

Research Article

How Much Feedback Is Required for TDD Multi-Antenna Broadcast Channels with User Selection?

Umer Salim and Dirk Slock

Mobile Communications Department, EURECOM, BP 193, F-06904 Sophia Antipolis Cedex, France

Correspondence should be addressed to Umer Salim, umer.salim@eurecom.fr

Received 7 October 2009; Revised 4 February 2010; Accepted 5 May 2010

Academic Editor: Ana Pérez-Neira

Copyright © 2010 U. Salim and D. Slock. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The enormous gains in a multi-antenna transmitter broadcast channel require the Channel State Information at the Transmitter (CSIT). Although the fundamental question “How much feedback is required for a broadcast channel?” has been treated in the literature to some extent, a more comprehensive treatment is certainly desirable. We study the time-division duplex broadcast channel with initial assumption of channel state information (CSI) neither at the base station (BS) nor at the users’ side. We provide two transmission strategies through which the BS and the users get necessary CSI. We derive novel lower and upper bounds for the sum rate reflecting the rate loss compared to a perfect CSIT system. Corresponding approximate sum rate expressions are also developed for both schemes. These expressions fully capture the benefits of the CSIT feedback, allowing multi-user diversity gain and better inter-user interference cancellation, and the cost of exchange of information required. These expressions can be optimized for any set of system parameters to unveil the trade-off between the cost and the gains associated to feedback. Thus they allow to characterize the optimal amount of feedback which maximizes the sum rate of the broadcast channel, a well-accepted metric of system performance at the physical layer.

1. Introduction

1.1. Background and Motivation. In a broadcast channel (BC) having a base station (BS) equipped with n_t transmit antennas and K ($K \geq n_t$) single antenna users, the dominant term of the sum capacity is $n_t \log(\text{SNR})$ [1–3], where SNR denotes the signal-to-noise ratio of the received signal and n_t is called the multiplexing gain. The BC enjoys another gain, coined as multiuser diversity [4], which is due to the possibility of user selection from a larger ($K > n_t$) pool of users. It has been shown in [5, 6] that the sum capacity of the Gaussian broadcast channel scales with the number of users as $n_t \log(\log(K))$, where K is the total number of users in the system whose channel information is available at the BS. Apart from the multiplexing gain and the multiuser diversity benefit, the BC enjoys two other advantages over the single user multiple-input multiple-output (SU MIMO) channel. It allows mobile users to have a single antenna each so user terminals can remain quite inexpensive and simple. The second advantage is that the BC channel matrix

is often much better conditioned as compared to that of an SU MIMO link which may suffer from line-of-sight channel conditions and strong spatial correlation [7]. These same advantages typically continue to hold compared to multiuser (MU) MIMO systems in which the total number of receive antennas equals or exceeds n_t .

These promising advantages of broadcast MIMO do not come for free as without channel state information at the transmitter (CSIT) and perfect channel state information at the receiver (CSIR), the dominant term of the sum capacity is only $\log(\text{SNR})$ because of the optimality of transmitting to a single user in that case [8–10]. Thus the CSIT of n_t users is indispensable to achieve the full multiplexing gain [1, 3]. Furthermore, to capture the multiuser diversity benefit of $n_t \log(\log(K))$ in the sum rate, the BS should know the CSI of all these K users, where normally K could be much larger than n_t .

With perfect CSIT, [11] shows that zero-forcing (ZF) precoding achieves the full multiplexing gain of n_t and the full multiuser diversity gain of $n_t \log(\log(K))$ of the broadcast

channel if the number of users is asymptotically large, although the optimal transmission strategy for the Gaussian BC is dirty paper coding (DPC) [12]. In [5], the authors introduced a very innovative scheme coined as Orthogonal Random Beam Forming (ORBF), where only a few bits of feedback are required from every user and the sum rate was shown to converge to the optimal DPC sum capacity [6] for a large number of users.

There is an enormous volume of research publications analyzing CSIT acquisition techniques and the associated feedback gains in different scenarios but the fundamental issue, which is usually ignored, is the feedback overhead cost of providing CSI to the BS which leads to reduced sum capacity. Both the gain and the acquisition overhead increase with the amount of feedback but there is an optimal operating point (optimal amount of feedback) that maximizes the net gain. To the best of our knowledge, there is no single contribution that has properly analyzed the net gain of feedback in a general broadcast setting (with $K > n_t$) which can be defined as the gain in downlink (DL) sum rate due to feedback, taking into account the uplink (UL) feedback load. For a given feedback load, the problem boils down to a secondary problem, namely, partitioning of this feedback, either to improve the CSIT quality (better interference cancellation for a given user selection) or to get the multiuser diversity gain (select from a larger pool). We take a step back and address the more fundamental problem of how much feedback should be there to achieve the feedback gain-cost tradeoff. A very simple example showing the importance of this absolute gain would be the ORBF transmission scheme which requires as little feedback as $\log(n_t)$ bits plus a scalar from each user in the system, but considering the fact that ORBF requires the presence and the feedback from a large number of users, the absolute gain would become questionable.

The second fundamental aspect which often gets overlooked in the analysis of multiuser systems is the consideration of channel coherence time. Wireless channels have a finite coherence time and when multiuser transmission strategies with multiple rounds of training, feedback and data are devised, there is the possibility that the channel has significantly changed during the preliminary training and feedback intervals and that the channel information attained during these phases has become meaningless.

1.2. Contribution. In this paper, we do not make any assumption of CSI. Hence initially, the BS and the users are ignorant of the channel realization but they can estimate/feedback the CSI as is done in practice. To analyze the cost incurred and the attainable benefit of feedback in a meaningful and tractable fashion, we simplify the problem by selecting a time-division duplex (TDD) broadcast channel with perfect reciprocity. TDD reciprocal channels simplify the CSIT acquisition through UL pilot transmission [13, 14] in contrast to frequency-division duplex (FDD) systems in which the users first estimate the DL channel and then send its quantized version in an UL slot. We restrict the CSIT acquisition to be training based only, thanks to TDD reciprocity [13, 14]. In the sequel, we use the terms training

and feedback synonymously due to this restriction [14]. So we have a fixed resource (bandwidth and time slot) available, a BS having n_t transmit antennas and K ($K > n_t$) single antenna users. Now this fixed resource can be used for UL/DL data transmission or training/feedback. We assume that the users have no data to transmit in the UL direction. So the UL is solely reserved for channel training/feedback. But due to the TDD mode of operation, any UL transmission will come at the expense of having reduced DL transmission in the overall time slot, hence the cost of training/feedback gets properly accounted for. In this paper, we propose two transmission schemes. In the first scheme, the users, who feed back, are chosen independently of their channel realizations (whence oblivious users). In the second scheme, the users first learn their channel information and decide to feedback based upon their channel realizations (whence informed users). We derive a novel lower bound for the sum rate, capturing the gains and the costs of the CSIT acquisition, which shows explicitly the rate loss with respect to (w.r.t.) a perfect CSI system. We furthermore introduce a sum rate upper bound, that turns out to be closely related to the lower bound. The simplified expressions obtained for the two schemes allow us to maximize the DL sum rate (the performance metric considered here) achieving the cost-benefit tradeoff of the feedback.

1.3. Related Work. Caire et al. [15] studied the achievable rates for multiuser MIMO DL removing any assumption of CSIR or CSIT for FDD systems. They gave transmission schemes incorporating all the necessary training and feedback stages and compared achievable rates for analog and digital feedback schemes. This work was conducted under the assumptions of extremely large channel coherence lengths (which permits to neglect the training and feedback overhead) and of a number of users (K) equal to n_t . Later in [16], training and feedback parameters were optimized as a function of channel coherence time and SNR, although the number of users was still restricted to n_t . In [15], the sum rate for a system with training-based CSIT is lower bounded in terms of the sum rate of a related system with perfect CSIT. We provide here a novel similar lower bound in which the channel distribution corresponding to the perfect CSIT is actually that of the estimated channels, thus providing a much tighter lower bound compared to the one in [15], where the usual perfect CSIT channel distribution is that of the true channel. Furthermore, our novel lower bound is much easier to derive than the one in [15].

In [17], the authors analyze the tradeoff between multiuser diversity and the accuracy of quantized channel information at the BS. Under the restriction of a fixed number of feedback bits, they conclude that accurate channel information is more important than having multiuser diversity.

In another recent work [18] treating essentially UL systems, the authors considered chunk size optimization (amount of feedback) taking UL and DL of a SU MIMO channel into account.

References [13, 19] are related to our work as they also treat the TDD broadcast channel without any assumption

of CSI. But there are major differences in the scope. They treat the case when the number of users in the system is less than the number of BS antennas and try to exploit the channel hardening effect [20] due to a large number of BS antennas, which eliminates the multiuser diversity gain completely. Moreover, in both of these references, the users are never trained about their effective channels and the data is transmitted on the expected value of the effective channel. In contrast, our analysis is for systems with a larger number of users than BS transmit antennas because this setting is certainly more practical than its opposite counterpart. And in both of our transmission strategies (oblivious and informed users), the users are explicitly trained about their effective channel after precoding. The other major difference is in the achievable sum rate. The sum rate in [13, 19] saturates with DL SNR, giving zero multiplexing gain, even if DL and UL SNRs are of the same order, whereas our schemes achieve full multiplexing gain in this setting. A very recent related reference is [21] which is similar to [19] for the most part. Section VII of [21] gives a scheme similar to our scheme with oblivious users (see Section 3), but their sum rate lower bound, given in Theorem 3, involving four expectation operations, neither brings any insight about the sum rate behavior nor seems amenable to any further analysis.

