Capacity of a Multiple-Antenna Fading Channel With a Quantized Precoding Matrix

Wiroonsak Santipach, *Member, IEEE,* and Michael L. Honig, *Fellow, IEEE*

Abstract—Given a multiple-input multiple-output (MIMO) channel, feedback from the receiver can be used to specify a transmit precoding matrix, which selectively activates the strongest channel modes. Here we analyze the performance of *Random Vector Quantization (RVQ)*, in which the precoding matrix is selected from a random codebook containing independent, isotropically distributed entries. We assume that channel elements are *i.i.d.* and known to the receiver, which relays the optimal (rate-maximizing) precoder codebook index to the transmitter using B bits. We first derive the large system capacity of beamforming (rank-one precoding matrix) as a function of B, where large system refers to the limit as B and the number of transmit and receive antennas all go to infinity with fixed ratios. RVQ for beamforming is asymptotically optimal, i.e., no other quantization scheme can achieve a larger asymptotic rate. We subsequently consider a precoding matrix with arbitrary rank, and approximate the asymptotic RVQ performance with optimal and linear receivers (matched filter and Minimum Mean Squared Error (MMSE)). Numerical examples show that these approximations accurately predict the performance of finite-size systems of interest. Given a target spectral efficiency, numerical examples show that the amount of feedback required by the linear MMSE receiver is only slightly more than that required by the optimal receiver, whereas the matched filter can require significantly more feedback.

Index Terms—Beamforming, large system analysis, limited feedback, Multi-Input Multi-Output (MIMO), precoding, vector quantization.

I. INTRODUCTION

GIVEN a multi-input multi-output (MIMO) channel, providing channel information at the transmitter can increase the achievable rate and simplify the coder and decoder. Namely, this channel information can specify a precoding matrix, which aligns the transmitted signal along the strongest channel modes (i.e., singular vectors corresponding to the

largest singular values). In practice, the precoding matrix must be quantized at the receiver, and relayed to the transmitter via a feedback channel. The corresponding achievable rate is therefore limited by the accuracy of the quantizer.

The design and performance of quantized precoding matrices for multi-input single-output (MISO) and MIMO channels has been considered in numerous references, including [1]–[12]. In those references, and in this paper, the channel is assumed to be stationary, known at the receiver, and the performance is evaluated as a function of the number of quantization bits B. (This is in contrast with other work, which models estimation error at the receiver, but does not explicitly account for quantization error (e.g., [13], [14]), and which assumes a time-varying channel with feedback of second-order statistics [15]–[18].) Optimization of vector quantization codebooks is discussed in [1]–[3], [5], [6] for beamforming, and in [4], [7], [12] for MIMO channels with precoding matrices that provide multiplexing gain (i.e., have rank larger than one). It is shown in [2], [3] that this optimization can be interpreted as maximizing the minimum distance between points in a Grassmannian space. (See also [9].) The performance of this class of Grassmannian codebooks is also studied in [8]–[10].

In this paper, we evaluate the performance of a *Random Vector Quantization (RVQ)* scheme for the precoding matrix. Namely, given B feedback bits, the precoding matrix is selected from a random codebook containing 2^B matrices, which are independent and isotropically distributed. RVQ has been analyzed in other source coding contexts (e.g., see [19] and the related discussion in [20]), and achieves the rate-distortion bound for ergodic Gaussian sources. This work is motivated by prior work [21] in which RVQ is considered for signature quantization in Code-Division Multiple Access (CDMA). In that scenario, limited feedback is used to select a signature for a particular user, which maximizes the received Signal-to-Interference-Plus-Noise-Ratio (SINR). RVQ has the attractive properties of being tractable and asymptotically optimal. Namely, in [21] the received SINR with RVQ is evaluated in the asymptotic (large system) limit as processing gain, number of users, and feedback bits all tend to infinity with fixed ratios. Furthermore, it is shown that no other quantization scheme can achieve a larger asymptotic SINR.

Here we assume an *i.i.d.* block Rayleigh fading channel model with independent channel gains, and take ergodic capacity as the performance criterion. The receiver relays B bits to the transmitter (per codeword) via a reliable feedback channel (i.e., no feedback errors) with no delay. We start by evaluating the capacity of MISO and MIMO channels with a

quantized *beamformer*, i.e., rank-one precoding matrix. Our results are asymptotic as the number of transmit antennas N_t and feedback bits B both tend to infinity with fixed B/N_t (feedback bits per degree of freedom). For the MIMO channel the number of receive antennas N_r also tends to infinity in proportion with N_t and B. The asymptotic expressions accurately predict the performance of finite-size systems of interest as a function of normalized feedback and background Signal-to-Noise Ratio (SNR). In analogy with the optimality result shown in [21], RVQ is also asymptotically optimal in this scenario, i.e., no other quantization scheme can achieve a larger asymptotic rate. Furthermore, numerical examples for small N_t show that RVQ performance averaged over codebooks is essentially the same as that obtained from codebooks optimized via the Lloyd-Max algorithm [1], [5], [6]. (See also the numerical examples in [22], which compare RVQ performance with the optimized (Grassmannian) codebooks in [3].)

We then consider quantization of a precoding matrix with arbitrary rank. Namely, a rank K precoding matrix multiplexes K independent streams of transmitted information symbols onto the N_t transmit antennas. In that case, the capacity with limited feedback is approximated in the limit as B, N_t, N_r, and K all tend to infinity with fixed ratios K/N_t, N_r/N_t, and B/N_r^2. That is, the number of feedback bits again scales linearly with the number of degrees of freedom, which is proportional to N_r^2. Although our results for beamforming suggest that RVQ is also asymptotically optimal in this scenario, this remains an open question.

The asymptotic results for a precoder matrix with arbitrary rank K can be used to determine the normalized rank, or multiplexing gain K/N_t, which maximizes the capacity. This optimized rank in general depends on the normalized feedback, the ratio of antennas N_t/N_r, and the SNR. For example, if $N_t/N_r \geq 1$ and the SNR is sufficiently large, then as the feedback increases from zero to infinity, the optimized rank decreases from one to N_r/N_t. Numerical results are presented, which illustrate the effect of normalized rank on achievable rate, and also show that the asymptotic results accurately predict simulated results for finite-size systems of interest.

We also evaluate the performance of RVQ with linear receivers (i.e., the matched filter and linear Minimum Mean Squared Error (MMSE) receivers), and compare their performance with the optimal (capacity-achieving) receiver. With the optimal precoding matrix, corresponding to infinite feedback, both linear receivers are optimal. With limited feedback the two linear receivers are simpler than the optimal receiver, but require more feedback to achieve a target rate. Numerical results show that this additional feedback required by the linear MMSE receiver is quite small, whereas the additional feedback required by the matched filter can be significant (e.g., about one bit per precoding matrix element).

In addition to quantizing the optimal precoding matrix, power for each data stream can also be optimized, quantized, and fed back to the transmitter (e.g., see [23], [24]). Asymptotically, the amount of feedback required to specify the power is negligible compared to the feedback required for the precoding matrix. Furthermore, uniform power over the set of activated

channel typically performs close to the optimal (water-filling) performance [8]. We therefore only consider quantization of the precoding matrix.

Other related work on RVQ for MIMO channels has been presented in [25]–[27]. Namely, exact expressions for the ergodic capacity with beamforming and RVQ for a finite-size MISO channel are derived in [25]. The performance of RVQ for precoding over a broadcast MIMO channel is analyzed in [26], [27]. A closely related random beamforming scheme for the multiuser MIMO broadcast channel was previously presented in [28]. In that work, the growth in sum capacity is characterized asymptotically as the number of users becomes large with a *fixed* number of antennas. (Random beamforming was previously proposed in [29], although there the main focus is to improve fairness among users.)

The paper is organized as follows. Section II describes the channel model, Section III considers the capacity of beamforming with limited feedback, and sections IV and V examine the capacity of a quantized precoding matrix with optimal and linear receivers, respectively. Derivations of the main results are given in the appendices.

II. Channel Model

We consider a point-to-point, flat Rayleigh fading channel with N_t transmit antennas and N_r receive antennas. Let $\boldsymbol{x} = [x_k]$ be a $K \times 1$ vector of transmitted symbols with covariance matrix $\boldsymbol{I}_K$, where $\boldsymbol{I}_K$ is the $K \times K$ identity matrix, and K is the number of independent data streams. The received $N_r \times 1$ vector is given by

$$\boldsymbol{y} = \frac{1}{\sqrt{K}}\boldsymbol{H}\boldsymbol{V}\boldsymbol{x} + \boldsymbol{n} \qquad (1)$$

where $\boldsymbol{H} = [h_{n_r,n_t}]$ is an $N_r \times N_t$ channel matrix, $\boldsymbol{V} = [\boldsymbol{v}_1 \; \boldsymbol{v}_2 \; \ldots \; \boldsymbol{v}_K]$ is an $N_t \times K$ precoding matrix, and $\boldsymbol{n}$ is a complex Gaussian noise $N_r \times 1$ vector with covariance matrix $\sigma_n^2 \boldsymbol{I}_{N_r}$. Assuming rich scattering and Rayleigh fading, the elements of $\boldsymbol{H}$ are independent, and the channel coefficient between the n_tth transmit antenna and the n_rth receive antenna, h_{n_r,n_t}, is a circularly symmetric complex Gaussian random variable with zero mean and unit variance ($E[|h_{n_r,n_t}|^2] = 1$).

We assume *i.i.d.* block fading, i.e., the channel is static within a fading block, and the channels across blocks are independent. The ergodic capacity is achieved by coding the transmitted symbols across an infinitely large number of fading blocks. With perfect channel knowledge at the receiver and a given precoding matrix $\boldsymbol{V}$, the ergodic capacity is the mutual information between $\boldsymbol{x}$ and $\boldsymbol{y}$ with a complex Gaussian distributed input, averaged over the channel, given by

$$I(\boldsymbol{x}; \boldsymbol{y}) = E_{\boldsymbol{H}}\left[\log\det\left(\boldsymbol{I} + \frac{\rho}{K}\boldsymbol{H}\boldsymbol{V}\boldsymbol{V}^\dagger\boldsymbol{H}^\dagger\right)\right] \qquad (2)$$

where $\rho = 1/\sigma_n^2$ is the background SNR. We wish to specify the precoding matrix $\boldsymbol{V}$ that maximizes the mutual information, subject to a power constraint $\|\boldsymbol{v}_k\| \leq 1$, for $1 \leq k \leq K$.

With unlimited feedback, the columns of the optimal precoding matrix, which maximizes (2), are eigenvectors of the channel covariance matrix $\boldsymbol{H}^\dagger\boldsymbol{H}$. With B feedback bits per

fading block, we can specify the precoding matrix from a quantization set or codebook $\mathcal{V} = \{\boldsymbol{V}_1, \cdots, \boldsymbol{V}_{2^B}\}$ known *a priori* to both the transmitter and receiver. The receiver chooses the $\boldsymbol{V}_j$ that maximizes the sum mutual information, and relays the corresponding index back to the transmitter. Of course, the performance (ergodic capacity) depends on the codebook $\mathcal{V}$.

III. BEAMFORMING WITH LIMITED FEEDBACK

We start with a rank-one precoding matrix, corresponding to a single data stream ($K = 1$). In that case, the precoding matrix is specified by an $N_t \times 1$ beamforming vector $\boldsymbol{v}$, which ideally corresponds to the strongest channel mode. That is, the optimal $\boldsymbol{v}$, which maximizes the ergodic capacity in (2), is the eigenvector of $\boldsymbol{H}^\dagger \boldsymbol{H}$ corresponding to the largest eigenvalue. This vector is computed at the receiver and a quantized version is relayed back to the transmitter.

Let $\mathcal{V} = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{2^B}\}$ denote the quantization codebook for $\boldsymbol{v}$, given B feedback bits. Optimization of this codebook has been considered in [2], [3], [6] with outage capacity and ergodic capacity as performance metrics. The performance of an optimized codebook is difficult to evaluate exactly, and is approximated in [2], [3], [6], [9]–[11]. Here we consider RVQ in which $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_{2^B}$ are independent, isotropically distributed random vectors, each with unit norm. This is motivated by the observation that given a channel matrix $\boldsymbol{H}$ with *i.i.d.* elements, the eigenvectors of $\boldsymbol{H}^\dagger \boldsymbol{H}$ are isotropically distributed [30], hence the codebook entries should be uniformly distributed over the space of beamforming vectors.

A. MISO Channel

We first consider a MISO channel, corresponding to a single receive antenna ($N_r = 1$). In that case, $\boldsymbol{H}$ is an $N_t \times 1$ channel vector, which we denote as $\boldsymbol{h}$. The optimal beamformer, which maximizes the mutual information in (2), is the normalized channel vector $\boldsymbol{h}/\|\boldsymbol{h}\|$ and the corresponding mutual information is $E_{\boldsymbol{h}}[\log(1 + \rho \boldsymbol{h}^\dagger \boldsymbol{h})]$. The receiver selects the quantized precoding vector to maximize the mutual information, i.e.,

$$\hat{\boldsymbol{v}} = \arg \max_{1 \leq j \leq 2^B} \left\{ I_j = \log(1 + \rho |\boldsymbol{h}^\dagger \boldsymbol{v}_j|^2) \right\} \quad (3)$$

and the corresponding achievable rate is

$$I_{\mathsf{rvq}}^{N_t} \triangleq \max_{1 \leq j \leq 2^B} I_j. \quad (4)$$

where the superscript N_t denotes the system size. The achievable rate depends on the codebook $\mathcal{V}$ and the channel vector $\boldsymbol{h}$, and is therefore random. Rather than averaging $I_{\mathsf{rvq}}^{N_t}$ over $\mathcal{V}$ and $\boldsymbol{h}$ to find the ergodic capacity, we instead evaluate the limiting performance as N_t and B tend to infinity with fixed $\bar{B} = B/N_t$ (feedback bits per transmit antenna). In this limit, $I_{\mathsf{rvq}}^{N_t}$ converges to a deterministic constant. This is illustrated in Fig. 1, which shows the pdf of $|\boldsymbol{h}^\dagger \hat{\boldsymbol{v}}|^2/\|\boldsymbol{h}\|^2$ for different N_t with no feedback ($\bar{B} = 0$), and for RVQ with $\bar{B} = 2$. The figure shows that convergence of the pdf to a point mass is faster with feedback than without.

