

Research Article

Mode Switching for the Multi-Antenna Broadcast Channel Based on Delay and Channel Quantization

Jun Zhang, Robert W. Heath Jr., Marios Kountouris, and Jeffrey G. Andrews

Wireless Networking and Communications Group, Department of Electrical and Computer Engineering, The University of Texas at Austin, 1 University Station C0803, Austin, TX 78712-0240, USA

Correspondence should be addressed to Jun Zhang, jzhang06@mail.utexas.edu

Received 16 December 2008; Revised 12 March 2009; Accepted 23 April 2009

Recommended by Markus Rupp

Imperfect channel state information degrades the performance of multiple-input multiple-output (MIMO) communications; its effects on single-user (SU) and multiuser (MU) MIMO transmissions are quite different. In particular, MU-MIMO suffers from residual interuser interference due to imperfect channel state information while SU-MIMO only suffers from a power loss. This paper compares the throughput loss of both SU and MU-MIMO in the broadcast channel due to delay and channel quantization. Accurate closed-form approximations are derived for achievable rates for both SU and MU-MIMO. It is shown that SU-MIMO is relatively robust to delayed and quantized channel information, while MU-MIMO with zero-forcing precoding loses its spatial multiplexing gain with a fixed delay or fixed codebook size. Based on derived achievable rates, a mode switching algorithm is proposed, which switches between SU and MU-MIMO modes to improve the spectral efficiency based on average signal-to-noise ratio (SNR), normalized Doppler frequency, and the channel quantization codebook size. The operating regions for SU and MU modes with different delays and codebook sizes are determined, and they can be used to select the preferred mode. It is shown that the MU mode is active only when the normalized Doppler frequency is very small, and the codebook size is large.

Copyright © 2009 Jun Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. Introduction

Over the last decade, the point-to-point multiple-input multiple-output (MIMO) link (SU-MIMO) has been extensively researched and has transitioned from a theoretical concept to a practical technique [1, 2]. Due to space and complexity constraints, however, current mobile terminals only have one or two antennas, which limits the performance of the SU-MIMO link. Multiuser MIMO (MU-MIMO) provides the opportunity to overcome such a limitation by communicating with multiple mobiles simultaneously. It effectively increases the number of equivalent spatial channels and provides spatial multiplexing gain proportional to the number of transmit antennas at the base station even with single-antenna mobiles. In addition, MU-MIMO has higher immunity to propagation limitations faced by SU-MIMO, such as channel rank loss and antenna correlation [3].

There are many technical challenges that must be overcome to exploit the full benefits of MU-MIMO. A major

one is the requirement of channel state information at the transmitter (CSIT), which is difficult to get especially for the broadcast channel. For the multiantenna broadcast channel with N_t transmit antennas and N_r receive antennas, with full CSIT the sum throughput can grow linearly with N_t even when $N_r = 1$, but without CSIT the spatial multiplexing gain is the same as for SU-MIMO, that is, the throughput grows linearly with $\min(N_t, N_r)$ at high SNR [4]. Limited feedback is an efficient way to provide partial CSIT, which feeds back the quantized channel information to the transmitter via a low-rate feedback channel [5, 6]. However, such imperfect CSIT will degrade the throughput gain provided by MU-MIMO [7, 8]. Besides quantization, there are other imperfections in the available CSIT, such as estimation error and feedback delay. With imperfect CSIT, it is not clear whether—or more to the point, when—MU-MIMO can outperform SU-MIMO. In this paper, we compare SU and MU-MIMO transmissions in the multiantenna broadcast channel with CSI delay and channel quantization, and propose to switch between SU and MU-MIMO modes based on the

achievable rate of each technique with practical receiver assumptions. Note that “mode” in this paper refers to the single-user mode (SU-MIMO transmission) or multiuser mode (MU-MIMO transmission). This differs from use of the term in some related recent work (all for single user MIMO), for example switching between spatial multiplexing and diversity mode [9] or between different numbers of data streams per user [10–12]

1.1. Related Work. For the MIMO broadcast channel, CSIT is required to separate the spatial channels for different users. To obtain the full spatial multiplexing gain for MU-MIMO systems employing zero-forcing (ZF) or block-diagonalization (BD) precoding, it was shown in [7, 13] that the quantization codebook size for limited feedback needs to increase linearly with SNR (in dB) and the number of transmit antennas. Zero-forcing dirty-paper coding and channel inversion systems with limited feedback were investigated in [8], where a sum rate ceiling due to a fixed codebook size was derived for both schemes. In [14], it was shown that to exploit multiuser diversity for ZF, both channel direction and information about signal-to-interference-plus-noise ratio (SINR) must be fed back. In [15], it was shown that the feedback delay limits the performance of joint precoding and scheduling schemes for the MIMO broadcast channel at moderate levels of Doppler. More recently, a comprehensive study of the MIMO broadcast channel with ZF precoding was done in [16], which considered downlink training and explicit channel feedback and concluded that significant downlink throughput is achievable with efficient CSI feedback. For a compound MIMO broadcast channel, the information theoretic analysis in [17] showed that scaling the CSIT quality such that the CSIT error is dominated by the inverse of SNR is both necessary and sufficient to achieve the full spatial multiplexing gain.

Although previous studies show that the spatial multiplexing gain of MU-MIMO can be achieved with limited feedback, it requires the codebook size to increase with SNR and the number of transmit antennas. Even if such a requirement is satisfied, there is an inevitable rate loss due to quantization error, plus other CSIT imperfections such as estimation error and delay. In addition, most of prior work focused on the achievable spatial multiplexing gain, mainly based on the analysis of the rate loss due to imperfect CSIT, which is usually a loose bound [7, 13, 17]. Such analysis cannot accurately characterize the throughput loss, and no comparison with SU-MIMO has been made.

There are several related studies comparing space division multiple access (SDMA) and time division multiple access (TDMA) in the multiantenna broadcast channel with limited feedback and with a large number of users. TDMA and SDMA with different scalar feedback schemes for scheduling were compared in [18], which shows that SDMA outperforms TDMA as the number of users becomes large while TDMA outperforms SDMA at high SNR. TDMA and SDMA with opportunistic beamforming were compared in [19], which proposed to adapt the number of beams to the number of active users to improve the throughput. A distributed mode selection algorithm switching between TDMA

and SDMA was proposed in [20], where each user feeds back its preferred mode and the channel quality information.

1.2. Contributions. In this paper, we derive good approximations for the achievable throughput for both SU and MU-MIMO systems with fixed channel information accuracy, that is, with a fixed delay and a fixed quantization codebook size. We are interested in the following question: *With imperfect CSIT, including delay and channel quantization, when can MU-MIMO actually deliver a throughput gain over SU-MIMO?* Based on this, we can select the one with the higher throughput as the transmission technique. The main contributions of this paper are as follows.

(i) *SU versus MU Analysis.* We investigate the impact of imperfect CSIT due to delay and channel quantization. We show that the SU mode is more robust to imperfect CSIT as it only suffers a constant rate loss, while MU-MIMO suffers more severely from residual inter-user interference. We characterize the residual interference due to delay and channel quantization, which shows that these two effects are equivalent. Based on an independence approximation of the interference terms and the signal term, accurate closed-form approximations are derived for ergodic achievable rates for both SU and MU-MIMO modes.

(ii) *Mode Switching Algorithm.* An SU/MU mode switching algorithm is proposed based on the ergodic sum rate as a function of average SNR, normalized Doppler frequency, and the quantization codebook size. This transmission technique only requires a small number of users to feed-back instantaneous channel information. The mode switching points can be calculated from the previously derived approximations for ergodic rates.

(iii) *Operating Regions.* Operating regions for SU and MU modes are determined, from which we can determine the active mode and find the condition that activates each mode. With a fixed delay and codebook size, if the MU mode is possible at all, there are two mode switching points, with the SU mode preferred at both low and high SNRs. The MU mode will only be activated when the normalized Doppler frequency is very small and the codebook size is large. From the numerical results, the minimum feedback bits per user to get the MU mode activated grow approximately linearly with the number of transmit antennas.

The rest of the paper is organized as follows. The system model and some assumptions are presented in Section 2. The transmission techniques for both SU and MU-MIMO modes are described in Section 3. The rate analysis for both SU and MU modes and the mode switching are done in Section 4. Numerical results and conclusions are in Sections 5 and 6, respectively. In this paper, we use uppercase boldface letters for matrices (\mathbf{X}) and lowercase boldface for vectors (\mathbf{x}). $\mathbb{E}[\cdot]$ is the expectation operator. The conjugate transpose of a matrix \mathbf{X} (vector \mathbf{x}) is \mathbf{X}^* (\mathbf{x}^*). Similarly, \mathbf{X}^\dagger denotes the pseudo-inverse, $\tilde{\mathbf{x}}$ denotes the normalized vector of \mathbf{x} , i.e. $\tilde{\mathbf{x}} = \mathbf{x}/\|\mathbf{x}\|$, and $\hat{\mathbf{x}}$ denotes the quantized vector of $\tilde{\mathbf{x}}$.

2. System Model

We consider a multiantenna broadcast channel, where the transmitter (the base station) has N_t antennas and each mobile user has a single antenna. The system parameters are listed in Table 1. During each transmission period, which is less than the channel coherence time and the channel is assumed to be constant, the base station transmits to one (SU-MIMO mode) or multiple (MU-MIMO mode) users. For the MU-MIMO mode, we assume that the number of active users is $U = N_t$, and the users are scheduled independently of their channel conditions, for example, through round-robin scheduling, random user selection, or scheduling based on the queue length. The discrete-time complex baseband received signal at the u th user at time n is given as

$$y_u[n] = \mathbf{h}_u^*[n] \sum_{u'=1}^U \mathbf{f}_{u'}[n] x_{u'}[n] + z_u[n], \quad (1)$$

where $\mathbf{h}_u[n]$ is the $N_t \times 1$ channel vector from the transmitter to the u th user, and $z_u[n]$ is the normalized complex Gaussian noise vector, that is, $z_u[n] \sim \mathcal{CN}(0, 1)$. $x_u[n]$ and $\mathbf{f}_u[n]$ are the transmit signal and the normalized $N_t \times 1$ precoding vector for the u th user, respectively. The transmit power constraint is $\mathbb{E}\{\mathbf{x}^*[n]\mathbf{x}[n]\} = P$, where $\mathbf{x}[n] = [x_1^*, x_2^*, \dots, x_{N_t}^*]^*$. As the noise is normalized, P is also the average transmit SNR. To assist the analysis, we assume that the channel $\mathbf{h}_u[n]$ is well modeled as a spatially white Gaussian channel, with entries $h_{i,j}[n] \sim \mathcal{CN}(0, 1)$, and the channels are i.i.d. over different users. Note that in the case of line of sight MIMO channel, fewer feedback bits are required compared to the Rayleigh channel [21].

