

# Training in Multiple-Antenna Rician Fading Wireless Channels with Deterministic Specular Component

Mahesh Godavarti, *Member, IEEE*, and Alfred O. Hero-III, *Fellow, IEEE*

**Abstract**—We determine the optimum training strategy for a multiple-antenna wireless link in a Rician fading channel using a training based lower bound on capacity. We consider the standard Rician block fading channel where the channel coefficients are modeled as independent circular Gaussian random variables with non-zero means (the specular component). The specular component is known to both the transmitter and receiver. The channel coefficients of this model are constant over a block of  $T$  symbol periods but, independent over different blocks. For such a model, it is shown that the training based capacity, the optimum training signals, the training period, transmit and training energy are dependent on the Rician factor  $r$  along with SNR  $\rho$ , the number of transmit antennas  $M$ , the number of receive antennas  $N$  and the coherence interval  $T$ . Also, unlike in the case of Rayleigh fading channels, it can be shown using the lower bound for Rician fading channels that for low SNR Rician fading channels behave like a purely AWGN channel and the optimum strategy is to spend no effort in learning the channel. When SNR is not low and training is required then the optimum training period is as many symbol intervals as there are transmit antennas.

**Index Terms**—Capacity, asymptotic capacity, training, information theory, Rician fading, multiple antennas.

## I. INTRODUCTION

DEPLOYING multiple antennas at the transmitter and receiver has been demonstrated to be a viable solution to the demand for high data rate in wireless communications [6], [7], [16], [20], [22]. A straightforward way for the transmitter and the receiver to transmit data would be for the receiver to first learn the channel and then use the channel estimate to decode the transmitted symbols. Such training methods are prevalent in wireless communication systems like IS-95 CDMA and GSM and have been investigated in [13], [14], [17], [22].

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 optimum training signal is like. Hassibi and Hochwald [14] have addressed these

issues for the popular case of Rayleigh block fading channels. They used a capacity lower bound based on MMSE channel estimate to find the optimum training strategy by maximizing the lower bound. They showed that 1) pilot symbol training based communication schemes are highly suboptimum for low SNR but practically optimum for high SNR; 2) when practical the optimum amount of time<sup>1</sup> devoted to training is equal to  $M$  symbol intervals, where  $M$  is the number of transmitters, when the fraction of power devoted to training is allowed to vary and 3) the orthonormal signal is the optimum signal for training.

In [13], the authors investigated the same problem for a more general fading model and a more general training strategy using a generalized mutual information lower bound based on Gaussian codebooks with modified nearest neighbor decoding. Also, unlike in [14], they relaxed the assumption of identity transmit signal covariance matrix and included the covariance matrix to be one of the parameters to be optimized. The authors showed that for the special case of piecewise block fading channels with Gaussian fading and additive Gaussian noise the training based scheme as in [14] with minimum mean-squared error channel estimator is optimum in the sense that it maximizes the generalized mutual information lower bound. They showed that for low SNR the transmit signal covariance matrix has only one non-zero eigenvalue. For high SNR, the results agree with the assumption on transmit signal covariance matrix made in [14]. In this case, as expected, for Rayleigh block fading the optimum lower bound based on generalized mutual information is equal to the lower bound derived in [14].

However, 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 [1], [4], [5], [18], [19]. 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.

In this paper, we investigate how much training is necessary for a wireless link operating in a Rician fading channel under the average energy constraint on the input signal. We use the standard Rician fading channel throughout the paper, that is, we assume that the specular component is deterministic,

Manuscript received DATE; revised DATE; accepted DATE. The associate editor coordinating the review of this paper and approving it for publication was NAME. Parts of this work were presented at ISIT 2002 held in Lausanne, Switzerland.

M. Godavarti is with Ditech Networks, Inc., 825 E. Middlefield Rd., Mountain View, CA 94043 (email: mgodavarti@ditechnetworks.com, mgodavar@hotmail.com). This work was performed while M. Godavarti was a Ph.D. candidate at the University of Michigan.

A. Hero is with the Dept. of Electrical Engineering, University of Michigan, 1301 Beal Ave., Ann Arbor, MI 48107 (email: hero@eecs.umich.edu).

Digital Object Identifier 10.1109/TWC.2006.04579.

<sup>1</sup>Note: time is always measured in terms of number of symbol intervals. This is tied to the initial assumption that channel coherence interval (the amount of time channel coefficients remain constant) is itself measured in terms of number of symbol intervals.

of general rank and known to both the transmitter and the receiver. The Rayleigh component is never known to the transmitter. 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*.

Keeping in mind the results obtained in [13], we use the same training signal based approach as that of [14]. However, we relax the assumption of identity transmit signal covariance matrix and we leave the matrix to be one of the parameters to be optimized. We draw similar conclusions about training for non-coherent communications as in [14] with one big difference in the regime of low SNR. For Rayleigh fading channels the optimum training period is not zero for all values of SNR. However, for Rician fading channels there exists a threshold dependent on the Rician factor  $r$  such that for all SNRs below the threshold the optimum strategy is to have no training at all. An interesting find regarding the optimum transmit strategy, in this paper, is that the optimum strategy for low SNR is to concentrate all the available energy in the direction of strongest specular component whereas for high SNR it is to spread the energy equally in all directions. Note that this finding, for high SNR, is the same as that of [13] and [14] because the transmit signal covariance matrix in this region is an identity matrix. However, for low SNR even though the optimum transmit covariance matrix consists of a single non-zero eigenvalue as in [13] the eigenvector corresponding to this eigenvalue can not be arbitrary. The eigenvector should point in the same direction as the strongest specular component.

The training based lower bound derived in [14] and adapted here for Rician fading channels is suboptimum for low SNR as the capacity for low SNR for block fading channels is a linear function of SNR [21]. The training based lower bound for Rayleigh fading channels turns out to be a quadratic function of SNR [14]. From this paper we find that for Rician fading channels it is a linear function of SNR. However, the bound is still suboptimum since the constant multiplying SNR in the capacity expression is purely a function of the specular component instead of the whole channel. That is, if  $\rho$  denotes SNR,  $\lambda_{max}(A)$  the largest eigenvalue of matrix  $A$ ,  $H$  the Rician fading channel and  $H_m$  the specular component of  $H$  then for low SNR the capacity of block fading Rician channel behaves as  $\rho\lambda_{max}(E[HH^\dagger])$  [12] whereas the training based lower bound on capacity of block fading Rician channel behaves as  $\rho\lambda_{max}(E[H_mH_m^\dagger]) = \rho\lambda_{max}(H_mH_m^\dagger)$ . For high SNRs and large coherent periods, the ratio of the training bound and the actual capacity tends to one and indicates that training based schemes can achieve rates close to capacity.

