

Power-Controlled Feedback and Training for Two-way MIMO Channels

Vaneet Aggarwal, and Ashutosh Sabharwal, *Senior Member, IEEE*

Abstract—Most communication systems use some form of feedback, often related to channel state information. The common models used in analyses either assume perfect channel state information at the receiver and/or noiseless state feedback links. However, in practical systems, neither is the channel estimate known perfectly at the receiver and nor is the feedback link perfect. In this paper, we study the achievable diversity multiplexing tradeoff using i.i.d. Gaussian codebooks, considering the errors in training the receiver and the errors in the feedback link for FDD systems, where the forward and the feedback are independent MIMO channels.

Our key result is that the maximum diversity order with one-bit of feedback information is identical to systems with more feedback bits. Thus, asymptotically in SNR, more than one bit of feedback does not improve the system performance at constant rates. Furthermore, the one-bit diversity-multiplexing performance is identical to the system which has perfect channel state information at the receiver along with noiseless feedback link. This achievability uses novel concepts of power controlled feedback and training, which naturally surface when we consider imperfect channel estimation and noisy feedback links. In the process of evaluating the proposed training and feedback protocols, we find an asymptotic expression for the joint probability of the SNR exponents of eigenvalues of the actual channel and the estimated channel which may be of independent interest.

Index Terms—Channel state information, diversity multiplexing tradeoff, feedback, multiple access channel, outage probability, power-controlled, training.

I. INTRODUCTION

Channel state information at the transmitter has been well established to improve communication performance, measured either as increased capacity (see e.g. [1–6]), improved diversity-multiplexing performance [7–18] or higher signal-to-noise ratio at the receiver [2, 19, 20] among many possible metrics. Several of the aforementioned works have considered the impact of incomplete channel knowledge at the transmitter, by considering quantized channel information at the transmitter which can be visualized to be made available by the receiver through a noiseless finite-capacity feedback link. While the model and subsequent analysis clearly shows that reduced channel information at the transmitter can lead to significant performance gains due to channel knowledge, a key requirement is that the receiver knows what the transmitter knows (even if there is an error in feedback link). In practice, the communicating nodes are distributed and have no way of aligning their channel knowledge perfectly.

In this paper, we systematically analyze the impact of mismatch in channel knowledge at the transmitter and receiver. For clarity of presentation, we will largely focus on a single-user point-to-point link (one transmitter and one receiver) and only at the end of the paper, extend the results to the case of multiple-access channels (many transmitters and one receiver). The key departure from prior work is that we explicitly model both the forward and feedback links as fading wireless channels. A little thought immediately shows that all practical wireless networks have “two-way” communication links, that is, the nodes are transceivers such that all the received and transmitted packets (control, data, feedback) travel over noisy fading channels. The two-way model to analyze feedback was first proposed in [10] for TDD systems, where it was shown to enable accurate resource accounting of the feedback link resources (power and spectrum) and analyze the important case of transmitter and receiver channel knowledge mismatch. A key message from [10] was that in the general case when the transmitter and receiver knowledge is mismatched, both feedback and forward communication has to be *jointly* designed. The simple two-way training protocol proposed in [10] led to the concept of power-controlled training which enabled joint estimation of transmitter’s intended power control and actual channel realization. We will continue the line of thought initiated in [10] for the case of FDD (Frequency Division Duplex) systems and more importantly, systematically show how mismatch between information at the transmitter and receiver impacts the overall system performance.

For the case of FDD systems, we will continue to model the forward and feedback links as fading channels. However, unlike [10], the forward and feedback channel will be assumed to be completely uncorrelated since they use different frequency bands. In this case, the two-way training protocol proposed in [10] will not be applicable and we will use the quantized feedback model, in which the receiver sends a quantized version of its own channel information over the feedback link.

Our main result is an achievable diversity-multiplexing tradeoff for an $m \times n$ MIMO (Multiple Input Multiple Output) system, with measured channel information at the receiver which is used to send quantized channel information via a noisy link using i.i.d. Gaussian codebooks. We show that, in the scope of modeling assumptions, diversity-multiplexing tradeoff with one-bit of noisy feedback with estimated CSIR (channel state information at receiver) is identical to diversity-multiplexing tradeoff with perfect CSIR and noiseless one-bit feedback. Even more importantly, more bits of feedback do *not* improve the maximum diversity order in the imperfect

V. Aggarwal is with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544, USA (email: vaggarwa@princeton.edu). A. Sabharwal is with the Department of Electrical and Computer Engineering, Rice University, Houston, TX, 77005, USA (email: ashu@rice.edu).

system (estimated CSIR with noisy feedback) in contrast of the idealized system where the maximum diversity order increases exponentially [8, 13, 17]. The conclusion holds for multiuser systems in general.

The encouraging news from our analysis is two-fold. First, even noisy and mismatched information about the channel at both transmitter and receiver is sufficient to improve the whole diversity-multiplexing tradeoff when compared to the case with no feedback. Second, very few bits of feedback, in fact one-bit, is all one needs to build in a practical system. However, to achieve any diversity order gain over the no-feedback system, a straightforward application of the Genie-aided feedback analysis [8, 13, 17] fails. That is, conventional training followed by conventionally coded quantized feedback signal leads to no improvement in diversity order. A combination of power-controlled training (much like in [10]) and the new concept of power-controlled feedback (proposed in this paper) appear essential for significant improvement in diversity-multiplexing gain compared to the no-feedback system.

We will build our main result in three major steps. In the first step, we assume that the receiver knowledge of the channel is perfect and the feedback channel is error-prone. Thus, the transmitter information about the channel is potentially different from that sent by the receiver. In this case, the coding of the feedback signal becomes important. If the feedback information about the channel is sent using a codebook where each codeword has equal power, then the maximum diversity order is $2mn$ irrespective of the amount of feedback information. In contrast, b -bit noiseless feedback leads to a maximum diversity order grows exponentially in number of bits [8, 13]. The reason for such dramatic decrease (from unbounded growth with b in the noiseless case to maximum $2mn$ in the noisy case) is that the transmitter cannot distinguish between rare channel states, especially those where it is supposed to send large power to overcome poor channel conditions. As a result, it becomes conservative in its power allocation and the gain from power control does not increase unboundedly as in the case of noiseless feedback. A careful examination of the outage events on the *feedback path* naturally points to unequal error protection in the form of *power controlled feedback* coding scheme, where the rare states are encoded with higher power codewords. With power controlled feedback, the diversity order can be improved to $(mn)^2 + 2mn$, a substantial increase compared to constant power feedback.

As our second step, we assume that the channel knowledge at the receiver is imperfect but the feedback path is error-free. We first analyze the commonly used protocol setup, where the receiver is trained at the average available power SNR (assuming noise variance is one) to obtain a channel estimate and then a quantized version of the estimated channel is fed back to the transmitter. While both the transmitter and receiver have identical information, the error in receiver information leads to *no* gain in the diversity order compared to a no-feedback system. We take a cue from an earlier work in [10] and the power-controlled feedback mechanism, and propose a two-round training protocol, where the receiver is trained twice.

First round training uses an average power SNR and then after obtaining the channel feedback, the receiver is trained again with a training power dependent on the channel estimate. This implies that in poor estimated channel conditions, second round training power is higher than the first and in good conditions, it is lower. The adaptation of training power is labeled *power-controlled training* and allows a substantial increase in maximum diversity order $((mn)^2 + mn)$ compared to no-feedback case (mn). For general multiplexing gains, the achievable diversity order is identical to that obtained when the receiver knows perfect channel state information with 1 bit of perfect feedback in [13]. Further, additional bits of feedback do not help in improving the maximum diversity order at constant rates when the receiver is trained to obtain a channel estimate.

Finally, we put the first two steps together to analyze the general case of noisy channel estimates with noisy feedback. The key conclusion is that the receiver channel estimate errors are the bottleneck in the maximum achievable diversity order. Thus, the maximum achievable diversity order at multiplexing gain $r = 0$ is $(mn)^2 + mn$, which is less than $(mn)^2 + 2mn$ obtained in the case of perfect receiver information. For general multiplexing, we get the same tradeoff as with 1 bit of perfect feedback and the receiver having perfect channel state information as in [13]. This diversity multiplexing tradeoff can be achieved with a single bit of noisy feedback from the receiver.

The results on MIMO point-to-point channels are extended to the multiple access channel in which all the transmitters have m antennas each and the receiver has n antennas, and all the conclusions drawn earlier for the single user also holds in a multiple access system. The channel information for each transmitter is measured at the receiver and a global quantized feedback is sent from the receiver to all the transmitters. We use the similar combination of power control training and feedback as in single-user systems to achieve a diversity order of $(mn)^2 + mn$ with a single bit of power-controlled feedback in the main case in which the errors in the channel estimate and the feedback link are accounted.

We note that the improvement in diversity order is completely dependent on our use of power control as adaptation mechanism. For example, in MIMO systems, feedback can be used for beamforming (e.g. [19, 20]). However, beamforming does not lead to any change the diversity order and only increases the receiver SNR by a constant amount. Since we have focused on the asymptotic regime, we do not consider schemes like beamforming which have identical diversity-multiplexing performance as a non-feedback based system. In this paper, we find new achievable schemes to improve the diversity multiplexing tradeoff which are better than the traditional approaches but do not claim globally optimality of these schemes.

The rest of the paper is organized as follows. Section II describes preliminaries on channel model and diversity multiplexing tradeoff. Section III summarizes the known results for the case when the receiver knows perfect channel state information and the feedback is sent over a noiseless channel [13]. Section IV describes the diversity order when the receiver knows the channel perfectly while the transmitter

receives feedback on a noisy channel. Section V describes the diversity multiplexing tradeoff when the receiver is trained to get channel estimate while the feedback link to the transmitter is noiseless. Section VI consider both the above errors, i.e., it considers receiver estimating the channel and the feedback link is also noisy. Section VII presents some numerical results. We consider the extension to multiple access channels in Section VIII. Section IX concludes the paper.

II. PRELIMINARIES

A. Two-way Channel Model

We will primarily focus on single-user multiple input-output channel with the transmitting node denoted by T and receiving node denoted by R. Later, we will extend the main result to the multiple access channel in Section VIII. For the single-user channel, we will assume that there are m transmit antennas at the source node and n receive antennas at the destination node, such that the input-output relation is given by

$$\text{T} \rightarrow \text{R} : Y = HX + W, \quad (1)$$

where the elements of H and W are assumed to be i.i.d. with complex Normal distribution of zero mean and unit variance, $CN(0, 1)$. The matrices Y, H, X and W are of dimension $n \times T_{\text{coh}}, n \times m, m \times T_{\text{coh}}$ and $n \times T_{\text{coh}}$, respectively. Here T_{coh} is coherence interval such that the channel H is fixed during a fading block of T_{coh} consecutive channel uses, and statistically independent from one block to another. The transmitter is assumed to be power-limited, such that the long-term power is upper bounded, i.e., $\frac{1}{T_{\text{coh}}} \text{trace}(\mathbb{E}[XX^\dagger]) \leq \text{SNR}$.

Since our focus will be studying feedback over noisy channels, we assume that the same multiple antennas at the transmitter and receiver are available to send feedback on an orthogonal frequency band. For the feedback path, the receiver will act as a transmitter and the transmitter as a receiver. As a result, the feedback source (which is destination for data bits) will have n transmit antennas and feedback destination (which is source of data bits) will be assumed to have m receive antennas. Furthermore, a block fading channel model is assumed

$$\text{R} \rightarrow \text{T} : Y_f = H_f X_f + W_f, \quad (2)$$

where H_f is the MIMO fading channel for the feedback link, and the W_f is the additive noise at the receiver of the feedback; both are assumed to have i.i.d. $CN(0, 1)$ elements. The feedback transmissions are also assumed to be power-limited with a long-term power constraint given by $\frac{1}{T_{\text{coh}}} \text{trace}(\mathbb{E}[X_f X_f^\dagger]) \leq \text{SNR}_f$. Without loss of generality, we will assume the case where the transmitter and receiver have *symmetric resources*, such that $\text{SNR} = \text{SNR}_f$.

We note that a phase-symmetric two-way channel model with $H = H_f^T$ was studied earlier in [10, 11]. The phase-symmetric two-way channel is a good model for slow-fading time division duplex (TDD) systems. On the other hand, the above SNR-symmetric ($\text{SNR} = \text{SNR}_f$) model is well-suited for symmetrically resourced FDD systems.

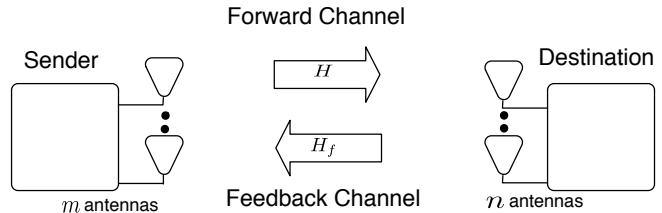


Fig. 1. Two-way MIMO channel. The forward and feedback channels use the same antennas. If the forward and feedback are frequency-division duplex (FDD), then H and H_f are not equal. On the other hand, if the forward and feedback channels are time-division duplex (TDD), then $H = H_f^T$ is often a reasonable assumption.

B. Obtaining Channel State Information

The two-way channel model allows two-way protocols, as depicted in Figure 2, where the transmitter and receiver can conduct multiple back-and-forth transactions to complete transmission of one codeword. The model was used in [10, 11] to study the diversity-multiplexing performance of two-way training method. In this paper, we will focus on another instance of the two-way protocols which will involve channel estimation and quantized feedback. However, unlike [10, 11], we will *not* account for resources spent on feedback path since the methods developed in [10] directly apply to the current case. Instead, we will focus on the more important issue of understanding how the mismatch in the transmitter and receiver channel information affects system performance.

To develop a systematic understanding of the two-way fading channel described in Section II-A, we will consider two forms of receiver knowledge about the channel H . The receiver will either be assumed to know the channel H perfectly or have a noisy estimate \hat{H} obtained using a minimum mean-squared channel estimate (MMSE) via a training sequence (described in detail in Section V).

For the transmitter knowledge, the receiver will quantize its own knowledge (H or \hat{H}) and map it to an index J_R , where $J_R \in \{0, 1, \dots, K-1\}$ where $K \geq 1$ is the number of feedback levels used by the receiver for feedback. The receiver then transmits the quantized channel information J_R over the feedback channel and the transmitter knowledge of the index is denoted by J_T . For the case when the feedback channel is assumed to be noiseless, $J_T = J_R$, else $J_T \neq J_R$ with finite probability. Here the error probability will be depend on the channel signal-to-noise ratio SNR and the transmission scheme. Based on the quantized information about the channel, in the form of J_T , the transmitter adapts its codeword to minimize the probability of outage (defined in the next section).

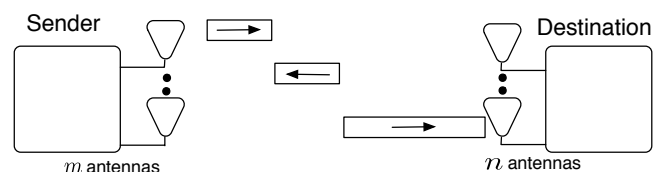


Fig. 2. Two-way protocols: In a two-way channel, the protocols can involve many exchanges between the transmitter and receiver. The arrows indicate the direction of transmission.

C. Diversity-Multiplexing Tradeoff

We will consider the case when then the codeword X spans a single fading block. Based on the transmitter channel knowledge $J_T = i$, the transmitted codeword is chosen from the codebook $C_i = \{X_i(1), X_i(2), \dots, X_i(2^{RT_{\text{coh}}})\}$, where the codebook rate is R . All $X_i(k)$'s (for $1 \leq k \leq 2^{RT_{\text{coh}}}$) are matrices of size $m \times T_{\text{coh}}$. We assume that T_{coh} is finite and does not scale with SNR. In this paper, we will only consider single rate transmission where the rate of the codebooks does not depend on the feedback index. Furthermore, we will assume codebooks C_i are derived from the same base codebook C by power scaling of the codewords. In other words $C_i = \sqrt{P_i}C$, where the product implies that each element of every codeword is multiplied by $\sqrt{P_i}$ where each element of codeword $X(k) \in C$ has unit power. Thus, P_i is the power of the transmitted codewords. Recall that there is an average transmit power constraint, such that $\mathbb{E}(P_i) \leq \text{SNR}$. We further assume that the base codebook C consists of Gaussian entries, in other words we only focus on Gaussian inputs in this paper.

In point-to-point channels, outage is defined as the event that the mutual information of the channel for a channel knowledge $H \in \xi$ with some distribution $P_{H|\xi}$ (for ξ a subset of all $n \times m$ matrices) at the receiver, $I_\xi(X; Y)$ is less than the desired rate R [21]. If ξ is a singleton set containing H , the channel H is known at the receiver in which case $I_\xi(X; Y) = \log \det \left(I + \frac{P(J_T)}{m} H Q H^\dagger \right)$ is the mutual information of a point-to-point link with m transmit and n receive antennas, transmit signal to noise ratio $P(J_T)$ and input distribution Gaussian with covariance matrix Q [22]. The dependence of the index at the transmitter is made explicit by writing the transmit SNR as a function of J_T . Note that P_i and $P(i)$ will mean the same thing in this paper.

Let $\Pi(\mathcal{O})$ denote the probability of outage, where \mathcal{O} is the set of all the channels where the maximum supportable rate $I_\xi(X; Y)$ is less than the transmitted rate R [21]. The system is said to have diversity order of d if $\Pi(\mathcal{O}) \doteq \text{SNR}^{-d}$ ¹. Note that all the index mappings, codebooks, rates, powers are dependent on the average signal to noise ratio, SNR. Specifically, the dependence of rate R on SNR is explicitly given by $R = r \log \text{SNR}$, where r is labeled as the multiplexing gain. The diversity-multiplexing tradeoff is then described as the maximum diversity order $d(r)$ that can be achieved for a given multiplexing gain r .

The following result captures the diversity-multiplexing tradeoff for the case of perfect receiver information and no transmitter information. For a given rate $R \doteq r \log \text{SNR}$ and power $P \doteq \text{SNR}^p$, define the outage set $\mathcal{O}(R, P) = \{H : (I + \frac{P}{m} H Q H^\dagger) < R\}$. Denote the result diversity order as $G(r, p)$, which is $\Pi(\mathcal{O}(R, P)) \doteq \text{SNR}^{-G(r, p)}$. The following result completely characterizes $G(r, p)$ and is a straightforward extension of the result in [22].

Lemma 1 ([16]). *The diversity-multiplexing tradeoff for $P \doteq$*

SNR^p and $r \leq p \min(m, n)$ is given by the $(r, G(r, p))$ where

$$G(r, p) = \inf \sum_{i=1}^{\min(m, n)} (2i - 1 + \max(m, n) - \min(m, n)) \alpha_i$$

where the inf is over all $\alpha_1, \dots, \alpha_{\min(m, n)}$ satisfying

$$\{\alpha_1 \geq \dots \alpha_{\min(m, n)} \geq 0, \sum_{i=0}^{\min(m, n)} (p - \alpha_i)^+ < r\}.$$

Remark 1. $G(r, p)$ is a piecewise linear curve connecting the points $(r, G(r, p)) = (kp, p(m - k)(n - k))$, $k = 0, 1, \dots, \min(m, n)$ for fixed m, n and $p > 0$. This follows directly from Lemma 2 of [13]. $G(r, 1)$ is the diversity multiplexing tradeoff with perfect CSIR and the transmit signal to noise ratio $\doteq \text{SNR}$ [22]. Further, $G(r, p) = pG(\frac{r}{p}, 1)$.