We consider two of the main sources of the CSIT imperfection-delay and quantization error, specified as follows. For a practical system, the feedback bits for each user is usually fixed, and there will inevitably be delay in the available CSI, both of which are difficult or even impossible to adjust. Other effects such as channel estimation error can be made small such as by increasing the transmit power or the number of pilot symbols.

2.1. CSI Delay Model. We consider a stationary ergodic Gauss-Markov block fading process [22, Section 16.1], where the channel stays constant for a symbol duration and changes from symbol to symbol according to

$$\mathbf{h}[n] = \rho \mathbf{h}[n-1] + \mathbf{e}[n], \quad (2)$$

where $\mathbf{e}[n]$ is the channel error vector, with i.i.d. entries $e_i[n] \sim \mathcal{CN}(0, \epsilon_e^2)$, and it is uncorrelated with $\mathbf{h}[n-1]$. We assume that the CSI delay is of one symbol. It is straightforward to extend the results to the scenario with a delay of multiple symbols. For the numerical analysis, the classical Clarke's isotropic scattering model will be used as an example, for which the correlation coefficient is $\rho = J_0(2\pi f_d T_s)$ with Doppler spread f_d [23], where $J_0(\cdot)$ is the zeroth-order Bessel function of the first kind. The variance of the error vector is $\epsilon_e^2 = 1 - \rho^2$. Therefore, both ρ and ϵ_e are determined by the normalized Doppler frequency $f_d T_s$.

TABLE 1: System parameters.

Symbol	Description
N_t	Number of transmit antennas
U	Number of mobile users
B	Number of feedback bits
L	Quantization codebook size, $L = 2^B$
P	Average SNR
n	Time index
T_s	The length of each symbol
f_d	The Doppler frequency

The channel in (2) is widely used to model the time-varying channel. For example, it is used to investigate the impact of feedback delay on the performance of closed-loop transmit diversity in [24] and the system capacity and bit error rate of point-to-point MIMO link in [25]. It simplifies the analysis, and the results can be easily extended to other scenarios with the channel model of the form

$$\mathbf{h}[n] = \mathbf{g}[n] + \mathbf{e}[n], \quad (3)$$

where $\mathbf{g}[n]$ is the available CSI at time n with an uncorrelated error vector $\mathbf{e}[n]$, $\mathbf{g}[n] \sim \mathcal{CN}(\mathbf{0}, (1 - \epsilon_e^2)\mathbf{I})$, and $\mathbf{e}[n] \sim \mathcal{CN}(\mathbf{0}, \epsilon_e^2\mathbf{I})$. It can be used to consider the effect of other imperfect CSITs, such as estimation error and analog feedback. The difference is in $\mathbf{e}[n]$, which has different variance ϵ_e^2 for different scenarios. Some examples are given as follows.

(a) *Estimation Error.* If the receiver obtains the CSI through minimum mean-squared error (MMSE) estimation from τ_p pilot symbols, the error variance is $\epsilon_e^2 = 1/(1 + \tau_p \gamma_p)$, where γ_p is the SNR of the pilot symbol [16].

(b) *Analog Feedback.* For analog feedback, the error variance is $\epsilon_e^2 = 1/(1 + \tau_{ul} \gamma_{ul})$, where τ_{ul} is the number of channel uses per channel coefficient and γ_{ul} is the SNR on the uplink feedback channel [26].

(c) *Analog Feedback with Prediction.* As shown in [27], for analog feedback with a d -step MMSE predictor and the Gauss-Markov model, the error variance is $\epsilon_e^2 = \rho^{2d} \epsilon_0 + (1 - \rho^2) \sum_{l=0}^{d-1} \rho^{2l}$, where ρ is the same as in (2) and ϵ_0 is the Kalman filtering mean-square error.

Therefore, the results in this paper can be easily extended to these systems. In the following parts, we focus on the effect of CSI delay.

2.2. Channel Quantization Model. We consider frequency-division duplexing (FDD) systems, where limited feedback techniques provide partial CSIT through a dedicated feedback channel from the receiver to the transmitter. The channel direction information for the precoder design is fed back using a quantization codebook known at both the transmitter and receiver. The quantization is chosen from a codebook of unit norm vectors of size $L = 2^B$. We

assume that each user uses a different codebook to avoid the same quantization vector. The codebook for user u is $\mathcal{C}_u = \{\mathbf{c}_{u,1}, \mathbf{c}_{u,2}, \dots, \mathbf{c}_{u,L}\}$. Each user quantizes its channel to the closest codeword, where closeness is measured by the inner product. Therefore, the index of channel for user u is

$$I_u = \arg \max_{1 \leq \ell \leq L} \left| \tilde{\mathbf{h}}_u^* \mathbf{c}_{u,\ell} \right|. \quad (4)$$

Each user needs to feed-back B bits to denote this index, and the transmitter has the quantized channel information $\hat{\mathbf{h}}_u = \mathbf{c}_{u,I_u}$. As the optimal vector quantizer for this problem is not known in general, random vector quantization (RVQ) [28] is used, where each quantization vector is independently chosen from the isotropic distribution on the N_t -dimensional unit sphere. It has been shown in [7] that RVQ can facilitate the analysis and provide performance close to the optimal quantization. In this paper, we analyze the achievable rate averaged over both RVQ-based random codebooks and fading distributions.

An important metric for the limited feedback system is the squared angular distortion, defined as $\sin^2(\theta_u) = 1 - |\tilde{\mathbf{h}}_u^* \hat{\mathbf{h}}_u|^2$, where $\theta_u = \angle(\tilde{\mathbf{h}}_u, \hat{\mathbf{h}}_u)$. With RVQ, it was shown in [7, 29] that the expectation in i.i.d. Rayleigh fading is given by

$$\mathbb{E}[\sin^2(\theta_u)] = 2^B \cdot \beta\left(2^B, \frac{N_t}{N_t - 1}\right), \quad (5)$$

where $\beta(\cdot)$ is the beta function [30]. It can be tightly bounded as [7]

$$\frac{N_t - 1}{N_t} 2^{-B/(N_t - 1)} \leq \mathbb{E}[\sin^2(\theta_u)] \leq 2^{-B/(N_t - 1)}. \quad (6)$$

3. Transmission Techniques

In this section, we describe the transmission techniques for both SU and MU-MIMO systems with perfect CSIT, which will be used in the subsequent sections for imperfect CSIT systems. By doing this, we focus on the impacts of imperfect CSIT on the conventional transmission techniques. Throughout this paper, we use the achievable ergodic rate as the performance metric for both SU and MU-MIMO systems. The base station transmits to a single user ($U = 1$) for the SU-MIMO system and to N_t users ($U = N_t$) for the MU-MIMO system. The SU/MU mode switching algorithm is also described.

3.1. SU-MIMO System. With perfect CSIT, it is optimal for the SU-MIMO system to transmit along the channel direction [1], that is, selecting the beamforming (BF) vector as $\mathbf{f}[n] = \tilde{\mathbf{h}}[n]$, denoted as *eigen-beamforming* in this paper. The ergodic capacity of this system is the same as that of a maximal ratio combining diversity system, given by [31]

$$\begin{aligned} R_{\text{BF}}(P) &= \mathbb{E}_{\mathbf{h}} \left[\log_2(1 + P \|\mathbf{h}[n]\|^2) \right] \\ &= \log_2(e) e^{1/P} \sum_{k=0}^{N_t-1} \frac{\Gamma(-k, 1/P)}{P^k}, \end{aligned} \quad (7)$$

where $\Gamma(\cdot, \cdot)$ is the complementary incomplete gamma function defined as $\Gamma(\alpha, x) = \int_x^\infty t^{\alpha-1} e^{-t} dt$.

3.2. MU-MIMO System. For multiantenna broadcast channels, although dirty-paper coding (DPC) [32] is optimal [33–37], it is difficult to implement in practice. As in [7, 16], ZF precoding is used in this paper, which is a linear precoding technique that precancels inter-user interference at the transmitter. There are several reasons for us to use this simple transmission technique. Firstly, due to its simple structure, it is possible to derive closed-form results, which can provide helpful insights. Second, the ZF precoding is able to provide full spatial multiplexing gain and only has a power offset compared to the optimal DPC system [38]. In addition, it was shown in [38] that the ZF precoding is optimal among the set of all linear precoders at asymptotically high SNR. In Section 5, we will show that our results for the ZF system also apply for the regularized ZF precoding (aka MMSE precoding) [39], which provides a higher throughput than the ZF precoding at low to moderate SNRs.

With precoding vectors $\mathbf{f}_u[n]$, $u = 1, 2, \dots, U$, assuming equal power allocation, the received SINR for the u th user is given as

$$\gamma_{\text{ZF},u} = \frac{(P/U) |\mathbf{h}_u^*[n] \mathbf{f}_u[n]|^2}{1 + (P/U) \sum_{u' \neq u} |\mathbf{h}_u^*[n] \mathbf{f}_{u'}[n]|^2}. \quad (8)$$

This is true for a general linear precoding MU-MIMO system. With perfect CSIT, this quantity can be calculated at the transmitter, while with imperfect CSIT, it can be estimated at the receiver and fed back to the transmitter given knowledge of $\mathbf{f}_u[n]$. At high SNR, equal power allocation performs closely to the system employing optimal water-filling, as power allocation mainly benefits at low SNR.

Denote $\tilde{\mathbf{H}}[n] = [\tilde{\mathbf{h}}_1[n], \tilde{\mathbf{h}}_2[n], \dots, \tilde{\mathbf{h}}_U[n]]^*$. With perfect CSIT, the ZF precoding vectors are determined from the pseudoinverse of $\tilde{\mathbf{H}}[n]$, as $\mathbf{F}[n] = \tilde{\mathbf{H}}^\dagger[n] = \tilde{\mathbf{H}}^*[n] (\tilde{\mathbf{H}}[n] \tilde{\mathbf{H}}^*[n])^{-1}$. The precoding vector for the u th user is obtained by normalizing the u th column of $\mathbf{F}[n]$. Therefore, $\mathbf{h}_u^*[n] \mathbf{f}_{u'}[n] = 0$, $\forall u \neq u'$, that is, there is no inter-user interference. The received SINR for the u th user becomes

$$\gamma_{\text{ZF},u} = \frac{P}{U} |\mathbf{h}_u^*[n] \mathbf{f}_u[n]|^2. \quad (9)$$

As $\mathbf{f}_u[n]$ is independent of $\mathbf{h}_u[n]$, and $\|\mathbf{f}_u[n]\|^2 = 1$, the effective channel for the u th user is a single-input single-output (SISO) Rayleigh fading channel. Therefore, the achievable sum rate for the ZF system is given by

$$R_{\text{ZF}}(P) = \sum_{u=1}^U \mathbb{E}_\gamma \left[\log_2(1 + \gamma_{\text{ZF},u}) \right]. \quad (10)$$

Each term on the right-hand side of (10) is the ergodic capacity of an SISO system in Rayleigh fading, given in [31] as

$$\begin{aligned} R_{\text{ZF},u} &= \mathbb{E}_\gamma \left[\log_2(1 + \gamma_{\text{ZF},u}) \right] \\ &= \log_2(e) e^{U/P} E_1 \left(\frac{U}{P} \right), \end{aligned} \quad (11)$$

where $E_1(\cdot)$ is the exponential-integral function of the first order, $E_1(x) = \int_1^\infty (e^{-xt}/t)dt$.

