Achieving Sum-Capacity of the MIMO BC with Large Transmit Array using One-Shot Scalable Feedback Protocol

Rajiv Agarwal and John Cioffi
Department of Electrical Engineering
Stanford University, CA-94305
{rajivag,cioffi@stanford.edu}

***Abstract*— We consider a MIMO broadcast channel (BC) with large M and K, where M is the number of transmit antennas and K is the number of single-antenna users; and propose a scalable feedback protocol that achieves the sum-capacity with limited feedback of channel state information (CSI). In our earlier work, we showed that if feedback time slots correspond to channel gains and not to users, the sum-capacity $M \log\log K$ can be achieved with feedback resources growing only as $M \log K$, unlike linearly as MK for conventional schemes. In this work, we show that feedback requirement can further be reduced by half or more, while still achieving the sum-capacity. A scalable feedback (SF) protocol using Random Beamforming (RBF) is proposed, which a) requires finite number of feedback slots upper bounded by a constant, b) is fully distributed, c) needs finite transmission energy during feedback and d) achieves the sum-capacity. Numerical results show that feedback load is substantially reduced as compared to conventional schemes, as well as our own previously proposed SF scheme.**

I. Introduction

Multiple-input multiple-output (MIMO) systems have great potential to achieve high throughput in wireless systems [1]. In a BC, multiple antennas can be easily deployed at the base station to enhance system capacity. The capacity achieving strategy for MIMO BC's is dirty paper coding (DPC) and it was shown in [2] that the sum-capacity of the MIMO BC scales as $M \log(\log(K))$, where M is the number of transmit antennas at the base station, K is the total number of single-antenna users in the system and each entry of the vector channel is i.i.d zero mean circularly symmetric complex Gaussian (ZMCSCG) distributed for all the users. Thus the sum-capacity of the MIMO BC scales linearly with the number of transmit antennas (multiplexing gain) and only as $\log(\log(K))$ with the number of users (diversity gain). Although the sum-capacity of DPC with perfect CSIT can be approached by Tomlinson Harashima precoding [3], its large computational complexity hinders its application in practice. Recently, a number of techniques have been proposed [4], [5], [6], which are much simpler than dirty paper coding in terms of transmit complexity and still achieve the same $O(M \log(\log(K))$ asymptotic sum-capacity although they require perfect CSIT as well. If the feedback link has limited rate, it was shown in [7] that feeding back only one bit per user per transmit antenna is asymptotically optimal for achieving sum-capacity. Also, for ZFBF, it was shown in [8] that only a finite number of bits per user achieves sum-capacity. However, in all the schemes, all users feedback their channel, hence feedback requirement grows linearly with both K and M. The sum-capacity grows only as $M \log\log K$; so for large M and K, the uplink channel for CSI feedback may consume most of the system resources.

In the above mentioned schemes for achieving sum-capacity of the MIMO BC, CSI from all users is required at the base station. Consequently the amount of feedback (measured as total number of bits) increases linearly with the K, the total number of users in the system; even if each user feeds back just one bit for CSIT. It was shown in [9], a feedback method based on opportunistic p-persistent CSMA [10] can limit the number of users who feedback to be a constant that is independent of K in a single antenna system. In a recent result [11], we proposed SF protocols for the MIMO BC that achieve the asymptotic sum-capacity with only a finite number of users (independent of K) feeding back their CSI to the base station. To generalize proofs in [9], a new method for calculating threshold was provided. In addition, we proposed an SF protocol with ZFBF (SF-ZFBF) which uses semi-orthogonal user selection as proposed in [8] and achieves improved sum-rate for small K as compared to SF protocol with random beamforming (SF-RBF). For the SF protocols in [11], the feedback load was shown to increase as $M \log K$, unlike as MK for conventional schemes.

In this paper, a one-shot SF protocol using RBF (OSF-RBF) is proposed that achieves sum-capacity of the MIMO BC with the feedback resources reduced by half or more in comparison to that for the SF protocols in [11]. In order to improve performance, a new way for the calculation of the thresholds is provided. Also, the SF protocol is modified to reduce control signal overheads by a factor of M, which is substantial when along with K, M too is large. The proposed OSF-RBF protocol retains the optimality properties of the previously proposed SF protocols [11]. Specifically, it is shown analytically that a) the expected number of feedback slots is independent of K and bounded above by a small constant which is a design parameter, b) the energy consumption during CSI feedback is independent of K as well and bounded by a small constant and c) the asymptotic sum-capacity of the MIMO BC is achieved. Besides asymptotic optimality in the regime of large K, interesting tradeoffs between the *actual* values of the three objectives a), b) and c) are also pointed out for the benefit of a system designer. Numerical results show that the proposed OSF-RBF protocol achieves the asymptotic sum-rate capacity $O(M \log(\log(K))$ with very little feedback resources.

