

Optimization of Training and Feedback Overhead for Beamforming over Block Fading Channels

Wiroonsak Santipach and Michael L. Honig

Abstract

We examine the capacity of beamforming over a single-user, multi-antenna link taking into account the overhead due to channel estimation and limited feedback of channel state information. Multi-input single-output (MISO) and multi-input multi-output (MIMO) channels are considered subject to block Rayleigh fading. Each coherence block contains L symbols, and is spanned by T training symbols, B feedback bits, and the data symbols. The training symbols are used to obtain a Minimum Mean Squared Error estimate of the channel matrix. Given this estimate, the receiver selects a transmit beamforming vector from a codebook containing 2^B *i.i.d.* random vectors, and sends the corresponding B bits back to the transmitter. We derive bounds on the beamforming capacity for MISO and MIMO channels and characterize the optimal (rate-maximizing) training and feedback overhead (T and B) as L and the number of transmit antennas N_t both become large. The optimal N_t is limited by the coherence time, and increases as $L/\log L$. For the MISO channel the optimal T/L and B/L (fractional overhead due to training and feedback) are asymptotically the same, and tend to zero at the rate $1/\log N_t$. For the MIMO channel the optimal feedback overhead B/L tends to zero faster (as $1/\log^2 N_t$).

This work was supported by the U.S. Army Research Office under grant W911NF-07-1-0028 and the National Science Foundation under grant CCR-0310809, and was presented in part at the IEEE International Symposium on Information Theory, Seattle, WA, July 2006, and the IEEE Wireless Communications and Networking Conference, Hong Kong, China, March, 2007.

W. Santipach is with the Department of Electrical Engineering, Faculty of Engineering, Kasetsart University, Bangkok 10900, Thailand; Email: wiroonsak.s@ku.ac.th.

M. L. Honig is with the Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, Illinois 60208 USA; Email: mh@eecs.northwestern.edu

I. INTRODUCTION

With perfect channel knowledge at the transmitter and receiver, the capacity of a multi-antenna system with independent Rayleigh fading increases with the number of antennas [1], [2]. In practice, the channel estimate at the receiver will not be perfect, and furthermore, this estimate must be quantized before it is relayed back to the transmitter. This has motivated work on the performance of feedback schemes with imperfect channel knowledge [3]–[9], and the design and performance of limited feedback schemes for Multi-Input Multi-Output (MIMO) and Multi-Input Single-Output (MISO) channels (e.g., see [9]–[17] and the recent survey paper [18]). All of the previous work on limited feedback assumes perfect channel knowledge at the receiver. Here we consider a model that takes into account *both* imperfect channel estimation at the receiver *and* limited channel state feedback.

We focus on single-user MISO and MIMO links with rank-one precoders (beamforming), and study the achievable rate as a function of overhead for channel estimation and channel state feedback. Our objective is to characterize the optimal amount of overhead and the associated achievable rate, and to show how those scale with the system size (i.e., as the number of transmit and/or receive antennas become large). Motivated by practical systems, a pilot-based scheme for channel estimation is assumed. Given a finite coherence time, the number of antennas that can be used effectively is limited by the channel estimation error and quantization error associated with the transmit beam. We show how the optimal (rate-maximizing) number of transmit antennas scales with the system size.

More specifically, an independent identically distributed (*i.i.d.*) block Rayleigh fading channel is considered in which the channel parameters are stationary within each coherence block, and are independent from block to block. The block length L is assumed to be constant, and the transmitted codewords span many blocks, so that the maximum achievable rate is the ergodic capacity. Each coherence block contains T training symbols and D data symbols. Furthermore, we assume that after transmission of the training symbols, the transmitter waits for the receiver to relay B bits over a feedback channel, which specify a particular beamforming vector. This delay, in addition to the T training symbols, must occur within the coherence block, and is therefore counted as

part of the packet overhead.¹

We assume that the receiver computes a Minimum Mean Square Error (MMSE) estimate of the channel, based on the training symbols, and uses the noisy channel estimate to choose a transmit beamforming vector. The Random Vector Quantization (RVQ) scheme in [14], [16], [21] is assumed in which the beamformer is selected from a codebook consisting of 2^B random vectors, which are independent and isotropically distributed, and known *a priori* at the transmitter and receiver. The associated codebook index is relayed using B bits via a noiseless feedback channel to the transmitter. The capacity of this scheme with perfect channel estimation is analyzed in [14], [16], [17], [21], [22]. It is shown in [14] that the RVQ codebook is optimal (i.e., maximizes the capacity) in the large system limit in which number of transmit antennas N_t and B tend to infinity with fixed ratio $\bar{B} = B/N_t$. In [14], [23], RVQ has been observed to give essentially optimal performance for systems with small N_t . Furthermore, for the MISO channel the performance averaged over the random codebooks can be explicitly computed [16].

The capacity with MMSE channel estimates at the receiver (with or without limited feedback) is unknown. We derive upper and lower bounds on the capacity with RVQ and limited feedback, which are functions of the number of training symbols T and feedback bits B . Given a fixed block size, or coherence time L , we then optimize the capacity bounds over B and T . Namely, small T leads to a poor channel estimate, which decreases capacity, whereas large T leads to an accurate channel estimate, but leaves few symbols in the packet for transmitting the message. This trade-off has been studied in [24], [25] for MIMO channels without feedback. Here there is also an optimal amount of feedback B , which increases with the training interval T . That is, more feedback is needed to quantize more accurate channel estimates.

We characterize the optimal overhead due to training and feedback in the large system limit as the coherence time L and number of transmit antennas N_t both tend to infinity with fixed ratio $\bar{L} = L/N_t$. For the MIMO channel we also let the number of receiver antennas $N_r \rightarrow \infty$ with fixed N_t/N_r . This allows a characterization of the achievable rate as a function of the number of feedback bits per degree of freedom [14].²

¹An implicit assumption is that the transmitter cannot learn the channel by detecting a received signal in the reverse direction, as in some Time-Division Duplex systems (e.g., see [19]). Although the feedback overhead is counted as part of the coherence time, a similar penalty arises with a Frequency-Division Duplex model [20].

²See also the tutorial on large random matrix theory [26].

For both MISO and MIMO channels the optimal normalized training $\bar{T} = T/L$, which maximizes the bounds on capacity, tends to zero at the rate $1/\log N_t$. For the MISO channel the normalized feedback $\bar{B} = B/L$ also tends to zero at this rate. Moreover, the training and feedback require the same asymptotic overhead. For the MIMO channel the optimal $\bar{B} = B/L$ tends to zero at the rate $1/\log^2 N_t$. Hence the overhead due to feedback is lower for the MIMO channel than for the MISO channel. This is apparently due to the additional degrees of freedom at the receiver, which can compensate for the performance loss associated with quantization error.

For both MISO and MIMO channels, the optimal T increases as $N_t/\log N_t$, and we observe that the associated capacity can be achieved by activating only $N_t/\log N_t$ antennas (assuming N_t increases linearly with L). Equivalently, for this pilot-based scheme with limited feedback, the optimal number of (active) transmit antennas increases as $L/\log L$. Hence the training and feedback overhead pose a fundamental limit on the number of antennas that can be effectively used. The capacity with optimized overhead grows as $\log N_t$. This is the same as with perfect channel knowledge; however, there is a second-order loss term, which increases as $\log \log N_t$.

A similar type of model for optimizing feedback overhead has been previously considered in [20]. A key difference is that here the relation between training and channel estimation error is explicitly taken into account. The model we present is also closely related to the two-way limited feedback system considered in [27], [28] (see also [19]). However, here the feedback channel is simply modeled with a fixed rate (i.e., is not the result of an optimization), and reflects the likelihood that the forward channel may be quite different from the reverse (feedback) channel. Also, the scaling of the optimal overhead and capacity with system size, given a fixed coherence time and fixed feedback rate, is not addressed in the preceding references. Similar types of overhead and capacity scaling results to those presented here are presented in [29] for a single-user wideband multi-carrier channel and in [30] for the cellular downlink based on Orthogonal Frequency Division Multiple Access.

The rest of the paper is organized as follows. Section II describes the multi-antenna channel model. Bounds on the beamforming capacity for the MISO channel with channel estimation and limited feedback are presented in Section III along with a characterization of the optimal (capacity-maximizing) training and feedback lengths in the large system limit. Corresponding results for the MIMO channel are presented in Section IV. Numerical results for finite-size MISO

and MIMO channels are shown in Section V, and conclusions are presented in Section VI.

II. SYSTEM MODEL

We consider a point-to-point *i.i.d.* block fading channel with N_t transmit antennas and N_r receive antennas. A rich scattering environment is assumed so that the channel gains corresponding to different pairs of transmit/receive antennas are independent and Rayleigh distributed. The i th $N_r \times 1$ received vector in a particular block is given by

$$\mathbf{r}(i) = \mathbf{H}\mathbf{v}b(i) + \mathbf{n}(i) \quad \text{for } 1 \leq i \leq D \quad (1)$$

where \mathbf{H} is an $N_r \times N_t$ channel matrix whose elements are independent, complex Gaussian random variables with zero mean and unit variance, \mathbf{v} is an $N_t \times 1$ unit-norm beamforming vector, b is the transmitted symbol with unit variance, \mathbf{n} is additive white Gaussian noise (AWGN) with covariance $\sigma_n^2 \mathbf{I}$, and D is the number of data (information) symbols in a block.

A. Random Vector Quantization

In prior work [14], we have analyzed the channel capacity with perfect channel knowledge at the receiver, but with *limited* channel knowledge at the transmitter. Specifically, the optimal beamformer is quantized at the receiver, and the quantized version is relayed back to the transmitter. Given the quantization codebook $\mathcal{V} = \{\mathbf{v}_1, \dots, \mathbf{v}_{2^B}\}$, which is also known *a priori* at the transmitter, and the channel \mathbf{H} , the receiver selects the quantized beamforming vector to maximize the instantaneous rate,

$$\mathbf{v}(\mathbf{H}) = \arg \max_{\mathbf{v}_j \in \mathcal{V}} \{\log(1 + \rho \|\mathbf{H}\mathbf{v}_j\|^2)\} \quad (2)$$

where $\rho = 1/\sigma_n^2$ is the background signal-to-noise ratio (SNR). The (uncoded) index for the rate-maximizing beamforming vector is relayed to the transmitter via an error-free feedback link. The capacity depends on the beamforming codebook \mathcal{V} and B . With unlimited feedback ($B \rightarrow \infty$) the $\mathbf{v}(\mathbf{H})$ that maximizes the capacity is the eigenvector of $\mathbf{H}^\dagger \mathbf{H}$, which corresponds to the maximum eigenvalue.

We will assume that the codebook vectors are independent and isotropically distributed over the unit sphere. It is shown in [14], [21] that this RVQ scheme is optimal (i.e., maximizes the achievable rate) in the large system limit in which $(B, N_t, N_r) \rightarrow \infty$ with fixed normalized

feedback $\bar{B} = B/N_t$ and $\bar{N}_r = N_r/N_t$. (For the MISO channel $N_r = 1$.) Furthermore, the corresponding capacity grows as $\log(\rho N_t)$, which is the same order-growth as with perfect channel knowledge at the transmitter. Although strictly speaking, RVQ is suboptimal for a finite-size system, numerical results indicate that the average performance is often indistinguishable from the performance with optimized codebooks [14], [23].