As $N_t \to \infty$, $(\boldsymbol{h}^\dagger \boldsymbol{h})/N_t \to 1$ almost surely, so that $\log(1 + \rho \boldsymbol{h}^\dagger \boldsymbol{h}) - \log(\rho N_t) \to 0$. That is, with perfect channel
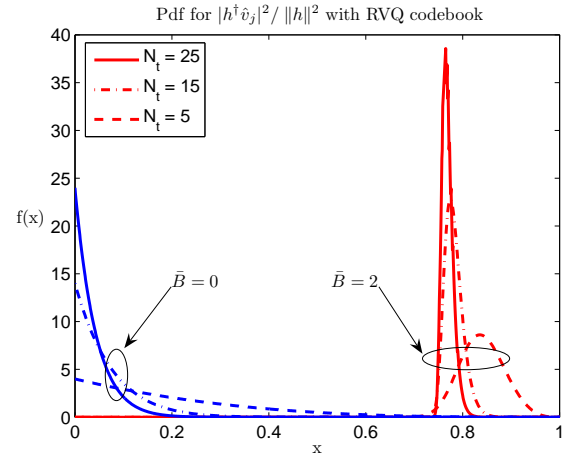

Fig. 1. pdf of $|\boldsymbol{h}^\dagger \hat{\boldsymbol{v}}|^2/\|\boldsymbol{h}\|^2$ with RVQ for different values of N_t.

knowledge at the transmitter, the ergodic capacity increases as $\log(\rho N_t)$. With finite feedback there is a rate loss, which is defined as

$$I_{\mathsf{rvq}}^\triangle = I_{\mathsf{rvq}}^{N_t} - \log(\rho N_t). \quad (5)$$

For finite N_t, $I_{\mathsf{rvq}}^\triangle$ is random; however, in the large system limit $I_{\mathsf{rvq}}^\triangle$ converges to a deterministic constant.

Theorem 1. *As $(N_t, B) \to \infty$ with fixed $\bar{B} = B/N_t$, the rate difference $I_{\mathsf{rvq}}^\triangle$ converges in the mean square sense to*

$$\mathcal{I}_{\mathsf{rvq}}^\triangle = \log(1 - 2^{-\bar{B}}). \quad (6)$$

The proof is given in Appendix A. For $\bar{B} > 0$, the rate loss due to finite feedback is a constant. As $\bar{B} \to 0$, this rate loss tends to infinity, since with $\bar{B} = 0$, the capacity tends to a constant as $N_t \to \infty$, whereas the capacity grows as $\log N_t$ for $\bar{B} > 0$. Of course, as $\bar{B} \to \infty$ (unlimited feedback), the rate loss vanishes.

RVQ is asymptotically optimal in the following sense. Suppose that $\{\mathcal{V}_{N_t}\}$ is an arbitrary sequence of codebooks for the beamforming vector where

$$\mathcal{V}_{N_t} = \left\{ \boldsymbol{v}_1^{N_t}, \boldsymbol{v}_2^{N_t}, \ldots, \boldsymbol{v}_{2^B}^{N_t} \right\} \quad (7)$$

is the codebook for a particular N_t and $\|\boldsymbol{v}_j^{N_t}\|^2 = 1$ for each j. The associated rate is given by

$$I_{\mathcal{V}_{N_t}} = \max_{1 \leq j \leq 2^B} \log(1 + \rho |\boldsymbol{h}^\dagger \boldsymbol{v}_j^{N_t}|^2) \quad (8)$$

and the rate difference $I_{\mathcal{V}_{N_t}}^\triangle = I_{\mathcal{V}_{N_t}} - \log(\rho N_t)$.

Theorem 2. *For any sequence of codebooks $\{\mathcal{V}_{N_t}\}$,*

$$\limsup_{(N_t, B) \to \infty} E_{\boldsymbol{h}}[I_{\mathcal{V}_{N_t}}^\triangle] \leq \mathcal{I}_{\mathsf{rvq}}^\triangle. \quad (9)$$

The proof is given in Appendix B.

Although the optimality of RVQ holds only in the large system limit, numerical results in Section III-C show that for finite-size systems of interest RVQ performs essentially the same as optimized quantization codebooks.

B. Multi-Input Multi-Output (MIMO) Channel

We now consider quantized beamforming for a MIMO channel, i.e., with multiple transmit *and* receive antennas. Taking the rank $K = 1$ maximizes the diversity gain [31], but the corresponding capacity grows only as $\log N_t$ instead of linearly with N_t, which is the case when K grows proportionally with N_t. (This is true with both unlimited and limited feedback, assuming a fixed number of feedback bits per precoder element.) Also, a beamformer is significantly less complex than a matrix precoder with $K > 1$, and requires less feedback to specify.

We again consider an RVQ codebook $\mathcal{V}$ with 2^B independent unit-norm vectors, where each vector is uniformly distributed over the N_t-dimensional unit sphere. The achievable rate is $E_{\boldsymbol{H}}[I_{\mathsf{rvq}}^{N_t}]$, where

$$
\begin{aligned}
I_{\mathsf{rvq}}^{N_t} &= E_{\mathcal{V}}\left[\max_{1 \le j \le 2^B} \log\left(1 + \rho\|\boldsymbol{H}\boldsymbol{v}_j\|^2\right)\Big| \boldsymbol{H}\right] \quad (10)\\
&= E_{\mathcal{V}}\left[\log(1 + \rho \max_{1 \le j \le 2^B}\|\boldsymbol{H}\boldsymbol{v}_j\|^2)\Big| \boldsymbol{H}\right]. \quad (11)
\end{aligned}
$$

As for the MISO channel, with unlimited feedback the achievable rate increases as $\log(\rho N_t)$. We again define the rate difference due to quantization as

$$
I_{\mathsf{rvq}}^{\triangle} \triangleq I_{\mathsf{rvq}}^{N_t} - \log(\rho N_t) = E_{\mathcal{V}}\left[\log\left(\frac{1}{\rho N_t} + \max_j \gamma_j\right)\Big| \boldsymbol{H}\right] \quad (12)
$$

where

$$
\gamma_j = \frac{1}{N_t}\boldsymbol{v}_j^{\dagger}\boldsymbol{H}^{\dagger}\boldsymbol{H}\boldsymbol{v}_j. \quad (13)
$$

Evaluating the expectation in (12) is difficult for finite N_t, N_r, and B, so that we again resort to a large system analysis. Namely, we let N_t, N_r, and B each tend to infinity with fixed $\bar{B} = B/N_t$ and $\bar{N}_r = N_r/N_t$. For each N_t and N_r the channel matrix $\boldsymbol{H}$ is chosen as the $N_r \times N_t$ upper-left corner of a matrix $\bar{\boldsymbol{H}}$ with an infinite number of rows and columns, and with *i.i.d.* complex Gaussian entries.

The received power in this large system limit is given by

$$
\gamma_{\mathsf{rvq}}^{\infty} = \lim_{(N_t,N_r,B)\to\infty}\left[\max_{1 \le j \le 2^B}\gamma_j\Big| \bar{\boldsymbol{H}}\right] \quad (14)
$$

where convergence to the deterministic limit can be shown in the mean square sense. Conditioned on $\bar{\boldsymbol{H}}$, the γ_j's are *i.i.d.* since the beamforming vectors $\boldsymbol{v}_j$ are *i.i.d.*, and applying [32, Theorem 2.1.2], it can be shown that

$$
\gamma_{\mathsf{rvq}}^{\infty} = \lim_{(N_t,N_r,B)\to\infty} F_{\gamma|\bar{\boldsymbol{H}}}^{-1}\left(1 - 2^{-B}\right) \quad (15)
$$

where $F_{\gamma|\bar{\boldsymbol{H}}}(\cdot)$ is the cdf of γ_j given $\bar{\boldsymbol{H}}$. Analogous results for the interference power in CDMA with quantized signatures have been presented in [21], so that we omit the proofs of (14) and (15). Note that $\bar{N}_r \le \gamma_{\mathsf{rvq}}^{\infty} \le (1 + \sqrt{\bar{N}_r})^2$, where the lower and upper bounds correspond to $\bar{B} = 0$ and $\bar{B} = \infty$, respectively. That is, $(1 + \sqrt{\bar{N}_r})^2$ is the asymptotic maximum eigenvalue of the channel covariance matrix $\frac{1}{N_t}\boldsymbol{H}^{\dagger}\boldsymbol{H}$ [33]. The asymptotic rate difference is given by

$$
\mathcal{I}_{\mathsf{rvq}}^{\Delta} = \lim_{(N_t,N_r,B)\to\infty} I_{\mathsf{rvq}}^{\Delta} = \log(\gamma_{\mathsf{rvq}}^{\infty}) \quad (16)
$$

The limit in (15) can be explicitly evaluated, and is independent of the channel realization $\bar{\boldsymbol{H}}$.

Theorem 3. *For $0 \le \bar{B} \le \bar{B}^*$, $\gamma_{\mathsf{rvq}}^{\infty}$ satisfies*

$$
\left(\gamma_{\mathsf{rvq}}^{\infty}\right)^{\bar{N}_r} \mathrm{e}^{-\gamma_{\mathsf{rvq}}^{\infty}} = 2^{-\bar{B}}\left(\frac{\bar{N}_r}{\mathrm{e}}\right)^{\bar{N}_r} \quad (17)
$$

and for $\bar{B} \ge \bar{B}^*$,

$$
\begin{aligned}
\gamma_{\mathsf{rvq}}^{\infty} = (1 + \sqrt{\bar{N}_r})^2 - \exp\Big\{&\frac{1}{2}\bar{N}_r\log(\bar{N}_r)\\
&- (\bar{N}_r - 1)\log(1 + \sqrt{\bar{N}_r}) + \sqrt{\bar{N}_r} - \bar{B}\log(2)\Big\}
\end{aligned} \quad (18)
$$

where

$$
\bar{B}^* = \frac{1}{\log(2)}\left(\bar{N}_r\log\left(\frac{\sqrt{\bar{N}_r}}{1 + \sqrt{\bar{N}_r}}\right) + \sqrt{\bar{N}_r}\right). \quad (19)
$$

The proof is given in Appendix C and is motivated by an analogous result for CDMA, presented in [34]. As stated in Theorem 3, $\gamma_{\mathsf{rvq}}^{\infty}$ depends only on $\bar{B}$ and $\bar{N}_r$. Letting $\bar{N}_r \to 0$ gives the the asymptotic capacity of the MISO channel with RVQ. As for the MISO channel, RVQ is asymptotically optimal.

Theorem 4. *As $(N_t, N_r, B) \to \infty$ with fixed $\bar{N}_r = N_r/N_t$ and $\bar{B} = B/N_t$,*

$$
\limsup_{(N_t,N_r,B)\to\infty} I_{\mathcal{V}_{N_t}}^{N_t} - \log(\rho N_t) \le \mathcal{I}_{\mathsf{rvq}}^{\Delta} \quad (20)
$$

for any sequence of codebooks $\{\mathcal{V}_{N_t}\}$.

The proof is similar to the proof of Theorem 2 in [21] and is therefore omitted.

C. Numerical Results

Figs. 2 and 3 show $\mathcal{I}_{\mathsf{rvq}}^{\triangle}$ for MISO and MIMO channels, respectively, with beamforming and RVQ versus normalized feedback bits $(\bar{B})$ with $\rho = 5$ and 10 dB. Also shown for comparison are achievable rates with a quantization codebook optimized via the Lloyd-Max algorithm [1], [5], [6], and the capacity with perfect beamforming, corresponding to unlimited feedback. The results for RVQ are averaged over codebook realizations, and are essentially the same as those shown for the optimized Lloyd-Max codebooks. For the MISO channel, the asymptotic capacity (6) accurately predicts the simulated results shown even with a relatively small number of transmit antennas ($N_t = 3$ and 6). For the MIMO results $\bar{N}_r = 1.5$, and simulation results are shown for 4×6 and 16×24 channels. The asymptotic results accurately predict the performance for the larger channel, and are somewhat less accurate for the smaller channel.

Comparing finite feedback with perfect beamforming, the results show that one feedback bit per complex entry ($\bar{B} = 1$) provides more than 50% of the potential gain due to feedback. For both the MISO and MIMO examples shown, the perfect beamforming capacity is nearly achieved with two feedback bits per complex coefficient.
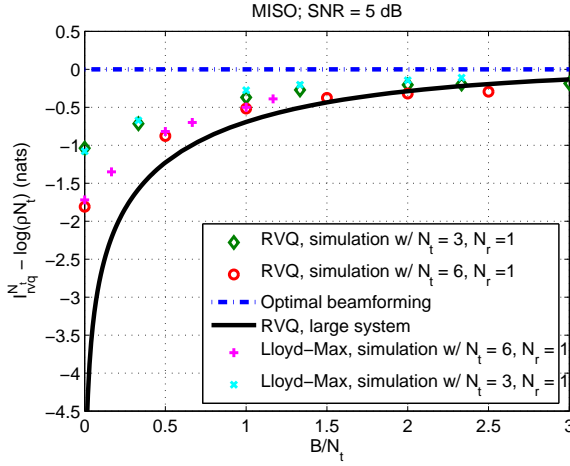
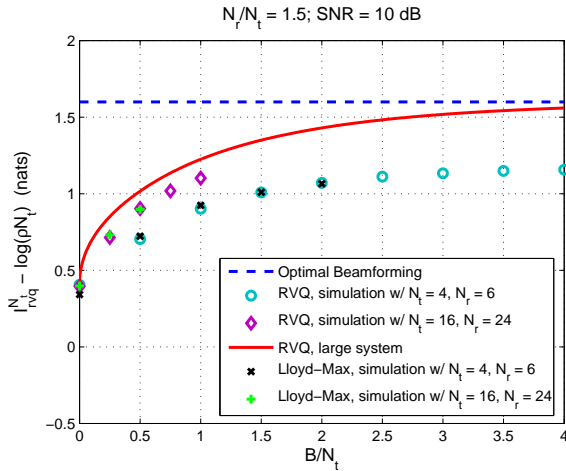Fig. 2. Asymptotic and simulated rate differences versus feedback bits for a MISO channel with beamforming.

Fig. 3. Asymptotic and simulated rate differences versus feedback bits for a MIMO channel with beamforming ($\bar{N}_r = 1.5$).

IV. PRECODING MATRIX WITH ARBITRARY RANK

In this section we consider the performance of a single-user MIMO channel with precoding matrix $\boldsymbol{V}$ having rank $K > 1$. We wish to determine the asymptotic capacity with RVQ as in the previous section. Here we consider the large system limit as $(N_t, N_r, B, K) \to \infty$ with fixed ratios $\bar{N}_r = N_r/N_t$, $\hat{B} = B/N_r^2$, and $\bar{K} = K/N_t$. That is, we scale the rank of the precoding matrix with N_t. The number of feedback bits is normalized by N_r^2, instead of N_r, since the feedback must scale linearly with degrees of freedom (in this case the number of channel elements $N_r N_t$).) Given a fixed number of feedback bits per channel coefficient, the capacity grows *linearly* with the number of antennas (N_t or N_r).

Given a rank $K \leq N_t$, the precoding matrix is chosen from the RVQ set

$$\mathcal{V} = \{\boldsymbol{V}_j, 1 \leq j \leq 2^B\}, \qquad (21)$$

where the entries are independent $N_t \times K$ random unitary matrices, i.e., $\boldsymbol{V}_j^\dagger \boldsymbol{V}_j = \boldsymbol{I}_K$. This codebook is an extension of

the RVQ codebook for beamforming. Letting

$$J_j^{N_r} = \frac{1}{N_r} \log \det \left(\boldsymbol{I}_{N_r} + \frac{\rho}{K} \boldsymbol{H} \boldsymbol{V}_j \boldsymbol{V}_j^\dagger \boldsymbol{H}^\dagger \right), \qquad (22)$$

the receiver again selects the quantized precoding matrix, which maximizes the mutual information

$$\hat{\boldsymbol{V}} = \arg \quad \max_{1 \leq j \leq 2^B} J_j^{N_r}. \qquad (23)$$

For finite N_r, we define

$$I_{\text{rvq}}^{N_r} = E_\mathcal{V}[\quad \max_{1 \leq j \leq 2^B} J_j^{N_r} | \bar{\boldsymbol{H}}] \qquad (24)$$

$$= E_\mathcal{V} \left[\frac{1}{N_r} \log \det \left(\boldsymbol{I}_{N_r} + \frac{\rho}{K} \boldsymbol{H} \hat{\boldsymbol{V}} \hat{\boldsymbol{V}}^\dagger \boldsymbol{H}^\dagger \right) \middle| \bar{\boldsymbol{H}} \right] \qquad (25)$$

and the average sum mutual information per receive antenna with B feedback bits is then $E_{\boldsymbol{H}}[I_{\text{rvq}}^{N_r}]$.

