# GMI and Mismatched-CSI Outage Exponents in MIMO Block-Fading Channels

A. Taufiq Asyhari
National Chiao Tung University
Hsinchu, Taiwan, R.O.C.
taufiq-a@ieee.org

Albert Guillén i Fàbregas
ICREA & Universitat Pompeu Fabra
University of Cambridge
guillen@ieee.org

***Abstract*—We study transmission over multiple-antenna block-fading channels with imperfect channel state information at both the transmitter and receiver. Specifically, we investigate achievable rates based on the generalized mutual information. We then analyze the corresponding outage probability in the high signal-to-noise ratio regime.**

## I. Introduction

The block-fading channel is a widely-used model to represent transmission over slowly-varying fading scenarios. The channel is non-ergodic and the reliability of transmission is limited by the information outage probability [1]. The majority of works on block-fading channels studied the outage probability assuming perfect channel state information (CSI) (see, e.g., [2], [3]). This idealistic assumption seems to be too optimistic in practical scenarios as the CSI is usually imperfect due to imperfections in estimating the channel.

This paper studies imperfect CSI in block-fading channels. Specifically, each communicating terminal is assumed to acquire a noisy version of the actual CSI. The imperfect CSI at the transmitter (CSIT) is used to adaptively allocate power across different fading blocks (subject to a long-term power constraint) with the aim to improve the performance. At the receiving end, the receiver perceives the noisy CSI as if it was noiseless and uses a nearest neighbor decoder. Because of the noisy CSI at the receiver (CSIR), the decoder is mismatched. With this setup, we investigate the generalized mutual information (GMI) [4]—which is an achievable rate under mismatched decoding—to examine the effect of imperfect CSI on the reliability of transmission.

This study was initiated in [5] for a single antenna setup. In this work, we extend the results of [5] to multiple-input multiple-output (MIMO) channels. and provide upper and lower bounds on the GMI of a MIMO block-fading channel. These bounds are then used to find the outage exponent and to determine diversity-achieving input distributions. (The outage exponent is defined as the high signal-to-noise ratio (SNR) slope of the outage probability curve plotted in a logarithmic-logarithmic scale against the SNR.)

The work of A.T. Asyhari has been supported in part by the Yousef Jameel Scholarship at University of Cambridge and in part by the National Science Council (NSC) under Grant NSC 102-2218-E-009-001. The work by A. Guillén i Fàbregas has been supported in part by the European Research Council under ERC grant agreement 259663 and by the Spanish Ministry of Economy and Competitiveness under grant TEC2012-38800-C03-03.

## II. System Model

Consider a MIMO block-fading channel with $n_{\rm t}$ transmit antennas, $n_{\rm r}$ receive antennas and $B$ fading blocks. The output of the channel for block $b$, $b = 1,\dots,B$, is an $n_{\rm r} \times J$-dimensional random matrix

$$\mathbb{Y}_b = \mathbb{H}_b \mathsf{P}_b^{\frac{1}{2}} \mathsf{X}_b + \mathbb{Z}_b \tag{1}$$

where $\mathbb{Z}_b$ is the $n_{\rm r} \times J$-dimensional noise matrix and $\mathsf{X}_b \in \mathcal{X}^{n_{\rm t}\times J}$ is the transmitted signal matrix; $\mathsf{P}_b \in \mathbb{R}^{n_{\rm t}}$, $J$ and $\mathcal{X}$ denote the transmission power matrix, the block length and the constellation set, respectively. The entries of $\mathbb{Z}_b$ are assumed to be independent and identically distributed (i.i.d.) complex-Gaussian random variables with zero mean and unit variance. The $(n_{\rm r} \times n_{\rm t})$-dimensional matrix $\mathbb{H}_b$ is the $b$-th fading matrix and is assumed to be i.i.d. across $b = 1,\dots,B$. We further assume that the entries of $\mathbb{H}_b$ are i.i.d. zero-mean unit-variance complex-Gaussian random variables.

The codeword for message $m \in \{1,\dots,2^{BJR}\}$ is denoted by $\mathsf{X}(m) = [\mathsf{X}_1(m),\dots,\mathsf{X}_B(m)]$ where $R$ is the data rate. The entries of $\mathsf{X}$ are drawn i.i.d. from a probability distribution over $\mathcal{X}^{n_{\rm t}}$. We assume that $R$ is a fixed positive number implying zero multiplexing gain [2]. We further assume that codewords are normalized such that $\frac{1}{BJ}\mathbb{E}[\|\mathbb{X}\|_F^2] = n_{\rm t}$ where $\|\cdot\|_F$ denotes the Frobenius norm.

