

# Multuser MIMO Achievable Rates with Downlink Training and Channel State Feedback

Giuseppe Caire, University of Southern California  
Los Angeles CA, 90089 USA

Nihar Jindal, University of Minnesota  
Minneapolis MN, 55455 USA

Mari Kobayashi, SUPELEC  
Gif-sur-Yvette, France

Niranjay Ravindran, University of Minnesota  
Minneapolis MN, 55455 USA

## Abstract

We consider a MIMO fading broadcast channel and compute achievable ergodic rates when channel state information is acquired at the receivers via downlink training and it is provided to the transmitter by channel state feedback. Unquantized (analog) and quantized (digital) channel state feedback schemes are analyzed and compared under various assumptions. Digital feedback is shown to be potentially superior when the feedback *channel uses* per channel state coefficient is larger than 1. Also, we show that by proper design of the digital feedback link, errors in the feedback have a minor effect even if simple uncoded modulation is used on the feedback channel. We discuss first the case of an unfaded AWGN feedback channel with orthogonal access and then the case of fading MIMO multi-access (MIMO-MAC). We show that by exploiting the MIMO-MAC nature of the uplink channel, a much better scaling of the feedback channel resource with the number of base station antennas can be achieved. Finally, for the case of delayed feedback, we show that in the realistic case where the fading process has (normalized) maximum Doppler frequency shift  $0 \leq F < 1/2$ , a fraction  $1 - 2F$  of the optimal multiplexing gain is achievable. The general conclusion of this work is that *very significant* downlink throughput is achievable with *simple and efficient* channel state feedback, provided that the feedback link is properly designed.

## I. INTRODUCTION

In the downlink of a cellular-like system, a base station equipped with multiple antennas communicates with a number of terminals, each possibly equipped with multiple receive antennas. If a traditional orthogonalization technique such as TDMA is used, the base station transmits to a single receiver on each time-frequency resource and thus is limited to point-to-point MIMO techniques [1], [2]. Alternatively, the base station can use multi-user MIMO to *simultaneously* transmit to multiple receivers on the same time-frequency resource. Under the assumption of perfect channel state information at the transmitter (CSIT) and at the receivers (CSIR), a combination of single-user Gaussian codes, linear beamforming and “Dirty-Paper Coding” (DPC) [3] is known to achieve the capacity of the MIMO downlink channel [4], [5], [6], [7], [8]. When the number of base station antennas is larger than the number of antennas at each terminal, the capacity of the MIMO downlink channel is significantly larger than the rates achievable with point-to-point MIMO techniques [4], [9], [10].

Given the widespread applicability of the MIMO downlink channel model (e.g., to cellular, WiFi, and DSL), it is of great interest to design systems that can operate near the capacity limit. Although realizing the optimal DPC coding strategy still remains a formidable challenge (see for example [11], [12], [13]), it has been shown that linear beamforming without DPC performs quite close to capacity when combined with user selection, again under the simplifying assumption of perfect channel state information (see for example [14], [15]).

In real systems, however, channel state information is not *a priori* provided and must be acquired, e.g., through training. Acquiring the channel state is a challenging and resource-consuming task in time-varying systems, and the obtained information is inevitably imperfect. It is therefore critical to understand what rates are achievable under realistic channel state information assumptions, and in particular to understand the sensitivity of achievable rates to such imperfections. To emphasize the importance of channel state information, note that in the extreme case of no CSIT at the BS and identical fading statistics (and perfect CSIR) at all terminals, the multi-user MIMO benefit is completely lost and point-to-point MIMO becomes optimal [4].

### A. Contributions of this work

The focus of this paper is a rigorous information theoretic characterization of the *ergodic* achievable rates of a fading multiuser MIMO downlink channel in which the UTs and the BS obtain imperfect CSIR/CSIT via downlink training and channel state feedback.<sup>1</sup> Converse results on the capacity region of the MIMO broadcast channel with imperfect channel knowledge are essentially open (see for example [16] and [17] for some partial results). Here, we focus on the achievable rates of a specific signaling strategy, zero-forcing (ZF) linear beamforming. Consistently with contemporary wireless system technology, we assume that each UT estimates its own channel during a downlink training phase and then feeds back its estimate over the reverse uplink channel to the BS. The BS designs beamforming vectors on the basis of the received channel feedback, after which an additional round of downlink training is performed (essentially to inform the UTs of the selected beamformers). Our results tightly bound the rate that is achievable after this process in terms of the resources (i.e., channel symbols) used for training and feedback and the channel feedback technique.

The analysis of this paper inscribes itself in the line of works dealing with “training capacity” [18] of block-fading channels. Several previous and concurrent works have treated training and channel feedback for point-to-point MIMO systems (see for example [19], [20], [21], [22], [23], [24], [25]) and, more recently, for MIMO broadcast channels (see for example [26], [27], [28], [29], [30], [31], [32]). However, this paper presents a number of novelties relative to prior/concurrent work:

- Rather than assuming perfect CSIR at the UT’s, we consider the realistic scenario where the UTs have imperfect CSIR obtained via downlink training. Because the imperfect CSIR is the basis for the channel feedback from the UTs, this degrades the quality of the CSIT provided to the BS in a non-negligible manner.
- Instead of idealizing the feedback channel as a fixed-rate, error-free bit-pipe, we explicitly consider transmission from each UT to the BS over the noisy feedback channel. This reveals the fundamental joint source-channel coding nature of channel feedback. In addition, this allows us to meaningfully measure the uplink resources dedicated to channel feedback and also allows for a comparison between analog (unquantized) and digital (quantized) feedback. We begin by modeling the feedback channel as an AWGN channel (orthogonal across UTs), and later generalize to a multiple-antenna uplink channel that is shared by the UTs. In this way, we precisely quantify the fundamental advantage of using the multiple BS antennas for efficient channel state feedback.
- A fundamental property of the system is that UTs are unaware of the chosen beamforming vectors, because the beamformers depend on all channels whereas each UT only has an estimate of its own channel. Several previous works (e.g., [26], [33], [34]) have resolved this uncertainty by making the unstated assumption that each UT has perfect knowledge of the post-beamforming SINR. In contrast, we make no such assumption and rigorously show that this ambiguity can be resolved by an additional round of (dedicated) training.
- Most prior work has used a worst-case uncorrelated noise argument [35][36][18] to show that imperfect CSI, at worse, leads to the introduction of additional Gaussian noise and thus the achievable rate is lower bounded by the mutual information with ideal channel state information and reduced SNR. In our case, however, this same argument yields a largely uncomputable quantity and a further step must be taken that yields a tractable lower bound in terms of the rate difference between the ideal and actual cases, rather than in terms of a SNR penalty.
- We consider delayed feedback and quantify in a simple and appealing form the loss of degrees of freedom (pre-log factor in the achievable rate) in terms of the fading channel Doppler bandwidth, which is ultimately related to UT velocity.

The analysis presented in this paper is relevant from at least two related but different viewpoints. On one hand, it provides accurate bounds on the achievable ergodic rates of the linear zero-forcing beamforming scheme with realistic channel estimation and feedback. These bounds are useful *at any operating SNR* (not necessarily large), and in subsequent work have been used to optimize the system resources allocated for training and feedback [37][38]. On the other hand, it yields sufficient conditions on the training and feedback such that the system achieves the same *multiplexing gain* (also referred to as “pre-log factor”, or “degrees of freedom”) of the optimal DPC-based scheme under perfect CSIR/CSIT. Perhaps the most striking fact about this second aspect is that the full multiplexing gain

<sup>1</sup>Since this work considers feedback schemes where the role of transmitter and receiver are reversed, we avoid using “transmitter” and “receiver” and prefer the use of BS and UT instead, in order to avoid ambiguity.

of the ideal MIMO broadcast channel can be achieved with simple pilot-based channel estimation and feedback schemes that consume a relatively small fraction of the system capacity. Indeed, a fundamental property of the MIMO broadcast channel is that the quality of the CSIT must increase with signal-to-noise ratio (SNR), regardless of what coding strategy is used, in order for the full multiplexing gain to be achievable [16], [17]. Under the reasonable assumption that the uplink channel quality is in some sense proportional to the downlink channel, our work shows that this requirement can be met using a fixed number of downlink and uplink channel symbols (i.e., system resources used for training and feedback need not increase with SNR).

When there is a significant delay in the feedback loop, the simple scheme analyzed in this paper does not attain full multiplexing gain. However, for fading processes with normalized Doppler bandwidth  $F$  strictly less than  $1/2$ , we show the achievability of a multiplexing gain equal to  $M(1 - 2F)$ , where  $M$  is the number of BS antennas. This result follows from a fundamental property of the noisy prediction error of the channel process and is closed related to Lapidath's high-SNR capacity of single-user fading channels without the perfect CSIR assumption [39].

The paper is organized as follows. Section II introduces the system model, describes linear beamforming, and defines the baseline estimation, feedback, and beamforming strategy. Section III develops bounds on the ergodic rates achievable by the baseline scheme. In Section IV we consider an AWGN feedback channel and particularize the rate bounds to analog and digital feedback (incorporating the effect of decoding errors for digital feedback), and compare the different feedback options. Section V generalizes the results to the setting where the feedback link is a fading MIMO multiple access channel (MAC). Section VI considers time-correlated fading and the effect of delay in the feedback link. Some concluding remarks are provided in Section VII.

## II. SYSTEM MODEL

We consider a multi-input multi-output (MIMO) Gaussian broadcast channel modeling the downlink of a system where a Base Station (BS) has  $M$  antennas and  $K$  User Terminals (UTs) have one antenna each. A channel use of such channel is described by

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

where  $y_k$  is the channel output at UT  $k$ ,  $z_k \sim \mathcal{CN}(0, N_0)$  is the corresponding Additive White Gaussian Noise (AWGN),  $\mathbf{h}_k \in \mathbb{C}^M$  is the vector of channel coefficients from the  $k$ -th UT antenna to the BS antenna array (the superscript H refers to the Hermitian, or conjugate transpose) and  $\mathbf{x}$  is the vector of channel input symbols transmitted by the BS. The channel input is subject to the average power constraint  $\mathbb{E}[|\mathbf{x}|^2] \leq P$ .

We assume that the channel *state*, given by the collection of all channel vectors  $\mathbf{H} = [\mathbf{h}_1, \dots, \mathbf{h}_K] \in \mathbb{C}^{M \times K}$ , varies in time according to a block-fading model [40], where  $\mathbf{H}$  is constant over each *frame* of length  $T$  channel uses, and evolves from frame to frame according to an ergodic stationary spatially white jointly Gaussian process, where the entries of  $\mathbf{H}$  are Gaussian i.i.d. with elements  $\sim \mathcal{CN}(0, 1)$ . Our bounds on the ergodic achievable rate do not directly depend on the frame size  $T$ ; rather, these bounds depend only on whether the training, feedback, and data phases all occur within a frame or in different frames. In Sections IV - V we consider the simplified scenario where the three phases all occur within a single frame (i.e., the channel is constant across the phases) and fading is independent across blocks, but we remove these simplifications in Section VI. It should also be noticed that the rate lower bounds given in the following should be multiplied by the factor  $(1 - \Delta/T)$ , where  $\Delta$  denotes the total number of channel uses per frame dedicated to training and feedback. This factor is neglected in this paper since it is common to all rate bounds and since  $\Delta \ll T$  in a typical slowly-fading system scenario. However, in the general case where  $\Delta$  is not necessarily small with respect to  $T$ , the amount of training and feedback should be optimized by taking this multiplicative factor into account. Based on the bounds developed in the present paper, this system optimization is carried out in the follow-up works [37], [38].

### A. Linear beamforming

Because of simplicity and robustness to non-perfect CSIT, simple *linear precoding* schemes with standard Gaussian coding have been extensively considered: the transmit signal is formed as  $\mathbf{x} = \mathbf{V}\mathbf{u}$ , such that  $\mathbf{V} \in \mathbb{C}^{M \times K}$  is a *linear beamforming* matrix and  $\mathbf{u} \in \mathbb{C}^K$  contains the symbols from  $K$  independently generated Gaussian codewords. In particular, for  $K \leq M$  Zero-Forcing beamforming chooses the  $k$ -th column  $\mathbf{v}_k$  of  $\mathbf{V}$  to be a unit vector orthogonal to the subspace  $\mathcal{S}_k = \text{span}\{\mathbf{h}_j : j \neq k\}$ .

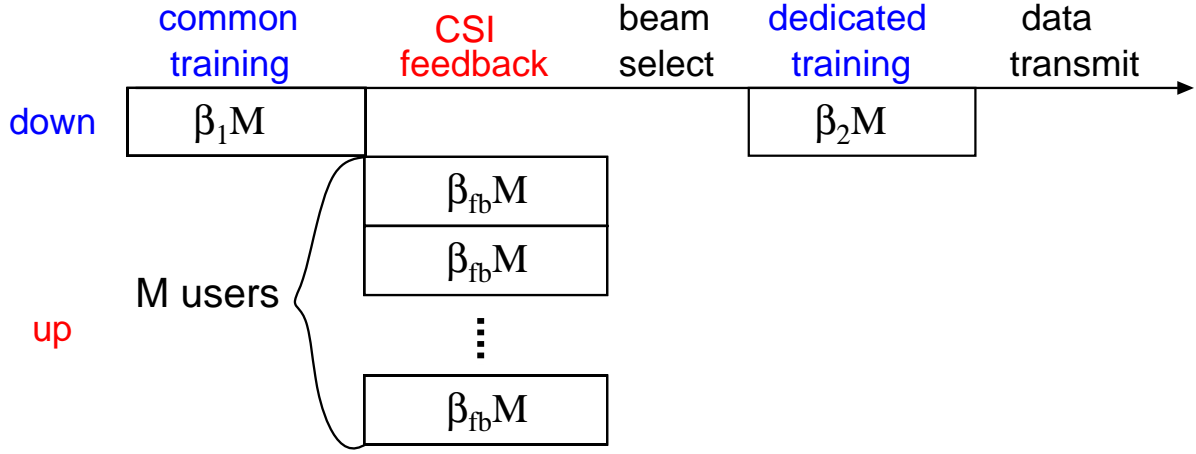


Fig. 1. Channel estimation and feedback model

We focus on the achievable *ergodic* rates under ZF linear beamforming and Gaussian coding. In this case, the achievable rate-sum is given by

$$\max_{\sum_k \mathbb{E}[\mathcal{P}_k(\mathbf{H})] \leq P} \sum_{k=1}^K \mathbb{E} \left[ \log \left( 1 + \frac{|\mathbf{h}_k^H \mathbf{v}_k|^2 \mathcal{P}_k(\mathbf{H})}{N_0} \right) \right]. \quad (2)$$

where the optimal power allocation is obtained by waterfilling over the set of channel gains  $\{|\mathbf{h}_k \mathbf{v}_k|^2 : k = 1, \dots, K\}$ . Performance can further be improved by using a *user scheduling* algorithm to select in each frame an *active* user subset not larger than  $M$  (if  $K > M$ , such selection must be performed if ZF is used). Schemes for user scheduling have been extensively discussed, for example in [41], [32], [15], [42].

We focus, however, on the case  $K = M$  with uniform power allocation (across users and frames:  $\mathcal{P}_k(\mathbf{H}) = \frac{P}{M}$ ) and without user selection, in which case the per-user ergodic rate is

$$R_k^{\text{ZF}}(P) = \mathbb{E} \left[ \log \left( 1 + \frac{|\mathbf{h}_k^H \mathbf{v}_k|^2 P}{N_0 M} \right) \right]. \quad (3)$$

Because  $\mathbf{h}_k$  is spatially white and  $\mathbf{v}_k$  is selected independent of  $\mathbf{h}_k$  (by the ZF procedure), it follows that  $\mathbf{h}_k^H \mathbf{v}_k$  is  $\sim \mathcal{CN}(0, 1)$ . As a result,  $R_k^{\text{ZF}}$  is the ergodic capacity of a point-to-point channel in Rayleigh fading with average SNR  $\frac{P}{N_0 M}$ , and thus can be written in closed form as [43]  $R_k^{\text{ZF}} = \exp\left(\frac{N_0 M}{P}\right) \text{E}_i\left(1, \frac{N_0 M}{P}\right)$  where  $\text{E}_i(n, x) = \int_1^\infty \frac{e^{-xt}}{t^n} dt$ ,  $x > 0$  [44]. In the remainder of the paper  $R_k^{\text{ZF}}$  serves as a benchmark against which we compare the achievable rates with imperfect CSI.

This restriction is dictated by a few reasons. On one hand, the case  $K = M$  without selection makes closed-form analysis (in the presence of imperfect CSI) possible. In addition, the maximum multiplexing gain is  $M$  for all  $K \geq M$  and hence the case  $K = M$  suffices to capture the fundamental aspects of the problem (particularly at high SNR). Finally, recent results [33], [45] show that the dependence on CSI quality is roughly the same even when user selection is performed.

### B. Channel state estimation and feedback

We assume that each UT estimates its channel vector from *downlink training symbols* and then feeds this information back to the BS. This scenario, referred to as “closed-loop” CSIT estimation, is relevant for Frequency-Division Duplexed (FDD) systems. Our baseline system is depicted in Fig. 1 and consists of the following phases:

- 1) **Common Training:** The BS transmits  $\beta_1 M$  shared pilots ( $\beta_1 \geq 1$  symbols per antenna) on the downlink<sup>2</sup>. Each UT  $k$  estimates its channel from the observation

$$\mathbf{s}_k = \sqrt{\beta_1 P} \mathbf{h}_k + \mathbf{z}_k \quad (4)$$

<sup>2</sup>If  $\beta_1$  is an integer, pilot symbols can be orthogonal in time, i.e.,  $\beta_1$  pilots are successively transmitted from each of the  $M$  BS antennas for a total of  $\beta_1 M$  channel uses. More generally, it is sufficient for  $\beta_1 M$  to be an integer and to use a unitary  $M \times \beta_1 M$  spreading matrix as described in [28]; in either case the effective received SNR is  $\beta_1 \frac{P}{N_0}$ .

corresponding to the common training (downlink) channel output, where  $\mathbf{z}_k \sim \mathcal{CN}(0, N_0 \mathbf{I})$ . The MMSE estimate  $\tilde{\mathbf{h}}_k$  of  $\mathbf{h}_k$  given the observation  $\mathbf{s}_k$  is given by [46]:

$$\tilde{\mathbf{h}}_k = \mathbb{E}[\mathbf{h}_k \mathbf{s}_k^H] \mathbb{E}[\mathbf{s}_k \mathbf{s}_k^H]^{-1} \mathbf{s}_k = \frac{\sqrt{\beta_1 P}}{N_0 + \beta_1 P} \mathbf{s}_k \quad (5)$$

The channel  $\mathbf{h}_k$  can be written in terms of the estimate  $\tilde{\mathbf{h}}_k$  and estimation noise  $\mathbf{n}_k$  as:

$$\mathbf{h}_k = \tilde{\mathbf{h}}_k + \mathbf{n}_k, \quad (6)$$

where  $\mathbf{n}_k$  is independent of the estimate and is Gaussian with covariance  $\sigma_1^2 \mathbf{I}$  with

$$\sigma_1^2 = \frac{1}{1 + \beta_1 P / N_0} \quad (7)$$

- 2) **Channel State Feedback:** Each UT feeds back its channel estimate  $\tilde{\mathbf{h}}_k$  to the BS immediately after completion of the common training phase. We use  $\hat{\mathbf{H}} = [\hat{\mathbf{h}}_1, \dots, \hat{\mathbf{h}}_K] \in \mathbb{C}^{M \times K}$  to denote the (imperfect) CSIT available at the BS; the feedback is thus a mapping, possibly probabilistic, from  $\tilde{\mathbf{h}}_k$  to  $\hat{\mathbf{h}}_k$ . For now we leave the feedback scheme unspecified to allow development of general achievability bounds in Section III, and particularize to specific feedback schemes from Section IV onwards.