Here the power allocation over channel modes is "on-off". Namely, active modes are assigned equal powers. This simplifies the analysis, and it has been observed that the additional gain due to an optimal power allocation (water pouring) is quite small [8].

Since the entries of the RVQ codebook are *i.i.d.*, the mutual informations $J_j^{N_r}, j = 1, \ldots, 2^B$, are also *i.i.d.* for a given $\boldsymbol{H}$. In principle, the large system limit of $I_{\text{rvq}}^{N_r}$ can be evaluated, in analogy with (15), given the cdf of $J_j^{N_r}$ given $\boldsymbol{H}$, denoted as $F_{J;N_r|\boldsymbol{H}}$. This cdf appears to be difficult to determine in closed-form for general (N_r, N_t, K), so that we are unable to derive the exact asymptotic capacity with RVQ. Still, we can provide an accurate approximation for this large system limit. Before presenting this approximation, we first compare the capacity with no channel information at the transmitter ($\hat{B} = 0$) to the capacity with perfect channel information ($\hat{B} = \infty$).

If $\hat{B} = 0$, then the optimal transmit covariance matrix $\boldsymbol{V} \boldsymbol{V}^\dagger = \boldsymbol{I}_{N_t}$ and $K = N_t$ [35]. That is, all channel modes are allocated equal power. As $(N_t, N_r) \to \infty$ with fixed $\bar{N}_r = N_r/N_t$, the capacity per receive antenna is given by

$$\frac{1}{N_r} \log \det \left(\boldsymbol{I}_{N_r} + \frac{\rho}{N_t} \boldsymbol{H} \boldsymbol{H}^\dagger \right) \to \int_0^\infty \log \left(1 + \rho\lambda \right) g(\lambda) \, d\lambda \qquad (26)$$

$$= \mathcal{I}_{\text{rvq}}(\hat{B} = 0) \qquad (27)$$

where convergence is in the almost sure sense, and $g(\lambda)$ is the asymptotic probability density function for a randomly chosen eigenvalue of $\frac{1}{N_t} \boldsymbol{H} \boldsymbol{H}^\dagger$, and is given by [33]

$$g(\lambda) = \frac{\sqrt{(\lambda - a)(b - \lambda)}}{2\pi\lambda\bar{N}_r} \quad \text{for} \quad a \leq \lambda \leq b, \qquad (28)$$

$$a = \left(1 - \sqrt{\bar{N}_r} \right)^2 \quad \text{and} \quad b = \left(1 + \sqrt{\bar{N}_r} \right)^2 \qquad (29)$$

for $\bar{N}_r \leq 1$. The integral in (26) has been evaluated in [36], which gives the closed-form expression

$$\mathcal{I}_{\text{rvq}}(\hat{B} = 0) = \log \rho y + \frac{1 - \bar{N}_r}{\bar{N}_r} \log \left(\frac{1}{1 - z} \right) - \frac{z}{\bar{N}_r} \qquad (30)$$

where

$$y = \frac{1}{2}\left(1 + \bar{N}_r + \frac{1}{\rho} + \sqrt{\left(1 + \bar{N}_r + \frac{1}{\rho}\right)^2 - 4\bar{N}_r}\right) \quad (31)$$

$$z = \frac{1}{2}\left(1 + \bar{N}_r + \frac{1}{\rho} - \sqrt{\left(1 + \bar{N}_r + \frac{1}{\rho}\right)^2 - 4\bar{N}_r}\right). \quad (32)$$

If $\hat{B} = \infty$, then the K columns of the optimal $\boldsymbol{V}$ are the eigenvectors of the channel covariance matrix corresponding to the K largest eigenvalues. As $(N_t, N_r, B) \to \infty$, we have

$$\mathcal{I}_{\mathsf{rvq}}(\hat{B} = \infty) = \int_\eta^\infty \log\left(1 + \frac{\rho}{K}\lambda\right) g(\lambda) \, \mathrm{d}\lambda \quad (33)$$

where η satisfies

$$\int_\eta^\infty g(\lambda) \, \mathrm{d}\lambda = \min\{1, \frac{\bar{K}}{\bar{N}_r}\} \quad (34)$$

for $\bar{N}_r \le 1$. We emphasize that this corresponds to a uniform allocation of power over the set of K active eigenvectors. (This result has also been presented in [8].) The rank of the optimal $\boldsymbol{V}$, or optimal multiplexing gain, is at most $\min\{N_t, N_r\}$ and can be obtained by differentiating (33) with respect to $\bar{K}$. It can be verified that $\mathcal{I}_{\mathsf{rvq}}(\hat{B} = 0) \le \mathcal{I}_{\mathsf{rvq}}(\hat{B} = \infty)$.

To illustrate the increase in capacity with feedback, in Fig. 4 we plot the rate ratio $\mathcal{I}_{\mathsf{rvq}}(\hat{B} = \infty)/\mathcal{I}_{\mathsf{rvq}}(\hat{B} = 0)$ versus SNR for different values of $\bar{N}_r$, where $\mathcal{I}_{\mathsf{rvq}}(\hat{B} = \infty)$ is optimized over rank K. For large SNR ρ, we can expand

$$\mathcal{I}_{\mathsf{rvq}}(\hat{B} = 0) = \log(\rho) + o(\log(\rho)) \quad (35)$$

$$\mathcal{I}_{\mathsf{rvq}}(\hat{B} = \infty) = \log(\rho) \int_\eta^\infty g(\lambda) \, \mathrm{d}\lambda + o(\log(\rho)). \quad (36)$$

Therefore

$$\lim_{\rho \to \infty} \frac{\mathcal{I}_{\mathsf{rvq}}(\hat{B} = \infty)}{\mathcal{I}_{\mathsf{rvq}}(\hat{B} = 0)} = \int_\eta^\infty g(\lambda) \, \mathrm{d}\lambda = \min\{1, \frac{\bar{K}}{\bar{N}_r}\} \quad (37)$$

which implies that the optimal rank $K^* = \min\{N_t, N_r\}$, and the corresponding asymptotic rate ratio is one. The increase in achievable rate from feedback is small in this case, since for large SNRs, the transmitter excites all channel modes, and the uniform power allocation asymptotically gives the same capacity as water pouring. Of course, although the increase in rate is small, feedback can simplify coding and decoding.

For small ρ, we can expand $\log(1 + \rho\lambda)$ and $\log(1 + \rho\lambda/\bar{K})$ in Taylor series. Taking $\rho \to 0$ gives

$$\lim_{\rho \to 0} \frac{\mathcal{I}_{\mathsf{rvq}}(\hat{B} = \infty)}{\mathcal{I}_{\mathsf{rvq}}(\hat{B} = 0)}$$

$$= \frac{1}{\bar{K}}\int_\eta^\infty \lambda g(\lambda) \, \mathrm{d}\lambda \le \frac{1}{\bar{N}_r}\frac{\int_\eta^\infty \lambda g(\lambda) \, \mathrm{d}\lambda}{\int_\eta^\infty g(\lambda) \, \mathrm{d}\lambda} \quad (38)$$

$$\le \frac{1}{\bar{N}_r}\frac{b\int_\eta^b g(\lambda) \, \mathrm{d}\lambda}{\int_\eta^b g(\lambda) \, \mathrm{d}\lambda} = \left(1 + \frac{1}{\sqrt{\bar{N}_r}}\right)^2 \quad (39)$$

where b is the asymptotic maximum eigenvalue given by (29). The inequality in (38) follows from (34), which implies $\bar{K} \ge \bar{N}_r \int_\eta^\infty g(\lambda) \, \mathrm{d}\lambda$. Note that (39) corresponds to allocating all

transmission power to the strongest channel mode, which is known to maximize capacity at low SNRs. The maximal rate ratio (39) can also be obtained from Theorem 3.

The rate increase due to feedback is substantial when $\bar{N}_r$ is small, and the rate ratio tends to infinity as $\bar{N}_r \to 0$. This is because the channel becomes a MISO channel, in which case the capacity is a constant with $\bar{B} = 0$ and increases as $\log(\rho N_t)$ with $\bar{B} = \infty$.
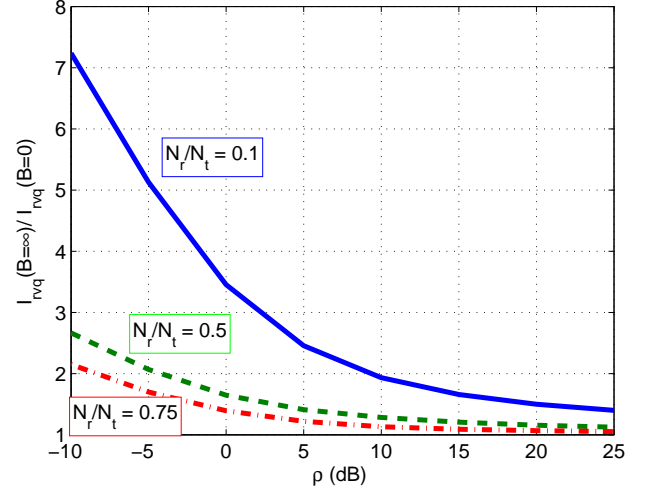

Fig. 4. The rate ratio $\mathcal{I}_{\mathsf{rvq}}(\hat{B} = \infty)/\mathcal{I}_{\mathsf{rvq}}(\hat{B} = 0)$ versus SNR (dB) for various values of $\bar{N}_r$.

To evaluate the asymptotic capacity with arbitrary $\hat{B}$, we approximate $J_j^{N_r}$ given $\bar{H}$ as a Gaussian random variable. This is motivated by the fact that $J_j^{N_r}$ is Gaussian in the large system limit [37], since $\boldsymbol{HV}$ is *i.i.d.* Conditioning on $\boldsymbol{H}$ introduces dependence among the elements of $\boldsymbol{HV}$; however, numerical examples indicate that the Gaussian assumption is still valid for large N_t and N_r. Alternatively, if we do not condition on $\boldsymbol{H}$, then the rates $\{J_j^{N_r}\}$ are dependent. Application of the results from extreme statistics, assuming the rates $\{J_j^{N_r}\}$ are independent, gives an upper bound on the asymptotic achievable rate (e.g., see the proof of Theorem 2 in [21]). This is illustrated by subsequent numerical results.

Evaluating the large system limit of $I_{\mathsf{rvq}}^{N_r}$, assuming that the cdf of $J_j^{N_r}$ is Gaussian, gives the approximate rate

$$\tilde{\mathcal{I}}_{\mathsf{rvq}} = \mu_J + \sigma_J \sqrt{2\hat{B}\log 2} \quad (40)$$

independent of the channel realization, where μ_J and σ_J^2 are the asymptotic mean of $J_j^{N_r}$, and variance of $N_r^2 J_j^{N_r}$, respectively. The derivation of (40) is a straightforward extension of [32, Sec. 2.3.2] and is not shown here. As $\hat{B} \to 0$, this approximation becomes exact. However, as $\hat{B} \to \infty$, the approximate rate $\tilde{\mathcal{I}}_{\mathsf{rvq}} \to \infty$, whereas the actual rate $\mathcal{I}_{\mathsf{rvq}}(\hat{B} = \infty)$ is finite, and can be computed from (33) and (34). This is because $J_j^{N_r}$ is bounded for all N_r, whereas a Gaussian random variable can assume arbitrarily large values. Therefore the Gaussian approximation gives an inaccurate estimate of $\mathcal{I}_{\mathsf{rvq}}$ for large $\hat{B}$. (This implies that we should approximate $\mathcal{I}_{\mathsf{rvq}}$ as $\min\{\tilde{\mathcal{I}}_{\mathsf{rvq}}, \mathcal{I}_{\mathsf{rvq}}(\hat{B} = \infty)\}$.)

The asymptotic mean and variance of $J_j^{N_r}$ are computed in Appendix D. The asymptotic mean is given by

$$\mu_J = \frac{\bar{K}}{\bar{N}_r} \log\left(1 + \frac{\bar{N}_r}{\bar{K}}\rho - \frac{\bar{N}_r}{\bar{K}}\rho v\right)$$
$$+ \log\left(1 + \rho - \frac{\bar{N}_r}{\bar{K}}\rho v\right) - v \qquad (41)$$

where

$$v = \frac{1}{2} + \frac{\bar{K}}{2\bar{N}_r} + \frac{\bar{K}}{2\bar{N}_r\rho} - \frac{1}{2}\sqrt{\left(1 + \frac{\bar{K}}{\bar{N}_r} + \frac{\bar{K}}{\bar{N}_r\rho}\right)^2 - \frac{4\bar{K}}{\bar{N}_r}}. \qquad (42)$$

The asymptotic variance is approximated for $0 \leq \bar{K} = \bar{N}_r \leq 1$ and small SNR ($\rho \leq -5$ dB) as

$$\sigma_J^2 \approx \rho^2(1 - \bar{N}_r). \qquad (43)$$

The asymptotic variance for moderate SNRs and normalized rank $\bar{K} \neq \bar{N}_r$ can be computed easily via numerical simulation.[1]

In contrast with the beamforming results in the preceding section, we are unable to show that RVQ is asymptotically optimal when the precoding matrix has arbitrary rank. The corresponding argument for beamforming relies on the evaluation of the asymptotic rate difference $\mathcal{I}_{\text{rvq}}^\Delta$. Since here we are unable to evaluate $\mathcal{I}_{\text{rvq}}$ exactly, we cannot apply that argument. Nevertheless, numerical results have indicated that the performance of RVQ matches that of optimized codebooks (e.g., see [22]).

Fig. 5 shows $\tilde{\mathcal{I}}_{\text{rvq}}$ with normalized rank $\bar{K} = \bar{N}_r$ versus $\hat{B}$ for $\rho = -5, 0, 5$ dB and $\bar{N}_r = 0.5$. The dashed lines show the unlimited feedback capacity $\mathcal{I}_{\text{rvq}}(\hat{B} = \infty)$, which is computed from (33) with optimized $\bar{K}$. The asymptotic rate with RVQ is computed from (40), where σ_J for $\rho = -5$ dB is approximated by (43), and σ_J is determined from simulation with $N_t = 20$ for $\rho = 0$ and 5 dB. Also shown in Fig. 5 are simulation results for $\mathcal{I}_{\text{rvq}}$ with $N_t = 8$ and $N_r = 4$. Because the size of the RVQ codebook increases exponentially with $\hat{B}$, it is difficult to generate simulation results for moderate to large values of $\hat{B}$. Hence simulation results are shown only for $\hat{B} \leq 0.8$. The asymptotic results accurately approximate the simulated results shown. The accuracy increases as the feedback $\tilde{B}$ decreases.

