

# Rethinking Network MIMO: Cost of CSIT, Performance Analysis, and Architecture Comparisons

Giuseppe Caire<sup>†</sup> and Sean A. Ramprasad\* and Haralabos C. Papadopoulos\*

\* DoCoMo Communications Labs USA, Inc., Palo Alto, CA 94304

<sup>†</sup> EE Dept., University of Southern California, Los Angeles, CA 90089

**Abstract**—We compare the downlink throughput of various cellular architectures with multi-antenna base stations and multiple single-antenna users per cell, by considering a number of inherent physical layer issues such as path-loss and time and frequency selective fading. In particular, we focus on Multiuser MIMO (MU-MIMO) downlink techniques that require channel state information at the transmitter (CSIT). Our analysis takes explicit account of the cost of CSIT estimation and illuminates the tradeoffs between CSIT, estimation error, and system resource dedicated to training. This tradeoff shows that the number of antennas that can be jointly coordinated (either on the same base station or across multiple base stations) is intrinsically limited not just by “external factors,” such as complexity and rate of the backbone wired network, but by the inherent time and frequency variability of the fading channels. Our analysis, in agreement with a number of recent simulation results, shows that conventional MU-MIMO cellular architectures may outperform schemes based on coordinated transmission from base stations (referred to as Network MIMO schemes, NW-MIMO), at the negligible cost of a few extra antennas per station. In light of these results, it appears that the inherent bottleneck of NW-MIMO systems is not the backbone network (which here is assumed ideal with infinite capacity) but the intrinsic dimensional limitation of estimating the channels.

## I. INTRODUCTION

Network MIMO (NW-MIMO) has recently generated much interest as a potentially attractive technique to increase overall wireless efficiency [1]. As already noticed by Wyner [2]–[4], if all base-stations (BSs) are jointly coordinated through an ideal backbone network such that their signals can be jointly processed, then inter-cell interference (ICI) disappears and all signal power becomes “useful”. Even in the case where base-station cooperation is limited to small clusters, the system throughput improvement of NW-MIMO can be very significant, specially if combined with clever opportunistic multiuser scheduling schemes and with overlapped architectures where different overlapping clusters of coordinated base-stations operate on different frequency subchannels [5], [6].

Such system performance benefits, however, are often based on studies that ignore inherent and fundamental overheads that are necessary to enable the downlink signaling by MU-MIMO. Even without assuming any limitation in the number of antennas per base-station and/or in the number of jointly coordinated base-stations, a fundamental intrinsic limitation is represented by the ability of providing channel state information at the transmitters (CSIT), when the fading channel is varying in time and frequency.

For example, one such overhead is the training resources required in these multi-cell environments to obtain sufficiently

accurate CSIT to support efficient downlink signaling by MU-MIMO. This CSIT overhead is often greater for NW-MIMO systems than for conventional cellular systems given the need for a user to estimate channels with respect to multiple base-stations. In some recent works, the tradeoff between CSIT accuracy and system throughput has been studied by simulation, under some rather crude assumptions on the overhead necessary to reliably estimating the fading channels [7], [8]. Instead of comparing systems for a fixed number of antennas at the base-station, as done in most existing works, [7], [8] compared different system architectures assuming only the following fundamental limitations: 1) the transmit power per base-station and 2) a fixed dimensional cost of estimating the fading channels, which grows with the number of jointly coordinated antennas. Subject to these constraints, the number of antennas per BS becomes a parameter to be optimized, as well as the number of BSs per coordination cluster.

Such studies demonstrate a number of interesting system tradeoffs. For example, CSIT accuracy requirements can differ between NW-MIMO and conventional cellular systems<sup>1</sup>. Nonetheless, in general NW-MIMO does require more pilot training overhead. The result is that for a given CSIT overhead, a cellular system, possibly with frequency reuse factors greater than one, may provide larger throughput than a corresponding NW-MIMO system in certain conditions of path-loss and fading time-frequency selectivity.

The works in [5]–[8] are largely based on simulations since an exact theoretical analysis of such systems, involving non-perfect CSIT, ICI and multiuser scheduling, appears to be very difficult if not impossible. One fundamental aspect of cellular systems considered in these studies that complicates analysis is the distance-dependent path-loss. This creates an inherent asymmetry of the system with respect to users near their base-station and users located at the edge of the cells. Downlink scheduling schemes strive to maximize the system capacity subject to some fairness criteria [9], [10]. Disregarding fairness and focusing only on system capacity would lead to misleading conclusions, since the system would allocate resource only to the users in favorable locations and neglect the bottleneck effect of “cell-edge” users. Therefore, fairness and scheduling must be considered as a relevant part of the problem.

In this paper we focus on certain relevant models for

<sup>1</sup>NW-MIMO does at times require less accurate CSIT than conventional cellular with respect to some stations. For example, a user at the edge of a cluster experiencing high ICI may only require accurate CSI for a few closest signaling stations.

conventional cellular and NW-MIMO architectures. We assume explicit downlink training and channel estimation. Our analysis applies to mainly to FDD, since TDD can exploit channel reciprocity and train a large number of base stations antennas with the same uplink pilot signals [11]. Furthermore, TDD is affected by a more significant pilot contamination problem [12] that is not so severe in FDD. For the considered models, we provide bounds on the long-term average system throughput subject to fairness criteria, under some simplified scheduling assumptions that make the analysis feasible. In particular, the fraction of time (activity factor) with which each user is served by the downlink scheduler is the result of a convex optimization problem, for a given concave network utility function that reflects the fairness criterion. In this work we focus on Proportional Fair Scheduling (PFS). While not fully representing all the techniques examined in [5]–[8], our theoretical analysis leads to similar trends and conclusions as the simulation results of the above referenced works.

## II. SYSTEM MODELS

We review here the system setup of [5], [7], [8]. Here we focus on a 1-dimensional cellular layout. Simulation results for 2-dimensional layouts can be found in [6].

We assume that the system total bandwidth spans  $F$  fading coherence bandwidths of  $W$  Hz each. On each fading coherence bandwidth, the channels are assumed frequency-flat. The coherence bands are indexed by  $f = 0, \dots, F - 1$ . We assume the usual block-fading model, according to which the channel is constant in time over intervals of duration  $T$  seconds (channel coherence time). Furthermore, we assume that the product  $WT$  is sufficiently large such that by sending a codeword in a time-frequency block at a rate below the mutual information of the resulting Gaussian channel (corresponding to the time-frequency constant fading coefficient), an arbitrarily small probability of error is achieved. In this work we compare three system architectures reviewed in Section II-B and referred to as: 1) Conventional cellular; 2) Expanded Cellular; and 3) Overlapped Cluster.

### A. Cellular layout, SNR and pathloss

We indicate all base-stations in the system by a unique index  $b \in \{0, \dots, B - 1\}$  and all users in the system by a unique index  $k \in \{0, \dots, K - 1\}$ . The cellular layout spans the interval  $[0, B]$  in the real line, where base-station  $b$  is located at the spatial coordinate  $b + \frac{1}{2}$  and user  $k$  is located at a point  $v_k$  belonging to a regular uniform grid of points such that the number of users per cell is  $K/B$  for all cells and users are symmetrically located in each cell.<sup>2</sup> The channel-path gain for a transmitter-receiver pair located at  $u, v \in [0, B]$  is given by

$$g(u, v) = \frac{G_0}{1 + (d_B(u, v)/\delta)^\alpha} \quad (1)$$

where we define the modulo- $B$  distance  $d_B(u, v) = \min_{z \in \mathbb{Z}} |u - v + Bz|$ . This has the advantage of eliminating the

<sup>2</sup>The grid is such that no user is at a cell boundary. Thus each user is uniquely associated with a cell.

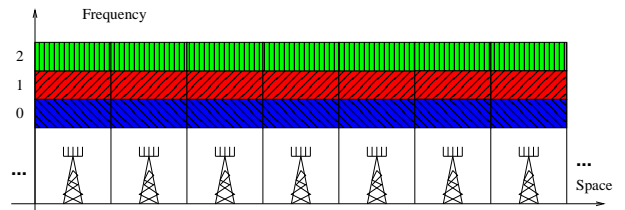


Fig. 1. An example of cellular system with  $F = 3$ .

boundary effects. The parameters  $G_0$ ,  $\alpha$  and  $\delta$  in (1) denote the SNR at the cell center, the path-loss exponent and the 3dB-loss distance. The function (1) can approximate pathloss models used in LTE where gains are  $\propto 1/d^\alpha$  but are clipped to a maximum value for  $d$  below some minimum distance [13, Table 5.1]. All systems operate under strict per base-station power constraints, where  $p_b[f] \geq 0$  is the maximum power that station  $b$  can use on subchannel  $f$ . The powers are relative to the SNR given by (1), assuming a normalized noise power spectral density equal to 1, where  $\frac{1}{F} \sum_{f=0}^{F-1} p_b[f] = 1$ .

