Multiple Antenna Channels With Partial Channel State Information at the Transmitter

June Chul Roh, *Student Member, IEEE,* and Bhaskar D. Rao, *Fellow, IEEE*

***Abstract*—We investigate transmission strategies for flat-fading multiple antenna channels with t transmit and r receive antennas, and with channel state information (CSI) partially known to the transmitter. We start with an assumption that the first n eigenvectors of $H^{\dagger}H$, where $0 \leq n \leq \min(t, r)$ and H is the channel matrix in $\mathbb{C}^{r \times t}$, are available at the transmitter as partial spatial information of the channel. A beamforming method is proposed in which a beamforming matrix is determined from the n eigenvectors in some predefined way; as a result, the receiver also knows the beamforming matrix. With this beamforming scheme, we develop a new multiple antenna system concept that provides a mechanism to reduce the amount of channel feedback information. This paper focuses on deriving the channel capacity of the multiple antenna channels employing the proposed beamforming and feedback methods. An important task for achieving capacity is the solution of interesting optimization problems for the optimal power allocation over the transmit symbols. The results show that the proposed methods lead to systems wherein the amount of feedback information can be significantly reduced with a minor sacrifice of achievable transmission rate.**

***Index Terms*—Channel capacity, channel state information (CSI), multiple-input multiple-output (MIMO) systems, multiple antennas, power allocation, transmit beamforming, wireless communication.**

I. Introduction

In recent years, communication systems with multiple antennas at both the transmitter and the receiver have gathered much attention for high-rate data transmission. The information-theoretical capacity of the multiple antenna channels has been studied by many researchers, immediately following the promising results by Telatar [1] and Foschini [2]. Many previous studies have focused on the following two assumptions about channel state information (CSI): the first is the case where CSI is known to both the receiver and the transmitter [1], [3]; and the second is where CSI is available only at the receiver, not at the transmitter [1], [2], [4]. We will refer to the former as *complete CSIT* and the latter as *no CSIT*[1] ; these will be used as the two references in comparing channel capacities.

We remark that there are gaps between the capacities of the two cases, in particular, when the transmit power is relatively low, or when the number of transmit antennas is greater than the number of receive antennas. In order to achieve the higher capacity of the complete CSIT, the transmitter should perfectly know instantaneous channel information. In general, the channel state information at the transmitter, even *partial* information, can be utilized to increase the channel capacity. This research was motivated by a natural insight that there is a tradeoff between the improvement in channel capacity and the degree of completeness of the CSI available at the transmitter. In practical situations, particularly, in systems with a feedback channel for the channel state information, the amount of channel information that is required at the transmitter can be too large to handle, since the channel has $t \times r$ number of fading parameters. In this paper, we consider the cases where the channel information is *partially* known to the transmitter in a way that enables a reduction in the amount of the feedback information.

There were several studies that have considered partial CSIT for different systems or in different forms of CSIT other than what this paper considers. We mention a few works in this content. Caire and Shamai [5] investigated single-input single-output channels. Multiple-input single-output (MISO) channels were studied by Visotsky and Madhow [6] considering the second order statistics of channel as partial CSIT. Jafar *et al.* [7] considered extension of Visotsky's result to multiple-input multiple-output (MIMO) channels, Bhashyam, Sabharwal, and Aazhang [8] explored MISO channels considering the received SNR as partial channel information. Beamforming combined with space-time coding in MIMO channels were investigated by Jongren, Skoglund and Ottersten [9] where an estimate of the channel realization is assumed to be available at the transmitter.

There are many applications in which there exists a feedback channel for the channel state information. However, in many real systems, the channel information can not be fully provided to the transmitter, for example, due to a limited transmission capacity of feedback channel or rapid channel variation. In designing such systems, it is important to determine what type of the channel information to feed back while minimizing the loss of channel capacity. In this paper, we consider flat-fading channels with t transmit and r receive antennas which is modeled by an $r \times t$ complex matrix H. We propose a beamforming method in which the beamforming matrix is determined from a subset of the eigenvectors of $H^{\dagger}H$ in some predefined way; as a result, the receiver also knows the beamforming matrix. With this beamforming scheme, we introduce a new multiple antenna

Manuscript received October 19, 2002; revised February 3, 2003. The editor coordinating the review of this paper and approving it for publication is M. Shafi. This work was supported by a CoRe Research Grant Cor00-10074 and by a Research Grant from Ericsson. The material in this paper was presented in part at the 36th Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, November 2002, and at the IEEE International Conference on Communications, Anchorage, AK, May 2003.

The authors are with the Department of Electrical and Computer Engineering, University of California at San Diego, La Jolla, CA 92093 USA (e-mail: jroh@ece.ucsd.edu; brao@ece.ucsd.edu).

Digital Object Identifier 10.1109/TWC.2003.821144

[1]CSIT: channel state information at the transmitter. In this paper, we assume that in all cases the channel information is completely known to the receiver.

system concept that provides a mechanism to reduce the amount of channel feedback information. It is shown that this system concept leads to schemes wherein the amount of feedback information can be significantly reduced with a minor sacrifice of achievable transmission rate.

In fast time-varying channels, it may be more realistic that averaged channel information is provided to the transmitter as in [10]. Since the basic principle in beamforming and feedback of channel information except averaging the channel in obtaining the feedback information, the proposed beamforming and system concept can be adapted in that situation. Therefore, this paper focuses on the cases that feedback the instantaneous channel information.

This paper is organized as follows. In Section II, the channel model is described and the capacity results for complete and no CSIT are summarized. The new beamforming method and a novel multiple antenna system concept are introduced in Section III. It is shown that a MIMO channel can be decomposed into two parts, n parallel independent channels and a new smaller coupled MIMO channel. In Sections IV and V, the channel capacities for *partial CSIT* are derived under two different conditions: i) unequal and ii) equal power allocated over the symbols for the coupled MIMO channel. Numerical results are presented in Section VI.

II. System Model and Background

A. Channel Model

We consider multiple antenna systems with t antennas at the transmitter and r at the receiver. Assuming slow flat-fading, the MIMO channel is modeled by the channel matrix $H \in \mathbb{C}^{r\times t}$. That is, the channel input $x \in \mathbb{C}^t$ and the channel output $y \in \mathbb{C}^r$ have the following relationship:

$$y = Hx + \eta \tag{1}$$

where $\eta \in \mathbb{C}^r$ is the complex additive white Gaussian noise (AWGN) vector with each element being assumed *i.i.d.* complex Gaussian random variable with zero-mean and unit variance, i.e., $E\{\eta\eta^\dagger\} = I_r$, where $E\{\cdot\}$ denotes the expectation operation and I_r is the $r \times r$ identity matrix. We denote the rank of H by m. And the singular value decomposition (SVD) of H is given by $H = U\Sigma V^\dagger$, where $A^\dagger$ denotes the conjugate transpose of a matrix A; unitary matrices $V \in \mathbb{C}^{t\times t}$ and $U \in \mathbb{C}^{r\times r}$ span the input space $\mathbb{C}^t$ and the output space $\mathbb{C}^r$, respectively; and $\Sigma \in \mathbb{R}^{r\times t}$ contains the singular values with σ_i representing the ith singular value of H and $\sigma_1 \geq \cdots \geq \sigma_m > 0$. We impose a constraint on the transmit power, $E\{x^\dagger x\} \leq P_T$.

In this paper, we assume that in all cases perfect CSI is known to the receiver. In addition, it is assumed that the transmitter knows the first n column vectors of V, where $0 \leq n \leq m$, or the first n eigenvectors of $H^\dagger H$, as partial *spatial* information of the channel. This assumption includes the two extreme cases: i) $n = m$ is the case that the transmitter has same spatial information as in the *complete CSIT* case; and ii) $n = 0$ accounts that no spatial information is available at the transmitter as in the *no CSIT* case. This paper mainly considers the cases of $0 < n < m$; these corresponds to *partial CSIT* cases. For notational convenience, let us define $V_1 = [v_1, \ldots, v_n]$ where v_i is the ith column vector of V, and $V_2 = [v_{n+1}, \ldots, v_t]$, i.e., $V = [V_1, V_2]$.