Since $\tilde{\mathcal{I}}_{\text{rvq}}$ is a function of both rank $\bar{K}$ and feedback $\hat{B}$, for a given $\hat{B}$, we can select $\bar{K}$ to maximize $\tilde{\mathcal{I}}_{\text{rvq}}$. Fig. 6 shows mutual information per receive antenna versus normalized rank from (40) with $\bar{N}_r = 0.2$, $\rho = 5$ dB, and different values of $\hat{B}$. (σ_J is obtained from numerical simulations.) The maximal rates are attained at $\bar{K} = 1$, 0.3, and 0.2 for $\hat{B} = 0$, 0.5, and 2, respectively. In general, the optimal rank is approximately $\bar{N}_r$ for large enough $\hat{B}$ and SNR. The results in Fig. 6 indicate that taking $\bar{K} = \bar{N}_r$ achieves near-optimal performance, independent of $\hat{B}$ when $\hat{B} > 0$. As $\hat{B}$ increases, the rate increases and the difference between the rate with optimized rank and full-rank ($\bar{K} = 1$) also increases. For the
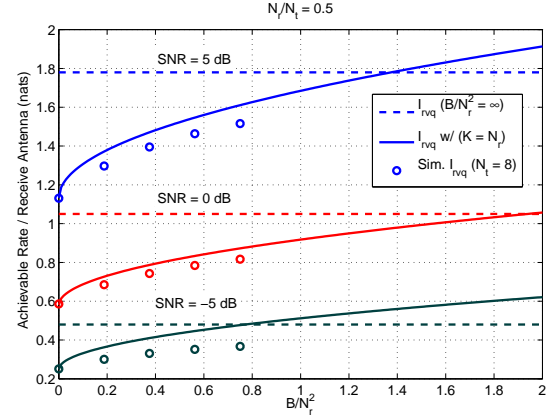

Fig. 5. Sum mutual information per receive antenna with RVQ and an optimal receiver versus normalized feedback. The asymptotic approximation is shown along with Monte Carlo simulation results for an 8×4 channel.
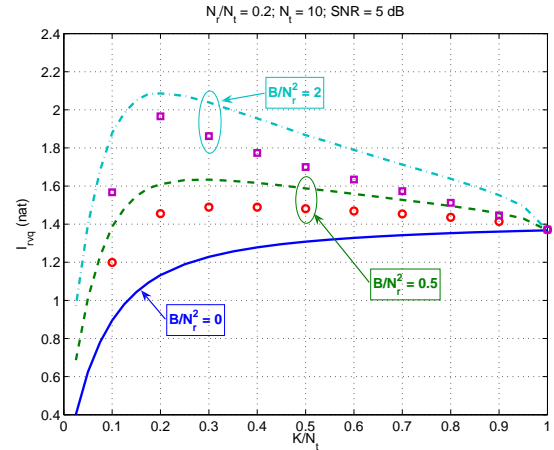

Fig. 6. Mutual information per receive antenna versus normalized rank with different normalized feedback. Discrete points correspond to simulation with $N_t = 10$.

example shown, the rate increase from selecting the optimal rank is as high as 50% when $\hat{B} = 2$.

V. QUANTIZED PRECODING WITH LINEAR RECEIVERS

In this section we evaluate the performance of a quantized precoding matrix with linear receivers (matched filter and MMSE), and compare with the performance of the optimal receiver. As $\hat{B} \to \infty$, the optimal precoding matrix eliminates the cross-coupling among channel modes, and the optimal receiver becomes the linear matched filter. Hence the corresponding achievable rates should be the same in this limit. However, for finite $\hat{B}$ the optimal receiver is expected to perform better than the linear receiver. Given a target rate, increasing the feedback therefore enables a reduction in receiver complexity.

We again assume that there are K independent data streams, which are multiplexed by the linear precoder onto N_t transmit antennas. To detect the transmitted symbols in data stream k, the received signal $\boldsymbol{y}$ is passed through the $N_r \times 1$ receive

---

[1]We note that the simulation needed to compute this variance is much simpler than the simulation, which would be required to obtain the RVQ rate directly, especially with a moderate to large number of feedback bits.

filter c_k. The matched filter is given by

$$c_k = \frac{1}{\sqrt{K}} H v_k \qquad (44)$$

where v_k is the kth column of the precoding matrix V, and the linear MMSE filter is given by

$$c_k = \frac{1}{\sqrt{K}} \left(\frac{1}{K} H V V^\dagger H^\dagger + \sigma_n^2 I_{N_r} \right)^{-1} H v_k. \qquad (45)$$

The SINR at the output of the linear filter c_k is

$$\mathsf{SINR}_k = \frac{|c_k^\dagger H v_k|^2}{c_k^\dagger \left(\sum_{i \neq k} H v_i v_i^\dagger H^\dagger + K \sigma_n^2 I_{N_r} \right) c_k}. \qquad (46)$$

Of course, the interference among data streams can significantly decrease the channel capacity.

The performance measure is again mutual information between the transmitted symbol x_k and the output of the filter c_k, denoted by $\hat{x}_k$. In what follows, we assume independent coders and decoders for each data stream. Assuming that the interference plus noise at the output of the linear filter has a Gaussian distribution, which is true in the large system limit to be considered, the sum mutual information of all data streams per receive antenna is given by

$$R^{N_r} = \frac{1}{N_r} \sum_{k=1}^{K} I(x_k, \hat{x}_k) \qquad (47)$$

$$= \frac{1}{N_r} \sum_{k=1}^{K} \log(1 + \gamma_k). \qquad (48)$$

where γ_k is the SINR for the kth data stream. Given a channel matrix H, the sum rate R^{N_r} depends on the precoding matrix V. We are interested in maximizing R^{N_r} subject to the power constraint $\|v_k\| \leq 1$, $\forall k$, assuming that the power is allocated equally across streams.

Given the codebook of precoding matrices $\mathcal{V} = \{V_j, \; 1 \leq j \leq 2^B\}$, the receiver selects the precoding matrix

$$\hat{V} = \arg \max_{1 \leq j \leq 2^B} R^{N_r}(V_j). \qquad (49)$$

We again consider RVQ, in which the V_j's are *i.i.d.* unitary matrices.

A. Matched filter

Substituting (44) into (46), the SINR at the output of the matched filter is given by

$$\gamma_{k;\mathsf{mf}} = \frac{(v_k^\dagger H^\dagger H v_k)^2}{K \sigma_n^2 (v_k^\dagger H^\dagger H v_k) + \sum_{i=1, i \neq k}^{K} |v_k^\dagger H^\dagger H v_i|^2} \qquad (50)$$

where subscript k denotes the kth data stream. The average sum rate per receive antenna is given by

$$E_{H, \mathcal{V}} \left[\max_{1 \leq j \leq 2^B} \{ R_{\mathsf{mf}}^{N_r}(V_j) = \frac{1}{N_r} \sum_{k=1}^{K} \log(1 + \gamma_{k;\mathsf{mf}}) \} \right] \qquad (51)$$

where the expectation is over the channel matrix and codebook. Since the pdf of $R_{\mathsf{mf}}^{N_r}$ is unknown for finite (N_t, N_r, K), we are unable to evaluate (51). Motivated by the central limit theorem,[2] in what follows we approximate the cdf of $R_{\mathsf{mf}}^{N_r}$ as

---

[2]The terms in the sum in (51) are not *i.i.d.*, which prevents a direct application of the central limit theorem.

PDF for $R_{\mathsf{mf};N_r}$; $N_r = 10$; $\bar{N}/N_t = 1$; $K/N_t = 0.3$; SNR = 5dB
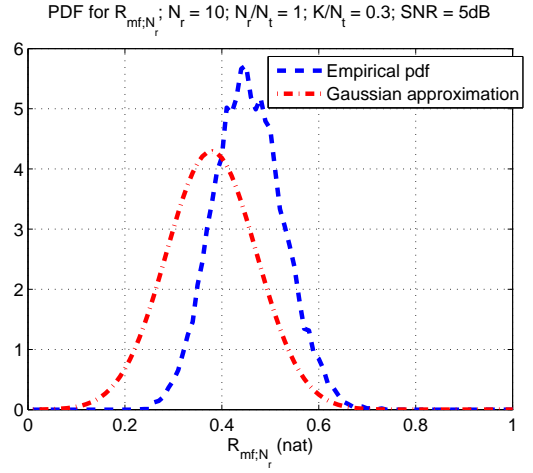
Fig. 7. Comparison of the empirical pdf for $R_{\mathsf{mf};N_r}$ with the Gaussian approximation.

Gaussian. The mean is taken to be the asymptotic limit

$$\mu_{\mathsf{mf}} = \lim_{(N_t, N_r, K) \to \infty} R_{\mathsf{mf}}^{N_r} = \frac{\bar{K}}{\bar{N}_r} \log \left(1 + \frac{\bar{N}_r}{\bar{K}(1 + \sigma_n^2)} \right). \qquad (52)$$

This limit follows from the fact that $\gamma_{k;\mathsf{mf}}$ converges almost surely to $[\bar{K}(1 + \sigma_n^2)/\bar{N}_r]^{-1}$ as $(N_t, N_r, K) \to \infty$ with fixed $\bar{N}_r$ and $\bar{K}$. As for the optimal receiver,

$$N_r^2 \mathrm{var}[R_{\mathsf{mf}}^{N_r} | V_j] \to \sigma_{\mathsf{mf}}^2 \qquad (53)$$

where σ_{mf}^2 can be easily obtained by numerical simulations.

The accuracy of the Gaussian approximation for $R_{\mathsf{mf}}^{N_r}$ is illustrated in Fig. 7, which compares the empirical pdf with the Gaussian approximation for $N_r = 10$, $\bar{N}_r = 1$, $K/N_r = 0.3$ and $\mathrm{SNR} = 5$ dB. The difference between the empirical and asymptotic means vanishes as $(N_t, N_r, K) \to \infty$.

We wish to apply the theory of extreme order statistics [32] to evaluate the large system limit

$$\mathcal{R}_{\mathsf{rvq};\mathsf{mf}} = \lim_{(N_t, N_r, K, B) \to \infty} [\max_{1 \leq j \leq 2^B} R_{\mathsf{mf}}^{N_r}(V_j) | \mathcal{V}]. \qquad (54)$$

Given $\mathcal{V}$, the sum rates $\{R_{\mathsf{mf}}^{N_r}(V_1), \ldots, R_{\mathsf{mf}}^{N_r}(V_{2^B})\}$ are identically distributed. However, the $R_{\mathsf{mf}}^{N_r}(V_j)$'s are not independent since each depends on H. This makes an exact calculation of the asymptotic rate difficult. Nevertheless, for a small number of entries in the codebook (small B), assuming that the rates for a given codebook are independent leads to an accurate approximation. We therefore replace the rates $R_{\mathsf{mf}}^{N_r}(V_j)$, $j = 1, \cdots, 2^B$, with *i.i.d.* Gaussian variables with mean μ_{mf} and variance $\sigma_{\mathsf{mf}}^2 / N_r^2$. In analogy with the analysis of the optimal receiver in the preceding section, this gives the approximate asymptotic rate

$$\tilde{\mathcal{R}}_{\mathsf{rvq};\mathsf{mf}} = \mu_{\mathsf{mf}} + \sigma_{\mathsf{mf}} \sqrt{2 \hat{B} \log 2}. \qquad (55)$$

Numerical results, to be presented, show that this asymptotic approximation is very accurate for small to moderate values of normalized feedback $\hat{B}$. As $\hat{B} \to 0$, this approximation becomes exact. However, as $\hat{B} \to \infty$, $\tilde{\mathcal{R}}_{\mathsf{rvq};\mathsf{mf}} \to \infty$, whereas $\mathcal{R}_{\mathsf{rvq};\mathsf{mf}}$ with $\hat{B} = \infty$ is the same as the asymptotic rate with

RVQ and an optimal receiver, given by (33) and (34). Hence $\mathcal{R}_{\text{rvq;mf}}$ with $\hat{B} = \infty$ is finite. As for the analysis of the optimal receiver, this discrepancy is again due to the fact that the cdf of $R_{\text{mf}}^{N_r}$, which has compact support, is being approximated by a Gaussian cdf with infinite support, and also because the dependence among the sum rates $R_{\text{mf}}^{N_r}(\boldsymbol{V}_j)$ is being ignored.

B. MMSE receiver

Substituting (45) into (46) gives the SINR at the output of MMSE receiver for the kth symbol stream

$$\gamma_{k;\text{mmse}} = \boldsymbol{v}_k^\dagger \boldsymbol{H}^\dagger \left(\sum_{i \neq k} \boldsymbol{H} \boldsymbol{v}_i \boldsymbol{v}_i^\dagger \boldsymbol{H}^\dagger + K\sigma_n^2 \boldsymbol{I}_{N_r} \right)^{-1} \boldsymbol{H} \boldsymbol{v}_k. \tag{56}$$

As for the matched filter receiver, given a codebook $\mathcal{V}$, we approximate the pdf of the instantaneous sum rate

$$R_{\text{mmse}}^{N_r} = \frac{1}{N_r} \sum_{k=1}^{K} \log(1 + \gamma_{k;\text{mmse}}) \tag{57}$$

as a Gaussian pdf with mean

$$\mu_{\text{mmse}} = \lim_{(N_t, N_r) \to \infty} R_{\text{mmse}}^{N_r} = \frac{\bar{K}}{\bar{N}_r} \log(1 + \gamma_{\text{mmse}}) \tag{58}$$

where the large system SINR is given by [30]

$$\gamma_{\text{mmse}} = \frac{1 - \bar{K}/\bar{N}_r}{2\sigma_n^2} - \frac{1}{2} \\ + \sqrt{\frac{(1 - \bar{K}/\bar{N}_r)^2}{4\sigma_n^4} + \frac{1 + \bar{K}/\bar{N}_r}{2\sigma_n^2} + \frac{1}{4}}. \tag{59}$$

As for the matched filter, the asymptotic variance σ_{mmse}^2 can be obtained via numerical simulation.

In analogy with (55), the asymptotic rate with RVQ and the MMSE receiver is given by

$$\mathcal{R}_{\text{rvq;mmse}} \approx \tilde{\mathcal{R}}_{\text{rvq;mmse}} = \mu_{\text{mmse}} + \sigma_{\text{mmse}} \sqrt{2\hat{B} \log 2}. \tag{60}$$

As for the matched filter receiver, when $\hat{B}$ is large, $\tilde{\mathcal{R}}_{\text{mmse}}$ over-estimates $\mathcal{R}_{\text{mmse}}$. For $\hat{B} = \infty$, $\mathcal{R}_{\text{mmse}} = \mathcal{R}_{\text{mf}} = \mathcal{I}_{\text{rvq}}$ with the optimal receiver, given by (33).

