

# Optimization of the MIMO Compound Capacity

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

**Abstract**—In this paper, we consider the optimization of the compound capacity in a rank one Ricean multiple input multiple output channel using partial channel state information at the transmitter side. We model the channel as a deterministic matrix within a known ellipsoid, and address the compound capacity defined as the maximum worst case mutual information in the set. We find that the optimal transmit strategy is always beamforming, and can be found using a simple one dimensional search. Similar results are derived for the worst case sum-rate of a multiple access channel with individual power constraints and a total power constraint. In this multiuser setting we assume equal array response at the receiver for all users. These results motivate the growing use of systems using simple beamforming transmit strategies.

**Index Terms**—MIMO, compound capacity, beamforming.

## I. 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) at the transmitter. On the other hand, when perfect CSI is available, the standard technique is to use MIMO precoding algorithms, such as beamforming (BF). 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 complex Gaussian noise is signaling using random complex 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. BF is defined as the transmit strategy in which all of the transmitted power is allocated to one eigenvector. In this case, the MIMO system is transformed into a simple single input single output channel. This allows the system designer to resort to well known conventional scalar coding strategies.

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. Next, in [2] this work was generalized and conditions for the optimality of BF in rank one Ricean MIMO systems were found in closed form. Recently, an extension of these results to the multiple access channel (MAC) of a multiuser system was presented in [3]. In [4], [5] 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 [6]. Recently, in [7], the outage capacity with no available CSI was analyzed. A competing stochastic CSI model was introduced in [8] 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 [9]–[11]. 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 maximum worst case mutual information in the set [12]. (See also [13] 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

Manuscript received May 22, 2005; revised November 13, 2005; accepted November 26, 2005. The associate editor coordinating the review of this paper and approving it for publication was E. Serpedin.

The authors are with the Technion, Israel Institute of Technology (email: amiw@tx.technion.ac.il, {yonina, sshlomo}@ee.technion.ac.il). The authors were supported by the EU 6th Framework Programme, via the NEWCOM Network of Excellence. This research was also supported by the Israel Science Foundation, founded by the Israel Academy of Sciences and Humanities. Conference versions of this paper were presented at the ITG-IEEE Workshop on Smart Antennas (WSA-2005) and the IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC-2005).

Digital Object Identifier 10.1109/TWC.2007.05381.

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 [14]. For completeness, we mention that the compound capacity is also related to the problem of optimizing the capacity of the worst case noise covariance [15].

Due to its importance, the compound MIMO capacity recently gained a considerable attention. In [16], [17] it was shown that under different uncertainty sets the optimal transmit strategy is uniform power allocation. However, the uncertainty sets used in these papers are very different from the structure of the stochastic CSI models used in [1]–[4]. Therefore, the results are different and it is difficult to compare these two approaches. In another work [18], 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 this paper, we follow the deterministic approach, but use an uncertainty set with a structure that is very similar to the CSI model used in [1]–[4]. Specifically, 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.

After characterizing the MIMO compound capacity for point to point systems, we consider the problem in a multiuser setting. In particular, we address the optimization of the sum-rate (throughput) in a MAC system. It was recently shown that when perfect CSI is available BF maximizes the sum-rate of the MAC as the number of users increases [19]. Next, in [3], the MAC ergodic sum-rate with a mean or a covariance feedback model was considered under the assumption of equal array response at the receiver for all users. Conditions for the optimality of BF in this stochastic CSI model were derived. We continue with our new approach and characterize the worst case sum-rate of the MAC under a deterministic CSI model that resembles the stochastic CSI model of [3]. We consider two alternative power allocation policies: individual power constraints, and a total power constraint. Again, we show that BF strategies maximize the compound sum-rate under both policies.

The paper is organized as follows. We begin in Section II by defining our channel model and introducing the compound optimization problem. We provide our main result in Theorem 1 and discuss its consequences. Next in Section III, we extend this result to the MAC. The connection between our work and previous results based on stochastic CSI models is addressed in Section IV. In particular, we discuss the relation between

the compound capacity and the outage mutual information.

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  $(\cdot)^H$  and  $(\cdot)^{-1}$  denote the Hermitian 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, and by  $\mathbf{I}$  we denote the identity matrix of appropriate size.  $|\cdot|$  denotes the determinant, and  $\|\cdot\|$  denotes the standard Euclidean norm. Finally,  $\mathbf{X} \succeq 0$  means that  $\mathbf{X}$  is a Hermitian positive semidefinite matrix.

## II. OPTIMIZATION OF THE COMPOUND CAPACITY

In this section, we address the optimization of the compound capacity in point to point, rank one Ricean MIMO channels. Consider a point to point communication system with  $K$  transmit antennas and  $N$  receive antennas. Mathematically, the system can be represented using the following MIMO channel model:

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

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

$$\begin{aligned} \mathbf{H} &= \mathbf{h}_r\mathbf{h}_t^H + \mathbf{D}; \\ \text{Tr}\{\mathbf{D}\mathbf{W}\mathbf{D}^H\} &\leq 1, \end{aligned} \quad (2)$$

where  $\mathbf{h}_r \neq \mathbf{0}$  is a length  $N$  vector denoting the array response at the receiver's side,  $\mathbf{h}_t$  is a length  $K$  vector denoting the array response at the transmitter's side, and  $\mathbf{W} \succeq 0$  is a weight matrix. In our terminology,  $\mathbf{h}_r\mathbf{h}_t^H$  is the rank one specular component of the channel, and  $\mathbf{D}$  is the scattering component. We assume that the transmitter knows  $\mathbf{h}_r\mathbf{h}_t^H$  and  $\mathbf{W}$ , but does not have access to the specific realization of  $\mathbf{H}$  within the set. In Section IV, we will show that this CSI model is the deterministic analog of the stochastic CSI model used in [1]–[4], where  $\mathbf{h}_r\mathbf{h}_t^H$  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 [12], [13], [18]:

$$C(\mathbf{h}_r\mathbf{h}_t^H, \mathbf{W}) = \max_{\substack{\mathbf{Q} \succeq 0 \\ \text{Tr}\{\mathbf{Q}\} \leq P}} \min_{\text{Tr}\{\mathbf{D}\mathbf{W}\mathbf{D}^H\} \leq 1} I(\mathbf{Q}, \mathbf{D}), \quad (3)$$

where

$$I(\mathbf{Q}, \mathbf{D}) = \log \left| \mathbf{I} + \frac{1}{\sigma^2} (\mathbf{h}_r\mathbf{h}_t^H + \mathbf{D}) \mathbf{Q} (\mathbf{h}_r\mathbf{h}_t^H + \mathbf{D})^H \right|, \quad (4)$$

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

In the following theorem we provide the solution to the optimization of the point to point compound capacity:

**Theorem 1.** Consider the optimization of the rank one Ricean MIMO compound capacity of  $C(\mathbf{h}_r \mathbf{h}_t^H, \mathbf{W})$ :

$$C(\mathbf{h}_r \mathbf{h}_t^H, \mathbf{W}) = \max_{\substack{\mathbf{Q} \succeq 0 \\ \text{Tr}\{\mathbf{Q}\} \leq P}} \min_{\text{Tr}\{\mathbf{D}\mathbf{W}\mathbf{D}^H\} \leq 1} \log \left| \mathbf{I} + \frac{1}{\sigma^2} (\mathbf{h}_r \mathbf{h}_t^H + \mathbf{D}) \mathbf{Q} (\mathbf{h}_r \mathbf{h}_t^H + \mathbf{D})^H \right|. \quad (5)$$

Its optimal value is

$$C(\mathbf{h}_r \mathbf{h}_t^H, \mathbf{W}) = \log \left( 1 + \frac{P}{\sigma^2} \|\mathbf{h}_r\|^2 c \right), \quad (6)$$

where

$$c = \begin{cases} \mathbf{h}_t^H \left[ \mathbf{I} - (\mathbf{I} + \lambda \mathbf{W})^{-1} \right]^2 \mathbf{h}_t, & \mathbf{h}_t^H \mathbf{W} \mathbf{h}_t > \frac{1}{\|\mathbf{h}_r\|^2}; \\ 0, & \text{else,} \end{cases} \quad (7)$$

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

$$\mathbf{h}_t^H (\mathbf{I} + \lambda \mathbf{W})^{-1} \mathbf{W} (\mathbf{I} + \lambda \mathbf{W})^{-1} \mathbf{h}_t = \frac{1}{\|\mathbf{h}_r\|^2}. \quad (8)$$

If  $\mathbf{h}_t^H \mathbf{W} \mathbf{h}_t > \frac{1}{\|\mathbf{h}_r\|^2}$  then the optimal  $\mathbf{Q}$  is

$$\mathbf{Q} = P \frac{\mathbf{q}(\lambda) \mathbf{q}^H(\lambda)}{\|\mathbf{q}(\lambda)\|^2}, \quad (9)$$

where

$$\mathbf{q}(\lambda) = \left[ \mathbf{I} - (\mathbf{I} + \lambda \mathbf{W})^{-1} \right] \mathbf{h}_t. \quad (10)$$

Otherwise, any feasible  $\mathbf{Q}$  will result in a zero compound capacity.

*Proof:* We begin the proof by showing that the optimal argument  $\mathbf{D}$  of the inner minimization in (5) can be chosen as  $\mathbf{D} = \mathbf{h}_r \mathbf{d}^H$  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  $\tilde{\mathbf{D}} = \mathbf{h}_r \mathbf{d}^H$  with  $\mathbf{d}^H = \frac{\mathbf{h}_r^H}{\|\mathbf{h}_r\|^2} \mathbf{D}$  is also optimal. Now, assume that  $\mathbf{D}$  is feasible, then  $\tilde{\mathbf{D}}$  is also feasible since

$$\begin{aligned} \text{Tr}\{\tilde{\mathbf{D}}\mathbf{W}\tilde{\mathbf{D}}^H\} &= \text{Tr}\left\{ \mathbf{h}_r \frac{\mathbf{h}_r^H}{\|\mathbf{h}_r\|^2} \mathbf{D} \mathbf{W} \mathbf{D}^H \frac{\mathbf{h}_r}{\|\mathbf{h}_r\|^2} \mathbf{h}_r^H \right\} \\ &= \frac{\mathbf{h}_r^H \mathbf{D} \mathbf{W} \mathbf{D}^H \mathbf{h}_r}{\mathbf{h}_r^H \mathbf{h}_r} \leq \max_{\mathbf{v} \neq 0} \frac{\mathbf{v}^H \mathbf{D} \mathbf{W} \mathbf{D}^H \mathbf{v}}{\mathbf{v}^H \mathbf{v}} \\ &= \lambda_{\max}(\mathbf{D} \mathbf{W} \mathbf{D}^H) \leq \text{Tr}\{\mathbf{D} \mathbf{W} \mathbf{D}^H\} \leq 1, \end{aligned} \quad (11)$$

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

$$\begin{aligned} & \left| \mathbf{I} + \frac{1}{\sigma^2} (\mathbf{h}_r \mathbf{h}_t^H + \mathbf{D}) \mathbf{Q} (\mathbf{h}_r \mathbf{h}_t^H + \mathbf{D})^H \right| \\ &= \prod_{i=1}^N \left[ 1 + \frac{1}{\sigma^2} \lambda_i \left( (\mathbf{h}_r \mathbf{h}_t^H + \mathbf{D}) \mathbf{Q} (\mathbf{h}_r \mathbf{h}_t^H + \mathbf{D})^H \right) \right] \\ &\geq 1 + \frac{1}{\sigma^2} \lambda_{\max} \left( (\mathbf{h}_r \mathbf{h}_t^H + \mathbf{D}) \mathbf{Q} (\mathbf{h}_r \mathbf{h}_t^H + \mathbf{D})^H \right) \\ &= 1 + \frac{1}{\sigma^2} \max_{\mathbf{v} \neq 0} \frac{\mathbf{v}^H (\mathbf{h}_r \mathbf{h}_t^H + \mathbf{D}) \mathbf{Q} (\mathbf{h}_r \mathbf{h}_t^H + \mathbf{D})^H \mathbf{v}}{\mathbf{v}^H \mathbf{v}} \\ &\geq 1 + \frac{1}{\sigma^2} \frac{\mathbf{h}_r^H (\mathbf{h}_r \mathbf{h}_t^H + \mathbf{D}) \mathbf{Q} (\mathbf{h}_r \mathbf{h}_t^H + \mathbf{D})^H \mathbf{h}_r}{\mathbf{h}_r^H \mathbf{h}_r} \\ &= 1 + \frac{\mathbf{h}_r^H \mathbf{h}_r}{\sigma^2} \left( \mathbf{h}_t^H + \frac{\mathbf{h}_r^H}{\|\mathbf{h}_r\|^2} \mathbf{D} \right) \mathbf{Q} \left( \mathbf{h}_t + \mathbf{D}^H \frac{\mathbf{h}_r}{\|\mathbf{h}_r\|^2} \right) \\ &= \left| \mathbf{I} + \frac{1}{\sigma^2} \left( \mathbf{h}_r \mathbf{h}_t^H + \mathbf{h}_r \frac{\mathbf{h}_r^H}{\|\mathbf{h}_r\|^2} \mathbf{D} \right) \mathbf{Q} \left( \mathbf{h}_r \mathbf{h}_t^H + \mathbf{h}_r \frac{\mathbf{h}_r^H}{\|\mathbf{h}_r\|^2} \mathbf{D} \right)^H \right| \\ &= \left| \mathbf{I} + \frac{1}{\sigma^2} (\mathbf{h}_r \mathbf{h}_t^H + \tilde{\mathbf{D}}) \mathbf{Q} (\mathbf{h}_r \mathbf{h}_t^H + \tilde{\mathbf{D}})^H \right|, \end{aligned} \quad (12)$$

