Multiple-Antenna Capacity in a Deterministic Rician Fading Channel

MAHESH GODAVARTI AND ALFRED O. HERO-III

July 13, 2003

Submitted to *IEEE Transactions on Information Theory*, Sept. 2001

Abstract

We calculate the capacity of a multiple-antenna wireless link with M-transmit and N-receive antennas in a Rician fading channel. We consider the standard Rician fading channel where the channel coefficients are modeled as independent circular Gaussian random variables with non-zero means (non-zero specular component). The channel coefficients of this model are constant over a block of T symbol periods but, independent over different blocks. For such a model, the capacity and capacity achieving signals are dependent on the specular component. We obtain asymptotic expressions for capacity in the low and high SNR (ρ) scenarios. We establish that for low SNR the optimum signal is determined completely by the specular component of the channel and beamforming is the optimum signaling strategy. For high SNR, we show that Rayleigh optimal signaling (the signal that ignores the specular component) achieves the optimal rate of increase of capacity with respect to $\log \rho$. Further, as $\rho \to \infty$, the signals that are optimum for Rayleigh fading channel are also optimum for Rician channel in the case of coherent commumincations and in the case of non-coherent communications when $T \to \infty$. Finally, we show that the the number of degrees of freedom of a Rician fading channel is atleast as much as that of a Rayleigh fading channel namely $K(1 - \frac{K}{T})$ where $K = \min\{M, N, \lfloor T/2 \rfloor\}$.

Keywords: capacity, degrees of freedom, Rician fading, multiple antennas.

Corresponding author:

Mahesh Godavarti
Ditech Communications, Inc.
825 E Middlefield Rd
Mountain View, CA 94085.
Tel: (650) 623-1479.
Fax: (949) 564-9599.
e-mail: mgodavarti@ditechcom.com.

[1]This work was partially supported by the Army Research Office under grant ARO DAAH04-96-1-0337 and was completed while Mahesh Godavarti was a Ph.D. student under the supervision of Prof. Alfred O. Hero-III at the University of Michigan, Ann Arbor.

1 Introduction

The demand for high date rates in wireless channels has led to the investigation of employing multiple antennas at the transmitter and the receiver [6, 7, 13, 16, 18]. Telatar [16], Marzetta and Hochwald [13] and Zheng and Tse [18] have analyzed the maximum achievable rates possible for multiple antenna wireless channels in the presence of Rayleigh fading.

Rayleigh fading models are not sufficient to describe many channels found in the real world. It is important to consider other models and investigate their performance as well. Rician fading is one such model [2, 4, 5, 14, 15]. Rician fading model is applicable when the wireless link between the transmitter and the receiver has a direct path component in addition to the diffused Rayleigh component. Farrokhi et. al. [5] calculate the coherent capacity (when the channel is known at the receiver) for a Rician fading model where they assume that the transmitter has no knowledge of the specular component. Godavarti et. al [8] extend the results to non-coherent capacity (unknown channel at both the transmitter and the receiver) for the same model. In [10], the authors consider another non-traditional model where the specular component is also modeled as random with isotropic distribution and varying over time. They establish results similar to those reported by Marzetta and Hochwald for Rayleigh fading in [13].

In this paper, we analyze the standard Rician fading model for channel capacity under the average energy constraint on the input signal. Throughout the paper, we assume that the specular component is deterministic and is known to both the transmitter and the receiver. The specular component in this paper is of general rank except in Section 2 where it is restricted to be of rank one. The Rayleigh component is never known to the transmitter. There are some cases we consider where the receiver has complete knowledge of the channel. In such cases, the receiver has knowledge about the Rayleigh as well as the specular component whereas the transmitter has knowledge only about the specular component. The capacity when the receiver has complete knowledge about the channel will be referred to as *coherent capacity* and the capacity when the receiver has no knowledge about the Rayleigh component will be referred to as *non-coherent capacity*. This paper is organized as follows. In Section 2 we deal with the special case of a rank-one specular component with the characterization of coherent capacity in Section 2.1. The general case of no restrictions on the rank of the specular component is dealt with in Section 3. For this case, the coherent and non-coherent capacity for low SNR is considered in Section 3.1, coherent capacity for high SNR in Section 3.2, and the non-coherent capacity for high SNR in Section 3.4.

2 Rank-one Specular Component

We adopt the following model for the Rician fading channel

$$X = \sqrt{\frac{\rho}{M}} SH + W \tag{2.1}$$

where X is the $T \times N$ matrix of received signals, H is the $M \times N$ matrix of propagation coefficients, S is the $T \times M$ matrix of transmitted signals, W is the $T \times N$ matrix of additive noise components and ρ is the expected signal to noise ratio at the receivers.

A deterministic rank one Rician channel is defined as

$$H = \sqrt{1-r}G + \sqrt{rNM}H_m \tag{2.2}$$

where G is a matrix of independent $\mathcal{CN}(0,1)$ random variables, H_m is an $M \times N$ deterministic matrix of rank one such that $\text{tr}\{H_m^\dagger H_m\} = 1$ and r is a non-random constant lying between 0 and 1. Without loss of generality we can assume that $H_m = \alpha\beta^\dagger$ where α is a length M vector and β is a length N vector such that

$$H_m = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} [1\ 0 \ldots\ 0] \tag{2.3}$$

where the column and row vectors are of appropriate lengths.

In this case, the conditional probability density function of X given S is given by,

$$p(X|S) = \frac{e^{-\text{tr}\{[I_T + (1-r)(\rho/M)SS^\dagger]^{-1}(X-\sqrt{rNM}SH_m)(X-\sqrt{rNM}SH_m)^\dagger\}}}{\pi^{TN}\det^N[I_T + (1-r)(\rho/M)SS^\dagger]}.$$

The conditional probability density enjoys the following properties

1. For any $T \times T$ unitary matrix ϕ

$$p(\phi X|\phi S) = p(X|S)$$

2. For any $(M-1) \times (M-1)$ unitary matrix ψ

$$p(X|S\Psi) = p(X|S)$$

where

$$\Psi = \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0}^\dagger & \psi \end{bmatrix}. \tag{2.4}$$

2.1 Coherent Capacity

The mutual information (MI) expression for the case where H is known by the receiver has already been derived in [6]. The informed receiver capacity-achieving signal S is zero mean Gaussian independent from time instant to time instant. For such a signal the MI is

$$I(X;S|H) = T \cdot E \log\det\left[I_N + \frac{\rho}{M} H^\dagger \Lambda H\right]$$

where $\Lambda = E[S_t^\tau S_t^*]$ for $t = 1, \ldots, T$, S_t is the t^{th} row of the $T \times M$ matrix S. S_t^τ denotes the transpose of S_t and $S_t^* \stackrel{\text{def}}{=} (S_t^\tau)^\dagger$.

Theorem 1 *Let the channel H be Rician (2.2) and be known to the receiver. Then the capacity is*

$$C_H = \max_d T E \log\det[I_N + \frac{\rho}{M} H^\dagger \Lambda_d H] \tag{2.5}$$

where the signal covariance matrix Λ_d is of the form

$$\Lambda_d = \begin{bmatrix} M-(M-1)d & \underline{0}_{M-1} \\ \underline{0}_{M-1}^\tau & dI_{M-1} \end{bmatrix}$$

where d is a positive real number such that $0 \leq d \leq M/(M-1)$. I_{M-1} is the identity matrix of dimension $M-1$ and $\underline{0}_{M-1}$ is the all zeros column vector of length $M-1$.

Proof: This proof is a modification of the proof in [16]. Using the property that $\Psi^\dagger H$ has the same distribution as H where Ψ is of the form given in (2.4) we conclude that

$$T \cdot E \log\det\left[I_N + \frac{\rho}{M} H^\dagger \Lambda H\right] = T \cdot E \log\det\left[I_N + \frac{\rho}{M} H^\dagger \Psi \Lambda \Psi^\dagger H\right].$$

If Λ is written as

$$\Lambda = \begin{bmatrix} c & A \\ A^\dagger & B \end{bmatrix}$$

where c is a positive number such that $c \geq A^\dagger B^{-1} A$ (to ensure positive semi-definiteness of the covariance matrix Λ), A is a row vector of length $M-1$ and B is a positive definite matrix of size $(M-1) \times (M-1)$. Then

$$\Psi \Lambda \Psi^\dagger = \begin{bmatrix} c & A\psi^\dagger \\ \psi A^\dagger & \psi B \psi^\dagger \end{bmatrix}.$$

Since $B = UDU^\dagger$ where D is a diagonal matrix and U is a unitary matrix of size $(M-1) \times (M-1)$, choosing $\psi = \Pi U$ or $\psi = -\Pi U$ where Π is a $(M-1) \times (M-1)$ permutation matrix, we obtain that

$$\begin{aligned} T \cdot E \log\det\left[I_N + \frac{\rho}{M} H^\dagger \Lambda H\right] &= T \cdot E \log\det\left[I_N + \frac{\rho}{M} H^\dagger \Lambda^{\Pi} H\right] \\ &= T \cdot E \log\det\left[I_N + \frac{\rho}{M} H^\dagger \Lambda^{-\Pi} H\right] \end{aligned}$$

where

$$\Lambda^{\Pi} = \left[\begin{array}{cc} c & AU^{\dagger}\Pi^{\dagger} \\ \Pi U A^{\dagger} & \Pi D \Pi^{\dagger} \end{array} \right].$$

Since log det is a concave (convex cap) function we have

$$\begin{aligned} T \cdot E \log\det\left[I_N + \frac{\rho}{M} H^{\dagger}\ \overline{\Lambda^{\Pi}}\ H\right] &\geq T \cdot \frac{1}{2(M-1)!} \sum_{\Pi} E \log\det\left[I_N + \frac{\rho}{M} H^{\dagger} \Lambda^{\Pi} H\right] \\ &= I(X;S) \end{aligned}$$

where $\overline{\Lambda^{\Pi}} = \frac{1}{2(M-1)!} \sum_{\Pi} \left[\Lambda^{\Pi} + \Lambda^{-\Pi}\right]$ and the summation is over all $(M-1)!$ possible permutation matrices Π. Therefore, the capacity achieving Λ is given by $\overline{\Lambda^{\Pi}}$ and is of the form

$$\Lambda = \left[\begin{array}{cc} c & \underline{0}_{M-1} \\ \underline{0}_{M-1}^{\tau} & dI_{M-1} \end{array} \right]$$

where $d = \text{tr}\{B\}/(M-1)$. Now, the capacity achieving signal matrix has to satisfy $\text{tr}\{\Lambda\} = M$ since MI is monotonically increasing in $\text{tr}\{\Lambda\}$. Therefore, $c = M - (M-1)d$. □

The problem remains to find the d that achieves the maximum in (2.5). This problem has an analytical solution for the special cases of: 1) $r = 0$ for which $d = 1$; and 2) $r = 1$ for which $d = 0$ (rank 1 signal S). In general, the optimization problem (2.5) can be solved by using the method of steepest descent over the space of parameters that satisfy the average power constraint (See Appendix A). Results for $\rho = 100, 10, 1, 0.1$ are shown in Figure 1. As can be seen from the plot the optimum value of d stays close to 1 for high SNR and close to 0 for low SNR. That is, the optimum covariance matrix is close to an identity matrix for high SNR. For low SNR, all the energy is concentrated in the direction of the specular component or in other words the optimal signaling strategy is beamforming. These observations are proven in Section 3.2.