Notation: $\mathbb{E}$ denotes statistical expectation. $P\{\mathcal{S}\}$ denotes the probability of an event $\{\mathcal{S}\}$. $\mathcal{C}^{x\times y}$ denotes the matrix of dimension $x \times y$ and with entries being complex numbers. $\|\mathbf{x}\|$ denotes the ℓ^2-norm of a vector $\mathbf{x}$. $f = O(\phi)$ means that $|f| < A\phi$ for some constant A and for all n with a sequence f_n.

II. Broadcast Channel Model

We consider a MIMO broadcast channel with M transmit antennas at the base station and $K \geq M$ users. For simplicity, we assume users with single receive antenna. We assume users are homogenous and experience flat Rayleigh fading. The signal received by a user k can be represented as

$$y_k = \mathbf{h}_k \mathbf{x} + z_k, \quad , k = 1, 2, \ldots, K \tag{1}$$

where $\mathbf{h}_k \in \mathcal{C}^{1\times M}$ is the channel gain vector with i.i.d. ZMCSCG entries. Hence, the squared-norm of the channel or the channel gain is distributed as $Gamma(M, 1)$ for all the users[1]. The transmit vector is denoted by $\mathbf{x} \in \mathcal{C}^{M\times 1}$ which contains the information symbols for the chosen users, with an average power constraint $\mathbb{E}||\mathbf{x}||^2 \leq P$ and z_k is the additive Gaussian noise with unit variance. The set of chosen users is denoted by $\mathcal{S}$. To model temporal channel variations, a slowly-varying block-fading channel is assumed i.e. the channel vector is unchanged during a feedback period plus a BC packet transmission period (defined in Section III), and it changes independently of the previous channel realization after a packet transmission.

The transmit symbol vector $\mathbf{x}$ is related to information symbols $s_i, i \in \mathcal{S}$, via linear beamforming $\mathbf{x} = \sum_{i\in\mathcal{S}} \mathbf{w}_i s_i$. Therefore, the received symbol at one of the chosen users can be written as

$$y_k = (\mathbf{h}_k \mathbf{w}_k) s_k + \sum_{j\in\mathcal{S}, j\neq k} (\mathbf{h}_k \mathbf{w}_j) s_j + z_k, \quad k \in \mathcal{S} \tag{2}$$

The user set $\mathcal{S}$ is chosen to maximize the sum-rate as described in Section IV.

III. Feedback Protocol

This section proposes a feedback protocol, which will be shown in the next section to maintain the amount of feedback a constant and to achieve the asymptotic sum-rate capacity.

The operation of the OSF-RBF protocol is shown in Figure 1.

The uplink feedback channel is divided into short slots. All the users are synchronized to the base station such that any feedback transmission initiates precisely at the beginning of a slot. Slot duration is long enough to transmit one user's CSI x_k, $k = 1, 2, \ldots, K$ without packet errors. The base station chooses M random orthonormal directions $\mathbf{w}_m^{\mathrm{RBF}}, m = 1, 2, \ldots, M$ from an isotropic distribution as proposed in [4]. It then broadcasts a training sequence with pilots in each of

[1]Norm-squared of each entry of the channel vector is exponentially distributed and sum of exponential random variables is the *Gamma* distribution.