C. Numerical Results

Fig. 8 compares the approximation for asymptotic RVQ performance with a matched filter receiver from (55) with simulated results for $N_t = 12$, $\bar{N}_r = 0.75$, $K/N_t = 1/2$, and SNR $= 5$ dB. Also shown for comparison are the asymptotic rate for RVQ with an optimal receiver, derived in Section IV, the water-filling capacity ($\hat{B} = \infty$), and the rate achieved with a scalar quantizer for each coefficient.[3] For the case shown, the analytical approximation gives an accurate estimate of the performance of the finite size system with limited feedback. The capacity with the water-filling power allocation is only slightly greater than that achieved with the on-off power allocation. The optimal receiver requires $\hat{B} \approx 0.6$ bit/dimension

[3]For the scalar quantization results the available bits are spread evenly over the corresponding fraction of precoding coefficients. The remaining coefficients are set to one.
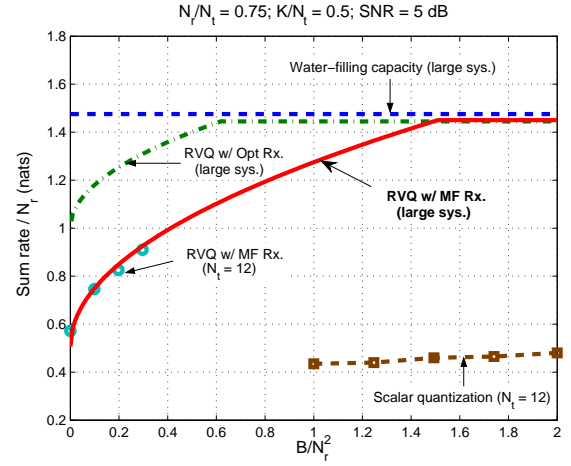

Fig. 8. Sum rate per receive antenna versus normalized feedback bits with a matched filter receiver. Results are shown for RVQ (asymptotic and $N_t = 12$) and scalar quantization. Also shown are results for the optimal receiver, and the water-filling capacity with infinite feedback.
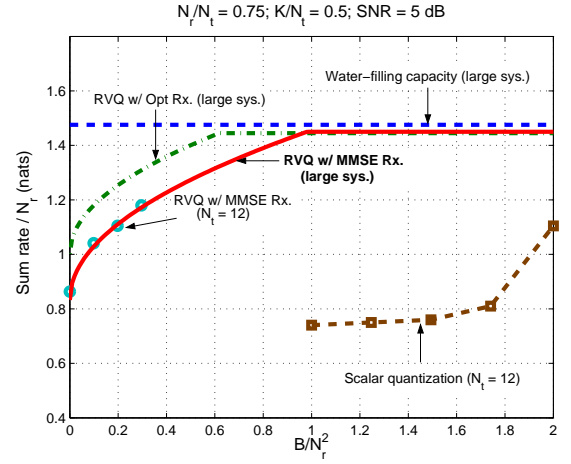

Fig. 9. Sum rate per receive antenna versus normalized feedback bits with a linear MMSE receiver. Results are shown for RVQ (asymptotic and $N_t = 12$) and scalar quantization. Also shown are results for the optimal receiver, and the water-filling capacity with infinite feedback.

to achieve the capacity corresponding to unlimited feedback (33), whereas the matched filter requires 1.2 feedback bits per dimension to reach that capacity. For other target rates, these curves illustrate the trade-off between feedback and receiver complexity.

Fig. 9 shows the same set of results as those shown in Fig. 8, but with an MMSE receiver. These results show that for the parameters selected, the MMSE receiver performs nearly as well as the optimal receiver, and requires substantially less feedback than the matched filter to achieve a target rate. Again the asymptotic approximation accurately predicts the performance of a system with a relatively small number of antennas.

VI. CONCLUSIONS

We have studied the capacity of single-user MISO and MIMO fading channels with limited feedback. The feedback

specifies a transmit precoding matrix, which can be optimized for a given channel realization. We first considered the performance with a rank-one precoding matrix (beamformer), and showed that the RVQ codebook is asymptotically optimal. Exact expressions for the asymptotic mutual information for MISO and MIMO channels were presented, and reveal how much feedback is required to achieve a desired performance. For the cases considered, one feedback bit for each precoder coefficient can achieve close to the water-filling capacity. Perhaps more important than the increase in capacity provided by this feedback is the associated simplification in the coding and decoding schemes that can achieve a rate close to capacity.

The performance of a precoding matrix with rank $K > 1$ was also evaluated with RVQ. Although numerical examples and our beamforming results ($K = 1$) suggest that RVQ is also asymptotically optimal in this case, proving this is an open problem. To compute the asymptotic achievable rate for RVQ with both optimal and linear receivers, the achievable rate with a random channel and fixed precoding matrix is approximated as a Gaussian random variable. The asymptotic rate then depends on the asymptotic mean and variance of this random variable. Although the asymptotic variance appears to be difficult to compute analytically, it can be easily obtained by simulation. Numerical results have shown that the resulting approximation accurately estimates the achievable rate with limited feedback for finite-size systems of interest.

Numerical examples comparing the performance of optimal and linear receivers have shown that the linear MMSE receiver requires little additional feedback, relative to the optimal receiver, to achieve a target rate close to the water-filling capacity. The matched filter requires significantly more feedback than the MMSE receiver (more than 0.5 bit per degree of freedom for the cases shown). At low feedback rates the achievable rate with RVQ is generally much greater than that associated with scalar quantization. Of course, this comes at a price of high complexity, since the receiver is assumed to compute the performance metric for every entry in the codebook. Other reduced complexity schemes for quantizing a beamforming vector are presented in [21], [38], [39].

Key assumptions for our results are that the channel is stationary and known at the receiver, and that the channel elements are *i.i.d.* Depending on user mobility and associated Doppler shifts, the channel may change too fast to allow reliable channel estimation and feedback. In that case, feedback of channel *statistics*, as proposed in [15]–[18], [40], can exploit correlation among channel elements. The design of quantization codebooks for precoders, which takes correlation into account, is addressed in [40]. The effect of channel estimation error on the performance of limited feedback beamforming with finite coherence time (i.e., block fading) is presented in [41], [42].

We have also assumed that the channel gains are not frequency-selective. Limited feedback schemes for frequency-selective scalar channels are discussed in [43], and could be combined with the quantization schemes considered here. Finally, the approach presented here for a single-user MIMO channel can also be applied to multi-user models. Quantization of beamformers for the MIMO downlink have been considered

in [26]–[28]. In that scenario the potential capacity gain due to feedback is generally much more than for the single-user channel considered here. The benefits of limited feedback for related models (e.g., frequency-selective MIMO downlink) are currently being studied.

APPENDIX

A. Proof of Theorem 1

Given $\boldsymbol{h}$, the receiver selects the quantized beamforming vector $\hat{\boldsymbol{v}}$ to maximize the instantaneous rate in (3). Since $\log$ is monotonically increasing, the quantized beamforming vector is given by

$$\hat{\boldsymbol{v}} = \arg \max_{1 \leq j \leq 2^B} \left\{ Y_j = |\boldsymbol{v}_j^\dagger \boldsymbol{h}|^2 / \|\boldsymbol{h}\|^2 \right\}. \tag{61}$$

Since the codebook entries $\boldsymbol{v}_j$, $j = 1, \cdots, 2^B$, are *i.i.d.*, the Y_j's, given $\boldsymbol{h}$, are also *i.i.d.* with the cdf [25]

$$F_{Y|\boldsymbol{h}}(y) = 1 - (1 - y)^{N_t - 1}, \quad 0 \leq y \leq 1. \tag{62}$$

We wish to determine the distribution of $\max_j Y_j$ given $\boldsymbol{h}$. From [32, Theorem 2.1.2] it follows that

$$\frac{\max_j Y_j - a_n}{b_n} \xrightarrow{\mathcal{D}} \mathcal{Y} \tag{63}$$

where $\mathcal{Y}$ is a Weibull random variable having distribution

$$H_\gamma(x) = \begin{cases} 1, & x \geq 0 \\ \exp(-(-x)^\gamma), & x < 0 \end{cases}, \tag{64}$$

$\mathcal{D}$ denotes convergence in distribution, and a_n and b_n are normalizing sequences, where $n = 2^B$. Specifically, the theorem requires that $\omega(F_{Y|\boldsymbol{h}}) = \sup\{y : F_{Y|\boldsymbol{h}}(y) < 1\}$ be finite, and that the distribution function $F_{Y|\boldsymbol{h}}^*(y) = F_{Y|\boldsymbol{h}}(\omega(F_{Y|\boldsymbol{h}}) - 1/y), y > 0$ satisfies, for all $y > 0$,

$$\lim_{t \to \infty} \frac{1 - F_{Y|\boldsymbol{h}}^*(ty)}{1 - F_{Y|\boldsymbol{h}}^*(t)} = y^{-\gamma} \tag{65}$$

where the constant $\gamma > 0$.

Substituting the expression for $F_{Y|\boldsymbol{h}}$ in (62) into (65), where $\omega(F_{Y|\boldsymbol{h}}) = 1$, gives

$$\lim_{t \to \infty} \frac{1 - F_{Y|\boldsymbol{h}}^*(ty)}{1 - F_{Y|\boldsymbol{h}}^*(t)} = \lim_{t \to \infty} \frac{(1/(ty))^{N_t - 1}}{(1/t)^{N_t - 1}} \tag{66}$$

$$= y^{-(N_t - 1)} \tag{67}$$

so that [32, Theorem 2.1.2] applies when $N_t > 1$. Furthermore, the normalizing constants are given by

$$a_n = \omega(F_{Y|\boldsymbol{h}}) = 1 \tag{68}$$

and

$$b_n = \omega(F_{Y|\boldsymbol{h}}) - \inf \left\{ y : 1 - F_{Y|\boldsymbol{h}}(y) \leq \frac{1}{n} \right\} \tag{69}$$

$$= 1 - F_{Y|\boldsymbol{h}}^{-1} \left(1 - \frac{1}{n} \right) = \left(\frac{1}{n} \right)^{\frac{1}{N_t - 1}}. \tag{70}$$

To take the limit as $N_t \to \infty$, we will assume that the channel vector $\boldsymbol{h}$ contains the first N_t elements of an infinite-length *i.i.d.* complex Gaussian vector $\bar{\boldsymbol{h}}$. Rearranging terms in (63) and taking the large system limit gives

$$\lim_{(N_t,n)\to\infty} E_{\mathcal{V}}[\max_j Y_j | \bar{\boldsymbol{h}}] \tag{71}$$

$$= \lim_{(N_t,n)\to\infty} a_n + b_n E[\mathcal{Y}] \tag{72}$$

$$= 1 - \lim_{(N_t,n)\to\infty} \left(\frac{1}{n}\right)^{\frac{1}{N_t-1}} \Gamma\left(1 - \frac{1}{N_t-1}\right) \tag{73}$$

$$= 1 - \lim_{(N_t,B)\to\infty} 2^{-\frac{B}{N_t-1}} \tag{74}$$

$$= 1 - 2^{-\bar{B}} \tag{75}$$

where the gamma function $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t}\, dt$ and we have used the fact that $E[\mathcal{Y}] = -\Gamma(1 - 1/(N_t - 1))$ [44].

From (63), as $(n, N_t) \to \infty$,

$$\text{var}[\max_j Y_j | \bar{\boldsymbol{h}}] - b_n^2 \text{var}[\mathcal{Y}] \to 0. \tag{76}$$

Since $\text{var}[\mathcal{Y}] = \Gamma(1 - 2/(N_t-1)) - \Gamma^2(1 - 1/(N_t-1)) \to 0$, it follows that $\text{var}[\max_j Y_j | \boldsymbol{h}] \to 0$. This establishes that given $\bar{\boldsymbol{h}}$,

$$\max_j Y_j \to 1 - 2^{-\bar{B}} \tag{77}$$

in the mean square sense. The asymptotic rate difference is given by

$$\mathcal{I}_{\text{rvq}}^{\triangle} = \lim_{(N_t,B)\to\infty} \left(\log(1 + \rho\|\boldsymbol{h}\|^2 \max_{1 \le j \le 2^B} Y_j) \right) - \log(\rho N_t) \tag{78}$$

$$= \lim_{(N_t,B)\to\infty} \left(\log\left(\frac{1}{\rho N_t} + \frac{1}{N_t}\|\boldsymbol{h}\|^2 \max_{1 \le j \le 2^B} Y_j\right) \right) \tag{79}$$

$$= \log(1 - 2^{-\bar{B}}) \tag{80}$$

in the mean square sense, since $\|\boldsymbol{h}\|^2/N_t \to 1$ almost surely.

B. Proof of Theorem 2

The rate difference associated with codebook $\mathcal{V}_{N_t}$ is

$$I_{\mathcal{V}_{N_t}}^{\triangle} = \max_{1 \le j \le 2^B} \log\left(\frac{1}{\rho N_t} + \frac{1}{N_t}|\boldsymbol{h}^\dagger \boldsymbol{v}_j^{N_t}|^2\right) \tag{81}$$

$$= \log\left(\frac{1}{\rho N_t} + \max_{1 \le j \le 2^B} \frac{1}{N_t}|\boldsymbol{h}^\dagger \boldsymbol{v}_j^{N_t}|^2\right). \tag{82}$$

Taking expectation of the rate difference with respect to $\boldsymbol{h}$ and applying Jensen's inequality, we obtain

$$E[I_{\mathcal{V}_{N_t}}^{\triangle}] \le \log\left(\frac{1}{\rho N_t} + E[\frac{1}{N_t}|\boldsymbol{h}^\dagger \hat{\boldsymbol{v}}^{N_t}|^2]\right) \tag{83}$$

$$= \log\left(\frac{1}{\rho N_t} + E[\frac{1}{N_t}\|\boldsymbol{h}\|^2] E[\mu]\right) \tag{84}$$

where the optimal beamforming vector

$$\hat{\boldsymbol{v}}^{N_t} = \arg \max_{1 \le j \le 2^B} |\boldsymbol{h}^\dagger \boldsymbol{v}_j^{N_t}|^2, \tag{85}$$

$\mu = |\boldsymbol{h}^\dagger \hat{\boldsymbol{v}}^{N_t}|^2 / \|\boldsymbol{h}\|^2$, and (84) follows from the fact that $\|\boldsymbol{h}\|^2$ and μ are independent [25].

We now derive an upper bound for $E[|\boldsymbol{h}^\dagger \hat{\boldsymbol{v}}^{N_t}|^2 / \|\boldsymbol{h}\|^2]$. From (30) in [2] we have

$$\Pr\{|\boldsymbol{h}^\dagger \hat{\boldsymbol{v}}^{N_t}|^2 > \gamma s \mid \|\boldsymbol{h}\|^2 = \gamma\}$$

$$\le \begin{cases} 1, & 0 \le s < s^* \\ 2^B(1-s)^{N_t-1}, & s^* \le s \le 1 \end{cases} \tag{86}$$

where

$$s^* = 1 - 2^{-\frac{B}{N_t-1}}. \tag{87}$$

Since the right-hand side of (86) is independent of γ, averaging over γ gives

$$\Pr\{\mu > s\} \le \begin{cases} 1, & 0 \le s < s^* \\ 2^B(1-s)^{N_t-1}, & s^* \le s \le 1 \end{cases}. \tag{88}$$

Integrating by parts, we have that

$$E[\mu] = \int_0^1 \Pr\{\mu > x\}\, dx \tag{89}$$

$$= \int_0^{s^*} \Pr\{\mu > x\}\, dx + \int_{s^*}^1 \Pr\{\mu > x\}\, dx. \tag{90}$$

Substituting (88) into (90) and evaluating both integrals gives

$$E[\mu] \le 1 - 2^{-\frac{B}{N_t-1}} + \frac{1}{N_t} 2^{-\frac{B}{N_t-1}}. \tag{91}$$

Substituting $E[\|\boldsymbol{h}\|^2] = N_t$ and (91) into (84) gives

$$E[I_{\mathcal{V}_{N_t}}^{\triangle}] \le \log\left(1 - 2^{-\frac{B}{N_t-1}} + \frac{1}{N_t} 2^{-\frac{B}{N_t-1}} + \frac{1}{\rho N_t}\right) \tag{92}$$

and taking the large system limit gives

$$\lim_{(N_t,B)\to\infty} E[I_{\mathcal{V}_{N_t}}^{\triangle}] \le \log(1 - 2^{-\bar{B}}). \tag{93}$$

Theorem 1 states that RVQ achieves this upper bound, and therefore upper bounds the asymptotic rate difference corresponding to any quantization scheme.

