Min-Capacity of a Multiple-Antenna Wireless Channel in a Static Ricean Fading Environment

Mahesh Godavarti, *Member, IEEE*, Alfred O. Hero-III, *Fellow, iEEE*, and Thomas L. Marzetta, *Fellow, IEEE*,

***Abstract*— This paper presents the optimal guaranteed performance for a multiple-antenna wireless compound channel with M antennas at the transmitter and N antennas at the receiver on a Ricean fading channel with a static specular component. The channel is modeled as a compound channel with a Rayleigh component and an unknown rank-one deterministic specular component. The Rayleigh component remains constant over a block of T symbol periods, with independent realizations over each block. The rank one deterministic component is modeled as an outer product of two unknown deterministic vectors of unit magnitude. Under this scenario to guarantee service it is required to maximize the worst-case capacity (*min-capacity*). It is shown that for computing *min-capacity* instead of optimizing over the joint density of $T \cdot M$ complex transmitted signals it is sufficient to maximize over a joint density of $\min\{T, M\}$ real transmitted signal magnitudes. The optimal signal matrix is shown to be equal to the product of three independent matrices, a $T \times T$ unitary matrix, a $T \times M$ real nonnegative diagonal matrix and a $M \times M$ unitary matrix. A tractable lower bound on capacity is derived for this model which is useful for computing achievable rate regions. Finally, it is shown that the average capacity (*avg-capacity*) computed under the assumption that the specular component is constant but, random with isotropic distribution is equal to *min-capacity*. This means that *avg-capacity* which in general has no practical meaning for non-ergodic scenarios, has a coding theorem associated with it in this particular case.**

***Index Terms*— capacity, compound channel, information theory, Ricean fading, multiple antennas.**

I. Introduction

NEED for higher rates in wireless communications has never been greater than in the present. Due to this need and the dearth of extra bandwidth available for communication multiple antennas have attracted considerable attention [6], [7], [16], [20], [21]. Multiple antennas at the transmitter and the receiver provide spatial diversity that can be exploited to improve spectral efficiency of wireless communication systems and to improve performance.

Two kinds of models widely used for describing fading in wireless channels are Rayleigh and Ricean models. For wireless links in Rayleigh fading environment, it has been shown by Foschini et. al. [6], [7] and Telatar [20] that with perfect channel knowledge at the receiver, for high SNR a capacity gain of $\min(M, N)$ bits/second/Hz, where M is the number of antennas at the transmitter and N is the number of antennas at the receiver, can be achieved with every 3 dB increase in SNR. The assumption of complete knowledge about the channel might not be true in the case of fast mobile receivers and large number of transmit antennas because of insufficient training. Marzetta and Hochwald [16] considered the case when neither the receiver nor the transmitter has any knowledge of the fading coefficients. They consider a model where the fading coefficients remain constant for T symbol periods and instantaneously change to new independent realizations after that. They derive the structure of capacity achieving signals and also show that under this model the complexity for capacity calculations is considerably reduced.

In contrast, the attention paid to Ricean fading models has been fairly limited. Ricean fading components traditionally have been modeled as independent Gaussian components with a deterministic non-zero mean [1], [4], [5], [12], [17], [19]. Farrokhi et. al. [5] used this model to analyze the capacity of a MIMO channel with a specular component. They assume that the specular component is deterministic and unchanging and unknown to the transmitter but known to the receiver. They also assume that the receiver has complete knowledge about the fading coefficients (i.e. has knowledge about the Rayleigh component as well). They work with the premise that since the transmitter has no knowledge about the specular component the signaling scheme has

Manuscript received January 20, 2002; revised November 18, 2002.

Mahesh Godavarti is with Ditech Communications Inc., Mountain View. Alfred Hero is with the Dept of Electrical Engineering, the University of Michigan, Ann Arbor. Thomas L. Marzetta is with Bell Laboratories, Lucent Technologies, Murray Hill. This work was performed in part while M. Godavarti was a summer intern at the Mathematical Sciences Research Center, Bell Laboratories and in part while M. Godavarti was a Ph.D. candidate at the University of Michigan. Parts of this work were presented at ISIT 2001 held in Washington, D.C., USA

to be designed to guarantee a given rate irrespective of the value of the deterministic specular component. They conclude that the signal matrix has to be composed of independent circular Gaussian random variables of mean 0 and equal variance to maximize the rate.

Godavarti et. al [13] consider a non-conventional ergodic model for the case of Ricean fading where the fading channel consists of a Rayleigh component, modeled as in [16] and an independent rank-one isotropically distributed specular component. The fading channel is assumed to remain constant over a block of T consecutive symbol periods but take a completely independent realization over each block. They derive similar results on optimal capacity achieving signal structures as in [16]. They also establish a lower bound to capacity that can be easily extended to the model considered in this paper. The model described in [13] is applicable to a mobile-wireless link where both the direct line of sight component (specular component) and the diffuse component (Rayleigh component) change with time.

In [12], Godavarti et. al. consider the standard Ricean fading model. The capacity calculated for the standard Ricean fading model is a function of the specular component since the specular component is deterministic and known to both the transmitter and the receiver. The authors establish asymptotic results for capacity and conclude that beamforming is the optimum signaling strategy for low SNR whereas for high SNR the optimum signaling strategy is same as that for purely Rayleigh fading channels.

In this paper, we consider a quasi-static Ricean model where the specular component is non-changing while the Rayleigh component is varying over time. The only difference between this model and the standard Ricean fading model is that in this model the specular component is of single rank and is not known to the transmitter. We can also contrast the formulation here to that in [13] where the specular component is also modeled as stochastic and the ergodic channel capacity is clearly defined. In spite of a completely different formulation, we obtain surprisingly similar results as [13].