3.3. SU/MU Mode Switching. Imperfect CSIT will degrade the performance of the MIMO communication. In this case, it is unclear whether and when the MU-MIMO system can actually provide a throughput gain over the SU-MIMO system. Based on the analysis of the achievable ergodic rates in this paper, we propose to switch between SU and MU modes and select the one with the higher achievable rate.

The channel correlation coefficient ρ , which captures the CSI delay effect, usually varies slowly. The quantization codebook size is normally fixed for a given system. Therefore, it is reasonable to assume that the transmitter has knowledge of both delay and channel quantization, and can estimate the achievable ergodic rates of both SU and MU-MIMO modes. Then it can determine the active mode and select one (SU mode) or N_t (MU mode) users to serve. This is a low-complexity transmission strategy, and can be combined with random user selection, round-robin scheduling, or scheduling based on queue length rather than channel status. It only requires the selected users to feed-back instantaneous channel information. Therefore, it is suitable for a system that has a constraint on the total feedback bits and only allows a small number of users to send feedback, or a system with a strict delay constraint that cannot employ opportunistic scheduling based on instantaneous channel information.

To determine the transmission rate, the transmitter sends pilot symbols, from which the active users estimate the received SINRs and feed-back them to the transmitter. In this paper, we assume that the transmitter knows perfectly the actual received SINR at each active user, and so there will be no outage in the transmission.

4. SU versus MU with Delayed and Quantized CSIT

In this section, we investigate the achievable ergodic rates for both SU and MU-MIMO modes. We first analyze the average received SNR for the BF system and the average residual interference for the ZF system, which provide insights on the impact of imperfect CSIT. To select the active mode, accurate closed-form approximations for achievable rates of both SU and MU modes are then derived.

4.1. SU Mode: Eigen-Beamforming. First, if there is no delay and only channel quantization, the BF vector is based on the quantized feedback, $\mathbf{f}^{(Q)}[n] = \hat{\mathbf{h}}[n]$. The average received SNR is

$$\begin{aligned} \overline{\text{SNR}}_{\text{BF}}^{(Q)} &= \mathbb{E}_{\mathbf{h}, \mathbf{e}} \left[P \left| \mathbf{h}^* [n] \hat{\mathbf{h}}[n] \right|^2 \right] \\ &= \mathbb{E}_{\mathbf{h}, \mathbf{e}} \left[P \|\mathbf{h}[n]\|^2 \left| \tilde{\mathbf{h}}^* [n] \hat{\mathbf{h}}[n] \right|^2 \right] \\ &\stackrel{(a)}{\leq} P N_t \left(1 - \frac{N_t - 1}{N_t} 2^{-B/(N_t - 1)} \right), \end{aligned} \quad (12)$$

where (a) follows by the independence between $\|\mathbf{h}[n]\|^2$ and $|\tilde{\mathbf{h}}^* [n] \hat{\mathbf{h}}[n]|^2$, together with the result in (6).

With both delay and channel quantization, the BF vector is based on the quantized channel direction with delay, that is, $\mathbf{f}^{(QD)}[n] = \hat{\mathbf{h}}[n - 1]$. The instantaneous received SNR for the BF system

$$\text{SNR}_{\text{BF}}^{(QD)} = P \left| \mathbf{h}^* [n] \mathbf{f}^{(QD)}[n] \right|^2. \quad (13)$$

Based on (12), we get the following theorem on the average received SNR for the SU mode.

Theorem 1. *The average received SNR for a BF system with channel quantization and CSI delay is*

$$\overline{\text{SNR}}_{\text{BF}}^{(QD)} \leq P N_t \left(\rho^2 \Delta_{\text{BF}}^{(Q)} + \Delta_{\text{BF}}^{(D)} \right), \quad (14)$$

where $\Delta_{\text{BF}}^{(Q)}$ and $\Delta_{\text{BF}}^{(D)}$ show the impact of channel quantization and feedback delay, respectively, given by

$$\Delta_{\text{BF}}^{(Q)} = 1 - \frac{N_t - 1}{N_t} 2^{-B/(N_t - 1)}, \quad \Delta_{\text{BF}}^{(D)} = \frac{\epsilon_e^2}{N_t}. \quad (15)$$

Proof. See Appendix B. \square

From Jensen's inequality, an upper bound of the achievable rate for the BF system with both quantization and delay is given by

$$\begin{aligned} R_{\text{BF}}^{(QD)} &= \mathbb{E}_{\mathbf{h}, \mathbf{e}} \left[\log_2 \left(1 + \text{SNR}_{\text{BF}}^{(QD)} \right) \right] \\ &\leq \log_2 \left[1 + \overline{\text{SNR}}_{\text{BF}}^{(QD)} \right] \\ &\leq \log_2 \left[1 + P N_t \left(\rho^2 \Delta_{\text{BF}}^{(Q)} + \Delta_{\text{BF}}^{(D)} \right) \right]. \end{aligned} \quad (16)$$

Remark 1. Note that $\rho^2 = 1 - \epsilon_e^2$, so the average SNR decreases with ϵ_e^2 . With a fixed B and fixed delay, the SNR degradation is a constant factor independent of P . At high SNR, the imperfect CSIT introduces a constant rate loss $\log_2(\rho^2 \Delta_{\text{BF}}^{(Q)} + \Delta_{\text{BF}}^{(D)})$.

The upper bound provided by Jensen's inequality is not tight. To get a better approximation for the achievable rate, we first make the following approximation on the instantaneous received SNR

$$\begin{aligned} \text{SNR}_{\text{BF}}^{(QD)} &= P \left| \mathbf{h}^* [n] \hat{\mathbf{h}}[n - 1] \right|^2 \\ &= P \left| (\rho \mathbf{h}[n - 1] + \mathbf{e}[n])^* \hat{\mathbf{h}}[n - 1] \right|^2 \\ &\approx P \rho^2 \left| \mathbf{h}^* [n - 1] \hat{\mathbf{h}}[n - 1] \right|^2, \end{aligned} \quad (17)$$

that is, we remove the term with $\mathbf{e}[n]$ as it is normally insignificant compared to $\rho \mathbf{h}[n - 1]$. This will be verified later by simulation. In this way, the system is approximated as the one with limited feedback and with equivalent SNR $\rho^2 P$.

From [29], the achievable rate of the limited feedback BF system is given by

$$R_{\text{BF}}^{(Q)}(P) = \log_2(e) \left(e^{1/P} \sum_{k=0}^{N_t-1} E_{k+1} \left(\frac{1}{P} \right) - \int_0^1 (1-(1-x)^{N_t-1})^{2B} \frac{N_t}{x} e^{1/Px} E_{N_t+1} \left(\frac{1}{Px} \right) dx \right), \quad (18)$$

where $E_n(x) = \int_1^\infty e^{-xt} x^{-n} dt$ is the n th order exponential integral. So $R_{\text{BF}}^{(QD)}$ can be approximated as

$$R_{\text{BF}}^{(QD)}(P) \approx R_{\text{BF}}^{(Q)}(\rho^2 P). \quad (19)$$

As a special case, considering a system with delay only, for example, the time-division duplexing (TDD) system which can estimate the CSI from the uplink with channel reciprocity but with propagation and processing delay, the BF vector is based on the delayed channel direction, that is, $\mathbf{f}^{(D)}[n] = \tilde{\mathbf{h}}[n-1]$. We provide a good approximation for the achievable rate for such a system as follows.

The instantaneous received SNR is given as

$$\begin{aligned} \text{SNR}_{\text{BF}}^{(D)} &= P \left| \mathbf{h}^*[n] \mathbf{f}^{(D)}[n] \right|^2 \\ &= P \left| (\rho \mathbf{h}[n-1] + \mathbf{e}[n])^* \tilde{\mathbf{h}}[n-1] \right|^2 \\ &\stackrel{(a)}{\approx} P \rho^2 \|\mathbf{h}[n-1]\|^2 + P \left| \mathbf{e}^*[n] \tilde{\mathbf{h}}[n-1] \right|^2. \end{aligned} \quad (20)$$

In step (a) we eliminate the cross terms since $\mathbf{e}[n]$ is normally small, for example, its variance is $\epsilon_e^2 = 0.027$ with carrier frequency at 2 GHz, mobility of 20 km/hr and delay of 1 msec. As $\mathbf{e}[n]$ is independent of $\tilde{\mathbf{h}}[n-1]$, $\mathbf{e}[n] \sim \mathcal{CN}(\mathbf{0}, \epsilon_e^2 \mathbf{I})$ and $\|\tilde{\mathbf{h}}[n-1]\|^2 = 1$, we have $|\mathbf{e}^*[n] \tilde{\mathbf{h}}[n-1]|^2 \sim \chi_2^2$, where χ_M^2 denotes chi-square distribution with M degrees of freedom. In addition, $\|\mathbf{h}[n-1]\|^2 \sim \chi_{2N_t}^2$, and it is independent of $|\mathbf{e}^*[n] \tilde{\mathbf{h}}[n-1]|^2$. Then the following theorem can be derived.

Theorem 2. *The achievable ergodic rate of the BF system with delay can be approximated as*

$$\begin{aligned} R_{\text{BF}}^{(D)} &\approx \log_2(e) a_0^{N_t} e^{1/\eta_2} E_1 \left(\frac{1}{\eta_2} \right) \\ &\quad - \log_2(e) (1 - a_0) \sum_{i=0}^{N_t-1} \frac{a_0^{N_t-1-i}}{(i-l)!} \eta_1^{-(i-l)} I_1 \left(\frac{1}{\eta_1}, 1, i-l \right), \end{aligned} \quad (21)$$

where $\eta_1 = P\rho^2$, $\eta_2 = P\epsilon_e^2$, $a_0 = \eta_2/(\eta_2 - \eta_1)$, and $I_1(\cdot, \cdot, \cdot)$ is given in (A.3) in Appendix A.

