

BEAMFORMING MAXIMIZES THE RANK ONE RICEAN MIMO COMPOUND CAPACITY

Ami Wiesel, Yonina C. Eldar and Shlomo Shamai (Shitz)

Department of Electrical Engineering
Technion – Israel Institute of Technology
Technion City, Haifa 32000, Israel.

ABSTRACT

In this paper, we consider the optimization of the compound capacity in a rank one Ricean multiple input multiple output channels using partial channel state information at the transmitter side. We model the channel as a deterministic matrix within a known ellipsoid, and maximize the compound capacity defined as the worst case capacity within this set. We find that the optimal transmit strategy is always beamforming, and can be found using a simple one dimensional search. These results motivate the growing use of systems using simple beamforming transmit strategies.

1. INTRODUCTION

The use of multiple transmit and receive antennas is known to improve the capacity and reliability of wireless communication links. The two common techniques for exploiting this multiple input multiple output (MIMO) channel are space time coding, and MIMO precoding. Space time coding is a technique that allows for spatial diversity without any channel state information (CSI) in the transmitter. On the other hand, when perfect CSI is available, the standard technique is to use MIMO precoding algorithms, such as beamforming. These two strategies are based on two extreme assumptions on the availability of CSI at the transmitter side. In many practical applications only partial CSI is available, in which case it is not clear what the optimal transmit strategy is.

The capacity achieving transmit technique in MIMO channels with additive Gaussian noise is signaling using random Gaussian vectors. The strategy is therefore defined by the covariance matrix of these vectors. The eigenvectors of this matrix can be visualized as the *directions* in which the transmitter signals. Due to its importance, the optimization of the covariance has been extensively studied. Different optimization criteria were considered, as well as different models for the CSI. Most of the research in this area is devoted to stochastic models of the CSI, i.e., scenarios in which the transmitter has access to the statistics of the channel. Typically, the channel is modelled as a complex normal random vector with known mean and covariance. In this stochastic CSI model, the mutual information is also a random quantity and must be treated appropriately, either by considering its ensemble average known as the *ergodic capacity*, or by considering its cumulative distribution function (CDF) via the *outage mutual information*. One of the first papers in this field is [1] where a multiple input single output (MISO) channel was considered. In this work, the structure of the

optimal transmit strategy in the sense of maximizing the ergodic capacity was derived. The basic result was that if a non zero mean is available then the optimal strategy is to transmit in its direction and uniformly in all other directions. If a non trivial covariance matrix is available, then the optimal strategy is to transmit along its eigenvectors. An extension of this work was presented in [2] where conditions for the optimality of beamforming (BF) in rank one Ricean MIMO systems were found in closed form. In [3,4] the ergodic capacity and the outage mutual information were derived analytically and the optimal transmit strategies were found numerically. One of the interesting results was that a system which switches between BF and uniformly transmitting in all directions is close to optimal. The impact of correlation between the antennas and more details on the optimal power allocation strategies for maximizing the ergodic capacity were discussed in [5]. Recently, in [6], the outage capacity with no available CSI was analyzed. A competing stochastic CSI model was introduced in [7] where the channel was modelled using the probability distribution function of the phase shifts between the antennas. Similarly to the previous references, here too the ergodic capacity was optimized.

A different approach for describing partial CSI is using a deterministic model for the channel, i.e., assuming that the channel is a deterministic variable within a known set of possible values. When the set is a singleton, the CSI is complete and perfect. The bigger the set is, the more uncertainty there is on the actual realization of the channel. The use of deterministic CSI models is common in the signal processing community for designing algorithms which are robust to the uncertainty [8,9]. In the context of information theory, the maximal achievable rate of reliable communication over such channels is the *compound capacity* and is defined as the capacity of the worst case realization within the set [10] (See also [11] for a tutorial on the topic). A possible application is in communication through a slow fading channel. In such channels, the system cannot average over the realizations of the channel, and must cope with the specific realization. Assuming a strict constraint on the quality of service, the system must be designed for the worst case scenario. In this sense, compound capacity is related to outage mutual information which also aims at designing communication systems over slow fading channels. More details on the compound capacity and its relation to the outage capacity and other information theoretic notions can be found in [12]. Fortunately, the compound capacity is much easier to handle than the cumbersome outage capacity. For completeness, we mention that the compound capacity is also related to the problem of optimizing the capacity of the worst case noise covariance [13].

Due to its importance, the compound MIMO capacity recently gained a considerable attention. In [14,15] it was shown that under different uncertainty sets the optimal transmit strategy is uni-

The authors were supported by the EU 6th framework programme, via the NEWCOM network of excellence. This research was also supported by ISRAEL SCIENCE FOUNDATION founded by the Israel Academy of Sciences and Humanities.

form power allocation. However, the uncertainty sets used in these papers are very different from the structure of the stochastic CSI models used in [1–3]. Therefore, the results are different and it is difficult to compare these two classes of works. In another work [16], the compound capacity was analyzed and bounded under a rank one Ricean MIMO model when the specular component was unknown. It was shown that if this component is random with an isotropic distribution then the compound capacity is equal to the average capacity.