B. Channel Estimation

In addition to limited channel information at the transmitter, here we also account for channel estimation error at the receiver. Letting $\hat{\mathbf{H}}$ be the estimated channel matrix, the receiver selects $\mathbf{v}(\hat{\mathbf{H}})$ assuming that $\hat{\mathbf{H}}$ is the actual channel, i.e.,

$$\mathbf{v}(\hat{\mathbf{H}}) = \arg \max_{\mathbf{v}_j \in \mathcal{V}} \left\{ \log(1 + \rho \|\hat{\mathbf{H}} \mathbf{v}_j\|^2) \right\}. \quad (3)$$

We will assume that the receiver computes the linear MMSE estimate of \mathbf{H} given the received vectors corresponding to T training vectors. Specifically, the transmitter transmits T training symbols $b_T(1), \dots, b_T(T)$, where the training symbol $b_T(i)$ modulates the corresponding beamforming vector $\mathbf{v}_T(i)$. For the MISO channel the row vector of T received samples is given by

$$\mathbf{r}_T = \mathbf{h} \mathbf{V}_T \mathbf{B}_T + \mathbf{n}_T \quad (4)$$

where the channel \mathbf{h} is a $1 \times N_t$ row vector, $\mathbf{V}_T = [\mathbf{v}_T(1) \cdots \mathbf{v}_T(T)]$, $\mathbf{B}_T = \text{diag}\{b_T(i)\}$, and $\mathbf{n}_T = [n(1) \cdots n(T)]$. The channel estimate is $\hat{\mathbf{h}} = \mathbf{r}_T \mathbf{C}$, where the $T \times N_t$ linear MMSE channel estimation filter is given by

$$\mathbf{C} = \arg \min_{\tilde{\mathbf{C}}} E[\|\mathbf{h} - \mathbf{r}_T \tilde{\mathbf{C}}\|^2] \quad (5)$$

$$= \mathbf{V}_T \mathbf{B}_T (\mathbf{V}_T^\dagger \mathbf{V}_T + \sigma_n^2 \mathbf{I})^{-1}. \quad (6)$$

The MSE

$$\sigma_w^2 = E[\|h_i - \hat{h}_i\|^2] = 1 - \frac{1}{N_t} \text{trace}\{\mathbf{C}^\dagger \mathbf{R}_T \mathbf{C}\} \quad (7)$$

where h_i and \hat{h}_i are i th elements of \mathbf{h} and $\hat{\mathbf{h}}$, respectively, and the received covariance matrix

$$\mathbf{R}_T = E[\mathbf{r}_T^\dagger \mathbf{r}_T] = \mathbf{B}_T \mathbf{V}_T^\dagger \mathbf{V}_T \mathbf{B}_T + \sigma_n^2 \mathbf{I}. \quad (8)$$

The preceding expressions also apply to the MIMO channel where the estimation is for a particular row of \mathbf{H} . That is, \mathbf{C} is replaced by \mathbf{C}_i , which is applied to the i th receiver antenna,

and used to estimate the i th row of \mathbf{H} . The MSE for each element of \mathbf{H} therefore remains the same.

Because the elements of \mathbf{H} are assumed to be complex *i.i.d.* Gaussian random variables, we have

$$\mathbf{H} = \hat{\mathbf{H}} + \mathbf{w} \quad (9)$$

where the estimate $\hat{\mathbf{H}}$ and the error matrix \mathbf{w} are independent, and each contain *i.i.d.* complex Gaussian elements. The elements of \mathbf{w} have zero mean and variance σ_w^2 , so that $\hat{\mathbf{H}}$ has zero mean and covariance $(1 - \sigma_w^2)\mathbf{I}$.

The variance σ_w^2 clearly decreases as T increases. Furthermore, since the beamforming vectors during training \mathbf{V}_T are known *a priori* to the transmitter and receiver, those can be chosen to minimize the MSE. It is shown in [24] that the corresponding set of (unit-norm) beamforming vectors achieves the Welch bound with equality. We therefore have that [31]

$$\mathbf{V}_T \mathbf{V}_T^\dagger = \bar{T} \mathbf{I} \quad \text{if } T > N_t, \quad (10)$$

$$\mathbf{V}_T^\dagger \mathbf{V}_T = \mathbf{I} \quad \text{if } T \leq N_t. \quad (11)$$

Applying (6)-(11), we obtain the variance of the estimation error

$$\sigma_w^2 = \begin{cases} 1 - \frac{\bar{T}}{1+\rho^{-1}}, & \bar{T} < 1 \\ \frac{1}{1+\rho T}, & \bar{T} \geq 1 \end{cases}. \quad (12)$$

C. Ergodic Capacity

In what follows, we assume that the forward and feedback links are time-division multiplexed, and each block consists of T training symbols, B feedback bits, and D data symbols. Given that the size of each block is L symbols, we have the constraint

$$L = T + \mu B + D \quad (13)$$

where μ is a conversion factor, which relates bits to symbols. Our objective is to maximize the ergodic capacity, which is the maximum mutual information between b and \mathbf{r} ,

$$\max_{T,B} \{C = E[\max_{p_b} I(\mathbf{r}; b | \mathbf{H}, \hat{\mathbf{H}}, \mathbf{v}(\hat{\mathbf{H}}))]\} \quad (14)$$

subject to (13), where p_b is the probability density function (pdf) for the transmitted symbol b , and the expectation is over the channel \mathbf{H} , the estimation error \mathbf{w} , and the RVQ codebook

\mathcal{V} . Determining the ergodic capacity of RVQ with channel estimation appears to be intractable, so instead we derive upper and lower bounds, which are functions of D , B , and T . We then maximize both bounds over $\{D, B, T\}$, subject to (13).

III. MULTI-INPUT SINGLE-OUTPUT CHANNEL

A. Capacity Bounds

We first consider a MISO channel with $1 \times N_t$ channel vector \mathbf{h} . Applying Jensen's inequality, we obtain the upper bound on ergodic capacity

$$C = E[\max_{p_b} I(b; \mathbf{r} | \hat{\mathbf{h}}, \mathbf{v}(\hat{\mathbf{h}}), \mathbf{h})] \quad (15)$$

$$= E[\log(1 + \rho |\mathbf{h}\mathbf{v}(\hat{\mathbf{h}})|^2)] \quad (16)$$

$$\leq \log(1 + \rho E[|\mathbf{h}\mathbf{v}(\hat{\mathbf{h}})|^2]) \quad (17)$$

where the maximizing pdf is Gaussian, and the expectation is over \mathbf{h} , the estimation error \mathbf{w} , and the random codebook \mathcal{V} . Substituting $\mathbf{h} = \hat{\mathbf{h}} + \mathbf{w}$ into the expectation in (17) and simplifying gives

$$E[|\mathbf{h}\mathbf{v}(\hat{\mathbf{h}})|^2] = \sigma_w^2 + E[|\hat{\mathbf{h}}\mathbf{v}(\hat{\mathbf{h}})|^2]. \quad (18)$$

Since $\|\hat{\mathbf{h}}\|^2$ and $\nu \triangleq |\hat{\mathbf{h}}\mathbf{v}(\hat{\mathbf{h}})|^2 / \|\hat{\mathbf{h}}\|^2$ are independent [13], [16], we have

$$E[|\hat{\mathbf{h}}\mathbf{v}(\hat{\mathbf{h}})|^2] = E[\|\hat{\mathbf{h}}\|^2] E[\nu] = (1 - \sigma_w^2) N_t E[\nu]. \quad (19)$$

With RVQ we have

$$\nu = \max_{1 \leq j \leq 2^B} \{\nu_j = |\hat{\mathbf{h}}\mathbf{v}_j|^2 / \|\hat{\mathbf{h}}\|^2\} \quad (20)$$

where the ν_j 's are *i.i.d.* with pdf given in [12]. The pdf for ν and associated mean can be explicitly computed [16]. The mean is given by

$$E[\nu] = 1 - 2^B B \left(2^B, \frac{N_t}{N_t - 1} \right) \quad (21)$$

where the beta function $B(m, n) = \int_0^1 t^{m-1} (1-t)^{n-1} dt$ for m and $n > 0$. We can bound $E[\nu]$ as follows.

Lemma 1: For $\bar{B} \geq 0$ and $N_t \geq 2$,

$$E[\nu] \leq 1 - 2^{-\bar{B}} + \frac{1 + (\gamma - 1)2^{-\bar{B}} + 2^{-\bar{B}N_t}}{N_t - 1} \quad (22)$$

$$E[\nu] \geq 1 - 2^{-\bar{B}} \quad (23)$$

where $\gamma = 0.5772\dots$ is the Euler constant.

The proof is given in Appendix A. We note that $E[\nu] \rightarrow 1 - 2^{-\bar{B}}$ as $N_t \rightarrow \infty$. Substituting (18)-(22) into (17) gives an upper bound on capacity.

To derive a lower bound on capacity, we use the estimation error equation $\mathbf{h} = \hat{\mathbf{h}} + \mathbf{w}$ to write

$$r(i) = (\hat{\mathbf{h}}\mathbf{v}(\hat{\mathbf{h}}))b(i) + \underbrace{(\mathbf{w}\mathbf{v}(\hat{\mathbf{h}}))b(i) + n(i)}_{z(i)}. \quad (24)$$

Since \mathbf{w} and $\hat{\mathbf{h}}$ are independent, it follows that $E[z(i)b(i)] = 0$. It is shown in [24], [32] that replacing $z(i)$ with a zero-mean Gaussian random variable minimizes the mutual information $I(r; b|\hat{\mathbf{h}}, \mathbf{v}(\hat{\mathbf{h}}))$ and therefore gives a lower bound on the capacity with channel estimation and quantized beamforming. The lower bound is maximized when $b(i)$ has a Gaussian pdf, i.e.,

$$C \geq E[\max_{p_b} \min_{p_z} I(r; b|\hat{\mathbf{h}}, \mathbf{v}(\hat{\mathbf{h}}))] = E \left[\log \left(1 + \frac{|\hat{\mathbf{h}}\mathbf{v}(\hat{\mathbf{h}})|^2}{\sigma_z^2} \right) \right] \quad (25)$$

where p_z and σ_z^2 denote the pdf and variance for z , respectively. We derive the following lower bound on C by applying the inequality in [33].