Modeling the specular component to be of rank one is fairly common in the literature [14], [15], [18]. The rank of the specular component is determined by the number of direct line of sight paths between the transmitter and the receiver, which is typically much lower than the number of transmit and receive antennas leading to ill-conditioned specular components [10], [11]. Furthermore, if the distance between the transmit and receive antennas is much greater than the distance between individual antenna elements then the rank of the specular component can only be one [3], [4].

The Ricean channel models considered in [12], [13] and in this paper are all extensions of the Rayleigh model considered in [16] in the sense that all models reduce to the Rayleigh model of [16] when the specular component goes to zero.

The channel model considered here is applicable to the case where the transmitter and receiver are either fixed in space or are in motion but sufficiently far apart with a single direct path so that the specular component has single rank and is practically constant while the diffuse multipath component changes rapidly. This can be contrasted with the channel model of [13] where the specular component is changing as rapidly as the diffuse multipath component. This allows modeling the channel of [13] as an ergodic channel. On the other hand, the channel model of [12] is almost exactly the same as the model proposed in this paper except that in [12] it is assumed that there is a feedback path from the receiver to the transmitter and as a result the transmitter can be modeled to have complete knowledge of the specular component.

In this paper, since the transmitter has no knowledge about the specular component the transmitter can either maximize the worst-case rate over the ensemble of values that the specular component can take or maximize the average rate by establishing a prior distribution on the ensemble. We address both approaches in this paper. Note that when the transmitter has no knowledge about the specular component, knowledge of it at the receiver makes no difference on the worst-case capacity [2]. We however assume the knowledge as it makes it easier to analyze the fading channel.

Similar to [5] the specular component is an outer product of two vectors of unit magnitude that are non-changing and unknown to the transmitter but known to the receiver. The difference between our approach and that of [5] is that in [5] the authors consider the channel to be known completely to the receiver. We assume that the receiver's extent of knowledge about the channel is limited to the specular component. That is, the receiver has no knowledge about the Rayleigh component of the model. Considering the absence of knowledge at the transmitter it is important to design a signal scheme that guarantees the largest overall rate for communication irrespective of the value of the specular component. This is formulated as the problem of determining the worst-case capacity in Section II. This is followed by derivation of upper and lower bounds on the worst-case capacity in Section III and optimum signal properties in Section IV. In Section V the average capacity is considered instead of worst-case capacity and it is shown that both formulations imply the same optimal signal structure and the same maximum possible rate. In Section VI we use the results derived in this paper to compute capacity regions for some Ricean fading channels. For interested readers, we show the existence, in the appendix, of a coding theorem corresponding to the worst-case capacity for the fading model considered here.

II. Signal Model and Problem Formulation

Let there be M transmit antennas and N receive antennas. It is assumed that the fading coefficients remain constant over a block of T consecutive symbol periods but are independent from block to block. Keeping that in mind, the channel is modeled as carrying a $T \times M$ signal matrix S over a $M \times N$ MIMO channel H, producing X at the receiver according to the model:

$$X = \sqrt{\frac{\rho}{M}} SH + W \tag{1}$$

where the elements, w_{tn} of W are independent circular complex Gaussian random variables with mean 0 and variance 1 ($\mathcal{CN}(0,1)$).

The MIMO Ricean model for the matrix H is $H = \sqrt{1-r}G + \sqrt{rNM}\alpha\beta^\dagger$ where G consists of independent $\mathcal{CN}(0,1)$ random variables and α and β are deterministic vectors of length M and N, respectively, such that $\alpha^\dagger\alpha = 1$ and $\beta^\dagger\beta = 1$. The parameter r, $0 \le r \le 1$ denotes the fraction of the energy propagated via the specular component. $r = 0$ and $r = 1$ correspond to purely Rayleigh and purely specular fading, respectively. Irrespective of the value of r, the average variance of the elements of H is 1, that is H satisfies $E[\mathrm{tr}\{HH^\dagger\}] = M \cdot N$.

It is assumed α and β are known to the receiver. Since the receiver is free to apply a co-ordinate transformation by post multiplying X by a unitary matrix, without loss of generality β can be taken to be identically equal to $[1\ 0\ \ldots\ 0]^T$. We will sometimes write H as H_α to highlight the dependence of H on α. G remains constant for T symbol periods and takes on a completely independent realization every T^{th} symbol period.

The problem in this section is to find the distribution $p^*(S)$ that attains the maximum in the following maximization defining the worst case channel capacity

$$C^* = \max_{p(S)} I^*(X;S) = \max_{p(S)} \inf_{\alpha \in A} I^\alpha(X;S)$$

and also to find the maximum value, C^*.

$$I^\alpha(X;S) = \int p(S)p(X|S,\alpha\beta^\dagger) \log \frac{p(X|S,\alpha\beta^\dagger)}{\int p(S)p(X|S,\alpha\beta^\dagger)\, dS}\, dSdX$$

is the mutual information between X and S when the specular component is given by $\alpha\beta^\dagger$ and $A \stackrel{\text{def}}{=} \{\alpha : \alpha \in \mathbb{C}^M \text{ and } \alpha^\dagger\alpha = 1\}$. Since A is compact the "inf" in the problem can be replaced by "min". For convenience we will refer to $I^*(X;S)$ as the *min-mutual information* and C^* as *min-capacity*.

The above formulation is justified for the Ricean fading channel considered here because there exists a corresponding coding theorem that we prove in the appendix. However, the existence of a coding theorem can also be obtained from [2, chapter 5, pp. 172-178]. *Min-capacity* defined above is just the capacity of a *compound channel*. We will use the notation in this paper to briefly describe the concept of compound channels given in [2]. Let $\alpha \in A$ denote a candidate channel. Let $C^* = \max_{p(S)} \min_\alpha I^\alpha(X;S)$ and $P^*(e,n) = \max_\alpha P^\alpha(e,n)$ where $P^\alpha(e,n)$ is the maximum probability of decoding error for channel α when a code of length n is used. Then for every $R < C^*$ there exists a sequence of ($2^{nR}, n$) codes such that

$$\lim_{n\to\infty} P^*(e,n) = 0.$$

It is also shown in [2, Prob. 13, p. 183] that *min-capacity* does not depend on the receiver's knowledge of the channel. Hence, it is not necessary for us to assume that the specular component is known to the receiver. However, we do so because it facilitates easier computation of *min-capacity* and *avg-capacity* in terms of the conditional probability distribution $p(X|S)$.

