Two-Stage Channel Feedback for Beamforming and Scheduling in Network MIMO Systems

Behrouz Khoshnevis and Wei Yu
Department of Electrical and Computer Engineering
University of Toronto, Toronto, Ontario, Canada
Email: {bkhoshnevis, weiyu}@comm.utoronto.ca

Yves Lostanlen
Siradel, Wireless Expertise and Research Centre
Toronto, Ontario, Canada
Email: YLostanlen@siradel.com

Abstract—This paper proposes an efficient two-stage beam-forming and scheduling algorithm for the limited-feedback cooperative multi-point (CoMP) systems. The system includes multiple base-stations cooperatively transmitting data to a pool of users, which share a rate-limited feedback channel for sending back the channel state information (CSI). The feedback mechanism is divided into two stages that are used separately for scheduling and beamforming. In the first stage, the users report their best channel gain from all the base-station antennas and the base-stations schedule the best user for each of their antennas. The scheduled users are then polled in the second stage to feedback their quantized channel vectors. The paper proposes an analytical framework to derive the bit allocation between the two feedback stages and the bit allocation for quantizing each user's CSI. For a total number of feedback bits B, it is shown that the number of bits assigned to the second feedback stage should scale as $\log B$. Furthermore, in quantizing channel vectors from different base-stations, each user should allocate its feedback budget in proportion to the logarithm of the corresponding channel gains. These bit allocation are then used to show that the overall system performance scales double-logarithmically with B and logarithmically with the transmit SNR. The paper further presents several numerical results to show that, in comparison with other beamforming-scheduling algorithms in the literature, the proposed scheme provides a consistent improvement in downlink sum rate and network utility. Such improvements, in particular, are achieved in spite of a significant reduction in the beamforming-scheduling computational complexity, which makes the proposed scheme an attractive solution for practical system implementations.

I. Introduction

The performance of cooperative cellular communication systems is highly dependent on the amount of channel state information (CSI) available at the base-stations. Whether the base-stations cooperate on a beamforming and interference management level or a full-cooperative joint transmission level, also referred to as network multiple-input multiple-output (MIMO) systems, a certain degree of network-wide CSI feedback is necessary for any efficient beamforming-scheduling algorithm to operate [1].

Unfortunately, the feasibility of CSI estimation and feedback is limited in practice either by the limited feedback capacity or the intrinsic delay associated with feedback leading to outdated CSI [2], [3]. Although some innovative techniques are developed to take advantage of the outdated CSI, as in [4], such methods face significant practical challenges.

This paper studies the performance of network MIMO systems and addresses the CSI feedback issues by assuming a limit on the total number of available feedback bits throughout the network. In particular, we propose an efficient beamforming-scheduling algorithm by partitioning the feedback process into two stages that are separately utilized for scheduling and beamforming purposes. The paper develops an analytical framework to optimize the feedback bit allocation between the two stages and also the bit allocation associated with quantizing direct vs. interfering channels. The latter problem can be considered as an extension of the work in [5], which considers a single-user per-cell model.

As the comparison benchmark, the paper considers a single-stage zero-forcing beamforming scheme with greedy user selection, where the number of users participating in CSI feedback is optimized [6]. In comparison with this scheme, the two-stage feedback mechanism is shown to considerably improve the downlink sum rate and the individual per-user average rates for a majority of the users in the network. More importantly, these improvements are shown to be achievable with a significantly reduced beamforming-scheduling computational complexity, which makes the proposed scheme an attractive solution for practical implementations.

II. System Model

The system model consists of I base-stations each with M antennas and a total of J single-antenna users uniformly distributed within the network boundaries. All the users within the network share a common CSI feedback channel with a total capacity of B bits per fading block.

Let us define $\mathcal{I} \overset{\text{def}}{=} \{1, 2, \cdots, I\}$ and $\mathcal{J} \overset{\text{def}}{=} \{1, 2, \cdots, J\}$ as the index sets for the base-stations and users. The channel vector from BS i to user j, denoted by $\mathbf{g}_{ij} \in \mathbb{C}^M$, is defined as

$$\mathbf{g}_{ij} = \sqrt{\alpha_{ij}} \mathbf{h}_{ij}, \qquad (1)$$

where $\mathbf{h}_{ij} \in \mathbb{C}^M$ has i.i.d. $\mathcal{CN}(0,1)$ entries and models the Rayleigh fading component and

$$\sqrt{\alpha_{ij}} = \frac{G_0}{1 + (d_{ij}/\delta)^\zeta} \qquad (2)$$

denotes the path-loss component [2]. Here, d_{ij} is the distance between BS i and user j, G_0 is the transmission gain, δ is the 3dB breakpoint distance, and ζ is the path-loss exponent.