### B. System architectures

a) *Conventional cellular system.*: In this case, base-stations are equipped with  $M$  antennas. Base-station  $b$  serves only users within its cell, i.e., users in the interval  $[b, b + 1]$ . We restrict to the case where base-stations allocate their power equally over all frequency bands, i.e.,  $p_b[f] = 1 \forall b, f$ . Fig. 1 shows an example for  $M = 3$  and  $F = 3$ .

b) *Expanded cellular system.*: An expanded cellular system is a system where only one in every  $R$ -th,  $R > 1$ , station is active on a subchannel. We assume that  $R$  divides  $F$ , such that each station is assigned  $F/R$  subchannels in a round-robin fashion. Thus  $p_b[f] = R$  for all  $f$  on which station  $b$  is active, and zero otherwise. Differently from the conventional frequency reuse  $R$ , stations are able to serve users in adjacent cells. Specifically, base-station  $b$  is allowed to server users in the *expanded* interval  $[b - (R - 1)/2, b + (R + 1)/2]$  (wrapped modulo  $[0, B]$ ) on its active frequency subchannels. Thus, even though  $R > 1$ , all users in a given cell can be served on all frequency subchannels and therefore the system has frequency reuse 1. Fig. 2 shows an example for  $F = R = 3$  with the overlapped coverage areas. Note, given the ability of users to use all subchannels, this system with appropriate user scheduling outperforms a conventional cellular system using a frequency reuse factor of  $R > 1$ .

c) *Overlapped clustering.*: In this case, clusters of  $C$  adjacent base stations are connected to a central cluster controller, such that they act as a distributed joint transmitter with  $CM$  antennas. Clusters serve the users within the cells of constituent stations (with the modulo  $[0, B]$  wrapping). Different clustering patterns are active on different subchannels. We assume that  $C$  divides  $F$  such that  $F/C$  subchannels are allocated to any given clustering pattern. Furthermore, clustering patterns are shifted versions of each other and overlap in space as proposed in [5]. In this way no user is permanently in a disadvantaged location (near a cluster edge) on all subchannels. Fig. 3 shows an example for  $C = F = 3$ .

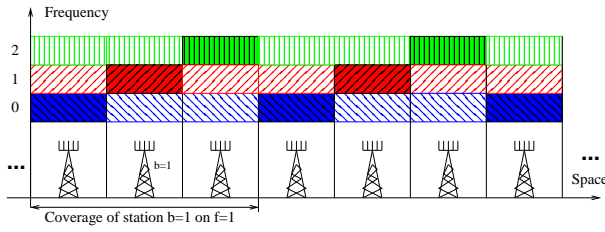


Fig. 2. An example of expanded cellular system with  $R = F = 3$ .

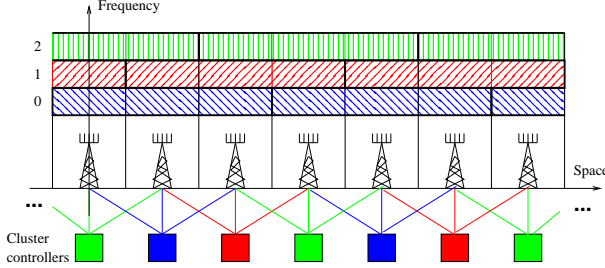


Fig. 3. An example of intertwined clustering system with  $C = F = 3$ .

### C. Received signal model

Channel vectors are assumed i.i.d. from block to block in time and for different coherence bands (time-frequency i.i.d. block-fading model). We shall omit the time index and focus on a given subchannel within a coherence interval. The vector of channel coefficients between base station  $b$  and the single-antenna user  $k$ , located at  $v_k \in [0, B]$ , on subchannel  $f = 0, \dots, F-1$ , is indicated by  $\mathbf{h}_{k,b}[f] \in \mathbb{C}^{M \times 1}$ , where  $\mathbf{h}_{k,b}[f] \sim \mathcal{CN}(\mathbf{0}, g_{k,b}\mathbf{I})$  with  $g_{k,b} = g(v_k, b)$ . We denote by  $\mathcal{C}_{(k,f)}$  the set of base stations that may serve user  $k$  on subchannel  $f$ . This depends on the system architecture as described before. With this notation, a block of received signal samples on subchannel  $f$  of user  $k$  receiver is given by

$$\mathbf{y}_k[f] = \underbrace{\sum_{b \in \mathcal{C}_{(k,f)}} \mathbf{X}_b[f] \mathbf{h}_{k,b}[f]}_{\text{desired}} + \underbrace{\sum_{b \notin \mathcal{C}_{(k,f)}} \mathbf{X}_b[f] \mathbf{h}_{k,b}[f]}_{\text{interference}} + \mathbf{z}_k[f] \quad (2)$$

where  $\mathbf{y}_k[f] \in \mathbb{C}^{L \times 1}$ ,  $\mathbf{z}_k[f]$  is an  $L$ -dimensional i.i.d. Gaussian noise vector with components  $\sim \mathcal{CN}(0, 1)$ , and where  $\mathbf{X}_b[f] \in \mathbb{C}^{L \times M}$  denotes the matrix of symbols sent by base-station  $b$ . Consistently with the system assumptions of before, the number of complex dimensions (channel uses) per block is given by  $L = \lfloor WT \rfloor$ . Base-station  $b$ , at every channel use, sends one row of the matrix  $\mathbf{X}_b[f]$  on subchannel  $f$ , in parallel from its  $M$  antennas.

### III. DOWNLINK TRAINING AND CHANNEL ESTIMATION

We assume explicit training and estimation of the downlink channel vectors. The scheme is described for the general case  $C \geq 1$ , understanding that  $C = 1$  corresponds to the cellular systems. In order to estimate a  $CM$ -dimensional downlink channel vector,  $CM$  orthogonal training sequences per fading time-frequency block (spanning  $L = WT$  dimensions) are needed. In the case of FDD,  $CM$  orthogonal training sequences per fading block are broadcasted by the base-stations in the cluster. Users estimate their own channel and feed back some noisy or quantized version of their estimate. Several

alternatives have been proposed in order to implement this CSIT feedback (see for example [14]–[16]) and these may require different uplink overhead. Since it is not clear how to take into account the uplink CSIT feedback overhead as a “cost” for the downlink spectral efficiency, we shall assume a genie-aided CSIT feedback where the channel estimates at the user terminals are simultaneously known by all base-stations in the cluster at no cost. Therefore, in our optimistic analysis of CSIT training/feedback overhead, we take into account only the overhead and the estimation error resulting from the training phase. Clearly, in both cases of TDD and FDD, this provides a best-case for any system based on explicit training, channel estimation and CSIT feedback.

#### A. Pilot reuse scheme

Each block of  $L$  channel uses is divided into a training and a data transmission phase. The training phase spans  $\beta Q M$  channel uses, using a scaled training sequence  $\theta_i = \sqrt{a} \times \phi_i$ , where  $\phi_i \in \mathbb{C}^{\beta Q M \times 1}$ ,  $\phi_i^H \phi_i = 1$ ,  $a$  is a power normalization factor,  $\beta \geq 1$ , and  $Q$  is an integer that defines the *pilot reuse factor*. The system has  $QM$  orthonormal training sequences  $\{\phi_0, \dots, \phi_{Q-1}\}$ , forming a unitary matrix  $\Phi \in \mathbb{C}^{\beta Q M \times Q M}$ . The matrix is partitioned into  $Q$  non-overlapping blocks of size  $\beta Q M \times M$ , denoted by  $\Phi_0, \dots, \Phi_{Q-1}$ . On each frequency subchannel  $f$  the blocks are assigned to base-stations in a round-robin fashion, such that the same training matrix is reused every  $Q$  stations. Also, we assume that training phases are aligned in time and frequency such that pilots interfere only with pilots from other stations. In the training phase, the  $b$ -th base station on channel  $f$  transmits the matrix  $\Theta_b[f] = \sqrt{a} \Phi_{b \bmod Q}$ .

The per-base station power constraint  $p_b[f]$  applies also to the training phase and yields

$$\text{tr}(\Theta_b^H[f] \Theta_b[f]) = aM = p_b[f] \beta Q M$$

For the conventional cellular systems and cluster systems, since all base-stations are active on all frequencies, we have  $p_b[f] = 1 \forall f, b$ , leading to  $a = \beta Q$ . For the expanded cellular system, since the base-station transmits only on a fraction  $F/R$  of subchannels, then  $p_b[f] = R$ , leading to  $a = R\beta Q$ .

#### B. MMSE channel estimation

Consider user  $k$  and assume  $c \in \mathcal{C}_{(k,f)}$  is a base station potentially serving user  $k$  on subchannel  $f$ . We denote by  $\bar{\mathcal{P}}(c, f)$  the set of all base stations (including  $c$ ) sharing the same pilot matrix with  $c$  and which are active on subchannel  $f$ , and by  $\mathcal{P}(c, f) = \bar{\mathcal{P}}(c, f) - \{c\}$  the same set excluding  $c$ . In the conventional cellular and overlapped cluster systems, each training matrix is effectively reused every  $Q$ -th station. In the expanded cellular case, each training matrix is reused every  $RQ$ -th station since only one every  $R$ -th station is active on the same subchannel.

After projecting the received signal onto the training matrix for station  $c$ , the received signal for channel estimation is

$$\mathbf{r}_{k,c}[f] = a \mathbf{h}_{k,c}[f] + \sum_{b \in \mathcal{P}(c,f)} a \mathbf{h}_{k,b}[f] + \mathbf{w}_{k,c}[f] \quad (3)$$

where  $\mathbf{w}_{k,c}[f] \sim \mathcal{CN}(\mathbf{0}, a\mathbf{I})$ . The estimation MMSE for the linear Gaussian model (3) is given by

$$\epsilon_{k,c}[f] = \frac{g_{k,c}}{1 + \frac{g_{k,c}}{a^{-1} + \sum_{b \in \mathcal{P}(c,f)} g_{k,b}}} \quad (4)$$

The above formula applies to all systems considered here, by particularizing  $\mathcal{P}(c,f)$  and the value of  $a$  according to the cases. Note also that the MMSE is the same for all the components of the channel vector  $\mathbf{h}_{k,c}[f]$ , and that the estimation error vector is Gaussian and white. From well-known results on MMSE estimation, we can write the channel vectors as

$$\mathbf{h}_{k,c}[f] = \hat{\mathbf{h}}_{k,c}[f] + \mathbf{e}_{k,c}[f] \quad (5)$$

where  $\mathbf{e}_{k,c}[f] \sim \mathcal{CN}(\mathbf{0}, \epsilon_{k,c}[f]\mathbf{I})$  is the estimation error vector,  $\hat{\mathbf{h}}_{k,c}[f] \sim \mathcal{CN}(\mathbf{0}, (g_{k,c} - \epsilon_{k,c}[f])\mathbf{I})$  is the channel estimate, and where  $\mathbf{e}_{k,c}[f]$  and  $\hat{\mathbf{h}}_{k,c}[f]$  are statistically independent.