D. Summary of Results

We will study the following four systems with different accuracy of channel state information (CSI) at the transmitter and receiver.

- 1) CSIRT_q: In this case, the receiver knowledge about H is assumed to be perfect and the transmitter is assumed to receive a noiseless quantized feedback from the receiver about channel H . This case was first studied in [23] and the diversity order increase was first proved in [17] with later extensions in [13]. The channel quantizer maps H to an index J_R which is then communicated over the noiseless feedback channel. Since the feedback is noiseless, $J_T = J_R$. Based on the index J_T , the transmitter adapts its transmission power as described in Section III. The diversity order for K levels of feedback is defined recursively as $d_{\text{RT}_q}(r, K) = G(r, 1 + d_{\text{RT}_q}(r, K - 1))$ where $d_{\text{RT}_q}(r, 0) = 0$ and grows exponentially in number of feedback bits.
- 2) CSIRT_q: Our first set of results analyze the case of imperfect channel knowledge at the transmitter, where the errors are caused by errors in the received quantized feedback index. In this case, $J_T \neq J_R$ with finite probability. We show that if a MIMO scheme optimized for equally likely input messages is used to send the feedback information, then the diversity order is limited to $d_{\text{RT}_q}(r, K)$ given in Table I. However, the feedback information is not equally-likely. Hence, we propose a natural unequal error protection method labeled *power controlled feedback*, where the power control is performed based on input probabilities. The power-controlled feedback results in a diversity-order $\bar{d}_{\text{RT}_q}(r, K)$, specified in Table I. With power-controlled feedback, the maximum diversity order increases from $d_{\text{RT}_q}(0, K) = 2mn$ to $\bar{d}_{\text{RT}_q}(0, K) = mn(mn + 2)$ for all $K > 2$. For $K = 2$, the maximum diversity order increases from $d_{\text{RT}_q}(0, K) = 2mn$ to $\bar{d}_{\text{RT}_q}(0, K) = mn(mn + 1)$.
- 3) CSIRT_q: Our next step will be to isolate the effect of errors in receiver knowledge of the channel. Thus, the receiver will estimate \hat{H} , which will be mapped to J_R . Since the feedback is assumed to be perfect,

¹We adopt the notation of [22] to denote \doteq to represent exponential equality. We similarly use \prec, \succ, \leq, \geq to denote exponential inequalities.

the transmitter knowledge is same as that of receiver, $J_T = J_R$. We will show that for a three-phase power-controlled training based protocol can achieve a diversity order of $\bar{d}_{\widehat{RT}_q}(r, K) = G(r, 1 + G(r, 1))$. In fact, if the power-controlled training is not performed then channel estimation errors completely dominate and feedback is rendered useless; the resultant tradeoff collapses to $d_{\widehat{RT}_q}(r, K) = G(r, 1)$ for all K . So analogous to power-controlled feedback, power-controlled training appears essential to improve the diversity order in this case.

- 4) CSIRT $_q$: Finally, we put the above two cases together and derive the maximum achievable diversity order as $\bar{d}_{\widehat{RT}_q}(r, K) = G(r, 1 + G(r, 1))$, which is equivalent to the case with only receiver errors. Thus, we conclude that the receiver errors dominate the achievable diversity order.

Remark 2. An important word of caution for all the results (new and known) summarized in Table I. Unlike the previous work in [10, 21], we do not account for the resources spent in channel training and feedback in this paper. Resource accounting can be performed using the procedure developed in [10], by scaling the multiplexing r appropriately. More precisely, the multiplexing r should be replaced by $rT_{\text{coh}}/(T_{\text{coh}} - 1)$ for CSIRT $_q$, $rT_{\text{coh}}/(T_{\text{coh}} - 2m)$ for the three-phase protocol in power-controlled training of CSIRT $_q$, $rT_{\text{coh}}/(T_{\text{coh}} - m)$ for two-phase protocol in constant power training of CSIRT $_q$, and $rT_{\text{coh}}/(T_{\text{coh}} - 2m - 1)$ for the three-phase protocol of CSIRT $_q$. The resource accounting multipliers assumes that the feedback requires one channel use and the training requires m channel uses. The details are further explained in Remarks 3 and 5. Further, the use of the number of antennas to use and the protocols would need to be optimized for each multiplexing as in [10, 21] and is omitted in this paper for readability.

III. CSIRT $_Q$: PERFECT CSIR WITH NOISELESS QUANTIZED FEEDBACK

The diversity-multiplexing tradeoff for the case of receiver with perfect information and noiseless quantized information has been extensively studied in [8, 13, 14]. In this section, we will discuss the main ideas for the single user MIMO channel model stated in Section II-A. We start off with an example to illustrate the main idea.

Example 1 (SISO): Consider the case where $m = n = 1$. Without any feedback, the maximum diversity (at $r \rightarrow 0$) is 1. The space of channels $\{H\} = \mathbb{C}$ can be divided into two sets: the outage set $\mathcal{O}_0 = \{H : \|H\|^2 < \frac{2^R - 1}{\text{SNR}}\}$ and its complement $\bar{\mathcal{O}}_0 = \{H\} \setminus \mathcal{O}_0$; see Figure 3(a). The probability of the set $\Pi(\mathcal{O}_0) \approx \text{SNR}^{-1}$ and $\Pi(\bar{\mathcal{O}}_0) \approx (1 - \text{SNR}^{-1})$.

With one-bit of noiseless feedback, the receiver can convey whether the current channel H belongs to \mathcal{O}_0 or $\bar{\mathcal{O}}_0$. Since the outage event is rare, the transmitter can send much larger power than usual to reduce the outage probability as follows. When the feedback index is $J_R = 1$ representing event \mathcal{O}_0 , the transmitter will use transmit power $\frac{1}{2}\text{SNR}/\Pi(\mathcal{O}_0) \approx \frac{1}{2}\text{SNR}^2$. For the feedback index $J_R = 0$, representing $\bar{\mathcal{O}}_0$, power

$\frac{1}{2}\text{SNR}$ is used. The average power used in the above two-level power control is $\frac{1}{2}\text{SNR}/\Pi(\mathcal{O}_0) \cdot \Pi(\mathcal{O}_0) + \frac{1}{2}\text{SNR} \cdot (1 - \text{SNR}^{-1}) \leq \text{SNR}$. With the above one-bit power control, the outage probability is $\approx \text{SNR}^{-2}$ since the set of channels in outage is reduced to $\mathcal{O}_1 = \left\{H : \|H\|^2 < \frac{(2^R - 1) \cdot 2 \cdot \Pi(\mathcal{O}_0)}{\text{SNR}}\right\} \approx \left\{H : \|H\|^2 < \frac{(2^R - 1) \cdot 2}{\text{SNR}^2}\right\}$; see Figure 3(b).

If the feedback rate was $\log_2(3)$ bits/channel state, then the receiver could convey information about three events $\{\mathcal{O}_1, \mathcal{O}_0 \setminus \mathcal{O}_1, \bar{\mathcal{O}}_0\}$. In state \mathcal{O}_1 , the transmitter can send power $\approx \text{SNR}^3$ since $\Pi(\mathcal{O}_1) \approx \text{SNR}^{-2}$ and thus reduce the outage probability to SNR^{-3} . If we had two bits of feedback or equivalently $K = 4$ levels, then the diversity order will be 4 and the outage region will be given by \mathcal{O}_3 which will be the set of all those channels which could not support rate R with power $\approx \text{SNR}^4$; see Figure 3(c). Using the above recursive argument, for K levels of feedback, a diversity order of $K = 2^b$ can be achieved, where b is the number of feedback bits per channel realization. ■

The above example captures the essence of the general result for the MIMO channels, given by the following theorem.

Theorem 1 ([13]). *Suppose that $K \geq 1$ and $r < \min(m, n)$. Then, the diversity-multiplexing tradeoff for the case of perfect CSIR with noiseless quantized feedback is $d_{RT_q}(r, K)$ defined recursively as*

$$d_{RT_q}(r, K) = G(r, 1 + d_{RT_q}(r, K - 1)),$$

where $d_{RT_q}(r, 0) = 0$.

The main idea of the proof is along the lines of Example 1 and is summarized as follows. Based on its knowledge of H , the receiver decides the feedback index $J_R = J_T \in \{0, 1, \dots, K - 1\}$, where $K = 2^b$ is the number of quantization levels and b is the number of feedback bits. If the transmitter receives feedback index J_T , it sends data at power level P_{J_T} . Without loss of generality, $P_0 \leq P_1 \leq \dots \leq P_{K-1}$. The optimal deterministic index mapping has the following form [13],

$$J_R = \begin{cases} \arg \min_{i \in \mathbb{I}} i, & \mathbb{I} = \{k : \log \det(I + HH^\dagger P_k) \geq R, \\ & k \in \{0, \dots, K - 1\}\} \\ 0, & \text{if the set } \mathbb{I} \text{ is empty} \end{cases}.$$

Based on the above assignment of feedback indices, the optimal power levels can be found out as $P_i \doteq \text{SNR}^{1+p_i}$ where p_i are recursively defined as: $p_0 = 0$, $p_j = G(r, 1 + p_{j-1}) \forall j \geq 1$. Using the above recursion, the optimal diversity is given by $G(r, 1 + p_{K-1})$ which reduces to $d_{RT_q}(r, K)$, thus proving Theorem 1. We note that special cases of Theorem 1 were also proved in [8].

We note that event that $\{H : \log \det(I + HH^\dagger P_{K-1}) < R\}$ corresponds to the outage event, because none of the power levels can be used to reach a mutual information of R . Thus any state can be assigned to the outage event and in fact, the diversity-multiplexing tradeoff is unaffected by which index is used to represent this state. Since the receiver knowledge is perfect, using the index $J_R = 0$ ensures that the overall power consumption is minimized. However, when the receiver

TABLE I
SUMMARY OF DIVERSITY-MULTIPLEXING TRADEOFFS. SEE REMARK 2 FOR A CAUTION IN USING THIS TABLE.

Case	Main Characteristic	D-M Tradeoff
CSIRT	Perfect Information at T and R	$d_{RT}(r) = \infty \forall r < \min(m, n)$
CSIRT _q	Quantized Information at T	$d_{RT_q}(r, K) = G(r, 1 + d_{RT_q}(r, K - 1)), d_{RT_q}(r, 0) = 0$
CSIRT _q [̂]	Noisy information at T (Constant Power Feedback)	$d_{RT_q}(r, K) = \min(B_K(r), mn + G(r, 1))$
CSIRT _q [̄]	Noisy information at T (Power-controlled Feedback)	$\bar{d}_{RT_q}(r, K) = \min(d_{RT_q}(r, K), \max_{q_j \leq 1 + d_{RT_q}(r, j)} \min_{i=1}^{K-1} (mn((q_i)^+ - (q_{i-1})^+) + d_{RT_q}(r, i)))$
CSIRT _q [̂]	Noisy information at R (Constant Power Training)	$d_{RT_q}(r, K) = G(r, 1)$
CSIRT _q [̄]	Noisy information at R (Power-controlled training)	$\bar{d}_{RT_q}(r, K) = G(r, 1 + G(r, 1))$
CSIRT _q ^{̄̂}	No Genie-aided information (Power-controlled training & feedback)	$d_{RT_q}(r, K) = G(r, 1 + G(r, 1))$

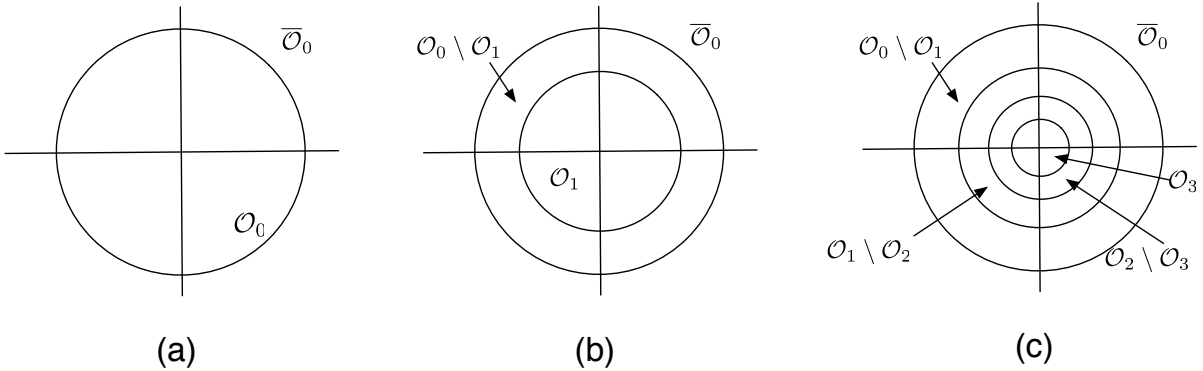


Fig. 3. Example 1: Channel events for the case of (a) no feedback (\mathcal{O}_0 is the set of channels in outage) and (b) one-bit feedback (\mathcal{O}_1 is the outage set) and (c) two-bit feedback (\mathcal{O}_3 is the outage set).

or transmitter knowledge is not perfect (which will be the case in the rest of the paper), we will assign the perceived outage state the highest power level P_{K-1} , which helps reduce the outage probability due to misestimation of channel.

The maximum diversity order in Theorem 1 increases very rapidly with the number of feedback levels K . The maximum diversity order [13] $d_{RT_q}(0, K) = \sum_{g=1}^K (mn)^g$, which grows exponentially fast in the number of levels K . In fact, as $K \rightarrow \infty$, the diversity order also increases unboundedly. Since as $K \rightarrow \infty$, the feedback approaches the perfect feedback and as a result, perfect channel inversion becomes possible. In [3], it was shown that for perfect channel state information at transmitter and receiver, zero outage can be obtained at finite SNR if $\max(m, n) > 1$. Perfect channel inversion essentially converts the fading channel into a vector Gaussian channel, whose error probability goes to zero for all rates less than its capacity. The unbounded growth of diversity order is rather unsettling and is in fact, a fragile result as shown by our results in the following sections.

IV. CSIRT_q[̂]: PERFECT CSIR WITH NOISY QUANTIZED FEEDBACK

In this section, we will analyze the case when the receiver knows the channel H perfectly but the quantization index J_R is conveyed to the transmitter over a noisy feedback link, resulting in the event $J_T \neq J_R$ with non-zero probability. We will consider two feedback designs. In the first design, the feedback channel will use a constant power transmission scheme designed for equally-likely symbols (which is the commonly studied case for i.i.d. data). Learning from the limitations of constant power feedback transmission, we will then construct a new power-controlled feedback strategy which will exploit the unequal probability of different channel events quantized at the receiver.

A. Constant Power Feedback Transmission

First, we observe the impact of using the power control described in Section III by considering Example 1.

Example 2 (Impact of Feedback Errors, SISO): Consider the case of $b = 2$ feedback bits which allows $K = 4$ feedback indices. In this case, the receiver can convey four events, which

we choose to be $\mathcal{O} = \{\bar{\mathcal{O}}_0, \mathcal{O}_0 \setminus \mathcal{O}_1, \mathcal{O}_1 \setminus \mathcal{O}_2, \mathcal{O}_2\}$ (shown in Figure 3(c)), where the events are described as

$$\mathcal{O}_i = \{H : \log(1 + |H|^2 \text{SNR}^{1+i}) < R\}, i = 0, 1, 2. \quad (3)$$

As discussed in the previous section, the power control in the forward channel uses four different power levels, $P_{J_T} \doteq \text{SNR}^{1+J_T}$ for $J_T = 0, 1, 2, 3$. The probability of each of the above events and the associated power control is defined in Table II. The last column in Table II shows the impact of probability of events as seen by the transmitter when there are errors in the feedback link such that $\Pi(J_T = j | J_R = i) \doteq \text{SNR}^{-1}$ for $j \neq i$ as shown in Figure 4(b). For example, with noisy feedback, $\Pi(J_T = 3) = \sum_{j=0}^3 \Pi(J_T = 3 | J_R = j) \Pi(J_R = j) \doteq \text{SNR}^{-1} + \text{SNR}^{-1} \text{SNR}^{-1} + \text{SNR}^{-1} \text{SNR}^{-2} + \text{SNR}^{-3} \doteq \text{SNR}^{-1}$.

In fact, there are two dominant error events, first being $\bar{\mathcal{O}}_0$ ($J_R = 0$) being confused as either $\mathcal{O}_0 \setminus \mathcal{O}_1$ or $\mathcal{O}_1 \setminus \mathcal{O}_2$ or \mathcal{O}_2 ($J_T = 1, 2$ or 3) which amounts to limiting the maximum power that can be used and the second being the probability that $\bar{\mathcal{O}}_1$ ($J_R = 1$) being confused as $\bar{\mathcal{O}}_0$ ($J_T = 0$). The second error event has approximate probability of SNR^{-2} which limits the maximum diversity to 2. Thus, the event probabilities at transmitter are no longer the same in the presence of the feedback errors. As a result, the transmitter *cannot* use the power control $P_{J_T} = \text{SNR}^{1+J_T}$ without exceeding the average power constraint. In fact, the highest power the transmitter can send is of the order $\doteq \text{SNR}^2$, and the maximum diversity order of the constant power feedback mechanism is limited to 2 *irrespective* of the number of feedback bits. ■

We now generalize the above example to MIMO for arbitrary multiplexing gain. Define $\bar{B}_j(r)$ be defined by the recursive equation

$$\bar{B}_j(r) = \begin{cases} 0, & j = 0 \\ G(r, 1 + \min(mn, \bar{B}_{j-1}(r))), & j \geq 1 \end{cases}$$

Theorem 2. [16] *Suppose that $K > 1$ and $r < \min(m, n)$. Then, the diversity-multiplexing tradeoff for the constant power feedback transmission with K indices of feedback is given by*

$$d_{R\hat{T}_q}(r, K) = \min(\bar{B}_K(r), mn + G(r, 1)).$$

Remark 3. *Accounting for the feedback resources can be done as follows. In the limit of high SNR, the feedback will consume one channel use and hence to get the rate $R \doteq r \log(\text{SNR})$, r should be replaced by $rT_{\text{coh}}/(T_{\text{coh}} - 1)$ in the above expression. However, the reader must note that there are implicit time factor terms which can be easily integrated and one may carry out an optimization on the diversity obtained as a function of these time loss terms for each multiplexing. For example, for $r > \min(m, n)(T_{\text{coh}} - 1)/T_{\text{coh}}$, the feedback would not be useful due to the time spent and it is better to use a non-feedback based strategy.*

Much like in Example 2, the maximum diversity order for $r \rightarrow 0$ is $2mn$ for all $K \geq 2$. That is one-bit of feedback is sufficient to achieve the maximum diversity order with constant-power feedback. However, for $r > 0$, as the number

of feedback levels K increases, the diversity order $d_{R\hat{T}_q}(r, K)$ also increases, such that $d_{R\hat{T}_q}(r, K) \leq G(r, 1) + mn$ for all r . Note that $G(r, 1)$ is the diversity order without any feedback. Coincidentally, the diversity order of the feedback link is mn (feedback link is non-coherent where the error probability can decay no faster than SNR^{-mn}), and ends up determining the maximum gain possible beyond $G(r, 1)$.

The key bottleneck in the above result is that the feedback link is using a transmission scheme optimized for equally-likely signals, which is appropriate if the feedback link was being used to send equally-likely messages like data packets. However, the information being conveyed in the feedback link is not equiprobable and hence the usual MIMO schemes optimized for equally-likely messages are not well suited. Again, in the context of Example 2, we will first show an alternate design of *power-controlled feedback* can improve the diversity order in the next section. Before we proceed, we note that our analysis in [10] for a TDD two-way channel yields the same maximum diversity order of $2mn$. It is satisfying to see how two different feedback methods have identical behavior; their relationship is further explored in [24].