Finally, the collective channel vector of user j from all base-stations is denoted by $\underline{\mathbf{g}}_j \in \mathbb{C}^{MI}$:

$$\underline{\mathbf{g}}_j = \left[\mathbf{g}_{1j}^T, \mathbf{g}_{2j}^T, \cdots, \mathbf{g}_{Ij}^T\right]^T. \qquad (3)$$

The system is assumed to operate in the network MIMO mode, which with a collective number of MI antennas, can support MI users in the downlink. The exact method of choosing those users depends on the scheduling algorithm and is described in the next section. Let us denote the scheduling function by $\pi(\cdot) : \mathcal{M} \times \mathcal{I} \to \mathcal{J}$ and the scheduled users by $\pi(mi)$, where the double index mi belongs to $\mathcal{M} \times \mathcal{I}$ and $\mathcal{M} \stackrel{\text{def}}{=} \{1, 2, \cdots, M\}$. With this notation, the received signal-to-interference-plus-noise ratio (SINR) at user $\pi(mi)$ can be expressed as $\mathrm{SINR}_{\pi(mi)} = \mathcal{N}_{\pi(mi)} / \mathcal{D}_{\pi(mi)}$, where

$$\mathcal{N}_{\pi(mi)} \stackrel{\text{def}}{=} \rho \left| \sum_{k \in \mathcal{I}} \mathbf{g}_{k\pi(mi)}^\dagger \mathbf{v}_{k\pi(mi)} \right|^2 \qquad (4)$$

$$\mathcal{D}_{\pi(mi)} \stackrel{\text{def}}{=} \rho \sum_{j \neq \pi(mi)} \left| \sum_{k \in \mathcal{I}} \mathbf{g}_{k\pi(mi)}^\dagger \mathbf{v}_{kj} \right|^2 + 1. \qquad (5)$$

Here, $\rho \stackrel{\text{def}}{=} \frac{P}{MI}$, P is the transmit power at each BS, and $\mathbf{v}_{kj}$ is the beamforming vector used at BS $k \in \mathcal{I}$ to transmit the data corresponding to user $j \in \mathcal{J}$. With these notations, the downlink sum rate of the network is given by

$$R = MI \mathbb{E}[\log(1 + \mathrm{SINR}_{\pi(mi)})]. \qquad (6)$$

This sum-rate expression is not so amenable to analysis. For convenience, this paper uses a modified performance measure, referred to as the *virtual sum rate*. This approach is partly justified in [7]. If we define *virtual SINR* as

$$\widetilde{\mathrm{SINR}} = \frac{\mathbb{E}\left[\mathcal{N}_{\pi(mi)}\right]}{\mathbb{E}\left[\mathcal{D}_{\pi(mi)}\right]}, \qquad (7)$$

the virtual sum rate is defined as

$$\widetilde{R} = MI \log(1 + \widetilde{\mathrm{SINR}})]. \qquad (8)$$

The next section describes the proposed two-stage beamforming and scheduling algorithm.

III. Two-Stage CSI Feedback Mechanism

The feedback mechanism is shown in Fig. 1. Prior to the CSI feedback, the base-stations transmit pilot signals that allow the users estimate their channels from all base-stations. All the users in the network share a common feedback channel with a capacity of B bits per fading block. The feedback process is divided into two stages, which consume B_1 and B_2 bits such that $B_1 + B_2 = B$. These stages are separately used for scheduling and beamforming purposes as described in the following.

A. Scheduling Process

In the first feedback stage, each user estimates its channel from each antenna of each BS and feeds back $\log(MI)$ bits as the index of the antenna with the highest channel gain along with the gain itself. Each base-station then chooses the user with the highest gain feedback for each of its antennas.
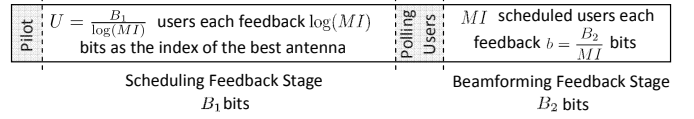

Fig. 1. Two-stage feedback mechanism.

Given B_1 feedback bits in the first stage, the total number of users that can participate in this process is given by

$$U = \frac{B_1}{\log(MI)}, \qquad (9)$$

which are randomly chosen from the pool of J users in the network ensuring that each cell gets an equal share. Denote the set of such users by $\mathcal{U}$. Let $\mathbf{g}_{ij,m}$ denote the m'th entry of the channel vector $\mathbf{g}_{ij}$, i.e., the scalar channel from the m'th antenna of BS i to user j. The user scheduled for the m'th antenna of BS i is therefore determined as follows:

$$\pi(mi) = \arg \max_{j \in \mathcal{U}} |\mathbf{g}_{ij,m}|. \qquad (10)$$

We denote the set of MI scheduled users by $\mathcal{U}^\star \subseteq \mathcal{U}$. Clearly $|\mathcal{U}^\star| = MI$.

After completing the scheduling process, the MI scheduled users are polled by the base-stations to explicitly quantize and feedback their CSI in the second stage. This information is then used for calculating the downlink beams.