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

Therefore, we can restrict ourselves to solutions of the form  $\mathbf{D} = \mathbf{h}_r \mathbf{d}^H$  which yield the following program

$$\max_{\substack{\mathbf{Q} \succeq 0 \\ \text{Tr}\{\mathbf{Q}\} \leq P}} \min_{\text{Tr}\{\mathbf{h}_r \mathbf{d}^H \mathbf{W} \mathbf{d} \mathbf{h}_r^H\} \leq 1} \quad (13)$$

$$\log \left| \mathbf{I} + \frac{1}{\sigma^2} (\mathbf{h}_r \mathbf{h}_t^H + \mathbf{h}_r \mathbf{d}^H) \mathbf{Q} (\mathbf{h}_r \mathbf{h}_t^H + \mathbf{h}_r \mathbf{d}^H)^H \right|,$$

Using  $|\mathbf{I} + \mathbf{A}\mathbf{B}| = |\mathbf{I} + \mathbf{B}\mathbf{A}|$  yields

$$\max_{\substack{\mathbf{Q} \succeq 0 \\ \text{Tr}\{\mathbf{Q}\} \leq P}} \min_{\mathbf{d}^H \mathbf{W} \mathbf{d} \leq \frac{1}{\|\mathbf{h}_r\|^2}} \quad (14)$$

$$\log \left( 1 + \frac{1}{\sigma^2} \|\mathbf{h}_r\|^2 (\mathbf{h}_t + \mathbf{d})^H \mathbf{Q} (\mathbf{h}_t + \mathbf{d}) \right).$$

Now, introducing a simple change of variables  $\mathbf{Q}' = \frac{1}{P} \mathbf{Q}$ , we can normalize  $\text{Tr}\{\mathbf{Q}\}$  (In order to keep the notations simple, we continue to use  $\mathbf{Q}$  instead of  $\mathbf{Q}'$ ):

$$\max_{\substack{\mathbf{Q} \succeq 0 \\ \text{Tr}\{\mathbf{Q}\} \leq 1}} \min_{\mathbf{d}^H \mathbf{W} \mathbf{d} \leq \frac{1}{\|\mathbf{h}_r\|^2}} \quad (15)$$

$$\log \left( 1 + \frac{P}{\sigma^2} \|\mathbf{h}_r\|^2 (\mathbf{h}_t + \mathbf{d})^H \mathbf{Q} (\mathbf{h}_t + \mathbf{d}) \right).$$

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

$$\max_{\substack{\mathbf{Q} \succeq 0 \\ \text{Tr}\{\mathbf{Q}\} \leq 1}} \min_{\mathbf{d}^H \mathbf{W} \mathbf{d} \leq \frac{1}{\|\mathbf{h}_r\|^2}} (\mathbf{h}_t + \mathbf{d})^H \mathbf{Q} (\mathbf{h}_t + \mathbf{d}). \quad (16)$$

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 [20] 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

$$\max_{\substack{\mathbf{Q} \succeq 0 \\ \text{Tr}\{\mathbf{Q}\} \leq 1}} (\mathbf{h}_t + \bar{\mathbf{d}})^H \mathbf{Q} (\mathbf{h}_t + \bar{\mathbf{d}}), \quad (17)$$

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

$$\min_{\mathbf{d}^H \mathbf{W} \mathbf{d} \leq \frac{1}{\|\mathbf{h}_r\|^2}} (\mathbf{h}_t + \mathbf{d})^H \bar{\mathbf{Q}} (\mathbf{h}_t + \mathbf{d}). \quad (18)$$

The Lagrangian associated with program (17) is

$$L_1 = -(\mathbf{h}_t + \bar{\mathbf{d}})^H \mathbf{Q} (\mathbf{h}_t + \bar{\mathbf{d}}) - \text{Tr}\{\mathbf{Y}\mathbf{Q}\} + \nu [\text{Tr}\{\mathbf{Q}\} - 1], \quad (19)$$

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

$$\begin{aligned} -(\mathbf{h}_t + \bar{\mathbf{d}})(\mathbf{h}_t + \bar{\mathbf{d}})^H - \mathbf{Y} + \nu \mathbf{I} &= \mathbf{0}; \\ \text{Tr}\{\mathbf{Y}\bar{\mathbf{Q}}\} &= 0; \\ \nu [\text{Tr}\{\bar{\mathbf{Q}}\} - 1] &= 0. \end{aligned} \quad (20)$$

It is easy to check that

$$\bar{\mathbf{Q}} = \frac{(\mathbf{h}_t + \bar{\mathbf{d}})(\mathbf{h}_t + \bar{\mathbf{d}})^H}{\|\mathbf{h}_t + \bar{\mathbf{d}}\|^2}, \quad (21)$$

along with

$$\begin{aligned} \mathbf{Y} &= \|\mathbf{h}_t + \bar{\mathbf{d}}\|^2 \mathbf{I} - (\mathbf{h}_t + \bar{\mathbf{d}})(\mathbf{h}_t + \bar{\mathbf{d}})^H; \\ \nu &= \|\mathbf{h}_t + \bar{\mathbf{d}}\|^2, \end{aligned} \quad (22)$$

satisfy these exact conditions. In addition, the saddle point must satisfy the optimality conditions associated with program (18). The Lagrangian of this problem is

$$L_2 = (\mathbf{h}_t + \mathbf{d})^H \bar{\mathbf{Q}} (\mathbf{h}_t + \mathbf{d}) + \lambda \left[ \mathbf{d}^H \mathbf{W} \mathbf{d} - \frac{1}{\|\mathbf{h}_r\|^2} \right], \quad (23)$$

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

$$(\bar{\mathbf{Q}} + \lambda \mathbf{W}) \bar{\mathbf{d}} = -\bar{\mathbf{Q}} \mathbf{h}_t; \quad (24)$$

$$\lambda \left[ \bar{\mathbf{d}}^H \mathbf{W} \bar{\mathbf{d}} - \frac{1}{\|\mathbf{h}_r\|^2} \right] = 0. \quad (25)$$

Plugging  $\bar{\mathbf{Q}}$  from (21) into (24) results in:

$$(\mathbf{h}_t + \bar{\mathbf{d}}) + \lambda \mathbf{W} \bar{\mathbf{d}} = \mathbf{0}. \quad (26)$$

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

$$\bar{\mathbf{d}} = -(\mathbf{I} + \lambda \mathbf{W})^{-1} \mathbf{h}_t. \quad (27)$$

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

$$\mathbf{h}_t^H (\mathbf{I} + \lambda \mathbf{W})^{-1} \mathbf{W} (\mathbf{I} + \lambda \mathbf{W})^{-1} \mathbf{h}_t = \frac{1}{\|\mathbf{h}_r\|^2}. \quad (28)$$

It is easy to see that the left hand side of (28) is monotonically decreasing in  $\lambda$  from  $\mathbf{h}_t^H \mathbf{W} \mathbf{h}_t > \frac{1}{\|\mathbf{h}_r\|^2}$  when  $\lambda = 0$  to 0 when  $\lambda \rightarrow \infty$ . Therefore, a unique solution for  $\lambda$  in (28) always exists. Finally, plugging the optimal  $\bar{\mathbf{d}}$  and  $\bar{\mathbf{Q}}$  into (16) and (21) yields (7) and (9), respectively. This concludes the proof.