In Section IV we consider the simplified setting where the feedback channel is an unfaded AWGN channel SNR  $\frac{P}{N_0}$ , orthogonal across UTs, but in Section V we consider the more realistic setting where the uplink channel is a MIMO-MAC with fading. Furthermore, the baseline model of Fig. 1 assumes no delay in the feedback, i.e., the channel is constant across the training, feedback, and data phases. In Section VI we remove this assumption and consider the case where feedback has delay and the channel state changes from frame to frame according to a time-correlation model.

We assume each UT transmits its feedback over  $\beta_{\text{fb}} M$  feedback channel symbols.

- 3) **Beamformer Selection:** The BS selects the beamforming vectors by treating the estimated CSIT  $\hat{\mathbf{H}}$  as if it was the true channel (we refer to this approach as “naive” ZF beamforming). Following the ZF recipe,  $\hat{\mathbf{v}}_k$  is a unit vector orthogonal to the subspace  $\mathcal{S}_k = \text{span}\{\hat{\mathbf{h}}_j : j \neq k\}$ . We use the notation  $\hat{\mathbf{V}} = [\hat{\mathbf{v}}_1, \dots, \hat{\mathbf{v}}_K]$ . Since  $K = M$  and the BS channel estimates  $\hat{\mathbf{h}}_1, \dots, \hat{\mathbf{h}}_M$  are independent, the subspace  $\mathcal{S}_k$  is  $M - 1$  dimensional (with probability one) and is independent of  $\hat{\mathbf{h}}_k$ . The beamforming vector  $\hat{\mathbf{v}}_k$  is chosen in the one-dimensional nullspace of  $\mathcal{S}_k$ ; as a result  $\hat{\mathbf{v}}_k$  is independent of the channel estimate  $\hat{\mathbf{h}}_k$  and of the true channel vector  $\mathbf{h}_k$ .
- 4) **Dedicated Training:** Once the the BS has computed the beamforming vectors  $\hat{\mathbf{V}}$ , coherent detection of data at each UT is enabled by an additional round of downlink training transmitted along each beamforming vector. This additional round of training is required because the beamforming vectors  $\{\hat{\mathbf{v}}_k\}$  are functions of the channel state information  $\{\hat{\mathbf{h}}_1, \dots, \hat{\mathbf{h}}_K\}$  at the BS, while UT  $k$  knows only  $\tilde{\mathbf{h}}_k$  or, at best,  $\hat{\mathbf{h}}_k$  (if error-free digital feedback is used). Therefore, the coupling coefficients between the beamforming vectors and the UT channel vector are unknown.

Let the set of the coefficients affecting the signal received by UT  $k$  be denoted by

$$\mathcal{A}_k \triangleq \{a_{k,j} : j = 1, \dots, M\}$$

where  $a_{k,j} = \mathbf{h}_k^H \hat{\mathbf{v}}_j$  is the coupling coefficient between the  $k$ -th channel and the  $j$ -th beamforming vector. The received signal at the  $k$ -th UT is given by

$$\begin{aligned} y_k = \mathbf{h}_k^H \hat{\mathbf{V}} \mathbf{u} + z_k &= a_{k,k} u_k + \sum_{j \neq k} a_{k,j} u_j + z_k \\ &= a_{k,k} u_k + I_k + z_k \end{aligned} \quad (8)$$

where the interference at UT  $k$  is denoted as

$$I_k = \sum_{j \neq k} a_{k,j} u_j \quad (9)$$

and  $a_{k,k}$  is the *useful signal* coefficient. The dedicated training is intended to allow the estimation of the coefficients in  $\mathcal{A}_k$  at each UT  $k$ . This is accomplished by transmitting  $\beta_2$  orthogonal training symbols along

each of the beamforming vectors on the downlink, thus requiring a total of  $\beta_2 M$  downlink channel uses.<sup>3</sup> The relevant observation model for the estimation of  $\mathcal{A}_k$  is given by

$$r_{k,j} = \sqrt{\beta_2 P} a_{k,j} + z_{k,j}, \quad j = 1, \dots, M \quad (10)$$

We denote the full set of observations available to UT  $k$  as:

$$\mathcal{R}_k \triangleq \{r_{k,j} : j = 1, \dots, M\}.$$

In particular, we shall consider explicitly the case where UT  $k$  estimates its useful signal coefficient using linear MMSE estimation based on  $r_{k,k}$ , i.e.,

$$\hat{a}_{k,k} = \frac{\sqrt{\beta_2 P}}{N_0 + \beta_2 P} r_{k,k}. \quad (11)$$

Because  $\hat{\mathbf{v}}_k$  is a unit vector independent of  $\mathbf{h}_k$ , the useful signal coefficient  $a_{k,k} = \mathbf{h}_k^H \hat{\mathbf{v}}_k$  is complex Gaussian with unit variance. As a result we have the representation

$$a_{k,k} = \hat{a}_{k,k} + f_k \quad (12)$$

where  $f_k$  and  $\hat{a}_{k,k}$  are independent and Gaussian with variance  $\sigma_2^2$  and  $1 - \sigma_2^2$ , respectively, with

$$\sigma_2^2 = \frac{1}{1 + \beta_2 P / N_0}. \quad (13)$$

5) **Data Transmission:** After the dedicated downlink training phase, the BS sends the coded data symbols  $u_1, \dots, u_K$  for the rest of the frame duration. The effective channel output for this phase is therefore given by the sequence of corresponding channel output symbols  $y_k$  given by (8), and by the observation of the dedicated training phase  $\mathcal{R}_k$  given by (10).

When considering the *ergodic* rates achievable by the proposed scheme, we implicitly assume that coding is performed over a long sequence of frames, each frame comprising a common training phase, channel state feedback phase, dedicated training phase and data transmission.

We conclude this section with a few remarks. First, we would like to observe that two phases of training, a common ‘‘pilot channel’’ and dedicated per-user training symbols is common practice in some wireless cellular systems, as for example in the downlink of the 3rd generation Wideband CDMA standard [47] and in the MIMO component of future 4th generation systems [48]. Second, we note that an alternative to FDD is Time-Division Duplexing (TDD), where uplink and downlink share in time-division the same frequency band. In this case, provided that the coherence time is significantly larger than the concatenation of an uplink and downlink slot and hardware calibration, the downlink channel can be learned by the BS from uplink training symbols [28], [49]. Although we focus on FDD systems, in Remark 4.2 we note the straightforward extension of our results to TDD systems.

### III. ACHIEVABLE RATE BOUNDS

We assume that the user codes are independently generated according to an i.i.d. Gaussian distribution, i.e., the input symbols are  $u_k \sim \mathcal{CN}(0, P/M)$ . The remainder of this section is dedicated to deriving upper and lower bounds on the mutual information achieved by such Gaussian inputs, indicated by  $R_k \triangleq I(u_k; y_k, \mathcal{R}_k)$ .

#### A. Lower Bounds

The following lower bound is obtained by using techniques similar to those in [35], [18], [36].

*Theorem 1:* The achievable rate for ZF beamforming with Gaussian inputs and CSI training and feedback as described in Section II-B can be bounded from below by:

$$R_k \geq \mathbb{E} \left[ \log \left( 1 + \frac{|\hat{a}_{k,k}|^2 P / (N_0 M)}{1 + \sigma_2^2 P / (N_0 M) + \mathbb{E}[|I_k|^2 |\hat{a}_{k,k}| / N_0]} \right) \right] \quad (14)$$

**Proof:** See Appendix I. □

<sup>3</sup>If  $\beta_2 M$  is an integer but  $\beta_2$  is not, the unitary spreading approach used for common training can also be used here.

The conditional interference second moment  $\mathbb{E}[|I_k|^2|\hat{a}_{k,k}]$  in (14) may be difficult to compute even by Monte Carlo simulation, due to the complicated dependency of  $I_k$  on  $\hat{a}_{k,k}$  (this dependence is unknown even if the dedicated training is perfect, i.e.,  $\hat{a}_{k,k} = a_{k,k}$ ). However, we will not need to compute this explicitly, as is seen in our next results.

A very useful measure is the difference between  $R_k$  and  $R_k^{\text{ZF}}$ , the achievable rate with ZF beamforming and ideal CSI defined in (3). The rate gap is defined as follows

$$\Delta R \triangleq R_k^{\text{ZF}} - R_k, \quad (15)$$

and is upper bounded in the following theorem.

*Theorem 2:* The rate gap incurred by ZF beamforming with training and feedback as described in Section II-B with respect to ideal ZF with equal power allocation is upperbounded by:

$$\Delta R \leq \log \left( 1 + \sigma_2^2 \frac{P}{N_0 M} + \frac{\mathbb{E}[|I_k|^2]}{N_0} \right) \quad (16)$$

**Proof:** See Appendix II. □

For clarity of notation, we denote the RHS of the above, referred to as the rate gap upper bound, as  $\overline{\Delta R}$ :

$$\overline{\Delta R} \triangleq \log \left( 1 + \sigma_2^2 \frac{P}{N_0 M} + \frac{\mathbb{E}[|I_k|^2]}{N_0} \right) \quad (17)$$

$$= \log \left( 1 + \frac{P}{N_0 M} \left( \sigma_2^2 + \sum_{j \neq k} \mathbb{E}[|\mathbf{h}_k^H \hat{\mathbf{v}}_j|^2] \right) \right) \quad (18)$$

where the latter follows from a simple calculation of  $\mathbb{E}[|I_k|^2]$ . The term  $\sigma_2^2$  depends only on dedicated training; on the other hand,  $\mathbb{E}[|\mathbf{h}_k^H \hat{\mathbf{v}}_j|^2]$  is determined by the mismatch between  $\mathbf{h}_k$  and the BS estimate  $\hat{\mathbf{h}}_k$  (because  $\hat{\mathbf{v}}_j$  is chosen orthogonal to  $\hat{\mathbf{h}}_k$  rather than  $\mathbf{h}_k$ ) and therefore depends on the common training and feedback phases.

An obvious result of the rate gap upper bound is the following lower bound to  $R_k$ :

*Corollary 3.1:* The achievable rate for ZF beamforming with Gaussian inputs and CSIT training and feedback as described in Section II-B can be bounded from below by:

$$R_k \geq R_k^{\text{ZF}} - \overline{\Delta R} \quad (19)$$

Because only the estimate of  $a_{k,k}$  is used in the derivation, Corollary 3.1 is also a lower bound to  $I(u_k; y_k, r_{k,k})$ .

## B. Upper Bounds

A useful upper bound to  $R_k$  is reached by providing each UT  $k$  with exact knowledge of the interference coefficients  $\mathcal{A}_k$ . Thus, this is referred to as the ‘‘genie-aided upper-bound’’.

*Theorem 3:* The achievable rate for ZF beamforming with Gaussian inputs and CSI training and feedback is upper bounded by the rate achievable when, after the beamforming matrix  $\hat{\mathbf{V}}$  is chosen, a genie provides the  $k$ -th UT with perfect knowledge of the coefficients  $\mathcal{A}_k = \{a_{k,j} = \mathbf{h}_k^H \hat{\mathbf{v}}_j : j = 1, \dots, M\}$ :

$$R_k \leq \mathbb{E} \left[ \log \left( 1 + \frac{|a_{k,k}|^2 P / (N_0 M)}{1 + \sum_{j \neq k} |a_{k,j}|^2 P / (N_0 M)} \right) \right]. \quad (20)$$

**Proof:** Since  $\mathcal{R}_k$  is a noisy version of  $\mathcal{A}_k$ , the data-processing inequality yields

$$R_k = I(u_k; y_k, \mathcal{R}_k) \leq I(u_k; y_k, \mathcal{A}_k) \quad (21)$$

Because  $y_k$  conditioned on  $\mathcal{A}_k$  is complex Gaussian with variance  $N_0 + \sum_{j=1}^M |a_{k,j}|^2 P / M$  while  $y_k$  conditioned on  $(\mathcal{A}_k, u_k)$  is complex Gaussian with variance  $N_0 + \sum_{j \neq k} |a_{k,j}|^2 P / M$ , we immediately obtain (20). □

The practical relevance of Theorem 3 is two-fold: on one hand, (20) is easy to evaluate by Monte Carlo simulation.<sup>4</sup> On the other hand, this bound can be approached for large  $\beta_2$ , since in this case each UT can accurately estimate all interference coupling coefficients and not only the useful signal coefficient.

<sup>4</sup>It is usually difficult if not impossible to obtain in closed form the joint distribution of the coefficients  $\mathcal{A}_k$ .

#### IV. CHANNEL STATE FEEDBACK OVER AN AWGN CHANNEL

In this section we quantify the rate gap upper bound for different feedback strategies under the assumption that the feedback channel is an unfaded AWGN channel with the same SNR as the downlink, i.e.,  $P/N_0$ , and that the UTs access the channel orthogonally. Each UT uses  $\beta_{\text{fb}}M$  feedback channel symbols, and therefore the total number of feedback channel uses is  $\beta_{\text{fb}}M^2$ .

##### A. Analog feedback

Analog feedback refers to transmission (on the feedback link) of the estimated downlink channel coefficients by each UT using unquantized quadrature-amplitude modulation [28], [32], [50], [51]. More specifically, each UT transmits on the feedback channel a scaled version of its common downlink training observation  $\mathbf{s}_k$  defined in (4). The resulting feedback channel output (BS observation) relative to UT  $k$  is given by:

$$\mathbf{g}_k = \frac{\sqrt{\beta_{\text{fb}}P}}{\sqrt{\beta_1P + N_0}} \mathbf{s}_k + \tilde{\mathbf{w}}_k \quad (22)$$

$$= \frac{\sqrt{\beta_{\text{fb}}\beta_1P}}{\sqrt{\beta_1P + N_0}} \mathbf{h}_k + \frac{\sqrt{\beta_{\text{fb}}P}}{\sqrt{\beta_1P + N_0}} \mathbf{z}_k + \tilde{\mathbf{w}}_k \quad (23)$$

$$= \frac{\sqrt{\beta_{\text{fb}}\beta_1P}}{\sqrt{\beta_1P + N_0}} \mathbf{h}_k + \mathbf{w}_k \quad (24)$$

where  $\tilde{\mathbf{w}}_k$  represents the AWGN noise on the uplink feedback channel (variance  $N_0$ ) and  $\mathbf{z}_k$  is the noise during the common training phase. The power scaling  $\beta_{\text{fb}}$  corresponds to the number of channel uses per channel coefficient (we require  $\beta_{\text{fb}} \geq 1$  so that each coefficient is transmitted at least once), assuming that transmission in the feedback channel has per-symbol power  $P$  (averaged over frames) and that the channel state vector is modulated by a  $\beta_{\text{fb}}M \times M$  unitary spreading matrix [28]. Because  $\tilde{\mathbf{w}}_k$  and  $\mathbf{z}_k$  are each complex Gaussian with covariance  $N_0\mathbf{I}$  and are independent,  $\mathbf{w}_k$  is complex Gaussian with covariance  $\sigma_w^2\mathbf{I}$  with:

$$\sigma_w^2 = N_0 \left( 1 + \frac{\beta_{\text{fb}}P/N_0}{1 + \beta_1P/N_0} \right) \quad (25)$$

The BS computes the MMSE estimate of the channel vector  $\mathbf{h}_k$  based on  $\mathbf{g}_k$  as:

$$\hat{\mathbf{h}}_k = \frac{\sqrt{\beta_{\text{fb}}\beta_1P}}{\sqrt{\beta_1P + N_0}(\beta_{\text{fb}}P + N_0)} \mathbf{g}_k. \quad (26)$$

Using (24), the channel can be written in terms of the BS estimate and estimation error  $\mathbf{e}_k$  as:

$$\mathbf{h}_k = \hat{\mathbf{h}}_k + \mathbf{e}_k \quad (27)$$

where  $\mathbf{e}_k$  is independent of the estimate and is Gaussian with covariance  $\sigma_e^2\mathbf{I}$  with:

$$\sigma_e^2 = \frac{\sigma_w^2}{\sigma_w^2 + \frac{\beta_{\text{fb}}\beta_1P^2}{\beta_1P + N_0}} = \frac{1}{1 + \beta_{\text{fb}}\frac{P}{N_0}} + \frac{\beta_{\text{fb}}\frac{P}{N_0}}{(1 + \beta_{\text{fb}}\frac{P}{N_0})(1 + \beta_1\frac{P}{N_0})}. \quad (28)$$

This characterization of  $(\mathbf{h}_k, \hat{\mathbf{h}}_k)$  can be used to derive the rate gap upper bound for analog feedback:

*Theorem 4:* If each UT feeds back its channel coefficients in analog fashion over  $\beta_{\text{fb}}M$  channel uses of an AWGN uplink channel with SNR  $\frac{P}{N_0}$ , the rate gap upper bound is given by (“AF” standing for Analog Feedback):

$$\overline{\Delta R}^{\text{AF}} = \log \left( 1 + \frac{P}{N_0M} \left( \frac{1}{1 + \beta_2\frac{P}{N_0}} + (M-1) \left( \frac{1}{1 + \beta_{\text{fb}}\frac{P}{N_0}} + \frac{\beta_{\text{fb}}\frac{P}{N_0}}{(1 + \beta_{\text{fb}}\frac{P}{N_0})(1 + \beta_1\frac{P}{N_0})} \right) \right) \right). \quad (29)$$

**Proof:** See Appendix III.  $\square$

It is straightforward to see that  $\overline{\Delta R}^{\text{AF}}$  can be upper bounded as

$$\overline{\Delta R}^{\text{AF}} \leq \log \left( 1 + \frac{1}{M\beta_2} + \frac{M-1}{M} \left( \frac{1}{\beta_1} + \frac{1}{\beta_{\text{fb}}} \right) \right), \quad (30)$$

Hence, the rate gap is uniformly bounded for all SNRs and therefore the multiplexing gain is preserved (i.e.,  $\lim_{P \rightarrow \infty} \frac{R_k}{\log_2 P} = 1$ ) in spite of the imperfect CSI.