B. Beamforming Process

Given B_2 bits in the second stage, each of the MI scheduled users receives a share of

$$b \stackrel{\text{def}}{=} \frac{B_2}{MI} \qquad (11)$$

bits for quantizing its channel vectors from all base-stations. Consider an arbitrary user $j \in \mathcal{U}^\star$. Let $\mathcal{C}_{kj}$ denote the codebook used for quantizing the channel direction $\hat{\mathbf{g}}_{kj} = \mathbf{g}_{kj}/\|\mathbf{g}_{kj}\|$ from BS k to user j and let b_{kj} be the corresponding number of quantization bits: $b_{kj} \stackrel{\text{def}}{=} \log |\mathcal{C}_{kj}|$, which should satisfy

$$\sum_{k \in \mathcal{I}} b_{kj} = b. \qquad (12)$$

Also let $\mathbf{u}_{kj}$ denote the quantized direction corresponding to $\hat{\mathbf{g}}_{kj}$:

$$\mathbf{u}_{kj} \stackrel{\text{def}}{=} \arg \max_{\mathbf{w} \in \mathcal{C}_{kj}} \left| \mathbf{w}^\dagger \hat{\mathbf{g}}_{kj} \right|. \qquad (13)$$

With this notation, the quantized version of the overall channel vector $\underline{\mathbf{g}}_j$, denoted by $\underline{\tilde{\mathbf{g}}}_j$, is formed as follows:

$$\underline{\tilde{\mathbf{g}}}_j \stackrel{\text{def}}{=} \left[\tilde{\mathbf{g}}_{1j}^T, \tilde{\mathbf{g}}_{2j}^T, \cdots, \tilde{\mathbf{g}}_{Ij}^T\right]^T, \qquad (14)$$

where

$$\tilde{\mathbf{g}}_{kj} \stackrel{\text{def}}{=} r_{kj} e^{i\theta_{kj}} \mathbf{u}_{kj}, \qquad (15)$$

and $r_{kj} \stackrel{\text{def}}{=} \|\mathbf{g}_{kj}\| \cos \phi_{kj}$, $\phi_{kj} \stackrel{\text{def}}{=} \angle(\hat{\mathbf{g}}_{kj}, \mathbf{u}_{kj})$, and $\theta_{kj} \stackrel{\text{def}}{=}$ phase$(\hat{\mathbf{g}}_{kj}^\dagger \mathbf{u}_{kj})$.

Considering the definition of the quantized channel in (14), it is easy to verify that as the codebook sizes increase and ϕ_{kj}'s diminish to zero, $\underline{\tilde{\mathbf{g}}}_j$ approaches the actual channel $\underline{\mathbf{g}}_j$. As it should be clear, the mere feedback of the channel directions

$\hat{\mathbf{g}}_{kj}$ is not sufficient by itself. The users should also feedback the magnitude and phase information, r_{kj} and θ_{kj}, so that the base-stations can form the quantized vector $\tilde{\mathbf{g}}_j$. These magnitude and phase information are scalar variables and can be quantized and reconstructed using a few quantization bits. We therefore assume that this information is unquantized and focus on the quantization bit allocation for the channel directions $\hat{\mathbf{g}}_{kj}$.

Having the quantized information at the base-stations, the unit-norm downlink beams $\underline{\mathbf{v}}_j$'s for the scheduled users $j \in \mathcal{U}^\star$ are formed as the zero-forcing directions for the quantized channel vectors $\tilde{\mathbf{g}}_j$. If we write

$$\underline{\mathbf{v}}_j = \left[\mathbf{v}_{1j}^T, \mathbf{v}_{2j}^T, \cdots, \mathbf{v}_{Ij}^T\right]^T, \qquad (16)$$

then $\mathbf{v}_{kj}$ will denote the beam used at base-station k for transmitting user j's data.

Consider a user $s \in \mathcal{U}^\star$ among the scheduled users. Due to the CSI imperfection, the designed zero-forcing beams cannot completely cancel out the multi-user interference. User s will therefore experience an interference from the signals intended for users $j \neq s$. By applying the zero-forcing principle and after some calculations, one can bound such interference in the average sense as follows:

$$\mathbb{E}\left[\left|\sum_{k \in \mathcal{I}} \mathbf{g}_{ks}^\dagger \mathbf{v}_{kj}\right|^2\right] \leq \mathbb{E}\left[\sum_{k \in \mathcal{I}} \|\mathbf{g}_{ks}\|^2 \sin^2 \phi_{ks}\right] \qquad (17)$$

$$\leq \mathbb{E}\left[\sum_{k \in \mathcal{I}} \|\mathbf{g}_{ks}\|^2 2^{-b_{ks}/(M-1)}\right], \quad (18)$$

where in driving (18), we have used the bound for the quantization error with respect to the quantization codebook size, as described in [8]. The bound in (18) is used in the next section to analyze and optimize the system performance.

IV. SYSTEM ANALYSIS

As mentioned in Section II, we use virtual rate defined in (8) as the metric for system analysis. To this end, we start by approximating and bounding the expressions for $\mathcal{N}_{\pi(mi)}$ and $\mathcal{D}_{\pi(mi)}$ defined in (4) and (5).