For completeness, we mention that an alternative proof can be obtained using the well known minimax theorem and interchanging the order of the minimization and maximization in (16). More details on this procedure can be found in [10].

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 (10). This direction is defined by the  $\lambda$  which satisfies (8). Finding this  $\lambda$  is very easy. Using the eigenvalue decomposition of  $\mathbf{W} = \mathbf{U} \mathbf{\Delta} \mathbf{U}^H$  where  $\mathbf{U}$  is an orthogonal matrix and  $\mathbf{\Delta}$  is a diagonal matrix with the elements  $\delta_i$ , we rewrite the condition as

$$\sum_i \frac{\delta_i}{(1 + \lambda \delta_i)^2} |[\mathbf{U}^H \mathbf{h}_t]_i|^2 = \frac{1}{\|\mathbf{h}_r\|^2}. \quad (29)$$

As explained in the proof, the left hand side of (29) is monotonically decreasing in  $\lambda \geq 0$ . Therefore, any simple bisection can efficiently solve for  $\lambda$ . Moreover, (29) belongs to a well known family of non linear equations called *secular equations* for which there are highly efficient root finding algorithms [21].

An important practical case is when the optimal BF is in the direction of  $\mathbf{h}_t$ . This is probably the standard technique in many applications due to its simplicity. The following corollary provides a sufficient optimality condition for this transmit strategy:

**Corollary 1.** *If  $\mathbf{h}_t$  is a scaled eigenvector of  $\mathbf{W}$  then BF along the mean  $\mathbf{h}_t$  maximizes the compound capacity  $C(\mathbf{h}_r, \mathbf{h}_t^H, \mathbf{W})$  in (5). A common example where this condition holds is  $\mathbf{W} = \alpha \mathbf{I}$  for some  $\alpha \geq 0$ .*

*Proof:* Theorem 1 shows that BF along  $\mathbf{q}(\lambda)$  of (10) is optimal. Now, if  $\mathbf{h}_t$  is a scaled eigenvector of  $\mathbf{W}$  then it is also a scaled eigenvector of  $[\mathbf{I} - (\mathbf{I} + \lambda \mathbf{W})^{-1}]$ , and  $\mathbf{q}(\lambda)$  is equal to  $\mathbf{h}_t$  up to a scaling.

One of the interesting properties of Theorem 1 is the existence of a saddle point in our problem. As explained in the proof, interchanging the order of the maximization and minimization in (5) does not change the optimal value of the optimization. This means that there is no advantage in better feedback in such channels. Suppose perfect feedback was available, then in the worst case scenario the capacity would be the minimum of the maximal mutual information. The saddle point property shows that this minimum is equal to the compound capacity (which is the maximum of the minimal mutual information). More details on this property can be found in [13].

Another of property 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  $\mathbf{h}_r$  only through its norm. This result resembles previous results in [16], [18] where it was shown that under similar models the compound capacity is invariant to multiplying the channel by a unitary matrix on the left.

### III. MULTIPLE ACCESS CHANNEL COMPOUND CAPACITY

In this section, we extend Theorem 1 to a multiuser system. In particular, following [3], we consider a MAC with  $M$  users signaling to a single base station. We assume that each user is equipped with  $K$  transmit antennas, and that the base station is

equipped with  $N$  receive antennas. In this system, the channel model is

$$\mathbf{y} = \sum_{i=1}^M \mathbf{H}_i \mathbf{x}_i + \mathbf{w}, \quad (30)$$

where  $\mathbf{y}$  is a length  $N$  received vector,  $\mathbf{H}_i$  for  $i = 1, \dots, M$  are  $N \times K$  channel matrices,  $\mathbf{x}_i$  for  $i = 1, \dots, M$  are the length  $K$  transmit vectors of the multiple users, and  $\mathbf{w}$  is a length  $N$  complex Gaussian noise vector of covariance  $\sigma^2 \mathbf{I}$ .

As before, we model the rank one Ricean MIMO channels  $\mathbf{H}_i$  for  $i = 1, \dots, M$  as unknown deterministic matrices:

$$\begin{aligned} \mathbf{H}_i &= \mathbf{h}_{r,i} \mathbf{h}_{t,i}^H + \mathbf{D}_i; \\ \text{Tr} \{ \mathbf{D}_i \mathbf{W}_i \mathbf{D}_i^H \} &\leq 1. \end{aligned} \quad (31)$$

We assume that the transmitters know  $\mathbf{h}_{r,i} \mathbf{h}_{t,i}^H$  and  $\mathbf{W}_i$ , but do not have access to the specific realizations of  $\mathbf{H}_i$ . Following [3], we assume that the array response at the receiver  $\mathbf{h}_{r,i} = \mathbf{h}_r \neq \mathbf{0}$  is the same for all users, while the array responses at the multiple transmitters  $\mathbf{h}_{t,i}$  are different. Physically, this means that the multiple signals arrive at the receiver in phase (see [3] for more details). However, it is important to remember that this model has its limitations. For example, consider the case when  $K = 1$ , then  $\mathbf{h}_{r,i} = \mathbf{h}_r$  means that the channels of all the users are just scaled versions of each other. Clearly, this does not reflect the system. Therefore, although outside the scope of this paper, we believe that the more general problem is also very interesting.

In general, the main performance measure in the MAC is the capacity region, i.e., the region of all the sets of rates that allow simultaneous reliable communication for all users. Following [3], we will restrict ourselves to the sum-rate (throughput), i.e., the set of rates within this region that has the maximal sum. It can be shown that when the transmitters have full CSI the sum-rate is equal to:

$$\log \left| \mathbf{I} + \frac{1}{\sigma^2} \sum_{i=1}^M \mathbf{H}_i \mathbf{Q}_i \mathbf{H}_i^H \right|, \quad (32)$$

where  $\mathbf{Q}_i = E\{\mathbf{x}_i \mathbf{x}_i^H\}$  for  $i = 1, \dots, M$  is the transmit covariance matrix of the  $i$ 'th user. As before, when only partial deterministic CSI is available, we propose to optimize the worst case sum-rate of MAC. For more details on the compound capacity of the MAC we refer the reader to [12], [13]. We now characterize this worst case sum-rate under two alternative power policies. Our first result follows [3] and provides the worst case sum-rate under individual power constraints:

**Theorem 2.** Consider the optimization of the worst case sum-rate of the MAC with individual power constraints:

$$\begin{aligned} C^I &= \max_{\substack{\mathbf{Q}_i \succeq 0 \\ \text{Tr}\{\mathbf{Q}_i\} \leq P_i}} \min_{\text{Tr}\{\mathbf{D}_i \mathbf{W}_i \mathbf{D}_i^H\} \leq 1} \log \left| \mathbf{I} + \frac{1}{\sigma^2} \sum_{i=1}^M (\mathbf{h}_r \mathbf{h}_{t,i}^H + \mathbf{D}_i) \mathbf{Q}_i (\mathbf{h}_r \mathbf{h}_{t,i}^H + \mathbf{D}_i)^H \right|. \end{aligned} \quad (33)$$