#### IV. MU-MIMO ZERO-FORCING BEAMFORMING

This section describes the MU-MIMO linear zero-forcing beamforming (LZFB) downlink scheme with user selection and scheduling based on the available CSIT at the base-stations, as in [5]–[8].

##### A. Linear zero-forcing beamforming

For brevity, let  $N_t = CM$  denote the number of jointly coordinated antennas and consider a reference cluster  $\mathcal{C}^* = \{c_0, \dots, c_{C-1}\}$  active in subchannel  $f$ . We denote by  $\mathcal{K}_{\mathcal{C}^*}$  the set of users  $k$  potentially served by the reference cluster on subchannel  $f$ , i.e., such that  $\mathcal{C}_{(k,f)} = \mathcal{C}^*$ . For  $k \in \mathcal{K}_{\mathcal{C}^*}$ , we define the composite channel vectors

$$\underline{\mathbf{h}}_{k,\mathcal{C}^*}[f] = [\mathbf{h}_{k,c_0}^\top[f], \mathbf{h}_{k,c_1}^\top[f], \dots, \mathbf{h}_{k,c_{C-1}}^\top[f]]^\top \quad (6)$$

of length  $N_t$ , and let  $\hat{\underline{\mathbf{h}}}_{k,\mathcal{C}^*}[f]$  denote its MMSE estimate, obtained by stacking the MMSE estimates  $\hat{\mathbf{h}}_{k,c}[f]$  of the individual subvectors in the same order. Furthermore, define the channel matrices  $\underline{\mathbf{H}}_{\mathcal{C}^*}[f]$  and  $\hat{\underline{\mathbf{H}}}_{\mathcal{C}^*}[f]$  collecting all the composite channel vectors arranged by columns,  $\forall k \in \mathcal{K}_{\mathcal{C}^*}$ .

Define the set of active users  $\mathcal{S}_{\mathcal{C}^*}[f] \subseteq \mathcal{K}_{\mathcal{C}^*}$  in cluster  $\mathcal{C}^*$  as those users who are served with positive powers by the cluster on subchannel  $f$ . Note, this selection of users changes with time, but for simplicity we drop the dependence on time in the notation. For LZFB, it must be  $|\mathcal{S}_{\mathcal{C}^*}[f]| \leq \min\{N_t, |\mathcal{K}|\}$ . We focus on the case  $M \leq K/B$  so that the number of active users per cluster is limited by the number of antennas  $N_t$ . We define the matrices  $\underline{\mathbf{H}}_{\mathcal{S}_{\mathcal{C}^*}[f]}[f]$  and  $\hat{\underline{\mathbf{H}}}_{\mathcal{S}_{\mathcal{C}^*}[f]}[f]$  as the submatrices of  $\underline{\mathbf{H}}_{\mathcal{C}^*}[f]$  and  $\hat{\underline{\mathbf{H}}}_{\mathcal{C}^*}[f]$ , respectively, restricted to the columns in the active user set  $\mathcal{S}_{\mathcal{C}^*}[f] = \{k_0, k_1, \dots, k_{|\mathcal{S}_{\mathcal{C}^*}[f]|-1}\}$ . Define the Moore-Penrose pseudo-inverse of  $\hat{\underline{\mathbf{H}}}_{\mathcal{S}_{\mathcal{C}^*}[f]}[f]$  as

$$\hat{\underline{\mathbf{H}}}_{\mathcal{S}_{\mathcal{C}^*}[f]}^+[f] = \hat{\underline{\mathbf{H}}}_{\mathcal{S}_{\mathcal{C}^*}[f]}[f] \left( \hat{\underline{\mathbf{H}}}_{\mathcal{S}_{\mathcal{C}^*}[f]}[f] \hat{\underline{\mathbf{H}}}_{\mathcal{S}_{\mathcal{C}^*}[f]}[f] \right)^{-1} \quad (7)$$

With these definitions, the transmit signal from the cluster is given by

$$\mathbf{X}_{\mathcal{C}^*}[f] = \mathbf{U}_{\mathcal{C}^*}[f] \mathbf{V}_{\mathcal{C}^*}^H[f]$$

where  $\mathbf{X}_{\mathcal{C}^*}[f] \in \mathbb{C}^{(L-\beta QM) \times M}$  and where  $\mathbf{U}_{\mathcal{C}^*}[f] \in \mathbb{C}^{(L-\beta QM) \times |\mathcal{S}_{\mathcal{C}^*}[f]|}$  is a matrix of complex coded modulation symbols containing the user codewords arranged by columns. The precoding matrix  $\mathbf{V}_{\mathcal{C}^*}[f] \in \mathbb{C}^{N_t \times |\mathcal{S}_{\mathcal{C}^*}[f]|}$  is formed by the unit-norm columns obtained by normalizing the columns of  $\hat{\underline{\mathbf{H}}}_{\mathcal{S}_{\mathcal{C}^*}[f]}^+[f]$ .

Each active user  $k \in \mathcal{S}_{\mathcal{C}^*}[f]$  is allocated a power  $P_k[f] > 0$ . Define  $\mathbf{V}_{c_i}[f] \in \mathbb{C}^{M \times |\mathcal{S}_{\mathcal{C}^*}[f]|}$  as the submatrix of  $\mathbf{V}_{\mathcal{C}^*}[f]$  containing the  $M$  rows corresponding to base-station  $c_i \in \mathcal{C}^*$ . The transmit power constraint is given by

$$\frac{1}{L-\beta QM} \text{tr} \left( \mathbb{E} [\mathbf{U}_{\mathcal{C}^*}[f] \mathbf{V}_{c_i}^H[f] \mathbf{V}_{c_i}[f] \mathbf{U}_{\mathcal{C}^*}^H[f]] \right) \leq p_{c_i}[f] \quad (8)$$

It is assumed that user codes are generated according to an i.i.d. Gaussian distribution  $\sim \mathcal{CN}(0, P_k[f])$ . For  $k_j \in \mathcal{S}_{\mathcal{C}^*}[f]$ , let  $\mathbf{v}_{k_j, c_i}[f]$  be the  $j^{\text{th}}$  column of  $\mathbf{V}_{c_i}[f]$ . Then, (8) yields<sup>3</sup>

$$\sum_{k \in \mathcal{S}_{\mathcal{C}^*}[f]} P_k[f] \|\mathbf{v}_{k, c_i}[f]\|^2 \leq p_{c_i}[f] \quad (9)$$

Letting  $u_{k,\mathcal{C}^*}[f]$  denote a generic element of the  $k$ -th column of  $\mathbf{U}_{\mathcal{C}^*}[f]$ , and letting  $\mathbf{x}_{\mathcal{C}^*}^H[f]$  denote a generic row of the transmitted signal matrix  $\mathbf{X}_{\mathcal{C}^*}[f]$ , the received signal for  $k \in \mathcal{S}_{\mathcal{C}^*}[f]$  on a generic channel use of the data phase is given by

$$\begin{aligned} y_k[f] = & \underbrace{(\mathbf{v}_{k,\mathcal{C}^*}^H[f] \underline{\mathbf{h}}_{k,\mathcal{C}^*}[f]) u_{k,\mathcal{C}^*}[f]}_{\text{desired}} + \underbrace{\sum_{c' \neq \mathcal{C}^*} \mathbf{x}_{c'}^H[f] \underline{\mathbf{h}}_{k,c'}[f]}_{\text{inter-cell interference}} \\ & + \underbrace{\sum_{j \in \mathcal{S}_{\mathcal{C}^*}[f]: j \neq k} (\mathbf{v}_{j,\mathcal{C}^*}^H[f] \underline{\mathbf{h}}_{k,\mathcal{C}^*}[f]) u_{j,\mathcal{C}^*}[f]}_{\text{intra-cluster interference}} + z_k[f] \quad (10) \end{aligned}$$

If the CSIT was perfect, the intra-cluster interference term in (10) would be zero given ideal zero-forcing. With non-perfect CSIT, the intra-cluster interference is generally non-zero.

##### B. Scheduling and rate allocation

The active user subset in each cluster is chosen on every scheduling slot according to the Proportional Fair Scheduling (PFS) rule [9]. The scheduler updates the average throughput  $\mathcal{T}[t]$  of user  $k$  according to the well-known rule

$$\mathcal{T}_k[t] = \left( 1 - \frac{1}{t_c} \right) \mathcal{T}_k[t-1] + \frac{1}{t_c F} \sum_{f=0}^{F-1} R_k[t, f]$$

where  $t$  denotes the scheduling slot index and  $R_k[t, f]$  is the *service rate* of user  $k$  on subchannel  $f$  at slot  $t$ . The problem of downlink scheduling with imperfect CSIT is addressed in [19] where the following scheduling rule is studied: the scheduler at each time  $t$  selects the active user subset  $\mathcal{S}_{\mathcal{C}^*}[f]$  and the

<sup>3</sup>We remark that while the Moore-Penrose pseudo-inverse is the optimal LZFB precoder subject to the per-cluster (sum) power constraint, it is not generally optimal for per-base station power constraint [17]. The performance of the cluster system can be improved by using the optimal LZFB solution subject to the ‘‘per-BS power constraint’’ problem [18].

corresponding transmit powers  $\{P_k[f]\}$  in order to maximize the conditional mean weighted sum rate

$$\sum_{k \in \mathcal{S}_{\mathcal{C}^*}[f]} W_k[t] \mathbb{E} \left[ \mathcal{R}_k(\mathbf{V}_{\mathcal{C}^*}[f], \mathbf{H}_{\mathcal{C}^*}[f], \{P_k[f]\}, \chi_k[f]) \mid \widehat{\mathbf{H}}_{\mathcal{C}^*}[f] \right] \quad (11)$$

where  $W_k[t] = 1/\mathcal{T}_k[t]$  and where  $\mathcal{R}_k(\cdot, \cdot, \cdot, \cdot)$  is the rate user  $k$  on time-frequency slot  $[t, f]$ , and is a function of the beamforming matrix  $\mathbf{V}_{\mathcal{C}^*}[f]$  (which in turns depends on  $\widehat{\mathbf{H}}_{\mathcal{C}^*}[f]$  and on  $\mathcal{S}_{\mathcal{C}^*}[f]$ ), of the channel matrix  $\widehat{\mathbf{H}}_{\mathcal{C}^*}[f]$ , of the power allocation  $\{P_k[f]\}$  and of the total instantaneous ICI power coming from all clusters  $\mathcal{C}' \neq \mathcal{C}^*$ .<sup>4</sup> Notice also that the scheduler weights are independent of  $f$ , i.e., the same set of weights is used across all frequency subchannels. Downlink scheduling across the frequency subchannels was shown to provide significant improvements with respect to independent scheduling on each subchannel, especially for the expanded cellular and for the overlapped cluster systems [6].

