Capacity of fading MIMO channels with channel estimation error

Taesang Yoo and Andrea Goldsmith
Dept. of Electrical Engineering, Stanford University
Stanford, CA 94305
{yoots,andrea}@stanford.edu

Abstract

Multiple input multiple output (MIMO) systems provide increased spectral efficiency: in an i.i.d. Rayleigh flat-fading channel, MIMO capacity increases linearly with the smaller of the number of transmit and receive antennas. This dramatic increase of capacity, however, requires perfect channel knowledge at the transmitter and the receiver, which rarely happens in real systems. We thus study lower and upper bounds of mutual information under channel estimation error, and show that the two bounds are tight for Gaussian inputs. We also derive a tight capacity lower bound and the optimal transmitter strategy to achieve this bound. Numerical results show that some capacity gain is obtained by shaping transmit power over the subchannels. However, temporal power adaptation appears to give negligible gain in terms of ergodic capacity.

1 Introduction

Multiple input multiple output (MIMO) systems have been shown to provide dramatic capacity gain. Specifically, in an i.i.d. Rayleigh fading channel, the capacity of a system with t transmit and r receive antennas grows $\min\{r, t\}$ times faster than that of a single input single output (SISO) system. This large capacity gain, however, requires perfect knowledge of the instantaneous channel fading at both the transmitter and receiver, which rarely happens in real systems. Thus, the impact of imperfect channel knowledge and channel estimation error on capacity is an important area to investigate.

Various assumptions about channel state information (CSI) and channel fading lead to different capacity results. In [1], bounds on mutual information with imperfect CSIR and no CSIT have been derived. This result has been extended in [2] to flat fading channels and a capacity lower bound has been derived assuming the feedback link is perfect. Caire and Shamai [3] have studied some channels with imperfect CSIT and perfect CSIR. These studies all use genie-provided (partial) CSI's and independent fading processes. In the absence of CSI, Lapidoth [4] has shown that at high SNR capacity grows double-logarithmically in the SNR.

Fading correlation over time can increase capacity; in [5], capacity of noncoherent multiple antenna fading channels has been studied for a Rayleigh block-fading environment. This capacity increases logarithmically with SNR, and approaches the perfect channel knowledge capacity as the coherence interval increases. Achievable rates of practical channel estimation schemes are upper-bounded by the noncoherent capacity since they use suboptimal ways to probe the channel. In [6], bounds of achievable rates have been derived for typical MIMO flat-fading channels using pilot symbols and an interleaver. These studies assume no CSI but exploit fading correlation in order to directly or indirectly obtain it.

This paper, which extends the results of [1, 2] from SISO to MIMO channels, uses the same assumptions as [1]-[4] about the channel fading and CSI. Specifically, we study lower and upper bounds of mutual information for an i.i.d. Rayleigh flat fading channel with a genie-provided MMSE channel estimate at the receiver and perfect feedback. Then, we derive optimal transmitter strategies that maximize the lower bound of mutual information to obtain the corresponding capacity lower bound.

The rest of this paper is organized as follows. In Section 2, our system model is introduced. In Section 3, lower and upper bounds of mutual information under channel estimation error are derived. The capacity bounds are also derived subject to an average power constraint. In Section 4, the optimal power allocation for different values of mutual information is determined. Finally, numerical results are presented in Section 5.

2 System Model

Consider a MIMO system with t transmit and r receive antennas. The discrete-time channel is modeled as $\mathbf{Y}_n = \mathbf{H}_n\mathbf{X}_n + \mathbf{Z}_n$, where $\mathbf{Y}_n$ is an $r \times 1$ channel output, $\mathbf{H}_n$ is an $r \times t$ channel transfer matrix, $\mathbf{X}_n$ is a $t \times 1$ channel input, and $\mathbf{Z}_n$ is an $r \times 1$ vector of additive white Gaussian noise (AWGN). We assume both $\mathbf{H}_n$ and $\mathbf{Z}_n$ are ergodic and stationary, and their entries are i.i.d. and zero-mean circularly symmetric complex Gaussian (ZMCSCG). We normalize the channel and noise variance such that the entries of $\mathbf{H}_n$ and $\mathbf{Z}_n$ have unit variance. By properly scaling the transmit power, this normalization does not change the mutual information of the channel. We also assume that both the channel fading process and the noise process have no correlation between time instances. This would be the case if we use a perfect interleaver and do not exploit information contained in the fading correlation. In such a system, and for more general classes, if the fading process is *perfectly* known to the receiver, the mutual information between the channel input and output is given by [7]

$$I(\mathbf{X};\mathbf{Y}) = E\left\{\log_2 \left|\mathbf{I} + \mathbf{H}_n^*\mathbf{H}_n\mathbf{Q}\right|\right\} \tag{1}$$

