002, Tech. Rep., Jan. 2010.

[2] 3GPP technical specification group radio access network, "Further advancements for E-UTRA: LTE-Advanced feasibility studies in RAN WG4," 3GPP TR 36.815, Tech. Rep., March 2010.

[3] S. Parkvall, E. Dahlman, A. Furuskar, Y. Jading, M. Olsson, S. Wanstedt, and K. Zangi, "LTE-Advanced – Evolving LTE towards IMT-Advanced," in *Proc. IEEE Vehic. Tech. Conf. (VTC)*, Calgary, Alberta, Sept. 2008.

[4] J.-B. Landre, A. Saadani, and F. Ortolan, "Realistic performance of HSDPA MIMO in macro-cell environment," in *Proc. IEEE Int. Symp. on Personal, Indoor and Mobile Radio Commun. (PIMRC)*, Tokyo, Japan, Sept. 2009.

[5] A. Farajidana, W. Chen, A. Damnjanovic, T. Yoo, D. Malladi, and C. Lott, "3GPP LTE downlink system performance," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Honolulu, HI, Nov. 2009.

[6] 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.

[7] P. Viswanath and D. Tse, "Sum capacity of the vector Gaussian broadcast channel and uplink-downlink duality," *IEEE Trans. on Inform. Theory*, vol. 49, pp. 1912–1921, Aug. 2003.

[8] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels," *IEEE Trans. on Inform. Theory*, vol. 49, pp. 2658–2668, Oct. 2003.

[9] W. Yu and J. Cioffi, "Sum capacity of Gaussian vector broadcast channels," *IEEE Trans. on Inform. Theory*, vol. 50, pp. 1875–1892, Sept. 2004.

[10] 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.

[11] G. Foschini, K. Karakayali, and R. A. Valenzuela, "Coordinating multiple antenna cellular networks to achieve enormous spectral efficiency," *IEE Proc. Commun.*, vol. 152, pp. 548–555, Aug 2006.

[12] S. Jing, D. N. C. Tse, J. B. Soriaga, J. Hou, J. E. Smee, and R. Padovani, "Downlink macro-diversity in cellular networks," in *Proc. IEEE Int. Symp. on Inform. Theory (ISIT)*, Nice, France, June 2007.

[13] F. Boccardi and H. Huang, "Limited downlink network coordination in cellular networks," in *Proc. IEEE Int. Symp. on Personal, Indoor and Mobile Radio Commun. (PIMRC)*, Athens, Greece, Sept. 2007.

[14] G. Caire, S. Ramprashad, H. Papadopoulos, C. Pepin, and C.-E. Sundberg, "Multiuser MIMO downlink with limited inter-cell cooperation: approximate interference alignment in time, frequency and space," in *Proc. Allerton Conf. on Commun., Control, and Computing*, Urbana-Champaign, IL, Sept. 2008.

[15] H. Dahrouj and W. Yu, "Coordinated beamforming for the multicell, multi-antenna wireless system," *IEEE Trans. on Wireless Commun.*, vol. 9, no. 5, pp. 1748–1759, May 2010.

[16] H. Huh, H. C. Papadopoulos, and G. Caire, "Multiuser MISO transmitter optimization for intercell interference mitigation," *IEEE Trans. on Sig. Proc.*, vol. 58, no. 8, pp. 4272–4285, Aug. 2010.

[17] G. Boudreau, J. Panicker, N. Guo, R. Chang, N. Wang, and S. Vrzic, "Interference coordination and cancellation for 4G networks," *IEEE Communications Magazine*, vol. 47, no. 4, pp. 74–81, April 2009.

[18] WiMAX Forum, "Mobile WiMAX – Part I: A technical overview and performance evaluation," Tech. Rep., Aug. 2006.

[19] P. Viswanath, D. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Trans. on Inform. Theory*, vol. 48, pp. 1277–1294, June 2002.

[20] 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.

[21] H. Shirani-Mehr, G. Caire, and M. J. Neely, "MIMO downlink scheduling with non-perfect channel state knowledge," *IEEE Trans. on Commun.*, vol. 58, no. 7, pp. 2055–2066, July 2010.

[22] 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.

[23] H. Huang and R. Valenzuela, "Fundamental simulated performance of downlink fixed wireless cellular networks with

multiple antennas," in *Proc. IEEE Int. Symp. on Personal, Indoor and Mobile Radio Commun. (PIMRC)*, Berlin, Germany, Sept. 2005.

[24] J. Zhang, R. Chen, J. Andrews, A. Ghosh, and R. Heath, "Networked MIMO with clustered linear precoding," *IEEE Trans. on Wireless Commun.*, vol. 8, pp. 1910–1921, April 2009.

[25] H. Huang, M. Trivellato, A. Hottinen, M. Shafi, P. Smith, and R. Valenzuela, "Increasing downlink cellular throughput with limited network MIMO coordination," *IEEE Trans. on Wireless Commun.*, vol. 8, pp. 2983–2989, June 2009.

[26] S. A. Ramprashad and G. Caire, "Cellular vs. network MIMO: a comparison including the channel state information overhead," in *Proc. IEEE Int. Symp. on Personal, Indoor and Mobile Radio Commun. (PIMRC)*, Tokyo, Japan, Sept. 2009.