From (10), the rate function for active user  $k$  is given by

$$\mathcal{R}_k(\mathbf{V}_{\mathcal{C}^*}[f], \mathbf{H}_{\mathcal{C}^*}[f], \{P_k[f]\}, \chi_k[f]) = \log \left( 1 + \frac{|\mathbf{v}_{k,\mathcal{C}^*}^H[f] \mathbf{h}_{k,\mathcal{C}^*}[f]|^2 P_k[f]}{1 + \chi_k[f] + \sum_{j \in \mathcal{S}_{\mathcal{C}^*}[f]: j \neq k} |\mathbf{v}_{j,\mathcal{C}^*}^H[f] \mathbf{h}_{j,\mathcal{C}^*}[f]|^2 P_j[f]} \right) \quad (12)$$

where

$$\chi_k[f] = \sum_{\mathcal{C}' \neq \mathcal{C}^*} \mathbf{h}_{k,\mathcal{C}'}[f]^H \mathbf{\Sigma}_{\mathcal{C}'}[f] \mathbf{h}_{k,\mathcal{C}'}[f] \quad (13)$$

is the instantaneous ICI power at user  $k$  receiver coming from the set of clusters  $\{\mathcal{C}' : \mathcal{C}' \neq \mathcal{C}^*\}$ , and where  $\mathbf{\Sigma}_{\mathcal{C}'}[f] = \text{cov}(\mathbf{x}_{\mathcal{C}'}^H[f])$  is the transmit covariance matrix of cluster  $\mathcal{C}'$ .

The conditional expectation in (11) is hard if not impossible to compute in closed form. Furthermore, the resulting maximization is a complicated non-convex problem, involving a combinatorial search over all possible active user sets  $\mathcal{S}_{\mathcal{C}^*}[f]$  and, for each given set, a non-convex maximization due to the presence of the powers  $P_j[f]$  at the denominator inside the log function. Therefore, we consider the following low-complexity approximation, which results in a very simple scheduling rule.

Let  $\mathbf{e}_{k,\mathcal{C}^*}[f]$  denote the estimation MMSE vector for the composite vector channel  $\mathbf{h}_{k,\mathcal{C}^*}[f]$ , with block diagonal covariance matrix

$$\text{cov}(\mathbf{e}_{k,\mathcal{C}^*}[f]) = \text{diag}(\epsilon_{k,c_0}[f] \mathbf{I}, \dots, \epsilon_{k,c_{C-1}}[f] \mathbf{I})$$

(each block is  $M \times M$ ), we can include the self-interference due to the non-perfect CSIT into the intra-cell interference and write (10) as

$$y_k[f] = (\mathbf{v}_{k,\mathcal{C}^*}^H[f] \widehat{\mathbf{h}}_{k,\mathcal{C}^*}[f]) u_{k,\mathcal{C}^*}[f] + \sum_{\mathcal{C}' \neq \mathcal{C}^*} \mathbf{x}_{\mathcal{C}'}^H[f] \mathbf{h}_{k,\mathcal{C}'}[f] + \sum_{j \in \mathcal{S}_{\mathcal{C}^*}[f]} (\mathbf{v}_{j,\mathcal{C}^*}^H[f] \mathbf{e}_{k,\mathcal{C}^*}[f]) u_{j,\mathcal{C}^*}[f] + z_k[f] \quad (14)$$

Using standard bounding techniques (see for example [20] and references therein), we can lower bound the mutual

<sup>4</sup>All these quantities are time-dependent and change from slot to slot. However, for notation consistency we omit time index.

information (12) by the quantity

$$\log \left( 1 + \frac{|\mathbf{v}_{k,\mathcal{C}^*}^H[f] \widehat{\mathbf{h}}_{k,\mathcal{C}^*}[f]|^2 P_k[f]}{1 + \mathbb{E}[\xi_k[f] \mid \widehat{\mathbf{H}}_{\mathcal{C}^*}[f]] + \mathbb{E}[\chi_k[f] \mid \widehat{\mathbf{H}}_{\mathcal{C}^*}[f]]} \right) \quad (15)$$

where  $\xi_k[f] = \sum_{j \in \mathcal{S}_{\mathcal{C}^*}[f]} |\mathbf{v}_{j,\mathcal{C}^*}^H[f] \mathbf{e}_{k,\mathcal{C}^*}[f]|^2 P_j[f]$  is the instantaneous intra-cluster interference power. Let's consider the terms in the denominator of (15) separately. For the intra-cluster interference, using the block-diagonal structure of the covariance of the MMSE error vector and the fact that  $\mathbf{e}_{k,\mathcal{C}^*}[f]$  is independent of  $\widehat{\mathbf{H}}_{\mathcal{C}^*}[f]$ , we obtain

$$\mathbb{E}[\xi_k[f] \mid \widehat{\mathbf{H}}_{\mathcal{C}^*}[f]] = \sum_{c \in \mathcal{C}} \epsilon_{k,c}[f] \left( \sum_{j \in \mathcal{S}_{\mathcal{C}^*}[f]} \|\mathbf{v}_{j,c}[f]\|^2 P_j[f] \right)$$

Notice that the term  $\sum_{j \in \mathcal{S}_{\mathcal{C}^*}[f]} \|\mathbf{v}_{j,c}[f]\|^2 P_j[f]$  is the transmit total power from base station  $c \in \mathcal{C}^*$ . Under a strict per-base station constraint, this is not larger than  $p_c[f]$ . Therefore, the intra-cluster interf. terms is upperbounded by  $\sum_{c \in \mathcal{C}^*} \epsilon_{k,c}[f] p_c[f]$ .

For the ICI term, neglecting the conditioning in the expectation, we can use the approximation

$$\mathbb{E}[\chi_k[f] \mid \widehat{\mathbf{H}}_{\mathcal{C}^*}[f]] \approx \sum_{b \notin \mathcal{C}^*} g_{k,b} p_b[f] \quad (16)$$

Using the above simplifications, the proposed (suboptimal) scheduling rule finds the active user set and the power allocation that maximizes the approximated weighted rate sum

$$\sum_{k \in \mathcal{S}_{\mathcal{C}^*}[f]} W_k[t] \log \left( 1 + \frac{|\mathbf{v}_{k,\mathcal{C}^*}^H[f] \widehat{\mathbf{h}}_{k,\mathcal{C}^*}[f]|^2 P_k[f]}{1 + \sum_{c \in \mathcal{C}^*} \epsilon_{k,c}[f] p_c[f] + \sum_{b \notin \mathcal{C}^*} g_{k,b} p_b[f]} \right) \quad (17)$$

subject to the base-station power constraints (9) for  $c_i \in \mathcal{C}^*$ . The maximization of (17) has the same form of the user selection and power allocation optimization with LZFB and perfect CSIT, after replacing the actual channel matrix by  $\widehat{\mathbf{H}}_{\mathcal{C}^*}[f]$  and augmenting the noise variance at each given user  $k$  by the corresponding intra and inter cell interference terms. Therefore, the optimization of (17) can be easily obtained by using efficient user selection algorithms as in [21], [22].

Finally, we have to specify the actual user service rates  $R_k[t, f]$ . We assume that after  $\mathcal{S}_{\mathcal{C}^*}[f]$  and  $\{P_k[f]\}$  have been optimized as explained above, given knowledge of estimated channels, that the users  $k \in \mathcal{S}_{\mathcal{C}^*}[f]$  achieve on the current slot an "optimistic" rate equal to the instantaneous mutual information (12), even though this rate is actually not known to the scheduler. In [23], [24] we showed that this "optimistic" user rates can be approached by combining scheduling with incremental redundancy hybrid ARQ.

## V. SIMULATION RESULTS

It is instructive to first revisit some simulation results from our previous work [8]. Here simulations use  $B = 60$  stations with  $K/B = 36$  users per cell, uniformly and symmetrically

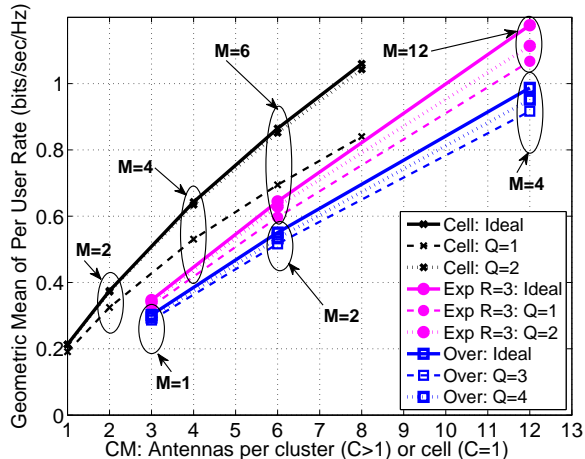


Fig. 4. Geometric Mean of Per-User Rate under the limiting assumption  $\eta \rightarrow 1$ , i.e., zero effective overhead by pilots.  $\beta = 1$  in all cases.

placed in each cell. Expanded cellular and overlapped cluster systems use  $C = R = F = 3$ . The pathloss parameters are  $G_0 = 10^6$ ,  $\kappa = 3.76$  and  $\delta = 0.05$ . We consider user rates  $r_k$ , where  $r_k$  is the average value of the “optimistic” rates (12) over all slots and subchannels achieved by simulation [8].