where $\mathbf{Q}$ is the input covariance matrix $\mathbf{Q} = E(\mathbf{X}_n\mathbf{X}_n^*)$, and $E\{\bullet\}$ is an expectation operator. In the remainder of this paper, we will not write the time index n explicitly for notational convenience. This is possible because the random processes we are dealing with are i.i.d. over time. Besides the aforementioned assumptions about the channel, we further assume the following throughout this paper: (1) At each time instance, the receiver has genie-provided channel state information $\tilde{\mathbf{H}}$, and performs MMSE estimation of the channel fading process, $\hat{\mathbf{H}} = E(\mathbf{H}|\tilde{\mathbf{H}})$. Let $\mathbf{H} = \hat{\mathbf{H}} + \mathbf{E}$, then, $\hat{\mathbf{H}}$ and $\mathbf{E}$ are uncorrelated, and the entries of $\mathbf{E}$ are ZMCSCG with variance $\sigma_{\mathbf{E}}^2 = \text{MMSE} = E(\mathbf{H}_{ij}^2) - E(\hat{\mathbf{H}}_{ij}^2)$. $\sigma_{\mathbf{E}}^2$ dictates the quality of the channel estimation, and is assumed to be known to both the transmitter and the receiver. (2) There exists a perfect and instantaneous feedback from the receiver to the transmitter so that whatever CSI the receiver has is also available at the transmitter. (3) The transmitter is constrained in its total power P, and it can adapt its power to the channel fading to maximize capacity. That is, $E(P) = E\left(\text{Tr}(\mathbf{Q})\right) \leq \bar{P}$. $\text{Tr}(\bullet)$ stands for trace.

3 Bounds of Mutual Information with Imperfect CSI

3.1 Lower bound of mutual information

The ergodic capacity of the fading channel model in Section 2 with an estimated channel $\hat{\mathbf{H}}$ at the transmitter and the receiver is given by [3]

$$C = \max_{p(\mathbf{X}|\hat{\mathbf{H}})} E\left[I(\mathbf{X};\mathbf{Y}|\hat{\mathbf{H}})\right] \tag{2}$$

where $p(\mathbf{X}|\hat{\mathbf{H}})$ is the probability distribution of $\mathbf{X}$ given $\hat{\mathbf{H}}$. In this section, we focus on evaluating the mutual information $I(\mathbf{X};\mathbf{Y}|\hat{\mathbf{H}})$: its maximization with respect to power allocation will be the main topic of Section 4.

We begin by expanding the mutual information into differential entropies.

$$I(\mathbf{X};\mathbf{Y}|\hat{\mathbf{H}}) = h(\mathbf{X}|\hat{\mathbf{H}}) - h(\mathbf{X}|\mathbf{Y},\hat{\mathbf{H}}) \tag{3}$$

Denoting as $\mathbf{Q}$ the covariance matrix of $\mathbf{X}$ given $\hat{\mathbf{H}}$, and choosing $\mathbf{X}|\hat{\mathbf{H}}$ to be Gaussian (which is not necessarily the capacity achieving distribution when the CSIR is not perfect [1, 4]), the first term on the RHS becomes $E\left[\log_2(\pi e\mathbf{Q})\right]$. Evaluation of the second term, on the other hand, is not trivial because the conditional distribution $p(\mathbf{X}|\mathbf{Y},\hat{\mathbf{H}})$ is in general not Gaussian. However, we can derive an upper bound to it as [1]

$$h(\mathbf{X}|\mathbf{Y},\hat{\mathbf{H}}) \leq E\left[\log_2\left(\left|\pi e\Sigma_{\mathbf{X}-\mathbf{AY}|\hat{\mathbf{H}}}\right|\right)\right] \tag{4}$$

for any $t \times r$ matrix $\mathbf{A}$, where $\Sigma_{\mathbf{X}-\mathbf{AY}|\hat{\mathbf{H}}}$ represents the covariance matrix of $\mathbf{X}-\mathbf{AY}$ given $\hat{\mathbf{H}}$. Since (4) holds for any $\mathbf{A}$, we pick $\mathbf{A}$ so that the RHS is minimized to yield the tightest bound. This corresponds to the case when $\mathbf{AY}$ is the linear MMSE estimate of $\mathbf{X}$, in which case the lower bound is given by

$$I_{lower}(\mathbf{X};\mathbf{Y}|\hat{\mathbf{H}}) = E\left[\log_2\left|\mathbf{I} + \hat{\mathbf{H}}^*(\mathbf{I}+\mathbf{\Sigma_{EX}})^{-\mathbf{1}}\hat{\mathbf{H}}\mathbf{Q}\right|\right] \tag{5}$$

Note that (5) is equivalent to the mutual information of the MIMO channel with an effective gain $(\mathbf{I}+\mathbf{\Sigma_{EX}})^{-\mathbf{1/2}}\hat{\mathbf{H}}$ and perfect CSIR. When the entries of $\mathbf{H}$ are independent and identically distributed, we have that the channel estimation errors $\{\mathbf{E}_{ij}\}$ are also independent and identically distributed, i.e., $E(\mathbf{E}_{ij}\mathbf{E}_{mn}) = \sigma^2_{\mathbf{E}}\delta_{i-m,j-n}$, and $\Sigma_{\mathbf{EX}} = \sigma^2_{\mathbf{E}}P\mathbf{I}$. Then, (5) becomes

$$I_{lower}(\mathbf{X};\mathbf{Y}|\hat{\mathbf{H}}) = E\left[\log_2\left|\mathbf{I} + \frac{1}{1+\sigma^2_{\mathbf{E}}P}\hat{\mathbf{H}}^*\hat{\mathbf{H}}\mathbf{Q}\right|\right] \tag{6}$$