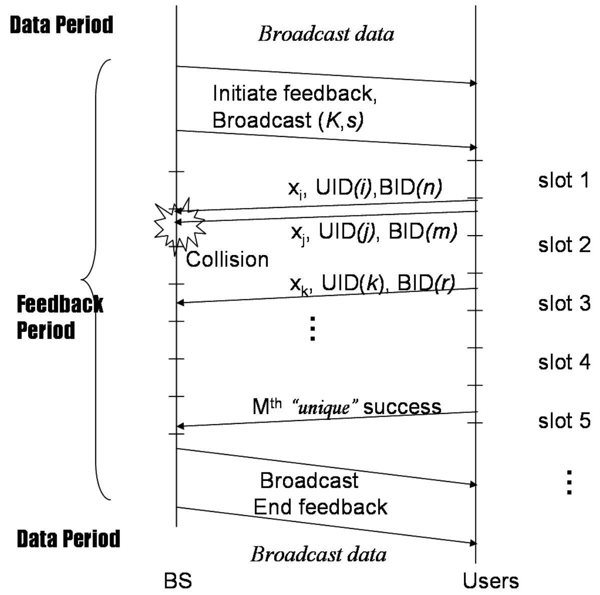

Fig. 1. The feedback period of the OSF-RBF protocol.

the chosen direction, which is used by the user k to estimate M effective SINR's

$$\begin{aligned} \mathrm{SINR}_{k,m} &= \frac{\mathbf{w}_m^{\mathrm{RBF}*}\mathbf{h}_k}{\frac{1}{\rho} + \sum n \neq m \mathbf{w}_n^{\mathrm{RBF}*}\mathbf{h}_k}, \\ k &= 1, \ldots, K, m = 1, \ldots, M \end{aligned} \tag{3}$$

where $\rho = \frac{P}{M}$. Subsequently, the base station broadcasts an *Initiate Feedback* message to begin the feedback period. The total number of users K and the non-collision probability s are also attached to this message for the users to calculate the thresholds $T_i, i = 1, \ldots, N$. The relationship between T_i and K, s will be discussed later in this section. If no user sends a feedback message in a slot, the feedback period continues till the base station receives CSI from a user successfully, where success happens when there is no collision. At slot i, any user k compares its channel gain $\mathrm{SINR}_{k,m}, \forall m = 1, \ldots, M$ with the thresholds T_i and T_{i-1}. If $T_i < \mathrm{SINR}_{k,m} < T_{i-1}$ for any m, then user k transmits its CSI x_k in slot i. The CSI x_k is the unique user ID along with the beam number (for which the number of bits needed will be $\log_2(MK)$). If two users, j and k, are such that $T_i < \mathrm{SINR}_{j,m}, \mathrm{SINR}_{k,n} < T_{i-1}$ for any two beam directions m, n, then their feedback messages will collide. These collided messages are assumed to be irrecoverable for simplicity of the strategy, although the base station may decode both their messages by standard MAC decoding strategies like successive cancelation. If only one user transmits a feedback message in slot i, the base station decodes x_k successfully. The base station waits to successfully recieve a user for each beam $m = 1, \ldots, M$ and thereby broadcasts an *End Feedback* message to stop the feedback period. Finally the base station transmits data to the selected users with a capacity achieving code and transmit beamforming using the same randomly selected M beams.

Analysis that the SF-RBF protocol achieves the asymptotic sum-capacity is done in Section IV.

Notice that the SINR values used for calculation of transmission rate at the base station are T_i, where $T_i < \text{SINR}_{k,m} < T_{i-1}$ as the actual SINR value (unquantized or quantized) is not transmitted. If the probability of collisions reduces to 0 exponentially with KM, the SF protocol 1) selects the user with the highest channel gain, 2) with its channel gain known to a high accuracy, in an autonomous and distributed manner. Though proved rigorously in Section IV, it can be noticed that if the number of thresholds (denoted by N) is large, both 1) and 2) are achieved.

If the distribution function for the SINR values (same for all users since homogenous fading) is given by $F_\Gamma(\gamma)$, then the probability that a user transmits its feedback message in the ith slot for any of the M beams is given as

$$P\{T_i < \gamma < T_{i-1}\} = F_\Gamma(T_{i-1}) - F_\Gamma(T_i), \quad i = 1, 2, \ldots, N \tag{4}$$

To maintain the probability that a user transmits a feedback message for any beam in any slot to be p, we require $P\{T_i < \gamma < T_{i-1}\} = p, \forall i$. The transmission probability at the 1st slot is

$$P\{T_1 < \gamma\} = 1 - F_\Gamma(T_1) = p \tag{5}$$

and so the transmission threshold for any slot i is given as

$$T_i = F_\Gamma^{-1}(1 - ip), i = 1, 2, \ldots, N \tag{6}$$

It is easy to see that $N = \frac{1}{p}$.