Fig. 4 shows the tradeoff in the geometric mean (what PFS maximizes) of  $r_k$  vs.  $CM$  under various pilot overheads under a limiting assumption that  $\eta = (1 - \beta QM/L) \rightarrow 1$ , i.e., effectively zero overhead as  $L \rightarrow \infty$  for finite  $\beta QM$ . “Ideal” cases assume perfect (genie aided) CSIT knowledge. “Cell” and “Exp” denote conventional and expanded cellular, and “Over” denotes the overlapped cluster architecture. Note, the x-axis  $CM$  represents the minimum value of  $\beta QM$  since  $Q \geq C$  and  $\beta \geq 1$  for each system.

The plot demonstrates that cellular systems outperform NW-MIMO systems as a function of pilot overhead. For example, an expanded cellular system with  $M = 12$  and  $Q = 1$  and the NW-MIMO system with  $M = 4$ ,  $C = 3$ , and  $Q = 3$  have the same pilot overhead  $\beta QM = 12$ . Here  $Q = 1$  provides effective estimation for the expanded cellular while  $Q = 3$  is the minimum required by the  $C = 3$  NW-MIMO system.

It is also important to note that on each subchannel in NW-MIMO the use of  $Q = 3$ , and even  $Q = 4$  or  $5$ , only provides cluster-center users with accurate CSIT with respect to all  $C = 3$  stations of the corresponding cluster. Many cluster-edge users are effectively only obtaining CSIT from the two nearest stations in their cluster on that subchannel [8]. The close performance of  $Q = 3$  and  $Q = 4$  cases to the “Ideal” case for NW-MIMO demonstrates the more relaxed CSIT requirements for some user locations in an overlapped NW-MIMO architecture.

The effect of training overhead becomes apparent in Fig. 5 which shows the case of  $1/L = 0.045$ , i.e., where each pilot takes 4.5% of the transmission resources. The plot shows the geometric mean of the net rate  $\eta r_k$  vs  $\beta QM$ . Here a conventional cellular architecture with  $M = 4$  and  $Q = 1$  exceeds the performance of a cluster system with  $M = 3$  and  $Q = 3$ . The plot also demonstrates that reducing errors in

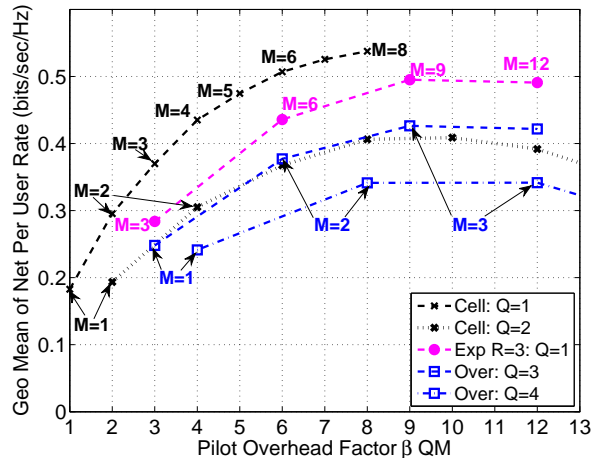


Fig. 5. Geometric Mean of Net Per-User Rates with  $1/L = 0.045$ .  $\beta = 1$  in all cases.

CSIT estimates by increasing  $Q$  to 2 for cellular and to 4 for cluster systems (despite the fact many users are not getting accurate CSIT from all stations) results in a net overall loss in rate. As  $\eta \rightarrow 1$  user rates approaches those of Figure 4.

## VI. SYSTEM CAPACITY APPROXIMATED ANALYSIS

A major difficulty of any theoretical analysis is represented by the fact that, in the presence of user-scheduling, subsets of users are selected and therefore the joint statistics of their channel vectors, after selection, is neither i.i.d. nor Gaussian. On the other hand, the results of [25] show that when both the number of antennas and the number of users grow large, the probability of selecting a set of channel vectors that are quasi-orthogonal, i.e., that significantly differ from random i.i.d., vanishes to zero. These asymptotic results suggest that opportunistic user selection has a less and less significant impact as the system dimensions become large.

Motivated by this consideration, we consider a simplified model where the scheduler selects  $S \leq C \min\{K/B, M\}$  users at random, with uniform probability, to be scheduled on each sub-channel  $f$ . The number of selected users per cluster,  $S$ , is the same for all clusters. As before, we let  $\mathcal{S}_{C^*}[f]$  denote the subset of selected users for a reference cluster  $C^*$ . Finally, we simplify the power allocation problem and consider equal power for all selected users, i.e., we have  $P_k[f] = \frac{1}{SC} \sum_{c \in C^*} p_c[f]$  for all  $k \in \mathcal{S}_c[f]$ . We denote by  $\bar{\mathcal{B}}[f]$  the set of all base-stations which are active on frequency subchannel  $f$ , and by  $\mathcal{B}[c, f]$  the same set excluding  $c$ .

### A. Cellular and Expanded Cellular Systems

We consider an expanded cellular system with  $R \geq 1$ . The results of the conventional cellular system are obtained by letting  $R = 1$  and particularizing definitions of  $\mathcal{B}[f]$  and  $p_c[f]$ . In the expanded cellular system, we consider  $R$  adjacent cells, denoted by  $C^* = \{c_0, \dots, c_{R-1}\}$ , such that cell  $c_f$  is active on subchannel  $f$  for  $0 \leq f \leq R - 1$ . As described in Section II-B, each station covers a symmetric interval of size  $R$  centered around the stations position, and wrapped modulo to  $[0, B)$ . Thus a station serves  $KR/B$  users on subchannel

$f$ . We focus on the set of users served by station  $c_f$  in band  $f$ , denoted by  $\mathcal{K}_{c_f} = \{k : \mathcal{C}(k, f) = c_f\}$ . Station  $c_f$  picks  $S \leq \min\{KR/B, M\}$  users in its extended coverage to be served simultaneously, at random and with possibly different probabilities corresponding to user-specific activity factors to be chosen later. When active, stations transmit at a power  $p_{c_f}[f] = R$ .

Assuming uniform power allocation  $R/S$  to each scheduled user we can obtain the following conditional mutual information lower bound for user  $k \in \mathcal{K}_{c_f}$  (given the CSIT available to base station  $c_f$ ):

$$I_{k,c_f}^{\text{lb}} = I_{k,c_f}^{\text{lb}} \left( \widehat{\mathbf{H}}_{c_f}[f], \{\mathbf{h}_{k,b}[f] : b \in \mathcal{B}[c_f, f]\} \right) = \quad (18)$$

$$\log \left( 1 + \frac{|\mathbf{v}_{k,c_f}^{\text{H}}[f] \widehat{\mathbf{h}}_{k,c_f}[f]|^2 R/S}{1 + R\epsilon_{k,c_f}[f] + \mathbb{E}[\chi_k[f] | \widehat{\mathbf{H}}_{c_f}[f]]} \right)$$

where  $\chi_k[f]$  is given by (13) is the ICI coming from all cells  $b \in \mathcal{B}[c_f, f]$ , and where  $\text{tr}(\mathbf{\Sigma}_b[f]) = R$ . As shown in Appendix A, by using the correlation structure between  $\mathbf{h}_{k,b}[f]$  and  $\widehat{\mathbf{H}}_{c_f}[f]$  this mutual information bound is a random variable distributed as

$$I_{k,c_f}^{\text{lb}}(\mathbf{t}) = \log(1 + \tau_k[f] \|\mathbf{t}_1\|^2 + \rho_k[f] \|\mathbf{t}_2\|^2) - \log(1 + \rho_k[f] \|\mathbf{t}\|^2) \quad (19)$$

where  $\mathbf{t}_1 \sim \mathcal{CN}(0, \mathbf{I}_{M-S+1})$ ,  $\mathbf{t}_2 \sim \mathcal{CN}(0, \mathbf{I}_{S-1})$ ,  $\mathbf{t} = [\mathbf{t}_1^T \ \mathbf{t}_2^T]^T \sim \mathcal{CN}(0, \mathbf{I}_M)$ , and where the user-dependent parameters  $\rho_k[f]$  and  $\tau_k[f]$  are given by

$$\rho_k[f] = \frac{\frac{1}{M} \sum_{\ell \neq 0} g_{k,c_f+RQ\ell}^2}{(R^{-1} + \sum_{\ell} g_{k,c_f+R\ell})(a^{-1} + \sum_{\ell} g_{k,c_f+RQ\ell}) - \sum_{\ell} g_{k,c_f+RQ\ell}^2} \quad (20)$$

and

$$\tau_k[f] = \frac{\frac{1}{S} g_{k,c_f}^2 + \frac{1}{M} \sum_{\ell \neq 0} g_{k,c_f+RQ\ell}^2}{(R^{-1} + \sum_{\ell} g_{k,c_f+R\ell})(a^{-1} + \sum_{\ell} g_{k,c_f+RQ\ell}) - \sum_{\ell} g_{k,c_f+RQ\ell}^2} \quad (21)$$

where  $a = \beta RQ$ , (with the understanding that all sums are restricted over indices in the range  $\{0, 1, \dots, B-1\}$  and that  $c_f$  denotes the base-station that is closer to user  $k$  among the base stations active on channel  $f$ ).

