# Capacity of a Multi-Antenna Fading Channel with a Quantized Precoding Matrix

Wiroonsak Santipach and Michael L. Honig

**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. With beamforming RVQ is asymptotically optimal, i.e., no other quantization scheme can achieve a larger asymptotic rate. The performance of RVQ is also compared with that of a simpler reduced-rank scalar quantization scheme in which the beamformer is constrained to lie in a random subspace. 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**

Multiple antennas, limited feedback, beamforming, precoding, vector quantization, large system limit, reduced-rank optimization.

This work was supported by the U.S. Army Research Office under grant DAAD19-99-1-0288 and the National Science Foundation under grant CCR-0310809, and was presented in part at IEEE MILCOM, Boston, MA, Oct. 2003, IEEE ISIT, Chicago, IL, June 2004, and IEEE ISSSTA, Sydney, Australia, Aug. 2004.

W. Santipach was with the Department of Electrical Engineering and Computer Science; Northwestern University, Evanston, IL 60208 USA. He is currently with the Department of Electrical Engineering; Faculty of Engineering; Kasetsart University, Bangkok, Thailand, email: `wiroonsak.s@ku.ac.th`.

M. L. Honig is with the Department of Electrical Engineering and Computer Science; Northwestern University, Evanston, IL 60208 USA; Tel/Fax: (847) 491-7803/467-3550, email: `mh@eecs.northwestern.edu`.

# 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 is motivated by related work on quantized signature optimization for Code-Division Multiple Access (CDMA) [19]. 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 [19] 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 [19], RVQ is also asymptotically optimal in this scenario, i.e., no other quantization scheme can achieve a larger asymptotic rate.[1]

We then consider quantization of a precoding matrix with arbitrary rank. That is, 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

[1]Numerical examples in [20] also show that RVQ can achieve the performance of optimized (Grassmannian) codebooks in [3] for small systems.

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).

Although generating an RVQ codebook is simple, selecting the codebook entry that optimizes performance generally requires an exhaustive search over all codebook entries. The complexity therefore grows exponentially with the number of feedback bits. We therefore compare RVQ performance with that of a simpler *reduced-rank* quantization scheme, motivated by a similar scheme proposed for CDMA signature quantization [19]. For a rank-one precoding matrix, the beamforming vector is constrained to lie in a lower dimensional (random) subspace, which is known *a priori* at both the transmitter and receiver. The receiver optimizes the beamformer within the subspace, and relays the *scalar-quantized* coefficients back to the transmitter. Numerical results show that for a fixed number of quantization bits this reduced-rank scalar quantization scheme performs substantially better than direct scalar quantization of the beamforming coefficients. For precoding matrices with arbitrary rank this scheme can be applied to each column of the precoding matrix. Although this technique performs substantially better than scalar quantization of each precoding matrix element, there is a substantial degradation in performance relative to RVQ.

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 [21], [22]). 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 [23]–[25]. Namely, exact expressions for the ergodic capacity with beamforming and RVQ for a finite-size MISO channel are derived in [23]. The performance of RVQ for precoding over a broadcast MIMO channel is analyzed in [24], [25]. A closely related random beamforming scheme for the multiuser MIMO broadcast channel was previously presented in [26]. In that work, the growth in sum capacity is characterized asymptotically as the number of users becomes large with a *fixed* number of antennas.

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

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

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 $n_t$th transmit antenna and the $n_r$th 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

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

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 $\|\boldsymbol{v}_k\| \le 1$, for $1 \le k \le K$.

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 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 [27], 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 $E_{\boldsymbol{h}}[\log(1+\rho(\boldsymbol{h}^\dagger \boldsymbol{h})^2)]$. The receiver selects the quantized precoding vector to maximize the mutual information, i.e.,

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

and the corresponding achievable rate is

$$I_{\mathsf{rvq}}^{N_t} \triangleq \max_{1 \le j \le 2^B} I_j. \tag{4}$$

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 $I_{\textsf{rvq}}^{N_t}$ 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, $I_{\textsf{rvq}}^{N_t}$ converges to a deterministic constant.

As $N_t \to \infty$, $(\boldsymbol{h}^\dagger \boldsymbol{h})/N_t \to 1$ almost surely, so that $\log[1 + \rho(\boldsymbol{h}^\dagger \boldsymbol{h})^2] - \log(\rho N_t) \to 0$. That is, with perfect channel 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

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

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

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

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

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

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

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

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

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 $\{\mathcal{V}_{N_t}\}$, $\limsup_{(N_t, B) \to \infty} I_{\mathcal{V}_{N_t}}^{\triangle} \le \mathcal{I}_{\textsf{rvq}}^{\triangle}$.*

*Proof:* Since $\boldsymbol{v}_j$ is a unit-norm vector, maximizing $I_j$ in (3) is equivalent to minimizing $\|\boldsymbol{h} - \boldsymbol{v}_j\|^2$ for a given channel vector $\boldsymbol{h}$. This is a vector quantization problem for an *i.i.d.*

Gaussian source with squared Euclidean distance as the distortion measure. From Shannon's source coding theorem [28], the minimum expected distortion achieved with bit rate $\bar{B}$ is $2^{-\bar{B}}$ as $(B, N_t) \to \infty$. Theorem 1 states that RVQ achieves this average distortion and, therefore, upper bounds the asymptotic mutual information corresponding to any quantization scheme. ■

Although the optimality of RVQ holds only in the large system limit, numerical results in [20] show that for finite-size systems of interest RVQ can perform 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, 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

$$\begin{aligned} I_{\mathsf{rvq}}^{N_t} &= E_{\mathcal{V}}\left[\max_{1\le j\le 2^B} \log\left(1+\rho\|\boldsymbol{H}\boldsymbol{v}_j\|^2\right)\middle|\,\boldsymbol{H}\right] && (9)\\ &= E_{\mathcal{V}}\left[\log(1+\rho \max_{1\le j\le 2^B}\|\boldsymbol{H}\boldsymbol{v}_j\|^2)\middle|\,\boldsymbol{H}\right]. && (10)\end{aligned}$$

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

$$I_{\mathsf{rvq}}^{\triangle} \triangleq I_{\mathsf{rvq}}^{N_t} - \log(\rho N_t) = E_{\mathcal{V}}\left[\log\left(\frac{1}{\rho N_t} + \max_j \gamma_j\right)\middle|\,\boldsymbol{H}\right] \tag{11}$$

where

$$\gamma_j = \frac{1}{N_t}\boldsymbol{v}_j^\dagger \boldsymbol{H}^\dagger \boldsymbol{H}\boldsymbol{v}_j. \tag{12}$$

Evaluating the expectation in (11) 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 $N_t \times N_r$ 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

$$\gamma_{\mathsf{rvq}}^{\infty} = \lim_{(N_t,N_r,B)\to\infty} \left[ \max_{1\le j\le 2^B} \gamma_j \middle| \bar{\boldsymbol{H}} \right] \tag{13}$$

where convergence to the deterministic limit can be shown in the mean square sense. Conditioned on $\bar{\boldsymbol{H}}$, the $\gamma_j$'s are *i.i.d.* since the beamforming vectors $\boldsymbol{v}_j$ are *i.i.d.*, and applying [29, Theorem 2.1.5], it can be shown that

$$\gamma_{\mathsf{rvq}}^{\infty} = \lim_{(N_t,N_r,B)\to\infty} F_{\gamma|\bar{\boldsymbol{H}}}^{-1}\left(1-2^{-B}\right) \tag{14}$$

where $F_{\gamma|\bar{\boldsymbol{H}}}(\cdot)$ is the cdf of $\gamma_j$ given $\bar{\boldsymbol{H}}$. Analogous results for the interference power in CDMA with quantized signatures have been presented in [19], so that we omit the proofs of (13) and (14). 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 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 $\frac{1}{N_t}\boldsymbol{H}^\dagger\boldsymbol{H}$ [30]. The asymptotic rate difference is given by

$$\mathcal{I}_{\mathsf{rvq}}^{\Delta} = \lim_{(N_t,N_r,B)\to\infty} I_{\mathsf{rvq}}^{\Delta} = \log(\gamma_{\mathsf{rvq}}^{\infty}) \tag{15}$$

The limit in (14) can be explicitly evaluated, using the expression for $F_{\gamma|\boldsymbol{H}}$ in [31], 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*

$$\left(\gamma_{\mathsf{rvq}}^{\infty}\right)^{\bar{N}_r} \mathrm{e}^{-\gamma_{\mathsf{rvq}}^{\infty}} = 2^{-\bar{B}} \left(\frac{\bar{N}_r}{\mathrm{e}}\right)^{\bar{N}_r} \tag{16}$$

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

$$\gamma_{\mathsf{rvq}}^{\infty} = (1+\sqrt{\bar{N}_r})^2 - \exp\left\{\frac{1}{2}\bar{N}_r\log(\bar{N}_r) - (\bar{N}_r-1)\log(1+\sqrt{\bar{N}_r}) + \sqrt{\bar{N}_r} - \bar{B}\log(2)\right\} \tag{17}$$

*where*

$$\bar{B}^* = \frac{1}{\log(2)}\left(\bar{N}_r \log\left(\frac{\sqrt{\bar{N}_r}}{1+\sqrt{\bar{N}_r}}\right) + \sqrt{\bar{N}_r}\right). \tag{18}$$

The proof is given in Appendix B and is motivated by an analogous result for CDMA, presented in [32]. 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* $\bar{B} = B/N_t$*,*

$$\limsup_{(N_t,N_r,B)\to\infty} I^{N_t}_{\mathcal{V}_{N_t}} - \log(\rho N_t) \leq \mathcal{I}^{\Delta}_{\textsf{rvq}} \tag{19}$$

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

The proof is similar to the proof of Theorem 2 in [19] and is therefore omitted.

### C. Numerical Results

Figs. 1 and 2 show $\mathcal{I}^{\triangle}_{\textsf{rvq}}$ for MISO and MIMO channels with beamforming versus normalized feedback bits ($\bar{B}$) with $\rho = 5$ and 10 dB, respectively. 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 $4 \times 6$ and $16 \times 24$ channels. The asymptotic results accurately predict the performance for the larger channel, and are somewhat less accurate for the smaller channel. The capacity with perfect beamforming, corresponding to unlimited feedback, is also shown. 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 beamforming capacity with infinite feedback is nearly achieved with two feedback bits per complex coefficient.

RVQ requires that for each channel realization, the receiver perform an exhaustive search over the beamforming codebook. Hence the complexity of this scheme increases exponentially with the number of feedback bits. We therefore include in Fig. 3 a comparison of RVQ performance with a simpler reduced-rank (RR) beamforming scheme, analogous to that proposed in [19] for CDMA signature optimization. Namely, the RR beamforming vector is constrained to lie in a $D$-dimensional subspace spanned by the columns of an $N_t \times D$ partial unitary random matrix $\boldsymbol{F}$. We therefore have $\boldsymbol{v} = \boldsymbol{F}\boldsymbol{a}$, where $\boldsymbol{a}$ is is the $D \times 1$ vector of combining coefficients, and the matrix $\boldsymbol{F}$ is assumed to be known at both the transmitter and receiver. It is straightforward to show that the optimal $\boldsymbol{a}$, which maximizes the SINR, is the eigenvector of $\boldsymbol{F}^\dagger \boldsymbol{H}^\dagger \boldsymbol{H} \boldsymbol{F}$ corresponding to the maximum eigenvalue.

The receiver computes the optimal RR beamforming vector for a given channel and quantizes the real and imaginary parts of $\boldsymbol{a}$ independently with the same scalar quantizer optimized using the Lloyd-Max algorithm [28]. Based on numerical simulations, we assume that both real and

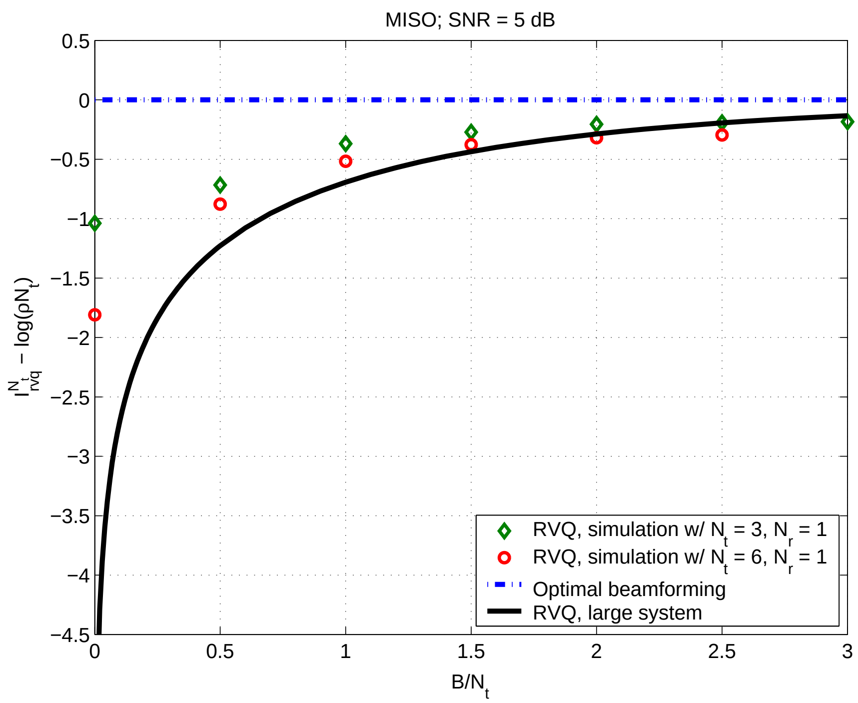



Fig. 1. Asymptotic and simulated rate differences versus feedback bits for a MISO channel with beamforming.

