Multiple-Antenna Capacity in a Deterministic Rician Fading Channel

Mahesh Godavarti and Alfred O. Hero-III

September 20, 2001

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

Abstract

We calculate the capacity of a multiple-antenna wireless link 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 scenarios. We also consider the capacity of the wireless system that uses pilot symbol-based training to estimate the channel. We establish that for low SNR the specular component of channel coefficients completely determines the form of the optimum signal whereas for high SNR it has no effect on the optimum signal structure. We further conclude that beamforming is the optimum signaling strategy for low SNR whereas for high SNR the optimum signal structure is same as that for purely Rayleigh fading channels. Finally, we establish that training is not effective at low SNR.

Keywords: capacity, information theory, Rician fading, multiple antennas.

Corresponding author:

Mahesh Godavarti
Altra Broadband, Inc.
Irvine Technology Center
16275 Laguna Canyon Rd
Irvine, CA 92618.
Tel: (949) 341-0106 x230.
Fax: (949) 341-0226.
e-mail: godavarti@altrabroadband.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 [7, 8, 14, 17, 19]. Telatar [17], Marzetta and Hochwald [14] and Zheng and Tse [19] 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 [3, 5, 6, 15, 16]. 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. [6] 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 [9] extend the results to non-coherent capacity (unknown channel at both the transmitter and the receiver) for the same model. In [11], 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 [14].

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. The coherent capacity for this case is considered in Section 3.1, the non-coherent capacity for low SNR in Section 3.3 and the non-coherent capacity for high SNR in Section 3.4. Finally, in Section 4 we consider the performance of a Rician channel in terms of capacity when pilot symbol based training is used in the communication system.

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 $\mathrm{tr}\{H_m^\dagger H_m\} = 1$ 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,

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

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 [7]. 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*

$$C_H = \max_{l,d} T E \log\det[I_N + \frac{\rho}{M} H^\dagger \Lambda^{(l,d)} H] \tag{2.5}$$

where the signal covariance matrix $\Lambda^{(l,d)}$ is of the form

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

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

Proof: This proof is a modification of the proof in [17]. 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 $c \ge A^\dagger B^{-1} A$ (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 $\psi = \Pi U$ where Π is a $(M-1)\times(M-1)$ permutation matrix, we obtain that

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

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

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

where $\overline{\Lambda^{\Pi}} = \frac{1}{(M-1)!}\sum_{\Pi} \Lambda^{\Pi}$ 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

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

where $d = \mathrm{tr}\{B\}/(M-1)$. Now, the capacity achieving signal matrix has to satisfy $\mathrm{tr}\{\Lambda\} = M$ since MI is monotonically increasing in $\mathrm{tr}\{\Lambda\}$. Therefore, $c = M - (M-1)d$. And since $c \geq L^{\dagger}D^{-1}L$ this implies $M - (M-1)d \geq \frac{(M-1)|l|^2}{d}$ and we obtain the desired signal covariance structure. $\square$

The problem remains to find the l and d that achieve the maximum in (2.5). This problem has an analytical solution for the special cases of: 1) $r = 0$ for which $d = 1$ and $l = 0$ (rank M signal S); and 2) $r = 1$ for which $d = l = 0$ (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. The optimum values of l for different values of ρ turned out to be zero, i.e. the signal energy transmitted is uncorrelated over different antenna elements and over time. 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 3.1.

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 $\mathrm{tr}\{H_m H_m^{\dagger}\} = MN$ 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.

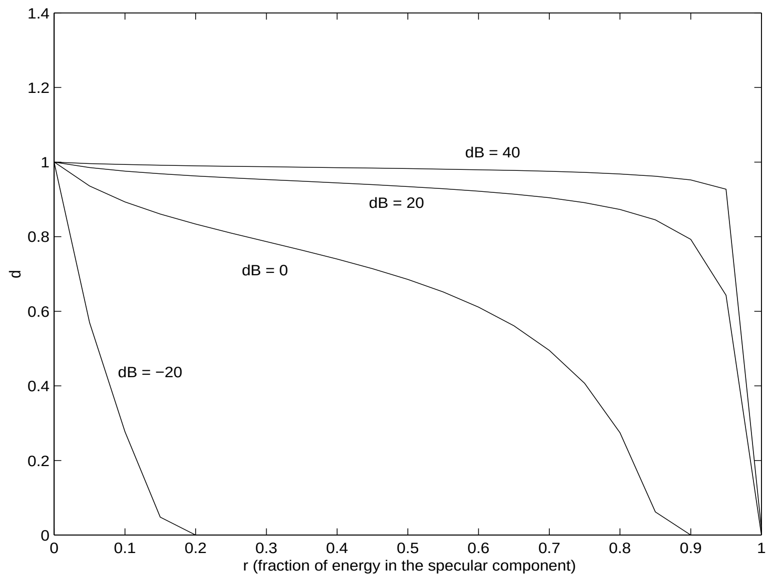

Figure 1: Optimum value of d as a function of r for different values of ρ

3.1 Coherent Capacity

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

Proposition 1 *Let H be Rician (3.1). Let C_H be the capacity for H known at the receiver. For high SNR ρ, 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 H have SVD $H = \Phi\Sigma\Psi^\dagger$ then

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

Let $\Phi^\dagger \Lambda \Phi = D$. Then

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

The right hand side expression is maximized by choosing Λ such that D is diagonal [4, page 255] (We will show finally that the optimum D does not depend on the specific realization of H). Let $D = \text{diag}\{d_1,\ d_2,\ \ldots, d_M\}$ and 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

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

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

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

since

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

and

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

On account of log being a convex cap function the first term in the expression on the last line above is maximized by choosing $d_i = d$ for $i = 1, \ldots, M$ such that $M \cdot d = M$. $\square$

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

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

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.

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

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

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 $\|H\| < 1/\rho^\gamma$ $\log\det[I_N + \frac{\rho}{M} H^\dagger \Lambda H] = \mathrm{tr}[\frac{\rho}{M} H^\dagger \Lambda H] + O(\rho^{2-2\gamma})$. Since γ is arbitrary

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

Now

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

Therefore, we have to choose Λ to maximize $\mathrm{tr}\{(1-r)N\Lambda + r\Lambda H_m H_m^\dagger\}$. 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, $\mathrm{tr}\{\Lambda\} \leq M$, to maximize the expression we choose $\mathrm{tr}\{\Lambda\} = M$. 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. □

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 [9] and [11] 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.

3.2 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 3 *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 [9] and [11].

Proposition 4 *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 [11] and won't be repeated here. □

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

3.3 Non-Coherent Capacity: Expressions for Low SNR

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

For low SNR, it will be shown that the channel behaves as an AWGN channel. Calculation of capacity for the special case of peak power constraint has been shown in Appendix B.

Theorem 2 *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: 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

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

$\square$

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

The peak constraint results in Appendix B and the Gaussian input results imply that for low SNR, Rayleigh fading channels are at a capacity disadvantage as compared to Rician fading channels for equal values of ρ. But, it has been shown in [2, 18] for single antenna transmit and receive channel Rayleigh fading provides as much capacity as a Gaussian channel for low SNR. We next extend that result to multiple transmit and receive antenna channel for the general case of Rician fading. The result for Rayleigh fading will follow as a special case.

Theorem 3 *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: First, absorb $\sqrt{\frac{\rho}{M}}$ into $\tilde{S}$ and rewrite the channel as

$$\tilde{X} = \hat{H}\tilde{S} + W$$

with the average power constraint on the signal $\tilde{S}$ $E[\mathrm{tr}\{\tilde{S}\tilde{S}^\dagger\}] \leq \frac{\rho}{M}TM = \rho T$.

It has been shown [18] that if the input alphabet includes the value "0" (symbol with 0 power) for a channel with output X, and conditional probability denoted by $p(X|S)$, then

$$\lim_{P_C \to 0} \frac{C}{P_C} = \sup_{s \in \mathcal{S}} \frac{D(\ p(X|S=s) \ \|\ p(X|S=0)\ )}{P_s}$$

where $\mathcal{S}$ is the set of values that the input can take, P_C is the average power constraint on the input (in our case, $E[\mathrm{tr}\{SS^\dagger\}] \leq P_C = \rho T$) and $P_s = \mathrm{tr}\{ss^\dagger\}$ is the energy in the specific realization of the input $S = s$ and $D(p_A \| p_B)$ is the Kullback-Leibler distance for continuous density functions with argument x defined as

$$D(p_A \| p_B) = \int p_A(x) \log \frac{p_A(x)}{p_B(x)} dx.$$

Applying the above result to the case of Rician fading channels, we obtain

$$\lim_{\rho \to 0} \frac{C}{\rho T} = \sup_{\tilde{S}} \frac{D(\ p(\tilde{X}|\tilde{S}) \ \|\ p(\tilde{X}|0)\ )}{\mathrm{tr}\{\tilde{S}\tilde{S}^\dagger\}}.$$