Note that since A is unitarily invariant it means that no preference is attached to the direction of the line of sight component and therefore, it is intuitive to expect the optimum signal to attach no significance to the direction of the line of sight component as well. Moreover, since all $\alpha \in A$ have the same strength it is intuitive to expect the optimum signal to be such that it generates the same mutual information irrespective of the choice of the specular component. This intuition is made concrete in the following sections.

III. Capacity Upper and Lower Bounds

Theorem 1: Min-capacity, C_H^* when the channel matrix H is known to the receiver but not to the transmitter is given by

$$C_H^* = TE \log\det\left[I_N + \frac{\rho}{M} H_{e_1}^\dagger H_{e_1}\right] \tag{2}$$

where $e_1 = [1\ 0\ \ldots\ 0]^T$ is a unit vector in $\mathbb{C}^M$. Note that e_1 in (2) can be replaced by any $\alpha \in A$ without changing the answer.

Proof: The idea for this proof has been taken from the proof of Theorem 1 in [20]. First note that for $T > 1$, given H the channel is memoryless and hence the rows of the input signal matrix S are independent of each other. That means the mutual information $I^\alpha(X;S) = \sum_{i=1}^T I^\alpha(X_i;S_i)$ where X_i and S_i denote the i^{th} row of X and S, respectively. The maximization over each term can be done separately and it is easily seen that each term will be maximized individually for the same density on S_i. That is $p(S_i) = p(S_j)$ for $i \neq j$ and $\max_{p(S)} I^\alpha(X;S) = T\max_{p(S_1)} I^\alpha(X_1;S_1)$. Therefore, WLOG assume $T = 1$.

Given H the channel is an AWGN channel therefore, capacity is attained by Gaussian signal vectors. Let Λ_S be the input signal covariance. Since the transmitter does not know α, Λ_S can not depend on α and the *min-capacity* is given by

$$\begin{aligned}&\max_{\Lambda_S:\mathrm{tr}\{\Lambda_S\}\leq M} \mathcal{F}(\Lambda_S) = \\ &\max_{\Lambda_S:\mathrm{tr}\{\Lambda_S\}\leq M} \min_{\alpha\in A} E\log\det\left[I_N + \tfrac{\rho}{M}H_\alpha^\dagger\Lambda_S H_\alpha\right]\end{aligned} \quad (3)$$

where $\mathcal{F}(\Lambda_S)$ is implicitly defined in an obvious manner. First note that $\mathcal{F}(\Lambda_S)$ in (3) is a concave function of Λ_S (This follows from the fact that $\log\det K$ is a concave function of K). Also, $\mathcal{F}(\Psi^\dagger\Lambda_S\Psi) = \mathcal{F}(\Lambda_S)$ for any $M\times M$ $\Psi : \Psi^\dagger\Psi = I_M$ since $\Psi^\dagger\alpha \in A$ for any $\alpha \in A$ and G has i.i.d. zero mean complex Gaussian entries. Let $Q^\dagger DQ$ be the SVD of Λ_S then we have $\mathcal{F}(D) = \mathcal{F}(Q^\dagger DQ) = \mathcal{F}(\Lambda_S)$. Therefore, we can choose Λ_S to be diagonal. Moreover, $\mathcal{F}(P_k^\dagger\Lambda_S P_k) = \mathcal{F}(\Lambda_S)$ for any permutation matrix P_k, $k = 1,\ldots,M!$. Therefore, if we choose $\Lambda'_S = \frac{1}{M!}\sum_{k=1}^{M!} P_k^\dagger\Lambda_S P_k$ then by concavity and Jensen's inequality we have

$$\mathcal{F}(\Lambda'_S) \geq \frac{1}{M!}\sum_{k=1}^{M!}\mathcal{F}(P_k^\dagger\Lambda_S P_k) = \mathcal{F}(\Lambda_S)$$

Therefore, it can be concluded that the maximizing input signal covariance Λ_S is a multiple of the identity matrix. It is quite obvious to see that to maximize the expression in (3) we need to choose $\mathrm{tr}\{\Lambda_S\} = M$ or $\Lambda_S = I_M$ and since $E\log\det[I_N + \frac{\rho}{M}H_{\alpha_1}^\dagger H_{\alpha_1}] = E\log\det[I_N + \frac{\rho}{M}H_{\alpha_2}^\dagger H_{\alpha_2}]$ for any $\alpha_1,\alpha_2 \in A$, (2) easily follows. ■

By the data processing theorem additional information at the receiver does not decrease capacity. Therefore:

Proposition 1: An upper bound on the channel *min-capacity* when neither the transmitter nor the receiver has any knowledge about the channel is given by

$$C^* \leq T\cdot E\log\det\left[I_N + \frac{\rho}{M}H_{e_1}^\dagger H_{e_1}\right] \quad (4)$$

Now, we establish a lower bound.

Proposition 2: A lower bound on *min-capacity* when the transmitter has no knowledge about H and the receiver has no knowledge about G is given by

$$\begin{aligned}C^* &\geq C_H^* - NE\left[\log_2\det\left(I_T + (1-r)\frac{\rho}{M}SS^\dagger\right)\right] && (5)\\ &\geq C_H^* - NM\log_2(1+(1-r)\frac{\rho}{M}T) && (6)\end{aligned}$$

Proof: Proof is a slight modification of the proof of Theorem 3 in [13] therefore, only the essential steps will be shown here.

