Downlink Overhead Reduction for Multi-Cell Cooperative Processing enabled Wireless Networks

Agisilaos Papadogiannis†, Hans Jørgen Bang∗, David Gesbert‡ and Eric Hardouin†
†France Telecom, Research & Development Division
38-40 rue du Général Leclerc, 92794 Issy Moulineaux cedex 9, France
Email: {agisilaos.papadogiannis, eric.hardouin}@orange-ftgroup.com
∗Unik - University Graduate Centre, University of Oslo
Instituttveien 25, P. O. Box 70, N-2027 Kjeller, Norway. Email: hans@unik.no
‡Mobile Communications Department, Eurécom Institute
06904 Sophia-Antipolis, France. Email: gesbert@eurecom.fr

***Abstract*—Multi-cell Cooperative Processing (MCP) has been recognised as an efficient technique for increasing spectral efficiency of future cellular systems. However the provided benefits come at the cost of increased overhead and computational complexity; Mobile Stations (MSs) need to feed back to their assigned Base Station (BS) their local channel state information (CSI) which in turn needs to be transmitted to the Control Unit that coordinates the cooperating BSs. Furthermore user data needs to be routed to and from all cooperating BSs on the downlink and uplink respectively. Therefore in order for the overhead to be affordable, it is admitted that cooperating BSs shall be organised in clusters of a limited size. Nevertheless, it is still crucial that CSI feedback and inter-base information exchange be reduced. In this paper linear precoding is considered with the target of overhead minimisation of the downlink. A novel technique is proposed which allows MSs not to feed back the channel coefficients related to the cluster BSs that provide weak channel quality. This is shown to provide a good trade-off between performance and overhead.**

I. Introduction

Multi-cell Cooperative Processing (MCP) is well acknowledged as an effective means of co-channel interference (CCI) mitigation and spectral efficiency increase. In MCP enabled networks CCI can be alleviated with no extra bandwidth cost but with the addition of some infrastructure complexity and communication overhead [1],[2]. Base Stations (BSs) that cooperate are inter-connected with high capacity backhaul links and form a distributed antenna array. The extra infrastructure needed consists of the costly backhaul links and the Control Units (CUs) necessary for coordinating BSs and performing scheduling and signal processing operations. In order for the cost and the complexity to be affordable, cooperation clusters in practice need to be of a limited size [3]-[5]. Thus, the conception of MCP has moved from *full cooperation* where all BSs of the network cooperate to *static limited cooperation*. In the latter, cooperation clusters comprise a limited number of BSs and they do not change in time [3],[4]. Furthermore, important gains in performance can be obtained from *dynamic cluster formation* [5].

Feedback of the CSI by the Mobile Stations (MSs) is one of the setbacks of multi-user MIMO Frequency Division Duplexing (FDD) systems, since its existence and quality greatly affects the downlink performance (radio signaling overhead). Techniques for CSI feedback load reduction of non-cooperative single cell processing networks have been investigated [6]. CSI feedback reduction is even more demanding in MCP enabled networks since MSs need to estimate and feed back an increased number of channel coefficients (the channel of all cooperating antennas). Furthermore BSs need to exchange local CSI with the cluster CU (backhaul signaling) and also buffer user data corresponding to the chosen set of all MSs to be served by the cluster. This results to a significant infrastructure cost (need for costly high capacity links) and to an increased processing complexity (processing of an increased number of MSs). Therefore overhead reduction is highly desirable even in small cooperation clusters.

In this paper a novel framework is proposed that allows significant reduction of MCP overhead. Inside a cooperation cluster each MS selects its Master BS, the one that it receives the maximum SNR from. The rest of the BSs belonging to the same cluster are considered Slave BSs for this MS. Signaling reduction is achieved either with the aid of absolute or relative thresholds. In the first case, if the SNR related to any BS is below an absolute threshold, the corresponding coefficient is not fed back by the MS. In the second case, if the SNR related to a Slave BS is below a relative threshold (relative to the SNR of the Master BS), its associated channel coefficient is not fed back. These techniques can greatly reduce not only radio signaling, but also backhaul overhead. They can also simplify user data routing and they prove to be a good trade-off between overhead and performance.