In [11], the SF protocol using Random beamforming was presented as an M fold extension of the SF protocol for the single-antenna case. Specifically, the feedback period was divided into M sub-feedback periods. Each sub-feedback period corresponded to a particular beam $m = 1, \ldots, M$. During a sub-feedback period, the users transmitted just their unique ID as x_k or the CSIT. During a particular sub-feedback period (say corresponding to $m = 3$), the users compared just the values of $\text{SINR}_{k,3}$ against the thresholds T_i and T_{i-1} and decided to transmit x_k or not. The one-shot SF protocol, on the other hand, has a single feedback period thus reducing the control overhead of coordinating M feedback periods a lot when M is large. When the number of users K is large, the maximum values of $\text{SINR}_{k,m}$ will be close-by for all m; and hence not just the control overhead, the actual number of feedback slots till success are also reduced. The number of bits required for x_k however increases from $\log_2(K)$ in [11] to $\log_2(KM)$. Since $K \gg M$, this increase is negligible. The performance improvement of the OSF-RBF protocol over SF protocols in [11] is shown in Section V.

IV. Performance Analysis

In this section, the achievable sum-rate and the amount of required feedback resources are analyzed. The consumption of feedback resources is measured by the expected number of slots per feedback period ($\mathbb{E}[S]$) and the expected number of feedback messages transmitted during a feedback period ($\mathbb{E}[M']$). Notice that $\mathbb{E}[M']$ is representative of the total energy spent by the users during the feedback period. Also, the sum-rate proof is given for any general distribution for the channel gain (or effective SINR as the case may be depending upon SF-RBF or SF-ZFBF is used) $F_\Gamma(\gamma)$. The $M \log\log(K)$ result follows when $F_\Gamma(\gamma)$ is the distribution derived in [4] for SINR[2].

Since each user compares its SINR for any of the M for transmission in a feedback slot, effectively we have KM users each with an SINR value. Hence on in this section, we will assume that there are KM effective users each with an SINR value. Since the users have independent fading, the probability that any user transmits in a given feedback slot i is p. Hence, by binomial probability, the probability that a single user transmits in the first feedback slot is

$$r = KMpq^{KM-1} \tag{7}$$

In case more than one user transmits or no user transmits, the SF protocol proceeds to the second feedback slot. This happens with probability $(1-r)$. Now in the second feedback slot, a single user transmits with the same probability r if no user transmitted in the previous slot or with a lower probability $(KM - n_1)pq^{(KM-n_1)-1}$, where $n_1 \geq 2$ users transmitted in the first feedback slot. As shown later in the section, for the choice of $q = s^{1/(KM)}$, the probability that two or more users transmit in a single slot goes to 0 exponentially in KM, where $s \simeq 1$. So we assume that the number of users remains the same after a feedback slot in the rest of this section. Under this assumption (valid for KM large), the number of feedback slots till successful reception is geometrically distributed with probability r. The feedback slot is terminated after M successes. Although, the success have to be unique, as in a success corresponding to each beam M; for the sake of analysis we assume that the first M successes is the same as first M unique successes, which holds fairly true when the SINR's are i.i.d. distributed. Numerical results presented in Section V corroborate this fact. Using the union bound for probability, the expected value of required feedback slots can then be given as

$$\mathbb{E}[S] \leq M\frac{1-r}{r} \tag{8}$$

Now using the fact that $s = q^K$, we can upper-bound $\mathbb{E}[S]$ as

$$\mathbb{E}[S] \leq M\frac{1}{s(1-s)} \tag{9}$$

where the details of the steps can be found in [11]. Thus we have shown that the expected number of feedback slots is upper bounded by a finite number independent of K.

Using the union bound argument, the proofs in [11] for the asymptotic sum-rate optimality and for showing that $\mathbb{E}[M'] \leq \frac{M}{s}$ can be arrived at in a straight forward manner and so are skipped for avoiding repetition and conciseness.