First, we have

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

and

$$p(\tilde{X}|0) = \frac{1}{\pi^{TN}} e^{-\tilde{X}^\dagger \tilde{X}}.$$

Therefore,

$$\begin{aligned} D(\ p(\tilde{X}|\tilde{S}) \parallel p(\tilde{X}|0)\ ) &= \int p(\tilde{X}|\tilde{S}) \Bigg[\log \frac{1}{|\Lambda_{\tilde{X}|\tilde{S}}|} \ + \ \tilde{X}^\dagger \tilde{X} \ - \\ & \left(\tilde{X} - \sqrt{r}\hat{H}_m \tilde{S}\right)^\dagger \ \Lambda_{\tilde{X}|\tilde{S}}^{-1} \ \left(\tilde{X} - \sqrt{r}\hat{H}_m \tilde{S}\right)\Bigg] d\tilde{X} \\ &= \log \frac{1}{|\Lambda_{\tilde{X}|\tilde{S}}|} \ + \ \mathrm{tr}\left\{r\hat{H}_m \tilde{S}\tilde{S}^\dagger \hat{H}_m^\dagger + \Lambda_{\tilde{X}|\tilde{S}}\right\} - TN \\ &= \log \frac{1}{\det(I_{TN} + (1-r)SS^\dagger \otimes I_N)} \ + \\ & \mathrm{tr}\left\{r\hat{H}_m \tilde{S}\tilde{S}^\dagger \hat{H}_m^\dagger + (1-r)SS^\dagger \otimes I_N\right\}. \end{aligned}$$

This gives,

$$\begin{aligned} \frac{D(\ p(\tilde{X}|\tilde{S}) \parallel p(\tilde{X}|0)\ )}{\mathrm{tr}\{\tilde{S}\tilde{S}^\dagger\}} &= -N\frac{\sum_{i=1}^T \log(1+\lambda_i(SS^\dagger))}{\sum_{i=1}^T \lambda_i(SS^\dagger)} \ + \ \frac{\mathrm{tr}\{r\hat{H}_m \tilde{S}\tilde{S}^\dagger \hat{H}_m^\dagger\}}{\sum_{i=1}^T \mathrm{tr}\{S_i S_i^\dagger\}} \ + \\ & N(1-r) \end{aligned}$$

where we have used the facts that $\det(I_{TN}+(1-r)SS^\dagger \otimes I_N) = \det(I_T+(1-r)SS^\dagger)^N$, $\mathrm{tr}\{\tilde{S}\tilde{S}^\dagger\} = \mathrm{tr}\{SS^\dagger\} = \sum_{i=1}^T \mathrm{tr}\{S_i^\tau S_i^*\}$ where S_i is the i^{th} row in the matrix S.

Since,

$$\hat{H}_m \tilde{S}\tilde{S}^\dagger \hat{H}_m^\dagger = \begin{bmatrix} H_m^\tau S_1^\tau S_1^* H_m^* & H_m^\tau S_1^\tau S_2^* H_m^* & \ldots & H_m^\tau S_1^\tau S_T^* H_m^* \\ H_m^\tau S_2^\tau S_1^* H_m^* & H_m^\tau S_2^\tau S_2^* H_m^* & \ldots & H_m^\tau S_2^\tau S_T^* H_m^* \\ \vdots & \vdots & \ddots & \vdots \\ H_m^\tau S_T^\tau S_1^* H_m^* & H_m^\tau S_T^\tau S_2^* H_m^* & \ldots & H_m^\tau S_T^\tau S_T^* H_m^* \end{bmatrix}$$

we have $\mathrm{tr}\{\hat{H}_m \tilde{S}\tilde{S}^\dagger \hat{H}_m^\dagger\} = \sum_{i=1}^T \mathrm{tr}\{H_m^\tau S_i^\tau S_i^* H_m^*\} = \mathrm{tr}\{H_m^* H_m^\tau \sum_{i=1}^T S_i^\tau S_i^*\}$. Therefore,

$$\frac{D(\ p(\tilde{X}|\tilde{S}) \parallel p(\tilde{X}|0)\ )}{\mathrm{tr}\{\tilde{S}\tilde{S}^\dagger\}} = -N\frac{\sum_{i=1}^N \log(1+\lambda_i(S^\dagger S))}{\sum_{i=1}^N \lambda_i(S^\dagger S)} \ + \ r\frac{\mathrm{tr}\{H_m^* H_m^\tau \sum_{i=1}^T S_i^\tau S_i^*\}}{\sum_{i=1}^T \mathrm{tr}\{S_i^\tau S_i^*\}} + N(1-r).$$

Note that since H_m is a diagonal matrix only the diagonal elements of $S_i S_i^\dagger$ affect the second term. Therefore, for a given $\sum_{i=1}^T S_i^\tau S_i^*$ the second term in right hand side of the expression above can be maximized by choosing S_i such that $S_i^\tau S_i^*$ is diagonal. In addition the non-zero values of $S_i S_i^\dagger$ should be located at the same diagonal positions as the maximum entries of $H_m^* H_m^\tau$. In such a case the expression above evaluates to

$$\frac{D(\ p(\tilde{X}|\tilde{S}) \parallel p(\tilde{X}|0)\ )}{\mathrm{tr}\{\tilde{S}\tilde{S}^\dagger\}} = -N\frac{\log(1+\mathrm{tr}\{S^\dagger S\})}{\mathrm{tr}\{S^\dagger S\}} \ + \ r\lambda_{max}(H_m^* H_m^\tau) + N(1-r).$$

The first term can be made arbitrarily small by letting $\mathrm{tr}\{S^\dagger S\} \to \infty$. Therefore, we have $\lim_{\rho\to 0} \frac{C}{\rho T} = r\lambda_{max}(H_m H_m^\dagger) + N(1-r)$. □

Theorem 3 suggests that at low SNR all the energy has to be concentrated in the strongest directions of the specular component. In [2] it is shown that the optimum signaling scheme for Rayleigh fading channels is an "on-off" signaling scheme. We conjecture that the capacity achieving signaling scheme for low SNR in the case of the Rician fading is also a similar "on-off" signaling scheme.

3.4 Non-Coherent Capacity: Expressions for High SNR

In this section we apply the method developed in [19] for the analysis of Rayleigh fading channels. The only difference between the models considered in [19] 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 $\mathrm{tr}\{H_m H_m^\dagger\} = MN$. G satisfies $E[\mathrm{tr}\{GG^\dagger\}] = MN$ and r is a number between 0 and 1 so that $E[\mathrm{tr}\{HH^\dagger\}] = MN$.

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 [19] 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 [19] 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 $G(T,M)$ [19] 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 $(s_i^\sigma, i = 1,\dots,M)$ be the optimal input signal of each antenna at when the noise power at the receive antennas is given by σ^2. If $T \geq 2M$,*

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

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 $i = 1,\dots,M$ 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 4 *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 $i = 1, \ldots, M$

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

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 5 *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 [19] 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 $T \to \infty$

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

$\square$

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 6 *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. $\square$

4 Training in Non-Coherent Communications

It is important to know whether training based signal schemes are practical and if they are how much time can be spent in learning the channel and what the optimal training signal is like. Hassibi and Hochwald [12] have addressed these issues for the case of Rayleigh fading channels. They showed that 1) pilot symbol training based communication schemes are highly suboptimal for low SNR and 2) when practical the optimal amount of time devoted to training is equal to the number of transmitters, M when the fraction of power devoted to training is allowed to vary and 3) the orthonormal signal is the optimal signal for training.

In [19] the authors demonstrate a very simple training method that achieves the optimal rate of increase with SNR. The same training method can also be easily applied to the Rician fading model with deterministic specular component. The training signal is the $M \times M$ diagonal matrix dI_M. d is chosen such that the same power is used in the training and the communication phase. Therefore, $d = \sqrt{M}$. Using $S = dI_M$, the output of the MIMO channel in the training phase is given by

$$X = \sqrt{M}\sqrt{r}H_m + \sqrt{M}\sqrt{1-r}G + W.$$

The Rayleigh channel coefficients G can be estimated independently using scalar minimum mean squared error (MMSE) estimates since the elements of W and G are i.i.d. Gaussian random variables

$$\hat{G} = \frac{\sqrt{1-r}\sqrt{M}}{(1-r)M + \sigma^2}[X - \sqrt{M}\sqrt{r}H_m],$$

where we recall that σ^2 is the variance of the components of W. The elements of the estimate $\hat{G}$ are i.i.d. Gaussian with variance $\frac{(1-r)M}{(1-r)M+\sigma^2}$. Similarly, the estimation error matrix $G - \hat{G}$ has i.i.d Gaussian distributed elements with zero mean and variance $\frac{\sigma^2}{(1-r)M+\sigma^2}$.

The output of the channel in the communication phase is given by

$$\begin{aligned} X &= SH + W \\ &= \sqrt{r}SH_m + \sqrt{1-r}S\hat{G} + \sqrt{1-r}S(G - \hat{G}) + W, \end{aligned}$$

where S consists of zero mean i.i.d circular Gaussian random variables with zero mean and unit variance. This choice of S is sub-optimal as this might not be the capacity achieving signal, but this choice gives us a lower bound on capacity. Let $\hat{W} = \sqrt{1-r}S(G - \hat{G}) + W$. For the choice of S given above the entries of $\hat{W}$ are uncorrelated with each other and also with $S(\sqrt{r}H_m + \sqrt{1-r}\hat{G})$. The variance of each of the entries of $\hat{W}$ is given by $\sigma^2 + (1-r)M\frac{\sigma^2}{(1-r)M+\sigma^2}$. If $\hat{W}$ is replaced with a white Gaussian noise with the same covariance matrix then the resulting mutual information is a lower bound on the actual mutual information [4, p. 263]. This result is formally stated in Proposition 5. In this section we deal with normalized capacity C/T instead of capacity C. The lower bound on the normalized capacity is given by

$$C/T \geq \frac{T - T_t}{T} E \log\det\left(I_M + \frac{\rho_{eff}}{M}H_1H_1^\dagger\right)$$

where ρ_{eff} in the expression above is the effective SNR at the output (explained at the end of this paragraph), and H_1 is a Rician channel with a new Rician parameter r_{new} where $r_{new} = \frac{r}{r+(1-r)\frac{(1-r)M}{(1-r)M+\sigma^2}}$. This lower bound can be easily calculated because the lower bound is essentially the coherent capacity with H replaced by $\sqrt{r_{new}}H_m + \sqrt{1-r_{new}}\hat{G}$. The signal covariance structure was chosen to be an identity matrix as this is the optimum covariance matrix for high SNR. The effective SNR is now given by the ratio of the energy of the elements of $S(\sqrt{r}H_m + \sqrt{1-r}\hat{G})$ to the energy of the elements of $\hat{W}$. The energy in the elements of $S(\sqrt{r}H_m + \sqrt{1-r}\hat{G})$ is given by $M(r + (1-r)^2\frac{M}{(1-r)M+\sigma^2})$ and the energy in the elements of $\hat{W}$ are given by $\sigma^2 + \frac{(1-r)M\sigma^2}{(1-r)M+\sigma^2}$. Therefore, the effective SNR, ρ_{eff} is given by $\frac{\rho[r+r(1-r)\rho+(1-r)^2\rho]}{[1+2(1-r)\rho]}$ where $\rho = \frac{M}{\sigma^2}$ is the actual SNR. Note, for $r = 1$ no training is required since the channel is completely known.