In our work, we follow the deterministic approach, but use an uncertainty set with a structure which is very similar to the CSI model used in [1–3]. We model the rank one Ricean MIMO channel as a matrix within a known ellipsoid defined using the deterministic analogs of the channel's mean and covariance. We find that the optimal transmit strategy for maximizing the compound capacity in such CSI models is always BF. If the ellipsoid is symmetric with respect to its center (mean), then the optimal direction is the right singular vector of the center (mean) matrix. In more general scenarios, we provide a simple strategy for finding the optimal direction based on a one dimensional search. These results motivate the growing use of simple BF transmit strategies.

The paper is organized as follows. We begin in Section 2 by defining our channel model and introducing the compound optimization problem. In Section 3, we provide our main result in Theorem 1 and discuss its consequences. The connection between our work and previous works based on stochastic CSI models is addressed in Section 4. In particular, we discuss the relation between the compound capacity and the outage mutual information. In Section 5, we illustrate our results using a simple numerical example.

The following notation is used. Boldface upper case letters denote matrices, boldface lower case letters denote column vectors, and standard lower case letters denote scalars. The superscripts $(\cdot)^T$ and $(\cdot)^{-1}$ denote the transpose and the matrix inverse operators, respectively. $[\mathbf{x}]_i$ denotes the i 'th element of the vector \mathbf{x} . By $\text{Tr}\{\cdot\}$ we denote the trace operator, by $\text{vec}(\cdot)$ we denote the operator that stacks the elements of a matrix into a single column vector, and by \mathbf{I} we denote the identity matrix of appropriate size. \otimes denotes the Kronecker product, $|\cdot|$ denotes the determinant, $\|\cdot\|$ denotes the standard Euclidean norm. Finally, $\mathbf{X} \succeq 0$ means that \mathbf{X} is a Hermitian positive semidefinite matrix.

2. PROBLEM FORMULATION

Consider the following MIMO channel model:

$$\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{w}, \quad (1)$$

where \mathbf{y} is a received vector of length N , \mathbf{H} is a size $N \times K$ channel matrix, \mathbf{s} is a length K transmitted random vector of covariance $E\{\mathbf{s}\mathbf{s}^T\} = \mathbf{Q}$ satisfying $\text{Tr}\{\mathbf{Q}\} \leq P$, and \mathbf{w} is a length N Gaussian noise vector of covariance $\sigma^2\mathbf{I}$. We model the rank one Ricean MIMO channel \mathbf{H} as an unknown deterministic matrix within the following set

$$\mathbf{H} = \mathbf{a}\mathbf{x}^T + \mathbf{D}; \quad \text{Tr}\{\mathbf{D}\mathbf{W}\mathbf{D}^T\} \leq 1, \quad (2)$$

where \mathbf{a} is a length N vector, \mathbf{x} is a length K vector, and $\mathbf{W} \succeq 0$ is a weight matrix. In our terminology, $\mathbf{a}\mathbf{x}^T$ is the rank one specular component of the channel, and \mathbf{D} is the scattering component. We assume that the transmitter knows $\mathbf{a}\mathbf{x}^T$ and \mathbf{W} , but does not have access to the specific realization of \mathbf{H} within the set. In Section

4, we will show that this CSI model is the deterministic analog of the stochastic CSI model used in [1–3], where $\mathbf{a}\mathbf{x}^T$ is the rank one mean channel and \mathbf{W} is related to its covariance.

A classical result in information theory states that the following compound capacity is the maximal achievable rate of reliable communication over the above channel [10, 11, 16]:

$$C(\mathbf{a}\mathbf{x}^T, \mathbf{W}) = \max_{\substack{\mathbf{Q} \succeq 0 \\ \text{Tr}\{\mathbf{Q}\} \leq P}} \min_{\text{Tr}\{\mathbf{D}\mathbf{W}\mathbf{D}^T\} \leq 1} I(\mathbf{Q}, \mathbf{D}), \quad (3)$$

where

$$I(\mathbf{Q}, \mathbf{D}) = \log \left| \mathbf{I} + \frac{1}{\sigma^2} (\mathbf{a}\mathbf{x}^T + \mathbf{D}) \mathbf{Q} (\mathbf{a}\mathbf{x}^T + \mathbf{D})^T \right|, \quad (4)$$

is the mutual information between \mathbf{y} and \mathbf{s} . It can be achieved by signaling with Gaussian vectors \mathbf{s} of covariance $E\{\mathbf{s}\mathbf{s}^T\} = \mathbf{Q} \succeq 0$. BF is defined as the transmit strategy when $\mathbf{Q} = \mathbf{q}\mathbf{q}^T$ is rank one, i.e., $\mathbf{s} = s\mathbf{q}$ where s is a Gaussian random variable.