1.4. Organization. This paper is structured as follows. First the system model is described in Section 2. Then in Section 3, the transmission scheme with oblivious users is detailed and the novel lower bound for the sum rate is derived. Section 4 gives the parallel developments for the scheme with informed users. The tightness of the sum rate lower bound and the accuracy of the approximate expressions are illustrated in Section 5. The behavior of the sum rate for oblivious users strategy under various asymptotic regimes is investigated in Section 6 followed by its counterpart for informed users in Section 7. The results for optimal feedback load (optimal number of users) with finite system parameters are explored in Section 8. Section 9 presents the conclusions and some directions for future research.

Notation. \mathbb{E} denotes statistical expectation. Lowercase letters represent scalars, boldface lowercase letters represent vectors, and boldface uppercase letters denote matrices. Moreover \mathbf{A}^\dagger denotes the Hermitian transpose of matrix \mathbf{A} . The identity matrix of n_t dimensions is denoted by \mathbf{I}_{n_t} . The logarithm with base 2 is denoted by $\log(\cdot)$. The cardinality of a set \mathcal{S} is expressed as $|\mathcal{S}|$.

2. System Model

The frequency-flat system we consider consists of a BS having n_t transmit antennas and K ($K > n_t$) single-antenna user terminals. In the DL, the signal received by k th user can be expressed as

$$y_k = \mathbf{h}_k^\dagger \mathbf{x} + n_k, \quad k = 1, 2, \dots, K, \quad (1)$$

where $\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_K$ are the (complex conjugated) channel vectors of users 1 through K , with $\mathbf{h}_k \in \mathbb{C}^{n_r \times 1}$ ($\mathbb{C}^{n_r \times 1}$

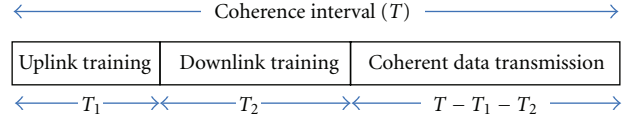


FIGURE 1: Transmission Phases for Oblivious Users.

denotes the n_t -dimensional complex space), $\mathbf{x} \in \mathbb{C}^{n_t \times 1}$ denotes the n_t -dimensional signal transmitted by the BS, and n_1, n_2, \dots, n_K are independent complex Gaussian additive noise terms with zero mean and unit variances. We denote the concatenation of the channels by $\mathbf{H}_F^\dagger = [\mathbf{h}_1 \mathbf{h}_2 \dots \mathbf{h}_K]$, so \mathbf{H}_F is the $K \times n_t$ forward channel matrix. The channel input from the BS must satisfy an (average) transmit power constraint of P , that is, $\mathbb{E}[\|\mathbf{x}\|^2] \leq P$. In this setting, the transmit power is equal to the true signal-to-noise ratio at each user due to normalized noise variances.

The channel is assumed to follow a block-fading model having a coherence length of T symbol intervals without channel variation, with independent fading from one block to the next [22]. The entries of the forward channel matrix \mathbf{H}_F are independent and identically distributed (i.i.d.) complex Gaussian with zero mean and unit variance. Due to the no CSI assumption, initially all the users and the BS are oblivious of the channel realizations in each block.

For the power constraint at the user terminals, we mainly treat the case of peak power constraint. The peak power per user per channel use is bounded by P_{pk} . The sum rate bounds for average power constrained users are provided in the appendix. The noise in the UL at the BS is also assumed to be spatiotemporally white complex Gaussian with unit variance.

3. Transmission Scheme with Oblivious Users

In this scheme, the users who feed back are unaware of their channel state. So they might be selected in a round-robin or any other fashion, independently of their channel realizations. For the block fading channel with coherence length of T symbol intervals, we divide this interval in three phases (see Figure 1): (1) uplink training, (2) downlink training, and (3) coherent data transmission with imperfect CSI. The first phase is the uplink training phase in which a certain number of users train the BS about their forward channels and the BS makes an estimate of the associated forward channel matrix. Based upon this channel information, the BS does the scheduling and chooses the transmit precoding which in general could be simple linear ZF, some nonlinear strategy like vector perturbation or the optimal DPC. The second phase is the downlink training phase, where the BS transmits pilots so that the scheduled users estimate their corresponding effective channels. When this second phase ends, both sides of the broadcast channel have necessary CSI, albeit imperfect. Hence in the third data phase, the BS transmits simultaneously to the selected users who can decode the data coherently.

Remark. When transmission is switched from UL to DL or vice versa, a guard interval must be inserted in practice. We do not take this guard interval into account as it does not affect our absolute feedback gain analysis.

Below we provide a detailed analysis of the three transmission phases and the necessary BS processing steps.

3.1. Uplink Training Phase. In a TDD system with perfect channel reciprocity, CSIT can be provided to the BS just by transmitting pilots from the users. The BS estimates the users' uplink channels and these are then also the forward channels due to the perfect reciprocity assumption. Suppose K^{obl} (superscript obl stands for oblivious users) of the K users transmit pilots, then the length of this uplink training interval is $T_1 = \beta K^{\text{obl}}$, where $\beta \geq 1$ should assume a real value so that T_1 is an integer ($\beta > 1$ can be used when we want a given number of users to transmit for more time and improve their channel estimates at the BS, $\beta < 1$ leads to significantly degraded channel estimates). Assuming orthogonal codes (which is optimal here) of length T_1 , the users can transmit simultaneously with transmit energy per user equal to $P_{\text{pk}} T_1$. As each antenna at the BS receives the transmitted code from a particular user through the channel coefficient which links this antenna to that user, the energy received for each channel coefficient (of CSIT) would be $P_{\text{pk}} T_1$. Assuming the BS employs MMSE channel estimation, the resulting channel coefficient estimation errors are i.i.d. complex Gaussian with zero mean and mean-square error (MSE)

$$\sigma_h^2 = \frac{1}{P_{\text{pk}} T_1 + 1} = \frac{1}{P_{\text{pk}} \beta K^{\text{obl}} + 1}. \quad (2)$$

For the k th user, with channel \mathbf{h}_k , the channel estimate is denoted as $\hat{\mathbf{h}}_k$ and the corresponding estimation error is $\tilde{\mathbf{h}}_k$ according to $\mathbf{h}_k = \hat{\mathbf{h}}_k + \tilde{\mathbf{h}}_k$. The $\hat{\mathbf{h}}_k$ also have i.i.d. complex Gaussian entries with zero mean and variance $1 - \sigma_h^2$ (due to the orthogonality property of MMSE estimation). Note that due to training code orthogonality, the channel estimation quality improves with the number of users K^{obl} .

The training length $T_1 \geq K^{\text{obl}}$ is basically the price of obtaining CSIT at the BS through feedback which reduces the effective channel coherence time to $T - T_1$. Hence, the CSIT acquisition from a very large number of users may be very suboptimal.

3.2. BS Transmission Strategy: ZF Precoding with Semiorthogonal User Selection. It is known that DPC allows to achieve the full capacity region of the MIMO broadcast channel [12] but this scheme is complex and its implementation is quite tedious. ZF linear precoding with user selection has been shown to behave quite optimally at high SNR achieving full multiplexing gain in the sum rate [23]. Furthermore in [11], the authors showed that ZF preceded by semiorthogonal user selection (SUS) achieves both the multiplexing gain and the multiuser diversity gain. SUS has been modified in [24] to work with imperfect CSIT in a robust manner. Due to its simplicity, analytical tractability and attractive performance,

we choose SUS and ZF precoding as the BS transmission strategy.

We adopt the SUS algorithm of [11] where user orthogonality is imposed at each selection stage. Suppose \mathcal{S} denotes the set of selected users having cardinality $|\mathcal{S}| = n_t$ and $\hat{\mathbf{H}}(\mathcal{S})$ denotes the BS estimate of the channel matrix of the selected users. In ZF precoding, the unit-norm beamforming vector for the k th selected user (denoted as $\bar{\mathbf{v}}_k$) is chosen to be orthogonal to the channel vectors of all other selected users, that is, $\hat{\mathbf{h}}_j^\dagger \bar{\mathbf{v}}_k = 0, j \in \mathcal{S} \setminus \{k\}$. If $\mathbf{W}(\mathcal{S})$ is the pseudoinverse of $\hat{\mathbf{H}}(\mathcal{S})$, that is

$$\mathbf{W}(\mathcal{S}) = \hat{\mathbf{H}}(\mathcal{S})^\dagger [\hat{\mathbf{H}}(\mathcal{S}) \hat{\mathbf{H}}(\mathcal{S})^\dagger]^{-1}, \quad (3)$$

then the precoding matrix $\bar{\mathbf{V}} = [\bar{\mathbf{v}}_1 \bar{\mathbf{v}}_2 \cdots \bar{\mathbf{v}}_{|\mathcal{S}|}]$ can be obtained from $\mathbf{W}(\mathcal{S})$ by normalizing all of its columns. For ZF with perfect CSIT, each user receives only the beam directed to it and no multiuser interference is experienced. For the imperfect CSIT case, there is some residual interference. If \mathbf{u} represents the vector of information symbols (u_k intended for the k th user), the transmitted signal \mathbf{x} becomes $\mathbf{x} = \bar{\mathbf{V}}\mathbf{u}$ and the signal received by the k th selected user (1) can be expressed as follows:

$$y_k = \mathbf{h}_k^\dagger \bar{\mathbf{V}}\mathbf{u} + n_k = \mathbf{h}_k^\dagger \bar{\mathbf{v}}_k u_k + \sum_{j \neq k} \mathbf{h}_k^\dagger \bar{\mathbf{v}}_j u_j + n_k. \quad (4)$$

3.3. Downlink Training Phase. It was remarked in [25] that only one symbol interval is sufficient to let the $|\mathcal{S}|$ selected users learn their effective scalar channels $\mathbf{h}_k^\dagger \bar{\mathbf{v}}_k$. In a very recent reference [26], the authors show that this minimal training becomes optimal with joint pilot and data processing. As this DL training length has no relation with the number of users K present in the system or the number of BS antennas (n_t), we assume that the selected users are able to estimate their effective scalar channels perfectly even though we ignore the overhead of this phase. This simplifies the analysis without influencing the underlying cost-benefit tradeoff of the feedback.