This simple scheme achieves the optimum increase of capacity with SNR and uses only M of the T symbols for training. The performance of this scheme is plotted with respect to different SNR values for comparison with the asymptotic upper bound to capacity in the proof of Theorem 5. The plot also verifies the result of Theorem 5. The plots are for $M = N = 5$, $r = 0.9$ and $T = 50$ in Figure 2 the specular component is a rank-one specular component given by (2.3).

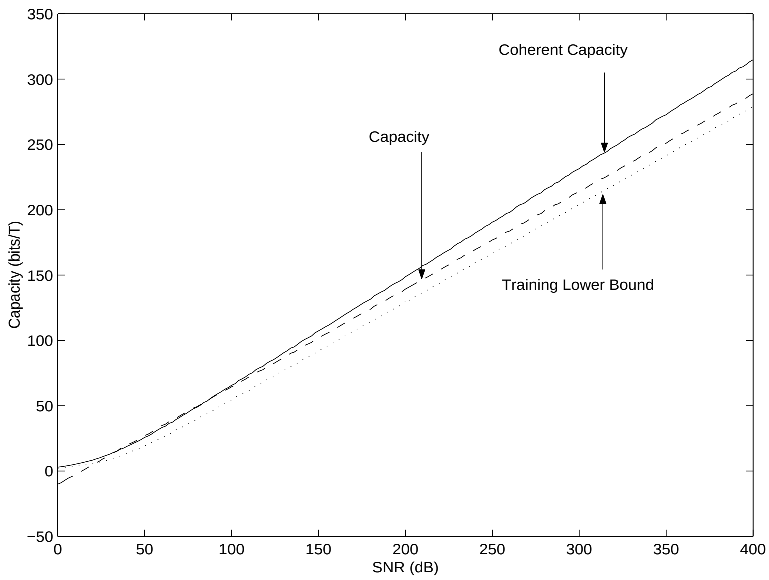

Figure 2: Asymptotic capacity, Capacity Upper and Lower bounds for different values of SNR

We can quantify the amount of training required using the techniques in [12]. In [12], the authors use the optimization of the lower bound on capacity to find the optimal allocation of training as compared to communication. Let T_t denote the amount of time devoted to training and T_c the amount of time devoted to actual communication. Let S_t be the $T_t \times M$ signal used for training and S_c the $T_c \times M$ signal used for communication.

Let the "energy allocation factor" κ denote the fraction of the energy used for communication. Then $T = T_t + T_c$ and $\mathrm{tr}\{S_t S_t^\dagger\} = (1-\kappa)TM$ and $\mathrm{tr}\{S_c S_c^\dagger\} = \kappa TM$.

$$\begin{aligned} X_t &= S_t(\sqrt{r}H_m + \sqrt{1-r}G) + W_t \\ X_c &= S_c(\sqrt{r}H_m + \sqrt{1-r}G) + W_c \end{aligned}$$

where X_t is $T_t \times N$ and X_c is $T_c \times N$. G is estimated from the training phase. For that we need $T_t \geq M$. Since G and W_t are Gaussian the MMSE estimate of G is also the linear MMSE estimate conditioned on S. The optimal estimate is given by

$$\hat{G} = \sqrt{1-r}(\sigma^2 I_M + (1-r)S_t^\dagger S_t)^{-1} S_t^\dagger (X_t - \sqrt{r} S_t H_m).$$

Let $\bar{G} = G - \hat{G}$ then

$$X_c = S_c(\sqrt{r}H_m + \sqrt{1-r}\hat{G}) + \sqrt{1-r}S_c\bar{G} + W_c.$$

Let $\hat{W}_c = \sqrt{1-r}S_t\bar{G} + W$. Note that elements of $\hat{W}_c$ are uncorrelated with each other and have the same marginal densities when the elements of S_c are chosen to be i.i.d Gaussian. If we replace $\hat{W}_c$ with Gaussian noise that is zero-mean and spatially and temporally independent the elements of which have the same variance as the elements of $\hat{W}_c$ then the resulting mutual information is a lower bound to the actual mutual information in the above channel. This is stated formally in the following proposition.

Proposition 5 (Theorem 1 in [12]) *Let*

$$X = SH + W$$

be a Rician fading channel with H known to the receiver. Let S and W satisfy $\frac{1}{M}E[SS^\dagger] = 1$ and $\frac{1}{M}E[WW^\dagger] = \sigma^2$ and be uncorrelated with each other. Then the worst case noise has i.i.d. zero mean Gaussian distribution, i.e. $W \sim \mathcal{CN}(0, I_N)$. Moreover, this distribution has the following minimax property

$$I_{W\sim\mathcal{CN}(0,\sigma^2 I_N),S}(X;S) \leq I_{W\sim\mathcal{CN}(0,\sigma^2 I_N),S\sim\mathcal{CN}(0,I_M)}(X;S) \leq I_{W,S\sim\mathcal{CN}(0,I_M)}(X;S)$$

where $I_{W\sim\mathcal{CN}(0,\sigma^2 I_N),S}(X;S)$ denotes the mutual information between X and S when W has a zero mean complex circular Gaussian distribution and S has any arbitrary distribution.

The variance of the elements of $\hat{W}_c$ is given by

$$\sigma^2_{w_c} = \sigma^2 + \frac{1-r}{NT_c}\mathrm{tr}\{E[\bar{G}\bar{G}^\dagger]\kappa T I_M\}$$

$$
\begin{aligned}
&= \sigma^2 + \frac{(1-r)\kappa T M}{T_c}\frac{1}{NM}\mathrm{tr}\{E[\bar{G}\bar{G}^\dagger]\} \\
&= \sigma^2 + \frac{(1-r)\kappa T M}{T_c}\sigma^2_{\bar{G}}
\end{aligned}
$$

and the lower bound is

$$
C_t/T \geq \frac{T-T_t}{T} E \log\det\left(I_M + \frac{\rho_{eff}}{M} H_1 \Lambda H_1^\dagger\right), \tag{4.1}
$$

where the "post training SNR" ρ_{eff}, is the ratio of the energies in the elements of $S_c\hat{H}$ and energies in the elements of $\hat{W}_c$ and $H_1 = \sqrt{r_{new}}H_m + \sqrt{1-r_{new}}\hat{G}$ where $r_{new} = \frac{r}{r+(1-r)\sigma^2_{\hat{G}}}$. Λ is the optimum signal correlation matrix the form of which depends on the distribution of H_1 according to Proposition 2 for low SNR and Proposition 1 for high SNR and $M \leq N$ as given in Section 3.1.

To calculate ρ_{eff}, the energy in the elements of $S\hat{H}$ is given by

$$
\begin{aligned}
\sigma^2_{SH} &= \frac{1}{NT_c}[r\mathrm{tr}\{H_m H_m^\dagger \kappa T I_M\} + (1-r)\mathrm{tr}\{\hat{G}\hat{G}^\dagger \kappa T I_M\}] \\
&= \frac{\kappa T M}{T_c}\frac{1}{NM}[rNM + (1-r)\mathrm{tr}\{\hat{G}\hat{G}^\dagger\}] \\
&= \frac{\kappa T M}{T_c}[r + (1-r)\sigma^2_{\hat{G}}],
\end{aligned}
$$

which gives us

$$
\rho_{eff} = \frac{\kappa T \rho[r + (1-r)\sigma^2_{\hat{G}}]}{T_c + (1-r)\kappa T \rho \sigma^2_{\hat{G}}}.
$$

4.1 Optimization of S_t, κ and T_t

We will optimize S_t, κ and T_t to maximize the lower bound (4.1). In this section we merely state the main results and their interpretations. Derivations and details are given in the Appendices.

Optimization of the lower bound over S_t is difficult as S_t effects the distribution of $\hat{H}$, the form of Λ as well as ρ_{eff}. To make the problem simpler we will just find the value of S_t that maximizes ρ_{eff}.

Theorem 7 *The signal S_t that maximizes ρ_{eff} satisfies the following condition*

$$
S_t^\dagger S_t = (1-\kappa)T I_M
$$

and the corresponding ρ_{eff} is

$$
\rho^*_{eff} = \frac{\kappa T \rho[Mr + \rho(1-\kappa)T]}{T_c(M + \rho(1-\kappa)T) + (1-r)\kappa T \rho M}.
$$

Proof: See Appendix E. $\square$

The optimum signal derived above is the same as the optimum signal derived in [12].