3. OPTIMIZATION OF THE COMPOUND CAPACITY

In this section, we provide our main results in the following theorem:

Theorem 1. Consider the optimization of the rank one Ricean MIMO compound capacity of $C(\mathbf{a}\mathbf{x}^T, \mathbf{W})$ in (3). If $\mathbf{x}^T \mathbf{W} \mathbf{x} \leq \frac{1}{\|\mathbf{a}\|^2}$ then $C(\mathbf{a}\mathbf{x}^T, \mathbf{W}) = 0$ and any feasible \mathbf{Q} will attain it. Otherwise, its optimal value is

$$C(\mathbf{a}\mathbf{x}^T, \mathbf{W}) = \log \left(1 + \frac{\|\mathbf{a}\|^2}{\sigma^2} c \right), \quad (5)$$

where

$$c = P \mathbf{x}^T \left[\mathbf{I} - \left(\mathbf{I} + \frac{\lambda}{P} \mathbf{W} \right)^{-1} \right]^2 \mathbf{x}, \quad (6)$$

and $\lambda > 0$ is the unique root of the following non linear equation

$$\mathbf{x}^T \left(\mathbf{I} + \frac{\lambda}{P} \mathbf{W} \right)^{-1} \mathbf{W} \left(\mathbf{I} + \frac{\lambda}{P} \mathbf{W} \right)^{-1} \mathbf{x} = \frac{1}{\|\mathbf{a}\|^2}. \quad (7)$$

In this case, the optimal \mathbf{Q} is

$$\mathbf{Q} = P \frac{\mathbf{q}(\lambda) \mathbf{q}^T(\lambda)}{\|\mathbf{q}(\lambda)\|^2}, \quad (8)$$

where

$$\mathbf{q}(\lambda) = \left[\mathbf{I} - \left(\mathbf{I} + \frac{\lambda}{P} \mathbf{W} \right)^{-1} \right] \mathbf{x}. \quad (9)$$

Proof. We begin the proof by showing that the optimal argument \mathbf{D} of the inner minimization in (3) has the structure $\mathbf{D} = \mathbf{a}\mathbf{d}^T$ for some \mathbf{d} . This will allow us to optimize over the vector \mathbf{d} instead of the matrix \mathbf{D} . In particular, we prove that if \mathbf{D} is optimal then $\tilde{\mathbf{D}} = \mathbf{a}\mathbf{d}^T$ with $\mathbf{d}^T = \frac{\mathbf{a}^T}{\|\mathbf{a}\|^2} \mathbf{D}$ is also optimal. Now, assume that \mathbf{D} is feasible, then $\tilde{\mathbf{D}}$ is also feasible since

$$\begin{aligned} \text{Tr}\{\tilde{\mathbf{D}}\mathbf{W}\tilde{\mathbf{D}}^T\} &= \text{Tr}\left\{ \mathbf{a} \frac{\mathbf{a}^T}{\|\mathbf{a}\|^2} \mathbf{D}\mathbf{W}\mathbf{D}^T \frac{\mathbf{a}}{\|\mathbf{a}\|^2} \mathbf{a}^T \right\} \\ &= \frac{\mathbf{a}^T \mathbf{D}\mathbf{W}\mathbf{D}^T \mathbf{a}}{\mathbf{a}^T \mathbf{a}} \leq \max_{\mathbf{v} \neq 0} \frac{\mathbf{v}^T \mathbf{D}\mathbf{W}\mathbf{D}^T \mathbf{v}}{\mathbf{v}^T \mathbf{v}} \\ &= \lambda_{\max}(\mathbf{D}\mathbf{W}\mathbf{D}^T) \leq \text{Tr}\{\mathbf{D}\mathbf{W}\mathbf{D}^T\} \leq 1, \end{aligned} \quad (10)$$

where $\lambda_{\max}(\mathbf{A})$ is the maximal eigenvalue of \mathbf{A} , and we used the identity $\lambda_{\max}(\mathbf{A}) = \max_{\mathbf{v} \neq \mathbf{0}} \frac{\mathbf{v}^T \mathbf{A} \mathbf{v}}{\mathbf{v}^T \mathbf{v}}$. In addition, $\tilde{\mathbf{D}}$ results in an equal or better objective value than that of \mathbf{D} since

