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

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

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

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

I. INTRODUCTION

GIVEN a multi-input multi-output (MIMO) channel, providing channel information at the transmitter can increase the achievable rate and simplify the coder and decoder. Namely, this channel information can specify a precoding matrix, which aligns the transmitted signal along the strongest channel modes (i.e., singular vectors corresponding to the largest singular values). In practice, the precoding matrix must be quantized at the receiver, and relayed to the transmitter via a feedback channel. The corresponding achievable rate is therefore limited by the accuracy of the quantizer.

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

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

Here we assume an i.i.d. block Rayleigh fading channel model with independent channel gains, and take ergodic capacity as the performance criterion. The receiver relays \( B \) bits to the transmitter (per codeword) via a reliable feedback channel (i.e., no feedback errors) with no delay. We start by evaluating the capacity of MISO and MIMO channels with a
quantized beamformer, i.e., rank-one precoding matrix. Our results are asymptotic as the number of transmit antennas $N_t$ and feedback bits $B$ both tend to infinity with fixed $B/N_t$ (feedback bits per degree of freedom). For the MIMO channel the number of receive antennas $N_r$ also tends to infinity in proportion with $N_t$ and $B$. The asymptotic expressions accurately predict the performance of finite-size systems of interest as a function of normalized feedback and background Signal-to-Noise Ratio (SNR). In analogy with the optimality interest as a function of normalized feedback and background accurately predict the performance of finite-size systems of rank $K$ remains an open question.

Furthermore, numerical examples for small $N_t$ show that RVQ performance averaged over codebooks is essentially the same as that obtained from codebooks optimized via the Lloyd-Max algorithm [1], [5], [6]. (See also the numerical examples in [22], which compare RVQ performance with the optimized (Grassmannian) codebooks in [3].)

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

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

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

In addition to quantizing the optimal precoding matrix, power for each data stream can also be optimized, quantized, and fed back to the transmitter (e.g., see [23], [24]). Asymptotically, the amount of feedback required to specify the power is negligible compared to the feedback required for the precoding matrix. Furthermore, uniform power over the set of activated channel typically performs close to the optimal (water-filling) performance [8]. We therefore only consider quantization of the precoding matrix.

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

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

## II. Channel Model

We consider a point-to-point, flat Rayleigh fading channel with $N_t$ transmit antennas and $N_r$ receive antennas. Let $x = [x_k]$ be a $K \times 1$ vector of transmitted symbols with covariance matrix $I_K$, where $I_K$ 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

$$y = \frac{1}{\sqrt{K}}HVx + n$$

where $H = [h_{n_r,n_t}]$ is an $N_r \times N_t$ channel matrix, $V = [v_1 v_2 \ldots v_K]$ is an $N_t \times K$ precoding matrix, and $n$ is a complex Gaussian noise $N_r \times 1$ vector with covariance matrix $\sigma_n^2 I_{N_r}$. Assuming rich scattering and Rayleigh fading, the elements of $H$ 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 $V$, the ergodic capacity is the mutual information between $x$ and $y$ with a complex Gaussian distributed input, averaged over the channel, given by

$$I(x;y) = E_H \left[ \log \det \left( I + \rho \frac{HVV^\dagger H^\dagger}{K} \right) \right]$$

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

With unlimited feedback, the columns of the optimal precoding matrix, which maximizes (2), are eigenvectors of the channel covariance matrix $H^\dagger H$. With $B$ feedback bits per
fading block, we can specify the precoding matrix from a quantization set or codebook \( \mathcal{V} = \{V_1, \cdots, V_{2^B}\} \) known \textit{a priori} to both the transmitter and receiver. The receiver chooses the \( V_j \) 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 \( \mathbf{v} \), which ideally corresponds to the strongest channel mode. That is, the optimal \( \mathbf{v} \), which maximizes the ergodic capacity in (2), is the eigenvector of \( \mathbf{H}^\dagger \mathbf{H} \) corresponding to the largest eigenvalue. This vector is computed at the receiver and a quantized version is relayed back to the transmitter.

Let \( \mathcal{V} = \{v_1, \cdots, v_{2^B}\} \) denote the quantization codebook for \( \mathbf{v} \), 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 \( v_1, \cdots, v_{2^B} \) are independent, isotropically distributed random vectors, each with unit norm. This is motivated by the observation that given a channel matrix \( \mathbf{H} \) with \( i.i.d. \) elements, the eigenvectors of \( \mathbf{H}^\dagger \mathbf{H} \) are isotropically distributed [30], hence the codebook entries should be uniformly distributed over the space of beamforming vectors.

#### A. MISO Channel

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

\[
\hat{\mathbf{v}} = \arg \max_{1 \leq j \leq 2^B} \{I_j = \log(1 + \rho \|\hat{\mathbf{h}}\|^2 v_j^\dagger v_j)\}
\]

and the corresponding achievable rate is

\[
I^N_{\text{rvq}} = \max_{1 \leq j \leq 2^B} I_j.
\]

The proof is given in Appendix A. For \( B > 0 \), the rate loss due to finite feedback is a constant. As \( B \to 0 \), this rate loss tends to infinity, since with \( \hat{\mathbf{h}} = \hat{\mathbf{h}}/\|\hat{\mathbf{h}}\| \), the capacity grows as \( \log N_t \). Of course, as \( \hat{\mathbf{h}} \to \infty \) (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

\[
\mathcal{V}_{N_t} = \{v_1^{N_t}, v_2^{N_t}, \ldots, v_{2^B}^{N_t}\}
\]

is the codebook for a particular \( N_t \) and \( \|v_j^{N_t}\|^2 = 1 \) for each \( j \). The associated rate is given by

\[
I_{\mathcal{V}_{N_t}} = \max_{1 \leq j \leq 2^B} \log(1 + \rho \|\hat{\mathbf{h}}\|^2 v_j^\dagger v_j)
\]

and the rate difference \( I^\Delta_{\text{rvq}} = I_{\mathcal{V}_{N_t}} - \log(\rho N_t) \).

#### Theorem 2.

For any sequence of codebooks \( \{\mathcal{V}_{N_t}\} \),

\[
\lim_{(N_t,B) \to \infty} \max_{N_t} E_{\mathbf{h}}[I^\Delta_{\text{rvq}}] \leq I^\Delta_{\text{rvq}}.
\]

The proof is given in Appendix B. Although the optimality of RVQ holds only in the large system limit, numerical results in Section III-C show that for finite-size systems of interest RVQ performs essentially the same as optimized quantization codebooks.
B. Multi-Input Multi-Output (MIMO) Channel

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

We again consider an RVQ codebook \( \mathcal{V} \) with \( 2^B \) independent unit-norm vectors, where each vector is uniformly distributed over the \( N_t \)-dimensional unit sphere. The achievable rate is \( \bar{I}_{rvq}^N_i \), where

\[
\bar{I}_{rvq}^N_i = E_Y \left[ \max_{1 \leq j \leq 2^B} \log \left( 1 + \rho \| H v_j \|^2 \right) \right] \tag{10}
\]

\[
= E_Y \left[ \log(1 + \rho \max_{1 \leq j \leq 2^B} \| H v_j \|^2) \right] \tag{11}
\]

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

\[
\Delta I_{rvq}^N_i \triangleq I_{rvq}^N_i - \log(\rho N_t) = E_Y \left[ \log \left( \frac{1}{\rho N_t} + \max_j \gamma_j \right) \right] \tag{12}
\]

where

\[
\gamma_j = \frac{1}{N_t} v_j^\dagger H^\dagger H v_j. \tag{13}
\]

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

\[
\bar{\gamma}_{rvq}^\infty = \lim_{(N_t, N_r, B) \to \infty} \max_{1 \leq j \leq 2^B} \gamma_j \tag{14}
\]

where convergence to the deterministic limit can be shown in the mean square sense. Conditioned on \( H \), the \( \gamma_j \)'s are i.i.d. since the beamforming vectors \( v_j \) are i.i.d., and applying [32, Theorem 2.1.2], it can be shown that

\[
\bar{\gamma}_{rvq}^\infty = \lim_{(N_t, N_r, B) \to \infty} F^{-1}_{CH} \left( 1 - 2^{-B} \right) \tag{15}
\]