The corresponding capacity lower bound using the S_t obtained above is

$$C_t/T \geq \frac{T - T_t}{T} E \log\det\left(I_M + \frac{\rho_{eff}}{M} H_1 \Lambda H_1^\dagger\right),$$

where ρ_{eff} is as given above and $H_1 = \sqrt{r_{new}}H_m + \sqrt{1-r_{new}}G$ where $r_{new} = r\frac{1+(1-r)(1-\kappa)\frac{\rho}{M}T}{r+(1-r)(1-\kappa)\frac{\rho}{M}T}$ and as before G is a matrix consisting of i.i.d. Gaussian circular random variables with mean zero and unit variance. Now, Λ is the covariance matrix of the source S_c when the channel is Rician and known to the receiver. The form of Λ was derived for $\rho_{eff} \to 0$ and $\rho_{eff} \to \infty$ in Section 3.1.

Optimization of (4.1) over the energy allocation factor κ, is straightforward as κ affects the lower bound only through the post training SNR ρ_{eff}, and can be stated as the following proposition.

Theorem 8 *For fixed T_t and T_c the optimal power allocation κ in a training based scheme is given by*

$$\kappa = \begin{cases} \min\{\gamma - \sqrt{\gamma(\gamma-1-\eta)}, 1\} & \text{for } T_c > (1-r)M \\ \min\{\frac{1}{2} + \frac{rM}{2T\rho}, 1\} & \text{for } T_c = (1-r)M \\ \min\{\gamma + \sqrt{\gamma(\gamma-1-\eta)}, 1\} & \text{for } T_c < (1-r)M \end{cases}$$

where $\gamma = \frac{MT_c + T\rho T_c}{T\rho[T_c - (1-r)M]}$ and $\eta = \frac{rM}{T\rho}$. The corresponding lower bound is given by

$$C_t/T \geq \frac{T - T_t}{T} E \log\det\left(I_M + \frac{\rho_{eff}}{M} H_1 \Lambda H_1^\dagger\right)$$

where for $T_c > (1-r)M$

$$\rho_{eff} = \begin{cases} \frac{T\rho}{T_c - (1-r)M}(\sqrt{\gamma} - \sqrt{\gamma - 1 - \eta})^2 & \text{when} \kappa = \gamma - \sqrt{\gamma(\gamma-1-\eta)} \\ \frac{r\rho}{1+(1-r)\rho} & \text{when} \kappa = 1 \end{cases}$$

for $T_c = (1-r)M$

$$\rho_{eff} = \begin{cases} \frac{T^2\rho^2}{4(1-r)M(M+T\rho)}(1 + \frac{rM}{T\rho})^2 & \text{when} \kappa = \frac{1}{2} + \frac{rM}{2T\rho} \\ \frac{rT\rho}{(1-r)(M+T\rho)} & \text{when} \kappa = 1 \end{cases}$$

and for $T_c < (1-r)M$

$$\rho_{eff} = \begin{cases} \frac{T\rho}{(1-r)M - T_c}(\sqrt{-\gamma} - \sqrt{-\gamma + 1 + \eta})^2 & \text{when} \kappa = \gamma + \sqrt{\gamma(\gamma-1-\eta)} \\ \frac{r\rho}{1+(1-r)\rho} & \text{when} \kappa = 1 \end{cases}$$

and r_{new} is given by substituting the appropriate value of κ in the expression

$$r\frac{1 + (1-r)(1-\kappa)\frac{\rho}{M}T}{r + (1-r)(1-\kappa)\frac{\rho}{M}T}.$$

Proof: See Appendix F. □

For optimization over T_t we draw similar conclusions as in [12]. In [12] the optimal setting for T_t was shown to be $T_t = M$ for all values of SNR. We however show that for small SNR the optimal setting is $T_t = 0$ i.e., no training is required. When training is required, the intuition is that increasing T_t linearly decreases the capacity through the term $(T - T_t)/T$, but only logarithmically increases the capacity through the higher effective SNR ρ_{eff} [12]. Therefore, it makes sense to make T_t as small as possible. For small SNR we show that $\kappa = 1$ i.e., all energy is allocated to communications. It is clear that optimization of T_t makes sense only when κ is strictly less than 1. When $\kappa = 1$ no power is devoted to training and T_t can be made as small as possible which is zero. When $\kappa < 1$ the smallest value T_t can be is M since it takes at least that many intervals to completely determine the unknowns.

Theorem 9 *The optimal length of the training interval is $T_t = M$ whenever $\kappa < 1$ for all values of ρ and $T > M$, and the capacity lower bound is*

$$C_t/T \geq \frac{T-M}{T} E \log\det\left(I_M + \frac{\rho_{eff}}{M} H_1 \Lambda H_1^\dagger\right) \tag{4.2}$$

where

$$\rho_{eff} = \begin{cases} \frac{T\rho}{T-(2-r)M}(\sqrt{\gamma} - \sqrt{\gamma - 1 - \eta})^2 & \text{for } T > (2-r)M \\ \frac{T^2\rho^2}{4(1-r)M(M+T\rho)}(1 + \frac{rM}{T\rho})^2 & \text{for } T = (2-r)M \\ \frac{T\rho}{T-(2-r)M}(\sqrt{-\gamma} - \sqrt{-\gamma + 1 + \eta})^2 & \text{for } T < (2-r)M \end{cases}$$

The optimal power allocations are easily obtained from Theorem 8 by simply setting $T_c = T - M$.

Proof: See Appendix G □

4.2 Equal training and data power

As stated in [12], sometimes it is difficult for the transmitter to assign different powers for training and communication phases. In this section, we will concentrate on setting the training and communication powers equal to each other in the following sense

$$\frac{(1-\kappa)T}{T_t} = \frac{\kappa T}{T_c} = \frac{\kappa T}{T - T_t} = 1$$

this means $\kappa = 1 - T_t/T$ and that the power transmitted in T_t and T_c are equal.

In this case,

$$\rho_{eff} = \frac{\rho[r + \rho\frac{T_t}{M}]}{1 + \rho[\frac{T_t}{M} + (1-r)]}$$

and the capacity lower bound is

$$C_t/T \geq \frac{T-T_t}{T} E \log\det(I_M + \frac{\rho_{eff}}{M} H_1 \Lambda H_1^\dagger)$$

where ρ_{eff} is as given above and $H_1 = \sqrt{r_{new}} H_m + \sqrt{1-r_{new}} G$ where $r_{new} = r\frac{1+(1-r)\frac{\rho}{M}T_t}{r+(1-r)\frac{\rho}{M}T_t}$.

4.3 Numerical Comparisons

Throughout the section we have chosen the number of transmit antennas M, and receive antennas N, to be equal and $H_m = I_M$.

The Figures 3 and 4 show r_{new} and κ respectively as a function of r for different values of SNR. The plots have been calculated for a block length given by $T = 40$ and the number of transmit and receive antennas given by $M = N = 5$. Figure 3 shows that for low SNR values the channel behaves like a purely AWGN channel given by $\sqrt{r}H_m$ and for high SNR values the channel behaves exactly like the original Rician fading channel. Figure 4 shows that as the SNR goes to zero less and less power is allocated for training. This agrees with the plot in Figure 3.

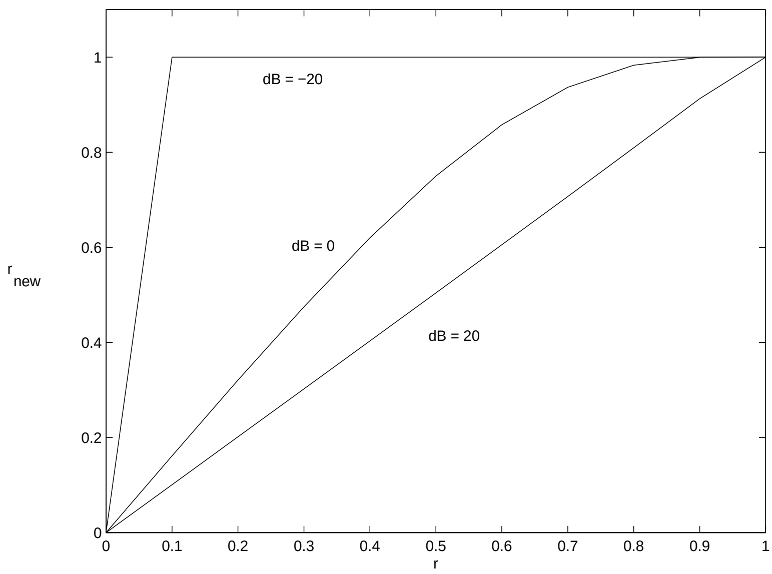

Figure 3: Plot of r_{new} as a function of Rician parameter r

In Figure 5 we plot the training and communication powers for $M = N = 10$ and $dB = 18$ for different values of r. We see that as r goes to 1 less and less power is allocated to the training phase. This makes sense as the proportion of the energy through the specular component increases there is less need for the system to estimate the unknown Rayleigh component.