3.4. Coherent Data Phase. We adopt uniform power allocation. So the k th user input signal u_k is i.i.d. Gaussian, $u_k \sim \mathcal{CN}(0, p)$, where p is the power allocated to k th user data stream. The BS is bound to satisfy an average power constraint of P but it does not transmit during the entire coherence block due to the initial UL training phase of length T_1 . Hence, for the rest of the coherence block, the BS is able to transmit an average per symbol power of $PT/(T - T_1)$ instead of P . So the power p allocated to each of the $|\mathcal{S}|$ selected users becomes

$$p = \frac{P}{|\mathcal{S}|} \frac{T}{T - T_1}. \quad (5)$$

3.5. Sum Rate Lower Bound. We are interested in getting an expression for the achievable sum rate of this broadcast channel which captures the gain and the cost associated with

feedback. The received signal from (4) can be further written as

$$y_k = \hat{\mathbf{h}}_k^\dagger \bar{\mathbf{v}}_k u_k + \tilde{\mathbf{h}}_k^\dagger \bar{\mathbf{v}}_k u_k + \sum_{j \neq k} \tilde{\mathbf{h}}_k^\dagger \bar{\mathbf{v}}_j u_j + n_k. \quad (6)$$

This is obtained by exploiting the fact that $\mathbf{h}_k^\dagger \bar{\mathbf{v}}_j = \tilde{\mathbf{h}}_k^\dagger \bar{\mathbf{v}}_j$ for $k \neq j$ due to ZF beamforming ($\hat{\mathbf{h}}_k^\dagger \bar{\mathbf{v}}_j = 0$) and by splitting the effective channel $\mathbf{h}_k^\dagger \bar{\mathbf{v}}_k$ into two parts, one of which, $\hat{\mathbf{h}}_k^\dagger \bar{\mathbf{v}}_k$, is perfectly known at the BS. The above equation can be rewritten as

$$y_k = \hat{\mathbf{h}}_k^\dagger \bar{\mathbf{v}}_k u_k + \sum_{j \in \mathcal{S}} \tilde{\mathbf{h}}_k^\dagger \bar{\mathbf{v}}_j u_j + n_k. \quad (7)$$

Now we transition from the above exact signal model to the following degraded model:

$$y_k^{\text{lb}} = \hat{\mathbf{h}}_k^\dagger \bar{\mathbf{v}}_k u_k + w_k, \quad \text{i.i.d. } w_k \sim \mathcal{CN}(0, \sigma_{w_k}^2), \quad (8)$$

$$\sigma_{w_k}^2 = 1 + \sum_{j \in \mathcal{S}} p \mathbb{E}_{\tilde{\mathbf{h}}_k} \left| \tilde{\mathbf{h}}_k^\dagger \bar{\mathbf{v}}_j \right|^2.$$

In this degraded model, we have relegated, as in [27, 28], the signal part $\tilde{\mathbf{h}}_k^\dagger \bar{\mathbf{v}}_k u_k$ into the interference. Considering this interference or noise w_k as independent from the signal u_k leads to a first reduction in the capacity of system (8) w.r.t. the true system (7). Furthermore, for a noise term w_k with given variance $\sigma_{w_k}^2$, we get a further capacity reduction by taking the worst case noise distribution, namely, a Gaussian distribution. Now, for the system model (8) with Gaussian noise, the optimal input is Gaussian: $u_k \sim \mathcal{CN}(0, p)$. So the (instantaneous) capacity expression for system (8) is that of an AWGN channel. Note that the channel that we need to average over in order to get the ergodic capacity is now $\hat{\mathbf{h}}_k^\dagger \bar{\mathbf{v}}_k$, or in other words average over $\hat{\mathbf{h}}_{1:|\mathcal{S}|}$ (on which $\sigma_{w_k}^2$ may in principle also still depend). So, the capacity of model (8) is

$$R_k^{\text{lb,obl}} = \mathbb{E}_{\hat{\mathbf{h}}_{1:|\mathcal{S}|}} \log(1 + \text{SINR}_k^{\text{lb,obl}}), \quad \text{SINR}_k^{\text{lb,obl}} = \frac{p \left| \hat{\mathbf{h}}_k^\dagger \bar{\mathbf{v}}_k \right|^2}{\sigma_{w_k}^2} \quad (9)$$

which is a lower bound for the capacity R_k^{obl} of (7). Now, to compute $\sigma_{w_k}^2$, the variance of each interference coefficient $\tilde{\mathbf{h}}_k^\dagger \bar{\mathbf{v}}_j$ can be computed based upon the fact that the BS performs MMSE estimation which makes the estimation error $\tilde{\mathbf{h}}_k$ (with variance σ_h^2 per channel entry) independent of any function of the channel estimates $\hat{\mathbf{h}}_k$ [29], of which beamforming vectors are one particular instance:

$$\mathbb{E}_{\tilde{\mathbf{h}}_k} \left| \tilde{\mathbf{h}}_k^\dagger \bar{\mathbf{v}}_j \right|^2 = \bar{\mathbf{v}}_j^\dagger \mathbb{E}_{\tilde{\mathbf{h}}_k} (\tilde{\mathbf{h}}_k \tilde{\mathbf{h}}_k^\dagger) \bar{\mathbf{v}}_j = \bar{\mathbf{v}}_j^\dagger (\sigma_h^2 \mathbf{I}_{n_t}) \bar{\mathbf{v}}_j = \sigma_h^2. \quad (10)$$

Furthermore, by introducing $\hat{\mathbf{h}}_k = \sqrt{1 - \sigma_h^2} \mathbf{g}_k$ and $\mathbf{g} = \mathbf{g}_{1:|\mathcal{S}|}$, we get $\mathbf{g}_k \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{n_t})$ and the SINR lower bound becomes

$$\text{SINR}_k^{\text{lb,obl}} = \frac{1 - \sigma_h^2}{1 + p|\mathcal{S}|\sigma_h^2} p \left| \mathbf{g}_k^\dagger \bar{\mathbf{v}}_k \right|^2. \quad (11)$$

For a system with perfect CSIT ($\sigma_h = 0$), $p|\mathbf{g}_k^\dagger \bar{\mathbf{v}}_k|^2 = p|\mathbf{h}_k^\dagger \bar{\mathbf{v}}_k|^2$ would be the SINR and hence its coefficient in the above expression represents the SINR loss factor w.r.t. a system with perfect CSIT. So during the data phase, the lower bound for the per symbol sum rate can be written as

$$R^{\text{lb,obl}} = \sum_{k \in \mathcal{S}} R_k^{\text{lb,obl}} = \sum_{k \in \mathcal{S}} \mathbb{E}_{\mathbf{g}} \log \left(1 + \frac{1 - \sigma_h^2}{1 + p|\mathcal{S}|\sigma_h^2} p \left| \mathbf{g}_k^\dagger \bar{\mathbf{v}}_k \right|^2 \right). \quad (12)$$

Due to the average power constraint and the reduction of the transmission time to $T - T_1$, the transmit power during the data phase gets boosted by a factor $T/(T - T_1)$. Putting now the value of $p = (T/(T - T_1)) (P/|\mathcal{S}|)$, we get

$$R^{\text{lb,obl}} = \sum_{k \in \mathcal{S}} \mathbb{E}_{\mathbf{g}} \log \left(1 + \frac{1 - \sigma_h^2}{1 + P(T/(T - T_1))\sigma_h^2} \frac{T}{T - T_1} \frac{P}{|\mathcal{S}|} \left| \mathbf{g}_k^\dagger \bar{\mathbf{v}}_k \right|^2 \right). \quad (13)$$

If the same system had perfect CSI ($\sigma_h^2 = 0, T_1 = 0$), the sum rate obtained through SUS and ZF beamforming with equal power allocation would be [3, 11]

$$R_{\text{ZF}}(K, n_t, P) = \sum_{k \in \mathcal{S}} \mathbb{E}_{\mathbf{g}} \log \left(1 + \frac{P}{|\mathcal{S}|} \left| \mathbf{g}_k^\dagger \bar{\mathbf{v}}_k \right|^2 \right). \quad (14)$$

