# Capacity and Power Allocation for Fading MIMO Channels With Channel Estimation Error

Taesang Yoo, *Student Member, IEEE,* and Andrea Goldsmith, *Fellow, IEEE*

***Abstract*—In this correspondence, we investigate the effect of channel estimation error on the capacity of multiple-input–multiple-output (MIMO) fading channels. We study lower and upper bounds of mutual information under channel estimation error, and show that the two bounds are tight for Gaussian inputs. Assuming Gaussian inputs we also derive tight lower bounds of ergodic and outage capacities and optimal transmitter power allocation strategies that achieve the bounds under perfect feedback. For the ergodic capacity, the optimal strategy is a modified waterfilling over the spatial (antenna) and temporal (fading) domains. This strategy is close to optimum under small feedback delays, but when the delay is large, equal powers should be allocated across spatial dimensions. For the outage capacity, the optimal scheme is a spatial waterfilling and temporal truncated channel inversion. Numerical results show that some capacity gain is obtained by spatial power allocation. Temporal power adaptation, on the other hand, gives negligible gain in terms of ergodic capacity, but greatly enhances outage performance.**

***Index Terms*—Capacity, channel estimation error, feedback delay, multiple-input–multiple-output (MIMO), mutual information, outage capacity, power allocation, waterfilling.**

## I. Introduction

Multiple-input–multiple-output (MIMO) systems provide dramatic capacity gain through an increased spatial dimension [1], [2]. However, the capacity gain is reduced if the channel state information (CSI) is not perfect. Since perfect CSI is difficult to obtain in MIMO systems due to an increased number of channel parameters to estimate at the receiver and to be fed back to the transmitter, the channel capacity with imperfect CSI is an important problem to investigate. There have been several approaches in this area. In [3], [4], Lapidoth, *et al.* show that in the absence of CSI the MIMO capacity only grows double-logarithmically as a function of signal-to-noise ratio (SNR), and we do not benefit from the increased spatial dimensions. However, under certain conditions, MIMO systems with a reasonable channel estimation accuracy can achieve linear increase of capacity at practical SNR values [5]. For example, in [6], [7], the authors study capacity of MIMO channels under a block fading assumption and show that the capacity increases logarithmically in the SNR but with a reduced slope. Thus, it is important to specify the channel fading condition since it affects the accuracy of CSI and the way CSI is obtained. CSI at the receiver (CSIR) is typically obtained via channel estimation. Pilot or training based channel estimation schemes are studied in [7]–[9] for block fading channels and channels with a band-limited Doppler spectrum, respectively, and their achievable rates are derived. It is seen that the capacity is decreased by the reduced SNR due to the channel estimation error and by the loss of available dimensions used for channel estimation.

In this correspondence, we assume that channel estimation is done in a separate background channel and does not consume any dimension, i.e., partial CSI is provided by a genie at the receiver. Then, we concentrate on the effect of channel estimation error on the capacity of MIMO fading channels. This approach is taken in [10], where the author derives a lower bound of mutual information with channel estimation error for single-input–single-output (SISO) channels. In our work, we extend this result to MIMO channels and derive a lower and upper bound of mutual information. Since MIMO channels provide spatial dimensions, an appropriate spatial filter should be designed to shape the input distribution to maximize mutual information. Fading channels provide another degree of freedom for the transmitter optimization; transmit power can be adapted to the channel variation to increase the capacity. In SISO channels, the optimal power adaptation that maximizes the ergodic capacity is given by waterfilling over the channel fading [11]. This optimal power adaptation, however, should be modified when we have channel estimation error; in [12], the author revises the optimal power adaptation based on the mutual information lower bound in [10]. In this correspondence, these results are extended to MIMO channels. Specifically, we study lower and upper bounds of mutual information for a Rayleigh flat fading channel with a genie-provided MMSE channel estimate at the receiver. Then, assuming a perfect and instantaneous feedback, we derive optimal transmitter strategies that maximize the lower bound of mutual information to obtain the corresponding lower bounds of ergodic and outage capacities. The optimal transmitter strategies involve power allocations over both spatial (antenna) and temporal (fading) domains. We derive those strategies for both ergodic and outage capacities. We also consider the effect of feedback delay and show that the strategies designed for perfect feedback are still applicable with little capacity loss when the delay is small.

The rest of this correspondence is organized as follows. In Section II, our system model is introduced. In Section III, 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 IV, optimal power allocation strategies are determined for both ergodic and outage capacities. Finally, numerical results are presented in Section V.

## II. System Model

Consider a MIMO system with $t$ transmit and $r$ receive antennas. The discrete-time channel at time $n$ 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). Throughout this correspondence, we will use upper case boldface letters for matrices and lower case boldface for vectors. We assume both $\mathbf{H}_n$ and $\mathbf{z}_n$ are ergodic and stationary, and their entries are independent, identically distributed (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. In such a system, if the fading process $\{\mathbf{H}_n\}$ is *perfectly* known to the receiver, the mutual information between the channel input and output is given by [1]

$$I(\mathbf{x};\mathbf{y}) = \mathcal{E}\left\{\log_2\left|\mathbf{I} + \mathbf{H}_n^*\mathbf{H}_n\mathbf{Q_n}\right|\right\} \tag{1}$$

where $\mathbf{Q_n}$ is the input covariance matrix $\mathbf{Q_n} = \mathcal{E}(\mathbf{x}_n\mathbf{x}_n^*)$, and $\mathcal{E}\{\bullet\}$ is an expectation operator. $\mathbf{H}_n^*$ denotes the conjugate transpose of $\mathbf{H}_n$.

Now, consider the situation where $\{\mathbf{H}_n\}$ is *imperfectly* known to the receiver, i.e., the receiver is provided with some partial information $\{\tilde{\mathbf{H}}_n\}$ of the channel, with which the receiver performs minimum mean square error (MMSE) estimation of $\{\mathbf{H}_n\}$. We assume $\{\mathbf{H}_n\}$

Manuscript received August 1, 2004; revised December 13, 2005. This work was supported in part by the National Science Foundation (NSF) under Grant CCR-0325639-001 and by LG electronics. The material in this correspondence was presented in part at the IEEE International Conference on Communications (ICC), Paris, France, June 2004.

The authors are with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA (e-mail: yoots@wsl.stanford.edu; andrea@wsl.stanford.edu).

Communicated by M. Médard, Associate Editor for Communications.
Digital Object Identifier 10.1109/TIT.2006.872984

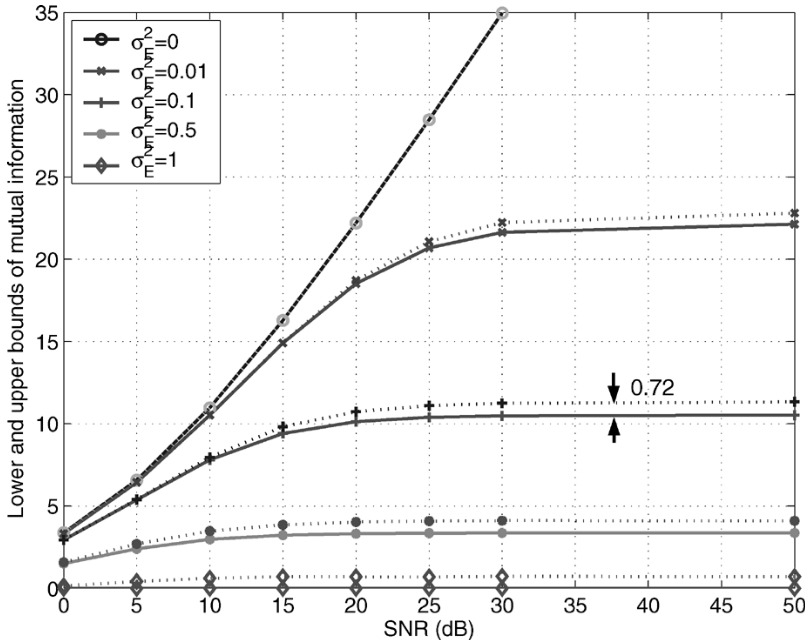


Fig. 1. Lower and upper bounds of mutual information under Gaussian inputs for $4 \times 4$ MIMO channel versus SNR ($P$) for several channel estimation accuracies $\sigma_{\mathbf{E}}^2$.

and $\{\tilde{\mathbf{H}}_n\}$ are jointly ergodic and stationary Gaussian processes, and the entries of $\tilde{\mathbf{H}}_n$ are independent. Denote the MMSE estimation as $\hat{\mathbf{H}}_n = \mathcal{E}(\mathbf{H}_n|\tilde{\mathbf{H}}_n, \tilde{\mathbf{H}}_{n-1}, \ldots)$, and the estimation error as $\mathbf{E}_n = \mathbf{H}_n - \hat{\mathbf{H}}_n$. Then, by the property of MMSE estimation, $\hat{\mathbf{H}}_n$ and $\mathbf{E}_n$ are uncorrelated, and the entries of $\mathbf{E}_n$ are ZMCSCG with variance $\sigma_{\mathbf{E}}^2 = \mathrm{MMSE} = \mathcal{E}(|\mathbf{H}_{n,ij}|^2) - \mathcal{E}(|\hat{\mathbf{H}}_{n,ij}|^2)$, where $\mathbf{H}_{n,ij}$ and $\hat{\mathbf{H}}_{n,ij}$ represent the $(i,j)$th elements of $\mathbf{H}_n$ and $\hat{\mathbf{H}}_n$, respectively. The entries of $\hat{\mathbf{H}}_n$ are also i.i.d. ZMCSCG with variance $1-\sigma_{\mathbf{E}}^2$. Note that $\sigma_{\mathbf{E}}^2$ is a parameter that captures the quality of the channel estimation and is assumed to be known to both the transmitter and the receiver. $\sigma_{\mathbf{E}}^2$ can be appropriately chosen depending on the channel dynamics and channel estimation schemes. Some examples are as follows.

1) In [8], the authors consider a block Rayleigh fading channel of coherence time $T$ and use orthonormal training signals which is optimal for spatially white inputs. With this setting they obtain $\sigma_{\mathbf{E}}^2 = 1/\left(1 + \frac{T_\tau}{t} P_\tau\right)$, where $T_\tau$ is the training interval length and $P_\tau$ is the transmit power (average received SNR) of the training symbols.
2) In [9], the authors consider a continuously time-varying Rayleigh-fading channel with a bandlimited lowpass spectrum with the cutoff frequency $F$. Using pilot symbols with a sampling rate $1/L \geq 2F$, they show that $\sigma_{\mathbf{E}}^2 = 1/\left(1 + \frac{1}{2FL} P_\tau\right)$.
3) In [5], the channel is modeled as a Gauss-Markov process. The authors propose a decision-oriented training scheme which gives an upper bound $\sigma_{\mathbf{E}}^2 \leq t2^t\epsilon + t/\bar{P}$, where $\epsilon$ is a parameter related to the channel coherence time and $\bar{P}$ is the average transmit power. The mean square error (MSE) resulting from practical MIMO channel estimators can often be approximated in the form $\sigma_{\mathbf{E}}^2 = \sigma_u + \sigma_n^2/\bar{P}$, e.g., [13].