As is observed in [6], comparing (6) to (1), we can see that the channel estimation error affects the mutual information by two separate mechanisms. First, the estimation error increases the effective noise power from unity to $1+\sigma^2_{\mathbf{E}}P$. Secondly, it reduces the average channel power gain from $E(\mathbf{H}^*\mathbf{H}) = r\mathbf{I}$ to $E(\hat{\mathbf{H}}^*\hat{\mathbf{H}}) = r(1-\sigma^2_{\mathbf{E}})\mathbf{I}$. These two effects result in an SNR loss factor of $\delta = (1-\sigma^2_{\mathbf{E}})/(1+\sigma^2_{\mathbf{E}}P)$.

3.2 Upper bound of mutual information

Expanding the mutual information (3) in an alternate way, we have

$$I(\mathbf{X};\mathbf{Y}|\hat{\mathbf{H}}) = h(\mathbf{Y}|\hat{\mathbf{H}}) - h(\mathbf{Y}|\mathbf{X},\hat{\mathbf{H}}) \tag{7}$$

Using the fact that the Gaussian distribution maximizes the entropy over all distributions with the same covariance, we obtain an upper bound of the first term on the RHS as

$$h(\mathbf{Y}|\hat{\mathbf{H}}) \leq E\left[log_2\left|\pi e\left(\hat{\mathbf{H}}\mathbf{Q}\hat{\mathbf{H}}^* + \left(1+\sigma_{\mathbf{E}}^2 P\right)\mathbf{I}\right)\right|\right] \tag{8}$$

Since $\mathbf{E}$ is complex Gaussian by assumption, $(\mathbf{Y}|\mathbf{X},\hat{\mathbf{H}})$ is also complex Gaussian with $N_C(\hat{\mathbf{H}}\mathbf{X}, \mathbf{\Sigma}_{\mathbf{EX}|\mathbf{X}} + \mathbf{I})$. Thus, the second term on the RHS in (7) becomes

$$h(\mathbf{Y}|\mathbf{X},\hat{\mathbf{H}}) = E_{\mathbf{X}}\left[\log_2\left|\pi e\left(1+\sigma_{\mathbf{E}}^2\|\mathbf{X}\|^2\right)\mathbf{I}\right|\right] \tag{9}$$

Combining (7)-(9), we have

$$I_{upper}(\mathbf{X};\mathbf{Y}|\hat{\mathbf{H}}) = I_{lower}(\mathbf{X};\mathbf{Y}|\hat{\mathbf{H}}) + rE_{\mathbf{X}}\left[\log_2\frac{\sigma_{\mathbf{E}}^2 P+1}{\sigma_{\mathbf{E}}^2\|\mathbf{X}\|^2+1}\right] \tag{10}$$

The upper bound is related to the lower bound by Jenson's inequality, i.e., noting that $E(\|\mathbf{X}\|^2) = P$, the second term on the RHS in (10) is nonnegative [6]. The following lemma shows that the gap between the two bounds is usually small for a *Gaussian* input unless $r \gg t$. In other words, the two bounds are approximately equal to the exact *Gaussian mutual information.*

Lemma 1 *In the limit of high SNR and a large number of antennas, the second term in (10) approaches* $(r/t)\log_2\sqrt{e} \approx 0.72(r/t)$ *for Gaussian inputs. (Refer to [9] for proof.)*

3.3 Capacity bounds for Gaussian input

In this subsection, we study ergodic capacity bounds by finding optimal input covariance matrices $\mathbf{Q}$ that maximize the bounds of the previous subsections. We first consider the maximization of the lower bound in (6). Let the singular value decomposition of the estimated channel matrix be $\hat{\mathbf{H}} = \mathbf{UDV}^*$, where $\mathbf{U}$ and $\mathbf{V}$ are unitary and $\mathbf{D}$ is diagonal, and let us define two quantities, $\tilde{\mathbf{Q}} = \mathbf{V}^*\mathbf{QV}$ and $\mathbf{\Lambda} = \mathbf{D}^*\mathbf{D}$. Then

$$I_{lower}(\mathbf{X};\mathbf{Y}|\hat{\mathbf{H}}) = E\left(\log_2\left|\mathbf{I} + \frac{1}{1+\sigma_{\mathbf{E}}^2 P}\mathbf{\Lambda}\tilde{\mathbf{Q}}\right|\right) \tag{11}$$