We assume an additive-noise imperfect CSI model, i.e.,

$$\text{CSIT} \quad \tilde{\mathbb{H}}_b = \mathbb{H}_b + \tilde{\mathbb{E}}_b \tag{2}$$

$$\text{CSIR} \quad \hat{\mathbb{H}}_b = \mathbb{H}_b + \hat{\mathbb{E}}_b, \tag{3}$$

where error estimation matrices $\tilde{\mathbb{E}}_b$ and $\hat{\mathbb{E}}_b$ are independent. The entries of $\tilde{\mathbb{E}}_b$ and $\hat{\mathbb{E}}_b$ are assumed to be independent from $\mathbb{H}_b$ and i.i.d. complex-Gaussian random variables with zero mean and variances $\tilde{\varepsilon}^2 = P^{-\tilde{d}_e}$ and $\hat{\varepsilon}^2 = P^{-\hat{d}_e}$, respectively, where $P$ is the average SNR. The positive parameters $\tilde{d}_e$ and $\hat{d}_e$ are incorporated to denote the CSIT-error and the CSIR-error diversities, respectively. This model corresponds to a system that exploits channel reciprocity [6], for which the fading realization stays constant within a block and fading estimation is possible at both communicating ends at approximately the same time. Furthermore, it captures widely-used pilot-aided channel estimators with error variance inversely proportional to the pilot SNR [7].

For a given CSIT matrix $\tilde{\mathsf{H}} \triangleq [\tilde{\mathsf{H}}_1,\dots,\tilde{\mathsf{H}}_B]$, the power matrix $\mathsf{P}_b$ is assumed to be the scaling of the $n_\mathrm{t} \times n_\mathrm{t}$ identity matrix $\mathsf{I}_{n_\mathrm{t}}$

$$\mathsf{P}_b(\tilde{\mathsf{H}}) = \frac{P_b(\tilde{\mathsf{H}})}{n_\mathrm{t}} \mathsf{I}_{n_\mathrm{t}}. \tag{4}$$

We further denote the power allocation matrix as $\mathsf{P} \triangleq [\mathsf{P}_1, \cdots, \mathsf{P}_B]$ and the CSIR matrix as $\hat{\mathsf{H}} \triangleq [\hat{\mathsf{H}}_1, \ldots, \hat{\mathsf{H}}_B]$. We further consider an average power constraint

$$\mathbb{E}\left[\frac{1}{B}\sum_{b=1}^{B} \mathrm{tr}\Big(\mathsf{P}_b(\tilde{\mathbb{H}})\Big)\right] \le P \tag{5}$$

where $\mathrm{tr}(\cdot)$ denotes the trace operator. For each block, power adaptation uses noisy fading information for all blocks $b = 1, \ldots, B$ to adapt the scalar power coefficient $P_b$ in order to improve performance. This model is relevant for multi-carrier transmission, where the channel is estimated in the time domain, but transmission is in the frequency domain.

By treating the imperfect CSIR as if it was perfect, the receiver infers the transmitted message using a nearest neighbor rule, which outputs the message $\hat{m}$ such that

$$\hat{m} = \underset{m \in \{1, \ldots, 2^{BJR}\}}{\arg\max} \; Q(\mathsf{X}(m), \mathsf{Y}, \hat{\mathsf{H}}, \mathsf{P}) \tag{6}$$

where

$$Q(\mathsf{X}, \mathsf{Y}, \hat{\mathsf{H}}, \mathsf{P}) \propto \exp\left(-\sum_{b=1}^{B} \left\| \mathsf{Y}_b - \hat{\mathsf{H}}_b \mathsf{P}_b^{\frac{1}{2}} \mathsf{X}_b \right\|_F^2\right). \tag{7}$$

## III. GMI of the MIMO Block-Fading Channel

For a given fading matrix $\mathsf{H}$, CSIR estimate $\hat{\mathsf{H}}$ and power allocation matrix $\mathsf{P}$, the average error probability for the ensemble of i.i.d. random codes of rate $R$ can be bounded [8] as

$$\bar{P}_e(\mathsf{H}, \hat{\mathsf{H}}, \mathsf{P}) \le 2^{-BJE_r^Q(R, \mathsf{H}, \hat{\mathsf{H}}, \mathsf{P})} \tag{8}$$

where

$$E_r^Q(R, \mathsf{H}, \hat{\mathsf{H}}, \mathsf{P}) = \sup_{\substack{s>0 \\ 0 \le \rho \le 1}} \frac{1}{B}\sum_{b=1}^{B} E_0^Q(s, \rho, \mathsf{H}_b, \hat{\mathsf{H}}_b, \mathsf{P}_b) - \rho R \tag{9}$$

is the mismatched error exponent and where

$$E_0^Q(s, \rho, \mathsf{H}_b, \hat{\mathsf{H}}_b, \mathsf{P}_b) = -\log_2 \mathbb{E}\left[\left(\mathbb{E}\left[\left(\frac{Q(\boldsymbol{X}', \boldsymbol{Y}, \hat{\mathsf{H}}_b, \mathsf{P}_b)}{Q(\boldsymbol{X}, \boldsymbol{Y}, \hat{\mathsf{H}}_b, \mathsf{P}_b)}\right)^s \middle| \boldsymbol{X}, \boldsymbol{Y}, \mathsf{H}_b, \hat{\mathsf{H}}_b, \mathsf{P}_b\right]\right)^\rho \middle| \mathsf{H}_b, \hat{\mathsf{H}}_b, \mathsf{P}_b\right] \tag{10}$$