[27] H. Huh, G. Caire, S. H. Moon, Y.-T. Kim, and I. Lee, "Multi-cell MIMO downlink with cell cooperation and fair scheduling: a large-system limit analysis," *submitted to IEEE Trans. on Inform. Theory*, 2010. [Online]. Available: http://arxiv.org/abs/1006.2162

[28] G. Caire, S. A. Ramprashad, and H. C. Papadopoulos, "Rethinking network MIMO: cost of CSIT, performance Analysis, and architecture Comparisons," in *Proc. Inform. Theory and Appl. Workshop (ITA)*, San Diego, CA, Feb. 2010.

[29] A. Tulino and S. Verdu, "Asymptotic analysis of improved linear receivers for BPSK-CDMA subject to fading," *IEEE J. Select. Areas Commun.*, vol. 19, no. 8, pp. 1544–1555, aug 2001.

[30] R. Muller and S. Verdu, "Design and analysis of low-complexity interference mitigation on vector channels," *IEEE J. Select. Areas Commun.*, vol. 19, no. 8, pp. 1429–1441, aug 2001.

[31] G. Dimic and N. Sidiropoulos, "On downlink beamforming with greedy user selection: Performance analysis and simple new algorithm," *IEEE Trans. on Sig. Proc.*, vol. 53, pp. 3857–3868, Oct. 2005.

[32] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE J. Select. Areas Commun.*, vol. 24, pp. 528–541, March 2006.

[33] B. Hochwald, T. Marzetta, and V. Tarokh, "Multiple-antenna channel hardening and its implications for rate feedback and scheduling," *IEEE Trans. on Inform. Theory*, vol. 50, no. 9, pp. 1893–1909, Sept. 2004.

[34] A. Tomasoni, G. Caire, M. Ferrari, and S. Bellini, "On the selection of semi-orthogonal users for zero-forcing beamforming," in *Proc. IEEE Int. Symp. on Inform. Theory (ISIT)*, Seoul, Korea, June 2009.

[35] P. Ding, D. J. Love, and M. D. Zoltowski, "Multiple antenna broadcast channels with shape feedback and limited feedback," *IEEE Trans. on Sig. Proc.*, vol. 55, pp. 3417–3428, July 2007.

[36] J. Jose, A. Ashikhmin, P. Whiting, and S. Vishwanath, "Precoding methods for multi-user TDD MIMO systems," in *Proc. Allerton Conf. on Commun., Control, and Computing*, Urbana-Champaign, IL, Sept. 2008.

[37] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, "Multiuser MIMO achievable rates with downlink training and channel state feedback," *IEEE Trans. on Inform. Theory*, vol. 56, no. 6, pp. 2845–2866, June 2010.

[38] S. Wagner, R. Couillet, M. Debbah, and D. T. M. Slock, "Large system analysis of linear precoding in MISO broadcast channels with limited feedback," *submitted to IEEE Trans. on Inform. Theory*, 2009. [Online]. Available: http://arxiv.org/abs/0906.3682

[39] J. Hoydis, M. Kobayashi, and M. Debbah, "On the optimal number of cooperative base stations in network MIMO systems," *submitted to IEEE Trans. on Sig. Proc.*, 2010. [Online]. Available: http://arxiv.org/abs/1003.0332

[40] R. Zakhour and S. V. Hanly, "Base station cooperation on the downlink: large system analysis," *submitted to IEEE J. Select. Areas Commun.*, 2010. [Online]. Available: http://arxiv.org/abs/1006.3360

[41] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.

[42] A. Wiesel, Y. Eldar, and S. Shamai (Shitz), "Zero-forcing precoding and generalized inverses," *IEEE Trans. on Sig. Proc.*, vol. 56, pp. 4409–4418, Sept. 2008.

[43] H. Huh, H. C. Papadopoulos, and G. Caire, "MIMO broadcast channel optimization under general linear constraints," in *Proc. IEEE Int. Symp. on Inform. Theory (ISIT)*, Seoul, Korea, June 2009.

[44] R. Zhang, "Cooperative multi-cell block diagonalization with per-base-station power constraints," *submitted to IEEE J. Select. Areas Commun.*, 2009. [Online]. Available: http://arxiv.org/abs/0910.2304

[45] A. Lapidoth and S. Shamai, "Channel state feedback schemes for multiuser MIMO-OFDM downlink," *IEEE Trans. on Commun.*, vol. 57, no. 9, pp. 2713–2723, Sept. 2009.

[46] K. R. Kumar and G. Caire, "Channel state feedback over the MIMO-MAC," in *Proc. IEEE Int. Symp. on Inform. Theory (ISIT)*, Seoul, Korea, June 2009.

[47] M. Kobayashi, N. Jindal, and G. Caire, "Training and feedback optimization for multiuser MIMO downlink," *To appear in IEEE Trans. on Inform. Theory*, 2010. [Online]. Available: http://arxiv.org/abs/0912.1987

[48] T. Marzetta and B. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Trans. on Inform. Theory*, vol. 45, no. 1, pp. 139–157, Jan. 1999.

[49] L. Zheng and D. Tse, "Communication on the Grassmann manifold: a geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. on Inform. Theory*, vol. 48, no. 2, pp. 359–383, Feb. 2002.

[50] B. Hassibi and B. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. on Inform. Theory*, vol. 49, no. 4, pp. 951–963, April 2003.

[51] C. Wang, T. Gou, and S. A. Jafar, "Aiming perfectly in the dark - blind interference alignment through staggered antenna switching," *submitted to IEEE Trans. on Inform. Theory*, 2010. [Online]. Available: http://arxiv.org/abs/1002.2720

[52] A. M. Tulino and S. Verdu, *Random Matrix Theory and Wireless Communications*. Foundations and Trends in Communications and Information Theory, 2004, vol. 1, no. 1.

[53] 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.