Notice that  $I_{k,c_f}^{\text{lb}}$  in (18) is a function of the instantaneous CSIT available at base station  $c_f$  as well as of the random channel vectors from all interfering base stations  $b \neq c_f$  to the user  $k$  receiver. The scheduler wishes to optimize some suitable concave and componentwise non-decreasing network utility function  $\Gamma(\overline{\mathbf{R}})$ , function of the users' long-term average (or ergodic) rates. To this purpose, the scheduler may choose users with different "activity factors", that is, some users may be scheduled more or less frequently than others. Suppose that user  $k \in \mathcal{K}_{c_f}$  is scheduled with activity factor  $q_k[f]$ , where the vector of utility factors  $\mathbf{q}[f]$  is subject to the constraint  $\sum_{k \in \mathcal{K}_c} q_k[f] = S$ ,  $q_k \geq 0$  for all  $k$  and all  $f$ . By ergodicity,

the long-term average rate of user  $k$  is lower-bounded by

$$\lim_{t \rightarrow \infty} \frac{1}{t} \sum_{\tau=1}^t R_k[\tau, f] \geq q_k[f] \overline{R}_k^{\text{lb}}[f] \quad (22)$$

where  $\overline{R}_k^{\text{lb}}[f] = \mathbb{E} \left[ I_{k,c_f}^{\text{lb}}(\mathbf{t}) \right]$

The set of values  $\{\overline{R}_k^{\text{lb}}[f]\}$  can be evaluated for a reference station,  $c^*$ , active on a reference subchannel  $f^*$ . Users are located on a fixed symmetric grid on the coverage and thus for every active user in any cell there is an equivalent user in  $c^*$  on  $f^*$  having the same long-term average rate. Specifically, these are users with the same relative location with respect to their active serving base-station. Evaluation of this reference station can therefore be used to define an array  $\overline{\mathbf{R}}$  of  $F \times KR/B$  rate lower bound values such that the  $(f, k)$  element of this array, denoted by  $\overline{R}_k^{\text{lb}}[f]$ , is the rate lower bound for user  $k \in \mathcal{K}_{c_0}$  on subchannel  $f$  (with  $0 \leq f \leq F-1$ ). The average total rate (over all frequency subchannels) for user  $k$  is therefore lower bounded as  $\overline{R}_k \geq \frac{\eta}{F} \sum_{f=0}^{F-1} q_k[f] \overline{R}_k^{\text{lb}}[f]$ . In order to take into account also the overhead due to CSIT training, the rates must be weighted by the efficiency factor  $\eta = (1 - \beta QM/L)$ . For PFS, the utility function is  $\Gamma(\{R_k\}) = \sum_k \log R_k$ , resulting in the following network utility function maximization problem

$$\text{maximize} \quad \sum_k \log \left( \eta \frac{1}{F} \sum_{f=0}^{F-1} q_k[f] \overline{R}_k^{\text{lb}}[f] \right) \quad (23)$$

$$\text{subject to} \quad \sum_{k \in \mathcal{K}_{c_0}} q_k[f] \leq S, \quad q_k[f] \geq 0, \quad \forall f = 0, \dots, F-1$$

The associated expressions for the conventional cellular case are readily obtained by specializing all the expressions in this section to the case  $R = 1$ .

## B. Overlapped Cluster

The analysis in the overlapped cluster case poses additional challenges. Unlike the expanded cellular case, the composite channel vectors  $\mathbf{h}_{k,C}[f]$  defined in (6) are not isotropic, since as they are formed by  $C$  vectors of length  $M$  (one per base station in the cluster) each with path-gain specific (i.e., different) power. Hence, assuming  $C > 1$ , it is not enough to assume that the cluster picks a set of  $S$  users at random throughout the cluster, since this would not happen under the actual user selection algorithm. For example, it is very unlikely that the cluster selects  $S$  users all belonging to the same cell; if that were the case (and  $S > M$ ) the resulting channel matrix would be almost rank-deficient (all the large-variance coefficients are confined in a submatrix of size  $M \times S$ ).

Here we consider a simplified user selection rule that selects  $S/C$  users in each cell in the cluster, but inside each cell, this selection is random with possibly different probabilities which determine the user activity factors. Again, we choose  $S \leq C \min\{K/B, M\}$ . This assumption imposes a given set of joint statistics of the composite channels  $\mathbf{H}_{C^*}[f]$  for a given cluster  $C^*$  active on subchannel  $f$ . Strictly speaking, the joint channel statistics depend on the probability with which users

are chosen. We neglect this dependency here and optimize the user activity factors separately, as if they had no influence on the joint channel statistics. We also assume without loss of generality  $C = F$ .

Assuming uniform power allocation  $C/S$  for each active user, we have the conditional mutual information lower bound

$$I_{k,C^*}^{\text{lb}}(\widehat{\mathbf{H}}_{C^*}[f], \{\mathbf{h}_{k,C}[f] : C \neq C^*\}) = \log \left( 1 + \frac{|\mathbf{v}_{k,C^*}^{\text{H}}[f] \widehat{\mathbf{h}}_{k,C^*}[f]|^2 C/S}{1 + \sum_{c \in C^*} \epsilon_{k,c}[f] + \mathbb{E}[\chi_k[f] | \widehat{\mathbf{H}}_{C^*}[f]]} \right) \quad (24)$$

where  $\chi_k[f]$  is the ICI in (13), coming from clusters  $C' \neq C^*$ , and where  $\text{tr}(\mathbf{\Sigma}_C[f]) = C$ .

Similarly to the expanded cellular case we have the long-term average rate bound (22) where

$$\overline{R}_k^{\text{lb}}[f] = \mathbb{E} \left[ I_{k,C^*}^{\text{lb}}(\widehat{\mathbf{H}}_{C^*}[f], \{\mathbf{h}_{k,C}[f] : C \neq C^*\}) \right]$$

Since the users are on a fixed symmetric grid, the average rates that any user  $k$  gets on any subchannel  $f = 0, \dots, C-1$  can be obtained by appropriately mapping users-rates determined for users on the reference cluster  $C^*$  active on reference subchannel  $f = 0$  to equivalent users on other clusters and subchannels. Such equivalent users have the same relative position with respect to their serving active cluster. Exploiting this symmetry, we can obtain an array  $\overline{\mathbf{R}}^{\text{lb}}$  of  $C \times KC/B$  rate lower bound values such that the  $(f, k)$  element of this array, denoted by  $\overline{R}_k^{\text{lb}}[f]$ , is the rate lower bound for user  $k \in \mathcal{K}_{C^*}$  on subchannel  $f$ . The average total rate (over all frequency subchannels) for user  $k$  is therefore lower bounded as  $\overline{R}_k \geq \frac{\eta}{F} \sum_{f=0}^{F-1} q_k[f] \overline{R}_k^{\text{lb}}[f]$  and the network utility maximization to be solved is takes on the same form as in (23).

Finally, we obtain a simplified expression for the average mutual information lower bound, by exploiting Jensen's inequality. Taking expectation of both sides of (24) we have

$$\begin{aligned} \overline{R}_k^{\text{lb}}[0] &= \mathbb{E} \left[ \log \left( 1 + \frac{|\mathbf{v}_{k,C^*}^{\text{H}}[0] \widehat{\mathbf{h}}_{k,C^*}[0]|^2 C/S}{1 + \sum_{c \in C^*} \epsilon_{k,c}[0] + \mathbb{E}[\chi_k[0] | \widehat{\mathbf{H}}_{C^*}[0]]} \right) \right] \\ &\geq \mathbb{E} \left[ \log \left( 1 + \sum_{c \in C^*} \epsilon_{k,c}[0] + \frac{|\mathbf{v}_{k,C^*}^{\text{H}}[0] \widehat{\mathbf{h}}_{k,C^*}[0]|^2 C}{S} \right) \right] \\ &\quad - \log \left( 1 + \sum_{c \in C^*} \epsilon_{k,c}[0] + \sum_{b \notin C^*} g_{k,b} \right) \quad (25) \end{aligned}$$

Note that the expectation in the bounding expression in (25) is evaluated by Monte Carlo simulation, due to the fact that  $\widehat{\mathbf{h}}_{k,C^*}[0]$ , and therefore  $\mathbf{v}_{k,C^*}^{\text{H}}[0]$ , have non-isotropic statistics.

## VII. RESULTS FROM ERGODIC ANALYSIS

We repeat the scenarios described in Section V. Figure 6 shows the geometric mean of the bounds on ergodic per-user rates for various architectures plotted against  $CM$  as done for figure 4. The rates for all systems are lower, as expected since the ergodic analysis does show lowerbound rate

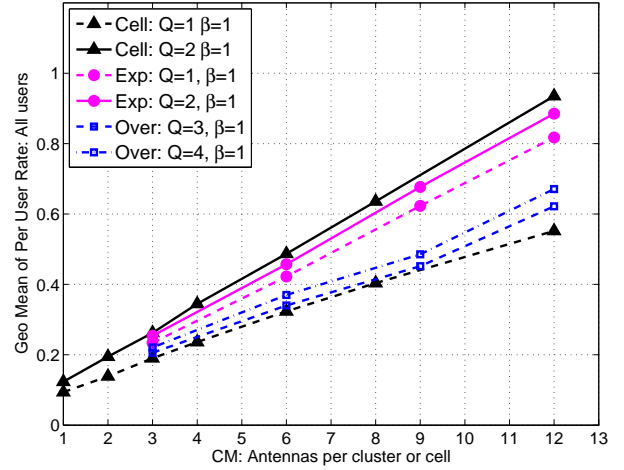


Fig. 6. Geometric Mean of Ergodic Per-User Rates under the limiting assumption  $\eta \rightarrow 1$

and ignores benefits from user-scheduling. The comparison between expanded cellular and overlapped cluster systems is similar, with the expanded cellular case  $Q = 1$  exceeding the cluster architecture requiring a larger pilot reuse of  $Q \geq 3$ .