In practice, the channel estimation process itself may incur loss of bandwidth and energy. However, since the focus of this correspondence is to study the effect of channel estimation error, we assume that $\hat{\mathbf{H}}_n$ and $\sigma_{\mathbf{E}}^2$ are given *a priori*, through, for example, channel estimation in a separate background channel, and we concentrate on analyzing its effect on capacity. See [8] for the discussion regarding trade-offs between channel estimation accuracy and bandwidth/energy loss due to channel estimation in training based systems.

The transmitter typically obtains CSI via channel feedback. To simplify analysis, we assume that the feedback is instantaneous and error-free as in [12], which implies that whatever CSI the receiver has is also available at the transmitter. This analysis could serve as a basis for more practical situations where the feedback is imperfect or limited as in Section IV-D, where we consider the case that CSIT is different from CSIR. This situation can arise either due to feedback delays or in time division duplex (TDD) systems where CSIT can be obtained directly using the reciprocity of the channel. We will assume that the transmitter is constrained in its total power $P_n$, and it can adapt its power to the channel fading to maximize capacity, i.e., $\mathcal{E}(P_n) = \mathcal{E}\left(\mathrm{Tr}(\mathbf{Q_n})\right) \leq \bar{P}$, where $\mathrm{Tr}(\bullet)$ stands for trace. Under $\mathbf{Q}_n = (\bar{P}/t)\mathbf{I}$, $\bar{P}$ is equal to the average SNR at each receive antenna [14]. Therefore, we will simply refer to $\bar{P}$ as the average SNR. In the remainder of this correspondence, we will often omit the time index $n$ when it is obvious from the context[1] and use the subscript for other indexing purposes; $\mathbf{M}_{ij}$ and $\mathbf{v}_k$ represent the $(i,j)$th element of the matrix $\mathbf{M}$ and the $k$th element of the vector $\mathbf{v}$, respectively. This use of subscript should not be confused with the time indices and will be obvious from the context.

## III. Capacity Bounds With Channel Estimation Error

### A. Lower and Upper Bounds of Mutual Information

In this section, we develop lower and upper bounds of mutual information $I(\mathbf{x};\mathbf{y}|\hat{\mathbf{H}})$ given an estimated channel knowledge $\hat{\mathbf{H}}$.

*Lemma 1:* A lower bound of mutual information is given by [8]–[10]

$$I_{\mathrm{lower}}(\mathbf{x};\mathbf{y}|\hat{\mathbf{H}}) = \mathcal{E}\left[\log_2 \left|\mathbf{I} + \frac{1}{1+\sigma_{\mathbf{E}}^2 P}\hat{\mathbf{H}}^*\hat{\mathbf{H}}\mathbf{Q}\right|\right]. \quad (2)$$

*Proof:* See Appendix I. □

Comparing (2) to (1), we observe that the channel estimation error results in an SNR loss factor of at most $\delta = (1-\sigma_{\mathbf{E}}^2)/(1+\sigma_{\mathbf{E}}^2 P)$ [9].

[1]This is possible because the random processes we are dealing with are stationary.

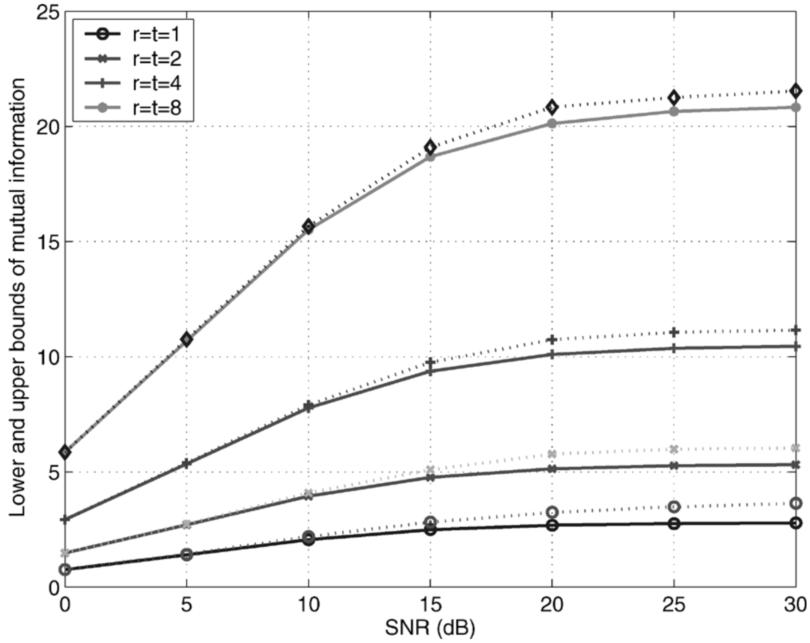


Fig. 2. Lower and upper bounds of mutual information for several antenna configurations versus SNR ($P$) for $\sigma_{\mathbf{E}}^2 = 0.1$.

*Lemma 2:* An upper bound of mutual information is given by [9]

$$I_{\rm upper}(\mathbf{x};\mathbf{y}|\hat{\mathbf{H}}) = I_{\rm lower}(\mathbf{x};\mathbf{y}|\hat{\mathbf{H}}) + r\ \mathcal{E}\left[\log_2 \frac{1+\sigma_{\mathbf{E}}^2 P}{1+\sigma_{\mathbf{E}}^2\|\mathbf{x}\|^2}\right] \quad (3)$$

where the expectation is taken over the joint distribution of $\mathbf{x}$. □

*Proof:* See Appendix II. □

Note that the second term in (3) is nonnegative by Jensen's inequality with $\mathcal{E}(\|\mathbf{x}\|^2) \le P$. Also note that we have not restricted the input distribution $p(\mathbf{x})$ to be Gaussian in deriving the upper bound (3). However, the upper bound (3), which we need to maximize over $p(\mathbf{x})$, is quite loose for certain non-Gaussian input distributions and thus may not be useful. Therefore, we are more interested in evaluating the performance of Gaussian inputs, although not optimal, when used under channel estimation error. The following lemma shows that the gap between the lower (2) and the upper (3) bounds is usually small for a Gaussian input unless $r \gg t$. In other words, the two bounds are approximately equal to the exact mutual information under Gaussian inputs.

*Lemma 3:* In the limit of high SNR and a large number of antennas, the second term in (3) approaches $(r/t)\log_2\sqrt{e} \approx 0.72(r/t)$ for Gaussian inputs.

*Proof:* See Appendix III. □

In Figs. 1 and 2 we plot the lower (2) and upper (3) bounds of mutual information. The plots not only confirm Lemma 3 but also show that the gap is small for any SNR values.

### B. Lower Bound of Ergodic Capacity

The ergodic capacity of the fading channel model in Section II with an estimated channel $\hat{\mathbf{H}}$ known at the transmitter and the receiver is given by [15]

$$C = \mathcal{E}\left[\max_{p(\mathbf{x}|\hat{\mathbf{H}})} I(\mathbf{x};\mathbf{y}|\hat{\mathbf{H}})\right] \quad (4)$$

where $p(\mathbf{x}|\hat{\mathbf{H}})$ is the probability distribution of $\mathbf{x}$ given $\hat{\mathbf{H}}$.[2] In this section, we obtain a lower bound of (4) by finding the optimal input distribution $p(\mathbf{x}|\hat{\mathbf{H}})$ that maximizes (2). Since the lower bound (2) of $I(\mathbf{x};\mathbf{y}|\hat{\mathbf{H}})$ depends on $p(\mathbf{x}|\hat{\mathbf{H}})$ only through $\mathbf{Q}$, it remains to find the optimal $\mathbf{Q}$ as a function of $\hat{\mathbf{H}}$. Let the singular value decomposition (SVD) 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 $\boldsymbol{\Lambda}\mathbf{D}^*\mathbf{D}$. Then

$$I_{\rm lower}(\mathbf{x};\mathbf{y}|\hat{\mathbf{H}}) = \mathcal{E}\left[\log_2\left|\mathbf{I} + \frac{1}{1+\sigma_{\mathbf{E}}^2 P}\boldsymbol{\Lambda}\tilde{\mathbf{Q}}\right|\right]. \quad (5)$$

Under an average power constraint $\mathcal{E}(P) = \mathcal{E}(\mathrm{Tr}(\mathbf{Q})) \le \bar{P}$, observing that $\mathrm{Tr}(\mathbf{Q}) = \mathrm{Tr}(\tilde{\mathbf{Q}})$, (5) is maximized with $\tilde{\mathbf{Q}}$ a diagonal matrix, $\tilde{\mathbf{Q}} = \mathrm{diag}(P_1,\ldots,P_t)$, 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

$$C_{\rm lower} = \mathcal{E}\left\{\max_{\{P_i\}}\sum_{i=1}^t \log_2\left(1+\frac{P_i\lambda_i}{1+\sigma_{\mathbf{E}}^2 P}\right)\right\}$$
$$\text{subject to}\quad \mathcal{E}(P) = \mathcal{E}\left(\sum_{i=1}^t P_i\right) \le \bar{P} \quad (6)$$

where $\lambda_i$ is the $(i,i)$th element of $\boldsymbol{\Lambda}$ and thus the $i$th eigenvalue of $\hat{\mathbf{H}}^*\hat{\mathbf{H}}$. The above expectations are performed over the joint distribution of $(\lambda_1,\ldots,\lambda_t)$. The input to the channel that achieves the capacity has covariance matrix of the form $\mathbf{Q} = \mathbf{V}\mathrm{diag}(P_1,\ldots,P_t)\mathbf{V}^*$ whose optimal subchannel powers $\{P_i\}$ are determined as functions of $(\lambda_1,\ldots,\lambda_t)$. Finding the $P_i$'s will be the main topic of Section IV.

The capacity lower bound (6) can be achieved by the following. First, 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. After all, as a result of using imperfect channel information $\hat{\mathbf{H}}$ to construct a precoding and receiver shaping matrices,

[2] Note that (4) should more precisely be written as $C = \mathcal{E}[\max_{p(\mathbf{x}|\hat{\mathbf{H}}_0)} I(\mathbf{x};\mathbf{y}|\hat{\mathbf{H}} = \hat{\mathbf{H}}_0)]$ but we used a simplified notation to keep subsequent notations simple.

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

$$\tilde{\mathbf{y}}_i = \mathbf{D}_{ii}\tilde{\mathbf{x}}_i + \tilde{\mathbf{E}}_{ii}\tilde{\mathbf{x}}_i + \left(\sum_{j=1, j\neq i}^{t} \tilde{\mathbf{E}}_{ij}\tilde{\mathbf{x}}_j\right) + \tilde{\mathbf{z}}_i \tag{7}$$

where $\mathbf{D}_{ii}$ can be interpreted as an estimated subchannel gain, $\tilde{\mathbf{E}}_{ii}$ as its channel estimation error, and the third and the last term together are viewed as an additive non-Gaussian noise with an average power of $N_i = 1 + \sigma_{\mathbf{E}}^2(P - P_i)$. The mutual information $I(\tilde{\mathbf{x}}_i; \tilde{\mathbf{y}}_i|\mathbf{D}_{ii})$ of this channel 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 [3]

$$I(\tilde{\mathbf{x}}_i; \tilde{\mathbf{y}}_i|\mathbf{D}_{ii}) = \log_2\left(1 + \frac{P_i\mathbf{D}_{ii}^2}{N_i + \sigma_{\mathbf{E}}^2 P_i}\right) \tag{8}$$

$$= \log_2\left(1 + \frac{P_i\lambda_i}{1 + \sigma_{\mathbf{E}}^2 P}\right). \tag{9}$$

Summing over $m = \min\{r, t\}$ subchannels and using the optimal power allocation, we obtain the capacity lower bound in (6). Thus, the lower bound can be interpreted as the maximum achievable data rate of communication systems that are designed to perform optimally with perfect channel knowledge but ignore channel estimation error. Those systems will typically use SVD, Gaussian inputs, and nearest neighbor decoders to achieve capacity, but fail to give optimal performance in the presence of channel estimation error, and only achieve the lower bound.