Proof. See Appendix C. \square

4.2. MU Mode: Zero-Forcing

4.2.1. Average Residual Interference. If there is no delay but only channel quantization, the precoding vectors for the ZF system are designed based on $\hat{\mathbf{h}}_1[n], \hat{\mathbf{h}}_2[n], \dots, \hat{\mathbf{h}}_U[n]$ to achieve $\hat{\mathbf{h}}_u^*[n] \mathbf{f}_{u'}^{(Q)}[n] = 0, \forall u \neq u'$. With random vector quantization, it is shown in [7] that the average noise plus interference for each user is

$$\begin{aligned} \Delta_{\text{ZF},u}^{(Q)} &= \mathbb{E}_{\mathbf{h}, \mathbf{c}} \left[1 + \frac{P}{U} \sum_{u' \neq u} \left| \mathbf{h}_u^*[n] \mathbf{f}_{u'}^{(Q)}[n] \right|^2 \right] \\ &= 1 + 2^{-B/(N_t-1)} P. \end{aligned} \quad (22)$$

With both channel quantization and CSI delay, precoding vectors are designed based on $\hat{\mathbf{h}}_1[n-1], \hat{\mathbf{h}}_2[n-1], \dots, \hat{\mathbf{h}}_U[n-1]$ and achieve $\hat{\mathbf{h}}_u^*[n-1] \mathbf{f}_{u'}^{(QD)}[n] = 0, \forall u \neq u'$. The received SINR for the u th user is given as

$$\gamma_{\text{ZF},u}^{(QD)} = \frac{(P/U) \left| \mathbf{h}_u^*[n] \mathbf{f}_u^{(QD)}[n] \right|^2}{1 + (P/U) \sum_{u' \neq u} \left| \mathbf{h}_u^*[n] \mathbf{f}_{u'}^{(QD)}[n] \right|^2}. \quad (23)$$

As $\mathbf{f}_u^{(QD)}[n]$ is in the nullspace of $\hat{\mathbf{h}}_{u'}[n-1] \forall u' \neq u$, it is isotropically distributed in \mathbb{C}^{N_t} and independent of $\tilde{\mathbf{h}}_u[n-1]$ as well as $\tilde{\mathbf{h}}_u[n]$, so $|\mathbf{h}_u^*[n] \mathbf{f}_u^{(QD)}[n]|^2 \sim \chi_2^2$. The average noise plus interference is given in the following theorem.

Theorem 3. *The average noise plus interference for the u th user of the ZF system with both channel quantization and CSI delay is*

$$\Delta_{\text{ZF},u}^{(QD)} = 1 + (U-1) \frac{P}{U} (\rho_u^2 \Delta_{\text{ZF},u}^{(Q)} + \Delta_{\text{ZF},u}^{(D)}), \quad (24)$$

where $\Delta_{\text{ZF},u}^{(Q)}$ and $\Delta_{\text{ZF},u}^{(D)}$ are the degradations brought by channel quantization and feedback delay, respectively, given by

$$\Delta_{\text{ZF},u}^{(Q)} = \frac{U}{U-1} 2^{-B/(N_t-1)}, \quad \Delta_{\text{ZF},u}^{(D)} = \epsilon_{e,u}^2. \quad (25)$$

Proof. The proof is similar to the one for Theorem 1 in Appendix B. \square

Remark 2. From Theorem 3 we see that the average residual interference for a given user consists of three parts.

- (i) *The number of interferers, $U-1$.* The more users the system supports, the higher the mutual interference.
- (ii) *The transmit power of the other active users, P/U .* As the transmit power increases, the system becomes interference-limited.
- (iii) *The CSIT accuracy for this user,* which is reflected from $\rho_u^2 \Delta_{\text{ZF},u}^{(Q)} + \Delta_{\text{ZF},u}^{(D)}$. The user with a larger delay or a smaller codebook size suffers a higher residual interference.

From this remark, the residual inter-user interference equivalently comes from $U-1$ virtual interfering users,

each with equivalent SNR as $(P/U)(\rho_u^2 \Delta_{ZF,u}^{(Q)} + \Delta_{ZF,u}^{(D)})$. With a high P and a fixed $\epsilon_{e,u}$ or B , the system is interference-limited and cannot achieve the full spatial multiplexing gain. Therefore, to keep a constant rate loss, that is, to sustain the spatial multiplexing gain, the channel error due to both quantization and delay needs to be reduced as SNR increases. Similar to the result for the limited feedback system in [7], for the ZF system with both delay and channel quantization, we can get the following corollary for the condition to achieve the full spatial multiplexing gain.

Corollary 1. *To keep a constant rate loss of $\log_2 \delta_0$ bps/Hz for each user, the codebook size and CSI delay need to satisfy the following condition:*

$$\rho_u^2 \Delta_{ZF,u}^{(Q)} + \Delta_{ZF,u}^{(D)} = \frac{U}{U-1} \cdot \frac{\delta_0 - 1}{P}. \quad (26)$$

Proof. As shown in [7, 16], the rate loss for each user due to imperfect CSIT is upper bounded by $\Delta R_u \leq \log_2 \Delta_{ZF,u}^{(QD)}$. The corollary follows from solving $\log_2 \Delta_{ZF,u}^{(QD)} = \log_2 \delta_0$. \square

Equivalently, this means that for a given ρ^2 , the feedback bits per user needs to scale as

$$B = (N_t - 1) \log_2 \left(\frac{\delta_0 - 1}{\rho_u^2 P} - \frac{U-1}{U} \cdot \left(\frac{1}{\rho_u^2} - 1 \right) \right)^{-1}. \quad (27)$$

As $\rho_u^2 \rightarrow 1$, that is, there is no CSI delay, the condition becomes $B = (N_t - 1) \log_2 (P/(\delta_0 - 1))$, which agrees with the result in [7] with limited feedback only.

4.2.2. Achievable Rate. For the ZF system with imperfect CSI, the genie-aided upper bound for the ergodic achievable rate is given by [16]

$$R_{ZF}^{(QD)} \leq \sum_{u=1}^U \mathbb{E}_\gamma \left[\log_2 \left(1 + \gamma_{ZF,u}^{(QD)} \right) \right] = R_{ZF,ub}^{(QD)}. \quad (28)$$

This upper bound is achievable only when a genie provides users with perfect knowledge of all interference and the transmitter knows perfectly the received SINR at each user. We assume that the mobile users can perfectly estimate the noise and interference and feed-back it to the transmitter, and so the upper bound is chosen as the performance metric, that is, $R_{ZF}^{(QD)} = R_{ZF,ub}^{(QD)}$, as in [7, 8, 14].

The following lower bound based on the rate loss analysis is used in [7, 16]:

$$R_{ZF}^{(QD)} \geq R_{ZF} - \sum_{u=1}^U \log_2 \Delta_{ZF,u}^{(QD)}, \quad (29)$$

where R_{ZF} is the achievable rate with perfect CSIT, given in (10). However, this lower bound is very loose. In the following, we will derive a more accurate approximation for the achievable rate for the ZF system.

To get a good approximation for the achievable rate for the ZF system, we first approximate the instantaneous SINR as

$$\begin{aligned} \gamma_{ZF,u}^{(QD)} &= \frac{(P/U) \left| \mathbf{h}_u^* [n] \mathbf{f}_u^{(QD)} [n] \right|^2}{1 + (P/U) \sum_{u' \neq u} \left| (\rho_u \mathbf{h}_u [n-1] + \mathbf{e}_u [n])^* \mathbf{f}_{u'}^{(QD)} [n] \right|^2} \\ &\approx \frac{(P/U) \left| \mathbf{h}_u^* [n] \mathbf{f}_u^{(QD)} [n] \right|^2}{1 + (P/U) (I^{(Q)} + I^{(D)})}, \end{aligned} \quad (30)$$

where $I^{(Q)} = \sum_{u' \neq u} \rho_u^2 \left| \mathbf{h}_u^* [n-1] \mathbf{f}_{u'}^{(QD)} [n] \right|^2$ and $I^{(D)} = \sum_{u' \neq u} \left| \mathbf{e}_u^* [n] \mathbf{f}_{u'}^{(QD)} [n] \right|^2$ are interference due to channel quantization and delay, respectively. Essentially, we eliminate interference terms which have both $\mathbf{h}_u [n-1]$ and $\mathbf{e}_u [n]$ as $\mathbf{e}_u [n]$ is normally very small.

For the interference term due to delay, $\left| \mathbf{e}_u^* [n] \mathbf{f}_u^{(QD)} [n] \right|^2 \sim \chi_2^2$, as $\mathbf{e}_u [n]$ is independent of $\mathbf{f}_u^{(QD)} [n]$ and $\|\mathbf{f}_u^{(QD)} [n]\|^2 = 1$. For the interference term due to quantization, it was shown in [7] that $\left| \tilde{\mathbf{h}}_u^* [n-1] \mathbf{f}_u^{(QD)} [n] \right|^2$ is equivalent to the product of the quantization error $\sin^2 \theta_u$ and an independent $\beta(1, N_t - 2)$ random variable. Therefore, we have

$$\begin{aligned} &\left| \mathbf{h}_u^* [n-1] \mathbf{f}_u^{(QD)} [n] \right|^2 \\ &= \|\mathbf{h}_u [n-1]\|^2 (\sin^2 \theta_u) \cdot \beta(1, N_t - 2). \end{aligned} \quad (31)$$

In [14], with a quantization cell approximation [40, 41], the quantization cell approximation is based on the ideal assumption that each quantization cell is a Voronoi region on a spherical cap with the surface area 2^{-B} of the total area of the unit sphere for a B bits codebook. The detail can be found in [14, 40, 41], it was shown that $\|\mathbf{h}_u [n-1]\|^2 (\sin^2 \theta_u)$ has a Gamma distribution with parameters $(N_t - 1, \delta)$, where $\delta = 2^{-B/(N_t - 1)}$. As shown in [14] the analysis based on the quantization cell approximation is close to the performance of random vector quantization, and so we use this approach to derive the achievable rate.

The following lemma gives the distribution of the interference term due to quantization.

Lemma 1. *Based on the quantization cell approximation, the interference term due to quantization in (30), $\left| \mathbf{h}_u [n-1] \mathbf{f}_u^{(QD)} [n] \right|^2$, is an exponential random variable with mean δ , that is, its probability distribution function (pdf) is*

$$p(x) = \frac{1}{\delta} e^{-x/\delta}, \quad x \geq 0. \quad (32)$$

Proof. See Appendix D. \square

Remark 3. From this lemma, we see that the residual interference terms due to both delay and quantization are exponential random variables, which means that the delay and quantization error have equivalent effects, only with different means. By comparing the means of these two

terms, that is, comparing ϵ_e^2 and $2^{-B/(N_t-1)}$, we can find the dominant one. In addition, with this result, we can approximate the achievable rate of the ZF-limited feedback system, which will be provided later in this section.

Based on the distribution of the interference terms, the approximation for the achievable rate for the MU mode is given in the following theorem.