The conventional cellular case in Figure 6 does not predict the advantages shown in the simulations in Figure 4. One reason for this is that the lower bounding of the ergodic rates is looser for cluster-edge users with high ICI, thereby giving them very low rate bounds. For the conventional cellular case this has a strong effect on the overall geometric mean since any cell-edge user suffers from this effect on all sub-channels. For expanded cellular and cluster cases, cell-edge users have favorable ICI conditions (and rate bounds) on at least one of the sub-channels. The activity factor optimization  $\{q_k[f]\}$  guarantees that cell-edge users are scheduled less (on not at all) in clusters where they are cluster-edge users.<sup>5</sup> This makes the expanded cellular and cluster architectures less sensitive to cluster-edge effects.

To eliminate this edge-cluster effect, we next compare the three architectures by dropping from each cell the 2 edge users from each cell side. Figure 7 shows for each architecture the (geometric mean) rate performance based on the rates of the remaining 32 cell users per cell (top 89-percentile of users). The ‘‘architecture’’ trends in Figure 7 match more closely those in Figure 4, revealing that, if properly adapted, the theoretical analysis can also be applied to cases of  $\eta < 1$  and predict the trends of the full-scale system simulations.

## VIII. CLOSING REMARKS

We conducted a performance comparison of three types of coordinated and non-coordinated cellular architectures in the context of MU-MIMO downlink transmission to multiple single-antenna users per cell. Our study considers inherent physical layer issues such as path-loss, time and frequency

<sup>5</sup>In addition, the multiuser diversity effect provided in the simulations by the greedy ZFBF scheduler yields rate improvements (with respect to the performance of the random scheduler assumed in obtaining the lower-rate bounds) that are proportionally larger for edge users.

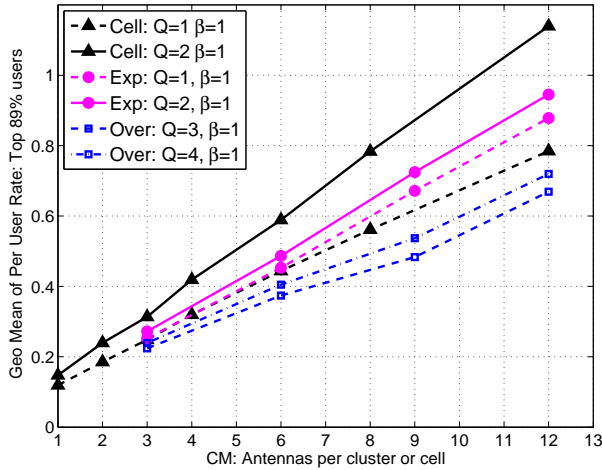


Fig. 7. Geometric Mean of Ergodic Per-User Rates ignoring CSI overhead for top 89-percentile users under the limiting assumption  $\eta \rightarrow 1$

selective fading, and the need for channel state information at the transmitter (CSIT). By taking explicit account of the inherent cost of CSIT estimation, the analysis illuminates the tradeoffs between CSIT estimation error, system resources dedicated to training, and net system performance.

The paper provides a theoretical analysis of such architectures. We show that the predictions from theory generally agree with simulation based studies [5]–[8], provided that we account for the pessimistic performance predictions for cell-edge user rates by the theory. In the context of MU-MIMO downlink, our study reveals that conventional cellular architectures can yield higher system performance than Network MIMO architectures for a given CSIT training overhead and/or by the use of a few extra antennas per station. In light of these results, it appears that the inherent bottleneck of Network MIMO systems is the intrinsic requirement in number of dimensions dedicated to estimating user channels.

## REFERENCES

- [1] 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.
- [2] A. D. Wyner, "Shannon-theoretic approach to a Gaussian cellular multiple-access channel," *IEEE Trans. on Inform. Theory*, vol. 40, pp. 1713–1727, Nov. 1994.
- [3] S. Shamai and A. D. Wyner, "Information-theoretic considerations for symmetric, cellular, multiple-access fading channels - Part I," *IEEE Trans. on Inform. Theory*, vol. 43, pp. 1877–1894, Nov. 1997.
- [4] O. Simeone, O. Simeone, Y. Bar-Ness, A. M. Haimovich, U. Spagnolini, and S. Shamai (Shitz), *An Information Theoretic View of Distributed Antenna Systems*. Auerbach Publications, CRC Press, 2007.
- [5] G. Caire, S. A. Ramprasad, H. C. Papadopoulos, C. Pepin, and C.-E. W. Sundberg, "Multiuser MIMO downlink with limited inter-cell cooperation: Approximate interference alignment in time, frequency and space," in *Proc. 46th Allerton Conf. Commun., Control and Computing*, (Monticello, IL), Oct. 2008.
- [6] S. Ramprasad, G. Caire, and H. Papadopoulos, "A joint scheduling and cell clustering scheme for MU-MIMO downlink with limited coordination," in *To be presented in ICC*, (Cape Town, South Africa), May 2010.
- [7] S. Ramprasad and G. Caire, "Cellular vs Network MIMO: A comparison including channel state information overhead," in *Proc. Personal, Indoor and Mobile Radio Communications Symposium 2009 (PIMRC'09)*, (Tokyo, Japan), Sept. 2009.

- [8] S. Ramprasad, G. Caire, and H. Papadopoulos, "Cellular and Network MIMO Architectures: MU-MIMO Spectral Efficiency and Costs of Channel State Information," in *Proc. IEEE Asilomar Conference on Signals, Systems, and Computers, ACSSC*, (Pacific Grove, CA), Nov. 2009.
- [9] P. Viswanath, D. N. C. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Trans. on Inform. Theory*, vol. 48, no. 6, pp. 1277–1294, 2002.
- [10] P. Bender, P. Black, M. Grob, R. Padovani, N. Sindhushayana, and A. Viterbi, "CDMA/HDR: a bandwidth-efficient high-speed wireless data service for nomadic users," *IEEE Commun. Mag.*, vol. 38, pp. 70–77, Jul. 2000.
- [11] T. Marzetta, "Multi-Cellular wireless with base stations employing unlimited number of antennas," in *Information Theory and Application (ITA) Workshop*, (UCSD, La Jolla, CA), Feb. 2010.
- [12] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, "Pilot Contamination Problem in Multi-Cell TDD Systems," *Arxiv preprint arXiv:0901.1703*, 2009.
- [13] 3GPP, "3GPP TR 25.996 V6.1.0: Spatial channel model for Multiple Input Multiple Output (MIMO) simulations (Release 6)," 3GPP, Sept 2003.
- [14] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, "Multiuser MIMO achievable rates with downlink training and channel state feedback," *submitted to IEEE Trans. on Inform. Theory*, Nov. 2007. Arxiv preprint <http://arxiv.org/abs/0711.2642>.
- [15] 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.
- [16] H. Shirani-Mehr and G. Caire, "Channel state feedback schemes for multiuser MIMO-OFDM downlink," *Accepted for publication in IEEE Trans on Commun.*, 2009.
- [17] 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.
- [18] H. Huh, H. 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.
- [19] H. Shirani-Mehr, G. Caire, and M. J. Neely, "MIMO Downlink Scheduling with Non-Perfect Channel State Knowledge," *submitted to IEEE Trans. on Commun.*, 2009.
- [20] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, "Quantized vs. analog feedback for the MIMO broadcast channel: a comparison between zero-forcing based achievable rates," in *Proc. IEEE Int. Symp. on Inform. Theory, ISIT*, (Nice, France), June 2007.
- [21] 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.
- [22] 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.
- [23] H. Shirani-Mehr, H. Papadopoulos, S. Ramprasad, and G. Caire, "Joint scheduling and Hybrid-AQR for MU-MIMO downlink in the presence of inter-cell interference," in *To be presented in ICC*, (Cape Town, South Africa), May 2010.
- [24] H. Shirani-Mehr, H. Papadopoulos, S. Ramprasad, and G. Caire, "Joint Scheduling and ARQ for MU-MIMO Downlink in the Presence of Inter-Cell Interference," *Under review by IEEE Trans. on Com.*, Jan 2010. Arxiv preprint <http://arxiv.org/abs/1001.1187>.
- [25] 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.

## APPENDIX

In this section we derive the expressions for the mutual-information lower bounds used in Section VI. Note that in the process we also derive the conditional user-ICI expressions for the expanded cellular system.

First, for any base-station  $c$  that is active on channel  $f$ , we derive the conditional PDF for the channel  $\mathbf{h}_{k,b}[f]$  between user  $k$  and the antennas at BS with index  $b$ , given  $\hat{\mathbf{h}}_{k,c}[f]$ , the MMSE estimate (obtained via pilot-on-pilot training) of

$\mathbf{h}_{k,c}[f]$  the channel between user  $k$  and BS with index  $c$ . Recall that  $\hat{\mathbf{h}}_{k,c}[f]$  is the MMSE estimate of  $\mathbf{h}_{k,c}[f]$  based on (3). First note that for  $b \notin \mathcal{P}(c, f)$ , the vectors  $\mathbf{h}_{k,b}[f]$  and  $\hat{\mathbf{h}}_{k,c}[f]$  are independent, and thus  $\mathbf{h}_{k,b}[f]|\hat{\mathbf{h}}_{k,c}[f] \sim \mathcal{CN}(0, g_{k,b}\mathbf{I})$ . Consider next the case  $b \in \mathcal{P}(c, f)$ . First, note that the estimate of  $\mathbf{h}_{k,b}[f]$  based on  $\mathbf{r}_{k,c}[f]$  in (3) for any  $b \in \mathcal{P}(c, f) = \mathcal{P}(c, f) \cup \{c\}$  is given by

$$\hat{\mathbf{h}}_{k,b}[f] = \frac{g_{k,b}}{a \sum_{d \in \bar{\mathcal{P}}(c,f)} g_{k,d} + 1} \mathbf{r}_{k,c}[f] = \frac{g_{k,b}}{g_{k,c}} \hat{\mathbf{h}}_{k,c}[f]$$