$$\begin{aligned} & \left| \mathbf{I} + \frac{1}{\sigma^2} (\mathbf{a}\mathbf{x}^T + \mathbf{D}) \mathbf{Q} (\mathbf{a}\mathbf{x}^T + \mathbf{D})^T \right| \\ &= \prod_{i=1}^N \left[1 + \frac{1}{\sigma^2} \lambda_i \left((\mathbf{a}\mathbf{x}^T + \mathbf{D}) \mathbf{Q} (\mathbf{a}\mathbf{x}^T + \mathbf{D})^T \right) \right] \\ &\geq 1 + \frac{1}{\sigma^2} \lambda_{\max} \left((\mathbf{a}\mathbf{x}^T + \mathbf{D}) \mathbf{Q} (\mathbf{a}\mathbf{x}^T + \mathbf{D})^T \right) \\ &= 1 + \frac{1}{\sigma^2} \max_{\mathbf{v} \neq \mathbf{0}} \frac{\mathbf{v}^T (\mathbf{a}\mathbf{x}^T + \mathbf{D}) \mathbf{Q} (\mathbf{a}\mathbf{x}^T + \mathbf{D})^T \mathbf{v}}{\mathbf{v}^T \mathbf{v}} \\ &\geq 1 + \frac{1}{\sigma^2} \frac{\mathbf{a}^T (\mathbf{a}\mathbf{x}^T + \mathbf{D}) \mathbf{Q} (\mathbf{a}\mathbf{x}^T + \mathbf{D})^T \mathbf{a}}{\mathbf{a}^T \mathbf{a}} \\ &= 1 + \frac{\mathbf{a}^T \mathbf{a}}{\sigma^2} \left(\mathbf{x}^T + \frac{\mathbf{a}^T \mathbf{D}}{\|\mathbf{a}\|^2} \right) \mathbf{Q} \left(\mathbf{x} + \mathbf{D}^T \frac{\mathbf{a}}{\|\mathbf{a}\|^2} \right) \\ &= \left| \mathbf{I} + \frac{1}{\sigma^2} \left(\mathbf{a}\mathbf{x}^T + \mathbf{a} \frac{\mathbf{a}^T \mathbf{D}}{\|\mathbf{a}\|^2} \right) \mathbf{Q} \left(\mathbf{a}\mathbf{x}^T + \mathbf{a} \frac{\mathbf{a}^T \mathbf{D}}{\|\mathbf{a}\|^2} \right)^T \right| \\ &= \left| \mathbf{I} + \frac{1}{\sigma^2} (\mathbf{a}\mathbf{x}^T + \tilde{\mathbf{D}}) \mathbf{Q} (\mathbf{a}\mathbf{x}^T + \tilde{\mathbf{D}})^T \right|, \end{aligned} \quad (11)$$

where $\lambda_i(\mathbf{A})$ are the eigenvalues of \mathbf{A} , and we used the identities $|\mathbf{A}| = \prod_i \lambda_i(\mathbf{A})$, $\lambda_{\max}(\mathbf{A}) = \max_{\mathbf{v} \neq \mathbf{0}} \frac{\mathbf{v}^T \mathbf{A} \mathbf{v}}{\mathbf{v}^T \mathbf{v}}$ and $|\mathbf{I} + \mathbf{A}\mathbf{B}| = |\mathbf{I} + \mathbf{B}\mathbf{A}|$.

Therefore, solving (3) reduces to the following program

$$\begin{aligned} & \max_{\substack{\mathbf{Q} \succeq \mathbf{0} \\ \text{Tr}\{\mathbf{Q}\} \leq P}} \min_{\text{Tr}\{\mathbf{a}\mathbf{d}^T \mathbf{W} \mathbf{d} \mathbf{a}^T\} \leq 1} \\ & \log \left| \mathbf{I} + \frac{1}{\sigma^2} (\mathbf{a}\mathbf{x}^T + \mathbf{a}\mathbf{d}^T) \mathbf{Q} (\mathbf{a}\mathbf{x}^T + \mathbf{a}\mathbf{d}^T)^T \right|, \end{aligned} \quad (12)$$

or using $|\mathbf{I} + \mathbf{A}\mathbf{B}| = |\mathbf{I} + \mathbf{B}\mathbf{A}|$ to

$$\max_{\substack{\mathbf{Q} \succeq \mathbf{0} \\ \text{Tr}\{\mathbf{Q}\} \leq P}} \min_{\mathbf{d}^T \mathbf{W} \mathbf{d} \leq \frac{1}{\|\mathbf{a}\|^2}} \log \left(1 + \frac{\|\mathbf{a}\|^2}{\sigma^2} (\mathbf{x} + \mathbf{d})^T \mathbf{Q} (\mathbf{x} + \mathbf{d}) \right). \quad (13)$$

If $\mathbf{x}^T \mathbf{W} \mathbf{x} \leq \frac{1}{\|\mathbf{a}\|^2}$ then $\mathbf{d} = -\mathbf{x}$, $C(\mathbf{a}\mathbf{x}^T, \mathbf{W}) = 0$ and any feasible \mathbf{Q} will attain it. We now consider the case when $\mathbf{x}^T \mathbf{W} \mathbf{x} > \frac{1}{\|\mathbf{a}\|^2}$. Due to the monotonicity of the objective function in the quadratic form $(\mathbf{x} + \mathbf{d})^T \mathbf{Q} (\mathbf{x} + \mathbf{d})$, we can optimize it instead:

$$\max_{\substack{\mathbf{Q} \succeq \mathbf{0} \\ \text{Tr}\{\mathbf{Q}\} \leq P}} \min_{\mathbf{d}^T \mathbf{W} \mathbf{d} \leq \frac{1}{\|\mathbf{a}\|^2}} (\mathbf{x} + \mathbf{d})^T \mathbf{Q} (\mathbf{x} + \mathbf{d}). \quad (14)$$