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

\[
\Delta I_{rvq}^\Delta = \lim_{(N_t, N_r, B) \to \infty} \bar{I}_{rvq}^\Delta = \log(\bar{\gamma}_{rvq}^\infty) \tag{16}
\]

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

**Theorem 3.** For \( 0 \leq \bar{B} \leq \bar{B}^\ast \), \( \bar{\gamma}_{rvq}^\infty \) satisfies

\[
(\bar{\gamma}_{rvq}^\infty) \bar{N}_r e^{-\bar{\gamma}_{rvq}^\infty} = 2^{-\bar{B}} \left( \frac{\bar{N}_r}{e} \right) \tag{17}
\]

and for \( \bar{B} \geq \bar{B}^\ast \),

\[
\bar{\gamma}_{rvq}^\infty = (1 + \sqrt{\bar{N}_r})^2 - \exp \left( \frac{1}{2} \bar{N}_r \log(\bar{N}_r) \right) - (\bar{N}_r - 1) \log(1 + \sqrt{\bar{N}_r}) + \sqrt{\bar{N}_r} - \bar{B} \log(2) \tag{18}
\]

where

\[
\bar{B}^\ast = \frac{1}{\log(2)} \left( \bar{N}_r \log \left( \frac{\sqrt{\bar{N}_r}}{1 + \sqrt{\bar{N}_r}} \right) + \sqrt{\bar{N}_r} \right) \tag{19}
\]

The proof is given in Appendix C and is motivated by an analogous result for CDMA, presented in [34]. As stated in Theorem 3, \( \bar{\gamma}_{rvq}^\infty \) 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 \( B = \bar{B}/N_t \),

\[
\limsup_{(N_t, N_r, B) \to \infty} \bar{I}_{rvq}^{\Delta} - \log(\rho N_t) \leq \Delta I_{rvq} \tag{20}
\]

for any sequence of codebooks \( \{ V_{N_t} \} \).

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

**C. Numerical Results**

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

Comparing finite feedback with perfect beamforming, the results show that one feedback bit per complex entry \( (\bar{B} = 1) \) provides more than 50% of the potential gain due to feedback. For both the MISO and MIMO examples shown, the perfect beamforming capacity is nearly achieved with two feedback bits per complex coefficient.
the RVQ codebook for beamforming. Letting
\[ J_j^{N_r} = \frac{1}{N_r} \log \det \left( I_{N_r} + \frac{\rho}{K} H V_j V_j^\dagger H^\dagger \right), \]
the receiver again selects the quantized precoding matrix, which maximizes the mutual information
\[ \hat{V} = \arg \max_{1 \leq j \leq 2^B} J_j^{N_r}. \]

For finite \( N_r \), we define
\[ I_{\text{rvq}}^{N_r} = E_{H} \left[ \max_{1 \leq j \leq 2^B} J_j^{N_r} | H \right] \]
\[ = E_{V} \left[ \frac{1}{N_r} \log \det \left( I_{N_r} + \frac{\rho}{K} H \hat{V} \hat{V}^\dagger H^\dagger \right) \right] \]
and the average sum mutual information per receive antenna with \( B \) feedback bits is then \( E_{H} [I_{\text{rvq}}^{N_r}] \).

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

Since the entries of the RVQ codebook are i.i.d., the mutual informations \( J_j^{N_r}, j = 1, \ldots, 2^B \), are also i.i.d. for a given \( H \). In principle, the large system limit of \( I_{\text{rvq}}^{N_r} \) can be evaluated, in analogy with (15), given the cdf of \( J_j^{N_r} \) given \( H \), denoted as \( F_{J_j^{N_r}|H} \). 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 \( V V^\dagger = I_{N_t} \) and \( K = N_t \) [35]. That is, all channel modes are allocated equal power. As \((N_r, N_t) \rightarrow \infty \) with fixed \( N_r = N_t / N_t \), the capacity per receive antenna is given by
\[ \frac{1}{N_r} \log \det \left( I_{N_r} + \frac{\rho}{N_t} H H^\dagger \right) \rightarrow \int_0^\infty \log (1 + \rho \lambda) g(\lambda) \, d\lambda \]
\[ = I_{\text{rvq}}(\hat{B} = 0) \]
where convergence is in the almost sure sense, and \( g(\lambda) \) is the asymptotic probability density function for a randomly chosen eigenvalue of \( \frac{1}{N_t} H H^\dagger \), and is given by [33]
\[ g(\lambda) = \frac{\sqrt{(\lambda - a)(b - \lambda)}}{2\pi \lambda N_t} \]
for \( a \leq \lambda \leq b \),
\[ a = \left( 1 - \sqrt{N_r} \right)^2 \quad \text{and} \quad b = \left( 1 + \sqrt{N_r} \right)^2 \]
for \( N_r \leq 1 \). The integral in (26) has been evaluated in [36], which gives the closed-form expression
\[ I_{\text{rvq}}(\hat{B} = 0) = \log \rho y + \frac{1 - N_r}{N_r} \log \left( \frac{1}{1 - z} \right) - \frac{z}{N_r} \]
where
\[
y = \frac{1}{2} \left( 1 + \tilde{N}_r + \frac{1}{\rho} + \sqrt{\left( 1 + \tilde{N}_r + \frac{1}{\rho} \right)^2 - 4\tilde{N}_r} \right)
\]
(31)
\[
z = \frac{1}{2} \left( 1 + \tilde{N}_r + \frac{1}{\rho} - \sqrt{\left( 1 + \tilde{N}_r + \frac{1}{\rho} \right)^2 - 4\tilde{N}_r} \right).
\]
(32)

If \( \hat{B} = \infty \), then the \( K \) columns of the optimal \( \mathbf{V} \) are the eigenvectors of the channel covariance matrix corresponding to the \( K \) largest eigenvalues. As \( (N_t, N_r, B) \to \infty \), we have
\[
\mathcal{I}_{\text{req}}(\hat{B} = \infty) = \int_{\eta}^{\infty} \log \left( 1 + \frac{\rho}{K} \lambda \right) g(\lambda) \, d\lambda
\]
(33)
where \( \eta \) satisfies
\[
\int_{\eta}^{\infty} g(\lambda) \, d\lambda = \min \left\{ 1, \frac{K}{\tilde{N}_r} \right\}
\]
(34)
for \( \tilde{N}_r \leq 1 \). We emphasize that this corresponds to a uniform allocation of power over the set of \( K \) active eigenvectors. (This result has also been presented in [8].) The rank of the optimal \( \mathbf{V} \), or optimal multiplexing gain, is at most \( \min \{N_t, N_r\} \) and can be obtained by differentiating (33) with respect to \( K \). It can be verified that \( \mathcal{I}_{\text{req}}(\hat{B} = 0) \leq \mathcal{I}_{\text{req}}(\hat{B} = \infty) \).