denotes the generalized Gallager function for a given $\mathsf{H}_b, \hat{\mathsf{H}}_b$ and $\mathsf{P}_b$ [8]. Since $E_0^Q(s, \rho, \mathsf{H}_b, \hat{\mathsf{H}}_b, \mathsf{P}_b)$ is concave in $\rho$, $\rho \in [0, 1]$, the largest slope of the Gallager function occurs at $\rho = 0$. The GMI is the resulting rate [8]

$$I^{\mathrm{gmi}}(\mathsf{H}, \hat{\mathsf{H}}, \mathsf{P}) = \sup_{s>0} \frac{1}{B}\sum_{b=1}^{B} I_b^{\mathrm{gmi}}(\mathsf{P}_b, \mathsf{H}_b, \hat{\mathsf{H}}_b, s) \tag{11}$$

where

$$I_b^{\mathrm{gmi}}(\mathsf{P}_b, \mathsf{H}_b, \hat{\mathsf{H}}_b, s) = \mathbb{E}\left[\log_2 \frac{Q^s\Big(\boldsymbol{X}, \boldsymbol{Y}, \hat{\mathsf{H}}_b, \mathsf{P}_b\Big)}{\mathbb{E}\left[Q^s\Big(\boldsymbol{X}', \boldsymbol{Y}, \hat{\mathsf{H}}_b, \mathsf{P}_b\Big) \middle| \boldsymbol{Y}, \mathsf{H}_b, \hat{\mathsf{H}}_b, \mathsf{P}_b\right]} \middle| \mathsf{H}_b, \hat{\mathsf{H}}_b, \mathsf{P}_b\right]. \tag{12}$$

Convexity analysis of $E_0^Q(s, \rho, \mathsf{H}_b, \hat{\mathsf{H}}_b, \mathsf{P}_b)$ reveals that the exponent $E_r^Q(R, \mathsf{H}, \hat{\mathsf{H}}, \mathsf{P})$ is only positive whenever $R \le I^{\mathrm{gmi}}(\mathsf{H}, \hat{\mathsf{H}}, \mathsf{P}) - \epsilon$. Then, for rates below $I^{\mathrm{gmi}}(\mathsf{H}, \hat{\mathsf{H}}, \mathsf{P})$, $\bar{P}_e(\mathsf{H}, \hat{\mathsf{H}}, \mathsf{P})$ can be made arbitrarily small by increasing the block length $J$. The ensemble average implies that there exists a code with rate $R$ such that its error probability vanishes with $J$ as long as $R \le I^{\mathrm{gmi}}(\mathsf{H}, \hat{\mathsf{H}}, \mathsf{P}) - \epsilon$, proving the achievability of $I^{\mathrm{gmi}}(\mathsf{H}, \hat{\mathsf{H}}, \mathsf{P})$. When $\hat{\mathsf{H}} = \mathsf{H}$, the GMI gives the mutual information of the perfect CSIR case, as $s = 1$ is optimal.

For a given input distribution, expressing the right-hand side (RHS) of (11) may in general be difficult due to the optimization over $s > 0$ across $B$ fading blocks. Therefore, non-trivial upper and lower bounds are relevant in the analysis. A GMI upper bound can be obtained by exchanging the supremum and the average on the RHS of (11), i.e.,

$$I^{\mathrm{gmi}}(\mathsf{H}, \hat{\mathsf{H}}, \mathsf{P}) \le \frac{1}{B}\sum_{b=1}^{B} \sup_{s>0} I_b^{\mathrm{gmi}}(\mathsf{P}_b, \mathsf{H}_b, \hat{\mathsf{H}}_b, s). \tag{13}$$

This upper bound leads to the exact GMI in a number of cases, e.g., when the optimizing $s$ on the RHS of (13) does not depend on $\mathsf{P}_b$, $\mathsf{H}_b$ and $\hat{\mathsf{H}}_b$ or when $B = 1$ (quasi-static channel). Also, for single-input single-output (SISO) channels with Gaussian inputs, the optimal value of $s$ can be found analytically. A GMI lower bound can be obtained by choosing a particular $s$. As shown in [9, App. D], a good choice is

$$\hat{s} = \frac{B}{Bn_\mathrm{r} + \sum_{b=1}^{B} \left\| \hat{\mathsf{E}}_b \mathsf{P}_b^{\frac{1}{2}} \right\|_F^2}. \tag{14}$$

Using the non-negativity of the GMI [9, Prop. 2], we obtain

$$I^{\mathrm{gmi}}(\mathsf{H}, \hat{\mathsf{H}}, \mathsf{P}) \ge \max\left[0, \frac{1}{B}\sum_{b=1}^{B} I_b^{\mathrm{gmi}}(\mathsf{P}_b, \mathsf{H}_b, \hat{\mathsf{H}}_b, \hat{s})\right]. \tag{15}$$

The RHS of (15) is an achievable rate as all rates below $I^{\mathrm{gmi}}(\mathsf{H}, \hat{\mathsf{H}}, \mathsf{P})$ are achievable.