B. Ergodic Channel Capacity

We consider the ergodic capacity as a performance measure. More details of this section can be found in [1]. The ergodic capacity of a random MIMO channel with transmit power constraint P_T is given by

$$C(P_T) = E_H\{C(P_T; H)\}$$

where $E_H\{\cdot\}$ indicates the expectation over channel realizations; and $C(P_T; H)$ is the conditional capacity for a given channel realization H with a power constraint P_T. That is

$$C(P_T; H) = \max_{p(x):E\{x^\dagger x\}\leq P_T} I(x; y).$$

The mutual information $I(x; y)$ satisfies the following inequality:

$$I(x; y) \leq \log\det(I_r + H\Phi_x H^\dagger) \tag{2}$$

where $\log(\cdot)$ is the base-2 logarithm, $\det(A)$ denotes the determinant of A, $\Phi_x = E\{xx^\dagger\}$, and the equality holds if and only if x is a circularly symmetric complex Gaussian random vector. In summary, the ergodic channel capacity is expressed as

$$C(P_T) = E_H\left\{\max_{\mathrm{tr}\Phi_x \leq P_T} \log\det(I_r + H\Phi_x H^\dagger)\right\}$$

C. MIMO Channels With Complete and No CSIT

Let us denote by C_{HH} the capacity of MIMO channels with CSI fully known to both the transmitter and the receiver (complete CSIT); by $C_{\phi H}$ the capacity with CSI known only to the receiver (no CSIT). When the transmitter knows the channel information, the optimum power allocation can be solved by water-filling [11] over m independent spatial channels [1]. The capacity is given by

$$\begin{aligned} C_{HH}(P_T) &= E_H\left\{\max_{\substack{P_1\geq 0,\ldots,P_m\geq 0\\ P_1+\cdots+P_m\leq P_T}} \sum_{i=1}^{m} \log(1 + P_i\lambda_i)\right\} \\ &= E_H\left\{\sum_{i=1}^{m}[\log(\nu\lambda_i)]^+\right\} \end{aligned} \tag{3}$$

where $\lambda_i = \sigma_i^2$ is the ith largest eigenvalue of $H^\dagger H$ (or $HH^\dagger$), and $\{P_i\}$ are the transmit powers allocated on the transmit symbol $s \in \mathbb{C}^t$ ($x = Vs$), i.e., $\Phi_s = E\{ss^\dagger\} = \mathrm{diag}(P_1, \ldots, P_t)$. $[a]^+$ is defined as $\max\{a, 0\}$ and ν is the level of water-filling satisfying the power constraint

$$\sum_{i=1}^{m}\left[\nu - \frac{1}{\lambda_i}\right]^+ = P_T.$$

On the other hand, when the transmitter has no knowledge about the channel, it is optimal to use an equal power allocation [1], i.e., $P_i = P_T/t$ for $1 \leq i \leq t$. Then, the capacity is given by

$$C_{\phi H}(P_T) = E_H\left\{\sum_{i=1}^{m}\log\left(1 + \frac{P_T}{t}\lambda_i\right)\right\}. \tag{4}$$

The difference between C_{HH} and $C_{\phi H}$ at high transmit power is summarized in the following Lemma.

Lemma 1: At high transmit power P_T (to satisfy $P_T \gg t/\lambda_i$ for all $1 \leq i \leq m$)

$$C_{HH} - C_{\phi H} \simeq m \log \frac{t}{m} \begin{cases} = 0, & \text{when } t = m \\ > 0, & \text{when } t > m \end{cases} \quad (5)$$

where $m = \min(t, r)$ if H is full rank. □

Proof: See Appendix A.

That is, when $t = m$, the two capacities (C_{HH} and $C_{\phi H}$) is asymptotically same; and when $t > m$, and also when $t > r$, there exist fundamental gaps between C_{HH} and $C_{\phi H}$ even as $P_T \to \infty$.

III. The Beamforming Method and New System Concept

A. Extended Maximal-Ratio Transmission

In a system where *single* data stream is transmitted over t transmit antennas after passing through a beamformer $w \in \mathbb{C}^t$, the optimum choice of w is the first eigenvector of $H^\dagger H$, or, v_1. This transmission scheme is called maximal-ratio transmission (MRT). The choice of the beamformer as $w = v_1$ is optimal in terms of maximizing the received signal-to-noise ratio [12]. In addition, it can be easily shown that the beamformer choice of $w = v_1$ is also optimal in the sense of maximizing the mutual information.

A generalization of the MRT for a system with $n > 1$ data streams is to employ n different beamformers for each data stream. In this scheme, the transmitted signal $x \in \mathbb{C}^t$ can be modeled as

$$x = w_1 s_1 + \cdots + w_n s_n = Ws; \quad |w_i| = 1, \quad 1 \leq i \leq n \quad (6)$$

where $W = [w_1, \ldots, w_n] \in \mathbb{C}^{t \times n}$ is referred to as beamformer matrix, $s = [s_1, \ldots, s_n]^T \in \mathbb{C}^n$, and $|\cdot|$ is the l_2 (Euclidean) norm on $\mathbb{C}^n$. In [13], it was discussed that the optimum beamformer, for the case when $n < m$, is likely to be V_1. The transmitter uses only the n known spatial channels, that is, transmitting n data streams using the n eigenvectors. We will call this scheme *extended MRT*. Since the receiver knows the channel parameters, the channel can be decomposed into n parallel channels with different channel gains. As a result, with this strategy, one can employ conventional *scalar* coding to each spatial channel. But, because the inherent multiplexing capability of the multiple antenna channel is not fully exploited, it will be shown later that this strategy is inferior to the transmission strategies we propose in the following.

B. The Beamforming Method

To fully exploit potential multiplexing capability of the channel, we propose a new and improved beamforming method that also utilizes the orthogonal complement of the space spanned by V_1. A beamforming matrix $W \in \mathbb{C}^{t \times t}$ is generated as a function of V_1 in a predefined manner. Since the receiver has knowledge of V_1, the receiver is also aware of the beamforming matrix that the transmitter will use. This property enables us to conceive of a new multiple antenna system concept which is described in Section III-C. One reasonable way to generate the beamforming matrix is the following:

1) Choose $t - n$ column vectors, namely, $\tilde{V}_2 = [\tilde{v}_{n+1}, \ldots, \tilde{v}_t]$, that are mutually orthogonal and also orthogonal to the space spanned by V_1, i.e.,

$$\tilde{V}_2^\dagger \tilde{V}_2 = I_{t-n}, \quad {V_1}^\dagger \tilde{V}_2 = 0 \quad (7)$$

where I_p is $p \times p$ identity matrix and 0 is $n \times (t-n)$ zero matrix.

2) Concatenate $\tilde{V}_2$ to V_1 to form a beamforming matrix $W = [V_1, \tilde{V}_2]$.

It can be easily shown that, if H is full rank, W spans the same input space as V does. The beamforming matrix W is used in transmitting the information vector $s \in \mathbb{C}^t$ in a manner similar to the use of V in the complete CSIT case. The procedure for selecting $\tilde{V}_2$ satisfying (7) can be defined in various ways, e.g., $\tilde{V}_2$ are the eigenvectors corresponding to the nonzero eigenvalues of $I_t - V_1 V_1^\dagger$. Whatever be the mechanism for generating $\tilde{V}_2$ at the transmitter, the generating mechanism is assumed to be known at the receiver so that the receiver can also independently generate $\tilde{V}_2$ and, hence, W.

C. New Multiple Antenna System Concept

With the proposed beamforming scheme, we develop a new multiple antenna system concept that can potentially lead to a reduction in the amount of channel feedback information. It involves

1) Based on W, calculation for optimal power allocation over transmit symbols is performed at the receiver.
2) The power allocation result is provided to the transmitter as additional CSIT.

The first step will be described in detail in Sections IV and V. Two approaches will be discussed, each of which results in t and $n+1$ real values, respectively. These values are bounded between 0 and 1, and sum up to be 1. Then, the total channel feedback information is n t-dimensional complex vectors, i.e., V_1, plus t (for the first approach) or $n+1$ (for the second approach) real values in $[0, 1]$. Thus, in most systems, in particular, when the number of transmit antennas t is large, the amount of feedback information can be significantly reduced.

Table I summarizes the transmission strategies that will be used as references. Table II is a summary of the proposed transmission strategies including the spatial information available at the transmitter, the beamforming matrix, and the power allocation results as additional CSIT. The first scheme and corresponding capacity $C_{V_1 H}^{(\mathrm{opt})}$ in Table I will be dealt in Section IV, the second is a special case of the first ($n = 0$), and the third scheme denoted by $C_{V_1 H}$ will be discussed in Section V.

As shown in the tables, in the proposed schemes, the amount of feedback information required at the transmitter is more than in the no CSIT case and much less than in the complete CSIT case; therefore, it is expected that the capacities of the proposed schemes lie between the two extremes. But, it will be shown later that at moderate transmit power region, the proposed schemes achieve the most gain of channel knowledge of the complete CSIT. Comparing with the extended MRT scheme, the capacity is much higher particularly at high transmit power, although the amount of feedback information is comparable. This results from the transmission strategy that utilizes the or-

TABLE I
SUMMARY OF THE TRANSMISSION STRATEGIES: REFERENCE SCHEMES

	Complete CSIT	No CSIT	Extended MRT
Capacity notation	C_{HH}	$C_{\phi H}$	$C_{V_1 H}^{(\mathrm{EMRT})}$
CSIT (spatial info.)	$V = [v_1, \ldots, v_t]$	$-$	$V_1 = [v_1, \ldots, v_n]$
Beamforming	$W = V$	$W = I$	$W = V_1$
Additional CSIT	$(\gamma_1, \ldots, \gamma_t)$	$-$	$(\gamma_1, \ldots, \gamma_n)$
Power allocation	$P_i = \gamma_i P_T,\ 1 \le i \le t$	$P_i = P_T/t,\ 1 \le i \le t$	$P_i = \gamma_i P_T,\ 1 \le i \le n$

info.: information; EMRT: extended MRT.

$\gamma_i = P_i/P_T$, therefore, $0 \le \gamma_i \le 1\ \forall i$ and $\sum_i \gamma_i = 1$.