This paper is organized as follows. First, in Section II the model used in the paper is established and a simple training based lower bound is discussed. Then in Section III, a more general training based lower bound on capacity is established and its optimization is performed over various parameters (choosing the parameters that maximize the lower bound). More precisely, in Sub-section III-A, optimization is performed for training over the transmitting signal, energy distribution and the training period. This is followed by

optimization over the same parameters under the constraint of equal training and transmit signal powers in Sub-section III-B. Additional insights into the optimization problem are obtained from numerical simulation in Section IV. Then in Section V, optimization is performed in the regimes of low and high SNR followed finally by Section VI in which the results of this paper are summarized.

## II. SIGNAL MODEL AND A SIMPLE TRAINING BASED LOWER BOUND

We adopt the following model, which is the same as the one used in [12], for the Rician fading channel:

$$X = SH + W \quad (1)$$

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. Note that  $T$  denotes the coherence interval,  $M$  denotes the number of transmit antennas and  $N$  the number of receive antennas. The Rician fading channel  $H$  is defined as

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

where  $H_m$  is the deterministic 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  $I_{MN}$  and  $\sigma^2 I_{TN}$ , respectively.  $H_m$  is a deterministic matrix satisfying  $\text{tr}\{H_mH_m^\dagger\} = MN$ .  $G$  satisfies  $E[\text{tr}\{GG^\dagger\}] = MN$  and  $r$  is the Rician parameter between 0 and 1 so that  $E[\text{tr}\{HH^\dagger\}] = MN$ . We assume that both the transmitter and receiver have complete knowledge of the probability density function of  $H$ . This means that  $H_m$  and the probability density function of  $G$  are known to both the transmitter and receiver. We also assume that communication is taking place under the average energy constraint on the input signal given by  $E[\text{tr}\{SS^\dagger\}] \leq MT$ . The SNR of the channel, given by the ratio of the energy of the elements of  $SH$  to the energy of the elements of  $W$ , is therefore  $\rho = \frac{M}{\sigma^2}$  when the constraint is satisfied with equality.

### A. A Simple Training Based Lower Bound

In [22] the authors demonstrated a very simple training method that achieves the optimum 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<sup>2</sup> using scalar minimum mean squared error

<sup>2</sup>It might be useful to remind the readers at this point that  $H_m$  is already known to both the transmitter and receiver

(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  $\sigma^2$  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 suboptimum 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 [2, p. 263], [14, Theorem 1]. In this section we deal with normalized capacity  $C/T$  instead of capacity  $C$ . The lower bound on the normalized capacity is given by

$$C/T \geq \frac{T - T_t}{T} E \log \det \left( I_M + \frac{\rho_{eff}}{M} H_1 H_1^\dagger \right) \quad (2)$$

where  $T_t$  is number of symbol intervals devoted to training and  $\rho_{eff}$  in the expression above is the effective SNR at the output (explained at the end of the next paragraph) given as follows:

$$\rho_{eff} = \frac{\rho[r + r(1-r)\rho + (1-r)^2\rho]}{[1 + 2(1-r)\rho]} \quad (3)$$

where  $\rho = \frac{M}{\sigma^2}$  is the actual SNR.  $H_1$  in (2) is a Rician channel with a new Rician parameter  $r_{eff}$  where

$$r_{eff} = \frac{r}{r + (1-r)\frac{(1-r)M}{(1-r)M + \sigma^2}}. \quad (4)$$

Note that  $r_{eff} > r$  in the effective channel because part of the energy from the unknown Rayleigh component has been diverted to the additive noise in the new effective channel model.

The lower bound in (2) can be easily calculated because the lower bound is essentially the coherent capacity with  $H$  replaced by  $\sqrt{r_{eff}}H_m + \sqrt{1-r_{eff}}\hat{G}$ . The signal covariance structure was chosen to be an identity matrix as this is the optimum covariance matrix for high SNR (refer to Result 2 in the following section). 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

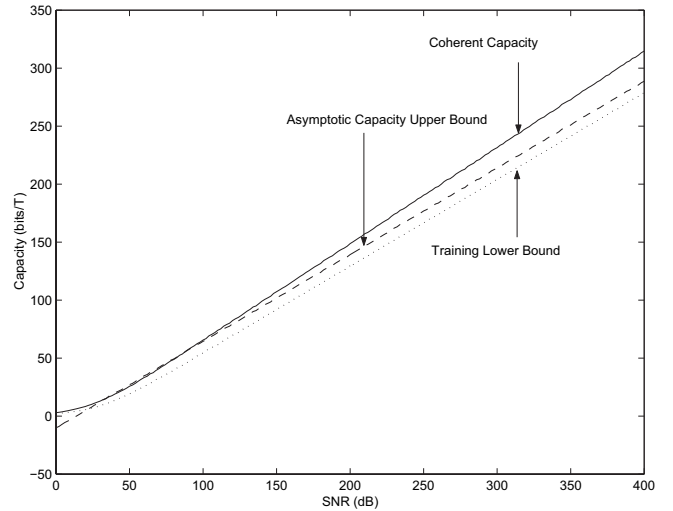


Fig. 1. Asymptotic capacity, capacity upper and lower bounds for different values of SNR.

given by  $\sigma^2 + \frac{(1-r)M\sigma^2}{(1-r)M + \sigma^2}$ . Therefore, the effective SNR,  $\rho_{eff}$  is given as in (3). 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 following asymptotic upper bound on non-coherent capacity as  $\rho \rightarrow \infty$  [12]

$$\begin{aligned} C \leq \log |G(T, M)| + (T - M)E[\log \det H^\dagger H] \quad (5) \\ + M(T - M) \log \frac{T\rho}{M\pi e} - M^2 \log(1 - r). \end{aligned}$$

where  $|G(T, M)|$  is the volume of the Grassmann manifold [22] and is equal to

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

As can be seen the lower bound and the asymptotic upper bound agree well with each other for large SNR values. Figure 1 demonstrates this for  $M = N = 5$ ,  $r = 0.9$  and  $T = 50$ . The specular component in Figure 1 is a rank-one specular component given by

$$H_m = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} [1 \ 0 \ \dots \ 0].$$

### III. TRAINING BASED LOWER BOUND AND OPTIMIZATION

In this section, instead of fixing the training signal and the amount of training as done in the previous section we optimize over these parameters using the techniques in [14]. In [14], the authors use the optimization of the lower bound on capacity to find the optimum allocation of training as compared to communication. Let  $T_t$  denote the amount of time, in terms of number of symbol intervals, 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”  $\kappa$  denote the fraction of the energy used for communication. Then  $T = T_t + T_c$  and  $\text{tr}\{S_t S_t^\dagger\} = (1 - \kappa)TM$  and  $\text{tr}\{S_c S_c^\dagger\} = \kappa TM$ .