## IV. Outage and the High-SNR Behavior

The average error probability of long codes when the fading varies from codeword to codeword is obtained by averaging (8) with respect to $\mathsf{H}, \hat{\mathsf{H}}, \mathsf{P}$ in the limit as $J$ tends to infinity,

$$\bar{P}_e \le \lim_{J\to\infty} \mathbb{E}\left[2^{-BJE_r^Q(R, \mathbb{H}, \hat{\mathbb{H}}, \mathsf{P})}\right] \tag{16}$$

$$= \inf_{\epsilon>0} \Pr\left\{I^{\mathrm{gmi}}(\mathbb{H}, \hat{\mathbb{H}}, \mathsf{P}) - \epsilon < R\right\} \tag{17}$$

$$= \Pr\left\{I^{\mathrm{gmi}}(\mathbb{H}, \hat{\mathbb{H}}, \mathsf{P}) < R\right\} \triangleq P_{\mathrm{gout}}(R) \tag{18}$$

where (18) defines the generalized outage probability.

An important outage performance indicator is the outage exponent, defined as

$$d \triangleq \lim_{P\to\infty} -\frac{\log P_{\rm gout}(R)}{\log P}. \tag{19}$$

In words, the $d$ is the high-SNR slope of $P_{\rm gout}(R)$ plotted in a logarithmic-logarithmic scale against the SNR.

The following result gives the outage exponent of a system that optimizes the power to minimize $P_{\rm gout}(R)$ subject to the power constraint (5). The proof uses the bounds (13) and (15) and interestingly, it does not require an explicit characterization of the minimum-outage power allocation scheme.

*Theorem 1:* Consider transmission over the MIMO channel (1) with CSIT and CSIR error diversities $\tilde{d}_e$ and $\hat{d}_e$ as described in Section II. The outage exponent $d_{\rm icsi}$ for both Gaussian and discrete constellations of size $2^M$ is

$$d_{\rm icsi} = \begin{cases} d^{\rm u}_{\rm csir}\hat{d}_e, & \text{if } \hat{d}_e \le 1 + d^{\rm u}_{\rm csir}\tilde{d}_e \\ d^{\rm u}_{\rm csir}\left(1 + d^{\rm u}_{\rm csir}\tilde{d}_e\right), & \text{if } \hat{d}_e > 1 + d^{\rm u}_{\rm csir}\tilde{d}_e \end{cases} \tag{20}$$

where $d^{\rm u}_{\rm csir}$ is the perfect-CSIR outage exponent with uniform power allocation, given by

$$d^{\rm u}_{\rm csir} = \begin{cases} B n_{\rm t} n_{\rm r}, & \text{for Gaussian inputs} \\ n_{\rm r}\left(1 + \left\lfloor B\left(n_{\rm t} - \frac{R}{M}\right)\right\rfloor\right), & \text{for discrete inputs.} \end{cases} \tag{21}$$

*Proof:* (Sketch) The proof for the MIMO case follows the same general lines as that for the SISO case [5, App. A], which involves analyzing the asymptotic behavior of optimal power allocation, evaluating the asymptotic $P_{\rm gout}(R)$ and characterizing the asymptotic outage set, and evaluating the exponent using Varadhan's lemma [10].

There are three key steps to extend the proof to the MIMO case. The first step obtains the asymptotic power allocation by evaluating the expectation (5) using the joint density of the entries of $\tilde{\mathbb{H}}$, which are independent by assumption. The second one is to obtain an asymptotic expression $P_{\rm gout}(R)$ using the joint density of the entries of $(\mathbb{H}, \hat{\mathbb{E}} = \hat{\mathbb{H}} - \mathbb{H}, \tilde{\mathbb{H}})$. The third one is to characterize the asymptotic outage set using the GMI bounds (13) and (15). ■

The result captures the following important observations.

1) The case $\hat{d}_e \le 1$ corresponds to the CSIR being too unreliable. Any available CSIT cannot improve the outage exponent via power adaptation. The poor quality of the CSIR makes power adaptation as good as uniform power allocation in terms of outage exponent (see [9] for the result with uniform power allocation).
2) The case $\hat{d}_e > 1$ corresponds to the CSIR allowing for power adaptation gains.
    a) Case $\hat{d}_e \le 1 + d^{\rm u}_{\rm csir}\tilde{d}_e$: Though power adaptation improves the performance, the outage exponent is limited by the quality of the CSIR. The outage exponent improves as the quality of CSIR improves.
    b) Case $\hat{d}_e > 1 + d^{\rm u}_{\rm csir}\tilde{d}_e$: Full performance improvement is achieved with power adaptation as the achievable outage exponent is identical to the perfect-CSIR outage exponent in [11], [12].