TABLE II
SUMMARY OF THE TRANSMISSION STRATEGIES: PROPOSED SCHEMES

	Partial CSIT		
Capacity notation	$C_{V_1 H}^{(\mathrm{opt})}$	$C_{\phi H}^{(\mathrm{opt})}$	$C_{V_1 H}$
CSIT (spatial info.)	$V_1 = [v_1, \ldots, v_n]$	$-$	$V_1 = [v_1, \ldots, v_n],\ n < t$
Beamforming	$W = [V_1, \tilde{V}_2]$	$W = I$	$W = [V_1, \tilde{V}_2]$
Additional CSIT	$(\gamma_1, \ldots, \gamma_t)$	$(\gamma_1, \ldots, \gamma_t)$	$(\gamma_1, \ldots, \gamma_n, \gamma_{n+1})$
Power allocation	$P_i = \gamma_i P_T,\ 1 \le i \le t$	$P_i = \gamma_i P_T,\ 1 \le i \le t$	$P_i = \gamma_i P_T,\ 1 \le i \le n$; $P_i = \frac{\gamma_{n+1}}{t-n} P_T,\ n+1 \le i \le t$

thogonal complement space, i.e., H_2 channel, which can be obtained from the partial spatial information at the transmitter.

D. Channel Decomposition

Now, we will show that by using the proposed beamforming method the original MIMO channel is decomposed into two parts: n parallel independent channels, and a new smaller coupled MIMO channel. The transmitted signal x is given by

$$x = Ws = [V_1, \tilde{V}_2] \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} = V_1 s_1 + \tilde{V}_2 s_2$$

where $s = [s_1, s_2]^T \in \mathbb{C}^t$, $s_1 \in \mathbb{C}^n$, and $s_2 \in \mathbb{C}^{t-n}$.

The receiver pre-multiplies the received signal $y = Hx + \eta$ by $U^\dagger$ to have $\tilde{y} = U^\dagger y$. Using the partitioned matrices of compatible size to $W = [V_1, \tilde{V}_2]$, $\tilde{y}$ can be written as follows:

$$\begin{aligned} \tilde{y} = \begin{bmatrix} \tilde{y}_1 \\ \tilde{y}_2 \end{bmatrix} &= \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} \begin{bmatrix} V_1^\dagger V_1 & V_1^\dagger \tilde{V}_2 \\ V_2^\dagger V_1 & V_2^\dagger \tilde{V}_2 \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} + \begin{bmatrix} \tilde{\eta}_1 \\ \tilde{\eta}_2 \end{bmatrix} \\ &= \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 V_2^\dagger \tilde{V}_2 \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} + \begin{bmatrix} \tilde{\eta}_1 \\ \tilde{\eta}_2 \end{bmatrix} \end{aligned} \tag{8}$$

where $\tilde{y}_1 \in \mathbb{C}^n$, $\tilde{y}_2 \in \mathbb{C}^{r-n}$, diagonal matrices $\Sigma_1 \in \mathbb{R}^{n \times n}$ and $\Sigma_2 \in \mathbb{R}^{(r-n)\times(t-n)}$ contain $\sigma_1, \ldots, \sigma_n$ and $\sigma_{n+1}, \ldots, \sigma_m$, respectively, and the zero matrices are of suitable size. Equation (8) results from the facts that $V_1^\dagger V_1 = I_n$, $V_1^\dagger \tilde{V}_2 = 0$, and $V_2^\dagger V_1 = 0$.

We can see that the MIMO channel has been decomposed into n noninterfering parallel channels and a new coupled MIMO channel with a channel matrix $H_2 = \Sigma_2 V_2^\dagger \tilde{V}_2$ in $\mathbb{C}^{(r-n)\times(t-n)}$. That is

$$\tilde{y}_1 = \Sigma_1 s_1 + \tilde{\eta}_1 \tag{9}$$
$$\tilde{y}_2 = H_2 s_2 + \tilde{\eta}_2. \tag{10}$$

We will refer to the first channel of (9) as the Σ_1 channel, and the second channel of (10) as the H_2 channel. Note that the covariance of $\tilde{\eta} = U^\dagger \eta = [\tilde{\eta}_1^T, \tilde{\eta}_2^T]^T$ is unchanged as $E\{\tilde{\eta}\tilde{\eta}^\dagger\} = I_r$. An interesting property about the singular values of the channel matrix is summarized in the following Lemma.

Lemma 2: The singular values of the channel matrix $H_2 = \Sigma_2 V_2^\dagger \tilde{V}_2$ is preserved as $\mathrm{diag}(\Sigma_2)$.

Proof: See Appendix B. □

By the following Lemma, we show that the mutual information $I(x; y)$ is preserved with the linear operations $x = Ws$ and $\tilde{y} = U^\dagger y$. Furthermore, $I(x; y)$ can be given by the sum of the mutual information expressions for two decomposed channels.

Lemma 3: For a given channel realization H, the mutual information between the input and the output of the MIMO channel can be expressed as

$$I(x; y) = I(s; \tilde{y}) \tag{11}$$
$$= I(s_1; \tilde{y}_1) + I(s_2; \tilde{y}_2). \tag{12}$$

Proof: See Appendix C. □

IV. MIMO CHANNELS WITH PARTIAL CSIT: OPTIMUM TRANSMISSION STRATEGY

Here and in Sections V and VI, we derive the ergodic channel capacity and the optimum transmit power allocation schemes for the MIMO system with partial CSIT described in Sections II and III.

As we discussed in Sections II and III, by using a beamforming matrix generated in a predefined way, the beamforming matrix W is also known to the receiver. We discuss the transmission schemes and accompanying optimum power allocation solutions that take advantage of receiver's knowledge of the channel. For simplicity, we define an equivalent channel matrix $A = U^\dagger HW$, or $\Sigma V^\dagger W$, which represents the channel between s and $\tilde{y}$, i.e., $\tilde{y} = As + \tilde{\eta}$. By Lemma 3 , maximizing $I(x;y)$ is equivalent to maximizing $I(s;\tilde{y})$. Thus, since $I(s;\tilde{y}) \le \log\det(I_r + A\Phi_s A^\dagger)$, we have

$$C_{V_1H}^{(\mathrm{opt})}(P_T;H) = \max_{\substack{P_1\ge 0,\ldots,P_t\ge 0\\ P_1+\cdots+P_t=P_T}} \log\det(I_r + A\Phi_s A^\dagger). \quad (13)$$

The maximization problem is complex, and except in some special cases it requires the use of numerical optimization methods. Fortunately, this optimization problem is matched to the so-called *determinant maximization* problem [14] and can be solved numerically by using MAXDET algorithm [15]. The optimal power allocation for the H_2 channel, in general cases, turned out to be unequal even for the symbols over the H_2 channel. In this scenario, the channel feedback information is V_1 and the power allocation $(\gamma_1,\ldots,\gamma_t)$, where γ_i is defined as P_i/P_T. Compared to the equal power allocation cases which will be discussed in Section V, although this optimal scheme requires complicated calculation for the power allocation and a little bit more feedback information, i.e., $(t-n-1)$ real numbers in $[0,1]$, it provides higher channel capacity. We can find closed-form expressions in the following practically important cases.

A. Case I: When $r = 1$ and $n = 0$

In this case, since no eigenvector of $H^\dagger H$ is provided to the transmitter, we consider a natural choice of the beamforming matrix, $W = I_t$. Then, since the equivalent channel is now a t-dimensional row vector denoted by $A = [a_1,\ldots,a_t]$, the maximization problem of (13) becomes a simple form of the following:

$$C_{\phi H}^{(\mathrm{opt})}(P_T;H) = \max_{\substack{P_1\ge 0,\ldots,P_t\ge 0\\ P_1+\cdots+P_t=P_T}} \log\left(1+\sum_{i=1}^{t}|a_i|^2P_i\right).$$

By the monotonicity of log function and using a well-known solution in conventional linear programming, we have the following solution

$$C_{\phi H}^{(\mathrm{opt})}(P_T;H) = \log\left(1+|a_{i^*}|^2P_T\right) \quad (14)$$

where $i^* = \arg\max_i |a_i|^2$. This means that the total transmit power should be allocated such that $P_{i^*} = P_T$ and $P_i = 0$ for all $i \ne i^*$.

The analysis indicates that if the index for i^* is available at the transmitter via some feedback we can obtain higher channel capacity than that of no CSIT case. This can be interpreted as *transmit antenna selection* method in which at a given time only one antenna that provides the best link to the receiver is used in transmitting data.