First note that,

$$\begin{aligned}I^\alpha(X;S) &= I(X;S|\alpha)\\ &= I(X;S,H|\alpha) - I(X;H|S,\alpha)\\ &= I(X;H|\alpha) + I(X;S|H,\alpha) - I(X;H|S,\alpha)\\ &\geq I(X;S|H,\alpha) - I(X;H|S,\alpha)\end{aligned}$$

where the last inequality follows from the fact that $I(X;H|\alpha) \geq 0$. Therefore,

$$C^*(X;S) \geq \max_{p(S)}\min_\alpha[I(X;S|H,\alpha) - I(X;H|S,\alpha)]$$

The lower bound is obtained by observing that the second term is the mutual information between the "input" $H = \sqrt{1-r}G + \sqrt{rNM}\alpha e_1^\dagger$, and the "output" X through the "channel" $X = \sqrt{\frac{\rho}{M}}SH + W$. Since α is fixed and α and S are known at the "receiver", the mutual information between H and X is same as the mutual information between G and X' where $X' = X - \sqrt{\frac{\rho}{M}}\sqrt{rNM}S\alpha e_1^\dagger$. Therefore, the second term can be evaluated, irrespective of the value of α, as

$$NE\left[\log_2\det\left(I_T + (1-r)\frac{\rho}{M}SS^\dagger\right)\right]$$

and the first term can be maximized by choosing $p(S)$ such that the elements of S are independent $\mathcal{CN}(0,1)$ random variables. ■

Notice that the second term in right hand side of the lower bound is

$$NE\left[\log_2\det\left(I_T + (1-r)\frac{\rho}{M}SS^\dagger\right)\right]$$

instead of $NE\left[\log_2\det\left(I_T + \frac{\rho}{M}SS^\dagger\right)\right]$ which occurs in the lower bound derived for the model in [13]. The second term $I(X;H|S)$, is the mutual information between the output and the channel given the transmitted signal. In other words this is the information carried in the transmitted signal about the channel. Therefore, the second term in the lower bound can be viewed as a penalty term for using part of the available rate to learn the channel. When $r = 1$ or when the channel is purely specular it can be seen that the penalty term for training goes to zero. This makes perfect sense because the specular component is known to the receiver and the penalty for learning the specular component is zero in the current model as contrasted to the model in [13].

Combining (4) and (6) gives us the following

Corollary 1: The normalized *min-capacity*, $C_n^* = C^*/T$ in bits per channel use as $T \to \infty$ is given by

$$C_n^* = E \log\det\left[I_N + \frac{\rho}{M} H_{e_1}^\dagger H_{e_1}\right]$$

Note that this is same as the capacity when the receiver knows H, so that as $T \to \infty$ perfect channel estimation can be performed.

IV. Properties of capacity achieving signals

In this section, the optimum signal structure for achieving *min-capacity* is derived. The optimization is being done under the power constraint $E[\mathrm{tr}\{SS^\dagger\}] \leq TM$.

The results in this section theoretically establish what can be gauged intuitively. It has already been established in [16] that when the specular component is zero the optimum signal density is invariant to unitary transformations. This is no longer true if the specular component is non-zero. However, if the set of non-zero specular components (called A in this paper), from which the worst case specular component is selected and the corresponding performance maximized, is invariant to unitary transformations then it is natural to expect the optimum signal to be invariant to unitary transformations as well.

The basic ideas for showing invariance of optimum signals to unitary transformations in this section, and also in the next, have been taken from [16].

Lemma 1: $I^*(X;S)$ as a functional of $p(S)$ is concave in $p(S)$.

Proof: First note that $I^\alpha(X;S)$ is a concave functional of $p(S)$ for every $\alpha \in A$. Let $I^*(X;S)_{p(S)}$ denote $I^*(X;S)$ evaluated using $p(S)$ as the signal density. Then,

$$\begin{aligned}
&I^*(X;S)_{\delta p_1(S)+(1-\delta)p_2(S)} \\
&= \min_{\alpha\in A} I^\alpha(X;S)_{\delta p_1(S)+(1-\delta)p_2(S)} \\
&\geq \min_{\alpha\in A}[\delta I^\alpha(X;S)_{p_1(S)} + (1-\delta)I^\alpha(X;S)_{p_2(S)}] \\
&\geq \delta \min_{\alpha\in A} I^\alpha(X;S)_{p_1(S)} + (1-\delta)\min_{\alpha\in A} I^\alpha(X;S)_{p_2(S)} \\
&= \delta I^*(X;S)_{p_1(S)} + (1-\delta) I^*(X;S)_{p_2(S)}
\end{aligned}$$

■

Lemma 2: For any $T \times T$ unitary matrix Φ and any $M \times M$ unitary matrix Ψ, if $p(S)$ generates $I^*(X;S)$ then so does $p(\Phi S \Psi^\dagger)$.

Proof: 1) Note that $p(\Phi X|\Phi S) = p(X|S)$, therefore $I^\alpha(X;\Phi S) = I^\alpha(X;S)$ for any $T \times T$ unitary matrix Φ and all $\alpha \in A$.

2) Also, $\Psi\alpha \in A$ for any $\alpha \in A$ and any $M \times M$ unitary matrix Ψ. Therefore, if $I^*(X;S)$ achieves its minimum value at $\alpha_0 \in A$ then $I^*(X;S\Psi^\dagger)$ achieves its minimum value at $\Psi\alpha_0$ because $I^\alpha(X;S) = I^{\Psi\alpha}(X;S\Psi^\dagger)$ for $\alpha \in A$ and Ψ an $M \times M$ unitary matrix.

Combining 1) and 2) we get the lemma. ■

Lemma 3: The *min-capacity* achieving signal distribution, $p(S)$ is unchanged by any pre- and post- multiplication of S by unitary matrices of appropriate dimensions.