Lemma 2:

$$E \left[\log \left(1 + \frac{1}{\sigma_z^2} |\hat{\mathbf{h}}\mathbf{v}(\hat{\mathbf{h}})|^2 \right) \right] \geq (1 - d_{N_t}) \log \left(1 + \frac{1}{\sigma_z^2} E[|\hat{\mathbf{h}}\mathbf{v}(\hat{\mathbf{h}})|^2] \right) \quad (26)$$

where

$$d(N_t) = \frac{1}{2} \sqrt{\frac{\frac{1}{N_t} + \left(1 + \frac{1}{N_t}\right) \frac{\Gamma\left(1 + \frac{2}{N_t-1}\right) - \Gamma^2\left(1 + \frac{1}{N_t-1}\right) (1 + 2^{-\bar{B}N_t})^{-\frac{2}{N_t-1}}}{2^{\bar{B} + \frac{\bar{B}}{N_t-1}} - \Gamma\left(1 + \frac{1}{N_t-1}\right)}}} \quad (27)$$

and the gamma function $\Gamma(m) = \int_0^\infty t^{m-1} e^{-t} dt$ for $m > 0$.

The proof is given in Appendix A. We note that $d(N_t) \rightarrow 0$ as $N_t \rightarrow \infty$.

To obtain a lower bound on capacity C , we substitute $\sigma_z^2 = \sigma_w^2 + \sigma_n^2$, (23), and (26)-(27) into (25). The capacity bounds are summarized as follows.

Theorem 1: The capacity for a MISO channel with channel estimation variance σ_w^2 and normalized feedback \bar{B} satisfies

$$C_l \leq C \leq C_u \quad \text{for } \bar{B} \geq 0 \text{ and } N_t \geq 2 \quad (28)$$

where

$$C_l = (1 - d(N_t)) \log \left(1 + \rho \frac{1 - \sigma_w^2}{1 + \rho \sigma_w^2} (1 - 2^{-\bar{B}}) N_t \right), \quad (29)$$

$$C_u = \log \left(1 + \rho \sigma_w^2 + \rho (1 - \sigma_w^2) N_t \left(1 - 2^{-\bar{B}} + \frac{1 + (\gamma - 1) 2^{-\bar{B}} + 2^{-\bar{B} N_t}}{N_t - 1} \right) \right). \quad (30)$$

The gap between the two bounds tends to zero as $\rho \rightarrow 0$ (since both C_u and C_l tend to zero), and as $N_t \rightarrow \infty$. With fixed \bar{B} and σ_w^2 the bounds (and the capacity) grow as $O(\log N_t)$ as $N_t \rightarrow \infty$. Substituting (12) for σ_w^2 gives the bounds as a function of training T .

Fig. 1 compares the bounds in Theorem 1 with (16) and the tighter lower bound (25). The bounds are plotted versus N_t with parameters $B/N_t = 1$ (one bit per antenna coefficient), $\sigma_w^2 = 0.15$, and SNR $\rho = 5$ dB. The tighter bounds, which are analytically intractable, are evaluated by Monte Carlo simulation and shown as \circ 's and \times 's in the figure. The plots show that the upper bound in Theorem 1 is close to (16) even for small N_t while the lower bound in the Theorem is close to (25) for much larger N_t . Since RVQ requires an exhaustive search over the codebook, and the number of entries in the codebook grows exponentially with the number of antennas, simulation results are not shown for $N_t > 12$. As expected, both the upper and lower bounds grow at the same rate as N_t increases.

B. Asymptotic Behavior

We now study the behavior of the optimal T , B and D , and the capacity as $N_t \rightarrow \infty$. With D transmitted symbols in an L -symbol packet the effective capacity $\mathcal{C} = (\bar{D}/\bar{L})C$ where $\bar{D} = D/N_t$ and $\bar{L} = L/N_t$. The associated bounds are $\mathcal{C}_u = (\bar{D}/\bar{L})C_u$ and $\mathcal{C}_l = (\bar{D}/\bar{L})C_l$. From Theorem 1 and (12), we can write C_l and C_u as functions of $\{\bar{T}, \bar{B}, \bar{D}\}$ and optimize, i.e., for the lower bound we wish to

$$\max_{\bar{T}, \bar{B}, \bar{D}} C_l \quad (31)$$

$$\text{subject to } \bar{T} + \mu \bar{B} + \bar{D} = \bar{L}. \quad (32)$$

Let $\{\bar{T}_l^o, \bar{B}_l^o, \bar{D}_l^o\}$ denote the optimal values of \bar{T} , \bar{B} , and \bar{D} , respectively, and let \mathcal{C}_l^o denote the maximized lower bound on capacity. Similarly, maximizing the upper bound gives the optimal parameters $\{\bar{T}_u^o, \bar{B}_u^o, \bar{D}_u^o\}$ and the corresponding bound \mathcal{C}_u^o . These optimized values can be easily

computed numerically, and also allow us to characterize the asymptotic behavior of the *actual* capacity.³

Theorem 2: Let $\{\bar{T}^o, \bar{B}^o, \bar{D}^o\} = \arg \max_{\{\bar{T}, \bar{B}, \bar{D}\}} \mathcal{C}$ subject to (32). As $N_t \rightarrow \infty$,

$$\bar{T}^o \log N_t \rightarrow \bar{L} \quad (33)$$

$$\bar{B}^o \log N_t \rightarrow \frac{1}{\mu} \bar{L} \quad (34)$$

and the capacity satisfies

$$\mathcal{C}^o - \log(\rho N_t) + 2 \log \log N_t \rightarrow \zeta \quad (35)$$

where ζ is a constant bounded by

$$\zeta^* - \log(1 + \rho) \leq \zeta \leq \zeta^* \quad (36)$$

where $\zeta^* = \log(\bar{L}^2 \log(2)) - \log(\mu(1 + \rho^{-1})) - 2$.

The proof is given in Appendix C. Combining (33) and (34) with (32) gives the corresponding behavior of the data segment

$$\frac{\bar{D}^o}{\bar{L}} = 1 - \delta(N_t) \quad (37)$$

where $\delta(N_t) \log N_t / 2 \rightarrow 1$.

According to the theorem, as N_t becomes large, to maximize the achievable rate the fraction of \bar{L} devoted to training and feedback tends to zero, in which case the rate increases as $\log(\rho N_t) - 2 \log \log N_t$. The achievable rate with RVQ and perfect channel estimation is $E[\log(1 + \rho \|\mathbf{h}\|^2)]$, which grows as $\log(\rho N_t)$. Hence the loss of $2 \log \log N_t$ is due to imperfect channel estimation.⁴ Theorem 2 also implies that $\mu B/T \rightarrow 1$, i.e., the fraction of the packet devoted to feedback is asymptotically the same as that for training. This equal allocation therefore balances the reductions in capacity due to estimation and quantization.

The preceding analysis applies if the beamforming vectors during training are chosen to be unit vectors. Namely, the matrix \mathbf{V}_T can be taken to be diagonal, which corresponds to transmitting the sequence of training symbols over the transmit antennas successively one at a time. Hence the fact that the optimal T increases as $N_t / \log N_t$ implies that only $N_t / \log N_t$ antennas are

³In what follows all logarithms are assumed to be natural.

⁴The capacity estimate in the theorem becomes accurate when N_t is large enough so that $\bar{L} / \log N_t$ is small, in which case the loss term $2 \log \log N_t$ is greater than the constant offset ζ .

activated. Since $\bar{L} = L/N_t$ is fixed, we conclude that as the coherence time L increases, the optimal number of transmit antennas should increase as $L/\log L$. The training and feedback overhead therefore reduces the number of antennas that can be effectively used by a factor of $1/\log L$.

IV. MULTI-INPUT MULTI-OUTPUT CHANNEL

In this section, we let the number of receive antennas N_r scale with N_t . As for the MISO channel, we can bound the capacity with limited training and feedback as follows,

$$C \leq C_u = \log(1 + \rho\sigma_w^2 + \rho E[\eta]) \quad (38)$$

$$C \geq C_l = (1 - c(N_t)) \log \left(1 + \frac{\rho}{1 + \rho\sigma_w^2} E[\eta] \right) \quad (39)$$

where $\eta = \mathbf{v}(\hat{\mathbf{H}})^\dagger \hat{\mathbf{H}}^\dagger \hat{\mathbf{H}} \mathbf{v}(\hat{\mathbf{H}})$ and

$$c(N_t) = \frac{\sigma_\eta}{2E[\eta]} \quad (40)$$

where σ_η is the standard deviation of η .

We would like to express the bounds (38) and (39) as functions of \bar{T} and \bar{B} . As discussed in Section II, the variance of the estimation error is again given by (12). Although it is difficult to evaluate $E[\eta]$ explicitly for finite (N_t, N_r, B) , it can be computed in the large system limit as the parameters tend to infinity with fixed ratios $\bar{N}_r = N_r/N_t$ and \bar{B} . Specifically, since $\hat{\mathbf{H}}$ has *i.i.d.* elements with variance $1 - \sigma_w^2$, we have

$$\frac{1}{N_t} \eta \longrightarrow (1 - \sigma_w^2) \gamma_{\text{rvq}} \quad (41)$$

in the mean square sense, where the asymptotic received signal power with RVQ γ_{rvq} is evaluated in [14], and is a function of \bar{N}_r and \bar{B} . Therefore

$$E[\eta] = (1 - \sigma_w^2) \gamma_{\text{rvq}} N_t + \kappa(N_t) \quad (42)$$

where $\kappa(N_t)/N_t \rightarrow 0$. Characterizing $\kappa(N_t)$ explicitly appears to be difficult, but this is not needed to prove the following theorem.⁵ Substituting (42) and (12) into (38) and (39) gives upper and lower bounds on the capacity, C_l and C_u , respectively, as functions of \bar{T} and \bar{B} .

⁵We will assume that $\kappa(N_t)$ is a smooth function of \bar{T} and \bar{B} for all N_t , and that $\kappa(N_t)/N_t$ converges to zero uniformly over all \bar{T} and \bar{B} .

Maximizing both bounds over \bar{T} and \bar{B} leads to the following theorem, which characterizes the asymptotic behavior of the actual capacity.

Theorem 3: Let $\{\bar{T}^o, \bar{B}^o, \bar{D}^o\} = \arg \max_{\{\bar{T}, \bar{B}, \bar{D}\}} \mathcal{C}$ subject to (32). As $(N_t, N_r) \rightarrow \infty$ with fixed $\bar{N}_r = N_r/N_t$,

$$\bar{T}^o \log N_t \longrightarrow \bar{L} \quad (43)$$

$$\bar{B}^o \log^2 N_t \longrightarrow \frac{\bar{L}^2 \log 2}{2\mu^2 \bar{N}_r} \quad (44)$$

and the capacity satisfies

$$\mathcal{C}^o - \log(\rho N_t) + \log \log N_t \rightarrow \xi \quad (45)$$

where

$$\xi^* - \log(1 + \rho) \leq \xi \leq \xi^* \quad (46)$$

and $\xi^* = \log(\bar{L}\bar{N}_r) - \log(1 + \rho^{-1}) - 1$.

The proof is given in Appendix D. Combining (43), (44), and (32) gives the corresponding behavior of the optimized data segment

$$\frac{\bar{D}^o}{\bar{L}} = 1 - \epsilon_1(N_t) - \epsilon_2(N_t) \quad (47)$$

where $\epsilon_1(N_t) \log N_t \rightarrow 1$ and $\frac{2\bar{N}_r\mu}{L \log 2} \epsilon_2(N_t) \log^2 N_t \rightarrow 1$.