3 General Rank Specular Component

In this case the channel matrix can be written as

$$H = \sqrt{1-r}G + \sqrt{r}H_m \tag{3.1}$$

where G is the Rayleigh Fading component and H_m is a deterministic matrix such that $\text{tr}\{H_m H_m^{\dagger}\} = MN$ with no restriction on its rank. Without loss of generality, we can assume H_m to be an $M \times N$ diagonal matrix with positive real entries.

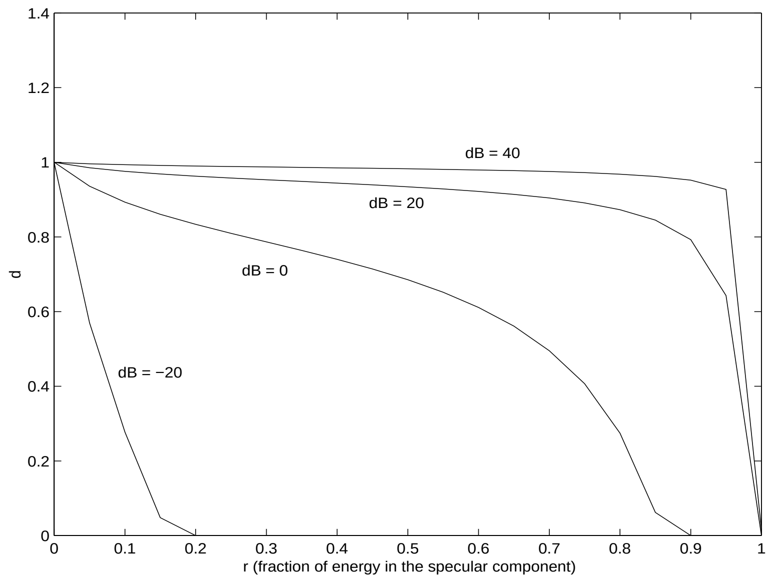

Figure 1: Optimum value of d as a function of r for different values of ρ

3.1 Non-Coherent and Coherent Capacity: Expressions for Low SNR

We next show that for low SNR the Rician fading channel essentially behaves like an AWGN channel in the sense that the Rayleigh fading component has no effect on the structure of the optimum covariance structure. We would like to point out that we proved this result before [17] was published. From [17, Section IV.C, Theorem 1] we can conclude that irrespective of whether the receiver knows H:

$$\lim_{\rho \to 0} \frac{C}{\rho} = T\left[r\lambda_{max}(H_m H_m^\dagger) + N(1-r)\right].$$

The results in this section can be considered simple corollaries of the result in [17]. However, for the coherent case we do have a slightly stronger result.

Proposition 1 *Let H be Rician (3.1) and let the receiver have complete knowledge of the Rayleigh component G. For low SNR, C_H is attained by the same signal covariance matrix that attains capacity when $r = 1$, irrespective of the value of M and N, and*

$$C_H = T\rho[r\lambda_{max}(H_m H_m^\dagger) + (1-r)N] + O(\rho^2).$$

Proof: See Appendix B.1. □

Note that if we choose $\Lambda = I_M$ then varying r has no effect on the value of the capacity. Therefore, this explains the trends seen in [8] and [10] where we have seen how for low SNR the change in capacity is not as pronounced as for high SNR when the channel varies from a purely Rayleigh fading channel to a purely specular one.

Proposition 2 *Let H be Rician (3.1) and the receiver have no knowledge of G. For fixed M, N and T*

$$\lim_{\rho \to 0} \frac{C}{\rho} = T\left[r\lambda_{max}(H_m H_m^\dagger) + N(1-r)\right].$$

Proof: A more general result has been proved in [17, Theorem 1]. □

Proposition 2 suggests that at low SNR all the energy has to be concentrated in the strongest directions of the specular component. For low SNR, we have seen that the channel behaves as an AWGN channel in the sense that it has the same capacity as an equivalent AWGN channel $Y = \sqrt{\frac{\rho}{M}} S H_{awgn} + W$ with $H_{awgn} H_{awgn}^\dagger = E[HH^\dagger]$. In that respect, it would be of interest to know what the capacity would be if we choose the source distribution to be Gaussian.

Proposition 3 *Let the channel H be Rician (3.1) and the receiver have no knowledge of G. For fixed M, N and T if S is a Gaussian distributed source then as $\rho \to 0$*

$$I(X; S_G) = rT\rho\lambda_{max}(H_m H_m^\dagger) + O(\rho^2)$$

where $I(X; S_G)$ is the mutual information between the output and the Gaussian source.

Proof: See Appendix B.2. □

A related version of this result appears in [12, Proposition 5.2.1].

Corollary 1 *For purely Rayleigh fading channels when the receiver has no knowledge of G a Gaussian transmitted signal satisfies $\lim_{\rho \to 0} I(X; S_G)/\rho = 0$.*

3.2 Coherent Capacity: Expression for high SNR

For high SNR, we show that the capacity achieving signal structure basically ignores the specular component.

Theorem 2 *Let H be Rician (3.1). Let C_H be the capacity for H known at the receiver. As $\rho \to \infty$, C_H is attained by an identity signal covariance matrix when $M \leq N$ and*

$$C_H = T \cdot E \log\det[\frac{\rho}{M} HH^\dagger] + O(\frac{\log(\sqrt{\rho})}{\sqrt{\rho}}).$$

Proof: The expression for capacity, C_H is

$$C_H = T \cdot E \log\det[I_N + \frac{\rho}{M} H^\dagger \Lambda H].$$

Let the optimum signal covariance matrix have eigenvalues $\{d_i, i = 1, \ldots, M\}$. We will show that $d_i \to 1$ for $i = 1, \ldots, M$ as $\rho \to \infty$.

Let $D = \text{diag}\{d_1,\ d_2,\ \ldots, d_M\}$ and H have SVD $H = \Phi\Sigma\Psi^\dagger$. Let $C_H^{\Phi,\Psi,D}$ denote the capacity when Φ and Ψ but not Σ are known to the transmitter and the optimum signal covariance matrix is chosen to have eigenvalues $\{d_i\}$ then $C_H \leq C_H^{\Phi,\Psi,D}$. Also,

$$\begin{aligned} C_H^{\Phi,\Psi,D} &= E \log\det[I_N + \frac{\rho}{M} H^\dagger \Lambda^{\Phi,\Psi} H] \\ &= E \log\det[I_N + \frac{\rho}{M} \Sigma^\dagger \Phi^\dagger \Lambda^{\Phi,\Psi} \Phi \Sigma]. \end{aligned}$$

Let $\Phi^\dagger \Lambda^{\Phi,\Psi} \Phi = D^\Phi$. Then

$$\log\det[I_N + \frac{\rho}{M} \Sigma^\dagger D^\Phi \Sigma] = \log\det[I_M + \frac{\rho}{M} D^\Phi \Sigma\Sigma^\dagger].$$

The right hand side expression is maximized by choosing Λ such that D^Φ is diagonal [3, page 255]. In other words, we choose Λ such that $D^\Phi = D$. Let σ_i be the eigenvalues of $\Sigma\Sigma^\dagger$ and define

$$E_A[f(x)] \stackrel{\text{def}}{=} E[f(x)\chi_A(x)] \tag{3.2}$$

where $\chi_A(x)$ is the indicator function for the set A ($\chi_A(x) = 0$ if $x \notin A$ and $\chi_A(x) = 1$ otherwise). Then for large ρ

$$\begin{aligned} C_H^{\Phi,\Psi,D} &= E \log\det[I_M + \frac{\rho}{M} D\Sigma\Sigma^\dagger] \\ &= \sum_{i=1}^M E_{\sigma_i < 1/\sqrt{\rho}} \log[1 + \frac{\rho}{M} d_i \sigma_i] + \sum_{i=1}^M E_{\sigma_i \geq 1/\sqrt{\rho}} \log[1 + \frac{\rho}{M} d_i \sigma_i]. \end{aligned}$$

Let K denote the first term in the right hand side of the expression above and L denote the second term. It is easy to show that

$$\begin{aligned} C_H^{\Phi,\Psi,D} &= E \log\det[I_M + \frac{\rho}{M} D\Sigma\Sigma^\dagger] \\ &= \log\frac{\rho}{M} + \sum_{i=1}^M \log(d_i) + \sum_{i=1}^M E_{\sigma_i > 1/\sqrt{\rho}}[\log(\sigma_i)] + O(\log(\sqrt{\rho})/\sqrt{\rho}) \end{aligned}$$

since

$$K \leq \log[1 + \sqrt{\rho}] \sum_{i=1}^M P(\sigma_i < 1/\sqrt{\rho}) = O(\log(\sqrt{\rho})/\sqrt{\rho})$$

and

$$L = \log \frac{\rho}{M} + \sum_{i=1}^{M} \log(d_i) + \sum_{i=1}^{M} E_{\sigma_i > 1/\sqrt{\rho}}[\log(\sigma_i)] + O(1/\sqrt{\rho}).$$

Let C_H^I denote the capacity achieved when the transmission signal covariance matrix is chosen to be an identity matrix. Then $C_H \geq C_H^I$. Also,

$$\begin{aligned} C_H^I &= E \log\det[I_N + \frac{\rho}{M} H^\dagger H] \\ &= E \log\det[I_N + \frac{\rho}{M} \Sigma^\dagger \Sigma] \\ &= E \log\det[I_M + \frac{\rho}{M} \Sigma\Sigma^\dagger]. \end{aligned}$$

Proceeding as before, we obtain

$$\begin{aligned} C_H^I &= E \log\det[I_M + \frac{\rho}{M} D\Sigma\Sigma^\dagger] \\ &= \log \frac{\rho}{M} + M \log d + \sum_{i=1}^{M} E_{\sigma_i > 1/\sqrt{\rho}}[\log(\sigma_i)] + O(\log(\sqrt{\rho})/\sqrt{\rho}) \end{aligned}$$

where $d = 1 \Rightarrow M \log d = 0$ in the previous expression.