Proof: It will be shown that for any signal density $p_0(S)$ generating *min-mutual information* I_0^* there exists a density $p_1(S)$ generating $I_1^* \geq I_0^*$ such that $p_1(S)$ is invariant to pre- and post- multiplication of S by unitary matrices of appropriate dimensions. By Lemma 2, for any pair of permutation matrices, Φ $(T \times T)$ and Ψ $(M \times M)$ $p_0(\Phi S \Psi^\dagger)$ generates the same *min-mutual information* as $p(S)$. Define $u_T(\Phi)$ to be the isotropically random unitary density function of a $T \times T$ unitary matrix Φ. Similarly define $u_M(\Psi)$. Let $p_1(S)$ be a mixture density given as follows

$$p_1(S) = \int\int p_0(\Phi S \Psi^\dagger) u(\Phi) u(\Psi)\; d\Phi d\Psi$$

It is easy to see that $p_1(S)$ is invariant to any pre- and post- multiplication of S by unitary matrices and if I_1^* is the *min-mutual information* generated by $p_1(S)$ then from Jensen's inequality and concavity of $I^*(X;S)$ we have $I_1^* \geq I_0^*$. ■

Corollary 2: $p^*(S)$, the optimal *min-capacity* achieving signal density lies in $\mathcal{P} = \cup_{I>0}\mathcal{P}_I$ where

$$\mathcal{P}_I = \{p(S) : I^\alpha(X;S) = I \quad \forall \alpha \in A\} \tag{7}$$

Proof: Follows immediately from Lemma 3 because any signal density that is invariant to pre- and post-multiplication of S by unitary matrices generates the same mutual information $I^\alpha(X;S)$ irrespective of the value of α. ■

The above result is intuitively obvious because all $\alpha \in A$ are identical to each other except for unitary transformations. Therefore, any density function that is invariant to unitary transformations is expected to behave the same way for all α.

Theorem 2: The signal matrix that achieves *min-capacity* can be written as $S = \Phi V \Psi^\dagger$, where Φ and Ψ are $T \times T$ and $M \times M$ isotropically distributed matrices independent of each other, and V is a $T \times M$ real, nonnegative, diagonal matrix, independent of both Φ and Ψ.

Proof: From the singular value decomposition (SVD) we can write $S = \Phi V \Psi^\dagger$, where Φ is a $T \times T$ unitary matrix, V is a $T \times M$ nonnegative real diagonal matrix, and Ψ is an $M \times M$ unitary matrix. In general, Φ, V and Ψ

are jointly distributed. Suppose S has probability density $p_0(S)$ that generates *min-mutual information* I_0^*. Let Θ_1 and Θ_2 be isotropically distributed unitary matrices of size $T \times T$ and $M \times M$ independent of S and of each other. Define a new signal $S_1 = \Theta_1 S \Theta_2^\dagger$, generating *min-mutual information* I_1^*. Now conditioned on Θ_1 and Θ_2, the *min-mutual information* generated by S_1 equals I_0^*. From the concavity of the *min-mutual information* as a functional of $p(S)$, and Jensen's inequality we conclude that $I_1^* \geq I_0^*$.

Since Θ_1 and Θ_2 are isotropically distributed $\Theta_1\Phi$ and $\Theta_2\Psi$ are also isotropically distributed when conditioned on Φ and Ψ respectively. This means that both $\Theta_1\Phi$ and $\Theta_2\Psi$ are isotropically distributed making them independent of Φ, V and Ψ. Therefore, S_1 is equal to the product of three independent matrices, a $T \times T$ unitary matrix Φ, a $T \times M$ real nonnegative matrix V and an $M \times M$ unitary matrix Ψ.

Now, it will be shown that the density $p(V)$ on V is unchanged by rearrangements of diagonal entries of V. There are $\min\{M!, T!\}$ ways of arranging the diagonal entries of V. This can be accomplished by pre- and post-multiplying V by appropriate permutation matrices, P_{Tk} and P_{Mk}, $k = 1, \ldots, \min\{M!, T!\}$. The permutation does not change the *min-mutual information* because ΦP_{Tk} and ΨP_{Mk} have the same density functions as Φ and Ψ. By choosing an equally weighted mixture density for V, involving all $\min\{M!, T!\}$ arrangements a higher value of *min-mutual information* can be obtained because of concavity and Jensen's inequality. This new density is invariant to the rearrangements of the diagonal elements of V. ■

V. Average Capacity Criterion

In this section, we will investigate how much worse the worst-case performance is compared to the average performance. To find the average performance, we maximize $I_E(X;S) = E_\alpha[I^\alpha(X;S)]$, where I^α is as defined earlier and E_α denotes expectation over $\alpha \in A$ under the assumption that all α are equally likely. That is, under the assumption that α is unchanging over time, isotropically random and known to the receiver. Note that this differs from the model considered in [13] where the authors consider the case of a piecewise constant, time varying, i.i.d. specular component.

Therefore, the problem can be stated as finding $p_E(S)$ the probability density function on the input signal S that achieves the following maximization

$$C_E = \max_{p(S)} E_\alpha[I^\alpha(X;S)] \tag{8}$$

and also to find the value C_E. We will refer to $I_E(X;S)$ as *avg-mutual information* and C_E as *avg-capacity*.

Like in the previous section, we would expect the optimum signal to be such that it generates the same mutual information irrespective of the choice of the specular component because the density function attaches no significance to any particular $\alpha \in A$. In other words, since the set of non-specular components A, and the density function on A are such that the density function is invariant under unitary transformations we would expect the optimum signal density to be invariant to unitary transformations as well. Moreover, since all $\alpha \in A$ are identical to each other except for unitary transformations, intuition tells us that Corollary 2 in the previous section should hold here also. Therefore, the average mutual information over all α should be equal to the mutual information for a single α. That is the average performance should be equal to the worst-case performance.

Formally, it will be shown that the signal density $p^*(S)$ that attains C^* also attains C_E. For that we need to establish the following Lemmas. We omit the proofs because the proofs are very similar to the proofs in Section IV.