So the lower bound of the sum rate from (13) can be written in terms of the sum rate of a perfect CSI system as

$$\text{LB} = R_{\text{ZF}}(K^{\text{obl}}, n_t, P_m), \quad (15)$$

where P_m , the reduced transmission power due to imperfect CSIT, is given by

$$P_m = \frac{(1 - \sigma_h^2)(T/(T - T_1))}{1 + P(T/(T - T_1))\sigma_h^2} P. \quad (16)$$

An important subtlety however is that in (14), the expectation is over the $\mathbf{g}_k = \mathbf{h}_k$, the true channels, whereas in (13) the \mathbf{g}_k are a per-coefficient variance normalized version of the \mathbf{h}_k , the channel estimates. With the assumptions taken here, of i.i.d. channel coefficients and noise elements, this does not make any difference here. However, this issue could potentially make a big difference in the case of spatially correlated channels and/or receiver noise at the BS. In that case also, (13) would allow for a more straightforward analysis of SUS, which is also based on the $\hat{\mathbf{h}}_k$! By taking now into account the loss of coherence interval T due to feedback (training) interval of length $T_1 = \beta K^{\text{obl}}$, the per symbol average sum rate lower bound for this oblivious scheme becomes

$$\begin{aligned} \text{LB}^{\text{obl}} &= \frac{T - \beta K^{\text{obl}}}{T} \\ &\times \sum_{k \in \mathcal{S}} \mathbb{E}_{\mathbf{g}} \log \left(1 + \frac{(1 - \sigma_h^2)(T/(T - \beta K^{\text{obl}}))}{1 + P\sigma_h^2(T/(T - \beta K^{\text{obl}}))} \frac{P}{|\mathcal{S}|} \left| \mathbf{g}_k^\dagger \bar{\mathbf{v}}_k \right|^2 \right). \end{aligned} \quad (17)$$

The biggest virtue of this lower bound is that it gives the achievable sum rate of this scheme in terms of the sum rate of a perfect CSI system (employing SUS and ZF precoding) with loss appearing as an SNR reduction factor and as a reduced multiplexing gain due to the feedback interval.

At this point, we introduce a large user regime approximation for LB^{obl} . As the \mathbf{g}_k and $\bar{\mathbf{v}}_k$ are perfectly known at the BS (perfect CSIT but with a different channel distribution), we can invoke Theorem 1 from [11] for a large number of users. This allows to approximate $|\mathbf{g}_k^\dagger \bar{\mathbf{v}}_k|^2$ by $\log(K^{\text{obl}})$ in the above expression. (The origin of this approximation can easily be traced as follows. The effective channel strength $|\mathbf{g}_k^\dagger \bar{\mathbf{v}}_k|^2$ can be written as the product of the channel norm squared $\|\mathbf{g}_k\|^2$ and the inner product $|\bar{\mathbf{g}}_k^\dagger \bar{\mathbf{v}}_k|^2$. This inner product has value close to one as the selected users are close to orthogonal. The channel norm squared $\|\mathbf{g}_k\|^2$, a chi-square distributed random variable with varying degrees of

freedom depending upon the SUS selection stage, shows a growth with $\log(K^{\text{obl}})$ (see, e.g., (A10) in [5]) using results from order statistics.) Theoretically this scaling kicks in only when K^{obl} is sufficiently large but we show in Section 5 that this starts to hold very well for K^{obl} being a reasonable multiple of n_t . Using this approximation, the fact that for large user regime $|\mathcal{S}| = n_t$ and putting the value of σ_h^2 from (2), the sum rate lower bound (17) becomes the following approximation(Though at low SNR, strictly speaking $|\mathcal{S}| = 1$ is optimal, for large K the sum rate becomes essentially insensitive to $|\mathcal{S}|$ since for small P , $|\mathcal{S}| \ln(1 + (P/|\mathcal{S}|)\gamma) = P\gamma$. Hence $|\mathcal{S}| = n_t$ can be maintained at all SNR. At this point we wish to add that perhaps the main characteristic of user power optimization, namely that it leads to varying $|\mathcal{S}|$, is captured by our simplified approach with uniform power loading over a set of $|\mathcal{S}|$ selected users of varying size):

$$SR^{\text{obl}} = \frac{T - \beta K^{\text{obl}}}{T} n_t \log \left(1 + \frac{(P/n_t)(T/(T - \beta K^{\text{obl}}))(P_{\text{pk}}\beta K^{\text{obl}}/(P_{\text{pk}}\beta K^{\text{obl}} + 1)) \log(K^{\text{obl}})}{1 + P(T/(T - \beta K^{\text{obl}}))(1/(P_{\text{pk}}\beta K^{\text{obl}} + 1))} \right). \quad (18)$$

Due to the approximation made at this final step, this sum rate expression is not necessarily a lower bound but, as we will see, it closely follows both the lower bound and the true sum rate of the system.

3.6. Sum Rate Upper Bound. Consider the signal model in (7), in which certain channel realizations $\hat{\mathbf{h}}_{1:|\mathcal{S}|}$ appear. We get for the mutual information between input and output for user k

$$I(y_k; u_k) = h(y_k) - h(y_k | u_k), \quad (19)$$

where $h(\cdot)$ denotes entropy. For a given distribution (including the optimal one) of the inputs u_i and of the channel estimation errors, the output y_k with variance $\sigma_{y_k}^2$ has an entropy that is upper bounded by that of a Gaussian distribution. Hence

$$I(y_k; u_k) \leq \log(\sigma_{y_k}^2) - \mathbb{E}_{u_k} \log(\sigma_{y_k|u_k}^2). \quad (20)$$

The conditional distribution of y_k given u_k (and $\hat{\mathbf{h}}_{1:|\mathcal{S}|}$) is Gaussian with (zero mean and) variance $\sigma_{y_k|u_k}^2$. We have

$$\begin{aligned} \sigma_{y_k}^2 &= 1 + p \left| \hat{\mathbf{h}}_k^\dagger \bar{\mathbf{v}}_k \right|^2 + \sum_{j \in \mathcal{S}} p \mathbb{E}_{\tilde{\mathbf{h}}_k} \left| \tilde{\mathbf{h}}_k^\dagger \bar{\mathbf{v}}_j \right|^2, \\ \sigma_{y_k|u_k}^2 &= 1 + |u_k|^2 \mathbb{E}_{\tilde{\mathbf{h}}_k} \left| \tilde{\mathbf{h}}_k^\dagger \bar{\mathbf{v}}_k \right|^2 + \sum_{j \neq k} p \mathbb{E}_{\tilde{\mathbf{h}}_k} \left| \tilde{\mathbf{h}}_k^\dagger \bar{\mathbf{v}}_j \right|^2. \end{aligned} \quad (21)$$

Hence, we get

$$\begin{aligned} R_k^{\text{ub,obl}} &= \mathbb{E}_{\hat{\mathbf{h}}_{1:|\mathcal{S}|}} \mathbb{E}_{|u_k|} \\ &\times \log \left(\frac{1 + p \left| \hat{\mathbf{h}}_k^\dagger \bar{\mathbf{v}}_k \right|^2 + \sum_{j \in \mathcal{S}} p \mathbb{E}_{\tilde{\mathbf{h}}_k} \left| \tilde{\mathbf{h}}_k^\dagger \bar{\mathbf{v}}_j \right|^2}{1 + |u_k|^2 \mathbb{E}_{\tilde{\mathbf{h}}_k} \left| \tilde{\mathbf{h}}_k^\dagger \bar{\mathbf{v}}_k \right|^2 + \sum_{j \neq k} p \mathbb{E}_{\tilde{\mathbf{h}}_k} \left| \tilde{\mathbf{h}}_k^\dagger \bar{\mathbf{v}}_j \right|^2} \right) \\ &\geq \mathbb{E}_{\hat{\mathbf{h}}_{1:|\mathcal{S}|}} \log \left(1 + \frac{p \left| \hat{\mathbf{h}}_k^\dagger \bar{\mathbf{v}}_k \right|^2}{1 + \sum_{j \in \mathcal{S}} p \mathbb{E}_{\tilde{\mathbf{h}}_k} \left| \tilde{\mathbf{h}}_k^\dagger \bar{\mathbf{v}}_j \right|^2} \right) = R_k^{\text{lb,obl}}, \end{aligned} \quad (22)$$

where the inequality follows from Jensen's inequality ($-\mathbb{E}_{|u_k|} \log(x) \geq -\log(\mathbb{E}_{|u_k|} x)$). Hence, a simple application of Jensen's inequality leads to a lower bound for the upper bound which coincides with the rate lower bound we had before in (9). Note that one cannot say that upper and lower bounds coincide with the case of constant $|u_k|$ because the lower bound assumes Gaussian signal and interference. Nevertheless, one can sense that upper and lower bound are fairly close. In particular, they behave similarly as far as the rate degradations due to imperfect CSIT as considered here are concerned. This implies in particular that the lower bound should also be a good rate approximation.

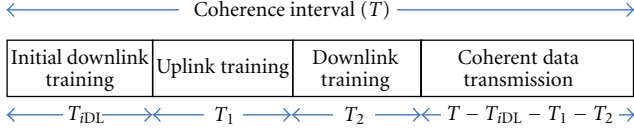


FIGURE 2: Transmission Phases for Informed Users.