Theorem 4. *The ergodic achievable rate for the u th user in the MU mode with both delay and channel quantization can be approximated as*

$$R_{ZF,u}^{(QD)} \approx \log_2(e) \sum_{i=0}^{M-1} \sum_{j=1}^2 \left[a_i^{(j)} i! \cdot I_3\left(\frac{1}{\alpha}, \frac{1}{\delta_j}, i+1\right) \right], \quad (33)$$

where $\alpha = P/U$, $\delta_1 = \rho_u^2 \delta$, $\delta_2 = \epsilon_{e,u}^2$, $M = N_t - 1$, $a_i^{(1)}$ and $a_i^{(2)}$ are given in (E.3), and $I_3(\cdot, \cdot, \cdot)$ is given in (A.5) in Appendix A.

Proof. See Appendix E. \square

The ergodic sum throughput is

$$R_{ZF}^{(QD)} = \sum_{u=1}^U R_{ZF,u}^{(QD)}. \quad (34)$$

As a special case, for a ZF system with delay only, we can get the following approximation for the ergodic achievable rate.

Corollary 2. *The ergodic achievable rate for the u th user in the ZF system with delay is approximated as*

$$R_{ZF,u}^{(D)} \approx \log_2(e) \epsilon_{e,u}^{2(M-1)} \cdot I_3\left(\frac{1}{\alpha}, \frac{1}{\epsilon_{e,u}^2}, M-1\right), \quad (35)$$

where $\alpha = P/U$, $M = N_t - 1$, and $I_3(\cdot, \cdot, \cdot)$ is given in (A.5) in Appendix A.

Proof. Following the same steps in Appendix E with $\delta_1 = 0$. \square

Remark 4. As shown in Lemma 1, the effects of delay and channel quantization are equivalent, and so the approximation in (35) also applies for the limited feedback system. This is verified by simulation in Figure 1, which shows that this approximation is very accurate and can be used to analyze the limited feedback system.

4.3. Mode Switching. We first verify the approximation (33) in Figure 2, which compares the approximation with simulation results and the lower bound (29), with $B = 10$ bits, $v = 20$ km/hr, $f_c = 2$ GHz, and $T_s = 1$ msec. We see that the lower bound is very loose, while the approximation is accurate especially for $N_t = 2$. In fact, the approximation turns out to be a lower bound. Note that due to the imperfect CSIT, the sum rate reduces with N_t .

In Figure 3, we compare the BF and ZF systems, with $B = 18$ bits, $f_c = 2$ GHz, $v = 10$ km/hr, and $T_s = 1$ msec. We

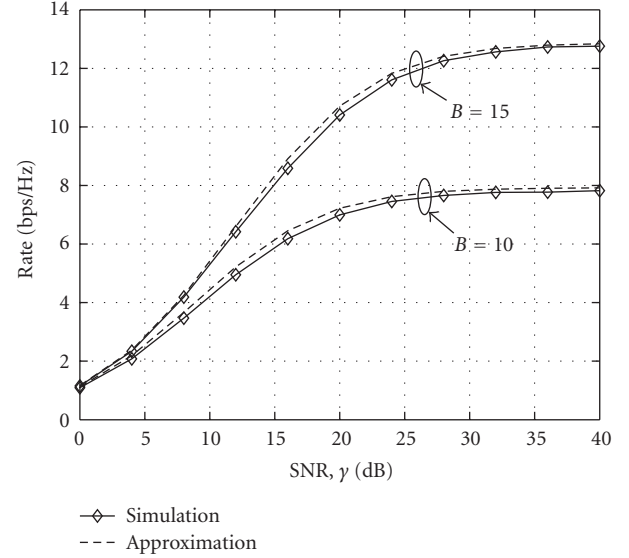


FIGURE 1: Approximated and simulated ergodic rates for the ZF precoding system with limited feedback, $N_t = U = 4$.

see that the approximation for the BF system almost matches the simulation exactly. The approximation for the ZF system is accurate at low to medium SNRs, and becomes a lower bound at high SNR, which is approximately 0.7 bps/Hz in total, or 0.175 bps/Hz per user, lower than the simulation. The throughput of the ZF system is limited by the residual inter-user interference at high SNR, where it is lower than the BF system. This motivates to switch between the SU and MU-MIMO modes. The approximations (19) and (33) will be used to calculate the mode switching points. There may be two switching points for the system with imperfect CSIT, as the SU mode will be selected at both low and high SNR. These two points can be calculated by providing different initial values to the nonlinear equation solver, such as *fsolve* in MATLAB.

5. Numerical Results

In this section, numerical results are presented. First, the operating regions for different modes are plotted, which show the impact of different parameters, including the normalized Doppler frequency, the codebook size, and the number of transmit antennas. Then the extension of our results for ZF precoding to MMSE precoding is demonstrated.

5.1. Operating Regions. As shown in Section 4.3, finding mode switching points requires solving a nonlinear equation, which does not have a closed-form solution and gives little insight. However, it is easy to evaluate numerically for different parameters, from which insights can be drawn. In this section, with the calculated mode switching points for different parameters, we plot the operating regions for both SU and MU modes. The active mode for the given parameter and the condition to activate each mode can be found from such plots.

In Figure 4, the operating regions for both SU and MU modes are plotted, for different normalized Doppler frequencies and different number of feedback bits in Figures 4(a) and 4(b), respectively, and with $U = N_t = 4$. There are analogies between the two plots. Some key observations are as follows.

- (i) For the delay plot in Figure 4(a), comparing the two curves for $B = 16$ bits and $B = 20$ bits, we see that the smaller the codebook size, the smaller the operating region for the ZF mode. For the ZF mode to be active, $f_d T_s$ needs to be small, specifically we need $f_d T_s < 0.055$ and $f_d T_s < 0.046$ for $B = 20$ bits and $B = 16$ bits, respectively. These conditions are not easily satisfied in practical systems. For example, with carrier frequency $f_c = 2$ GHz, mobility $v = 20$ km/hr, the Doppler frequency is 37 Hz, and then to satisfy $f_d T_s < 0.055$ the delay should be less than 1.5 msec.
- (ii) For the codebook size plot in Figure 4(b), comparing the two curves with $v = 10$ km/hr and $v = 20$ km/hr, as $f_d T_s$ increases (v increases), the ZF operating region shrinks. For the ZF mode to be active, we should have $B \geq 12$ bits and $B \geq 14$ bits for $v = 10$ km/hr and $v = 20$ km/hr, respectively, which means a large codebook size. Note that for BF we only need a small codebook size to get the near-optimal performance [5].
- (iii) For a given $f_d T_s$ and B , the SU mode will be active at both low and high SNRs, which is due to its array gain and the robustness to imperfect CSIT, respectively.

The operating regions for different N_t values are shown in Figure 5. We see that as N_t increases, the operating region for the MU mode shrinks. Specifically, we need $B > 12$ bits for $N_t = 4$, $B > 19$ bits for $N_t = 6$, and $B > 26$ bits for $N_t = 8$ to get the MU mode activated. Note that the minimum required feedback bits per user for the MU mode grow approximately linearly with N_t .

5.2. ZF versus MMSE Precoding. It is shown in [39] that the regularized ZF precoding, denoted as *MMSE precoding* in this paper, can significantly increase the throughput at low SNR. In this section, we show that our results on mode switching with ZF precoding can also be applied to MMSE precoding.

Denote $\hat{\mathbf{H}}[n] = [\hat{\mathbf{h}}_1[n], \hat{\mathbf{h}}_2[n], \dots, \hat{\mathbf{h}}_U[n]]^*$. Then the MMSE precoding vectors are chosen to be the normalized columns of the matrix [39]

$$\hat{\mathbf{H}}^*[n] \left(\hat{\mathbf{H}}[n] \hat{\mathbf{H}}^*[n] + \frac{U}{P} \mathbf{I} \right)^{-1}. \quad (36)$$

From this, we see that the MMSE precoders converge to ZF precoders at high SNR. Therefore, our derivations for the ZF system also apply to the MMSE system at high SNR.

In Figure 6, we compare the performance of ZF and MMSE precoding systems with delay. Such a comparison can also be done in the system with both delay and quantization, which is more time-consuming. As shown in Lemma 1,

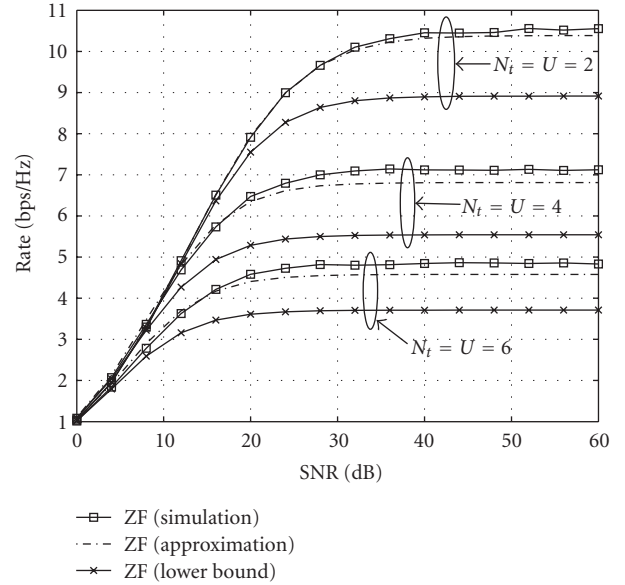


FIGURE 2: Comparison of approximation in (33), the lower bound in (29), and the simulation results for the ZF system with both delay and channel quantization. $B = 10$ bits, $f_c = 2$ GHz, $v = 20$ km/hr, and $T_s = 1$ msec.