To illustrate the increase in capacity with feedback, in Fig. 4 we plot the rate ratio \( \mathcal{I}_{\text{req}}(\hat{B} = \infty)/\mathcal{I}_{\text{req}}(\hat{B} = 0) \) versus SNR for different values of \( \tilde{N}_r \), where \( \mathcal{I}_{\text{req}}(\hat{B} = \infty) \) is optimized over rank \( K \). For large SNR \( \rho \), we can expand
\[
\mathcal{I}_{\text{req}}(\hat{B} = 0) = \log(\rho) + o(\log(\rho))
\]
(35)
\[
\mathcal{I}_{\text{req}}(\hat{B} = \infty) = \log(\rho) \int_{\eta}^{\infty} g(\lambda) \, d\lambda + o(\log(\rho)).
\]
(36)
Therefore
\[
\lim_{\rho \to \infty} \frac{\mathcal{I}_{\text{req}}(\hat{B} = \infty)}{\mathcal{I}_{\text{req}}(\hat{B} = 0)} = \int_{\eta}^{\infty} g(\lambda) \, d\lambda = \min \left\{ 1, \frac{K}{\tilde{N}_r} \right\}
\]
(37)
which implies that the optimal rank \( K^* = \min \{N_t, N_r\} \), and the corresponding asymptotic rate ratio is one. The increase in achievable rate from feedback is small in this case, since for large SNRs, the transmitter excites all channel modes, and the pure 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 \( \log(1 + \rho \lambda/\hat{K}) \) in Taylor series. Taking \( \rho \to 0 \) gives
\[
\lim_{\rho \to 0} \frac{\mathcal{I}_{\text{req}}(\hat{B} = \infty)}{\mathcal{I}_{\text{req}}(\hat{B} = 0)} = \frac{1}{\hat{K}} \int_{\eta}^{\infty} \lambda g(\lambda) \, d\lambda \leq \frac{1}{\tilde{N}_r} \int_{\eta}^{\infty} \frac{\lambda g(\lambda)}{\sqrt{\tilde{N}_r}} \, d\lambda \leq \frac{1}{\tilde{N}_r} \int_{\eta}^{\infty} \frac{b \lambda g(\lambda)}{\sqrt{\tilde{N}_r}} \, d\lambda = \left( 1 + \frac{1}{\sqrt{\tilde{N}_r}} \right)^2
\]
(38)
(39)
where \( b \) is the asymptotic maximum eigenvalue given by (29). The inequality in (38) follows from (34), which implies \( \hat{K} \geq \tilde{N}_r \int_{\eta}^{\infty} g(\lambda) \, d\lambda \). Note that (39) corresponds to allocating transmission power to the strongest channel mode, which is known to maximize capacity at low SNRs. The maximal rate ratio (39) can also be obtained from Theorem 3.

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

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

Evaluating the large system limit of \( \mathcal{I}_{\text{req}}^{\hat{B}} \), assuming that the cdf of \( J_{N_r}^{\hat{B}} \) is Gaussian, gives the approximate rate
\[
\tilde{\mathcal{I}}_{\text{req}} = \mu_j + \sigma_j \sqrt{2\hat{B} \log 2}
\]
(40)
independent of the channel realization, where \( \mu_j \) and \( \sigma^2_j \) are the asymptotic mean of \( J_{N_r}^{\hat{B}} \), and variance of \( N_r^2 J_{N_r}^{\hat{B}} \), respectively. The derivation of (40) is a straightforward extension of [32, Sec. 2.3.2] and is not shown here. As \( \hat{B} \to 0 \), this approximation becomes exact. However, as \( \hat{B} \to \infty \), the approximate rate \( \tilde{\mathcal{I}}_{\text{req}} \to \infty \), whereas the actual rate \( \mathcal{I}_{\text{req}}(\hat{B} = \infty) \) is finite, and can be computed from (33) and (34). This is because \( J_{N_r}^{\hat{B}} \) 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 \( \mathcal{I}_{\text{req}} \) for large \( \hat{B} \). (This implies that we should approximate \( \tilde{\mathcal{I}}_{\text{req}} \) as \( \min \{\mathcal{I}_{\text{req}}, \mathcal{I}_{\text{req}}(\hat{B} = \infty)\} \).)
The asymptotic mean and variance of $J^N_r$ are computed in Appendix D. The asymptotic mean is given by
\[
\mu_j = \frac{K}{N_r} \log \left( 1 + \frac{N_r \rho - \bar{N}_r}{N_r} \right) + \log \left( 1 + \rho - \bar{N}_r \rho \right) v - v \tag{41}
\]
where
\[
v = \frac{1}{2} + \frac{K}{2N_r} \left( 1 + \frac{K}{N_r} \right) \left( 1 + \frac{K}{N_r \rho} \right)^{-2} \frac{2K}{N_r} \tag{42}
\]
The asymptotic variance is approximated for $0 \leq \bar{K} = N_r \leq 1$ and small SNR ($\rho \leq -5$ dB) as
\[
\sigma_j^2 \approx \rho^2 (1 - \bar{N}_r). \tag{43}
\]
The asymptotic variance for moderate SNRs and normalized rank $\bar{K} \neq N_r$ can be computed easily via numerical simulation.\(^1\)

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

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

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

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 $\bar{B} \rightarrow \infty$, the optimal precoding matrix eliminates the cross-coupling among channel modes, and the optimal receiver becomes the linear matched filter. Hence the corresponding achievable rates should be the same in this limit. However, for finite $\bar{B}$ 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 $y$ is passed through the $N_r \times 1$ receive
filter \( c_k \). The matched filter is given by

\[
c_k = \frac{1}{\sqrt{K}} H v_k
\]

where \( v_k \) is the \( k \)th column of the precoding matrix \( V \), and the linear MMSE filter is given by

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

The SINR at the output of the linear filter \( c_k \) is

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

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

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

\[
R^{N_r} = \frac{1}{N_r} \sum_{k=1}^K I(x_k, \hat{x}_k) = \frac{1}{N_r} \sum_{k=1}^K \log(1 + \gamma_k).
\]

where \( \gamma_k \) is the SINR for the \( k \)th data stream. Given a channel matrix \( H \), the sum rate \( R^{N_r} \) depends on the precoding matrix \( V \). We are interested in maximizing \( R^{N_r} \) subject to the power constraint \( \|v_k\|^2 \leq 1 \), \( \forall k \), assuming that the power is allocated equally across streams.

Given the codebook of precoding matrices \( V = \{V_j, 1 \leq j \leq 2^B\} \), the receiver selects the precoding matrix \( \hat{V} = \arg \max_{1 \leq j \leq 2^B} R^{N_r}(V_j) \). (49)

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

A. Matched filter

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

\[
\gamma_{k;mt} = \frac{(v_k^\dagger H^\dagger H v_k)^2}{K \sigma_n^2 (v_k^\dagger H^\dagger H v_k) + \sum_{i=1,i \neq k}^K |v_k^\dagger H^\dagger H v_i|^2}
\]

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

\[
E_{H,V} \left[ \max_{1 \leq j \leq 2^B} \left\{ R^{N_r}_{mt}(V_j) = \frac{1}{N_r} \sum_{k=1}^K \log(1 + \gamma_{k;mt}) \right\} \right]
\]

where the expectation is over the channel matrix and codebook. Since the pdf of \( R^{N_r}_{mt} \) is unknown for finite \( (N_t, N_r, K) \), we are unable to evaluate (51). Motivated by the central limit theorem,\(^2\) in what follows we approximate the cdf of \( R^{N_r}_{mt} \) as Gaussian. The mean is taken to be the asymptotic limit

\[
\mu_{mt} = \lim_{(N_t, N_r, K) \to \infty} R^{N_r}_{mt} = \frac{K}{N_r} \log \left( 1 + \frac{N_r}{K(1 + \sigma_n^2)} \right).
\]

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

\[
N_r^2 \text{var}[R^{N_r}_{mt}(V_j)] \to \sigma^2_{mt}
\]

where \( \sigma^2_{mt} \) can be easily obtained by numerical simulations.

The accuracy of the Gaussian approximation for \( R^{N_r}_{mt} \) is illustrated in Fig. 7, which compares the empirical pdf with the Gaussian approximation for \( N_r = 10, N_r = 1, K \) or \( N_r = 0.3 \) and \( \text{SNR} = 5 \) dB. The difference between the empirical and asymptotic means vanishes as \( (N_t, N_r, K) \to \infty \).

