Mutual Information Bounds for MIMO Channels under Imperfect Receiver CSI

Adriano Pastore and Michael Joham
Associate Institute for Signal Processing
Technische Universität München
80290 Munich, Germany
Telephone: +49 89 289-28510, Fax: +49 89 289-28504
Email: adriano@mytum.de, joham@tum.de

***Résumé*—We propose an upper and a lower bound on the mutual information of a pilot-aided MIMO link operating under imperfect receiver CSI. Said bounds depend on the linear channel estimator employed. They can be understood as a generalization of results from [1] and [2]. Furthermore, we prove that any bijective operation performed on the channel estimation does not modify the value of the mutual information bounds. Also, we prove that any sufficient channel estimator is equivalently optimal in terms of maximizing the upper and lower mutual information bounds. These two properties result to be identical to known properties of the *true* mutual information, namely the invariance against bijective (invertible) operations performed on the random variables, and the optimality of sufficient estimation (e.g., [3]). The study of these bound properties offers new insights and a deeper understanding of them, as compared to their original derivation in [1] and [2]. To complete the analysis, we show that the gap between the upper and lower bound (and thus also between the true mutual information and the lower bound) is generally upper-bounded by a constant which only depends on the number of receive antennas. This can be seen as a generalization of findings from [1]. Additionally, we show that by taking certain parameter interdependencies more accurately into account, this bound gap becomes significantly smaller, and the mutual information bounds are thus shown to be even tighter. Finally, we focus on the optimization of the pilot symbols as a function of the channel/noise statistics and derive the optimal pilot sequence for a SISO channel.**

I. Introduction

Most of the papers on *multiple-input multiple-output* (MIMO) communications make the assumption that the *channel state information* (CSI) at the receiver side is error-free (e.g., [4], [5]). The resulting $\log\text{-}\det$ expression for the data rate enabled the waterfilling solution in [4], the efficient solution for the MIMO *multiple access channel* (MAC) in [6], the duality results in [7], [8], [9], [10], the result for the MIMO BC capacity region in [11], and the efficient solution for the MIMO *broadcast channel* (BC) in [12], [13] for example.

Unfortunately, the receiver does not know the CSI *a priori* and must therefore estimate it. Clearly, the resulting estimate is only error-free approximately. In this paper, the case is considered that the errors cannot be neglected. Consequently, the classical $\log\text{-}\det$ expression for the mutual information is not applicable anymore. The treatment of such channels is difficult in general, especially due to the fact that the output distribution for typical input alphabets such as Gaussian alphabets—and thus the computation of mutual information—is intractable and costly to estimate via simulation, let alone the capacity-achieving distribution. While for less realistic assumptions (e.g., perfectly known channel statistics at the transmitter and the receiver), partial characterizations of the capacity-achieving distributions have been achieved [14], results are still scarce for the more realistic assumption of erroneous CSI. In fact, accurate characterizations of these distributions do not even exist for single-antenna systems, as of now.

Lower and upper bounds for the mutual information in the case of erroneous CSI for a *single-input single-output* (SISO) system were proposed in [15]. The results from [15] were then generalized to MIMO channels and employed in [2] for predicting the amount of necessary channel estimation.

In [1] and [2], it is assumed that a *minimum mean square error* (MMSE) channel estimator is used. By following the ideas of these two works, we derive upper and lower mutual information bounds for erroneous CSI that are valid for any linear[1] channel estimator. We will show that any sufficient channel estimator is optimal for simultaneously maximizing the upper and lower mutual information bound. Moreover, unlike [1], we do not assume that the channel estimation error covariance is a scaled identity matrix that does not depend on the noise covariance, but take into account the existing dependency between those two. In particular, this allows us to show that the bounds are substantially tighter than in the analysis of [1]. Finally, we derive the pilot sequence for a SISO channel that maximizes the mutual information lower bound.

II. System Model

We use the standard MIMO model with an $M \times N$ channel matrix $\boldsymbol{H}$, whose entries are zero-mean circularly symmetric complex Gaussian distributed and can have any correlations, i.e., $\boldsymbol{h} = \operatorname{vec}(\boldsymbol{H}) \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{0}, \boldsymbol{C}_{\boldsymbol{h}})$ with the covariance matrix $\boldsymbol{C}_{\boldsymbol{h}} = \mathrm{E}[\boldsymbol{h}\boldsymbol{h}^{\mathrm{H}}]$. The received signal for data transmission is

$$\boldsymbol{x} = \boldsymbol{H}\boldsymbol{s} + \boldsymbol{\eta} \in \mathbb{C}^M \tag{1}$$

where $\boldsymbol{s} \in \mathbb{C}^N$ is the MIMO channel input and the noise is $\boldsymbol{\eta} \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{0}, \boldsymbol{C}_{\boldsymbol{\eta}})$. Under the assumption of perfect CSI at