The paper is structured in the following way: In section II the signal and system model are presented. In section III the algorithms for signaling reduction are described and in section IV numerical results are presented and discussed. In section V the paper is concluded.

Notation: Lower and upper case boldface symbols denote vectors and matrices respectively, $(.)^T$ and $(.)^H$ denote the transpose and the transpose conjugate respectively. $\|.\|_F$ represents the Frobenius norm, $|.|$ the cardinality of a set and $\mathbb{C}^k$ the complex space with k dimensions. Let ${}_nC_k$ denote the binomial coefficient.

II. Signal and System Model

A BS cluster is considered which comprises B base stations with M antennas each and K single antenna MSs overall. If MCP is enabled, the antennas of the cluster jointly combine and can serve at most $B \times M$ mobile stations simultaneously, under a linear precoding framework. Flat fading uncorrelated channels are considered. The complete channel matrix is

$$\mathbf{H} = \left[\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_K\right]^T \tag{1}$$

where $\mathbf{h}_i \in \mathbb{C}^{BM \times 1}$ is the channel vector of the i-th MS.

Let $\mathcal{S}$ be the set of MSs scheduled to be served at a specific time slot, where $|\mathcal{S}| \leq BM$. Therefore $\mathbf{H}(\mathcal{S})$ is the channel matrix related to these MSs, $\mathbf{y}(\mathcal{S})$ is the received signal vector, $\mathbf{u}(\mathcal{S})$ is a vector of independent complex Gaussian transmit symbols with unit variance, $\mathbb{E}\left\{\mathbf{u}\mathbf{u}^H\right\} = \mathbf{I}_{|S|}$. $\mathbf{n}(\mathcal{S})$ is a vector of independent complex circularly symmetric additive Gaussian noise components, $n \sim \mathcal{NC}\left(0, \sigma^2\right)$. Therefore $\mathbb{E}\left\{\mathbf{n}\mathbf{n}^H\right\} = \sigma^2 \mathbf{I}_{|S|}$. Equal power allocation across users is considered throughout the paper for simplicity.

A. Single Cell Processing

In the case of single cell processing and one antenna per BS ($|\mathcal{S}| \leq B$), the diagonal power allocation matrix $\mathbf{A}(\mathcal{S})$ is

$$\mathbf{A}(\mathcal{S}) = \sqrt{P} \times \mathbf{I}_{|S|} \tag{2}$$

and the received signal of the MSs is

$$\mathbf{y}(\mathcal{S}) = \mathbf{H}(\mathcal{S})\, \mathbf{A}(\mathcal{S})\, \mathbf{u}(\mathcal{S}) + \mathbf{n}(\mathcal{S}). \tag{3}$$

The Signal to Interference plus Noise Ratio (SINR) of the i-th MS, when k is its associated BS, is

$$SINR_i = \frac{\|h_{ik}\|^2}{\sum_{j \neq k} \|h_{ij}\|^2 + \sigma^2/P} \tag{4}$$

where $\|h_{ik}\|^2$ corresponds to the channel gain of the useful signal and $\sum_{j \neq k} \|h_{ij}\|^2$ corresponds to the detrimental CCI.

B. MCP with Linear Precoding

In this paper, linear precoding has been considered for MCP since it provides a good trade-off between performance and complexity. $\mathbf{W}(\mathcal{S})$ is the precoding matrix of size $BM \times |\mathcal{S}|$ and $\mathbf{y}(\mathcal{S})$ is the received signal vector. The signal model can be represented in the following way,

$$\mathbf{y}(\mathcal{S}) = \mathbf{H}(\mathcal{S})\, \mathbf{x} + \mathbf{n}(\mathcal{S}) \tag{5}$$

where $\mathbf{x}$ is the transmit signal vector of size $BM \times 1$. The transmit symbols are mapped to the transmit antennas,

$$\mathbf{x} = \mathbf{W}(\mathcal{S})\mathbf{A}(\mathcal{S})\mathbf{u}(\mathcal{S}). \tag{6}$$