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

\[
\bar{R}_{\text{rvq;mt}} = \lim_{(N_t, N_r, K) \to \infty} \left[ \max_{1 \leq j \leq 2^B} \left\{ R^{N_r}_{mt}(V_j) \right\} \right].
\]

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

\[
\bar{R}_{\text{rvq;mt}} = \mu_{mt} + \sigma_{mt} \sqrt{2B \log 2}.
\]

Numerical results, to be presented, show that this asymptotic approximation is very accurate for small to moderate values of normalized feedback \( B \). As \( B \to 0 \), this approximation becomes exact. However, as \( B \to \infty \), \( \bar{R}_{\text{rvq;mt}} \to \infty \), whereas \( \bar{R}_{\text{rvq;mt}} \) with \( B = \infty \) is the same as the asymptotic rate with
RVQ and an optimal receiver, given by (33) and (34). Hence \( R_{rvq,mt} \) with \( B = \infty \) is finite. As for the analysis of the optimal receiver, this discrepancy is again due to the fact that the cdf of \( R_{mf}^{N_r} \), which has compact support, is being approximated by a Gaussian cdf with infinite support, and also because the dependence among the sum rates \( R_{mf}^{N_r} (\mathbf{V}_j) \) is being ignored.

B. MMSE receiver

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

\[
\gamma_{k:mmse} = v_k^H \left( \sum_{i \neq k} H v_i v_i^H + K \sigma_n^2 I_{N_r} \right)^{-1} H v_k.
\]

(56)

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

\[
R_{\text{mmse}}^{N_r} = \frac{1}{N_r} \sum_{k=1}^{K} \log(1 + \gamma_{k:mmse})
\]

(57)
as a Gaussian pdf with mean

\[
\mu_{\text{mmse}} = \lim_{(N_r,N_r) \to \infty} R_{\text{mmse}}^{N_r} = \frac{K}{N_r} \log(1 + \gamma_{\text{mmse}})
\]

(58)

where the large system SINR is given by [30]

\[
\gamma_{\text{mmse}} = \frac{1 - K/N_r}{2\sigma_n^2} - \frac{1}{2} + \sqrt{\left( \frac{1 - K/N_r}{2\sigma_n^2} \right)^2 + \frac{1 + K/N_r}{2\sigma_n^2} + \frac{1}{4}}.
\]

(59)

As for the matched filter, the asymptotic variance \( \sigma_{\text{mmse}}^2 \) can be obtained via numerical simulation.

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

\[
R_{rvq,mmse} \approx \tilde{R}_{rvq,mmse} = \mu_{\text{mmse}} + \sigma_{\text{mmse}} \sqrt{2B \log 2}.
\]

(60)

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

C. Numerical Results

Fig. 8 compares the approximation for asymptotic RVQ performance with a matched filter receiver from (55) with simulated results for \( N_t = 12, N_r = 0.75, K/N_t = 1/2 \), and SNR = 5 dB. Also shown for comparison are the asymptotic rate for RVQ with an optimal receiver, derived in Section IV, the water-filling capacity (\( B = \infty \)), and the rate achieved with a scalar quantizer for each coefficient.\(^3\) For the case shown, the analytical approximation gives an accurate estimate of the performance of the finite size system with limited feedback. The capacity with the water-filling power allocation is only slightly greater than that achieved with the on-off power allocation. The optimal receiver requires \( B \approx 0.6 \) bit/dimension to achieve the capacity corresponding to unlimited feedback (33), whereas the matched filter requires 1.2 feedback bits per dimension to reach that capacity. For other target rates, these curves illustrate the trade-off between feedback and receiver complexity.

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

VI. Conclusions

We have studied the capacity of single-user MISO and MIMO fading channels with limited feedback. The feedback
specifies a transmit precoding matrix, which can be optimized for a given channel realization. We first considered the performance with a rank-one precoding matrix (beamformer), and showed that the RVQ codebook is asymptotically optimal. Exact expressions for the asymptotic mutual information for MISO and MIMO channels were presented, and reveal how much feedback is required to achieve a desired performance. For the cases considered, one feedback bit for each precoder coefficient can achieve close to the water-filling capacity.

Perhaps more important than the increase in capacity provided by this feedback is the associated simplification in the coding and decoding schemes that can achieve a rate close to capacity.

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

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

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

We have also assumed that the channel gains are not frequency-selective. Limited feedback schemes for frequency-selective scalar channels are discussed in [43], and could be combined with the quantization schemes considered here. Finally, the approach presented here for a single-user MIMO channel can also be applied to multi-user models. Quantization of beamformers for the MIMO downlink have been considered in [26]–[28]. In that scenario the potential capacity gain due to feedback is generally much more than for the single-user channel considered here. The benefits of limited feedback for related models (e.g., frequency-selective MIMO downlink) are currently being studied.

APPENDIX

A. Proof of Theorem 1

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

$$\hat{v} = \arg \max_{1 \leq j \leq 2^B} \left\{ Y_j = |v_j^T h|^2/\|h\|^2 \right\}. \quad (61)$$

Since the codebook entries $v_j$, $j = 1, \ldots, 2^B$, are i.i.d., the $Y_j$’s, given $h$, are also i.i.d. with the cdf [25]

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

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

$$\max_j Y_j - a_n \xrightarrow{D} Y \quad (63)$$

where $Y$ is a Weibull random variable having distribution

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

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

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

where the constant $\gamma > 0$.

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

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

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

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

and

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

$$= 1 - F_{Y|h}^{-1} \left( 1 - \frac{1}{n} \right) = \left( \frac{1}{n} \right)^{N_t - 1}. \quad (70)$$
To take the limit as $N_t \to \infty$, we will assume that the channel vector $\mathbf{h}$ contains the first $N_t$ elements of an infinite-length i.i.d. complex Gaussian vector $\mathbf{h}$. Rearranging terms in (63) and taking the large system limit gives
\begin{equation}
\lim_{(N_t, n) \to \infty} E[Y_j \max_j Y_j^2 | \mathbf{h}] 
= \lim_{(N_t, n) \to \infty} \frac{a_n}{a_n + b_n} E[\mathcal{Y}]
= 1 - \lim_{(N_t, n) \to \infty} \left( \frac{1}{n} \right)^{\frac{1}{N_t}} \Gamma \left( 1 - \frac{1}{N_t} \right)
= 1 + \frac{2 B}{N_t} - 2^B - \frac{B}{N_t} - 1
\end{equation}
where the gamma function $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} \, dt$ and we have used the fact that $E[Y] = -\Gamma(1 - 1/(N_t - 1))$ [44].

From (63), as $(n, N_t) \to \infty$,
\begin{equation}
\text{Var}\left[\max_j Y_j^2 | \mathbf{h}\right] - b_n^2 \text{Var}[\mathcal{Y}] \to 0.
\end{equation}
Since $\text{Var}[\mathcal{Y}] = \Gamma(1 - 2/(N_t - 1)) - 1 - 2(1 - 1/(N_t - 1)) \to 0$, it follows that $\text{Var}\left[\max_j Y_j^2 | \mathbf{h}\right] \to 0$. This establishes that given $\mathbf{h}$,
\begin{equation}
\max_j Y_j^2 \to 1 - 2^B
\end{equation}
in the mean square sense. The asymptotic rate difference is given by
\begin{equation}
\Delta I_{\text{RVQ}} = \lim_{(N_t, B) \to \infty} \left( \log(1 + \rho \|\mathbf{h}\|^2 \max_{1 \leq j \leq 2^B} Y_j) \right) - \log(\rho N_t)
= \lim_{(N_t, B) \to \infty} \left( \log \left( \frac{1}{\rho N_t} \|\mathbf{h}\|^2 \max_{1 \leq j \leq 2^B} Y_j \right) \right)
= \log(1 - 2^B)
\end{equation}
in the mean square sense, since $\|\mathbf{h}\|^2/N_t \to 1$ almost surely.