Its optimal value is

$$C^I = \log \left( 1 + \sum_{i=1}^M \frac{P_i}{\sigma^2} \|\mathbf{h}_r\|^2 c_i \right), \quad (34)$$

where  $c_i$  for  $i = 1, \dots, M$  are defined as

$$c_i = \begin{cases} \mathbf{h}_{t,i}^H \left[ \mathbf{I} - (\mathbf{I} + \lambda_i \mathbf{W}_i)^{-1} \right]^2 \mathbf{h}_{t,i}, & \mathbf{h}_{t,i}^H \mathbf{W}_i \mathbf{h}_{t,i} > \frac{1}{\|\mathbf{h}_r\|^2}; \\ 0, & \text{else,} \end{cases} \quad (35)$$

and  $\lambda_i > 0$  are the unique roots of the following non linear equations

$$\mathbf{h}_{t,i}^H (\mathbf{I} + \lambda_i \mathbf{W}_i)^{-1} \mathbf{W}_i (\mathbf{I} + \lambda_i \mathbf{W}_i)^{-1} \mathbf{h}_{t,i} = \frac{1}{\|\mathbf{h}_r\|^2}. \quad (36)$$

In this case, the optimal  $\mathbf{Q}_i$  are

$$\mathbf{Q}_i = P_i \frac{\mathbf{q}_i(\lambda_i) \mathbf{q}_i^H(\lambda_i)}{\|\mathbf{q}_i(\lambda_i)\|^2}, \quad (37)$$

where

$$\mathbf{q}_i(\lambda_i) = \left[ \mathbf{I} - (\mathbf{I} + \lambda_i \mathbf{W}_i)^{-1} \right] \mathbf{h}_{t,i}. \quad (38)$$

*Proof:* The proof is very similar to the proof of Theorem 1, therefore we only provide the main differences. We begin as in (12) by showing that if  $\mathbf{D}_i$  for  $i = 1, \dots, M$  are optimal then  $\tilde{\mathbf{D}}_i = \mathbf{h}_r \mathbf{d}_i^H$  with  $\mathbf{d}_i^H = \frac{\mathbf{h}_r^H}{\|\mathbf{h}_r\|^2} \mathbf{D}_i$  yield an equal or better objective value:

$$\begin{aligned} & \left| \mathbf{I} + \frac{1}{\sigma^2} \sum_{i=1}^M (\mathbf{h}_r \mathbf{h}_{t,i}^H + \mathbf{D}_i) \mathbf{Q}_i (\mathbf{h}_r \mathbf{h}_{t,i}^H + \mathbf{D}_i)^H \right| \\ & \geq 1 + \frac{1}{\sigma^2} \lambda_{\max} \left( \sum_{i=1}^M (\mathbf{h}_r \mathbf{h}_{t,i}^H + \mathbf{D}_i) \mathbf{Q}_i (\mathbf{h}_r \mathbf{h}_{t,i}^H + \mathbf{D}_i)^H \right) \\ & = 1 + \frac{1}{\sigma^2} \max_{\mathbf{v} \neq \mathbf{0}} \frac{\mathbf{v}^H \left[ \sum_{i=1}^M (\mathbf{h}_r \mathbf{h}_{t,i}^H + \mathbf{D}_i) \mathbf{Q}_i (\mathbf{h}_r \mathbf{h}_{t,i}^H + \mathbf{D}_i)^H \right] \mathbf{v}}{\mathbf{v}^H \mathbf{v}} \\ & \geq 1 + \frac{1}{\sigma^2} \frac{\mathbf{h}_r^H \left[ \sum_{i=1}^M (\mathbf{h}_r \mathbf{h}_{t,i}^H + \mathbf{D}_i) \mathbf{Q}_i (\mathbf{h}_r \mathbf{h}_{t,i}^H + \mathbf{D}_i)^H \right] \mathbf{h}_r}{\mathbf{h}_r^H \mathbf{h}_r} \\ & = 1 + \frac{1}{\sigma^2} \sum_{i=1}^M \mathbf{h}_r^H \mathbf{h}_r \left( \mathbf{h}_{t,i}^H + \frac{\mathbf{h}_r^H}{\|\mathbf{h}_r\|^2} \mathbf{D}_i \right) \mathbf{Q}_i \left( \mathbf{h}_{t,i} + \mathbf{D}_i^H \frac{\mathbf{h}_r}{\|\mathbf{h}_r\|^2} \right) \\ & = \left| \mathbf{I} + \frac{1}{\sigma^2} \sum_{i=1}^M (\mathbf{h}_r \mathbf{h}_{t,i}^H + \tilde{\mathbf{D}}_i) \mathbf{Q}_i (\mathbf{h}_r \mathbf{h}_{t,i}^H + \tilde{\mathbf{D}}_i)^H \right|. \end{aligned} \quad (39)$$

Therefore, solving (33) reduces to the following program

$$\begin{aligned} & \max_{\substack{\mathbf{Q}_i \succeq 0 \\ \text{Tr}\{\mathbf{Q}_i\} \leq P}} \min_{\text{Tr}\{\mathbf{h}_r \mathbf{d}_i^H \mathbf{W}_i \mathbf{d}_i \mathbf{h}_r^H\} \leq 1} \log \left| \mathbf{I} + \frac{1}{\sigma^2} \sum_{i=1}^M (\mathbf{h}_r \mathbf{h}_{t,i}^H + \mathbf{h}_r \mathbf{d}_i^H) \mathbf{Q}_i (\mathbf{h}_r \mathbf{h}_{t,i}^H + \mathbf{h}_r \mathbf{d}_i^H)^H \right|. \end{aligned} \quad (40)$$

Using  $|\mathbf{I} + \mathbf{A}\mathbf{B}| = |\mathbf{I} + \mathbf{B}\mathbf{A}|$ , and normalizing the covariances as in (14)-(15) yields

$$\begin{aligned} & \max_{\substack{\mathbf{Q}_i \succeq 0 \\ \text{Tr}\{\mathbf{Q}_i\} \leq 1}} \min_{\mathbf{d}_i^H \mathbf{W}_i \mathbf{d}_i \leq \frac{1}{\|\mathbf{h}_r\|^2}} \log \left( 1 + \sum_{i=1}^M \frac{P_i}{\sigma^2} \|\mathbf{h}_r\|^2 (\mathbf{h}_{t,i} + \mathbf{d}_i)^H \mathbf{Q}_i (\mathbf{h}_{t,i} + \mathbf{d}_i) \right). \end{aligned} \quad (41)$$

Due to the monotonicity of the objective function in the sum of quadratic forms, we can optimize the sum instead:

$$\begin{aligned} & \max_{\substack{\mathbf{Q}_i \succeq 0 \\ \text{Tr}\{\mathbf{Q}_i\} \leq 1}} \min_{\mathbf{d}_i^H \mathbf{W}_i \mathbf{d}_i \leq \frac{1}{\|\mathbf{h}_r\|^2}} \quad (42) \\ & \sum_{i=1}^M \frac{P_i}{\sigma^2} \|\mathbf{h}_r\|^2 (\mathbf{h}_{t,i} + \mathbf{d}_i)^H \mathbf{Q}_i (\mathbf{h}_{t,i} + \mathbf{d}_i). \end{aligned}$$

It is easy to see that the above problem is separable and can be reduced to  $M$  independent optimization problems of the form (16) which we have already solved in the previous section.