## IV. Optimal Power Allocation and Capacity Lower Bounds

As we have seen, $C_{\text{lower}}$ 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 $C_{\text{lower}}$ and the exact Gaussian capacity is small unless $r \gg t$. Hence, in this section, we treat $C_{\text{lower}}$ as a performance measure and concentrate on deriving the optimal power allocation strategy to achieve it for each of the following cases.

### A. Rayleigh Fading Channels, No CSIT

In Rayleigh fading channels, $\hat{\mathbf{H}}$ is ZMCSCG with i.i.d. entries as is stated in Section II. Without CSIT, the optimal $\mathbf{Q}$ that maximizes (6) is spatially white, i.e., $\mathbf{Q} = (P/t)\mathbf{I}$. Moreover, it can be verified by using Jensen's inequality that the temporal power adaptation does not increase the capacity, i.e., $P = \bar{P}$. Therefore, the capacity lower bound is given by

$$C_{\text{lower, NP}} = \sum_{i=1}^{t} \mathcal{E}\left[\log_2\left(1 + \frac{\bar{P}/t}{1 + \sigma_{\mathbf{E}}^2\bar{P}}\lambda_i\right)\right]. \tag{10}$$

### B. CSIT, Spatio–Temporal Power Allocation for Ergodic Capacity

In this section, we derive the optimal spatio–temporal power allocation that maximizes ergodic capacity in (6) when the transmitter knows $\hat{\mathbf{H}}$ through feedback. Note that the function $\sum_{i=1}^{t}\log_2\left(1 + \frac{P_i\lambda_i}{1+\sigma_{\mathbf{E}}^2\sum_{j=1}^{t}P_j}\right)$ in (6) is not concave in $[P_1, \ldots, P_t]^{\mathrm{T}}$ for $\sigma_{\mathbf{E}}^2 > 0$, due to the presence of $P_k$ in the denominator. Hence it cannot be directly optimized. However, for a fixed $P = \sum_{j=1}^{t} P_j$, the function becomes concave. This suggests a way to find the optimal spatio-temporal power allocation for (6) by a two-step approach as follows.

Let us define a temporal power adaptation policy $P(\hat{\mathbf{H}})$ as the choice of a transmit power at the transmitter with estimated channel knowledge $\hat{\mathbf{H}}$. The power adaptation policy shall be called valid if it satisfies an average power constraint, $\mathcal{E}[P(\hat{\mathbf{H}})] \leq \bar{P}$. Let us denote the set of all valid power adaptation policies as $\mathcal{P} = \left\{P(\hat{\mathbf{H}}) \mid \mathcal{E}[P(\hat{\mathbf{H}})] \leq \bar{P}\right\}$. Now consider a fixed power adaptation policy $P(\hat{\mathbf{H}}) \in \mathcal{P}$. It can be easily shown that the optimal *spatial* power allocation for a given estimated channel $\hat{\mathbf{H}}$ under the fixed power adaptation policy $P(\hat{\mathbf{H}})$ is given by a water-filling over subchannels with the subchannel gains scaled by $\sigma_{\mathbf{E}}^2$ as

$$P_i(P(\hat{\mathbf{H}}), \hat{\mathbf{H}}) = \left(\mu(P(\hat{\mathbf{H}}), \hat{\mathbf{H}}) - \frac{1 + \sigma_{\mathbf{E}}^2 P(\hat{\mathbf{H}})}{\lambda_i}\right)^+ \tag{11}$$

and the (instantaneous) capacity lower bound is

$$C_{\text{lower}}(P(\hat{\mathbf{H}}), \hat{\mathbf{H}}) = \sum_{i=1}^{m}\left[\log_2\left(\frac{\mu(P(\hat{\mathbf{H}}), \hat{\mathbf{H}})\lambda_i}{1 + \sigma_{\mathbf{E}}^2 P(\hat{\mathbf{H}})}\right)\right]^+ \tag{12}$$

where $\mu(P(\hat{\mathbf{H}}), \hat{\mathbf{H}})$ represents the *local* water-level associated with $P(\hat{\mathbf{H}})$ and $\hat{\mathbf{H}}$, and is chosen to satisfy $\sum_{i=1}^{t} P_i(P(\hat{\mathbf{H}}), \hat{\mathbf{H}}) = P(\hat{\mathbf{H}})$. Then, it remains to find the optimal *temporal* power adaptation policy $P(\hat{\mathbf{H}}) \in \mathcal{P}$ that maximizes the expected value of (12)

$$C_{\text{lower}} = \max_{P(\hat{\mathbf{H}}) \in \mathcal{P}} \mathcal{E}\left[C_{\text{lower}}(P(\hat{\mathbf{H}}), \hat{\mathbf{H}})\right]. \tag{13}$$

An interesting suboptimal choice is $P(\hat{\mathbf{H}}) = \bar{P}$, which corresponds to the case with only spatial power allocation and gives the following capacity lower bound:

$$C_{\text{lower, SP}} = \mathcal{E}\left(\sum_{i=1}^{m}\left[\log_2\left(\frac{\mu(\bar{P}, \hat{\mathbf{H}})\lambda_i}{1 + \sigma_{\mathbf{E}}^2\bar{P}}\right)\right]^+\right). \tag{14}$$

The optimal solution to (13) can be found using a Lagrangian method. Forming Lagrange multipliers and differentiating both sides with respect to $P(\hat{\mathbf{H}})$, we get the following condition:

$$\frac{\partial C_{\text{lower}}(P(\hat{\mathbf{H}}), \hat{\mathbf{H}})}{\partial P(\hat{\mathbf{H}})} = \frac{1}{\nu \ln 2} \tag{15}$$

where $\nu$ is a constant that represents the *global* water level. In Appendix IV, we show that the partial derivative in the LHS of (15) is given by

$$\frac{\partial C_{\text{lower}}(P(\hat{\mathbf{H}}), \hat{\mathbf{H}})}{\partial P(\hat{\mathbf{H}})} = \frac{1}{\ln 2}\frac{1}{\mu(P(\hat{\mathbf{H}}), \hat{\mathbf{H}})\left(1 + \sigma_{\mathbf{E}}^2 P(\hat{\mathbf{H}})\right)} \tag{16}$$

which is a decreasing function of $P(\hat{\mathbf{H}})$, implying that $C_{\text{lower}}(P(\hat{\mathbf{H}}), \hat{\mathbf{H}})$ is concave in $P(\hat{\mathbf{H}})$. Thus, (15) becomes both necessary and sufficient, and achieves the global maximum of (13).

Equations (15)–(16) suggest that the marginal capacity gain for a specific channel realization $\hat{\mathbf{H}}$ always decreases as we assign more power, and that the optimal temporal power adaptation strategy $P(\hat{\mathbf{H}})$ is to pour power until the marginal capacity gain for each channel realization drops to a constant value $1/(\nu \ln 2)$ that is determined by the average available transmit power $\bar{P}$. These facts are illustrated in Fig. 3. The following theorem gives the closed-form expression for the optimal transmitter power adaptation policy as a function of channel estimate $\hat{\mathbf{H}}$.

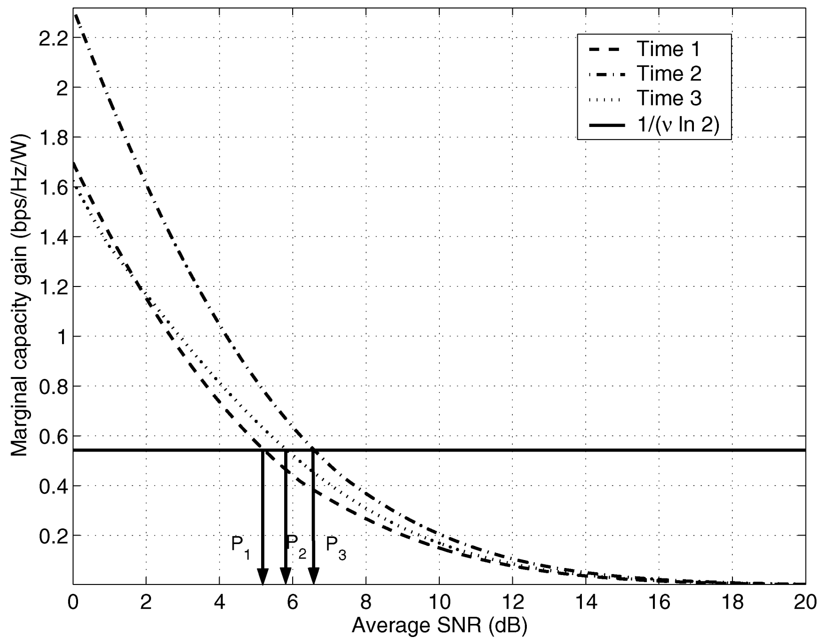


Fig. 3. Marginal capacity gains at three different $\hat{\mathbf{H}}$. The horizontal line represents $1/(\nu \ln 2)$ for $\bar{P} = 6$ dB and $\sigma_{\mathbf{E}}^2 = 0.1$. For each $\hat{\mathbf{H}}$ the transmit power is determined at the point where the marginal capacity gain intersects this horizontal line.

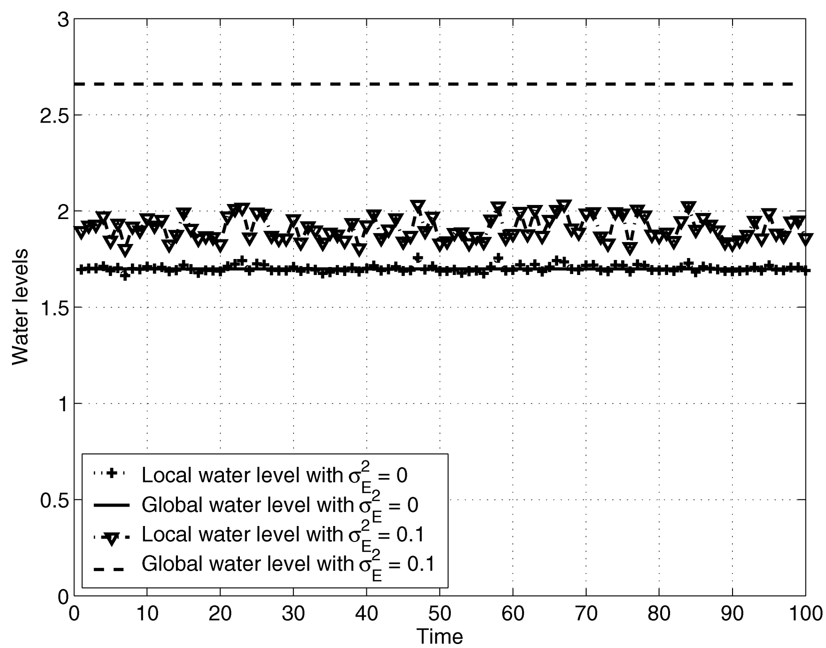


Fig. 4. Global ($\nu$) and local ($\mu$) water levels. These two levels coincide when $\sigma_{\mathbf{E}}^2 = 0$.