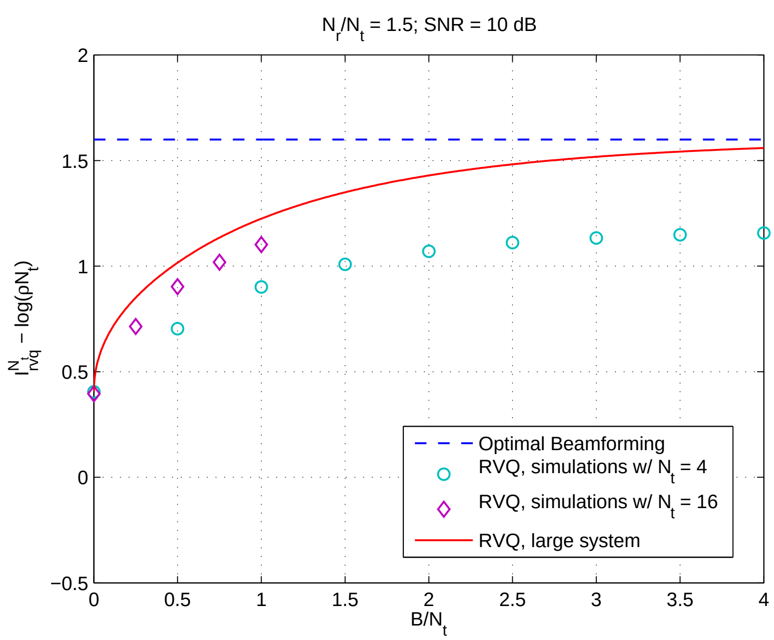



Fig. 2. Asymptotic and simulated rate differences versus feedback bits for a MIMO channel with beamforming ($\bar{N}_r = 1.5$).

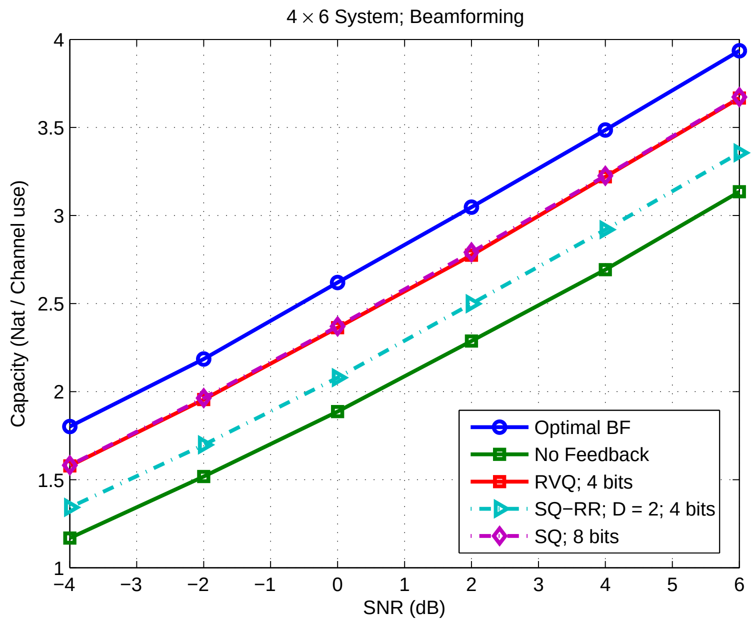



Fig. 3. Capacity of quantized beamformers versus SNR with $N_t = 4$ and $N_r = 6$. SQ-RR refers to a scalar-quantized reduced-rank beamformer.

imaginary parts have a Gaussian distribution with zero mean and variance $1/(2D)$. The number of bits is allocated equally among all coefficients. In comparison with RVQ, this scheme is suboptimal, but avoids the search over codebook entries.

Fig. 3 shows capacity versus SNR with $N_t = 4$ and $N_r = 6$. Results are shown for a RR beamformer with dimension $D = 2$, and RVQ with four-bit quantization. Also shown are results for the RR beamformer with eight-bit quantization (four bits per complex coefficient). Results for optimal beamforming and no feedback are also shown. Relative to optimal beamforming, RVQ has a loss of about 1.2 dB, the RR beamformer with four bits has a loss of about three dB, and no feedback loses about four dB. With eight-bit quantization, the performance of the RR scheme matches that of RVQ.

## 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

$$\mathcal{V} = \{\boldsymbol{V}_j, 1 \leq j \leq 2^B\}, \tag{20}$$

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

$$J_j^{N_r} = \frac{1}{N_r} \log\det\left(\boldsymbol{I}_{N_r} + \frac{\rho}{K}\boldsymbol{H}\boldsymbol{V}_j\boldsymbol{V}_j^\dagger\boldsymbol{H}^\dagger\right), \tag{21}$$

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

$$\hat{\boldsymbol{V}} = \arg \max_{1\leq j\leq 2^B} J_j^{N_r}. \tag{22}$$

For finite $N_r$, we define

$$\begin{aligned} I_{\mathsf{rvq}}^{N_r} &= E_{\mathcal{V}}[\max_{1\leq j\leq 2^B} J_j^{N_r} | \bar{\boldsymbol{H}}] \qquad (23)\\ &= E_{\mathcal{V}}\left[\frac{1}{N_r}\log\det\left(\boldsymbol{I}_{N_r} + \frac{\rho}{K}\boldsymbol{H}\hat{\boldsymbol{V}}\hat{\boldsymbol{V}}^\dagger\boldsymbol{H}^\dagger\right)\middle| \bar{\boldsymbol{H}}\right] \qquad (24)\end{aligned}$$

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

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 $I_{\mathsf{rvq}}^{N_r}$ can be evaluated, in analogy with (14), 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 $K = N_t$ [33]. 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

$$\frac{1}{N_r} \det \log \left( \boldsymbol{I}_{N_r} + \frac{\rho}{N_t} \boldsymbol{H}\boldsymbol{H}^\dagger \right) \to \int_0^\infty \log\left(1 + \rho\lambda\right) g(\lambda)\, \mathrm{d}\lambda \tag{25}$$

$$= \mathcal{I}_{\mathsf{rvq}}(\hat{B} = 0) \tag{26}$$

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 [30]

$$g(\lambda) = \frac{\sqrt{(\lambda - a)(b - \lambda)}}{2\pi\lambda\bar{N}_r} \quad \text{for} \quad a \le \lambda \le b, \tag{27}$$

$$a = \left(1 - \sqrt{\bar{N}_r}\right)^2 \quad \text{and} \quad b = \left(1 + \sqrt{\bar{N}_r}\right)^2. \tag{28}$$

The integral in (25) has been evaluated in [34], which gives the closed-form expression

$$\mathcal{I}_{\mathsf{rvq}}(\hat{B} = 0) = \log \rho y + \frac{1 - \bar{N}_r}{\bar{N}_r} \log \left( \frac{1}{1 - z} \right) - \frac{z}{\bar{N}_r} \tag{29}$$

where

$$y = \frac{1}{2} \left( 1 + \bar{N}_r + \frac{1}{\rho} + \sqrt{\left(1 + \bar{N}_r + \frac{1}{\rho}\right)^2 - 4\bar{N}_r} \right) \tag{30}$$

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

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

$$\mathcal{I}_{\mathsf{rvq}}(\hat{B} = \infty) = \int_\eta^\infty \log\left(1 + \frac{\rho}{\bar{K}}\lambda\right) g(\lambda)\, \mathrm{d}\lambda \tag{32}$$

where $\eta$ satisfies

$$\int_\eta^\infty g(\lambda)\, \mathrm{d}\lambda = \min\{1, \frac{\bar{K}}{\bar{N}_r}\}. \tag{33}$$

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 (32) with respect to $\bar{K}$. It can be verified that $\mathcal{I}_{\mathsf{rvq}}(\hat{B}=0) \leq \mathcal{I}_{\mathsf{rvq}}(\hat{B}=\infty)$.

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 $\rho$, we can expand

$$\mathcal{I}_{\mathsf{rvq}}(\hat{B}=0) = \log(\rho) + o(\log(\rho)) \tag{34}$$

$$\mathcal{I}_{\mathsf{rvq}}(\hat{B}=\infty) = \log(\rho)\int_{\eta}^{\infty} g(\lambda)\,\mathrm{d}\lambda + o(\log(\rho)). \tag{35}$$

Therefore

$$\begin{aligned} \lim_{\rho\to\infty} \frac{\mathcal{I}_{\mathsf{rvq}}(\hat{B}=\infty)}{\mathcal{I}_{\mathsf{rvq}}(\hat{B}=0)} &= \int_{\eta}^{\infty} g(\lambda)\,\mathrm{d}\lambda && (36)\\ &= \min\{1, \frac{\bar{K}}{\bar{N}_r}\} && (37)\end{aligned}$$

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

$$\begin{aligned} \lim_{\rho\to 0} \frac{\mathcal{I}_{\mathsf{rvq}}(\hat{B}=\infty)}{\mathcal{I}_{\mathsf{rvq}}(\hat{B}=0)} &= \frac{1}{\bar{K}}\int_{\eta}^{\infty} \lambda g(\lambda)\,\mathrm{d}\lambda && (38)\\ &\leq \frac{1}{\bar{N}_r}\frac{\int_{\eta}^{\infty}\lambda g(\lambda)\,\mathrm{d}\lambda}{\int_{\eta}^{\infty} g(\lambda)\,\mathrm{d}\lambda} && (39)\\ &\leq \left(1+\frac{1}{\sqrt{\bar{N}_r}}\right)^2. && (40)\end{aligned}$$

The inequality (39) follows from (33), which implies $\bar{K} \geq \bar{N}_r \int_{\eta}^{\infty} g(\lambda)\,\mathrm{d}\lambda$. Differentiating the upper bound (39) with respect to normalized rank $\bar{K}$ shows that the optimal rank $\bar{K}^* = 0$. Applying L'Hôpital's rule to take the limit $\bar{K}\to 0$ in (39) gives the upper bound (40). That is, for small SNR allocating all transmission power to the strongest channel mode, or beamforming, maximizes the capacity. The maximal rate ratio (40) 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 $\bar{B} = \infty$.

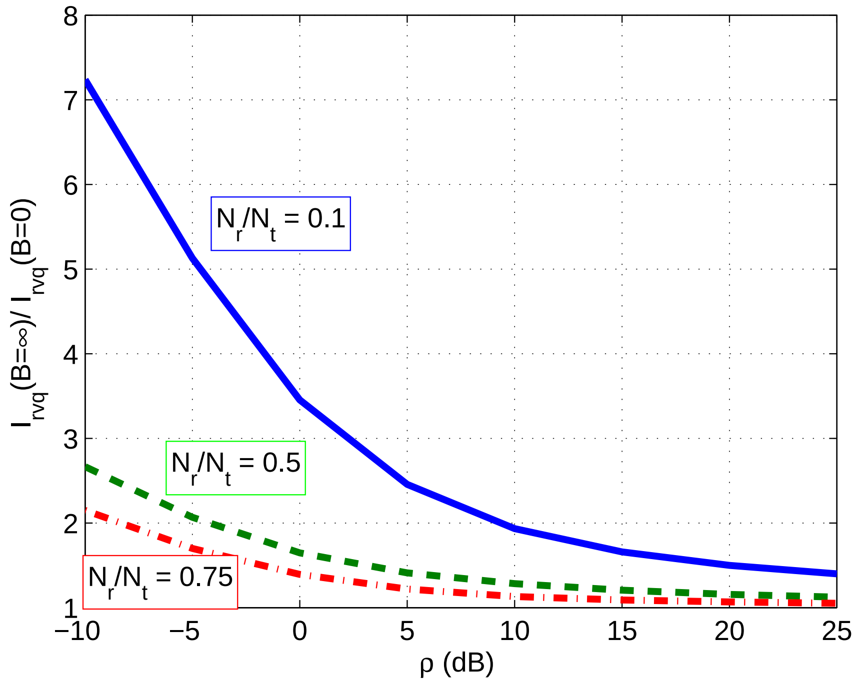



Fig. 4. The rate ratio $\mathcal{I}_{\text{rvq}}(\hat{B} = \infty)/\mathcal{I}_{\text{rvq}}(\hat{B} = 0)$ versus SNR (dB) for various values of $\bar{N}_r$.

To evaluate the asymptotic capacity with arbitrary $\hat{B}$, we approximate $J_j^{N_r}$ given $\bar{\boldsymbol{H}}$ as a Gaussian random variable. This has been shown to be valid in the large system limit when $\boldsymbol{HV}$ is replaced by a matrix with *i.i.d.* elements [35]. Conditioned on $\boldsymbol{H}$, the elements of $\boldsymbol{HV}$ are not *i.i.d.*; however, numerical examples indicate that this assumption is still valid for large $N_t$ and $N_r$.

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

$$\tilde{\mathcal{I}}_{\text{rvq}} = \mu_J + \sigma_J \sqrt{2\hat{B}\log 2} \tag{41}$$

independent of the channel realization, where $\mu_J$ and $\sigma_J^2$ are the asymptotic mean and variance of $J_j^{N_r}$. The derivation of (41) is a straightforward extension of [29, 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 (32) and (33). 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 in Appendix C. The asymptotic mean is given by

$$\mu_J = \frac{\bar{K}}{\bar{N}_r} \log\left(1 + \frac{\bar{N}_r}{\bar{K}}\rho - \frac{\bar{N}_r}{\bar{K}}\rho v\right) + \log\left(1 + \rho - \frac{\bar{N}_r}{\bar{K}}\rho v\right) - v \tag{42}$$

where

$$v = \frac{1}{2} + \frac{\bar{K}}{2\bar{N}_r} + \frac{\bar{K}}{2\bar{N}_r\rho} - \frac{1}{2}\sqrt{\left(1 + \frac{\bar{K}}{\bar{N}_r} + \frac{\bar{K}}{\bar{N}_r\rho}\right)^2 - \frac{4\bar{K}}{\bar{N}_r}}. \tag{43}$$

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

$$\sigma_J^2 \approx \rho^2(1 - \bar{N}_r)(1 - 2\rho). \tag{44}$$

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

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 $\mathcal{I}^{\triangle}_{\mathsf{rvq}}$. Since here we are unable to evaluate $\mathcal{I}_{\mathsf{rvq}}$ exactly, we cannot apply that argument. Nevertheless, numerical results have indicated that the performance of RVQ matches that of optimized codebooks (e.g., see [20]).