C. Proof of Theorem 3

We first prove the theorem for $\bar{N}_r \ge 1$. Let $z = F_{\gamma|\bar{\boldsymbol{H}}}^{-1}(1 - 2^{-B})$. Rearranging (15) gives

$$\lim_{\substack{(N_t,N_r)\to\infty \\ z \to \gamma_{\text{rvq}}^\infty}} \left[1 - F_{\gamma|\bar{\boldsymbol{H}}}(z)\right]^{\frac{1}{N_t}} = 2^{-\bar{B}}. \tag{94}$$

Next, we derive upper and lower bounds for the left-hand side of (94) and show that they are the same. The derivation of the upper bound is motivated by the evaluation of a similar bound for CDMA signature optimization in [34]. That is,

$$1 - F_{\gamma|\bar{\boldsymbol{H}}}(z) = \Pr\left\{\gamma_j > z | \bar{\boldsymbol{H}}\right\} \tag{95}$$

$$= \Pr\left\{\boldsymbol{v}_j^\dagger \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\dagger \boldsymbol{v}_j > z | \boldsymbol{\Lambda}, \boldsymbol{U}\right\} \tag{96}$$

$$= \Pr\left\{\frac{\boldsymbol{w}_j^\dagger \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\dagger \boldsymbol{w}_j}{\boldsymbol{w}_j^\dagger \boldsymbol{w}_j} > z \mid \boldsymbol{\Lambda}, \boldsymbol{U}\right\} \tag{97}$$

where $\gamma_j = \frac{1}{N_t} \boldsymbol{v}_j^\dagger \boldsymbol{H}^\dagger \boldsymbol{H} \boldsymbol{v}_j$ and we have applied the singular value decomposition $\frac{1}{N_t} \boldsymbol{H}^\dagger \boldsymbol{H} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\dagger$, where $\boldsymbol{U}$ is an $N_t \times N_t$ unitary matrix, $\boldsymbol{\Lambda} = \text{diag}\{\lambda_1, \cdots, \lambda_{N_t}\}$, and the eigenvalues are ordered as $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_{N_t}$. Also, $\boldsymbol{w}_j$

is an $N_t \times 1$ vector with independent, circularly symmetric, zero-mean and unit-variance Gaussian elements. Both v_j and $w_j/\|w_j\|$ are isotropically distributed, i.e., $Uw_j/\|w_j\|$ and $w_j/\|w_j\|$ have the same distribution, so that

$$1 - F_{\gamma|\bar{H}}(z) = 1 - F_{\gamma|\Lambda}(z) \tag{98}$$

$$= \Pr\left\{ \frac{\sum_{i=1}^{N_t} \lambda_i w_i^2}{\sum_{i=1}^{N_t} w_i^2} > z \middle| \Lambda \right\} \tag{99}$$

$$= \Pr\left\{ -\sum_{i=1}^{N_t} (z - \lambda_i) w_i^2 > 0 \middle| \Lambda \right\} \tag{100}$$

$$= \Pr\left\{ -\rho \sum_{i=1}^{N_t} (z - \lambda_i) w_i^2 > 0 \middle| \Lambda \right\}, \quad \forall \rho > 0 \tag{101}$$

$$= \Pr\left\{ \exp\left(-\rho \sum_{i=1}^{N_t} (z - \lambda_i) w_i^2 \right) > 1 \middle| \Lambda \right\} \tag{102}$$

where $\{w_i\}$ are elements of w_j. (We omit the index j to simplify the notation.) Applying Markov's inequality and the independence of the w_i's gives

$$1 - F_{\gamma|\Lambda}(z) \leq E_{\{w_i\}}\left[\exp\left\{ -\rho \sum_{i=1}^{N_t} (z - \lambda_i) w_i^2 \right\} \middle| \Lambda \right] \tag{103}$$

$$= \prod_{i=1}^{N_t} E_{w_i}\left[\exp\left\{ -\rho(z - \lambda_i) w_i^2 \right\} | \lambda_i \right] \tag{104}$$

$$= \prod_{i=1}^{N_t} \int_0^\infty \exp\left\{ -\rho(z - \lambda_i) x \right\} e^{-x} dx \tag{105}$$

$$= \prod_{i=1}^{N_t} \int_0^\infty \exp\left\{ -(1 + \rho(z - \lambda_i)) x \right\} dx \tag{106}$$

$$= \prod_{i=1}^{N_t} \frac{1}{1 + \rho(z - \lambda_i)} \tag{107}$$

$$= \exp\left\{ -\sum_{i=1}^{N_t} \log(1 + \rho(z - \lambda_i)) \right\} \tag{108}$$

when $1 + \rho(z - \lambda_i) > 0$ for all i, or $\rho < 1/(\lambda_1 - z)$. Taking the large system limit, we obtain

$$\lim_{\substack{(N_t, N_r) \to \infty \\ z \to \gamma_{\text{rvq}}^\infty}} \left[1 - F_{\gamma|\Lambda}(z) \right]^{\frac{1}{N_t}} \leq \exp\{-\Phi(\gamma_{\text{rvq}}^\infty, \rho)\} \tag{109}$$

for $0 < \rho < \frac{1}{\lambda_{\max}^\infty - \gamma_{\text{rvq}}^\infty}$, where

$$\Phi(\gamma_{\text{rvq}}^\infty, \rho) \triangleq \int_a^b \log(1 + \rho(\gamma_{\text{rvq}}^\infty - \lambda)) g(\lambda) d\lambda, \tag{110}$$

$g(\lambda)$ is given by (28)-(29), and $\lambda_{\max}^\infty = \lim_{(N_t, N_r) \to \infty} \lambda_1 = (1 + \sqrt{N_r})^2$. To tighten the upper bound, we minimize (109) with respect to ρ, i.e.,

$$\lim_{\substack{(N_t, N_r) \to \infty \\ z \to \gamma_{\text{rvq}}^\infty}} \left[1 - F_{\gamma|\Lambda}(z) \right]^{\frac{1}{N_t}} \leq \exp\{-\Phi(\gamma_{\text{rvq}}^\infty, \rho^*)\} \tag{111}$$

where

$$\rho^* = \arg \max_{0 < \rho < \frac{1}{\lambda_{\max}^\infty - \gamma_{\text{rvq}}^\infty}} \Phi(\gamma_{\text{rvq}}^\infty, \rho). \tag{112}$$

A similar expression for RVQ performance when used to quantize signatures for CDMA is derived in [34].

To derive the lower bound, we use a change of measure. (A similar approach was used in [45, Section 1.2].) Let $y_i \triangleq (\lambda_i - z) w_i^2$, which is a scaled exponential random variable with cdf $F_i(\cdot)$. We define the new distribution

$$G_i(x) \triangleq \frac{1}{M_i(\rho^*)} \int_{-\infty}^x e^{\rho^* y} dF_i(y), \tag{113}$$

so that

$$M_i(\rho^*) dG_i(x) = e^{\rho^* x} dF_i(x), \tag{114}$$

where ρ^* is given in (112), and the moment generating function for y_i is

$$M_i(\theta) \triangleq E[e^{\theta y_i}] \tag{115}$$

$$= \int_0^\infty e^{\theta(\lambda_i - z) x} e^{-x} dx \tag{116}$$

$$= \frac{1}{1 + \theta(z - \lambda_i)}. \tag{117}$$

Applying the change of measure (114), we have

$$1 - F_{\gamma|\Lambda}(z)$$

$$= \Pr\{\sum_{i=1}^{N_t} y_i > 0\} \tag{118}$$

$$= \int \cdots \int \mathbf{1}[\sum_{i=1}^{N_t} y_i > 0] dF_1(y_1) \cdots dF_{N_t}(y_{N_t}) \tag{119}$$

$$= \int \cdots \int \mathbf{1}[\sum_{i=1}^{N_t} y_i > 0] e^{-\rho^* \sum_{i=1}^{N_t} y_i} e^{\rho^* y_1} dF_1(y_1) \tag{120}$$

$$\cdots e^{\rho^* y_{N_t}} dF_{N_t}(y_{N_t})$$

$$= \prod_{i=1}^{N_t} M_i(\rho^*) \int \cdots \int \mathbf{1}[\sum_{i=1}^{N_t} y_i > 0] e^{-\rho^* \sum_{i=1}^{N_t} y_i} dG_1(y_1) \tag{121}$$

$$\cdots dG_{N_t}(y_{N_t})$$

where

$$\mathbf{1}[x > 0] = \left\{ \begin{array}{lll} 1 & : & x > 0 \\ 0 & : & x \leq 0 \end{array} \right. . \tag{122}$$

For any $\epsilon > 0$,

$$1 - F_{\gamma|\Lambda}(z)$$

$$\geq \prod_{i=1}^{N_t} M_i(\rho^*) \int \cdots \int \mathbf{1}[\epsilon N_t \geq \sum_{i=1}^{N_t} y_i > 0] e^{-\rho^* \sum_{i=1}^{N_t} y_i} \tag{123}$$

$$dG_1(y_1) \cdots dG_{N_t}(y_{N_t})$$

$$\geq \prod_{i=1}^{N_t} M_i(\rho^*) e^{-\rho^* \epsilon N_t} \int \cdots \int \mathbf{1}[\epsilon N_t \geq \sum_{i=1}^{N_t} y_i > 0] \tag{124}$$

$$dG_1(y_1) \cdots dG_{N_t}(y_{N_t})$$

$$= \prod_{i=1}^{N_t} M_i(\rho^*) e^{-\rho^* \epsilon N_t} \Pr\{\epsilon N_t \geq \sum_{i=1}^{N_t} \tilde{y}_i > 0\} \tag{125}$$

where the $\tilde{y}_i$'s are independent random variables with cdf $G_i(\cdot)$, and the second inequality follows since $\rho^* > 0$. To determine the probability on the right-hand side of (125), we first compute the mean of $\tilde{y}_i$,

$$m_i = \int y \, dG_i(y) \tag{126}$$

$$= \frac{1}{M_i(\rho^*)} \int y e^{\rho^* y} dF_i(y) \tag{127}$$

$$= (1 + \rho^*(z - \lambda_i)) \int_0^\infty (\lambda_i - z) x e^{\rho^*(\lambda_i - z)x} e^{-x} \, dx \tag{128}$$

$$= \frac{\lambda_i - z}{1 + \rho^*(z - \lambda_i)}. \tag{129}$$

Therefore

$$\frac{1}{N_t} \sum_{i=1}^{N_t} m_i = \frac{1}{N_t} \sum_{i=1}^{N_t} \frac{\lambda_i - z}{1 + \rho^*(z - \lambda_i)} \tag{130}$$

and the asymptotic mean

$$m_\infty = \lim_{\substack{(N_t, N_r) \to \infty \\ z \to \gamma_{\mathrm{rvq}}^\infty}} \frac{1}{N_t} \sum_{i=1}^{N_t} m_i \tag{131}$$

$$= \int_a^b \frac{\lambda - \gamma_{\mathrm{rvq}}^\infty}{1 + \rho^*(\gamma_{\mathrm{rvq}}^\infty - \lambda)} g(\lambda) \, d\lambda < \infty. \tag{132}$$

Similarly, since $\tilde{y}_i$ is exponentially distributed, the variance of $\tilde{y}_i$ is

$$\sigma_i^2 = \left(\frac{\lambda_i - z}{1 + \rho^*(z - \lambda_i)} \right)^2 < \infty \tag{133}$$

and the asymptotic variance

$$\sigma_\infty^2 = \lim_{\substack{(N_t, N_r) \to \infty \\ z \to \gamma_{\mathrm{rvq}}^\infty}} \frac{1}{N_t} \sum_{i=1}^{N_t} \sigma_i^2 \tag{134}$$

$$= \int_a^b \left(\frac{\lambda - \gamma_{\mathrm{rvq}}^\infty}{1 + \rho^*(\gamma_{\mathrm{rvq}}^\infty - \lambda)} \right)^2 g(\lambda) \, d\lambda < \infty. \tag{135}$$

Both the asymptotic mean and variance are finite.

Since the $\tilde{y}_i$'s are independent with finite mean and variance, the central limit theorem implies that the cdf for

$$T \triangleq \frac{\sum_{i=1}^{N_t} \tilde{y}_i - \sum_{i=1}^{N_t} m_i}{\sqrt{\sum_{i=1}^{N_t} \sigma_i^2}} \tag{136}$$

converges to a Gaussian cdf with zero mean and unit variance. Therefore we have

$$\mathrm{Pr}\{0 < \sum_{i=1}^{N_t} \tilde{y}_i \leq \epsilon N_t\}$$

$$= \mathrm{Pr}\{-\sqrt{N_t} a_{N_t} < T \leq -\sqrt{N_t} a_{N_t} + \sqrt{N_t} \epsilon \frac{1}{b_{N_t}}\} \tag{137}$$

$$= F_T(-\sqrt{N_t} a_{N_t} + \sqrt{N_t} \epsilon \frac{1}{b_{N_t}}) - F_T(-\sqrt{N_t} a_{N_t}) \tag{138}$$

where $F_T(\cdot)$ is the cdf for T and

$$a_{N_t} \triangleq \frac{\sum_{i=1}^{N_t} m_i / N_t}{\sqrt{\sum_{i=1}^{N_t} \sigma_i^2 / N_t}} \to \frac{m_\infty}{\sigma_\infty}, \tag{139}$$

$$b_{N_t} \triangleq \sqrt{\sum_{i=1}^{N_t} \sigma_i^2 / N_t} \to \sigma_\infty. \tag{140}$$

Let $\phi(\cdot)$ denote a Gaussian cdf with zero mean and unit variance. We can rewrite (138) as

$$\mathrm{Pr}\{0 < \sum_{i=1}^{N_t} \tilde{y}_i \leq \epsilon N_t\} = \phi(-\sqrt{N_t} a_{N_t} + \sqrt{N_t} \epsilon \frac{1}{b_{N_t}}) \tag{141}$$
$$- \phi(-\sqrt{N_t} a_{N_t}) + \zeta_{N_t} - \xi_{N_t}$$

where

$$\zeta_{N_t} \triangleq F_T(-\sqrt{N_t} a_{N_t} + \sqrt{N_t} \epsilon \frac{1}{b_{N_t}})$$
$$- \phi(-\sqrt{N_t} a_{N_t} + \sqrt{N_t} \epsilon \frac{1}{b_{N_t}}), \tag{142}$$

$$\xi_{N_t} \triangleq F_T(-\sqrt{N_t} a_{N_t}) - \phi(-\sqrt{N_t} a_{N_t}). \tag{143}$$

Applying the Berry-Esséen theorem [46], we can bound both ζ_{N_t} and ξ_{N_t} for large N_t as

$$|\zeta_{N_t}|, \quad |\xi_{N_t}| \leq \frac{C}{\sqrt{N_t}} \tag{144}$$

where C is a positive constant that depends on the variance and third moment of $\tilde{y}$. Similar to the mean and variance, we can show that the third moment is also finite.