First, we mention without proof that the scheduling rule in (10) guarantees the scheduled users' channels to be almost orthogonal in $\mathbb{C}^{MI}$. Assuming sufficiently accurate CSI feedback, the corresponding zero-forcing beams are therefore expected to be almost aligned with the users' channels. We can therefore use the following approximation for $\mathcal{N}_{\pi(mi)}$:

$$\mathcal{N}_{\pi(mi)} = \rho \left|\sum_{k \in \mathcal{I}} \mathbf{g}_{k\pi(mi)}^\dagger \mathbf{v}_{k\pi(mi)}\right|^2 = \rho \left|\underline{\mathbf{g}}_{\pi(mi)}^\dagger \mathbf{v}_{\pi(mi)}\right|^2 \quad (19)$$

$$\approx \rho \left\|\underline{\mathbf{g}}_{\pi(mi)}\right\|^2 = \sum_{k \in \mathcal{I}} \left\|\mathbf{g}_{k\pi(mi)}\right\|^2. \qquad (20)$$

Therefore,

$$\mathbb{E}\left[\mathcal{N}_{\pi(mi)}\right] \approx \rho \sum_{k \in \mathcal{I}} \mathbb{E}\left[\lambda_{k\pi(mi)}\right], \qquad (21)$$

where

$$\lambda_{k\pi(mi)} \overset{\text{def}}{=} \left\|\mathbf{g}_{k\pi(mi)}\right\|^2. \qquad (22)$$

Also, considering (18) and (5) and summing over $MI - 1$ scheduled indices $j \neq \pi(mi)$, we have

$$\mathbb{E}\left[\mathcal{D}_{\pi(mi)}\right] < 1 + \rho MI \mathbb{E}\left[\sum_{k \in \mathcal{I}} \left\|\mathbf{g}_{k\pi(mi)}\right\|^2 2^{-b_{k\pi(mi)}/(M-1)}\right]. \qquad (23)$$

Combining (21) and (23), we achieve the following lower bound for the virtual SINR:

$$\widetilde{\text{SINR}} \geq \frac{\rho \sum_{k \in \mathcal{I}} \mathbb{E}\left[\lambda_{k\pi(mi)}\right]}{1 + \rho MI \mathbb{E}\left[\sum_{k \in \mathcal{I}} \lambda_{k\pi(mi)} 2^{-b_{k\pi(mi)}/(M-1)}\right]}. \quad (24)$$

Our objective is to maximize the virtual SINR lower bound in (24). To this end, we need to optimize a two-level bit allocation:

1) CSI quantization bit allocation problem: optimizing the number of quantization bits $b_{k\pi(mi)}$ as a function of channel gains $\left\|\mathbf{g}_{k\pi(mi)}\right\|$ subject to the constraint $\sum_{k \in \mathcal{I}} b_{k\pi(mi)} = b = B_2/(MI)$.
2) Beamforming-scheduling bit allocation problem: optimizing the bit allocation between the two feedback stages subject to the constraint $B_1 + B_2 = B$.

A. CSI Quantization Bit Allocation

The bit allocation problem is stated as minimization of the residual interference in the denominator of (24):

$$\min_{\substack{b_{k\pi(mi)} \\ 1 \leq k \leq I}} \sum_{k \in \mathcal{I}} \lambda_{k\pi(mi)} 2^{-b_{k\pi(mi)}/(M-1)} \qquad (25)$$

$$\text{s.t.} \quad \sum_{k \in \mathcal{I}} b_{k\pi(mi)} = b, \quad b_{k\pi(mi)} \geq b_0, \qquad (26)$$

where $b_0 \overset{\text{def}}{=} \log M$ is the minimum number of bits required for quantizing an M-dimensional channel vector. This problem has a waterfilling-type solution as follows:

$$b_{k\pi(mi)} = b_0 + (M - 1)\left(\log \frac{\lambda_{k\pi(mi)}}{\mu}\right)^+, \qquad (27)$$

where $\mu > 0$ is such that $\sum_{k \in \mathcal{I}} b_{k\pi(mi)} = b$. The bit allocation rule in (27) shows that the scheduled user $\pi(mi)$ should allocate its CSI quantization bits in proportion to the logarithm of the channel gains $\lambda_{k\pi(mi)}$.

If B_2 and therefore b are large enough, this bit allocation can be approximated as follows:

$$b_{k\pi(mi)} \approx \frac{b}{I} + (M - 1)\log \frac{\lambda_{k\pi(mi)}}{\mathcal{G}\left(\lambda_{k\pi(mi)}\big|_{k \in \mathcal{I}}\right)}, \qquad (28)$$

where $\mathcal{G}\left(\lambda_{k\pi(mi)}\big|_{k \in \mathcal{I}}\right)$ is the geometric mean of $\left\{\lambda_{k\pi(mi)}\right\}_{k \in I}$. By substituting this back in the objective function in (25), we obtain the following residual interference:

$$\sum_{k \in \mathcal{I}} \lambda_{k\pi(mi)} 2^{-b_{k\pi(mi)}/(M-1)} = I\mathcal{G}\left(\lambda_{k\pi(mi)}\big|_{k \in \mathcal{I}}\right) 2^{-b/(M-1)I}. \qquad (29)$$