Fig. 5 shows $\tilde{\mathcal{I}}_{\mathsf{rvq}}$ 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 $\mathcal{I}_{\mathsf{rvq}}(\hat{B} = \infty)$, which is computed from (32) with optimized $\bar{K}$. The asymptotic rate with RVQ is computed from (41), where $\sigma_J$ for $\rho = -5\,$dB is approximated by (44), and $\sigma_J$ is determined from simulation with $N_t = 20$ for $\rho = 0$ and $5$ dB. Also shown in Fig. 5 are simulation results for $\mathcal{I}_{\mathsf{rvq}}$ 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 $\hat{B} \le 0.8$. The asymptotic results accurately approximate the simulated results shown. The accuracy increases as the SNR and feedback $\hat{B}$ both decrease.

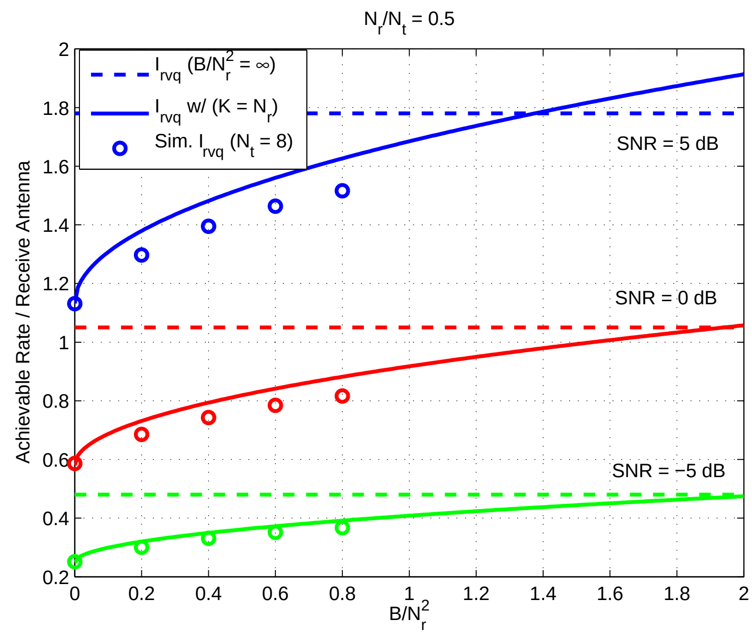



Fig. 5. Sum mutual information per receive antenna with RVQ and an optimal receiver versus normalized feedback. The asymptotic approximation is shown along with Monte Carlo simulation results for an $8 \times 4$ channel.

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 (41) with $\bar{N}_r = 0.2$, $\rho = 5$ dB, and different values of $\hat{B}$. ($\sigma_J$ 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 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

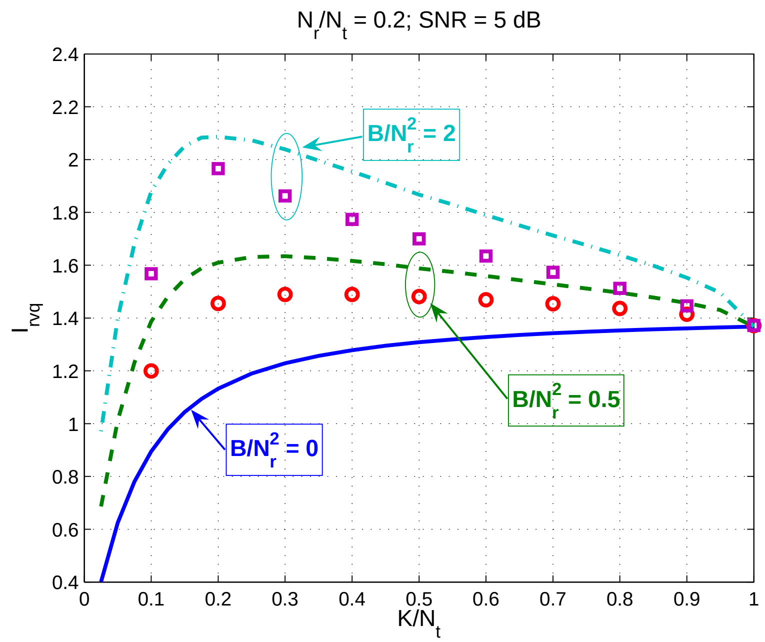



Fig. 6. Mutual information per receive antenna versus normalized rank with different normalized feedback. Discrete points correspond to simulation with $N_t = 10$.

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 filter $\boldsymbol{c}_k$. The matched filter is given by

$$\boldsymbol{c}_k = \frac{1}{\sqrt{K}} \boldsymbol{H} \boldsymbol{v}_k \tag{45}$$

where $\boldsymbol{v}_k$ is the $k$th column of the precoding matrix $\boldsymbol{V}$, and the linear MMSE filter is given by

$$\boldsymbol{c}_k = \frac{1}{\sqrt{K}} \left( \frac{1}{K} \boldsymbol{H} \boldsymbol{V} \boldsymbol{V}^\dagger \boldsymbol{H}^\dagger + \sigma_n^2 \boldsymbol{I}_{N_r} \right)^{-1} \boldsymbol{H} \boldsymbol{v}_k. \tag{46}$$

The SINR at the output of the linear filter $\boldsymbol{c}_k$ is

$$\mathsf{SINR}_k = \frac{|\boldsymbol{c}_k^\dagger \boldsymbol{H} \boldsymbol{v}_k|^2}{\boldsymbol{c}_k^\dagger \left( \sum_{i \neq k} \boldsymbol{H} \boldsymbol{v}_i \boldsymbol{v}_i^\dagger \boldsymbol{H}^\dagger + K \sigma_n^2 \boldsymbol{I}_{N_r} \right) \boldsymbol{c}_k}. \tag{47}$$

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 $\boldsymbol{c}_k$, 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

$$R^{N_r} = \frac{1}{N_r}\sum_{k=1}^{K} I(x_k, \hat{x}_k) \tag{48}$$

$$= \frac{1}{N_r}\sum_{k=1}^{K} \log(1+\gamma_k). \tag{49}$$

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

Given the codebook of precoding matrices $\mathcal{V} = \{\boldsymbol{V}_j,\ 1 \le j \le 2^B\}$, the receiver selects the precoding matrix

$$\hat{\boldsymbol{V}} = \arg\max_{1\le j\le 2^B} R^{N_r}(\boldsymbol{V}_j). \tag{50}$$

We again consider RVQ, in which the $\boldsymbol{V}_j$'s are *i.i.d.* unitary matrices.

### A. Matched filter

Substituting (45) into (47), the SINR at the output of the matched filter is given by

$$\gamma_{k;\mathsf{mf}} = \frac{(\boldsymbol{v}_k^\dagger \boldsymbol{H}^\dagger \boldsymbol{H}\boldsymbol{v}_k)^2}{K\sigma_n^2(\boldsymbol{v}_k^\dagger \boldsymbol{H}^\dagger \boldsymbol{H}\boldsymbol{v}_k) + \sum_{i=1,i\neq k}^{K} |\boldsymbol{v}_k^\dagger \boldsymbol{H}^\dagger \boldsymbol{H}\boldsymbol{v}_i|^2} \tag{51}$$

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

$$E_{\boldsymbol{H},\mathcal{V}}\left[\max_{1\le j\le 2^B}\{R_{\mathsf{mf}}^{N_r}(\boldsymbol{V}_j) = \frac{1}{N_r}\sum_{k=1}^{K}\log(1+\gamma_{k;\mathsf{mf}})\}\right] \tag{52}$$

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 (52). Motivated by the central limit theorem,[2] in

[2]The terms in the sum in (52) are not *i.i.d.*, which prevents a direct application of the central limit theorem.

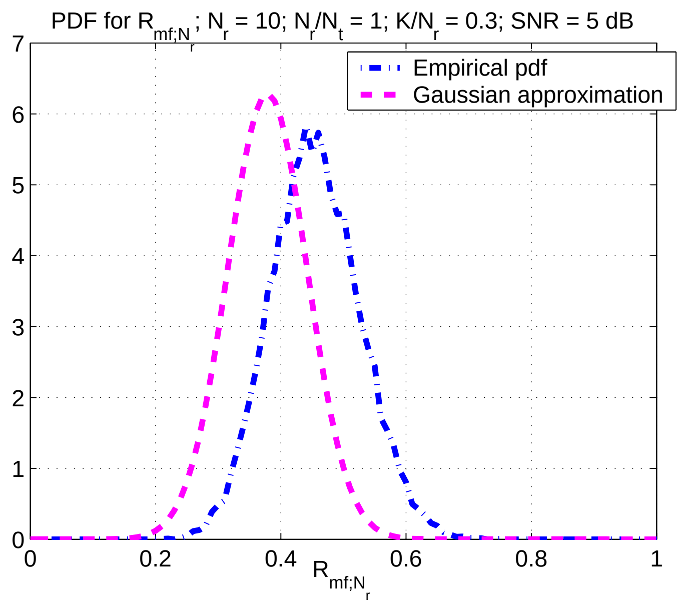



Fig. 7. Comparison of the empirical pdf for $R_{\mathsf{mf};N_r}$ with the Gaussian approximation.

what follows we approximate the cdf of $R_{\mathsf{mf}}^{N_r}$ as Gaussian. The mean is taken to be the asymptotic limit

$$\mu_{\mathsf{mf}} = \lim_{(N_t,N_r,K)\to\infty} R_{\mathsf{mf}}^{N_r} = \frac{\bar{K}}{\bar{N}_r}\log\left(1+\frac{\bar{N}_r}{\bar{K}(1+\sigma_n^2)}\right). \tag{53}$$

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 $\bar{K}$. We show in Appendix D that

$$N_r^2\mathrm{var}[R_{\mathsf{mf}}^{N_r}|\boldsymbol{V}_j] \to \sigma_{\mathsf{mf}}^2 = \frac{\bar{N}_r^2+\bar{K}\bar{N}_r\,(4+6\sigma_n^2+3\sigma_n^4)-2\bar{K}\bar{N}_r^2}{\bar{K}^2(1+\sigma_n^2)^2(1+\sigma_n^2+\bar{N}_r/\bar{K})^2}. \tag{54}$$

Given a codebook $\mathcal{V}$, we approximate the variance of $R_{\mathsf{mf}}^{N_r}$ as $\sigma_{\mathsf{mf}}^2/N_r^2$.

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 $\mathrm{SNR} = 5\,\mathrm{dB}$. The estimated variance is quite close to the empirical value. 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 [29] to evaluate the large system limit

$$\mathcal{R}_{\mathsf{rvq;mf}} = \lim_{(N_t,N_r,K,B)\to\infty}[\max_{1\le j\le 2^B} R_{\mathsf{mf}}^{N_r}(\boldsymbol{V}_j)|\mathcal{V}]. \tag{55}$$

Given $\mathcal{V}$, the sum rates $\{R_{\mathsf{mf}}^{N_r}(\boldsymbol{V}_1),\ldots,R_{\mathsf{mf}}^{N_r}(\boldsymbol{V}_{2^B})\}$ are identically distributed. However, the $R_{\mathsf{mf}}^{N_r}(\boldsymbol{V}_j)$'s are not independent since each depends on $\boldsymbol{H}$. 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 $R_{\mathsf{mf}}^{N_r}(\boldsymbol{V}_j)$, $j = 1, \cdots, 2^B$, with *i.i.d.* Gaussian variables with mean $\mu_{\mathsf{mf}}$ 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

$$\tilde{\mathcal{R}}_{\mathsf{rvq;mf}} = \mu_{\mathrm{mf}} + \sigma_{\mathsf{mf}}\sqrt{2\hat{B}\log 2}. \tag{56}$$

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 $\hat{B} \to \infty$, $\tilde{\mathcal{R}}_{\mathsf{rvq;mf}} \to \infty$, whereas $\mathcal{R}_{\mathsf{rvq;mf}}$ with $\hat{B} = \infty$ is the same as the asymptotic rate with RVQ and an optimal receiver, given by (32) and (33). Hence $\mathcal{R}_{\mathsf{rvq;mf}}$ 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 $R_{\mathsf{mf}}^{N_r}$, which has compact support, is being approximated by a Gaussian cdf with infinite support, and also because the dependence among the sum rates $R_{\mathsf{mf}}^{N_r}(\boldsymbol{V}_j)$ is being ignored.