The precoding matrix is

$$\mathbf{W}(\mathcal{S}) = \left[\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_{|\mathcal{S}|}\right] \tag{7}$$

where $\mathbf{w}_i \in \mathbb{C}^{BM \times 1}$. Therefore the scheduled MSs receive

$$\mathbf{y}(\mathcal{S}) = \mathbf{H}(\mathcal{S})\, \mathbf{W}(\mathcal{S})\mathbf{A}(\mathcal{S})\, \mathbf{u}(\mathcal{S}) + \mathbf{n}(\mathcal{S}). \tag{8}$$

The precoding matrix is chosen in order to meet the Zero-Forcing criteria, $\mathbf{H}(\mathcal{S})\, \mathbf{W}(\mathcal{S}) = \mathbf{I}_{|\mathcal{S}|}$, where $\mathbf{I}_{|\mathcal{S}|}$ is an identity matrix with the dimension equal to the number of selected users. Hence, the selected precoding matrix is the Moore-Penrose pseudoinverse of the channel matrix,

$$\mathbf{W}(\mathcal{S}) = \mathbf{H}^H(\mathcal{S}) \left[\mathbf{H}(\mathcal{S})\, \mathbf{H}^H(\mathcal{S})\right]^{-1}. \tag{9}$$

Note that other choices of precoding (MMSE etc.) can be considered. In practice each column of $\mathbf{W}$ is normalised to unity, which is equivalent to adding an additional scaling factor to the power allocation matrix $\mathbf{A}(\mathcal{S})$.

Realistic per-antenna power constraints are considered. It is assumed that each antenna has an average power constraint P. Thus $\mathbb{E}\left\{|x_i|^2\right\} \leq P$ for $i = 1, \ldots, BM \Rightarrow \left[\mathbf{W}\mathbf{W}^H\right]_{ii} \mathbf{A}_{ii}^2 \leq P$. In order to guarantee that the power constraints are always met, the power allocation matrix is

$$\mathbf{A}(\mathcal{S}) = \sqrt{P / \max_{k=1,\ldots,BM} \left\|\mathbf{W}^{[k]}\right\|_F^2} \times \mathbf{I}_{|S|} \tag{10}$$

where $\mathbf{W}^{[k]}$ is the row vector of $\mathbf{W}$ which corresponds to the k-th antenna. The power allocation matrix is computed by the CU that gathers CSI and selects users. The SINR of the i-th MS, where $i \in \mathcal{S}$, when linear precoding is employed is

$$SINR_i = \frac{\|\mathbf{h}_i \mathbf{w}_i\|^2}{\sum_{j \in \mathcal{S}, j \neq i} \|\mathbf{h}_i \mathbf{w}_j\|^2 + \left(\max_{k=1,\ldots,BM} \left\|\mathbf{W}^{[k]}\right\|_F^2 \sigma^2\right)/P} \tag{11}$$

where $\mathbf{w}_m$ is the beamforming vector for the m-th MS and $\mathbf{h}_m$ is the channel vector between the m-th MS and all the antennas of the cooperation cluster. The term $\sum_{j \in \mathcal{S}, j \neq i} \|\mathbf{h}_i \mathbf{w}_j\|^2$ corresponds to the intra-cluster interference. With zero-forcing precoding intra-cluster interference is eliminated and the SINR becomes,

$$SINR_i = \frac{P}{\max_{k=1,\ldots,BM} \left\|\mathbf{W}^{[k]}\right\|_F^2 \sigma^2}. \tag{12}$$

The evaluation metric is the average achieved sum-rate per cell given by the following expression

$$C = \frac{1}{B} \mathbb{E}_H \left\{\sum_{i \in \mathcal{S}} \log_2\left(1 + SINR_i\right)\right\}. \tag{13}$$