The outage result in Theorem 1 corresponds to a system where the transmitter estimates the channel independently of the receiver. This is a typical model for two-way training for which both the transmitter and receiver transmit pilot symbols. One might also consider a CSIT model for which the transmitter obtains a noisy version of the CSIR via a dedicated feedback channel. This is commonly referred to as the mean-feedback model [13]. For this model, we have for block $b$, transmit antenna $t$ and receive antenna $r$ that

$$\text{CSIR} \quad \hat{H}_{b,r,t} = H_{b,r,t} + \hat{E}_{b,r,t}, \tag{22}$$

$$\text{CSIT} \quad \tilde{H}_{b,r,t} = \hat{H}_{b,r,t} + \tilde{E}_{b,r,t}. \tag{23}$$

The CSIT can then be rewritten as

$$\tilde{H}_{b,r,t} = H_{b,r,t} + \hat{E}_{b,r,t} + \tilde{E}_{b,r,t}. \tag{24}$$

The CSIT noise has zero mean and variance $P^{-\tilde{d}_e} + P^{-\hat{d}_e}$. The CSIT-error diversity is then obtained from the exponent of $P^{-\tilde{d}_e} + P^{-\hat{d}_e}$, which is given by $\min(\tilde{d}_e, \hat{d}_e)$. Thus, for the mean-feedback model, the outage exponent can be obtained by replacing $\tilde{d}_e$ with $\min(\tilde{d}_e, \hat{d}_e)$ in Theorem 1.

## V. Discussion

### A. Connections with Previous Works

The technique used to derive the outage exponent is based on the GMI, which is the largest achievable rate for i.i.d. generated codebooks [4], [14], [15]. Therefore, the outage result in Theorem 1 is the optimal diversity for i.i.d. codebooks (Gaussian or discrete) and a nearest neighbor decoder. An improved achievable rate (LM rate) can be obtained with codewords satisfying a good cost constraint [4], [14].

Several works in the literature studied a similar problem, but used different information rates. In particular, we refer to the works in [16]–[18] for comparison on the validity as achievable rates. For simplicity and for the sake of comparison, we consider a SISO quasi-static channel ($B = 1$). References [16]–[18] assumed Gaussian inputs and linear minimum mean-squared error (LMMSE) channel estimation at the receiver, where the estimate $\hat{H}$ is related to the actual fading $H$ as

$$H = \hat{H} + \hat{E} \tag{25}$$

where $\hat{E}$ is the scalar fading estimation error having zero mean and variance $P^{-\hat{d}_e}$. Thus, from (1) and (25) we can write the input-output relationship as

$$\boldsymbol{Y} = \sqrt{P}\hat{H}\boldsymbol{x} + \sqrt{P}(H - \hat{H})\boldsymbol{x} + \boldsymbol{Z} \tag{26}$$

where $\boldsymbol{Y}$ and $\boldsymbol{Z}$ are the random received and noise vectors, respectively, which take values on $\mathbb{C}^J$; $\boldsymbol{x}$ is the $J$-dimensional input vector; $P$ is the transmission power. Note that since every realization of $\hat{H}$ is known at the receiver, the argument in [16]–[18] is that one can treat the term $\sqrt{P}(H - \hat{H})\boldsymbol{x}$ as an additional noise term. It was further argued in [16], [18] that by modeling the signal-dependent noise

$$\boldsymbol{Z}' = \sqrt{P}(H - \hat{H})\boldsymbol{x} + \boldsymbol{Z} \tag{27}$$

as a zero-mean Gaussian noise with i.i.d. entries independent of $\boldsymbol{x}$ and each having variance $1+P|H-\hat{H}|^2$ one can obtain a rate that is claimed to be a lower bound to the instantaneous mutual information as [18]

$$\underline{I}(H,\hat{H},P)=\log_2\left(1+\frac{P|\hat{H}|^2}{1+P|H-\hat{H}|^2}\right). \tag{28}$$

Note that the above expression leads to an outage exponent that is obtained by solving

$$\Pr\left\{\underline{I}(H,\hat{H},P)<R\right\} = \Pr\left\{\log_2\left(1+\frac{P|\hat{H}|^2}{1+P|H-\hat{H}|^2}\right)<R\right\}. \tag{29}$$

Interestingly, using the lower bound in (15) and following the steps used in [9, App. D] for $B=1$, the GMI can be lower-bounded by

$$I^{\mathrm{gmi}}(H,\hat{H},P)\geq\log_2\left(1+\frac{P|\hat{H}|^2}{1+P|H-\hat{H}|^2}\right)-\frac{1}{\log 2}. \tag{30}$$

In the high-SNR regime, the constant difference between (28) and the RHS of (30) does not affect the outage exponent. Thus, it is not surprising that for the case under consideration, our results are identical to the results in [16], [18]. Remark that although the model (25) differs to (3) in terms of which variables correlate to each other, it can be shown that such a difference does not affect the outage exponent as long as we perform a proper chain rule on the joint density of $H$ and $\hat{E}$.

Rate (28) seems to be easier to evaluate than the GMI. However, there are some technical problems associated with the derivation of (28), which we explain in the following.