4. Transmission Scheme with Informed Users

Similar to the previous scheme with “oblivious users,” this scheme comprises transmission phases through which both the BS and the users get necessary CSI. We call this the scheme with “informed users” as the users who feed back are no longer randomly selected. These users are selected based upon their channel realizations in a manner to be described shortly. This scheme divides the coherence length of T symbol intervals in four phases (see Figure 2): (1) initial downlink training, (2) uplink training, (3) downlink training, and (4) coherent data transmission.

In the first phase the BS transmits DL pilots based upon which all the users estimate their corresponding channel vectors. As the BS has n_t antennas, this training interval length is lower bounded by n_t , $T_{IDL} \geq n_t$, and is independent of the number of users K . The channel estimation quality will depend on the transmit energy $P_{IDL} T_{IDL}$ spent by the BS, where P_{IDL} is the BS power during T_{IDL} . The contribution of P_{IDL} to the BS average transmit power constraint is also through the product $P_{IDL} T_{IDL}$. On the other hand, as T_{IDL} leads to reduced data transmission time, the optimal choice is the minimal $T_{IDL} = n_t$. As P_{IDL} can be sufficiently high to allow good channel estimates, we do not take into account the estimation error during this phase but only subtract $T_{IDL} = n_t$ from the coherence length. Also, we will neglect the effect of $P_{IDL} T_{IDL} = P_{IDL} n_t$ on the BS power constraint. Another point is that the channel estimation quality in phase 1) will only affect the selection of informed users, but not the quality of CSIT acquisition in phase 2).

Once the users have acquired the information regarding their respective channels, there could be plenty of criteria to prioritize users depending upon their channel realizations

but we restrict ourselves to the simple scheme where K^{inf} best users with largest channel norm are selected for feedback. Hence the BS receives the CSIT from the K^{inf} of the K users that have the largest channel norms for the current channel coherence block.

The next three transmission phases are exactly similar as those for the transmission scheme with oblivious users.

Important Remark. In this transmission scheme involving informed users, we select the strongest users (having largest channel norms) who train the BS about their channels. Strictly speaking, this is impractical as how can the users know about being the strongest or not with only information about their own channels. But the underlying idea is to evaluate how much feedback load (how many users) should be there to maximize the DL sum rate if good users feed back. Then, in practice, those users can be made to feed back, on the average, by intelligent selection of a threshold with which users compare their channel strength locally as detailed in [30] and decide to feed back or not, and by designing a proper UL channel access protocol. This threshold will be a function of the total number of users, their channel statistics, the number of BS antennas, and the optimal number of users who should feed back. In practice also then, users should use pseudorandom instead of orthogonal codes in phase T_1 , with the resulting nonorthogonality influencing the ensuing analysis.

4.1. Sum Rate Lower Bound. We will be quite brief here as the treatment resembles a lot the one for the oblivious users. If every user is constrained with a peak per symbol power constraint of P_{pk} and K^{inf} users transmit pilots in the UL direction, the feedback length would be $T_1 = \beta K^{\text{inf}}$ where $\beta \geq 1$. The MSE of CSIT per channel coefficient at the BS is given by

$$\sigma_h^2 = \frac{1}{P_{pk} \beta K^{\text{inf}} + 1}. \quad (23)$$

For this scheme with informed users, we have an initial step of DL training so the length of the data phase reduces to $T - n_t - \beta K^{\text{inf}}$. Thus the DL sum rate for the informed scheme with peak power constrained users is given by

$$SR^{\text{inf}} = \frac{T - n_t - \beta K^{\text{inf}}}{T} n_t \log \left(1 + \frac{(P/n_t) (T/(T - \beta K^{\text{inf}})) (P_{pk} \beta K^{\text{inf}} / (P_{pk} \beta K^{\text{inf}} + 1)) \log(K)}{1 + P(T/(T - \beta K^{\text{inf}})) (1/(P_{pk} \beta K^{\text{inf}} + 1))} \right), \quad (24)$$

One striking difference from the oblivious users case is that the channel strength factor due to multiuser diversity of $\log(K^{\text{obl}})$ now becomes $\log(K)$, where K is the total number of users in the system. This difference arises due to the fact that in the informed users case, the optimization eventually involves all users.

5. Accuracy of the Approximate Sum Rate Expressions

To obtain the approximate sum rate expressions, we derived the novel sum rate lower bound and then used the large user regime approximation in the final step. To see how closely

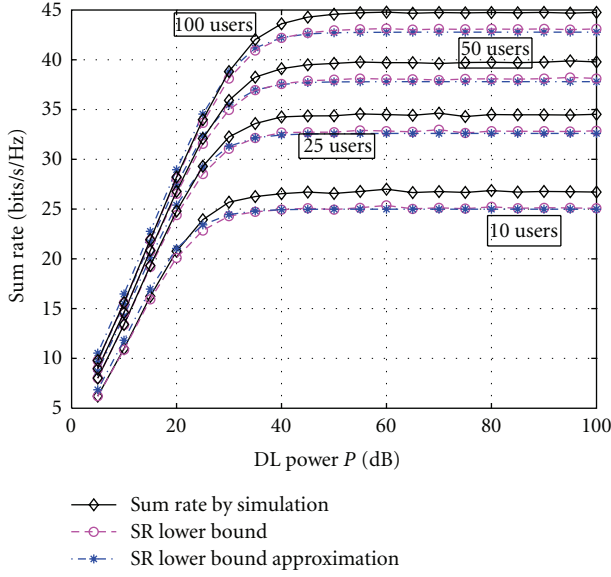


FIGURE 3: Sum Rate versus DL power P for various numbers of users feeding back.

these sum rate expressions capture the true sum rate behavior of the two schemes, it is sufficient to show the accuracy for one of the rate expressions because the same approximation is made for both.

We choose the approximate sum rate expression for oblivious peak power constrained users and we compare it with the true sum rate. We use the version of (18) for the data transmission segment

SR

$$= n_t \log \left(1 + \frac{(P/n_t) (P_{pk} \beta K^{obl} / (P_{pk} \beta K^{obl} + 1)) \log(K^{obl})}{1 + P (1 / (P_{pk} \beta K^{obl} + 1))} \right) \quad (25)$$

which is the part where the approximation occurs. For fixed values of T , β , and K^{obl} , we absorb the constant factor of $T/(T - \beta K^{obl})$ in the BS power constraint both in numerator and denominator of SINR and leaving the constant multiplying factor of $(T - \beta K^{obl})/T$ outside of the logarithm. So in this form it captures the behavior of the sum rate expression independent of the coherence length T . The corresponding saturation level is

$$SR = n_t \log \left(\frac{P_{pk} \beta K^{obl} \log(K^{obl})}{n_t} \right). \quad (26)$$

To obtain the true sum rate for various system parameter settings, we use Monte-Carlo simulations in which all the steps in the transmission strategy, for example, feedback, SUS scheduling, and ZF beam formation are replicated and then the SINR at each user is evaluated.

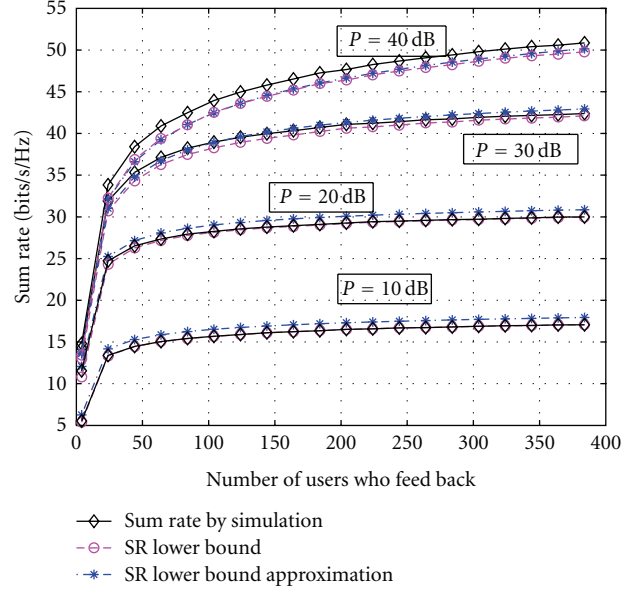


FIGURE 4: Sum Rate versus number of users feeding back, for various DL powers P .

Figure 3 shows plots of the sum rate versus DL power constraint P . The uplink power constraint for each user has been fixed to 10 dB and the BS is equipped with 4 antennas. The curves for the true sum rate, the lower bound from (17) (but without the factors $T/(T - \beta K^{obl})$) and the approximation (25) of the LB are plotted when 10, 25, 50, and 100 users feed back their channel information to the BS. The SR LB and its approximation almost coincide. The approximate expression captures very closely the true sum rate behavior for any SNR, even at saturation. This saturation of the sum rate is caused by the imperfect CSIT based upon which ZF beamforming vectors are computed (shown for quantized FB in [3] and for analog feedback in [25]). Note that squaring the number of users from 10 to 100 leads to roughly a doubling of the sum rate at high SNR and hence the multiuser diversity aspect is indeed operating.

Figure 4 shows the plot of the sum rate with varying numbers of users feeding back K^{obl} and for DL power constraint levels of 10, 20, 30, and 40 dB. These curves further illustrate the tightness of the lower bound (17) and the approximate expression (25) in capturing the multiuser diversity benefit.