Under an average power constraint $E(P) = E\left(\text{Tr}(\mathbf{Q})\right) \leq \bar{P}$, observing that $\text{Tr}(\mathbf{Q}) = \text{Tr}(\tilde{\mathbf{Q}})$, (11) is maximized with $\tilde{\mathbf{Q}}$ a diagonal matrix, $\tilde{\mathbf{Q}} = \text{diag}(P_1,\cdots,P_t)$, with an optimal power distribution $\{P_i\}$ such that $\sum_{i=1}^t P_i = P$. Thus, the lower bound of ergodic capacity is given by

$$\begin{array}{c} C_{lower} = \max_{\{P_i\}} E\left(\sum_{i=1}^t \log_2\left(1+\frac{P_i\lambda_i}{1+\sigma_{\mathbf{E}}^2 P}\right)\right) \\ \text{subject to} \quad E(P) = E\left(\sum_{i=1}^t P_i\right) \leq \bar{P} \end{array} \tag{12}$$

where λ_i is the $(i,i)^{th}$ element of $\mathbf{\Lambda}$ and thus the i^{th} eigenvalue of $\hat{\mathbf{H}}^*\hat{\mathbf{H}}$. The above expectations are performed over the joint distribution of $(\lambda_1,\cdots,\lambda_t)$. The input to the channel that achieves the capacity has covariance matrix of the form $\mathbf{Q} = \mathbf{V}\text{diag}(P_1,\cdots,P_t)\mathbf{V}^*$ whose optimal subchannel powers $\{P_i\}$ are determined as functions of $(\lambda_1,\cdots,\lambda_t)$.

Let us now consider the upper bound. In maximizing the upper bound (10), unlike the lower bound, the distribution of $\mathbf{X}$ should be taken into account. The resultant capacity bound, however, exceeds the trivial perfect-CSI upper bound if we allow $\mathbf{X}$ to have an arbitrary distribution. To obtain a practical bound, therefore, we restrict $\mathbf{X}$ to be jointly complex Gaussian, acquiring the upper bound of Gaussian ergodic capacity. From (10) and (12), we have

$$\begin{array}{c} C_{upper} = \max_{\{P_i\}} E\left(\sum_{i=1}^{t} \log_2\left(1+\frac{P_i\lambda_i}{1+\sigma_{\mathbf{E}}^2 P}\right) + r\log_2\left(\frac{\sigma_{\mathbf{E}}^2 P+1}{\sigma_{\mathbf{E}}^2\|\mathbf{X}\|^2+1}\right)\right) \\ \text{subject to} \quad E(P) = E\left(\sum_{i=1}^{t} P_i\right) \le \bar{P} \end{array} \tag{13}$$

The expectation is performed both over the joint distribution of $(\lambda_1, \cdots, \lambda_t)$ and over the joint Gaussian distribution of $\mathbf{X} \sim N_C\left(\mathbf{0}, \mathrm{diag}(P_1, \cdots, P_t)\right)$.

The capacity lower bound (12) can be achieved by the following. First, singular value decomposition (SVD) is performed on the estimated channel, $\hat{\mathbf{H}} = \mathbf{UDV}^*$, which would diagonalize the MIMO channel if the channel estimation were correct. However, with $\hat{\mathbf{H}}$ different from $\mathbf{H}$, the channel is not fully decomposed into independent SISO links. To see this, let $\tilde{\mathbf{E}} = \mathbf{U}^*\mathbf{EV}$. Then we have $\mathbf{H} = \hat{\mathbf{H}}+\mathbf{E} = \mathbf{U}(\mathbf{D}+\tilde{\mathbf{E}})\mathbf{V}^*$. Thus, transmit precoding and receiver shaping by $\mathbf{V}$ and $\mathbf{U}^*$ will produce an equivalent channel matrix $\mathbf{D}+\tilde{\mathbf{E}}$. It can be easily verified that $\tilde{\mathbf{E}}$ is zero mean with uncorrelated entries with variance $\sigma_{\mathbf{E}}^2$. Hence, the decomposed channel, $\mathbf{D}+\tilde{\mathbf{E}}$, is not diagonal unless $\hat{\mathbf{H}} = \mathbf{H}$. After all, as a result of using imperfect channel information $\hat{\mathbf{H}}$ to construct a precoding and receiver shaping matrices, we have obtained subchannels that are not independent, but instead behave like an interference channel with $\tilde{\mathbf{E}}$ representing channel gains from interferers. Specifically, the i^{th} SISO link is described by $y_i = D_{ii}x_i + \tilde{E}_{ii}x_i + \left(\sum_{j=1,j\neq i}^{t} \tilde{E}_{ij}x_j\right) + n_i$ where D_{ii} can be interpreted as an estimated subchannel gain, $\tilde{E}_{ii}$ as its channel estimation error, and the third and the last term together are viewed as a non-Gaussian noise process with an average power of $\sigma_{\mathbf{E}}^2(P-P_i)+1$. The mutual information $I(x_i; y_i|D_{ii})$ is in general difficult to compute because of the non-Gaussian nature of the noise plus interference, but the generalized mutual information (GMI), which is the achievable rate under an i.i.d. Gaussian input distribution and nearest neighbor decoding rule, is known in this case and given by [4]

$$I_i = \log_2\left(1+\frac{P_i D_{ii}^2}{(\sigma_{\mathbf{E}}^2(P-P_i)+1)+(\sigma_{\mathbf{E}}^2 P_i)}\right) = \log_2\left(1+\frac{P_i\lambda_i}{1+\sigma_{\mathbf{E}}^2 P}\right) \tag{14}$$