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

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 optimum estimate in the MSE sense 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 [14, Theorem 1].

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

$$\begin{aligned} \sigma_{\hat{w}_c}^2 &= \sigma^2 + \frac{1-r}{NT_c} \text{tr}\{E[\bar{G}\bar{G}^\dagger]\kappa T I_M\} \\ &= \sigma^2 + \frac{(1-r)\kappa TM}{T_c} \frac{1}{NM} \text{tr}\{E[\bar{G}\bar{G}^\dagger]\} \\ &= \sigma^2 + \frac{(1-r)\kappa TM}{T_c} \sigma_{\bar{G}}^2 \end{aligned} \quad (6)$$

where  $\sigma_{\bar{G}}^2$  is given in (33) and the lower bound is

$$C_t/T \geq \frac{T - T_t}{T} E \log \det \left( I_M + \frac{\rho_{eff}}{M} H_1 \Lambda H_1^\dagger \right), \quad (7)$$

where the “post training SNR”  $\rho_{eff}$ , is the ratio of the energies in the elements of  $S_c \hat{H}$  and energies in the elements of  $\hat{W}_c$  and  $H_1 = \sqrt{r_{eff}}H_m + \sqrt{1-r_{eff}}\hat{G}$  where  $r_{eff} = \frac{r}{r+(1-r)\sigma_{\bar{G}}^2}$ . To calculate  $\rho_{eff}$ , the energy in the elements of  $S_c \hat{H}$  is given by

$$\begin{aligned} \sigma_{S_c \hat{H}}^2 &= \frac{1}{NT_c} [r \text{tr}\{H_m H_m^\dagger \kappa T I_M\} + \\ &\quad (1-r) \text{tr}\{\hat{G} \hat{G}^\dagger \kappa T I_M\}] \\ &= \frac{\kappa TM}{T_c} \frac{1}{NM} [r NM + (1-r) \text{tr}\{\hat{G} \hat{G}^\dagger\}] \\ &= \frac{\kappa TM}{T_c} [r + (1-r)\sigma_{\bar{G}}^2], \end{aligned}$$

which gives us

$$\rho_{eff} = \frac{\kappa T \rho [r + (1-r)\sigma_{\bar{G}}^2]}{T_c + (1-r)\kappa T \rho \sigma_{\bar{G}}^2}. \quad (8)$$

$\Lambda$  in (7) is the optimum signal covariance matrix the structure of which is determined from the following results obtained from [12].

*Result 1:* [12, Proposition 1] Let  $H$  be Rician (1) and let the receiver have complete knowledge of the Rayleigh component  $G$ . For low SNR, the coherent capacity  $C_H$ , is attained by the same signal covariance matrix that attains capacity when  $r = 1$  and

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

*Result 2:* [12, Theorem 2] Let  $H$  be Rician (1) then as  $\rho \rightarrow \infty$ , the coherent capacity  $C_H$ , is attained by an identity signal covariance matrix and

$$\frac{C_H}{T \cdot E \log \det \left[ \frac{\rho}{M} H H^\dagger \right]} \rightarrow 1.$$

Therefore, For low SNR,  $\Lambda$  has only one non-zero eigenvalue such that all energy is concentrated in the direction of the largest eigenvalue of  $H_m$  and for high SNR  $\Lambda$  is an identity matrix.

#### A. Optimization of $S_t$ , $\kappa$ and $T_t$

We will optimize  $S_t$ ,  $\kappa$  and  $T_t$  to maximize the lower bound (7).

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

*Proposition 1:* The signal  $S_t$  that maximizes  $\rho_{eff}$  satisfies the following condition

$$S_t^\dagger S_t = (1 - \kappa) T I_M \quad (9)$$

and the corresponding  $\rho_{eff}$  is

$$\rho_{eff}^* = \frac{\kappa T \rho [Mr + \rho(1-r)(1-\kappa)T]}{T_c(M + \rho(1-r)(1-\kappa)T) + (1-r)\kappa T \rho M}. \quad (10)$$

*Proof:* Refer to Appendix I.

The optimum signal derived above is the same as the optimum signal derived in [14] and the corresponding capacity lower bound using the  $S_t$  obtained above is given by (7) but where  $\rho_{eff}$  is as given below

$$\rho_{eff} = \frac{\kappa T \rho [Mr + \rho(1-r)(1-\kappa)T]}{T_c(M + \rho(1-r)(1-\kappa)T) + (1-r)\kappa T \rho M} \quad (11)$$

and  $H_1 = \sqrt{r_{eff}}H_m + \sqrt{1-r_{eff}}G$  where  $r_{eff} = \frac{r + (1-r)(1-\kappa)\frac{\rho}{M}T}{r + (1-r)(1-\kappa)\frac{\rho}{M}T}$  and as before  $G$  is a matrix consisting of i.i.d. Gaussian circular random variables with mean zero and unit variance. Now,  $\Lambda$  is the covariance matrix of the source  $S_c$  when the channel is Rician and known to the receiver. Therefore from Results 1 and 2,  $\Lambda$  is an identity matrix for  $\rho_{eff} \rightarrow \infty$  and is a diagonal matrix with only one non-zero diagonal element for  $\rho_{eff} \rightarrow 0$ .

Optimization of (7) over the energy allocation factor  $\kappa$ , is straightforward as  $\kappa$  affects the lower bound only through the post training SNR  $\rho_{eff}$ , and can be stated as the following proposition.