Notice that $\mathbb{E}[M']$ is less than $\mathbb{E}[S]$ as $s < 1$. This happens because q (the probability of no transmission) goes to 1 as $K \to \infty$ and so slots with no transmission are more frequent

[2]The CDF of SINR in [4] was derived to be $F_\Gamma^{\text{RBF}}(\gamma) = 1 - \frac{e^{-\gamma/\rho}}{(1+\gamma)^{M-1}}$.

as we want the probability that two or more users transmit in a single slot to go to 0 exponentially in KM. Due to this, when the user with the highest $\text{SINR}_{k,m}$ for any m, transmits in a feedback slot, there is no collision with high probability and the SF protocols in [11] perform the same as a scheme with perfect CSIT. Thus, by keeping q very high, although we get high sum-rate and low $\mathbb{E}[M']$, due to the large number of idle slots, $\mathbb{E}[S]$ goes up. In a sense this is over-partitioning of the SINR pdf as we are fine with empty slots but do not want slots with two or more users. Based on this observation, the expected number of feedback slots can be reduced in comparison to OSF-RBF. For this we consider a scheme where KM is replaced by K during the calculation of q i.e. $q = s^{1/K}$ instead of $q = s^{1/KM}$. This decreases the expected number of feedback slots required by reducing the idle slots, at the same time increasing the chances of two or more users in a single slot. The proof for boundedness of $\mathbb{E}[S]$ and $\mathbb{E}[M']$ and for sum-rate optimality still continue to hold as s is fixed; though the actual value of $\mathbb{E}[S]$ is considerably reduced as seen in Section V.

The over-partitioning of the SINR pdf can also be avoided by maximizing r for given K, M. Specifically, each user computes p as

$$p = \arg \max_{z \in [0,1]} (KM) z (1-z)^{KM-1} \quad (10)$$

and then $q = 1 - p$. Firstly, in this case, the *Initiate Feedback* message just needs to carry K and there is no s. Also, the probability of success or a single user transmitting in a feedback slot is explicitly maximized; hence the expected number of feedback slots till M successes is reduced. Unfortunately, this scheme is hard to analyze and only numerical results for it are presented in Section V.

In the next section, we refer to these two heuristic schemes as OSF-RBF-Mod1 and OSF-RBF-Mod2 respectively, where Mod stands for modified. Also, in our efforts to reduce the expected number of feedback slots by reducing the idle slots, we pay the price of a higher $E[M']$ and a lower sum-rate. This is because now the slots with two or more transmissions will increase leading to comparatively more collisions. These design trade-offs are important for actual system implementation and one or the other solution can be preferred by the designer.

V. Numerical Results

In this section, numerical results are presented. In Figure 2, we plot the sum-rate performance of the proposed OSF-RBF protocols as function of the number of users K for $s = 0.95$, $P = 10$dB and $M = 10$. We compare the performance of the SF-RBF protocol with that of the RBF with unquantized SINR feedback from all the users as proposed in [4]. It is seen that the OSF-RBF protocols lose very little performance in terms of sum-rate by just the user and beam ID feedback. The performance loss in comparison to the SF-RBF protocol in [11], which requires M times more control overhead, is also negligible. Also visible is the performance loss of the OSF-RBF-Mod1 and OSF-RBF-Mod2 protocols in comparison to the OSF-RBF protocol. The sum-capacity asymptote $M \log \log(K)$ is also plotted and as seen all the schemes achieve asymptotic sum-capacity.

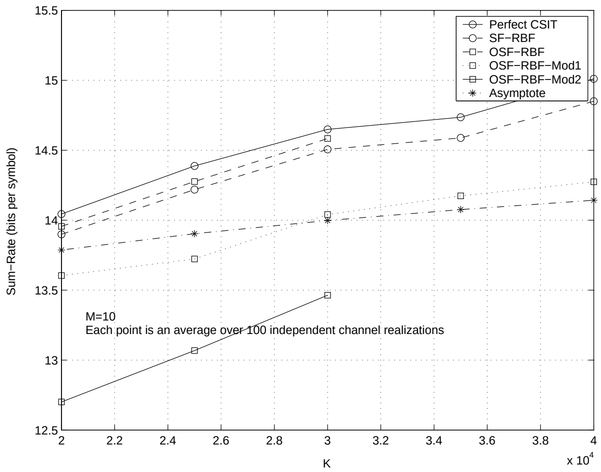

Fig. 2. Sum-rate comparison of the proposed OSF-RBF protocols for M=10.