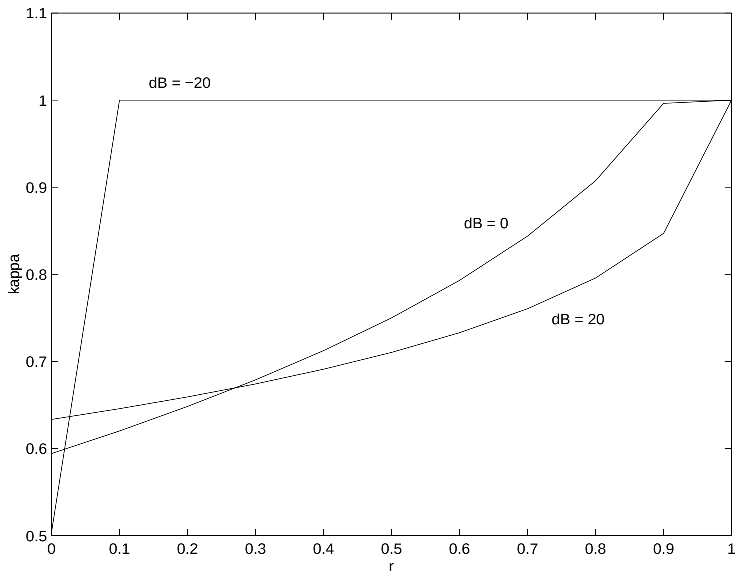

Figure 4: Plot of optimal energy allocation κ as a function of Rician parameter r

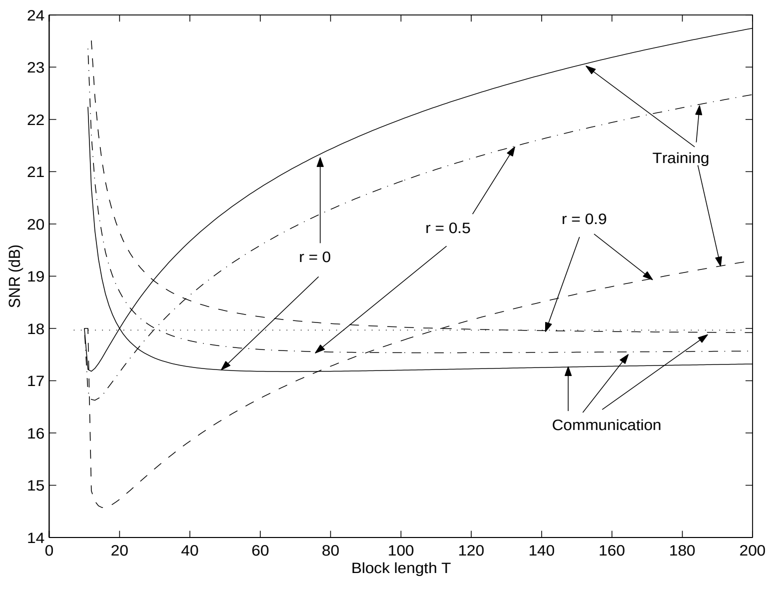

Figure 5: Plot of optimal power allocation as a function of T

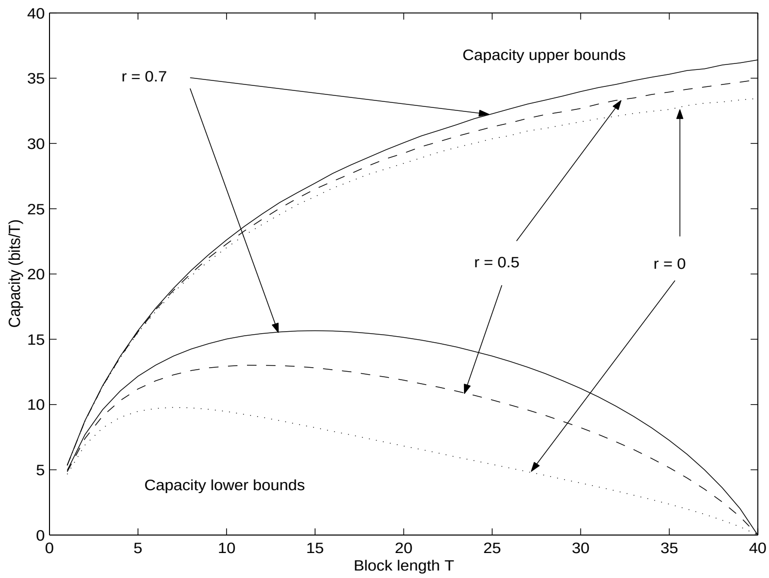

Figure 6: Plot of capacity as a function of number of transmit antennas for a fixed T

Figure 6 shows capacity as a function of the number of transmit antennas for a fixed block length $T = 40$ when $dB = 0$ and $N = 40$. We can easily see calculate the optimum number of transmit antennas from the figure. In this case, we see that for a fixed T the optimum number of transmit antennas increases as as r increases. This shows that as r goes to 1 there is a lesser need to estimate the unknown Rayleigh part of the channel and this agrees very well with Figure 5 and Figure 7 as well which shows that the optimal amount of training decreases as r increases. Figure 7 shows the optimal training period, constrained to integer multiples of the symbol interval T, as a function of the block length for the case of equal transmit and training powers.

4.4 Effect of Low SNR on Capacity Lower Bound

Let's consider the effect of low SNR on the optimization of κ when $r \neq 0$. For $T_c > (1-r)M$, as $\rho \to 0$ it is easy to see that $\gamma - \sqrt{\gamma(\gamma - 1 - \eta)} \to \infty$. Therefore, we conclude that for small ρ we have $\kappa = 1$. Similarly, for $T_c = (1-r)M$ and $T_c < (1-r)M$. Therefore, the lower bound tells us that no energy need be spent on training for small ρ. Also, the form of Λ is known from Section 3.1.

Evaluating the case where the training and transmission powers are equal we come to a similar conclusion. For small ρ, $\rho_{eff} \approx r\rho$ which is independent of T_t. Therefore, the best value of T_t is $T_t = 0$. Which also means that we spend absolutely no time on training. This is in stark contrast to the case when $r = 0$. In this case, for low SNR $T_t = T/2$ [12] and ρ_{eff} behaves as $O(\rho^2)$.

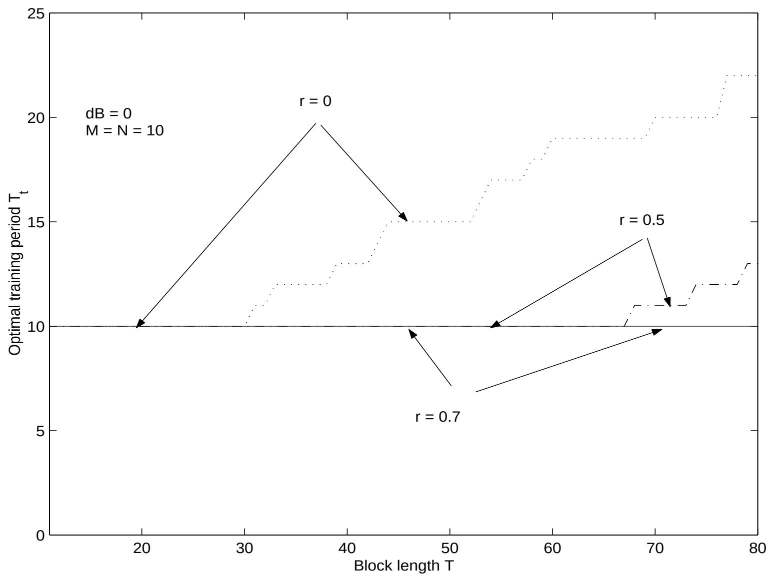

Figure 7: Optimal T_t as a function of T for equal transmit and training powers

Note that in both cases of equal and unequal power distribution between training and communication phases the signal distribution during data transmission phase is Gaussian. Therefore, the lower bound behaves as $r\rho\lambda_{max}\{H_m H_m^\dagger\}$. Also, $r_{new} = 1$ for small ρ showing that the channel behaves as a purely Gaussian channel.

These conclusions mimic those in Section 3.3 for capacity results with Gaussian input. The low SNR non-coherent capacity results for the case of a Gaussian input tell us that the capacity behaves as $r\rho\lambda_{max}$ with Gaussian input. Moreover, the results in [12] also agree with the results derived in Section 3.3. We showed that for purely Rayleigh fading channels with Gaussian input the capacity behaves as ρ^2 which is what the lower bound results in [12] also show. This makes sense because the capacity lower bound assumes that the signaling input during communication period is Gaussian. This shows that the lower bound derived in [12] and extended here is quite tight for low SNR values.

4.5 Effect of High SNR on Capacity Lower Bound

For high SNR, γ becomes $\frac{T_c}{T_c-(1-r)M}$ and the optimal power allocation κ becomes

$$\kappa = \frac{\sqrt{T_c}}{\sqrt{T_c} + \sqrt{(1-r)M}}$$

and

$$\rho_{eff} = \frac{T}{(\sqrt{T_c} + \sqrt{(1-r)M})^2}\rho.$$

In the case of equal training and transmit powers, we have for high ρ

$$\rho_{eff} = \rho \frac{T_t}{T_t + M(1-r)}.$$

For high SNR, the channel behaves as if it is completely known to the receiver. Note that in this case $r_{new} = r$ and Λ is an identity matrix for the case $M \leq N$. From the expressions for ρ_{eff} given above we conclude that unlike the case of low SNR the value of r affects the amount of time and power devoted for training.