B. Power-controlled Feedback

In this section, we exploit the unequal probabilities of the outage events at different power levels to develop a power-controlled feedback scheme to reduce overall outage probabilities. Our feedback transmission will be designed for a non-coherent channel, since the feedback channel H_f is not known at the transmitter or the receiver.

Example 3 (Power-controlled Feedback): For the constant-power feedback transmission scheme described in the Section IV-A, each input is mapped to a codeword with the same power and is pictorially depicted in Figure 5(a). However, since the events in set \mathcal{O} are not equi-likely, we assign different power levels to each event. Now consider the power assignments shown in Table III, also depicted in Figure 5(b) for the feedback channel. The average power over the feedback channel is

$$\text{SNR}^0(1 - \text{SNR}^{-1}) + \text{SNR}^2 \text{SNR}^{-1} + \text{SNR}^3 \text{SNR}^{-2} + \text{SNR}^{-3} \text{SNR}^4 \doteq \text{SNR}$$

Here the rare events are conveyed with more power, which allows a more reliable delivery of the feedback information without violating feedback power constraint. Thus, the power-controlled scheme resembles an amplitude shift keying.

The reason that the probabilities at the transmitter and the receiver are same for power-controlled is because the events $J_T > J_R$ occur with exponentially small probability. The outage probability can be lower bounded by $\Pi(J_T = 0, J_R = 1)$ which is an event when the receiver requested a higher power level than what it received. Note that $\Pi(J_R = 1) \doteq \text{SNR}^{-1}$ and $\Pi(J_T = 0 | J_R = 1) \doteq \text{SNR}^{0-2}$. Hence, $\Pi(J_T = 0, J_R = 1) \doteq \text{SNR}^{-1} \text{SNR}^{-2} = \text{SNR}^{-3}$ is the probability of this dominant error event. Note that both the forward and feedback channel power constraints are satisfied. Hence, we can obtain a diversity order of 3, which is higher than the diversity order

TABLE II

EXAMPLE 2: PROBABILITY OF EVENTS AT THE TRANSMITTER AND RECEIVER WITH AND WITHOUT NOISE IN THE FEEDBACK CHANNEL. RECEIVER KNOWLEDGE IS ASSUMED TO BE PERFECT. (CAUTION: PROBABILITIES ARE ONLY REPORTED UP TO THEIR ORDER, AND CONSTANTS SUCH THAT THEY SUM TO ONE ARE OMITTED.)

Event	Feedback Index (J_R)	Prob of Event at Receiver	Prob at Transmitter (noiseless feedback)	Prob at Transmitter (noisy feedback)
$\bar{\mathcal{O}}_0$	0	$(1 - \text{SNR}^{-1})$	$(1 - \text{SNR}^{-1})$	$(1 - \text{SNR}^{-1})$
$\mathcal{O}_0 \setminus \mathcal{O}_1$	1	SNR^{-1}	SNR^{-1}	SNR^{-1}
$\mathcal{O}_1 \setminus \mathcal{O}_2$	2	SNR^{-2}	SNR^{-2}	SNR^{-1}
\mathcal{O}_2	3	SNR^{-3}	SNR^{-3}	SNR^{-1}

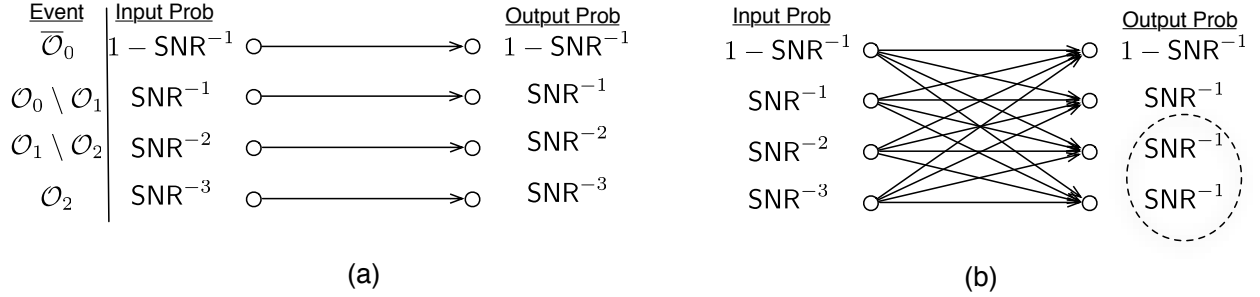


Fig. 4. Example 2: Input-output probabilities of the feedback channel, when (a) the channel is noiseless and (b) when the channel is noisy causing mismatch between transmitter and receiver knowledge, assuming that a code optimized for equally-likely source is used. The noise in the feedback channel changes the output probabilities, as shown by the circled output probabilities. (Note that the sum of probabilities is more than one since we have omitted the constants for ease of understanding).

TABLE III

EXAMPLE 3: POWER ASSIGNMENT FOR NOISY FEEDBACK CHANNEL. (CAUTION: PROBABILITIES ARE ONLY REPORTED UP TO THEIR ORDER, AND CONSTANTS SUCH THAT THEY SUM TO ONE ARE OMITTED.)

Event	Feedback Index	Prob at Receiver	Feedback Transmit Power	Prob at Transmitter (noisy feedback)
$\bar{\mathcal{O}}_0$	0	$(1 - \text{SNR}^{-1})$	SNR^0	$(1 - \text{SNR}^{-1})$
$\mathcal{O}_0 \setminus \mathcal{O}_1$	1	SNR^{-1}	SNR^2	SNR^{-1}
$\mathcal{O}_1 \setminus \mathcal{O}_2$	2	SNR^{-2}	SNR^3	SNR^{-2}
\mathcal{O}_2	3	SNR^{-3}	SNR^4	SNR^{-3}

of 2 obtained via constant-power feedback design (Example 2) but lower than 4 that can be achieved with a noiseless feedback channel (Example 1). ■

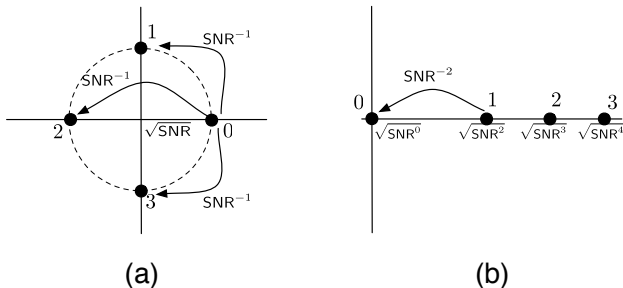


Fig. 5. Examples 2 and 3: Dominant error events when (a) the feedback channel uses a transmission suited for equally-likely messages and (b) a power-controlled feedback design where codewords are power-controlled based on their probability of occurrence.

We will generalize the above example to the case of general MIMO channels by using the following power-control for the feedback channel. Let the feedback power level for different

feedback messages be $Q_0 \doteq \text{SNR}^{q_0}$, $Q_1 \doteq \text{SNR}^{q_1}$, ..., $Q_{K-1} \doteq \text{SNR}^{q_{K-1}}$ with $q_j > q_{j-1}$. Assume for now that the set of powers $\{Q_i\}$ satisfy the feedback power constraint of SNR.

We will use the maximum a posteriori probability (MAP) detection rule to detect the message transmitted by the receiver on the feedback channel. If the transmitted signal is at signal to noise ratio level of Q , the received power is $S_f \doteq \text{trace}(HH^\dagger Q + WW^\dagger + \sqrt{Q}(HW^\dagger + WH^\dagger))$. Since we are encoding the feedback index information in signal power, we will compare the received power S_f with different threshold power levels. We first note that for all $q_i \leq 0$, the received power will be dominated by the noise term $\text{trace}(WW^\dagger)$, which is $\doteq 1$ with high probability. Thus, we assume that $q_j > 0$ for $j > 0$.

Suppose the eigenvalues of HH^\dagger are $(\lambda_1, \dots, \lambda_{m_N})$ and $\lambda_i \doteq \text{SNR}^{-\alpha_i}$ where α_i are the negative SNR exponents of the λ_i , $m_N = \min(m, n)$ and $\alpha = (\alpha_1, \dots, \alpha_{m_N})$. Then, the distribution of α_i is given by the following result.

Lemma 2. [22] Assume $\alpha_1 \geq \alpha_2 \geq \dots \geq \alpha_{m_N}$ are the power

exponents as described above. In the limit of high SNR, the probability density function of the SNR exponents, α , of the eigenvalues of HH^\dagger is given by

$$p(\alpha) \doteq \prod_{i=1}^{mN} \text{SNR}^{-(2i-1+|n-m|)\alpha_i} \mathbf{1}_{\min(\alpha) \geq 0}. \quad (4)$$

We now show that the thresholds for MAP decoding are $\text{SNR}^{\max(q_0,0)+\epsilon}$, $\text{SNR}^{q_1+\epsilon}$, \dots , $\text{SNR}^{q_{K-2}+\epsilon}$ for some small $\epsilon > 0$. This can be derived by observing reason is that $\Pi(J_T > i | J_R = i)$ decays faster than polynomially² in SNR and is hence $\doteq 0$ as long as the threshold for detecting $J_T = i$ is above $\text{SNR}^{q_i+\epsilon}$ for any $\epsilon > 0$. To see this, for $0 \leq j < K-1$,

$$\begin{aligned} & \Pi(J_T > j | J_R = j) \\ &= \Pi\left(\text{trace}(HH^\dagger \text{SNR}^{\max(q_j,0)}) + 1 \geq \text{SNR}^{\max(q_j,0)+\epsilon}\right) \\ &\doteq \Pi\left(\sum (\text{SNR}^{q_j - \alpha_i}) + 1 \geq \text{SNR}^{\max(q_j,0)+\epsilon}\right) \\ &\doteq \Pi\left(\text{SNR}^{\max(q_j - \min \alpha_i, 0)} \geq \text{SNR}^{\max(q_j,0)+\epsilon}\right) \\ &\doteq \Pi(\max(q_j - \min \alpha_i, 0) \geq \max(q_j, 0) + \epsilon) \end{aligned} \quad (5)$$

We see that for the last expression to occur with finite probability, $\min \alpha_i < 0$ which cannot happen with polynomially decreasing probability by Lemma 2 and hence $\Pi(J_T > j | J_R = j) \doteq 0$ irrespective of $\epsilon > 0$. Now, for MAP detection, we would find a threshold between q_i and q_{i+1} so as to minimize $\Pi(J_T = i+1, J_R = i) + \Pi(J_T = i, J_R = i+1)$. Since the first term $\doteq 0$ and the second term is minimized by choosing ϵ as small as possible, choosing ϵ small enough gives the desired threshold for MAP decoding.

Further, for all $Q \gtrsim \text{SNR}$, $\text{trace}(HH^\dagger Q + WW^\dagger + \sqrt{Q}(HW^\dagger + WH^\dagger)) \doteq \text{trace}(HH^\dagger)Q + 1$. Hence, we obtain $\Pi(J_T = 0 | J_R = 1) = \Pi(\text{trace}(HH^\dagger \text{SNR}^{p_1}) + 1 \lesssim \text{SNR}^{\max(p_0,0)+\epsilon})$. Let λ_i be the eigenvalues of HH^\dagger and $\lambda_i \doteq \text{SNR}^{-\alpha_i}$. Then,

$$\begin{aligned} & \Pi(J_T = 0 | J_R = 1) \\ &= \Pi\left(\text{trace}(HH^\dagger \text{SNR}^{q_1}) + 1 \lesssim \text{SNR}^{\max(q_0,0)+\epsilon}\right) \\ &\doteq \Pi\left(\sum_{i=1}^{\min(m,n)} \text{SNR}^{q_1 - \alpha_i} + 1 < \text{SNR}^{\max(q_0,0)+\epsilon}\right) \\ &\doteq \Pi\left(\text{SNR}^{\max(q_1 - \min \alpha_i, 0)} \lesssim \text{SNR}^{\max(q_0,0)+\epsilon}\right) \\ &\doteq \Pi(\max(q_1 - \min \alpha_i, 0) < \max(q_0, 0) + \epsilon) \\ &\doteq \Pi(\min \alpha_i > q_1 - \max(q_0, 0) - \epsilon) \\ &\doteq \text{SNR}^{-mn(q_1 - \max(q_0,0))} \end{aligned}$$

where the last step follows from Lemma 2. Since $q_1 \leq 1 + G(r, 1)$ (as $\Pi(J_R = 1) \gtrsim \text{SNR}^{-G(r,1)}$) and there is power constraint for feedback link of SNR), the outage probability is lower bounded by $\Pi(J_T = 0, J_R = 1) \geq \text{SNR}^{-G(r,1)} \text{SNR}^{-mn(1+G(r,1)-0)} \doteq \text{SNR}^{-mn(1+G(r,1))-G(r,1)}$. Further, we can use the same technique to find that for any $j < i$,

$$\Pi(J_T = j | J_R = i) \doteq \text{SNR}^{-mn(q_i - \max(q_j,0))}. \quad (6)$$

²Any function $f(\text{SNR})$ which decays faster than any polynomial in SNR, like exponential or super-exponential functions of SNR, are $\doteq 0$ since $\lim_{\text{SNR} \rightarrow \infty} \frac{\log(f(x))}{\log(\text{SNR})} = 0$.

Thus we obtain the following result.

Theorem 3 (Power-controlled Feedback). *Suppose that $K > 1$ and $r < \min(m, n)$. Then, the following diversity-multiplexing tradeoff can be achieved with power-controlled feedback*

$$\bar{d}_{R\hat{T}_q}(r, K) = \min\left(d_{RT_q}(r, K), \max_{q_j \leq 1 + d_{RT_q}(r, j)} \min_{i=1, \dots, K-1} (mn((q_i)^+ - (q_{i-1})^+) + d_{RT_q}(r, i))\right).$$

Proof: The proof is provided in Appendix B. ■

Corollary 1. *The diversity multiplexing tradeoff with one bit of imperfect feedback is same as the diversity multiplexing tradeoff with one bit of perfect feedback. In other words, for $K = 2$, $\bar{d}_{R\hat{T}_q}(r, 2) = G(r, 1 + G(r, 1)) = d_{RT_q}(r, 2)$.*

Proof: For $K = 2$, the optimal choice of q is to maximize $q_1^+ - q_0^+$ which gives optimal choice of $q_0 = 0$ and $q_1 = 1 + G(r, 1)$. Using this for one bit of feedback, diversity of $G(r, 1 + G(r, 1))$ can be achieved with power-controlled feedback which is the optimal considering the upper bound of perfect feedback is $d_{RT_q}(r, 2) = G(r, 1 + G(r, 1))$. ■

Thus, we find that there is no loss of diversity with one imperfect bit of feedback compared to the case of perfect feedback. Hence, power controlled feedback scheme is better than the constant power scheme which was limited to a maximum of $2mn$ diversity even as $r \rightarrow 0$.

Corollary 2. *As $K \rightarrow \infty$ and $r \rightarrow 0$, the maximum diversity that can be obtained with power-controlled feedback is $mn(mn + 2)$, i.e.,*

$$\lim_{K \rightarrow \infty, r \rightarrow 0} \bar{d}_{R\hat{T}_q}(r, K) = mn(mn + 2).$$

Proof: Note that in this case, choosing $q_0 = 0$ and $q_i = 1 + d_{RT_q}(r, i) \approx 1 + mn \frac{mn^i - 1}{mn - 1} \mathbf{1}_{mn > 1} + i \mathbf{1}_{mn = 1}$ gives the optimal diversity multiplexing point for $r \rightarrow 0$ and thus, the diversity is $mn(mn + 2)$. We further note that $K = 3$ is enough to get this point. ■

Thus, even with power-controlled feedback, arbitrary number of feedback bits do not yield unbounded increase in diversity order as in the case of CSIRT_q, where the diversity order increases unbounded with K .

Note that however, we restricted our attention to an ordering $q_j > q_{j-1}$ which can be relaxed giving better results as in the following Lemma. However, for our objective of the achievability for imperfect channel state at the receiver and imperfect channel state at the transmitter, the achievability strategy in Theorem 3 is enough. Further note that the new achievability strategy which relaxes the assumption of $q_j > q_{j-1}$ does give improved diversity, but still the diversity remain bounded with increase of the feedback levels.

Lemma 3. *Let $q_i \geq 0$ and $q_i \neq q_j$ for any $i \neq j$, $0 \leq i, j \leq K-1$. Then, the following diversity can be achieved by power-controlled feedback.*

$$\min \left(d_{RT_q}(r, K), \max_{(q_0, \dots, q_{K-1}) \in Q} \min_{i, j \in \{0, \dots, K-1\}: j < i \text{ and } q_j < q_i} (mn(q_i - q_j) + d_{RT_q}(r, i)) \right),$$

where $Q = \{(q_0, \dots, q_{K-1}) : q_j \leq 1 + \min(d_{RT_q}(r, j), \min_{i \in \{0, \dots, j-1\}: q_i > q_j} (d_{RT_q}(r, i) + (q_i - q_j)mn)) \forall 0 \leq j \leq K-1\}$.

Proof: The proof is a simple generalization of the proof of Theorem 3. The constraint on Q keeps the $\Pi(J_T = i) \text{SNR}^{q_i} \leq \text{SNR}$ for the power constraint and the diversity expression records the possible dominant outage events. ■

C. A Source Coding Interpretation

The idea of power-controlled feedback transmission is akin to source-coding the feedback information. In conventional lossless source coding (e.g. Huffman coding), the currency of representation is bits. There, to minimize average code-length, the codeword length is (approximately) inversely proportional to the probability of an event. Rare events are represented by longer codewords while more frequent events are represented by shorter length codewords, thereby minimizing the average code length.

Analogously, our currency is average transmit power and objective is to minimize average error probability. Thus rare events get higher power and frequent events lower transmit power. Note that there are many power allocations which will meet the power constraint but they will all result in different error probabilities. Our proposed feedback transmit power allocation minimizes the error probability (in asymptotic sense).

V. CSIR \hat{T}_Q : ESTIMATED CSIR WITH NOISELESS QUANTIZED FEEDBACK

In this section, we will consider the case when the receiver obtains its channel information from an MMSE estimate, \hat{H} , using training. As a result, the receiver index J_R is not always equal to the optimal index, J , based on actual channel H . We will model the relationship between J and J_R via an effective channel (described in Section V-C). The feedback link, on the other hand, is assumed to be noiseless. As a result, the transmitter index $J_T = J_R$.

We will analyze two protocols. The first protocol, labeled constant-power training, trains the receiver once at the beginning using a constant power level. We show that feedback, even if noiseless, is completely useless in providing any gains in diversity order compared to a no-feedback system. Inspired by our understanding of the noisy feedback channel in Section IV, we propose a second protocol, labeled *power-controlled training* which utilizes the feedback and trains the receiver twice, where the second training is power-controlled based on feedback information. The second protocol has a higher diversity order than any non-feedback system.