*Theorem 1:* The optimal *temporal* power adaptation policy is given as

$$P(\hat{\mathbf{H}}) = \frac{\left[-(\lambda_0 + 2\sigma_{\mathbf{E}}^2) + \sqrt{\lambda_0^2 + 4k(\hat{\mathbf{H}})\nu\lambda_0\sigma_{\mathbf{E}}^2(\lambda_0 + \sigma_{\mathbf{E}}^2)}\right]^+}{2\sigma_{\mathbf{E}}^2(\lambda_0 + \sigma_{\mathbf{E}}^2)} \quad (17)$$

where $k(\hat{\mathbf{H}})$ is the number of subchannels that have positive powers, i.e., $k(\hat{\mathbf{H}}) = \sum_{i=1}^{t} I\left(P_i(\hat{\mathbf{H}}) > 0\right)$, $I(\cdot)$ is an indicator function, and $\lambda_0$, which is a scalar value that represents the matrix channel, satisfies $\lambda_0^{-1} = \sum_{i=1}^{k(\hat{\mathbf{H}})} \lambda_i^{-1}$.

*Proof:* See Appendix V. □

Equation (17) is a direct extension of the similar result for SISO channels in [12]. Practical systems could use either (17) or an iterative algorithm based on (15)–(16) [16]. The iterative algorithm will always converge due to the concavity of (12). Note that as $\sigma_{\mathbf{E}}^2 \to 0$, $\mu(P(\hat{\mathbf{H}}), \hat{\mathbf{H}}) \to \nu$ and thus (11) and (17) become a two-dimensional water-filling with a single water level $\nu$. When $\sigma_{\mathbf{E}}^2 > 0$, however, we need two levels of power allocation as is shown in Fig. 4; the *global* water-level $\nu$ determines how much power to use through (17) given the channel estimation $\hat{\mathbf{H}}$, and the *local* water-level $\mu(\hat{\mathbf{H}})$ dictates how to distribute the power to the subchannels through (11). Then the capacity is given by (12)–(13), and is achieved using input covariance $\mathbf{Q} = \mathbf{V}\mathrm{diag}(P_1, \ldots, P_t)\mathbf{V}^*$.

To gain further intuition, we examine special cases and asymptotic behaviors of the optimal power allocation and capacity lower bounds.

*1) 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 (17) simplifies to

$$P(\lambda) = \left( \frac{-(\lambda + 2\sigma_{\mathbf{E}}^2) + \sqrt{\lambda^2 + 4\nu\lambda\sigma_{\mathbf{E}}^2(\lambda + \sigma_{\mathbf{E}}^2)}}{2\sigma_{\mathbf{E}}^2(\lambda + \sigma_{\mathbf{E}}^2)} \right)^+ \tag{18}$$

where $\lambda = \|\hat{\mathbf{H}}\|^2$ is the squared norm of the channel estimation vector $\hat{\mathbf{H}}$. This formula is the same as the SISO result in [12] except that $\lambda$ is redefined as an effective scalar gain of the MISO (SIMO) channel. For MISO channels, the optimal input covariance matrix that achieves $C_{\text{lower}}$ 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[3] along the direction given by $\hat{\mathbf{H}}^*/\|\hat{\mathbf{H}}\|$ and a power adaptation $P(\lambda)$ over time. The power adaptation modulates the transmit power in time; whereas the beamformer allocates the given power at any time instant over the transmit antennas to maximize the instantaneous rate.

*2) Low SNR or Large Estimation Error Regime:* When $P_i\lambda_i \ll 1 + \sigma_{\mathbf{E}}^2 P$, we can approximate (6) and (13) as

$$C_{\text{lower}} \doteqdot \max_{P(\hat{\mathbf{H}})\in\mathcal{P}} \mathcal{E}\left( \max_{\{P_i\}} \frac{1}{\ln 2} \frac{\sum_{i=1}^t P_i\lambda_i}{1 + \sigma_{\mathbf{E}}^2 P(\hat{\mathbf{H}})} \right), \tag{19}$$

where the approximation symbol $\doteqdot$ denotes that the ratio of its two sides converges to one as $(P_i\lambda_i)/(1 + \sigma_{\mathbf{E}}^2 P) \to 0$. The inner maximization is achieved by assigning all the power to the subchannel with the highest gain. Therefore, we have

$$C_{\text{lower}} \doteqdot \max_{P(\hat{\mathbf{H}})\in\mathcal{P}} \mathcal{E}\left( \frac{1}{\ln 2} \frac{P(\hat{\mathbf{H}})\lambda_{\max}}{1 + \sigma_{\mathbf{E}}^2 P(\hat{\mathbf{H}})} \right) \tag{20}$$

where $\lambda_{\max} = \max\{\lambda_1, \ldots, \lambda_t\}$. This means that beamforming is the asymptotically optimal *spatial* power allocation at low SNR or large estimation error. Intuitively, this is because the effective subchannel gains $\lambda_i/(1 + \sigma_{\mathbf{E}}^2 P)$ are so low in this region that the waterfill covers only the strongest subchannel. Using Lagrange multipliers, the optimal *temporal* power adaptation policy is found to be

$$P(\hat{\mathbf{H}}) = \left( \frac{\sqrt{\nu\lambda_{\max}} - 1}{\sigma_{\mathbf{E}}^2} \right)^+. \tag{21}$$

*3) 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 both spatial subchannels and time domain is asymptotically optimal. Thus, (6) and (13) can be approximated as

$$C_{\text{lower}} \risingdotseq m \log_2 \frac{\bar{P}}{m\left(1 + \sigma_{\mathbf{E}}^2 \bar{P}\right)} + \mathcal{E}(\log_2 |\mathbf{W}|) \tag{22}$$

where $m = \min\{r, t\}$, the number of spatial subchannels, and the $m \times m$ Wishart 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 $t \le r$ [17]. The symbol $\risingdotseq$ means that the difference of its two sides vanishes as $(P_i\lambda_i)/(1 + \sigma_{\mathbf{E}}^2 P) \to \infty$. An analytical expression for (22) can be obtained by noting that $\hat{\mathbf{H}}$ is complex Gaussian. Following similar steps as in [18], we have

$$C_{\text{lower}} \risingdotseq m \log_2 \frac{(1 - \sigma_{\mathbf{E}}^2)\bar{P}}{m(1 + \sigma_{\mathbf{E}}^2 \bar{P})} + \frac{1}{\ln 2}\left( \sum_{j=1}^{m}\sum_{k=1}^{n-j} \frac{1}{k} - \gamma m \right) \tag{23}$$

where $m = \min\{r, t\}$, $n = \max\{r, t\}$, and $\gamma \approx 0.577$ denotes Euler's constant.

[3]Beamforming in this correspondence is defined as the use of a single spatial subchannel along one eigen-direction.

*4) High SNR and Non-Vanishing Estimation Error Regime:* [4]When $\bar{P} \to \infty$ and $\lim_{\bar{P}\to\infty} \sigma_{\mathbf{E}}^2 \triangleq \sigma_u^2 > 0$, as in the example (3) in Section II, it can be shown that the uniform power allocation over time domain is asymptotically optimal, and that (13) becomes

$$\lim_{\bar{P}\to\infty} C_{\text{lower}} = \mathcal{E}\left[ \sum_{i=1}^{k(\hat{\mathbf{H}})} \log_2 \left( \frac{\mu'(\hat{\mathbf{H}})\lambda_i}{\sigma_u^2} \right) \right] \tag{24}$$

where $\mu'(\hat{\mathbf{H}})$ is a water level that determines subchannel powers as

$$\frac{P_i}{\bar{P}} = \left( \mu'(\hat{\mathbf{H}}) - \frac{\sigma_u^2}{\lambda_i} \right)^+. \tag{25}$$

Thus, as $\bar{P} \to \infty$, the capacity lower bound does not go to infinity but approaches a finite rate. However, some caution is required in interpreting this result. In particular, Lapidoth *et al.* in [3], [4] show that the true Shannon capacity without any input restriction grows double-logarithmically in the SNR. However, they also point out that under Gaussian inputs the mutual information in the high SNR regime is bounded by the channel uncertainty and becomes independent of SNR. This explains why $C_{\text{lower}}$, which is based on a Gaussian input, is upper bounded as (24) at high SNR. Second, note that here we have assumed that $\sigma_{\mathbf{E}}^2$ does not vanish as SNR increases. The result, however, becomes entirely different when the channel estimation improves proportionally to SNR, which is the case in [7]–[9] 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. This fact can also be confirmed from (22) by using $\sigma_{\mathbf{E}}^2$ in the examples (1) or (2) in Section II. Thus, assumptions on channel estimation error greatly impact the high SNR behavior of capacity lower bounds.

### *C. CSIT, Spatio–Temporal Power Allocation for Outage Capacity*

In the previous sections, we have derived the optimal spatio–temporal power allocation to maximize the *ergodic* capacity of the channel. In this section, we use the transmitter channel knowledge $\hat{\mathbf{H}}$ to obtain a full spatio–temporal power allocation that maximizes the *outage* capacity. We declare an outage if for a given channel estimation $\hat{\mathbf{H}}$, a power adaptation policy $P(\hat{\mathbf{H}})$ cannot support a specified rate threshold $R_{\text{out}}^0$. Note that the optimal spatial power allocation in (11) is still applied to the outage capacity, since it maximizes the instantaneous data rate for a given $\hat{\mathbf{H}}$ and $P(\hat{\mathbf{H}})$. Therefore, an outage will be declared if and only if $C_{\text{lower}}(P(\hat{\mathbf{H}}), \hat{\mathbf{H}}) < R_{\text{out}}^0$, where $C_{\text{lower}}(P(\hat{\mathbf{H}}), \hat{\mathbf{H}})$ is as defined in (12), and the outage probability is given by

$$q = \Pr\left\{ C_{\text{lower}}(P(\hat{\mathbf{H}}), \hat{\mathbf{H}}) < R_{\text{out}}^0 \right\}. \tag{26}$$

Our objective is to find $P(\hat{\mathbf{H}}) \in \mathcal{P}$ that minimizes this outage.

Since $C_{\text{lower}}$ in (12) is a strictly increasing function of $P$ for a given $\hat{\mathbf{H}}$, an inverse function exists

$$P = f(C_{\text{lower}}, \hat{\mathbf{H}}) \tag{27}$$

that gives the minimum required transmit power to achieve the capacity lower bound $C_{\text{lower}}$ for a given channel estimation $\hat{\mathbf{H}}$. Thus, for the outage rate $R_{\text{out}}^0$, we need a transmit power of $P_{\min} = f(R_{\text{out}}^0, \hat{\mathbf{H}})$. This is the *minimum* transmit power to avoid an outage with $\hat{\mathbf{H}}$. Let us now define the set of channel estimation matrices $\mathcal{H}(P_{\text{cutoff}}, R_{\text{out}}^0)$ for which $P_{\min}$ is less than a certain threshold $P_{\text{cutoff}}$

$$\mathcal{H}(P_{\text{cutoff}}, R_{\text{out}}^0) = \left\{ \hat{\mathbf{H}} \,\middle|\, f(R_{\text{out}}^0, \hat{\mathbf{H}}) < P_{\text{cutoff}} \right\}. \tag{28}$$

[4]Note that the regimes 3) and 4) overlap.