It is easy to see that the objective is convex in \mathbf{d} and concave (linear) in \mathbf{Q} . Moreover, the constraint set of the minimization is convex, and the constraint set of the maximization is convex and compact. Therefore, minimax theory [17] states that there is a saddle point, i.e., a point $\{\bar{\mathbf{d}}, \bar{\mathbf{Q}}\}$ such that $\bar{\mathbf{Q}}$ solves the problem

$$\max_{\substack{\mathbf{Q} \succeq \mathbf{0} \\ \text{Tr}\{\mathbf{Q}\} \leq P}} (\mathbf{x} + \bar{\mathbf{d}})^T \mathbf{Q} (\mathbf{x} + \bar{\mathbf{d}}), \quad (15)$$

and $\bar{\mathbf{d}}$ solves the problem

$$\min_{\mathbf{d}^T \mathbf{W} \mathbf{d} \leq \frac{1}{\|\mathbf{a}\|^2}} (\mathbf{x} + \mathbf{d})^T \bar{\mathbf{Q}} (\mathbf{x} + \mathbf{d}). \quad (16)$$

The Lagrangian associated with program (15) is

$$L_1 = -(\mathbf{x} + \bar{\mathbf{d}})^T \mathbf{Q} (\mathbf{x} + \bar{\mathbf{d}}) - \text{Tr}\{\mathbf{Y}\mathbf{Q}\} + \nu [\text{Tr}\{\mathbf{Q}\} - P], \quad (17)$$

where $\mathbf{Y} \succeq \mathbf{0}$ and $\nu \geq 0$ are the dual variables. The matrix $\bar{\mathbf{Q}}$ is optimal if and only if it satisfies:

$$\begin{aligned} & -(\mathbf{x} + \bar{\mathbf{d}}) (\mathbf{x} + \bar{\mathbf{d}})^T - \mathbf{Y} + \nu \mathbf{I} = \mathbf{0}; \\ & \text{Tr}\{\mathbf{Y}\bar{\mathbf{Q}}\} = 0; \\ & \nu [\text{Tr}\{\bar{\mathbf{Q}}\} - P] = 0. \end{aligned} \quad (18)$$

It is easy to check that

$$\bar{\mathbf{Q}} = P \frac{(\mathbf{x} + \bar{\mathbf{d}}) (\mathbf{x} + \bar{\mathbf{d}})^T}{\|\mathbf{x} + \bar{\mathbf{d}}\|^2}, \quad (19)$$

along with

$$\begin{aligned} \mathbf{Y} &= \|\mathbf{x} + \bar{\mathbf{d}}\|^2 \mathbf{I} - (\mathbf{x} + \bar{\mathbf{d}}) (\mathbf{x} + \bar{\mathbf{d}})^T; \\ \nu &= \|\mathbf{x} + \bar{\mathbf{d}}\|^2, \end{aligned} \quad (20)$$

satisfy these exact conditions. In addition, the saddle point must satisfy the optimality conditions associated with program (16). The Lagrangian of this problem is

$$L_2 = (\mathbf{x} + \mathbf{d})^T \bar{\mathbf{Q}} (\mathbf{x} + \mathbf{d}) + \lambda \left[\mathbf{d}^T \mathbf{W} \mathbf{d} - \frac{1}{\|\mathbf{a}\|^2} \right], \quad (21)$$

where $\lambda \geq 0$ is a Lagrange multiplier. The necessary and sufficient optimality conditions are

$$(\bar{\mathbf{Q}} + \lambda \mathbf{W}) \bar{\mathbf{d}} = -\bar{\mathbf{Q}} \mathbf{x}; \quad (22)$$

$$\lambda \left[\bar{\mathbf{d}}^T \mathbf{W} \bar{\mathbf{d}} - \frac{1}{\|\mathbf{a}\|^2} \right] = 0. \quad (23)$$

Plugging $\bar{\mathbf{Q}}$ from (19) into (22) results in:

$$P (\mathbf{x} + \bar{\mathbf{d}}) + \lambda \mathbf{W} \bar{\mathbf{d}} = \mathbf{0}. \quad (24)$$

Solving for $\bar{\mathbf{d}}$ yields

$$\bar{\mathbf{d}} = - \left(\mathbf{I} + \frac{\lambda}{P} \mathbf{W} \right)^{-1} \mathbf{x}. \quad (25)$$