*Proposition 2:* For given  $T_t$  and  $T_c$  the optimum power allocation  $\kappa$  in a training based scheme is given by

$$\kappa = \begin{cases} \min\{\gamma - \sqrt{\gamma(\gamma - 1 - \eta)}, 1\} & \text{for } T_c > (1-r)M \\ \min\{\frac{1}{2} + \frac{rM}{2T\rho}, 1\} & \text{for } T_c = (1-r)M \\ \min\{\gamma + \sqrt{\gamma(\gamma - 1 - \eta)}, 1\} & \text{for } T_c < (1-r)M \end{cases} \quad (12)$$

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

$$C_t/T \geq \frac{T - T_t}{T} E \log \det \left( I_M + \frac{\rho_{eff}}{M} H_1 \Lambda H_1^\dagger \right) \quad (13)$$

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

$$\rho_{eff} = \begin{cases} \frac{\frac{T\rho}{T_c - (1-r)M} (\sqrt{\gamma} - \sqrt{\gamma - 1 - \eta})^2}{\text{when } \kappa = \gamma - \sqrt{\gamma(\gamma - 1 - \eta)}} & (14) \\ \frac{r\rho}{1 + (1-r)\rho} & \text{when } \kappa = 1 \end{cases}$$

for  $T_c = (1-r)M$ :

$$\rho_{eff} = \begin{cases} \frac{\frac{T^2\rho^2}{4(1-r)M(M+T\rho)} (1 + \frac{rM}{T\rho})^2}{\text{when } \kappa = \frac{1}{2} + \frac{rM}{2T\rho}} & (15) \\ \frac{rT\rho}{(1-r)(M+T\rho)} & \text{when } \kappa = 1 \end{cases}$$

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

$$\rho_{eff} = \begin{cases} \frac{\frac{T\rho}{(1-r)M - T_c} (\sqrt{-\gamma} - \sqrt{-\gamma + 1 + \eta})^2}{\text{when } \kappa = \gamma + \sqrt{\gamma(\gamma - 1 - \eta)}} & (16) \\ \frac{r\rho}{1 + (1-r)\rho} & \text{when } \kappa = 1 \end{cases}$$

and  $r_{eff}$  is given by substituting the appropriate value of  $\kappa$  in the expression

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

*Proof:* Refer to Appendix II.

We see from Proposition 2 that for  $\rho < \frac{rM}{T}$  the optimum setting for  $\kappa$  is  $\kappa = 1$ . That is,  $\frac{rM}{T}$  is the threshold such that for all SNRs below it the optimum strategy is to have no training at all. Note that even though the threshold increases with  $r$  it is not tight as can be gauged from the case  $r = 1$ . For  $r = 1$ , the threshold should actually be  $\infty$  as it is obvious that  $\kappa = 1$  for all  $\rho$  whereas the threshold turns out to be only  $\frac{M}{T}$ .

For optimization over  $T_t$  we draw similar conclusions as in [14]. In [14] the optimum setting for  $T_t$  was shown to be  $T_t = M$  for all values of SNR. However, in this paper for  $\rho < \frac{rM}{T}$  the optimum setting is  $T_t = 0$ . The argument follows from the fact that for these values of  $\rho$  we have  $\kappa = 1$  i.e., all energy is allocated to communications. It is clear that optimization of  $T_t$  makes sense only when  $\kappa$  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 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  $\rho_{eff}$  [14]. Therefore, it makes sense to make  $T_t$  as small as possible. Therefore, 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.

*Proposition 3:* The optimum length of the training interval is  $T_t = M$  whenever  $\kappa < 1$  for all values of  $\rho$  and  $T > M$ , and the capacity lower bound is

$$C_t/T \geq \frac{T - M}{T} E \log \det \left( I_M + \frac{\rho_{eff}}{M} H_1 \Lambda H_1^\dagger \right) \quad (17)$$

where

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

The optimum power allocations are easily obtained from Proposition 2 by simply setting  $T_c = T - M$ .

*Proof:* Refer to Appendix III.

### B. Equal Training and Data Power

As stated in [14], 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

$$C_t/T \geq \frac{T - T_t}{T} E \log \det \left( I_M + \frac{\rho_{eff}}{M} H_1 \Lambda H_1^\dagger \right) \quad (18)$$

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

We can derive the optimum training period using the same procedure as in the proof of Proposition 3. Using the definition of  $C_l$  in Proposition 3, consider the following

$$\frac{dC_l}{dT_c} = \sum_{i=1}^Q \left\{ \frac{1}{T} E \log(1 + \rho_{eff} \lambda_i) + \frac{T_c}{T} \frac{d\rho_{eff}}{dT_c} E \left[ \frac{\lambda_i}{1 + \rho_{eff} \lambda_i} \right] \right\}. \quad (19)$$

However, since

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

we have

$$\frac{d\rho_{eff}}{dT_c} = - \frac{\frac{\rho^2}{M} (1-r)(1+\rho)}{[1 + \rho(\frac{T - T_c}{M} + (1-r))]^2}. \quad (20)$$

Therefore,

$$\frac{dC_l}{dT_c} = \frac{1}{T} \sum_{i=1}^Q E \left[ \log(1 + \rho_{eff} \lambda_i) - \left\{ \frac{\rho_{eff} \lambda_i}{1 + \rho_{eff} \lambda_i} \times \frac{(1+\rho) \frac{\rho}{M} (1-r)}{(r + \rho \frac{T - T_c}{M}) [1 + \rho(\frac{T - T_c}{M} + (1-r))]} \right\} \right]. \quad (21)$$

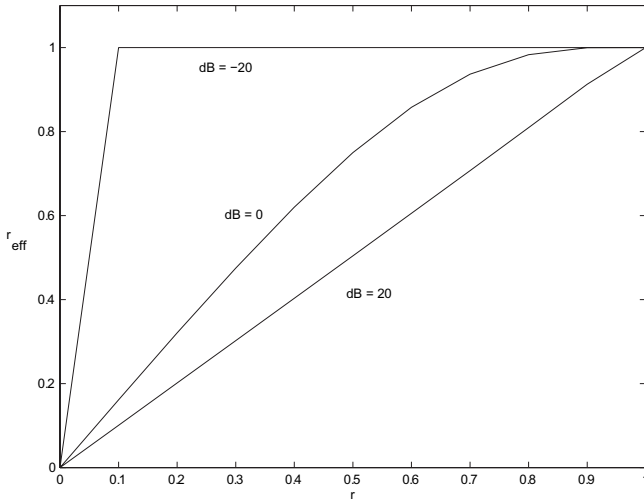


Fig. 2. Plot of  $r_{eff}$  as a function of Rician parameter  $r$ .