An *intuitive* understanding of this rate loss is obtained if one re-examines the UT received signal in the form used in Theorem 1:

$$y_k = \hat{a}_{k,k}u_k + \underbrace{f_k u_k}_{\text{Self Noise}} + \underbrace{\sum_{j \neq k} (\mathbf{h}_k^H \hat{\mathbf{v}}_j) u_j}_{\text{Interference}} + \underbrace{z_k}_{\text{Noise}} \quad (31)$$

The imperfect channel state information (at the UT and BS) effectively increases the noise from the thermal noise level  $N_0$  to the sum of the thermal noise, self-noise, and interference power, and the rate gap upper bound  $\overline{\Delta R}^{\text{AF}}$  is precisely the logarithm of the ratio of the effective noise to the thermal noise power.

**Remark 4.1:** In many systems, the uplink SNR is smaller than the downlink SNR because UT's transmit with reduced power. If the uplink SNR is  $\Gamma \frac{P}{N_0}$  rather than  $\frac{P}{N_0}$ ,  $\overline{\Delta R}^{\text{AF}}$  is equal to the expression in Theorem 4 with  $\beta_{\text{fb}}$  replaced with  $\Gamma \beta_{\text{fb}}$ . This does not change the multiplexing gain, but can have a significant effect on the rate gap.  $\diamond$

**Remark 4.2:** It is easy to see that a TDD system with perfectly reciprocal uplink-downlink channels where each UT transmits  $\beta_{TDD}$  pilots (a single pilot trains all  $M$  BS antennas) in an orthogonal manner corresponds exactly to an FDD system with perfect feedback ( $\beta_{\text{fb}} \rightarrow \infty$ ) and  $\beta_1 = \beta_{TDD}$ , because the downlink training in an FDD system is equivalent to the uplink training in a TDD system. Therefore, as a byproduct of our analysis, we obtain a result for TDD open loop CSIT estimation:

$$\overline{\Delta R}^{\text{TDD}} = \log \left[ 1 + \frac{P}{N_0 M} \left( \frac{1}{1 + \beta_2 \frac{P}{N_0}} + \frac{M-1}{1 + \beta_{TDD} \frac{P}{N_0}} \right) \right] \quad (32)$$

$$\leq \log \left( 1 + \frac{1}{M \beta_2} + \frac{M-1}{M} \frac{1}{\beta_{TDD}} \right). \quad (33)$$

Dedicated training is necessary even in TDD systems because UT's do not know the channels of other UT's and thus are not aware of the beamforming vectors used by the BS. Finally, note that in TDD a total of  $M\beta_{TDD}$  uplink training symbols and  $M\beta_2$  downlink (dedicated) training symbols are needed.  $\diamond$

## B. Digital feedback

We now consider “digital” feedback, where the estimated channel vector is quantized at each UT and represented by  $B$  bits. The packet of  $B$  bits is fed back by each UT to the BS. We begin by computing the rate gap upper bound in terms of *bits*, and later in the section relate this to feedback *channel uses*.

Following [21], [20], [19], [26], we consider a specific scheme for channel state quantization based on a quantization codebook  $\mathcal{C} = \{\mathbf{p}_1, \dots, \mathbf{p}_{2^B}\}$  of unit-norm vectors in  $\mathbb{C}^M$ . The quantization  $\hat{\mathbf{h}}_k$  of the estimated channel vector  $\tilde{\mathbf{h}}_k$  is found according to the decision rule:

$$\hat{\mathbf{h}}_k = \arg \max_{\mathbf{p} \in \mathcal{C}} |\tilde{\mathbf{h}}_k^H \mathbf{p}|^2 \quad (34)$$

and thus  $\hat{\mathbf{h}}_k$  is the quantization vector forming the minimum angle with  $\tilde{\mathbf{h}}_k$ . The corresponding  $B$ -bits quantization index is fed back to the BS. Because  $\hat{\mathbf{h}}_k$  is unit-norm, no channel magnitude information is conveyed.

In [24], [26] it is shown that for a random ensemble of quantization codebooks referred to as *Random Vector Quantization* (RVQ), obtained by generating  $2^B$  quantization vectors independently and uniformly distributed on the unit sphere in  $\mathbb{C}^M$  (see [26] and references therein), the average (angular) distortion is given by:

$$\mathbb{E} \left[ \sin^2 \left( \tilde{\mathbf{h}}_k, \hat{\mathbf{h}}_k \right) \right] = 2^B \beta \left( 2^B, \frac{M}{M-1} \right) \leq 2^{-\frac{B}{M-1}} \quad (35)$$

where  $\beta(\cdot)$  is the beta function and  $\sin^2 \left( \tilde{\mathbf{h}}_k, \hat{\mathbf{h}}_k \right) = 1 - \frac{|\tilde{\mathbf{h}}_k^H \hat{\mathbf{h}}_k|^2}{\|\tilde{\mathbf{h}}_k\|^2}$ . As in [26] we assume each UT uses an independently generated codebook. For this particular quantization scheme, we can compute the rate gap upper bound:

*Theorem 5:* If each UT quantizes its channel to  $B$  bits (using RVQ) and conveys these bits in an error-free fashion to the BS, the rate gap upper bound is given by (“DF” standing for Digital Feedback):

$$\overline{\Delta R}^{\text{DF}} = \log \left( 1 + \frac{P}{N_0 M} \left( \frac{1}{1 + \beta_2 \frac{P}{N_0}} + \frac{M}{1 + \beta_1 \frac{P}{N_0}} \left[ \frac{M-1}{M} + \frac{\beta_1 P}{N_0} 2^B \beta \left( 2^B, \frac{M}{M-1} \right) \right] \right) \right). \quad (36)$$

**Proof:** See Appendix IV.  $\square$

Using (35), the rate gap upper bound is further upper bounded as:

$$\overline{\Delta R}^{\text{DF}} \leq \log \left( 1 + \frac{1}{M\beta_2} + \frac{M-1}{M} \frac{1}{\beta_1} + \left( \frac{P}{N_0} \right) 2^{-\frac{B}{M-1}} \right) \quad (37)$$

Comparing this to the rate gap in the analog feedback case (30), we notice that the dependence on  $\beta_1$  and  $\beta_2$  are precisely the same for both analog and digital feedback.

The next step is translating the rate gap upper bound so that it is in terms of feedback symbols rather than bits. For the time being, we shall make the very unrealistic assumption that the feedback link can operate error-free at capacity, i.e., it can reliably transmit  $\log_2(1 + P/N_0)$  bits per symbol.<sup>5</sup>

The analog feedback considered before provides a noisy version of the channel vector norm in addition to its direction. Although this information is irrelevant for the ZF beamforming considered here, it might be useful in some user selection algorithms such as those proposed in [41], [32], [15], [42]. In contrast, digital feedback based on unit-norm quantization vectors provides no norm information. Thus, for fair comparison, we assume that  $\beta_{\text{fb}}M$  feedback symbols in the analog feedback scheme correspond to  $\beta_{\text{fb}}(M-1)$  feedback symbols for the digital feedback scheme; i.e., a system using digital feedback could use one feedback symbol to transmit channel norm information. An alternative justification for this is to notice that the analog feedback system could be modified to operate in  $\beta_{\text{fb}}(M-1)$  channel symbols by transmitting only the  $M-1$  *relative* phases and amplitudes of the channel coefficients, since the absolute norm and phase are irrelevant to the ZF beamforming considered here.

Under this assumption, the number of feedback bits per mobile is  $B = \beta_{\text{fb}}(M-1) \log_2(1 + P/N_0)$ . Plugging this into (37) gives:

$$\overline{\Delta R}^{\text{DF}} \leq \log \left( 1 + \frac{1}{M\beta_2} + \frac{M-1}{M} \frac{1}{\beta_1} + \frac{\frac{P}{N_0}}{\left( 1 + \frac{P}{N_0} \right)^{\beta_{\text{fb}}}} \right) \quad (38)$$

Similar to analog feedback, if  $\beta_{\text{fb}} \geq 1$  then the rate gap is upper bounded and full multiplexing gain is preserved. However, it should be noticed that for  $\beta_{\text{fb}}$  strictly larger than 1 digital feedback yields a term  $\left( \frac{P}{N_0} \right)^{1-\beta_{\text{fb}}}$  that vanishes as  $P/N_0 \rightarrow \infty$ . This should be contrasted with the constant term  $\frac{1}{\beta_{\text{fb}}}$  for the case of analog feedback.

### C. Effects of feedback errors

We now remove the optimistic assumption that the digital feedback channel can operate error-free at capacity. In general, coding for the CSIT feedback channel should be regarded as a joint source-channel coding problem, made particularly interesting by the non-standard distortion measure and by the fact that a very short block length is required. A thorough discussion of this subject is out of the scope of the present paper and is the matter of current investigation (see for example [52], [53]). Here, we restrict ourselves to the detailed analysis of a particularly simple scheme based on uncoded QAM. Perhaps surprisingly, this scheme is *sufficient* to achieve a vanishing rate gap in the high SNR region, for an appropriate choice of the system parameters.

In the proposed scheme, the UTs perform quantization using RVQ and transmit the feedback bits using plain uncoded QAM. The quantization bits are randomly mapped onto the QAM symbols (i.e., no intelligent bit-labeling or mapping is used). Therefore, even a single erroneous feedback bit from UT  $k$  makes the BS's CSIT vector  $\hat{\mathbf{h}}_k$  essentially useless. Also, no particular error detection strategy is used and thus the BS computes the beamforming matrix on the basis of the received feedback, although this may be in error.

We again let  $\beta_{\text{fb}}(M-1)$  denote the number of channel uses to transmit the feedback bits (per UT). Interestingly, even for this very simple scheme there is a non-trivial tradeoff between quantization distortion and channel errors. In order to maintain a bounded rate gap, the number of feedback bits must be scaled at least as  $(M-1) \log_2 \left( 1 + \frac{P}{N_0} \right) \approx (M-1) \log_2 \frac{P}{N_0}$ . Therefore, we consider sending  $B = \alpha(M-1) \log_2 \frac{P}{N_0}$  bits for  $1 \leq \alpha \leq \beta_{\text{fb}}$  in  $\beta_{\text{fb}}(M-1)$  channel uses, which corresponds to  $\frac{\alpha}{\beta_{\text{fb}}} \log_2 \frac{P}{N_0}$  bits per QAM symbol.

<sup>5</sup>This assumption is unrealistic in the context of this model because the feedback channel coding block length is very small and because the need for very fast feedback (essentially delay-free) prevents grouping blocks of channel coefficients and using larger coding block length.

The symbol error rate for square QAM with  $q$  constellation points is bounded by [54]:

$$P_s = 1 - \left(1 - 2 \left(1 - \frac{1}{\sqrt{q}}\right) Q\left(\frac{3(P/N_0)}{q-1}\right)\right)^2 \leq 2 \exp\left(-\frac{3}{2} \frac{P/N_0}{q-1}\right). \quad (39)$$

where  $Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$  is the Gaussian probability tail function. Using the fact that  $q = (P/N_0)^{\frac{\alpha}{\beta_{\text{fb}}}}$ , we obtain the upper bound

$$P_s \leq 2 \exp\left(-\frac{3}{2} \left(\frac{P}{N_0}\right)^{1-\frac{\alpha}{\beta_{\text{fb}}}}\right) \quad (40)$$

If  $\alpha = \beta_{\text{fb}}$ , which corresponds to signaling at capacity with uncoded modulation,  $P_s$  does not decrease with SNR and system performance is very poor. However, for  $\alpha < \beta_{\text{fb}}$ , which corresponds to transmitting at a fraction of capacity,  $P_s \rightarrow 0$  as  $\frac{P}{N_0} \rightarrow \infty$ . The error probability of the entire feedback message (transmitted in  $\beta_{\text{fb}}(M-1)$  QAM symbols) is given by

$$P_{e,\text{fb}} = 1 - (1 - P_s)^{\beta_{\text{fb}}(M-1)} \leq \beta_{\text{fb}}(M-1)P_s, \quad (41)$$

where the inequality follows from the union bound. Note the tradeoff between distortion and feedback error:  $\alpha$  large yields finer quantization but larger  $P_{e,\text{fb}}$ , while  $\alpha$  small provides poorer quantization but smaller  $P_{e,\text{fb}}$ .

*Theorem 6:* If each UT quantizes its estimated channel using  $B = \alpha(M-1) \log P/N_0$  bits (using RVQ), and transmits on the feedback link using  $\beta_{\text{fb}}(M-1)$  channel uses with uncoded QAM modulation, the resulting rate gap can be upperbounded by

$$\overline{\Delta R}^{\text{DF-ERRORS}} \leq \log\left(1 + \frac{1}{M\beta_2} + (1 - P_{e,\text{fb}}) \left(\left(\frac{P}{N_0}\right)^{1-\alpha} + \frac{M-1}{M} \frac{1}{\beta_1}\right) + \frac{P}{N_0} P_{e,\text{fb}}\right), \quad (42)$$

where  $P_{e,\text{fb}}$  is given by (40) and (41).

**Proof:** See Appendix V.  $\square$

If  $1 < \alpha < \beta_{\text{fb}}$ , then the effect of feedback vanishes as  $\frac{P}{N_0} \rightarrow \infty$ , somewhat similar to the case of error-free feedback. This is because the feedback error probability decays exponentially as  $(P/N_0)^{1-\frac{\alpha}{\beta_{\text{fb}}}}$ , so that the term  $\frac{P}{N_0} P_{e,\text{fb}}$  vanishes as  $\frac{P}{N_0} \rightarrow \infty$  for all  $\alpha < \beta_{\text{fb}}$ , while obviously  $(P/N_0)^{1-\alpha}$  vanishes for all  $\alpha > 1$ .

A number of simple improvements are possible. For example, each UT may estimate its interference coefficients  $\{a_{k,j} : j \neq k\}$  from the dedicated training phase, and decide if its feedback message was correctly received or was received in error by setting a threshold on the interference power: if the interference power is  $\approx (M-1)P$ , then it is likely that a feedback error occurred. If, on the contrary, it is  $\approx 2^{-B/(M-1)}P$ , then it is likely that the feedback message was correctly received. Interestingly, for  $B = \alpha(M-1) \log_2 \frac{P}{N_0}$  with  $\alpha > 1$ , detecting feedback error events becomes easier and easier as  $\frac{P}{N_0}$  increases and/or as the number of antennas  $M$  increases. In brief, for a large number of antennas any terminal whose feedback message was received in error is completely drowned into interference and should be able to detect this event with high probability. Assuming that the UTs can perfectly detect their own feedback error events as described above, then they can simply discard the frames corresponding to feedback errors. The resulting achievable rate in this case is lowerbounded by

$$R_k^{\text{DF-ERRORS-DETECT}} \geq (1 - P_{e,\text{fb}}) \left[ R_k^{\text{ZF}} - \log\left(1 + \frac{1}{M\beta_2} + \frac{M-1}{M} \frac{1}{\beta_1} + \left(\frac{P}{N_0}\right)^{1-\alpha}\right) \right] \quad (43)$$

in light of (38) (after replacing  $\alpha$  instead of  $\beta_{\text{fb}}$ ) and of Corollary 3.1. Note that this rate lies between the achievable rate lower bound obtained via the rate gap in (42) and the genie-aided upper bound from Theorem 3.

**Remark 4.3:** It is interesting to notice that feedback errors make the residual interference behave as an impulsive noise: it has very large variance with small probability  $P_{e,\text{fb}}$ . It is therefore clear that detecting the feedback errors and discarding the corresponding frames yields significant improvements. Using this knowledge at the receiver (as in the rate bound (43)), avoids the large ‘‘Jensen’s penalty’’ incurred by the rate gap in (42), where the expectation with respect to the feedback error events is taken *inside* the logarithm.  $\diamond$

**Remark 4.4:** We notice here that the naive ZF strategy examined in this paper is *robust to feedback errors* in the following sense: the residual interference experienced by a given UT depends only on that particular UT

feedback error probability. Therefore, a small number of users with poor feedback channel quality (very high feedback error probability) does not destroy the overall system performance. This observation goes against the conventional wisdom that feedback errors are “catastrophic”.  $\diamond$

#### D. Comparison between analog and digital channel feedback

Based upon the bounds developed in the previous subsections as well as the genie-aided upper bounds (computed using Monte Carlo simulation) we can now compare analog, error-free digital, and QAM-based digital feedback. Because the effect of downlink and common training is effectively the same for all feedback strategies, we pursue this comparison under the assumption of perfect CSIR, i.e., perfect common and dedicated training corresponding to  $\beta_1 = \beta_2 \rightarrow \infty$ . From (30) and (38) we have:

$$\Delta R_{\text{CSIR}}^{\text{AF}} \leq \log \left( 1 + \frac{1}{\beta_{\text{fb}}} \right) \quad (44)$$

$$\Delta R_{\text{CSIR}}^{\text{DF}} \leq \log \left( 1 + \frac{\frac{P}{N_0}}{\left( 1 + \frac{P}{N_0} \right)^{\beta_{\text{fb}}}} \right) \quad (45)$$

If  $\beta_{\text{fb}} = 1$  then analog and error-free digital feedback both achieve essentially the same rate gap of 1 bit per channel user (per UT). However, if  $\beta_{\text{fb}} > 1$ , the rate gap for quantized feedback vanishes for  $\frac{P}{N_0} \rightarrow \infty$ . This conclusion finds an appealing interpretation in the context of rate-distortion theory. It is well-known (see for example [55] and references therein) that “analog transmission” (the source signal is input directly to the channel after suitable power scaling) is an optimal strategy to send an i.i.d. Gaussian source over a AWGN channel with the same bandwidth under quadratic distortion. In our case, the source vector is  $\mathbf{h}_k$  (Gaussian and i.i.d.) and the feedback channel is AWGN with SNR  $\frac{P}{N_0}$ . Hence, the fact that analog feedback cannot be essentially outperformed for  $\beta_{\text{fb}} = 1$  is expected. However, it is also well-known that if the channel bandwidth is larger than the source bandwidth (which corresponds to the case where a block of  $M$  source coefficients are transmitted over  $\beta_{\text{fb}}M$  channel uses with  $\beta_{\text{fb}} > 1$ ), then analog transmission is strictly suboptimal with respect to a digital scheme operating at the rate-distortion bound, because the distortion with analog transmission is  $O((P/N_0)^{-1})$  whereas it is  $O((P/N_0)^{-\beta_{\text{fb}}})$  for digital transmission.