Summing over $m = \min\{r, t\}$ subchannels and using the optimal power allocation, we obtain the capacity lower bound in (12). Thus, the lower bound can be interpreted as the maximum achievable data rates of communication systems that are designed to perform optimally with perfect channel knowledge but ignore the channel estimation error. Those systems will typically use SVD, a Gaussian input, and a nearest neighbor decoder to achieve capacity, but fail to give optimal performance in the presence of channel estimation error, and only achieve the lower bound (12).

4 Optimal Power Allocation

As we have seen, C_{lower} is the supremum of achievable data rates in practical transmission systems that employ Gaussian codebooks and nearest neighbor decoders. Moreover, we have seen that the difference between C_{lower} and the exact capacity is small for Gaussian inputs. Hence, in this section, we treat C_{lower} as a performance measure and concentrate on deriving the optimal power allocation strategy to achieve it.

4.1 Capacity in Rayleigh fading without CSIT

In this subsection, we derive the lower bound of capacity when the transmitter does not know the instantaneous value of the channel estimation. Since $\hat{\mathbf{H}}$ is a complex Gaussian matrix with i.i.d. entries, the optimal $\mathbf{Q}$ that maximizes I_{lower} is $(P/t)\,\mathbf{I}$ with a certain power adaptation P [7]. Using Jenson's inequality it can be verified that the temporal power adaptation does not increase the capacity, i.e., $P = \bar{P}$. Therefore, the maximum mutual information is given by

$$C_{lower} = \sum_{i=1}^{t} E\left(\log_2\left(1 + \frac{\bar{P}/t}{1+\sigma_{\mathbf{E}}^2\bar{P}}\lambda_i\right)\right) \tag{15}$$

4.2 Optimal spatial power allocation with CSIT

In this and the next subsection, we derive the optimal power allocation for the capacity lower bound (12). Towards this end, we will derive the lower bound of capacity for transmitters using a full spatio-temporal power allocation. In this subsection, however, as an intermediate step, we do not allow adapting power over the channel variation: we fix the total transmit power to be constant over time and distribute the power optimally over the subchannel domain. The objective function is then, from (12),

$$\begin{array}{c} C_{lower} = \max_{\{P_i\}} E\left(\sum_{i=1}^{t}\log_2\left(1 + \frac{P_i\lambda_i}{1+\sigma_{\mathbf{E}}^2 P}\right)\right) \\ \text{subject to} \quad P = \sum_{i=1}^{t} P_i \le \bar{P} \end{array} \tag{16}$$

It can be easily verified that it is always better to use full power $P = \bar{P}$. Thus, the optimal power allocation and the corresponding capacity bound will be given by a water-filling over subchannels with the total power scaled by the variance of the estimation error

$$P_i = \left(\mu - \frac{1+\sigma_{\mathbf{E}}^2\bar{P}}{\lambda_i}\right)^+, \qquad C_{lower} = E\left(\sum_{i=1}^{m}\left[\log_2\left(\frac{\mu\lambda_i}{1+\sigma_{\mathbf{E}}^2\bar{P}}\right)\right]^+\right) \tag{17}$$

with μ chosen to satisfy $P = \sum_{i=1}^{t} P_i = \bar{P}$.

4.3 Optimal spatio-temporal power allocation with CSIT

It follows from the previous subsection that the optimal *spatial* power allocation for a given estimated channel $\hat{\mathbf{H}}$ with a given power budget $P(\hat{\mathbf{H}})$ is

$$P_i\left(P(\hat{\mathbf{H}}), \hat{\mathbf{H}}\right) = \left(\mu\left(P(\hat{\mathbf{H}}), \hat{\mathbf{H}}\right) - \frac{1+\sigma_{\mathbf{E}}^2 P(\hat{\mathbf{H}})}{\lambda_i}\right)^+ \tag{18}$$

$$C_{lower}\left(P(\hat{\mathbf{H}}), \hat{\mathbf{H}}\right) = \sum_{i=1}^{m}\left[\log_2\left(\frac{\mu\left(P(\hat{\mathbf{H}}), \hat{\mathbf{H}}\right)\lambda_i}{1+\sigma_{\mathbf{E}}^2 P(\hat{\mathbf{H}})}\right)\right]^+ \tag{19}$$

where $\mu\left(P(\hat{\mathbf{H}}), \hat{\mathbf{H}}\right)$ represents the water-level associated with $P(\hat{\mathbf{H}})$ and $\hat{\mathbf{H}}$. Then, it remains to find the optimal *temporal* power adaptation $P(\hat{\mathbf{H}})$ that maximizes the expected value of (19):

$$\begin{array}{c} C_{lower} = \max_{P(\hat{\mathbf{H}})} E_{\hat{\mathbf{H}}}\left\{C_{lower}\left(P(\hat{\mathbf{H}}), \hat{\mathbf{H}}\right)\right\} \\ \text{subject to} \quad E_{\hat{\mathbf{H}}}\left(P(\hat{\mathbf{H}})\right) \le \bar{P} \end{array} \tag{20}$$