Since $\sum_{i=1}^{M} d_i = M$ and log is a convex cap function, we have $\sum_{i=1}^{M} \log d_i \leq M \log d$ with equality only if $d_i = d$ for all $i = 1, \ldots, M$. If we assume that the limit, as $\rho \rightarrow, \infty$ of the optimum signal covariance matrix is not an identity matrix then we obtain a contradiction since $C_H^{\Phi,\Psi,D} - C_H^I = \sum_{i=i}^{M} \log d_i - M \log d + O(\log(\sqrt{\rho})/\sqrt{\rho})$ can be made negative by choosing ρ to be large enough. □

For $M > N$, optimization using the steepest descent algorithm similar to the one described in Appendix A shows that for high SNR the capacity achieving signal matrix is an identity matrix as well and the capacity is given by

$$C_H \approx T \cdot E \log\det[I_N + \frac{\rho}{M} H^\dagger H].$$

3.3 Non-Coherent Capacity Upper and Lower Bounds

It follows from the data processing theorem that the non-coherent capacity, C can never be greater than the coherent capacity C_H, that is, the uninformed capacity is never decreased when the channel is known to the receiver.

Proposition 4 *Let H be Rician (3.1) and the receiver have no knowledge of the Rayleigh component. Then*

$$C \leq C_H.$$

Now, we establish a lower bound which is similar in flavor to those derived in [8] and [10].

Proposition 5 *Let H be Rician (3.1). A lower bound on capacity when the receiver has no knowledge of G is*

$$\begin{aligned} C &\geq C_H - NE\left[\log_2 \det\left(I_T + (1-r)\frac{\rho}{M}SS^\dagger\right)\right] && (3.3) \\ &\geq C_H - NM\log_2(1 + (1-r)\frac{\rho}{M}T). && (3.4) \end{aligned}$$

Proof: Proof is similar to that of Proposition 2 in [10] and won't be repeated here. □

We notice that the second term in the lower bound goes to zero when $r = 1$: as expected.

As $\rho \to \infty$, the normalized lower bound is dominated by the following term

$$\left(\min\{M, N\} - \frac{NM}{T}\right)\log \rho.$$

Note that choosing $M' < M$ or $N' < N$ transmit antennas we get another lower bound to the capacity. Therefore, we can improve on the lower bound above by maximing the coefficient of $\log \rho$ over the number of transmit and receive antennas. We obtain

$$\max_{M' \leq M, N' \leq N}\left(\min\{M, N\} - \frac{NM}{T}\right) = K(1 - \frac{K}{T})$$

where $K = \min\{M, N, \lfloor T/2 \rfloor\}$. Therefore, we see that the number of degrees of freedom for a non-coherent Rician block fading channel is at least as much as that of a non-coherent Rayleigh block fading channel.

3.4 Non-Coherent Capacity: Expressions for High SNR

In this section we apply the method developed in [18] for the analysis of Rayleigh fading channels. The only difference between the models considered in [18] and here is that we assume H has a deterministic non-zero mean. For convenience, we use a different notation for the channel model:

$$X = SH + W$$

with $H = \sqrt{r}H_m + \sqrt{1-r}G$ where H_m is the specular component of H and G denotes the Rayleigh component. G and W consist of Gaussian circular independent random variables and the covariance matrices of G and W are given by $(1-r)I_{MN}$ and $\sigma^2 I_{TN}$, respectively. H_m is a deterministic matrix satisfying $\text{tr}\{H_m H_m^\dagger\} = MN$. G satisfies $E[\text{tr}\{GG^\dagger\}] = MN$ and r is a number between 0 and 1 so that $E[\text{tr}\{HH^\dagger\}] = MN$.

Lemma 1 *Let the channel be Rician (3.1) and the receiver have no knowledge of G. Then the capacity achieving signal, S can be written as $S = \Phi V \Psi^\dagger$ where Φ is a $T \times M$ unitary matrix independent of V and Ψ. V and Ψ are $M \times M$.*

Proof: Follows from the fact that $p(\Phi X|\Phi S) = p(X|S)$. □

In [18] the requirement for $X = SH + W$ was that X had to satisfy the property that in the singular value decomposition of X, $X = \Phi V \Psi^\dagger$ Φ be independent of V and Ψ. This property holds for the case of Rician fading too because the density functions of X, SH and S are invariant to pre-multiplication by a unitary matrix. Therefore, the leading unitary matrix in the SVD decomposition of any of X, SH and S is independent of the other two components in the SVD and isotropically distributed. This implies that Lemma 6 in [18] holds and we have

Lemma 2 *Let $R = \Phi_R \Sigma_R \Psi_R^\dagger$ be such that Φ_R is independent of Σ_R and Ψ_R. Then*

$$\mathcal{H}(R) = \mathcal{H}(Q\Sigma_R \Psi_R^\dagger) + \log|G(T,M)| + (T-M)E[\log\det \Sigma_R^2],$$

where Q is an $M \times M$ unitary matrix independent of V and Ψ and $|G(T,M)|$ is the volume of the Grassmann manifold and is equal to

$$\frac{\prod_{i=T-M+1}^{T} \frac{2\pi^i}{(i-1)!}}{\prod_{i=1}^{M} \frac{2\pi^i}{(i-1)!}}.$$

The Grassmann manifold $G(T,M)$ [18] is the set of equivalence classes of all $T \times M$ unitary matrices such that if P, Q belong to an equivalence class then $P = QU$ for some $M \times M$ unitary matrix U.

3.4.1 $M = N$, $T \geq 2M$

To calculate $I(X;S)$ we need to compute $\mathcal{H}(X)$ and $\mathcal{H}(X|S)$. To compute $\mathcal{H}(X|S)$ we note that given S, X is a Gaussian random vector with columns of X independent of each other. Each row has the common covariance matrix given by $(1-r)SS^\dagger + \sigma^2 I_T = \Phi V^2 \Phi^\dagger + \sigma^2 I_T$. Therefore

$$\mathcal{H}(X|S) = ME[\sum_{i=1}^{M} \log(\pi e((1-r)\|s_i\|^2 + \sigma^2)] + M(T-M)\log(\pi e \sigma^2).$$

To compute $\mathcal{H}(X)$, we write the SVD: $X = \Phi_X \Sigma_X \Psi_X^\dagger$. Note that Φ_X is isotropically distributed and independent of $\Sigma_X \Psi_X^\dagger$, therefore from Lemma 2 we have

$$\mathcal{H}(X) = \mathcal{H}(Q\Sigma_X \Psi_X^\dagger) + \log|G(T,M)| + (T-M)E[\log\det \Sigma_X^2].$$

We first characterize the optimal input distribution in the following lemma.

Lemma 3 *Let H be Rician (3.1) and the receiver have no knowledge of G. Let $(s_i^\sigma, i = 1, \ldots, M)$ be the optimal input signal of each antenna at when the noise power at the receive antennas is given by σ^2. If $T \geq 2M$,*

$$\frac{\sigma}{\|s_i^\sigma\|} \xrightarrow{P} 0, \text{ for } i = 1, \ldots, M \tag{3.5}$$

where $\xrightarrow{P}$ denotes convergence in probability.

Proof: See Appendix C. □

Lemma 4 *Let H be Rician (3.1) and the receiver have no knowledge of G. The maximal rate of increase of capacity, $\max_{p(S):E[\mathrm{tr}\{SS^\dagger\}]\leq TM} I(X;S)$ with SNR is $M(T-M)\log\rho$ and the constant norm source $\|s_i\|^2 = T$ for $i = 1, \ldots, M$ attains this rate.*

Proof: See Appendix C. □

Lemma 5 *Let H be Rician (3.1) and the receiver have no knowledge of G. As $T \to \infty$ the optimal source in Lemma 4 is the constant norm input*

Proof: See Appendix C. □

From now on, we assume that the optimal input signal is the constant norm input. For the constant norm input $\Phi V \Psi^\dagger = \Phi V$ since Φ is isotropically distributed.

Theorem 3 *Let the channel be Rician (3.1) and the receiver have no knowledge of G. For the constant norm input, as $\sigma^2 \to 0$ the capacity is given by*

$$\begin{aligned} C &= \log|G(T,M)| + (T-M)E[\log\det H^\dagger H] - M(T-M)\log\pi e\sigma^2 - \\ &\quad M^2\log\pi e + \mathcal{H}(QVH) + (T-2M)M\log T - M^2\log(1-r) \end{aligned}$$

where Q, V and $|G(T,M)|$ are as defined in Lemma 2.