Now, we are ready to bound the virtual SINR using the optimal bit allocations. By substituting the minimized residual interference in (29) in the denominator of the virtual SINR

lower bound in (24), the denominator will satisfy

$$1 + \rho M I^2 \mathbb{E}\left[\mathcal{G}\left(\lambda_{k\pi(mi)}\big|_{k\in\mathcal{I}}\right)\right] 2^{-b/(M-1)I} \tag{30}$$

$$\leq 1 + \rho M I^2 \mathcal{G}\left(\mathbb{E}\left[\lambda_{k\pi(mi)}\right]\big|_{k\in\mathcal{I}}\right) 2^{-b/(M-1)I}, \tag{31}$$

where we have used the Jensen's inequality and the concavity of the geometric mean with respect to its arguments. If we define

$$\beta_{ki} \stackrel{\text{def}}{=} \mathbb{E}\left[\lambda_{k\pi(mi)}\right], \tag{32}$$

and combine (31) and (24), we obtain the following lower bound for the virtual SINR:

$$\widetilde{\text{SINR}} \geq \frac{\rho \sum_{k\in\mathcal{I}} \beta_{ki}}{1 + \rho M I^2 \overline{\beta}_i \; 2^{-b/(M-1)I}}, \tag{33}$$

where $\overline{\beta}_i \stackrel{\text{def}}{=} \mathcal{G}\left(\beta_{ki}\big|_{k\in\mathcal{I}}\right)$ for notation convenience. This bound is used in the next subsection to optimize the beamforming-scheduling bit allocation.

B. Beamforming-Scheduling Bit Allocation

First, we note that

$$\lambda_{k\pi(mi)} = \left\|\mathbf{g}_{k\pi(mi)}\right\|^2 \geq \left|\mathbf{g}_{k\pi(mi),m}\right|^2. \tag{34}$$

Therefore, $\beta_{ki} = \mathbb{E}\left[\lambda_{k\pi(mi)}\right] \geq \chi_{ki}$, where $\chi_{ki} \stackrel{\text{def}}{=} \mathbb{E}\left[\left|\mathbf{g}_{k\pi(mi),m}\right|^2\right]$. Considering this notation and noting that the lower bound in (33) is monotonic in β_{ki}, we have

$$\widetilde{\text{SINR}} \geq \frac{\rho \sum_{k\in\mathcal{I}} \chi_{ki}}{1 + \rho M I^2 \overline{\chi}_i \; 2^{-b/(M-1)I}}, \tag{35}$$

where $\overline{\chi}_i$ is the geometric mean of $\{\chi_{ki}\}_{k\in\mathcal{I}}$. Finally, according to the scheduling rule in (10), we have $\chi_{ii} \geq \chi_{ki}, \forall k \in \mathcal{I}$. Therefore, $\overline{\chi}_i \leq \chi_{ii}$ and clearly $\sum_{k\in\mathcal{I}} \chi_{ki} \geq \chi_{ii}$. Combining these with the expression in (35), we obtain a more simplified lower bound for the virtual SINR:

$$\widetilde{\text{SINR}} \geq \frac{\rho \chi_{ii}}{1 + \rho M I^2 \chi_{ii} \; 2^{-b/(M-1)I}}. \tag{36}$$

We can now maximize the lower bound in (36) with respect to the beamforming-scheduling bit allocation subject to the constraint $B_1 + B_2 = B$. For this purpose, we go back to the scheduling rule in (10) and note that

$$\left|\mathbf{g}_{i\pi(mi),m}\right| = \max_{j\in\mathcal{U}} \left|\mathbf{g}_{ij,m}\right|, \tag{37}$$

where $\mathcal{U}$ is the set of users participating in the first feedback stage with $|\mathcal{U}| = B_1/\log(MI)$. If we define $\mathcal{U}_i$ as the subset of users in $\mathcal{U}$ that reside in cell i then clearly

$$\left|\mathbf{g}_{i\pi(mi),m}\right| \geq \max_{j\in\mathcal{U}_i} \left|\mathbf{g}_{ij,m}\right| \geq \sqrt{\alpha_d} \max_{j\in\mathcal{U}_i} \left|\mathbf{h}_{ij,m}\right|, \tag{38}$$

where

$$\sqrt{\alpha_d} = \frac{G_0}{1 + (C/\delta)^\zeta} \tag{39}$$

is the highest path-loss possible within the boundaries of cell i, where C is the cell radius. By taking the expectation of the both sides of (38), we obtain

$$\chi_{ii} = \mathbb{E}\left[\left|\mathbf{g}_{i\pi(mi),m}\right|^2\right] \geq \alpha_d \mathbb{E}\left[\max_{j\in\mathcal{U}_i} \left|\mathbf{h}_{ij,m}\right|^2\right] \tag{40}$$

$$\stackrel{(a)}{\approx} \alpha_d \ln|\mathcal{U}_i| \stackrel{(b)}{=} \alpha_d \ln \frac{B_1}{I\log(MI)}, \tag{41}$$

where the equality in (b) uses the fact that $|\mathcal{U}_i| = |\mathcal{U}|/I$ and the approximation in (a) uses the result in [9] regarding the behavior of the maximum of multiple exponential random variables.