B. Proof of Theorem 2

The rate difference associated with codebook $\mathcal{V}_{N_t}$ is
\begin{equation}
I_{\mathcal{V}_{N_t}}^\Delta = \max_{1 \leq j \leq 2^B} \log \left( \frac{1}{\rho N_t} + \frac{1}{N_t} \|\mathbf{h}\|^2 Y_j^2 \right)
= \log \left( \frac{1}{\rho N_t} + \max_{1 \leq j \leq 2^B} \frac{1}{N_t} \|\mathbf{h}\|^2 Y_j^2 \right).
\end{equation}
Taking expectation with respect to $\mathbf{h}$ and applying Jensen’s inequality, we obtain
\begin{equation}
E[I_{\mathcal{V}_{N_t}}^\Delta] \leq \log \left( \frac{1}{\rho N_t} + E[\|\mathbf{h}\|^2 Y_j^2] \right)
= \log \left( \frac{1}{\rho N_t} + E[\|\mathbf{h}\|^2] E[\|\mathbf{h}\|^2] \right)
= \log \left( \frac{1}{\rho N_t} + \frac{1}{N_t} \|\mathbf{h}\|^2 \mu \right)
\end{equation}
where the optimal beamforming vector
\begin{equation}
\hat{\mathbf{v}}_{N_t} = \arg \max_{1 \leq j \leq 2^B} \|\mathbf{h}|^2 Y_j^2 |\mathbf{v}_j\|^2, \mu = \|\mathbf{h}|^2 Y_j^2 |\mathbf{v}_j\|^2 / \|\mathbf{h}\|^2, \text{and} \mu \text{ are independent [25].}
\end{equation}

We now derive an upper bound for $E[\|\hat{\mathbf{v}}_{N_t}\|^2 / \|\mathbf{h}\|^2]$.

From (30) in [2] we have
\begin{equation}
\Pr\{\|\hat{\mathbf{v}}_{N_t}\|^2 > \gamma s | \|\mathbf{h}\|^2 = \gamma\}
\leq \left\{ \begin{array}{ll}
1, & 2^B (1 - s) N_t - 1, \quad 0 \leq s < s^* \leq 1
\end{array} \right.
\end{equation}
where
\begin{equation}
s^* = 1 - 2^{-\frac{B}{N_t}}.
\end{equation}
Since the right-hand side of (86) is independent of $\gamma$, averaging over $\gamma$ gives
\begin{equation}
\Pr\{\mu > s\} \leq \left\{ \begin{array}{ll}
1, & 2^B (1 - s) N_t - 1, \quad 0 \leq s < s^* \leq 1
\end{array} \right.
\end{equation}
Integrating by parts, we have that
\begin{equation}
E[\mu] = \int_0^1 \Pr\{\mu > x\} \, dx
= \int_0^{s^*} \Pr\{\mu > x\} \, dx + \int_{s^*}^1 \Pr\{\mu > x\} \, dx.
\end{equation}
Substituting (88) into (90) and evaluating both integrals gives
\begin{equation}
E[\mu] \leq 1 - 2^{-\frac{B}{N_t}} + \frac{1}{N_t} 2^{-\frac{B}{N_t}}.
\end{equation}
Substituting $E[\|\mathbf{h}\|^2] = N_t$ and (91) into (84) gives
\begin{equation}
E[I_{\mathcal{V}_{N_t}}^\Delta] \leq \log \left( 1 - 2^{-\frac{B}{N_t}} + \frac{1}{N_t} 2^{-\frac{B}{N_t}} + \frac{1}{\rho N_t} \right)
\end{equation}
and taking the large system limit gives
\begin{equation}
\lim_{(N_t, B) \to \infty} E[I_{\mathcal{V}_{N_t}}^\Delta] \leq \log(1 - 2^{-B}).
\end{equation}

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

C. Proof of Theorem 3

We first prove the theorem for $N_t \geq 1$. Let $z = F_{\gamma | H}(1 - 2^{-B})$. Rearranging (15) gives
\begin{equation}
\lim_{z \to \gamma N_t} \left[ 1 - F_{\gamma | H}(z) \right] \frac{1}{N_t} = 2^{-B}.
\end{equation}
Next, we derive upper and lower bounds for the left-hand side of (94) and show that they are the same. The derivation of the upper bound is motivated by the evaluation of a similar bound for CDMA signature optimization in [34]. That is,
\begin{equation}
1 - F_{\gamma | H}(z) = \Pr\{ \gamma_j > z | \hat{\mathbf{H}} \}
= \Pr\{ \mathbf{v}^* \mathbf{U} \mathbf{A} \mathbf{U}^* \mathbf{v}^*_j > z | \mathbf{A}, \mathbf{U} \}
= \Pr\{ \mathbf{w}_j \mathbf{U} \mathbf{A} \mathbf{U}^* \mathbf{w}_j > z | \mathbf{A}, \mathbf{U} \}
\end{equation}
where $\gamma_j = \frac{1}{N_t} \|\mathbf{h}|^2 Y_j^2 \mathbf{v}^*_j$ and we have applied the singular value decomposition $\frac{1}{N_t} \mathbf{H}^* \mathbf{H} = \mathbf{U} \mathbf{A} \mathbf{U}^*$, where $\mathbf{U}$ is an $N_t \times N_t$ unitary matrix, $\mathbf{A} = \text{diag}\{\lambda_1, \ldots, \lambda_{N_t}\}$, and the eigenvalues are ordered as $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{N_t}$. Also, $\mathbf{w}_j$
is an $N_i \times 1$ vector with independent, circularly symmetric, zero-mean and unit-variance Gaussian elements. Both $v_j$ and $w_j/\|w_j\|$ are isotropically distributed, i.e., $U(w_j/\|w_j\|)$ and $w_j/\|w_j\|$ have the same distribution, so that

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

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

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

$$= \frac{N_i}{\prod_{i=1}^{N_i} \left[ \exp \left\{ -\rho (z - \lambda_i)w_i^2 \right\} \right| \Lambda]$$

$$= \frac{N_i}{\prod_{i=1}^{N_i} \int_0^\infty \exp \{-\rho (z - \lambda_i) x \} e^{-x} dx}$$

$$= \frac{N_i}{\prod_{i=1}^{N_i} \int_0^\infty \exp \{-(1 + \rho (z - \lambda_i)) x \} dx}$$

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

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

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

$$\lim_{(N_i, N_r) \to \infty, \rho \gamma_{\text{rvq}} \to \infty} [1 - F_{\gamma | \Lambda}(z)] \xrightarrow{\text{D}} \exp \{-\Phi(\gamma_{\text{rvq}}, \rho)\}$$

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

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

$g(\lambda)$ is given by (28)-(29), and $\lambda_{\text{max}} = \lim_{(N_i, N_r) \to \infty} \lambda_1 = (1 + \sqrt{N_i})^2$. To tighten the upper bound, we minimize (103) with respect to $\rho$, i.e.,

$$\lim_{(N_i, N_{r}) \to \infty, \rho \gamma_{\text{rvq}} \to \infty} [1 - F_{\gamma | \Lambda}(z)] \xrightarrow{\text{D}} \exp \{-\Phi(\gamma_{\text{rvq}}, \rho^*)\}$$

where

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

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

To derive the lower bound, we use a change of measure. (A similar approach was used in [45, Section 1.2].) Let $y_i \triangleq (\lambda_i - z)w_i^2$, which is a scaled exponential random variable with cdf $F_1(\cdot)$. We define the new distribution

$$G_i(x) \triangleq \frac{1}{M_i(\rho^*)} \int_{-\infty}^x e^{\rho^* x} dF_1(y),$$

so that

$$M_i(\rho^*)dG_i(x) = e^{\rho^* x} dF_1(x),$$

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

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

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

where

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

$$= \prod_{i=1}^{N_i} \left[ \exp \left\{ -\rho \sum_{i=1}^{N_i} (z - \lambda_i)w_i^2 \right\} \right| \Lambda]$$

$$= \prod_{i=1}^{N_i} \int_0^\infty \exp \{-\rho (z - \lambda_i) x \} e^{-x} dx$$

$$= \prod_{i=1}^{N_i} \int_0^\infty \exp \{-(1 + \rho (z - \lambda_i)) x \} dx$$

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

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

$$\leq \exp \left\{ -\Phi(\gamma_{\text{rvq}}, \rho)\right\}$$

$$\geq \prod_{i=1}^{N_i} M_i(\rho^*) \int \int \int_0^\infty \frac{e^{-\rho^* \gamma_{\text{rvq}}}}{M_i(\rho^*)} dG_1(y_1) \cdots dG_{N_i}(y_{N_i})$$

$$= \prod_{i=1}^{N_i} M_i(\rho^*) e^{-\rho^* \gamma_{\text{rvq}}} \Pr \{ \epsilon N_i \geq \sum_{i=1}^{N_i} y_i > 0 \}$$