Roughly, $\mathcal{H}$ represents the set of *good* channels. Then, the power adaptation policy that minimizes the outage probability is given by

$$P(\hat{\mathbf{H}}) = \begin{cases} f(R_{\rm out}^0, \hat{\mathbf{H}}) & \text{if } \hat{\mathbf{H}} \in \mathcal{H} \\ 0 & \text{if } \hat{\mathbf{H}} \notin \mathcal{H} \end{cases} \tag{29}$$

with $P_{\rm cutoff}$ chosen to satisfy

$$\int_{\mathcal{H}} f(R_{\rm out}^0, \hat{\mathbf{H}}) p(\hat{\mathbf{H}}) d\hat{\mathbf{H}} = \bar{P}. \tag{30}$$

This form of power adaptation is a truncated channel inversion. This minimizes the average transmit power for a given outage rate $R_{\rm out}^0$ by turning the transmitter off while the channel is in a deep fade and using just enough power to avoid outage when the channel is in a good state. Since this minimizes the transmit power, it makes the set $\mathcal{H}$ the largest possible ($P_{\rm cutoff}$ is maximized) and thus yields minimum outage probability $q = \Pr\left\{\hat{\mathbf{H}} \notin \mathcal{H}\right\}$.

A lower bound of outage capacity, $C_{\rm outage}$, for a given outage probability $q_0$, can be obtained by numerically finding $R_{\rm out}$ such that $q = q_0$.

### D. Effect of Feedback Delay

The developments so far have assumed an instantaneous and error-free feedback. With feedback delay, however, blindly adapting the transmit power using (11) and/or (17) may hurt the data rate when the feedback delay is significant, since the channel has changed to a new state by the time the transmitter has the channel estimate. Clearly, the power allocation policies should be modified accordingly. Let $d$ denote the feedback delay in number of symbols and define the correlation coefficient $\rho$ between CSIT and CSIR as

$$\rho = \frac{\mathcal{E}(\hat{\mathbf{H}}_n \hat{\mathbf{H}}_{n-d})}{\sqrt{\mathcal{E}(|\hat{\mathbf{H}}_n|^2)\mathcal{E}(|\hat{\mathbf{H}}_{n-d}|^2)}} = \frac{\mathcal{E}(\hat{\mathbf{H}}_n \hat{\mathbf{H}}_{n-d})}{1-\sigma_{\mathbf{E}}^2}. \tag{31}$$

The transmitter best estimates $\mathbf{H}_n$ from $\hat{\mathbf{H}}_{n-d}$ in a MMSE sense as

$$\begin{aligned} \bar{\mathbf{H}}_n &= \mathcal{E}(\mathbf{H}_n | \hat{\mathbf{H}}_{n-d}) & (32)\\ &= \mathcal{E}(\hat{\mathbf{H}}_n + \mathbf{E}_n | \hat{\mathbf{H}}_{n-d}) & (33)\\ &= \mathcal{E}(\hat{\mathbf{H}}_n | \hat{\mathbf{H}}_{n-d}) & (34)\\ &= \rho \hat{\mathbf{H}}_{n-d}. & (35) \end{aligned}$$

Thus, the channel is decomposed as $\mathbf{H}_n = \bar{\mathbf{H}}_n + \mathbf{G}_n + \mathbf{E}_n$ where $\mathbf{G}_n = \hat{\mathbf{H}}_n - \rho\hat{\mathbf{H}}_{n-d}$ is the channel estimation error at the transmitter additionally incurred due to the delay. In fact, this formulation is valid not only for the systems where CSIT is a delayed version of CSIR, but also for the general cases where CSIT can be obtained independently from CSIR, such as in TDD systems, with an appropriate definition of $\rho$. Note that $\bar{\mathbf{H}}_n$ is known to both the transmitter and the receiver, while $\hat{\mathbf{H}}_n$, or equivalently, $\mathbf{G}_n$, is available only at the receiver. In this case, the ergodic capacity lower bound is obtained by

$$\begin{aligned} C_{\rm lower}^{(\rho)} &= \mathcal{E}\left[\max_{p(\mathbf{x}|\bar{\mathbf{H}})} I_{\rm lower}(\mathbf{x};\mathbf{y}|\hat{\mathbf{H}})\right] & (36)\\ &= \mathcal{E}\left\{\max_{\mathbf{Q}} \log_2 \left|\mathbf{I} + \frac{(\bar{\mathbf{H}}+\mathbf{G})\mathbf{Q}(\bar{\mathbf{H}}+\mathbf{G})^*}{1+\sigma_{\mathbf{E}}^2 P(\bar{\mathbf{H}})}\right|\right\} & (37) \end{aligned}$$

where $\mathbf{Q}$ is now a function of $\bar{\mathbf{H}}$, and $P(\bar{\mathbf{H}}) = \text{Tr}(\mathbf{Q})$. Unfortunately, there is no known closed-form solution to the above problem. However, the following theorem shows that the optimal $\mathbf{Q}$ has a simple structure, and the optimal power allocation therein can be determined numerically.

*Theorem 2:* Without loss of generality, (37) is maximized by the covariance matrix $\mathbf{Q} = \mathbf{V}\tilde{\mathbf{Q}}\mathbf{V}^*$, where $\mathbf{V}$ is a unitary matrix obtained from the SVD of $\bar{\mathbf{H}} = \mathbf{U}\mathbf{D}\mathbf{V}^*$, and $\tilde{\mathbf{Q}}$ is a diagonal matrix whose entries can be determined by solving the following convex optimization problem:

$$C_{\rm lower}^{(\rho)}(P(\bar{\mathbf{H}}), \bar{\mathbf{H}}) = \max_{\tilde{\mathbf{Q}}} \mathcal{E}\left\{\log_2 \left|\mathbf{I} + \frac{(\mathbf{D}+\tilde{\mathbf{G}})\tilde{\mathbf{Q}}(\mathbf{D}+\tilde{\mathbf{G}})^*}{1+\sigma_{\mathbf{E}}^2 P(\bar{\mathbf{H}})}\right|\right\}$$
$$\text{subject to } \tilde{\mathbf{Q}} \geq 0,\ \text{Tr}(\tilde{\mathbf{Q}}) \leq P(\bar{\mathbf{H}}),\ \tilde{\mathbf{Q}} \text{ is diagonal} \tag{38}$$

where $\tilde{\mathbf{G}}$ is a random matrix with the same distribution as $\mathbf{G}$, i.e., the entries of $\tilde{\mathbf{G}}$ are independently generated from ZMCSCG distribution with variance $(1-\sigma_{\mathbf{E}}^2)(1-\rho^2)$. The above expectation is taken over $\tilde{\mathbf{G}}$. Once $C_{\rm lower}^{(\rho)}(P(\bar{\mathbf{H}}), \bar{\mathbf{H}})$ is found, $C_{\rm lower}^{(\rho)}$ is obtained by optimizing it over feasible $P(\bar{\mathbf{H}})$ as

$$C_{\rm lower}^{(\rho)} = \max_{P(\bar{\mathbf{H}}):\mathcal{E}[P(\bar{\mathbf{H}}) \leq \bar{P}]} \mathcal{E}\left[C_{\rm lower}^{(\rho)}(P(\bar{\mathbf{H}}), \bar{\mathbf{H}})\right]. \tag{39}$$

*Proof:* The essential part of the proof is to observe that (37) is equivalent to the capacity of a MIMO channel with an AWGN variance $1+\sigma_{\mathbf{E}}^2 P(\bar{\mathbf{H}})$ and a mean feedback (or a Rician component) $\bar{\mathbf{H}}$. This problem has recently been solved in [19], [20], where the authors show that the optimal transmitter strategy is the eigenmode transmission along the eigenvectors of $\bar{\mathbf{H}}^*\bar{\mathbf{H}}$. □

However, finding the optimal $P(\bar{\mathbf{H}})$ is computationally intensive. Therefore, in the numerical discussion in Section V, we consider a suboptimal lower bound, $C_{\rm lower,\ SP}^{(\rho)} = \mathcal{E}[C_{\rm lower}^{(\rho)}(\bar{P}, \bar{\mathbf{H}})]$, obtained using a constant transmit power $P(\bar{\mathbf{H}}) = \bar{P}$. Even finding $C_{\rm lower,\ SP}^{(\rho)}$ requires the optimization in (38) to be performed at each symbol time, i.e., for each realization of $\bar{\mathbf{H}}$, and thus may not be realizable in practice. More discussion on this will follow in the next section with a numerical example.

## V. Numerical Results

In this section, simulation results are presented for $4 \times 4$ MIMO channels. Throughout this section, we use two different models for the channel estimation error: a) $\sigma_{\mathbf{E}}^2$ is independent of the SNR $\bar{P}$, and b) $\sigma_{\mathbf{E}}^2$ is a decreasing function of $\bar{P}$. In the latter case we use a simple model $\sigma_{\mathbf{E}}^2 = \sigma_u^2 + \sigma_n^2/\bar{P}$, where $\sigma_u^2$ can be interpreted as a prediction error caused by the time-varying nature of the channel, and $\sigma_n^2/\bar{P}$ as a measurement error caused by AWGN. Thus, $\sigma_u^2$ does not depend on the SNR $\bar{P}$, while $\sigma_n^2/\bar{P}$ improves with $\bar{P}$.[5] This model appears, for example, in [5] where the authors show that the channel estimation error of their decision-oriented training scheme is upper bounded by $\sigma_{\mathbf{E}}^2 \leq \sigma_u^2 + \sigma_n^2/\bar{P} = t2^t\epsilon + t/\bar{P}$, where $\epsilon$ is a constant that is related to the Doppler shift of the fading channel.

In Fig. 5, we compare capacity lower bounds using three different power allocation strategies; no CSIT ($C_{\rm lower,\ NP}$), spatial power allocation ($C_{\rm lower,\ SP}$), and spatio-temporal power allocation ($C_{\rm lower}$). The calculation is performed for three channel estimation error values; $\sigma_{\mathbf{E}}^2 = 0$, $\sigma_{\mathbf{E}}^2 = 0.1$, and $\sigma_{\mathbf{E}}^2 = 0.03 + 0.8/\bar{P}$. In the figure we observe that the difference between $C_{\rm lower,\ SP}$ and $C_{\rm lower}$ is negligible, which implies that temporal power adaptation gives little capacity gain, as has been shown in the literature [11], [15] for a single antenna case. Comparing the performance of $C_{\rm lower,\ NP}$ and $C_{\rm lower,\ SP}$, however, we observe that spatial power allocation does help. Moreover, the capacity gain of spatial power allocation becomes more pronounced as $\sigma_{\mathbf{E}}^2$ increases. Typically, when $\sigma_{\mathbf{E}}^2 = 0$, the capacity gain of knowing the

[5]Ideally, $\sigma_u^2$ would be zero if the Doppler spectrum is strictly bandlimited and a perfect noncausal filter is used for MMSE estimation.