- To the best of our knowledge, there is no explicit proof on the achievability of $\underline{I}(H,\hat{H},P)$ for a given $H,\hat{H}$. The argument to derive (28) follows from [19], where LMMSE channel estimation is used at the receiver to derive a lower bound to the blockwise-ergodic capacity. In this blockwise-ergodic setup, the block length $J$ is finite, and the capacity expression is obtained via coding over infinitely many blocks, where the estimate $\hat{H}$ and the error $(H-\hat{H})$ have uncorrelated statistics over these many blocks. A lower bound to the blockwise-ergodic capacity can then be obtained using the steps in [20, Sec. III] via averaging over all states of fading and its corresponding estimate.
  It is not clear whether the technique in [19] can directly be applied to non-ergodic fading channels. As supposed to coding over infinitely many blocks, in a quasi-static channel, coding is performed for only one block and the block length $J$ is taken to infinity to recover the information outage probability [2], [3]. Note that during a single block, both fading $H$ and fading estimate $\hat{H}$ are constant. Hence, rate (28) may not be an accurate lower bound to the instantaneous mutual information for the block of interest as both $H$ and $\hat{H}$ (and thus $(H-\hat{H})$) are constant within a single block alone. It therefore follows that there is no guarantee that transmitting codeword at rate $R=\underline{I}(H,\hat{H},P)-\epsilon$ for any $\epsilon>0$ has a vanishing error probability as the block length $J$ tends to infinity. This is in contrast with $I^{\mathrm{gmi}}(H,\hat{H},P)$ for which the achievability has been proven in [8].
- For some $H$ and $\hat{H}$, we may find $\underline{I}(H,\hat{H},P)$ that is larger than the perfect-CSIR mutual information
  $$I(H,P)=\log_2\left(1+P|H|^2\right). \tag{31}$$
  We illustrate this in Fig. 1 where we assume power $P=1$ and fading realization $H=1$, and we use estimation (25) to compute the density of $I^{\mathrm{gmi}}(H,\hat{H},P)$ and $\underline{I}(H,\hat{H},P)$. For a given $H=1$, the probability that $\underline{I}(H,\hat{H},P)$ is greater than $I(H,P)$ is non zero, which implies that the lower bound (28) violates the data-processing inequality. This result indirectly disproves the achievability of $\underline{I}(H,\hat{H},P)$, in contrast to $I^{\mathrm{gmi}}(H,\hat{H},P)$, which is always smaller than $I(H,P)$ as shown in [8].
- It is not clear whether modeling $\boldsymbol{Z}'$ in (27) as a signal-independent Gaussian noise would still result in the correct exponent for discrete inputs.

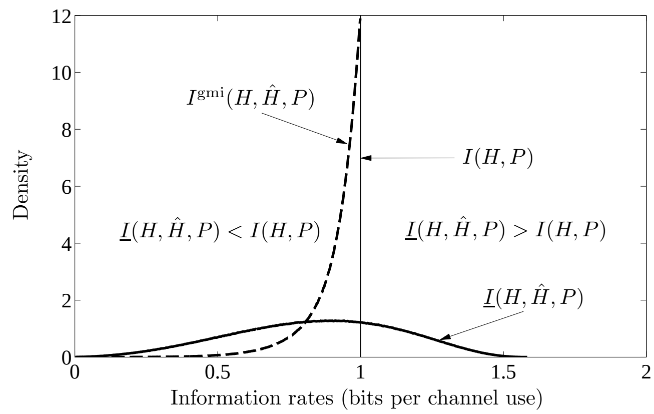


Fig. 1. Comparison of the densities of the GMI and the lower bound (28) with fading realization $H=1$, transmission power $P=1$ (unit power) and CSIR-error variance $\hat{\varepsilon}^2=0.1$.

Based on the preceding comparisons, we observe that rate characterization in [16]–[18] may fail to guarantee achievable outage performance. On the other hand, mismatched decoding approach via GMI will always provide an accurate characterization on achievable outage performance.

### *B. Diversity-Achieving Input Distributions*

Gaussian inputs are no longer optimal for the channel (1) when CSIR is imperfect. We can show using (12) that the outage exponent for Gaussian inputs is a lower bound to the outage exponent for some input distributions satisfying certain conditions. We first assume that the input vector is i.i.d. over all transmit antennas and all channel uses and is such that $\mathbb{E}[|X|^2]=1$. The expression in (12) (in natural-base logarithm) can be decomposed into two terms as follows

$$I_b^{\rm gmi}(\mathsf{P}_b,\mathsf{H}_b,\hat{\mathsf{H}}_b,s) = \mathbb{E}\left[\log Q^s\left(\boldsymbol{X},\boldsymbol{Y},\hat{\mathsf{H}}_b,\mathsf{P}_b\right)\Big|\,\mathsf{H}_b,\hat{\mathsf{H}}_b,\mathsf{P}_b\right] - \mathbb{E}\left[\log\mathbb{E}\left[Q^s\left(\boldsymbol{X}',\boldsymbol{Y},\hat{\mathsf{H}}_b,\mathsf{P}_b\right)\Big|\,\boldsymbol{Y},\mathsf{H}_b,\hat{\mathsf{H}}_b,\mathsf{P}_b\right]\Big|\,\mathsf{H}_b,\hat{\mathsf{H}}_b,\mathsf{P}_b\right]. \tag{32}$$