For any $\epsilon > 0$,

$$\lim_{(N_i, N_r) \to \infty, \rho \gamma_{\text{rvq}} \to \infty} [1 - F_{\gamma | \Lambda}(z)] \xrightarrow{\text{D}} \exp \{-\Phi(\gamma_{\text{rvq}}, \rho^*)\}$$
where the \( \tilde{y}_i \)'s are independent random variables with cdf \( G_i(\cdot) \), and the second inequality follows since \( \rho^* > 0 \). To determine the probability on the right-hand side of (125), we first compute the mean of \( \tilde{y}_i \).

\[
m_i = \int y \, dG_i(y) = \frac{1}{M_i(\rho^*)} \int y e^{\rho^* y} dF_i(y) = (1 + \rho^*(z - \lambda_i)) \int_0^\infty (\lambda_i - z) x e^{(\lambda_i - z)x - x} \, dx
\]

\[
= \frac{\lambda_i - z}{1 + \rho^*(z - \lambda_i)}.
\]

Therefore

\[
\frac{1}{N_t} \sum_{i=1}^{N_t} m_i = \frac{1}{N_t} \sum_{i=1}^{N_t} \frac{\lambda_i - z}{1 + \rho^*(z - \lambda_i)}
\]

and the asymptotic mean

\[
m_\infty = \lim_{N_t \to \infty} \frac{1}{N_t} \sum_{i=1}^{N_t} m_i = \int_a^b \frac{\lambda - \gamma_{eq} - \lambda}{1 + \rho^*(\gamma_{eq} - \lambda)} g(\lambda) \, d\lambda < \infty.
\]

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

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

and the asymptotic variance

\[
\sigma_\infty^2 = \lim_{N_t \to \infty} \frac{1}{N_t} \sum_{i=1}^{N_t} \sigma_i^2 = \int_a^b \left( \frac{\lambda - \gamma_{eq} - \lambda}{1 + \rho^*(\gamma_{eq} - \lambda)} \right)^2 g(\lambda) \, d\lambda < \infty.
\]

Both the asymptotic mean and variance are finite.

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

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

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

\[
\Pr\{0 < \sum_{i=1}^{N_t} \tilde{y}_i \leq \epsilon N_t\}
= \Pr\{-\sqrt{N_t} a_{N_t} \leq T \leq \sqrt{N_t} a_{N_t} + \sqrt{N_t} \epsilon \frac{1}{b_{N_t}}\}
= F_T(-\sqrt{N_t} a_{N_t}) - F_T(-\sqrt{N_t} a_{N_t} + \sqrt{N_t} \epsilon \frac{1}{b_{N_t}})
\]

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

\[
a_{N_t} \triangleq \sum_{i=1}^{N_t} \frac{m_i}{N_t} \rightarrow m_\infty, \quad \sigma_\infty \quad \text{as} \quad N_t \to \infty.
\]

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

\[
\Pr\{0 < \sum_{i=1}^{N_t} \tilde{y}_i \leq \epsilon N_t\} = \phi(-\sqrt{N_t} a_{N_t} + \sqrt{N_t} \epsilon \frac{1}{b_{N_t}}) - \phi(-\sqrt{N_t} a_{N_t} + \epsilon N_t - \xi_{N_t})
\]

where

\[
\xi_{N_t} \triangleq F_T(-\sqrt{N_t} a_{N_t} + \sqrt{N_t} \epsilon \frac{1}{b_{N_t}}),
\]

\[
\zeta_{N_t} \triangleq F_T(-\sqrt{N_t} a_{N_t} + \epsilon N_t - \xi_{N_t}).
\]

Applying the Berry-Esseen theorem [46], we can bound both \( \zeta_{N_t} \) and \( \xi_{N_t} \) for large \( N_t \) as

\[
|\zeta_{N_t}|, |\xi_{N_t}| \leq \frac{C}{\sqrt{N_t}}
\]

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

We can now evaluate

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

\[
= \frac{1}{\sqrt{2\pi}} \int_{-\sqrt{N_t} a_{N_t}}^{\sqrt{N_t} \epsilon \frac{1}{b_{N_t}}} e^{-t^2/2} \, dt
\]

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

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

\[
\Pr\{0 < \sum_{i=1}^{N_t} \tilde{y}_i \leq \epsilon N_t\} = O(1/\sqrt{N_t})
\]

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

\[
\lim_{(N_t, N_r) \to \infty} \Pr\{0 < \sum_{i=1}^{N_t} \tilde{y}_i \leq \epsilon N_t\} = 1.
\]

Taking the \( N_t \)th root and large system limit on both sides of (125) gives

\[
\lim_{(N_t, N_r) \to \infty} \left[ 1 - F_{\gamma\|A(\cdot)}(\epsilon) \right]^{1/N_t} \geq \exp\{-\Phi(\gamma_{eq} \rho^*)\}
\]

where we use (148) and let \( \epsilon \to 0 \).
The lower bound in (149) is exactly the upper bound (111). Therefore,
\[
\lim_{(N_r, N_q) \to \infty} \left( 1 - F_{|A|}(z) \right)^{\frac{1}{N_r}} = \exp\{-\Phi(\gamma_{rvq}^\infty, \rho^*)\}
\]
(150)
\[
= 2^{-B}
\]
(151)
and the asymptotic RVQ received power satisfies the fixed-point equation
\[
\Phi(\gamma_{rvq}^\infty, \rho^*) = \bar{B} \log(2)
\]
(152)
where \( \rho^* \) is given by (112). The goal of the rest of the proof is to simplify (152).

To determine \( \rho^* \), we first compute
\[
\frac{\partial \Phi(\gamma_{rvq}^\infty, \rho)}{\partial \rho} = \int_a^b \left[ \frac{\gamma_{rvq}^\infty - 1}{(\gamma_{rvq}^\infty - \lambda)^2} \right] g(\lambda) \, d\lambda
\]
(153)
\[
= \frac{1}{\rho} + \rho^2 \int_a^b \lambda \left( \frac{1}{\rho - (\gamma_{rvq}^\infty)} \right) g(\lambda) \, d\lambda
\]
(154)
\[
= \frac{1}{\rho} + \rho^2 \mathcal{S}_\Lambda(y)
\]
(155)
where \( \mathcal{S}_\Lambda(\cdot) \) is the Stieltjes Transform of the asymptotic eigenvalue distribution of \( \Lambda \). Setting the derivative to zero and solving for \( \rho \) gives
\[
\mathcal{S}_\Lambda(y) = -\rho = \frac{1}{\gamma_{rvq}^\infty - y}
\]
(156)
as the only valid solution. Substituting the expression for \( \mathcal{S}_\Lambda \) given in (33) into (156) gives
\[
\left( \hat{N}_r - y \right) \pm \sqrt{y^2 - 2(\hat{N}_r + 1)y + (\hat{N}_r - 1)^2}
\]
\[
= \frac{1}{\gamma_{rvq}^\infty - y}
\]
(157)
which simplifies to the quadratic equation
\[
(\hat{N}_r - \gamma_{rvq}^\infty)y^2 + [\gamma_{rvq}^\infty + \hat{N}_r - \gamma_{rvq}^\infty + (\gamma_{rvq}^\infty)^2]y = 0.
\]
(158)
Solving for \( y \) gives \( y = 0 \) or \( y = \gamma_{rvq}^\infty \left[ 1 + 1/(\gamma_{rvq}^\infty - \hat{N}_r) \right] \), or equivalently, \( \rho = -1/\gamma_{rvq}^\infty \) or \( \rho = (\gamma_{rvq}^\infty - \hat{N}_r)/\gamma_{rvq}^\infty \). Since \( \rho > 0 \), we must have
\[
\rho^* = \frac{\gamma_{rvq}^\infty - \hat{N}_r}{\gamma_{rvq}^\infty}.
\]
(159)
Since
\[
\frac{\partial^2 \Phi(\gamma_{rvq}^\infty, \rho)}{\partial \rho^2} = -\int_a^b \left( \frac{\lambda - \gamma_{rvq}^\infty}{1 + \rho^*(\gamma_{rvq}^\infty - \lambda)} \right)^2 g(\lambda) \, d\lambda
\]
(160)
\[
< 0,
\]
(161)
therefore \( \rho^* \) achieves a maximum.