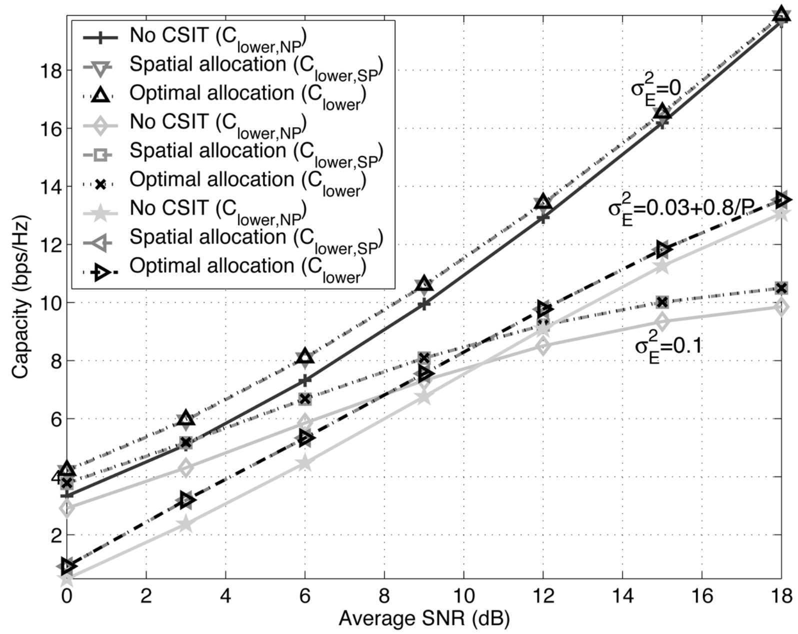


Fig. 5. Ergodic capacity for $4 \times 4$ MIMO channel with different power allocation strategies for $\sigma_{\mathbf{E}}^2 = 0$, $\sigma_{\mathbf{E}}^2 = 0.1$, and $\sigma_{\mathbf{E}}^2 = 0.03 + 0.8/\bar{P}$.

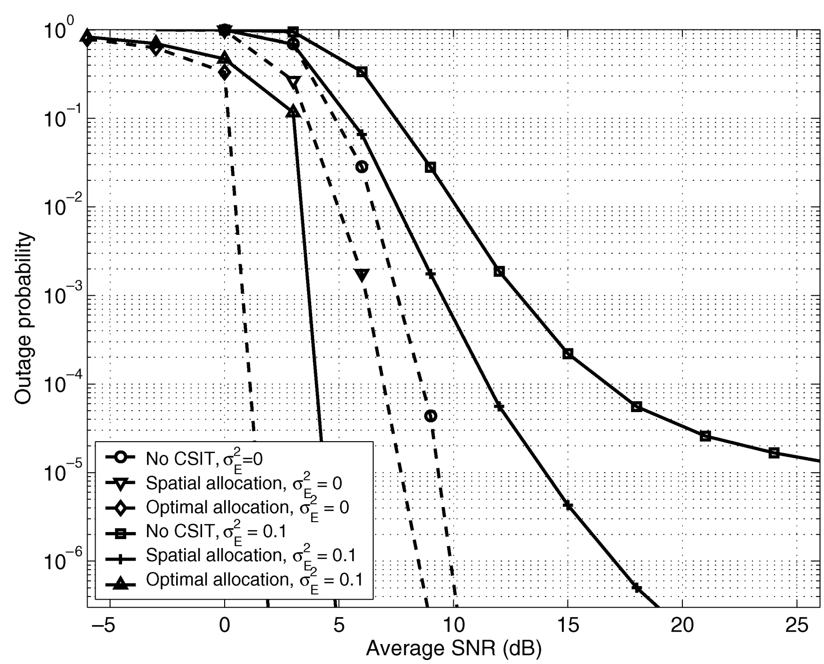


Fig. 6. Outage probability for $4 \times 4$ MIMO channel with different power allocation strategies for $\sigma_{\mathbf{E}}^2 = 0$ and $\sigma_{\mathbf{E}}^2 = 0.1$. $R_{\text{out}} = 5.5$ bps/Hz.

channel at the transmitter decreases as SNR increases [14]. This is because when $\sigma_{\mathbf{E}}^2 = 0$ and as $\bar{P} \to \infty$, the optimal spatial power allocation policy given by (11) approaches the uniform power allocation $P_i = \bar{P}/t$, which is also the optimal solution to the case of no CSIT. Therefore, when $\sigma_{\mathbf{E}}^2 = 0$, transmitter channel knowledge becomes unimportant at high SNR, and consequently, the spatial power allocation is only useful at low and medium SNR values. When $\sigma_{\mathbf{E}}^2 > 0$, however, the figure shows that the capacity gain of spatial power allocation does not reduce much at high SNRs. This is because the channel estimation error causes saturation in the effective subchannel SNR, $P_i\lambda_i/(1 + \sigma_{\mathbf{E}}^2 P)$, thereby eliminating the high SNR capacity region where transmitter channel knowledge becomes unimportant.

Fig. 6 shows outage probabilities using three different power allocation strategies; no CSIT ($P_i = \bar{P}/t$), spatial power allocation ($P(\hat{\mathbf{H}}) = \bar{P}$ and (11)), and spatio-temporal power allocation ((29) and (11)). We observe that CSIT greatly reduces outage probability. Spatial power allocation seems to give a constant SNR improvement, and temporal power adaptation additionally provides significant improvement in outage probability. Similar observations appear in [21] for MISO systems with perfect channel estimation. Thus, additional degrees of freedom in adapting transmit power over the time domain seems to provide diversity gain and improves the outage slope. Also note that with $\sigma_{\mathbf{E}}^2 = 0.1$ channel estimation error produces an irreducible error floor, i.e., the outage probability cannot be further reduced by increasing SNR. This is because the performance in the high SNR region is limited by channel estimation error. In Fig. 7 we plot corresponding outage capacities using the three different power allocations. We note that the effect of temporal power adaptation (29) on

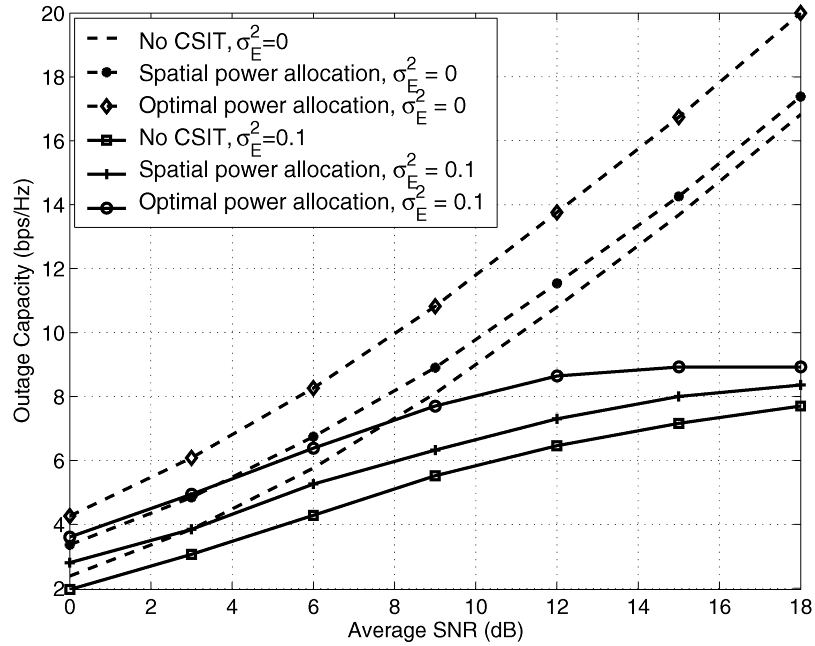


Fig. 7. Outage capacity for $4 \times 4$ MIMO channel with different power allocation strategies for $\sigma_{\mathrm{E}}^2 = 0$ and $\sigma_{\mathrm{E}}^2 = 0.1$ with an outage probability $q = 5\%$.

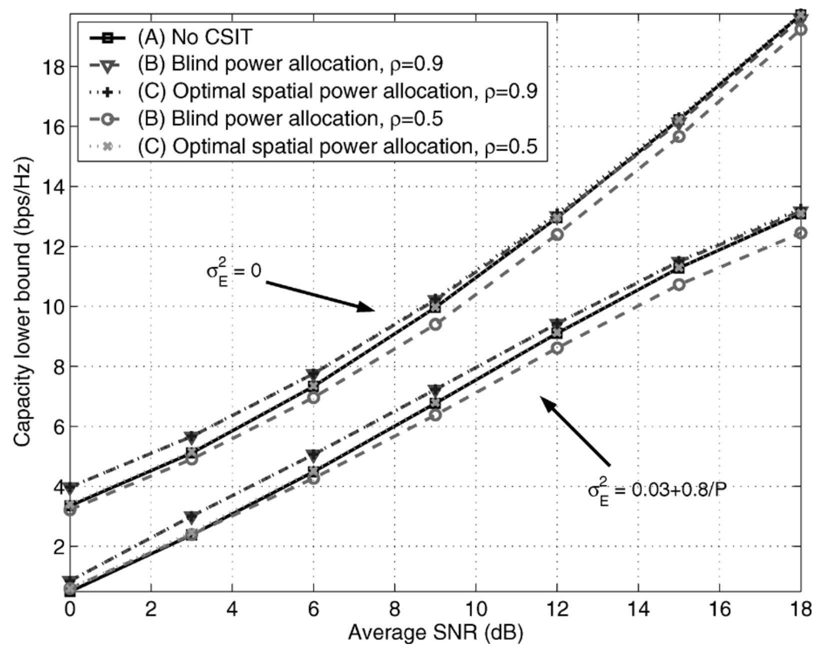


Fig. 8. Ergodic capacity for $4 \times 4$ MIMO channel with different power allocation strategies for combined channel estimation error and feedback delay.

the outage capacity is rather large. This is in contrast to the ergodic capacity where we obtain little gain by adapting the transmit power to the channel fading. Thus, the temporal power adaptation may be especially important for delay-constrained traffic, where the outage rate is the most useful performance measure [22]. Spatial power allocation seems to give a moderate increase to both the outage and ergodic capacity. Similarly, the channel estimation error severely limits the outage performance and causes saturation at high SNR if $\lim_{\bar{P}\to\infty} \sigma_{\mathrm{E}}^2 > 0$.

Fig. 8 shows the combined effect of channel estimation error and feedback delay on the ergodic capacities, and the performance of various power allocation strategies: (*A*) equal power allocation over space and time in case of no CSIT, (*B*) blind spatio–temporal power allocation using (11) and (17) ignoring the delay, and (*C*) the optimal spatial power allocation considering the delay, i.e., $C_{\mathrm{lower,\ SP}}^{(\rho)}$ in Section IV-D.[6] In the figure, it is seen that at high $\rho$ or low delay ($\rho = 0.9$) the performance of the blind power allocation is similar to that of the optimal spatial power allocation in $C_{\mathrm{lower,\ SP}}^{(\rho)}$, meaning that the power allocations in (11) and (17) are still applicable with little capacity loss when the delay is small. To see how much delay is tolerable, assume a uniform scattering environment. The correlation is given by $\rho = J_0(2\pi f_D \tau)$, where $J_0$ is a Bessel function of the first kind of order 0, $f_D$ is the maximum Doppler shift, and $\tau$ is the feedback delay measured in seconds. Then, $\rho = 0.9$ translates to

[6]We consider only the spatial power allocation, since the optimal spatio–temporal power allocation is difficult to compute. However, we expect the performance of the two to be very close to each other, because we have seen in other contexts that temporal power adaptation does not significantly increase capacity [11], [12], [23].