This conclusion is confirmed by the numerical results shown in Figures 2 and 3. In Figure 2 the lower and genie-aided upper bounds are plotted for analog feedback, digital feedback without error, and digital feedback with error (QAM) versus SNR for an  $M = 4$  system with  $\beta_{\text{fb}} = 1$ . For digital feedback with error, the error detection bound in (43) is also included. The analog and error-free digital feedback schemes perform virtually identically and achieve a rate approximately 3 dB away from the perfect channel state information benchmark. Note also that the gap between the upper and lower bounds is not very large. For digital feedback with uncoded QAM<sup>6</sup>, however, there is a substantial gap between the upper and lower bounds; this gap and the performance with error detection is explained by Remark 4.3. In Figure 3 only the genie-aided upper bounds are plotted (because the lower and upper are nearly identical and thus are difficult to distinguish) for the same setting with  $\beta_{\text{fb}} = 2$ . We see that digital feedback with uncoded QAM outperforms analog feedback above approximately 5 dB, and that the rate with digital feedback (with or without errors) converges to the ideal rate as predicted earlier. This figure confirms that the effect of feedback vanishes when digital feedback is used, with or without errors, and  $\beta_{\text{fb}} > 1$ . Finally, in Figure 4 the bounds are plotted as a function of  $\beta_{\text{fb}}$  for fixed SNR  $\frac{P}{N_0} = 10$  dB and  $\frac{P}{N_0} = 20$  dB. When  $\beta_{\text{fb}} \approx 1$  analog and error-free digital feedback are nearly equivalent, but as  $\beta_{\text{fb}}$  is increased the rate with error-free digital quickly approaches the perfect channel state information rate. When feedback errors are introduced, digital feedback does eventually outperform analog and also approaches the ideal rate, but a larger  $\beta_{\text{fb}}$  is required. It is also worth noticing that as the SNR is increased, the value of  $\beta_{\text{fb}}$  at which digital (with or without errors) begins to outperform analog decreases toward 1: this is to expected based upon the fact that the effect of feedback vanishes as  $\frac{P}{N_0} \rightarrow \infty$  for any  $\beta_{\text{fb}} > 1$  for digital, whereas it does not for analog feedback.

It is worth noting that the same basic conclusion, i.e., that digital feedback (with or without errors) outperforms analog for sufficiently large  $\beta_{\text{fb}}$ , also holds in the presence imperfect CSIR. However, because imperfect CSIR

<sup>6</sup>These results are obtained by optimizing the value of  $1 \leq \alpha \leq \beta_{\text{fb}}$  for each SNR. We refer to this as “envelope”, that is, the plotted curve is the pointwise maximum of the rate vs. SNR curves for all  $\alpha$ .

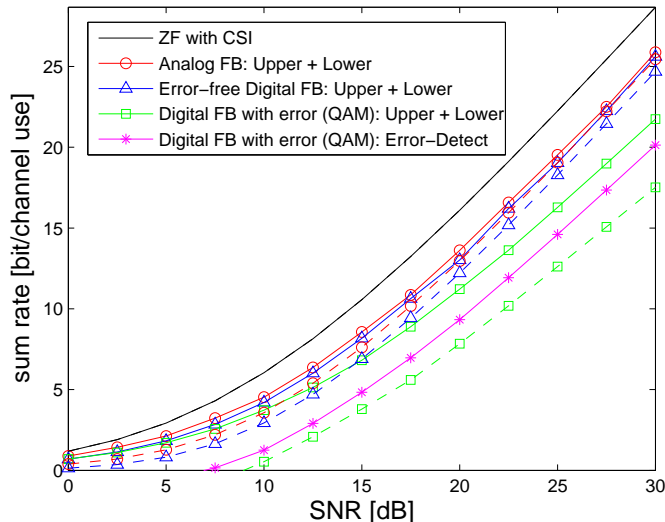


Fig. 2. Achievable rate lower (dotted lines) and upper (solid lines) bounds for analog, error-free digital, and QAM-based digital feedback for  $M = 4$  and  $\beta_{fb} = 1$ .

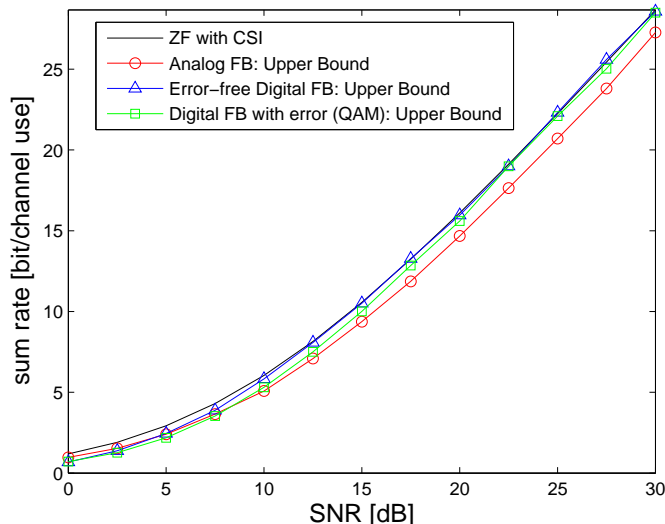


Fig. 3. Achievable rate upper bounds for analog, error-free digital, and QAM-based digital feedback for  $M = 4$  and  $\beta_{fb} = 2$ .

leads to a residual term in the rate gap expression that does not vanish (even for large  $\frac{P}{N_0}$ ), the absolute difference between digital and analog feedback is reduced.

## V. CHANNEL STATE FEEDBACK OVER THE MIMO-MAC

Orthogonal access in the feedback link requires  $O(M^2)$  channel uses for the feedback, while the downlink capacity scales at best as  $O(M)$ . When the number of antennas  $M$  grows large, such a system would not scale well with  $M$ . On the other hand, the inherent MIMO-MAC nature of the physical uplink channel suggests an alternative approach, where multiple UT's simultaneously transmit on the MIMO uplink (feedback) channel and the spatial dimension is exploited for channel state feedback too. This idea was considered for an FDD system in [28] and analyzed in terms of the mean square error of the channel estimate provided to the BS.

As in [28], we partition the  $M$  users into  $\frac{M}{L}$  groups of size  $L$ , and let UTs belonging to the same group transmit their feedback signal simultaneously, in the same time frame. Each UT transmits its  $M$  channel coefficients over  $\beta_{fb}M$  channel uses, with  $\beta_{fb} \geq 1$ . Therefore, each group uses  $\beta_{fb}M$  channel symbols and the total number of channel uses spent in the feedback is  $\beta_{fb}\frac{M^2}{L}$ . Choosing  $L \propto M$  (e.g.,  $L = M/2$ ) yields a total number of feedback

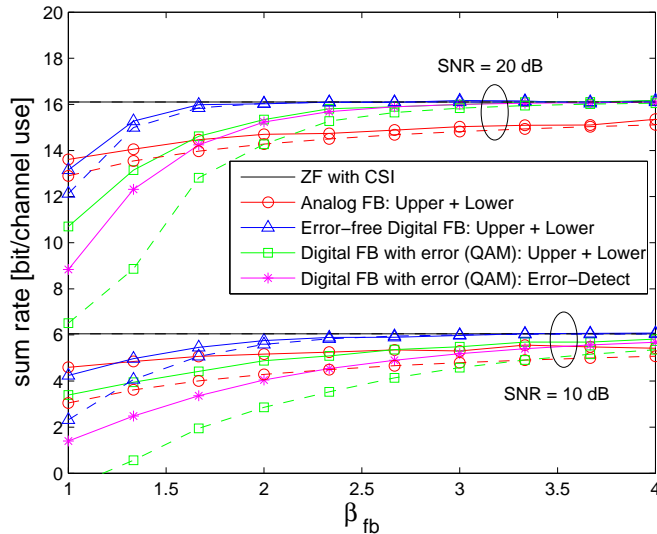


Fig. 4. Achievable rate lower (dotted lines) and upper (solid lines) bounds for analog, error-free digital, and QAM-based digital feedback for  $M = 4$  and  $\frac{P}{N_0} = 10$  dB and  $\frac{P}{N_0} = 20$  dB.

channel uses that grows linearly with  $M$ , such that the feedback resource converges to a fixed fraction of the downlink capacity. We assume that the uplink feedback channel is affected by i.i.d. block fading (i.e., has the same distribution as the downlink channel) and that there is no feedback delay.

With respect to the analysis provided in [28], the present work differs in a few important aspects: 1) we consider both analog and digital feedback; 2) although our analog feedback model is essentially identical to the FDD scheme of [28], we consider optimal MMSE estimation rather than Least-Squares estimation (zero-forcing pseudo-inverse); 3) we put out results in the context of the rate gap framework, that yields directly fundamental lower bounds on achievable rates, rather than in terms of channel state estimation error.

#### A. Analog Feedback

In an analog feedback scheme, each UT feeds back a scaled noisy version of its downlink channel, given by  $\frac{\sqrt{\beta_{\text{fb}}P}}{\sqrt{\beta_1 P + N_0}} \mathbf{s}_k$  where  $\mathbf{s}_k$  is the observation provided by the common training phase, defined in (4). Due to the symmetry of the problem, we can focus on the simultaneous transmission of a single group of  $L$  UTs. Let  $\mathbf{A} = [\mathbf{a}_1 \cdots \mathbf{a}_L] \in \mathbb{C}^{M \times L}$  denote the uplink fading matrix for this group of UTs (with i.i.d. entries,  $\sim \mathcal{CN}(0, 1)$ ) and let for  $k = 1, \dots, L$

$$b_{k,j} = \frac{\sqrt{\beta_{\text{fb}}P}}{\sqrt{\beta_1 P + N_0}} s_{k,j} = \frac{\sqrt{\beta_{\text{fb}}\beta_1 P}}{\sqrt{\beta_1 P + N_0}} h_{k,j} + \frac{\sqrt{\beta_{\text{fb}}P}}{\sqrt{\beta_1 P + N_0}} z_{k,j} \quad (46)$$

denote the transmitted symbol by UT  $k$  for its  $j$ -th channel coefficient, where  $s_{k,j}$  is the  $j$ -th component of  $\mathbf{s}_k$  and, from (4),  $z_{k,j}$  is the common training AWGN. For simplicity, we assume that the BS has perfect knowledge of the uplink channel state  $\mathbf{A}$ ; we later consider the more general case and see that the main conclusions are unchanged.

The  $M$ -dimensional received vector  $\mathbf{g}_j$ , upon which the BS estimates the  $j$ -th antenna downlink channel coefficients  $h_{1,j}, \dots, h_{L,j}$  of all users in the group, is given by:

$$\mathbf{g}_j = \sum_{i=1}^L \mathbf{a}_i b_{i,j} + \tilde{\mathbf{w}}_j = \mathbf{A} \mathbf{b}_j + \tilde{\mathbf{w}}_j \quad (47)$$

where  $\tilde{\mathbf{w}}_j$  is an AWGN vector with i.i.d. elements  $\sim \mathcal{CN}(0, N_0)$ . From the i.i.d. jointly Gaussian statistics of the channel coefficients, downlink noise and uplink (feedback noise) it is immediate to obtain the MMSE estimator for the downlink channel coefficient  $h_{k,j}$  in the form

$$\hat{h}_{k,j} = c \mathbf{a}_k^H \left[ \beta_{\text{fb}} P \mathbf{A} \mathbf{A}^H + N_0 \mathbf{I} \right]^{-1} \mathbf{g}_j \quad (48)$$

where we define the constant  $c = \frac{\sqrt{\beta_{\text{fb}}\beta_1 P}}{\sqrt{\beta_1 P + N_0}}$ . The corresponding MMSE, for given feedback channel matrix  $\mathbf{A}$ , is given by

$$\sigma_k^2(\mathbf{A}) = 1 - c^2 \mathbf{a}_k^H \left[ \beta_{\text{fb}} P \mathbf{A} \mathbf{A}^H + N_0 \mathbf{I} \right]^{-1} \mathbf{a}_k \quad (49)$$

*Theorem 7:* If each UT feeds back its channel coefficients in analog fashion over the MIMO MAC uplink channel, with groups of  $L$  users simultaneously feeding back and  $\beta_{\text{fb}} M$  channel uses per group, the rate gap upper bound is given by:

$$\overline{\Delta R}_{\text{MIMO-MAC}}^{\text{AF}} = \log \left( 1 + \frac{1}{M} \frac{\frac{P}{N_0}}{1 + \beta_2 \frac{N_0}{P}} + \frac{M-1}{M} \frac{P}{N_0} \left[ \frac{1}{1 + \frac{\beta_1 P}{N_0}} + \frac{\frac{\beta_1 P}{N_0}}{1 + \frac{\beta_1 P}{N_0}} \text{mmse} \left( \frac{\beta_{\text{fb}} P}{N_0} \right) \right] \right) \quad (50)$$

where we define the average channel state information estimation MMSE as

$$\text{mmse}(\rho) \triangleq \frac{1}{L} \sum_{k=1}^L \mathbb{E} \left[ \frac{1}{1 + \rho \lambda_k} \right] \quad (51)$$

and where  $\{\lambda_1, \dots, \lambda_L\}$  denote the eigenvalues of the  $L \times L$  central Wishart matrix  $\mathbf{A}^H \mathbf{A}$ .

Furthermore, if  $L < M$  the rate gap is bounded and converges at high SNR to the constant

$$\lim_{P/N_0 \rightarrow \infty} \overline{\Delta R}_{\text{MIMO-MAC}}^{\text{AF}} = \log \left( 1 + \frac{1}{\beta_2 M} + \frac{M-1}{M} \left( \frac{1}{\beta_1} + \frac{1}{\beta_{\text{fb}}(M-L)} \right) \right). \quad (52)$$

**Proof:** See Appendix VI.  $\square$

Comparing this expression to the rate gap for analog feedback over an AWGN channel (30), we notice that an SNR (array) gain of  $M-L$  is achieved (on the feedback channel) when the feedback is performed over the MIMO MAC because the feedback (of  $L$  users) is received over  $M$  antennas.<sup>7</sup> In addition, a factor of  $L$  fewer feedback symbols are required when the feedback is performed over the MIMO MAC ( $\beta_{\text{fb}} \frac{M^2}{L}$  vs.  $\beta_{\text{fb}} M^2$ ). On the other hand, using the second line of the RHS of (89) in Appendix III it is immediate to show that for  $L = M$  the rate gap upper bound grows unbounded as  $\log \log \left( \frac{P}{N_0} \right)$ .

From (52) we can optimize the value of  $L$  (assuming  $L < M$ ) for a fixed number of feedback channel uses, which we denote by  $aM$  for some  $a \geq 2$  (if  $L < M$  there must be at least two groups and thus we must have at least  $2M$  feedback symbols). By letting  $aM = \beta_{\text{fb}} \frac{M^2}{L}$ , we obtain  $\beta_{\text{fb}} = a \frac{L}{M}$ . Using this in (52), we have that minimizing the rate gap bound is equivalent to maximizing the term  $L(M-L)$  for fixed  $M$  and  $L < M$ . Therefore, the optimal group size is given by  $L^* = \frac{M}{2}$ . Substituting this value in (52) yields

$$\log \left( 1 + \frac{1}{\beta_2 M} + \frac{2(M-1)}{M^2 \beta_{\text{fb}}} + \frac{M-1}{M} \frac{1}{\beta_1} \right) \quad (53)$$

and the corresponding total number of feedback symbols is  $2\beta_{\text{fb}} M$ . Interestingly, we notice that in the regime of large  $M$  the term that dominates the optimized rate gap bound (53) corresponds to the downlink common training phase. In fact, the terms corresponding to dedicated training and feedback vanish as  $M$  increases.

When the total number of feedback symbols is larger or equal to  $2M$  (i.e.,  $a \geq 2$ ) numerical results verify that also at finite SNR the choice  $L^* = \frac{M}{2}$  yields the best performance both in terms of the achievable rate lower bound and of the genie-aided upper bound. Hence, the optimal MIMO-MAC feedback strategy is a combination of TDMA and SDMA. In contrast, when total number of feedback symbols is strictly smaller than  $2M$  (i.e.,  $1 \leq a < 2$ ), choosing  $L = M$  with  $\beta_{\text{fb}} = a$  is the only option. Although this choice yields an unbounded rate gap, it does provide reasonable performance at finite SNR's.

A legitimate question at this point is the following: is the condition  $L < M$  a fundamental limit of the MIMO-MAC analog feedback in order to achieve a bounded rate gap, or is it due to the looseness of Theorem 2? In order to address this question, we examine the genie-aided rate upper bound of Theorem 3 and obtain the following rate upper bound:

<sup>7</sup>At high SNR the feedback from a particular UT is effectively received over an interference-free  $1 \times (M-L+1)$  channel because  $L-1$  interfering signals are nulled. However, this results in only a  $M-L$  multiplicative gain because  $\mathbb{E}[1/\chi_{2k}^2] = 1/(k-1)$ .

*Theorem 8:* When a group of  $L = M$  UTs feed back the channel coefficients simultaneously over  $\beta_{\text{fb}}M$  channels uses of the fading MIMO-MAC, the difference between  $R_k^{\text{ZF}}$  and the genie-aided upper bound of Theorem 3 is uniformly bounded for all SNRs.

**Proof** See Appendix VII.  $\square$

Theorem 8 suggests that if the UTs are able to obtain an estimate of their *instantaneous* residual interference level in each frame, up to  $M$  UTs can feedback their channel state information at the same time. The ability of estimating the interference coefficients  $\mathcal{A}_k$  (see (8) and the comment following Theorem 3) depends critically on the quality of the dedicated training. Hence, the dedicated training has a direct impact on the design and efficiency of the channel state feedback. Such inter-dependencies between the different system components can be illuminated thanks to the comprehensive system analysis carried out in this work and are missed by making overly simplifying assumptions (e.g., genie-aided coherent detection with perfect knowledge of the coefficients  $\mathcal{A}_k$ ).