Forming Lagrange multipliers and differentiating both sides with respect to $P(\hat{\mathbf{H}})$, we get the following condition

$$\frac{\partial C_{lower}\left(P(\hat{\mathbf{H}}),\hat{\mathbf{H}}\right)}{\partial P(\hat{\mathbf{H}})} = \frac{1}{\nu \ln 2} \tag{21}$$

where ν is a constant that represents the global water level. By the following lemma, the above condition becomes both necessary and sufficient, and thus achieves the global maximum of (20).

Lemma 2 *The partial derivative on LHS of (21) is given by*

$$\frac{\partial C_{lower}\left(P(\hat{\mathbf{H}}),\hat{\mathbf{H}}\right)}{\partial P(\hat{\mathbf{H}})} = \frac{1}{\ln 2}\frac{1}{\mu\left(P(\hat{\mathbf{H}}),\hat{\mathbf{H}}\right)\left(1+\sigma_{\mathbf{E}}^2 P(\hat{\mathbf{H}})\right)} \tag{22}$$

which is a decreasing function of $P(\hat{\mathbf{H}})$. *Thus,* $C_{lower}\left(P(\hat{\mathbf{H}}),\hat{\mathbf{H}}\right)$ *is concave in* $P(\hat{\mathbf{H}})$. *(Refer to [9] for proof.)*

Equations (21)-(22) suggest that the marginal capacity gain for a specific channel realization always decreases as we assign more power, and that the optimal temporal power adaptation strategy is to pour power until the marginal capacity gain for each channel realization drops to a constant value that is determined by the average available transmit power. Solving (21)-(22) for $P(\hat{\mathbf{H}})$, we obtain the optimal *temporal* power adaptation

$$P(\hat{\mathbf{H}}) = \left(\frac{-(\lambda_0+2\sigma_{\mathbf{E}}^2)+\sqrt{\lambda_0^2+4k(\hat{\mathbf{H}})\nu\lambda_0\sigma_{\mathbf{E}}^2(\lambda_0+\sigma_{\mathbf{E}}^2)}}{2\sigma_{\mathbf{E}}^2(\lambda_0+\sigma_{\mathbf{E}}^2)}\right)^+ \tag{23}$$

where $k(\hat{\mathbf{H}})$ is the number of subchannels that have positive power $P_i(\hat{\mathbf{H}}) > 0$, and λ_0, which is a scalar value that represents the matrix channel, satisfies $\lambda_0^{-1} = \sum_{i=1}^{k(\hat{\mathbf{H}})} \lambda_i^{-1}$. Equation (23) is a direct extension of the similar result for SISO channels in [2]. Note that as $\sigma_{\mathbf{E}}^2 \to 0$, $\mu\left(P(\hat{\mathbf{H}}),\hat{\mathbf{H}}\right) \to \nu$ and thus (18) and (23) become a two-dimensional water-filling with a single water level ν. When $\sigma_{\mathbf{E}}^2 > 0$, however, we need two levels of power allocation; the *global* water-level ν determines how much power to use through (23) given the channel estimation, and the *local* water-level $\mu(\hat{\mathbf{H}})$ dictates how to distribute the power to the subchannels through (18). Then the capacity is given by (20), and is achieved using input covariance $\mathbf{Q} = \mathbf{V}^*\text{diag}(P_1,\cdots,P_t)\mathbf{V}$. In the following subsections, we study some special cases to gain further intuition.

4.4 MISO and SIMO channels

In multiple input single output (MISO) and single input multiple output (SIMO) channels where there is only one spatial dimension, the optimal power allocation in (23) simplifies to

$$P(\lambda) = \left(\frac{-(\lambda+2\sigma_{\mathbf{E}}^2)+\sqrt{\lambda^2+4\nu\lambda\sigma_{\mathbf{E}}^2(\lambda+\sigma_{\mathbf{E}}^2)}}{2\sigma_{\mathbf{E}}^2(\lambda+\sigma_{\mathbf{E}}^2)}\right)^+ \tag{24}$$

where $\lambda = \|\hat{\mathbf{H}}\|^2$ is the squared norm of the channel estimation vector $\hat{\mathbf{H}}$. This formula is exactly the same as the SISO result in [2]. For MISO channels, the optimal input covariance matrix that achieves C_{lower} is $\mathbf{Q} = P(\lambda)\hat{\mathbf{H}}^*\hat{\mathbf{H}}/\|\hat{\mathbf{H}}\|^2$, which means that the optimal transmission scheme employs both a beamforming along the direction given by $\hat{\mathbf{H}}^*/\|\hat{\mathbf{H}}\|$ and an optimal power adaptation $P(\lambda)$. The power adaptation modulates the transmit power in time; whereas the beamformer allocates the given power at any time instance over the transmit antennas to maximize the transmission rate.