$f_D\tau \approx 0.1$, and under carrier frequency of 2.4 GHz and a pedestrian speed of 4.5 Km/h, this corresponds to a delay of $\tau \approx 10$ ms, which can be achieved in practical systems. At larger delay, however, the blind scheme begins to perform poorly, and as is shown in the figure, it performs even worse than equal power allocation at $\rho = 0.5$, at which value the equal power allocation performs nearly as well as the optimal scheme. As mentioned in Section IV-D, the optimal power allocation requires a large amount of computation. The numerical results presented here suggest that a much simpler yet suboptimal scheme that uses the blind scheme at high $\rho$ and the equal power allocation at low $\rho$ can perform nearly as well as the optimal power allocation.

## VI. Conclusion

We have investigated the effect of channel estimation error in multiple antenna fading channels. We have developed lower and upper bounds of mutual information for systems with MMSE channel estimation, and have shown that these bounds are close to the exact mutual information for Gaussian inputs. The lower bound is the maximum achievable data rate of communication systems that ignore channel estimation error and use Gaussian inputs and nearest neighbor decoders. Despite the channel estimation error, the lower bound increases linearly with the number of antennas, but it approaches a certain limit (24) as SNR goes to infinity unless the channel estimation error vanishes and thus cannot achieve a double-logarithmic increase in the SNR. This is the price of using Gaussian inputs and having channel estimation quality that does not improve with SNR.

When the channel estimate is available at the transmitter through perfect and instantaneous feedback, the transmitter can adapt its strategy to the current channel estimate to maximize the lower bound of mutual information, thereby achieving the capacity lower bound. The optimal transmitter strategy for the capacity lower bound employs spatial filters resulting from singular value decomposition (SVD) of the channel estimate as well as appropriate power allocations over space and time. For the ergodic capacity maximization, the optimal power allocation is a modified waterfilling (11) over the spatial domain and a temporal power adaptation (17). For the outage capacity, the optimal scheme is a spatial modified waterfilling (11) and a temporal truncated channel inversion (29). When feedback delay is present, the optimal spatial beam directions are unchanged, but the power allocations along the spatial channels no longer have a closed form formula, and have to be determined numerically (38). At small delays, however, the modified waterfilling scheme (11) is close to optimum. Numerical results show that spatial power allocation becomes more important under channel estimation error and helps even at high SNR. Temporal power adaptation mainly increases outage capacity, but has little effect on ergodic capacity. Overall, the numerical results show that feedback of imperfect CSI is still useful in enhancing the MIMO system performance.

## Appendix I
## Proof of Lemma 1

To obtain a lower bound, we expand the mutual information into differential entropies as

$$I(\mathbf{x};\mathbf{y}|\hat{\mathbf{H}}) = h(\mathbf{x}|\hat{\mathbf{H}}) - h(\mathbf{x}|\mathbf{y},\hat{\mathbf{H}}). \tag{40}$$

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 [3], [10]), the first term on the RHS becomes $\mathcal{E}\left[\log_2|\pi e\mathbf{Q}|\right]$. The second term, on the other hand, is upper bounded by the entropy of a Gaussian random variable whose variance is equal to the mean square error of the linear MMSE estimate of $\mathbf{x}$ given $\mathbf{y}$ and $\hat{\mathbf{H}}$ as

$$h(\mathbf{x}|\mathbf{y},\hat{\mathbf{H}}) \le \mathcal{E}\left[\log_2\left|\pi e\left(\mathbf{Q} - \mathbf{Q}\hat{\mathbf{H}}^* \cdot\left(\hat{\mathbf{H}}\mathbf{Q}\hat{\mathbf{H}}^* + \Sigma_{\mathbf{Ex}} + \mathbf{I}\right)^{-1}\hat{\mathbf{H}}\mathbf{Q}\right)\right|\right] \tag{41}$$

where $\Sigma_{\mathbf{v}}$ denotes the covariance matrix of a random vector $\mathbf{v}$. Combining (40) – (41) we arrive at the following lower bound of mutual information:

$$I(\mathbf{x};\mathbf{y}|\hat{\mathbf{H}}) \ge \mathcal{E}\left[\log_2\left|\mathbf{I} + \hat{\mathbf{H}}^*(\mathbf{I}+\Sigma_{\mathbf{Ex}})^{-1}\hat{\mathbf{H}}\mathbf{Q}\right|\right]. \tag{42}$$

Since the entries of $\mathbf{H}$ are i.i.d., we have that the entries of the channel estimation error matrix $\mathbf{E}$ are also i.i.d., i.e., $E(\mathbf{E}_{ij}\mathbf{E}_{mn}) = \sigma_{\mathbf{E}}^2\delta_{i-m,j-n}$. Therefore, $\Sigma_{\mathbf{Ex}} = \mathcal{E}\left(\mathbf{E}\mathbf{x}\mathbf{x}^*\mathbf{E}^*\right) = \sigma_{\mathbf{E}}^2 P\mathbf{I}$, and thus (2) follows.

A more elegant proof appears in [8], based on the worst case noise interpretation of the additive noise plus residual channel estimation error.

## Appendix II
## Proof of Lemma 2

To obtain an upper bound, we expand the mutual information (40) in an alternate way

$$I(\mathbf{x};\mathbf{y}|\hat{\mathbf{H}}) = h(\mathbf{y}|\hat{\mathbf{H}}) - h(\mathbf{y}|\mathbf{x},\hat{\mathbf{H}}). \tag{43}$$

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

$$\begin{aligned} h(\mathbf{y}|\hat{\mathbf{H}}) &\le \mathcal{E}\left[\log_2\left|\pi e\Sigma_{\mathbf{y}|\hat{\mathbf{H}}}\right|\right] &(44)\\ &= \mathcal{E}\left[\log_2\left|\pi e\left(\hat{\mathbf{H}}\mathbf{Q}\hat{\mathbf{H}}^* + \left(1+\sigma_{\mathbf{E}}^2P\right)\mathbf{I}\right)\right|\right]. &(45)\end{aligned}$$

Since $\mathbf{E}$ is complex Gaussian, $(\mathbf{y}|\mathbf{x},\hat{\mathbf{H}})$ is also complex Gaussian with mean $\hat{\mathbf{H}}\mathbf{x}$ and variance $\Sigma_{\mathbf{Ex}|\mathbf{x}} + \mathbf{I}$. Thus, the second term on the RHS in (43) becomes

$$\begin{aligned} h(\mathbf{y}|\mathbf{x},\hat{\mathbf{H}}) &= \mathcal{E}\left[h(\mathbf{Ex}+\mathbf{z}|\mathbf{x})\right] &(46)\\ &= \mathcal{E}\left[\log_2\left|\pi e(\Sigma_{\mathbf{Ex}|\mathbf{x}}+\mathbf{I})\right|\right] &(47)\\ &= \mathcal{E}\left[\log_2\left|\pi e\left(1+\sigma_{\mathbf{E}}^2\|\mathbf{x}\|^2\right)\mathbf{I}\right|\right]. &(48)\end{aligned}$$

Combining (43)–(48), we have (3).

## Appendix III
## Proof of Lemma 3

At high SNRs, the second term $\Delta$ in (3) is approximated as

$$\begin{aligned} \lim_{P\to\infty}\Delta &= \lim_{P\to\infty} r\mathcal{E}\left[\log_2\frac{1+\sigma_{\mathbf{E}}^2P}{1+\sigma_{\mathbf{E}}^2\|\mathbf{x}\|^2}\right] &(49)\\ &= \lim_{P\to\infty} r\left[\log_2 P - \mathcal{E}\left(\log_2\|\mathbf{x}\|^2\right)\right]. &(50)\end{aligned}$$

The expected logarithm of the chi-square distributed random variable $\|\mathbf{x}\|^2$ with $\mathcal{E}\left(\|\mathbf{x}\|^2\right) = P$ is given by

$$\mathcal{E}\left(\log_2\|\mathbf{x}\|^2\right) = \log_2\frac{P}{t} + \frac{1}{\ln 2}\left(-\gamma + \sum_{i=1}^{t-1}\frac{1}{i}\right) \tag{51}$$

where $\gamma = \lim_{n\to\infty}\left(\sum_{k=1}^{n} 1/k - \ln n\right) \approx 0.577$ is Euler's constant. Therefore, the limiting value of $\Delta$ becomes

$$\lim_{P\to\infty} \Delta = \frac{r}{\ln 2}\left(\ln t + \gamma - \sum_{i=1}^{t-1}\frac{1}{i}\right). \tag{52}$$

The difference between $t^{th}$ convergent $\sum_{i=1}^{t} 1/i$ and $\gamma$ is bounded by [24]

$$\frac{1}{2(t+1)} < \sum_{i=1}^{t}\frac{1}{i} - \ln t - \gamma < \frac{1}{2t}. \tag{53}$$

Using this bound, (52) in the limit of a large number of antennas becomes

$$\lim_{t,r\to\infty}\lim_{P\to\infty} \Delta = \frac{r}{t}\log_2 \sqrt{e} \tag{54}$$

which proves the lemma.

## Appendix IV
## Proof of (16)

We first differentiate (12) with respect to $P(\hat{\mathbf{H}})$. Using the chain rule, we obtain

$$\frac{\partial C_{\mathrm{lower}}}{\partial P} = \frac{1}{\ln 2}\sum_{i=1}^{k}\left(\frac{1}{\mu}\frac{\partial \mu}{\partial P} - \frac{\sigma_{\mathbf{E}}^2}{1+\sigma_{\mathbf{E}}^2 P}\right) \tag{55}$$

where $k$ is the number of subchannels with nonzero power allocation, i.e., $k = \sum_{i=1}^{t} I(P_i > 0)$, where $I(\cdot)$ is an indicator function. From (11), the water-level $\mu$ can be expressed in terms of $P(\hat{\mathbf{H}})$ and subchannel gains as

$$\mu = \frac{1}{k}\left(P + (1+\sigma_{\mathbf{E}}^2 P)\sum_{i=1}^{k}\frac{1}{\lambda_i}\right) \tag{56}$$

and

$$\frac{\partial \mu}{\partial P} = \frac{1}{k}\left(1 + \sigma_{\mathbf{E}}^2 \sum_{i=1}^{k}\frac{1}{\lambda_i}\right). \tag{57}$$

Defining $\lambda_0^{-1} = \sum_{i=1}^{k}\lambda_i^{-1}$ and substituting (56), (57) into (55) we have

$$\frac{\partial C_{\mathrm{lower}}}{\partial P} = \frac{1}{\ln 2}\sum_{i=1}^{k}\left(\frac{1+\lambda_0^{-1}\sigma_{\mathbf{E}}^2}{P+\lambda_0^{-1}(1+\sigma_{\mathbf{E}}^2 P)} - \frac{\sigma_{\mathbf{E}}^2}{1+\sigma_{\mathbf{E}}^2 P}\right) \tag{58}$$

$$= \frac{1}{\ln 2}\frac{1}{\mu(1+\sigma_{\mathbf{E}}^2 P)} \tag{59}$$

which proves (16).

## Appendix V
## Proof of *Theorem* 1

From (15)–(16), we have

$$\mu\left(1+\sigma_{\mathbf{E}}^2 P\right) = \nu. \tag{60}$$

Substituting $\mu$ with (56) we obtain a quadratic equation of $P$ as

$$\left(P + (1+\sigma_{\mathbf{E}}^2 P)\frac{1}{\lambda_0}\right)\left(1+\sigma_{\mathbf{E}}^2 P\right) = k\nu \tag{61}$$

which has two solutions given as

$$P = \frac{-(\lambda_0 + 2\sigma_{\mathbf{E}}^2) \pm \sqrt{\lambda_0^2 + 4k\nu\lambda_0\sigma_{\mathbf{E}}^2(\lambda_0+\sigma_{\mathbf{E}}^2)}}{2\sigma_{\mathbf{E}}^2(\lambda_0+\sigma_{\mathbf{E}}^2)}. \tag{62}$$

The solution in (17) follows from (62), after applying the nonnegativity of the transmit power and the Kuhn–Tucker condition.