**Remark 5.1:** In [28], the same model in (46) for analog channel state feedback over the MIMO-MAC uplink channel is considered. Instead of the linear MMSE estimator considered here, a zero-forcing approach (via the pseudo-inverse of the matrix  $\mathbf{A}$ ) is examined. In the case of  $L = M$  this yields an infinite error variance, which does not make sense in light of the fact that each channel coefficient has unity variance. This odd behavior can be avoided by performing an additional component-wise MMSE step. As a matter of fact, performance very similar to what we have found for the full MMSE estimator can be obtained for  $L < M$  by using a zero-forcing receiver for the channel state feedback, followed by individual (componentwise) MMSE scaling. We omit the analysis of such suboptimal scheme for the sake of brevity.  $\diamond$

**Remark 5.2:** It is also possible to analyze the more realistic scenario where the uplink channel matrix  $\mathbf{A}$  is known imperfectly at the BS. We consider the following simple training-based scheme: the  $L$  UTs within a feedback group transmit a preamble of  $\beta_{\text{up}}L$  training symbols, where  $\beta_{\text{up}} \geq 1$  defines the uplink training length (per UT). Without repeating all steps in the details, the uplink channel  $\mathbf{A}$  admits the following decomposition:

$$\mathbf{A} = \hat{\mathbf{A}} + \tilde{\mathbf{A}} \quad (54)$$

where the channel estimate and estimation error ( $\hat{\mathbf{A}}, \tilde{\mathbf{A}}$ ) are jointly jointly Gaussian and independent, with per-component variances  $1 - \sigma_{\text{up}}^2$  and  $\sigma_{\text{up}}^2$ , respectively, with  $\sigma_{\text{up}}^2 = \frac{1}{1 + \beta_{\text{up}} \frac{P}{N_0}}$ . Now, the MMSE estimation of the downlink channel coefficients  $h_{k,j}$  is conditional with respect to  $\hat{\mathbf{A}}$ . By repeating all previous steps, after a lengthy calculation that we do not report here for the sake of brevity, we obtain the average estimation error in the form

$$\mathbb{E}[\sigma_k^2(\hat{\mathbf{A}})] = \frac{1}{1 + \beta_1 \frac{P}{N_0}} + \frac{\beta_1 \frac{P}{N_0}}{1 + \beta_1 \frac{P}{N_0}} \text{mmse} \left( \frac{\beta_{\text{up}} \frac{P}{N_0}}{1 + \beta_{\text{up}} \frac{P}{N_0} + L\beta_{\text{fb}} \frac{P}{N_0}} \frac{\beta_{\text{fb}} P}{N_0} \right) \quad (55)$$

where  $\text{mmse}(\cdot)$  was defined in (51). By comparing (55) with (87), we notice that they differ only in the argument of the function  $\text{mmse}(\cdot)$ . The two expressions coincide for  $\beta_{\text{up}} \rightarrow \infty$ , consistent with the fact that  $\beta_{\text{up}} \rightarrow \infty$  corresponds to perfect estimation of the channel matrix  $\mathbf{A}$ . Furthermore, for large SNR, the two arguments differ by a constant multiplicative factor. Hence, apart from this constant factor that depends on the uplink training parameter  $\beta_{\text{up}}$ , the conclusions about the rate gap obtained for the case of perfect uplink channel knowledge also hold for the case of training-based uplink channel estimation.  $\diamond$

## B. Digital Feedback

In the case of digital feedback, we let  $L \leq M$  UTs multiplex their channel state feedback codewords at the same time. The resulting MIMO-MAC channel model is again given by (47), but now the vector  $\mathbf{b}_j$  contains the  $j$ -th symbols of the feedback codewords of the  $L$  UTs sharing the same feedback frame. As in Section IV-B, we assume that feedback messages of  $\alpha(M-1) \log_2 \frac{P}{N_0}$  bits are sent in  $\beta_{\text{fb}}(M-1)$  channel uses. Hence, the feedback symbols transmitted by the  $L$  UT's can be grouped in a  $L \times \beta_{\text{fb}}(M-1)$  matrix, while the BS has an  $M \times \beta_{\text{fb}}(M-1)$  observation upon which to estimate the transmitted symbols. We again assume each feedback symbol has average energy  $P$ .

Suppose that the BS receiver operates optimally, by using a joint ML decoder for all the simultaneously transmitting UTs. The high-SNR error probability performance of the MIMO-MAC channel was characterized in terms of the diversity-multiplexing tradeoff in [56]. In particular, when each user transmits at rate  $r \log_2 \frac{P}{N_0}$

bits/symbol (i.e., with *multiplexing gain*  $r$ ) over the MIMO-MAC with i.i.d. channel fading (as considered here), the optimal ML decoder achieves an individual user average error probability

$$P_{e,\text{fb}} \doteq \left( \frac{P}{N_0} \right)^{-d^*(r)}$$

where the ‘‘dot-equality’’ notation, introduced in [57], [56], indicates that  $\lim_{P/N_0 \rightarrow \infty} \frac{-\log P_{e,\text{fb}}}{\log P/N_0} = d^*(r)$ . The error probability SNR exponent  $d^*(r)$  is referred to as the optimal *diversity gain* of the system. Particularizing the results of [56] to the case of  $L \leq M$  users with 1 antenna each, transmitting to a receiver with  $M$  antennas, the optimal diversity gain is given by

$$d^*(r) = \begin{cases} M(1-r) & \text{for } 0 \leq r \leq 1 \\ 0 & \text{otherwise} \end{cases} \quad (56)$$

This is the same exponent of a channel with a single user with a single antenna, transmitting to a receiver with  $M$  antennas (single-input multiple-output with receiver antenna diversity). In other words, under our system parameters, each UT achieves an error probability that decays with SNR as if TDMA on the feedback link was used (as if the UT transmitted its feedback message alone on the MIMO uplink channel). From what is said above, it follows that the multiplexing gain of all UTs is given by  $r = \frac{\alpha}{\beta_{\text{fb}}}$ . Furthermore, from the derivation of Section IV-C, we require that  $1 < \alpha < \beta_{\text{fb}}$ . It follows that the average feedback error message probability in the MIMO-MAC fading channel is given by

$$P_{e,\text{fb}} = \left( \frac{P}{N_0} \right)^{-M(1-\alpha/\beta_{\text{fb}})} \times g \left( \frac{P}{N_0} \right) \quad (57)$$

where  $g(x)$  is some sub-polynomial function, such that  $\lim_{x \rightarrow \infty} x^{-\epsilon} g(x) = 0$  for all fixed  $\epsilon > 0$ .

If we examine the rate-gap expression with digital feedback (42), we see that in order to achieve a bounded rate gap the error probability  $P_{e,\text{fb}}$  must go to zero at least as fast as  $\left( \frac{P}{N_0} \right)^{-1}$ . From (57) we have that for all  $1 < \alpha < \beta_{\text{fb}}$  such that  $M(1 - \alpha/\beta_{\text{fb}})$  is strictly larger than 1, the resulting rate gap is bounded and the effect of feedback errors vanishes. This imposes the condition  $\beta_{\text{fb}} > \frac{M}{M-1}$  and  $\alpha < \frac{M-1}{M} \beta_{\text{fb}}$ , which is stricter than the condition  $\beta_{\text{fb}} > 1$  and  $\alpha < \beta_{\text{fb}}$  needed in the case of TDMA an unfaded feedback channel previously analyzed in Section IV-C.

We conclude that a bounded rate gap can also be achieved with digital feedback on the MIMO-MAC uplink channel. Therefore, also in this case we can achieve a number of feedback channel uses that scales linearly with the number of the BS antennas  $M$ . Explicit design of codes that achieve the optimal diversity-multiplexing tradeoff of MIMO-MAC channels is not an easy task in general. In the particular case of  $M$  users with one antenna each, simple explicit constructions of MIMO-MAC codes for the digital channel state feedback are presented [53]. These codes can be optimally decoded by using a Sphere Decoder [58], [59] and achieve the performance promised by the above analysis. It should be noticed, however, that while in the AWGN case the term  $\frac{P}{N_0} P_{e,\text{fb}}$  in the rate gap expression vanishes rapidly (faster than polynomially, in  $P/N_0$ ), in the MIMO-MAC fading case it vanishes only as  $(P/N_0)^{1-M(1-\alpha/\beta_{\text{fb}})}$ . Thus, for finite SNR the rate gap may be significantly larger than in the case of unfaded feedback channel and the optimal tradeoff between quantization distortion and the feedback error probability must be sought by careful optimization of the parameters  $\alpha$  and  $\beta_{\text{fb}}$  (see details in [60]). Also, the same observations about detecting feedback errors at the UTs and discarding the corresponding frames made at the end of Section IV-C apply here.

### C. Numerical example

Fig. 5 shows both the genie-aided upper bound of Theorem 3 and the lower bound based on (50) of analog feedback over a fading MIMO-MAC for  $M = 4$  and  $L = 2, 4$ . We assume perfect CSIR. We notice that for  $L = 2$ , the lower bound coincides with the genie-aided upper bound and comes very close to the performance of ZF with ideal CSIT. For  $L = M$ , the rate gap of the lower bound (50) is unbounded but the double logarithmic growth  $\log \log(P/N_0)$  yields a very small gap for a wide range of practical SNRs. The genie-aided bound achieves a constant rate gap even for  $L = M$ , in accordance with Theorem 8. Although not shown here, a system using  $M = 4, L = 2$  and  $\beta_{\text{fb}} = 1$  does outperform  $M = L = 4, \beta_{\text{fb}} = 2$  (both configurations use a total of 8 feedback

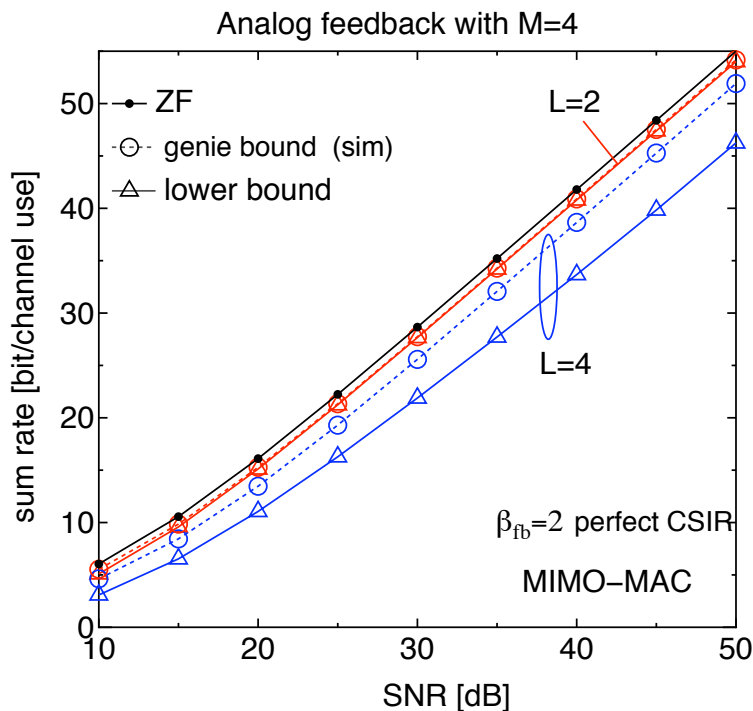


Fig. 5. Impact of  $L$  with analog feedback over MIMO-MAC

symbols per frame) in terms of the lower bound and the genie-aided upper bound throughout the SNR range shown; this validates our earlier claim about the optimality of  $L = \frac{M}{2}$  whenever at least  $2M$  feedback symbols are used.

Fig. 6 compares the achievable rates of analog and digital feedback schemes based on the rate gap (50), (42), over a fading MIMO-MAC for  $M = 4$ . For the digital feedback we assume that there exist some code achieving the outage probability (57) with  $g(P/N_0) = 1$ . We compare both schemes for the same total amount of the feedback symbols (24 symbols). For the analog feedback we choose  $L = 2, \beta_{fb} = 3$ , while for the digital feedback we let  $L = 4, \beta_{fb} = 8, \alpha = 4$ . We observe that the digital feedback achieves near-optimal sum rate over the all SNR ranges while the analog feedback achieves a constant gap of roughly 0.7 bit/channel use. Surprisingly, the digital feedback is able to let  $M$  users transmit simultaneously while vanishing both the quantization error and the feedback error.

## VI. EFFECTS OF CSIT FEEDBACK DELAY

In this section we wish to take into account the effect of feedback delay in a setting where the fading is temporally correlated. We assume that the fading is constant within each frame, but changes from frame to frame according to a stationary random process. In particular, assuming spatial independence, each entry of  $\mathbf{h}_k$  evolves independently according to the same complex circularly symmetric Gaussian stationary ergodic random process, denoted by  $\{h(t)\}$ , with mean zero, unit variance and power spectral density (Doppler spectrum) denoted by  $S_h(\xi)$ ,  $\xi \in [-1/2, 1/2]$ , and satisfying  $\int_{-1/2}^{1/2} S_h(\xi) d\xi = 1$ . Notice that the discrete-time process  $\{h(t)\}$  has time that ticks at the frame rate.

Because of symmetry and spatial independence, we can neglect the UT index  $k$  and the antenna index and consider scalar rather than vector processes. Generalizing (4), the observation available at each UT at time  $t - d$  from the common training phase takes on the form

$$\left\{ s(t - \tau) = \sqrt{\beta_1 P} h(t - \tau) + z(t - \tau) : \tau = d, d + 1, d + 2, \dots, \infty \right\} \quad (58)$$

where  $d$  indicates the feedback delay in frames. This means that the channel state feedback to be used by the BS at frame time  $t$  is formed from noisy observations of the channel up to time  $t - d$ . We consider a scheme where each UT at frame  $t - d$  produces the MMSE estimate of its channel at frame  $t$  and sends this estimate (using either analog or digital feedback) to the BS; the BS uses the received feedback to choose the beamforming vectors used for data transmission in frame  $t$ .

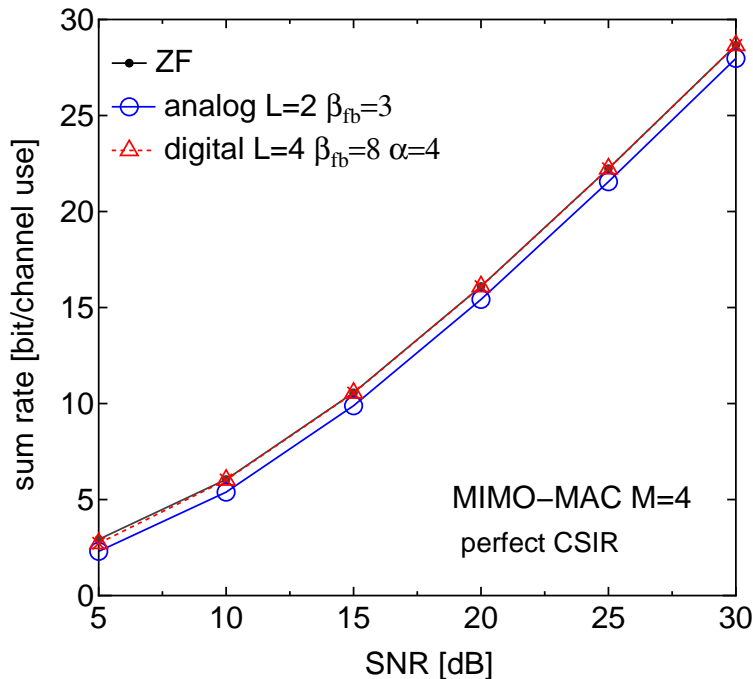


Fig. 6. Achievable rate lower bounds for analog and digital feedback for  $M = 4$  and 24 total feedback symbols.

#### A. Estimation Error at UT

The key quantity in the associated rate gap is the MMSE prediction error at the UT. Let  $\tilde{h}(t)$  denote the MMSE estimate of  $h(t)$  given the observations in (58). Given the joint Gaussianity of  $h$  and  $s$ , we can write

$$h(t) = \tilde{h}(t) + n(t) \quad (59)$$

where  $\mathbb{E}[|n(t)|^2] = \sigma_1^2$  is the estimation MMSE, and  $\tilde{h}(t)$  and  $n(t)$  are independent with  $\mathbb{E}[|\tilde{h}(t)|^2] = 1 - \sigma_1^2$ . From classical Wiener filtering theory [46], the one-step prediction ( $d = 1$ ) MMSE error is given by

$$\epsilon_1(\delta) = \exp\left(\int_{-1/2}^{1/2} \log(\delta + S_h(\xi)) d\xi\right) - \delta \quad (60)$$

where  $\delta = N_0/(\beta_1 P)$  is the observation noise variance. The filtering MMSE ( $d = 0$ ) is related to  $\epsilon_1(\delta)$  as

$$\epsilon_0(\delta) = \frac{\delta \epsilon_1(\delta)}{\delta + \epsilon_1(\delta)}. \quad (61)$$

The scenario considered in all previous sections corresponds to i.i.d. fading (across blocks) and  $d = 0$ , in which case  $\epsilon_1(\delta) = 1$  (past observations are useless) and thus  $\sigma_1^2 = \epsilon_0(\delta) = (1 + \beta_1 \frac{P}{N_0})^{-1}$ , which coincides with (7). More in general, in this section we shall consider <sup>8</sup>  $\sigma_1^2 = \epsilon_d(\delta)$  for  $d = 0, 1$ .

We distinguish two cases of channel fading statistics: *Doppler process* and *regular process*:

- We say that  $\{h(t)\}$  is a Doppler process if  $S_h(\xi)$  is strictly band-limited to  $[-F, F]$ , where  $F < 1/2$  is the maximum Doppler frequency shift, given by  $F = \frac{v f_c}{c} T_f$ , where  $v$  is the mobile terminal speed (m/s),  $f_c$  is the carrier frequency (Hz),  $c$  is light speed (m/s) and  $T_f$  is the frame duration (s) [40]. A Doppler process satisfies  $\int_{-F}^F \log S_h(\xi) d\xi > -\infty$ , and has prediction error<sup>9</sup>

$$\epsilon_1(\delta) = \delta^{1-2F} \exp\left(\int_{-F}^F \log(\delta + S_h(\xi)) d\xi\right) - \delta \quad (62)$$

<sup>8</sup>We focus on the case  $d = 1$ , because it is very relevant in practical applications. For example, high-data rate downlink systems such as 1xEV-Do [61] already implement a very fast channel state feedback with at most one frame delay. Furthermore, the one-step prediction case allows an elegant closed-form analysis.

<sup>9</sup>As in [39], the same result holds for a wider class of processes such that the Lebesgue measure of the set  $\{\xi \in [-1/2, 1/2] : S_h(\xi) = 0\}$  is equal to  $1 - 2F$ , and such that  $\int_{\mathcal{D}} \log(S_h(\xi)) d\xi > -\infty$  where  $\mathcal{D}$  is the support of  $S_h(\xi)$ .

Therefore,  $\lim_{\delta \rightarrow 0} \epsilon_1(\delta) = 0$ .

- We say that  $\{h(t)\}$  is a regular process if  $\epsilon_1(0) > 0$  (see [39] and references therein). In particular, a process satisfying the Paley-Wiener condition [46]  $\int_{-1/2}^{1/2} \log S_h(\xi) d\xi > -\infty$  is regular.

For the case of no delay ( $d = 0$ ), for either type of process the estimation error goes to zero with the observation noise, i.e.,  $\epsilon_0(\delta) \rightarrow 0$  as  $\delta \rightarrow 0$ . However, they differ sharply in terms of prediction error:  $\epsilon_1(\delta)$  is strictly positive for a regular process (even as  $\delta \rightarrow 0$ ), whereas  $\epsilon_1(\delta) \rightarrow 0$  for Doppler processes as quantified in the following:

*Lemma 1:* The noisy prediction error of a Doppler process satisfies

$$\epsilon_1(\delta) = \kappa \delta^{1-2F} + O(\delta) \quad (63)$$

for  $\delta \downarrow 0$ , where  $\kappa$  is a constant term independent of  $\delta$ .