C. Radio Signaling Overhead

Radio signaling overhead is measured by the number of channel coefficients fed back by the MSs. Let $\mathcal{E}(t)$ be the set of all channel coefficients in time slot t. With B BSs and K users there are a total of $|\mathcal{E}(t)| = BKM$ coefficients. In each time slot a subset $\mathcal{F}(t) \subseteq \mathcal{E}(t)$ is fed back. From a radio signaling overhead perspective we are interested in minimizing the instantaneous feedback load $N(t) = |\mathcal{F}(t)|$ in each slot. To measure the accomplishment of this goal we use the average number of coefficients fed back per MS of the system

$$\bar{L} = \frac{1}{K}\mathbb{E}_t\{N(t)\}. \tag{14}$$

D. Backhaul Overhead

An indicator of the backhaul overhead is the average number of user data streams transmitted per BS per time slot. This is determined by the number of zero elements of the precoding matrix (7). The number of transmitted streams per BS gives an indication of backhaul utilisation in order for these streams to be distributed to all BSs involved in the cooperation. Partial radio signaling results to the reduction of the number of transmitted streams per BS. Let $Z(t)$ be the number of zero elements of the beamforming matrix during slot t. The average number of transmitted data streams per BS is

$$\bar{S} = -\frac{1}{BM}\mathbb{E}_t\{Z(t)\} + BM. \tag{15}$$

III. Overhead Reduction Techniques

In this section some schemes allowing overhead reduction are described. It is assumed that each BS has one antenna for simplicity, although results can be easily generalized to the multiple antenna case. The signaling reduction algorithm dictates how many channel coefficients will be fed back which in turn can also reduce backhaul overhead.

A. Absolute thresholding

Next we formulate an algorithm based on absolute thresholding. In each time slot each MS estimates the channel coefficient to all BSs within the cooperating cluster. However only coefficients with a corresponding SNR exceeding a predefined threshold γ_{thr} are fed back.

Algorithm 1 Absolute thresholding

Require: Define power threshold γ_{thr}
  for all MSs $k = 1, \ldots, K$ **do**
    Define Master BS m, where $m \in \{1, \ldots, B\}$
    for all BSs $j = 1, \ldots, B$ **do**
      if $\gamma_{kj} \geq \gamma_{thr}$ **then**
        MS k feeds back h_{kj} to its Master BS m
      else
        MS k does not feed back h_{kj} to its Master BS m
      end if
    end for
  end for

The instantaneous feedback load $N(t) = |\mathcal{F}(t)|$ will be a random variable depending on the distributions of the channel coefficients and the threshold γ_{thr}. Let P_{γ_i} denote the cdf of the SNR associated with the i-th coefficient in $\mathcal{E}(t)$ (according to some arbitrary ordering). In general there are BK choose n (${}_{BK}C_n$) different combinations of coefficients belonging to $\mathcal{E}(t)$ that results in $N(t) = n$. Let $I(j)$ contain the indices of the coefficients belonging to the j-th combination. The corresponding probability is

$$T_j = \prod_{i \in I(j)} [1 - P_{\gamma_i}(\gamma_{thr})] \prod_{i \in [1,\ldots,BK] \setminus I(j)} P_{\gamma_i}(\gamma_{thr}). \tag{16}$$

Hence, the probability that $N(t) = n$ is

$$\Pr\{N(t) = n\} = \sum_{j=1}^{{}_{BK}C_n} T_j. \tag{17}$$

The probability of no feedback at all is $\Pr\{N(t) = 0\} = \prod_{i=1}^{BK} P_{\gamma_i}(\gamma_{thr})$. The average number of coefficients fed back per MS is

$$\bar{L} = \frac{1}{K}\sum_{n=0}^{BK} n \left[\sum_{j=1}^{{}_{BK}C_n} T_j\right] \tag{18}$$

which, in the case of identical average SNR $\bar{\gamma} = \mathbb{E}\{\gamma\}$ amongst all channel coefficients and Rayleigh fading, boils down to $\bar{L} = e^{-\gamma_{thr}/\bar{\gamma}}$. Even though the feedback load will fluctuate from slot to slot it will tend to stabilize around some mean value as the number of users increases [6].

The fact that MSs feed back a limited number of channel coefficients to their Master BS results to a partial inversion of the channel matrix in the precoding design phase, something that inevitably degrades performance. This also results to the transmission of a smaller number of data streams per BS and it can be exploited for reducing backhaul utilisation.

The algorithm above requires that the decisions about radio signaling reduction rely on short-term CSI. This inevitably requires high computational complexity since a new decision needs to be made by the MS when the small-scale fading realization changes. MSs need to continuously estimate the channels related to all BSs before deciding which ones will feed back. However the decision can be made taking into account long-term information about the channel ($\bar{\gamma_{kj}}$), something that can reduce complexity and channel estimation burden (the MS knows in advance which coefficients will not feed back and therefore it does not need to estimate them inside a long-term cycle).