In Figure 3, we plot $\log_2(MK)\mathbb{E}[S]$ or the required feedback resources for the proposed OSF-RBF protocols as function of the number of users K for $s = 0.95$, $P = 10$dB and $M = 10$ normalized by M. The growth rate of feedback resources for conventional schemes (linearly as MK) and for the SF-RBF in [11] (as $M \log_2(K)\mathbb{E}[S]$) is also plotted for the sake of comparison, all normalized by M. As seen the OSF-RBF-Mod1 scheme performs the best in terms of feedback reduction, closely followed by the OSF-RBF-Mod2 scheme. Thus we see that the proposed OSF-RBF protocol reduces the amount of feedback required substantially as compared to conventional schemes and by more than 50% as compared to our own previously proposed SF-RBF scheme in [11].

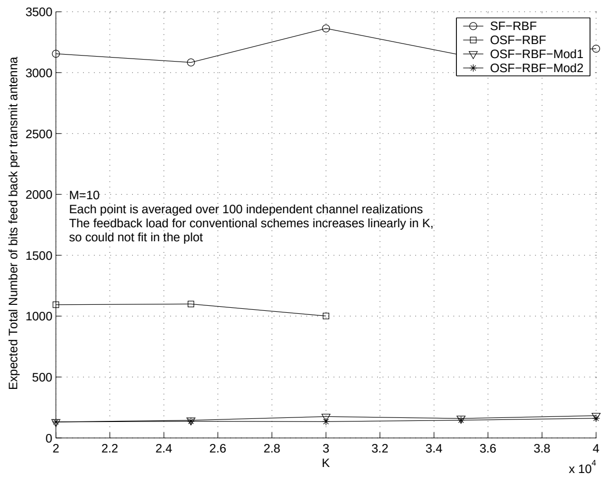

Fig. 3. Feedback resource comparison of the proposed OSF-RBF protocols for M=10.

In Figure 4, we plot the required feedback resources for the proposed OSF-RBF protocols as function of the number of transmit antennas M for $s = 0.95$, $P = 10$dB and $K = 20,000$. For the SF-RBF protocol in [11], $\mathbb{E}[S]$ normalized by M is a constant close to $\frac{1}{s(1-s)}$ as expected or $\mathbb{E}[S]$ grows linearly with M. The OSF-RBF-Mod1 scheme reduces $\mathbb{E}[S]$ considerably at the cost of increasing $\mathbb{E}[M']$. Also, as M increases, we have more effective users ($\sim KM$) within any feedback slot because for the OSF-RBF-Mod1 scheme, thresholds are a function of K only. Hence the number of idle slots decreases as M increases, resulting in decreasing $\mathbb{E}[S]$ and increasing $\mathbb{E}[M']$. Hence, the feedback resources for the OSF-RBF-Mod1 protocol grows sub-linearly than $M \log_2(K)$. The OSF-RBF-Mod2 scheme reduces $\mathbb{E}[S]$ substantially in the comparison of SF-RBF and OSF-RBF-Mod1 at the price of further increase in $\mathbb{E}[M']$ by an explicit maximization of r. Also, notice the design tradeoff. Moving from SF-RBF to OSF-RBF-Mod1 to OSF-RBF-Mod2, we see that $\mathbb{E}[S]$ can be reduced more and more at the price of increasing $\mathbb{E}[M']$ (and reduced sum-rate as seen in Figure 2).

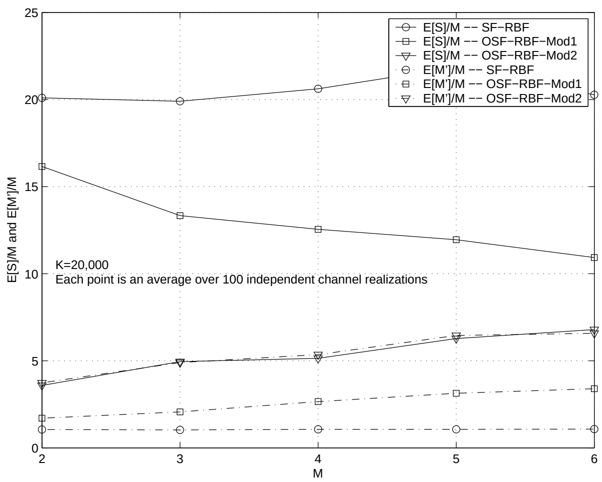

Fig. 4. Feedback resource comparison of the proposed OSF-RBF protocols for K=20,000.