**Proof:** Applying Jensen's inequality to (62) from the fact that  $\int S_h(\xi) d\xi = 1$ , we obtain the upper bound

$$\epsilon_1(\delta) \leq \delta^{1-2F} \left[ \left( \frac{1}{2F} + \delta \right)^{2F} - \delta^{2F} \right] \quad (64)$$

Using the fact that  $\log$  is increasing, we arrive at the lower bound

$$\epsilon_1(\delta) \geq \delta^{1-2F} \left[ \exp \left( \int_{-F}^F \log S_h(\xi) d\xi \right) - \delta^{2F} \right] \quad (65)$$

Combining these bounds we obtain the result.  $\square$

### B. Rate Gap Upper Bound

When analog feedback is used, each UT transmits a scaled version of its MMSE estimate  $\tilde{h}(t)$  over the feedback channel. The only difference from the scenarios studied in Sections V-A (AWGN feedback channel) and IV-A (MIMO MAC feedback channel) is that the estimation error at the UT is  $\epsilon_d(N_0/(\beta_1 P))$  rather than  $(1 + \beta_1 P/N_0)^{-1}$ . As a result, a simple calculation confirms that the expressions for the rate gap upper bound given in Theorems 4 (AWGN) and 7 (MIMO-MAC) apply to the present if  $\epsilon_d(N_0/(\beta_1 P))$  is substituted for  $(1 + \beta_1 P/N_0)^{-1}$ . The same equivalence holds for digital feedback: each UT quantizes its MMSE estimate  $\tilde{h}(t)$ , and as a result the rate gap upper bound given in Theorem 5 applies with the same substitution. For the sake of brevity, the expressions for the rate gap upper bound are not provided here.

In fact, the effect of feedback delay is most clearly illustrated by considering perfect feedback (i.e.,  $\beta_{fb} \rightarrow \infty$ ), in which case (at frame  $t$ ) the BS has perfect knowledge of  $\tilde{h}(t)$ , the UT's prediction of  $h(t)$  based on common training observations up to frame  $t - d$ . For the sake of simplicity we further assume perfect dedicated training (i.e.,  $\beta_2 \rightarrow \infty$ ), in which case the rate gap upper bound is

$$\overline{\Delta R}^{\text{PREDICT}} = \log \left( 1 + \frac{P}{N_0} \frac{M-1}{M} \epsilon_d \left( \frac{N_0}{\beta_1 P} \right) \right). \quad (66)$$

We now analyze the cases of no delay and one-step delay for both types of processes.

*a) No feedback delay ( $d = 0$ ):* Because using past observations can only help, the filtering error is no larger than the error if the past is ignored, i.e.,  $\epsilon_0(\delta) \leq (1 + \beta_1 P/N_0)^{-1}$ . It thus follows that for both Doppler and regular processes the rate gap is bounded. Based upon (61), Lemma 1, and the property  $\epsilon_0(0) > 0$  for regular processes, it is straightforward to see that  $(P/N_0)\epsilon_0(N_0/(\beta_1 P)) \rightarrow \frac{1}{\beta_1}$  as  $\frac{P}{N_0} \rightarrow \infty$  for either regular or Doppler processes. As a result, the rate gap upper bound in (66) converges to  $\log(1 + 1/\beta_1)$  at high SNR. This matches the high SNR expression for block-by-block estimation in (30), showing that filtering does not provide a significant advantage at asymptotically high SNR. However, as later illustrated through numerical results, this convergence occurs extremely slowly for Doppler processes or highly correlated regular processes, in which case filtering does provide a non-negligible gain over a wide range of SNR's.

*b) Feedback delay ( $d = 1$ ):* For regular fading process, since  $\epsilon_1(0) > 0$ , the quantity  $(P/N_0)\epsilon_1(N_0/(\beta_1 P))$  increases linearly with  $\frac{P}{N_0}$  and thus the rate gap upper bound  $\Delta \bar{R}^{\text{PREDICT}}$  grows like  $\log \frac{P}{N_0}$ . As a result, the achievable rate lower bound  $R_k^{\text{ZF}} - \Delta \bar{R}^{\text{PREDICT}}$  is bounded even as  $P/N_0 \rightarrow \infty$ . In addition, in Appendix VIII we show that the genie-aided upper bound is also bounded due to the fundamentally non-deterministic nature of regular processes. This shows that with delayed feedback and a channel that evolves according to a regular fading process, a system that makes use of zero-forcing naive beamforming to  $M$  users becomes interference limited.<sup>10</sup> This behavior holds even with CSIR (i.e., letting  $\beta_1 \rightarrow \infty$ ).

Fortunately, physically meaningful fading processes belong to the class of Doppler processes, at least over a time-span where they can be considered stationary. For a practical relative speed between BS and UT, such time span is much larger than any reasonable coding block length. Hence, we may say that Doppler processes are more the rule than the exception. In this case, the system behavior is radically different. Using Lemma 1, at high SNR the rate gap upper bound is

$$\log \left( 1 + \frac{M-1}{M} \frac{P}{N_0} \left( \kappa \left( \frac{\beta_1 P}{N_0} \right)^{2F-1} + O \left( \frac{N_0}{\beta_1 P} \right) \right) \right), \quad (67)$$

and thus the rate gap grows like  $2F \log \frac{P}{N_0}$ . Using this in the rate lower bound of Corollary (3.1), and considering the pre-log factor in high-SNR, we have that the system sum-rate is lowerbounded by

$$\sum_{k=1}^M R_k \geq M(1-2F) \log \frac{P}{N_0} + O(1) \quad (68)$$

This shows that a multiplexing gain of  $M(1-2F)$  is achievable.

**Remark 6.1:** If perfect CSIR is assumed, an interesting singularity is observed for Doppler processes. Under this assumption each UT is able to perform *perfect prediction* of its channel state on the basis of its past noiseless observations of the channel, by the definition of a Doppler process. Thus, it is as if there is no delay and the full multiplexing gain of  $M$  is achieved (even if the feedback link is imperfect). On the other hand, if perfect CSIR is not assumed and UT's learn their channel through  $\beta_1 M$  common training symbols, for any finite value of  $\beta_1$  a multiplexing gain of only  $M(1-2F)$  is achieved. This point illustrates, again, that neglecting some system aspects may yield to erroneous conclusions. In this case, by properly modeling imperfect CSIR we have illuminated the impact of the UTs speed (which determines the channel Doppler bandwidth) on the system achievable rates in a concise and elegant way.  $\diamond$

**Remark 6.2:** It is interesting to notice here the parallel with the results of [39] on the high-SNR capacity of the single-user scalar ergodic stationary fading channel with no CSIR and no CSIT, where it is shown that for a class of *non-regular* processes that includes the Doppler processes defined here, the high-SNR capacity grows like  $\mathcal{L} \log(P/N_0)$ , where  $\mathcal{L}$  is the Lebesgue measure of the set  $\{\xi \in [-1/2, 1/2] : S_h(\xi) = 0\}$ . In our case, it is clear that  $\mathcal{L} = 1 - 2F$ . These results, as ours, rely on the behavior of the noisy prediction error  $\epsilon_1(\delta)$  for small  $\delta$ .  $\diamond$

### C. Examples

We now present numerical results for the Jake's model and the Gauss-Markov model, which are two widely used Doppler and regular processes, respectively. The classical Jakes' correlation model has the following spectrum [62], [54]

$$S_h(\xi) = \frac{1}{\pi \sqrt{F^2 - \xi^2}}, \quad -F \leq \xi \leq F, \quad (69)$$

and auto-correlation function  $J_0(2\pi F\tau)$ . No closed-form solution is known for the prediction or filtering error. Under the Gauss-Markov model (i.e., auto regressive of order 1) the channel evolves in time as:

$$h(\tau) = rh(\tau-1) + \sqrt{1-r^2}\Delta(\tau) \quad (70)$$

<sup>10</sup>In order to have a non-interference limited system we can always use TDMA and serve one user at a time. However, in this case the sum-rate would grow like  $\log(P/N_0)$  instead of  $M \log(P/N_0)$  as promised by the MIMO downlink with perfect CSIT.

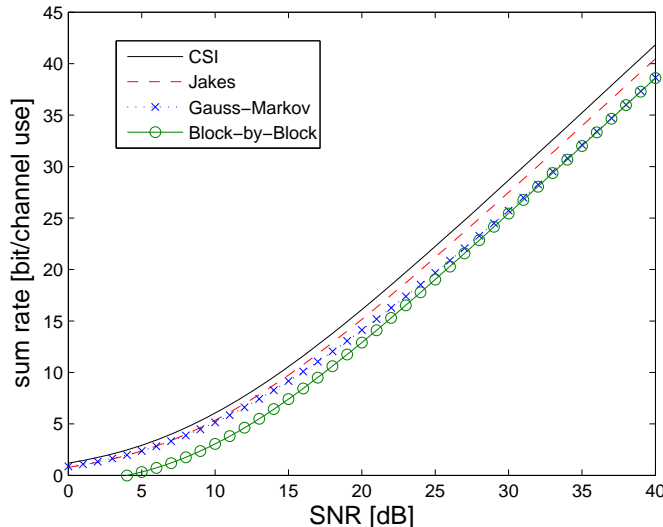


Fig. 7. Achievable rate lower bounds with optimal filtering for the Jake's and Gauss-Markov models for  $M = 4$  and  $v = 10$  km/hr ( $F = 0.0185$  and  $r = 0.9966$ ). Also shown are the rates with perfect CSI and with block-by-block estimation.

where  $r$  is the correlation coefficient ( $0 < r < 1$ ) and the innovation process  $\Delta(\tau)$  is unit-variance complex Gaussian, i.i.d. in time. The prediction error for such model can be written in closed-form and is given by (see for example [32])

$$\epsilon_1(\delta) = (1 - r^2) \left[ 1 + \frac{-(1 + \delta) + \sqrt{1 + \delta^2 + 2\delta \frac{1+r^2}{1-r^2}}}{2} \right] \quad (71)$$

For the Jakes' model we have  $F = \frac{vf_c}{c}T_f$ . In all results we consider  $f_c = 2$  GHz and  $T_f = 1$  msec. Motivated by the maximum-entropy principle [63], several works in wireless communication modeled channel fading as Gauss-Markov process with one-step correlation coefficient  $r = J_0(2\pi F)$ , given by Jakes' model. Comparing the performance of the true Jakes' model with its Gauss-Markov maximum-entropy approximation, we will point out that the latter may be overly pessimistic for high-speed mobile terminals.

In Fig. 7 the achievable rate lower bound with delay-free feedback ( $d = 0$ ) and optimal filtering is plotted versus SNR for the Jakes and Gauss-Markov models, for  $M = 4$ ,  $v = 10$  km/hr ( $F = 0.0185$  and  $r = 0.9966$ ), and  $\beta_1 = 1$ . Filtering is seen to provide an advantage with respect to block-by-block estimation for a wide range of SNR's. For the Gauss-Markov model this advantage vanishes around 30 dB, whereas for Jakes' model this advantage persists far beyond the range of this plot.

Using the same parameters, in Fig. 8 we plot the lower bound for one-step prediction ( $d = 1$ ) versus SNR for  $v = 3$  and 10 km/hr ( $F = 0.0056$  and  $F = 0.0185$ ). This figure illustrates the contrast between Doppler and regular processes: for Jakes' model the achieved rate is quite close to the perfect channel state information rate (although a slight loss in multiplexing gain is evident), whereas the rate for the Gauss-Markov model saturates at sufficiently high SNR due to the unpredictability inherent to the model. To further emphasize the difference in behavior, in Fig. 9 we plot the lower bound for one-step prediction ( $d = 1$ ) versus  $\beta_1$ , the number of common training symbols per block, for  $P/N_0 = 10$  and 15 dB and  $v = 10$  km/hr. As  $\beta_1$  increases (and thus the observation noise decreases) the rate for Jakes' model converges to the ideal case. On the other hand, the rate for the Gauss-Markov model saturates at a rate strictly smaller than the ideal channel state information rate because there is strictly positive prediction error even if noiseless past observations (i.e.,  $\beta_1 \rightarrow \infty$ ) are provided.

In conclusion, the most noteworthy result of this analysis is that under common fading models (Doppler processes), both analog and digital feedback scheme achieves a potentially high multiplexing gain even with realistic, noisy and delayed feedback.

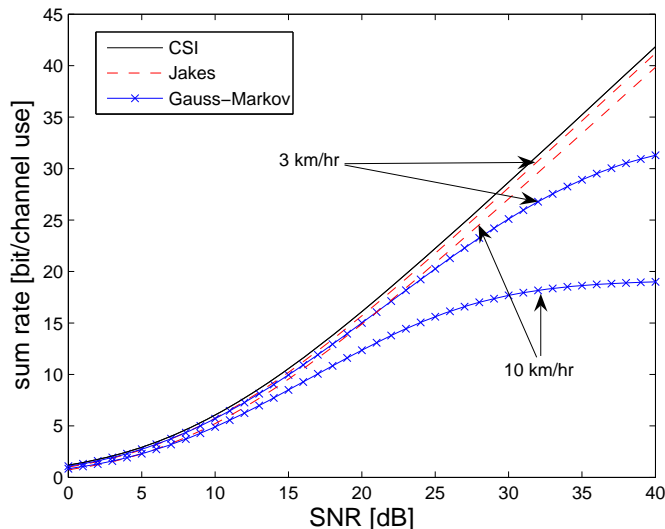


Fig. 8. Achievable rate lower bounds with optimal one-step prediction for the Jake's and Gauss-Markov models for  $M = 4$ .

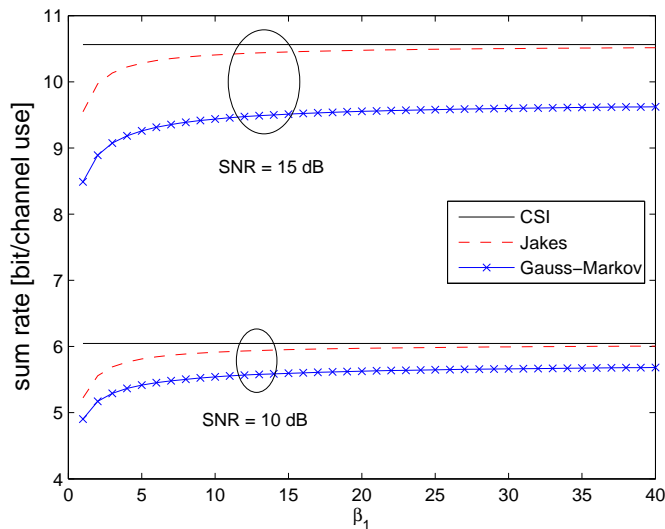


Fig. 9. Achievable rate lower bounds with optimal one-step prediction versus  $\beta_1$  for  $M = 4$  and  $v = 10$  km/hr.

## VII. CONCLUSIONS

This paper presents a comprehensive and rigorous analysis of the achievable performance of ZF beamforming under pilot-based channel estimation and explicit channel state feedback. We considered what we believe are the most relevant system aspects. In particular, the often neglected effect of explicit channel estimation at the UTs is taken into account, including both common training and dedicated training phases. As for the feedback, our closed-form bounds allow for a detailed comparison of analog and digital feedback schemes, including the effects of the MIMO-MAC fading channel, of digital feedback decoding errors, and of feedback delay.

Our results build on prior work, but generalize many results and models. We have focused on the case of FDD, but our results easily extend to TDD systems with channel reciprocity. It is perhaps important to point out here that our results show that, even in the case of FDD, a system with explicit CSIT feedback can be implemented, where the number of training and feedback channel uses scales linearly with the number of BS antennas, and eventually with the downlink throughput.

The throughput of the system analyzed here can be improved via the use of combined beamforming and user selection/scheduling. Simulation results show that a system with  $K = 10$  and  $M = 4$ , with a greedy scheduling as proposed in [15], [32], achieves a very small gap with respect to the optimal dirty-paper coding and perfect

CSIT case with the same parameters. Although a clean closed-form analytical characterization of a system with beamforming and user selection based on imperfect channel state information appears to be difficult, recent results [33], [45] indicate that the dependence on CSIT quality when user selection is performed is roughly the same as the equal-power/no selection scenario analyzed here.

We would like to conclude by noticing that some practically relevant extensions of the present work have been presented (by the same authors and by others) since the submission of this paper. In particular, the rate gap analysis was extended to the very relevant case of MIMO OFDM with frequency-correlated fading in [52], the optimal allocation of training and feedback resources is considered in [37], [38], explicit coding schemes for the CSIT digital feedback MIMO-MAC channel are presented in [53], and comparisons between single-user and multi-user MIMO (based on the bounds developed here and related approximations) are performed in [64].

## APPENDIX I PROOF OF THEOREM 1

The proof is closely inspired by that of Lemma B.0.1 of [36]. First, notice that since  $\hat{a}_{k,k}$  is a function of  $\mathcal{R}_k$ , by the data-processing inequality we have that

$$I(u_k; y_k, \mathcal{R}_k) \geq I(u_k; y_k, \hat{a}_{k,k})$$

Then, because  $I(u_k; y_k, \hat{a}_{k,k}) = h(u_k) - h(u_k|y_k, \hat{a}_{k,k})$  and  $h(u_k) = \log(\pi e \frac{P}{M})$ , a lower bound on mutual information is derived by upper bounding  $h(u_k|y_k, \hat{a}_{k,k})$  as follows:

$$\begin{aligned} h(u_k|y_k, \hat{a}_{k,k}) &\stackrel{(a)}{=} h(u_k - \alpha y_k|y_k, \hat{a}_{k,k}) \\ &\stackrel{(b)}{\leq} h(u_k - \alpha y_k|\hat{a}_{k,k}) \\ &\stackrel{(c)}{\leq} \mathbb{E} [\log(\pi e \cdot \mathbb{E}[|u_k - \alpha y_k|^2|\hat{a}_{k,k}])] \end{aligned} \quad (72)$$

where (a) holds for any deterministic function  $\alpha$  of  $y_k$  and  $\hat{a}_{k,k}$ , (b) follows from the fact that conditioning reduces entropy and (c) follows by the fact that differential entropy is maximized by a Gaussian RV with the same second moment. Substituting (12) in (8) we have

$$y_k = (\hat{a}_{k,k} + f_k)u_k + I_k + z_k \quad (73)$$

where  $\hat{a}_{k,k}u_k$  and  $f_k u_k + I_k + z_k$  are *uncorrelated* and zero-mean, even if we condition on  $\hat{a}_{k,k}$ , because  $\hat{a}_{k,k}, f_k, u_1, \dots, u_K, z_k$  are independent, zero-mean Gaussian's. Thus, we have

$$\mathbb{E}[|y_k|^2|\hat{a}_{k,k}] = |\hat{a}_{k,k}|^2 \mathbb{E}[|u_k|^2] + \sigma_2^2 \mathbb{E}[|u_k|^2] + \mathbb{E}[|I_k|^2|\hat{a}_{k,k}] + N_0, \quad (74)$$

Choosing  $\alpha$  that minimizes  $\mathbb{E}[|u_k - \alpha y_k|^2|\hat{a}_{k,k}]$  tightens the bound. This corresponds to setting  $\alpha y_k$  equal to the linear MMSE estimate of  $u_k$  given  $y_k$  and  $\hat{a}_{k,k}$ , i.e.,

$$\alpha = \frac{\mathbb{E}[u_k y_k^*|\hat{a}_{k,k}]}{\mathbb{E}[|y_k|^2|\hat{a}_{k,k}]} = \frac{\mathbb{E}[|u_k|^2|\hat{a}_{k,k}]}{\mathbb{E}[|y_k|^2|\hat{a}_{k,k}]} \quad (75)$$

Using (74), the corresponding MMSE is given by

$$\mathbb{E}[|u_k - \alpha y_k|^2|\hat{a}_{k,k}] = \mathbb{E}[|u_k|^2] \left(1 - \frac{\mathbb{E}[|u_k|^2|\hat{a}_{k,k}]}{\mathbb{E}[|y_k|^2|\hat{a}_{k,k}]}\right) \quad (76)$$

$$= \frac{P}{M} \frac{1 + \sigma_2^2 \frac{P}{N_0 M} + \mathbb{E}[|I_k|^2|\hat{a}_{k,k}]/N_0}{|\hat{a}_{k,k}|^2 \frac{P}{N_0 M} + 1 + \sigma_2^2 \frac{P}{N_0 M} + \mathbb{E}[|I_k|^2|\hat{a}_{k,k}]/N_0} \quad (77)$$

Replacing (77) into (72) and using  $h(u_k) = \log(\pi e \frac{P}{M})$ , we obtain (14).