Evaluating the first term of (32) yields

$$\begin{aligned}
&\mathbb{E}\left[\log Q^s\left(\boldsymbol{X},\boldsymbol{Y},\hat{\mathsf{H}}_b,\mathsf{P}_b\right)\Big|\,\mathsf{H}_b,\hat{\mathsf{H}}_b,\mathsf{P}_b\right]\\
&= -s\left(n_{\rm r} + \mathbb{E}\left[\left\|\hat{\mathsf{E}}_b\mathsf{P}_b^{\frac{1}{2}}\boldsymbol{X}\right\|^2\Big|\,\mathsf{H}_b,\hat{\mathsf{H}}_b,\mathsf{P}_b\right]\right) &(33)\\
&\geq -s\left(n_{\rm r} + \mathbb{E}\left[\left\|\hat{\mathsf{E}}_b\mathsf{P}_b^{\frac{1}{2}}\right\|_F^2\|\boldsymbol{X}\|^2\Big|\,\mathsf{H}_b,\hat{\mathsf{H}}_b,\mathsf{P}_b\right]\right) &(34)\\
&= -s\left(n_{\rm r} + n_{\rm t}\left\|\hat{\mathsf{E}}_b\mathsf{P}_b^{\frac{1}{2}}\right\|_F^2\right) &(35)
\end{aligned}$$

where we have denoted $\|\cdot\|$ as the Euclidean norm, and where inequality (34) is due to the property $\|\mathsf{AB}\|_F^2 \leq \|\mathsf{A}\|_F^2\cdot\|\mathsf{B}\|_F^2$ [21, Sec. 5.6]. The first expectation in the second term of (32) can be evaluated as follows

$$\mathbb{E}\left[Q^s\left(\boldsymbol{X}',\boldsymbol{Y},\hat{\mathsf{H}}_b,\mathsf{P}_b\right)\Big|\,\boldsymbol{Y},\mathsf{H}_b,\hat{\mathsf{H}}_b,\mathsf{P}_b\right] = \int_{\boldsymbol{x}'} P_{\boldsymbol{X}}(\boldsymbol{x}')e^{-s\left\|\boldsymbol{Y}-\hat{\mathsf{H}}_b\mathsf{P}_b^{1/2}\boldsymbol{x}'\right\|^2}d\boldsymbol{x}'. \tag{36}$$

Then, if the input density can be bounded as

$$P_{\boldsymbol{X}}(\boldsymbol{x}) \leq \frac{G}{\pi^{n_{\rm t}}}e^{-\|\boldsymbol{x}\|^2}, \qquad \boldsymbol{x}\in\mathbb{C}^{n_{\rm t}} \tag{37}$$

for some constant $G > 0$, independent of the SNR, then the above expectation can be bounded as

$$\begin{aligned}
&\int_{\boldsymbol{x}'} P_{\boldsymbol{X}}(\boldsymbol{x}')e^{-s\left\|\boldsymbol{Y}-\hat{\mathsf{H}}_b\mathsf{P}_b^{1/2}\boldsymbol{x}'\right\|^2}d\boldsymbol{x}'\\
&\leq G\int_{\boldsymbol{x}'}\frac{1}{\pi^{n_{\rm t}}}e^{-\|\boldsymbol{x}'\|^2}e^{-s\left\|\boldsymbol{Y}-\hat{\mathsf{H}}_b\mathsf{P}_b^{1/2}\boldsymbol{x}'\right\|^2}d\boldsymbol{x}' &(38)\\
&= \frac{G\cdot\exp\left[-s\boldsymbol{Y}^\dagger\left(\mathsf{I}_{n_{\rm r}}+s\hat{\mathsf{H}}_b\mathsf{P}_b\hat{\mathsf{H}}_b^\dagger\right)^{-1}\boldsymbol{Y}\right]}{\det\left(\mathsf{I}_{n_{\rm r}}+s\hat{\mathsf{H}}_b\mathsf{P}_b\hat{\mathsf{H}}_b^\dagger\right)}. &(39)
\end{aligned}$$

With $s > 0$, $I_b^{\rm gmi}(\mathsf{P}_b,\mathsf{H}_b,\hat{\mathsf{H}}_b,s)$ can then be lower-bounded as

$$\begin{aligned}
&I_b^{\rm gmi}(\mathsf{P}_b,\mathsf{H}_b,\hat{\mathsf{H}}_b,s)\\
&\geq \log\det\left(\mathsf{I}_{n_{\rm r}}+s\hat{\mathsf{H}}_b\mathsf{P}_b\hat{\mathsf{H}}_b^\dagger\right) - s\left(n_{\rm r}+n_{\rm t}\left\|\hat{\mathsf{E}}_b\mathsf{P}_b^{\frac{1}{2}}\right\|_F^2\right)\\
&\quad -\log G + \mathbb{E}\left[s\boldsymbol{Y}^\dagger\left(\mathsf{I}_{n_{\rm r}}+s\hat{\mathsf{H}}_b\mathsf{P}_b\hat{\mathsf{H}}_b^\dagger\right)^{-1}\boldsymbol{Y}\right]. &(40)
\end{aligned}$$