Theorem 3 states that the optimal training length for the MIMO channel grows as $N_t/\log N_t$, which is the same as for the MISO channel. Hence as N_t becomes large, only $N_t/\log N_t$ transmit antennas should be activated. (All receive antennas are used, since this does not change the training overhead.)

Theorem 3 also states that the capacity with limited training and feedback increases as $\log(\rho N_t) - \log \log N_t$. For large N_t the loss in achievable rate due to training and feedback therefore increases as $\log \log N_t$, as opposed to $2 \log \log N_t$ for the MISO channel. This gain is due to the smaller MIMO feedback overhead. Namely, because of the additional antennas for the MIMO channel, the optimal normalized feedback length tends to zero at the rate $1/\log^2 N_t$, as opposed to $1/\log N_t$ for the MISO channel. Note, however, that the training overhead is the same since the same training symbols are used to estimate the channel gains to all receive antennas simultaneously. Hence the ratio of optimized feedback to training overhead for the MIMO channel $\frac{\mu \bar{B}^o}{\bar{T}^o} \rightarrow 0$ as $1/\log N_t$.

V. NUMERICAL RESULTS

Fig. 2 shows achievable rates for the MISO channel versus normalized coherence time $\bar{L} = L/N_t$ with different assumptions about channel knowledge at the transmitter and receiver. Three curves are shown: (1) the optimized lower bound on capacity C_l^o , (2) the capacity assuming the receiver knows the channel, but with a quantized beamformer, and (3) the capacity with perfect channel knowledge at the transmitter and receiver (optimal beamforming). Parameters are $N_t = 10$, $\rho = 5$ dB, and $\mu = 1$ (BPSK feedback). As expected, the gaps between the curves diminishes to zero with increasing coherence time, albeit slowly. This reflects the fact that the training and feedback overhead tends to zero as $1/\log L$.

Fig. 3 illustrates the sensitivity of the capacity for the MISO channel to different choices for training and feedback overhead. The lower bound C_l^o is plotted versus the fractional overhead $(\bar{T} + \mu\bar{B})/\bar{L}$ with different relative allocations $\bar{T}/(\mu\bar{B})$. Parameters are $\bar{L} = 100$, $N_t = 6$, $\mu = 1$, and $\rho = 5$ dB. The solid line corresponds to optimized overhead T_l^o and B_l^o . The capacity is zero when $\bar{T} + \bar{B} = 0$, since the estimate is uncorrelated with the channel, and when $\bar{T} + \bar{B} = \bar{L}$, since $\bar{D} = 0$. With equal amounts of training and feedback the rate is essentially equal to that with optimized parameters. The peak is achieved when $(\bar{T} + \bar{B})/\bar{L} = 0.1$. The performance is relatively robust to this choice, i.e., small deviations from this value result in a relatively small performance loss, although the performance loss increases substantially as the deviations become larger. Likewise, the figure also shows that there is a significant performance degradation when \bar{B} deviates significantly from \bar{T} .

The optimized training, feedback, and data portions of the packet (normalized by the packet length L) versus N_t for the MIMO channel are shown in Fig. 4. These values were obtained by numerically optimizing the capacity lower bound, and are therefore denoted as B_l^o , T_l^o , and D_l^o in the figure. System parameters are $\bar{N}_r = 2$, $\bar{L} = 50$, $\mu = 1$, and $\rho = 5$ dB. As predicted by Theorem 3, both the optimal \bar{T} and \bar{B} decrease to zero, with \bar{B} decreasing somewhat faster than \bar{T} . The associated capacity lower bound is shown in Fig. 5. Also shown is the capacity lower bound with the heuristic choice of parameters $\bar{B} = 1$ (one feedback bit per coefficient) and $\bar{T} = 1.5$ (1.5 training symbols per coefficient). For $N_t = 3$, the bound with optimized parameters is approximately 10% greater than that with the heuristic choice. Those results are compared with the capacity with perfect channel knowledge at *both* the transmitter and receiver,

and the capacity with perfect channel knowledge at the receiver *only* with B_l^o feedback bits. This comparison indicates how much of the loss in achievable rate for the model considered is due to channel estimation at the receiver (including associated overhead), and how much is due to quantization of the precoding matrix.

The results show that for $N_t = 3$, the capacity with perfect channel knowledge at both the transmitter and receiver is about 40% larger than the rate with optimized feedback and training lengths. Knowing the channel at the receiver achieves most of this gain, largely due to the elimination of associated training overhead. Of course, this gap tends to zero as the block size $\bar{L} \rightarrow \infty$. Also shown in the figure for comparison is the capacity lower bound for a MISO channel with optimized training and feedback lengths. This is substantially lower than that shown for the MIMO channel. From Theorems 2 and 3 the gap between the optimized lower bounds for the MISO and MIMO channels increases as $\log \log N_t$.

Similar to Fig. 3, Fig. 6 shows the capacity lower bound versus total overhead $(\bar{T} + \mu\bar{B})/\bar{L}$ for a MIMO channel. The solid line corresponds to optimized parameters with $\bar{L} = 10$, $N_t = 9$, $\bar{N}_r = 2$, $\mu = 1$, and $\rho = 5$ dB. The curves are obtained by numerical optimization. For the case considered, these results show that the rate achieved with equal portions of training and feedback is close to the maximum (corresponding to optimized training and feedback). Allocating the overhead according to the asymptotic results in Theorem 3, i.e., taking $\mu\bar{B}/\bar{T} = \bar{L} \log 2 / (2\mu\bar{N}_r \log N_t)$, performs marginally better than allocating equal training and feedback. The total optimized overhead in this case is $(\bar{T} + \bar{B})/\bar{L} \approx 0.2$. The performance degrades when \bar{B} deviates significantly from \bar{T} (as shown by the curve corresponding to $\bar{B} = 2\bar{T}$). (The three curves shown are not extended to $(\bar{T} + \bar{B})/\bar{L} = 1$ since the simulation complexity associated with RVQ increases exponentially with \bar{B} .) Compared with the results for the MISO channel in Fig. 3, the capacity for the MIMO channel is somewhat more robust with respect to variations in overhead.

VI. CONCLUSIONS

We have presented bounds on the capacity of both MISO and MIMO block Rayleigh fading channels with beamforming, assuming limited training and feedback. For a large number of transmit antennas, we have characterized the optimal amount of training and feedback as a fraction of the packet duration, assuming linear MMSE estimation of the channel, and an RVQ

codebook for quantizing the beamforming vector. Our results show that the optimized training length for both MISO and MIMO channels increases as $N_t/\log N_t$, which can be interpreted as the optimal number of transmit antennas to activate. The ratio of optimized feedback to training overhead tends to one for the MISO channel, but tends to zero as $1/\log N_t$ for the MIMO channel, since additional receiver antennas improve robustness with respect to quantization error. The loss in capacity due to overhead increases as $\log \log N_t$ for the MIMO channel, and as $2 \log \log N_t$ for the MISO channel.

Although the pilot scheme considered is practical, it is most likely suboptimal. That is, in the absence of feedback such a pilot-based scheme is strictly suboptimal, although it is nearly optimal at high SNRs [24]. Computing the capacity of the block fading channel considered with feedback and no channel knowledge at the receiver and transmitter is an open problem. Consequently, although the optimal (capacity-maximizing) number of transmit antennas should still be limited by the coherence time, the growth rate may differ from the $L/\log L$ growth rate shown here for the pilot scheme.

The model and analysis presented here can be extended in a few different directions. A natural generalization of the MIMO beamforming model is to allow a general transmit precoding matrix with rank greater than one. The additional overhead should impose a limit on both the number of beams and antennas that can effectively be used. Also, the powers allocated to the training and data portions of the coherence block can be optimized in addition to the fraction of overhead symbols. Finally, feedback and training overhead becomes especially important in multi-user MIMO scenarios, such as the cellular downlink. The optimal overhead scaling with coherence time in those scenarios remains to be studied.

APPENDIX

A. Proof of Lemma 1

We need to evaluate (21). Letting $n = 2^B$, we first bound

$$nB \left(n, 1 + \frac{1}{N_t - 1} \right) = \frac{n\Gamma(n)\Gamma\left(1 + \frac{1}{N_t - 1}\right)}{\Gamma\left(n + 1 + \frac{1}{N_t - 1}\right)} \quad (48)$$

$$= \Gamma\left(1 + \frac{1}{N_t - 1}\right) \frac{\Gamma(n + 2)}{(n + 1)\Gamma\left(n + 1 + \frac{1}{N_t - 1}\right)} \quad (49)$$

$$\geq \Gamma\left(1 + \frac{1}{N_t - 1}\right) (n + 1)^{-\frac{1}{N_t - 1}} \quad (50)$$

$$= \Gamma\left(1 + \frac{1}{N_t - 1}\right) \left(1 + \frac{1}{n}\right)^{-\frac{1}{N_t - 1}} 2^{-\bar{B}\left(1 + \frac{1}{N_t - 1}\right)} \quad (51)$$

where we have used $B(p, q) = \Gamma(p)\Gamma(q)/\Gamma(p + q)$, the identity $\Gamma(k + 1) = k\Gamma(k)$ for $k \in \mathbb{N}$, and the inequality $\Gamma(k + 1)/\Gamma(k + x) \geq k^{1-x}$ for $0 \leq x \leq 1$ [34]. Since $\Gamma(x)$ is convex for $x \in [1, 2]$, for $N_t \geq 2$,

$$\Gamma\left(1 + \frac{1}{N_t - 1}\right) \geq \Gamma(1) + \frac{\Gamma'(1)}{N_t - 1} = 1 - \frac{\gamma}{N_t - 1} \quad (52)$$

where $\gamma = 0.5772\dots$ is the Euler constant. Expanding the second factor on the right-hand side of (51) in a Taylor series gives

$$\left(1 + \frac{1}{n}\right)^{-\frac{1}{N_t - 1}} = 1 - \frac{1}{N_t - 1} \frac{1}{n} + \frac{N_t}{2!(N_t - 1)^2} \frac{1}{n^2} - \frac{N_t(2N_t - 1)}{3!(N_t - 1)^3} \frac{1}{n^3} + \dots \quad (53)$$

$$\geq 1 - \frac{1}{n(N_t - 1)} \quad (54)$$

since the magnitude of each term in (53) is decreasing. We also expand

$$\begin{aligned} (2^{-\bar{B}})^{\frac{1}{N_t - 1}} &= 1 - \frac{1}{N_t - 1} (1 - 2^{-\bar{B}}) - \frac{N_t - 2}{2!(N_t - 1)^2} (1 - 2^{-\bar{B}})^2 \\ &\quad - \frac{(N_t - 2)(2N_t - 3)}{3!(N_t - 1)^3} (1 - 2^{-\bar{B}})^3 - \dots \end{aligned} \quad (55)$$

$$\geq 1 - \frac{1}{N_t - 1} \left[(1 - 2^{-\bar{B}}) + (1 - 2^{-\bar{B}})^2 + (1 - 2^{-\bar{B}})^3 + \dots \right] \quad (56)$$

$$= 1 - \frac{1}{N_t - 1} (2^{\bar{B}} - 1). \quad (57)$$

Substituting (52), (54), and (57) into (51) yields

$$\begin{aligned} nB \left(n, 1 + \frac{1}{N_t - 1} \right) & \\ & \geq 2^{-\bar{B}} \left(1 - \frac{\gamma}{N_t - 1} \right) \left(1 - \frac{1}{n(N_t - 1)} \right) \left(1 - \frac{2^{\bar{B}} - 1}{N_t - 1} \right) \end{aligned} \quad (58)$$

$$\geq 2^{-\bar{B}} \left[1 - \frac{1}{N_t - 1} (2^{\bar{B}} - 1 + \gamma + 2^{-B}) \right]. \quad (59)$$