Next consider the capacity lower bound for high SNR. The optimizing signal covariance matrix Λ, in this regime is an identity matrix. We know that at high SNR the optimal training period is M. Therefore, the resulting lower bound is given by

$$C_t/T \geq \frac{T-M}{T} E \log\det \left(I_M + \frac{\rho}{\left(\sqrt{1-\frac{M}{T}} + \sqrt{\frac{(1-r)M}{T}}\right)^2} \frac{HH^\dagger}{M} \right).$$

Note that the lower bound has H figuring in it instead of H_1. That is so because for high SNR, $r_{new} = r$. This lower bound can be optimized over the number of transmit antennas used in which case the lower bound can be rewritten as

$$C_t/T \geq \max_{M' \leq M} \max_{n \leq \binom{M}{M'}} \frac{T-M'}{T} E \log\det \left(I_{M'} + \frac{\rho}{\left(\sqrt{1-\frac{M'}{T}} + \sqrt{\frac{(1-r)M'}{T}}\right)^2} \frac{H^n {H^n}^\dagger}{M'} \right),$$

where now H^n is the n^{th} matrix out of a possible M choose M' (the number of ways to choose M' transmit elements out of a maximum M elements) matrices of size $M' \times N$. Let $Q = \min\{M', N\}$ and λ_i^n be an arbitrary nonzero eigenvalue of $\frac{1}{\left(\sqrt{1-\frac{M'}{T}}+\sqrt{\frac{(1-r)M'}{T}}\right)^2} \frac{H^n {H^n}^\dagger}{M'}$ then we have

$$C_t/T \geq \max_{M' \leq M} \max_{n \leq \binom{M}{M'}} \left(1 - \frac{M'}{T}\right) \sum_{i=1}^{Q} E \log(1 + \rho \lambda_i^n).$$

At high SNR, the leading term involving ρ in $\sum_{i=1}^{Q} E \log(1+\rho\lambda_i)$ is $Q \log \rho$ which is independent of n. Therefore,

$$C_t/T \geq \max_{M' \leq M} \begin{cases} (1-\frac{M'}{T})M' \log\rho & \text{if} M' \leq N \\ (1-\frac{M'}{T})N \log\rho & \text{if} M > N. \end{cases}$$

The expression $(1-\frac{M'}{T})M'$, is maximized by choosing $M' = T/2$ when $\min\{M,N\} \geq T/2$ and by choosing $M' = \min\{M,N\}$ when $\min\{M,N\} \leq T/2$. This means that the expression is maximized when $M' =$

$\min\{M, N, T/2\}$. This is a similar conclusion drawn in [12] and [19]. Also, the leading term in ρ for high SNR in the lower bound is given by

$$C_t/T \geq (1 - \frac{K}{T})K \log \rho$$

where $K = \min\{M, N, T/2\}$. This result suggests that the number of degrees of freedom available for communication is limited by the minimum of the number of transmit antennas, receive antennas and half the length of the coherence interval. Moreover, from the results in Section 3.4 we see that the lower bound is tight for the case when $M \leq N$ and large T in the sense that the leading term involving ρ in the lower bound is the same as the one in the expression for capacity.

4.6 Comparison of the training based lower bound (4.2) with the lower bound derived in Section 3.2

It is quite natural to use the lower bound to investigate training based techniques as the lower bound to the overall capacity of the system. Actually, using this "training" based lower bound it can be shown that the capacity as $T \to \infty$ converges to the capacity as if the receiver knows the channel. We will see how the new lower bound (3.3) derived in this work compares with this training based lower bound. The three figures (Figure 8, Figure 9 and Figure 10) below show that the new lower bound is indeed useful as it does better than the training based lower bound for $r = 0$. The plots are for $M = N = 1$ for different values of SNR.

However, we note that for $r = 1$ the training based lower bound and the lower bound derived in Section 3.2 agree perfectly with each other and are equal to the upper bound.

5 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.

A big part of the analysis e.g. the non-coherent capacity expression and training based lower bounds can be very easily extended to the non-standard Rician models considered in [9] and [11].

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,

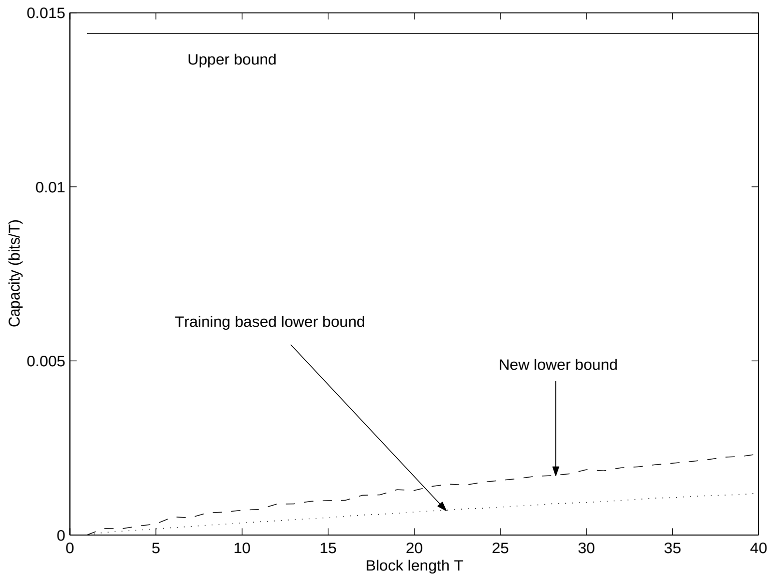

Figure 8: Comparison of the two lower bounds for $dB = -20$

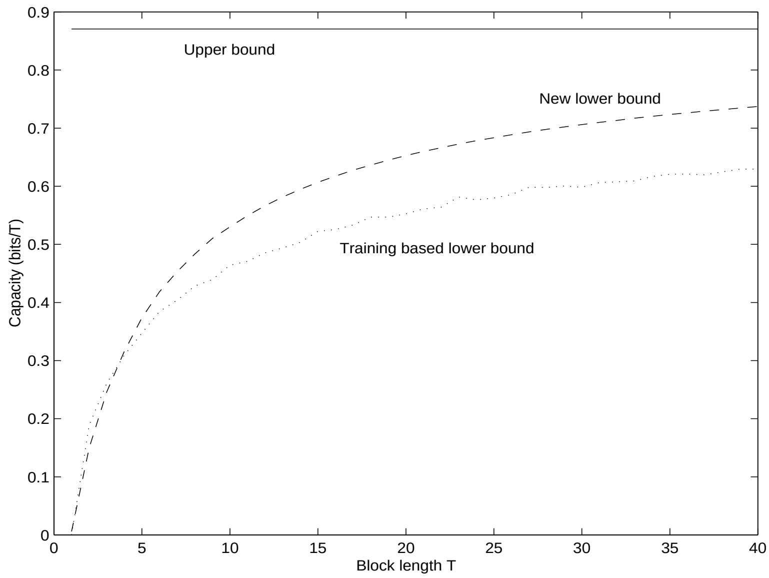

Figure 9: Comparison of the two lower bounds for $dB = 0$

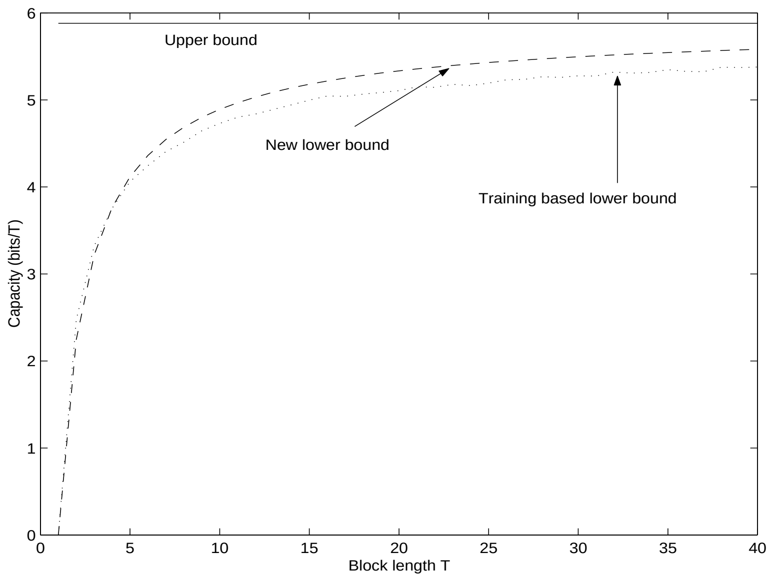

Figure 10: Comparison of the two lower bounds for $dB = 20$

one the standard codes designed for Rayleigh fading work for the case of Rician fading as well.

A lot more work needs to be done such as for case of $M > N$. We believe that more work along the lines of [19] is possible for the case of Rician fading. We conclude as in [19] that at least for the case $M = N$ the number of degrees of freedom is given by $M\frac{T-M}{T}$. The training based lower bound gives an indication that the number of degrees of freedom of a Rician channel is the same as that of a Rayleigh fading channel $\min\{M, N, T/2\}$ (derived in [19] and [12]). 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

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

Let l^r denote the real part of l and l^i the imaginary part. We can find the optimal value of d and l iteratively by using the method of steepest descent as follows

$$\begin{aligned} d_{k+1} &= d_k + \mu \frac{\partial C}{\partial d_k} \\ l^r_{k+1} &= l^r_k + \mu \frac{\partial C}{\partial l^r_k} \\ l^i_{k+1} &= l^i_k + \mu \frac{\partial C}{\partial l^i_k} \end{aligned}$$

where d_k, l^r_k and l^i_k are the values of d, l^r and l^i respectively 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

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

and similarly for l^r and l^i where

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

The derivative can be evaluated using monte carlo simulation.

B Non-coherent Capacity for low SNR values under Peak Power Constraint

In this section, we will use the notation introduced in Section 3.3. Here we concentrate on calculating the capacity under the constraint $\mathrm{tr}\{SS^\dagger\} \leq TM$.