The RHS of (40) is similar to $I_b^{\rm gmi}(\mathsf{P}_b,\mathsf{H}_b,\hat{\mathsf{H}}_b,s)$ for Gaussian inputs, except for the extra terms $-\log G$ and $n_{\rm t}$ in $n_{\rm t}\|\hat{\mathsf{E}}_b\mathsf{P}_b^{\frac{1}{2}}\|_F^2$ ($n_{\rm t}$ is replaced by 1 for Gaussian inputs). However, since those terms do not depend on the SNR, they do not affect the outage exponent. Then, noting that the outage exponent for Gaussian inputs derived using GMI upper and lower bounds are identical (as given in Theorem 1), it follows that for any input distribution meeting the condition (37), the outage exponent is lower-bounded by the outage exponent for Gaussian inputs. It is not yet clear whether this lower bound is tight because solving the GMI upper bound for input distributions such that (37) holds remains a challenge. This also implies that there may possibly exist other input distributions having a larger outage exponent than the Gaussian distribution.

## REFERENCES


[1] L. H. Ozarow, S. Shamai, and A. D. Wyner, "Information theoretic considerations for cellular mobile radio," *IEEE Trans. Veh. Technol.*, vol. 43, no. 2, pp. 359–378, May 1994.

[2] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.

[3] A. Guillén i Fàbregas and G. Caire, "Coded modulation in the block-fading channel: Coding theorems and code construction," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 91–114, Jan. 2006.

[4] N. Merhav, G. Kaplan, A. Lapidoth, and S. Shamai (Shitz), "On information rates for mismatched decoders," *IEEE Trans. Inf. Theory*, vol. 40, no. 6, pp. 1953–1967, Nov. 1994.

[5] A. T. Asyhari and A. Guillén i Fàbregas, "Mismatched CSI outage exponents of block-fading channels," in *Proc. IEEE Int. Symp. Inf. Theory*, St. Petersburg, Russia, Jul.–Aug. 2011.

[6] R. Knopp and G. Caire, "Power control and beamforming for systems with multiple transmit and receive antennas," *IEEE Trans. Wireless Commun.*, vol. 1, no. 4, pp. 638–648, Oct. 2002.

[7] G. Taricco and E. Biglieri, "Space-time decoding with imperfect channel estimation," *IEEE Trans. Wireless Commun.*, vol. 4, no. 4, pp. 1874–1888, Jul. 2005.

[8] G. Kaplan and S. Shamai (Shitz), "Information rates and error exponents of compound channels with application to antipodal signaling in a fading environment," *AEÜ Archiv für Elektronik und Übertragungstechnik*, vol. 47, no. 4, pp. 228–239, 1993.

[9] A. T. Asyhari and A. Guillén i Fàbregas, "Nearest neighbor decoding in MIMO block-fading channels with imperfect CSIR," *IEEE Trans. Inf. Theory*, vol. 58, no. 3, pp. 1483–1517, Mar. 2012.

[10] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, 2nd ed. New York: Springer-Verlag, 1998.

[11] T. T. Kim and G. Caire, "Diversity gains of power control with noisy CSIT in MIMO channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 4, pp. 1618–1626, Apr. 2009.

[12] T. T. Kim, K. D. Nguyen, and A. Guillén i Fàbregas, "Coded modulation with mismatched CSIT over MIMO block-fading channels," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5631–5640, Nov. 2010.

[13] E. Visotsky and U. Madhow, "Space-time transmit precoding with imperfect feedback," *IEEE Trans. Inf. Theory*, vol. 47, no. 6, pp. 2632–2639, Sep. 2001.

[14] A. Ganti, A. Lapidoth, and İ. E. Telatar, "Mismatched decoding revisited: General alphabets, channels with memory, and the wide-band limit," *IEEE Trans. Inf. Theory*, vol. 46, no. 7, pp. 2315–2328, Nov. 2000.

[15] A. Lapidoth, "Nearest neighbor decoding for additive non-Gaussian noise channels," *IEEE Trans. Inf. Theory*, vol. 42, no. 5, pp. 1520–1529, Sep. 1996.

[16] V. Aggarwal and A. Sabharwal, "Bits about the channel: Multiround protocols for two-way fading channels," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3352–3370, Jun. 2011.

[17] ——, "Power-controlled feedback and training for two-way MIMO channels," *IEEE Trans. Inf. Theory*, vol. 56, no. 7, pp. 3310–3331, Jul. 2010.

[18] X. J. Zhang, Y. Gong, and K. B. Letaief, "Power control and channel training for MIMO channels: A DMT perspective," *IEEE Trans. Wireless Commun.*, vol. 10, no. 7, pp. 2080–2089, Jul. 2011.

[19] L. Zheng and D. N. C. Tse, "Communication on the Grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. Inf. Theory*, vol. 48, no. 2, pp. 359–383, Feb. 2002.

[20] M. Médard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Trans. Inf. Theory*, vol. 46, no. 3, pp. 933–946, May 2000.

[21] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1985.