Proof: Since $\|s_i^2\| \gg \sigma^2$ for all $i = 1, \ldots, M$

$$\mathcal{H}(X|S) = ME[\sum_{i=1}^{M} \log\pi e((1-r)\|s_i\|^2 + \sigma^2)] + M(T-M)\log(\pi e\sigma^2)$$

$$
\begin{aligned}
&\approx \quad ME[\sum_{i=1}^{M} \log \pi e(1-r)\|s_i\|^2] + M(T-M)\log \pi e\sigma^2 \\
&= \quad ME[\log\det(1-r)V^2] + M^2 \log \pi e + M(T-M)\log \pi e\sigma^2
\end{aligned}
$$

and from Appendix D

$$
\begin{aligned}
\mathcal{H}(X) &\approx \mathcal{H}(SH) \\
&= \mathcal{H}(QVH) + \log|G(T,M)| + (T-M)E[\log\det(H^\dagger V^2 H)] \\
&= \mathcal{H}(QVH) + \log|G(T,M)| + (T-M)E[\log\det V^2] + \\
&\quad (T-M)E[\log\det HH^\dagger].
\end{aligned}
$$

Combining the two equations

$$
\begin{aligned}
I(X;S) &\approx \log|G(T,M)| + (T-M)E[\log\det H^\dagger H] - M(T-M)\log \pi e\sigma^2 + \\
&\quad \mathcal{H}(QVH) - M^2\log \pi e + (T-2M)E[\log\det V^2] - M^2\log(1-r).
\end{aligned}
$$

Now, since the optimal input signal is $\|s_i\|^2 = T$ for $i = 1, \ldots, M$, we have

$$
\begin{aligned}
C &= I(X;S) \\
&\approx \log|G(T,M)| + (T-M)E[\log\det H^\dagger H] - M(T-M)\log \pi e\sigma^2 - \\
&\quad M^2\log \pi e + \mathcal{H}(QVH) + (T-2M)M\log T - M^2\log(1-r).
\end{aligned}
$$

□

Theorem 4 *Let H be Rician (3.1) and the receiver have no knowledge of G. As $T \to \infty$ the normalized capacity $C/T \to E[\log\det \frac{\rho}{M} H^\dagger H]$ where $\rho = M/\sigma^2$.*

Proof: First, a lower bound to capacity as $\sigma^2 \to 0$ is given by

$$
\begin{aligned}
C &\geq \log|G(T,M)| + (T-M)E[\log\det H^\dagger H] + M(T-M)\log\frac{T\rho}{M\pi e} - \\
&\quad M^2\log T - M^2\log(1-r) - M^2\log \pi e.
\end{aligned}
$$

In [18] it's already been shown that $\lim_{T\to\infty}(\frac{1}{T}\log|G(T,M)| + M(1-\frac{M}{T})\log\frac{T}{\pi e}) = 0$. Therefore we have as $T \to \infty$

$$
C/T \geq ME[\log\det \frac{\rho}{M} H^\dagger H].
$$

Second, since $\mathcal{H}(QVH) \leq M^2 \log(\pi e T)$ an asymptotic upper bound on capacity is given by

$$\begin{aligned} C &\leq \log|G(T,M)| + (T-M)E[\log\det H^\dagger H] + M(T-M)\log\frac{T\rho}{M\pi e} - \\ &\quad M^2\log(1-r). \end{aligned}$$

Therefore, we have as $T \to \infty$

$$C/T \leq E[\log\det\frac{\rho}{M}H^\dagger H].$$

□

3.4.2 $M < N$, $T \geq M + N$

In this case we show that the optimal rate of increase is given by $M(T-M)\log\rho$. The higher number of receive antennas can provide only a finite increase in capacity for all SNRs.

Theorem 5 *Let the channel be Rician (3.1) and the receiver have no knowledge of G. Then the maximum rate of increase of capacity with respect to $\log\rho$ is given by $M(T-M)$.*

Proof: See Appendix C. □

For this case the lower bound on capacity tells us that the minimal rate of increase of capacity with respect to $\log\rho$ is also $M(T-M)$. Therefore, we conclude that the rate of increase is indeed $M(T-M)\log\rho$.

4 Conclusions and Future Work

In this paper, we have analyzed the standard Rician fading channel for capacity. Most of the analysis was for a general specular component but, for the special case of a rank-one specular component we were able to show more structure on the signal input. For the case of general specular component, we were able to derive asymptotic closed form expressions for capacity for low and high SNR scenarios.

One important result of the analysis is that for low SNRs beamforming is very desirable whereas for high SNR scenarios it is not. This result is very useful in designing space-time codes. For high SNR scenarios, the standard codes designed for Rayleigh fading work for the case of Rician fading as well. We conclude as in [18] that for the case $M \leq N$ and $T \geq M + N$ the number of degrees of freedom is given by $M\frac{T-M}{T}$. For the general case of M, N and T, we conclude from the lower bound on capacity that the number of degrees of freedom is atleast as much as $K\frac{T-K}{T}$ where $K = \min\{M, N, \lfloor T/2 \rfloor\}$.

A lot more work needs to be done such as for case of $M > N$ along the lines of [18]. It also seems reasonable that the work in [1] can be extended to the case of Rician fading.

APPENDICES

A Capacity Optimization in Section 2.1

We have the following expression for the capacity

$$C = E \log\det(I_N + \frac{\rho}{M} H^\dagger \Lambda H)$$

where Λ is of the form

$$\Lambda = \begin{bmatrix} M-(M-1)d & \underline{0}^\tau_{M-1} \\ \underline{0}_{M-1} & dI_{M-1} \end{bmatrix}.$$

We can find the optimal value of d iteratively by using the method of steepest descent as follows

$$d_{k+1} = d_k + \mu \frac{\partial C}{\partial d_k}$$

where d_k is the value of d at the k^{th} iteration. We use the following identity (Jacobi's formula) to calculate the partial derivatives.

$$\frac{\partial \log\det A}{\partial d} = \mathrm{tr}\{A^{-1}\frac{\partial A}{\partial d}\}.$$

Therefore, we obtain

$$\frac{\partial C}{\partial d} = E\mathrm{tr}\{[I_N + \frac{\rho}{M} H^\dagger \Lambda H]^{-1} \frac{\rho}{M} H^\dagger \frac{\partial \Lambda}{\partial d} H\}$$

where

$$\frac{\partial \Lambda}{\partial d} = \begin{bmatrix} -(M-1) & \underline{0}^\tau_{M-1} \\ \underline{0}_{M-1} & I_{M-1} \end{bmatrix}$$

The derivative can be evaluated using monte carlo simulation.

B Non-coherent Capacity for low SNR values

B.1 Proof of Proposition 1

Proof: Let $\|H\|$ denote the matrix 2-norm of H, γ be a positive number such that $\gamma \in (0,1)$ then

$$\begin{aligned} C_H &= T \cdot E \log\det[I_N + \frac{\rho}{M} H^\dagger \Lambda H] \\ &= T \cdot E_{\|H\| \geq 1/\rho^\gamma} \log\det[I_N + \frac{\rho}{M} H^\dagger \Lambda H] + E_{\|H\| < 1/\rho^\gamma} \log\det[I_N + \frac{\rho}{M} H^\dagger \Lambda H] \\ &= TE\text{tr}\{\frac{\rho}{M} H^\dagger \Lambda H\} + O(\rho^{2-2\gamma}) \end{aligned}$$

where $E_{\|H\| \geq 1/\rho^\gamma}[\cdot]$ is as defined in (3.2). This follows from the fact that $P(\|H\| \geq 1/\rho^\gamma) \leq O(e^{-\frac{1}{TM\rho^\gamma}})$ and for $\|H\| < 1/\rho^\gamma$ $\log\det[I_N + \frac{\rho}{M} H^\dagger \Lambda H] = \text{tr}[\frac{\rho}{M} H^\dagger \Lambda H] + O(\rho^{2-2\gamma})$. Since γ is arbitrary

$$E \log\det[I_N + \frac{\rho}{M} H^\dagger \Lambda H] = E\text{tr}[\frac{\rho}{M} H^\dagger \Lambda H] + O(\rho^2).$$

Now

$$\begin{aligned} E\text{tr}[H^\dagger \Lambda H] &= \text{tr}\{(1-r)E[G^\dagger \Lambda G] + rH_m^\dagger \Lambda H_m]\} \\ &= \text{tr}\{(1-r)\Lambda E[GG^\dagger] + r\Lambda H_m H_m^\dagger\}. \end{aligned}$$

Therefore, we have to choose Λ to maximize $\text{tr}\{(1-r)N\Lambda + r\Lambda H_m H_m^\dagger\}$. Since H_m is diagonal the trace depends only on the diagonal elements of Λ. Therefore, Λ can be chosen to be diagonal. Also, because of the power constraint, $\text{tr}\{\Lambda\} \leq M$, to maximize the expression we choose $\text{tr}\{\Lambda\} = M$. The maximizing Λ has as many non-zero elements as the multiplicity of the maximum eigenvalue of $(1-r)NI_M + rH_m H_m^\dagger$. The non-zero elements of Λ multiply the maximum eigenvalues of $(1-r)NI_M + rH_m H_m^\dagger$ and can be chosen to be of equal magnitude summing up to M. This is the same Λ maximizing the capacity for additive white Gaussian noise channel with channel H_m. And that completes the proof. □

For the rest of this section, we introduce some new notation for ease of description. If X is a $T \times N$ matrix then let $\tilde{X}$ denote the "unwrapped" $NT \times 1$ vector formed by placing the transposed rows of X in a single column in an increasing manner. That is, if $X_{i,j}$ denotes the element of X in the i^{th} row and j^{th} column then $\tilde{X}_{i,1} = X_{\lfloor i/N \rfloor, i\%N}$, where $\lfloor i/N \rfloor$ denotes the greater integer less than or equal to i/N and $i\%N$ denotes the operation i modulo N. The channel model $X = \sqrt{\frac{\rho}{M}} SH + W$ can now be written as $\tilde{X} = \sqrt{\frac{\rho}{M}} \hat{H}\tilde{S} + \tilde{W}$. $\hat{H}$ is given by $\hat{H} = I_T \otimes H^\tau$ where H^τ denotes the transpose of H. The notation $A \otimes B$ denotes the Kronecker product of the matrices A and B and is defined as follows. If A is a $I \times J$ matrix and

B a $K \times L$ matrix then $A \otimes B$ is a $IK \times JL$ matrix

$$A \otimes B = \begin{bmatrix} (A)_{11}B & (A)_{12}B & \dots & (A)_{1J}B \\ (A)_{21}B & (A)_{22}B & \dots & (A)_{2J}B \\ \vdots & \vdots & \ddots & \vdots \\ (A)_{I1}B & (A)_{I2}B & \dots & (A)_{IJ}B \end{bmatrix}.$$

This way, we can describe the conditional probability density function $p(X|S)$ as follows

$$p(X|S) = \frac{1}{\pi^{TN}|\Lambda_{\tilde{X}|\tilde{S}}|} e^{-(\tilde{X} - \sqrt{r\frac{\rho}{M}}\hat{H}_m\tilde{S})^\dagger \Lambda^{-1}_{\tilde{X}|\tilde{S}}(\tilde{X} - \sqrt{r\frac{\rho}{M}}\hat{H}_m\tilde{S})}$$

where $|\Lambda_{\tilde{X}|\tilde{S}}| = \det(I_{TN} + (1-r)SS^\dagger \otimes I_N)$.