A. Training the receiver

In this subsection, we will consider MMSE channel estimation for a single user MIMO channel. The channel is estimated

using a training signal that is known at the receiver. From the received signal, MMSE estimation is done as in [25] to get an estimate \hat{H} of the original channel H . Let X_T be the training signal of size $m \times N$ for some $m \leq N < T_{\text{coh}}$ that is known at the receiver and transmitter. The transmitter sends X_T and the destination receives $Y_T = HX_T + W_T$ where Y_T is a $n \times N$ received signal and W_T is the additive Gaussian noise with each entry from $CN(0, 1)$. Following [25], the optimal training signal is

$$X_T = \left[\sqrt{(n\mu_0 - 1)^+} I_m \quad 0_{m \times (N-m)} \right],$$

where μ_0 is tuned to satisfy power constraint, I_m denote $m \times m$ identity matrix and $0_{m \times (N-m)}$ represents $m \times (N-m)$ matrix having all entries 0. Hence, $n\mu_0 - 1 = \frac{N\text{SNR}}{m}$. Further, the channel estimate is given by

$$\hat{H} = Y_T (X_T^\dagger X_T + I_N)^{-1} X_T^\dagger = Y_T \begin{bmatrix} \frac{\sqrt{\frac{N\text{SNR}}{m}}}{1 + \frac{N\text{SNR}}{m}} I_m \\ 0_{N-m \times m} \end{bmatrix},$$

which can be rewritten as

$$\hat{H} = H \frac{1}{1 + \frac{m}{N\text{SNR}}} + W_2 \frac{\sqrt{\frac{N\text{SNR}}{m}}}{1 + \frac{N\text{SNR}}{m}} \quad (7)$$

where W_2 is the left $n \times m$ submatrix of W_T . We now note some properties of MMSE estimate. First, it is easy to see that expected value of $(H - \hat{H})\hat{H}$ is zero confirming the orthogonality of the error with the unbiased estimate. Further, the variance of any entry in $(H - \hat{H})$ is $\frac{1}{1 + \frac{N\text{SNR}}{m}}$. Thus, H and \hat{H} are matrices in which each corresponding element is highly correlated with the correlation coefficient between corresponding element of H and \hat{H} is $\rho = \frac{1}{\sqrt{1 + \frac{m}{N\text{SNR}}}}$. Further note that any $N \geq m$ channel uses are equivalent for analyzing the asymptotic performance.

In general, if G and H are correlated with correlation coefficient ρ , the joint probability distribution function of eigenvalues of HH^\dagger and GG^\dagger is given by the following result:

Lemma 4. [26] Consider two $n \times m$ random matrices $H = (h_{ij})$ and $G = (g_{ij})$, $i \in [1, n]$, $j \in [1, m]$, each with i.i.d complex zero-mean unit-variance Gaussian entries, i.e., $E[h_{ij}] = E[g_{ij}] = 0, \forall i, j, E[h_{ij}h_{pq}^\dagger] = E[g_{ij}g_{pq}^\dagger] = \delta_{ip}\delta_{jq}$, where the Kronecker symbol δ_{ij} is 1 or 0 when $i = j$ or $i \neq j$ respectively. Moreover, the correlation among the two random matrices is given by $E[h_{ij}g_{pq}^\dagger] = \rho\delta_{ip}\delta_{jq}, \forall i, j, p, q$, where $\rho = |\rho|e^{j\theta}$ is a complex number with $|\rho| < 1$. Let $n \leq m$ and $\nu = m - n$. The joint probability distribution function of the unordered eigenvalues of HH^\dagger and GG^\dagger is

$$p(\lambda, \hat{\lambda}) = \frac{\exp\left(-\frac{\sum_{k=1}^n \lambda_k + \hat{\lambda}_k}{1 - |\rho|^2}\right) \Delta(\lambda) \Delta(\hat{\lambda})}{n!n! \prod_{j=0}^{n-1} j! (j + \nu)! |\rho|^{mn-n} (1 - |\rho|^2)^n} \times \prod_{k=1}^n (\sqrt{\lambda_k \hat{\lambda}_k})^\nu \det \left| I_\nu \left(\frac{2|\rho| \sqrt{\lambda_k \hat{\lambda}_l}}{1 - |\rho|^2} \right) \right| \quad (8)$$

where $\Delta(\cdot)$ represents n -dimensional Vandermonde determinant, $I_k(\cdot)$ denotes the k^{th} order modified Bessel function of the first kind, the eigenvalues of HH^\dagger and GG^\dagger are given by

$\lambda = (\lambda_1, \dots, \lambda_n)$ and $\hat{\lambda} = (\hat{\lambda}_1, \dots, \hat{\lambda}_n)$ respectively.

Note that although Lemma 4 assumed $n \leq m$, it can be extended to the other case of $m > n$ since nonzero eigenvalues of HH^\dagger and $H^\dagger H$ are the same. Hence for all n and m , let $m_N = \min(m, n)$ and $\nu = |m - n|$. Then, the joint probability density function of the unordered eigenvalues of HH^\dagger and GG^\dagger is

$$p(\lambda, \hat{\lambda}) = \frac{\exp\left(-\frac{\sum_{k=1}^{m_N} \lambda_k + \hat{\lambda}_k}{1 - |\rho|^2}\right) \Delta(\lambda) \Delta(\hat{\lambda})}{m_N! m_N! \prod_{j=0}^{n-1} j! (j + \nu)! |\rho|^{m_N - m_N}} \times \frac{\prod_{k=1}^{m_N} (\sqrt{\lambda_k \hat{\lambda}_k})^\nu \det \left| I_\nu \left(\frac{2|\rho| \sqrt{\lambda_k \hat{\lambda}_k}}{1 - |\rho|^2} \right) \right|}{(1 - |\rho|^2)^{m_N}}. \quad (9)$$

Recall that the eigenvalues of HH^\dagger be $(\lambda_1, \dots, \lambda_{m_N})$, $\lambda_i \doteq \text{SNR}^{-\alpha_i}$ and $\alpha = (\alpha_1, \dots, \alpha_{m_N})$. Similarly, let the eigenvalues of $\hat{H}\hat{H}^\dagger$ be $(\hat{\lambda}_1, \dots, \hat{\lambda}_{m_N})$, $\hat{\lambda}_i \doteq \text{SNR}^{-\hat{\alpha}_i}$ and $\hat{\alpha} = (\hat{\alpha}_1, \dots, \hat{\alpha}_{m_N})$. The distribution of α_i 's is given earlier in Lemma 2. We will now find the joint distribution of α_i 's and $\hat{\alpha}_i$'s. Let $\alpha_1 \geq \alpha_2 \geq \dots \geq \alpha_{m_N}$ and $\hat{\alpha}_1 \geq \hat{\alpha}_2 \geq \dots \geq \hat{\alpha}_{m_N}$. Further, we define

$$E_k = \{(\alpha, \hat{\alpha}) : \min(\alpha_i, \hat{\alpha}_i) \geq 1 \ \forall i = 1, \dots, k, \text{ and } 0 \leq \alpha_i = \hat{\alpha}_i < 1 \ \forall i > k\} \quad (10)$$

for all $0 \leq k \leq m_N$.

Theorem 4. Let H be the channel and \hat{H} be the estimated channel. In the limit of high SNR, the probability density function of the SNR exponents of the eigenvalues of HH^\dagger and $\hat{H}\hat{H}^\dagger$ is given by

$$p(\alpha, \hat{\alpha}) \doteq \sum_{k=0}^{m_N} e_k \mathbf{1}_{E_k} \quad (11)$$

where $\alpha_1 \geq \alpha_2 \geq \dots \geq \alpha_{m_N}$, $\hat{\alpha}_1 \geq \hat{\alpha}_2 \geq \dots \geq \hat{\alpha}_{m_N}$, and

$$e_k = \text{SNR}^{k(n-m+k)} \prod_{i=1}^k \text{SNR}^{-(2i-1+|n-m|)\hat{\alpha}_i} \times \prod_{i=1}^{m_N} \text{SNR}^{-(2i-1+|n-m|)\alpha_i}. \quad (12)$$

Proof: Since all the results were symmetric about interchanging m and n , without loss of generality, we take $m \leq n$ for the purpose of this proof and the rest of the paper. The proof is provided in Appendix A. ■

Remark 4. Since the receiver is trained at SNR, we note that for all $\alpha_i < 1$, $\alpha_i = \hat{\alpha}_i$ with probability 1, which means that the channel estimate is a reliable proxy for the actual channel. On the other hand, if $\alpha_i \geq 1$, all we can state is that $\hat{\alpha}_i \geq 1$ with probability 1. For example, if $\hat{\alpha}_i \geq 100$, all we can reliably say about α_i is that it is ≥ 1 . In SISO case, the above property of $\hat{\alpha}_i$ implies that the channel cannot be resolved below the noise floor since the noise dominates the training signal. The interesting implication in MIMO is that this result of noise dominance holds for all the eigen-values. None of the eigen-value of the channel can be resolved beyond $\alpha_i \geq 1$ if $\hat{\alpha}_i \geq 1$.

Example 4 (Asymptotic Distribution): For the case of $m = n = 1$, which is our running example, $p(\alpha, \hat{\alpha}) \doteq$

$\text{SNR}^{-\alpha} \mathbf{1}_{0 \leq \alpha = \hat{\alpha} < 1} + \text{SNR}^{1-\alpha-\hat{\alpha}} \mathbf{1}_{\min(\alpha, \hat{\alpha}) \geq 1}$. This density function will be used to analyze the diversity tradeoffs in this section.

B. Constant-power Training

We first note the decoding scheme with trained channel estimate at the receiver. If the receiver has estimate \hat{H} trained with the power of SNR^p for some $p \geq 0$, the estimation error variance is SNR^{-p} . As $Y = HX + W = \hat{H}X + (H - \hat{H})X + W$, a lower bound on mutual information can be considered assuming that $(H - \hat{H})X + W$ is Gaussian noise and hence the expression for the coherent channel with actual channel as \hat{H} can be used as a lower bound [21].

The protocol is divided into three phases as described below:

Phase 1: Training from Tx to Rx: The training is done using transmit power of SNR to obtain the channel estimate \hat{H} . On the basis of this training, the receiver decides a feedback level J_R in the following way. Suppose that there are K power levels $P_i \doteq \text{SNR}^{1+p_i}$, $i \in \{0, \dots, K-1\}$. The feedback index chosen at the receiver is given by

$$J_R = \begin{cases} \arg \min_{i \in \mathbb{I}} i, & \mathbb{I} = \{k : \log \det(I + \hat{H}\hat{H}^\dagger P_k) \geq R, \\ & k \in \{0, \dots, K-1\}\} \\ K-1, & \text{if the set } \mathbb{I} \text{ is empty} \end{cases}.$$

Phase 2: The feedback index $J_R \in \{0, \dots, K-1\}$ is sent to the transmitter via the noiseless feedback channel. Thus, the feedback index received by the transmitter is $J_T = J_R$.

Phase 3: The transmitter receives a feedback power level J_T and sends data at power level $P_{J_T} \doteq \text{SNR}^{1+p_{J_T}}$.

Example 5 (Constant-power training): The only error in channel knowledge appears in the first phase, where the receiver is trained.

Note that for MMSE estimation \hat{H} and $\tilde{H} = H - \hat{H}$ are uncorrelated and since the receiver is trained with power of SNR, the variance of \tilde{H} is $\doteq 1/\text{SNR}$. Let $|\hat{H}|^2 = \text{SNR}^{-\hat{\alpha}}$ and $|\tilde{H}|^2 = \text{SNR}^{-1} \text{SNR}^{-\tilde{\alpha}}$ where the probability distribution function of $\hat{\alpha}$ and $\tilde{\alpha}$ is $p(\hat{\alpha}, \tilde{\alpha}) = \text{SNR}^{-\hat{\alpha}-\tilde{\alpha}} \mathbf{1}_{\hat{\alpha} \geq 0} \mathbf{1}_{\tilde{\alpha} \geq 0}$. The probability of outage is

$$\begin{aligned} \Pi(\mathcal{O}) &= \sum_{j=0}^{K-1} \Pi(\mathcal{O}, J_T = j) \\ &\geq \Pi(\mathcal{O}, J_T = 1) \\ &= \Pi(\mathcal{O}, J_R = 1) \\ &\geq \Pi(\mathcal{O}, \log(1 + |\hat{H}|^2 \text{SNR}^2) \geq r \log \text{SNR}, \\ &\quad \log(1 + |\tilde{H}|^2 \text{SNR}) < r \log \text{SNR}) \\ &\doteq \Pi\left(\log\left(1 + \frac{|\hat{H}|^2 \text{SNR}^2}{1 + \text{SNR}^2 |\tilde{H}|^2}\right) < r \log \text{SNR}, \right. \\ &\quad \left. \log(1 + |\hat{H}|^2 \text{SNR}^2) \geq r \log \text{SNR}, \right. \\ &\quad \left. \log(1 + |\tilde{H}|^2 \text{SNR}) < r \log \text{SNR}\right) \\ &\doteq \Pi((2 - \hat{\alpha} - (1 - \tilde{\alpha}))^+ < r, (2 - \hat{\alpha})^+ \geq r, \\ &\quad (1 - \tilde{\alpha})^+ < r) \\ &\doteq \text{SNR}^{-\min(\hat{\alpha}, \tilde{\alpha}) \in A(\hat{\alpha} + \tilde{\alpha})}, \end{aligned} \quad (13)$$

where $A = \{(\hat{\alpha}, \tilde{\alpha}) : \hat{\alpha} \geq 0, \tilde{\alpha} \geq 0, (2 - \hat{\alpha} - (1 - \tilde{\alpha})^+)^+ < r, (2 - \hat{\alpha})^+ \geq r, (1 - \tilde{\alpha})^+ < r\}$.

Hence, substituting $\tilde{\alpha} = 0$ and $\hat{\alpha} = 1 - r + \delta$ for δ small enough (Note that this choice of $(\hat{\alpha}, \tilde{\alpha})$ is in A .) gives a bound on the above probability as:

$$\begin{aligned} \Pi(\mathcal{O}) &\geq \text{SNR}^{-\min_{(\hat{\alpha}, \tilde{\alpha}) \in A} (\hat{\alpha} + \tilde{\alpha})} \\ &\geq \text{SNR}^{-(1-r)}. \end{aligned} \quad (14)$$

Thus, we see that the diversity of this scheme is at most same as the one without feedback. Hence, there is no advantage of feedback. ■

Theorem 5. Suppose that $K > 1$ and $r < \min(m, n)$. Then, the diversity-multiplexing tradeoff is given by $d_{\tilde{R}T_q}(r, K) = G(r, 1)$.

Proof: We will consider the third phase in this case, where we see that any increase in power levels do not help to increase the diversity. Achievability follows by not using any feedback. We will prove the converse here. Note that for MMSE estimation \hat{H} and $\tilde{H} = H - \hat{H}$ are uncorrelated and since the receiver is trained with power of SNR, the variance of \tilde{H} is $\doteq 1/\text{SNR}$. Let $\tilde{\lambda}_i$ be the eigenvalues of $\tilde{H}\tilde{H}^\dagger$ and $\hat{\lambda}_i \doteq \text{SNR}^{-\hat{\alpha}_i}$. Further, assume that $\tilde{\lambda}_i$ be the eigenvalues of $\tilde{H}\tilde{H}^\dagger$ and $\tilde{\lambda}_i \doteq \text{SNR}^{-1}\text{SNR}^{-\tilde{\alpha}_i}$. The probability distribution function of $\hat{\alpha} = (\hat{\alpha}_1, \dots, \hat{\alpha}_m)$ and $\tilde{\alpha} = (\tilde{\alpha}_1, \dots, \tilde{\alpha}_m)$ is $p(\hat{\alpha}, \tilde{\alpha}) = \prod_{i=1}^m \text{SNR}^{-(2i-1+|n-m|)(\hat{\alpha}_i+\tilde{\alpha}_i)} \mathbf{1}_{\min(\hat{\alpha}) \geq 0} \mathbf{1}_{\min(\tilde{\alpha}) \geq 0}$. The probability of outage is

$$\begin{aligned} \Pi(\mathcal{O}) &= \sum_{j=0}^{K-1} \Pi(\mathcal{O}, J_T = j) \\ &\geq \Pi(\mathcal{O}, J_T = 1) \\ &= \Pi(\mathcal{O}, J_R = 1) \\ &\geq \Pi(\mathcal{O}, \log \det(I + \hat{H}\hat{H}^\dagger P_1) \geq r \log \text{SNR}, \\ &\quad \log \det(I + \hat{H}\hat{H}^\dagger P_0) < r \log \text{SNR}) \\ &\doteq \Pi \left(\log \det \left(I + \frac{\hat{H}\hat{H}^\dagger P_1}{1 + P_1 \text{trace}(\tilde{H}\tilde{H}^\dagger)} \right) \right. \\ &\quad < r \log \text{SNR}, \\ &\quad \log \det(I + \hat{H}\hat{H}^\dagger P_1) \geq r \log \text{SNR}, \\ &\quad \left. \log \det(I + \hat{H}\hat{H}^\dagger P_0) < r \log \text{SNR} \right) \\ &\doteq \Pi \left(\sum_{i=1}^m (1 + p_1 - \hat{\alpha}_i - (p_1 - \min \tilde{\alpha})^+)^+ < r, \right. \\ &\quad \sum_{i=1}^m (1 + p_1 - \hat{\alpha}_i)^+ \geq r, \\ &\quad \left. \sum_{i=1}^m (1 + p_0 - \hat{\alpha}_i)^+ < r \right) \\ &\doteq \text{SNR}^{-\min_{(\hat{\alpha}, \tilde{\alpha}) \in A} \sum_{i=1}^m ((2i-1+|n-m|)(\hat{\alpha}_i+\tilde{\alpha}_i))} \end{aligned} \quad (15)$$

where $A = \{(\hat{\alpha}, \tilde{\alpha}) : \min \hat{\alpha} \geq 0, \tilde{\alpha} \geq 0, \sum_{i=1}^m (1 + p_1 - \hat{\alpha}_i - (p_1 - \min \tilde{\alpha})^+)^+ < r, \sum_{i=1}^m (1 + p_1 - \hat{\alpha}_i)^+ \geq r, \sum_{i=1}^m (1 + p_0 - \hat{\alpha}_i)^+ < r\}$. Hence, substituting $\tilde{\alpha}_i = 0$ and allocating $\hat{\alpha}$ as in [13] so that $\sum_{i=1}^m (1 + p_0 - \hat{\alpha}_i)^+ = r - \delta$ for δ small

enough gives a bound on the above probability as:

$$\begin{aligned} \Pi(\mathcal{O}) &\geq \text{SNR}^{-\min_{(\hat{\alpha}, \tilde{\alpha}) \in A} \sum_{i=1}^m ((2i-1+|n-m|)(\hat{\alpha}_i+\tilde{\alpha}_i))} \\ &\geq \text{SNR}^{-G(r, 1+p_0)} \\ &\geq \text{SNR}^{-G(r, 1)}. \end{aligned} \quad (16)$$

In the last step, $p_0 \leq 0$ else the transmit power constraint cannot be satisfied. Thus, we see that the diversity of this scheme is at most same as the one without feedback. Hence, there is no advantage of feedback for constant power training. This was observed for $r \rightarrow 0$ in [27]. ■

Remark 5. Analogous to Remark 3, the accounting for the training resources can be done. In the limit of high SNR, the training will consume m channel uses to train m antennas (as also seen in Section V-A) and hence to get the rate $R \doteq r \log(\text{SNR})$, r should be replaced by $rT_{\text{coh}}/(T_{\text{coh}} - m)$ in the above expression. However, these terms will be omitted in the sequel and can be similarly integrated. Much like in Remark 3, one may carry out an optimization on the diversity obtained as a function of these time loss terms for each multiplexing to optimize over the number of antennas that need to be trained like in [10, 21].