### *B. MMSE receiver*

Substituting (46) into (47) gives the SINR at the output of MMSE receiver for the $k$th symbol stream

$$\gamma_{k;\mathsf{mmse}} = \boldsymbol{v}_k^\dagger \boldsymbol{H}^\dagger \left( \sum_{i \neq k} \boldsymbol{H}\boldsymbol{v}_i\boldsymbol{v}_i^\dagger \boldsymbol{H}^\dagger + K\sigma_n^2 \boldsymbol{I}_{N_r} \right)^{-1} \boldsymbol{H}\boldsymbol{v}_k. \tag{57}$$

As $(N_r, N_t) \to \infty$, the distribution of $\gamma_{k;\mathsf{mmse}}$ converges to a Gaussian distribution [27], i.e.,

$$\sqrt{N_r}(\gamma_{k;\mathsf{mmse}} - \gamma_{\mathsf{mmse}}) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \sigma_\gamma^2\right) \tag{58}$$

where the large system SINR

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

and

$$\sigma_\gamma^2 = \frac{2\gamma_{\mathsf{mmse}}(1 + \gamma_{\mathsf{mmse}})^2}{\sigma_n^2(1 + \gamma_{\mathsf{mmse}})^2 + \bar{K}/\bar{N}_r} - \gamma_{\mathsf{mmse}}^2. \tag{60}$$

Hence $N_r \text{var}\left[\gamma_{k;\mathsf{mmse}}\right] \to \sigma_\gamma^2$ as $(N_t, N_r, K) \to \infty$. As for the matched filter receiver, given a codebook $\mathcal{V}$, we approximate the pdf of the instantaneous sum rate

$$R_{\mathsf{mmse}}^{N_r} = \frac{1}{N_r} \sum_{k=1}^{K} \log(1 + \gamma_{k;\mathsf{mmse}}) \tag{61}$$

as a Gaussian pdf with mean

$$\mu_{\mathsf{mmse}} = \lim_{(N_t, N_r)\to\infty} R_{\mathsf{mmse}}^{N_r} = \frac{\bar{K}}{\bar{N}_r} \log(1 + \gamma_{\mathsf{mmse}}). \tag{62}$$

It is shown in Appendix D that the variance satisfies

$$\frac{1}{N_r^2}\sigma_\gamma^2 = \frac{\bar{K}\bar{N}_r\sigma_{\mathsf{mmse}}^2 - \bar{K}[\mathrm{d}\gamma_{\mathsf{mmse}}/\mathrm{d}(\bar{K}/\bar{N}_r)]^2}{\bar{N}_r^2(1+\gamma_{\mathsf{mmse}})^2 N_r^2} \tag{63}$$

where

$$\frac{\mathrm{d}\gamma_{\mathsf{mmse}}}{\mathrm{d}(\bar{K}/\bar{N}_r)} = -\frac{1}{2\sigma_n^2} + \frac{1-(1-\bar{K}/\bar{N}_r)/\sigma_n^2}{8\sqrt{(1-\bar{K}/\bar{N}_r)^2 + 2\sigma_n^2(1+\bar{K}/\bar{N}_r) + \sigma_n^4}}. \tag{64}$$

In analogy with (56), the asymptotic rate with RVQ and the MMSE receiver is given by

$$\mathcal{R}_{\mathsf{rvq;mmse}} \approx \tilde{\mathcal{R}}_{\mathsf{rvq;mmse}} = \mu_{\mathsf{mmse}} + \sigma_{\mathsf{mmse}}\sqrt{2\hat{B}\log 2}. \tag{65}$$

As for the matched filter receiver, when $\hat{B}$ is large, $\tilde{\mathcal{R}}_{\mathsf{mmse}}$ over-estimates $\mathcal{R}_{\mathsf{mmse}}$. For $\hat{B} = \infty$, $\mathcal{R}_{\mathsf{mmse}} = \mathcal{R}_{\mathsf{mf}} = \mathcal{I}_{\mathsf{rvq}}$ with the optimal receiver, given by (32).

### *C. Reduced-Rank (RR) Precoding with Scalar Quantization*

Because of the high complexity associated with RVQ with moderate to large feedback, the numerical results in the next section include a comparison with a simpler RR precoding scheme. To extend the RR beamformer described in Section III to a precoding matrix with arbitrary rank, we constrain each $N_t \times 1$ column of the precoding matrix, $\boldsymbol{v}_k$, $k = 1, \cdots, K$, to lie in a $D$-dimensional subspace ($D \le N_t$). We therefore have

$$\boldsymbol{v}_k = \boldsymbol{F}_k \boldsymbol{\alpha}_k \tag{66}$$

where $\boldsymbol{F}_k$ is an $N_t \times D$ matrix of orthogonal basis vectors (known to the transmitter) and $\boldsymbol{\alpha}_k$ is a $D \times 1$ vector of combining coefficients. The $DK$ coefficients in $\boldsymbol{\alpha}_1, \cdots, \boldsymbol{\alpha}_K$ are then *scalar*-quantized and relayed to the transmitter.

For a general precoding matrix the performance depends on the set $\mathcal{F} = \{\boldsymbol{F}_k,\ 1 \le k \le K\}$. We select each $\boldsymbol{F}_k$ to be a random partial unitary matrix ($\boldsymbol{F}_k^\dagger \boldsymbol{F}_k = \boldsymbol{I}_D$). If $KD \le N_t$, then we can design the set such that the subspaces associated with each precoding matrix are orthogonal, i.e., $\boldsymbol{F}_k^\dagger \boldsymbol{F}_l = 0$ for all $k \neq l$. If $KD > N_t$, then we group all matrices in $\mathcal{F}$ into $\lceil KD/N_t \rceil$ subsets with at most $\lfloor N_t/D \rfloor$ matrices in each subset. The matrices within each subset are then chosen to be orthogonal.

To compute the set of $\boldsymbol{\alpha}_k$'s, which maximizes the sum instantaneous mutual information, we follow the iterative algorithm in [36]. Namely, for the MMSE receiver, we replace $\boldsymbol{\alpha}_k$ with the eigenvector of $\boldsymbol{F}_k^\dagger \boldsymbol{H}^\dagger \left( \sum_{i \neq k} \boldsymbol{H} \boldsymbol{v}_i \boldsymbol{v}_i^\dagger \boldsymbol{H}^\dagger + K\sigma_n^2 \boldsymbol{I}_{N_r} \right)^{-1} \boldsymbol{H} \boldsymbol{F}_k$ corresponding to the maximum eigenvalue. We do this for $k = 1$ to $k = K$, compute $R_{\mathsf{mmse}}^{N_r}$, and iterate until $R_{\mathsf{mmse}}^{N_r}$ converges.

For the matched filter, computing the jointly optimal $\boldsymbol{\alpha}_k$'s is more involved [36]. Here we construct a near-optimal set, which after quantization gives essentially the same performance. Namely, we approximate the optimization step in [36] by replacing each $\boldsymbol{\alpha}_k$, $k = 1, \cdots, K$, with the eigenvector of $\boldsymbol{F}_k^\dagger \boldsymbol{H}^\dagger \left( \sum_{i \neq k} \boldsymbol{H} \boldsymbol{v}_i \boldsymbol{v}_i^\dagger \boldsymbol{H}^\dagger \right) \boldsymbol{H} \boldsymbol{F}_k$ corresponding to the minimum eigenvalue. As with the MMSE receiver, we iterate until $R_{\mathsf{mf};N_r}$ converges.

Each real and imaginary part of the elements of $\boldsymbol{\alpha}_k$ are quantized with $B/(2DK)$ feedback bits. The Lloyd-Max quantizer [28] is used, where we assume that the distributions for both the real and imaginary parts are Gaussian with zero mean and variance $1/(2D)$. Numerical comparisons with empirical pdf's have shown that this is a reasonable assumption.

## *D. Numerical Results*

Fig. 8 compares the approximation for asymptotic RVQ performance with a matched filter receiver from (56) with simulated results for $N_t = 12$, $\bar{N}_r = 0.75$, $K/N_t = 1/2$, and SNR $= 5$ dB. For the case shown, the analytical approximation gives an accurate estimate of the performance of the finite size system with limited feedback. The asymptotic performance of RVQ with an optimal receiver, derived in Section IV, is also shown along with the water-filling capacity ($\hat{B} = \infty$). 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 to achieve the capacity corresponding to unlimited feedback (32), whereas the matched filter requires $1.8$ 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

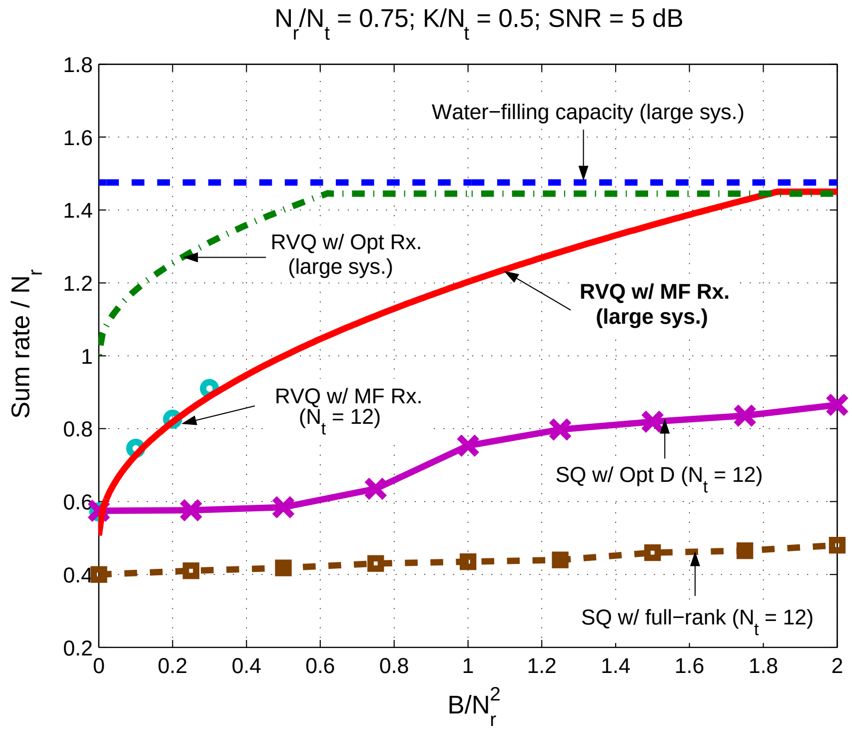



Fig. 8. Sum rate per receive antenna versus normalized feedback bits with a matched filter receiver. Results are shown for RVQ (both asymptotic and for $N_t = 12$), scalar quantization (full-rank), and reduced-rank scalar quantization of the precoding matrix. Also shown are results for the optimal receiver,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

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.

The limited-feedback performance with an RR scalar quantizer is also shown in Figs. 8 and 9, corresponding to matched filter and MMSE receivers, respectively. For a given feedback rate $\hat{B}$, the performance can be optimized over the rank $D$. The performance with a full-rank scalar quantizer ($D = N_t$) is also shown for comparison. With the MMSE receiver the achievable rate with the RR quantizer is about midway between the achievable rate for RVQ and the full-rank scalar quantizer. The RR scalar quantizer offers less of a performance gain (relative to the full-rank scalar quantizer) with the matched filter. The performance gains shown here are less than those observed for the MISO channel in Section III. This is because the subspaces spanned by the different $\boldsymbol{F}_k$'s can overlap substantially, which introduces correlation among the columns of

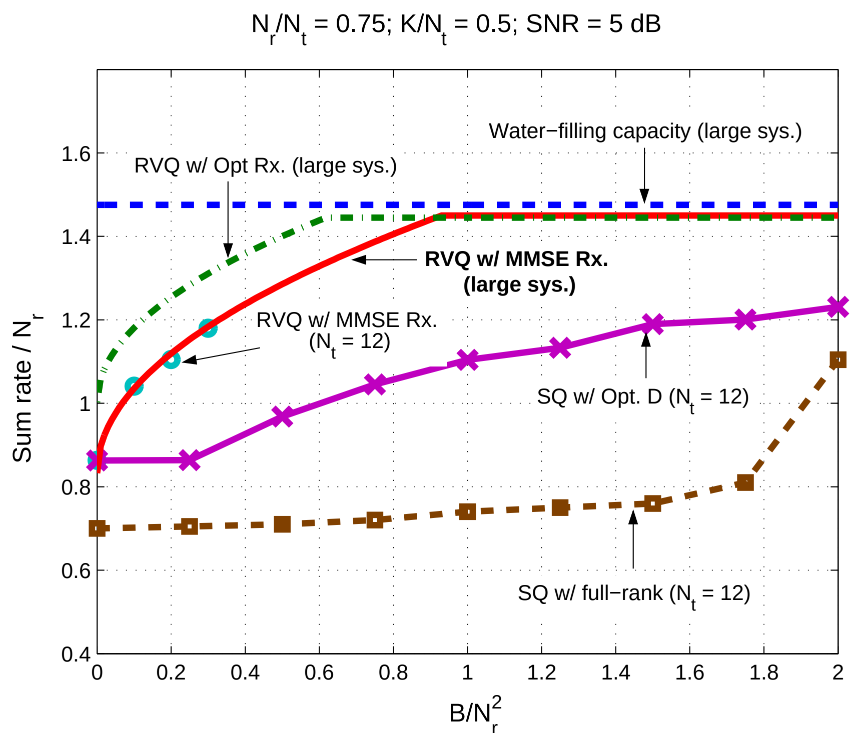



Fig. 9. Sum rate per receive antenna versus normalized feedback bits with a linear MMSE receiver. Results are shown for RVQ (both asymptotic and for $N_t = 12$), scalar quantization (full-rank), and reduced-rank scalar quantization of the precoding matrix. Also shown are results for the optimal receiver,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

the quantized precoding matrix.

## 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. Numerical results have shown that this approximation leads to an accurate estimate of 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). The comparison with the RR scalar quantizer has shown that this quantizer can provide a substantial increase in rate relative to full-rank scalar quantization; however, there is still a substantial gap between RR and RVQ performance.

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], [37], can exploit correlation among channel elements. The design of quantization codebooks for precoders, which takes correlation into account, is addressed in [37]. The effect of channel estimation error on the performance of limited feedback beamforming with finite coherence time (i.e., block fading) is presented [38], [39].

We have also assumed that the channel gains are not frequency-selective. Limited feedback schemes for frequency-selective scalar channels are discussed in [40], 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 [24]–[26]. 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

The inner product between the channel vector $\boldsymbol{h}$ and the unit-norm beamforming vector $\boldsymbol{v}_j$ can be written as

$$|\boldsymbol{h}^\dagger \boldsymbol{v}_j|^2 = \|\boldsymbol{h}\|^2 \cos^2 \Theta_j \tag{67}$$

where $\Theta_j$ is the angle between $\boldsymbol{h}$ and $\boldsymbol{v}_j$. For given $\boldsymbol{h}$, the receiver selects

$$\hat{\boldsymbol{v}} = \arg \max_{1 \le j \le 2^B} \{\log(1 + \rho \|\boldsymbol{h}\|^2 \cos^2 \Theta_j)\} \tag{68}$$

$$= \arg \max_{1 \le j \le 2^B} \{Y_j = \cos^2 \Theta_j\}. \tag{69}$$

since $\log$ is a monotonically increasing function.