Theorem 2 proves that under our definitions BF is also optimal in the MAC. Moreover, the multiuser optimization problem decouples into multiple single user optimization problems which are easy to solve. The optimal transmit strategy of each of the users does not depend on the CSI of the other users. This resembles the results of [3] where the transmit strategy of each user using stochastic CSI depends only on its own CSI.

Usually, the sum-rate of the MAC is optimized under individual power constraints as in Theorem 2. However, the problem of optimizing the sum-rate under a total power constraint has also been addressed [22]. Therefore, in the following theorem we derive the worst case sum-rate under a total power constraint:

**Theorem 3.** *Consider the optimization of the worst case sum-rate of the MAC with a total power constraint:*

$$\begin{aligned} C^T = & \max_{\substack{\mathbf{Q}_i \succeq 0 \\ \sum_{i=1}^M \text{Tr}\{\mathbf{Q}_i\} \leq P}} \min_{\text{Tr}\{\mathbf{D}_i \mathbf{W}_i \mathbf{D}_i^H\} \leq 1} \quad (43) \\ & \log \left| \mathbf{I} + \frac{1}{\sigma^2} \sum_{i=1}^M (\mathbf{h}_r \mathbf{h}_{t,i}^H + \mathbf{D}_i) \mathbf{Q}_i (\mathbf{h}_r \mathbf{h}_{t,i}^H + \mathbf{D}_i)^H \right|. \end{aligned}$$

Its optimal value is

$$C^T = \log \left( 1 + \frac{P}{\sigma^2} \|\mathbf{h}_r\|^2 \max_{i=1, \dots, M} c_i \right), \quad (44)$$

where  $c_i$  for  $i = 1, \dots, M$  are defined as

$$c_i = \begin{cases} \mathbf{h}_{t,i}^H \left[ \mathbf{I} - (\mathbf{I} + \lambda_i \mathbf{W}_i)^{-1} \right]^2 \mathbf{h}_{t,i}, & \mathbf{h}_{t,i}^H \mathbf{W}_i \mathbf{h}_{t,i} > \frac{1}{\|\mathbf{h}_r\|^2}; \\ 0, & \text{else,} \end{cases} \quad (45)$$

and  $\lambda_i > 0$  are the unique roots of the following non linear equations

$$\mathbf{h}_{t,i}^H (\mathbf{I} + \lambda_i \mathbf{W}_i)^{-1} \mathbf{W}_i (\mathbf{I} + \lambda_i \mathbf{W}_i)^{-1} \mathbf{h}_{t,i} = \frac{1}{\|\mathbf{h}_r\|^2}. \quad (46)$$

In this case, the optimal  $\mathbf{Q}_i$ 's are all zero except for the  $\mathbf{Q}_i$  with the index  $i = \max_{i'=1, \dots, M} c_{i'}$  which is defined as:

$$\mathbf{Q}_i = P \frac{\mathbf{q}_i(\lambda_i) \mathbf{q}_i^H(\lambda_i)}{\|\mathbf{q}_i(\lambda_i)\|^2}, \quad (47)$$

where

$$\mathbf{q}_i(\lambda_i) = \left[ \mathbf{I} - (\mathbf{I} + \lambda_i \mathbf{W}_i)^{-1} \right] \mathbf{h}_{t,i}. \quad (48)$$

*Proof:* Using the same steps as in the proof of Theorem 2 we can reduce (43) to the following program

$$\begin{aligned} & \max_{\substack{\mathbf{Q}_i \succeq 0 \\ \sum_i \text{Tr}\{\mathbf{Q}_i\} \leq 1}} \min_{\mathbf{d}_i^H \mathbf{W}_i \mathbf{d}_i \leq \frac{1}{\|\mathbf{h}_r\|^2}} \sum_{i=1}^M (\mathbf{h}_{t,i} + \mathbf{d}_i)^H \mathbf{Q}_i (\mathbf{h}_{t,i} + \mathbf{d}_i). \quad (49) \end{aligned}$$

The inner minimization can be separated into multiple minimizations over all  $i$ . By defining

$$f_i(\mathbf{Q}_i) = \min_{\mathbf{d}_i^H \mathbf{W}_i \mathbf{d}_i \leq \frac{1}{\|\mathbf{h}_r\|^2}} (\mathbf{h}_{t,i} + \mathbf{d}_i)^H \mathbf{Q}_i (\mathbf{h}_{t,i} + \mathbf{d}_i) \quad (50)$$

for  $i = 1, \dots, M$ , we have the following chain

$$\begin{aligned} & \max_{\substack{\mathbf{Q}_i \succeq 0 \\ \sum_{i=1}^M \text{Tr}\{\mathbf{Q}_i\} \leq 1}} \sum_{i=1}^M f_i(\mathbf{Q}_i) \quad (51) \end{aligned}$$

$$= \max_{\substack{\sigma_i \geq 0 \\ \sum_{i=1}^M \sigma_i \leq 1}} \max_{\substack{\mathbf{Q}_i \succeq 0 \\ \text{Tr}\{\mathbf{Q}_i\} \leq \sigma_i}} \sum_{i=1}^M f_i(\mathbf{Q}_i) \quad (52)$$

$$= \max_{\substack{\sigma_i \geq 0 \\ \sum_{i=1}^M \sigma_i \leq 1}} \max_{\substack{\mathbf{Q}_i \succeq 0 \\ \text{Tr}\{\mathbf{Q}_i\} \leq 1}} \sum_{i=1}^M \sigma_i f_i(\mathbf{Q}_i) \quad (53)$$

$$= \max_{\substack{\sigma_i \geq 0 \\ \sum_{i=1}^M \sigma_i \leq 1}} \sum_{i=1}^M \sigma_i \max_{\substack{\mathbf{Q}_i \succeq 0 \\ \text{Tr}\{\mathbf{Q}_i\} \leq 1}} f_i(\mathbf{Q}_i) \quad (54)$$

$$= \max_i \max_{\substack{\mathbf{Q}_i \succeq 0 \\ \text{Tr}\{\mathbf{Q}_i\} \leq 1}} f_i(\mathbf{Q}_i), \quad (55)$$

where we have added slack variables in (52), normalized the traces in (53) as we did in (14)-(15). We then separated the inner maximizations in (54), and used the fact that the solution of (54) is just the maximum element in the sum as in (55). Finally, each of the inner maximizations in (55) is identical to that of (16) which we have already solved.