[1] Here, linear estimators are optimal owing to the joint Gaussianity of the observation and the channel.

the receiver and for Gaussian channel inputs $\boldsymbol{s}$, the mutual information can be shown to be (e.g., [4])

$$\mathcal{I}(\boldsymbol{s};\boldsymbol{x}) = \log\det\left(\mathbf{I} + \boldsymbol{C}_{\boldsymbol{\eta}}^{-1}\boldsymbol{H}\boldsymbol{C}_{\boldsymbol{s}}\boldsymbol{H}^{\mathrm{H}}\right). \tag{2}$$

In this paper, however, it is assumed that the channel $\boldsymbol{H}$ is not perfectly known to the receiver. Instead, it is estimated via the pilot channel

$$\boldsymbol{x}_{\mathrm{T}} = \boldsymbol{S}\boldsymbol{h} + \boldsymbol{\eta}_{\mathrm{T}} \tag{3}$$

with the matrix of training symbols $\boldsymbol{S} \in \mathbb{C}^{MT\times MN}$, the noise $\boldsymbol{\eta}_{\mathrm{T}} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \boldsymbol{C}_{\boldsymbol{\eta}_{\mathrm{T}}}) = \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \boldsymbol{C}_{\boldsymbol{\eta},\text{time}}^{\mathrm{T}} \otimes \boldsymbol{C}_{\boldsymbol{\eta}})$, where $\boldsymbol{C}_{\boldsymbol{\eta},\text{time}}$ captures the temporal noise correlation, and the received signal $\boldsymbol{x}_{\mathrm{T}} \in \mathbb{C}^{MT}$ of the pilot channel. The duration of the training phase is T, and the sequence of training symbols is collected in a matrix $\boldsymbol{T} \in \mathbb{C}^{N\times T}$ such that $\boldsymbol{S} = \boldsymbol{T}^{\mathrm{T}} \otimes \mathbf{I}_M$, where '$\otimes$' denotes the Kronecker product.

Clearly, since the training symbols are known to the receiver, the received signal $\boldsymbol{x}_{\mathrm{T}}$ and the channel $\boldsymbol{h}$ are jointly Gaussian. Therefore, the optimal channel estimator is linear and its output $\hat{\boldsymbol{h}}$ is zero-mean Gaussian with covariance $\boldsymbol{C}_{\hat{\boldsymbol{h}}}$. With the linear estimator $\boldsymbol{G} \in \mathbb{C}^{MN\times MT}$, the estimate can be written as

$$\hat{\boldsymbol{h}} = \boldsymbol{G}\boldsymbol{x}_{\mathrm{T}} = \boldsymbol{G}(\boldsymbol{S}\boldsymbol{h} + \boldsymbol{\eta}_{\mathrm{T}}). \tag{4}$$

Accordingly, $\hat{\boldsymbol{H}} \in \mathbb{C}^{M\times N}$ is the estimate of $\boldsymbol{H}$ such that $\hat{\boldsymbol{h}} = \operatorname{vec}(\hat{\boldsymbol{H}})$. The channel estimation error is the difference

$$\boldsymbol{\Omega} = \boldsymbol{H} - \hat{\boldsymbol{H}}. \tag{5}$$

We will often refer to it in vectorized form, i.e., $\boldsymbol{\omega} = \operatorname{vec}(\boldsymbol{\Omega})$. With these variables, the equation (1) governing the MIMO system can be rewritten as

$$\boldsymbol{x} = \boldsymbol{H}\boldsymbol{s} + \boldsymbol{\eta} = \hat{\boldsymbol{H}}\boldsymbol{s} + \boldsymbol{\eta}', \tag{6}$$

where $\hat{\boldsymbol{H}}$ can be seen as a perfectly known channel variable, and $\boldsymbol{\eta}' = \boldsymbol{\Omega}\boldsymbol{s} + \boldsymbol{\eta} \in \mathbb{C}^M$ is called the *effective noise* [2]. Note that unlike $\boldsymbol{\eta}$, it is non-Gaussian and depends on the input signal $\boldsymbol{s}$.

III. Mutual Information Bounds

A. Mutual Information

The quantity of interest is the mutual information between the system input $\boldsymbol{s}$ and the output pair $(\boldsymbol{x}, \hat{\boldsymbol{H}})$. Due to the Bayes rule and the fact that $\boldsymbol{s}$ and $\hat{\boldsymbol{H}}$ are independent, this mutual information can be written as

$$\mathcal{I}(\boldsymbol{s};(\boldsymbol{x}|\hat{\boldsymbol{H}})) = \mathcal{I}(\boldsymbol{s};\boldsymbol{x}|\hat{\boldsymbol{H}}) + \mathcal{I}(\boldsymbol{s};\hat{\boldsymbol{H}}) = \mathcal{I}(\boldsymbol{s};\boldsymbol{x}|\hat{\boldsymbol{H}}). \tag{7}$$

From now on, we append a subscript G to any mutual information quantity, whenever it is assumed that the input alphabet is Gaussian, e.g., $\mathcal{I}_{\mathrm{G}}(\boldsymbol{s};\boldsymbol{x}|\hat{\boldsymbol{H}})$. The steps for obtaining an upper and a lower bound are essentially the same as in [2] for the lower bound and [1] for the upper bound, except that we generalize the derivation of these bounds so as to hold for *any* linear channel estimator.