By also evaluating \( \Phi(\gamma_{rvq}^\infty, \rho) \) at the boundary points \( \rho = 0 \) and \( \rho = 1/(\gamma_{max}^\infty - \gamma_{rvq}^\infty) \), we have
\[
\rho^* = \begin{cases} 
\frac{\gamma_{rvq}^\infty - \hat{N}_r}{\gamma_{rvq}^\infty} & \hat{N}_r \leq \gamma_{rvq}^\infty \leq \hat{N}_r + \sqrt{\hat{N}_r} \\
\frac{1}{(1 + \sqrt{\hat{N}_r})^2 - \gamma_{rvq}^\infty} & \hat{N}_r + \sqrt{\hat{N}_r} \leq \gamma_{rvq}^\infty < (1 + \sqrt{\hat{N}_r})^2.
\end{cases}
\]
(162)
To evaluate \( \Phi(\gamma_{rvq}^\infty, \rho^*) \), we re-write (110) as
\[
\Phi(\gamma_{rvq}^\infty, \rho^*)
\]
\[
= \int_a^b \left[ \log(\rho^*) + \log\left( \frac{1}{\rho^*} + \gamma_{rvq}^\infty - \lambda \right) \right] g(\lambda) \, d\lambda
\]
(163)
\[
= \log(\rho^*) + \int_a^b \log\left( \frac{1}{\rho^*} + \gamma_{rvq}^\infty - \lambda \right) g(\lambda) \, d\lambda.
\]
(164)
To evaluate the integral in (164), we apply the following Lemma.

**Lemma 1.** For \( x \geq (1 + \sqrt{N_r})^2 \),
\[
\Theta(x) \triangleq \int_a^b \log(x - \lambda) g(\lambda) \, d\lambda
\]
(165)
\[
= \log(w(x)) + \sqrt{N_r} u(x) - (\hat{N}_r - 1) \log(1 + \frac{u(x)}{\sqrt{N_r}})
\]
(166)
where
\[
w(x) = \frac{(x - 1 - \hat{N}_r) + \sqrt{(x - 1 - \hat{N}_r)^2 - 4N_r}}{2}
\]
(167)
\[
u(x) = \frac{(x - 1 - \hat{N}_r) - \sqrt{(x - 1 - \hat{N}_r)^2 - 4N_r}}{2N_r}
\]
(168)
The proof of this Lemma is similar to that given in [36] and is therefore omitted here.
For \( \hat{N}_r + \sqrt{N_r} \leq \gamma_{rvq}^\infty < (1 + \sqrt{N_r})^2 \), we substitute \( \rho^* = \frac{\gamma_{rvq}^\infty - \hat{N}_r}{\gamma_{rvq}^\infty} \) into (164) to obtain
\[
\Phi(\gamma_{rvq}^\infty, (1 + \sqrt{N_r})^2 - \gamma_{rvq}^\infty - 1)
\]
(169)
\[
= -\log[(1 + \sqrt{N_r})^2 - \gamma_{rvq}^\infty] + \Theta \left( 1 + \sqrt{N_r} \right)^2
\]
(170)
\[
= -\log[(1 + \sqrt{N_r})^2 - \gamma_{rvq}^\infty] + 1 \hat{N}_r \log(\hat{N}_r)
\]
(171)
\[
= \sqrt{\hat{N}_r} \log(2).
\]
(172)
Solving for \( \gamma_{rvq}^\infty \) gives (18). Taking \( \gamma_{rvq}^\infty = \hat{N}_r + \sqrt{N_r} \) and solving for \( B \) gives \( B^* \) in (19).
For \( \hat{N}_r \leq \gamma_{rvq}^\infty < \hat{N}_r + \sqrt{N_r} \), or \( 0 \leq B \leq B^* \), we substitute
\[
\rho^* = \frac{\gamma_{rvq}^\infty - \hat{N}_r}{\gamma_{rvq}^\infty} \quad \text{into (164) to obtain}
\]
\[
\Phi \left( \psi - \hat{N}_r, \frac{\psi}{\psi + N_r} \right) = \log(\psi - \hat{N}_r) - \log(\psi + \hat{N}_r)
\]
(173)
\[
+ \Theta \left( \frac{\gamma_{rvq}^\infty - \hat{N}_r}{\gamma_{rvq}^\infty} \right).
\]
(174)
To simplify (173), we let \( \psi \triangleq \gamma_{rvq}^\infty - \hat{N}_r \) and re-write (173) as
\[
\Phi \left( \psi - \hat{N}_r, \frac{\psi}{\psi + N_r} \right) = \log(\psi) - \log(\psi + \hat{N}_r)
\]
(175)
\[
+ \Theta \left( 1 + \hat{N}_r + \psi + \frac{N_r}{\psi} \right).
\]
(176)
and
\[
\Theta \left( 1 + \tilde{N}_r + \psi + \frac{\tilde{N}_r}{\psi} \right) = \log(\tilde{N}_r) - \log(\psi) \\
-(\tilde{N}_r - 1) \log \left( 1 + \frac{\psi}{\tilde{N}_r} \right) + \psi. 
\]
(177)
Substituting (177) into (174), we obtain
\[
\Phi \left( \psi - \tilde{N}_r, \psi + \frac{\tilde{N}_r}{\psi} \right) \]
\[
= \psi - \tilde{N}_r \log \left( 1 + \frac{\psi}{\tilde{N}_r} \right) \\
= \gamma_{\text{eq}} \tilde{N}_r - \tilde{N}_r \log(\gamma_{\text{eq}}) + \tilde{N}_r \log(\tilde{N}_r). 
\]
(178)
(179)
Setting this to \( \bar{B} \log(2) \) and simplifying gives (17).

For \( \tilde{N}_r < 1 \), the asymptotic eigenvalue density of \( \frac{1}{N_r} \mathbf{H}^\dagger \mathbf{H} \) is given by
\[
g(\lambda) = (1 - \tilde{N}_r) \delta(\lambda) + \frac{\sqrt{(\lambda - a)(b - \lambda)}}{2\pi \lambda}. 
\]
(180)
where \( a \) and \( b \) are given by (28)-(29). Following the same steps again from (110) gives (18) and (17). This completes the proof of Theorem 3.