Note that for large  $\rho$ ,

$$\frac{(1 + \rho) \frac{\rho}{M} (1 - r)}{\left(r + \rho \frac{T - T_c}{M}\right) \left[1 + \rho \left(\frac{T - T_c}{M} + (1 - r)\right)\right]} < 1 \quad (22)$$

if  $T - T_c \geq M$ . This means that  $\frac{dC_l}{dT_c} > 0$  and  $C_l$  is maximized by choosing  $T_c = M$ . And if  $\rho$  is made small enough,

$$\frac{(1 + \rho) \frac{\rho}{M} (1 - r)}{\left(r + \rho \frac{T - T_c}{M}\right) \left[1 + \rho \left(\frac{T - T_c}{M} + (1 - r)\right)\right]} < 1 \quad (23)$$

even for  $T_c = T$ . Therefore, for low SNR  $C_l$  is maximized by choosing  $T_c = T$ .

In this section we have seen that the optimum training period for low SNR is zero. This can be contrasted against the result for Rayleigh fading channels [14] where the optimum training period is equal to  $T/2$  for small  $\rho$ . But for large  $\rho$  the optimum training period here is  $M$  like in [14].

#### IV. 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 2 and 3 show  $r_{eff}$  and  $\kappa$  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 2 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 3 shows that as the SNR goes to zero less and less power is allocated for training. This agrees with the plot in Figure 2.

In Figure 4 we show the trend followed by the training and communication powers, in terms of the corresponding SNRs given by  $\frac{(1-\kappa)\rho T}{T_t}$  and  $\frac{\kappa\rho T}{T_c}$  respectively, as a function of block length  $T$ . The plots were generated for different values of  $r$  with fixed  $T_t = M$ ,  $M = N = 10$  and  $\rho = 18dB$ . The dependence of the SNRs on  $T$  is essentially driven by the dependence of  $\kappa$  on  $T_c$  as derived in Proposition 2. Therefore, the minimum for the amount of power devoted to training

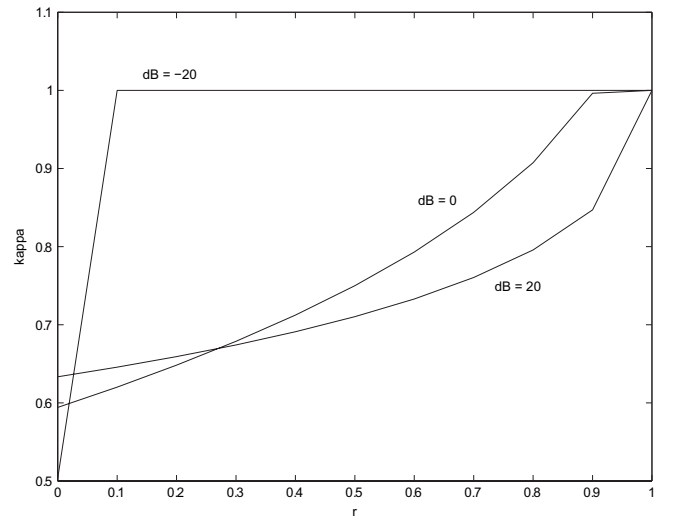


Fig. 3. Plot of optimum energy allocation  $\kappa$  as a function of Rician parameter  $r$ .

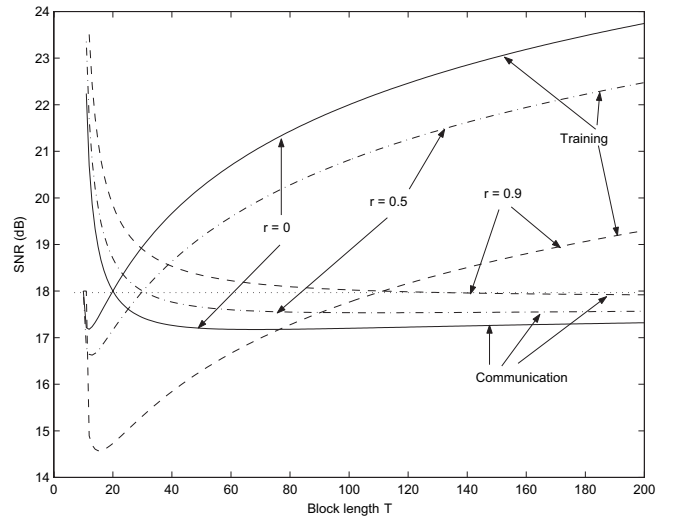


Fig. 4. Plot of SNRs in training and communication phases as a function of  $T$  for fixed  $T_t = M$ .

occurs at  $T = T_t + T_c = M + (1 - r)M = (2 - r)M$ . We decided to plot the SNRs instead of the actual powers because we wanted to contrast the behavior of Rician channels for different values of  $r$  against that of Rayleigh channel ( $r = 0$ ) as shown in Figure 4 of [14]. We see that as  $r$  goes to 1 less and less power is allocated to the training phase. This makes sense because as the proportion of signal energy transmitted through the specular component increases there is a lesser need for the system to estimate the unknown Rayleigh component.

Figure 5 shows capacity as a function of the number of transmit antennas in low-SNR regime ( $\rho \approx 0dB$ ) for a fixed block length  $T = 40$  and fixed number of receive antennas  $N = 40$ . We can easily estimate 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  $r$  increases. This shows that as  $r$  goes to 1, for a fixed  $T$ , we can tolerate larger uncertainty in the Rayleigh component. In other words, fewer resources can be devoted to estimate the unknown Rayleigh part of the channel because as pointed

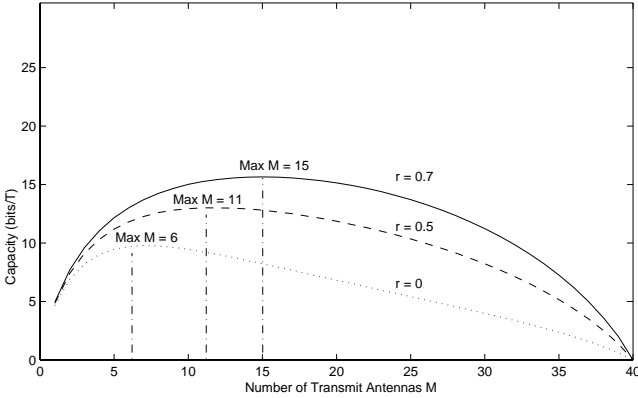


Fig. 5. Plot of capacity as a function of number of transmit antennas for a fixed  $T$ .

out in the previous paragraph in regards to Figure 4 there is less need to estimate the unknown Rayleigh part.