B.2 Proof of Proposition 3

First, $I(X;S) = \mathcal{H}(X) - \mathcal{H}(X|S)$. Since S is Gaussian distributed, $E[\log\det(I_N + \frac{\rho}{M}\hat{H}\Lambda_{\tilde{S}}\hat{H}^\dagger)] \leq \mathcal{H}(X) \leq \log\det(I_N + \frac{\rho}{M}\Lambda_{\tilde{X}})$ where the expectation is taken over the distribution of H and $\hat{H}\Lambda_{\tilde{S}}\hat{H}^\dagger = \Lambda_{\tilde{X}|H}$ is the covariance of $\tilde{X}$ for a particular H. Next, we show that $\mathcal{H}(X) = \frac{\rho}{M}\text{tr}\{\Lambda_{\tilde{X}}\} + O(\rho^2)$. First, the upper bound to $\mathcal{H}(X)$ can be written as $\frac{\rho}{M}\text{tr}\{\Lambda_{\tilde{X}}\} + O(\rho^2)$ because H is Gaussian distributed and the probability that $\|H\| > R$ is of the order e^{-R^2}. Second, using notation (3.2) $E[\log\det(I_{TN} + \frac{\rho}{M}\hat{H}\Lambda_{\tilde{S}}\hat{H}^\dagger)] = E_{\|H\|<(\frac{M}{\rho})^\gamma}[\cdot] + E_{\|H\|\geq(\frac{M}{\rho})^\gamma}[\cdot]$ where γ is a number such that $2-\gamma > 1$ or $\gamma < 1$. Then

$$\begin{aligned} E[\log\det(I_{TN} + \frac{\rho}{M}\hat{H}\Lambda_{\tilde{S}}\hat{H}^\dagger)] &= \frac{\rho}{M}E_{\|H\|<(\frac{M}{\rho})^\gamma}[\text{tr}\{\hat{H}\Lambda_{\tilde{S}}\hat{H}^\dagger\}] + \\ &\quad O(\rho^{2-\gamma}) + O(\log((\frac{M}{\rho})^\gamma)e^{-(\frac{M}{\rho})^\gamma}) \\ &= \frac{\rho}{M}E[\text{tr}\{\hat{H}\Lambda\hat{H}^\dagger\}] + O(\rho^{2-\gamma}). \end{aligned}$$

Since γ is arbitrary, we have $\mathcal{H}(X) = \frac{\rho}{M}E[\text{tr}\{\hat{H}\Lambda_{\tilde{S}}\hat{H}^\dagger\}] + O(\rho^2)$. Note that $\Lambda_{\tilde{X}} = E[\Lambda_{\tilde{X}|H}]$ and since $\mathcal{H}(X)$ is sandwiched between two expressions of the form $\frac{\rho}{M}\text{tr}\{\Lambda_{\tilde{X}}\} + O(\rho^2)$ the assertion follows.

Now $\mathcal{H}(X|S) = E[\log\det(I_{TN} + (1-r)\frac{\rho}{M}SS^\dagger \otimes I_N)]$. It can be shown similarly that $\mathcal{H}(X|S) = (1-r)\frac{\rho}{M}\text{tr}\{E[SS^\dagger \otimes I_N]\} + O(\rho^2)$.

Recall that $H = \sqrt{r}H_m + \sqrt{1-r}G$. Therefore, $\Lambda_{\tilde{X}} = E[\hat{H}\Lambda_{\tilde{S}}\hat{H}^\dagger] = r\hat{H}_m\Lambda_{\tilde{S}}\hat{H}_m^\dagger + (1-r)E[SS^\dagger]\otimes I_N$ and we have, for a Gaussian distributed input, $I(X;S_G) = r\frac{\rho}{M}\text{tr}\{\hat{H}_m\Lambda_{\tilde{S}}\hat{H}_m^\dagger\} + O(\rho^2)$. Since H_m is a diagonal matrix only the diagonal elements of $\Lambda_{\tilde{S}}$ matter and we we can choose the signals to be independent from time instant to time instant. Also, to maximize $\text{tr}\{\hat{H}_m\Lambda_{\tilde{S}}\hat{H}_m^\dagger\}$ under the condition $\text{tr}\{\Lambda\} \leq TM$ it is best to concentrate all the available energy on the largest eigenvalues of H_m. Therefore, we obtain

$$I(X;S_G) = r\frac{\rho}{M}TM\lambda_{max}(H_mH_m^\dagger) + O(\rho^2).$$

And that completes the proof.

C Proof of Lemma 3 in Section 3.4.1

In this section we will show that as $\sigma^2 \to 0$ or as $\rho \to \infty$ for the optimal input $(s_i^{(\sigma)}, i = 1, \ldots, M)$, $\forall \delta, \epsilon > 0$, $\exists \sigma_0$ such that for all $\sigma < \sigma_0$

$$P(\frac{\sigma}{\|s_i^\sigma\|} > \delta) < \epsilon \tag{C.1}$$

for $i = 1, \ldots, M$. $s_i^{(\sigma)}$ denotes the optimum input signal being transmitted over antenna i, $i = 1\ldots, M$ when the noise power at the receiver is σ^2. Also, throughout we use ρ to denote the average signal to noise ratio M/σ^2 present at each of the receive antennas.

The proof in this section has basically been reproduced from [18] except for some minor changes to account for the deterministic specular component (H_m) present in the channel. The proof is by contradiction. We need to show that if the distribution P of a source $s_i^{(\sigma)}$ satisfies $P(\frac{\sigma}{\|s_i\|} > \delta) > \epsilon$ for some ϵ and δ and for arbitrarily small σ^2, there exists σ^2 such that $s_i^{(\sigma)}$ is not optimal. That is, we can construct another input distribution that satisfies the same power constraint, but achieves higher mutual information. The steps in the proof are as follows

1. We show that in a system with M transmit and N receive antennas, coherence time $T \geq 2N$, if $M \leq N$, there exists a finite constant $k_1 < \infty$ such that for any fixed input distribution of S, $I(X;S) \leq k_1 + M(T-M)\log\rho$. That is, the mutual information increases with SNR at a rate no higher than $M(T-M)\log\rho$.

2. For a system with M transmit and receive antennas, if we choose signals with significant power only in M' of the transmit antennas, that is $\|s_i\| \leq C\sigma$ for $i = M'+1, \ldots, M$ and some constant C, we show that the mutual information increases with SNR at rate no higher than $M'(T-M')\log\rho$.

3. We show that for a system with M transmit and receive antennas if the input distribution doesn't satisfy (C.1), that is, has a positive probability that $\|s_i\| \leq C\sigma$, the mutual information achieved increases with SNR at rate strictly lower than $M(T-M)\log\rho$.

4. We show that in a system with M transmit and receive antennas for constant equal norm input $P(\|s_i\| = \sqrt{T}) = 1$, for $i = 1, \ldots, M$, the mutual information increases with SNR at rate $M(T - M)\log\rho$. Since $M(T-M) \geq M'(T-M')$ for any $M' \leq M$ and $T \geq 2M$, any input distribution that doesn't satisfy $(C.1)$ yields a mutual information that increases at lower rate than constant equal norm input, and thus is not optimal at high enough SNR level.

Step 1 For a channel with M transmit and N receive antennas, if $M < N$ and $T \geq 2N$, we write the conditional differential entropy as

$$\mathcal{H}(X|S) = N\sum_{i=1}^{M} E[\log((1-r)\|s_i\|^2 + \sigma^2)] + N(T-M)\log\pi e\sigma^2.$$

Let $X = \Phi_X\Sigma_X\Psi_X^\dagger$ be the SVD for X then

$$\begin{aligned}
\mathcal{H}(X) &\leq \mathcal{H}(\Phi_X) + \mathcal{H}(\Sigma_X|\Psi) + \mathcal{H}(\Psi) + E[\log J_{T,N}(\sigma_1,\ldots,\sigma_N)] \\
&\leq \mathcal{H}(\Phi_X) + \mathcal{H}(\Sigma_X) + \mathcal{H}(\Psi) + E[\log J_{T,N}(\sigma_1,\ldots,\sigma_N)] \\
&= \log|R(N,N)| + \log|R(T,N)| + \mathcal{H}(\Sigma_X) + E[\log J_{T,N}(\sigma_1,\ldots,\sigma_N)]
\end{aligned}$$

where $R(T,N)$ is the Steifel manifold for $T \geq N$ [18] and is defined as the set of all unitary $T \times N$ matrices. $|R(T,N)|$ is given by

$$|R(T,N)| = \prod_{i=T-N+1}^{T} \frac{2\pi^i}{(i-1)!}.$$

$J_{T,N}(\sigma_1,\ldots,\sigma_N)$ is the Jacobian of the transformation $X \to \Phi_X\Sigma_X\Psi_X^\dagger$ [18] and is given by

$$J_{T,N} = (\frac{1}{2\pi})^N \prod_{i<j\leq N}(\sigma_i^2 - \sigma_j^2)^2 \prod_{i=1}^{N}\sigma_i^{2(T-M)+1}.$$