Lemma 4: $I_E(X;S)$ is a concave functional of the signal density $p(S)$

Lemma 5: For any $T \times T$ unitary matrix Φ and any $M \times M$ unitary matrix Ψ, if $p(S)$ generates $I_E(X;S)$ then so does $p(\Phi S \Psi^\dagger)$.

Proof: We want to show if $p(S)$ generates $I_E(X;S)$ then so does $p(\Phi S\Psi^\dagger)$. Now since the density function of α, $p(\alpha) = \frac{\Gamma(M)}{\pi^M}\delta(\alpha^\dagger\alpha - 1)$ we have

$$I_E(X;S) = \frac{\pi^M}{\Gamma(M)} \int I^\alpha(X;S)\, d\alpha$$

Note that $I^\alpha(X;\Phi S) = I^\alpha(X;S)$ Therefore,

$$\begin{aligned} I'_E(X;S) &= \frac{\pi^M}{\Gamma(M)} \int I^\alpha(X;\Phi S\Psi^\dagger)\, d\alpha \\ &= \frac{\pi^M}{\Gamma(M)} \int I^\alpha(X;S\Psi^\dagger)\, d\alpha \end{aligned}$$

Also note that $I^{\Psi\alpha}(X;S\Psi^\dagger) = I^\alpha(X;S)$ which means $I^{\Psi^\dagger\alpha}(X;S) = I^\alpha(X;S\Psi^\dagger)$. Therefore,

$$\begin{aligned} I'_E(X;S) &= \frac{\pi^M}{\Gamma(M)} \int I^{\Psi^\dagger\alpha}(X;S)\, d\alpha \\ &= \frac{\pi^M}{\Gamma(M)} \int I^\omega(X;S)\, d\omega \\ &= I_E(X;S) \end{aligned}$$

where the last two equalities follow from the transformation $\omega = \Psi^\dagger\alpha$ and the fact the Jacobian of the transformation is equal to 1. ■

Lemma 6: The *avg-capacity* achieving signal distribution, $p(S)$ is unchanged by any pre- and post- multiplication of S by unitary matrices of appropriate dimensions.

Corollary 3: $p^*(S)$, the optimal *avg-capacity* achieving signal density lies in $\mathcal{P} = \cup_{I>0}\mathcal{P}_I$ where $\mathcal{P}_I$ is as defined in (7).

Based on the last corollary it can be concluded that for a given $p(S)$ in $\mathcal{P}$ we have $I^*(X;S) = \min_{\alpha \in A} I^\alpha(X;S) = E_\alpha[I^\alpha(X;S)] = I_E(X;S)$. Therefore, the maximizing densities for C_E and C^* are the same and also $C_E = C^*$. Therefore, designing the signal constellation with the objective of maximizing the worst-case performance is not more pessimistic than maximizing the average performance.

Finally, we have the following theorem similar to Theorem 2 in the previous section.

Theorem 3: The signal matrix that achieves *avg-capacity* can be written as $S = \Phi V \Psi^\dagger$, where Φ and Ψ are $T \times T$ and $M \times M$ isotropically distributed matrices independent of each other, and V is a $T \times M$ real, nonnegative, diagonal matrix, independent of both Φ and Ψ.

VI. Numerical Results

Plotting the upper and lower bounds on *min-capacity* leads to similar conclusions as in [13] except for the fact when $r = 1$ the upper and lower bounds coincide. A tighter lower bound can be obtained by first observing that a lower bound on *avg-capacity* is also a lower bound on *min-capacity* and then optimizing over the number of transmit and receive antennas. Therefore,

$$\begin{aligned} C^* &= C_E \\ &\geq \max_{m \leq M, n \leq N} \Bigg\{ T E_{\alpha'} \left[E \log_2 \det \left(I_n + \frac{\rho}{m} H_{\alpha'}^\dagger H_{\alpha'} \right) \right] \\ &\qquad - nE \left[\log_2 \det \left(I_T + (1-r)\frac{\rho}{m} S_m S_m^\dagger \right) \right] \Bigg\}. \end{aligned}$$

In the expression above the expectation in the second term is over the distribution of the $T \times m$ input signal matrix S_m. The outer expectation in the first term is over the distribution of $m \times 1$ vector α' and the inner expectation is over the distribution of the Rayleigh component of the $m \times n$ channel matrix $H_{\alpha'}$. $H_{\alpha'}$ is obtained from H_α by selecting the first $m \times n$ block of the original $M \times N$ channel matrix H_α. Therefore, α' is simply the vector of first m elements of the $M \times 1$ vector α.

In Figure 1 the *min-capacity* upper and lower bounds have been plotted as a function of the Ricean parameter r. It can be seen that the change in capacity is not drastic for low SNR as compared to larger SNR values. Also, from Figure 2 we conclude that this change in capacity is more prominent for larger number of antennas. We also conclude that for a purely specular channel increasing the number of transmit antennas has no effect on the capacity. This is due to the fact that with a rank-one specular component, only beamforming SNR gains can be exploited, no multiplexing gains are possible.

VII. Discussion and Conclusion

The idea of maximizing *min-capacity* can be traced back to [5], [7] where intuitive arguments were used to justify the choice of identity matrix as the optimum input signal covariance matrix. In both works the channel is assumed to be known at the receiver hence the optimum signal is a Gaussian signal with only its covariance matrix to be determined. Choosing the covariance matrix to be identity amounts to choosing a unitarily invariant density on the input signal.

The model considered in this paper can be extended in various ways. One way would be to assume the parameter r to be not fixed. In this case, for *min-capacity* in addition to minimizing over $\alpha \in A$ we would also have to minimize over the parameter $r \in [0, 1]$. For *avg-capacity* the average over r needs to be taken as well assuming a prior distribution like the uniform distribution over $[0, 1]$ on r. In both cases, the optimum signal density would still be unitarily invariant. However, *min-capacity* will no longer be equal to *avg-capacity*.