We can now evaluate

$$\phi(-\sqrt{N_t} a_{N_t} + \sqrt{N_t} \epsilon \frac{1}{b_{N_t}}) - \phi(-\sqrt{N_t} a_{N_t})$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\sqrt{N_t} a_{N_t}}^{-\sqrt{N_t} a_{N_t} + \sqrt{N_t} \epsilon \frac{1}{b_{N_t}}} e^{-t^2/2} \, dt \tag{145}$$

$$\leq \frac{1}{\sqrt{2\pi}} e^{-N_t(a_{N_t} - \epsilon/b_{N_t})^2} \sqrt{N_t} \epsilon \frac{1}{b_{N_t}}. \tag{146}$$

Substituting (144) and (146) into (141), we have

$$\mathrm{Pr}\{0 < \sum_{i=1}^{N_t} \tilde{y}_i \leq \epsilon N_t\} = O(1/\sqrt{N_t}) \tag{147}$$

Taking the large system limit and applying L'Hopital's rule, it follows that

$$\lim_{\substack{(N_t, N_r) \to \infty \\ z \to \gamma_{\mathrm{rvq}}^\infty}} [\mathrm{Pr}\{0 < \sum_{i=1}^{N_t} \tilde{y}_i \leq \epsilon N_t\}]^{\frac{1}{N_t}} = 1. \tag{148}$$

Taking the N_tth root and large system limit on both sides of (125) gives

$$\lim_{\substack{(N_t, N_r) \to \infty \\ z \to \gamma_{\mathrm{rvq}}^\infty}} [1 - F_{\gamma | \Lambda}(z)]^{\frac{1}{N_t}} \geq \exp\{-\Phi(\gamma_{\mathrm{rvq}}^\infty, \rho^*)\} \tag{149}$$

where we use (148) and let $\epsilon \to 0$.

The lower bound in (149) is exactly the upper bound (111). Therefore,

$$\lim_{\substack{(N_t, N_r) \to \infty \\ z \to \gamma_{\mathsf{rvq}}^\infty}} \left[1 - F_{\gamma|\mathbf{\Lambda}}(z)\right]^{\frac{1}{N_t}} = \exp\{-\Phi(\gamma_{\mathsf{rvq}}^\infty, \rho^*)\} \quad (150)$$

$$= 2^{-\bar{B}} \quad (151)$$

and the asymptotic RVQ received power satisfies the fixed-point equation

$$\Phi(\gamma_{\mathsf{rvq}}^\infty, \rho^*) = \bar{B}\log(2) \quad (152)$$

where ρ^* is given by (112). The goal of the rest of the proof is to simplify (152).

To determine ρ^*, we first compute

$$\frac{\partial \Phi(\gamma_{\mathsf{rvq}}^\infty, \rho)}{\partial \rho} = \int_a^b \left[\frac{\gamma_{\mathsf{rvq}}^\infty - \lambda}{1 + (\gamma_{\mathsf{rvq}}^\infty - \lambda)\rho}\right] g(\lambda) \, d\lambda \quad (153)$$

$$= \frac{1}{\rho} + \frac{1}{\rho^2} \int_a^b \frac{1}{\underbrace{\lambda - \left(\frac{1}{\rho} + \gamma_{\mathsf{rvq}}^\infty\right)}_{y}} g(\lambda) \, d\lambda \quad (154)$$

$$= \frac{1}{\rho} + \frac{1}{\rho^2} \mathcal{S}_{\mathbf{\Lambda}}(y) \quad (155)$$

where $\mathcal{S}_{\mathbf{\Lambda}}(\cdot)$ is the Stieltjés Transform of the asymptotic eigenvalue distribution of $\mathbf{\Lambda}$. Setting the derivative to zero and solving for ρ gives

$$\mathcal{S}_{\mathbf{\Lambda}}(y) = -\rho = \frac{1}{\gamma_{\mathsf{rvq}}^\infty - y} \quad (156)$$

as the only valid solution. Substituting the expression for $\mathcal{S}_{\mathbf{\Lambda}}$ given in [33] into (156) gives

$$\frac{(-1 + \bar{N}_r - y) \pm \sqrt{y^2 - 2(\bar{N}_r + 1)y + (\bar{N}_r - 1)^2}}{2y}$$
$$= \frac{1}{\gamma_{\mathsf{rvq}}^\infty - y}, \quad (157)$$

which simplifies to the quadratic equation

$$(\bar{N}_r - \gamma_{\mathsf{rvq}}^\infty)y^2 + [\gamma_{\mathsf{rvq}}^\infty + \bar{N}_r\gamma_{\mathsf{rvq}}^\infty + (\gamma_{\mathsf{rvq}}^\infty)^2]y = 0. \quad (158)$$

Solving for y gives $y = 0$ or $y = \gamma_{\mathsf{rvq}}^\infty[1 + 1/(\gamma_{\mathsf{rvq}}^\infty - \bar{N}_r)]$, or equivalently, $\rho = -1/\gamma_{\mathsf{rvq}}^\infty$ or $\rho = (\gamma_{\mathsf{rvq}}^\infty - \bar{N}_r)/\gamma_{\mathsf{rvq}}^\infty$. Since $\rho > 0$, we must have

$$\rho^* = \frac{\gamma_{\mathsf{rvq}}^\infty - \bar{N}_r}{\gamma_{\mathsf{rvq}}^\infty}. \quad (159)$$

Since

$$\frac{\partial^2 \Phi(\gamma_{\mathsf{rvq}}^\infty, \rho)}{\partial \rho^2} = -\int_a^b \left(\frac{\lambda - \gamma_{\mathsf{rvq}}^\infty}{1 + \rho^*(\gamma_{\mathsf{rvq}}^\infty - \lambda)}\right)^2 g(\lambda) \, d\lambda \quad (160)$$

$$< 0, \quad (161)$$

therefore ρ^* achieves a maximum.

By also evaluating $\Phi(\gamma_{\mathsf{rvq}}^\infty, \rho)$ at the boundary points $\rho = 0$ and $\rho = 1/(\lambda_{\max}^\infty - \gamma_{\mathsf{rvq}}^\infty)$, we have

$$\rho^*$$
$$= \begin{cases} \frac{\gamma_{\mathsf{rvq}}^\infty - \bar{N}_r}{\gamma_{\mathsf{rvq}}^\infty}, & \bar{N}_r \leq \gamma_{\mathsf{rvq}}^\infty \leq \bar{N}_r + \sqrt{\bar{N}_r} \\ \frac{1}{(1+\sqrt{\bar{N}_r})^2 - \gamma_{\mathsf{rvq}}^\infty}, & \bar{N}_r + \sqrt{\bar{N}_r} \leq \gamma_{\mathsf{rvq}}^\infty < (1 + \sqrt{\bar{N}_r})^2 \end{cases}. \quad (162)$$

To evaluate $\Phi(\gamma_{\mathsf{rvq}}^\infty, \rho^*)$, we re-write (110) as

$$\Phi(\gamma_{\mathsf{rvq}}^\infty, \rho^*)$$
$$= \int_a^b \left[\log(\rho^*) + \log\left(\left(\frac{1}{\rho^*} + \gamma_{\mathsf{rvq}}^\infty\right) - \lambda\right)\right] g(\lambda) \, d\lambda \quad (163)$$

$$= \log(\rho^*) + \int_a^b \log\left(\left(\frac{1}{\rho^*} + \gamma_{\mathsf{rvq}}^\infty\right) - \lambda\right) g(\lambda) \, d\lambda. \quad (164)$$

To evaluate the integral in (164), we apply the following Lemma.

Lemma 1. *For* $x \geq (1 + \sqrt{\bar{N}_r})^2$,

$$\Theta(x) \triangleq \int_a^b \log(x - \lambda)g(\lambda) \, d\lambda \quad (165)$$

$$= \log(w(x)) + \sqrt{\bar{N}_r}u(x) - (\bar{N}_r - 1)\log\left(1 + \frac{u(x)}{\sqrt{\bar{N}_r}}\right) \quad (166)$$

where

$$w(x) = \frac{(x - 1 - \bar{N}_r) + \sqrt{(x - 1 - \bar{N}_r)^2 - 4\bar{N}_r}}{2}, \quad (167)$$

$$u(x) = \frac{(x - 1 - \bar{N}_r) - \sqrt{(x - 1 - \bar{N}_r)^2 - 4\bar{N}_r}}{2\sqrt{\bar{N}_r}}. \quad (168)$$

The proof of this Lemma is similar to that given in [36] and is therefore omitted here.

For $\bar{N}_r + \sqrt{\bar{N}_r} \leq \gamma_{\mathsf{rvq}}^\infty < (1 + \sqrt{\bar{N}_r})^2$, we substitute $\rho^* = [(1 + \sqrt{\bar{N}_r})^2 - \gamma_{\mathsf{rvq}}^\infty]^{-1}$ into (164) to obtain

$$\Phi(\gamma_{\mathsf{rvq}}^\infty, [(1 + \sqrt{\bar{N}_r})^2 - \gamma_{\mathsf{rvq}}^\infty]^{-1}) \quad (169)$$

$$= -\log[(1 + \sqrt{\bar{N}_r})^2 - \gamma_{\mathsf{rvq}}^\infty] + \Theta\left((1 + \sqrt{\bar{N}_r})^2\right) \quad (170)$$

$$= -\log[(1 + \sqrt{\bar{N}_r})^2 - \gamma_{\mathsf{rvq}}^\infty] + \frac{1}{2}\bar{N}_r\log(\bar{N}_r)$$
$$- (\bar{N}_r - 1)\log(1 + \sqrt{\bar{N}_r}) + \sqrt{\bar{N}_r} \quad (171)$$

$$= \bar{B}\log(2). \quad (172)$$

Solving for $\gamma_{\mathsf{rvq}}^\infty$ gives (18). Taking $\gamma_{\mathsf{rvq}}^\infty = \bar{N}_r + \sqrt{\bar{N}_r}$ and solving for $\bar{B}$ gives $\bar{B}^*$ in (19).

For $\bar{N}_r \leq \gamma_{\mathsf{rvq}}^\infty < \bar{N}_r + \sqrt{\bar{N}_r}$, or $0 \leq \bar{B} \leq \bar{B}^*$, we substitute $\rho^* = \frac{\gamma_{\mathsf{rvq}}^\infty - \bar{N}_r}{\gamma_{\mathsf{rvq}}^\infty}$ into (164) to obtain

$$\Phi\left(\gamma_{\mathsf{rvq}}^\infty, \frac{\gamma_{\mathsf{rvq}}^\infty - \bar{N}_r}{\gamma_{\mathsf{rvq}}^\infty}\right) = \log(\gamma_{\mathsf{rvq}}^\infty - \bar{N}_r) - \log(\gamma_{\mathsf{rvq}}^\infty)$$
$$+ \Theta\left(\gamma_{\mathsf{rvq}}^\infty + \frac{\gamma_{\mathsf{rvq}}^\infty}{\gamma_{\mathsf{rvq}}^\infty - \bar{N}_r}\right). \quad (173)$$

To simplify (173), we let $\psi \triangleq \gamma_{\mathsf{rvq}}^\infty - \bar{N}_r$ and re-write (173) as

$$\Phi\left(\psi - \bar{N}_r, \frac{\psi}{\psi + \bar{N}_r}\right) = \log(\psi) - \log(\psi + \bar{N}_r)$$
$$+ \Theta\left(1 + \bar{N}_r + \psi + \frac{\bar{N}_r}{\psi}\right). \quad (174)$$

After some manipulation we have

$$w\left(1 + \bar{N}_r + \psi + \frac{\bar{N}_r}{\psi}\right) = \frac{\bar{N}_r}{\psi}, \quad (175)$$

$$u\left(1 + \bar{N}_r + \psi + \frac{\bar{N}_r}{\psi}\right) = \frac{\psi}{\sqrt{\bar{N}_r}}, \quad (176)$$

and

$$\Theta\left(1 + \bar{N}_r + \psi + \frac{\bar{N}_r}{\psi}\right) = \log(\bar{N}_r) - \log(\psi)$$
$$- (\bar{N}_r - 1)\log\left(1 + \frac{\psi}{\bar{N}_r}\right) + \psi. \quad (177)$$

Substituting (177) into (174), we obtain

$$\Phi\left(\psi - \bar{N}_r, \frac{\psi}{\psi + \bar{N}_r}\right)$$
$$= \psi - \bar{N}_r \log\left(1 + \frac{\psi}{\bar{N}_r}\right) \quad (178)$$
$$= \gamma_{\text{rvq}}^\infty - \bar{N}_r - \bar{N}_r \log(\gamma_{\text{rvq}}^\infty) + \bar{N}_r \log(\bar{N}_r). \quad (179)$$

Setting this to $\bar{B}\log(2)$ and simplifying gives (17).

For $\bar{N}_r < 1$, the asymptotic eigenvalue density of $\frac{1}{N_t}H^\dagger H$ is given by

$$g(\lambda) = (1 - \bar{N}_r)\delta(\lambda) + \frac{\sqrt{(\lambda - a)(b - \lambda)}}{2\pi\lambda}. \quad (180)$$

where a and b are given by (28)-(29). Following the same steps again from (110) gives (18) and (17). This completes the proof of Theorem 3.

D. Derivation of (41)-(43)

To compute μ_J, we first write

$$J_j^{N_r} = \frac{1}{N_r}\sum_{k=1}^{N_r}\log\left(1 + \rho\frac{\bar{N}_r}{K}v_k\right) \quad (181)$$

where v_k is the kth eigenvalue of $\Upsilon = \frac{1}{N_r}HV_jV_j^\dagger H^\dagger$. As $(N_t, N_r, K) \to \infty$, the empirical eigenvalue distribution converges to a deterministic function $F_\Upsilon(t)$. The asymptotic mean is given by

$$\mu_J = \lim_{(N_t,N_r,K)\to\infty} E[J_j^{N_r}] = \int_0^\infty \log\left(1 + \rho\frac{\bar{N}_r}{K}t\right)\mathrm{d}F_\Upsilon(t). \quad (182)$$

A similar integral has been evaluated in [36, Eq. (6)], and the result can be directly applied to (182), giving (41).