Also,  $\mathbf{h}_{k,b}[f] = \hat{\mathbf{h}}_{k,b}[f] + \mathbf{e}_{k,b}[f]$  with  $\hat{\mathbf{h}}_{k,b}[f]$  and  $\mathbf{e}_{k,b}[f]$  independent, and  $\mathbf{e}_{k,b}[f] \sim \mathcal{CN}(0, \epsilon_{k,b}[f])$ , with  $\epsilon_{k,b}[f]$  given by (4). Since  $\hat{\mathbf{h}}_{k,b}[f] = (g_{k,b}/g_{k,c})\hat{\mathbf{h}}_{k,c}[f]$ , we also have  $\mathbf{h}_{k,b}[f] = (g_{k,b}/g_{k,c})\hat{\mathbf{h}}_{k,c}[f] + \mathbf{e}_{k,b}[f]$ , with  $\hat{\mathbf{h}}_{k,c}[f]$  and  $\mathbf{e}_{k,b}[f]$  statistically independent. Hence we have

$$\mathbf{h}_{k,b}[f]|\hat{\mathbf{h}}_{k,c}[f] \sim \begin{cases} \mathcal{CN}\left(\frac{g_{k,b}}{g_{k,c}}\hat{\mathbf{h}}_{k,c}[f], \epsilon_{k,b}[f]\mathbf{I}\right) & \text{if } b \in \bar{\mathcal{P}}(c, f) \\ \mathcal{CN}(\mathbf{0}, g_{k,b}\mathbf{I}) & \text{if } b \notin \bar{\mathcal{P}}(c, f) \end{cases} \quad (26)$$

We next obtain an expression for the user-ICI in the expanded cellular case. Note that for all  $c \in \mathcal{B}(f)$ ,  $p_c[f] = R$ . Letting  $c_f$  denote the station serving user  $k$  in band  $f$ , we obtain an expression for  $\mathbb{E}\left[\mathbf{h}_{k,b}[f]^H \Sigma_b[f] \mathbf{h}_{k,b}[f] | \hat{\mathbf{H}}_{c_f}[f]\right]$  for  $b \neq c$ . Since base-station  $b$  is not serving on channel  $f$  any of the users served by cell  $c_f$  on channel  $f$  (including user  $k$ ), and since the channels (and their estimates) of the users served by cell  $c_f$  (in channel  $f$ ) are not exploited in the scheduler at base-station  $b$ , and these channel vectors and estimates are independent of the channels and estimates of the users served by base-station  $b$ , the transmitted signal  $\mathbf{x}_b[f]$  is independent of  $\mathbf{h}_{k,b}[f]$  and of  $\hat{\mathbf{H}}_{c_f}[f]$ . Using (26), we obtain

$$\mathbb{E}\left[\mathbf{h}_{k,b}[f]^H \Sigma_b[f] \mathbf{h}_{k,b}[f] | \hat{\mathbf{H}}_{c_f}[f]\right] = \frac{g_{k,b}^2}{g_{k,c_f}^2} \frac{R}{M} \|\hat{\mathbf{h}}_{k,c_f}[f]\|^2 + \epsilon_{k,b}[f] \quad (27)$$

if  $b \in \mathcal{P}(c_f, f)$ . This user-ICI dependency on the channel estimate is a manifestation of ‘‘pilot contamination,’’ also observed by [12] in the context of TDD-based training. Also the LHS of (27) equals  $Rg_{k,b}$  if  $b \notin \mathcal{P}(c, f)$  and  $b \neq c_f$ , and equals 0 for all  $b \notin \mathcal{B}(f)$ . We next rewrite  $\hat{\mathbf{h}}_{k,c_f}[f]$  in the form  $\hat{\mathbf{h}}_{k,c_f}[f] = \sqrt{g_{k,c_f} - \epsilon_{k,c_f}[f]} \mathbf{t}$  where  $\mathbf{t} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ . Then for all  $b \in \bar{\mathcal{P}}(c_f, f)$

$$\hat{\mathbf{h}}_b[f] = \frac{g_{k,b}}{g_{k,c_f}} \hat{\mathbf{h}}_c[f] = \sqrt{g_{k,b} - \epsilon_{k,b}[f]} \mathbf{t}$$

Summing both sides of (27) over all  $b \neq c$ , and using the preceding expression we get

$$\sum_{b \neq c_f} \mathbb{E}\left[\mathbf{h}_{k,b}[f]^H \Sigma_b[f] \mathbf{h}_{k,b}[f] | \hat{\mathbf{H}}_{c_f}[f]\right] = \sum_{b \in \mathcal{B}(f); b \neq c_f} Rg_{k,b} + R \left[ \frac{\|\mathbf{t}\|^2}{M} - 1 \right] \sum_{b \in \mathcal{P}(c_f, f)} (g_{k,b} - \epsilon_{k,b}[f]) \quad (28)$$

To derive the mutual-information lower bound expression for the expanded cellular case we start by using (28) to express

$I_{k,c}^{\text{lb}}[f]$  in (18) in the following form:

$$I_{k,c_f}^{\text{lb}} = \log \left( 1 + \frac{(g_{k,c_f} - \epsilon_{k,c_f}[f]) |\mathbf{t}^H \mathbf{v}_{k,c_f}[f]|^2 / S}{R^{-1} + \epsilon_{k,c_f}[f] + \sum_{\substack{b \in \mathcal{B}(f) \\ b \neq c_f}} g_{k,b} + \left[ \frac{\|\mathbf{t}\|^2}{M} - 1 \right] \sum_{b \in \mathcal{P}(c_f)} (g_{k,b} - \epsilon_{k,b}[f])} \right) \quad (29)$$

Writing  $\mathbf{t} = [\mathbf{t}_1^T \ \mathbf{t}_2^T]^T$  with  $\mathbf{t}_1$  being  $(M - S + 1) \times 1$ , first note that we can substitute  $\|\mathbf{t}_1\|^2$  for  $|\mathbf{t}^H \mathbf{v}_{k,c_f}[f]|^2$  in the expression above. To see this, assume without loss of generality that the nullspace of the other  $S - 1$  (scheduled) users is confined to the first  $M - S + 1$  dimensions of  $\mathbf{t}$ , i.e., in the space of  $\mathbf{t}_1$ . Then the projection of  $\mathbf{v}_{k,c_f}$  in that space is chosen for beamforming (with  $\mathbf{t}_1$ ), thereby  $|\mathbf{t}^H \mathbf{v}_{k,c_f}|^2 = \|\mathbf{t}_1\|^2$ . Hence (29) can be put in the form

$$I_{k,c_f}^{\text{lb}}[f] = \log \left( 1 + \frac{(\tau_k[f] - \rho_k[f]) \|\mathbf{t}_1\|^2 / S}{1 + \rho_k[f] \|\mathbf{t}\|^2} \right)$$

which, using  $\|\mathbf{t}\|^2 = \|\mathbf{t}_1\|^2 + \|\mathbf{t}_2\|^2$  can be expressed equivalently as in (19), and where

$$\rho_k[f] = \frac{\frac{1}{M} \sum_{b \in \mathcal{P}(c_f, f)} (g_{k,b} - \epsilon_{k,b}[f])}{R^{-1} + \sum_{b \in \mathcal{B}(f)} g_{k,b} - \sum_{b \in \bar{\mathcal{P}}(c_f, f)} (g_{k,b} - \epsilon_{k,b}[f])} \quad (30)$$

and

$$\tau_k[f] = \frac{\frac{1}{S} (g_{k,c_f} - \epsilon_{k,c_f}[f]) + \frac{1}{M} \sum_{b \in \mathcal{P}(c_f, f)} (g_{k,b} - \epsilon_{k,b}[f])}{R^{-1} + \sum_{b \in \mathcal{B}(f)} g_{k,b} - \sum_{b \in \bar{\mathcal{P}}(c_f, f)} (g_{k,b} - \epsilon_{k,b}[f])} \quad (31)$$

Note that for any  $b \in \bar{\mathcal{P}}(c_f, f)$ ,

$$g_{k,b} - \epsilon_{k,b}[f] = \frac{g_{k,b}^2}{a^{-1} + \sum_{\ell \in \bar{\mathcal{P}}(c_f, f)} g_{k,\ell}}, \quad (32)$$

which, when substituted in (30) and (31), yields

$$\rho_k[f] = \frac{\frac{1}{M} \sum_{b \in \mathcal{P}(c_f, f)} g_{k,b}^2}{\left( R^{-1} + \sum_{b \in \mathcal{B}(f)} g_{k,b} \right) \left( a^{-1} + \sum_{\ell \in \bar{\mathcal{P}}(c_f, f)} g_{k,\ell} \right) - \sum_{b \in \bar{\mathcal{P}}(c_f, f)} g_{k,b}^2} \quad (33)$$

and

$$\tau_k[f] = \frac{\frac{1}{S} g_{k,c_f}^2 + \frac{1}{M} \sum_{b \in \mathcal{P}(c_f, f)} g_{k,b}^2}{\left( R^{-1} + \sum_{b \in \mathcal{B}(f)} g_{k,b} \right) \left( a^{-1} + \sum_{\ell \in \bar{\mathcal{P}}(c_f, f)} g_{k,\ell} \right) - \sum_{b \in \bar{\mathcal{P}}(c_f, f)} g_{k,b}^2} \quad (34)$$

Without loss of generality we set the reuse factor  $R$  to equal  $F$ . Assuming that for  $0 \leq c \leq F - 1$  base station  $c$  is only active in channel  $c = f$ , we have

$$\bar{\mathcal{P}}(c, f) = \{c + QF\ell\}_\ell \cap \{0, 1, \dots, B - 1\} \quad (35)$$

if  $f = c \bmod F$ , and  $\bar{\mathcal{P}}(c, f)$  is empty otherwise. Using (35),  $\rho_k[f]$  in (33) and  $\tau_k[f]$  in (34) can be reexpressed in the forms (20) and (21), respectively. The cellular case expressions are given by evaluating the expanded cellular ones at  $R = 1$ .