B. Case II: When $t \ge r$ and $n = r-1$

This is a generalization of Case I. An example of this case is $r = 2, n = 1$, and $t \ge 2$. In this case, the equivalent channel A is given by

$$A = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0\\ 0 & \sigma_2 & 0 & \vdots & \vdots & \vdots & \vdots & 0\\ 0 & 0 & \ddots & 0 & 0 & 0 & \cdots & 0\\ 0 & \cdots & 0 & \sigma_{r-1} & 0 & 0 & \cdots & 0\\ 0 & \cdots & 0 & 0 & \sigma_r\alpha_{r,r} & \sigma_r\alpha_{r,r+1} & \cdots & \sigma_r\alpha_{r,t} \end{bmatrix}$$

where $\alpha_{i,j}$ is defined as the inner product between the ith column of V and the jth column of W, i.e., $\alpha_{r,j} = v_r^\dagger \tilde{v}_j$ for $r \le j \le t$. Therefore, the channel can be expresses as follows:

$$\tilde{y}_i = \begin{cases} \sigma_i s_i + \tilde{\eta}_i, & i = 1,\ldots,r-1\\ \sigma_r(\alpha_{r,r}s_r + \cdots + \alpha_{r,t}s_t) + \tilde{\eta}_r, & i = r \end{cases}$$

where $\tilde{y}_i$ denotes the ith component of the r-dimensional vector $\tilde{y}$ and s_i the ith symbol of s.[2]

Then, applying the same linear optimization technique as in Section IV-A to the last MISO channel (the input are $s_r,\ldots,s_t$ and the output is $\tilde{y}_r$), the maximization problem of (13) becomes the following:

$$\begin{aligned} &C_{V_1H}^{(\mathrm{opt})}(P_T;H)\\ &= \max_{\substack{P_1\ge 0,\ldots,P_{r-1}\ge 0,P_{j^*}\ge 0\\ P_1+\cdots+P_{r-1}+P_{j^*}=P_T}} \left\{\sum_{i=1}^{r-1}\log(1+\sigma_i^2P_i)\right.\\ &\qquad\left. + \log(1+\sigma_r^2|\alpha_{r,j^*}|^2P_{j^*})\right\} \end{aligned} \quad (15)$$

where $j^* = \arg\max_{r\le j\le t}|\alpha_{r,j}|^2$. Notice that the channel we consider consists of r parallel Gaussian channels with each channel having a channel gain, $\sigma_1,\ldots,\sigma_{r-1},\sigma_r|\alpha_{r,j^*}|$, respectively. Thus, the conditional channel capacity can be solved by conventional water-filling over the r noisy channels with equivalent noise levels of $(1/\sigma_1^2,\ldots,1/\sigma_{r-1}^2,1/\sigma_r^2|\alpha_{r,j^*}|^2)$.

The result implies that the spatial information to feed back is V_1 and the index for j^* that indicates the antenna, among the antenna set of rth to tth antenna, that provides the best link to last channel output $\tilde{y}_r$.

V. MIMO Channels With Partial CSIT: Suboptimum Transmission Strategy

By using Lemma 3, the conditional channel capacity can be expressed as follows:

$$\begin{aligned} &C_{V_1H}(P_T;H)\\ &= \max_{\substack{P_{T,1}\ge 0,P_{T,2}\ge 0\\ P_{T,1}+P_{T,2}\le P_T}} \{C(P_{T,1};\Sigma_1) + C(P_{T,2};H_2)\} \end{aligned} \quad (16)$$

where $P_{T,1}$ is the transmit power allocated to the Σ_1 channel of (9), and $P_{T,2}$ is the transmit power on the H_2 channel of (10). $C(P_{T,1};\Sigma_1)$ is the conditional channel capacity of the Σ_1 channel with transmit power $P_{T,1}$. Because the Σ_1 channel consists of n parallel Gaussian channels, for a given $P_{T,1}$, the capacity and the optimum power allocation can be obtained similarly to the complete CSIT case of (3). That is

$$\begin{aligned} C(P_{T,1};\Sigma_1) &= \max_{p(s_1):E\{s_1^\dagger s_1\}\le P_{T,1}} I(s_1;\tilde{y}_1)\\ &= \max_{\substack{P_1\ge 0,\ldots,P_n\ge 0\\ P_1+\cdots+P_n\le P_{T,1}}} \sum_{i=1}^{n}\log(1+P_i\lambda_i). \end{aligned} \quad (17)$$

The second term $C(P_{T,2};H_2)$ in (16) is the conditional channel capacity of the H_2 channel with transmit power $P_{T,2}$. In this Section, we confine our attention to a practically reasonable transmission strategy: an eqal power allocation for the H_2 channel. Compared to the optimum scheme in Section IV, the

[2]Note the definitions of $\tilde{y}_i$ and s_i are different from those in Section III.

power allocation results in $n+1$ real values in $[0,1]$; therefore, the amount of channel feedback information has been reduced, which is one of advantages of this transmission strategy. The analysis of this transmission scenario is also meaningful because it explains the limiting performance of the systems that comprises of n parallel channels (the Σ_1 channel) from beamforming, for which conventional time-domain only codes would be used; and a MIMO channel (the H_2 channel), for which a space-time code would be employed. In other words, this section assumes that the transmitter has no information about the H_2 channel except the total transmit power $P_{T,2}$ for the channel. Then, from (4), the conditional capacity expression is given by

$$C(P_{T,2};H_2) = \max_{p(s_2):E\{s_2^\dagger s_2\}\le P_{T,2}} I(s_2;\tilde{y}_2) = \sum_{i=n+1}^{m} \log\left(1+\frac{P_{T,2}}{t-n}\lambda_i\right). \quad (18)$$

Combining (17) and (18) with (16), the conditional capacity for a given channel realization H is obtained by solving the following maximization problem

$$C_{V_1H}(P_T;H) = \max_{\substack{P_1\ge 0,\ldots,P_n\ge 0,P_{T,2}\ge 0\\ P_1+\cdots+P_n+P_{T,2}\le P_T}} \Psi(P_1,\ldots,P_n,P_{T,2}) \quad (19)$$

where $\Psi(\cdot)$ is defined as

$$\Psi(P_1,\ldots,P_n,P_{T,2}) = \sum_{i=1}^{n}\log(1+P_i\lambda_i) + \sum_{i=n+1}^{m}\log\left(1+\frac{P_{T,2}}{t-n}\lambda_i\right). \quad (20)$$

We can write the constraint maximization using Lagrange multipliers as the maximization of

$$J = \Psi(P_1,\ldots,P_n,P_{T,2}) - \mu\log(e)\cdot\left(\sum_{i=1}^{n}P_i + P_{T,2} - P_T\right)$$

where $-\mu\log(e)$ is the Lagrange multiplier (a constant $-\log(e)$ is included here for a simplicity in the following derivations). Let us define two kinds of functions that will be referred many times in following discussion

$$f_i(P_i) = \frac{\lambda_i}{1+P_i\lambda_i}, \quad 1\le i\le n$$
$$g(P_{T,2}) = \sum_{i=n+1}^{m}\frac{\lambda_i}{t-n+P_{T,2}\lambda_i}.$$

Differentiate J with respect to P_i $(1\le i\le n)$ and $P_{T,2}$, and set the derivatives to zeros; that is, $\partial J/\partial P_i = f_i(P_i)-\mu = 0, \quad 1\le i\le n$, and $\partial J/\partial P_{T,2} = g(P_{H,2})-\mu = 0$. Then, we obtain the following equations

$$f_i(P_i) = \mu, \quad 1\le i\le n \quad (21)$$
$$g(P_{T,2}) = \mu \quad (22)$$

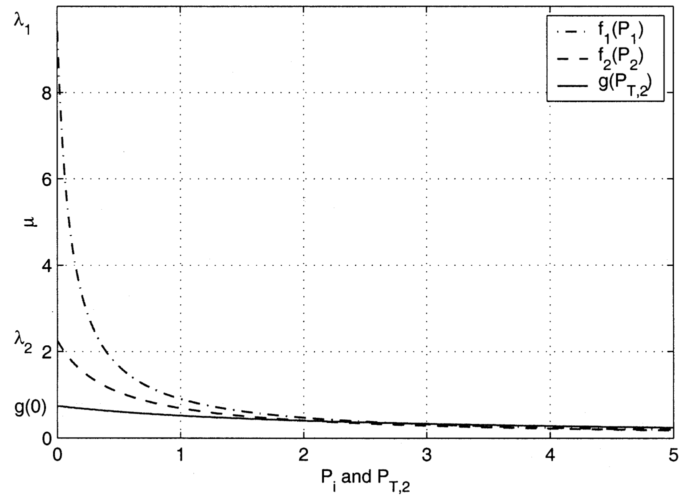

Fig. 1. An example of functions $f_i(P_i)$ and $g(P_{T,2})$ (when $t=4$, $r=4$, $n=2$, $\lambda_1=9.6303$, $\lambda_2=2.2467$, $\lambda_3=1.0682$, $\lambda_4=0.4174$).

The power constraint $(\sum_{i=1}^{n}P_i)+P_{T,2}=P_T$ can be rewritten using inverse functions[3] of $f_i(P_i), 1\le i\le n$, and $g(P_{T,2})$.

$$\sum_{i=1}^{n}\left[f_i^{-1}(\mu)\right]^+ + \left[g^{-1}(\mu)\right]^+ = P_T. \quad (23)$$

Here note that the channel capacity is achieved when the total transmit power equals to P_T. Since their shapes determine the optimum power allocation, to facilitate understanding the functions $\{f_i(P_i)\}$ and $g(P_{T,2})$ are shown in Fig. 1 as an example. Note also that the inverse function for $f_i(\cdot)$ is easily written by $f_i^{-1}(\mu) = 1/\mu - 1/\lambda_i$, while it is not easy to find a simple expression for $g^{-1}(\cdot)$. The following Theorem summarizes the steps to obtain the conditional channel capacity $C_{V_1H}(P_T;H)$.

Theorem 1: For a given channel, the channel capacity of MIMO channel with partial CSIT can be obtained by solving for μ satisfying

$$f_i(P_i)=\mu, 1\le i\le n;\ g(P_{T,2})=\mu;\ \text{and}$$
$$\sum_{i=1}^{n}\left[f_i^{-1}(\mu)\right]^+ + \left[g^{-1}(\mu)\right]^+ = P_T$$

where functions $f_i(\cdot)$ and $g(\cdot)$ are defined in (21) and (22). Once the solution μ^* is obtained, the optimum power allocation is given by

$$P_i^* = \left[\frac{1}{\mu^*}-\frac{1}{\lambda_i}\right]^+, \quad 1\le i\le n;\ \text{and}\ P_{T,2}^* = \left[g^{-1}(\mu^*)\right]^+ \quad (24)$$

and, the conditional channel capacity is given by

$$C_{V_1H}(P_T;H) = \Psi(P_1^*,\ldots,P_n^*,P_{T,2}^*) \quad (25)$$

where $\Psi(\cdot)$ is defined in (20).