6. Asymptotic Analysis with Oblivious Users

In this section, we analyze how the sum rate of oblivious users behaves in different asymptotic regimes. Although this analysis is asymptotic, it gives valuable insight about the optimal amount of feedback and its utilization.

6.1. Noise Limited Regime. For the noise limited regime, the power available to the BS is very limited, that is, $P \rightarrow 0$. In this regime, the noise in each user's received signal fully

dominates the interference coming from the beams of other selected users

SR^{obl}

$$= \frac{T - \beta K^{obl}}{T} n_t \log \left(1 + \frac{P}{n_t} \frac{T}{(T - \beta K^{obl})} (1 - \sigma_h^2) \log(K^{obl}) \right). \quad (27)$$

Using the approximation of $\log(1 + x) \approx x$ for very small x , the above sum rate becomes

$$SR^{obl} = P(1 - \sigma_h^2) \log(K^{obl}). \quad (28)$$

Plugging in the value of σ_h^2 shows that the above expression is an increasing function of K^{obl} . Hence all the users K should feed back provided that $T > K$ and then the BS chooses the strongest user for transmission with full power. This asymptote shows that, at low SNR, the multiplexing gain is lost but the multiuser diversity gives logarithmic instead of double logarithmic gain.

6.2. Interference Limited Regime. The interference power due to imperfect CSIT at the BS scales up with the increase in DL power. So when the DL SNR gets very large ($P \rightarrow \infty$), the interference completely dominates the noise and the sum rate saturates to

$$\begin{aligned} SR^{obl} &= \frac{T - \beta K^{obl}}{T} n_t \\ &\times \log \left(1 + \frac{(P/n_t)(T/(T - \beta K^{obl}))(1 - \sigma_h^2) \log(K^{obl})}{P(T/(T - \beta K^{obl}))\sigma_h^2} \right), \end{aligned} \quad (29)$$

which reduces to

$$SR^{obl} = \frac{T - \beta K^{obl}}{T} n_t \log \left(1 + \frac{\log(K^{obl})}{n_t} \left(\frac{1}{\sigma_h^2} - 1 \right) \right). \quad (30)$$

This equation shows that the sum rate saturation level can be increased by improving the CSIT quality, that is, by reducing σ_h^2 . The SNR switching point at which saturation starts is $P = P_{pk}\beta K^{obl}$, the point at which the interference equals the channel noise power and after which the interference starts dominating.

6.3. Asymptotically Large Number of Users. For the oblivious scheme with peak power constrained users, the sum rate expression is completely independent of the number of users present in the system and only depends upon the users who actually feed back.

7. Asymptotic Analysis with Informed Users

In this section, we analyze how the sum rate of the scheme with informed users behaves in different asymptotic regimes.

7.1. Noise Limited Regime. In the noise limited regime, the noise completely dominates the interference at each active user, so

$$\begin{aligned} SR^{inf} &= \frac{T - n_t - \beta K^{inf}}{T} n_t \\ &\times \log \left(1 + \frac{P}{n_t} \frac{T}{(T - \beta K^{inf})} (1 - \sigma_h^2) \log(K) \right). \end{aligned} \quad (31)$$

Again by using the approximation of $\log(1 + x) \approx x$ for very small x , the above sum rate becomes

$$SR^{inf} = \left(P \frac{T - n_t - \beta K^{inf}}{T - \beta K^{inf}} \right) (1 - \sigma_h^2) \log(K). \quad (32)$$

7.2. Interference Limited Regime. In this case the SR,

$$\begin{aligned} SR^{inf} &= \frac{T - n_t - \beta K^{inf}}{T} n_t \\ &\times \log \left(1 + \frac{(P/n_t)(T/(T - \beta K^{inf}))(1 - \sigma_h^2) \log(K)}{P(T/(T - \beta K^{inf}))\sigma_h^2} \right), \end{aligned} \quad (33)$$

saturates at

$$SR^{inf} = \frac{T - n_t - \beta K^{inf}}{T} n_t \log \left(1 + \frac{\log(K)}{n_t} \left(\frac{1}{\sigma_h^2} - 1 \right) \right). \quad (34)$$

Again, this saturation level can be increased by refining the CSIT quality.

7.3. Asymptotically Large Number of Users. For the informed scheme, for any power constraint imposed on user terminals, the effective signal strength increases with $\log(K)$ as only strong users feed back to the BS and hence get scheduled. For this reason, the sum rate shows unbounded growth with the number of users present in the system.

8. Feedback Load Optimization

As Section 5 showed that the approximate sum rate expressions match closely the true sum rate, they can be used to find the optimal number of users to train the BS about their channel information. There are two parameters, namely, the number of users who feed back and the β factor. A careful

observation of the final sum rate expressions (18) and (24) reveals that for any fixed value of the product β and the number of users who feedback, β can always be selected to be 1 without any loss of optimality of the sum rate. The optimization of β may affect other system parameters such as the total UL energy spent for training but we focus here on the DL sum rate maximization. Hence, with $\beta = 1$, the

$$\text{SR}^{\text{obl}} = \frac{T - K^{\text{obl}}}{T} n_t \log \left(1 + \frac{(P/n_t)(T/(T - K^{\text{obl}}))(P_{\text{pk}}K^{\text{obl}}/(P_{\text{pk}}K^{\text{obl}} + 1)) \log(K^{\text{obl}})}{1 + P(T/(T - K^{\text{obl}}))(1/(P_{\text{pk}}K^{\text{obl}} + 1))} \right). \quad (35)$$

For high SNR ($P \gg 1$), and also assuming $P_{\text{pk}}K^{\text{obl}} \gg 1$, (35) becomes

$$\text{SR}^{\text{obl}} = \left(1 - \frac{K^{\text{obl}}}{T}\right) \left[\text{SR}^{\text{csit}} - n_t \log \left(1 - \frac{K^{\text{obl}}}{T} + \frac{P}{P_{\text{pk}}K^{\text{obl}}}\right) \right], \quad (36)$$

where SR^{csit} denotes the sum rate for the case of perfect CSIT. We get for the derivative (denoting K^{obl} as K temporarily and ignoring the weak dependence of SR^{csit} on K)

$$\begin{aligned} \frac{\partial \text{SR}^{\text{obl}}}{\partial K} = & -\frac{1}{T} \left[\text{SR}^{\text{csit}} - n_t \log \left(1 - \frac{K}{T} + \frac{P}{P_{\text{pk}}K}\right) \right] \\ & + \frac{1 - (K/T)}{1 - K/T + P/P_{\text{pk}}} n_t \left(\frac{1}{T} + \frac{P}{P_{\text{pk}}K^2} \right). \end{aligned} \quad (37)$$

In (37), for large P , T , and K , we can ignore the second term in the square brackets compared to the first term SR^{csit} , and we can ignore the factor $(1 - (K/T))$ in front of the second term (as can be verified a posteriori from the resulting solution for K). We now distinguish two cases. In the first case we assume high downlink SNR ($P \rightarrow \infty$) and finite uplink SNR (P_{pk}). Solving (37) for K now yields the optimal

$$K^{\text{obl}} = \frac{T n_t}{\text{SR}^{\text{csit}}}, \quad \text{for } P \rightarrow \infty, P_{\text{pk}} \text{ finite.} \quad (38)$$

Hence in this case $K^{\text{obl}} \sim T$ and $K^{\text{obl}} \sim 1/\log(P)$. In the second case, $P, P_{\text{pk}} \rightarrow \infty$ with finite ratio P/P_{pk} . Solving for K now yields

$$K^{\text{obl}} = \sqrt{\frac{PT n_t}{P_{\text{pk}} \text{SR}^{\text{csit}}}}, \quad \text{for } P, P_{\text{pk}} \rightarrow \infty, \frac{P}{P_{\text{pk}}} \text{ finite.} \quad (39)$$

In this case $K^{\text{obl}} \sim \sqrt{T}$ and $K^{\text{obl}} \sim 1/\sqrt{\log(P)}$.

We also investigate the optimization w.r.t. K^{obl} by numerical optimization. First we check how the optimal number of users (feeding back) scales with DL power P . We plot the curves for the optimal number of users versus P in Figure 5 and plot corresponding sum rates achieved by using

amount of feedback load appears as the number of users who feedback.

8.1. Optimal Number of Users versus DL Power. We formulate the problem here for the scheme with oblivious users. The sum rate for this scheme was developed to be

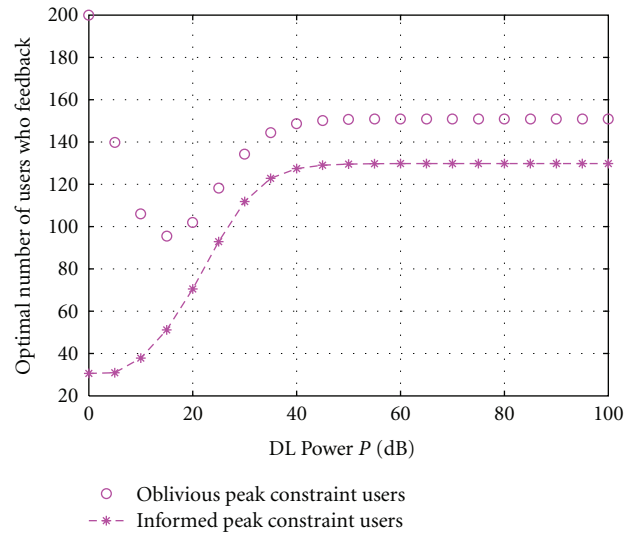


FIGURE 5: Optimal Number of Users versus DL Power P .

that optimal number of users for each value of P in Figure 6. The parameters are $T = 1000$ symbol intervals, $K = 200$ users in the system, per user peak power constraint P_{pk} of 5 dB, and the BS is equipped with $n_t = 4$ antennas. It is evident that the gains with optimal feedback are undeniable as the sum rate with only feedback from n_t users is much less than the sum rate with the optimal number of users. The saturation of the sum rate due to imperfect CSIT as depicted in Figure 6 and analyzed in (30) and (34) has previously been investigated in [3, 25].