Interestingly, Theorem 3 shows that BF remains optimal for maximizing the worst case sum-rate even when a total power constraint is applied. In this case, the system chooses only one user that signals with all of the available power.

#### IV. 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. In this section, we will discuss this relation. In order to keep the notations simple, we will address the point to point case but the extension to the multiuser system is straight forward.

The most common stochastic CSI model is the complex Gaussian model<sup>1</sup> [2]:

$$\begin{aligned} \mathbf{H}_s &= \mathbf{h}_{r,s} \mathbf{h}_{t,s}^H + \mathbf{D}_s; \\ \mathbf{D}_s &= \tilde{\mathbf{D}}_s \mathbf{W}_s^{-\frac{1}{2}}, \end{aligned} \quad (56)$$

where  $\mathbf{h}_{r,s} \neq \mathbf{0}$  is a length  $N$  vector,  $\mathbf{h}_{t,s}$  is a length  $K$  vector,  $\tilde{\mathbf{D}}_s$  is a  $N \times K$  matrix with independent, zero mean, and unit variance complex Gaussian random variables,  $\mathbf{W}_s \succ 0$  is a  $K \times K$  matrix (which we assume invertible for simplicity). It is

<sup>1</sup>We use the subscript  $s$  to denote variables of the stochastic model as opposed to variables of the deterministic model in (2)

easy to see the resemblance between the deterministic model in (2) and the stochastic model in (56). 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

$$I_{\text{out}} = \text{OUT}(P_{\text{out}}), \quad (57)$$

where

$$P_{\text{out}} = \Pr(I(\mathbf{Q}, \mathbf{D}_s) \leq I_{\text{out}}). \quad (58)$$

The inverse is unique due to the monotonicity of the CDF. The meaning of (57) is that there is a probability of  $P_{\text{out}}$  that in any realization of  $\mathbf{H}$  from the ensemble, we will obtain a mutual information  $I$  less than  $I_{\text{out}}$ . Therefore, the system is designed to maximize the outage mutual information [4].

In general, the calculation of the outage capacity is very difficult. In [4], it was derived for the MISO case using integrals over 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  $\mathbf{Q} = p_1 \mathbf{h}_{r,s} \mathbf{h}_{t,s}^H + p_2 \mathbf{I}$  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  $\mathbf{h}_{r,s} \mathbf{h}_{t,s}^H$  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 (56). Then,

$$\text{OUT}(P_{\text{out}}) \geq C \left( \mathbf{h}_{r,s} \mathbf{h}_{t,s}^H, \frac{1}{\text{CDF}_{\chi_{2NK}^2}^{-1}(1 - P_{\text{out}})} \mathbf{W}_s \right), \quad (59)$$

where  $C(\cdot, \cdot)$  is defined in (6) of Theorem 1, and  $\text{CDF}_{\chi_{2NK}^2}^{-1}(\cdot)$  is the inverse cumulative distribution function (CDF) of a Chi Squared random variable with  $2NK$  degrees of freedom.

*Proof:* Let us define the event  $A_\alpha$  as the event when the realization of  $\mathbf{H}_s$  falls within the ellipsoid set defined in (2) with  $\mathbf{h}_r = \mathbf{h}_{r,s}$ ,  $\mathbf{h}_t = \mathbf{h}_{t,s}$  and  $\mathbf{W} = \alpha \mathbf{W}_s$ . The probability of this event is

$$\begin{aligned} \Pr(A_\alpha) &= \Pr \left( \text{Tr} \{ \mathbf{D}_s \mathbf{W}_s \mathbf{D}_s^H \} \leq \frac{1}{\alpha} \right) \\ &= \Pr \left( \text{Tr} \{ \tilde{\mathbf{D}}_s \tilde{\mathbf{D}}_s^H \} \leq \frac{1}{\alpha} \right) \\ &= \text{CDF}_{\chi_{2NK}^2} \left( \frac{1}{\alpha} \right). \end{aligned} \quad (60)$$

By conditioning on  $A_\alpha$ , we have

$$\begin{aligned} \Pr(I(\mathbf{Q}, \mathbf{D}_s) \leq C(\mathbf{X}_s, \alpha \mathbf{W}_s)) \\ &\leq 0 \cdot \Pr(A_\alpha) + 1 \cdot (1 - \Pr(A_\alpha)) \\ &= 1 - \text{CDF}_{\chi_{2NK}^2} \left( \frac{1}{\alpha} \right). \end{aligned} \quad (61)$$

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

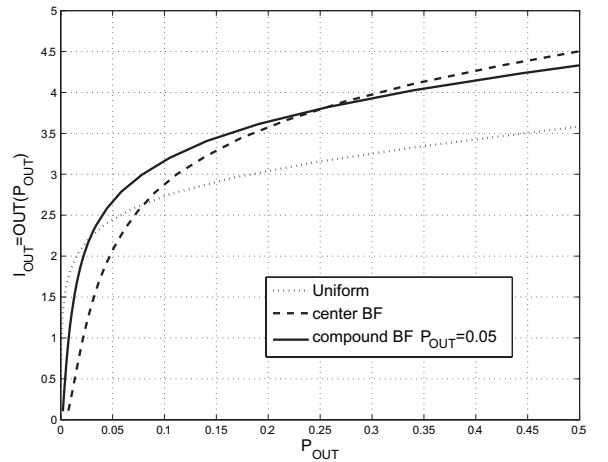


Fig. 1. Outage capacity as a function of the outage probability.

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  $P_{\text{out}}$ , one can solve the right hand side of (59) and then the optimal  $\mathbf{Q}$  will promise a lower bound on the outage mutual information of probability  $P_{\text{out}}$ . It is important to emphasize that this is only a lower bound which is not necessarily tight. For example, there may be circumstances in which the compound capacity is zero, but the outage capacity is not. Nonetheless, the proposed strategy 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.

#### A. Numerical example of the outage capacity

As explained, it is very difficult to handle the outage capacity analytically. Therefore, we now present a simple numerical example that illustrates the use of Lemma 1. We consider a system with  $K = 4$  transmit antennas and  $N = 1$  receive antennas. The transmitter has the following stochastic CSI: The channel is modeled as a random complex Gaussian vector with mean  $[1, 0, 0, 0]^H$  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 in Lemma 1 using the compound capacity. For comparison, we also simulate two other strategies: uniform power allocation, and BF 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.

## V. 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 BF, 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.

We also addressed the MAC compound capacity under a similar deterministic CSI model. We considered two problem formulations: MAC with individual power constraints, and MAC with a total power constraint. Assuming equal array response at the receiver, the optimal solution in both problems is based on BF. An interesting extension of this work is to use a more general CSI model. For example, one can relax the assumption of equal array response, or relax the rank one constraint on the mean matrix. In these cases, BF based solutions will not necessarily be optimal and therefore optimality conditions should be derived and analyzed.