C. Power-controlled Training

As we showed in the previous section, the receiver estimate was not good enough to help improve the diversity with feedback. We now propose a power-controlled training protocol which can improve the outage performance using feedback. The protocol is again divided into three phases as described below:

Phase 1: The transmitter sends the training signal using power SNR which is used at the receiver to obtain channel estimate \hat{H} . On the basis of this training, the receiver decides a feedback level J_R in the following way. Suppose that there are K power levels $P_i \doteq \text{SNR}^{1+p_i}$, $i \in \{0, \dots, K-1\}$. (We will give the exact constants in front while proving achievability.) The feedback index is

$$J_R = \begin{cases} \arg \min_{i \in \mathbb{I}} i, & \mathbb{I} = \{k : \log \det(I + \hat{H}\hat{H}^\dagger P_k) \geq \\ & R + \epsilon \log(\text{SNR}), \\ & k \in \{0, \dots, K-1\}\} \\ K-1, & \text{if the set } \mathbb{I} \text{ is empty} \end{cases},$$

where $\epsilon > 0$ is some small constant chosen. We will later substitute $\epsilon \rightarrow 0$.

Phase 2: A feedback level $J_R \in \{0, \dots, K-1\}$ is fed back over the noiseless feedback channel to the transmitter, which implies $J_T = J_R$.

Phase 3: If the feedback index J_T is received, the transmitter trains the receiver *again* at power level P_{J_T} which is followed by the data at power level P_{J_T} . The trained channel estimate is denoted by \hat{H}_2 and let $\tilde{H} = H - \hat{H}_2$.

The outage probability is defined as: $\Pi(\mathcal{O}) \triangleq$

$$\Pi \left(\log \det \left(I + \frac{P(J_T)}{m} \frac{\hat{H}_2 Q \hat{H}_2^\dagger}{1 + \frac{\text{SNR}}{mn} \text{trace}(\tilde{H}\tilde{H}^\dagger)} \right) < R \right) \quad (17)$$

TABLE IV

EXAMPLE 6: POWER ASSIGNMENT FOR CSIR \hat{T}_Q . (CAUTION: PROBABILITIES ARE ONLY REPORTED UP TO THEIR ORDER, AND CONSTANTS SUCH THAT THEY SUM TO ONE ARE OMITTED.)

Event	Prob at Receiver	Training and Transmit Power
$\hat{\mathcal{O}}_0$	$1 - \text{SNR}^{-1}$	SNR^1
$\hat{\mathcal{O}}_0 \setminus \mathcal{O}_1$	SNR^{-1}	SNR^2
$\hat{\mathcal{O}}_1 \setminus \mathcal{O}_2$	SNR^{-2}	SNR^3
$\hat{\mathcal{O}}_2$	SNR^{-3}	SNR^4

This is the effective outage probability using Gaussian codebooks and considering the channel estimation error as noise [10].

Example 6 (Power-controlled Training): Since the receiver now does not know the value of channel estimate H , but only an estimate \hat{H} , the events $\hat{\mathcal{O}}_i$ can in this case be defined as

$$\hat{\mathcal{O}}_i = \left\{ \hat{H} : \log \left(1 + |\hat{H}|^2 \text{SNR}^{1+i} \right) < R \right\}, i = 0, 1, 2. \quad (18)$$

Figure 6 depicts the relation between J and J_R . Let α be the negative SNR exponent of $|H|^2$ while $\hat{\alpha}$ be the negative SNR exponent of $|\hat{H}|^2$. Let

$$J = \begin{cases} \arg \min_{i \in \mathbb{I}} i, & \mathbb{I} = \{k : \log(1 + |H|^2 P_k) \geq R, \\ & k \in \{0, \dots, 3\}\} \\ 0, & \text{if the set } \mathbb{I} \text{ is empty} \end{cases}.$$

We assume that 3^{rd} phase is perfect for this example and hence $\tilde{H} = 0$. We will prove later that this interference error in (17) due to \tilde{H} does not make a difference asymptotically.

Consider the event ($J_R < 2, J = 2$) which would result in outage. The powers and the probabilities can be seen in Table IV. The outage probability for $r \rightarrow 0$ is then

$$\begin{aligned} \Pi(\mathcal{O}) &\geq \Pi(J_R < 2, J = 2) \\ &= \Pi(\log \det(I + \hat{H}\hat{H}^\dagger P_1) \geq R, \\ &\quad \log(1 + |H|^2 P_1) < R, \\ &\quad \log(1 + |H|^2 P_2) \geq R) \\ &\geq \Pi(2 - \hat{\alpha} > 0, 2 - \alpha < 0, 3 - \alpha > 0) \\ &\doteq \text{SNR}^{\max_{2 < \alpha < 3, 1 < \hat{\alpha} < 2(1 - \alpha - \hat{\alpha})}} \\ &\doteq \text{SNR}^{1-2-1} \doteq \text{SNR}^{-2} \end{aligned} \quad (19)$$

Thus, the maximum diversity order is 2 with any number of feedback levels. Note that this is more than constant power training which was limited to 1. We will show later in this Section that this can be achieved with a single feedback bit. ■

The above example can be generalized to MIMO systems as follows.

Theorem 6. For $K > 1$ and $r < \min(m, n)$, the diversity-multiplexing tradeoff of $\bar{d}_{\hat{R}T_q}(r, K) = G(r, 1 + G(r, 1))$ can be achieved with power-controlled training. Further, the above is optimal for zero multiplexing.

Proof: The proof of this Theorem is provided in Appendix

C. We will first show that the diversity cannot be greater than $mn(mn + 1)$ and later prove that the diversity multiplexing tradeoff of $G(r, 1 + G(r, 1))$ can be achieved for $K = 2$. ■

Note that $\bar{d}_{\hat{R}T_q}(r, K) = d_{RT_q}(r, 2)$ which means that diversity with imperfect receiver information with any number of feedback levels $K \geq 2$ is same as the diversity with perfect receiver information with 1 bit of feedback.

VI. CSIR \hat{T}_Q : ESTIMATED CSIR WITH NOISY QUANTIZED FEEDBACK

We observed in Section V that with CSIR obtained by MMSE training and perfect feedback, the diversity-multiplexing tradeoff of $\bar{d}_{\hat{R}T_q}(r, K) = G(r, 1 + G(r, 1))$ can be achieved with 1 bit of noiseless feedback. We also observed that the diversity-multiplexing tradeoff of $\bar{d}_{\hat{R}T_q}(r, K) = G(r, 1 + G(r, 1))$ is also achievable with 1 bit of noisy feedback with perfect CSIR. In this section, we will show that diversity-multiplexing tradeoff $\bar{d}_{\hat{R}T_q}(r, 2)$ can be achieved when both imperfections are present simultaneously: 1 bit of imperfect feedback based on noisy training-based receiver information. Thus, our main result is

Theorem 7. For $K = 2$ and $r < \min(m, n)$, an achievable diversity-multiplexing tradeoff is given by $\bar{d}_{\hat{R}T_q}(r, 2) = G(r, 1 + G(r, 1))$.

A. Protocol

The complete protocol with both power-controlled training and power-controlled feedback constitutes of three phases as described below.

Phase 1: The training is done using power SNR to get the channel estimate \hat{H} . On the basis of this training, the receiver decides a feedback level J_R in the following way. Since the feedback is assumed to be only one bit, there are only two power levels at the transmitter. Denote the two power levels as $P_i \doteq \text{SNR}^{1+p_i}, i \in \{0, 1\}$ with $p_0 = 0$ and $p_1 = G(r, 1)$. (We will give the exact constants in front while proving that the average power constraint will be satisfied.) The feedback index is

$$J_R = \begin{cases} 1 & \text{if } \log \det \left(I + \hat{H}\hat{H}^\dagger P_0 \right) < R + \epsilon \log(\text{SNR}) \\ 0 & \text{otherwise} \end{cases},$$

where $\epsilon > 0$ is an arbitrarily small constant chosen as before. The above index assignment is simply choosing the higher power level if the lower power level is estimated to be too low.

Phase 2: A feedback level $J_R \in \{0, 1\}$ is transmitted from the receiver which is received at the transmitter as $J_T \in \{0, 1\}$. The receiver employs the power-controlled encoding scheme described in Section IV to send the feedback index. The power levels used for sending the feedback information are denoted as $Q_0 = 0$ and $Q_1 \doteq \text{SNR}^{1+G(r, 1)}$.

Phase 3: The transmitter gets a feedback index J_T which is used to train the receiver at power level P_{J_T} and then send data at power level P_{J_T} . The trained channel estimate is denoted \hat{H}_2 and let $\tilde{H} = H - \hat{H}_2$.

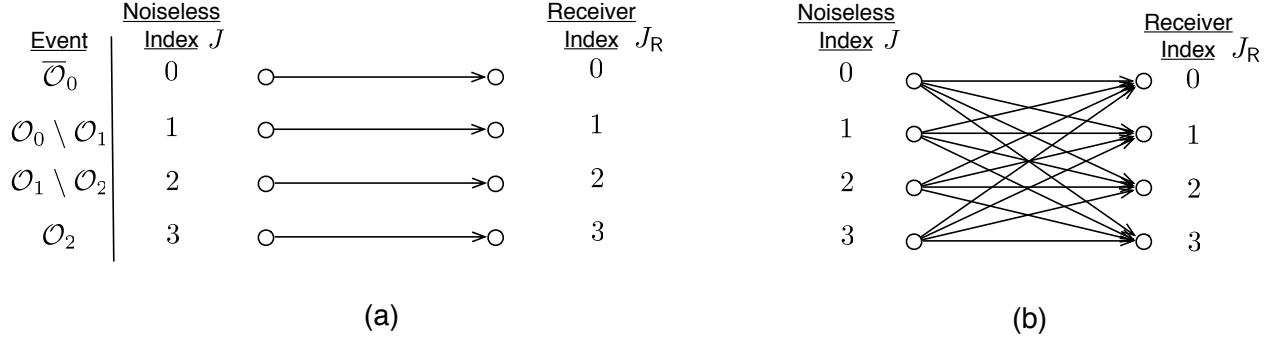


Fig. 6. Example 6: Obtaining channel information at the receiver: (a) if the receiver has perfect information, then it knows the correct feedback index, else (b) the receiver index is not known perfectly and can be viewed as an output of a noisy channel.

The outage probability of the above three-phase protocol is upper bounded by

$$\Pi(\mathcal{O}) = \Pi \left(\log \det \left(I + \frac{P_{J_T}}{m} \frac{\hat{H}_2 Q \hat{H}_2^\dagger}{1 + \frac{P_{J_T}}{mn} \text{trace}(\tilde{H} \tilde{H}^\dagger)} \right) < R \right). \quad (20)$$

The above expression for the effective outage probability considers the channel estimation error as noise [10] and hence is only an upper bound for the optimal scheme. Following identical steps as in Section V, we find that the interference error due to \tilde{H} does not impact the analysis asymptotically and can thus be ignored. Hence,

$$\Pi(\mathcal{O}) \doteq \Pi \left(\log \det \left(I + P(J_T) \hat{H}_2 \hat{H}_2^\dagger \right) < R \right). \quad (21)$$

Now define J as

$$J = \begin{cases} 1 & \text{if } \log \det \left(I + \hat{H}_2 \hat{H}_2^\dagger P_0 \right) < R \text{ and} \\ & \log \det \left(I + \hat{H}_2 \hat{H}_2^\dagger P_1 \right) \geq R \\ 0 & \text{otherwise} \end{cases}.$$

Using the analysis as in Sections III and IV, we observe that

$$\begin{aligned} \Pi(J_T = 0 | J_R = 1) &\doteq \text{SNR}^{-mn(1+G(r,1))}, \\ \Pi(J_T = 1 | J_R = 0) &\doteq 0, \\ \Pi(J_R = 0, J = 1) &\doteq 0. \end{aligned}$$

We now split $\Pi(\mathcal{O})$ in (21) into 8 terms depending on the values of J , J_R and J_T as follows,

$$\begin{aligned} \Pi(\mathcal{O}) &\doteq \Pi(\log \det(I + P(J_T) \hat{H}_2 \hat{H}_2^\dagger) < R) \\ &\doteq \sum_{i=0}^1 \sum_{j=0}^1 \sum_{k=0}^1 \Pi(\log \det(I + P(J_T) \hat{H}_2 \hat{H}_2^\dagger) < R, \\ &\quad J = i, J_R = j, J_T = k). \end{aligned} \quad (22)$$

Note that the terms with $J_R \neq J_T$ can be asymptotically upper bounded $\text{SNR}^{-G(r,1+G(r,1))}$. Also, the term corresponding to $i = j = k = 0$ can be upper bounded by $\text{SNR}^{-G(r,1+G(r,1))}$ since when $J = 0$, $\log \det(I + P(0) \hat{H}_2 \hat{H}_2^\dagger) < R$ happens when $\log \det(I + P_1 \hat{H}_2 \hat{H}_2^\dagger) < R$. Further, $J_R = 0$ and $J = 1$ happens with probability $\doteq 0$. Thus, the only remaining case

is when $J_T = J_R = 1$. Hence,

$$\begin{aligned} \Pi(\mathcal{O}) &\doteq \sum_{i=0}^1 \sum_{j=0}^1 \sum_{k=0}^1 \Pi(\log \det(I + P(J_T) \hat{H}_2 \hat{H}_2^\dagger) < R, \\ &\quad J = i, J_R = j, J_T = k) \\ &\leq \text{SNR}^{-G(r,1+G(r,1))} \\ &\quad + \Pi(\log \det(I + P(J_T) \hat{H}_2 \hat{H}_2^\dagger) < R, \\ &\quad J_T = 1, J_R = 1) \\ &\leq \text{SNR}^{-G(r,1+G(r,1))} \\ &\quad + \Pi(\log \det(I + P_1 \hat{H}_2 \hat{H}_2^\dagger) < R) \\ &\doteq \text{SNR}^{-G(r,1+G(r,1))}. \end{aligned} \quad (23)$$

Hence, the diversity order of $G(r, 1+G(r, 1))$ for multiplexing gain r can be achieved with only one bit of feedback. We now show that the power constraint is also satisfied which completes the proof. Recall that the power levels at the transmitter are denoted by P_0 and P_1 , while at the receiver are denoted by Q_0 and Q_1 .

Let $P_0 = \frac{\text{SNR}}{2}$ and $P_1 = \frac{\text{SNR}}{4(\Pi(J_R=1))}$. The average power used at the transmitter is $P_0 \Pi(J_T = 0) + P_1 \Pi(J_T = 1)$. To show that this average power $\leq \text{SNR}$, it is enough to prove that

$$\Pi(J_T = 1) \leq 2(\Pi(J_R = 1)). \quad (24)$$

Thus, the left hand side

$$\begin{aligned} \Pi(J_T = 1) &= \Pi(J_T = 1, J_R = 0) + \Pi(J_T = 1, J_R = 1) \\ &\leq \Pi(J_T = 1 | J_R = 0) + \Pi(J_R = 1). \end{aligned} \quad (25)$$

Note that the first decays faster with SNR than $\Pi(J_R = 1)$, thus the above $\leq 2\Pi(J_R = 1)$.

Further, let $Q_0 = 0$ and $Q_1 = \frac{\text{SNR}}{2(\Pi(J_R=1))}$. The average power used at the receiver is $Q_0 \Pi(J_R = 0) + Q_1 \Pi(J_R = 1) \leq \text{SNR}$.

Thus, diversity order of $G(r, 1+G(r, 1))$ can be achieved with imperfect feedback and imperfect CSIR. Hence, one bit of imperfect feedback and imperfect CSIR is same as one bit of perfect feedback and perfect CSIR except the time losses in trainings and the feedback.

VII. NUMERICAL RESULTS

First consider the case of $m = 1$ and $n = 2$. The diversity multiplexing tradeoff in the various cases for 1 or 2 bits of feedback can be seen in Figure 7. We will now go through all the different tradeoff curves in the order of the legend from top to bottom. The first line $G(r, 1)$ represents the diversity obtained with no feedback (CSIR and CSIR have identical performance since time lost in training is not accounted in our expressions) and also the diversity obtained in the case when the feedback is perfect and the receiver is trained through a constant power symbol (CSIRT_q with constant power training). The second piecewise linear curve represents the diversity obtained when the receiver knows perfect channel state information but one bit of constant power feedback is sent over a noisy feedback channel (CSIRT_q with constant power feedback). The third line $G(r, 1 + G(r, 1))$ represents the diversity obtained with 1 bit of perfect feedback (CSIRT_q). This also represents the diversity obtained when receiver knows the channel perfectly while 1 bit of power-controlled feedback is provided on a feedback channel (CSIRT_q with power-controlled feedback). This diversity is also obtained when the feedback link is perfect while the receiver is trained using power-controlled training symbols (CSIRT_q with power-controlled training). Further, the diversity obtained when the receiver does not know channel state information and the feedback link is noisy is also $G(r, 1 + G(r, 1))$ (CSIRT_q with power-controlled training and power-controlled feedback). Note that the second curve started from diversity of 4 at zero multiplexing, but then followed the third line after $r \approx 0.5$ since it cannot perform better than the perfect feedback case.

Next, we consider three levels of feedback. More bits of feedback do not increase the diversity in any case except when the receiver knows the channel perfectly. In this case, the next two curves lines show the effect of constant power feedback and the power controlled feedback on the diversity. The last line shows the performance with 3 levels of perfect feedback. The diversity multiplexing tradeoff curve achieved with constant power starts from 4 and hits the line of 3 levels of perfect feedback.

We will now see the effect of increase on SNR on the outage probabilities at a constant multiplexing gain. We will focus on 1 bit of feedback. In Figure 8, $m = n = 1$ and $r = 0.2$. Thus, the theoretical diversity for the case of no feedback is 0.8 while for all other cases considered in the Figure is 1.6 and we find that we obtain close to the expected diversity order at SNR of about 20 dB. Note that the differences in the higher diversity order curves is small and they seem on top of each other in the plot. In Figure 9, $m = 1$, $n = 2$ and $r = 0.5$. In this case, the theoretical diversity for the case of no feedback is 1 while with feedback is 3 which can be noted from the slope.

VIII. EXTENSION TO MULTIUSER MIMO

In this section, we extend our main result for point-to-point systems, i.e., Theorem 7, to the case of multiple access channel under the model of common feedback to all transmitters.

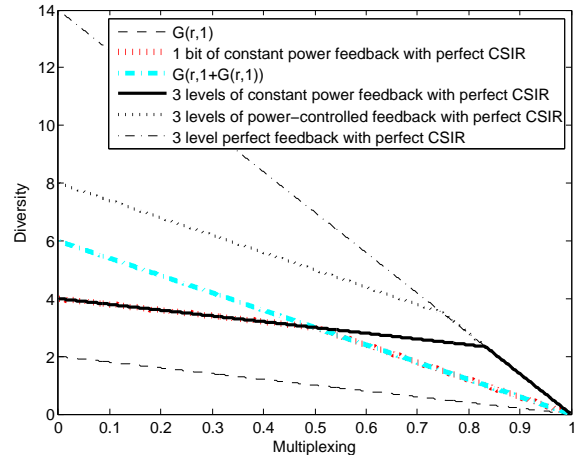


Fig. 7. Diversity multiplexing tradeoff in various scenarios.

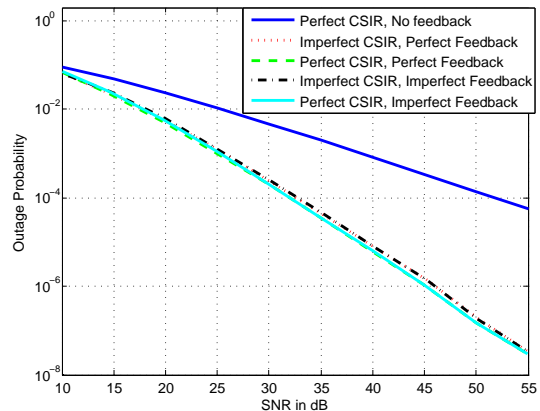


Fig. 8. Diversity multiplexing tradeoff in various scenarios. The various parameters chosen are: $m=1$, $n=1$, $r=0.2$. The channel estimates are obtained using a training time of 10 channel uses.