The inequality (59) holds for $N_t \geq 2$ and $\bar{B} \geq 0$. Therefore

$$E[\nu] = 1 - 2^B B \left(2^B, 1 + \frac{1}{N_t - 1} \right) \quad (60)$$

$$\leq 1 - 2^{-\bar{B}} + \frac{1 + (\gamma - 1)2^{-\bar{B}} + 2^{-\bar{B}N_t}}{N_t - 1}. \quad (61)$$

To show (23), we derive the following upper bound

$$nB \left(n, 1 + \frac{1}{N_t - 1} \right) = \Gamma \left(1 + \frac{1}{N_t - 1} \right) \frac{\Gamma(n + 1)}{\Gamma \left(n + 1 + \frac{1}{N_t - 1} \right)} \quad (62)$$

$$\leq \Gamma \left(1 + \frac{1}{N_t - 1} \right) \left(n + \frac{N_t}{2(N_t - 1)} \right)^{-\frac{1}{N_t - 1}} \quad (63)$$

$$= \Gamma \left(1 + \frac{1}{N_t - 1} \right) \left(1 + \frac{N_t}{2n(N_t - 1)} \right)^{-\frac{1}{N_t - 1}} 2^{-\frac{\bar{B}}{N_t - 1}} 2^{-\bar{B}}. \quad (64)$$

The inequality (63) is shown in [35]. Since every factor in (64) is less than or equal to one, we conclude that

$$nB \left(n, 1 + \frac{1}{N_t - 1} \right) \leq 2^{-\bar{B}}, \quad (65)$$

and combining with (60) gives the lower bound (23).

B. Proof of Lemma 2

Since $\log \left(1 + \frac{1}{\sigma_z^2} X \right)$ is concave for $X \in [0, \infty)$ and

$$\lim_{t \rightarrow \infty} \frac{1}{t} \log \left(1 + \frac{1}{\sigma_z^2} t \right) = 0, \quad (66)$$

we can apply the following inequality in [33]

$$E \left[\log \left(1 + \frac{1}{\sigma_z^2} X \right) \right] \geq \left(1 - \frac{E|X - E[X]|}{2E[X]} \right) \log \left(1 + \frac{1}{\sigma_z^2} E[X] \right). \quad (67)$$

Now set $X = A\nu$, where $A \triangleq \|\hat{\mathbf{h}}\|^2$ and $\nu \triangleq |\hat{\mathbf{h}}\mathbf{v}(\hat{\mathbf{h}})|^2/\|\hat{\mathbf{h}}\|^2$. Since A and ν are independent, and using the relation $(E|X - E[X]|)^2 \leq \text{var}[X]$, we obtain

$$\frac{E|X - E[X]|}{2E[X]} \leq \frac{\sqrt{\text{var}[X]}}{2E[X]} \quad (68)$$

$$= \frac{1}{2} \sqrt{\frac{E[A^2]}{E^2[A]} \frac{E[\nu^2]}{E^2[\nu]} - 1}. \quad (69)$$

Each element in $\hat{\mathbf{h}}$ is *i.i.d.* with a complex Gaussian distribution. Hence A is Gamma distributed so that

$$\frac{E[A^2]}{E^2[A]} = 1 + \frac{1}{N_t}. \quad (70)$$

To evaluate $E[\nu^2]/E^2[\nu]$ in (69) we first compute

$$E[(1 - \nu)^2] = \int_0^1 (1 - v)^2 f_\nu(v) dv \quad (71)$$

$$= \int_0^1 (1 - v)^2 \left[n(N_t - 1) (1 - (1 - v)^{N_t - 1})^{n-1} (1 - v)^{N_t - 2} \right] dv \quad (72)$$

where $f_\nu(\cdot)$ is the pdf for ν , and is given in [16]. Applying the change of variables $q = (1 - v)^{N_t - 1}$ gives

$$E[(1 - \nu)^2] = n \int_0^1 q^{\frac{2}{N_t - 1}} (1 - q)^{n-1} dq \quad (73)$$

$$= nB\left(n, 1 + \frac{2}{N_t - 1}\right). \quad (74)$$

Therefore

$$\text{var}[\nu] = E[\nu^2] - E^2[\nu] \quad (75)$$

$$= nB\left(n, 1 + \frac{2}{N_t - 1}\right) - (1 - E[\nu])^2 \quad (76)$$

$$= nB\left(n, 1 + \frac{2}{N_t - 1}\right) - n^2 B^2\left(n, 1 + \frac{1}{N_t - 1}\right). \quad (77)$$

Applying the inequality in [35], we have

$$nB\left(n, 1 + \frac{2}{N_t - 1}\right) = \Gamma\left(1 + \frac{2}{N_t - 1}\right) \frac{\Gamma(n + 1)}{\Gamma\left(n + 1 + \frac{2}{N_t - 1}\right)} \quad (78)$$

$$\leq \Gamma\left(1 + \frac{2}{N_t - 1}\right) \left(n + \frac{1}{N_t - 1} + \frac{1}{2}\right)^{-\frac{2}{N_t - 1}}. \quad (79)$$

Substituting (79) and (50) into (77) gives

$$\begin{aligned} \text{var}[\nu] &\leq \Gamma\left(1 + \frac{2}{N_t - 1}\right) \left(n + \frac{1}{N_t - 1} + \frac{1}{2}\right)^{-\frac{2}{N_t - 1}} - \Gamma^2\left(1 + \frac{1}{N_t - 1}\right) (n + 1)^{-\frac{2}{N_t - 1}} \end{aligned} \quad (80)$$

$$\begin{aligned} &= 2^{-2\bar{B}(1 + \frac{1}{N_t - 1})} \left[\Gamma\left(1 + \frac{2}{N_t - 1}\right) \left(1 + \frac{1}{n(N_t - 1)} + \frac{1}{2n}\right)^{-\frac{2}{N_t - 1}} \right. \\ &\quad \left. - \Gamma^2\left(1 + \frac{1}{N_t - 1}\right) \left(1 + \frac{1}{n}\right)^{-\frac{2}{N_t - 1}} \right] \end{aligned} \quad (81)$$

$$\leq 2^{-2\bar{B}(1 + \frac{1}{N_t - 1})} \left[\Gamma\left(1 + \frac{2}{N_t - 1}\right) - \Gamma^2\left(1 + \frac{1}{N_t - 1}\right) \left(1 + \frac{1}{n}\right)^{-\frac{2}{N_t - 1}} \right]. \quad (82)$$

Since the second factor in (64) is less than or equal to one, we have

$$E[\nu] \geq 1 - \Gamma\left(1 + \frac{1}{N_t - 1}\right) 2^{-\bar{B}(1 + \frac{1}{N_t - 1})}. \quad (83)$$

Finally, combining (69), (70), (82), and (83) gives $E|X - E[X]| / (2E[X]) \leq d(N_t)$ in (27), which completes the proof.

C. Proof of Theorem 2

We first maximize the upper bound given by

$$\mathcal{C}_u = \frac{\bar{D}}{\bar{L}} C_u \quad (84)$$

$$= \frac{\bar{D}}{\bar{L}} \log\left(\frac{\rho}{1 + \rho^{-1}} \bar{T}(1 - 2^{-\bar{B}}) N_t\right) + \frac{\bar{D}}{\bar{L}} \log(1 + r(N_t)) \quad (85)$$

where

$$r(N_t) = \frac{(1 + \rho^{-1})^2 - \bar{T}}{\bar{T}(1 - 2^{-\bar{B}}) N_t} + \frac{1 + (\gamma - 1)2^{-\bar{B}} + 2^{-\bar{B}N_t}}{(N_t - 1)(1 - 2^{-\bar{B}})}. \quad (86)$$

The expression for σ_w^2 in (12) with $\bar{T} \leq 1$ has been used in (85), since we will show that $\bar{T} \rightarrow 0$ as $N_t \rightarrow \infty$. We wish to characterize the behavior of the optimal parameters $\{\bar{T}_u^o, \bar{B}_u^o, \bar{D}_u^o\}$ as $N_t \rightarrow \infty$.

The Lagrangian is given by

$$\mathcal{L} = \mathcal{C}_u + \lambda(\bar{L} - \bar{T} - \mu\bar{B} - \bar{D}) \quad (87)$$

where λ is the Lagrangian multiplier. Setting the partial derivatives of \mathcal{L} with respect to \bar{D} , \bar{T} , \bar{B} , and λ to zero gives the necessary conditions

$$\log\left(\frac{\rho}{1+\rho^{-1}}\right) + \log(\bar{T}) + \log(1-2^{-\bar{B}}) + \log N_t + \log(1+r(N_t)) - \bar{L}\lambda = 0 \quad (88)$$

$$\frac{\bar{D}}{\bar{T}} + \left(\frac{\bar{D}}{1+r(N_t)}\right) \frac{\partial r(N_t)}{\partial \bar{T}} - \bar{L}\lambda = 0 \quad (89)$$

$$\frac{\bar{D} \log 2}{2^{\bar{B}} - 1} + \left(\frac{\bar{D}}{1+r(N_t)}\right) \frac{\partial r(N_t)}{\partial \bar{B}} - \bar{L}\mu\lambda = 0 \quad (90)$$

$$\bar{L} - \bar{T} - \mu\bar{B} - \bar{D} = 0. \quad (91)$$

Substituting (89), (91), and the expression for $\frac{\partial r(N_t)}{\partial \bar{T}}$ into (88) gives

$$\begin{aligned} \bar{T} \log N_t + \bar{T} \log\left(\frac{\rho}{1+\rho^{-1}}\right) + \bar{T} \log(1-2^{-\bar{B}}) + \bar{T} \log \bar{T} + \bar{T} \log(1+r(N_t)) \\ = (\bar{L} - \bar{T} - \mu\bar{B}) \left(1 - \frac{(1+\rho^{-1})^2}{(1+r(N_t))(1-2^{-\bar{B}})\bar{T}N_t}\right). \end{aligned} \quad (92)$$

We first observe that $(1-2^{-\bar{B}_u})\bar{T}_u^o N_t \rightarrow \infty$ as $N_t \rightarrow \infty$. Otherwise, it easily verified from (85) that C_u must be bounded by a constant. However, this is clearly suboptimal, since if \bar{B} and \bar{T} are constants, then C_u grows as $O(\log N_t)$. This observation implies that $r(N_t) \rightarrow 0$.