Theorem 10 *Let the channel H be Rician (3.1) and the receiver have no knowledge of G. For fixed M, N and T under the peak power constraint*

$$C = rT\rho\lambda_{max}(H_m H_m^\dagger) + O(\rho^{3/2}).$$

Proof: First, Define $p(\tilde{X}) = E[p(\tilde{X}|\tilde{S})]$ where

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

Now

$$\mathcal{H}(\tilde{X}) = E_{\|\tilde{X}\|<(\frac{M}{\rho})^\gamma}[\log p(\tilde{X})] + E_{\|\tilde{X}\|\geq(\frac{M}{\rho})^\gamma}[\log p(\tilde{X})]$$

$E_{\|\tilde{X}\|\geq(\frac{M}{\rho})^\gamma}$ is defined by (3.2). Since $P(\|\tilde{X}\| \geq (\frac{M}{\rho})^\gamma) < O(e^{-(\frac{M}{\rho})^\gamma/TM})$ where we have chosen γ such that $1-2\gamma > 1/2$ or $\gamma < 1/4$. We have

$$\mathcal{H}(\tilde{X}) = E_{\|\tilde{X}\|<(\frac{M}{\rho})^\gamma}[\log p(\tilde{X})] + O(e^{-\frac{1}{TM}(\frac{M}{\rho})^\gamma}).$$

For $\|\tilde{X}\| < (\frac{M}{\rho})^\gamma$

$$\begin{aligned} p(\tilde{X}|\tilde{S}) &= \frac{1}{\pi^{TN}} e^{-\tilde{X}^\dagger\tilde{X}} \Big[1 + \sqrt{r\frac{\rho}{M}}(\tilde{X}^\dagger\hat{H}_m\tilde{S} + \tilde{S}^\dagger\hat{H}_m^\dagger\tilde{X}) - \\ & \frac{\rho}{M}\left(\mathrm{tr}\{(1-r)SS^\dagger \otimes I_N\} + \mathrm{tr}\{r\tilde{S}^\dagger\hat{H}_m^\dagger\hat{H}_m\tilde{S}\}\right) + \\ & (1-r)\frac{\rho}{M}\tilde{X}^\dagger SS^\dagger \otimes I_N\tilde{X} + r\frac{1}{2}\frac{\rho}{M}\left(\tilde{S}^\dagger\hat{H}_m^\dagger\tilde{X}\tilde{S}^\dagger\hat{H}_m^\dagger\tilde{X} + \tilde{S}^\dagger\hat{H}_m^\dagger\tilde{X}\tilde{X}^\dagger\hat{H}_m\tilde{S} + \right. \\ & \left. \tilde{X}^\dagger\hat{H}_m\tilde{S}\tilde{S}^\dagger\hat{H}_m^\dagger\tilde{X} + \tilde{X}^\dagger\hat{H}_m\tilde{S}\tilde{X}^\dagger\hat{H}_m\tilde{S}\right) + O(\rho^{3/2-3\gamma})\Big]. \end{aligned}$$

Since the capacity achieving signal has zero mean, for $\|\tilde{X}\| < (\frac{M}{\rho})^\gamma$

$$\begin{aligned} p(\tilde{X}) &= \frac{1}{\pi^{TN}} e^{-\tilde{X}^\dagger\tilde{X}} \Big[1 - \frac{\rho}{M}\left(\mathrm{tr}\{(1-r)E[SS^\dagger] \otimes I_N\} + \mathrm{tr}\{r\hat{H}_m E[\tilde{S}\tilde{S}^\dagger]\hat{H}_m^\dagger\}\right) + \\ & \frac{\rho}{M}((1-r)\tilde{X}^\dagger E[SS^\dagger] \otimes I_N\tilde{X} + r\tilde{X}^\dagger\hat{H}_m E[\tilde{S}\tilde{S}^\dagger]\hat{H}_m^\dagger\tilde{X} + O(\rho^{3/2-3\gamma})\Big] \\ &= \frac{1}{\pi^{TN}\det(\Lambda_{\tilde{X}})} e^{-\tilde{X}^\dagger\Lambda_{\tilde{X}}^{-1}\tilde{X}} + \frac{1}{\pi^{TN}} e^{-\tilde{X}^\dagger\tilde{X}}[O(\rho^{3/2-3\gamma})] \end{aligned}$$

where $\Lambda_{\tilde{X}} = I_{TN} + \frac{\rho}{M}(1-r)E[SS^\dagger] \otimes I_N + \frac{\rho}{M}r\hat{H}_m E[\tilde{S}\tilde{S}^\dagger]\hat{H}_m^\dagger$. Also,

$$\begin{aligned} \mathcal{H}(\tilde{X}) &= \log\det(I_{TN} + \frac{\rho}{M}(1-r)E[SS^\dagger] \otimes I_N + \frac{\rho}{M}r\hat{H}_m E[\tilde{S}\tilde{S}^\dagger]\hat{H}_m^\dagger) + O(\rho^{3/2-3\gamma}) \\ &= \frac{\rho}{M}\mathrm{tr}\{(1-r)E[SS^\dagger] \otimes I_N + r\hat{H}_m E[\tilde{S}\tilde{S}^\dagger]\hat{H}_m^\dagger\} + O(\rho^{3/2-3\gamma}). \end{aligned}$$

Since $P(\|S\|^2 > TM) = 0$ we can show $\mathcal{H}(\tilde{X}|\tilde{S}) = (1-r)\frac{\rho}{M}\text{tr}\{E[SS^\dagger] \otimes I_N\} + O(\rho^2)$. Since $0 < \gamma < 1/4$, $I(X;S) = r\frac{\rho}{M}\text{tr}\{\hat{H}_m E[\tilde{S}\tilde{S}^\dagger]\hat{H}_m^\dagger\} + O(\rho^{3/2})$. It is very clear that to maximize C we need to choose $E[\tilde{S}\tilde{S}^\dagger]$ in such a way that all the energy is concentrated in the direction of the maximum eigenvalues of $H_m H_m^\dagger$. So that we obtain, $C = r\frac{\rho}{M}\lambda_{max}(H_m H_m^\dagger)\text{tr}E[\tilde{S}\tilde{S}^\dagger] + O(\rho^{3/2})$. $\text{tr}E[\tilde{S}\tilde{S}^\dagger]$ is maximized by choosing $\text{tr}\{\tilde{S}\tilde{S}^\dagger\}$ to be the maximum possible which is TM. Therefore,

$$C = r\rho T\lambda_{max}(H_m H_m^\dagger) + O(\rho^{3/2}).$$

□

Corollary 2 *For purely Rayleigh fading channels* $\lim_{\rho\to 0} C/\rho = 0$

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 [19] 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 $T \geq N$ [19] 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 $X \to \Phi_X\Sigma_X\Psi_X^\dagger$ [19] 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,

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

Next define $C_1 = \Phi_1\Sigma_1\Psi_1^\dagger$ where $\Sigma_1 = \mathrm{diag}(\sigma_1,\ldots,\sigma_M)$, 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

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

Note that C_1 has bounded total power

$$
\mathrm{tr}\{E[C_1C_1^\dagger]\} = \mathrm{tr}\{E[\sigma_i^2]\} = \mathrm{tr}\{E[XX^\dagger]\} \leq NT(M+\sigma^2).
$$

Therefore, the differential entropy of C_1 is bounded by the entropy of a random matrix with entries iid

Gaussian distributed with variance $\frac{T(M+\sigma^2)}{M}$ [4, 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.

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

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

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

Therefore, we obtain

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

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

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

By Jensen's inequality

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

For γ it will be shown that

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

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

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

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 $X_1 = V H_1 + W$

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 [13]

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

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

where $k' = 1 + kE[\|Z_1 Z_1^\dagger\|_2]$ is a finite constant. Next, we have

$$
\begin{aligned}
\sum_{i=1}^{M} E[\log \sigma_i^2 | S] - \sum_{i=1}^{M} \log((1-r)\|s_i\|^2 + \sigma^2) &\leq E[\log\det Z_1 Z_1^\dagger] + \\
&\quad \sum_{i=1}^{M} \log \frac{(1-r)\|s_i\|^2 + k'\sigma^2}{(1-r)\|s_i\|^2 + \sigma^2} \\
&\leq E[\log\det Z_1 Z_1^\dagger] + k''
\end{aligned}
$$

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 $E[Z_1 Z_1^\dagger] \to \frac{1}{1-r} E[H_1 H_1^\dagger] = \frac{1}{1-r} E[HH^\dagger]$.

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

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

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,*

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

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 4 in Section 3.4.1. We need the following two theorems, proved in [10], for proving the results in this section.

Theorem 11 *Let $\{X_i \in \mathbb{C}^P\}$ be a sequence of continuous random variables with probability density functions, $\{f_i\}$ and $X \in \mathbb{C}^P$ 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 12 *Let $\{X_i \in \mathbb{C}^P\}$ be a sequence of continuous random variables with probability density functions, f_i and $X \in \mathbb{C}^P$ 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 4 and then use the result to 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 $f(X|S)$.

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

then

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

Now

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

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,

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

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 $\to \infty$, 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 $\lim_{\sigma^2 \to 0} f(X)$ 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 11. 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 $T = M = N$. $\square$

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 12 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, $\mathcal{H}(X) \approx \mathcal{H}(SH)$.