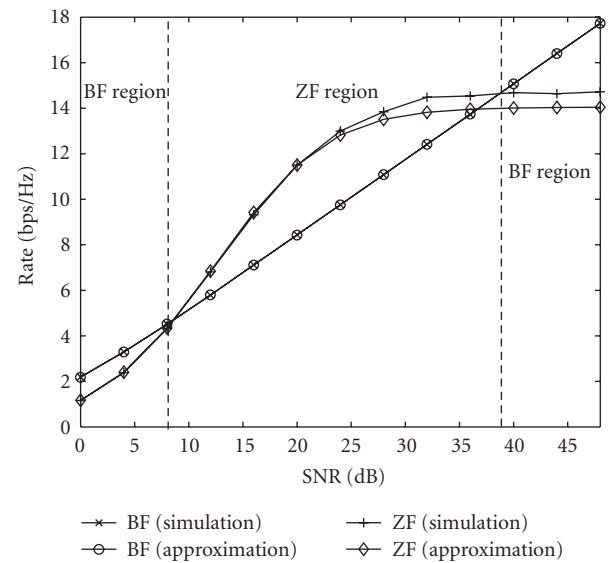
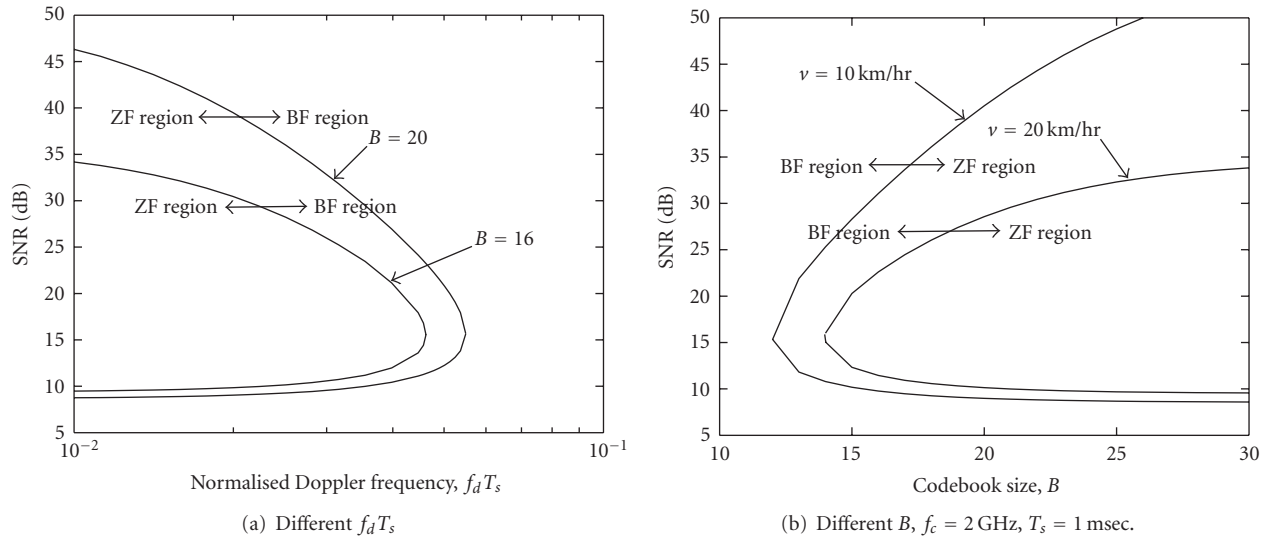
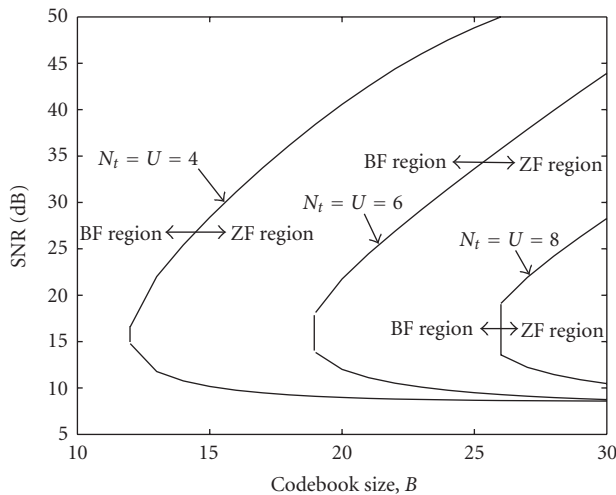
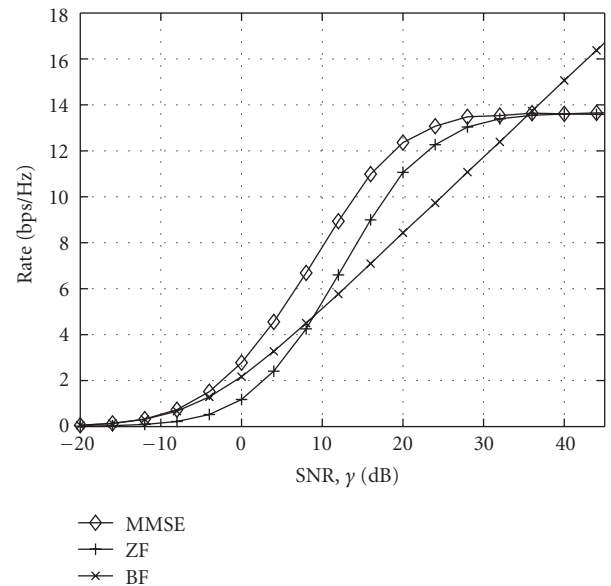


FIGURE 3: Mode switching between BF and ZF modes with both CSI delay and channel quantization, $B = 18$ bits, $N_t = 4$, $f_c = 2$ GHz, $T_s = 1$ msec, $v = 10$ km/hr.

the effects of delay and quantization are equivalent, so the conclusion will be the same. We see that the MMSE precoding outperforms ZF at low to medium SNRs, and converges to ZF at high SNR while converges to BF at low SNR. In addition, it has the same rate ceiling as the ZF system, and crosses the BF curve roughly at the same point, after which we need to switch to the SU mode. Based on this, we can use the second predicted mode switching point (the one at higher SNR) of the ZF system for the MMSE

TABLE 2: Mode switching points.

	$f_d T_s = 0.03$	$f_d T_s = 0.04$	$f_d T_s = 0.05$
MMSE (simulation)	44.2 dB	35.7 dB	29.5 dB
ZF (simulation)	44.2 dB	35.4 dB	28.6 dB
ZF (calculation)	41.6 dB	32.9 dB	26.1 dB

FIGURE 4: Operating regions for BF and ZF with both CSI delay and quantization, $N_t = 4$.FIGURE 5: Operating regions for BF and ZF with different N_t , $f_c = 2$ GHz, $v = 10$ km/hr, $T_s = 1$ msec.FIGURE 6: Simulation results for BF, ZF and MMSE systems with delay, $N_t = U = 4$, $f_d T_s = 0.04$.

system. We compare the simulation results and calculation results by (21) and (35) for the mode switching points in Table 2. For the ZF system, it is the second switching point; for the MMSE system, it is the only switching point. We see that the switching points for MMSE and ZF systems are very close, and the calculated ones are roughly 2.5 ~ 3 dB lower.

6. Conclusions

In this paper, we compare the SU and MU-MIMO transmissions in the broadcast channel with delayed and quantized

CSIT, where the amount of delay and the number of feedback bits per user are fixed. The throughput of MU-MIMO saturates at high SNR due to residual inter-user interference, for which an SU/MU mode switching algorithm is proposed. We derive accurate closed-form approximations for the ergodic rates for both SU and MU modes, which are then used to calculate the mode switching points. It is shown that the MU mode is only possible to be active in the medium SNR regime, with a small normalized Doppler frequency and a large codebook size.

In this paper, we assume that the transmitter knows perfectly the actual received SINR at each active user. In practice, there will inevitably be errors in such information due to estimation error and feedback delay, which will result in rate mismatch, that is, the transmission rate based on the estimated SINR does not match the actual SINR on the channel, so there will be outage events. How to deal with such rate mismatch is of practical importance and we mention several possible approaches as follows. The full investigation of this issue is left to future work. Considering the outage events, the transmission strategy can be designed based on the actual information symbols successfully delivered to the receiver, denoted as *goodput* in [42, 43]. With the estimated SINR, another approach is to back off on the transmission rate based on the variance of the estimation error, as did in [44, 45] for the single-antenna opportunistic scheduling system and in [46] for the multiple-antenna opportunistic beamforming system. Combined with user selection, the transmission rate can also be determined based on some lower bound of the actual SINR to make sure that no outage occurs, as did in [47] for the limited feedback system.

For other future work, the MU-MIMO mode studied in this paper is designed with zero-forcing criterion, which is shown to be sensitive to CSI imperfections, so robust precoding design is needed and the impact of the imperfect CSIT on nonlinear precoding should be investigated. As power control is an effective way to combat interference, it is interesting to consider the efficient power control algorithm rather than equal power allocation to improve the performance, especially in the heterogeneous scenario. It is also of practical importance to investigate possible approaches to improve the quality of the available CSIT with a fixed codebook size, for example, through channel prediction. In this paper, the mode switching algorithm only switches between the SU mode and the MU mode with N_t users, and how to extend it to allow more MU modes to further improve the performance is currently under investigation. For practical applications, the impact of more realistic channel models should also be investigated, such as channel correlation.

Appendices

A. Useful Results for Rate Analysis

In this appendix, we present some useful results that are used for rate analysis in this paper.

The following lemma will be used frequently in the derivation of the achievable rate.

Lemma 2. For a random variable X with probability distribution function (pdf) $f_X(x)$ and cumulative distribution function (cdf) $F_X(x)$, one has

$$\mathbb{E}_X[\ln(1+X)] = \int_0^\infty \frac{1-F_X(x)}{1+x} dx. \quad (\text{A.1})$$

Proof. The proof follows the integration by parts,

$$\begin{aligned} \mathbb{E}_X[\ln(1+X)] &= \int_0^\infty \ln(1+x)f_X(x)dx \\ &= -\int_0^\infty \ln(1+x)[1-F_X(x)]' dx \\ &\stackrel{(a)}{=} \int_0^\infty \frac{1-F_X(x)}{1+x} dx, \end{aligned} \quad (\text{A.2})$$

where g' is the derivative of the function g , and step (a) follows the integration by parts. \square

The following lemma provides some useful integrals for rate analysis, which can be derived using the results in [30].

Lemma 3.

$$\begin{aligned} I_1(a, b, m) &= \int_0^\infty \frac{x^m e^{-ax}}{x+b} dx \\ &= \sum_{k=1}^m (k-1)!(-b)^{m-k} a^{-k} \\ &\quad - (-1)^{m-1} b^m e^{ab} E_1(ab), \end{aligned} \quad (\text{A.3})$$

$$\begin{aligned} I_2(a, b, m) &= \int_0^\infty \frac{e^{-ax}}{(x+b)^m} dx \\ &= \begin{cases} e^{ab} E_1(ab) & m = 1, \\ \sum_{k=1}^{m-1} \frac{(k-1)! (-a)^{m-k-1}}{(m-1)! b^k} \\ \quad + \frac{(-a)^{m-1}}{(m-1)!} e^{ab} E_1(ab) & m \geq 2, \end{cases} \end{aligned} \quad (\text{A.4})$$

$$\begin{aligned} I_3(a, b, m) &= \int_0^\infty \frac{e^{-ax}}{(x+b)^m(x+1)} dx \\ &= \sum_{i=1}^m (-1)^{i-1} (1-b)^{-i} \cdot I_2(a, b, m-i+1) \\ &\quad + (b-1)^{-m} \cdot I_2(a, 1, 1), \end{aligned} \quad (\text{A.5})$$

where $E_1(x)$ is the exponential-integral function of the first order.