APPENDIX II  
PROOF OF THEOREM 2

Using the lower bound on  $R_k$  from Theorem 1 we have:

$$\begin{aligned} \Delta R &\leq \mathbb{E} \left[ \log \left( 1 + \frac{|\mathbf{h}^H \mathbf{v}_k|^2 P}{N_0 M} \right) \right] - \mathbb{E} \left[ \log \left( 1 + \frac{|\hat{a}_{k,k}|^2 P / (N_0 M)}{1 + \sigma_2^2 P / (N_0 M) + \mathbb{E}[|I_k|^2 |\hat{a}_{k,k}|] / N_0} \right) \right] \\ &\stackrel{(a)}{\leq} \mathbb{E} \left[ \log \left( 1 + \frac{|\mathbf{h}^H \mathbf{v}_k|^2 P}{N_0 M} \right) \right] - \mathbb{E} \left[ \log \left( 1 + \frac{P}{N_0 M} (|\hat{a}_{k,k}|^2 + \sigma_2^2) \right) \right] \\ &\quad + \mathbb{E} \left[ \log \left( 1 + \sigma_2^2 \frac{P}{N_0 M} + \frac{\mathbb{E}[|I_k|^2 |\hat{a}_{k,k}|]}{N_0} \right) \right] \\ &\stackrel{(b)}{\leq} \mathbb{E} \left[ \log \left( 1 + \sigma_2^2 \frac{P}{N_0 M} + \frac{\mathbb{E}[|I_k|^2 |\hat{a}_{k,k}|]}{N_0} \right) \right] \end{aligned} \quad (78)$$

$$\stackrel{(c)}{\leq} \log \left( 1 + \sigma_2^2 \frac{P}{N_0 M} + \frac{\mathbb{E}[|I_k|^2]}{N_0} \right) \quad (79)$$

where (a) follows by dropping the non-negative term  $\mathbb{E}[|I_k|^2 |\hat{a}_{k,k}|] / N_0$ . Using the fact that  $\mathbf{h}_k$  is spatially white and  $\mathbf{v}_k$  is selected independent of  $\mathbf{h}_k$  (by the ZF procedure), it follows that  $\mathbf{h}_k^H \mathbf{v}_k \sim \mathcal{CN}(0, 1)$  and  $\hat{a}_{k,k} \sim \mathcal{CN}(0, 1 - \sigma_2^2)$ . Direct application of Lemma 2, which is provided below, with  $A = P/(N_0 M)$ ,  $\lambda = \sigma_2^2$  and  $X = |\mathbf{h}_k^H \mathbf{v}_k|^2$ , thus proves (b). Finally, (c) follows from the concavity of  $\log(\cdot)$  and Jensen's inequality.

*Lemma 2:* If  $X$  is a non-negative random variable with  $\mathbb{E}[X] = 1$ , for any  $A > 0$  and any  $0 \leq \lambda \leq 1$ :

$$\mathbb{E}[\log(1 + XA)] \leq \mathbb{E}[\log(1 + (\lambda + (1 - \lambda)X)A)]. \quad (80)$$

**Proof:** For all  $0 \leq z \leq 1$ , define the function

$$\psi(z) = \mathbb{E}[\log(1 + zA + (1 - z)XA)]. \quad (81)$$

Then (80) is equivalent to the inequality  $\psi(0) \leq \psi(\lambda)$ . By the concavity of  $\log(\cdot)$  and Jensen's inequality we have

$$\psi(z) \leq \log(1 + zA + (1 - z)\mathbb{E}[X]A) = \psi(1). \quad (82)$$

In particular,  $\psi(0) \leq \psi(1)$ . Moreover,  $\psi(z)$  is an expectation of the composition of a concave function and a linear function of  $z$ , and is hence concave [65]. Thus, the concave function  $\psi(z)$  for  $z \in [0, 1]$  lies above the line joining the points  $(0, \psi(0))$  and  $(1, \psi(1))$ . Hence, we have  $\psi(0) \leq \psi(\lambda)$  for  $\lambda \in [0, 1]$ , which proves (80).  $\square$

APPENDIX III  
PROOF OF THEOREM 4

Using (18), to compute  $\overline{\Delta R}^{\text{AF}}$  we only need to find  $\mathbb{E}[|\mathbf{h}_k^H \hat{\mathbf{v}}_j|^2]$ :

$$\begin{aligned} \mathbb{E}[|\mathbf{h}_k^H \hat{\mathbf{v}}_j|^2] &\stackrel{(a)}{=} \mathbb{E}[|\hat{\mathbf{h}}_k^H \hat{\mathbf{v}}_j + \mathbf{e}_k^H \hat{\mathbf{v}}_j|^2] \\ &\stackrel{(b)}{=} \mathbb{E}[|\mathbf{e}_k^H \hat{\mathbf{v}}_j|^2] \\ &\stackrel{(c)}{=} \mathbb{E}[\hat{\mathbf{v}}_j^H \mathbb{E}[\mathbf{e}_k \mathbf{e}_k^H] \hat{\mathbf{v}}_j] \\ &\stackrel{(d)}{=} \sigma_e^2 \end{aligned} \quad (83)$$

where (a) follows from (27), (b) follows from the fact that  $\hat{\mathbf{h}}_k^H \hat{\mathbf{v}}_j = 0 \forall j \neq k$  by naive ZF, (c) is obtained from the independence of  $\mathbf{e}_k$  and  $\hat{\mathbf{v}}_j$  ( $\hat{\mathbf{v}}_j$  is a deterministic function of  $\{\hat{\mathbf{h}}_i\}_{i \neq j}$ ), and (d) follows from  $\mathbb{E}[\mathbf{e}_k \mathbf{e}_k^H] = \sigma_e^2 \mathbf{I}$  and  $\|\hat{\mathbf{v}}_j\| = 1$ .

APPENDIX IV  
PROOF OF THEOREM 5

To compute the rate gap upper bound, we determine  $\mathbb{E}[|\mathbf{h}_k^H \hat{\mathbf{v}}_j|^2]$  by writing the channel in terms of the UT channel estimate (which is quantized) and the UT estimation error:  $\mathbf{h}_k = \tilde{\mathbf{h}}_k + \mathbf{n}_k$  from (6). This yields:

$$\begin{aligned}
\mathbb{E}[|\mathbf{h}_k^H \hat{\mathbf{v}}_j|^2] &\stackrel{(a)}{=} \mathbb{E}[|\tilde{\mathbf{h}}_k^H \hat{\mathbf{v}}_j|^2] + \mathbb{E}[|\mathbf{n}_k^H \hat{\mathbf{v}}_j|^2] \\
&\stackrel{(b)}{=} \mathbb{E}[\|\tilde{\mathbf{h}}_k\|^2] \mathbb{E}\left[\frac{|\tilde{\mathbf{h}}_k^H \hat{\mathbf{v}}_j|^2}{\|\tilde{\mathbf{h}}_k\|^2}\right] + \mathbb{E}[|\mathbf{n}_k^H \hat{\mathbf{v}}_j|^2] \\
&\stackrel{(c)}{=} \frac{\mathbb{E}[\|\tilde{\mathbf{h}}_k\|^2]}{M-1} 2^B \beta \left(2^B, \frac{M}{M-1}\right) + \mathbb{E}[\hat{\mathbf{v}}_j^H \mathbb{E}[\mathbf{n}_k \mathbf{n}_k^H] \hat{\mathbf{v}}_j] \\
&\stackrel{(d)}{=} \frac{M}{M-1} \frac{\beta_1 P}{N_0 + \beta_1 P} 2^B \beta \left(2^B, \frac{M}{M-1}\right) + \sigma_1^2
\end{aligned} \tag{84}$$

where (a) is obtained from the representation  $\mathbf{h}_k = \tilde{\mathbf{h}}_k + \mathbf{n}_k$  and the fact that  $\mathbb{E}[\tilde{\mathbf{h}}_k^H \hat{\mathbf{v}}_j \hat{\mathbf{v}}_j^H \mathbf{n}_k] = 0$  because  $\mathbf{n}_k$  is zero-mean Gaussian and is independent of  $\tilde{\mathbf{h}}_k$  and  $\hat{\mathbf{v}}_j$ , (b) from the independence of the channel norm and direction of  $\tilde{\mathbf{h}}_k$ , (c) from (35) and from the property [26, Lemma 2]  $\mathbb{E}\left[\frac{|\tilde{\mathbf{h}}_k^H \hat{\mathbf{v}}_j|^2}{\|\tilde{\mathbf{h}}_k\|^2}\right] = \frac{1}{M-1} \mathbb{E}[\sin^2(\tilde{\mathbf{h}}_k, \hat{\mathbf{v}}_j)]$ , and finally (d) by computing the expected norm of  $\tilde{\mathbf{h}}_k = \frac{\sqrt{\beta_1 P}}{N_0 + \beta_1 P} \mathbf{s}_k$  using  $\mathbf{s}_k = \sqrt{\beta_1 P} \mathbf{h}_k + \mathbf{z}_k$ . The final result follows by using the above result in the expression (16) for the rate gap.

APPENDIX V  
PROOF OF THEOREM 6

We first decompose the interference variance term as

$$\mathbb{E}[|\mathbf{h}_k^H \hat{\mathbf{v}}_j|^2] = (1 - P_{e,\text{fb}}) \mathbb{E}[|\mathbf{h}_k^H \hat{\mathbf{v}}_j|^2 | \text{no fb. errors}] + P_{e,\text{fb}} \mathbb{E}[|\mathbf{h}_k^H \hat{\mathbf{v}}_j|^2 | \text{fb. errors}] \tag{85}$$

$$\leq (1 - P_{e,\text{fb}}) \frac{M}{M-1} \frac{\beta_1 P}{N_0 + \beta_1 P} 2^B \beta \left(2^B, \frac{M}{M-1}\right) + \sigma_1^2 + P_{e,\text{fb}}, \tag{86}$$

where  $\mathbb{E}[|\mathbf{h}_k^H \hat{\mathbf{v}}_j|^2 | \text{no fb. errors}]$  is the same as in the error-free case and is thus given in (84) while for the case of feedback errors we trivially have  $\mathbb{E}[|\mathbf{h}_k^H \hat{\mathbf{v}}_j|^2 | \text{fb. errors}] \leq 1$ . The final result is reached by simply substituting  $B = \alpha(M-1) \log_2 \frac{P}{N_0}$  and using the bound in the beta function (35).

APPENDIX VI  
PROOF OF THEOREM 7

Using the argument from the proof of Theorem III (analog FB over AWGN channel), the expected interference coefficient  $\mathbb{E}[|\mathbf{h}_k^H \hat{\mathbf{v}}_j|^2]$  is equal to the variance of the channel estimation error. This quantity must be averaged over the uplink channel matrix  $\mathbf{A}$ , and thus using symmetry and (49), is given by

$$\begin{aligned}
\mathbb{E}[\sigma_k^2(\mathbf{A})] &= \mathbb{E}\left[\frac{1}{L} \text{tr}\left(\mathbf{I} - c^2 \mathbf{A}^H \left[\beta_{\text{fb}} P \mathbf{A} \mathbf{A}^H + N_0 \mathbf{I}\right]^{-1} \mathbf{A}\right)\right] \\
&= \mathbb{E}\left[\frac{1}{L} \sum_{k=1}^L \frac{N_0 + (\beta_{\text{fb}} P - c^2) \lambda_k}{N_0 + \beta_{\text{fb}} P \lambda_k}\right] \\
&= \frac{1}{1 + \beta_1 \frac{P}{N_0}} + \frac{\beta_1 \frac{P}{N_0}}{1 + \beta_1 \frac{P}{N_0}} \text{mmse}\left(\beta_{\text{fb}} \frac{P}{N_0}\right)
\end{aligned} \tag{87}$$

where  $\text{mmse}(\rho)$  is defined in (51).

In order to obtain the high SNR result, we first state a closed-form expression for  $\text{mmse}(\rho)$  using well-known results from multivariate statistics (see for example [66]):

$$\text{mmse}(\rho) = \frac{e^{1/\rho}}{\rho} \sum_{k=0}^{L-1} \sum_{\ell=0}^k \sum_{m=0}^{2\ell} \mathbf{X}_{k,\ell,m} \text{E}_i(M - L + m + 1, 1/\rho) \tag{88}$$

where the coefficients  $X_{k,\ell,m}$  are given by

$$X_{k,\ell,m} = \frac{(-1)^m (2\ell)! (M-L+m)!}{L 2^{2k-m} \ell! m! (M-L+\ell)!} \binom{2(k-\ell)}{k-\ell} \binom{2(M-L+\ell)}{2\ell-m}$$

Based upon this we can characterize the asymptotic behavior of the product  $\rho \text{mmse}(\rho)$  for  $\rho \rightarrow \infty$ . Using the asymptotic expansion of  $e^{1/\rho} \text{E}_i(n, 1/\rho)$ , we have

$$\rho \text{mmse}(\rho) = \begin{cases} \frac{1}{M-L} + o(1) & \text{for } L < M \\ -\gamma + \log \rho + \sum_{k=0}^{L-1} \sum_{\ell=0}^k \sum_{m=1}^{2\ell} \frac{X_{k,\ell,m}}{m} + o(1) & \text{for } L = M \end{cases} \quad (89)$$

where we used the facts:

$$\text{E}_i(1, 1/\rho) e^{1/\rho} = -\gamma + \log \rho + o(1), \quad \rho \rightarrow \infty \quad (90)$$

$$\text{E}_i(n, 1/\rho) e^{1/\rho} = \frac{1}{n-1} + o(1), \quad \text{for } n > 1, \rho \rightarrow \infty \quad (91)$$

$$\sum_{k=0}^{L-1} \sum_{\ell=0}^k \sum_{m=0}^{2\ell} \frac{X_{k,\ell,m}}{M-L+m} = \frac{1}{M-L}, \quad \text{for } L < M \quad (92)$$

$$\sum_{k=0}^{L-1} \sum_{\ell=0}^k X_{k,\ell,0} = 1, \quad \text{for } L = M \quad (93)$$

#### APPENDIX VII PROOF OF THEOREM 8

We can lower bound the genie-aided rate of Theorem 3 as follows.

$$\begin{aligned} I(u_k; y_k, \mathcal{A}_k) &= \mathbb{E} \left[ \log \left( 1 + \frac{|a_{k,k}|^2 P / (N_0 M)}{1 + \sum_{j \neq k} |a_{k,j}|^2 P / (N_0 M)} \right) \right] \\ &= \mathbb{E} \left[ \log \left( 1 + \sum_j |a_{k,j}|^2 P / (N_0 M) \right) \right] - \mathbb{E} \left[ \log \left( 1 + \sum_{j \neq k} |a_{k,j}|^2 P / (N_0 M) \right) \right] \\ &\stackrel{(a)}{\geq} R_k^{\text{ZF}} - \mathbb{E} \left[ \log \left( 1 + \sum_{j \neq k} |a_{k,j}|^2 \frac{P}{N_0 M} \right) \right] \\ &\stackrel{(b)}{\geq} R_k^{\text{ZF}} - \mathbb{E} \left[ \log \left( 1 + \sum_{j \neq k} \mathbb{E}[|a_{k,j}|^2 | \mathbf{A}] \frac{P}{N_0 M} \right) \right] \\ &\stackrel{(c)}{=} R_k^{\text{ZF}} - \mathbb{E} \left[ \log \left( 1 + \frac{P}{N_0 M} \sigma_k^2(\mathbf{A}) \right) \right] \end{aligned}$$

where (a) follows by dropping the non-negative terms and (b) follows by conditioning with respect to the uplink channel matrix  $\mathbf{A}$  and then applying Jensen's inequality in the inner conditional expectation, (c) follows by noticing  $\mathbb{E}[|I_k|^2 | \mathbf{A}] = (M-1)P\sigma_k^2(\mathbf{A})$  where  $\sigma_k^2(\mathbf{A})$  is defined in (49). Then, we obtain an upper bound of for the gap

between the ideal ZF rate and the genie-aided rate given by

$$\begin{aligned}
R_k^{\text{ZF}} - I(u_k; y_k, \mathcal{A}_k) &\leq \mathbb{E} \left[ \log \left( 1 + \frac{P}{N_0} \frac{M-1}{M} \sigma_k^2(\mathbf{A}) \right) \right] \\
&\stackrel{\text{(a)}}{=} \frac{1}{M} \sum_{k=1}^M \mathbb{E} \left[ \log \left( 1 + \frac{P}{N_0} \frac{M-1}{M} \sigma_k^2(\mathbf{A}) \right) \right] \\
&\leq \mathbb{E} \left[ \log \left( 1 + \frac{P}{N_0} \frac{M-1}{M} \frac{1}{M} \sum_{k=1}^M \sigma_k^2(\mathbf{A}) \right) \right] \\
&\stackrel{\text{(b)}}{=} \mathbb{E} \left[ \log \left( 1 + \frac{P}{N_0} \frac{M-1}{M} \left( \frac{1}{1 + \beta_1 \frac{P}{N_0}} + \frac{\beta_1 \frac{P}{N_0}}{1 + \beta_1 \frac{P}{N_0}} \frac{1}{M} \sum_{k=1}^M \frac{1}{1 + \beta_{\text{fb}} \frac{P}{N_0} \lambda_k} \right) \right) \right] \\
&\leq \mathbb{E} \left[ \log \left( 1 + \frac{M-1}{M} \left( \frac{1}{\beta_1} + \frac{\frac{P}{N_0}}{1 + \beta_{\text{fb}} \frac{P}{N_0} \lambda_{\min}} \right) \right) \right] \tag{94}
\end{aligned}$$

where (a) follows because the term  $\mathbb{E} \left[ \log \left( 1 + \frac{P}{N_0} \frac{M-1}{M} \sigma_k^2(\mathbf{A}) \right) \right]$  is independent of  $k$  due to the symmetry over  $k$ , (b) follows by using the same derivation that leads to (87) and (51), and the last line follows by monotonicity of the log, where  $\lambda_{\min}$  denotes the minimum eigenvalue of  $\mathbf{A}^H \mathbf{A}$ .