D. Derivation of (41)-(43)

To compute \( \mu_j \), we first write
\[
J_{N_r}^{N_r} = \frac{1}{N_r} \sum_{k=1}^{N_r} \log \left( 1 + \rho \frac{N_r}{N_r} \eta_k \right) 
\]
(181)
where \( \eta_k \) is the \( k \)th eigenvalue of \( \mathbf{Y} = \mathbf{N}_r \mathbf{H} \mathbf{V}_j \mathbf{V}_j^\dagger \mathbf{H}^\dagger \).
As \( (N_l, N_r, K) \to \infty \), the empirical eigenvalue distribution converges to a deterministic function \( F_\mathbf{Y}(t) \). The asymptotic mean is given by
\[
\mu_j = \lim_{(N_l, N_r, K) \to \infty} E[J_{N_r}^{N_r}] 
= \int_0^\infty \log \left( 1 + \rho \frac{N_r}{N_r} \right) dF_\mathbf{Y}(t). 
\]
(182)
A similar integral has been evaluated in [36, Eq. (6)], and the result can be directly applied to (182), giving (41).

To compute the variance, we express \( J_{N_r}^{N_r} \) differently by first performing the singular value decomposition \( \mathbf{H} = \mathbf{V}_\mathbf{H} \Sigma_\mathbf{H} \mathbf{U}_\mathbf{H}^\dagger \), where \( \mathbf{V}_\mathbf{H} \) is the \( N_r \times N_r \) left singular matrix, \( \mathbf{U}_\mathbf{H} \) is the \( N_r \times N_r \) right singular matrix, and \( \Sigma_\mathbf{H} \) is an \( N_r \times N_r \) diagonal matrix. Here we assume that \( N_l \geq N_r \).
(For the result for \( N_l < N_r \) can be shown by a similar approach.)
We therefore have
\[
J_{N_r}^{N_r} = \frac{1}{N_r} \log \det (\mathbf{I}_N + \rho \mathbf{A} \mathbf{L}_j) 
\]
(183)
\[
= \frac{1}{N_r} \sum_{i=1}^{N_r} \log (1 + \rho \eta_i) 
\]
(184)
where \( \mathbf{A} = \frac{1}{K} \Sigma_\mathbf{H}^2 \), \( \mathbf{L}_j = \mathbf{U}_\mathbf{H}^j \mathbf{V}_j \mathbf{V}_j^\dagger \mathbf{U}_\mathbf{H}^\dagger \), and \( \eta_i \) is the \( i \)th eigenvalue of \( \mathbf{A} \mathbf{L}_j \). To compute \( \text{var}[J_{N_r}^{N_r}] \), 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 \( J_{N_r}^{N_r} \), we substitute a Taylor series expansion for \( \log(1 + \delta x) \) into (184) to write
\[
J_{N_r}^{N_r} = \frac{\rho}{N_r} \sum_{i=1}^{N_r} \eta_i - \frac{\rho^2}{2N_r} \sum_{i=1}^{N_r} \eta_i^2 + \frac{\rho^3}{3N_r} \sum_{i=1}^{N_r} \eta_i^3 + \ldots 
\]
(185)
\[
= \frac{\rho}{N_r} \text{tr}\{\mathbf{A} \mathbf{L}\} - \frac{\rho^2}{2N_r} \text{tr}\{(\mathbf{A} \mathbf{L})^2\} + \frac{\rho^3}{3N_r} \text{tr}\{(\mathbf{A} \mathbf{L})^3\} + \ldots 
\]
(186)
for \( \rho \eta_{\text{max}} < 1 \), where \( \eta_{\text{max}} = \max_i \eta_i \), and is the maximum eigenvalue of \( \mathbf{H} \mathbf{V}_j \mathbf{V}_j^\dagger \mathbf{H}^\dagger \). Since \( \mathbf{H} \mathbf{V}_j \) is \( N_r \times K \) and i.i.d., \( \eta_{\text{max}} \) has asymptotic value \( (1 + \sqrt{N_r/K})^2 \). If \( K/N_r = 1 \), then the condition asymptotically becomes \( \rho < 1/4 \) (-6 dB).
Ignoring the terms of order \( \rho^3 \) and higher, we can approximate the variance of \( J_{N_r}^{N_r} \) at low SNR as
\[
\text{var}[J_{N_r}^{N_r}] \approx \rho^2 \text{var} \left[ \frac{1}{N_r} \text{tr}\{\mathbf{A} \mathbf{L}\} \right]. 
\]
(187)
Letting \( \mathbf{A} = \text{diag} \{ \lambda_i \} \) and \( l_{ij} \) denote the \((i,j)\)th element of \( \mathbf{L} \), the first term in (187) can be expanded as
\[
\text{var}[\text{tr}\{\mathbf{A} \mathbf{L}\} | \mathbf{A}] = \sum_{i=1}^{N_r} \lambda_i^2 \left( E[t_{i,i}^2] - E^2[t_{i,i}] \right) 
+ \sum_{i \neq j} \lambda_i \lambda_j \left( E[t_{i,i} t_{i,j}] - E[t_{i,i}] E[t_{i,j}] \right). 
\]
(188)
For a given \( \mathbf{U}_\mathbf{H} \) and random unitary \( \mathbf{V} \) with \( K = N_r \), Theorem 3 in [7] states that \( \mathbf{L} \) has a multivariate beta distribution with parameters \( N_r \) and \( N_l - N_r \). (The distribution of \( \mathbf{L} \) is not known for general \( K \).) From Theorem 2 in [47], we have
\[
E[t_{i,i}] = \frac{N_r}{N_l} + 1 
\]
(189)
\[
E[t_{i,i}^2] = \frac{(N_r + 1)(N_r + 3)}{(N_r + 2)(N_l + 4)} 
\]
(190)
\[
E[t_{i,i} t_{i,j}] = \frac{N_r(N_r + 1)(N_l + 4) + (N_r + 1)(N_l - N_r + 1)}{(N_l + 1)(N_l + 2)(N_l + 4)}, 
\]
(191)
(\( i \neq j \)),
for \( 1 \leq i,j \leq N_r \). Substituting (189)-(191) into (188) gives
\[
\text{var}[\text{tr}\{\mathbf{A} \mathbf{L}\} | \mathbf{A}] \approx \frac{1}{N_r} \sum_{i=1}^{N_r} \lambda_i^2 \left( \tilde{N}_r^2 (1 - \tilde{N}_r) + O \left( \frac{1}{N_r} \right) \right) 
+ \frac{1}{N_r^2} \sum_{i \neq j} \lambda_i \lambda_j \left( (\tilde{N}_r - 1) \tilde{N}_r^3 + O \left( \frac{1}{N_r} \right) \right). 
\]
(192)
Taking expectation with respect to \( \mathbf{A} \), and the large system limit, we have
\[
E_\mathbf{A} \left( \text{var}[\text{tr}\{\mathbf{A} \mathbf{L}\} | \mathbf{A}] \right) \to 1 - \tilde{N}_r. 
\]
(193)
Also, in the large system limit
\[
\frac{1}{N_r} \sum_{i=1}^{N_r} \lambda_i^2 \to \int t^2 dF_\mathbf{A}(t) = \frac{1}{N_r} \left( 1 + \frac{1}{N_r} \right) 
\]
(194)
\[
\frac{1}{N_r^2} \sum_{i \neq j} \lambda_i \lambda_j \to \left[ \int t dF_\mathbf{A}(t) \right]^2 = \frac{1}{N_r^2}. 
\]
(195)
where $F_A(t)$ is the asymptotic distribution for the diagonal elements of $A$ or, equivalently, the asymptotic eigenvalue distribution of $HH^H/N_r$.

Substituting (193) into (187), we have

$$
\sigma^2 = \lim_{(N_r, \bar{N}) \to \infty} \frac{N_r^2 \text{var} [J_u]}{(N_r, \bar{N})} \approx \rho^2 (1 - N_r).
$$

ACKNOWLEDGEMENT

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

REFERENCES


Wiroonsak Santipach (S’00-M’06) received the B.S. (summa cum laude), M.S., and Ph.D. degrees all in electrical engineering from Northwestern University, Illinois, USA in 2000, 2001, and 2006, respectively. He is currently a lecturer at the Department of Electrical Engineering, Faculty of Engineering, Kasetsart University in Bangkok, Thailand. His research interests are in wireless communications, and include performance evaluation of CDMA and MIMO system.

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

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