## V. OPTIMIZATION IN THE LOW SNR AND HIGH SNR REGIMES

Let's consider the effect of low SNR on the optimization of  $\kappa$  in (12) when  $r \neq 0$ . For  $T_c > (1-r)M$ , as  $\rho \rightarrow 0$  it is easy to see that  $\gamma - \sqrt{\gamma(\gamma-1-\eta)} \rightarrow \infty$ . Therefore, we conclude that for small  $\rho$  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  $\rho$ . Also, the form of  $\Lambda$  is that of a diagonal matrix with only one non-zero diagonal element.

Evaluating the case where the training and transmission powers are equal we come to a similar conclusion. For small  $\rho$ ,  $\rho_{eff} \approx r\rho$  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  $T_t = T/2$  [14] and  $\rho_{eff}$  behaves as  $O(\rho^2)$ .

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,  $r_{eff} = 1$  for small  $\rho$  showing that the channel behaves as a purely Gaussian channel.

These conclusions mimic those of Proposition 3 in Section 3.1 of [12] for non-coherent capacity results with Gaussian input. The low SNR non-coherent capacity results in Section 3.1 of [12] for the case of a Gaussian input signal tell us that the capacity behaves as  $r\rho\lambda_{max}$  for Rician fading and behaves as  $\rho^2$  for Rayleigh fading which is what the lower bound results in [14] also show. We would expect the general behavior of the results obtained here to agree with those of [12] as both sets of results are obtained under the same assumption of Gaussian probability distribution on the input signal.

Next consider the case of high SNR. Now,  $\gamma$  becomes  $\frac{T_c}{T_c - (1-r)M}$  and the optimum power allocation  $\kappa$  becomes

$$\kappa = \frac{\sqrt{T_c}}{\sqrt{T_c} + \sqrt{(1-r)M}} \quad (24)$$

and

$$\rho_{eff} = \frac{T}{(\sqrt{T_c} + \sqrt{(1-r)M})^2} \rho. \quad (25)$$

In the case of equal training and transmit powers, we have for high  $\rho$

$$\rho_{eff} = \rho \frac{T_t}{T_t + M(1-r)}. \quad (26)$$

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

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

$$C_t/T \geq f(M). \quad (27)$$

where  $f(M)$  denotes a function dependent on  $M$  and is given by

$$f(M) = \frac{T-M}{T} E \log \det \left( I_M + \rho'(M) \frac{HH^\dagger}{M} \right).$$

with  $\rho'(M)$  another function of  $M$  given by

$$\rho'(M) = \frac{\rho}{\left( \sqrt{1 - \frac{M}{T}} + \sqrt{\frac{(1-r)M}{T}} \right)^2}.$$

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

$$C_t/T \geq \max_{M' \leq M} \max_{n \leq \binom{M}{M'}} f_2(M', n), \quad (28)$$

where now

$$f_2(M', n) = \frac{T-M'}{T} E \log \det \left( I_{M'} + \rho'(M') \frac{H(n)H(n)^\dagger}{M'} \right)$$

with  $\rho'(\cdot)$  as defined earlier.  $H(n)$  denotes the  $n^{\text{th}}$  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  $\lambda_i(n)$  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

$$C_t/T \geq \max_{M' \leq M} \max_{n \leq \binom{M}{M'}} f_3(M', n) \quad (29)$$

where now

$$f_3(M', n) = \frac{T - M'}{T} \sum_{i=1}^Q E \log(1 + \rho \lambda_i(n)).$$

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

$$C_t/T \geq \max_{M' \leq M} \begin{cases} (1 - \frac{M'}{T})M' \log \rho & \text{if } M' \leq N \\ (1 - \frac{M'}{T})N \log \rho & \text{if } M' > N. \end{cases} \quad (30)$$

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 [14] and [22]. Also, the leading term in  $\rho$  for high SNR in the lower bound is given by

$$C_t/T \geq (1 - \frac{K}{T})K \log \rho \quad (31)$$

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, the results obtained in Section 3.4 of [12] for the case when  $M \leq N$  and large  $T$  show that the asymptotic capacity behaves as  $M(T - M) \log \rho$  from which we see that the lower bound is tight in the sense that the leading term involving  $\rho$  in the lower bound is the same as the one in the expression for capacity.

## VI. CONCLUSIONS AND FUTURE WORK

In this paper, we investigated the utility of training based communication schemes in non-coherent communications over Rician fading channels with deterministic specular component. The findings in this paper can be compared with the findings on Rayleigh fading channels from [13] and [14]. The similarities and differences are summarized below:

- Our study showed that for Rician fading channels, similar to Rayleigh fading, at low SNR training based schemes are suboptimum whereas for high SNRs training can attain capacity.
- For high SNR and large coherence interval Rayleigh and Rician channels behave in a similar manner both in terms of the parameters maximizing the training based lower bound and the lower bound itself. This means that the number of degrees of freedom of a training based scheme over a Rician fading channel (Section V) is the same as that of a Rayleigh fading channel which is  $\min\{M, N, T/2\}$  [14], [22].
- The differences between Rayleigh fading and Rician fading show up at low SNR. For low SNR from the analysis itself we conclude that for Rician fading channels no time or energy should be spent in training the receiver to learn the channel. More precisely, for Rician fading channels there exists a threshold that is an increasing function of the Rician factor  $r$  such that for all SNRs below the threshold the best strategy is not to spend any effort in learning the channel.

- Another difference between Rayleigh and Rician fading channels is that the lower bound on training based capacity for Rayleigh fading channels is a quadratic function of SNR whereas for Rician fading it is a linear function. This linear function of SNR is still suboptimum except when  $r = 1$  at which point the Rician fading channel is simply an AWGN channel.
- Finally, there exists a difference between Rician and Rayleigh fading in the optimum transmit signal covariance matrix structure for low SNR. Eventhough, for low SNR the optimum transmit signal covariance matrix consists of a single non-zero eigenvalue in both cases; for the case of Rician fading the eigenvector corresponding to the non-zero eigenvalue has to point in the same direction as the specular component of maximum strength.

In the future, it would be very useful to obtain a solution to the problem of finding maximum achievable rates on non-coherent block fading Rician channels without explicit channel estimation along the lines of [15]. Some progress in this regard was made in [10], [11]. In addition, the assumption of deterministic specular component was relaxed and the specular component was treated as unknown. However, research along these lines is challenging and a breakthrough would require considerable efforts from the communications research community.