We first determine the distribution of $Y_j$ for a given channel vector $\boldsymbol{h}$. Since $\boldsymbol{v}_j$ is uniformly distributed on an $N_t$-dimensional unit-norm ball, the probability that $\Theta_j \le \theta$ is the ratio between the surface area of a spherical cap in an $N_t$-dimensional unit ball, defined by angle $\theta$, to the total surface area [2]. Therefore, we have

$$\Pr\{\Theta_j \le \theta | \boldsymbol{h}\} = \begin{cases} \frac{1}{2}(1 - \cos^2 \theta)^{N_t - 1}, & 0 \le \theta \le \pi/2 \\ 1 - \frac{1}{2}(1 - \cos^2 \theta)^{N_t - 1}, & \pi/2 \le \theta \le \pi \end{cases} \tag{70}$$

and after some algebraic manipulations, it can be shown that the cdf of $Y_j$ given $\boldsymbol{h}$ is

$$F_{Y|\boldsymbol{h}}(y) = 1 - (1 - y)^{N_t - 1}, \quad 0 \le y \le 1. \tag{71}$$

In addition to having identical distributions, the $Y_j$'s are independent, since the codebook entries $\boldsymbol{v}_j$, $j = 1, \cdots, 2^B$, are independent.

Now from [29, Theorem 2.1.2], it follows that

$$\frac{\max_j Y_j - a_n}{b_n} \xrightarrow{\mathcal{D}} \mathcal{Y} \tag{72}$$

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

$$H_\gamma(x) = \begin{cases} 1, & x \ge 0 \\ \exp(-(-x)^\gamma), & x < 0 \end{cases}, \tag{73}$$

$\mathcal{D}$ 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 $y > 0$,

$$\lim_{t\to\infty} \frac{1 - F^*_{Y|\boldsymbol{h}}(ty)}{1 - F^*_{Y|\boldsymbol{h}}(y)} = y^{-\gamma} \tag{74}$$

where the constant $\gamma > 0$.

Substituting the expression for $F_{Y|\boldsymbol{h}}$ in (71) into (74), where $\omega(F_{Y|\boldsymbol{h}}) = 1$, gives

$$\lim_{t\to\infty} \frac{1 - F^*_{Y|\boldsymbol{h}}(ty)}{1 - F^*_{Y|\boldsymbol{h}}(y)} = \lim_{t\to\infty} \frac{\left(1 - \frac{1}{ty}\right)^{N_t - 1}}{\left(1 - \frac{1}{t}\right)^{N_t - 1}} \tag{75}$$

$$= y^{-(N_t - 1)} \tag{76}$$

so that [29, Theorem 2.1.2] applies when $N_t > 1$. Furthermore, the normalizing constants are given by

$$a_n = \omega(F_{Y|\boldsymbol{h}}) = 1 \tag{77}$$

and

$$b_n = \omega(F_{Y|\boldsymbol{h}}) - \inf\left\{y : 1 - F_{Y|\boldsymbol{h}}(y) \le \frac{1}{n}\right\} \tag{78}$$

$$= 1 - F^{-1}_{Y|\boldsymbol{h}}\left(1 - \frac{1}{n}\right) = \left(\frac{1}{n}\right)^{\frac{1}{N_t - 1}}. \tag{79}$$

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 (72) and taking the large system limit gives

$$\lim_{(N_t,n)\to\infty} E_{\mathcal{V}}[\max_j Y_j | \bar{\boldsymbol{h}}] = \lim_{(N_t,n)\to\infty} a_n + b_n E[\mathcal{Y}] \tag{80}$$

$$= 1 - \lim_{(N_t,n)\to\infty} \left(\frac{1}{n}\right)^{\frac{1}{N_t - 1}} \Gamma\left(1 - \frac{1}{N_t - 1}\right) \tag{81}$$

$$= 1 - \lim_{(N_t,B)\to\infty} 2^{-\frac{B}{N_t - 1}} \tag{82}$$

$$= 1 - 2^{-\bar{B}} \tag{83}$$

where the gamma function $\Gamma(z) = \int_0^\infty t^{z-1} \mathrm{e}^{-t}\, \mathrm{d}t$ and we have used the fact that $E[\mathcal{Y}] = -\Gamma(1 - 1/(N_t - 1))$ [41].

From (72), as $(n, N_t) \to \infty$,

$$\mathrm{var}[\max_j Y_j|\bar{\boldsymbol{h}}] - b_n^2 \mathrm{var}[\mathcal{Y}] \to 0. \tag{84}$$

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

$$\max_j Y_j \to 1 - 2^{-\bar{B}} \tag{85}$$

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

$$\begin{aligned} \mathcal{I}_{\mathsf{rvq}}^{\triangle} &= \lim_{(N_t,B)\to\infty} \left( \log(1 + \rho\|\boldsymbol{h}\|^2 \max_{1\le j\le 2^B} Y_j) \right) - \log(\rho N_t) && (86)\\ &= \lim_{(N_t,B)\to\infty} \left( \log\left( \frac{1}{\rho N_t} + \frac{1}{N_t}\|\boldsymbol{h}\|^2 \max_{1\le j\le 2^B} Y_j \right)\right) && (87)\\ &= \log(1 - 2^{-\bar{B}}) && (88) \end{aligned}$$

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

### *B. Proof of Theorem 3*

We first prove the theorem for $\bar{N}_r \ge 1$. Let $z = F^{-1}_{\gamma|\bar{\boldsymbol{H}}}(1-2^{-B})$. Rearranging (14) gives

$$\lim_{\substack{(N_t,N_r)\to\infty\\ z\to\gamma_{\mathsf{rvq}}^{\infty}}} \left[1 - F_{\gamma|\bar{\boldsymbol{H}}}(z)\right]^{\frac{1}{N_t}} = 2^{-\bar{B}}. \tag{89}$$

Next, we derive upper and lower bounds for the left-hand side of (89) 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 [32]. That is,

$$\begin{aligned} 1 - F_{\gamma|\bar{\boldsymbol{H}}}(z) &= \Pr\left\{ \gamma_j > z |\, \bar{\boldsymbol{H}} \right\} && (90)\\ &= \Pr\left\{ \boldsymbol{v}_j^{\dagger}\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{\dagger}\boldsymbol{v}_j > z|\boldsymbol{\Lambda}, \boldsymbol{U} \right\} && (91)\\ &= \Pr\left\{ \left. \frac{\boldsymbol{w}_j^{\dagger}\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{\dagger}\boldsymbol{w}_j}{\boldsymbol{w}_j^{\dagger}\boldsymbol{w}_j} > z \,\right| \boldsymbol{\Lambda}, \boldsymbol{U} \right\} && (92) \end{aligned}$$

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, $\boldsymbol{\Lambda} = \mathrm{diag}\{\lambda_1, \cdots, \lambda_{N_t}\}$, 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. It can be shown that $\boldsymbol{v}_j$ and $\boldsymbol{w}_j/\|\boldsymbol{w}_j\|$ are both isotropically distributed, i.e., $\boldsymbol{U}\boldsymbol{w}_j$ and $\boldsymbol{w}_j$ have the same distribution, so that

$$\begin{aligned}
1-F_{\gamma|\bar{\boldsymbol{H}}}(z) &= 1-F_{\gamma|\boldsymbol{\Lambda}}(z) && (93)\\
&= \Pr\left\{ \frac{\sum_{i=1}^{N_t}\lambda_i w_i^2}{\sum_{i=1}^{N_t} w_i^2} > z \,\middle|\, \boldsymbol{\Lambda} \right\} && (94)\\
&= \Pr\left\{ -\sum_{i=1}^{N_t}(z-\lambda_i)w_i^2 > 0 \,\middle|\, \boldsymbol{\Lambda} \right\} && (95)\\
&= \Pr\left\{ -\rho\sum_{i=1}^{N_t}(z-\lambda_i)w_i^2 > 0 \,\middle|\, \boldsymbol{\Lambda} \right\}, \quad \forall\rho>0 && (96)\\
&= \Pr\left\{ \exp\left(-\rho\sum_{i=1}^{N_t}(z-\lambda_i)w_i^2\right) > 1 \,\middle|\, \boldsymbol{\Lambda} \right\} && (97)
\end{aligned}$$

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

$$\begin{aligned}
1-F_{\gamma|\boldsymbol{\Lambda}}(z) &\leq E_{\{w_i\}}\left[ \exp\left\{-\rho\sum_{i=1}^{N_t}(z-\lambda_i)w_i^2\right\} \,\middle|\, \boldsymbol{\Lambda} \right] && (98)\\
&\leq \prod_{i=1}^{N_t} E_{w_i}\left[\exp\left\{-\rho(z-\lambda_i)w_i^2\right\}|\lambda_i\right] && (99)\\
&= \prod_{i=1}^{N_t}\int_0^\infty \exp\left\{-\rho(z-\lambda_i)x\right\}\mathrm{e}^{-x}\mathrm{d}x && (100)\\
&= \prod_{i=1}^{N_t}\int_0^\infty \exp\left\{-(1+\rho(z-\lambda_i))x\right\}\mathrm{d}x && (101)\\
&= \prod_{i=1}^{N_t}\frac{1}{1+\rho(z-\lambda_i)} && (102)\\
&= \exp\left\{-\sum_{i=1}^{N_t}\log(1+\rho(z-\lambda_i))\right\} && (103)
\end{aligned}$$

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

$$\lim_{\substack{(N_t,N_r)\to\infty\\ z\to\gamma_{\mathsf{rvq}}^\infty}}\left[1-F_{\gamma|\boldsymbol{\Lambda}}(z)\right]^{\frac{1}{N_t}} \leq \exp\{-\Phi(\gamma_{\mathsf{rvq}}^\infty,\rho)\} \tag{104}$$

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

$$\Phi(\gamma_{\mathsf{rvq}}^\infty,\rho) \triangleq \int_a^b \log(1+\rho(\gamma_{\mathsf{rvq}}^\infty-\lambda))g(\lambda)\mathrm{d}\lambda, \tag{105}$$

$g(\lambda)$ is given by (27)-(28), and $\lambda_{\max}^{\infty} = \lim_{(N_t,N_r)\to\infty} \lambda_1 = (1+\sqrt{\bar{N}_r})^2$. To tighten the upper bound, we minimize (104) with respect to $\rho$, i.e.,

$$\lim_{\substack{(N_t,N_r)\to\infty \\ z\to\gamma_{\mathsf{rvq}}^{\infty}}} \left[1-F_{\gamma|\mathbf{\Lambda}}(z)\right]^{\frac{1}{N_t}} \le \exp\{-\Phi(\gamma_{\mathsf{rvq}}^{\infty},\rho^*)\} \tag{106}$$

where

$$\rho^* = \arg \max_{0<\rho<\frac{1}{\lambda_{\max}^{\infty}-\gamma_{\mathsf{rvq}}^{\infty}}} \Phi(\gamma_{\mathsf{rvq}}^{\infty},\rho). \tag{107}$$

A similar expression for RVQ performance when used to quantize signatures for CDMA is derived in [32].

To derive the lower bound, we use the following result in [31],

$$1-F_{\gamma|\mathbf{\Lambda}}(z) = \sum_{i=1}^{s} \frac{(\lambda_i - z)^{N_t-1}}{\prod_{\substack{l=1\\ l\ne i}}^{N_t}(\lambda_i-\lambda_l)} \tag{108}$$

where $\lambda_s \ge z \ge \lambda_{s+1}$. The terms in the preceding sum have alternating signs. We rearrange (108) as follows,

$$\begin{aligned} 1-F_{\gamma|\mathbf{\Lambda}}(z) &= \sum_{i=1}^{s} \frac{1}{\prod_{l\ne i}\left(1+(z-\lambda_l)\left(\frac{1}{\lambda_i - z}\right)\right)} && (109)\\ &= \sum_{i=1}^{s}(-1)^{i+1}\exp\left\{-\sum_{l\ne i}\log\left|1+(z-\lambda_l)\left(\frac{1}{\lambda_i-z}\right)\right|\right\} && (110)\\ &= \exp\left\{-\sum_{l\ne i}\log|1+(z-\lambda_l)\beta'|\right\} && (111)\end{aligned}$$

where $\beta'$ is a constant to be determined. Hence

$$\lim_{\substack{(N_t,N_r)\to\infty \\ z\to\gamma_{\mathsf{rvq}}^{\infty}}} \left[1-F_{\gamma|\mathbf{\Lambda}}(z)\right]^{\frac{1}{N_t}} = \exp\left\{-\int_a^b \log|1+\beta'(\gamma_{\mathsf{rvq}}^{\infty}-\lambda)|g(\lambda)\,\mathrm{d}\lambda\right\}. \tag{112}$$

To show that there exists a $\beta'$, which satisfies (110)-(111), from (106) we have

$$0=\exp\{-\Phi(\gamma_{\mathsf{rvq}}^{\infty},0)\} \le \lim_{\substack{(N_t,N_r)\to\infty \\ z\to\gamma_{\mathsf{rvq}}^{\infty}}} \left[1-F_{\gamma|\mathbf{\Lambda}}(z)\right]^{\frac{1}{N_t}} \le \exp\{-\Phi(\gamma_{\mathsf{rvq}}^{\infty},\rho^*)\}. \tag{113}$$

Since $\rho^* \in (0, 1/(\lambda_{\max}^{\infty}-\gamma_{\mathsf{rvq}}^{\infty}))$ and $\Phi(\gamma_{\mathsf{rvq}}^{\infty}, x)$ is continuous for $x \in (0, 1/(\lambda_{\max}^{\infty}-\gamma_{\mathsf{rvq}}^{\infty}))$, we conclude that $\beta' \in (0, 1/(\lambda_{\max}^{\infty}-\gamma_{\mathsf{rvq}}^{\infty}))$. With $\beta'$ in that range, the absolute value in (112) is unnecessary. Therefore

$$\lim_{\substack{(N_t,N_r)\to\infty \\ z\to\gamma_{\mathsf{rvq}}^{\infty}}} \left[1 - F_{\gamma|\mathbf{\Lambda}}(z)\right]^{\frac{1}{N_t}} = \exp\{-\Phi(\gamma_{\mathsf{rvq}}^{\infty}, \beta')\} \tag{114}$$

$$\geq \min_{0<\beta<\frac{1}{\lambda_{\max}^{\infty}-\gamma_{\mathsf{rvq}}^{\infty}}} \exp\{-\Phi(\gamma_{\mathsf{rvq}}^{\infty}, \beta)\}. \tag{115}$$

The lower bound in (115) is exactly the upper bound (106). Therefore

$$\lim_{\substack{(N_t,N_r)\to\infty \\ z\to\gamma_{\mathsf{rvq}}^{\infty}}} \left[1 - F_{\gamma|\mathbf{\Lambda}}(z)\right]^{\frac{1}{N_t}} = \exp\{-\Phi(\gamma_{\mathsf{rvq}}^{\infty}, \rho^*)\} \tag{116}$$

$$= 2^{-\bar{B}} \tag{117}$$

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

$$\Phi(\gamma_{\mathsf{rvq}}^{\infty}, \rho^*) = \bar{B}\log(2) \tag{118}$$

where $\rho^*$ is given by (107). The goal of the rest of the proof is to simplify (118).