B. Proof of Theorem 1

The average SNR is

$$\begin{aligned}
\overline{\text{SNR}}_{\text{BF}}^{(\text{QD})} &= \mathbb{E} \left[P \left| \mathbf{h}^*[n] \mathbf{f}^{(\text{QD})}[n] \right|^2 \right] \\
&= P \mathbb{E} \left[\left| (\rho \mathbf{h}[n-1] + \mathbf{e}[n]) * \hat{\mathbf{h}}[n-1] \right|^2 \right] \\
&\stackrel{\text{(a)}}{=} P \mathbb{E} \left[\left| \rho \mathbf{h}^*[n-1] \hat{\mathbf{h}}[n-1] \right|^2 \right] \\
&\quad + P \mathbb{E} \left[\left| \mathbf{e}^*[n] \hat{\mathbf{h}}[n-1] \right|^2 \right] \\
&\stackrel{\text{(b)}}{\leq} P N_t \rho^2 \left(1 - \frac{N_t - 1}{N_t} 2^{-B/(N_t-1)} \right) \\
&\quad + P \mathbb{E} \left[\left| \hat{\mathbf{h}}^*[n-1] \cdot [\mathbf{e}[n] \mathbf{e}^*[n]] \cdot \hat{\mathbf{h}}[n-1] \right|^2 \right] \\
&\stackrel{\text{(c)}}{=} P N_t \rho^2 \left(1 - \frac{N_t - 1}{N_t} 2^{-B/(N_t-1)} \right) + P \epsilon_e^2.
\end{aligned} \tag{B.1}$$

As $\mathbf{e}[n]$ is independent of $\mathbf{h}[n-1]$, it is also independent of $\hat{\mathbf{h}}[n-1]$, which gives (a). Step (b) follows (12). Step (c) is from the fact $\mathbf{e}[n] \sim \mathcal{CN}(\mathbf{0}, \epsilon_e^2 \mathbf{I}_{N_t})$ and $|\hat{\mathbf{h}}[n-1]|^2 = 1$.

C. Proof of Theorem 2

Denote $y_1 = \|\mathbf{h}[n-1]\|^2$ and $y_2 = (1/\epsilon_e^2) |\mathbf{e}^*[n] \tilde{\mathbf{h}}[n-1]|^2$, then $y_1 \sim \chi_{2N_t}^2$, $y_2 \sim \chi_2^2$, and they are independent. The received SNR can be written as $x = \eta_1 y_1 + \eta_2 y_2$, where $\eta_1 = P\rho^2$ and $\eta_2 = P\epsilon_e^2$. The cdf of X is given as [48]

$$\begin{aligned}
F_X(x) &= 1 - \left(\frac{\eta_2}{\eta_2 - \eta_1} \right)^{N_t} e^{-x/\eta_2} \\
&\quad + e^{-x/\eta_1} \left(\frac{\eta_1}{\eta_2 - \eta_1} \right) \\
&\quad \cdot \sum_{i=0}^{N_t-1} \sum_{l=0}^i \frac{1}{(i-l)!} \left(\frac{\eta_2}{\eta_2 - \eta_1} \right)^{N_t-1-i} \left(\frac{x}{\eta_1} \right)^{i-l}.
\end{aligned} \tag{C.1}$$

Denote $a_0 = \eta_2/(\eta_2 - \eta_1)$ and following Lemma 2 we have

$$\begin{aligned}
&\mathbb{E}_X[\ln(1+X)] \\
&= \int_0^\infty \frac{1 - F_X(x)}{1+x} dx \\
&= a_0^{N_t} \int_0^\infty \frac{e^{-x/\eta_2}}{1+x} dx - (1 - a_0) \sum_{i=0}^{N_t-1} \sum_{l=0}^i \frac{a_0^{N_t-1-i}}{(i-l)!} \left(\frac{1}{\eta_1} \right)^{i-l} \\
&\quad \times \int_0^\infty \frac{x^{i-l} e^{-x/\eta_1}}{1+x} dx \\
&= a_0^{N_t} I_2 \left(\frac{1}{\eta_2}, 1, 1 \right) - (1 - a_0) \\
&\quad \times \sum_{i=0}^{N_t-1} \sum_{l=0}^i \frac{a_0^{N_t-1-i}}{(i-l)!} \left(\frac{1}{\eta_1} \right)^{i-l} I_1 \left(\frac{1}{\eta_1}, 1, i-l \right),
\end{aligned} \tag{C.2}$$

where $I_1(\cdot, \cdot, \cdot)$ and $I_2(\cdot, \cdot, \cdot)$ are given in (A.3) and (A.4), respectively.

D. Proof of Lemma 1

Let $x = \|\mathbf{h}_u[n-D]\|^2 \sin^2 \theta \sim \Gamma(M-1, \delta)$, $y \sim \beta(1, M-2)$, and x is independent of y . Then the interference term due to quantization is $Z = XY$. The cdf of Z is

$$\begin{aligned}
P_Z(z) &= P(xy \leq z) \\
&= \int_0^\infty F_{Y|X} \left(\frac{z}{x} \right) f_X(x) dx \\
&= \int_0^z f_X(x) dx + \int_z^\infty \left(1 - \left(1 - \frac{z}{x} \right)^{M-2} \right) f_X(x) dx \\
&= \int_0^\infty f_X(x) dx - \int_z^\infty \left(1 - \frac{z}{x} \right)^{M-2} x^{M-2} \frac{e^{-x/\delta}}{(M-2)! \delta^{M-1}} dx \\
&= 1 - e^{-z/\delta} \int_z^\infty (x-z)^{M-2} \frac{e^{-(x-z)/\delta}}{(M-2)! \delta^{M-1}} dx \\
&\stackrel{\text{(a)}}{=} 1 - e^{-z/\delta},
\end{aligned} \tag{D.1}$$

where step (a) follows the equality $\int_0^\infty y^M e^{-\alpha y} = M! \alpha^{-(M+1)}$.

E. Proof of Theorem 4

Assuming that each interference term in (30) is independent of each other and independent of the signal power term, denote $\sum_{u' \neq u} \rho_{u'}^2 |\mathbf{h}_u^*[n-1] \mathbf{f}_{u'}^{(\text{QD})}[n]|^2 = \rho_{e,u}^2 \delta y_1$ and $\sum_{u' \neq u} |\mathbf{e}_u^*[n] \mathbf{f}_{u'}^{(\text{QD})}[n]|^2 = \epsilon_{e,u}^2 y_2$, then from Lemma 1 we have $y_1 \sim \chi_{2(N_t-1)}^2$, and $y_2 \sim \chi_{2(N_t-1)}^2$ as $\mathbf{e}_u[n]$ is complex Gaussian with variance $\epsilon_{e,u}^2$ and independent of the normalized vector $\mathbf{f}_{u'}^{(\text{QD})}[n]$. In addition, the signal power $|\mathbf{h}_u^*[n] \mathbf{f}_u^{(\text{QD})}[n]|^2 \sim \chi_2^2$. Then the received SINR for the u th user is approximated as

$$\gamma_{\text{ZF},u}^{(\text{QD})} \approx \frac{\alpha z}{1 + \beta(\delta_1 y_1 + \delta_2 y_2)} \triangleq x, \tag{E.1}$$

where $\alpha = \beta = P/U$, $\delta_1 = \rho_{e,u}^2 \delta$, $\delta_2 = \epsilon_{e,u}^2$, $y_1 \sim \chi_{2M}^2$, $y_2 \sim \chi_{2M}^2$, $M = N_t - 1$, $z \sim \chi_2^2$, and y_1, y_2, z are independent of each other.

Let $y = \delta_1 y_1 + \delta_2 y_2$, then the pdf of y , which is the sum of two independent chi-square random variables, is given as [48]

$$\begin{aligned}
p_Y(y) &= e^{-y/\delta_1} \sum_{i=0}^{M-1} a_i^{(1)} y^i + e^{-y/\delta_2} \sum_{i=0}^{M-1} a_i^{(2)} y^i \\
&= \sum_{j=1}^2 \sum_{i=0}^{M-1} e^{-y/\delta_j} a_i^{(j)} y^i,
\end{aligned} \tag{E.2}$$

where

$$\begin{aligned}
 a_i^{(1)} &= \frac{1}{\delta_1^{i+1}(M-1)!} \left(\frac{\delta_1}{\delta_1 - \delta_2} \right)^M \\
 &\quad \times \frac{(2(M-1)-i)!}{i!(M-1-i)!} \left(\frac{\delta_2}{\delta_2 - \delta_1} \right)^{M-1-i}, \\
 a_i^{(2)} &= \frac{1}{\delta_2^{i+1}(M-1)!} \left(\frac{\delta_2}{\delta_2 - \delta_1} \right)^M \\
 &\quad \times \frac{(2(M-1)-i)!}{i!(M-1-i)!} \left(\frac{\delta_1}{\delta_1 - \delta_2} \right)^{M-1-i}.
 \end{aligned} \tag{E.3}$$

The cdf of X is

$$\begin{aligned}
 F_X(x) &= P\left(\frac{\alpha z}{1 + \beta y} \leq x\right) \\
 &= \int_0^\infty F_{Z|Y}\left(\frac{x}{\alpha}(1 + \beta y)\right) p_Y(y) dy \\
 &= \int_0^\infty \left(1 - e^{-(x/\alpha)(1+\beta y)}\right) p_Y(y) dy \\
 &= 1 - e^{-x/\alpha} \int_0^\infty e^{-\beta xy/\alpha} p_Y(y) dy \\
 &= 1 - e^{-x/\alpha} \int_0^\infty \left\{ \sum_{j=1}^2 \sum_{i=0}^{M-1} \exp\left[-\left(\frac{\beta}{\alpha}x + \frac{1}{\delta_j}\right)y\right] a_i^{(j)} y^i \right\} dy \\
 &\stackrel{(a)}{=} 1 - e^{-x/\alpha} \sum_{j=1}^2 \sum_{i=0}^{M-1} \left[\frac{a_i^{(j)} i!}{\left((\beta/\alpha)x + 1/\delta_j\right)^{i+1}} \right],
 \end{aligned} \tag{E.4}$$

where step (a) follows the equality $\int_0^\infty y^M e^{-\alpha y} = M! \alpha^{-(M+1)}$.

Then the ergodic achievable rate for the u th user is approximated as

$$\begin{aligned}
 R_{ZF,u}^{(QD)} &= \mathbb{E}_y \left[\log_2 \left(1 + \gamma_{ZF,u}^{(QD)} \right) \right] \\
 &\approx \log_2(e) \mathbb{E}_X [\ln(1 + X)] \\
 &\stackrel{(a)}{=} \log_2(e) \int_0^\infty \frac{1 - F_X(x)}{x + 1} dx \\
 &= \log_2(e) \int_0^\infty \sum_{i=0}^{M-1} \sum_{j=1}^2 \left[a_i^{(j)} i! \left(\frac{\alpha}{\beta} \right) \frac{e^{-x/\alpha}}{(x + \alpha/\beta\delta_j)^{i+1} (x+1)} \right] dx \\
 &\stackrel{(b)}{=} \log_2(e) \sum_{i=0}^{M-1} \sum_{j=1}^2 \left[a_i^{(j)} i! \left(\frac{\alpha}{\beta} \right)^{i+1} I_3 \left(\frac{1}{\alpha}, \frac{\alpha}{\beta\delta_j}, i+1 \right) \right],
 \end{aligned} \tag{E.5}$$

where step (a) follows from Lemma 2, step (b) follows the expression of $I_3(\cdot, \cdot, \cdot)$ in (A.5). For equal power allocation, $\alpha = \beta = P/U$, and the expression can be simplified into (33).