In Figure 5, we plot the required feedback resources for the proposed OSF-RBF protocols as function of M for $s = 0.95$, $P = 10$dB and K increasing with M at a rate faster than e^M, so that RBF scheme is sum-rate optimal [4]. Specifically for $M = \{4, 5, 6, 7, 8, 9, 10\}$, $K = \{500, 1000, 1500, 2000, 5000, 10000, 20000\}$ respectively. By a comparison of the figures 4 and 5, we see that the feedback resources for all the SF protocols is independent of K.

VI. Acknowledgement

The authors wish to thank Chan-Soo Hwang for helpful discussions regarding Scalable Feedback protocols.

VII. Conclusion

This paper describes one-shot scalable feedback protocols, which is shown to asymptotically achieve the MIMO-BC sum-capacity while the amount of feedback (the expected number

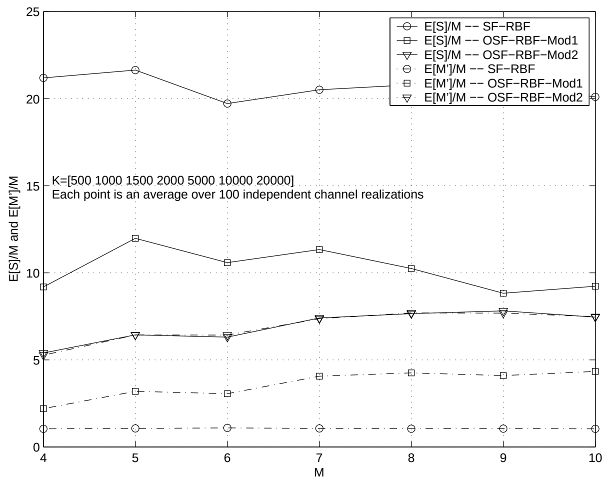

Fig. 5. Independence of feedback resource w.r.t. to K.

of slots and power consumptions) normalized by M, the number of transmit antennas, is shown to reduce with M. The feedback amount is also always less than a constant independent of the number of users K. Although the proposed protocols achieve asymptotic sum-rate optimality, interesting design tradeoffs between the actual values of various design objectives - sum-rate, feedback slots and power consumption during feedback were discussed and OSF-RBF protocols that achieve them were proposed. Numerically, it was shown that the proposed OSF-RBF protocol can reduce the required feedback resources by 50% or more as compared to SF-RBF protocol in [11].

References

[1] E. Telatar, "Capacity of multi-antenna gaussian channels," *Europ. Trans. Telecommun. (ETT)*, vol. 10, pp. 585–596, Nov. 1999.

[2] M. Sharif and B. Hassibi, "A comparison of time-sharing, dpc, and beamforming for mimo broadcast channels with many users," *IEEE Trans. Comm.* submitted for publication.

[3] W. Yu and J. Cioffi, "Sum capacity of gaussian vector broadcast channels," *IEEE Tran. on Info. Th.*, vol. 50, pp. 1875–1892, Sep. 2004.

[4] M. Sharif and B. Hassibi, "On the capacity of mimo broadcast channels with partial channel state information," *IEEE Trans. on Information Theory*, vol. 51, pp. 506–522, Feb. 2005.

[5] T. Yoo and A. Goldsmith, "Optimality of zero-forcing beamforming with multiuser diversity," in *IEEE ICC*, pp. 542–546, May 2005.

[6] A. Bayesteh and A. K. Khandani, "On the user selection for mimo broadcast channels," in *IEEE ISIT*, pp. 2325–2329, Sept. 2005.

[7] S. Sanayei and A. Nosnatinia, "Exploiting multiuser diversity with only 1-bit feedback," in *IEEE WCNC*, pp. 978–983, Mar. 2005.

[8] T. Yoo, N. Jindal, and A. Goldsmith, "Multi-antenna broadcast channels with limited feedback and user selection," *IEEE JSAC 2006*. Submitted.

[9] C. Hwang and J. Cioffi, "Achieving asymptotic broadcast channel sum-capacity with fixed feedback amount for scalable networks," *IEEE JSAC 2006*. Submitted.

[10] C. Hwang, K. Seong, and J. Cioffi, "Opportunistic p-persistent csma in wireless networks," in *IEEE ICC*, June 2006.

[11] R. Agarwal, C. Hwang, and J. Cioffi, "Scalable feedback protocol for achieving sum-capacity of the mimo bc with finite feedback," in *IEEE ICC*, 2007 (submitted).