A. Multiuser Channel Model

Consider a multiple access channel with L transmitters T_i where each transmitter has m transmit antennas and the receiver R has n receive antennas. The channel is constant during a fading block of T_{coh} channel uses, but changes independently from one block to the next.

The received signal can be written in the matrix form as

$$\{T_i\} \rightarrow R : Y = \sum_{1 \leq i \leq L} H_i X_i + W. \quad (26)$$

Here, W is of size $n \times T_{\text{coh}}$ represents additive white Gaussian noise at the receiver with all entries i.i.d. $CN(0, 1)$. We consider a Rayleigh fading environment, i.e. elements of H_i are assumed to be i.i.d. $CN(0, 1)$. The transmitters are subject to an average power constraint such that the long-term power

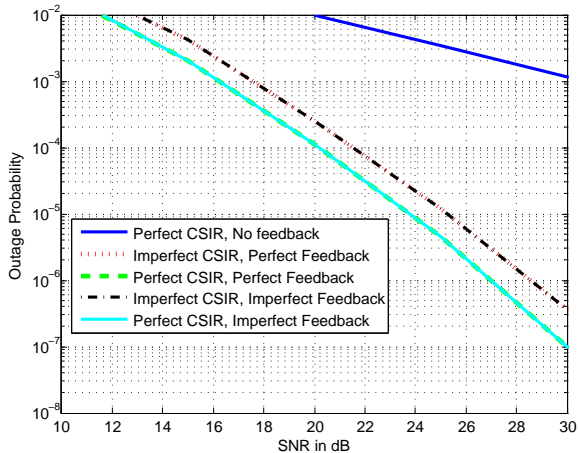


Fig. 9. Diversity multiplexing tradeoff in various scenarios. The various parameters chosen are: $m=1$, $n=2$, $r=0.5$. The channel estimates are obtained using a training time of 10 channel uses.

is upper bounded, i.e., $\frac{1}{T_{\text{coh}}}\text{trace}(\mathbb{E}[X_i X_i^\dagger]) \leq \text{SNR}$ for $1 \leq i \leq L$.

The feedback path to the transmitter in an orthogonal frequency band is given by

$$\mathbf{R} \rightarrow \mathbf{T}_i : Y_{f,i} = H_{f,i} X_{f,i} + W_f, \quad (27)$$

where $H_{f,i}$ is the MIMO fading channel for the feedback link to i^{th} user, normalized much like the forward link. The feedback transmissions are also assumed to be power-limited, that is the reverse link has a power budget of $\frac{1}{T_{\text{coh}}}\text{trace}(\mathbb{E}[X_{f,i} X_{f,i}^\dagger]) \leq \text{SNR}_f$. Without loss of generality, we will assume a *symmetry in resources*, such that $\text{SNR} = \text{SNR}_f$.

Finally, we assume that the receiver computes the common feedback index $J_R \in \{0, \dots, K-1\}$, which is broadcast over the downlink and received at transmitter i as $J_{T_i} \in \{0, \dots, K-1\}$.

B. Diversity-Multiplexing Tradeoff

The diversity multiplexing tradeoff for single user MIMO channels was described in Section II.C. Here, we extend that discussion to MIMO MAC channels. As before, we concentrate on single rate transmission. The dependence of rates on the SNRs is explicitly given by $R_s = r_s \log \text{SNR}_s$. We refer to $\mathbf{r} \triangleq (r_s)_{1 \leq s \leq L}$ as the multiplexing gains. Let $\mathbf{H} = \{H_1, \dots, H_L\}$. Further, let the channel estimates at the receiver be $\hat{\mathbf{H}} = \{\hat{H}_1, \dots, \hat{H}_L\}$.

In a multiple access channel, corresponding outage event is defined as the union over the events that the channel cannot support target data rate for some subset of the users [28], union over all the subsets. Hence, for a multiple access channel with L users, each equipped with m transmit antennas, and a receiver with n receive antennas, the outage event is $\mathcal{O} \triangleq \bigcup_S \mathcal{O}_S$. The union is taken over all subsets $S \subseteq \{1, 2, \dots, L\}$, and \mathcal{O}_S is the set of all the channels where the sum transmitted rate

by these $|S|$ users is less than the maximum supportable rate by the MIMO link from these $|S|$ users to the destination. The system is said to have diversity order of d if $\Pi(\mathcal{O}) \doteq \text{SNR}^{-d}$. The diversity multiplexing tradeoff for the multiple access channel can be described as follows: given the multiplexing gains \mathbf{r} for all the users, the diversity order that can be achieved describes the diversity-multiplexing tradeoff region.

The probability of outage with rate $\mathbf{R} = (R_1, R_2, \dots, R_L)$, transmit power $P(J_{T_i}) = P_i \forall J_{T_i}$ and perfect channel state information \mathbf{H} at the receiver is denoted by $\mathcal{S}(\mathbf{R}, \mathbf{P}) \triangleq \Pi\left(\bigcup_S \mathcal{O}_S(\mathbf{R}, \mathbf{P})\right)$. If we assume that the receiver knows the channel perfectly as \mathbf{H} , we denote the event $\bigcup_S \mathcal{O}_S(\mathbf{R}, \mathbf{P})$ by $U_{\mathbf{H}}(\mathbf{R}, \mathbf{P})$ where we assume that \mathbf{H} is the perfect channel knowledge at the receiver.

Let $D(\mathbf{r}, \mathbf{p})$ be defined as $\mathcal{S}(\mathbf{R}, \mathbf{P}) \doteq \text{SNR}^{-D(\mathbf{r}, \mathbf{p})}$ where $\mathbf{r} = (r_1, r_2, \dots, r_L)$ and $\mathbf{p} = (p_1, p_2, \dots, p_L)$. We further denote function $G(r, p)$ in Section II by $G_{m,n}(r, p)$ to explicitly depict that this is for m transmit and n receive antennas.

Lemma 5. [16] Let $p_s = p$ for all $1 \leq s \leq L$. Also, let $\sum_{i \in S} r_i \leq \min(|S|m, n)$ for all non-empty subsets S of $\{1, 2, \dots, L\}$. Then,

$$D(\mathbf{r}, \mathbf{p}) = \min_S G_{|S|m, n} \left(\sum_{i \in S} r_i, p \right). \quad (28)$$

C. $\widehat{\text{CSIR}}_q$: Estimated CSIR with Noisy Quantized Feedback

All the results related to quantized feedback in this paper can be extended to multiple access channel where there is a feedback level sent from the receiver and all the transmitters receive this signal and adjust the power accordingly. To demonstrate the extension, we consider a symmetric system where all transmitters have a statistically identical channel to the receiver with identical average SNR and employ the same power control thresholds. Furthermore, we will only consider a single-bit feedback, which implies that simultaneously all transmitters will be instructed to use the low power level or the high power level. However, since we assume independent errors in the feedback links, each transmitter may or may not transmit at the right power level. Under the above conditions, an achievable diversity-multiplexing tradeoff is given by

Theorem 8. For $K = 2$ and $\mathbf{r} = (r_1, \dots, r_L)$ with $\sum_{i \in S} r_i \leq \min(|S|m, n)$ for all non-empty subsets S of $\{1, 2, \dots, L\}$, an achievable diversity-multiplexing tradeoff for a multiple-access channel is given by $\widehat{d}_{\widehat{\text{RT}}_q}(\mathbf{r}, 2) = D(\mathbf{r}, \mathbf{1}(1 + D(\mathbf{r}, \mathbf{1})))$ where $\mathbf{1}x$ denotes (x, x, \dots, x) .

Proof: We will provide the main steps to prove the above result based on the following three-phase protocol. First define $p_0 = 0$ and $p_j = D(\mathbf{r}, \mathbf{1}(1 + p_{j-1})) \forall j \geq 1$.

Phase 1: Each transmitter trains the receiver using power SNR to get the channel estimate \hat{H}_i at the receiver. On the basis of this training, the receiver decides a feedback level J_R in the following way. We consider two power levels $P_i \doteq \text{SNR}^{1+p_i}$, $i \in [0, 1]$. We will state the exact constants for power control while proving that the average power constraint

will be satisfied. The feedback index is

$$J_R = \begin{cases} 1 & \text{if } U_{\hat{\mathbf{H}}}(\mathbf{R} + \mathbf{1}\epsilon \log(\text{SNR}), \mathbf{1}P_0) = 1 \\ 0 & \text{otherwise} \end{cases}.$$

Intuitively, we choose the higher of the two power levels if the lower power level is not sufficient to avoid outage (even for one of the users) based on the estimated channel.

Phase 2: A feedback level $J_R \in \{0, 1\}$ is sent from the receiver but each transmitter receives $J_{T_i} \in \{0, 1\}$ according to the power controlled feedback scheme in Section IV-B. Since the feedback links have i.i.d. errors, different transmitters may receive different feedback indices.

Phase 3: The transmitter k gets a feedback power level J_{T_k} , sends a training signal to the receiver at power level $P_{J_{T_k}}$ followed by data at power level $P_{J_{T_k}}$. The channel estimate based on this power-controlled training is denoted $\hat{H}_{2,k}$. Let $\tilde{H}_k = H_k - \hat{H}_{2,k}$. Further, $\hat{\mathbf{H}}_2 = \{\hat{H}_{2,1}, \dots, \hat{H}_{2,L}\}$. Also, denote \mathbf{H}_S as the $|S|m \times n$ matrix formed by concatenation of H_i in S . Similarly, define $\hat{\mathbf{H}}_{2S}$ and $\hat{\mathbf{H}}_S$.

The outage probability is bounded from above by sum of outage probabilities for each transmitter. The analysis of Phase 3 is similar to Appendix C-B1 since the estimation error in the third phase can be neglected for diversity multiplexing tradeoff purposes. Now define J as

$$J = \begin{cases} 1 & \text{if } U_{\hat{\mathbf{H}}_2}(\mathbf{R}, \mathbf{1}P_0) = 1 \text{ and } U_{\hat{\mathbf{H}}_2}(\mathbf{R}, \mathbf{1}P_1) = 0 \\ 0 & \text{otherwise} \end{cases}.$$

Since the third phase estimation error can be neglected, the outage probability is

$$\Pi(\mathcal{O}) \doteq \Pi(U_{\hat{\mathbf{H}}_2}(\mathbf{R}, \mathbf{P}(J_T)) = 1).$$

Hence, repeating the analysis of single user systems and using union bounds in Section VI, we get the same results as in single user systems, but with G replaced by D and single multiplexing gain replaced by multiplexing gain vector.

We show by example how to extend all the steps. $\Pi(J_R = 1, J = 0)$ can be written as

$$\begin{aligned} & \Pi(J_R = 0, J = 1) \\ &= \Pi \left(\log \det(I + \text{SNR}\hat{\mathbf{H}}_S\hat{\mathbf{H}}_S^\dagger) \geq \sum_{i \in S} R_i \forall S \text{ and} \right. \\ & \quad \left. \log \det(I + \text{SNR}\hat{\mathbf{H}}_{2S}\hat{\mathbf{H}}_{2S}^\dagger) < \sum_{i \in S} R_i \text{ for some } S \right. \\ & \quad \left. \text{and } \log \det(I + P_1\hat{\mathbf{H}}_{2S}\hat{\mathbf{H}}_{2S}^\dagger) \geq \sum_{i \in S} R_i \forall S \right) \quad (29) \end{aligned}$$

We will now bound the above statement as follows. First define the following events

$$A_S = \{\log \det(I + \text{SNR}\hat{\mathbf{H}}_S\hat{\mathbf{H}}_S^\dagger) \geq \sum_{i \in S} R_i\}, \quad (30)$$

$$C_S = \{\log \det(I + \text{SNR}\hat{\mathbf{H}}_{2S}\hat{\mathbf{H}}_{2S}^\dagger) < \sum_{i \in S} R_i\}, \quad (31)$$

$$D_S = \{\log \det(I + \text{SNR}^{1+p_1}\hat{\mathbf{H}}_{2S}\hat{\mathbf{H}}_{2S}^\dagger) \geq \sum_{i \in S} R_i\} \quad (32)$$

We note (29) is the probability of $\cap_S A_S \cap (\cup_S C_S) \cap$

$\cap_S D_S \subseteq \cup_S (A_S \cap C_S)$. Hence,

$$\begin{aligned} & \Pi(J_R = 0, J = 1) \\ & \leq \sum_S \Pi \left(\log \det(I + \text{SNR}\hat{\mathbf{H}}_S\hat{\mathbf{H}}_S^\dagger) \geq \sum_{i \in S} R_i \text{ and} \right. \\ & \quad \left. \log \det(I + \text{SNR}\hat{\mathbf{H}}_{2S}\hat{\mathbf{H}}_{2S}^\dagger) < \sum_{i \in S} R_i \right) \quad (33) \end{aligned}$$

Note that this term is similar to that in single user, and for each S is $\doteq 0$. Thus, the above probability $\doteq 0$.

Similarly, all other steps for single user MIMO channels can be extended to MIMO MAC system. Hence, the diversity order of $D(\mathbf{r}, \mathbf{1}(1 + D(\mathbf{r}, \mathbf{1})))$ can be achieved with imperfect feedback and imperfect CSIR for L transmitters. ■

In [28], diversity-multiplexing for multiple-access case was considered without feedback. The achievable diversity multiplexing without feedback is $D(\mathbf{r}, \mathbf{1})$. In [14], it was shown that with 1 bit of perfect feedback, the diversity multiplexing of $D(\mathbf{r}, \mathbf{1}(1 + D(\mathbf{r}, \mathbf{1})))$ can be achieved when the receiver knows perfect channel state information. In this paper, we show that the diversity multiplexing tradeoff of $D(\mathbf{r}, \mathbf{1}(1 + D(\mathbf{r}, \mathbf{1})))$ can be achieved even when the receiver is trained on a noisy channel and the feedback index is also sent on an orthogonal noisy channel with power-controlled training and feedback.

IX. CONCLUSIONS

In this paper, we find the diversity tradeoff for a non-symmetric FDD system in which the errors in MMSE channel estimate and the quantized feedback channel are accounted for a single user and a multiple access channel. We find that diversity multiplexing tradeoff of a system with 1 bit of feedback over a noisy channel and MMSE channel estimate at the receiver is the same as that of a system with 1 bit of perfect feedback and perfect channel estimate at the receiver. More importantly, we show that additional bits of feedback do not increase the diversity order of the system at constant rates.

The approach in this paper has also been used to improve the performance of a TDD system as is summarized in [24]. The two models, FDD and TDD, consider the two extreme cases of the correlations between the forward and the backward channel. As a next step, one can consider what happens if the forward and the feedback channel are correlated, but not exactly the same.

This paper suggests that one round of training provides a certain resolution to the channel gain which limits the diversity multiplexing tradeoff performance. The strategies can be extended to a multi-round communication between the sender and the receiver that allows better channel resolution at the nodes. This multi-round extension for both FDD and TDD models can be seen in [32].

Also, this paper assumes a Rayleigh fading channel model. The authors of [29] consider a general model for fading which includes Rayleigh, Rician, Nakagami and Weibull distributions to find the diversity multiplexing tradeoff for a system with no feedback and perfect channel estimate at the receiver. The

extension of the feedback cases to general fading models is still open.

Finally, the two way channel model can be extended to consider delays in the feedback channel. If there is a delay in the feedback process, the transmitter can decide to send some data as if there is no feedback till it receives feedback and then try to use the feedback to improve the diversity by sending power controlled data (possibly correlated with the data transmitted before feedback is received) in the remaining time.

X. ACKNOWLEDGEMENTS

The authors wish to thank Gajanana Krishna and Srikrishna Bhashyam for useful discussions related to this paper. We would also like to thank the anonymous reviewers for many suggestions that improved this paper.

APPENDIX A PROOF OF THEOREM 4

We first note some properties of $I_\nu(x)$, modified Bessel function of first kind, that will be used in the proof. The series expansion of $I_\nu(x)$ is given as [31, Equation 9.6.10][30],

$$I_\nu(x) = \sum_{i=0}^{\infty} \frac{1}{i!(i+\nu)!} \left(\frac{x}{2}\right)^{2i+\nu}. \quad (34)$$

When $|x|$ is large and $|\arg(x)| < \frac{\pi}{2}$, asymptotic expansion of $I_\nu(x)$ is given by [31, Equation 9.7.1]

$$I_\nu(x) \doteq \frac{e^x}{\sqrt{2\pi x}} \left\{ 1 - \frac{\mu-1}{x} + \frac{(\mu-1)(\mu-9)}{2!(8x)^2} - \frac{(\mu-1)(\mu-9)(\mu-25)}{3!(8x)^3} + \dots \right\}, \quad (35)$$

where $\mu = 4\nu^2$.

Now, for the proof of Theorem 4, we will use Lemma 4. We will further suppose that $m_N = m$ without loss of generality. Let $A = \{i : \alpha_i + \hat{\alpha}_i \leq 2\}$. We will evaluate $p(\alpha, \hat{\alpha})$ in the following five disjoint cases which comprise the whole space of possibilities.

- 1) $\min(\alpha_m, \hat{\alpha}_m) \geq 1$ (or $(\alpha, \hat{\alpha}) \in E_m$).
- 2) $\min(\alpha_m, \hat{\alpha}_m) < 1$, $\alpha_m + \hat{\alpha}_m \geq 2$.
- 3) $\min(\alpha_m, \hat{\alpha}_m) < 1$, $\alpha_m + \hat{\alpha}_m < 2$, $\alpha_i \neq \hat{\alpha}_i$ for some $i \in A$.
- 4) $\min(\alpha_m, \hat{\alpha}_m) < 1$, $\alpha_m + \hat{\alpha}_m < 2$, $\alpha_i = \hat{\alpha}_i$ for all $i \in A$, $(\alpha, \hat{\alpha}) \notin \bigcup_{k=0}^{m-1} (E_k)$.
- 5) $(\alpha, \hat{\alpha}) \in \bigcup_{k=0}^{m-1} (E_k)$.

Now, we consider all of the cases one by one as follows.