Now, we need to solve for μ that simultaneously satisfies (21), (22) and (23). Note that the function $f_i(P_i)$ is a monotonically decreasing function with $f_i(0)=\lambda_i$ and goes to zero

[3]For the existence of inverse functions, we limit the domains of functions $f_i(x)$ and $g(x)$ such that $f_i : (-1/\lambda_i,\infty)\to(0,\infty)$ and $g : (-(t-n)/\lambda_{n+1},\infty)\to(0,\infty)$.

as P_i increases; and, so is the function $g(P_{T,2})$ with $g(0) = \frac{1}{t-n}\sum_{i=n+1}^{m}\lambda_i$. A desirable fact is that

$$g(0) = \frac{1}{t-n}\sum_{i=n+1}^{m}\lambda_i \le \frac{1}{m-n}\sum_{i=n+1}^{m}\lambda_i = \text{average of } \{\lambda_{n+1}, \ldots, \lambda_m\}. \quad (26)$$

Hence, $g(0) \le \lambda_j$, for all $1 \le j \le n$. We now define some parameters to be used in the following discussion

$$\rho(\mu) = \sum_{i=1}^{n}\left[f_i^{-1}(\mu)\right]^+ + \left[g^{-1}(\mu)\right]^+; \text{ and } \rho_g = \rho(g(0)). \quad (27)$$

Then, we can solve for μ by considering two cases: i) when $P_T < \rho_g$, and ii) when $P_T \ge \rho_g$. When $P_T < \rho_g$, μ should be greater than $g(0)$; therefore, in (23), $\left[g^{-1}(\mu)\right]^+ = 0$. It means that $P_{T,2}$ should be zero, i.e., the H_2 should not be used. Then, the solution μ^* satisfying the three (21)–(23) and the channel capacity can be obtained by using normal water-filling just as in (3). The optimum power allocation is given by

$$P_i = \left[\frac{1}{\mu^*} - \frac{1}{\lambda_i}\right]^+, \quad 1 \le i \le n; \text{ and } P_{T,2} = 0. \quad (28)$$

The conditional channel capacity is given by

$$C_{V_1H}(P_T; H) = \sum_{i=1}^{n}\left[\log\left(\frac{\lambda_i}{\mu^*}\right)\right]^+. \quad (29)$$

In the second case when $P_T \ge \rho_g$, μ should be less than $g(0)$; therefore, $P_{T,2}$ is now positive. That is, $P_i = f_i^{-1}(\mu) > 0$ for $1 \le i \le n$, and also $P_{T,2} = \left[g^{-1}(\mu)\right]^+ = g^{-1}(\mu) \ge 0$. The H_2 channel is now being used. Therefore, from (21)–(23), we need to solve for μ satisfying

$$\sum_{i=1}^{m}\left(\frac{1}{\mu} - \frac{1}{\lambda_i}\right) + g^{-1}(\mu) = P_T$$

which is equivalent to

$$h(\mu) \triangleq g\left(P_T - \sum_{i=1}^{m}\left(\frac{1}{\mu} - \frac{1}{\lambda_i}\right)\right) - \mu = 0. \quad (30)$$

The solution μ^* satisfying (30) can be solved numerically by using a zero-finding algorithm for single-variable nonlinear functions. The following Lemma shows the range of μ^* which is helpful in setting up the zero-finding algorithm.

Lemma 4: $\mu^* \in (\mu_L, g(0)]$, and μ_L is given by

$$\mu_L = n\left[P_T + \sum_{i=1}^{n}\frac{1}{\lambda_i} + \frac{t-n}{\lambda_{n+1}}\right]^{-1}. \quad (31)$$

□

Proof: See Appendix D.

A. Equivalent Water-filling

From the above derivation of the optimum transmit power allocation, we can see that a MIMO channel with partial CSI at the transmitter has some characteristics of water-filling. In particular, the H_2 channel starts to be used when the transmit power P_T is greater than a certain threshold ρ_g and the power allocation on the Σ_1 channel is determined by the conventional water-filling method. In this subsection, we show that the power allocation on the H_2 channel also can be understood with an equivalent water-filling model, which is described in the following Theorem.

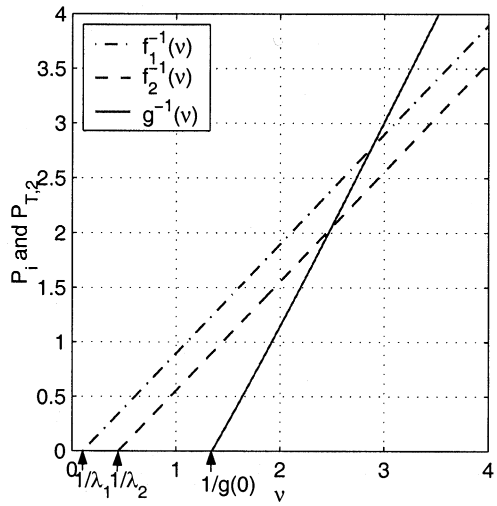

Fig. 2. Power allocation versus water-filling level ν ($t = 4$, $r = 4$, $n = 2$, $\lambda_1 = 9.6303$, $\lambda_2 = 2.2467$, $\lambda_3 = 1.0682$, $\lambda_4 = 0.4174$).

Theorem 2 [Equivalent Water-Filling]: The optimum power allocation over each channel can be viewed as the area determined by the the following function that defines the shape of the vessel for water-filling.

$$v(y) = \begin{cases} 0, & \text{if } 0 \le y < \frac{1}{\lambda_1}, \\ 1, & \text{if } \frac{1}{\lambda_1} \le y < \frac{1}{\lambda_2}, \\ \vdots & \vdots \\ n-1, & \text{if } \frac{1}{\lambda_{n-1}} \le y < \frac{1}{\lambda_n}, \\ n, & \text{if } \frac{1}{\lambda_n} \le y < \frac{1}{g(0)}, \\ f(y), & \text{if } y \ge g(0) \end{cases} \quad (32)$$

where $f(y)$ is given by

$$f(y) = \frac{1}{y^2}\left[\sum_{i=n+1}^{m}\frac{\lambda_i^2}{(t-n+g^{-1}\left(\frac{1}{y}\right)\lambda_i)^2}\right]^{-1} + n. \quad (33)$$

Then, the optimum power allocation can be written as follows.

$$P_i = \int_{i-1}^{i}\left[\nu - \frac{1}{\lambda_i}\right]^+ dx = \left[\nu - \frac{1}{\lambda_i}\right]^+, \quad \text{for } i = 1, \ldots, n \quad (34)$$

$$P_{T,2} = \begin{cases} \int_{1/g(0)}^{\nu}[v(y) - n]dy, & \text{if } \nu \ge \frac{1}{g(0)}, \\ 0, & \text{otherwise} \end{cases} \quad (35)$$

where $\nu = 1/\mu$ is the level of water-filling.

Proof: See Appendix E. □

We can obtain the relationship between the optimal power allocation to each channel and the water-filing level $\nu = 1/\mu$ from the definitions of functions $\mu = f_i(P_i)$ and $\mu = g(P_{T,2})$ given in (21) and (22). Fig. 2 is a numerical example for a given channel realization of $t = r = 4$ and $n = 2$. In this figure, we can see the characteristics of water-filling. That is, the thresholds are $1/\lambda_i$ and $1/g(0)$ as we expected. Note that $f_i^{-1}(\nu), 1 \le i \le n$, are linear functions with unit slope, and $g^{-1}(\nu)$ has an approximate slope of $m-n$. But, function $g^{-1}(\nu)$ is not a linear function. Fig. 3 shows an example of the equivalent water-filling

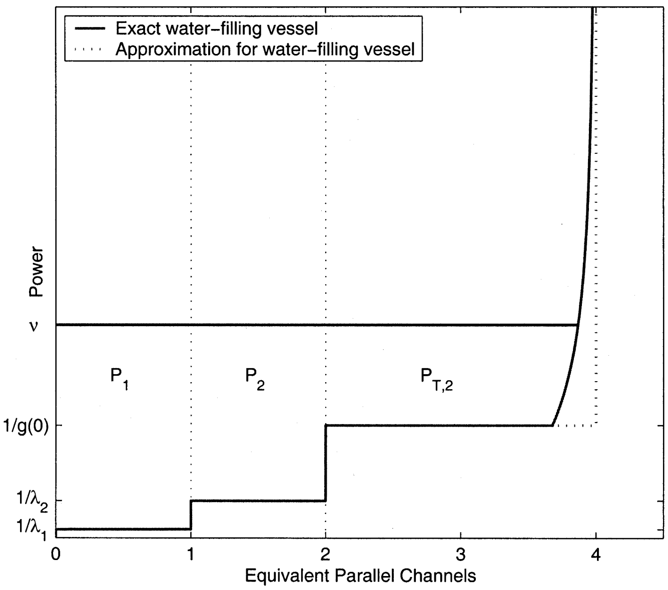

Fig. 3. Equivalent water-filling by Theorem 2 ($t = 4, r = 4, n = 2, \lambda_1 = 9.6303, \lambda_2 = 2.2467, \lambda_3 = 1.0682, \lambda_4 = 0.4174$).

shape that was calculated numerically from Theorem 2. The shape of the equivalent water-filling explains the water-filling characteristics discussed above. Since, for $1 \leq i \leq n$, the width of the ith channel is one, function $f_i^{-1}(\nu)$ has unit slope. And, the last H_2 channel is a nonlinear function which results in

$$P_{T,2}(\nu) < (m-n)\left(\nu - \frac{1}{g(0)}\right).$$

If we approximate the water-filling vessel to the rectangular one depicted in Fig. 3, then the calculation for the power allocation, therefore, also the channel capacity, will become much easier. The approximation also gives a meaningful insight in understanding the MIMO channel with partial CSIT. The following summarizes the results from the approximation.