To determine $\rho^*$, we first compute

$$\frac{\partial\Phi(\gamma_{\mathsf{rvq}}^{\infty}, \rho)}{\partial\rho} = \int_a^b \left[\frac{\gamma_{\mathsf{rvq}}^{\infty} - \lambda}{1 + (\gamma_{\mathsf{rvq}}^{\infty} - \lambda)\rho}\right] g(\lambda)\, \mathrm{d}\lambda \tag{119}$$

$$= \frac{1}{\rho} + \frac{1}{\rho^2}\int_a^b \frac{1}{\lambda - \underbrace{\left(\frac{1}{\rho} + \gamma_{\mathsf{rvq}}^{\infty}\right)}_{y}} g(\lambda)\, \mathrm{d}\lambda \tag{120}$$

$$= \frac{1}{\rho} + \frac{1}{\rho^2}\mathcal{S}_{\mathbf{\Lambda}}(y) \tag{121}$$

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 $\rho$ gives

$$\mathcal{S}_{\mathbf{\Lambda}}(y) = -\rho = \frac{1}{\gamma_{\mathsf{rvq}}^{\infty} - y} \tag{122}$$

as the only valid solution. Substituting the expression for $\mathcal{S}_{\mathbf{\Lambda}}$ given in [30] into (122) gives

$$\frac{(-1 + \bar{N}_r - y) \pm \sqrt{y^2 - 2(\bar{N}_r + 1)y + (\bar{N}_r - 1)^2}}{2y} = \frac{1}{\gamma_{\mathsf{rvq}}^{\infty} - y}, \tag{123}$$

which simplifies to the quadratic equation

$$(\bar{N}_r - \gamma_{\mathsf{rvq}}^{\infty})y^2 + [\gamma_{\mathsf{rvq}}^{\infty} + \bar{N}_r\gamma_{\mathsf{rvq}}^{\infty} + (\gamma_{\mathsf{rvq}}^{\infty})^2]y = 0. \tag{124}$$

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

$$\rho^* = \frac{\gamma_{\mathsf{rvq}}^{\infty} - \bar{N}_r}{\gamma_{\mathsf{rvq}}^{\infty}}. \tag{125}$$

It is easily verified that $\left.\frac{\partial^2 \Phi(\gamma_{\mathsf{rvq}}^{\infty}, \rho)}{\partial \rho^2}\right|_{\rho=\rho^*} < 0$, so that $\rho^*$ achieves a maximum.

From (125), $\rho^*$ increases with $\gamma_{\mathsf{rvq}}^{\infty} \in [\bar{N}_r, \lambda_{\max}^{\infty}]$. For large enough $\gamma_{\mathsf{rvq}}^{\infty}$, $\rho^*$ in (125) can exceed the range $(0, 1/(\lambda_{\max}^{\infty} - \gamma_{\mathsf{rvq}}^{\infty}))$. We can compute this value ${\gamma_{\mathsf{rvq}}^{\infty}}^*$ by solving

$$\frac{{\gamma_{\mathsf{rvq}}^{\infty}}^* - \bar{N}_r}{{\gamma_{\mathsf{rvq}}^{\infty}}^*} = \frac{1}{\lambda_{\max}^{\infty} - {\gamma_{\mathsf{rvq}}^{\infty}}^*}, \tag{126}$$

which gives ${\gamma_{\mathsf{rvq}}^{\infty}}^* = \bar{N}_r + \sqrt{\bar{N}_r}$. Therefore

$$\rho^* = \begin{cases} \frac{\gamma_{\mathsf{rvq}}^{\infty} - \bar{N}_r}{\gamma_{\mathsf{rvq}}^{\infty}}, & \bar{N}_r \le \gamma_{\mathsf{rvq}}^{\infty} \le \bar{N}_r + \sqrt{\bar{N}_r} \\ \frac{1}{(1+\sqrt{\bar{N}_r})^2 - \gamma_{\mathsf{rvq}}^{\infty}}, & \bar{N}_r + \sqrt{\bar{N}_r} \le \gamma_{\mathsf{rvq}}^{\infty} < (1 + \sqrt{\bar{N}_r})^2 \end{cases}. \tag{127}$$

To evaluate $\Phi(\gamma_{\mathsf{rvq}}^{\infty}, \rho^*)$, we re-write (105) as

$$\begin{aligned} \Phi(\gamma_{\mathsf{rvq}}^{\infty}, \rho^*) &= \int_a^b \left[\log(\rho^*) + \log\left(\left(\frac{1}{\rho^*} + \gamma_{\mathsf{rvq}}^{\infty}\right) - \lambda\right)\right] g(\lambda)\,\mathrm{d}\lambda & (128) \\ &= \log(\rho^*) + \int_a^b \log\left(\left(\frac{1}{\rho^*} + \gamma_{\mathsf{rvq}}^{\infty}\right) - \lambda\right) g(\lambda)\,\mathrm{d}\lambda. & (129) \end{aligned}$$

To evaluate the integral in (129), we apply the following Lemma.

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

$$\begin{aligned} \Theta(x) &\triangleq \int_a^b \log(x - \lambda) g(\lambda)\,\mathrm{d}\lambda & (130) \\ &= \log(w(x)) + \sqrt{\bar{N}_r}\, u(x) - (\bar{N}_r - 1)\log\left(1 + \frac{u(x)}{\sqrt{\bar{N}_r}}\right) & (131) \end{aligned}$$

*where*

$$\begin{aligned} w(x) &= \frac{(x - 1 - \bar{N}_r) + \sqrt{(x - 1 - \bar{N}_r)^2 - 4\bar{N}_r}}{2}, & (132) \\ u(x) &= \frac{(x - 1 - \bar{N}_r) - \sqrt{(x - 1 - \bar{N}_r)^2 - 4\bar{N}_r}}{2\sqrt{\bar{N}_r}}. & (133) \end{aligned}$$

The proof of this Lemma is similar to that given in [34] 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 (129) to obtain

$$\Phi(\gamma_{\mathsf{rvq}}^{\infty}, [(1+\sqrt{\bar{N}_r})^2 - \gamma_{\mathsf{rvq}}^{\infty}]^{-1}) \tag{134}$$

$$= -\log[(1+\sqrt{\bar{N}_r})^2 - \gamma_{\mathsf{rvq}}^{\infty}] + \Theta\left((1+\sqrt{\bar{N}_r})^2\right) \tag{135}$$

$$= -\log[(1+\sqrt{\bar{N}_r})^2 - \gamma_{\mathsf{rvq}}^{\infty}] + \frac{1}{2}\bar{N}_r \log(\bar{N}_r) - (\bar{N}_r - 1)\log(1+\sqrt{\bar{N}_r}) + \sqrt{\bar{N}_r} \tag{136}$$

$$= \bar{B}\log(2). \tag{137}$$

Solving for $\gamma_{\mathsf{rvq}}^{\infty}$ gives (17). Taking $\gamma_{\mathsf{rvq}}^{\infty} = \bar{N}_r + \sqrt{\bar{N}_r}$ and solving for $\bar{B}$ gives $\bar{B}^*$ in (18).

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 (129) to obtain

$$\Phi\left(\gamma_{\mathsf{rvq}}^{\infty}, \frac{\gamma_{\mathsf{rvq}}^{\infty} - \bar{N}_r}{\gamma_{\mathsf{rvq}}^{\infty}}\right) = \log(\gamma_{\mathsf{rvq}}^{\infty} - \bar{N}_r) - \log(\gamma_{\mathsf{rvq}}^{\infty}) + \Theta\left(\gamma_{\mathsf{rvq}}^{\infty} + \frac{\gamma_{\mathsf{rvq}}^{\infty}}{\gamma_{\mathsf{rvq}}^{\infty} - \bar{N}_r}\right). \tag{138}$$

To simplify (138), we let $\psi \triangleq \gamma_{\mathsf{rvq}}^{\infty} - \bar{N}_r$ and re-write (138) as

$$\Phi\left(\psi - \bar{N}_r, \frac{\psi}{\psi + \bar{N}_r}\right) = \log(\psi) - \log(\psi + \bar{N}_r) + \Theta\left(1 + \bar{N}_r + \psi + \frac{\bar{N}_r}{\psi}\right). \tag{139}$$

It can easily be shown that

$$w\left(1 + \bar{N}_r + \psi + \frac{\bar{N}_r}{\psi}\right) = \frac{\bar{N}_r}{\psi}, \tag{140}$$

$$u\left(1 + \bar{N}_r + \psi + \frac{\bar{N}_r}{\psi}\right) = \frac{\psi}{\sqrt{\bar{N}_r}}, \tag{141}$$

and

$$\Theta\left(1 + \bar{N}_r + \psi + \frac{\bar{N}_r}{\psi}\right) = \log(\bar{N}_r) - \log(\psi) - (\bar{N}_r - 1)\log\left(1 + \frac{\psi}{\bar{N}_r}\right) + \psi. \tag{142}$$

Substituting (142) into (139), we obtain

$$\Phi\left(\psi - \bar{N}_r, \frac{\psi}{\psi + \bar{N}_r}\right) = \psi - \bar{N}_r \log\left(1 + \frac{\psi}{\bar{N}_r}\right) \tag{143}$$

$$= \gamma_{\mathsf{rvq}}^{\infty} - \bar{N}_r - \bar{N}_r \log(\gamma_{\mathsf{rvq}}^{\infty}) + \bar{N}_r \log(\bar{N}_r). \tag{144}$$

Setting this to $\bar{B}\log(2)$ and simplifying gives (16).

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

$$g(\lambda) = (1 - \bar{N}_r)\delta(\lambda) + \frac{\sqrt{(\lambda - a)(b - \lambda)}}{2\pi\lambda}. \tag{145}$$

where $a$ and $b$ are given by (27)-(28). Following the same steps again from (105) gives (17) and (16). This completes the proof of Theorem 3.

### C. *Derivation of* (42)-(44)

To compute $\mu_J$, we first write

$$J_j^{N_r} = \frac{1}{N_r}\sum_{k=1}^{N_r}\log\left(1+\rho\frac{\bar{N}_r}{\bar{K}}\upsilon_k\right) \tag{146}$$

where $\upsilon_k$ is the $k$th eigenvalue of $\mathbf{\Upsilon} = \frac{1}{N_r}\boldsymbol{H}\boldsymbol{V}_j\boldsymbol{V}_j^\dagger\boldsymbol{H}^\dagger$. As $(N_t, N_r, K) \to \infty$, the empirical eigenvalue distribution converges to a deterministic function $F_{\mathbf{\Upsilon}}(t)$. The asymptotic mean is given by

$$\mu_J = \lim_{(N_t,N_r,K)\to\infty} E[J_j^{N_r}] = \int_0^\infty \log\left(1+\rho\frac{\bar{N}_r}{\bar{K}}t\right)\,\mathrm{d}F_{\mathbf{\Upsilon}}(t). \tag{147}$$

A similar integral has been evaluated in [34, eq. (6)], and the result can be directly applied to (147), giving (42).