Due to $\mathbf{x}^T \mathbf{W} \mathbf{x} > \frac{1}{\|\mathbf{a}\|^2}$, the optimal multiplier $\lambda > 0$ is strictly positive. Therefore, $\bar{\mathbf{d}}$ must satisfy the complementary slackness condition:

$$\mathbf{x}^T \left(\mathbf{I} + \frac{\lambda}{P} \mathbf{W} \right)^{-1} \mathbf{W} \left(\mathbf{I} + \frac{\lambda}{P} \mathbf{W} \right)^{-1} \mathbf{x} = \frac{1}{\|\mathbf{a}\|^2}. \quad (26)$$

It is easy to see that the left hand side of (26) is monotonically decreasing in λ from $\mathbf{x}^T \mathbf{W} \mathbf{x} > \frac{1}{\|\mathbf{a}\|^2}$ when $\lambda = 0$ to 0 when $\lambda \rightarrow \infty$. Therefore, a unique solution for λ in (26) always exists. Finally, plugging the optimal $\bar{\mathbf{d}}$ and $\bar{\mathbf{Q}}$ into (14) and (19) yields (6) and (8), respectively. This concludes the proof. \square

The main result of Theorem 1 is that the optimal transmit strategy for maximizing the compound capacity in our model is always BF in the direction of $\mathbf{q}(\lambda)$ in (9). This direction is defined by the λ which satisfies (7). Finding this λ is very easy. Using the

eigenvalue decomposition of $\mathbf{W} = \mathbf{U} \text{diag} \{ \delta_i \} \mathbf{U}^T$ we rewrite the condition as

$$\sum_i \frac{P^2 \delta_i}{(P + \lambda \delta_i)^2} \left(\left[\mathbf{U}^T \mathbf{x} \right]_i \right)^2 = \frac{1}{\|\mathbf{a}\|^2}. \quad (27)$$

As explained in the proof, the left hand side of (27) is monotonically decreasing in $\lambda \geq 0$. Therefore, any simple bisection can efficiently solve for λ . Moreover, (27) belongs to a well known family of non linear equations called *secular equations* for which there are highly efficient root finding algorithms [18].

An important practical case is when the optimal BF is in the direction of \mathbf{x} . This is probably the standard technique in many applications due to its simplicity. Theorem 1 shows that this strategy is optimal if \mathbf{W} has $\frac{\mathbf{x}}{\|\mathbf{x}\|}$ as an eigenvector, since in this case $\mathbf{q}(\lambda)$ is a scaled version of \mathbf{x} for all λ s. A common example where this condition holds is $\mathbf{W} = \alpha \mathbf{I}$ for some $\alpha \geq 0$.

One of the interesting properties of Theorem 1 is that the addition of antennas in the receiver does not change the basic structure of the compound capacity and that the capacity depends on \mathbf{a} only through its norm. This result resembles a previous result of [16] where it was shown that under a simplified model the compound capacity is invariant to multiplying the channel by a unitary matrix.

4. RELATION TO STOCHASTIC CSI MODELS

Most of the previous references regarding the optimality of BF examined the use of stochastic CSI models. As explained in the introduction there is an intimate relation between this model and our deterministic CSI model. The most common stochastic CSI model is the Gaussian model [2]:

$$\mathbf{H}_s = \mathbf{a}_s \mathbf{x}_s^T + \mathbf{D}_s; \quad E \left\{ \mathbf{D}_s \mathbf{W}_s \mathbf{D}_s^T \right\} = \mathbf{I}, \quad (28)$$

where \mathbf{a}_s is a length N vector, \mathbf{x}_s is a length K vector, $\mathbf{W}_s \succ 0$ is a $K \times K$ matrix, and the elements of \mathbf{D}_s are zero mean Gaussian random variables¹. For simplicity, we restrict \mathbf{W}_s to be invertible. It is easy to see the resemblance between the deterministic model in (2) and the stochastic model in (28). The only difference is that in the deterministic model \mathbf{D} is a deterministic matrix within an ellipsoid defined by \mathbf{W} , and in the stochastic model \mathbf{D}_s is a random matrix whose covariance is defined by \mathbf{W}_s .

In stochastic CSI models, the mutual information in (4) is a random quantity since it is a function of \mathbf{D}_s . One of the standard measures for analyzing such systems is the outage mutual information, i.e., the inverse function of the CDF of the mutual information

$$I_{\text{out}} = \text{OUT}(P_{\text{out}}), \quad (29)$$

where

$$P_{\text{out}} = \Pr(I(\mathbf{Q}, \mathbf{D}_s) \leq I_{\text{out}}). \quad (30)$$

The inverse is unique due to the monotonicity of the CDF. The meaning of (29) is that there is a probability of P_{out} that in any realization of \mathbf{H} from the ensemble, we will obtain a mutual information I less than I_{out} . Therefore, the system is designed to maximize the outage mutual information [3].