Corollary 1 (Approximation for C_{V_1H}): For a given channel realization,

$$C_{V_1H}(P_T;H) \geq \sum_{i=1}^{n} [\log(\nu\lambda_i)]^+ + (m-n)\,[\log(\nu g(0))]^+ \tag{36}$$

where the lower bound can be achieved by a rectangular water-filling vessel exemplified in Fig. 3 with the transmit power allocation:

$$P_i = \left[\nu - \frac{1}{\lambda_i}\right]^+, \quad i = 1, \ldots, n$$
$$P_{T,2} = (m-n)\left[\nu - \frac{1}{g(0)}\right]^+ \tag{37}$$

and ν is selected to satisfy the power constraint $(\sum_{i=1}^{n} P_i) + P_{T,2} = P_T$.

Proof: See Appendix F. □

VI. Numerical Results

For comparative studies, we consider the extended MRT scheme. The transmitter is also assumed to know V_1, and only n spatial channels are used for transmission. That is, all the transmit power P_T is allocated only to the Σ_1 channel in a water-filling manner. It is equivalent to the case where the H_2 channel in our system model is not used. Since the extended MRT scheme abandons a chance for the additional potential gain from the H_2 channel, it is surely inferior to the proposed scheme. By looking at the problems in points of optimization, we can easily see the following relationships between capacities with different CSI assumptions and transmission strategies.

$$C_{\phi H}(P_T) \leq C_{V_1H}(P_T) \leq C_{HH}(P_T)$$
$$C_{V_1H}^{(\mathrm{EMRT})}(P_T) \leq C_{V_1H}(P_T) \leq C_{HH}(P_T).$$

Although the proposed system does not depend on a specific channel model, for numerical comparisons, we considered the MIMO channel that was assumed in [1]. The channel gain matrix $H \in \mathbb{C}^{r \times t}$ is a random matrix independent to the transmit symbols s and the additive noise η, with *i.i.d.* entries, each having independent real and imaginary parts with zero-mean and variance 1/2.

Figs. 4 and 5 are ergodic capacities versus total transmit power with different CSI assumptions and transmission strategies that have been discussed in Section V, for MIMO channel with parameters $(t = 4, r = 2)$ and $(t = 6, r = 3)$, respectively. For details about CSI assumptions and transmission strategies, refer to Tables I and II in Section III. In simulation, for each channel realization, $C_{V_1H}(P_T;H)$ was calculated using Theorem 1. The figures include the results from the approximation for the water-filling vessel described in Corollary 1. We can see that the approximation is a good fit over a significant range of transmit power. Also, the difference between C_{HH} and $C_{\phi H}$ at high transmit power of Lemma 2 can be observed. In addition, the following asymptotical behaviors of C_{V_1H} can be seen.

$$C_{V_1H}(P_T) \rightarrow C_{HH}(P_T), \quad \text{as } P_T \rightarrow 0.$$

This is because when the transmit power is low, by water-filling, in most case the power is allocated only to first a few channels and nothing to the Σ_2 channel; therefore, the two capacities are similar. This shows that spatial information is more important in low range of transmit power.

By comparing C_{V_1H} and $C_{V_1H}^{(\mathrm{EMRT})}$, it is observed that at low transmit power the two are very close, and as the transmit power increases C_{V_1H} is getting higher than $C_{V_1H}^{(\mathrm{EMRT})}$. From this observation, we can say that the H_2 channel should be utilized to achieve higher capacity at medium and high range of transmit power. At high transmit power region, C_{V_1H} is higher than but close to $C_{\phi H}$. This can be understood by noticing that, at high transmit power, for the same reason as the no CSIT case, the transmit power is wasted with the equal-power transmission over $t-n$ transmit symbols for the H_2 channel.

Figs. 6 and 7 show ergodic capacities for the transmission schemes that have been dealt in Section IV, with parameters $(t = 4, r = 2)$ and $(t = 6, r = 3)$, respectively. The conditional capacity $C_{V_1H}^{(\mathrm{opt})}(P_T;H)$ was calculated, depending on the system parameters, using the MAXDET algorithm, or (14),

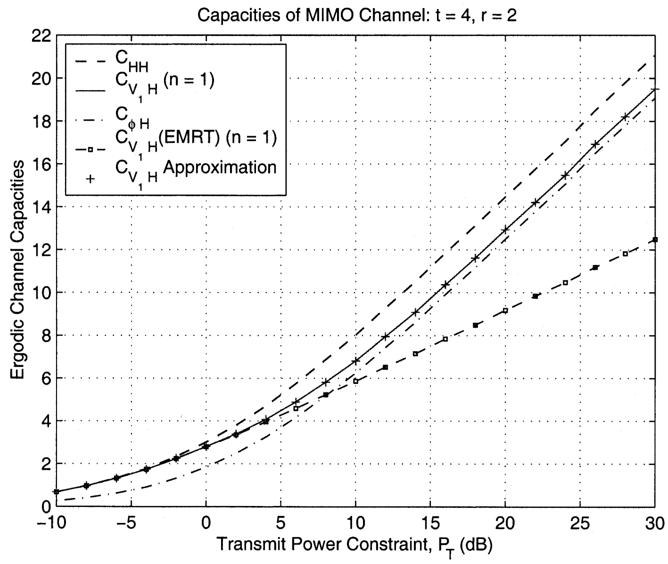

Fig. 4. Ergodic capacities of MIMO channel with different CSI assumptions and transmission strategies ($t = 4$ and $r = 2$).

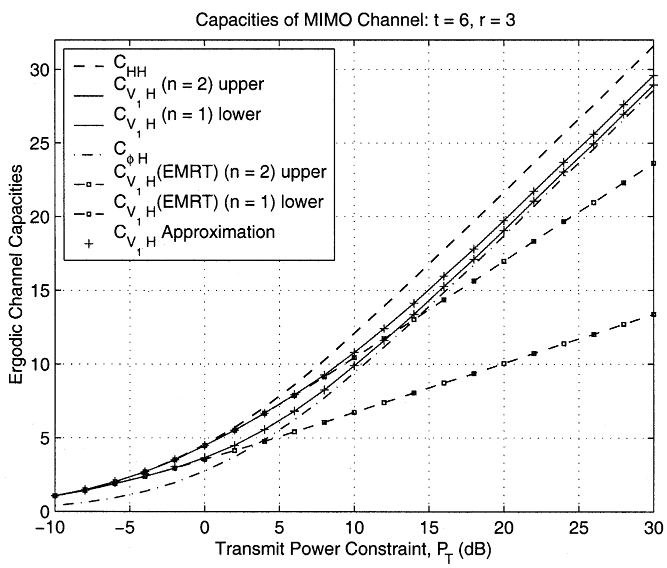

Fig. 5. Ergodic capacities of MIMO channel with different CSI assumptions and transmission strategies ($t = 6$ and $r = 3$).

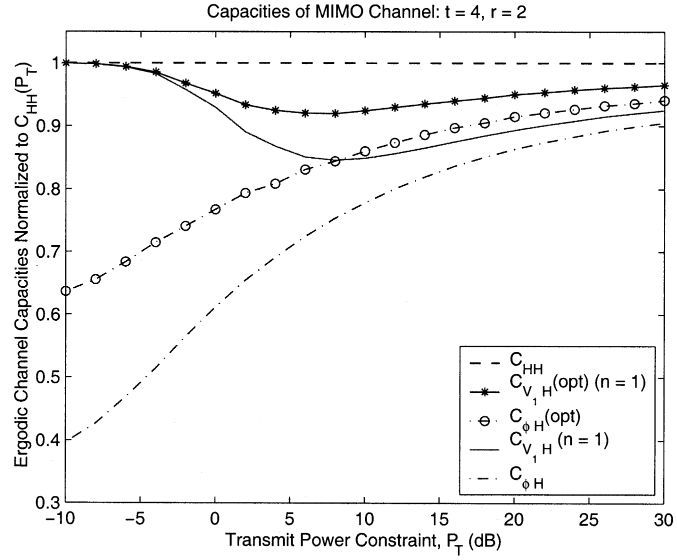

Fig. 6. Ergodic capacities of MIMO channel with different CSI assumptions and transmission strategies, normalized to the capacity for the complete CSIT ($t = 4$ and $r = 2$).