To compute the variance, we express $J_j^{N_r}$ differently by first performing the singular value decomposition $\boldsymbol{H} = \boldsymbol{V_H}\mathbf{\Sigma}_{\boldsymbol{H}}\boldsymbol{U}_{\boldsymbol{H}}^\dagger$, where $\boldsymbol{V_H}$ is the $N_r \times N_r$ left singular matrix, $\boldsymbol{U_H}$ is the $N_t \times N_r$ right singular matrix, and $\mathbf{\Sigma}_{\boldsymbol{H}}$ 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

$$\begin{aligned} J_j^{N_r} &= \frac{1}{N_r}\log\det\left(\boldsymbol{I}_{N_r} + \rho\frac{1}{K}\mathbf{\Sigma}_{\boldsymbol{H}}^2\boldsymbol{U}_{\boldsymbol{H}}^\dagger\boldsymbol{V}\boldsymbol{V}^\dagger\boldsymbol{U}_{\boldsymbol{H}}\right) && (148)\\ &= \frac{1}{N_r}\sum_{i=1}^{N_r}\log\left(1+\rho\eta_i\right) && (149)\end{aligned}$$

where $\eta_i$ is the $i$th eigenvalue of $\frac{1}{K}\mathbf{\Sigma}_{\boldsymbol{H}}^2\boldsymbol{U}_{\boldsymbol{H}}^\dagger\boldsymbol{V}\boldsymbol{V}^\dagger\boldsymbol{U}_{\boldsymbol{H}}$. To compute $\mathrm{var}[J_j^{N_r}]$, correlations between pairs of $\eta_i$'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 (149) to write

$$\begin{aligned} J_j^{N_r} &= \frac{\rho}{N_r}\sum_{i=1}^{N_r}\eta_i - \frac{\rho^2}{2N_r}\sum_{i=1}^{N_r}\eta_i^2 + \frac{\rho^3}{3N_r}\sum_{i=1}^{N_r}\eta_i^3 + \ldots && (150)\\ &= \frac{\rho}{N_r}\mathrm{tr}\{\mathbf{\Lambda}\boldsymbol{L}\} - \frac{\rho^2}{2N_r}\mathrm{tr}\{(\mathbf{\Lambda}\boldsymbol{L})^2\} + \frac{\rho^3}{3N_r}\mathrm{tr}\{(\mathbf{\Lambda}\boldsymbol{L})^3\} + \ldots && (151)\end{aligned}$$

for $\rho < [\max_i \eta_i]^{-1} = (4\bar{N}_r)^{-1}$ (approximately $-6$ dB for $\bar{N}_r = 1$), where $\boldsymbol{L} = \boldsymbol{U}_{\boldsymbol{H}}^\dagger\boldsymbol{V}\boldsymbol{V}^\dagger\boldsymbol{U}_{\boldsymbol{H}}$ and $\mathbf{\Lambda} = \frac{1}{K}\mathbf{\Sigma}_{\boldsymbol{H}}^2$. Ignoring the terms of order $\rho^4$ and higher, we can approximate the variance of $J_j^{N_r}$ at low SNR as

$$\mathrm{var}[J_j^{N_r}] \approx \rho^2\mathrm{var}\left[\frac{1}{N_r}\mathrm{tr}\{\mathbf{\Lambda}\boldsymbol{L}\}\right] - \frac{\rho^3}{2}\mathrm{cov}\left[\frac{1}{N_r}\mathrm{tr}\{\mathbf{\Lambda}\boldsymbol{L}\}\frac{1}{N_r}\mathrm{tr}\{(\mathbf{\Lambda}\boldsymbol{L})^2\}\right]. \tag{152}$$

Letting $\mathbf{\Lambda} = \mathrm{diag}\{\lambda_i\}$ and $l_{ij}$ denote the $(i,j)$th element of $\boldsymbol{L}$, the first term in (152) can be expanded as

$$\mathrm{var}[\mathrm{tr}\{\mathbf{\Lambda}\boldsymbol{L}\}|\mathbf{\Lambda}] = \sum_{i=1}^{N_r} \lambda_i^2 \left(E[l_{ii}^2] - E^2[l_{ii}]\right) + \sum_{i\neq j} \lambda_i\lambda_j \left(E[l_{ii}l_{jj}] - E[l_{ii}]E[l_{jj}]\right). \tag{153}$$

For a given $\boldsymbol{U_H}$ and random unitary $\boldsymbol{V}$ with $K = N_r$, Theorem 3 in [7] states that $\boldsymbol{L}$ has a multivariate beta distribution with parameters $N_r$ and $N_t - N_r$. (The distribution of $\boldsymbol{L}$ is not known for general $K$.) From Theorem 2 in [42], we have

$$E[l_{ii}] = \frac{N_r+1}{N_t+2} \tag{154}$$

$$E[l_{ii}^2] = \frac{(N_r+1)(N_r+3)}{(N_t+2)(N_t+4)} \tag{155}$$

$$E[l_{ii}l_{jj}] = \frac{N_r(N_r+1)(N_t+4)+(N_r+1)(N_t-N_r+1)}{(N_t+1)(N_t+2)(N_t+4)}, \quad i\neq j, \tag{156}$$

for $1 \leq i,j \leq N_r$. Substituting (154)-(156) into (153) gives

$$\begin{aligned}\mathrm{var}[\mathrm{tr}\{\mathbf{\Lambda}\boldsymbol{L}\}|\mathbf{\Lambda}] = &\left(\frac{1}{N_r}\sum_{i=1}^{N_r}\lambda_i^2\right)\left(\bar{N}_r^2(1-\bar{N}_r) + O\left(\frac{1}{N_r}\right)\right) \\ &+ \left(\frac{1}{N_r^2}\sum_{i\neq j}\lambda_i\lambda_j\right)\left((\bar{N}_r-1)\bar{N}_r^3 + O\left(\frac{1}{N_r}\right)\right).\end{aligned} \tag{157}$$

Taking expectation with respect to $\mathbf{\Lambda}$, and the large system limit, we have

$$E_{\mathbf{\Lambda}}\left(\mathrm{var}[\mathrm{tr}\{\mathbf{\Lambda}\boldsymbol{L}\}|\mathbf{\Lambda}]\right) \to 1 - \bar{N}_r. \tag{158}$$

Also, in the large system limit

$$\frac{1}{N_r}\sum_{i=1}^{N_r}\lambda_i^2 \to \int t^2 \mathrm{d}F_{\mathbf{\Lambda}}(t) = \frac{1}{\bar{N}_r}\left(1+\frac{1}{\bar{N}_r}\right) \tag{159}$$

$$\frac{1}{N_r^2}\sum_{i\neq j}\lambda_i\lambda_j \to \left[\int t\mathrm{d}F_{\mathbf{\Lambda}}(t)\right]^2 = \frac{1}{\bar{N}_r^2} \tag{160}$$

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

The computation of the second term in (152) follows similar steps; however, third-order terms involving $l_{ij}$ appear, in addition to the first- and second-order terms. Those can be evaluated by utilizing the independence property of the multivariate beta distribution shown in [42, Theorem 1]. It can then be shown that

$$E_{\boldsymbol{\Lambda}}\left(\operatorname{cov}\left[\operatorname{tr}\{\boldsymbol{\Lambda}\boldsymbol{L}\}\operatorname{tr}\{(\boldsymbol{\Lambda}\boldsymbol{L})^2\}|\boldsymbol{\Lambda}\right]\right) \to 4(1-\bar{N}_r). \tag{161}$$

Substituting (158) and (161) into (152), we have

$$\begin{aligned}\sigma_J^2 &= \lim_{(N_r,N_t)\to\infty} N_r^2 \operatorname{var}[J_j^{N_r}] && (162)\\ &\approx \rho^2(1-\bar{N}_r)(1-2\rho). && (163)\end{aligned}$$

### *D. Derivation of* (54) *and* (63)

Recall that $R_{\mathsf{mf}}^{N_r} = \frac{1}{N_r}\sum_{k=1}^{K}\log(1+\gamma_{k;\mathsf{mf}})$. To evaluate the variance of $R_{\mathsf{mf}}^{N_r}$, we first compute the variance of $\log(1+\gamma_{k;\mathsf{mf}})$. Letting $\tilde{\boldsymbol{v}}_k \triangleq \boldsymbol{H}\boldsymbol{v}_k$, the SINR at the output of the matched filter for substream $k$ is

$$\gamma_{k;\mathsf{mf}} \triangleq z_k^{-1} = \left[\frac{K\sigma_n^2}{\tilde{\boldsymbol{v}}_k^\dagger\tilde{\boldsymbol{v}}_k} + \frac{\sum_{i=1,i\neq k}^{K}|\tilde{\boldsymbol{v}}_k^\dagger\tilde{\boldsymbol{v}}_i|^2}{(\tilde{\boldsymbol{v}}_k^\dagger\tilde{\boldsymbol{v}}_k)^2}\right]^{-1}. \tag{164}$$

Letting $w_k \triangleq \tilde{\boldsymbol{v}}_k^\dagger\tilde{\boldsymbol{v}}_k$ and $y_k \triangleq \sum_{i=1,i\neq k}^{K}|\tilde{\boldsymbol{v}}_k^\dagger\tilde{\boldsymbol{v}}_i|^2$, the variance of $z_k$ is given by

$$\operatorname{var}[z_k] = K^2\sigma_n^4\operatorname{var}\left[\frac{1}{w_k}\right] + \operatorname{var}\left[\frac{y_k}{w_k^2}\right] + 2K\sigma_n^2 E\left[\frac{y_k}{w_k^3}\right] - 2K\sigma_n^2 E\left[\frac{1}{w_k}\right]E\left[\frac{y_k}{w_k^2}\right]. \tag{165}$$

Given the codebook $\mathcal{V}$, $\tilde{\boldsymbol{v}}_k$ has *i.i.d.* Gaussian elements with zero mean and unit variance. It follows that $w_k$ has a Gamma pdf given by

$$f_{w_k}(x) = \frac{1}{\Gamma(N_r)}x^{N_r-1}\mathrm{e}^{-x}. \tag{166}$$

Using (166), we can show that

$$\lim_{(N_t,N_r,K)\to\infty} N_r K^2 \operatorname{var}\left[\frac{1}{w_k}\right] = \frac{3\bar{K}^2}{\bar{N}_r^2}. \tag{167}$$

Since the $\tilde{\boldsymbol{v}}_i$'s are independent, we can apply the central limit theorem to show that both real and imaginary parts of $\frac{1}{N_r}\tilde{\boldsymbol{v}}_i^\dagger\tilde{\boldsymbol{v}}_k$ converge in distribution to a Gaussian random variable with zero mean and variance $1/2$ as $(N_t,N_r,K)\to\infty$. Furthermore, $y_k$ and $w_k$ become independent as the system size tends to infinity. It can then be shown that

$$\lim_{(N_t,N_r,K)\to\infty} N_r \operatorname{var}\left[\frac{y_k}{w_k^2}\right] = \frac{4\bar{K}^2}{\bar{N}_r^2} + \frac{\bar{K}}{\bar{N}_r} \tag{168}$$

and

$$\lim_{(N_t,N_r,K)\to\infty} K N_r\left(E\left[\frac{y_k}{w_k^3}\right] - E\left[\frac{1}{w_k}\right]E\left[\frac{y_k}{w_k^2}\right]\right) = \frac{3\bar{K}^2}{\bar{N}_r^2}. \tag{169}$$

Combining (167)-(169) and (165) gives

$$\lim_{(N_t,N_r,K)\to\infty} N_r \mathrm{var}[z_k] = \frac{\bar{K}}{\bar{N}_r}\left(1 + 4\frac{\bar{K}}{\bar{N}_r} + 6\sigma_n^2\frac{\bar{K}}{\bar{N}_r} + 3\sigma_n^4\frac{\bar{K}}{\bar{N}_r}\right). \tag{170}$$

Let $f(z_k) \triangleq \log(1+z_k^{-1})$ and $\bar{z}_k \triangleq E[z_k]$. Expanding $f(z_k)$ in a Taylor series around $\bar{z}_k$ gives

$$f(z_k) = f(\bar{z}_k) + f'(\bar{z}_k)(z_k - \bar{z}_k) + \frac{f''(\bar{z}_k)}{2!}(z_k - \bar{z}_k)^2 + \cdots . \tag{171}$$

The variance of $f(z_k)$ can then be expanded as

$$\mathrm{var}[f(z_k)] = E[(f(z_k) - E[f(z_k)])^2] \tag{172}$$

$$= [f'(\bar{z}_k)]^2\mathrm{var}[z_k] + f'(\bar{z}_k)f''(\bar{z}_k)E[(z_k-\bar{z}_k)^3] + \frac{1}{4}[f''(\bar{z}_k)]^2\mathrm{var}[(z_k-\bar{z}_k)^2] + \cdots . \tag{173}$$

By following steps similar to those used to derive (170), we can show that $N_r E[(z_k - \bar{z}_k)^n] \to 0$ for $n > 2$. Multiplying both sides of (173) by $N_r$ and taking the large system limit gives

$$\lim_{(N_t,N_r,K)\to\infty} N_r \mathrm{var}[\log(1+z_k^{-1})] = \lim_{(N_t,N_r,K)\to\infty} [f'(\bar{z}_k)]^2 N_r \mathrm{var}[z_k] \tag{174}$$

$$= \frac{\bar{N}_r^4}{\bar{K}^2(\sigma_n^2+1)^2(\sigma_n^2\bar{K} + \bar{K} + \bar{N}_r)^2} \lim_{(N_t,N_r,K)\to\infty} N_r \mathrm{var}[z_k] \tag{175}$$

where we have used the fact that $\bar{z}_k \to \bar{K}/\bar{N}_r + \sigma_n^2\bar{K}/\bar{N}_r$ and $f'(x) = [x(x+1)]^{-1}$.