As $N_t \rightarrow \infty$, the right-hand side of (92) converges to $\bar{L} - \bar{T} - \mu\bar{B}$, so that (92) implies $\bar{T} \rightarrow 0$. As $N_t \rightarrow \infty$, (92) therefore implies

$$\bar{T} \log N_t \rightarrow \bar{L} - \mu\bar{B}. \quad (93)$$

Combining (89) and (90) gives

$$\bar{B} = \frac{1}{\log 2} \log\left(1 + \frac{\log 2}{\mu} \bar{T} \left(\frac{1}{1+\xi(N_t)}\right)\right) \quad (94)$$

where

$$\xi(N_t) = \frac{\bar{T}}{1+r(N_t)} \left(\frac{\partial r(N_t)}{\partial \bar{T}} - \mu \frac{\partial r(N_t)}{\partial \bar{B}}\right). \quad (95)$$

Since $\bar{T} \rightarrow 0$, and $r(N_t) \rightarrow 0$ uniformly over \bar{T} and \bar{B} (so that the derivatives in (95) must also tend to zero), it follows that $\xi(N_t) \rightarrow 0$. Hence for large N_t (94) implies that

$$\bar{B} = \frac{1}{\mu} \bar{T} + O(\bar{T}^2), \quad (96)$$

where we have used the Taylor expansion $\log(1+x) = x + O(x^2)$ for small x . Combining (93) and (96), it follows that

$$\bar{T}_u^o \log N_t \rightarrow \bar{L}, \quad (97)$$

$$\bar{B}_u^o \log N_t \rightarrow \frac{1}{\mu} \bar{L}. \quad (98)$$

Substituting the optimal parameters in the capacity upper bound (85) gives

$$\begin{aligned} \mathcal{C}_u^o - \frac{\bar{D}_u^o}{\bar{L}} \log(\rho N_t) - \frac{\bar{D}_u^o}{\bar{L}} \log \bar{T}_u^o - \frac{\bar{D}_u^o}{\bar{L}} \log(1 - 2^{-\bar{B}_u^o}) \\ = -\frac{\bar{D}_u^o}{\bar{L}} \log(1 + \rho^{-1}) + \frac{\bar{D}_u^o}{\bar{L}} \log(1 + r(N_t)) \end{aligned} \quad (99)$$

where \mathcal{C}_u^o denotes the optimal \mathcal{C}_u . Taking $N_t \rightarrow \infty$ gives

$$\mathcal{C}_u^o - \log(\rho N_t) + 2 \log \log N_t \rightarrow \log(\bar{L}^2 \log 2) - 2 - \log[\mu(1 + \rho^{-1})]. \quad (100)$$

Following similar steps to optimize the lower bound (29) gives

$$\bar{T}_l^o \log N_t \rightarrow \bar{L} \quad (101)$$

$$\bar{B}_l^o \log N_t \rightarrow \frac{1}{\mu} \bar{L} \quad (102)$$

(Here we must show that $d(N_t)$ in (27) tends to zero uniformly over all \bar{T} and \bar{B} .) The optimized lower bound satisfies

$$\begin{aligned} \mathcal{C}_l^o - \log(\rho N_t) + 2 \log \log N_t \\ \rightarrow \log(\bar{L}^2 \log 2) - 2 - \log[\mu(1 + \rho^{-1})] - \log(1 + \rho). \end{aligned} \quad (103)$$

Since the optimized bounds grow with N_t at the same rate, the capacity must also grow at that rate. Hence we conclude that the parameters that maximize the capacity exhibit the asymptotic behavior stated in the theorem.

D. Proof of Theorem 3

Similar to the proof of Theorem 2 in Appendix C, we first optimize the upper bound given by

$$\mathcal{C}_u = \frac{\bar{D}}{\bar{L}} \log \left(\frac{\rho}{1 + \rho^{-1}} \bar{T} \gamma_{\text{rvq}} N_t \right) + \frac{\bar{D}}{\bar{L}} \log(1 + s(N_t)) \quad (104)$$

where

$$s(N_t) = \frac{(1 + \rho^{-1})^2 + (1 + \rho^{-1}) \kappa(N_t) - \bar{T}}{\bar{T} \gamma_{\text{rvq}} N_t}, \quad (105)$$

and we have substituted $\sigma_w^2 = 1 - \bar{T}/(1 + \rho^{-1})$, corresponding to $\bar{T} \leq 1$, since we will show that the optimal normalized training length $\bar{T}_u^o \rightarrow 0$ as $N_t \rightarrow \infty$.

The Lagrangian for this optimization problem is given by

$$\mathcal{L} = \mathcal{C}_u + \lambda(\bar{L} - \bar{T} - \mu\bar{B} - \bar{D}) \quad (106)$$

where λ is the Lagrange multiplier. The first-order necessary conditions are

$$\log\left(\frac{\rho}{1 + \rho^{-1}}\right) + \log(\bar{T}) + \log(\gamma_{\text{rvq}}) + \log N_t + \log(1 + s(N_t)) - \bar{L}\lambda = 0 \quad (107)$$

$$\frac{\bar{D}}{\bar{T}} + \left(\frac{\bar{D}}{1 + s(N_t)}\right) \frac{\partial s(N_t)}{\partial \bar{T}} - \bar{L}\lambda = 0 \quad (108)$$

$$\left(\frac{\bar{D}}{\gamma_{\text{rvq}}}\right) \frac{\partial \gamma_{\text{rvq}}}{\partial \bar{B}} + \left(\frac{\bar{D}}{1 + s(N_t)}\right) \frac{\partial s(N_t)}{\partial \bar{B}} - \bar{L}\mu\lambda = 0 \quad (109)$$

$$\bar{L} - \bar{T} - \mu\bar{B} - \bar{D} = 0. \quad (110)$$

Substituting (108) and (110) into (107) gives

$$\begin{aligned} \bar{T} \log N_t + \bar{T} \log\left(\frac{\rho}{1 + \rho^{-1}}\right) + \bar{T} \log(\gamma_{\text{rvq}}) + \bar{T} \log(\bar{T}) + \bar{T} \log(1 + s(N_t)) \\ = (\bar{L} - \bar{T} - \mu\bar{B}) \left(1 + \left(\frac{\bar{T}}{1 + s(N_t)}\right) \frac{\partial s(N_t)}{\partial \bar{T}}\right). \end{aligned} \quad (111)$$

Using an argument analogous to that used to show that $(1 - 2^{-\bar{B}_u^o})\bar{T}_u^o N_t \rightarrow \infty$ as $N_t \rightarrow \infty$ in Appendix C, we can show that as $N_t \rightarrow \infty$, $\bar{T}\gamma_{\text{rvq}}N_t \rightarrow \infty$, which implies that $s(N_t) \rightarrow 0$ uniformly in \bar{T} and \bar{B} , so that $\left(\frac{\bar{T}}{1 + s(N_t)}\right) \frac{\partial s(N_t)}{\partial \bar{T}} \rightarrow 0$. Taking $N_t \rightarrow \infty$ therefore gives

$$\bar{T}_u^o \log N_t - \bar{L} \rightarrow 0, \quad (112)$$

assuming that $\bar{B}_u^o \rightarrow 0$, which will be proved next.

Substituting (108) into (109) to eliminate λ and rearranging gives

$$\gamma_{\text{rvq}} \left(\frac{\partial \gamma_{\text{rvq}}}{\partial \bar{B}}\right)^{-1} = \frac{\bar{T}}{\mu} \left[1 + \frac{\bar{T}}{1 + s(N_t)} \left(\frac{\partial s(N_t)}{\partial \bar{T}} - \frac{1}{\mu} \frac{\partial s(N_t)}{\partial \bar{B}}\right)\right]^{-1}. \quad (113)$$

Since $\bar{T} \rightarrow 0$ and $s(N_t) \rightarrow 0$,

$$\gamma_{\text{rvq}} \left(\frac{\partial \gamma_{\text{rvq}}}{\partial \bar{B}}\right)^{-1} \rightarrow 0. \quad (114)$$

For $0 \leq \bar{B} \leq \bar{B}^*$ it is shown in [14] (Theorem 3) that γ_{rvq} satisfies (after some rearrangement)

$$\left(-\frac{\gamma_{\text{rvq}}}{\bar{N}_r}\right) e^{-\gamma_{\text{rvq}}/\bar{N}_r} = -\frac{1}{e} 2^{-\bar{B}/\bar{N}_r} \quad (115)$$

where \bar{B}^* is given by

$$\bar{B}^* = \frac{1}{\log 2} \left(\bar{N}_r \log(\sqrt{\bar{N}_r}) - \bar{N}_r \log(1 + \sqrt{\bar{N}_r}) + \sqrt{\bar{N}_r} \right). \quad (116)$$

We can therefore write $-\gamma_{\text{rvq}}/\bar{N}_r = W(-\frac{1}{e}2^{-\bar{B}/\bar{N}_r})$, where $W(x)$ is the Lambert- W function. It is straightforward to show that

$$\gamma_{\text{rvq}} \left(\frac{\partial \gamma_{\text{rvq}}}{\partial \bar{B}} \right)^{-1} = \left(\frac{\partial [\log \gamma_{\text{rvq}}]}{\partial \bar{B}} \right)^{-1} = \frac{\gamma_{\text{rvq}} - \bar{N}_r}{\log 2}. \quad (117)$$

Hence from (114), $\gamma_{\text{rvq}}/\bar{N}_r \rightarrow 1$ as $N_t \rightarrow \infty$, and substituting in (115) implies that $\bar{B} \rightarrow 0$.

To determine the first-order rate at which $\bar{B} \rightarrow 0$, we combine (113) and (117) to write

$$\frac{\gamma_{\text{rvq}}}{\bar{N}_r} - 1 = \frac{\log 2}{\mu \bar{N}_r} \bar{T} + O(\bar{T}^2) \quad (118)$$

The behavior of γ_{rvq} for small \bar{B} (equivalently, $\gamma_{\text{rvq}}/\bar{N}_r$ close to one) can be determined by expanding $W(x)$ around $x = -e^{-1}$. Such an expansion is given in [36], which we rewrite as

$$\gamma_{\text{rvq}} = \bar{N}_r \left(1 + \sqrt{\zeta_{\bar{B}}} + \frac{1}{3}\zeta_{\bar{B}} + \frac{11}{72}\zeta_{\bar{B}}\sqrt{\zeta_{\bar{B}}} + O(\zeta_{\bar{B}}^{5/2}) \right) \quad (119)$$

where $\zeta_{\bar{B}} = 2(1 - 2^{-\bar{B}/\bar{N}_r}) = (2 \log 2)(\bar{B}/\bar{N}_r) + O(\bar{B}^2)$ for small \bar{B} . Hence we have

$$\frac{\gamma_{\text{rvq}}}{\bar{N}_r} - 1 = \sqrt{\zeta_{\bar{B}}} + O(\zeta_{\bar{B}}) = \sqrt{\frac{2 \log 2}{\bar{N}_r}} \sqrt{\bar{B}} + O(\bar{B}). \quad (120)$$

Combining this with (118) gives

$$\sqrt{\bar{B}} = \frac{1}{\mu} \sqrt{\frac{\log 2}{2\bar{N}_r}} \bar{T} + O(\bar{T}^2) \quad (121)$$

and substituting for \bar{T} from (112), we conclude that the feedback overhead that maximizes the upper bound on achievable rate satisfies

$$\bar{B}_u^o = \frac{\bar{L}^2 \log 2}{2\mu^2 \bar{N}_r} \frac{1}{\log^2 N_t} + O\left(\frac{1}{\log^4 N_t}\right) \quad (122)$$

Substituting for the optimized \bar{T}_u^o and \bar{B}_u^o in \mathcal{C}_u gives

$$\mathcal{C}_u^o - \log(\rho N_t) + \log \log N_t \longrightarrow \log \left(\frac{\rho \bar{L} \bar{N}_r}{e(\rho + 1)} \right). \quad (123)$$

We can apply the same techniques to the lower bound on achievable rate to determine the behavior of the optimal parameters. (Here we must show that $c(N_t)$ in (40) tends to zero uniformly over all \bar{T} and \bar{B} .)