Our goal is to show that the term in the last line of (94) is bounded. To this purpose, we write the last line of (94) as the sum of three terms,

$$\log \left( 1 + \frac{M-1}{M\beta_1} + \frac{M-1}{M} \frac{P}{N_0} \right) + \mathbb{E} \left[ \log \left( 1 + \frac{\left(1 + \frac{M-1}{M\beta_1}\right) \frac{\beta_{\text{fb}} P}{N_0}}{1 + \frac{M-1}{M} \left( \frac{1}{\beta_1} + \frac{P}{N_0} \right)} \lambda_{\min} \right) \right] - \mathbb{E} \left[ \log \left( 1 + \beta_{\text{fb}} \frac{P}{N_0} \lambda_{\min} \right) \right] \tag{95}$$

For  $\mathbf{A} \in \mathbb{C}^{M \times M}$ , complex Gaussian with i.i.d. zero-mean components, it is well-known that  $\lambda_{\min}$  is chi-squared with 2 degrees of freedom and mean 1 [67]. Hence, the third term in (95) yields

$$\mathbb{E} \left[ \log \left( 1 + \beta_{\text{fb}} \frac{P}{N_0} \lambda_{\min} \right) \right] = e^{-\frac{N_0}{\beta_{\text{fb}} P}} \text{Ei} \left( 1, \frac{N_0}{\beta_{\text{fb}} P} \right) = -\gamma + \log \frac{\beta_{\text{fb}} P}{N_0} + o(1)$$

The second term in (95) is bounded by a constant, independent of  $P/N_0$ , and finally the first term in (95), for high SNR, can be written as  $\log \frac{P}{N_0} + O(1)$ . It follows that the  $\log(P/N_0)$  terms in the first and the third terms of the upper bound cancel, so that (95) is bounded. This establishes the result.

## APPENDIX VIII

### GENIE-AIDED UPPER BOUND FOR REGULAR PROCESSES WITH DELAYED FEEDBACK

We show that the genie-aided upper bound of Theorem 3, is uniformly bounded for any SNR when the noiseless prediction error is positive. For analytical simplicity, we assume perfect common training and perfect (delayed) feedback. Hence, the only source of “noise” in the CSIT is due to the prediction error. We can write  $\mathbf{h}_k(t) = \tilde{\mathbf{h}}_k(t) + \mathbf{n}_k(t)$ , where  $\tilde{\mathbf{h}}_k(t)$  is the one-step prediction of  $\mathbf{h}_k(t)$  from its (noiseless) past, and  $\mathbf{n}_k(t)$  is the prediction error. From what was stated earlier, we have that  $\mathbf{h}_k(t)$ ,  $\tilde{\mathbf{h}}_k(t)$  and  $\mathbf{n}_k(t)$  are jointly complex Gaussian, i.i.d. in the spatial domain, with mean zero and variance per component equal to 1,  $1 - \epsilon_1(0)$  and  $\epsilon_1(0)$ , respectively. It is useful to write the error as  $\mathbf{n}_k(t) = \sqrt{\epsilon_1(0)} \mathbf{\Delta}(t)$ , where  $\mathbf{\Delta}(t) \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ . From (20), the genie-aided upper bound is given by

$$R_k \leq \mathbb{E} \left[ \log \left( 1 + \frac{P |\mathbf{h}_k^H(t) \hat{\mathbf{v}}_k(t)|^2}{N_0 M + P \sum_{j \neq k} |\mathbf{h}_k^H(t) \hat{\mathbf{v}}_j(t)|^2} \right) \right]$$

where  $\hat{\mathbf{v}}_j(t)$  is orthogonal to  $\tilde{\mathbf{h}}_k(t)$ . Using the fact that the upper bound is non-decreasing in  $P/N_0$ , we let  $P/N_0 \rightarrow \infty$  in (96) and obtain

$$\begin{aligned}
R_k &\leq \mathbb{E} \left[ \log \left( |\mathbf{h}_k^H(t)\hat{\mathbf{v}}_k(t)|^2 + \sum_{j \neq k} |\mathbf{h}_k^H(t)\hat{\mathbf{v}}_j(t)|^2 \right) \right] - \mathbb{E} \left[ \log \left( \sum_{j \neq k} |\mathbf{h}_k^H(t)\hat{\mathbf{v}}_j(t)|^2 \right) \right] \\
&\stackrel{(a)}{\leq} \log(1 + \epsilon_1(0)(M-1)) - \mathbb{E} \left[ \log \left( \epsilon_1(0) \sum_{j \neq k} |\Delta_k^H(t)\hat{\mathbf{v}}_j(t)|^2 \right) \right] \\
&\stackrel{(b)}{=} \log \left( \frac{1}{\epsilon_1(0)} + M-1 \right) - \mathbb{E}[\log(|\Delta_k(t)|^2)] - \mathbb{E} \left[ \log \left( \sum_{j \neq k} \frac{|\Delta_k^H(t)\hat{\mathbf{v}}_j(t)|^2}{|\Delta_k(t)|^2} \right) \right] \\
&\stackrel{(c)}{=} \log \left( \frac{1}{\epsilon_1(0)} + M-1 \right) - \psi(M) + \frac{1}{2M-1} + \frac{1}{2M-2}
\end{aligned} \tag{96}$$

where (a) follows by applying Jensen's inequality to the first term and noticing that both  $\mathbf{h}_k^H(t)\hat{\mathbf{v}}_k(t)$  and  $\Delta_k^H(t)\hat{\mathbf{v}}_j(t)$  are  $\sim \mathcal{CN}(0, 1)$ , (b) follows by expressing  $|\Delta_k^H(t)\hat{\mathbf{v}}_j(t)|^2 = |\Delta_k^H(t)|^2 \frac{|\Delta_k^H(t)\hat{\mathbf{v}}_j(t)|^2}{|\Delta_k^H(t)|^2}$ , (c) is obtained by noticing that  $|\Delta_k(\tau)|^2$  is chi-square distributed with  $2M$  degrees of freedom and that  $\sum_{j \neq k} \frac{|\Delta_k^H(t)\hat{\mathbf{v}}_j(t)|^2}{|\Delta_k(t)|^2}$  is beta distributed with parameters  $(M-1, 1)$ , and finally  $\psi(M)$  is the Euler-Digamma function.  $\diamond$

#### Acknowledgment

The work of G. Caire was partially supported by NSF Grant CCF-0635326.

## REFERENCES

- [1] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Commun. : Kluwer Academic Press*, no. 6, pp. 311–335, 1998.
- [2] I. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–595, 1999.
- [3] M. Costa, "Writing on dirty paper," *IEEE Trans. on Inform. Theory*, vol. 29, pp. 439–441, May 1983.
- [4] G. Caire and S. Shamai, "On the achievable throughput of a multiantenna Gaussian broadcast channel," *IEEE Trans. on Inform. Theory*, vol. 49, no. 7, pp. 1691–1706, 2003.
- [5] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels," *IEEE Trans. on Inform. Theory*, vol. 49, no. 10, pp. 2658–2668, 2003.
- [6] P. Viswanath and D. Tse, "Sum capacity of the vector Gaussian broadcast channel and uplink-downlink duality," *IEEE Trans. on Inform. Theory*, vol. 49, no. 8, pp. 1912–1921, 2003.
- [7] W. Yu and J. Cioffi, "Sum capacity of Gaussian vector broadcast channels," *IEEE Trans. on Inform. Theory*, vol. 50, no. 9, pp. 1875–1892, 2004.
- [8] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Trans. on Inform. Theory*, vol. 52, no. 9, pp. 3936–3964, 2006.
- [9] N. Jindal and A. Goldsmith, "Dirty paper coding vs. TDMA for MIMO broadcast channels," *IEEE Trans. Inform. Theory*, vol. 51, no. 5, pp. 1783–1794, May 2005.
- [10] 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 Sig. Proc. Magazine*, 2007.
- [11] U. Erez and S. Ten Brink, "A close-to-capacity dirty paper coding scheme," *IEEE Trans. on Inform. Theory*, vol. 51, no. 10, pp. 3417 – 3432, October 2005.
- [12] A. Bennatan, D. Burshtein, G. Caire, and S. Shamai, "Superposition coding for side-information channels," *IEEE Trans. on Inform. Theory*, vol. 52, no. 5, pp. 1872 – 1889, May 2006.
- [13] Y. Sun, A. Liveris, V. Stankovic, and Zixiang Xiong, "Near-capacity dirty-paper code designs based on TCQ and IRA codes," in *proceedings of IEEE Int. Symp. on Inform. Theory, ISIT, Adelaide, Australia, September 2005*, pp. 184 – 188.
- [14] 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.
- [15] G. Dimic and N. Sidiropoulos, "On downlink beamforming with greedy user selection: performance analysis and simple new algorithm," *IEEE Trans. on Sig. Proc.*, vol. 53, no. 10, pp. 3857–3868, October 2005.
- [16] A. Lapidoth, S. S. (Shitz), and M. . A. Wigger, "On the capacity of fading MIMO broadcast channels with imperfect transmitter side-information," *Annual Allerton Conference on Communication, Control, and Computing*, September 2005.
- [17] S. Shamai, G. Caire, and N. Jindal, "On the required accuracy of transmitter channel state information in multiple-antenna broadcast channels," *Asilomar Conference on Signals, Systems and Computers*, 2007.
- [18] B. Hassibi and B. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. on Inform. Theory*, vol. 49, no. 4, pp. 951–963, 2003.
- [19] A. Narula, M. J. Lopez, M. D. Trott, and G. W. Wornell, "Efficient use of side information in multiple-antenna data transmission over fading channels," *IEEE Jour. Select. Areas in Commun.*, vol. 16, no. 8, pp. 1423–1436, October 1998.
- [20] D. Love, R. Heath Jr, and T. Strohmer, "Grassmannian beamforming for multiple-input multiple-output wireless systems," *IEEE Trans. on Inform. Theory*, vol. 49, no. 10, pp. 2735–2747, 2003.
- [21] K. Mukkavilli, A. Sabharwal, E. Erkip, and B. Aazhang, "On beamforming with finite rate feedback in multiple-antenna systems," *IEEE Trans. on Inform. Theory*, vol. 49, no. 10, pp. 2562–2579, 2003.
- [22] S. Srinivasa and S. A. Jafar, "The optimality of transmit beamforming: a unified view," *IEEE Trans. on Inform. Theory*, vol. 53, no. 4, pp. 1558–1564, April 2007.
- [23] S. A. Jafar and S. Srinivasa, "On the optimality of beamforming with quantized feedback," *To appear in the IEEE Trans. on Comm.*, 2007.
- [24] C. K. Au-Yeung and D. J. Love, "On the performance of random vector quantization limited feedback beamforming in a MISO system," *IEEE Trans. Wireless Comm.*, vol. 6, pp. 458 – 462, 2007.
- [25] A. D. Dabagh and D. J. Love, "Feedback rate-capacity loss tradeoff for limited feedback MIMO systems," *IEEE Trans. on Inform. Theory*, vol. 52, no. 5, pp. 2190–2202, May 2006.
- [26] N. Jindal, "MIMO broadcast channels with finite rate feedback," *IEEE Trans. on Inform. Theory*, vol. 52, no. 11, pp. 5045–5059, November 2006.
- [27] T. Marzetta, "How much training is required for multiuser MIMO ?" *Signals, Systems and Computers, 2006. ACSSC'06. Fortieth Asilomar Conference on*, pp. 359–363, 2006.
- [28] T.L.Marzetta and B.M.Hochwald, "Fast transfer of channel state information in wireless systems," *Submitted to "IEEE Transactions on Signal Processing"*, June 2004.
- [29] P. Ding, D. Love, and M. Zoltowski, "Multiple antenna broadcast channels with shape feedback and limited feedback," *IEEE Trans. on Sig. Proc.*, vol. 55, no. 7 Part 1, pp. 3417–3428, 2007.
- [30] K. Huang, R. Heath Jr, and J. Andrews, "Space division multiple access with a sum feedback rate constraint," *IEEE Trans. on Sig. Proc.*, vol. 55, no. 7, pp. 3879–3891, July 2007.
- [31] A. Dana, M. Sharif, and B. Hassibi, "On the capacity region of multi-antenna Gaussian broadcast channels with estimation error," in *IEEE Int. Symp. on Inform. Theory*, July 2006.
- [32] M. Kobayashi and G. Caire, "Joint beamforming and scheduling for a multi-antenna downlink with imperfect transmitter channel knowledge," *IEEE J. Select. Areas Commun.*, vol. 25, no. 7, 2007.

- [33] T. Yoo, N. Jindal, and A. Goldsmith, "Multi-antenna downlink channels with limited feedback and user selection," *IEEE J. Select. Areas Commun.*, vol. 25, pp. 1478–1491, 2007.
- [34] C. Swannack, G. W. Wornell, and E. Uysal-Biyikoglu, "MIMO broadcast scheduling with quantized channel state information," in *Proceedings of the IEEE Int. Symp. on Inform. Theory, ISIT*, Seattle, WA, 2006.
- [35] M. Médard, "Channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Trans. on Inform. Theory*, vol. 46, no. 3, pp. 933–946, May 2000.
- [36] A. Lapidoth and S. Shamai, "Fading channels: how perfect need "perfect" side information" be?" *IEEE Trans. on Inform. Theory*, vol. 48, no. 5, pp. 1118–1134, 2002.
- [37] M. Kobayashi, G. Caire, and N. Jindal, "How much training and feedback are needed in MIMO broadcast channels?" in *Proc. IEEE Int. Symp. Inform. Theory. Toronto, Canada*, 2008.
- [38] —, "Optimized training and feedback for MIMO dwnlink channels," in *Proc. IEEE Information Theory Workshop, Greece*, 2008.
- [39] A. Lapidoth, "On the asymptotic capacity of stationary Gaussian fading channels," *IEEE Trans. on Inform. Theory*, vol. 51, no. 2, p. 437, 2005.
- [40] E. Biglieri, J. Proakis, S. Shamai, and D. di Elettronica, "Fading channels: information-theoretic and communications aspects," *IEEE Trans. on Inform. Theory*, vol. 44, no. 6, pp. 2619–2692, 1998.
- [41] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE J. Select. Areas Commun.*, vol. 24, no. 3, pp. 528–541, 2006.
- [42] A. Bayesteh and A. K. Khandani, "An efficient method for user selection in MIMO broadcast channels," *Proceeding of CISS'2005*, March 2005.
- [43] M. Alouini and A. Goldsmith, "Capacity of Rayleigh fading channels under different adaptive transmission and diversity-combining techniques," *IEEE Trans. on Vehic. Tech.*, vol. 48, no. 4, pp. 1165–1181, 1999.
- [44] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. New York: Dover, 1964.
- [45] N. Ravindran and N. Jindal, "Multi-user diversity vs. accurate channel feedback for MIMO broadcast channels," *Arxiv preprint cs.IT/0710.1336*, 2007.
- [46] H. Poor, *An introduction to signal detection and estimation*. Springer-Verlag New York, Inc. New York, NY, USA, 1994.
- [47] R. Prasad, W. Mohr, and W. Konhauser, *Third Generation Mobile Communication Systems*. Artech House, Inc. Norwood, MA, USA, 2000.
- [48] "Pilots for MIMO communication systems," Patent EP 1556985/US 2004179627/WO 2004038988.
- [49] J. Jose, A. Ashikhmin, P. Whiting, and S. Vishwanath, "Scheduling and pre-conditioning in multi-user MIMO TDD systems," *Arxiv preprint cs.IT/0709.4513*, 2007.
- [50] T. Thomas, K. Baum, and P. Sartori, "Obtaining channel knowledge for closed-loop multi-stream broadband MIMO-OFDM communications using direct channel feedback," in *IEEE Global Telecommunications Conference, 2005. GLOBECOM '05*, vol. 6, November 2005.
- [51] D. Samardzija and N. Mandayam, "Unquantized and uncoded channel state information feedback on wireless channels," *Proceeding of IEEE WCNC'2005*, pp. New Orleans, LA, USA, March 2005.
- [52] H. Shirani-Mehr and G. Caire, "Channel state feedback schemes for multiuser MIMO-OFDM downlink," April 2008, see also: arXiv:0804.0611.
- [53] K. Raj Kumar and G. Caire, "Channel state feedback over the MIMO-MAC," in *IEEE Int. Symp. on Inform. Theory, ISIT09*, Seoul (Korea), June 2009.
- [54] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.
- [55] M. Gastpar, B. Rimoldi, and M. Vetterli, "To code, or not to code : Lossy source-channel communication revisited," *IEEE Trans. on Inform. Theory*, vol. 49, May 2003.
- [56] D. N. C. Tse, P. Viswanath, and L. Zheng, "Diversity-multiplexing tradeoff in multiple-access channels," *IEEE Trans. on Inform. Theory*, vol. 50, no. 9, pp. 1859–1874, 2004.
- [57] L. Zheng and D. Tse, "Diversity and multiplexing: a fundamental tradeoff in multiple-antenna channels," *IEEE Trans. on Inform. Theory*, vol. 49, no. 5, pp. 1073–1096, 2003.
- [58] E. Viterbo and J. Boutros, "A universal lattice code decoder for fading channels," *IEEE Transactions on Information Theory*, vol. 45, no. 5, pp. 1639–1642, 1999.
- [59] M. Damen, H. El Gamal, and G. Caire, "On maximum-likelihood detection and the search for the closest lattice point," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2389–2402, 2003.
- [60] R. Kumar and G. Caire, "Channel state feedback over the MIMO-MAC," in *Proc. Information Theory and Applications Workshop (ITA)*, San Diego, CA, 2009.
- [61] P. Bender, P. Black, M. Grob, R. Padovani, N. Sindhushayana, and A. Viterbi, "CDMA/HDR: A bandwidth-efficient high-speed wireless data service for nomadic users," *IEEE Commun. Mag.*, vol. 38, pp. 70–77, July 2000.
- [62] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.
- [63] T. Cover and J. Thomas, *Elements of Information Theory*. John Wiley, New York, 2005.
- [64] J. Zhang, R. H. Jr, M. Kountouris, and J. G. Andrews, "Mode Switching for MIMO Broadcast Channel Based on Delay and Channel Quantization," *Arxiv preprint arXiv:0812.3120*, December 2008.
- [65] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
- [66] H. Shin and J. Lee, "Capacity of multiple-antenna fading channels: spatial fading correlation, double scattering, and keyhole," *Information Theory, IEEE Transactions on*, vol. 49, no. 10, pp. 2636–2647, 2003.
- [67] A. Edelman, "Eigenvalues and Condition Numbers of Random Matrices," 1989, MIT PhD Dissertation.