Next, we compute the covariance between the rates of two substreams. This covariance is the same for any two substreams, so that we consider the first and second substreams, i.e.,

$$\begin{aligned}\mathrm{cov}\left[\log(1+\gamma_1)\log(1+\gamma_2)\right] = E\Big[ &(\log(1+\gamma_1) - E\log(1+\gamma_1)) \\ &(\log(1+\gamma_2) - E\log(1+\gamma_2))\Big]\end{aligned} \tag{176}$$

where we have dropped the subscript '$\mathsf{mf}$' from $\gamma_{k;\mathsf{mf}}$ for convenience. We express the SINR for substream 1 as

$$\gamma_1 = \frac{(\tilde{\boldsymbol{v}}_1^\dagger\tilde{\boldsymbol{v}}_1)^2}{K\sigma_n^2 + \sum_{i\neq 1,2}|\tilde{\boldsymbol{v}}_i^\dagger\tilde{\boldsymbol{v}}_1|^2 + |\tilde{\boldsymbol{v}}_1^\dagger\tilde{\boldsymbol{v}}_2|^2} \tag{177}$$

$$= \frac{\gamma_{1:2}}{1 + \frac{|\tilde{\boldsymbol{v}}_1^\dagger\tilde{\boldsymbol{v}}_2|^2}{(\tilde{\boldsymbol{v}}_1^\dagger\tilde{\boldsymbol{v}}_1)^2}\gamma_{1:2}} \tag{178}$$

where $\gamma_{1:2}$ is the output SINR for substream 1 without the interference from substream 2, i.e.,

$$\gamma_{1:2} \triangleq \frac{(\tilde{\boldsymbol{v}}_1^\dagger\tilde{\boldsymbol{v}}_1)^2}{K\sigma_n^2 + \sum_{i\neq 1,2}|\tilde{\boldsymbol{v}}_i^\dagger\tilde{\boldsymbol{v}}_1|^2}. \tag{179}$$

Since for large enough $N_t$, $E[\gamma_{1:2}|\tilde{\boldsymbol{v}}_1^\dagger\tilde{\boldsymbol{v}}_2|^2/(\tilde{\boldsymbol{v}}_1^\dagger\tilde{\boldsymbol{v}}_1)^2] < 1$, (178) can be expanded as

$$\gamma_1 = \gamma_{1:2} - \frac{|\tilde{\boldsymbol{v}}_1^\dagger\tilde{\boldsymbol{v}}_2|^2}{(\tilde{\boldsymbol{v}}_1^\dagger\tilde{\boldsymbol{v}}_1)^2}\gamma_{1:2} + O\left(|\tilde{\boldsymbol{v}}_1^\dagger\tilde{\boldsymbol{v}}_2|^4\right). \tag{180}$$

Therefore,

$$\log(1+\gamma_1) - E\log(1+\gamma_1) = \log\left(\frac{1+\gamma_1}{1+\gamma_{1:2}}\right) + \log(1+\gamma_{1:2}) - E\log(1+\gamma_1) \tag{181}$$

$$= -\frac{\gamma_{1:2}^2|\tilde{\boldsymbol{v}}_1^\dagger\tilde{\boldsymbol{v}}_2|^2}{(1+\gamma_{1:2})(\tilde{\boldsymbol{v}}_1^\dagger\tilde{\boldsymbol{v}}_1)^2} + O\left(|\tilde{\boldsymbol{v}}_1^\dagger\tilde{\boldsymbol{v}}_2|^4\right) + \log(1+\gamma_{1:2}) - E\log(1+\gamma_1) \tag{182}$$

where in (182) we use the Taylor series expansion $\log(1+x) = x + O(x^2)$ for $|x| < 1$. A similar expression can be derived for $\log(1+\gamma_2) - E\log(1+\gamma_2)$. Multiplying the two factors and taking expectation gives

$$\mathrm{cov}\left[\log(1+\gamma_1)\log(1+\gamma_2)\right] = -\frac{\gamma_{1:2}^2\gamma_{2:1}^2}{(1+\gamma_{1:2})(1+\gamma_{2:1})}E\left[\frac{|\tilde{\boldsymbol{v}}_1^\dagger\tilde{\boldsymbol{v}}_2|^4}{(\tilde{\boldsymbol{v}}_1^\dagger\tilde{\boldsymbol{v}}_1)^2(\tilde{\boldsymbol{v}}_2^\dagger\tilde{\boldsymbol{v}}_2)^2}\right] + O\left(|\tilde{\boldsymbol{v}}_1^\dagger\tilde{\boldsymbol{v}}_2|^6\right). \tag{183}$$

As $(N_t, N_r, K) \to \infty$,

$$\gamma_{1:2}, \gamma_{2:1} \longrightarrow \frac{\bar{N}_r}{\bar{K}(1+\sigma_n^2)}, \tag{184}$$

and it can be easily shown that

$$N_r^2 E\left[\frac{|\tilde{\boldsymbol{v}}_1^\dagger\tilde{\boldsymbol{v}}_2|^4}{(\tilde{\boldsymbol{v}}_1^\dagger\tilde{\boldsymbol{v}}_1)^2(\tilde{\boldsymbol{v}}_2^\dagger\tilde{\boldsymbol{v}}_2)^2}\right] \longrightarrow 2. \tag{185}$$

Substituting (184) and (185) into (183), we have

$$\lim_{(N_t,N_r,K)\to\infty} N_r^2\mathrm{cov}\left[\log(1+\gamma_1)\log(1+\gamma_2)\right] = \frac{-2\bar{N}_r^4}{\bar{K}^2(1+\sigma_n^2)^2(\bar{K}(1+\sigma_n^2)+\bar{N}_r)^2}. \tag{186}$$

Since

$$\mathrm{var}[R_{\mathsf{mf}}^{N_r}] = \frac{K}{N_r^2}\mathrm{var}[\log(1+\gamma_k)] + \frac{K^2-K}{N_r^2}\mathrm{cov}[\log(1+\gamma_1)\log(1+\gamma_2)], \tag{187}$$

combining (170), (175), (186), and (187) gives (54).

Similarly, we can show that for the MMSE filter,

$$\lim_{(N_t,N_r,K)\to\infty} N_r\mathrm{var}[\log(1+\gamma_{k;\mathsf{mmse}})] = \frac{\sigma_\gamma^2}{(1+\gamma_{\mathsf{mmse}})^2} \tag{188}$$

$$\lim_{(N_t,N_r,K)\to\infty} N_r^2\mathrm{cov}[\log(1+\gamma_{1;\mathsf{mmse}})\log(1+\gamma_{2;\mathsf{mmse}})] = -\frac{\xi^2}{(1+\gamma_{\mathsf{mmse}})^4} \tag{189}$$

where the expressions for $\gamma_{\mathsf{mmse}}$ and $\sigma_\gamma^2$ are given by (59) and (60), respectively, $\gamma_{1;\mathsf{mmse}}$ and $\gamma_{2;\mathsf{mmse}}$ denote the SINRs at the output of MMSE filters for substreams 1 and 2, respectively, and

$$\xi = \lim_{(N_t,N_r,K)\to\infty} N_r \frac{|\tilde{\boldsymbol{v}}_1^\dagger \boldsymbol{R}_{1,2}\tilde{\boldsymbol{v}}_2|^2}{(\tilde{\boldsymbol{v}}_1^\dagger\tilde{\boldsymbol{v}}_1)^2(\tilde{\boldsymbol{v}}_2^\dagger\tilde{\boldsymbol{v}}_2)^2} \tag{190}$$

where $\boldsymbol{R}_{1,2} = \sum_{i\neq 1,2} \tilde{\boldsymbol{v}}_i\tilde{\boldsymbol{v}}_i^\dagger + K\sigma_n^2\boldsymbol{I}_{N_r}$.

To compute $\xi$, we apply the matrix inversion lemma to write

$$\gamma_{1;\mathsf{mmse}} = \tilde{\boldsymbol{v}}_1^\dagger\boldsymbol{R}_{1,2}^{-1}\tilde{\boldsymbol{v}}_1 - \frac{|\tilde{\boldsymbol{v}}_1^\dagger\boldsymbol{R}_{1,2}^{-1}\tilde{\boldsymbol{v}}_2|^2}{1+\tilde{\boldsymbol{v}}_2^\dagger\boldsymbol{R}_{1,2}^{-1}\tilde{\boldsymbol{v}}_2} \tag{191}$$

and rearrange to obtain

$$\frac{\gamma_{1;\mathsf{mmse}} - \tilde{\boldsymbol{v}}_1^\dagger\boldsymbol{R}_{1,2}^{-1}\tilde{\boldsymbol{v}}_1}{\frac{K}{N_r}-\frac{K-1}{N_r}} = -N_r\frac{|\tilde{\boldsymbol{v}}_1^\dagger\boldsymbol{R}_{1,2}^{-1}\tilde{\boldsymbol{v}}_2|^2}{1+\tilde{\boldsymbol{v}}_2^\dagger\boldsymbol{R}_{1,2}^{-1}\tilde{\boldsymbol{v}}_2}. \tag{192}$$

Note that $\tilde{\boldsymbol{v}}_1^\dagger\boldsymbol{R}_{1,2}^{-1}\tilde{\boldsymbol{v}}_1$ is the SINR for stream 1 without stream 2. As $(N_r,N_t,K)\to\infty$, the left-hand side of (192) converges to $\mathrm{d}\gamma_{\mathsf{mmse}}/\mathrm{d}(\bar{K}/\bar{N}_r)$, and $\tilde{\boldsymbol{v}}_2^\dagger\boldsymbol{R}_{1,2}^{-1}\tilde{\boldsymbol{v}}_2 \to \gamma_{\mathsf{mmse}}$ almost surely. Therefore

$$N_r|\tilde{\boldsymbol{v}}_1^\dagger\boldsymbol{R}_{1,2}^{-1}\tilde{\boldsymbol{v}}_2|^2 \to -(1+\gamma_{\mathsf{mmse}})\frac{\mathrm{d}\gamma_{\mathsf{mmse}}}{\mathrm{d}\bar{K}/\bar{N}_r}. \tag{193}$$

Since both $(\tilde{\boldsymbol{v}}_1^\dagger\tilde{\boldsymbol{v}}_1)^2\to 1$ and $(\tilde{\boldsymbol{v}}_2^\dagger\tilde{\boldsymbol{v}}_2)^2\to 1$ almost surely,

$$\xi = -\frac{1+\gamma_{\mathsf{mmse}}}{\bar{N}_r}\frac{\mathrm{d}\gamma_{\mathsf{mmse}}}{\mathrm{d}\bar{K}} \tag{194}$$

where the derivative $\mathrm{d}\gamma_{\mathsf{mmse}}/\mathrm{d}\bar{K}$ is given by (64).

In analogy with the matched filter,

$$\mathrm{var}[R_{\mathsf{mmse}}^{N_r}] = \frac{K}{N_r^2}\mathrm{var}[\log(1+\gamma_{k;\mathsf{mmse}})] + \frac{K^2-K}{N_r^2}\mathrm{cov}[\log(1+\gamma_{1;\mathsf{mmse}})\log(1+\gamma_{2;\mathsf{mmse}})]. \tag{195}$$

Combining (188), (189), (194), and (195) gives (63).

## 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 of arbitrary dimensions and MIMO communications with feedback,” in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, vol. 3, St. Louis, MO, Dec. 2005, pp. 1–5.

[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] 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.

[20] 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.

[21] 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.

[22] 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.

[23] C. K. Au-Yeung and D. J. Love, "On the performance of random vector quantization limited feedback beamforming," *IEEE Trans. Wireless Commun.*, 2005, submitted for publication.

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

[25] T. Yoo, N. Jindal, and A. Goldsmith, "Finite-rate feedback MIMO broadcast channels with a large number of users," in *Proc. IEEE Int. Symp. on Info. Theory (ISIT)*, Seattle, WA, July 2006.

[26] 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.

[27] 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.

[28] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Kluwer Academic Press, 1992.

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

[30] 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.

[31] S. B. Provost and Y.-H. Cheong, "The distribution of Hermitian quadratic forms in elliptically contoured random vectors," *Journal of Statistical Planning and Inference*, vol. 102, no. 2, pp. 303–316, Apr. 2002.

[32] 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.

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

[34] 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.

[35] 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.

[36] G. S. Rajappan and M. L. Honig, "Signature sequence adaptation for DS-CDMA with multipath," *IEEE J. Select. Areas Commun.*, vol. 20, no. 2, pp. 384–395, Feb. 2002.

[37] 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.*, 2006, submitted for publication.

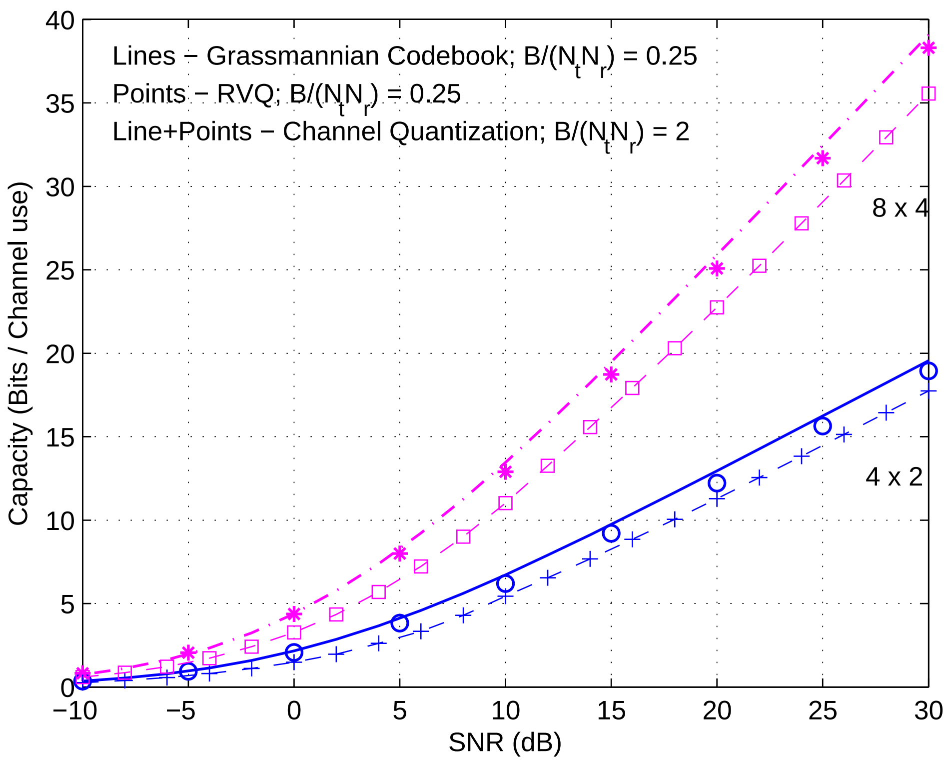
[38] W. Santipach and M. L. Honig, "Capacity of beamforming with limited training and feedback," in *Proc. IEEE Int. Symp. on Info. Theory (ISIT)*, Seattle, WA, July 2006.

[39] ——, "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.

[40] Y. Sun and M. L. Honig, "Asymptotic capacity of multi-carrier transmission over a fading channel with feedback," *IEEE Trans. Info. Theory*, 2004, submitted for publication.

[41] E. Castillo, *Extreme Value Thoery in Engineering*. Academic Press, 1988.

[42] 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.



Lines − Grassmannian Codebook; B/(N_t N_r) = 0.25
Points − RVQ; B/(N_t N_r) = 0.25
Line+Points − Channel Quantization; B/(N_t N_r) = 2
8 x 4
4 x 2
Capacity (Bits / Channel use)
SNR (dB)