The training and feedback overhead that maximize the lower bound on achievable rate satisfy

$$\bar{T}_l^o \log N_t \longrightarrow \bar{L} \quad (124)$$

$$\bar{B}_l^o \log^2 N_t \longrightarrow \frac{\bar{L}^2 \log 2}{2\mu^2 \bar{N}_r} \quad (125)$$

and substituting into the expression for C_l^o gives

$$C_l^o - \log(\rho N_t) + \log \log N_t \longrightarrow \log \left(\frac{\rho \bar{L} \bar{N}_r}{e(\rho + 1)} \right) - \log(1 + \rho). \quad (126)$$

Since the lower and upper bounds grow at the same rate, this establishes the theorem.

REFERENCES

- [1] İ. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. on Telecommun.*, vol. 10, pp. 585–595, Nov. 1999.
- [2] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Commun.*, vol. 6, no. 3, pp. 311–335, Mar. 1998.
- [3] E. Visotsky and U. Madhow, "Space-time transmit precoding with imperfect feedback," *IEEE Trans. Info. Theory*, vol. 47, no. 6, pp. 2632–2639, Sep. 2001.
- [4] S. Zhou and G. B. Giannakis, "Optimal transmitter eigen-beamforming and space-time block coding based on channel mean feedback," *IEEE Trans. Signal Processing*, vol. 50, no. 10, pp. 2599–2613, Oct. 2003.
- [5] S. H. Simon and A. L. Moustakas, "Optimizing MIMO antenna systems with channel covariance feedback," *IEEE J. Select. Areas Commun.*, vol. 21, no. 3, pp. 406–417, Apr. 2003.
- [6] T. Yoo and A. Goldsmith, "Capacity and power allocation for fading MIMO channels with channel estimation error," *IEEE Trans. Info. Theory*, vol. 52, no. 5, pp. 2203–2214, May 2006.
- [7] T. T. Kim and M. Skoglund, "On the expected rate of slowly fading channels with quantized side information," *IEEE Trans. Commun.*, vol. 55, no. 4, pp. 820–829, Apr. 2007.
- [8] R. S. Blum, "MIMO with limited feedback of channel side information," in *Proc. IEEE Int. Conf. Acoust., Speech and Sig. Proc.*, vol. 4, Apr. 2003, pp. 89–92.
- [9] V. K. N. Lau, Y. Liu, and T.-A. Chen, "On the design of MIMO block-fading channels with feedback-link capacity constraint," *IEEE Trans. Commun.*, vol. 52, no. 1, pp. 62–70, Jan. 2004.
- [10] 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 J. Select. Areas Commun.*, vol. 16, no. 8, pp. 1423–1436, Oct. 1998.
- [11] D. J. Love and R. W. Heath Jr., "Grassmannian beamforming for multiple-input multiple-output wireless systems," *IEEE Trans. Info. Theory*, vol. 49, no. 10, pp. 2735–2745, Oct. 2003.
- [12] K. K. Mukkavilli, A. Sabharwal, E. Erkip, and B. Aazhang, "On beamforming with finite rate feedback in multiple antenna systems," *IEEE Trans. Info. Theory*, vol. 49, no. 10, pp. 2562–2579, Oct. 2003.
- [13] J. C. Roh and B. D. Rao, "Transmit beamforming in multiple-antenna systems with finite rate feedback: A VQ-based approach," *IEEE Trans. Info. Theory*, vol. 52, no. 3, pp. 1101–1112, Mar. 2006.
- [14] W. Santipach and M. L. Honig, "Capacity of multiple-antenna fading channel with quantized precoding matrix," *IEEE Trans. Info. Theory*, vol. 55, no. 3, pp. 1218–1234, Mar. 2009.

- [15] W. Dai, Y. Liu, V. K. N. Lau, and B. Rider, "On the information rate of MIMO systems with finite rate channel state feedback and power on/off strategy," in *Proc. IEEE Int. Symp. on Inform. Theory (ISIT)*, Adelaide, Australia, Sep. 2005, pp. 1549–1553.
- [16] C. K. Au-Yeung and D. J. Love, "On the performance of random vector quantization limited feedback beamforming in a MISO system," *IEEE Trans. Wireless Commun.*, vol. 6, pp. 458–462, Feb. 2007.
- [17] A. D. Dabagh and D. J. Love, "Feedback rate-capacity loss tradeoff for limited feedback MIMO systems," *IEEE Trans. Info. Theory*, vol. 52, no. 5, pp. 2190–2202, May 2006.
- [18] D. J. Love, R. W. Heath, V. K. N. Lau, D. Gesbert, B. D. Rao, and M. Andrews, "An overview of limited feedback in wireless communications systems," *IEEE J. Select. Areas Commun.*, vol. 26, no. 8, pp. 1341–1365, Oct. 2008.
- [19] C. Steger and A. Sabharwal, "Single-input two-way SIMO channel: diversity-multiplexing tradeoff with two-way training," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 4877–4885, Jul. 2008.
- [20] Y. Xie, C. N. Georghiades, and K. Rohani, "Optimal bandwidth allocation for the data and feedback channels in MISO-FDD systems," *IEEE Trans. Commun.*, vol. 54, no. 2, pp. 197–203, Feb. 2006.
- [21] W. Santipach and M. L. Honig, "Signature optimization for CDMA with limited feedback," *IEEE Trans. Info. Theory*, vol. 51, no. 10, pp. 3475–3492, Oct. 2005.
- [22] N. Jindal, "Mimo broadcast channels with finite rate feedback," *IEEE Trans. Info. Theory*, vol. 52, no. 11, pp. 5045–5060, Nov. 2006.
- [23] D. J. Love, R. W. Heath Jr., W. Santipach, and M. L. Honig, "What is the value of limited feedback for MIMO channels?" *IEEE Commun. Mag.*, vol. 42, no. 10, pp. 54–59, Oct. 2004.
- [24] B. Hassibi and B. M. Hochwald, "How much training is needed in a multiple-antenna wireless link?" *IEEE Trans. Info. Theory*, vol. 49, no. 4, pp. 951–964, Apr. 2003.
- [25] Y. Sun and M. L. Honig, "Large system capacity of MIMO block fading channels with least squares linear adaptive receivers," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, St. Louis, MO, Nov. 2005.
- [26] A. M. Tulino and S. Verdú, "Random matrix theory and wireless communications," *Foundations and Trends in Communications and Information Theory*, vol. 1, no. 1, pp. 1–182, 2004.
- [27] D. J. Love and C. K. Au-Yeung, "On resource allocation in two-way limited feedback beamforming systems," in *Proc. Information Theory and Applications Workshop*, San Diego, CA, Jan. 2008, pp. 188–192.
- [28] —, "Optimization and tradeoff analysis of two-way limited feedback beamforming systems," *IEEE Trans. Wireless Commun.*, vol. 8, no. 5, pp. 2570–2579, May 2009.
- [29] M. Agarwal and M. Honig, "Wideband fading channel capacity with training and partial feedback," in *Proc. Allerton Conf. on Commun., Control, and Computing*, Monticello, IL, Sep. 2005.
- [30] J. Chen, R. A. Berry, and M. L. Honig, "Limited feedback schemes for downlink OFDMA based on sub-channel groups," *IEEE J. Select. Areas Commun.*, vol. 26, no. 8, pp. 1451–1461, Oct. 2008.
- [31] M. Rupf and J. L. Massey, "Optimum sequence multisets for synchronous code-division multiple-access channels," *IEEE Trans. Info. Theory*, vol. 40, no. 4, pp. 1261–1266, Jul. 1994.
- [32] M. Médard, "The effect upon channel capacity in wireless communication of perfect and imperfect knowledge of the channel," *IEEE Trans. Info. Theory*, vol. 46, no. 3, pp. 933–946, May 2000.
- [33] A. Ben-Tal and E. Hochman, "More bounds on the expectation of a convex function of a random variable," *Journal of Applied Probability*, vol. 9, pp. 803–812, 1972.

- [34] D. Kershaw, “Some extensions of W. Gautschi’s inequalities for the gamma function,” *Mathematics of Computation*, vol. 41, no. 164, pp. 607–611, 1983.
- [35] Wolfram Research Inc., “Gamma function: Inequalities,” [online], available: <http://functions.wolfram.com/06.05.29.0009.01>.
- [36] D. A. Barry and P. J. Culligan-Hensley, “Real values of the W -function,” *ACM Transactions on Mathematical Software*, vol. 21, no. 2, pp. 161–171, Jun. 1995.

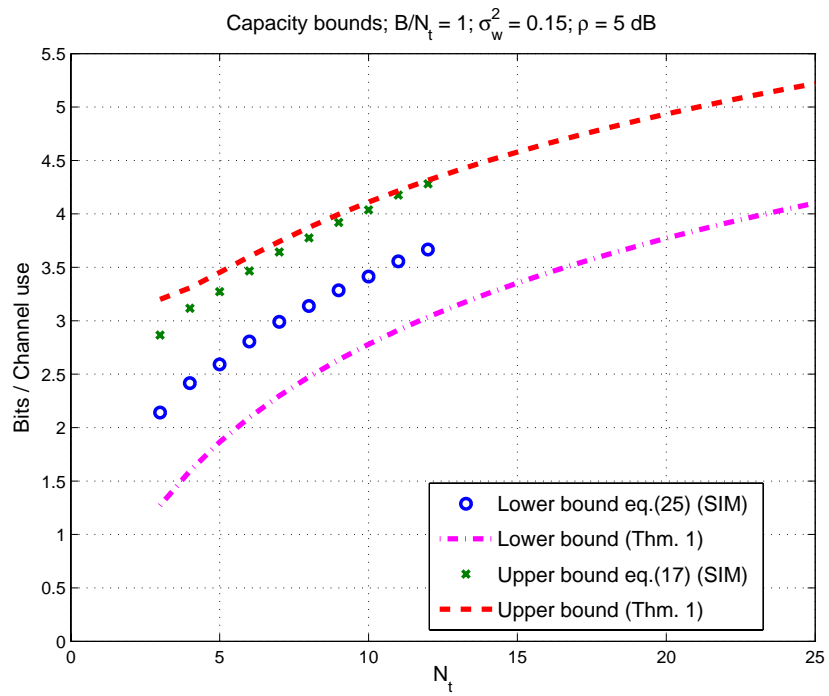


Fig. 1. The capacity bounds in Theorem 1 (bits/channel use) versus number of transmit antennas.