## REFERENCES

- [1] E. Visotsky and U. Madhow, "Space time transmit precoding with imperfect feedback," *IEEE Trans. Inf. Theory*, vol. 47, no. 6, pp. 2632–2639, Sept. 2001.
- [2] S. A. Jafar and A. Goldsmith, "Transmitter optimization and optimality of beamforming form multiple antenna systems," *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp. 1165–1175, July 2004.
- [3] A. Soysal and S. Ulukus, "Transmit directions and optimality of beamforming in MIMO-MAC with partial CSI at the transmitters," in *Proc. The Johns Hopkins University Conference on Information Sciences and Systems (CISS-2005)*.
- [4] A. Moustakas and S. H. Simon, "Optimizing multi-transmitter single-receiver (MISO) antenna systems with partial channel knowledge," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2770–2780, Oct. 2003.
- [5] S. H. Simon and A. Moustakas, "Optimizing MIMO antenna systems with channel covariance feedback," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 3, pp. 406–417, April 2003.
- [6] 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. Conference on Acoustics, Speech, and Signal Processing (ICASSP-2003)*.
- [7] H. Boche and E. A. Jorswieck, "Outage probability of multiple antenna systems: optimal transmission and impact of correlation," in *Proc. IEEE Int. Zurich Seminar (IZS-2004)*.
- [8] M. Vu and A. Paulraj, "Optimum transmission scheme for a MISO wireless system with partial channel knowledge and infinite K factor," in *Proc. IEEE Int. Conf. on Communications (ICC) 2004*.
- [9] A. Pascual Iserte, A. I. Perez Neira, and M. A. Lagunas, "A maximin approach for robust MIMO design: combining OSTBC and beamforming with minimum transmission power requirements," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP-2004)*.
- [10] A. Pascual Iserte, D. P. Palomar, A. I. Perez Neira, and M. A. Lagunas, "A robust maximin approach for MIMO communications with partial channel state information based on convex optimization," *IEEE Trans. Signal Processing*, vol. 54, no. 1, pp. 346–360, Jan. 2006.
- [11] A. Abdel-Samad, A. Gershman, and T. Davidson, "Robust transmit eigen-beamforming based on imperfect channel state information," in *Proc. ITG Workshop on Smart Antennas (WSA-2004)*.
- [12] I. Csiszar and J. Korner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic Press, 1981.
- [13] A. Lapidoth and P. Narayan, "Reliable communication under channel uncertainty," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2148–2177, Oct. 1998.
- [14] E. Biglieri, J. Proakis, and S. Shamai (Shitz), "Fading channels: Information-theoretic and communications aspects," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2619–2692, Oct. 1998.
- [15] S. Vishwanath, S. Boyd, and A. Goldsmith, "Worst-case capacity of gaussian vector channels," in *Proc. 2003 Canadian Workshop on Information Theory*.
- [16] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, "Uniform power allocation in MIMO channels: a game-theoretic approach," *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1707–1727, July 2003.
- [17] S. Wei and D. Goeckel, "On the minimax robustness of the uniform transmission power strategy in MIMO systems," *IEEE Commun. Lett.*, vol. 7, no. 11, pp. 523–524, Nov. 2003.
- [18] M. Godavarti, A. O. Hero, and T. L. Marzetta, "Min-capacity of a multiple-antenna wireless channel in a static Ricean fading environment," *IEEE Trans. Wireless Commun.*, vol. 4, no. 4, pp. 1715–1723, July 2005.

- [19] W. Rhee, W. Yu, and J. Cioffi, "The optimality of beam-forming in uplink multiuser wireless systems," *IEEE Trans. Wireless Commun.*, vol. 3, no. 1, pp. 86–96, Jan. 2004.
- [20] M. Sion, "On general minimax theorems," *Pac. J. Math*, vol. 8, pp. 1711–176, 1958.
- [21] 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.
- [22] Wei Yu, "A dual decomposition approach to the sum power gaussian vector multiple access channel sum capacity problem," in *Proc. The Johns Hopkins University Conference on Information Sciences and Systems (CISS-2003)*.



**Ami Wiesel (S02)** received the B.Sc. degree in electrical engineering and the M.Sc. degree in electrical engineering from Tel-Aviv University (TAU), Tel-Aviv, Israel, in 2000 and 2002, respectively. He is currently working toward the Ph.D. degree at the Department of Electrical Engineering, The Technion-Israel Institute of Technology, Haifa, Israel.

From June 2002 to March 2003, he was with the Department of Signal Theory and Communications of the Universitat Politècnica de Catalunya, Barcelona, Spain. His main areas of interests include

signal processing algorithms, convex optimization methods, and detection and estimation theory.

Mr. Wiesel was awarded the Weinstein Study Prize in 2002 and the Intel Award in 2005. He received the Student Paper Award for an SPAWC-2005 paper.



**Yonina C. Eldar (S'98–M'02)** received the B.Sc. degree in Physics in 1995 and the B.Sc. degree in Electrical Engineering in 1996 both from Tel-Aviv University (TAU), Tel-Aviv, Israel, and the Ph.D. degree in Electrical Engineering and Computer Science in 2001 from the Massachusetts Institute of Technology (MIT), Cambridge.

From January 2002 to July 2002 she was a Postdoctoral Fellow at the Digital Signal Processing Group at MIT. She is currently an Associate Professor in the Department of Electrical engineering at the Technion - Israel Institute of Technology, Haifa, Israel. She is also a Research Affiliate with the Research Laboratory of Electronics at MIT.

From 1992 through 1996 she was in the program for outstanding students in TAU. In 1998 she held the Rosenblith Fellowship for study in Electrical Engineering at MIT, and in 2000 she held an IBM Research Fellowship. She was a Horev Fellow in the Leaders in Science and Technology program at the Technion, and an Alon Fellow. In 2004, she was awarded the Wolf Foundation Krill Prize for Excellence in Scientific Research, and in 2005 the Andre and Bella Meyer Lectureship.

She is a member of the IEEE Signal Processing Theory and Methods technical committee, an Associate Editor for the *IEEE Transactions on Signal Processing* and the *EURASIP Journal on Signal Processing*, and on the Editorial Board of *Foundations and Trends in Signal Processing*.



**Shlomo Shamai (Shitz)** received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from the Technion—Israel Institute of Technology, in 1975, 1981 and 1986 respectively.

During 1975–1985 he was with the Communications Research Labs in the capacity of a Senior Research Engineer. Since 1986 he is with the Department of Electrical Engineering, Technion—Israel Institute of Technology, where he is now the William Fondiller Professor of Telecommunications. His research interests encompasses a wide spectrum

of topics in information theory and statistical communications.

Dr. Shamai (Shitz) is an IEEE Fellow and a member of the Union Radio Scientifique Internationale (URSI). He is the recipient of the 1999 van der Pol Gold Medal of URSI, and a co-recipient of the 2000 IEEE Donald G. Fink Prize Paper Award, the 2003, and the 2004 joint IT/COM societies paper award. He is also the recipient of 1985 Alon Grant for distinguished young scientists and the 2000 Technion Henry Taub Prize for Excellence in Research. He has served as Associate Editor for the Shannon Theory of the *IEEE Transactions on Information Theory*, and also serves on the Board of Governors of the Information Theory Society.