To compute the variance, we express $J_j^{N_r}$ differently by first performing the singular value decomposition $H = V_H\Sigma_H U_H^\dagger$, where V_H is the $N_r \times N_r$ left singular matrix, U_H is the $N_t \times N_r$ right singular matrix, and Σ_H is an $N_r \times N_r$ diagonal matrix. Here we assume that $N_t \geq N_r$. (The result for $N_t < N_r$ can be shown by a similar approach.) We therefore have

$$J_j^{N_r} = \frac{1}{N_r}\log\det(I_{N_r} + \rho\Lambda L_j) \quad (183)$$
$$= \frac{1}{N_r}\sum_{i=1}^{N_r}\log(1 + \rho\eta_i) \quad (184)$$

where $\Lambda = \frac{1}{K}\Sigma_H^2$, $L_j = U_H^\dagger V_j V_j^\dagger U_H$, and η_i is the ith eigenvalue of ΛL_j. To compute $\text{var}[J_j^{N_r}]$, correlations between pairs of η_i's are needed. Although the joint distribution of eigenvalues is known, it is complicated, so that computing the variance appears intractable.

To approximate the variance of $J_j^{N_r}$, we substitute a Taylor series expansion for $\log(1 + \delta x)$ into (184) to write

$$J_j^{N_r} = \frac{\rho}{N_r}\sum_{i=1}^{N_r}\eta_i - \frac{\rho^2}{2N_r}\sum_{i=1}^{N_r}\eta_i^2 + \frac{\rho^3}{3N_r}\sum_{i=1}^{N_r}\eta_i^3 + \ldots \quad (185)$$
$$= \frac{\rho}{N_r}\text{tr}\{\Lambda L\} - \frac{\rho^2}{2N_r}\text{tr}\{(\Lambda L)^2\} + \frac{\rho^3}{3N_r}\text{tr}\{(\Lambda L)^3\} + \ldots \quad (186)$$

for $\rho\eta_{\max} < 1$, where $\eta_{\max} = \max_i \eta_i$, and is the maximum eigenvalue of $HVV^\dagger H^\dagger/K$. Since HV is $N_r \times K$ and *i.i.d.*, $\eta_{\max}$ has asymptotic value $(1 + \sqrt{N_r/K})^2$. If $\bar{K}/\bar{N}_r = 1$, then the condition asymptotically becomes $\rho < 1/4$ (-6 dB). Ignoring the terms of order ρ^3 and higher, we can approximate the variance of $J_j^{N_r}$ at low SNR as

$$\text{var}[J_j^{N_r}] \approx \rho^2\text{var}\left[\frac{1}{N_r}\text{tr}\{\Lambda L\}\right]. \quad (187)$$

Letting $\Lambda = \text{diag}\{\lambda_i\}$ and l_{ij} denote the (i,j)th element of L, the first term in (187) can be expanded as

$$\text{var}[\text{tr}\{\Lambda L\}|\Lambda] = \sum_{i=1}^{N_r}\lambda_i^2\left(E[l_{ii}^2] - E^2[l_{ii}]\right)$$
$$+ \sum_{i\neq j}\lambda_i\lambda_j\left(E[l_{ii}l_{jj}] - E[l_{ii}]E[l_{jj}]\right). \quad (188)$$

For a given U_H and random unitary V with $K = N_r$, Theorem 3 in [7] states that L has a multivariate beta distribution with parameters N_r and $N_t - N_r$. (The distribution of L is not known for general K.) From Theorem 2 in [47], we have

$$E[l_{ii}] = \frac{N_r + 1}{N_t + 2} \quad (189)$$
$$E[l_{ii}^2] = \frac{(N_r + 1)(N_r + 3)}{(N_t + 2)(N_t + 4)} \quad (190)$$
$$E[l_{ii}l_{jj}] = \frac{N_r(N_r + 1)(N_t + 4) + (N_r + 1)(N_t - N_r + 1)}{(N_t + 1)(N_t + 2)(N_t + 4)},$$
$$i \neq j, \quad (191)$$

for $1 \leq i, j \leq N_r$. Substituting (189)-(191) into (188) gives

$$\text{var}[\text{tr}\{\Lambda L\}|\Lambda] = \left(\frac{1}{N_r}\sum_{i=1}^{N_r}\lambda_i^2\right)\left(\bar{N}_r^2(1 - \bar{N}_r) + O\left(\frac{1}{N_r}\right)\right)$$
$$+ \left(\frac{1}{N_r^2}\sum_{i\neq j}\lambda_i\lambda_j\right)\left((\bar{N}_r - 1)\bar{N}_r^3 + O\left(\frac{1}{N_r}\right)\right). \quad (192)$$

Taking expectation with respect to Λ, and the large system limit, we have

$$E_\Lambda\left(\text{var}[\text{tr}\{\Lambda L\}|\Lambda]\right) \to 1 - \bar{N}_r. \quad (193)$$

Also, in the large system limit

$$\frac{1}{N_r}\sum_{i=1}^{N_r}\lambda_i^2 \to \int t^2\mathrm{d}F_\Lambda(t) = \frac{1}{\bar{N}_r}\left(1 + \frac{1}{\bar{N}_r}\right) \quad (194)$$
$$\frac{1}{N_r^2}\sum_{i\neq j}\lambda_i\lambda_j \to \left[\int t\mathrm{d}F_\Lambda(t)\right]^2 = \frac{1}{\bar{N}_r^2} \quad (195)$$

where $F_{\boldsymbol{\Lambda}}(t)$ is the asymptotic distribution for the diagonal elements of $\boldsymbol{\Lambda}$ or, equivalently, the asymptotic eigenvalue distribution of $\boldsymbol{H}\boldsymbol{H}^{\dagger}/N_r$.

Substituting (193) into (187), we have

$$
\begin{aligned}
\sigma_J^2 &= \lim_{(N_r, N_t) \to \infty} N_r^2 \mathrm{var}[J_j^{N_r}] \qquad (196) \\
&\approx \rho^2 (1 - \bar{N}_r). \qquad (197)
\end{aligned}
$$

ACKNOWLEDGEMENT

The authors thank the anonymous reviewers for their detailed comments and for pointing out mistakes in the proofs of Theorems 2 and 3, which appeared in an earlier draft.

REFERENCES

[1] 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.

[2] 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.

[3] 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.

[4] ——, "Limited feedback unitary precoding for spatial multiplexing systems," *IEEE Trans. Info. Theory*, vol. 51, no. 8, pp. 2967–2976, Aug. 2005.

[5] 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.

[6] 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.

[7] ——, "Design and analysis of MIMO spatial multiplexing systems with quantized feedback," *IEEE Trans. Signal Processing*, vol. 54, no. 8, pp. 2874–2886, Aug. 2006.

[8] 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 Info. Theory (ISIT)*, Adelaide, Australia, Sept. 2005, pp. 1549–1553.

[9] W. Dai, Y. Liu, and B. Rider, "Quantization bounds on Grassmann manifolds and applications to MIMO communications," *IEEE Trans. Inf. Theory*, vol. 54, no. 3, pp. 1108–1123, Mar. 2008.

[10] S. Zhou, Z. Wang, and G. B. Giannakis, "Quantifying the power-loss when transmit-beamforming relies on finite rate feedback," *IEEE Trans. Wireless Commun.*, vol. 4, no. 4, pp. 1948–1957, July 2005.

[11] P. Xia and G. B. Giannakis, "Design and analysis of transmit-beamforming based on limited-rate feedback," *IEEE Trans. Signal Processing*, vol. 54, no. 5, pp. 1853–1863, May 2006.

[12] K. K. Mukkavilli, A. Sabharwal, and B. Aazhang, "Generalized beamforming for MIMO systems with limited transmitter information," in *Proc. Asilomar Conf. on Signals, Systems, and Computers*, vol. 1, Pacific Grove, CA, Nov. 2003, pp. 1052–1056.

[13] G. Jöngren, M. Skoglund, and B. Ottersten, "Combining beamforming and orthogonal space-time block coding," *IEEE Trans. Info. Theory*, vol. 48, no. 3, pp. 611–625, Mar. 2002.

[14] M. Skoglund and G. Jöngren, "On the capacity of a multiple-antenna communication link with channel side information," *IEEE J. Select. Areas Commun.*, vol. 21, no. 3, pp. 395–405, Apr. 2003.

[15] E. Visotsky and U. Madhow, "Space-time transmit precoding with imperfect feedback," *IEEE Trans. Info. Theory*, vol. 47, no. 6, pp. 2632–2639, Sept. 2001.

[16] 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.

[17] 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.

[18] S. A. Jafar and A. J. Goldsmith, "Transmitter optimization and optimality of beamforming for multiple antenna systems," *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp. 1165–1175, July 2004.

[19] P. Zador, "Asymptotic quantization error of continuous signals and the quantization dimension," *IEEE Trans. Info. Theory*, vol. 28, no. 2, pp. 139–149, Mar. 1982.

[20] R. M. Gray and D. L. Neuhoff, "Quantization," *IEEE Trans. Info. Theory*, vol. 44, no. 6, pp. 2325–2383, Oct. 1998.

[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] 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.

[23] S. Bhashyam, A. Sabharwal, and B. Aazhang, "Feedback gain in multiple antenna systems," *IEEE Trans. Commun.*, vol. 50, no. 5, pp. 785–798, May 2002.

[24] V. K. N. Lau, Y. Liu, and T.-A. Chen, "Role of transmit diversity for wireless communications – reverse link analysis with partial feedback," *IEEE Trans. Commun.*, vol. 50, no. 12, pp. 2082–2090, Dec. 2002.

[25] C. K. Au-Yeung and D. J. Love, "On the performance of random vector quantization limited feedback beamforming," *IEEE Trans. Wireless Commun.*, vol. 6, no. 2, pp. 458–462, Feb. 2005.

[26] N. Jindal, "MIMO broadcast channels with finite-rate feedback," *IEEE Trans. Info. Theory*, vol. 52, no. 11, pp. 5045–5060, Nov. 2006.

[27] T. Yoo, N. Jindal, and A. Goldsmith, "Multi-antenna downlink channels with limited feedback and user selection," *IEEE J. Select. Areas Commun.*, vol. 25, no. 7, pp. 1478–1491, Sep. 2007.

[28] M. Sharif and B. Hassibi, "On the capacity of MIMO broadcast channel with partial side information," *IEEE Trans. Info. Theory*, vol. 51, no. 2, pp. 506–522, Feb. 2005.

[29] P. Viswanath, D. N. C. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Trans. Info. Theory*, vol. 48, no. 6, pp. 1277–1294, June 2002.

[30] D. N. C. Tse and O. Zeitouni, "Linear multiuser receivers in random environments," *IEEE Trans. Info. Theory*, vol. 46, no. 1, pp. 171–188, Jan. 2000.

[31] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels," *IEEE Trans. Info. Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.

[32] J. Galambos, *The Asymptotic Theory of Extreme Order Statistics*, 2nd ed. Robert E. Krieger, 1987.

[33] V. A. Marčenko and L. A. Pastur, "Distribution of eigenvalues for some sets of random matrices," *Math. USSR-Sbornik*, vol. 1, pp. 457–483, 1967.

[34] W. Dai, Y. Liu, and B. Rider, "Performance analysis of CDMA signature optimization with finite rate feedback," in *Proc. Conf. on Info. Sciences and Systems (CISS)*, Princeton, NJ, Mar. 2006.

[35] İ. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. on Telecommun.*, vol. 10, pp. 585–595, Nov. 1999.

[36] P. B. Rapajic and D. Popescu, "Information capacity of a random signature multiple-input multiple-output channel," *IEEE Trans. Commun.*, vol. 48, no. 8, pp. 1245–1248, Aug. 2000.

[37] Z. D. Bai and J. W. Silverstein, "CLT for linear spectral statistics of large dimensional sample covariance matrices," *Annals of Probability*, vol. 32, no. 1A, pp. 553–605, 2004.

[38] W. Santipach and M. L. Honig, "Signature optimization for DS-CDMA with limited feedback," in *Proc. IEEE Int. Symp. on Spread-Spectrum Tech. and Appl. (ISSSTA)*, Prague, Czech Republic, Sept. 2002, pp. 180–184.

[39] D. J. Ryan, I. V. L. Clarkson, I. B. Collings, D. Guo, and M. L. Honig, "QAM and PSK codebooks for limited feedback MIMO beamforming," to appear in *IEEE Trans. on Commun.*, Feb. 2009.

[40] V. Raghavan, R. W. Heath, Jr., and A. M. Sayeed, "Systematic codebook designs for quantized beamforming in correlated MIMO channels," *IEEE J. Select. Areas Commun.*, vol. 25, no. 7, pp. 1298–1310, Sept. 2006.

[41] W. Santipach and M. L. Honig, "Capacity of beamforming with limited training and feedback," in *Proc. IEEE Int. Symp. on Info. Theory (ISIT)*, Seattle, WA, July 2006.

[42] ——, "Optimization of training and feedback for beamforming over a MIMO channel," in *Proc. IEEE Wireless Commun. and Networking Conf. (WCNC)*, Hong Kong, China, Mar. 2007.

[43] Y. Sun and M. L. Honig, "Asymptotic capacity of multicarrier transmission with frequency-selective fading and limited feedback," *IEEE Trans. Info. Theory*, vol. 54, no. 7, pp. 2879–2902, July 2008.

[44] E. Castillo, *Extreme Value Theory in Engineering*. Academic Press, 1988.

[45] A. Shwartz and A. Weiss, *Large Deviations for Performance Analysis. Queues, Communications, and Computing*. London, UK: Chapman & Hall, 1995.

[46] J. K. Patel and C. B. Read, *Handbook of the Normal Distribution*, 2nd ed., ser. Statistics: a Series of Textbooks and Monographs. New York: Marcel Dekker, 1996.

[47] C. G. Khatri and K. C. S. Pillai, "Some results on the non-central multivariate Beta distribution and moments of traces of two matrices," *Annals of Mathematical Statistics*, vol. 36, no. 5, pp. 1511–1520, Oct. 1965.

Wiroonsak Santipach (S'00-M'06) received the B.S. (*summa cum laude*), M.S., and Ph.D. degrees all in electrical engineering from Northwestern University, Illinois, USA in 2000, 2001, and 2006, respectively.

He is currently a lecturer at the Department of Electrical Engineering, Faculty of Engineering, Kasetsart University in Bangkok, Thailand. His research interests are in wireless communications, and include performance evaluation of CDMA and MIMO system.

Michael L. Honig (S'80-M'81-SM'92-F'97) received the B.S. degree in electrical engineering from Stanford University in 1977, and the M.S. and Ph.D. degrees in electrical engineering from the University of California, Berkeley, in 1978 and 1981, respectively. He subsequently joined Bell Laboratories in Holmdel, NJ, where he worked on local area networks and voiceband data transmission. In 1983 he joined the Systems Principles Research Division at Bellcore, where he worked on Digital Subscriber Lines and wireless communications. Since the Fall of 1994, he has been with Northwestern University where he is a Professor in the Electrical Engineering and Computer Science Department. He has held visiting scholar positions at the Technical University of Munich, Princeton University, the University of California, Berkeley, Naval Research Laboratory (San Diego), and the University of Sydney. He has also worked as a free-lance trombonist.

Dr. Honig has served as an editor for the IEEE Transactions on Information Theory (1998-2000), the IEEE Transactions on Communications (1990-1995), and was a guest editor for the European Transactions on Telecommunications and Wireless Personal Communications. He has also served as a member of the Digital Signal Processing Technical Committee for the IEEE Signal Processing Society, and as a member of the Board of Governors for the Information Theory Society (1997-2002). He is the recipient of a Humboldt Research Award for Senior U.S. Scientists, and the co-recipient of the 2002 IEEE Communications Society and Information Theory Society Joint Paper Award.