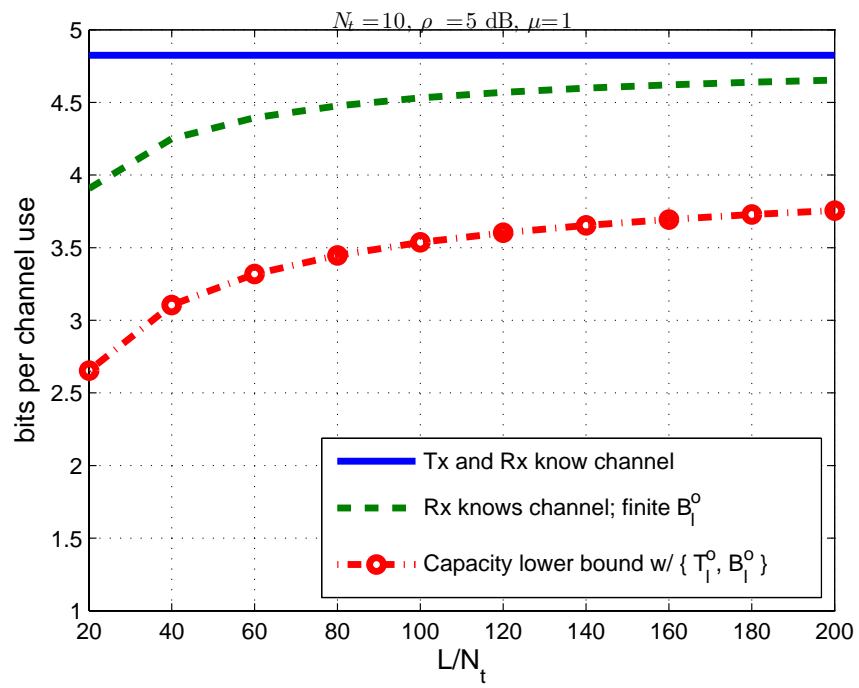


Fig. 2. Achievable rate versus normalized packet length \bar{L} .

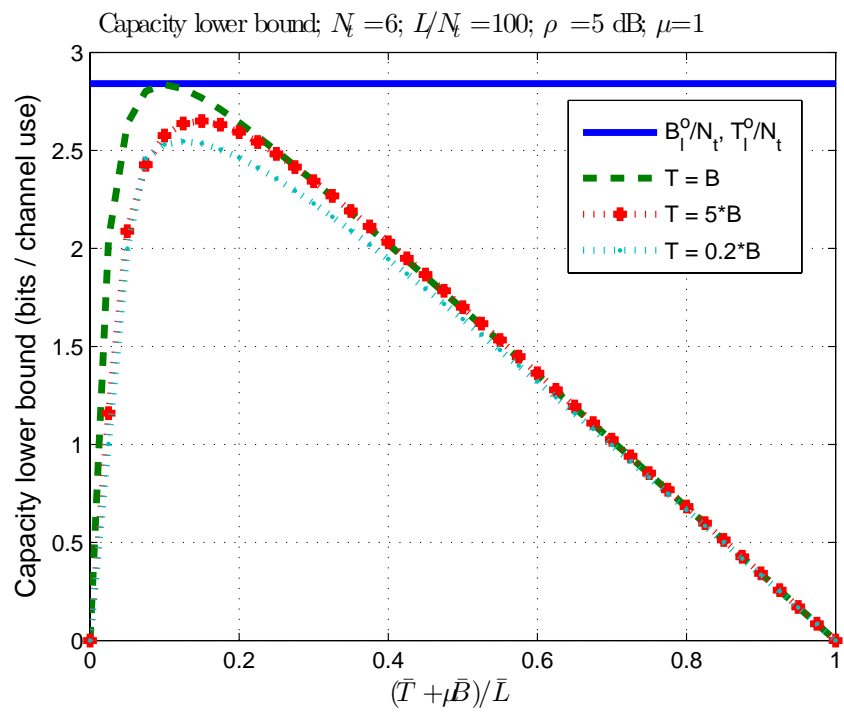


Fig. 3. Lower bound on capacity versus normalized training and feedback $(\bar{T} + \mu\bar{B})/\bar{L}$ with different allocations $\bar{T}/(\mu\bar{B})$.

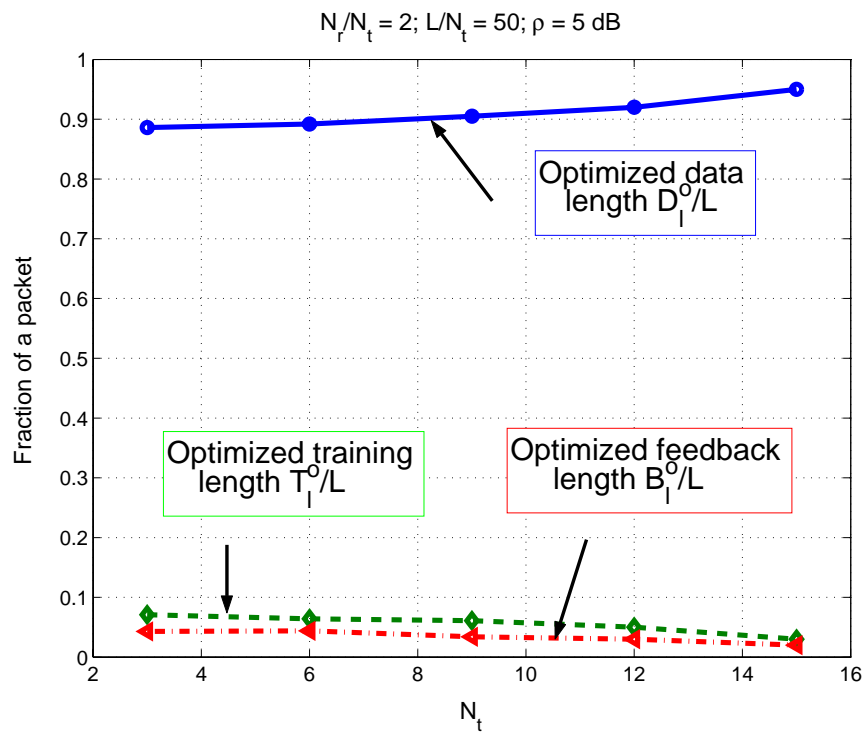


Fig. 4. Optimized training and feedback overhead, and fraction of data symbols $\{\bar{T}_1^o/\bar{L}, \bar{B}_1^o/\bar{L}, \bar{D}_1^o/\bar{L}\}$ versus number of transmit antennas N_t .

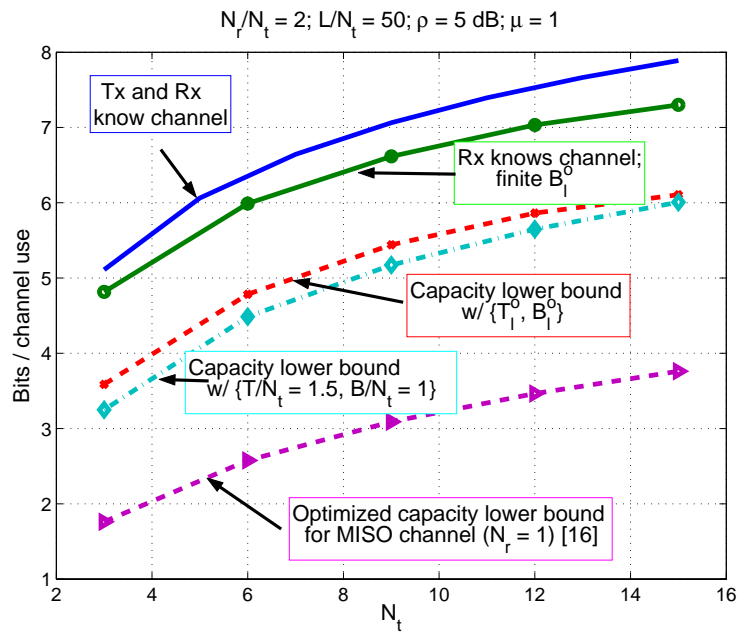


Fig. 5. Achievable rate for MIMO channel versus number of transmit antennas N_t with different assumptions about channel knowledge at the receiver and transmitter. Also shown is the optimized capacity lower bound for the corresponding MISO channel

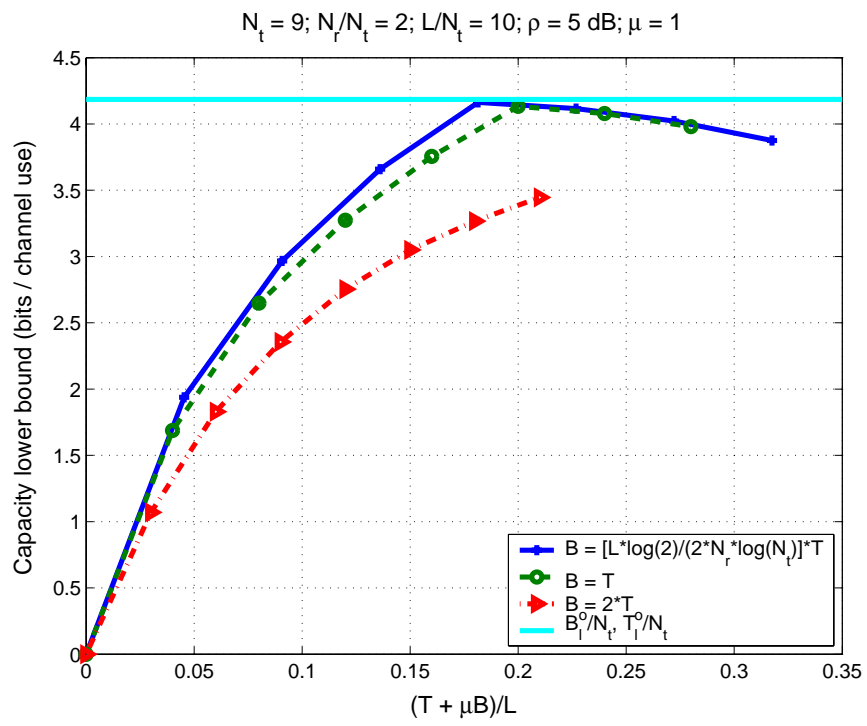


Fig. 6. Lower bound on beamforming capacity for MIMO channel versus normalized training and feedback $(\bar{T} + \mu\bar{B})/\bar{L}$.