B. Relative thresholding

In a cooperation cluster, each MS chooses its Master BS, which is the one that it receives the maximum SNR from. A possible solution for radio feedback reduction is that MSs always feedback the CSI related to their Master BS. They decide whether or not to feedback CSI related to Slave BSs depending on a threshold relative to the strongest channel (channel gain of the Master BS).

This algorithm makes its decision by taking into account the actual fading state of the channel a MS experiences. As in the case of absolute thresholding, complexity can be substantially reduced if feedback decision is made based on long term CSI ($\mathbb{E}\|h\|^2$).

Algorithm 2 Relative thresholding

Require: Define parameter ϵ, where $\epsilon \in [0, 1]$
  for all MSs $k = 1, \ldots, K$ **do**
    Step 1 Define Master BS m, where $m \in \{1, \ldots, B\}$
    Step 2 MS k feeds back h_{km} to its Master BS m
    for all BSs $j = 1, \ldots, B$, $j \neq m$ **do**
      if $\|h_{kj}\|^2 \geq \epsilon \|h_{km}\|^2$ **then**
        MS k feeds back h_{kj} to its Master BS m
      else
        MS k does not feed back h_{kj} to its Master BS m
      end if
    end for
  end for

By ranging a threshold (either an absolute or a relative one), the average radio signaling overhead can be controlled (see figure 1). Each value of the average number of fed back coefficients corresponds to a threshold value (figures 2,3). It also corresponds to an average value of user data streams needed per BS (figure 5). The latter is an indication of the backhaul overhead. By this the backhaul burden and also the complexity can be predicted.

IV. Numerical Results

A cooperation cluster of three cells has been considered ($B = 3$). BSs are located in the centre of each cell. Each BS has one omnidirectional antenna ($M = 1$). The channel coefficient between the i-th MS and the j-th BS is

$$h_{ij} = \Gamma_{ij} \sqrt{G \beta d_{ij}^{-\alpha} \gamma_{ij}} \tag{19}$$

where d_{ij} is the distance in km of the i-th antenna and the j-th MS. α is the pathloss exponent and β the pathloss constant. γ_{ij} is the corresponding log-normal coefficient which models the large-scale fading (shadowing), $\gamma_{dB} \sim \mathcal{N}(0\,dB, 8\,dB)$, and Γ is the complex Gaussian fading coefficient which models the small-scale fading, $\Gamma \sim \mathcal{NC}(0, 1)$. G is the BS antenna power gain which is assumed to be 9 dB (gain on the elevation). For the pathloss, the 3GPP Long Term Evolution (LTE) pathloss model has been used,

$$PL_{ij}^{dB} = 148.1 + 37.6 \log_{10}\left(d_{ij}^{km}\right). \tag{20}$$

MSs are selected in a round-robin fashion as we are interested in schemes that provide fairness (other scheduling techniques can be also considered). In figure 1 the average sum-rate is plotted as a function of the average number of channel coefficients fed back per MS for the proposed schemes (18). Results are compared with the case of static subclustering where the same subset of the overall number of BSs always forms a cooperation cluster and serves the users assigned to them. In the case of 3 BSs overall, there are three possibilities of static subclustering: no inter-BS cooperation, 2 of the 3 BSs cooperate, all the BSs cooperate. This corresponds to three distinct points on the curve. System SNR is the average SNR that a user experiences at the edge of the cell.