We have also chosen to arrange σ_i in decreasing order so that $\sigma_i > \sigma_j$ if $i < j$. Now

$$\begin{aligned}
\mathcal{H}(\Sigma_X) &= \mathcal{H}(\sigma_1,\ldots,\sigma_M,\sigma_{M+1},\ldots,\sigma_N) \\
&\leq \mathcal{H}(\sigma_1,\ldots,\sigma_M) + \mathcal{H}(\sigma_{M+1},\ldots,\sigma_N)
\end{aligned}$$

Also,

$$\begin{aligned}
E[\log J_{T,N}(\sigma_1,\ldots,\sigma_N)] &= \log\frac{1}{(2\pi)^N} + \sum_{i=1}^{N} E[\log\sigma_i^{2(T-N)+1}] + \sum_{i<j\leq N} E[\log(\sigma_i^2 - \sigma_j^2)^2] \\
&= \log\frac{1}{(2\pi)^M} + \sum_{i=1}^{M} E[\log\sigma_i^{2(T-N)+1}] + \\
&\quad \sum_{i<j\leq M} E[\log(\sigma_i^2-\sigma_j^2)^2] + \sum_{i\leq M, M<j\leq N} E[\underbrace{\log(\sigma_i^2-\sigma_j^2)^2}_{\leq\log\sigma_i^4}] +
\end{aligned}$$

$$
\begin{aligned}
& \log \frac{1}{(2\pi)^{N-M}} + \\
& \sum_{i=M+1}^{N} E[\log \sigma_i^{2(T-N)+1}] + \sum_{M<i<j\le N} E[\log(\sigma_i^2 - \sigma_j^2)^2] \\
\le\ & E[\log J_{N,M}(\sigma_1, \ldots, \sigma_M)] \\
& + E[\log J_{T-M,N-M}(\sigma_{M+1}, \ldots, \sigma_N)] \\
& + 2(T-M) \sum_{i=1}^{M} E[\log \sigma_i^2].
\end{aligned}
$$

Next define $C_1 = \Phi_1 \Sigma_1 \Psi_1^\dagger$ where $\Sigma_1 = \text{diag}(\sigma_1, \ldots, \sigma_M)$, Φ_1 is a $N \times M$ unitary matrix, Ψ_1 is a $M \times M$ unitary matrix. Choose Σ_1, Φ_1 and Ψ_1 to be independent of each other. Similarly define C_2 from the rest of the eigenvalues. Now

$$
\begin{aligned}
\mathcal{H}(C_1) &= \log|R(M,M)| + \log|R(N,M)| + \mathcal{H}(\sigma_1, \ldots, \sigma_M) + E[\log J_{N,M}(\sigma_1, \ldots, \sigma_M)] \\
\mathcal{H}(C_2) &= \log|R(N-M,N-M)| + \log|R(T-M,N-M)| \\
&\quad + \mathcal{H}(\sigma_{M+1}, \ldots, \sigma_N) + E[\log J_{T-M,N-M}(\sigma_{M+1}, \ldots, \sigma_N)].
\end{aligned}
$$

Substituting in the formula for $\mathcal{H}(X)$, we obtain

$$
\begin{aligned}
\mathcal{H}(X) &\le \mathcal{H}(C_1) + \mathcal{H}(C_2) + (T-M)\sum_{i=1}^{M} E[\log \sigma_i^2] + \log|R(T,N)| + \log|R(N,N)| \\
&\quad - \log|R(N,M)| - \log|R(M,M)| - \log|R(N-M,N-M)| - \\
&\quad \log|R(T-M,N-M)| \\
&= \mathcal{H}(C_1) + \mathcal{H}(C_2) + (T-M)\sum_{i=1}^{M} E[\log \sigma_i^2] + \log|G(T,M)|.
\end{aligned}
$$

Note that C_1 has bounded total power

$$
\text{tr}\{E[C_1 C_1^\dagger]\} = \text{tr}\{E[\sigma_i^2]\} = \text{tr}\{E[XX^\dagger]\} \le NT(M+\sigma^2).
$$

Therefore, the differential entropy of C_1 is bounded by the entropy of a random matrix with entries iid Gaussian distributed with variance $\frac{T(M+\sigma^2)}{M}$ [3, p. 234, Theorem 9.6.5]. That is

$$
\mathcal{H}(C_1) \le NM \log\left[\pi e \frac{T(M+\sigma^2)}{M}\right].
$$

Similarly, we bound the total power of C_2. Since $\sigma_{M+1}, \ldots, \sigma_N$ are the $N-M$ least singular values of X, for any $(N-M) \times N$ unitary matrix Q.

$$
\text{tr}\{E[C_2 C_2^\dagger]\} \le (N-M)T\sigma^2.
$$

Therefore, the differential entropy is maximized if C_2 has independent iid Gaussian entries and

$$\mathcal{H}(C_2) \leq (N-M)(T-M)\log\left[\pi e\frac{T\sigma^2}{T-M}\right].$$

Therefore, we obtain

$$\begin{aligned}\mathcal{H}(X) \quad \leq \quad & \log|G(T,M)| + NM\log\left[\pi e\frac{T(M+\sigma^2)}{M}\right] + (T-M)\sum_{i=1}^{M} E[\log\sigma_i^2] \\ & +(N-M)(T-M)\log\pi e\sigma^2 + (N-M)(T-M)\log\frac{T}{T-M}.\end{aligned}$$

Combining with $\mathcal{H}(X|S)$, we obtain

$$\begin{aligned}I(X;S) \quad \leq \quad & \underbrace{\log|G(T,M)| + NM\log\frac{T(M+\sigma^2)}{M} + (N-M)(T-M)\log\frac{T}{T-M}}_{\alpha} \\ & \underbrace{+\,(T-M-N)\sum_{i=1}^{M} E[\log\sigma_i^2]}_{\beta} + \\ & \underbrace{N\left(\sum_{i=1}^{M} E[\log\sigma_i^2] - \sum_{i=1}^{M} E[\log((1-r)\|s_i\|^2+\sigma^2)]\right)}_{\gamma} \\ & -M(T-M)\log\pi e\sigma^2.\end{aligned}$$

By Jensen's inequality

$$\begin{aligned}\sum_{i=1}^{M} E[\log\sigma_i^2] \quad & \leq \quad M\log\left(\frac{1}{M}\sum_{i=1}^{M} E[\sigma_i^2]\right) \\ & = \quad M\log\frac{NT(M+\sigma^2)}{M}.\end{aligned}$$

For γ it will be shown that

$$\sum_{i=1}^{M} E[\log\sigma_i^2] - \sum_{i=1}^{M} E[\log((1-r)\|s_i\|^2+\sigma^2)] \leq k$$

where k is some finite constant.

Given S, X has mean $\sqrt{r}SH_m$ and covariance matrix $I_N \otimes ((1-r)SS^\dagger + \sigma^2 I_T)$. If $S = \Phi V \Psi^\dagger$ then

$$\begin{aligned}X^\dagger X \quad & = \quad H^\dagger S^\dagger SH + W^\dagger SH + H^\dagger S^\dagger W + W^\dagger W \\ & \stackrel{d}{=} \quad H_1^\dagger V^\dagger V H_1 + W^\dagger V^\dagger H_1 + H_1^\dagger V W + W^\dagger W\end{aligned}$$

where H_1 has the covariance matrix as H but mean is given by $\sqrt{r}\Psi^\dagger H_m$. Therefore, $X^\dagger X = X_1^\dagger X_1$ where $X_1 = VH_1 + W$

Now, X_1 has the same distribution as $((1-r)VV^\dagger + \sigma^2 I_T)^{1/2} Z$ where Z is a random Gaussian matrix with mean $\sqrt{r}((1-r)VV^\dagger + \sigma^2 I_T)^{-1/2}\Psi^\dagger H_m$ and covariance I_{NT}. Therefore,

$$X^\dagger X \stackrel{d}{=} Z^\dagger((1-r)VV^\dagger + \sigma^2 I_T)Z.$$

Let $(X^\dagger X|S)$ denote the realization of $X^\dagger X$ given S then

$$(X^\dagger X|S) \stackrel{d}{=} Z^\dagger \begin{bmatrix} (1-r)\|s_1\|^2+\sigma^2 & & & & & \\ & \ddots & & & & \\ & & (1-r)\|s_M\|^2+\sigma^2 & & & \\ & & & \sigma^2 & & \\ & & & & \ddots & \\ & & & & & \sigma^2 \end{bmatrix} Z.$$

Let $Z = [Z_1|Z_2]$ be the partition of Z such that

$$(X^\dagger X|S) \stackrel{d}{=} Z_1^\dagger((1-r)V^2 + \sigma^2 I_M)Z_1 + \sigma^2 Z_2^\dagger Z_2$$

where Z_1 has mean $\sqrt{r}((1-r)V^2+\sigma^2 I_M)^{-1/2}V\Psi^\dagger H_m$ and covariance I_{NM} and Z_2 has mean 0 and covariance $I_{N(T-M)}$

We use the following Lemma from [11]

Lemma 6 *If C and B are both Hermitian matrices, and if their eigenvalues are both arranged in decreasing order, then*

$$\sum_{i=1}^{N}(\lambda_i(C) - \lambda_i(B))^2 \leq \|C-B\|_2^2$$

where $\|A\|_2^2 \stackrel{\text{def}}{=} \sum A_{ij}^2$, $\lambda_i(A)$ denotes the i^{th} eigenvalue of Hermitian matrix A.