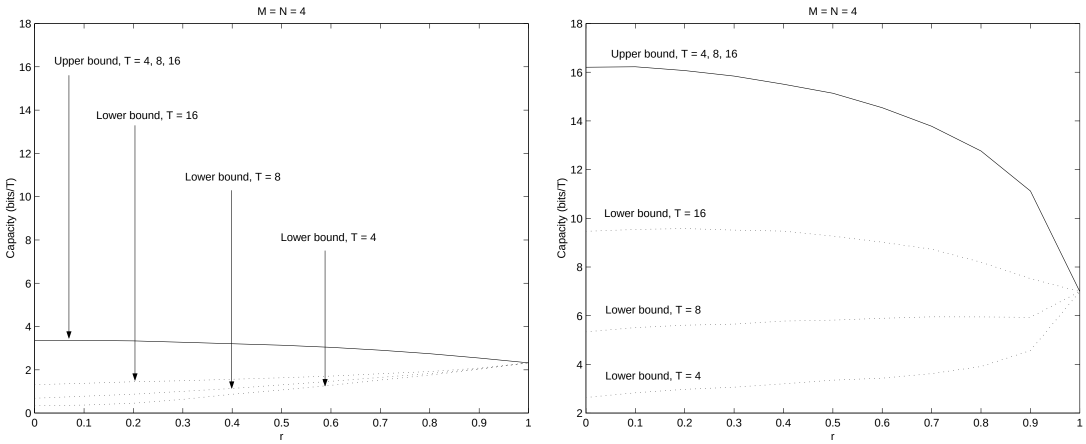
Another extension would be to assume the specular component to be composed of L rank-one components, that is

$$H = \sqrt{1-r}G + \sqrt{rNM}\sqrt{\frac{1}{L}}\sum_{l=1}^{L} v_l \alpha_l \beta_l^\dagger$$

where $v_l \geq 0$ with $\sum_{l=1}^{L} v_l = 1$ and α_l and β_l such that $\alpha_l \in \{\alpha : \alpha^\dagger\alpha = 1\}$ and $\beta_l \in \{\beta : \beta^\dagger\beta = 1\}$. Here also, the optimum signal structure turns out to be unitarily invariant with $C_E > C^*$.

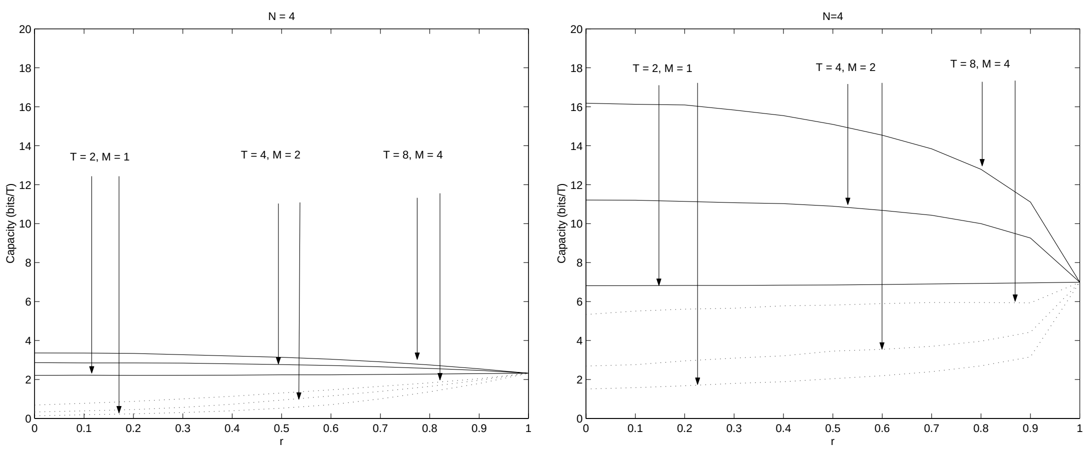
The summary of contributions in this paper is as follows. A non-ergodic but tractable model for Ricean fading channel different from the one in [13] but, along the lines of [5] has been proposed. For this channel, the worst case capacity was computed and it was shown that the optimal signal structure was unitarily invariant. As a result, the optimization effort is over a much smaller set of parameters of size $\min\{T, M\}$ than the set of size $T \times M$ originally begun with. Since the capacity is not in closed form a useful lower bound that illustrates the capacity trends as a function of the parameter r was derived.

Finally, it was shown that the approach of maximizing the worst case scenario is not pessimistic in the sense



(a) $\rho = 0dB$ (b) $\rho = 15dB$

Fig. 1. Capacity upper and lower bounds as the channel moves from purely Rayleigh to purely Ricean fading



(a) $\rho = 0dB$ (b) $\rho = 15dB$

Fig. 2. Capacity upper and lower bounds as the channel moves from purely Rayleigh to purely Ricean fading

that the signal density maximizing the worst-case performance also maximizes the average performance and the capacity value in both formulations turns out to be the same. The average capacity being equal to the worst-case capacity can also be interpreted in a different manner: it has been shown that the average capacity criterion is a quality of service guaranteeing capacity.

APPENDIX

Coding Theorem for *Min-capacity*]

To make this paper self-sufficient, we will prove the following theorem that is specific to the compound channel considered here. To understand the theorem the reader is not required to know the material in [2].

Theorem 4: For the quasi-static Ricean fading model, for every $R < C^*$ there exists a sequence of $(2^{nR}, n)$ codes with codewords, m_i^n, $i = 1, \dots, 2^{nR}$, satisfying the power constraint such that

$$\lim_{n\to\infty} \sup_{\alpha} P_{e\alpha,n} = 0$$

where $P_{e\alpha,n} = \max_{i=1}^{2^{nR}} P_e(m_i^n, \alpha)$ and $P_e(m_i)$ is the probability of incorrectly decoding the messages m_i when the channel is given by H_α.