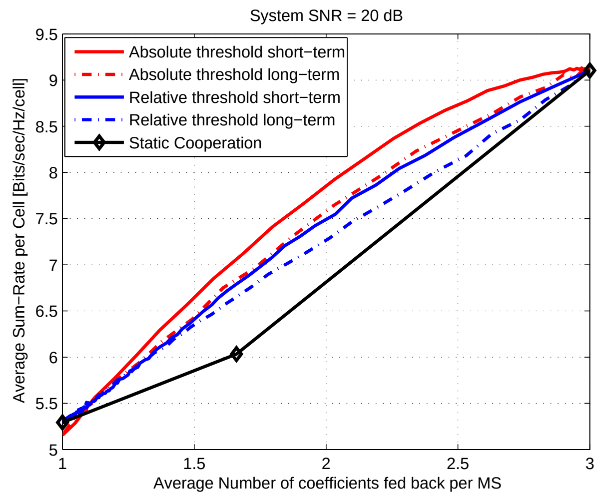

Fig. 1. A plot of the average sum-rate per cell versus the average number of fed back channel coefficients per MS for the proposed algorithms and also for static cooperation. The range of the average number of coefficients plotted is from 1 to 3 in order to compare absolute with relative thresholding.

It depends on the pathloss, antenna radiation patterns and the noise level. The plot shows the gain provided by the proposed algorithms in comparison to static cooperation. It can be seen that by increasing the number of the involved BSs, and therefore by increasing inter-base overhead, sum-rate performance increases and reaches its maximum when full cooperation is imposed. It is clear that absolute thresholding outperforms relative one. From figure 1, the radio signaling overhead of the system can be designed for a specific sum-rate target. The corresponding threshold values can be seen in figures 2 and 3.

In figure 4 the sum-rate performance is shown as a function of the system SNR. It can be clearly seen that the value of cooperation increases in the high SNR regime, since the system becomes more interference limited. Finally, figure 5 shows the relation between radio signaling reduction and the number of transmitted data streams per BS (15), a quantity which can impact backhaul overhead.

V. Conclusion

Multi-cell cooperative processing although very promising for future cellular systems, comes at the cost of increased overhead and complexity. On the downlink of FDD systems, MSs need to estimate and feed back several channel coefficients and BSs need to be interconnected in order to exchange CSI and user data for performing the needed signal processing operations. This creates the need for reducing signaling burden in order for the overhead to be practically affordable. A way for achieving this is by developping effective ways of forming cooperation clusters of a limited size, a fact that necessarily reduces information exchange. In addition, the intra-cluster overhead needs to be further reduced. In this contribution a novel technique has been proposed that effectively prevents MSs from feeding back channel coefficients related to BSs

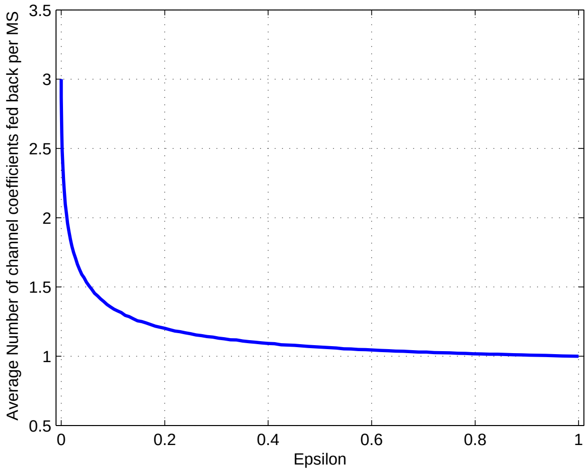

Fig. 2. Relative thresholding: a plot of the average number of fed back channel coefficients per MS versus ϵ. It is not a function of System SNR.

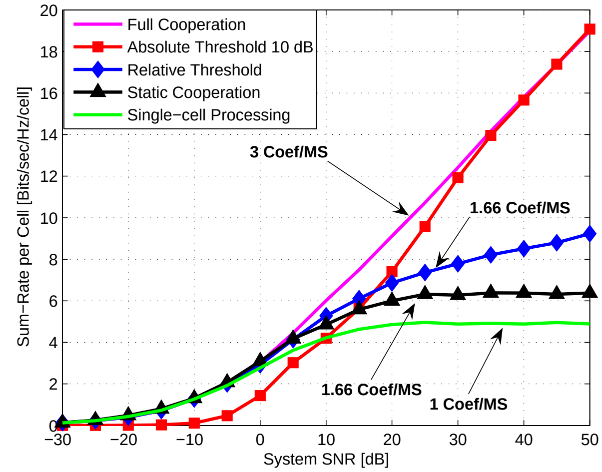

Fig. 4. A plot of the average sum-rate per cell versus the system SNR. All curves, appart from the one related to the absolute threshold, correspond to a specific value of average number of fed back coefficients per MS.