## APPENDIX I

### PROOF OF PROPOSITION 1

First we note that  $\sigma_{\bar{G}}^2 = 1 - \sigma_G^2$ . This means that

$$\rho_{eff} = \frac{\kappa T \rho + T_c}{(1 - r)\kappa T \rho \sigma_G^2 + T_c} - 1. \quad (32)$$

Therefore, to maximize  $\rho_{eff}$  we just need to minimize  $\sigma_G^2$ . Now,

$$\sigma_G^2 = \frac{1}{NM} \text{tr}\{E[\text{vec}(\bar{G})\text{vec}(\bar{G})^\dagger]\}$$

where  $\text{vec}(\bar{G})$  is a column vector obtained by stacking the columns of  $\bar{G}$  one on top of the other. Therefore,

$$\begin{aligned} \sigma_G^2 &= \frac{1}{NM} \text{tr}\{E[\text{vec}(\bar{G})\text{vec}(\bar{G})^\dagger]\} \\ &= \frac{1}{NM} \text{tr}\left\{(I_M + (1 - r)\frac{\rho}{M}S_t^\dagger S_t)^{-1} \otimes I_N\right\} \end{aligned} \quad (33)$$

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

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

The problem above can be restated as

$$\min_{\lambda_1, \dots, \lambda_M: \sum \lambda_m \leq (1 - \kappa)TM} \frac{1}{M} \sum_{m=1}^M \frac{1}{1 + (1 - r)\frac{\rho}{M}\lambda_m} \quad (34)$$

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

This gives  $\sigma_G^2 = \frac{1}{1 + (1 - r)\frac{1}{M}(1 - \kappa)T}$ . Also, for this choice of  $S_t$  we obtain the elements of  $\hat{G}$  to be zero mean independent with Gaussian distribution. This gives (11) as the expression for  $\rho_{eff}$ . ■



APPENDIX II  
 PROOF OF PROPOSITION 2

First, from Proposition 1

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

$$= \begin{cases} \frac{T \rho}{T_c - (1 - r)M} \frac{(1 - \kappa)\kappa + \kappa \frac{rM}{T \rho}}{\frac{MT_c + T \rho T_c}{T \rho [T_c - (1 - r)M]} - \kappa} & \text{when } T_c \neq (1 - r)M \\ \frac{T^2 \rho^2}{T_c(M + T \rho)} [(1 - \kappa)\kappa + \kappa \frac{rM}{T \rho}] & \text{when } T_c = (1 - r)M. \end{cases}$$

Consider the following three cases for the maximization of  $\rho_{eff}$  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

$$\rho_{eff} = \frac{T^2 \rho^2}{(1 - r)M(M + T \rho)} \left[ \kappa_0 \frac{rM}{T \rho} + \kappa_0(1 - \kappa_0) \right]. \quad (35)$$

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

In this case,

$$\rho_{eff} = \frac{T \rho}{T_c - (1 - r)M} \frac{(1 - \kappa)\kappa + \kappa \eta}{\gamma - \kappa} \quad (36)$$

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  $\kappa = \min\{\gamma - \sqrt{\gamma^2 - \gamma - \eta}, 1\}$ . Therefore,

$$\rho_{eff} = \frac{T \rho}{T_c - (1 - r)M} (\sqrt{\gamma} - \sqrt{\gamma - 1 - \eta})^2 \quad (37)$$

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

$$\rho_{eff} = \frac{\kappa T \rho [Mr + \rho(1 - \kappa)T]}{T_c(M + \rho(1 - \kappa)T) + (1 - r)\kappa T \rho M} \quad (38)$$

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

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

In this case,

$$\rho_{eff} = \frac{T \rho}{(1 - r)M - T_c} \frac{(1 - \kappa)\kappa + \kappa \eta}{\kappa - \gamma} \quad (39)$$

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

$$\rho_{eff} = \frac{T \rho}{T_c - (1 - r)M} (\sqrt{-\gamma} - \sqrt{-\gamma + 1 + \eta})^2 \quad (40)$$

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

 APPENDIX III  
 PROOF OF PROPOSITION 3

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  $\lambda_i$  denote the  $i^{th}$  non-zero eigenvalue of  $\frac{H_1 H_1^H}{M}$ ,  $i = 1, \dots, Q$ . 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

$$\frac{dC_l}{dT_c} = \sum_{i=1}^Q \left\{ \frac{1}{T} E \log(1 + \rho_{eff} \lambda_i) + \frac{T_c}{T} \frac{d\rho_{eff}}{dT_c} E \left[ \frac{\lambda_i}{1 + \rho_{eff} \lambda_i} \right] \right\}. \quad (41)$$

Now,  $\rho_{eff}$  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

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

Therefore,

$$\frac{dC_l}{dT_c} = \frac{1}{T} \sum_{i=1}^Q E \left[ \log(1 + \rho_{eff} \lambda_i) - \left\{ \frac{\rho_{eff} \lambda_i}{1 + \rho_{eff} \lambda_i} \frac{T_c}{T_c - (1 - r)M} \times \left( 1 - \sqrt{\frac{(1 - r)M(M + T \rho)}{T_c(T_c + T \rho + rM)}} \right) \right\} \right]. \quad (42)$$

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] N. Benvenuto, P. Bisaglia, A. Salloum, and L. Tomba, "Worst case equalizer for noncoherent HIPERLAN receivers," *IEEE Trans. Commun.*, vol. 48, no. 1, pp. 28–36, Jan. 2000.
- [2] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley Series in Telecommunications, 1995.
- [3] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic Press, 1981.
- [4] P. F. Driessen and G. J. Foschini, "On the capacity formula for multiple input-multiple output wireless channels: a geometric interpretation," *IEEE Trans. Commun.*, vol. 47, no. 2, pp. 173–176, Feb. 1999.
- [5] F. R. Farrokhi, G. J. Foschini, A. Lozano, and R. A. Valenzuela, "Link-optimal space-time processing with multiple transmit and receive antennas," *IEEE Commun. Lett.*, vol. 5, no. 3, pp. 85–87, Mar. 2001.
- [6] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multiple antennas," *Bell Labs Technical J.*, vol. 1, no. 2, pp. 41–59, 1996.