## Acknowledgment

The authors would like to thank the anonymous reviewers for their suggestions and insights, which greatly improved the manuscript.

## References

[1] E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. Telecommun.*, vol. 10, pp. 585–598, Nov. 1999.
[2] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Pers. Commun.*, vol. 6, pp. 311–335, Mar. 1998.
[3] A. Lapidoth and S. Shamai, "Fading channels: How perfect need "perfect side information" be?," *IEEE Trans. Inf. Theory*, vol. 48, pp. 1118–1134, May 2002.
[4] A. Lapidoth and S. Moser, "Capacity bounds via duality with applications to multiple-antenna systems on flat-fading channels," *IEEE Trans. Inf. Theory*, vol. 49, pp. 2426–2467, Oct. 2003.
[5] R. Etkin and D. N. C. Tse, "Degrees of freedom in underspread MIMO fading channels," *IEEE Trans. Inf. Theory*, vol. 52, pp. 1576–1608, Apr. 2006.
[6] T. Marzetta and B. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Trans. Inf. Theory*, vol. 45, pp. 139–157, Jan. 1999.
[7] L. Zheng and D. Tse, "Communication on the Grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. Inf. Theory*, vol. 48, pp. 359–383, Feb. 2002.
[8] B. Hassibi and B. Hochwald, "How much training is needed in multiple-antenna wireless links?," *IEEE Trans. Inf. Theory*, vol. 49, pp. 951–963, Apr. 2003.
[9] J. Baltersee, G. Fock, and H. Meyr, "Achievable rate of MIMO channels with data-aided channel estimation and perfect interleaving," *IEEE J. Sel. Areas Commun.*, vol. 19, pp. 2358–2368, Dec. 2001.
[10] M. Medard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Trans. Inf. Theory*, vol. 46, pp. 933–946, May 2000.
[11] A. Goldsmith and P. Varaiya, "Capacity of fading channels with channel side information," *IEEE Trans. Inf. Theory*, vol. 43, pp. 1986–1992, Nov. 1997.
[12] T. E. Klein and R. G. Gallager, "Power control for the additive white Gaussian noise channel under channel estimation errors," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2001, p. 304.
[13] H. Yang, "A road to future broadband wireless access: MIMO-OFDM-based air interface," *IEEE Commun. Mag.*, vol. 43, pp. 53–60, Jan. 2005.
[14] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*. Cambridge, U.K.: Cambridge University Press, 2003.
[15] G. Caire and S. Shamai, "On the capacity of some channels with channel state information," *IEEE Trans. Inf. Theory*, vol. 45, pp. 2007–2019, Sep. 1999.
[16] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge University Press, 2003.
[17] A. Edelman, "Eigenvalues and Condition Numbers of Random Matrices," PhD, Department of Mathematics, MIT, Cambridge, MA, 1989.
[18] O. Oyman, R. U. Nabar, H. Bolcskei, and A. J. Paulraj, "Tight lower bounds on the ergodic capacity of Rayleigh fading MIMO channels," in *Proc. IEEE GLOBECOM*, vol. 2, Nov. 2002, pp. 1172–1176.
[19] S. Venkatesan, S. Simon, and R. Valenzuela, "Capacity of a Gaussian MIMO channel with nonzero mean," in *Proc. IEEE Veh. Technol. Conf.*, vol. 3, Oct. 2003, pp. 1767–1771.
[20] D. Hösli and A. Lapidoth, "The capacity of a MIMO Ricean channel is monotonic in the singular values of the mean," in *Proc. 5th Int. ITG Conf. Source Channel Coding (SCC)*, Jan. 2004.

[21] S. Bhashyam, A. Sabharwal, and B. Aazhang, "Feedback gain in multiple antenna systems," *IEEE Trans. Commun.*, vol. 50, pp. 785–798, May 2002.
[22] S. V. Hanly and D. N. Tse, "Multiaccess fading channels. Part II: Delay-limited capacities," *IEEE Trans. Inf. Theory*, vol. 44, pp. 2816–2831, Nov. 1998.
[23] T. Yoo and A. Goldsmith, "Capacity of fading MIMO channels with channel estimation error," in *Proc. IEEE Int. Conf. Commun. (ICC)*, vol. 2, Jun. 2004, pp. 808–813.
[24] R. M. Young, "Euler's constant," *Math. Gaz.*, vol. 75, pp. 187–190, 1991.

# On Efficient Balanced Codes Over the $m$th Roots of Unity

Raffaele Mascella, Luca G. Tallini, Sulaiman Al-Bassam, and Bella Bose, *Fellow, IEEE*

***Abstract*—Let $\Phi_m \subseteq \mathbb{C}$ be the set of all $m$th roots of unity, $m \in \mathbb{N}$. A balanced code over $\Phi_m$ is a block code over the alphabet $\Phi_m$ such that each code word is balanced; that is, the complex sum of its components (or weight) is equal to 0. Let $\mathcal{B}_m(n)$ be the set of all balanced words of length $n$ over $\Phi_m$. In this correspondence, it is shown that when $m$ is a prime number, the set $\mathcal{B}_m(n)$ is not empty if, and only if, $m$ divides $n$. In this case, the minimum redundancy for a balanced code over $\Phi_m$ of length $n$ is**

$$\rho(\mathcal{B}_m(n)) = n - \lfloor \log_m |\mathcal{B}_m(n)| \rfloor \approx [(m-1)/2] \log_m(2\pi n) - m/2.$$

**On the other hand, it is shown that when $m = 4$, the set $\mathcal{B}_4(n)$ is not empty if, and only if, $n$ is even, and in this case, the minimum redundancy for a balanced code over $\Phi_4$ of length $n$ is**

$$\rho(\mathcal{B}_4(n)) = n - \lfloor \log_4 |\mathcal{B}_4(n)| \rfloor \approx \log_4 n + 0.326.$$

**Further, this correspondence completely solves the problem of designing efficient coding methods for balanced codes over $\Phi_m$, when $m = 4$. In fact, it reduces the problem of designing efficient coding schemes for balanced codes over $\Phi_4$ to the design of efficient balanced codes over the usual bipolar alphabet $\Phi_2 = \{-1, +1\}$.**

Manuscript received August 30, 2004; revised December 1, 2005. This work was supported by the Italian MIUR under Grant COFIN-2003013538 and the National Science Foundation under Grant CCR-0105204. The material in this correspondence was presented in part at the 2003 IEEE International Symposium on Information Theory, Yokohama, Japan, June/July 2003.

R. Mascella and L. G. Tallini are with the Dipartimento di Scienze della Comunicazione, Università di Teramo, Teramo 64100, Italy (e-mail: rmascella@unite.it, ltallini@unite.it).

S. Al-Bassam is with the ICS Department, KFUPM, Dhahran 31261, Saudi Arabia (e-mail: albassam@hotmail.com).

B. Bose is with the School of Electrical Engineering and Computer Science, Oregon State University, Corvallis, OR 97331 USA (e-mail: bose@eecs.oregonstate.edu).

Communicated by K. A. S. Abdel-Ghaffar, Associate Editor for Coding Theory.

Digital Object Identifier 10.1109/TIT.2006.872981

***Index Terms*—Balanced codes, bipolar alphabet, DC-free communication, digital communication, line codes, $m$-ary alphabet, $m$-ary communication, $m$th roots of unity.**

## I. Introduction

Let $i = \sqrt{-1} \in \mathbb{C}$ and consider the $m$-ary alphabet

$$\Phi_m \stackrel{\text{def}}{=} \{e^{2\pi i h/m} : h = 0, 1, \ldots, m-1\} \subseteq \mathbb{C}$$

given by the set of all $m$th roots of unity. For example, $\Phi_4 = \{+1, +i, -1, -i\}$.

Given $n \in \mathbb{N}$, the word $Z = z_1 z_2 \ldots z_n$ of length $n$ over the alphabet $\Phi_m$ is called balanced (over the $m$th roots of unity, or over $\Phi_m$) if, and only if, the weight of $Z$

$$w(Z) \stackrel{\text{def}}{=} \sum_{j=1}^{n} z_j = 0$$

where the sum is over the complex field. For example, the word

$$Z = (-i, +1, +i, +1, -1, -i, -1, +i)$$

($n = 8$ and $m = 4$) is balanced over $\Phi_4$ because

$$w(X) = -i + 1 + i + 1 - 1 - i - 1 + i = 0.$$

Let $\mathcal{B}_m(n)$ denote the set of all balanced words over $\Phi_m$. Any subset of $\mathcal{B}_m(n)$ is called a balanced code of length $n$ over $\Phi_m$. In particular, a code $\mathcal{C}$ is called **balanced code over $\Phi_m$ with $k$ information digits** and $r$ check digits (or briefly, DC$(\Phi_m, k + r, k)$) if, and only if

1) every codeword of $\mathcal{C}$ has length $n = k + r$;
2) $\mathcal{C}$ is a subset of $\mathcal{B}_m(n)$;
3) the code $\mathcal{C}$ contains $|\Phi_m|^k = m^k$ codewords.

For example, when $m = 4$ (see the code $\mathcal{C}$ at the bottom of the page) is a balanced code over $\Phi_4$ with $k = 2$ information digits and $r = 2$ **check digits** (or briefly, DC$(\Phi_4, 4, 2)$ code).

Note that, in the binary case ($m = 2$), the above balanced code definition coincides with the usual binary balanced codes over the bipolar alphabet $\Phi_2 = \{-1, +1\}$ [2], [5], [9]. The above generalization of balanced codes to the $m$-ary alphabet $\Phi_m$ given above, was recently considered in [3] to achieve DC-free communication in communication systems where the signal set (i.e., the communication alphabet) is given by the set of the $m$th roots of unity. Also in this case the problem is to find a $\mathrm{DC}(\Phi_m, k + r, k)$ code $\mathcal{C}$, and a one-to-one function (encoding function)

$$\mathcal{E} : \mathbb{Z}_m^k \to \mathcal{C} \subseteq \Phi_m^{k+r}$$

which, together with its inverse (decoding function), is very easy to compute. It is also desirable that the redundancy, $r$, of $\mathcal{C}$ be as small as possible.

In [3] the authors propose to use a generalization of Knuth's balancing method [7] and analyze some simple cases. In this correspondence we show some combinatorial properties of the set $\mathcal{B}_m(n)$ and

$$\mathcal{C} = \left\{ \begin{array}{llll} (+1,+1,-1,-1), & (+i,+1,-i,-1), & (-1,+1,+1,-1), & (-i,+1,+i,-1), \\ (+1,+i,-1,-i), & (+i,+i,-i,-i), & (-1,+i,+1,-i), & (-i,+i,+i,-i), \\ (+1,-1,-1,+1), & (+i,-1,-i,+1), & (-1,-1,+1,+1), & (-i,-1,+i,+1), \\ (+1,-i,-1,+i), & (+i,-i,-i,+i), & (-1,-i,+1,+i), & (-i,-i,+i,+i) \end{array} \right\}$$