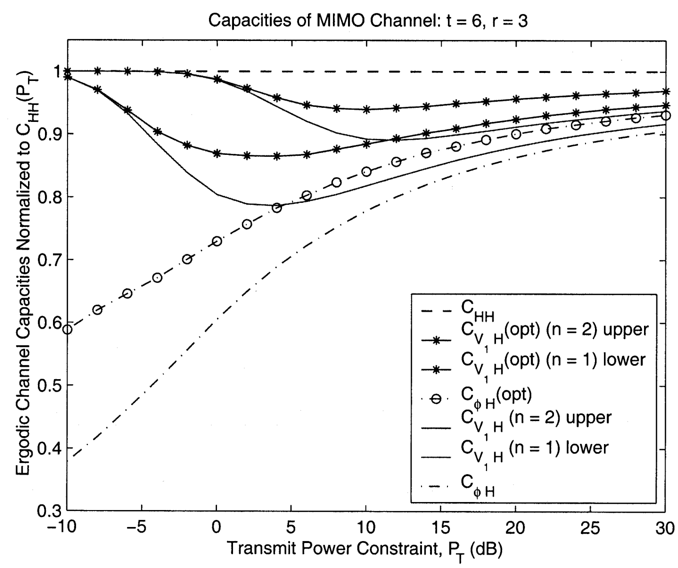

Fig. 7. Ergodic capacities of MIMO channel with different CSI assumptions and transmission strategies, normalized to the capacity for the complete CSIT ($t = 6$ and $r = 3$).

(15). In order to effectively compare the performance of different transmission strategies, each capacity has been normalized to $C_{HH}(P_T)$, the capacity of the complete CSIT. As expected, it is observed that for all range of transmit power

$$C_{V_1 H}(P_T) \le C_{V_1 H}^{(\mathrm{opt})}(P_T) \le C_{HH}(P_T)$$
$$C_{\phi H}(P_T) \le C_{\phi H}^{(\mathrm{opt})}(P_T) \le C_{V_1 H}^{(\mathrm{opt})}(P_T).$$

In comparing $C_{\phi H}^{(\mathrm{opt})}$ and $C_{V_1 H}(n = 1)$, it is noticeable that at low transmit power $C_{\phi H}^{(\mathrm{opt})} < C_{V_1 H}(n = 1)$, but at intermediate and high transmit power $C_{\phi H}^{(\mathrm{opt})}$ is superior. This observation implies that when the transmit power is low the spatial information of the channel is important, and as the transmit power increases the power allocation is becoming meaningful from a capacity point of view. Note that the channel feedback information required for the two strategies are different: for first strategy, t real values in $[0, 1]$, i.e., $(\gamma_1, \ldots, \gamma_t)$; and for the second one is one t-dimensional complex vector v_1 and one real value γ_1 in $[0, 1]$ (γ_2 is determined from γ_1 as $1 - \gamma_1$).

Generally speaking, in spatially correlated MIMO channels, the gains of spatial channels are more separated than in *i.i.d.* channels. Therefore, the proposed scheme is expected to be more beneficial in spatially correlated channels.

VII. Conclusion

We considered multiple antenna systems consisting of t transmit and r receive antennas, and partial channel state information available at the transmitter. When the multiple antenna channel is represented by a channel matrix H in $\mathbb{C}^{r \times t}$, the first n eigenvectors of $H^\dagger H$ are assumed to be available at the transmitter as partial spatial information of the channel. We proposed a novel transmission strategy in which a beamforming matrix is determined from the n eigenvectors in a predefined way. From the beamforming method, a new multiple antenna system concept was developed which enables better use of the MIMO channels. In particular, we have considered two methods, each having different degrees of additional channel information, i.e., power allocation results. The ergodic channel capacities and accompanying power allocation solutions for both cases have been derived. The simulation results have shown that, in moderate transmit power region, the proposed multiple antenna systems with partial channel information give channel capacities close to those with full channel information at the transmitter. In practical applications that need a feedback for the channel information, the amount of feedback information can be significantly reduced with a minor sacrifice of channel capacity by using the proposed schemes.

APPENDIX

A. Proof of Lemma 1

At high transmit power P_T,

$$C_{HH} - C_{\phi H} \simeq m \log \frac{t}{m} \begin{cases} = 0, & \text{when } t = m \\ > 0, & \text{when } t > m \end{cases}$$

where $m = \min(t, r)$ if H is full rank.

Proof: Assume P_T is high enough to satisfy $P_T \gg t\lambda_i^{-1}$ for all $1 \leq i \leq m$ and the level of water-filling ν can be approximated as $\nu \approx P_T/m$. Then

$$\begin{aligned} &C_{HH}(P_T) - C_{\phi H}(P_T) \\ &= E_H\left\{\sum_{i=1}^{m}\left\{[\log(\nu\lambda_i)]^+ - \log\left(1 + \frac{P_T}{t}\lambda_i\right)\right\}\right\} \\ &= E_H\left\{\sum_{i=1}^{m}\log\frac{\nu\lambda_i}{1 + (P_T/t)\,\lambda_i}\right\} \\ &\simeq E_H\left\{\sum_{i=1}^{m}\log\frac{\nu\lambda_i}{(P_T/t)\,\lambda_i}\right\} = E_H\left\{\sum_{i=1}^{m}\log\frac{\nu}{P_T/t}\right\} \\ &\simeq E_H\left\{\sum_{i=1}^{m}\log\frac{P_T/m}{P_T/t}\right\} = m\log\frac{t}{m} \end{aligned}$$

□

B. Proof of Lemma 2

The singular values of the channel matrix $H_2 = \Sigma_2 V_2^\dagger \tilde{V}_2$ is preserved as $\mathrm{diag}(\Sigma_2)$.

Proof: It suffices to show that $\tilde{V}_2 = V_2 Q$ for some unitary matrix $Q \in \mathbb{C}^{(t-n)\times(t-n)}$, because, if then, $H_2 = \Sigma_2 Q = I_{r-n}\Sigma_2 Q$ which is directly the SVD for H_2 with singular values given in Σ_2. Since the column spaces of $\tilde{V}_2$ and V_2 are same, i.e., $\mathcal{R}(\tilde{v}_2) = \mathcal{R}(V_2)$, we can write

$$\tilde{V}_2 = V_2 Q, \quad \text{for some full rank } Q \in \mathbb{C}^{(t-n)\times(t-n)}. \tag{38}$$

We will show that Q is unitary. Denote the spectral norm [16] of a matrix $A \in \mathbb{C}^{p\times q}$ as

$$\|A\| = \max_{|x|=1} |Ax| = \sigma_1(A)$$

where $\sigma_1(A)$ is the largest singular value of A. By using the submultiplicative property of matrix norm $\|\tilde{V}_2\| \leq \|V_2\|\|Q\|$ and $\|\tilde{V}_2\| = \|V_2\| = 1$, we arrive $\|Q\| \geq 1$. And, from (38), $V_2^\dagger \tilde{V}_2 = Q$. In a similar way, it can be shown that $\|Q\| \leq 1$. Therefore, $\|Q\| = 1$.

Since Q is invertible, $\tilde{V}_2 Q^{-1} = V_2$. It can be proved that $\|Q^{-1}\| = 1$ exactly in the same manner as the above. Since $\|Q\| = \|Q^{-1}\| = 1$, $\sigma_1(Q) = \cdots = \sigma_{t-n}(Q) = 1$, i.e., the SVD of Q is given by $Q = U_Q \Sigma_Q V_Q^\dagger = U_Q V_Q^\dagger$. Hence, $Q^\dagger Q = QQ^\dagger = I_{t-n}$, that is, Q is a unitary matrix. □

C. Proof of Lemma 3

For a given channel realization H, the mutual information between the input and the output of the MIMO channel can be expressed as

$$\begin{aligned} I(x; y) &= I(s; \tilde{y}) \\ &= I(s_1; \tilde{y}_1) + I(s_2; \tilde{y}_2). \end{aligned}$$

Proof: From the system model

$$\tilde{y} = U^\dagger H W s + U^\dagger \eta. \tag{39}$$

The first equation of (11) can be proved by

$$\begin{aligned} I(s; \tilde{y}) &= h(\tilde{y}) - h(\tilde{y}|s) = h(U^\dagger y) - h(U^\dagger \eta) \\ &= h(y) + \log|\det(U^\dagger)| - \big(h(\eta) + \log|\det(U^\dagger)|\big) \\ &= h(y) - h(\eta) = h(y) - h(y|x) = I(x; y) \end{aligned}$$

The second equation above cames from that $h(\tilde{y}|s) = h(U^\dagger \eta)$ since for given s the first term of (39) is just a constant vector, and by a property that the entropy is not changed with a shift.