- 1) $\min(\alpha_m, \hat{\alpha}_m) \geq 1$: Using Equation (9),

$$p(\lambda, \hat{\lambda}) = \frac{\exp\left(-\frac{\sum_{k=1}^m \lambda_k + \hat{\lambda}_k}{1-|\rho|^2}\right) \Delta(\lambda) \Delta(\hat{\lambda})}{m!m! \prod_{j=0}^{m-1} j!(j+\nu)! |\rho|^{mn-m} (1-|\rho|^2)^m} \prod_{k=1}^m \left(\sqrt{\lambda_k \hat{\lambda}_k} \right)^\nu \det \left| I_\nu \left(\frac{2|\rho| \sqrt{\lambda_k \hat{\lambda}_k}}{1-|\rho|^2} \right) \right|$$

Since $\rho = \frac{1}{\sqrt{1+\frac{m}{\text{SNR}}}}$, we substitute $\rho \doteq 1$ and $1-\rho^2 \doteq 1/\text{SNR}$ to obtain

$$p(\alpha, \hat{\alpha}) \doteq \exp\left(-\frac{\sum_{i=1}^m \text{SNR}^{-\alpha_i} + \text{SNR}^{-\hat{\alpha}_i}}{1/\text{SNR}}\right) \Delta(\text{SNR}^{-\alpha}) \Delta(\text{SNR}^{-\hat{\alpha}}) \text{SNR}^{-\sum_{i=1}^m (\alpha_i + \hat{\alpha}_i)} \left(\sqrt{\text{SNR}^{-\sum_{i=1}^m \alpha_i} \text{SNR}^{-\sum_{i=1}^m \hat{\alpha}_i}}\right)^{(n-m)} \frac{\det \left| I_{n-m} \left(\frac{2\sqrt{\text{SNR}^{-\alpha_k} \text{SNR}^{-\hat{\alpha}_k}}}{1/\text{SNR}} \right) \right|}{m!m! \prod_{j=0}^{m-1} j!(j+n-m)! \frac{1}{\text{SNR}^m}}. \quad (36)$$

As $\Delta(\text{SNR}^{-\alpha}) \doteq \text{SNR}^{-\sum_{i=1}^m (i-1)\alpha_i}$, we get

$$p(\alpha, \hat{\alpha}) \doteq \text{SNR}^m \text{SNR}^{-\sum_{i=1}^m (i+\frac{n-m}{2})(\alpha_i + \hat{\alpha}_i)} \det \left| I_{n-m} \left(\text{SNR}^{1-\frac{\alpha_k + \hat{\alpha}_k}{2}} \right) \right|. \quad (37)$$

We will now find $\det \left| I_\nu \left(\text{SNR}^{1-\frac{\alpha_k + \hat{\alpha}_k}{2}} \right) \right|$. Using (34), we get

$$\begin{aligned} & I_\nu \left(\text{SNR}^{1-\frac{\alpha_k + \hat{\alpha}_k}{2}} \right) \\ &= \sum_{i=0}^{\infty} \frac{1}{i!(i+\nu)!} \left(\frac{\text{SNR}^{1-\frac{\alpha_k + \hat{\alpha}_k}{2}}}{2} \right)^{2i+\nu} \\ &= \left(\frac{\text{SNR}^{1-\frac{\alpha_k + \hat{\alpha}_k}{2}}}{2} \right)^\nu \sum_{i=0}^{\infty} \frac{1}{i!(i+\nu)!} \left(\frac{\text{SNR}^{1-\frac{\alpha_k + \hat{\alpha}_k}{2}}}{2} \right)^{2i} \end{aligned} \quad (38)$$

Let $\text{per}(k_1, k_2, \dots, k_m)$ for (k_1, k_2, \dots, k_m) a permutation of $(1, \dots, m)$ be defined as follows

$$\begin{aligned} & \text{per}(k_1, k_2, \dots, k_m) \\ & \triangleq \begin{cases} 0 & \text{if } (k_1, k_2, \dots, k_m) \text{ is an even} \\ & \text{permutation of } (1, \dots, m) \\ 1 & \text{if } (k_1, k_2, \dots, k_m) \text{ is an odd} \\ & \text{permutation of } (1, \dots, m) \end{cases}. \end{aligned} \quad (39)$$

Thus,

$$\begin{aligned} & \det \left| I_\nu \left(\text{SNR}^{1-\frac{\alpha_k + \hat{\alpha}_k}{2}} \right) \right| \\ &= \sum_{\mathbf{k}} (-1)^{\text{per}(k_1, k_2, \dots, k_m)} \prod_{l=1}^m \left(\frac{\text{SNR}^{1-\frac{\alpha_{k_l} + \hat{\alpha}_l}{2}}}{2} \right)^\nu \\ & \quad \sum_{i=0}^{\infty} \frac{1}{i!(i+\nu)!} \left(\frac{\text{SNR}^{1-\frac{\alpha_{k_l} + \hat{\alpha}_l}{2}}}{2} \right)^{2i} \\ & \doteq \text{SNR}^{\nu \sum_{i=1}^m (1-\frac{\alpha_i + \hat{\alpha}_i}{2})} \sum_{\mathbf{k}} (-1)^{\text{per}(k_1, k_2, \dots, k_m)} \\ & \quad \prod_{l=1}^m \sum_{i=0}^{\infty} \frac{1}{i!(i+\nu)!} \left(\frac{\text{SNR}^{2-\alpha_{k_l} + \hat{\alpha}_l}}{4} \right)^i \end{aligned} \quad (40)$$

The above equation is same as Equation (56) in [33]

with $K = \frac{\sqrt{2}-1}{2}$, $\phi_i = \text{SNR}^{1-\alpha_i}$ and $\lambda_i = \text{SNR}^{1-\hat{\alpha}_i}$, and hence

$$\begin{aligned} & \det \left| I_\nu \left(\text{SNR}^{1-\frac{\alpha_k+\hat{\alpha}_l}{2}} \right) \right| \\ & \doteq \text{SNR}^\nu \sum_{i=1}^m (1-\frac{\alpha_i+\hat{\alpha}_i}{2}) \Delta(\text{SNR}^{1-\alpha_i}) \Delta(\text{SNR}^{1-\hat{\alpha}_i}) \\ & \doteq \text{SNR}^\nu \sum_{i=1}^m (1-\frac{\alpha_i+\hat{\alpha}_i}{2}) \text{SNR}^{2 \sum_{i=1}^m (i-1)(1-\frac{\alpha_i+\hat{\alpha}_i}{2})} \\ & \doteq \text{SNR}^{\sum_{i=1}^m (2i-2+\nu)} \left(1-\frac{\alpha_i+\hat{\alpha}_i}{2}\right). \end{aligned} \quad (41)$$

Substituting in Equation (37), we get

$$\begin{aligned} p(\alpha, \hat{\alpha}) & \doteq \text{SNR}^m \text{SNR}^{-\sum_{i=1}^m (i+\frac{n-m}{2})(\alpha_i+\hat{\alpha}_i)} \\ & \quad \text{SNR}^{\sum_{i=1}^m (2i-2+\nu)} \left(1-\frac{\alpha_i+\hat{\alpha}_i}{2}\right) \\ & \doteq \text{SNR}^{mn} \text{SNR}^{-\sum_{i=1}^m (i+\frac{n-m}{2})(\alpha_i+\hat{\alpha}_i)} \\ & \quad \text{SNR}^{-\sum_{i=1}^m (i-1+\frac{n-m}{2})(\alpha_i+\hat{\alpha}_i)} \\ & \doteq \text{SNR}^{mn} \text{SNR}^{-\sum_{i=1}^m (2i+n-m-1)(\alpha_i+\hat{\alpha}_i)} \\ & = e_m. \end{aligned} \quad (42)$$

2) $\min(\alpha_m, \hat{\alpha}_m) < 1$, $\alpha_m + \hat{\alpha}_m \geq 2$: Using Equation (36), we see that all terms except $\exp\left(-\frac{\sum_{i=1}^m \text{SNR}^{-\alpha_i} + \text{SNR}^{-\hat{\alpha}_i}}{1/\text{SNR}}\right)$ remain the same and hence are polynomial in SNR while $\exp\left(-\frac{\sum_{i=1}^m \text{SNR}^{-\alpha_i} + \text{SNR}^{-\hat{\alpha}_i}}{1/\text{SNR}}\right)$ decreases exponentially in SNR and hence $p(\alpha, \hat{\alpha}) \doteq 0$.

3) $\min(\alpha_m, \hat{\alpha}_m) < 1$, $\alpha_m + \hat{\alpha}_m < 2$, $\alpha_i \neq \hat{\alpha}_i$ for some $i \in A$: We will prove $p(\alpha, \hat{\alpha}) \doteq 0$ in this case. For this we consider $\exp\left(-\frac{\sum_{k=1}^m \lambda_k + \hat{\lambda}_k}{1-|\rho|^2}\right) \det \left| I_\nu \left(\frac{2|\rho|\sqrt{\lambda_k \hat{\lambda}_l}}{1-|\rho|^2} \right) \right|$ and prove this part to decrease exponentially and we would be done since rest of the terms are polynomial in SNR. Using (35),

$$\begin{aligned} & \det \left| I_\nu \left(\frac{2|\rho|\sqrt{\lambda_{k_l} \hat{\lambda}_l}}{1-|\rho|^2} \right) \right| \\ & \doteq \left(\sum_{\mathbf{k}} (-1)^{\text{per}(k_1, k_2, \dots, k_m)} \prod_{l=1, \alpha_{k_l} + \hat{\alpha}_l \leq 2} \right. \\ & \quad \left. \exp\left(\frac{2|\rho|\sqrt{\lambda_{k_l} \hat{\lambda}_l}}{1-\rho^2}\right) \frac{1 - \frac{\mu-1}{\sqrt{2\sqrt{\lambda_{k_l} \hat{\lambda}_l}}}}{\sqrt{2\sqrt{\lambda_{k_l} \hat{\lambda}_l}}} \left(1 - \frac{\mu-1}{\frac{16\sqrt{\lambda_{k_l} \hat{\lambda}_l}}{1-\rho^2}} + \dots\right) \right. \\ & \quad \left. \prod_{l=1, \alpha_{k_l} + \hat{\alpha}_l > 2} \text{Poly}(\text{SNR}) \right), \end{aligned} \quad (43)$$

where $\text{Poly}(\text{SNR})$ represents the term that are polynomial in SNR. First observe that

$$\exp\left(-\frac{\lambda_{k_l} + \hat{\lambda}_l}{1-\rho^2}\right) \exp\left(\frac{2|\rho|\sqrt{\lambda_{k_l} \hat{\lambda}_l}}{1-\rho^2}\right) \leq 1, \quad (44)$$

and hence the product of such terms cannot increase exponentially with SNR.

$$\begin{aligned} & \exp\left(-\frac{\sum_{k=1}^m \lambda_k + \hat{\lambda}_k}{1-|\rho|^2}\right) \det \left| I_\nu \left(\frac{2|\rho|\sqrt{\lambda_k \hat{\lambda}_l}}{1-|\rho|^2} \right) \right| \\ & \doteq \sum_{\mathbf{k}} (-1)^{\text{per}(k_1, k_2, \dots, k_m)} \\ & \quad \prod_{l=1}^m \exp(-\text{SNR}^{1-\min(\alpha_{k_l}, \hat{\alpha}_l)}) \\ & \quad \prod_{l=1, \alpha_{k_l} + \hat{\alpha}_l \leq 2} \exp(\text{SNR}^{1-\frac{\alpha_{k_l} + \hat{\alpha}_l}{2}}) \text{Poly}(\text{SNR}) \\ & \doteq \sum_{\mathbf{k}} (-1)^{\text{per}(k_1, k_2, \dots, k_m)} \\ & \quad \prod_{l=1, \alpha_{k_l} + \hat{\alpha}_l > 2} \exp(-\text{SNR}^{1-\min(\alpha_{k_l}, \hat{\alpha}_l)}) \\ & \quad \prod_{l=1, \alpha_{k_l} + \hat{\alpha}_l \leq 2} \left(\exp(-\text{SNR}^{1-\min(\alpha_{k_l}, \hat{\alpha}_l)}) \right. \\ & \quad \left. \exp(\text{SNR}^{1-\frac{\alpha_{k_l} + \hat{\alpha}_l}{2}}) \right) \text{Poly}(\text{SNR}) \\ & \leq \sum_{\mathbf{k}} (-1)^{\text{per}(k_1, k_2, \dots, k_m)} \\ & \quad \prod_{l=1, \alpha_{k_l} + \hat{\alpha}_l > 2} \mathbf{1}_{\min(\alpha_{k_l}, \hat{\alpha}_l) \geq 1} \\ & \quad \prod_{l=1, \alpha_{k_l} + \hat{\alpha}_l \leq 2} \mathbf{1}_{\alpha_{k_l} = \hat{\alpha}_l} \text{Poly}(\text{SNR}) \end{aligned} \quad (45)$$

Now, we will show that each term under the sum decays exponentially with SNR. For this, the product of the above indicators should be 0. Let us now consider all the scenarios when the above product of indicators is 1. If some $\hat{\alpha}_l < 1$, and pairs with α_{k_l} to give $\hat{\alpha}_l + \alpha_{k_l} \leq 2$, then the two must be equal and if it pairs to give $\hat{\alpha}_l + \alpha_{k_l} > 2$, the product of indicators will always be 0. If $\hat{\alpha}_l \geq 1$ and pairs with α_{k_l} to give $\hat{\alpha}_l + \alpha_{k_l} \leq 2$, it can only happen when the two are equal and if it pairs to give $\hat{\alpha}_l + \alpha_{k_l} > 2$, then $\alpha_{k_l} \geq 1$. Hence the only pairing that will work is that all elements of $\alpha_i < 1$ are matched to $\hat{\alpha}_j < 1$ and also $\alpha_i \geq 1$ is not mapped to $\hat{\alpha}_i < 1$. This can happen only when $\alpha_i = \hat{\alpha}_i$ whenever $\alpha_i < 1$ or $\hat{\alpha}_i < 1$ and all the rest are ≥ 1 . If $\alpha_i \neq \hat{\alpha}_i$ for some $i \in A$, the above condition do not hold. This proves Case 3.

4) $\min(\alpha_m, \hat{\alpha}_m) < 1$, $\alpha_m + \hat{\alpha}_m < 2$, $\alpha_i = \hat{\alpha}_i$ for all $i \in A$, $(\alpha, \hat{\alpha}) \notin \bigcup_{k=0}^{m-1} (E_k)$: We see that all the analysis of Case 3 holds for this case and hence if $\alpha_i = \hat{\alpha}_i$ for all $i \in A$ and $(\alpha, \hat{\alpha}) \notin \bigcup_{k=0}^{m-1} (E_k)$, an element ≥ 1 is mapped to an element < 1 which makes the product of indicators zero and hence the probability decreases exponentially with SNR.

5) $\alpha, \hat{\alpha} \in \bigcup_{k=0}^{m-1} (E_k)$. These are the cases we sum over in the statement of the Lemma. In each of these cases, $p(\alpha, \hat{\alpha})$ exist. When we integrate over $\hat{\alpha}$, we find that integral of $\sum_k e_k \mathbf{1}_{E_k}$ w.r.t. $\hat{\alpha}$ is $\prod_{i=1}^m \text{SNR}^{-(2i-1+n-m)\alpha_i} \mathbf{1}_{\alpha_1, \dots, \alpha_k \geq 1} \mathbf{1}_{0 \leq \alpha_{k+1}, \dots, \alpha_{m-N} < 1}$. We next note that in order to find the polynomial expressions associated, the two exponentials multiplication cannot decrease with SNR for which we would need $k_l = l$ in (43) (since all the λ 's and $\hat{\lambda}$'s are ordered) and hence from all the expressions, we find that $p(\alpha, \hat{\alpha}) = p_1(\alpha) p_2(\hat{\alpha})$ for any E_k . Using the

separability, we find that

$$e_k = \frac{\text{SNR}^{k(n-m+k)} \prod_{i=1}^k \text{SNR}^{-(2i-1+|n-m|)\hat{\alpha}_i}}{\prod_{i=1}^{m_N} \text{SNR}^{-(2i-1+|n-m|)\alpha_i}}. \quad (46)$$

APPENDIX B PROOF OF THEOREM 3

In this appendix, we prove that the diversity in the statement of Theorem 3 can be achieved. Let $S(R, P)$ be the probability of outage when the transmitter uses power level P and rate R is required. If $P \doteq \text{SNR}^p$, then $S(R, P) \doteq \text{SNR}^{-G(r,p)}$ and

$$\Pi(\mathcal{O}) \doteq S(R, P_{K-1}) + \sum_{i=1}^{K-1} \Pi(J_T < i, J_R = i).$$

We assign the feedback index J_R at the receiver as

$$J_R = \begin{cases} \arg \min_{i \in \mathbb{I}} i, & \mathbb{I} = \{k : \log \det(I + HH^\dagger P_k) \geq R, \\ & k \in \{0, \dots, K-1\}\} \\ K-1, & \text{if the set } \mathbb{I} \text{ is empty} \end{cases}.$$

We will now find $\Pi(J_T = i)$ for $i \geq 1$ as

$$\begin{aligned} & \Pi(J_T = i) \\ &= \sum_{j=0}^{K-1} \Pi(J_T = i, J_R = j) \\ &= \sum_{j=0}^{K-1} \Pi(J_T = i | J_R = j) \Pi(J_R = j) \\ &\doteq \Pi(J_R = i) + \sum_{j=i+1}^{K-1} \Pi(J_T = i | J_R = j) \Pi(J_R = j) \\ &\doteq \Pi(J_R = i). \end{aligned} \quad (47)$$

The above steps follow from the fact that $\Pi(J_T = i | J_R = j) \doteq 0$ if $j < i$. Further,

$$\begin{aligned} & \sum_{j=i+1}^{K-1} \Pi(J_T = i | J_R = j) \Pi(J_R = j) \\ &\leq \sum_{j=i+1}^{K-1} \Pi(J_R = j) \leq \Pi(J_R = i). \end{aligned} \quad (48)$$

Let the power levels be chosen as

$$P_i = \begin{cases} \frac{\text{SNR}}{K} & \text{when } i = 0 \\ \frac{\text{SNR}}{KS(R, P_{i-1})} & \text{when } i > 0 \end{cases}.$$

Further, $Q_i \leq P_i$. We first note that the power constraints are satisfied. Note that as before, $P_i = \text{SNR}^{1+p_i}$ and $Q_i = \text{SNR}^{q_i}$. Also, $P_i \doteq \text{SNR}^{1+d_{\text{RT}_q}(r,i)} \forall i$ and $S(R, P_i) \doteq \text{SNR}^{-d_{\text{RT}_q}(r,i+1)}$. Thus, $\Pi(J_R = i) = \Pi(J_T = i) \doteq \text{SNR}^{-d_{\text{RT}_q}(r,i)}$.

We will now show that using P_i as given above, we achieve the desired diversity multiplexing tradeoff using the following computation.

$$\begin{aligned} & \Pi(\mathcal{O}) \\ &\leq S(R, P_{K-1}) + \sum_{i=1}^{K-1} \Pi(J_T < i, J_R = i) \\ &\doteq \text{SNR}^{-d_{\text{RT}_q}(r,K)} + \sum_{i=1}^{K-1} \Pi(J_T < i | J_R = i) \Pi(J_R = i) \\ &\doteq \text{SNR}^{-d_{\text{RT}_q}(r,K)} + \sum_{i=1}^{K-1} \Pi(J_T < i | J_R = i) \text{SNR}^{-d_{\text{RT}_q}(r,i)} \\ &\doteq \text{SNR}^{-d_{\text{RT}_q}(r,K)} \\ &\quad + \sum_{i=1}^{K-1} \text{SNR}^{-mn(\max(q_i,0) - \max(q_{i-1},0))} \text{SNR}^{-d_{\text{RT}_q}(r,i)} \\ &\doteq \text{SNR}^{-\bar{d}_{\text{RT}_q}(r,K)}. \end{aligned} \quad (49)$$

APPENDIX C PROOF OF THEOREM 6

We will first show that the diversity cannot be greater than $mn(mn+1)$, and then show that diversity of $G(r, 1+G(r, 1))$ can be achieved with 1 bit of feedback.

A. Converse

In this subsection, we will prove that we cannot get more diversity with > 2 levels of feedback with imperfect CSIR as compared to 1 bit of feedback with perfect CSIR at zero multiplexing.