- [7] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Commun.*, vol. 6, no. 3, pp. 311–335, Mar. 1998.
- [8] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.
- [9] R. G. Gallager, "A simple derivation of the coding theorem and some applications," *IEEE Trans. Inf. Theory*, vol. 11, no. 1, pp. 3–18, Jan. 1965.
- [10] M. Godavarti, A. O. Hero, and T. Marzetta, "Min-capacity of a multiple-antenna wireless channel in a static Rician fading environment," *IEEE Trans. Wireless Commun.*, vol. 4, no. 4, pp. 1715–1723, July 2005.
- [11] M. Godavarti, T. Marzetta, and S. S. Shitz, "Capacity of a mobile multiple-antenna wireless link with isotropically random Rician fading," *IEEE Trans. Inf. Theory*, vol. 49, no. 12, pp. 3330–3334, Dec. 2003.
- [12] M. Godavarti and A. O. Hero III, "Multiple antenna capacity in a deterministic Rician fading channel," submitted to *IEEE Trans. Inf. Theory*.
- [13] H. Weingarten, Y. Steinberg, and S. Shamai (Shitz), "Gaussian codes and weighted nearest neighbor decoding in fading multi-antenna channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 9, pp. 1665–1686, Sept. 2004.
- [14] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, April 2003.
- [15] Z. B. Krusevac, P. B. Rapajic, and R. A. Kennedy, "On channel uncertainty modeling: an information theoretic approach," in *Proc. Global Telecommunications Conference 2004*, vol. 1, pp. 410–414.
- [16] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading channel," *IEEE Trans. Inf. Theory*, vol. 45, no. 1, pp. 139–157, Jan. 1999.
- [17] T. L. Marzetta, "BLAST training: estimating channel characteristics for high-capacity space-time wireless," in *Proc. 37th Annual Allerton Conference Communications, Control, and Computing 1999*.
- [18] S. Rice, "Mathematical analysis of random noise," *Bell Systems Technical J.*, vol. 23, 1944.
- [19] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: performance criterion and code construction," *IEEE Trans. Commun.*, vol. 44, no. 2, pp. 744–765, March 1998.
- [20] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. Telecommun.*, vol. 10, no. 6, pp. 585–596, Nov./Dec. 1999.
- [21] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1319–1343, June 2002.
- [22] L. Zheng and D. N. C. Tse, "Packing spheres in the Grassmann manifold: a geometric approach to the non-coherent multi-antenna channel," *IEEE Trans. Inf. Theory*, vol. 48, no. 2, pp. 359–383, Feb. 2002.



**Alfred O. Hero, III** was born in Boston, MA, in 1955. He received the B.S. degree in electrical engineering (summa cum laude) from Boston University in 1980 and the Ph.D. from Princeton University, Princeton, NJ, in 1984, both in electrical engineering. While at Princeton, he held the G.V.N. Lothrop Fellowship in Engineering. Since 1984 he has been with the University of Michigan, Ann Arbor, where he is a Professor in the Department of Electrical Engineering and Computer Science and, by courtesy, in the Department of Biomedical Engineering and the Department of Statistics. He has held visiting positions at Massachusetts Institute of Technology (2006), I3S University of Nice, Sophia-Antipolis, France (2001), Ecole Normale Supérieure de Lyon (1999), Ecole Nationale Supérieure des Télécommunications, Paris (1999), Scientific Research Labs of the Ford Motor Company, Dearborn, Michigan (1993), Ecole Nationale Supérieure des Techniques Avancées (ENSTA), Ecole Supérieure d'Electricité, Paris (1990), and M.I.T. Lincoln Laboratory (1987 - 1989). His recent research interests have been in areas including: inference for sensor networks, adaptive sensing, inverse problems, bioinformatics, and statistical signal and image processing.

He has served on the editorial boards of the *IEEE Transactions on Information Theory* (1995-1998, 1999), the *IEEE Transactions on Computational Biology and Bioinformatics* (2004-2006), and the *IEEE Transactions on Signal Processing* (2002, 2004). He was Chairman of the Statistical Signal and Array Processing (SSAP) Technical Committee (1997-1998) and Treasurer of the Conference Board of the IEEE Signal Processing Society. He was Chairman for Publicity of the 1986 IEEE International Symposium on Information Theory (Ann Arbor, MI) and General Chairman of the 1995 IEEE International Conference on Acoustics, Speech, and Signal Processing (Detroit, MI). He was co-chair of the 1999 IEEE Information Theory Workshop on Detection, Estimation, Classification and Filtering (Santa Fe, NM) and the 1999 IEEE Workshop on Higher Order Statistics (Caesaria, Israel). He chaired the 2002 NSF Workshop on Challenges in Pattern Recognition. He co-chaired the 2002 Workshop on Genomic Signal Processing and Statistics (GENSIPS). He was Vice President (Finance) of the IEEE Signal Processing Society (1999-2002). He was Chair of Commission C (Signals and Systems) of the US National Commission of the International Union of Radio Science (URSI) (1999-2002). He was member of the Signal Processing Theory and Methods (SPTM) Technical Committee of the IEEE Signal Processing Society (1999-2004). He is currently President of the IEEE Signal Processing Society (2006-2007) and a member of the IEEE TAB Periodicals Committee (2006-2008).

Alfred Hero is a Fellow of the Institute of Electrical and Electronics Engineers (IEEE), a member of Tau Beta Pi, the American Statistical Association (ASA), the Society for Industrial and Applied Mathematics (SIAM), and the US National Commission (Commission C) of the International Union of Radio Science (URSI). He has been plenary and keynote speaker and several major conferences and received a IEEE Signal Processing Society Meritorious Service Award (1998), a IEEE Signal Processing Society Best Paper Award (1998), a IEEE Third Millennium Medal (2000) and a 2002 IEEE Signal Processing Society Distinguished Lecturership.



**Mahesh Godavarti** was born in Jaipur, India in 1972. He received the B.Tech degree in Electrical and Electronics Engineering from the Indian Institute of Technology, Madras, India in 1993 and the M.S. degree in Electrical and Computer Engineering from the University of Arizona, Tucson, AZ in 1995. He was at the University of Michigan from 1997 to 2001 where he received the M.S. degree in Applied Mathematics and the Ph.D. degree in Electrical Engineering. Currently, he is employed as DSP Algorithm Manager with Ditech Networks,

Mountain View, CA, where he is researching new algorithms for speech enhancement. His research interests include topics in speech and signal processing, communications and information theory.