Finally, noting that the lower bound in (36) is monotonic in χ_{ii} and using (41) and (11), we obtain the following lower bound on the virtual SINR in terms of B_1 and B_2:

$$\widetilde{\text{SINR}} \geq \frac{\rho \alpha_d \ln\left(\frac{B_1}{I\log(MI)}\right)}{1 + \rho M I^2 \alpha_d \ln\left(\frac{B_1}{I\log(MI)}\right) \; 2^{-\frac{B_2}{M(M-1)I^2}}}. \tag{42}$$

We can now directly maximize the SINR lower bound subject to $B_1 + B_2 = B$. By doing so, one arrives at the following equations for the optimal values of B_1 and B_2:

$$\begin{cases} \frac{\rho \alpha_d \ln 2}{M-1} B_1 (\ln B_1 - \kappa)^2 = 2^{\frac{B_2}{M(M-1)I^2}} \\ B_1 + B_2 = B \end{cases} \tag{43}$$

where $\kappa = \ln(I\log(MI))$.

We can further simplify the bit allocation results by assuming an asymptotic regime where $B \to \infty$. by doing so, we get the following asymptotic bit allocations:

$$B_1 \doteq B \tag{44}$$

$$B_2 \doteq M(M-1)I^2 \log(\rho \alpha_d B), \tag{45}$$

where the notation $x \doteq y$ means $\lim_{B\to\infty} x/y = 1$. These results show that as the total feedback rate increases, higher percentage of bits should be used for the scheduling stage. In particular, the ratio B_2/B behaves $\log B/B$ as $B \to \infty$. Finally, by substituting the asymptotic bit allocations in the lower bound in (42) and the virtual rate expression in (8), we obtain

$$\widetilde{R} \doteq M I \log(\rho \alpha_d) + M I \log\log B, \tag{46}$$

which suggests that both multi-user multiplexing and multi-user diversity gains are realized.

To summarize:

1) We derived the CSI bit allocation rules in (27) and (28), which show that each scheduled user should allocate its feedback bit budget in proportion to the logarithm of the channel gains from base-stations.
2) We also derived asymptotic beamforming-scheduling bit allocation laws in (44) and (45) showing that the fraction of bits assigned to beamforming feedback stage scales as $\log B/B$.
3) Finally, we used the optimized bit allocations to show that the overall system performance scales double-logarithmically with B as $B \to \infty$.

In spite of the simplifications and approximate bounds used in the derivation process, simulation results in the next section show that these scaling laws give remarkably accurate predictions of the optimal bit allocations.

V. NUMERICAL RESULTS

Throughout the simulations, we assume a cell radius of $C = 0.5$km, a 3dB breakpoint distance of $\delta = 0.1$km, path-loss exponent $\zeta = 3.8$, a transmission gain of $G_0 = 20$dB,
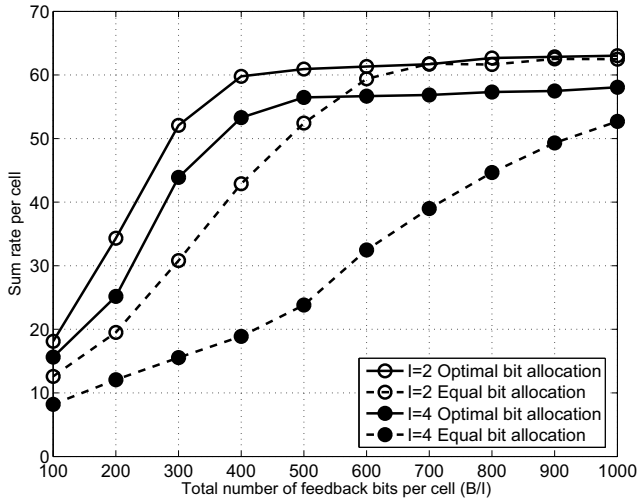
Fig. 2. CSI quantization bit allocation: optimal vs. equal bit allocation for $I = 2$ and 4 cells, SNR = 15dB, and 300 users per cell.
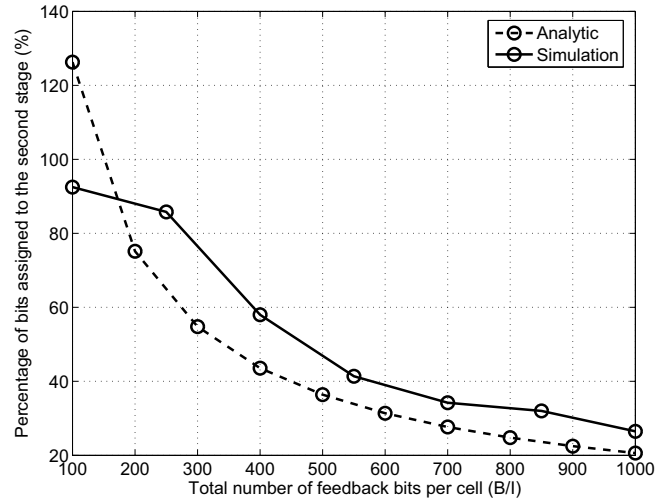