Proof: Proof follows if we can show that $P_{e\alpha,n}$ is bounded above by the same Gallager error exponent [8], [9] irrespective of the value of α. That follows from the following lemma (Lemma 7). ■

The intuition behind the existence of a coding theorem is that the *min-capacity* C^* achieving signal density is such that the mutual information, C^*, between the output and the input is the same irrespective of any particular realization of the channel H_α. Therefore, any codes generated from the random coding argument designed to achieve rates up to C^* for any particular channel H_α achieve rates up to C^* for all H_α.

For Lemma 7, we first need to briefly describe the Gallager error exponents [8], [9] for the quasi-static Ricean fading channel. For a system communicating at a rate R the upper bound on the maximum probability of error is given as follows

$$P_{e\alpha,n} \leq \exp\left(-n \max_{p(S)} \max_{0\leq\gamma\leq 1} [E_0(\gamma, p(S), \alpha) - \gamma R \log 2]\right)$$

where n is the length of the codewords in the codebook used and $E_0(\gamma, p(S), \alpha)$ is as follows

$$E_0(\gamma, p(S), \alpha) = -\log \int \left[\int p(S)p(X|S,\alpha)^{\frac{1}{1+\gamma}} dS\right]^\gamma dX$$

where S is the input to the channel and X is the observed output and

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

where β is simply $[1\ 0\ \ldots\ 0]^\tau$. Maximization over γ in the error exponent yields a value of γ such that $\frac{\partial E_0(\gamma,p(S),\alpha)}{\partial\gamma} = R$. Note that for $\gamma = 0$, $\frac{\partial E_0(\gamma,p(S),\alpha)}{\partial\gamma} = I^\alpha(X;S)$ [8], [9] where the mutual information has been evaluated when the input is $p(S)$. If $p(S)$ is the *min-capacity* achieving density, $p^*(S)$ then $\frac{\partial E_0(\gamma,p^*(S),\alpha)}{\partial\gamma} = C^*$. For more information refer to [8], [9].

Lemma 7: The $E_0(\gamma, p^*(S), \alpha)$ for the quasi-static Ricean fading model is independent of α.

Proof: First, note that

$$p^*(S) = p^*(S\Psi^\dagger)$$

for any $M \times M$ unitary matrix Ψ. Second,

$$\begin{aligned} E_0&(\gamma, p^*(S), \alpha) \\ &= -\log \int \left[\int p^*(S)p(X|S,\alpha)^{\frac{1}{1+\gamma}} dS\right]^\gamma dX \\ &= -\log \int \left[\int p^*(S\Psi^\dagger)p(X|S\Psi^\dagger,\alpha)^{\frac{1}{1+\gamma}} dS\right]^\gamma dX \\ &= -\log \int \left[\int p^*(S)p(X|S,\Psi^\dagger\alpha)^{\frac{1}{1+\gamma}} dS\right]^\gamma dX \\ &= E_0(\gamma, p^*(S), \Psi^\dagger\alpha) \end{aligned}$$

where the second equation follows from the fact that Ψ is a unitary matrix and its Jacobian is equal to 1 and the third equation follows from the fact that $p(X|S\Psi^\dagger,\alpha)^{\frac{1}{1+\gamma}} = p(X|S,\Psi^\dagger\alpha)^{\frac{1}{1+\gamma}}$. Since Ψ is arbitrary we obtain that $E_0(\gamma, p^*(S), \alpha)$ is independent of α. ■

REFERENCES

[1] 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.

[2] I. Csiszár and J. Kőrner, *Information Theory: Coding Theorems For Discrete Memoryless Systems*, New York and NY: Academic Press, 1981.

[3] H. Dai, A. F. Molisch, and H. V. Poor, "Downlink multiuser capacity of interference-limited mimo systems," *Proc. IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, vol. 2, pp. 849–853, Sep. 2002.

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

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

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

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

[8] R. G. Gallager, *Information Theory and Reliable Communication*, Wiley, New York, 1968.

[9] R. G. Gallager, "A simple derivation of the coding theorem and some applications," *IEEE Trans. on Inform. Theory*, vol. 11, no. 1, pp. 3–18, Jan. 1965.

[10] D. Gesbert, "Robust linear mimo receivers: A minimum error-rate approach," *Signal Processing*, vol. 51, no. 11, pp. 2863–2871, Nov. 2003.

[11] D. Gesbert, H. Bőlcskei, D. A. Gore, and A. J. Paulraj, "Outdoor mimo wireless channels: Models and performance prediction," *IEEE Trans. on Communications*, vol. 50, no. 12, pp. 1926–1934, Dec. 2002.

[12] M. Godavarti and A. O. Hero-III, "Multiple antenna capacity in a deterministic Rician fading channel," *Submitted to IEEE Trans. on Info. Theory*, 2001.

[13] M. Godavarti, T. Marzetta, and S. S. Shitz, "Capacity of a mobile multiple-antenna wireless link with isotropically random Rician fading," *IEEE Trans. on Inform. Theory*, vol. 49, no. 12, pp. 3330–3334, December 2003.

[14] J. Hansen and H. Bőlcskei, "A geometrical investigation of the rank-1 ricean mimo channel at high snr," *Submitted to ISIT 2004*.

[15] S. K. Jayaweera and H. V. Poor, "On the capacity of multi-antenna systems in the presence of rician fading," *Proc. of the IEEE Vehicular Technology Conference*, vol. 4, pp. 1963–1967, Sep. 2002.

[16] 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.

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

[18] S. Sandhu and A. Paulraj, "Space-time block codes: A capacity perspective," *IEEE Communications Letters*, vol. 4, no. 12, pp. 384–386, December 2000.

[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. on Communications*, vol. 44, no. 2, pp. 744–765, March 1998.

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

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