4.5 Low SNR or large estimation error regime

When $P_i\lambda_i \ll 1 + \sigma_{\mathbf{E}}^2 P$, (20) is approximated as

$$C_{lower} \approx \max_{P(\hat{\mathbf{H}})} E\left(\frac{P(\hat{\mathbf{H}})\lambda_{max}}{1+\sigma_{\mathbf{E}}^2 P(\hat{\mathbf{H}})}\right) \tag{25}$$

This means that beamforming or using a single subchannel is the asymptotically optimal spatial power allocation at low SNR or large estimation error. Using Lagrange multipliers, the optimal temporal power adaptation is found to be $P(\hat{\mathbf{H}}) = \left((\sqrt{\nu\lambda_{max}} - 1)/\sigma_{\mathbf{E}}^2\right)^+$.

4.6 High SNR and small estimation error regime

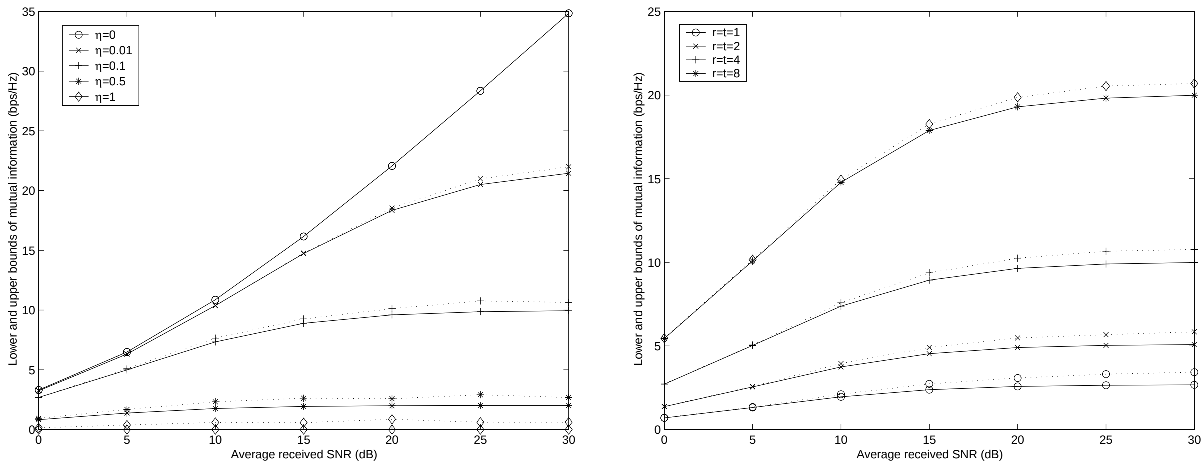
When $P_i\lambda_i \gg 1 + \sigma_{\mathbf{E}}^2 P$, it can be shown that the uniform power allocation over space and time is asymptotically optimal. Thus, (20) is approximated as

$$C_{lower} \approx m \log_2 \frac{\bar{P}}{m\left(1+\sigma_{\mathbf{E}}^2 \bar{P}\right)} + E(\log_2 |\mathbf{W}|) \tag{26}$$

where the $m \times m$ matrix $\mathbf{W}$ is defined as $\mathbf{W} = \hat{\mathbf{H}}\hat{\mathbf{H}}^*$ if $t > r$, and $\mathbf{W} = \hat{\mathbf{H}}^*\hat{\mathbf{H}}$ if $t \leq r$. Note that as $\bar{P} \to \infty$, C_{lower} does not go to infinity but approaches a finite rate, $-m\log_2(m\sigma_{\mathbf{E}}^2) + E(\log_2 |\mathbf{W}|)$. This is consistent with the result in [4] that the Gaussian mutual information in the high SNR regime is bounded by the channel uncertainty and becomes independent of SNR. However, some caution is required in interpreting this result. In particular, [4] show that the true Shannon capacity without any input restriction grows double-logarithmically in the SNR. Thus, the Gaussian input distribution is suboptimal in the presence of channel estimation error. Secondly, our result is only valid under the assumption that the quality of channel estimation does not improve as SNR increases. The result, however, becomes entirely different when the channel estimation improves with SNR, which is the case in [5] where it is shown that the noncoherent capacity of block fading channels and the capacity of pilot-based schemes still have a logarithmic dependence on the SNR but with a reduced slope compared to the coherent capacity. Finally, our result is different from [5, 6] in that we assume i.i.d. fading and genie-provided channel information, whereas in [5, 6] the channel knowledge is obtained through the fading correlation. In that case, the capacity will be much smaller in an i.i.d. fading channel because the receiver cannot estimate the channel reliably.