Acknowledgment

This work has been supported in part by AT&T Labs, Inc.

References

- [1] E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–595, 1999.
- [2] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 5, pp. 684–702, 2003.
- [3] D. Gesbert, M. Kountouris, R. W. Heath Jr., C. B. Chae, and T. Salzer, "Shifting the MIMO paradigm: from single user to multiuser communications," *IEEE Signal Processing Magazine*, vol. 24, no. 5, pp. 36–46, 2007.
- [4] B. Hassibi and M. Sharif, "Fundamental limits in MIMO broadcast channels," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 7, pp. 1333–1344, 2007.
- [5] D. J. Love, R. W. Heath Jr., W. Santipach, and M. L. Honig, "What is the value of limited feedback for MIMO channels?" *IEEE Communications Magazine*, vol. 42, no. 10, pp. 54–59, 2004.
- [6] D. J. Love, R. W. Heath Jr., V. K. N. Lau, D. Gesbert, B. D. Rao, and M. Andrews, "An overview of limited feedback in wireless communication systems," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 8, pp. 1341–1365, 2008.
- [7] N. Jindal, "MIMO broadcast channels with finite-rate feedback," *IEEE Transactions on Information Theory*, vol. 52, no. 11, pp. 5045–5060, 2006.
- [8] P. Ding, D. J. Love, and M. D. Zoltowski, "Multiple antenna broadcast channels with shape feedback and limited feedback," *IEEE Transactions on Signal Processing*, vol. 55, no. 7, pp. 3417–3428, 2007.
- [9] R. W. Heath Jr. and A. J. Paulraj, "Switching between diversity and multiplexing in MIMO systems," *IEEE Transactions on Communications*, vol. 53, no. 6, pp. 962–972, 2005.
- [10] D. J. Love and R. W. Heath Jr., "Multi-mode precoding using linear receivers for limited feedback MIMO systems," in *Proceedings of IEEE International Conference on Communications (ICC '04)*, vol. 1, pp. 448–452, Paris, France, June 2004.
- [11] R. W. Heath Jr. and D. J. Love, "Multimode antenna selection for spatial multiplexing systems with linear receivers," *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 3042–3056, 2005.
- [12] J. C. Roh and B. D. Rao, "Design and analysis of MIMO spatial multiplexing systems with quantized feedback," *IEEE Transactions on Signal Processing*, vol. 54, no. 8, pp. 2874–2886, 2006.
- [13] N. Ravindran and N. Jindal, "Limited feedback-based block diagonalization for the MIMO broadcast channel," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 8, pp. 1473–1482, 2008.
- [14] T. Yoo, N. Jindal, and A. Goldsmith, "Multi-antenna downlink channels with limited feedback and user selection," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 7, pp. 1478–1491, 2007.
- [15] G. Caire, "MIMO downlink joint processing and scheduling: a survey of classical and recent results," in *Proceedings of the Workshop on Information Theory and Its Applications (ITA '06)*, San Diego, Calif, USA, January 2006.
- [16] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, "Multiuser MIMO achievable rates with downlink training and channel state feedback," submitted to *IEEE Transactions on Information Theory*, November 2007.

- [17] G. Caire, N. Jindal, and S. Shamai, "On the required accuracy of transmitter channel state information in multiple antenna broadcast channels," in *Proceedings of IEEE Asilomar Conference on Signals, Systems and Computers*, pp. 287–291, Pacific Grove, Calif, USA, November 2007.
- [18] R. de Francisco and D. T. M. Slock, "A design framework for scalar feedback in MIMO broadcast channels," *EURASIP Journal on Advances in Signal Processing*, vol. 8, no. 1, pp. 1–12, 2008.
- [19] E. A. Jorswieck, P. Svedman, and B. Ottersten, "Performance of TDMA and SDMA based opportunistic beamforming," *IEEE Transactions on Wireless Communications*, vol. 7, no. 11, pp. 4058–4063, 2008.
- [20] M. Kountouris, D. Gesbert, and T. Sälzer, "Distributed transmit mode selection for MISO broadcast channels with limited feedback: switching from SDMA to TDMA," in *Proceedings of IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC '08)*, pp. 371–375, Recife, Brazil, July 2008.
- [21] N. Ravindran, N. Jindal, and H. C. Huang, "Beamforming with finite rate feedback for LOS MIMO downlink channels," in *Proceedings of IEEE Global Telecommunications Conference (GLOBECOM '07)*, pp. 4200–4204, Washington, DC, USA, November 2007.
- [22] S. Haykin, *Adaptive Filter Theory*, Prentice-Hall, Englewood Cliffs, NJ, USA, 3rd edition, 1996.
- [23] R. H. Clarke, "A statistical theory of mobile radio reception," *Bell System Technical Journal*, vol. 47, pp. 957–1000, 1968.
- [24] E. N. Onggosanusi, A. Gatherer, A. G. Dabak, and S. Hosur, "Performance analysis of closed-loop transmit diversity in the presence of feedback delay," *IEEE Transactions on Communications*, vol. 49, no. 9, pp. 1618–1630, 2001.
- [25] H. T. Nguyen, J. B. Andersen, and G. F. Pedersen, "Capacity and performance of MIMO systems under the impact of feedback delay," in *Proceedings of IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '04)*, vol. 1, pp. 53–57, Barcelona, Spain, September 2004.
- [26] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, "Quantized vs. analog feedback for the MIMO broadcast channel: a comparison between zero-forcing based achievable rates," in *Proceedings of IEEE International Symposium on Information Theory (ISIT '07)*, pp. 2046–2050, Nice, France, June 2007.
- [27] M. Kobayashi and G. Caire, "Joint beamforming and scheduling for a multi-antenna downlink with imperfect transmitter channel knowledge," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 7, pp. 1468–1477, 2007.
- [28] W. Santipach and M. L. Honig, "Asymptotic capacity of beamforming with limited feedback," in *Proceedings of IEEE International Symposium on Information Theory (ISIT '04)*, p. 290, Chicago, Ill, USA, June-July 2004.
- [29] C. K. Au-Yeung and D. J. Love, "On the performance of random vector quantization limited feedback beamforming in a MISO system," *IEEE Transactions on Wireless Communications*, vol. 6, no. 2, pp. 458–462, 2007.
- [30] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, Academic Press, San Diego, Calif, USA, 5th edition, 1994.
- [31] M.-S. Alouini and A. J. Goldsmith, "Capacity of Rayleigh fading channels under different adaptive transmission and diversity-combining techniques," *IEEE Transactions on Vehicular Technology*, vol. 48, no. 4, pp. 1165–1181, 1999.
- [32] M. H. M. Costa, "Writing on dirty paper," *IEEE Transactions on Information Theory*, vol. 29, no. 3, pp. 439–441, 1983.
- [33] G. Caire and S. Shamai (Shitz), "On the achievable throughput of a multiantenna Gaussian broadcast channel," *IEEE Transactions on Information Theory*, vol. 49, no. 7, pp. 1691–1706, 2003.
- [34] W. Yu and J. Cioffi, "The sum capacity of a Gaussian vector broadcast channel," *IEEE Transactions on Information Theory*, vol. 50, no. 9, pp. 1875–1892, 2004.
- [35] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2658–2668, 2003.
- [36] P. Viswanath and D. N. C. Tse, "Sum capacity of the vector Gaussian broadcast channel and uplink-downlink duality," *IEEE Transactions on Information Theory*, vol. 49, no. 8, pp. 1912–1921, 2003.
- [37] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Transactions on Information Theory*, vol. 52, no. 9, pp. 3936–3964, 2006.
- [38] N. Jindal, "A high SNR analysis of MIMO broadcast channels," in *Proceedings of IEEE International Symposium on Information Theory (ISIT '05)*, pp. 2310–2314, Adelaide, Australia, September 2005.
- [39] C. B. Peel, B. M. Hochwald, and A. L. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication—part I: channel inversion and regularization," *IEEE Transactions on Communications*, vol. 53, no. 1, pp. 195–202, 2005.
- [40] K. K. Mukkavilli, A. Sabharwal, E. Erkip, and B. Aazhang, "On beamforming with finite rate feedback in multiple-antenna systems," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2562–2579, 2003.
- [41] S. Zhou, Z. Wang, and G. B. Giannakis, "Quantifying the power loss when transmit beamforming relies on finite-rate feedback," *IEEE Transactions on Wireless Communications*, vol. 4, no. 4, pp. 1948–1957, 2005.
- [42] V. K. N. Lau and M. Jiang, "Performance analysis of multiuser downlink space-time scheduling for TDD systems with imperfect CSIT," *IEEE Transactions on Vehicular Technology*, vol. 55, no. 1, pp. 296–305, 2006.
- [43] T. Wu and V. K. N. Lau, "Robust rate, power and precoder adaptation for slow fading MIMO channels with noisy limited feedback," *IEEE Transactions on Wireless Communications*, vol. 7, no. 6, pp. 2360–2367, 2008.
- [44] A. Vakili, M. Sharif, and B. Hassibi, "The effect of channel estimation error on the throughput of broadcast channels," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '06)*, vol. 4, pp. 29–32, Toulouse, France, May 2006.
- [45] A. Vakili and B. Hassibi, "On the throughput of broadcast channels with imperfect CSI," in *Proceedings of IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC '06)*, pp. 1–5, Cannes, France, July 2006.
- [46] A. Vakili, A. F. Dana, and B. Hassibi, "On the throughput of opportunistic beamforming with imperfect CSI," in *Proceedings of the ACM International Wireless Communications and Mobile Computing Conference (IWCMC '07)*, pp. 19–23, Honolulu, Hawaii, USA, August 2007.

- [47] M. Kountouris, R. de Francisco, D. Gesbert, D. T. M. Slock, and T. Salzer, "Efficient metrics for scheduling in MIMO broadcast channels with limited feedback," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '07)*, vol. 3, pp. 109–112, Honolulu, Hawaii, USA, April 2007.
- [48] M. K. Simon, *Probability Distributions Involving Gaussian Random Variables: A Handbook for Engineers and Scientists*, Springer, New York, NY, USA, 2002.