The behavior in Figure 5 of the curves of optimal number of users feeding back for the two schemes versus P is not very straightforward. At high SNR (interference limited regime, see (30) and (34)), both schemes require very good quality CSIT and due to peak power constrained users, it translates to obtaining feedback from each user for longer intervals which comes out to be a lot of users transmitting feedback (users have orthogonal codes and hence can be separated). The numerical optimization seemingly giving a constant K at high SNR does not agree closely to (38), where K should behave as $1/\log(P)$, which indicates nevertheless a slow

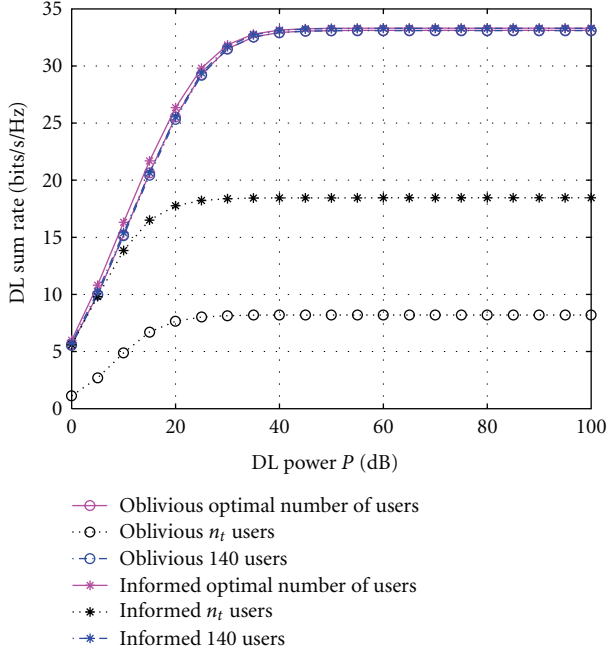


FIGURE 6: Sum rate with optimal number of users versus DL power P .

variation with P , but at high values of P , K , and T , the sensitivity of the sum rate to K is fairly low.

At low SNR both curves show very different behavior. At low SNR, the system is noise limited and the multiuser diversity factor is very important. Hence the users with very strong channels should be scheduled (see (28) and (32)). In the informed users scheme, only the very strong users feed back so it requires feedback from a small number of users. As SNR increases and interference, and hence CSIT quality start to become more important, more users start feeding back (which is the way to improve CSIT quality for peak-power constrained users). The scheme with oblivious users requires feedback from a large number of users initially to enjoy multiuser diversity. Since that consumes a lot of coherence time in feedback, the number of users who feed back decreases initially as SNR improves, and starts increasing again only at medium to high SNR, to provide higher quality CSIT.

Although the optimal number of users feeding back in the two schemes differs significantly in the lower to medium SNR range, the corresponding sum rates are very comparable, with just a tiny edge for the informed user strategy. In Figure 6, we have also plotted the sum rate curves when 140 users feed back (this number is close to optimal at high DL SNR for both schemes). The resulting sum rate curves are very close to the ones with optimized numbers of users. This indicates that for a fixed channel coherence length T , a fixed value of users feeding back (normally much larger than n_t) can basically achieve the cost-benefit tradeoff of feedback. In other words, the sum rate is not very sensitive to the number of users who feed back, for the whole SNR range.

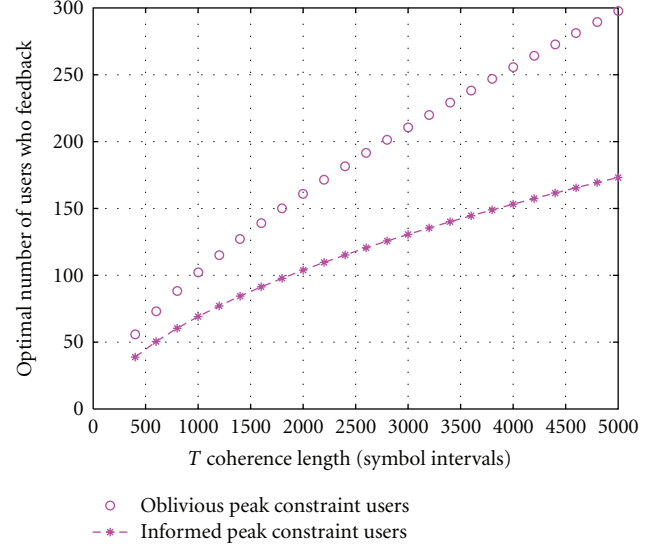


FIGURE 7: Optimal number of users versus coherence length.

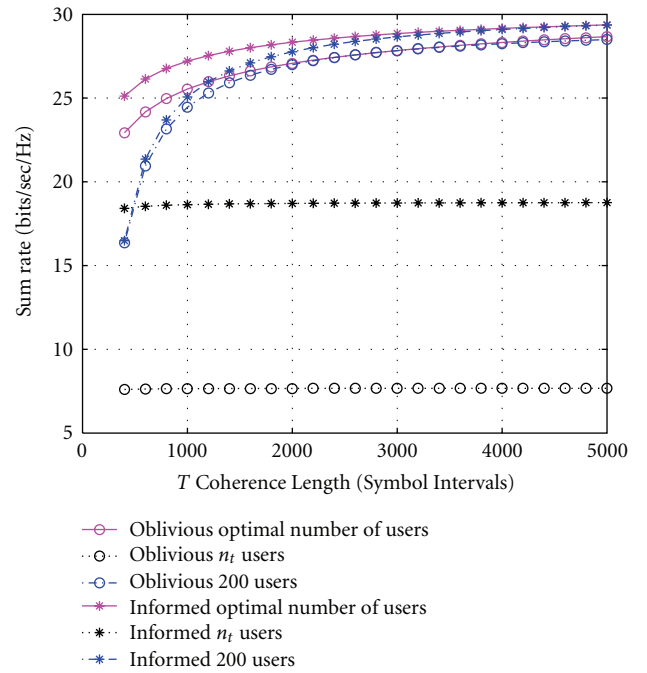


FIGURE 8: Sum rate with optimal number of users versus coherence length.

8.2. Optimal Number of Users versus Channel Coherence Time. We now analyze how the optimal number of users varies with the channel coherence time. We plot two figures, one showing the optimal number of users versus coherence interval in Figure 7 and the other showing the sum rate corresponding to the optimal number of users versus coherence interval in Figure 8. Here the BS has $n_t = 4$ antennas, its power constraint is 20 dB and there are 500 users in the system with each user restricted to a peak power constraint of 5 dB.

The curves of the optimal number of users vs. channel coherence time show eventually a linear behavior in T as predicted by (38) though the actual behavior appears to be intermediate between (38) and (39). For smaller values of the coherence interval, a small number of users is optimal so that not much of the coherence interval gets consumed in feedback. For very large values of the coherence interval, feedback from a large number of users is optimal so as to select the good users with good quality CSIT. Thus the number of users to feed back scales up with the increase in channel coherence time. The optimal number is always larger for the scheme with oblivious users than in the case of informed users. This behavior can be anticipated from Figure 5 which shows that from low to medium DL SNR values, the optimal number of users in the oblivious scheme is larger than that in the informed scheme.

Sum rate curves for optimal users are plotted at $P = 20$ dB in Figure 8, so the informed user scheme performs better as can be guessed from Figure 6. Sum rate curves have also been plotted for a fixed number of users (200) feeding back but contrary to the sum rate versus SNR curves where a single suitable number of users feeding back capture the gain of optimal feedback; here it is not possible to find one such number of users capturing the sum rate gains. So the sum rate as a function of T is relatively sensitive to the number of users who feed back.

9. Concluding Remarks

We studied the problem of determining the optimal amount of feedback/training for the sum rate maximization of the broadcast channel with no initial assumption of CSI. We introduced two transmission strategies for providing the CSIT to the BS and derived a novel tight lower bound which clearly shows the rate loss w.r.t. a perfect CSI system. The corresponding simplified sum rate expressions, incorporating the gains of the feedback and the cost of exchange of information, allow us to determine the optimal amount of feedback for any set of system parameters. Moreover, the asymptotic analysis carried out for both schemes gives us insight into the amount and the split of the optimal feedback between obtaining multiuser diversity and accurate channel information for better interuser interference cancellation. Apart from the system and channel parameters, the optimal split is also a function of the regime of operation of the system. The noise limited regime demands the use of feedback to benefit from multiuser diversity whereas the interference limited regime requires the use of feedback

resources to get fine quality CSIT because the MSE of CSIT is the principal factor to determine the saturation level of the sum rate versus SNR. In between these two regimes, the feedback split depends upon the contribution of multiuser diversity gain, the importance of CSIT quality in the sum rate and the fraction of the coherence interval used for feedback.