In general, the calculation of the outage capacity is very difficult. In [3], it was derived for the MISO case using integrals over

¹The complete characterization of the statistics of \mathbf{D}_s is given by the covariance $E\{\text{vec}(\mathbf{D}_s) \text{vec}(\mathbf{D}_s)^T\} = \mathbf{W}_s^{-1} \otimes \mathbf{I}$.

the complex plane. Using these integrals, the authors maximized $\text{OUT}(P_{\text{out}})$ with respect to \mathbf{Q} . In the special case of $\mathbf{W}_s = \alpha \mathbf{I}$, they found that the optimal \mathbf{Q} has the structure $\mathbf{Q} = p_1 \mathbf{x}_s \mathbf{x}_s^T + p_2 \mathbf{I}$ for some power allocation $p_1 \geq 0$ and $p_2 \geq 0$. To our knowledge, there is no solution for the general case of arbitrary $\mathbf{a}_s \mathbf{x}_s^T$ and \mathbf{W}_s . Fortunately, the following lemma shows that there is an intimate relationship between the compound capacity and the outage mutual information:

Lemma 1. *Let \mathbf{H}_s satisfy the stochastic model in (28). Then,*

$$C(\mathbf{a}_s \mathbf{x}_s^T, \alpha \mathbf{W}_s) \leq \text{OUT} \left(1 - \text{CDF}_{\chi_{NK}^2} \left(\frac{1}{\alpha} \right) \right), \quad (31)$$

where $\text{CDF}_{\chi_{NK}^2}(\cdot)$ is the cumulative distribution function of a Chi Squared random variable with NK degrees of freedom.

Proof. Let us define the event A_α as the event when the realization of \mathbf{H}_s falls within the ellipsoid set defined in (2) with $\mathbf{a} = \mathbf{a}_s$, $\mathbf{x} = \mathbf{x}_s$ and $\mathbf{W} = \alpha \mathbf{W}_s$. The probability of this event is

$$\begin{aligned} \Pr(A_\alpha) &= \Pr \left(\text{Tr} \left\{ \mathbf{D}_s \mathbf{W}_s \mathbf{D}_s^T \right\} \leq \frac{1}{\alpha} \right) \\ &= \text{CDF}_{\chi_{NK}^2} \left(\frac{1}{\alpha} \right). \end{aligned} \quad (32)$$

By conditioning on A_α , we have

$$\begin{aligned} \Pr(I(\mathbf{Q}, \mathbf{D}_s) \leq C(\mathbf{X}_s, \alpha \mathbf{W}_s)) &\leq 0 \cdot \Pr(A_\alpha) + 1 \cdot (1 - \Pr(A_\alpha)) \\ &= 1 - \text{CDF}_{\chi_{NK}^2} \left(\frac{1}{\alpha} \right). \end{aligned} \quad (33)$$

Applying $\text{OUT}(\cdot)$ on both sides and noting the monotonicity of $\text{OUT}(\cdot)$ yields the required inequality. \square

In other words, the compound capacity provides a lower bound on the outage capacity, and instead of maximizing the outage capacity we can maximize the bound. Given a target probability P_{out}^* , one can solve for α^* in

$$P_{\text{out}}^* = 1 - \text{CDF}_{\chi_{NK}^2} \left(\frac{1}{\alpha^*} \right), \quad (34)$$

and then the optimal \mathbf{Q} of $C(\mathbf{a}_s \mathbf{x}_s^T, \alpha^* \mathbf{W}_s)$ will promise a lower bound on the outage mutual information of probability P_{out}^* . This is a very simple ad hoc approach to the outage capacity problem. Due to Theorem 1, it will allow a BF based solution for this important problem.

5. SIMULATIONS

In this section, we provide a simple numerical example that illustrates our results. We consider a system with $K = 4$ transmit antennas and $N = 1$ receive antennas. The transmitter has the following stochastic CSI: The channel is modelled as a random Gaussian vector with mean $[1, 0, 0, 0]^T$ and covariance \mathbf{W} where $[\mathbf{W}]_{i,i} = 1$ for all i and $[\mathbf{W}]_{i,j} = 0.5$ for $i \neq j$. Our objective is to maximize the outage capacity for an outage probability of $P_{\text{out}} = 0.05$. To our knowledge, there is no known technique for this optimization. Therefore, we propose to maximize the lower bound presented in the previous section using the

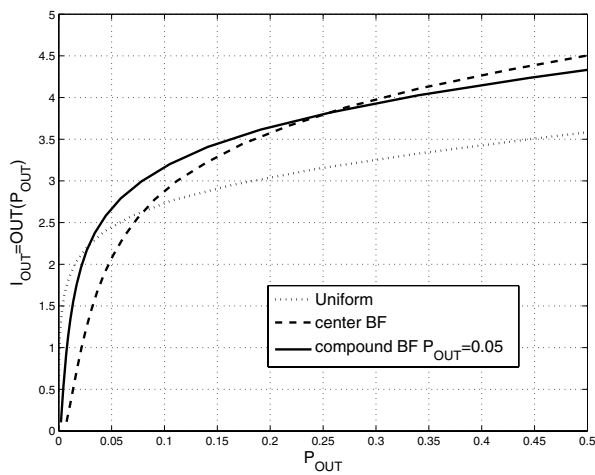


Fig. 1. Outage capacity as a function of the outage probability.

compound capacity. For comparison, we also simulate two other strategies: uniform power allocation, and beamforming along the mean (center). We numerically estimate the outage capacity using 100000 Monte Carlo simulations. The results are presented in Fig. 1. It is easy to see that at around the target outage probability our approach provides the highest outage capacity among the three transmit strategies.