Applying this Lemma with $C = (X^\dagger X|S)$ and $B = Z_1^\dagger(V^2 + \sigma^2 I_M)Z_1$ we obtain

$$\lambda_i(C) \leq \lambda_i(B) + \sigma^2\|Z_2^\dagger Z_2\|_2$$

for $i = 1, \ldots, M$ Note that $\lambda_i(B) = \lambda_i(B')$ where $B' = ((1-r)V^2 + \sigma^2 I_M)Z_1 Z_1^\dagger$. Let $k = E[\|Z_2^\dagger Z_2\|_2]$ be a finite constant. Now, since Z_1 and Z_2 are independent matrices (covariance of $[Z_1|Z_2]$ is a diagonal matrix)

$$\begin{aligned}
\sum_{i=1}^{M} E[\log\sigma_i^2|S] &\leq \sum_{i=1}^{M} E[\log(\lambda_i(((1-r)V^2 + \sigma^2 I_M)Z_1 Z_1^\dagger) + \sigma^2\|Z_2^\dagger Z_2\|_2)] \\
&= \sum_{i=1}^{M} E[E[\log(\lambda_i(((1-r)V^2 + \sigma^2 I_M)Z_1 Z_1^\dagger) + \sigma^2\|Z_2^\dagger Z_2\|_2) \mid Z_1]] \\
&\leq \sum_{i=1}^{M} E[\log(\lambda_i(((1-r)V^2 + \sigma^2 I_M)Z_1 Z_1^\dagger) + \sigma^2 k)] \\
&= E[\log\det(((1-r)V^2 + \sigma^2 I_M)Z_1 Z_1^\dagger + k\sigma^2 I_M)] \\
&= E[\log\det Z_1 Z_1^\dagger] + E[\log\det((1-r)V^2 + \sigma^2 I_M + k\sigma^2(Z_1 Z_1^\dagger)^{-1})]
\end{aligned}$$

where the second inequality follows from Jensen's inequality and taking expectation over Z_2. Using Lemma 6 again on the second term, we have

$$\begin{aligned}\sum_{i=1}^{M} E[\log \sigma_i^2|S] &\leq E[\log\det Z_1Z_1^\dagger] + E[\log\det((1-r)V^2+\sigma^2 I_M \\ &\quad +k\sigma^2\|(Z_1Z_1)^{-1}\|_2 I_M)] \\ &\leq E[\log\det Z_1Z_1^\dagger] + E[\log\det((1-r)V^2+k'\sigma^2 I_M)]\end{aligned}$$

where $k' = 1 + kE[\|Z_1Z_1^\dagger\|_2]$ is a finite constant. Next, we have

$$\begin{aligned}\sum_{i=1}^{M} E[\log \sigma_i^2|S] - \sum_{i=1}^{M}\log((1-r)\|s_i\|^2+\sigma^2) &\leq E[\log\det Z_1Z_1^\dagger] + \\ &\quad \sum_{i=1}^{M}\log\frac{(1-r)\|s_i\|^2+k'\sigma^2}{(1-r)\|s_i\|^2+\sigma^2} \\ &\leq E[\log\det Z_1Z_1^\dagger] + k''\end{aligned}$$

where k'' is another constant. Taking Expectation over S, we have shown that $\sum_{i=1}^{M} E[\log\sigma_i^2] - \sum_{i=1}^{M} E[\log((1-r)\|s_i\|^2+\sigma^2)]$ is bounded above by a constant.

Note that as $\|s_i\| \to \infty$, $Z_1 \to \sqrt{\frac{1}{1-r}}H_1$ so that $E[Z_1Z_1^\dagger] \to \frac{1}{1-r}E[H_1H_1^\dagger] = \frac{1}{1-r}E[HH^\dagger]$.

Step 2 Now assume that there are M transmit and receive antennas and that for $N-M' > 0$ antennas, the transmitted signal has bounded energy, that is, $\|s_i\|^2 < C\sigma^2$ for some constant C. Start from a system with only M' transmit antennas, the extra power we send on the rest $M-M'$ antennas accrues only a limited capacity gain since the SNR is bounded. Therefore, we conclude that the mutual information must be no more than $k_2 + M'(T-M')\log\rho$ for some finite k_2 that is uniform for all SNR level and all input distributions.

Particularly, if $M' = M-1$, ie we have at least 1 transmit antenna to transmit signal with finite SNR, under the assumption that $T \geq 2M$ (T greater than twice the number of receivers), we have $M'(T-M') < M(T-M)$. This means that the mutual information achieved has an upper bound that increases with log SNR at rate $M'(T-M')\log\rho$, which is a lower rate than $M(T-M)\log\rho$.

Step 3 Now we further generalize the result above to consider the input which on at least 1 antennas, the signal transmitted has finite SNR with a positive probability, that is $P(\|s_M\|^2 < C\sigma^2) = \epsilon$. Define the event $E = \{\|s_M\|^2 < C\sigma^2\}$, then the mutual information can be written as

$$\begin{aligned}I(X;S) &\leq \epsilon I(X;S|E) + (1-\epsilon)I(X;S|E^c) + I(E;X) \\ &\leq \epsilon(k_1 + (M-1)(T-M+1)\log\rho) + (1-\epsilon)(k_2 + M(T-M)\log\rho) + \log 2\end{aligned}$$

where k_1 and k_2 are two finite constants. Under the assumption that $T \geq 2M$, the resulting mutual information thus increases with SNR at rate that is strictly less than $M(T-M)\log\rho$.

Step 4 Here we will show that for the case of M transmit and receive antennas, the constant equal norm input $P(\|s_i\| = \sqrt{T}) = 1$ for $i = 1, \ldots, M$, achieves a mutual information that increases at a rate $M(T - M\log\rho$.

Lemma 7 *For the constant equal norm input,*

$$\lim \inf_{\sigma^2 \to 0} [I(X;S) - f(\rho)] \geq 0$$

where $\rho = M/\sigma^2$*, and*

$$f(\rho) = \log|G(T,M)| + (T-M)E[\log\det HH^\dagger] + M(T-M)\log\frac{T\rho}{M\pi e} - M^2\log[(1-r)T]$$

where $|G(T,M)|$ *is as defined in Lemma 2.*

Proof: Consider

$$\begin{aligned}
\mathcal{H}(X) &\geq \mathcal{H}(SH) \\
&= \mathcal{H}(QVH) + \log|G(T,M)| + (T-M)E[\log\det H^\dagger\Psi V^2\Psi^\dagger H] \\
&= \mathcal{H}(QVH) + \log|G(T,M)| + M(T-M)\log T + (T-M)E[\log\det HH^\dagger] \\
\mathcal{H}(X|S) &\leq \mathcal{H}(QVH) + M\sum_{i=1}^{M} E[\log((1-r)\|s_i\|^2 + \sigma^2)] + M(T-M)\log\pi e\sigma^2 \\
&\approx \mathcal{H}(QVH) + M^2\log[(1-r)T] + M^2\frac{\sigma^2}{(1-r)T} + M(T-M)\log\pi e\sigma^2.
\end{aligned}$$

Therefore,

$$\begin{aligned}
I(X;S) &\geq \log|G(T,M)| + (T-M)E[\log\det HH^\dagger] - M(T-M)\log\pi e\sigma^2 + \\
&\quad M(T-M)\log T - M^2\log[(1-r)T] - M^2\frac{\sigma^2}{(1-r)T} \\
&= f(\rho) - M^2\frac{\sigma^2}{(1-r)T} \to f(\rho).
\end{aligned}$$

□

Combining the result in step 4 with results in Step 3 we see that for any input that doesn't satisfy (C.1) the mutual information increases at a strictly lower rate than for the equal norm input. Thus at high SNR, any input not satisfying (C.1) is not optimal and this completes the proof of Lemma 3.

D Convergence of $\mathcal{H}(X)$ for $T > M = N$

The results in this section are needed in the proof of Theorem 3 in Section 3.4.1. We need the following two theorems, proved in [9], for proving the results in this section.

Theorem 6 *Let $\{X_i \in \mathcal{C}^P\}$ be a sequence of continuous random variables with probability density functions, $\{f_i\}$ and $X \in \mathcal{C}^P$ be a continuous random variable with probability density function f such that $f_i \to f$ pointwise. If 1) $\max\{f_i(x), f(x)\} \leq A < \infty$ for all i and 2) $\max\{\int \|x\|^\kappa f_i(x)dx, \int \|x\|^\kappa f(x)dx\} \leq L < \infty$ for some $\kappa > 1$ and all i then $\mathcal{H}(X_i) \to \mathcal{H}(X)$. $\|x\| = \sqrt{x^\dagger x}$ denotes the Euclidean norm of x.*

Theorem 7 *Let $\{X_i \in \mathcal{C}^P\}$ be a sequence of continuous random variables with probability density functions, f_i and $X \in \mathcal{C}^P$ be a continuous random variable with probability density function f. Let $X_i \xrightarrow{P} X$. If 1) $\int \|x\|^\kappa f_n(x)dx \leq L$ and $\int \|x\|^\kappa f(x)dx \leq L$ for some $\kappa > 1$ and $L < \infty$ 2) $f(x)$ is bounded then $\limsup_{i\to\infty} \mathcal{H}(X_i) \leq \mathcal{H}(X)$.*

First, we will show convergence for the case $T = M = N$ needed for Theorem 3 and then use the result to show convergence for the general case of $T > M = N$. We need the following lemma to establish the result for $T = M = N$.