For this, we assume that the third phase is perfect and thus $\tilde{H} = 0$, and $\hat{H}_2 = H$. Thus,

$$\Pi(\mathcal{O}) \doteq \Pi(\log \det(I + P(J_R)HH^\dagger) < R).$$

Define J as

$$J = \begin{cases} \arg \min_{i \in \mathbb{I}} i, & \mathbb{I} = \{k : \log \det(I + HH^\dagger P_k) \geq R, \\ & k \in \{0, \dots, K-1\}\} \\ 0, & \text{if the set } \mathbb{I} \text{ is empty} \end{cases}.$$

The probability of outage is then

$$\begin{aligned} \Pi(\mathcal{O}) &\doteq \Pi(\log \det(I + P(J_R)HH^\dagger) < R) \\ &\leq \Pi(\log \det(I + P(J_R)HH^\dagger) < R, J_R = 1) \\ &\doteq \Pi(\log \det(I + P_1 HH^\dagger) < R, \\ &\quad \log \det(I + P_0 \hat{H} \hat{H}^\dagger) < R + \epsilon \log(\text{SNR}), \\ &\quad \log \det(I + P_1 \hat{H} \hat{H}^\dagger) \geq \\ &\quad R + \epsilon \log(\text{SNR})). \end{aligned} \quad (50)$$

Let λ_j and $\hat{\lambda}_j$ be the eigenvalues of HH^\dagger and $\hat{H}\hat{H}^\dagger$ respectively. Further let α_j and $\hat{\alpha}_j$ be the negative SNR exponents of the corresponding eigenvalues. Then, Equation

(50) reduces to

$$\begin{aligned}
\Pi(\mathcal{O}) &\stackrel{\dot{\geq}}{\geq} \Pi(\log \det(I + P_1 H H^\dagger) < R, \\
&\quad \log \det(I + P_0 \hat{H} \hat{H}^\dagger) < R + \epsilon \log(\text{SNR}), \\
&\quad \log \det(I + P_1 \hat{H} \hat{H}^\dagger) \geq R + \epsilon \log(\text{SNR})) \\
&\stackrel{\dot{\geq}}{\geq} \Pi(\prod_{j=1}^m (1 + \lambda_j \text{SNR}^{1+p_1}) < \text{SNR}^r \text{ and} \\
&\quad \prod_{j=1}^m (1 + \hat{\lambda}_j \text{SNR}) < \text{SNR}^{r+\epsilon} \text{ and} \\
&\quad \prod_{j=1}^m (1 + \hat{\lambda}_j \text{SNR}^{1+p_1}) > \text{SNR}^{r+\epsilon}) \\
&\doteq \Pi(\sum_{j=1}^m (1 + p_1 - \alpha_j)^+ < r \text{ and} \\
&\quad \sum_{j=1}^m (1 - \hat{\alpha}_j)^+ < r + \epsilon \text{ and} \\
&\quad \sum_{j=1}^m (1 + p_1 - \hat{\alpha}_j)^+ > r + \epsilon). \tag{51}
\end{aligned}$$

For $r \rightarrow 0$, the above reduces to

$$\begin{aligned}
\Pi(\mathcal{O}) &\stackrel{\dot{\geq}}{\geq} \Pi(\sum_{j=1}^m (1 + p_1 - \alpha_j)^+ \leq 0 \text{ and} \\
&\quad \sum_{j=1}^m (1 - \hat{\alpha}_j)^+ < \epsilon \text{ and} \\
&\quad \sum_{j=1}^m (1 + p_1 - \hat{\alpha}_j)^+ > \epsilon). \tag{52}
\end{aligned}$$

Thus, the choice of $\alpha_j = 1 + p_1$ and all $\hat{\alpha}_j = 1$ for $\epsilon \rightarrow 0$ can be used for the outer bound on outage probability. This gives $\Pi(\mathcal{O}) \stackrel{\dot{\geq}}{\geq} \text{SNR}^{-mn(p_1+1)}$. As $p_1 \leq mn$, the diversity $\leq mn(mn + 1)$.

B. Achievability

In this subsection, we will show that the diversity gain of $G(r, 1 + G(r, 1))$ can be achieved with 1 bit of feedback with imperfect CSIR. We will first prove that the interference error due to \tilde{H} in third phase do not make a difference and thus can be removed.

1) *Analysis of Phase 3:* Let the actual channel be H while \tilde{H} be the error in the estimate of H in the third phase of training which is power controlled. As a result, we obtain

$$\begin{aligned}
\Pi(\mathcal{O}) &= \sum_{j=0}^{K-1} \Pi(\mathcal{O}, J_{\text{T}} = j) \\
&\doteq \sum_{j=0}^{K-1} \Pi(\log \det(1 + \frac{P_j \hat{H}_2 \hat{H}_2^\dagger}{1 + P_j \text{trace}(\tilde{H} \tilde{H}^\dagger)}) < R, \\
&\quad J_{\text{T}} = j).
\end{aligned}$$

Consider any term in the sum above, the receiver is trained with power P_j , and hence let eigen-values of $P_j \tilde{H} \tilde{H}^\dagger$ be $\tilde{\lambda}_i \doteq$

$\text{SNR}^{-\tilde{\alpha}_i}$. Let $\tilde{\alpha} = (\tilde{\alpha}_1, \dots, \tilde{\alpha}_m)$ Then,

$$\begin{aligned}
\Pi(\mathcal{O}) &\doteq \sum_{j=0}^{K-1} \Pi(\log \det(1 + \\
&\quad \frac{P_j \hat{H}_2 \hat{H}_2^\dagger}{1 + P_j \text{trace}(\tilde{H} \tilde{H}^\dagger)}) < R, J_{\text{T}} = j) \\
&\doteq \sum_{j=0}^{K-1} \Pi(\log \det(1 + P_j \hat{H}_2 \hat{H}_2^\dagger \text{SNR}^{-(\min \tilde{\alpha})^+}) < R, \\
&\quad J_{\text{T}} = j).
\end{aligned}$$

Now, since \hat{H}_2 and \tilde{H} are uncorrelated, probability that $(\min \tilde{\alpha})^+ > 0$ decreases higher than polynomial in SNR and hence,

$$\begin{aligned}
\Pi(\mathcal{O}) &\doteq \sum_{j=0}^{K-1} \Pi(\log \det(1 + P_j \hat{H}_2 \hat{H}_2^\dagger \text{SNR}^{-(\min \tilde{\alpha})^+}) \\
&\quad < R, J_{\text{T}} = j) \\
&\doteq \sum_{j=0}^{K-1} \Pi(\log \det(1 + P_j \hat{H}_2 \hat{H}_2^\dagger) < R, J_{\text{T}} = j) \\
&\doteq \Pi(\log \det(1 + P(J_{\text{T}}) \hat{H}_2 \hat{H}_2^\dagger) < R). \tag{53}
\end{aligned}$$

Thus,

$$\Pi(\mathcal{O}) \doteq \Pi(\log \det(I + P(J_{\text{T}}) \hat{H}_2 \hat{H}_2^\dagger) < R)$$

Note that the correlations between \hat{H}_2 and \hat{H} are same as between H and \hat{H} and thus, there is no difference in using H in place of \hat{H}_2 for the purpose of calculating the diversity multiplexing tradeoff. We will now only focus on $K = 2$.

2) *Analysis of Phase 1:* In this section, we see how the feedback error in first phase decays with SNR.

$$J_{\text{R}} = \begin{cases} 1 & \text{if } \log \det(I + \hat{H} \hat{H}^\dagger P_0) < R + \epsilon \log(\text{SNR}) \\ 0 & \text{if } \log \det(I + \hat{H} \hat{H}^\dagger P_0) \geq R + \epsilon \log(\text{SNR}) \end{cases}$$

Let us define J as

$$J = \begin{cases} 1 & \text{if } \log \det(I + \hat{H}_2 \hat{H}_2^\dagger P_0) < R \text{ and} \\ & \log \det(I + \hat{H}_2 \hat{H}_2^\dagger P_1) \geq R \\ 0 & \text{otherwise} \end{cases}.$$

The probability of outage is then

$$\begin{aligned}
\Pi(\mathcal{O}) &\doteq \Pi(\log \det(I + P(J_{\text{T}}) \hat{H}_2 \hat{H}_2^\dagger) < R) \\
&\doteq \Pi(\log \det(I + P(J_{\text{T}}) \hat{H}_2 \hat{H}_2^\dagger) < R, J = 0, J_{\text{R}} = 0) \\
&\quad + \Pi(\log \det(I + P(J_{\text{T}}) \hat{H}_2 \hat{H}_2^\dagger) < R, J_{\text{R}} = 1) \\
&\quad + \Pi(\log \det(I + P(J_{\text{T}}) \hat{H}_2 \hat{H}_2^\dagger) < R, J = 1, J_{\text{R}} = 0) \\
&\stackrel{\dot{\leq}}{\leq} \Pi(\log \det(I + P_1 \hat{H}_2 \hat{H}_2^\dagger) < R, J_{\text{R}} = 0) \\
&\quad + \Pi(\log \det(I + P_1 \hat{H}_2 \hat{H}_2^\dagger) < R, J_{\text{R}} = 1) \\
&\quad + \Pi(J_{\text{R}} = 0, J = 1) \\
&\doteq \Pi(\log \det(I + P_1 \hat{H}_2 \hat{H}_2^\dagger) < R) \\
&\quad + \Pi(J_{\text{R}} = 0, J = 1). \tag{54}
\end{aligned}$$

Denote eigenvalues of $\widehat{H}\widehat{H}^\dagger$ by $\widehat{\lambda}_i$ and the negative SNR exponents of $\widehat{\lambda}_i$ as $\widehat{\alpha}_i$. Also denote eigenvalues of $\widehat{H}_2\widehat{H}_2^\dagger$ by λ_i and the negative SNR exponents of λ_i as α_i .

Then, $\Pi(J_R = 0, J = 1)$ can be bounded as

$$\begin{aligned}
& \Pi(J_R = 0, J = 1) \\
&= \Pi \left(\log \det(I + \widehat{H}_2\widehat{H}_2^\dagger \text{SNR}) < R \text{ and} \right. \\
&\quad \log \det(I + \widehat{H}_2\widehat{H}_2^\dagger \text{SNR}^{1+p_1}) \geq R \text{ and} \\
&\quad \left. \log \det(I + \widehat{H}\widehat{H}^\dagger \text{SNR}) \geq R + \epsilon \log(\text{SNR}) \right) \\
&\doteq \Pi \left(\prod_{i=1}^m (1 + \lambda_i \text{SNR}) < \text{SNR}^r \text{ and} \right. \\
&\quad \prod_{i=1}^m (1 + \lambda_i \text{SNR}^{1+p_1}) \geq \text{SNR}^r \text{ and} \\
&\quad \left. \prod_{i=1}^m (1 + \widehat{\lambda}_i \text{SNR}) \geq \text{SNR}^{r+\epsilon} \right) \\
&\leq \Pi \left(\sum_{i=1}^m (1 - \alpha_i)^+ \leq r \text{ and} \right. \\
&\quad \sum_{i=1}^m (1 + p_1 - \alpha_i)^+ \geq r \text{ and} \\
&\quad \left. \sum_{i=1}^m (1 - \widehat{\alpha}_i)^+ \geq r + \epsilon \right) \\
&\doteq 0. \tag{55}
\end{aligned}$$

The last step follows since expanding in terms of E_k gives $\sum_{i=1}^m (1 - \alpha_i)^+ = \sum_{i=1}^m (1 - \widehat{\alpha}_i)^+$ for all cases of non exponentially decreasing probability.

We now show that the diversity order of $G(r, 1 + G(r, 1))$ can be achieved. From Equation (54), the probability of outage for $K = 2$ is

$$\begin{aligned}
\Pi(\mathcal{O}) &\leq \Pi(\log \det(I + P_1\widehat{H}_2\widehat{H}_2^\dagger) < R) \\
&\quad + \Pi(J_R = 0, J = 1) \\
&\leq \text{SNR}^{-G(r, 1+p_1)}. \tag{56}
\end{aligned}$$

Thus, diversity order of $G(r, 1 + G(r, 1))$ can be achieved which is the optimal diversity order for perfect training in [13].

Lastly, we show that the power constraint can be satisfied with the above choice of powers.

Lemma 6. *Power constraint is satisfied for $K = 2$.*

Proof: Let $P_0 = \frac{\text{SNR}}{2}$ and $P_1 = \frac{\text{SNR}}{2(\Pi(J_R=1))}$.

The power constraint is $P_0\Pi(J_T = 0) + P_1\Pi(J_T = 1) \leq \text{SNR}$, which trivially holds. ■

REFERENCES

- [1] A. J. Goldsmith and P. P. Varaiya, "Capacity of fading channels with channel side information," *IEEE Transactions on Information Theory*, vol. 43(6), pp. 1986-1992, Nov. 1997.
- [2] V. K. N. Lau and Y. R. Kwok, *Channel-Adaptive Technologies and Cross-Layer Designs for Wireless Systems with Multiple Antennas*, John Wiley & Sons, Inc., 2006.
- [3] G. Caire and S. Shamai, "On the capacity of some channels with channel state information," *IEEE Transactions on Information Theory*, vol. 45(6), pp. 2007-2019, Sep. 1999.
- [4] A. Kuhne and A. Klein, "Throughput Analysis of Multi-user OFDMA-Systems using Imperfect CQI Feedback and Diversity Techniques," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 8, pp. 1440-1450, Oct. 2008.
- [5] E. Biglieri, G. Caire and G. Taricco, "Limiting performance for block-fading channels with multiple antennas," *IEEE Trans. on Inform. Theory*, vol. 47, no. 4, pp. 1273-1289, May 2001.
- [6] M. Agarwal, D. Guo and M. Honig, "Limited-Rate Channel State Feedback for Multicarrier Block Fading Channels," Submitted to *IEEE Transactions on Information Theory*, Jan. 2009.
- [7] W. Shin, S. Chung and Y. H. Keem, "Outage analysis for MIMO Rician channels and channels with partial CSI," in *Proc. International Symposium on Information Theory*, Jul. 2006.
- [8] A. Khoshnevis and A. Sabharwal, "On the asymptotic performance of multiple antenna channels with quantized feedback," *IEEE Transactions on Wireless Communications*, 10 (7), pp. 3869 - 3877, October 2008.
- [9] V. Sharma, K. Premkumar and R. N. Swamy, "Exponential diversity achieving spatio-temporal power allocation scheme for fading channels," *IEEE Transactions on Information Theory*, vol. 54, no. 1, Jan. 2008.
- [10] C. Steger and A. Sabharwal, "Single-Input Two-Way SIMO Channel: Diversity-Multiplexing Tradeoff with Two-Way Training," to appear in *IEEE Transactions on Wireless Communications*, December 2008.
- [11] G. G. Krishna, S. Bhashyam and A. Sabharwal, "Decentralized power control with two-way training for multiple access," in *Proc. International Symposium on Information Theory*, July 2008, Toronto.
- [12] S. Ekbatani, F. Etemadi and H. Jafarkhani, "Outage behavior of slow fading channels with power control using noisy quantized CSIT," *arXiv:0804.0790v1*, Apr. 2008.
- [13] T. T. Kim and M. Skoglund, "Diversity-Multiplexing tradeoff in MIMO channels with partial CSIT," *IEEE Transactions on Information Theory*, vol. 53, Issue 8, pp. 2743-2759, Aug. 2007.
- [14] V. Aggarwal and A. Sabharwal, "Performance of multiple access channels with asymmetric feedback," *IEEE Journal on Selected Areas in Communication*, vol. 26, no. 8, pp. 1516-1525, Oct. 2008.
- [15] T. T. Kim and G. Caire, "Diversity gains of power control with noisy CSIT in MIMO channels," *IEEE Trans. Inf. Th.*, accepted for publication.
- [16] V. Aggarwal and A. Sabharwal, "Diversity order gain with noisy feedback in multiple access channels," in *Proc. International Symposium on Information Theory*, July 2008, Toronto.
- [17] A. Khoshnevis and A. Sabharwal, "Achievable diversity and multiplexing in multiple antenna systems with quantized power control," in *Proc. IEEE Intl. Conference on Communications*, May 2005.
- [18] H. El Gamal, G. Caire, M. O. Damen, "The MIMO ARQ Channel: Diversity-Multiplexing-Delay Tradeoff," *IEEE Transactions on Information Theory*, vol. 52, pp. 3601-3621, Aug. 2006.
- [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 JSAC*, vol. 16, pp. 1423-1436, Oct. 1998.
- [20] 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, pp. 2562-2579, Oct. 2003.
- [21] L. Zheng, "Diversity-Multiplexing Tradeoff: A Comprehensive View of Multiple Antenna Systems," *PhD Thesis*, University of California at Berkeley, 2002.
- [22] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: a fundamental tradeoff in multiple-antenna channels," *IEEE Transactions on Information Theory*, vol. 49, Issue 5, pp. 1073-1096, May 2003.
- [23] S. Bhashyam, A. Sabharwal and B. Aazhang, "Feedback gain in multiple antenna systems," *IEEE Transactions on Comm.*, vol. 50, Issue 5, pp. 785-798, May 2002.
- [24] V. Aggarwal, G. G. Krishna, S. Bhashyam and A. Sabharwal, "Two Models for Noisy Feedback in MIMO Channels," in *Proc. Asilomar Conference on Signals, Systems and Computers*, Oct. 2008, Pacific Grove, CA.
- [25] M. Biguesh and A. B. Gershman, "Training-based MIMO channel estimation: a study of estimator tradeoffs and optimal training signals," *IEEE Transactions on Signal Processing*, vol. 54, no. 3, pp. 884-893, Mar. 2006.
- [26] S. Wang, A. Abdi, "Joint singular value distribution of two correlated rectangular Gaussian matrices and its application," *SIAM Journal on Matrix Analysis and Applications*, volume 29, issue 3, pp. 972-981, Oct. 2007.
- [27] G. G. Krishna, "Feedback with resource accounting in MIMO systems," *B. Tech. Project Report*, Indian Institute of Technology Madras, May 2008.
- [28] D. N. C. Tse, P. Viswanath and L. Zheng, "Diversity-Multiplexing tradeoff in multiple-access channels," *IEEE Transactions on Information Theory*, vol. 50, Issue 9, pp. 1859-1874, Sept. 2004.
- [29] L. Zhao, W. Mo, Y. Ma and Z. Wang, "Diversity and multiplexing tradeoff in general fading channels," *IEEE Transactions on Information Theory*, vol. 53(4), pp. 1549-1557, Apr. 2007.

- [30] C. M. Bender, D. C. Brody and B. K. Meister, "On powers of Bessel functions," *Journal of Mathematical Physics*, vol. 44 no. 1, Jan. 2003.
- [31] M. Abramowitz and I. A. Stegun, "Handbook of Mathematical Functions." Courier Dover Publications, 1995.
- [32] V. Aggarwal and A. Sabharwal, "Bits About the Channel: Multi-round Protocols for Two-way Fading Channels," *Submitted to IEEE Trans. Inf. Th.*, Sept 2009, available at arXiv:0909.1011.
- [33] Won-Yong Shin, Sae-Young Chung and Yong H. Lee, "Diversity-Multiplexing Tradeoff and Outage Performance for Rician MIMO Channels," *IEEE Transactions on Information Theory*, Mar. 2008.

Vaneet Aggarwal received the B.Tech. degree in 2005 from the Indian Institute of Technology, Kanpur, India and the M.A. degree in 2007 from Princeton University, Princeton, NJ, USA, both in Electrical Engineering. He is currently pursuing the Ph.D. degree in Electrical Engineering at Princeton University, Princeton, NJ, USA.

His research interests are in applications of information and coding theory to wireless systems and quantum error correction. He was the recipient of Princeton University's Porter Ogden Jacobus Honorific Fellowship in 2009.

Ashutosh Sabharwal (S'91 - M'99 - SM'04) received the B.Tech. degree from the Indian Institute of Technology, New Delhi, in 1993 and the M.S. and Ph.D. degrees from The Ohio State University, Columbus, in 1995 and 1999, respectively.

He is currently an Assistant Professor in the Department of Electrical and Computer Engineering and also the Director of Center for Multimedia Communications at Rice University, Houston, TX. His research interests are in the areas of information theory and communication algorithms for wireless systems. Dr. Sabharwal was the recipient of Presidential Dissertation Fellowship award in 1998.