E Proof of Theorem 7 in Section 4.1

First we note that $\sigma^2_{\hat{G}} = 1 - \sigma^2_{\tilde{G}}$. This means that

$$\rho_{eff} = \frac{\kappa T\rho + T_c}{(1-r)\kappa T\rho\sigma^2_{\tilde{G}} + T_c} - 1.$$

Therefore, to maximize ρ_{eff} we just need to minimize $\sigma^2_{\tilde{G}}$. Now,

$$\sigma^2_{\tilde{G}} = \frac{1}{NM}\mathrm{tr}\{E[\tilde{\tilde{G}}\tilde{\tilde{G}}^\dagger]\}$$

where

$$E[\tilde{\tilde{G}}\tilde{\tilde{G}}^\dagger] = (I_M + (1-r)\frac{\rho}{M}S_t^\dagger S_t)^{-1} \otimes I_N$$

where $\rho = \frac{M}{\sigma^2}$. Therefore, the problem is the following

$$\min_{S_t:\mathrm{tr}\{S_t^\dagger S_t\}\leq(1-\kappa)TM} \frac{1}{M}\mathrm{tr}\{\left(I_M + (1-r)\frac{\rho}{M}S_t^\dagger S_t\right)^{-1}\}.$$

The problem above can be restated as

$$\min_{\lambda_1,\ldots,\lambda_M:\sum\lambda_m\leq(1-\kappa)TM} \frac{1}{M}\sum_{m=1}^{M}\frac{1}{1+(1-r)\frac{\rho}{M}\lambda_m}$$

where λ_m, $m = 1,\ldots,M$ are the eigenvalues of $S_t^\dagger S_t$. The solution to the above problem is $\lambda_1 = \ldots = \lambda_M = (1-\kappa)T$. Therefore, the optimum S_t satisfies $S_t^\dagger S_t = (1-\kappa)TI_M$.

This gives $\sigma^2_{\tilde{G}} = \frac{1}{1+(1-r)\frac{\rho}{M}(1-\kappa)T}$. Also, for this choice of S_t we obtain the elements of $\hat{G}$ to be zero mean independent with Gaussian distribution. This gives

$$\rho_{eff} = \frac{\kappa T\rho[Mr + \rho(1-\kappa)T]}{T_c(M+\rho(1-\kappa)T) + (1-r)\kappa T\rho M}.$$

F Proof of Theorem 8 in Section 4.1

First, from Theorem 7

$$\begin{aligned}\rho_{eff} &= \frac{\kappa T\rho[Mr + \rho(1-\kappa)T]}{T_c(M+\rho(1-\kappa)T) + (1-r)\kappa T\rho M} \\ &= \frac{T\rho}{T_c - (1-r)M}\frac{(1-\kappa)\kappa + \kappa\frac{rM}{T\rho}}{\frac{MT_c + T\rho T_c}{T\rho[T_c-(1-r)M]} - \kappa} \qquad T_c \neq (1-r)M\end{aligned}$$

$$= \frac{T^2\rho^2}{T_c(M+T\rho)}[(1-\kappa)\kappa + \kappa\frac{rM}{T\rho}] \qquad T_c = (1-r)M.$$

Consider the following three cases for the maximization of ρ_{eff} over $0 \leq \kappa \leq 1$.

Case 1. $T_c = (1-r)M$:

We need to maximize $(1-\kappa)\kappa + \kappa\frac{rM}{T\rho}$ over $0 \leq \kappa < 1$. The maximum occurs at $\kappa = \kappa_0 = \min\{\frac{1}{2} + \frac{rM}{2T\rho}, 1\}$. In this case

$$\rho_{eff} = \frac{T^2\rho^2}{(1-r)M(M+T\rho)}[\kappa_0\frac{rM}{T\rho} + \kappa_0(1-\kappa_0)].$$

Case 2. $T_c > (1-r)M$:

In this case,

$$\rho_{eff} = \frac{T\rho}{T_c-(1-r)M}\frac{(1-\kappa)\kappa+\kappa\eta}{\gamma-\kappa}$$

where $\eta = \frac{rM}{T\rho}$ and $\gamma = \frac{MT_c+T\rho T_c}{T\rho[T_c-(1-r)M]} > 1$. We need to maximize $\frac{(1-\kappa)\kappa+\kappa\eta}{\gamma-\kappa}$ over $0 \leq \kappa \leq 1$ which occurs at $\kappa = \min\{\gamma - \sqrt{\gamma^2-\gamma-\eta\gamma}, 1\}$. Therefore,

$$\rho_{eff} = \frac{T\rho}{T_c-(1-r)M}(\sqrt{\gamma} - \sqrt{\gamma-1-\eta})^2$$

when $\kappa < 1$. When $\kappa = 1$ we obtain $T_c = T$. Substituting $\kappa = 1$ in the expression for ρ_{eff}

$$\rho_{eff} = \frac{\kappa T\rho[Mr+\rho(1-\kappa)T]}{T_c(M+\rho(1-\kappa)T)+(1-r)\kappa T\rho M}$$

we obtain $\rho_{eff} = \frac{rT\rho}{T+(1-r)T\rho}$.

Case 3. $T_c < (1-r)M$:

In this case,

$$\rho_{eff} = \frac{T\rho}{(1-r)M-T_c}\frac{(1-\kappa)\kappa+\kappa\eta}{\kappa-\gamma}$$

where $\gamma = \frac{MT_c+T\rho T_c}{T\rho[T_c-(1-r)M]} < 0$. Maximizing $\frac{(1-\kappa)\kappa+\kappa\eta}{\gamma-\kappa}$ over $0 \leq \kappa \leq 1$ we obtain $\kappa = \min\{\gamma + \sqrt{\gamma^2-\gamma-\gamma\eta}, 1\}$. Therefore, when $\kappa < 1$

$$\rho_{eff} = \frac{T\rho}{T_c-(1-r)M}(\sqrt{-\gamma} - \sqrt{-\gamma+1+\eta})^2$$

Similar to the case $T_c < (1-r)M$, when $\kappa = 1$ we obtain $T_c = T$ and $\rho_{eff} = \frac{rT\rho}{T+(1-r)T\rho}$.

G Proof of Theorem 9 in Section 4.1

Note that optimization over T_c makes sense only when $\kappa < 1$. If $\kappa = 1$ then T_c obviously has to be set equal to T. First, we examine the case $T_c > (1-r)M$. The other two cases are similar. Let $Q = \min\{M, N\}$ and let λ_i denote the i^{th} non-zero eigenvalue of $\frac{H_1 H_1^\dagger}{M}$, $i = 1, \ldots, Q$. Then we have

$$C_t \geq \sum_{i=1}^{Q} \frac{T_c}{T} E \log(1 + \rho_{eff} \lambda_i).$$

Let C_l denote the RHS in the expression above. The idea is to maximize C_l as a function of T_c. We have

$$\frac{dC_l}{dT_c} = \sum_{i=1}^{Q} \left\{ \frac{1}{T} E \log(1 + \rho_{eff}\lambda_i) + \frac{T_c}{T}\frac{d\rho_{eff}}{dT_c} E\left[\frac{\lambda_i}{1 + \rho_{eff}\lambda_i}\right]\right\}.$$

Now, ρ_{eff} for $T_c > (1-r)M$ is given by

$$\rho_{eff} = \frac{T\rho}{T_c - (1-r)M}(\sqrt{\gamma} - \sqrt{\gamma - 1 - \eta})^2$$

where $\gamma = \frac{MT_c + T\rho T_c}{T\rho[T_c - (1-r)M]}$ and $\eta = \frac{rM}{T\rho}$. It can be easily verified that

$$\frac{d\rho_{eff}}{dT_c} = \frac{T\rho(\sqrt{\gamma} - \sqrt{\gamma - 1 - \eta})^2}{[T_c - (1-r)M]^2}\left[\sqrt{\frac{(1-r)M(M + T\rho)}{T_c(T_c + T\rho + rM)}} - 1\right].$$

Therefore,

$$\begin{aligned}\frac{dC_l}{dT_c} &= \frac{1}{T}\sum_{i=1}^{Q} E\Bigg[\log(1 + \rho_{eff}\lambda_i) - \\ &\qquad \frac{\rho_{eff}\lambda_i}{1 + \rho_{eff}\lambda_i}\frac{T_c}{T_c - (1-r)M}\left[1 - \sqrt{\frac{(1-r)M(M + T\rho)}{T_c(T_c + T\rho + rM)}}\right]\Bigg].\end{aligned}$$

Since, $\frac{T_c}{T_c - (1-r)M}\left[1 - \sqrt{\frac{(1-r)M(M+T\rho)}{T_c(T_c + T\rho + rM)}}\right] < 1$ and $\log(1+x) - x/(1+x) \geq 0$ for all $x \geq 0$ we have $\frac{dC_l}{dT_c} > 0$. Therefore, we need to increase T_c as much as possible to maximize C_l or $T_c = T - M$.

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] I. C. Abou-Faycal, M. D. Trott, and S. S. (Shitz), "The capacity of discrete-time memoryless Rayleigh fading channels," *To Appear in IEEE Transactions on Information Theory.*

[3] 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.

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

[5] 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.

[6] F. R. Farrokhi, G. J. Foschini, A. Lozano, and R. A. Valenzuela, "Link-optimal space-time processing with multiple transmit and receive antennas," *To appear in IEEE Communication Letters.*

[7] 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.

[8] 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.

[9] 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 Information Theory*, 2001.

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

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

[12] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?," *Bell Labs Technical Memorandum*, 2000.

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

[14] 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.

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

[16] 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.

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

[18] S. Verdú, "On channel capacity per unit cost," *IEEE Trans. on Inform. Theory*, vol. 36, no. 5, pp. 1019–1030, September 1990.

[19] L. Zheng and D. N. C. Tse, "Packing spheres in the Grassmann manifold: A geometric approach to the non-coherent multi-antenna channel," *Submitted to IEEE Trans. on Information Theory.*