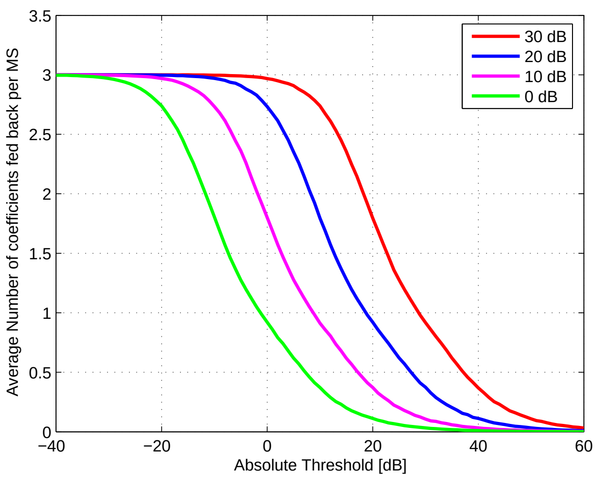

Fig. 3. Absolute thresholding: a plot of the average number of fed back channel coefficients per MS versus the absolute threshold for different values of System SNR.

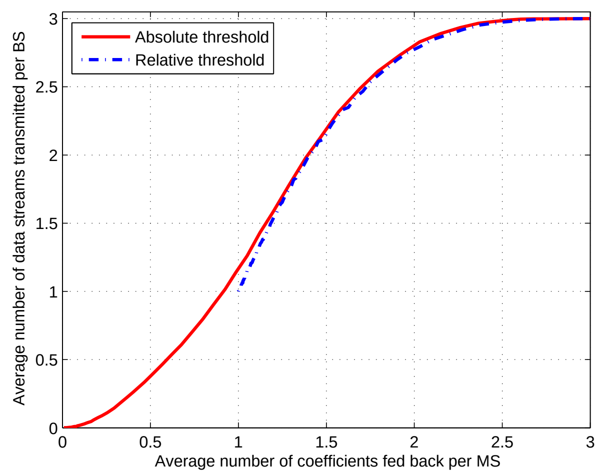

Fig. 5. A plot showing the relation between radio signaling reduction and the number of transmitted data streams per BS resulting from it.

which do not provide acceptable channel quality. The acceptability of the quality of a channel is determined by an absolute or a relative threshold. The decision can be made either by taking into account the instantaneous fading state of the channel or the long-term information. The latter reduces significantly the algorithmic complexity and the channel estimation load. It has been shown that this technique provides a good trade-off between sum-rate performance and overhead load of the network.

REFERENCES

[1] M.K. Karakayali, G.J. Foschini and R.A. Valenzuela, "Network Coordination for Spectrally Efficient Communications in Cellular Systems", *IEEE Wireless Communications Magazine*, August 2006.

[2] H. Zhang and H. Dai, "Cochannel Interference Mitigation and Cooperative Processing in Downlink Multicell Multiuser MIMO Networks," *EURASIP Journal on Wireless Communications and Networking*, pp. 222-235, February 2004.

[3] F. Boccardi and H. Huang, "Limited Downlink Network Coordination in Cellular Networks," in Proceedings of the *18th IEEE International Symposium on Personal*, *Indoor an d Mobile Radio Communications* (PIMRC 2007), September 2007, Athens, Greece.

[4] S. Venkatesan, "Coordinating Base Stations for Greater Uplink Spectral Efficiency in a Cellular Network", in Proceedings of the *18th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications* (PIMRC 2007), September 2007, Athens, Greece.

[5] A. Papadogiannis, D. Gesbert and E. Hardouin, "A Dynamic Clustering Approach in Wireless Networks with Multi-Cell Cooperative Processing", in Proceedings of the *IEEE International Conference on Communications*, (ICC 2008), May 2008, Beijing, China.

[6] D. Gesbert and M.S. Alouini, "How Much Feedback is Multi-User Diversity Really Worth?", in Proceedings of the *IEEE International Conference on Communications*, (ICC 2004), June 2004, Paris, France.