The second equation of (12) can be verified by noting that the two input-output relations, $s_1 \to \tilde{y}_1$ and $s_2 \to \tilde{y}_2$, are independent. Therefore

$$\begin{aligned} I(s; \tilde{y}) &= I(s_1, s_2; \tilde{y}_1, \tilde{y}_2) = E\left\{\log\frac{p(s_1, s_2, \tilde{y}_1, \tilde{y}_2)}{p(s_1, s_2)p(\tilde{y}_1, \tilde{y}_2)}\right\} \\ &= E\left\{\log\frac{p(s_1, \tilde{y}_1)}{p(s_1)p(\tilde{y}_1)}\frac{p(s_2, \tilde{y}_2)}{p(s_2)p(\tilde{y}_2)}\right\} \\ &= E\left\{\log\frac{p(s_1, \tilde{y}_1)}{p(s_1)p(\tilde{y}_1)}\right\} + E\left\{\log\frac{p(s_2, \tilde{y}_2)}{p(s_2)p(\tilde{y}_2)}\right\} \\ &= I(s_1; \tilde{y}_1) + I(s_2; \tilde{y}_2) \end{aligned}$$

□

D. Proof of Lemma 4

$\mu^* \in (\mu_L, g(0)]$, and μ_L is given by

$$\mu_L = n\left[P_T + \sum_{i=1}^{n}\frac{1}{\lambda_i} + \frac{t-n}{\lambda_{n+1}}\right]^{-1}$$

Proof: Since, by assumption, $P_T - \rho_g \geq 0$, and $g(x)$ is a decreasing function with $g(0) > 0$ for $x \geq 0$

$$h(g(0)) = g(P_T - \rho_g) - g(0) \leq 0.$$

From the definition of $g(x)$ in (22), we can see that $g(x)$ has the largest singular point at $x_0 = -(t-n)/\lambda_{n+1}$ with $g(x) \to \infty$ as $x \to x_0 + \epsilon$ for a small positive value ϵ. Therefore, when the argument of $g(\cdot)$ in (30) approaches to x_0, the output of $g(\cdot)$ goes to ∞. Setting the argument to x_0 and solving for μ gives μ_L in (31). Hence, the Lemma is proved by noting that $g(x)$ is continuous and monotonically deceasing for $x > x_0$. □

E. Proof of Theorem 2

Proof: Other things are straightforward except $f(y)$ of (33). We want to find a function $f(y)$ satisfying

$$\int_{1/g(0)}^{1/\mu} [f(y) - n] dy = P_{T,2}(\mu), \quad \text{for } \mu > g(0).$$

Differentiating both sides with respect to μ, we have

$$f\left(\frac{1}{\mu}\right) = -\mu^2 \frac{dP_{T,2}(\mu)}{d\mu} + n \tag{40}$$

We know that the relationship between μ and $P_{T,2}$ is given by $\mu = g(P_{T,2})$ which was presented in (22). We can obtain $dP_{T,2}(\mu)/d\mu$ by

$$\frac{d\mu}{dP_{T,2}} = \frac{dg(P_{T,2})}{dP_{T,2}} = -\sum_{i=n+1}^{m} \frac{\lambda_i^2}{(t-n+g^{-1}(\mu)\lambda_i)^2}.$$

where note that $g^{-1}(\mu)$ was written in denominator within the summation instead of $P_{T,2}$. By noticing that $dP_{T,2}(\mu)/d\mu$ in (40) is the inverse of $d\mu/dP_{T,2}$, and after changing variable, we reach the desired result of (33). □

F. Proof of Corollary 1

Proof: First, let us show the following inequality for the function $g(P_{T,2})$:

$$\begin{aligned} g(P_{T,2}) &= \sum_{i=n+1}^{m} \frac{\lambda_i}{(t-n) + P_{T,2}\lambda_i} \\ &\leq \frac{(m-n)g(0)}{(m-n) + P_{T,2}\, g(0)} \triangleq \tilde{g}(P_{T,2}) \end{aligned} \tag{41}$$

For a fixed $P_{T,2} \geq 0$, the function $\vartheta(\lambda) = \lambda/[(t-n) + P_{T,2}\lambda]$ is concave for $\lambda \geq 0$. For any concave function $\vartheta(x)$ defined over a set D, for all $x_1, \ldots, x_n \in D$, and for all nonnegative numbers $\alpha_1, \ldots, \alpha_n$ such that $\sum_{i=1}^n \alpha_i = 1$, it is easy to see that

$$\sum_{i=1}^{n} \alpha_i \vartheta(x_i) \leq \vartheta\left(\sum_{i=1}^{n} \alpha_i x_i\right).$$

By applying the above inequality with equal coefficients $\{\alpha_i\}$, we have

$$\begin{aligned} \frac{1}{m-n} \sum_{i=n+1}^{m} \frac{\lambda_i}{(t-n) + P_{T,2}\lambda_i} &\leq \frac{\bar{\lambda}}{(t-n) + P_{T,2}\bar{\lambda}} \\ &= \frac{g(0)}{(m-n) + P_{T,2}\, g(0)} \end{aligned}$$

where $\bar{\lambda} = \frac{1}{m-n}\sum_{i=n+1}^{m} \lambda_i$, and the last equality results from $\bar{\lambda} = \frac{t-n}{m-n} g(0)$. Therefore, the inequality (41) has been proved. It is interesting that the function $\tilde{g}(P_{T,2})$ is associated with the rectangular vessel depicted in Fig. 3. That is, if $\tilde{g}(P_{T,2}) = \mu$ is rewritten for the water-filling level $\nu = 1/\mu$, we have

$$\nu = \frac{P_{T,2}}{m-n} + \frac{1}{g(0)}$$

which explains the water-filling behavior of the rectangular shape. It is obvious that the mutual information obtained from the rectangular vessel is less than the maximum value, i.e., $C_{V_1 H}$. □

ACKNOWLEDGMENT

The authors would like to thank Prof. S. Boyd of Stanford University for his help in applying the MAXDET algorithm to problems with complex matrices, as well as the reviewers for their valuable comments and suggestions which improved the paper.

REFERENCES

[1] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," AT&T Bell Labs Tech. Memo., 1995.
[2] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Commun.*, vol. 6, no. 3, pp. 311–335, Mar. 1998.
[3] E. Biglieri, G. Caire, and G. Taricco, "Limiting performance of block-fading channels with multiple antennas," *IEEE Trans. Inform. Theory*, vol. 47, pp. 1273–1289, May 2001.
[4] A. Narula, M. D. Trott, and G. W. Wornell, "Performance limits of coded diversity methods for transmitter antenna arrays," *IEEE Trans. Inform. Theory*, vol. 45, pp. 2418–2433, Mar. 1999.
[5] G. Caire and S. Shamai, "On the capacity of some channels with channel state information," *IEEE Trans. Inform. Theory*, vol. 45, pp. 2007–2019, Sept 1999.
[6] E. Visotsky and U. Madhow, "Space-time transmit precoding with imperfect feedback," *IEEE Trans. Inform. Theory*, vol. 47, pp. 2632–2639, Sept. 2001.
[7] S. A. Jafar, S. Vishwanath, and A. Goldsmith, "Channel capacity and beamforming for multiple transmit and receive antennas with covariance feedback," *Proc. IEEE ICC 2001*, pp. 2266–2270, June 2001.
[8] S. Bhashyam, A. Sabharwal, and B. Aazhang, "Feedback gain in multiple antenna systems," *IEEE Trans. Commun.*, vol. 50, pp. 785–798, May 2002.
[9] G. Jongren, M. Skoglund, and B. Ottersten, "Combining beamforming and orthogonal space-time block coding," *IEEE Trans. Inform. Theory*, vol. 48, pp. 611–627, Mar. 2002.
[10] M. T. Ivrlac and J. A. Nossek, "Correlated fading in MIMO systems-blessing or curse?," in Proc. 39th Allerton Conf., 2001.
[11] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
[12] T. K. Y. Lo, "Maximum ratio transmission," *IEEE Trans. Commun.*, vol. 47, pp. 1458–1461, Oct. 1999.
[13] H. Sampath, P. Stoica, and A. Paulraj, "Generalized linear precoder and decoder design for MIMO channels using the weighted MMSE criterion," *IEEE Trans. Commun.*, vol. 49, pp. 2198–2206, Dec 2001.

[14] L. Vandenberghe, S. Boyd, and S.-P. Wu, "Determinant maximization with linear matrix inequality constraints," *SIAM J. Matrix Anal. Applic.*, vol. 19, no. 2, pp. 499–533, Apr. 1998.

[15] S.-P Wu, L. Vandenberghe, and S. Boyd, *MAXDET: Software for Determinant Maximization Problems: User's Guide, Alpha Version*, CA: Stanford Univ., 1996.

[16] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, MA: Cambridge Univ. Press, 1985.

June Chul Roh (S'03) received the B.S. and M.S. degrees in electronics engineering from the University of Seoul, Seoul, Korea, in 1993 and 1995, respectively.

Currently, he is working toward the Ph.D. degree with the University of California at San Diego, La Jolla. From 1995 to 2000, he was a Member of Technical Staff of the Korea Telecom R&D Group, where he worked on third-generation cellular systems and was involved in developing a wideband CDMA system. His research interests include communication theory, information theory, statistical signal processing, spread-spectrum communications, and multiple antenna systems.

Bhaskar D. Rao (S'80–M'83–SM'91–F'00) received the B.Tech. degree in electronics and electrical communication engineering from the Indian Institute of Technology, Kharagpur, in 1979 and the M.S. and Ph.D. degrees from the University of Southern California, Los Angeles, in 1981 and 1983, respectively.

Since 1983, he has been with the University of California at San Diego, La Jolla, where he is currently a Professor with the Electrical and Computer Engineering Department. His interests are in the areas of digital signal processing, estimation theory, and optimization theory, with applications to digital communications, speech signal processing, and human-computer interactions.

Dr. Rao is a member of the Statistical Signal and Array Processing Technical Committee of the IEEE Signal Processing Society. He is also a member of the Signal Processing Theory and Methods Technical Committee.