Fig. 3. Beamforming-scheduling bit allocation: percentage of bits assigned to the beamforming feedback stage, i.e. B_2/B, for $I = 2$ cells, SNR = 15dB, and 300 users per cell.
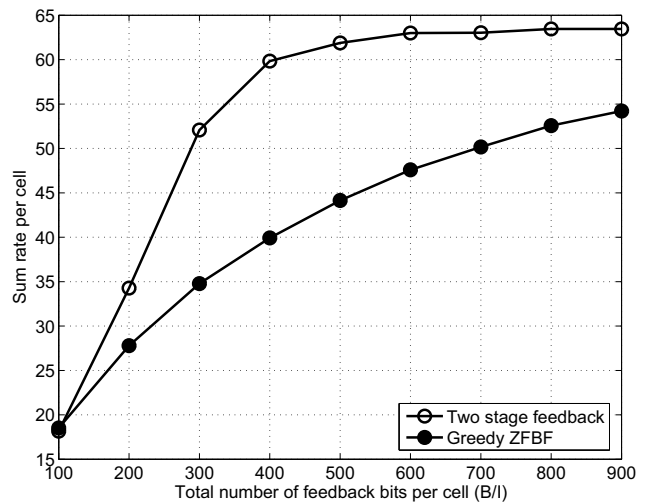

Fig. 4. Two-stage beamforming-scheduling algorithm vs. ZFBF with greedy user selection for $I = 2$ cells, SNR = 15dB, and 300 users per cell.

and $M = 4$ antennas per base-station. The number of cells, the number of users per cell, and the transmission SNR vary throughout the simulations and are specified separately for each simulation result[1].

We start with the CSI quantization bit allocation problem discussed in Section IV-A. Fig. 2 compares the downlink sum rate per cell as a function of total number of feedback bits per cell, when we use the optimal CSI quantization bit allocation rule in (27) vs. the case where one allocates equal number of bits for quantizing vector channels from different base-stations. As the figure shows, in comparison to equal bit allocation, optimal bit allocation provides significant gains in the sum rate. Furthermore, the gain in sum rate increases with the number of cells.

The next simulation result corresponds to the beamforming-scheduling bit allocation problem addressed in Section IV-B. In this regard, Fig. 3 plots the optimal percentage of bits that should be allocated to the second feedback stage, i.e. B_2/B, and compares it with the analytic bit allocation result in (45). As the figure verifies, the ratio B_2/B behaves as $\log B/B$ as B increases.

Next, we address the system performance in terms of the downlink sum rate. Fig. 4 compares the performance of the proposed two-stage feedback mechanism (with optimized bit allocations) with the performance of the ZFBF scheme with greedy user selection (with optimized number of users participating in the feedback process). The figure shows a significant improvement in the downlink sum rate. The reason is that the two-stage scheme allows more users to participate in the feedback process and is therefore much more efficient in finding users with better channel conditions, which in this case are the users closer to the base-stations.

In order to obtain a more balanced comparison between the two schemes, Fig. 5 compares the performances when

one implements a proportional fairness scheduler (PFS) on top of the two algorithms in order to ensure fairness among users. The PFS scheduler uses a logarithmic network utility function instead of sum rate as the scheduling criterion. As the figure shows, the two-stage approach still shows a considerable improvement in the downlink sum rate. Further results, not presented here due to space limits, show that the two-stage scheme also outperforms the one-stage ZFBF approach in terms of the PFS network utility.

In order to obtain a better understanding of the system performance, Fig. 6 plots the cumulative distribution function (CDF) for the downlink average rates per user in the presence of PFS. The figure shows that the two-stage scheme significantly improves the individual rates for a majority of users (in this case the higher 80%) at the expense of a decreased rate for a fraction of users (lower 20%).

It is important to note that all the rate improvements

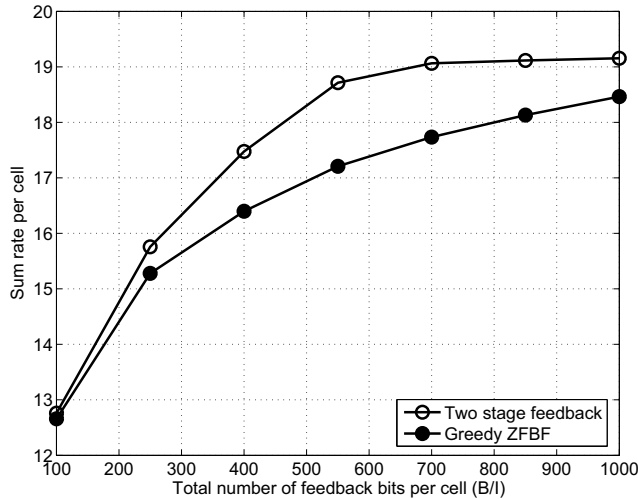[1]By transmission SNR, we mean the transmission power per BS denoted by P in earlier sections.

Fig. 5. Downlink sum rate with proportional fairness: two-stage beamforming-scheduling algorithm vs. ZFBF with greedy user selection for $I = 2$ cells, SNR $= 15$dB, and 300 users per cell.
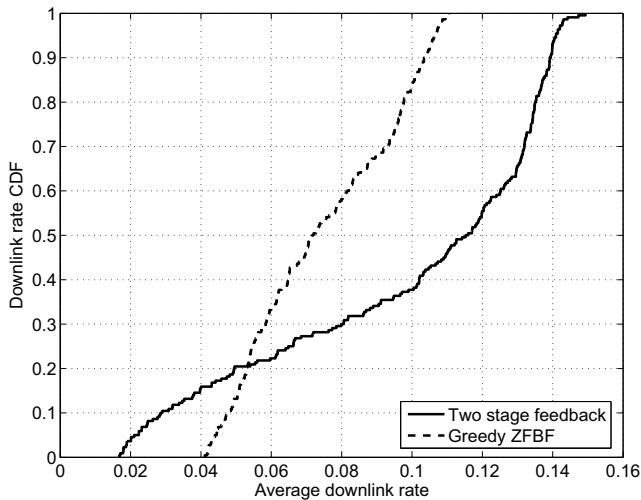

Fig. 7. Typical CPU time required for scheduling users with proportional fairness: $I = 2$ cells, SNR $= 15$dB, and 300 users per cell.