The analysis of net rate gain due to feedback for a TDD system with data transmission in both directions and for an FDD system is very interesting direction for future research. The practical implementation of the informed users scheme requires the design of a SNR threshold and a channel access protocol. The design of such an effective threshold metric as well as the analysis of the effect of using pseudorandom training codes could also be interesting research topics.

Appendix

Average Power Constrained Users

We treated the case when the users in the system are peak power constrained. For average power constrained users, the feedback behavior will change as the CSIT MSE changes. We keep the discussion to a minimum as we believe this power constraint to be unrealistic and impractical. If there are K users in the system having channel coherence length of T and each is constrained to an average power P_{avg} per channel use, the total UL energy available in each coherence block is $P_{\text{avg}}KT$. Now if K^{obl} users feed back, each one of these can transmit an energy of $P_{\text{avg}}KT/K^{\text{obl}}$. Here the use of orthogonal codes is not necessary because, due to more flexible power constraint, the users can transmit their available power in short intervals. Hence with this energy transmitted for every channel coefficient, the CSIT MSE at the BS will be

$$\sigma_h^2 = \frac{1}{P_{\text{avg}}KT/K^{\text{obl}} + 1}. \quad (\text{A.1})$$

Although the users feeding back will be able to transmit pilots with larger energy (if $K \gg K^{\text{obl}}$), they will be transmitting only occasionally, the probability of which will reduce with more users in the system, and hence the long term average power constraint will be satisfied. Such power constraint for the transmission in the UL direction was employed in [31]. The sum rate expressions for the two schemes when the users are average power constrained can be obtained by plugging in the MSE of CSIT from (A.1):

$$\begin{aligned} \text{SR}^{\text{obl}} &= \frac{T - \beta K^{\text{obl}}}{T} n_t \log \left(1 + \frac{(P/n_t)(T/(T - \beta K^{\text{obl}})) \left((P_{\text{avg}}KT/K^{\text{obl}}) / (P_{\text{avg}}KT/K^{\text{obl}} + 1) \right) \log(K^{\text{obl}})}{1 + P(T/(T - \beta K^{\text{obl}})) (1/(P_{\text{avg}}KT/K^{\text{obl}} + 1))} \right), \\ \text{SR}^{\text{inf}} &= \frac{T - n_t - \beta K^{\text{inf}}}{T} n_t \log \left(1 + \frac{(P/n_t)(T/(T - \beta K^{\text{inf}})) \left((P_{\text{avg}}KT/K^{\text{inf}}) / (P_{\text{avg}}KT/K^{\text{inf}} + 1) \right) \log(K)}{1 + P(T/(T - \beta K^{\text{inf}})) (1/(P_{\text{avg}}KT/K^{\text{inf}} + 1))} \right). \end{aligned} \quad (\text{A.2})$$

Then these sum rates can be optimized w.r.t. the amount of feedback (the number of users).

Acknowledgments

EURECOM's research is partially supported by its industrial members: BMW Group, Cisco, Monaco Telecom, ORANGE, SFR, SAP, STEricsson, Swisscom, Symantec, and Thales. The research reported in this paper has also been partially supported by the European FP7 project WHERE and NoE NEWCOM++ and by the French ANR project APOGEE.

References

- [1] G. Caire and S. Shamai (Shitz), "On the achievable throughput of a multiantenna Gaussian broadcast channel," *IEEE Transactions on Information Theory*, vol. 49, no. 7, pp. 1691–1706, 2003.
- [2] N. Jindal and A. Goldsmith, "Dirty-paper coding versus TDMA for MIMO broadcast channels," *IEEE Transactions on Information Theory*, vol. 51, no. 5, pp. 1783–1794, 2005.
- [3] N. Jindal, "MIMO broadcast channels with finite-rate feedback," *IEEE Transactions on Information Theory*, vol. 52, no. 11, pp. 5045–5060, 2006.
- [4] R. Knopp and P. A. Humblet, "Information capacity and power control in single-cell multiuser communications," in *Proceedings of the IEEE International Conference on Communications*, pp. 331–335, Seattle, Wash, USA, June 1995.
- [5] M. Sharif and B. Hassibi, "On the capacity of MIMO broadcast channels with partial side information," *IEEE Transactions on Information Theory*, vol. 51, no. 2, pp. 506–522, 2005.
- [6] M. Sharif and B. Hassibi, "A comparison of time-sharing, DPC, and beamforming for MIMO broadcast channels with many users," *IEEE Transactions on Communications*, vol. 55, no. 1, pp. 11–15, 2007.
- [7] D. Gesbert, M. Kountouris, J. R. W. Heath, C. B. Chae, and T. Salzer, "From single user to multiuser communications: shifting the MIMO paradigm," *IEEE Signal Processing Magazine*, vol. 24, no. 5, pp. 36–46, 2007.
- [8] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, New York, NY, USA, 1991.
- [9] T. Cover, "Broadcast channels," *IEEE Transactions on Information Theory*, vol. 18, no. 1, pp. 2–14, 1972.
- [10] D. Tse and P. Viswanath, *Fundamentals of Wireless Communications*, Cambridge University Press, Cambridge, UK, 2005.
- [11] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 3, pp. 528–541, 2006.
- [12] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Transactions on Information Theory*, vol. 52, no. 9, pp. 3936–3964, 2006.
- [13] T. L. Marzetta, "How much training is required for multiuser MIMO?" in *Proceedings of the 40th Asilomar Conference on Signals, Systems, and Computers (ACSSC '06)*, pp. 359–363, Pacific Grove, Calif, USA, November 2006.
- [14] T. L. Marzetta and B. M. Hochwald, "Fast transfer of channel state information in wireless systems," *IEEE Transactions on Signal Processing*, vol. 54, no. 4, pp. 1268–1278, 2006.
- [15] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, "Multiuser MIMO downlink made practical: achievable rates with simple channel state estimation and feedback schemes," submitted to *IEEE Transactions on Information Theory*.
- [16] M. Kobayashi, G. Caire, N. Jindal, and N. Ravindran, "How much training and feedback are needed in MIMO broadcast channels?" in *Proceedings of the IEEE International Symposium on Information Theory (ISIT '08)*, pp. 2663–2667, July 2008.
- [17] N. Ravindran and N. Jindal, "Multi-user diversity vs. accurate channel feedback for MIMO broadcast channels," in *Proceedings of the IEEE International Conference on Communications (ICC '08)*, pp. 3684–3688, Beijing, China, May 2008.
- [18] E. Jorswieck, A. Sezgin, B. Ottersten, and A. Paulraj, "Feedback reduction in uplink MIMO OFDM systems by chunk optimization," *EURASIP Journal on Advances in Signal Processing*, vol. 2008, Article ID 597072, 14 pages, 2008.
- [19] J. Jose, A. Ashikhmin, P. Whiting, and S. Vishwanath, "Scheduling and pre-conditioning in multi-user MIMO TDD systems," in *Proceedings of the IEEE International Conference on Communications (ICC '08)*, pp. 4100–4105, Beijing, China, May 2008.
- [20] B. M. Hochwald, T. L. Marzetta, and V. Tarokh, "Multiple-antenna channel hardening and its implications for rate feedback and scheduling," *IEEE Transactions on Information Theory*, vol. 50, no. 9, pp. 1893–1909, 2004.
- [21] J. Jose, A. Ashikhmin, P. Whiting, and S. Vishwanath, "Scheduling and precoding in multi-user multiple antenna time division duplex systems," preprint, 2008, <http://arxiv.org/abs/0812.0621v1>.
- [22] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Transactions on Information Theory*, vol. 45, no. 1, pp. 139–157, 1999.
- [23] N. Jindal, "High SNR analysis of MIMO broadcast channels," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT '05)*, pp. 2310–2314, Adelaide, Australia, September 2005.
- [24] O. Sjöbergh, E. A. Jorswieck, and E. G. Larsson, "Greedy user selection for zero-forcing and MMSE multiuser beamforming with channel estimation errors," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '08)*, pp. 3137–3140, Las Vegas, Nev, USA, March–April 2008.
- [25] U. Salim and D. Slock, "Broadcast channel: degrees of freedom with no CSIR," in *Proceedings of the 46th Annual Allerton Conference on Communication, Control, and Computing*, pp. 119–125, September 2008.
- [26] N. Jindal, A. Lozano, and T. L. Marzetta, "What is the value of joint processing of pilots and data in block-fading channels?" in *Proceedings of the IEEE International Symposium on Information Theory (ISIT '09)*, pp. 2189–2193, July 2009.
- [27] M. Médard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Transactions on Information Theory*, vol. 46, no. 3, pp. 933–946, 2000.
- [28] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Transactions on Information Theory*, vol. 49, no. 4, pp. 951–963, 2003.
- [29] S. M. Kay, *Fundamentals of Statistical Signal Processing—Estimation Theory*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1993.
- [30] D. Gesbert and M.-S. Alouini, "How much feedback is multi-user diversity really worth?" in *Proceedings of the IEEE International Conference on Communications*, pp. 234–238, Paris, France, June 2004.

- [31] S. Murugesan, E. Uysal-Biyikoglu, and P. Schniter, "Optimization of training and scheduling in the non-coherent SIMO multiple access channel," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 7, pp. 1446–1456, 2007.