5 Numerical Results

In this section, numerical results are presented based on Monte Carlo simulations. In Figures 1-(a) and 1-(b) we plot the lower (6) and upper (10) bound of mutual information. The plots confirm our previous observations that the two bounds are tight and that the



(a) (b)

Figure 1: (a) Lower and upper bounds of mutual information for 4×4 MIMO channel vs. SNR for several channel estimation accuracies η. (b) Lower and upper bounds of mutual information for several MIMO channels vs. SNR for 10% estimation error.

Gaussian mutual information is bounded in the SNR but still increases linearly with the number of transmit and receive antennas.

In Figure 2, we compare capacity lower bounds using different power allocation strategies; no CSIT (15), spatial power allocation (17), and spatio-temporal power allocation (20). First we observe that the difference between (17) and (20) is negligible, which implies that temporal power adaptation gives little capacity gain, as has been shown in the literature [3, 8] for a single antenna case.

Comparing (15) and (17), however, we observe that spatial power allocation does help. Typically, without channel estimation error, the capacity gain of knowing the channel at the transmitter reduces at higher SNR, because the optimal covariance matrix approaches the identity matrix, which is also the optimal covariance matrix when the channel is unknown at the transmitter. This trend, however, changes with channel estimation error. The capacity gain of exploiting transmitter channel knowledge becomes more important with increasing channel estimation error and doesn't reduce much at high SNRs. This is because the channel estimation error reduces the effective SNR and causes saturation, thereby eliminating the high SNR capacity region where transmitter channel knowledge becomes unimportant.

6 Conclusions

We have investigated the effect of channel estimation error in fading MIMO channels. We have developed lower and upper bounds of mutual information for systems with MMSE channel estimation and perfect feedback. We have shown that the lower bound is close to the exact mutual information for Gaussian inputs, and that it is in fact the supremum of achievable data rates for transceivers that are designed to be optimal in the absence of estimation error. Despite the channel estimation error, the mutual information increases linearly with the smaller of the number of transmit and receive antennas, but it is limited by the estimation error in the high SNR regime. Based on the lower bound, we have derived the optimal transmitter power allocation with and without estimated channel

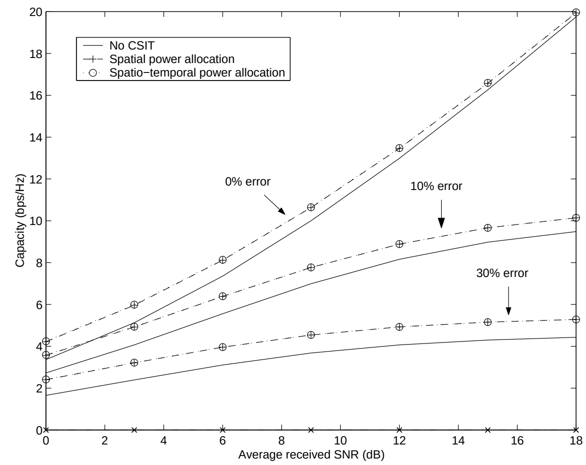


Figure 2: Capacity for 4 × 4 MIMO channel with different power allocation strategies for 0%, 10%, and 30% estimation error.

knowledge at the transmitter. Numerical results show that spatial power allocation becomes more important under channel estimation error and helps even at high SNR, whereas temporal power adaptation gives negligible gain in terms of ergodic capacity.

References

[1] M. Medard, "The Effect upon Channel Capacity in Wireless Communications of Perfect and Imperfect Knowledge of the Channel," *IEEE Trans. Information Theory*, vol. 46, pp. 933-946, May 2000.

[2] T. E. Klein, R. G. Gallager, "Power control for the additive white Gaussian noise channel under channel estimation errors," *Proceedings. 2001, IEEE International Symposium on Information Theory*, p. 304, 2001

[3] Caire, S. Shamai, "On the Capacity of Some Channels with Channel State Information," *IEEE Trans. Information Theory*, vol. 45, pp. 2007-2019, Sep. 1999

[4] A. Lapidoth, S. Shamai, "Fading Channels: How Perfect Need "Perfect Side Information" Be?" *IEEE Trans. Information Theory*, vol. 48, pp. 1118-1134, May 2002.

[5] L. Zheng, D. Tse, "Communication on the Grassmann Manifold: A Geometric Approach to the Noncoherent Multiple-Antenna Channel," *IEEE Trans. Information Theory*, vol. 48, pp. 359-383, Feb. 2002

[6] J. Baltersee, G. Fock, H. Meyr, "Achievable Rate of MIMO Channels With Data-Aided Channel Estimation and Perfect Interleaving," *IEEE Journal on Selected Areas in Communications*, vol. 19, pp. 2358-2368, Dec. 2001

[7] E. Telatar, "Capacity of Multi-antenna Gaussian Channels," *European Trans. on Telecom.*, vol. 10, pp 585-598, Nov. 1999

[8] A. Goldsmith, P. Varaiya, "Capacity of Fading Channels with Channel Side Information," *IEEE Trans. Information Theory*, vol. 43, pp. 1986-1992, Nov. 1997

[9] T. Yoo, A. Goldsmith, "Capacity of fading MIMO channels with channel estimation error," in preparation