6. CONCLUSION

We derived the compound capacity in a rank one Ricean MIMO channel using a deterministic CSI model. We showed that the optimal transmit strategy in this case is always beamforming, and can be found using a simple one dimensional search. These results strengthen previous results on the optimality of BF and motivate the growing use of systems using this practical transmit strategy. Due to its simplicity, we find that the compound capacity is an attractive alternative to the outage capacity as a design criterion in slow fading MIMO channels.

An interesting extension to this work is to use a more general deterministic CSI model and relax the rank one constraint on the center matrix. In such models, we conjecture that BF will not necessarily be optimal and therefore optimality conditions should be derived and analyzed.

7. REFERENCES

- [1] E. Visotsky and U. Madhow, "Space time transmit precoding with imperfect feedback," *IEEE Transactions on Information Theory*, vol. 47, no. 6, pp. 2632–2639, Sep. 2001.
- [2] S. A. Jafar and A. Goldsmith, "Transmitter optimization and optimality of beamforming for multiple antenna systems," *IEEE Transactions on Wireless Communications*, vol. 3, no. 4, pp. 1165–1175, July 2004.
- [3] A. Moustakas and S. H. Simon, "Optimizing multi-transmitter single-receiver (MISO) antenna systems with partial channel knowledge," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2770–2780, Oct. 2003.

- [4] S. H. Simon and A. Moustakas, "Optimizing MIMO antenna systems with channel covariance feedback," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 3, pp. 406–417, April 2003.
- [5] E. Jorswieck and H. Boche, "Optimal transmission strategies and impact of correlation in multi antenna systems with different types of channel state information," in *Proc. of conference on Acoustics, Speech, and Signal Processing (ICASSP-2003)*, April 2002.
- [6] H. Boche and E. A. Jorswieck, "Outage probability of multiple antenna systems: optimal transmission and impact of correlation," in *Proc. of IEEE Int. Zurich Seminar (IZS-2004)*, Feb. 2004.
- [7] M. Vu and A. Paulraj, "Optimum transmission scheme for a MISO wireless system with partial channel knowledge and infinite K factor," in *Proc. of IEEE Int. Conf. on Communications (ICC)*, June 2004.
- [8] A. Pascual Iserte, A. I. Perez Neira, and M. A. Lagunas, "A maximin approach for robust MIMO design: Combining OS-TBC and beamforming with minimum transmission power requirements," in *Proc. of IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP-2004)*, May 2004.
- [9] A. Abdel-Samad, A. Gershman, and T. Davidson, "Robust transmit eigen-beamforming based on imperfect channel state information," in *Proc. of the ITG Workshop on Smart Antennas (WSA-2004)*, March 2004.
- [10] I. Csiszar and J. Korner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, Academic Press, New York, 1981.
- [11] A. Lapidoth and P. Narayan, "Reliable communication under channel uncertainty," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2148–2177, Oct. 1998.
- [12] E. Biglieri, J. Proakis, and S. Shamai (Shitz), "Fading channels: Information-theoretic and communications aspects," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2619–2692, Oct. 1998.
- [13] S. Vishwanath, S. Boyd, and A. Goldsmith, "Worst-case capacity of gaussian vector channels," in *Proc. of the 2003 Canadian Workshop on Information Theory, Waterloo, ON, Canada.*, Feb. 2004.
- [14] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, "Uniform power allocation in MIMO channels: A game-theoretic approach," *IEEE Transactions on Information Theory*, vol. 49, no. 7, pp. 1707–1727, July 2003.
- [15] S. Wei and D. Goeckel, "On the minimax robustness of the uniform transmission power strategy in MIMO systems," *IEEE Communication Letters*, vol. 7, no. 11, pp. 523–524, November 2003.
- [16] M. Godavarti, A. O. Hero, and T. L. Marzetta, "Min-capacity of a multiple-antenna wireless channel in a static Ricean fading environment," *Submitted to IEEE Transactions on Information Theory*.
- [17] M. Sion, "On general minimax theorems," *Pac. J. Math.*, vol. 8, pp. 171176, 1958.
- [18] A. Melman, "A unifying convergence analysis of second-order methods for secular equations," *Mathematics Of Computation*, vol. 66, no. 217, pp. 333–344, Jan. 1997.