Fig. 6. Downlink average rate CDF with proportional fairness: two-stage beamforming-scheduling algorithm vs. ZFBF with greedy user selection for $I = 2$ cells, $B/I = 300$ bits per cell, SNR $= 10$dB, and 100 users per cell.

offered by the two-stage scheme are achieved in spite of its significantly lower computational complexity. To see this, Fig. 7 compares standard CPU times required for performing scheduling computations for the two-stage scheme and the one-stage ZFBF scheme. As the figure verifies, the two-stage scheme is $8-10$ times faster in making scheduling decisions.

VI. Conclusions

This paper proposes an efficient two-stage beamforming and scheduling algorithm for limited-feedback cooperative cellular communication systems. It is shown that the number of bits assigned to the beamforming feedback stage should scale logarithmically with the total number of feedback bits. Furthermore, in quantizing CSI from different base-stations, each user should allocate its feedback bit budget in proportion to the logarit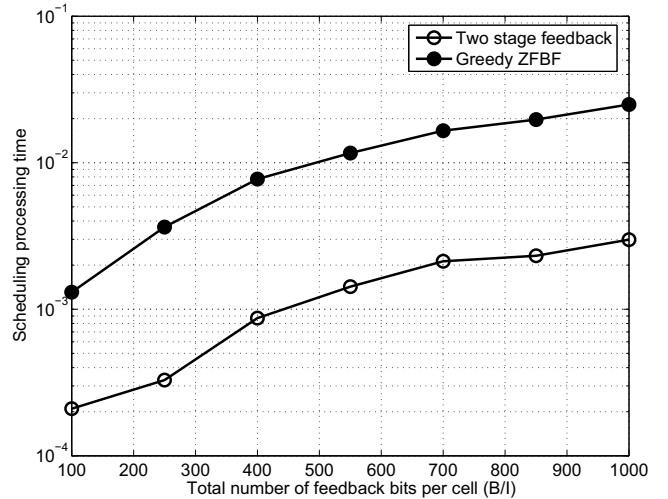hm of the corresponding channel gains. The bit allocation results are then used to show that the overall system performance scales double-logarithmically with the total feedback rate and logarithmically with the transmit SNR. The paper further presents several numerical results to show that the proposed scheme, in comparison to other available feedback mechanisms, improves both the downlink sum rate and the network utility. These advantages are achieved in spite of a significant reduction in the scheduling computational complexity, which makes the proposed scheme an attractive solution for practical system implementations.

References

[1] D. Gesbert, S. Hanly, H. Huang, S. Shamai, O. Simeone, and W. Yu, "Multicell MIMO cooperative networks: A new look at interference," *IEEE Journal on Selected Areas in Communications*, vol. 28, no. 9, pp. 1380-1408, Dec. 2010.

[2] G. Caire, A. Ramprashad, and C. Papadopoulos, "Rethinking network MIMO: Cost of CSIT, performance analysis, and architecture comparisons," *Inf. Theory Appl. Workshop (ITA)*, San Diego, CA, May 2010.

[3] J. Zhang, M. Kountouris, J. Andrews, and R. Heath, "Multi-mode transmission for the MIMO broadcast channel with imperfect channel state information," *IEEE Trans. on Communications*, vol. 59, no. 3, pp. 803-814, Mar. 2011.

[4] M. Maddah-Ali and D. Tse, "Completely stale transmitter channel state information is still very useful," in *Proc. of Allerton Conference*, Monticello, IL, Sep. 2010.

[5] R. Bhagavatula and R. Heath, "Adaptive limited feedback for sum-rate maximizing beamforming in cooperative multicell systems," *IEEE Trans. on Signal Process.*, vol. 59, no. 2, pp. 800-811, Feb. 2011.

[6] N. Ravindran and N. Jindal, "Multi-user diversity vs. accurate channel state information in MIMO downlink channels," submitted for publication. [Online]. Available: http://arxiv.org/abs/0907.1099

[7] B. Khoshnevis and W. Yu, "A limited-feedback scheduling and beamforming scheme for multi-user multi-antenna systems," in *Proc. Global Telecommun. Conf. (GLOBECOM)*, Houston, TX, 5-9 December 2011.

[8] N. Jindal, "MIMO broadcast channels with finite-rate feedback," *IEEE Trans. Inform. Theory*, vol. 52, no. 11, pp. 5045-5060, Nov. 2006.

[9] M. Sharif and B. Hassibi, "On the capacity of MIMO broadcast channels with partial side infonnation," *IEEE Trans. Inform. Theory*, vol. 51, no. 2, pp. 506-522, Feb. 2005.