Lemma 8 *If $\lambda_{min}(SS^\dagger) \geq \lambda > 0$ then $\forall n$ there exists an $\mathcal{M}$ such that $|f(X) - f(Z)| < \mathcal{M}\delta$ if $|X - Z| < \delta$.*

Proof: Let $Z = X + \Delta X$ with $|\Delta X| < \delta$ and $[\sigma^2 I_T + (1-r)SS^\dagger] = D$. First, we will fix S and show that for all S, $f(X|S)$ satisfies the above property. Therefore, it will follow that $f(X)$ also satisfies the same property. Consider $f^0(X|S)$ the density defined with zero mean which is just a translated version of $f(X|S)$.

$$f(X + \Delta X|S) = f(X|S)[1 - \text{tr}[D^{-1}(\Delta X X^\dagger + X\Delta X^\dagger + O(\|\Delta X\|_2^2))]]$$

then

$$|f(X + \Delta X|S) - f(X|S)| \leq f(X|S)|\text{tr}[D^{-1}(\Delta X X^\dagger + X\Delta X^\dagger)] + \text{tr}[D^{-1}\|\Delta X\|_2^2]|.$$

Now

$$f(X|S) \leq \frac{1}{\pi^{TN}\det^N[D]} \cdot \min\{\frac{1}{\sqrt{\text{tr}[D^{-1}XX^\dagger]}}, 1\}.$$

Next, make use of the following inequalities

$$\begin{aligned}\mathrm{tr}\{D^{-1}XX^\dagger\} &\geq \mathrm{tr}\{\lambda_{min}(D^{-1})XX^\dagger\} \\ &\geq \lambda_{min}(D^{-1})\lambda_{max}(XX^\dagger) = \lambda_{min}(D^{-1})\|X\|_2^2.\end{aligned}$$

Also,

$$\begin{aligned}|\mathrm{tr}\{D^{-1}(X\Delta X^\dagger + \Delta X X^\dagger + O(\|\Delta X\|_2^2)\}| &\leq \sum_i |\lambda_i(D^{-1}[\Delta X X^\dagger + X\Delta X^\dagger])| + \\ &\quad \|D^{-1}\|_2\|\Delta X\|_2^2 \\ &\leq T\|D^{-1}\|_2\|X\|_2\|\Delta X\|_2 + \\ &\quad T\|D^{-1}\|_2\|\Delta X\|_2^2.\end{aligned}$$

Therefore,

$$\begin{aligned}|f(X+\Delta X|S) - f(X|S)| &\leq \frac{1}{\pi^{TN}\det^N[D]} \cdot \min\{\frac{1}{\sqrt{\lambda_{min}(D^{-1})}\|X\|_2}, 1\} \cdot \\ &\quad T\|D^{-1}\|_2\|\Delta X\|_2(\|X\|_2 + \|\Delta X\|_2).\end{aligned}$$

Since, we have restricted $\lambda_{min}(SS^\dagger) \geq \lambda > 0$ we have for some constant $\mathcal{M}$

$$|f(X+\Delta X|S) - f(X|S)| \leq \mathcal{M}\|\Delta X\|_2.$$

From which the Lemma follows. Note that $\det[D]$ compensates for $\sqrt{\lambda_{min}(D^{-1})}$ in the denominator. $\square$

Let's consider the $T \times N$ random matrix $X = SH + W$. The entries of $M \times N$ matrix H, $T = M = N$, are independent circular complex Normal random variables with non-zero mean and unit variance whereas the entries of W are independent circular complex Normal random variables with zero-mean and variance σ^2.

Let S be a random matrix such that $\lambda_{min}(SS^\dagger) \geq \lambda > 0$ with distribution, $F_{max}(S)$ chosen in such a way to maximize $I(X;S)$. For each value of $\sigma^2 = 1/n$, n an integer $\rightarrow \infty$, the density of X is

$$f(X) = E_S\left[\frac{e^{-\mathrm{tr}\{[\sigma^2 I_T + (1-r)SS^\dagger]^{-1}(X-\sqrt{rNM}SH_m)(X-\sqrt{rNM}SH_m)^\dagger\}}}{\pi^{TN}\det^N[\sigma^2 I_T + (1-r)SS^\dagger]}\right].$$

where the expectation is over $F_{max}(S)$. It is easy to see that $f(X)$ as a function of σ^2 is a continuous function of σ^2. As $\lim_{\sigma^2\rightarrow 0} f(X)$ exists, let's call this limit $g(X)$.

Since we have imposed the condition that $\lambda_{min}(SS^\dagger) \geq \lambda > 0$ w.p. 1, $f(X)$ is bounded above by $\frac{1}{(\lambda\pi)^{TN}}$. Thus $f(X)$ satisfies the condition for Theorem 6. From Lemma 8 we also have that for all n there exists a common δ such that $|f(X) - f(Z)| < \epsilon$ for all $|X - Z| < \delta$. Therefore, $\mathcal{H}(X) \to \mathcal{H}_g$. Since λ is arbitrary we conclude that for all optimal signals with the restriction $\lambda_{min}(SS^\dagger) > 0$, $\mathcal{H}(X) \to \mathcal{H}_g$. Now, we claim that the condition $\lambda_{min} > 0$ covers all optimal signals. Otherwise, if $\lambda_{min}(SS^\dagger) = 0$ with finite probability then for all σ^2 we have $\min \|s_i\|^2 \leq L\sigma^2$ for some constant L with finite probability. This is a contradiction of the condition (3.5). This completes the proof of convergence of $\mathcal{H}(X)$ for $T = M = N$. □

Now, we show convergence of $\mathcal{H}(X)$ for $T > M = N$. We will show that $\mathcal{H}(X) \approx \mathcal{H}(SH)$ for small values of σ where $S = \Phi V \Psi^\dagger$ with Φ independent of V and Ψ.

Let $S_0 = \Phi_0 V_0 \Psi_0^\dagger$ denote a signal with its density set to the limiting optimal density of S as $\sigma^2 \to 0$.

$$\mathcal{H}(X) \geq \mathcal{H}(Y) = \mathcal{H}(Q\Sigma_Y \Psi_Y^\dagger) + \log|G(T, M)| + (T - M)E[\log\det \Sigma_Y^2]$$

where $Y = SH$ and Q is an isotropic matrix of size $N \times M$. Let

$$Y_Q = QV\Psi^\dagger H$$

Then $\mathcal{H}(Q\Sigma_Y \Psi_Y^\dagger) = \mathcal{H}(Y_Q)$.

From the proof of the case $T = M = N$, we have $\lim_{\sigma^2 \to 0} \mathcal{H}(Y_Q) = \mathcal{H}(QV_0\Psi_0^\dagger H)$. Also,

$$\lim_{\sigma^2 \to 0} E[\log\det \Sigma_Y^2] = E[\log\det \Sigma_{Y_0}^2]$$

where $Y_0 = S_0 H$ Therefore, $\liminf_{\sigma^2 \to 0} \mathcal{H}(X) \geq \lim_{\sigma^2 \to 0} \mathcal{H}(Y) = \mathcal{H}(S_0 H)$.

Now, to show $\lim_{\sigma^2 \to 0} \mathcal{H}(X) \leq \mathcal{H}(S_0 H)$. From before

$$\mathcal{H}(X) = \mathcal{H}(Q\Sigma_X \Psi_X^\dagger) + |G(T, N)| + (T - M)E[\log\det \Sigma_X^2].$$

Now $Q\Sigma_X\Psi_X^\dagger$ converges in distribution to $QV_0\Psi_0^\dagger H$. Since the density of $QV_0\Psi_0^\dagger H$ is bounded, from Theorem 7 we have $\limsup_{\sigma^2 \to 0} \mathcal{H}(Q\Sigma_X\Psi_X^\dagger) \leq \mathcal{H}(QV_0\Psi_0^\dagger H)$. Also, note that $\lim_{\sigma^2 \to 0} E[\log\det \Sigma_X^2] = E[\log\det \Sigma_{Y_0}^2] = \lim_{\sigma^2 \to 0} E[\log\det \Sigma_Y^2]$. Which leads to $\limsup_{\sigma^2 \to 0} \mathcal{H}(X) \leq \mathcal{H}(S_0 H) = \lim_{\sigma^2 \to 0} \mathcal{H}(SH)$.

Therefore, $\lim_{\sigma^2 \to 0} \mathcal{H}(X) = \lim_{\sigma^2 \to 0} \mathcal{H}(SH)$ and for small σ^2, $\mathcal{H}(X) \approx \mathcal{H}(SH)$.

References

[1] I. Abou-Faycal and B. Hochwald, "Coding requirements for multiple-antenna channels with unknown Rayleigh fading," *Bell Labs Technical Memorandum*, Mar. 1999.

[2] N. Benvenuto, P. Bisaglia, A. Salloum, and L. Tomba, "Worst case equalizer for noncoherent HIPERLAN receivers," *IEEE Trans. on Communications*, vol. 48, no. 1, pp. 28–36, January 2000.

[3] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley Series in Telecommunications, 1995.

[4] P. F. Driessen and G. J. Foschini, "On the capacity formula for multiple input-multiple output wireless channels: A geometric interpretation," *IEEE Trans. on Communications*, vol. 47, no. 2, pp. 173–176, February 1999.

[5] F. R. Farrokhi, G. J. Foschini, A. Lozano, and R. A. Valenzuela, "Link-optimal space-time processing with multiple transmit and receive antennas," *IEEE Communication Letters*, vol. 5, no. 3, pp. 85–87, Mar 2001.

[6] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multiple antennas," *Bell Labs Technical Journal*, vol. 1, no. 2, pp. 41–59, 1996.

[7] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6, no. 3, pp. 311–335, March 1998.

[8] M. Godavarti, A. O. Hero, and T. Marzetta, "Min-capacity of a multiple-antenna wireless channel in a static Rician fading environment," *Submitted to IEEE Trans. on Wireless Communications*, 2001.

[9] M. Godavarti and A. O. Hero-III, "Convergence of differential entropies," *Submitted to IEEE Trans. on Information Theory*, 2001.

[10] M. Godavarti, T. Marzetta, and S. S. Shitz, "Capacity of a mobile multiple-antenna wireless link with isotropically random Rician fading," *Accepted for publication - IEEE Trans. on Information Theory*, 2001.

[11] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, 1996.

[12] A. Lapidoth and S. S. (Shitz), "Fading channels: How perfect need "perfect side information" be?," *IEEE Trans. on Inform. Theory*, vol. 48, no. 5, pp. 1118–1134, May 2002.

[13] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading channel," *IEEE Trans. on Inform. Theory*, vol. 45, no. 1, pp. 139–157, Jan 1999.

[14] S. Rice, "Mathematical analysis of random noise," *Bell Systems Technical Journal*, vol. 23, , 1944.

[15] V. Tarokh, N. Seshadri, and A. R. Calderbank, “Space-time codes for high data rate wireless communication: Performance criterion and code construction,” *IEEE Trans. on Communications*, vol. 44, no. 2, pp. 744–765, March 1998.

[16] I. E. Telatar, “Capacity of multi-antenna Gaussian channels,” *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–596, Nov/Dec 1999.

[17] S. Verdú, “Spectral efficiency in the wideband regime,” *IEEE Trans. on Inform. Theory*, vol. 48, no. 6, pp. 1319–1343, June 2002.

[18] L. Zheng and D. N. C. Tse, “Packing spheres in the Grassmann manifold: A geometric approach to the non-coherent multi-antenna channel,” *IEEE Trans. on Inform. Theory*, vol. 48, no. 2, pp. 359–383, Feb 2002.