B. Lower Bound

Since the covariance $\boldsymbol{C}_{\boldsymbol{s}}$ may be rank-deficient, we substitute $\boldsymbol{s}$ by an appropriate dimension-reduced representation $\check{\boldsymbol{s}} \in \mathbb{C}^r$ with $r = \operatorname{rank}(\boldsymbol{C}_{\boldsymbol{s}}) \leq N$ related to $\boldsymbol{s}$ by $\boldsymbol{s} = \boldsymbol{R}\check{\boldsymbol{s}}$ with a tall subunitary matrix $\boldsymbol{R} \in \mathbb{C}^{N\times r}$. The covariance matrices of $\boldsymbol{s}$ and $\check{\boldsymbol{s}}$ are related via $\boldsymbol{C}_{\boldsymbol{s}} = \boldsymbol{R}\boldsymbol{C}_{\check{\boldsymbol{s}}}\boldsymbol{R}^{\mathrm{H}}$. The differential entropies of $\boldsymbol{s}$ and $\check{\boldsymbol{s}}$ are the same, i.e.,

$$\mathcal{H}(\boldsymbol{s}) = \mathcal{H}(\check{\boldsymbol{s}}) \qquad \mathcal{H}(\boldsymbol{s}|\hat{\boldsymbol{H}}) = \mathcal{H}(\check{\boldsymbol{s}}|\hat{\boldsymbol{H}}).$$

We first expand $\mathcal{I}_{\mathrm{G}}(\boldsymbol{s};\boldsymbol{x}|\hat{\boldsymbol{H}})$ as follows:

$$\begin{aligned}\mathcal{I}_{\mathrm{G}}(\boldsymbol{s};\boldsymbol{x}|\hat{\boldsymbol{H}}) &= \mathcal{H}(\check{\boldsymbol{s}}|\hat{\boldsymbol{H}}) - \mathcal{H}(\boldsymbol{s}|\hat{\boldsymbol{H}},\boldsymbol{x}) \\ &= \log\det(\pi\mathrm{e}\,\boldsymbol{C}_{\check{\boldsymbol{s}}}) - \mathcal{H}(\boldsymbol{s}|\hat{\boldsymbol{H}},\boldsymbol{x}),\end{aligned} \tag{8}$$

since the transmit signal $\boldsymbol{s}$ is assumed to be Gaussian, so its entropy equals $\log\det(\pi\mathrm{e}\,\boldsymbol{C}_{\check{\boldsymbol{s}}})$. Next, the subtrahend in (8) is upper-bounded in several steps until we obtain [2]

$$\mathcal{H}(\boldsymbol{s}|\hat{\boldsymbol{H}},\boldsymbol{x}) \leq \mathrm{E}_{\hat{\boldsymbol{H}}}\left[\log\det\left(\pi\mathrm{e}\left(\boldsymbol{C}_{\check{\boldsymbol{s}}} - \boldsymbol{C}_{\check{\boldsymbol{s}},\boldsymbol{x}|\hat{\boldsymbol{h}}}\boldsymbol{C}_{\boldsymbol{x}|\hat{\boldsymbol{h}}}^{-1}\boldsymbol{C}_{\boldsymbol{x},\check{\boldsymbol{s}}|\hat{\boldsymbol{h}}}\right)\right)\right]. \tag{9}$$

The matrix inside the determinant is regular, since it is a Schur complement of the full-rank covariance matrix of the Gaussian variable $\left[\check{\boldsymbol{s}}^{\mathrm{T}}|\hat{\boldsymbol{H}} \quad \boldsymbol{x}^{\mathrm{T}}|\hat{\boldsymbol{H}}\right]^{\mathrm{T}}$. We can thus invert it, so that the right-hand side of (9) reads as

$$-\mathrm{E}_{\hat{\boldsymbol{H}}}\left[\log\det\left(\frac{1}{\pi\mathrm{e}}\left(\boldsymbol{C}_{\check{\boldsymbol{s}}} - \boldsymbol{C}_{\check{\boldsymbol{s}},\boldsymbol{x}|\hat{\boldsymbol{h}}}\boldsymbol{C}_{\boldsymbol{x}|\hat{\boldsymbol{h}}}^{-1}\boldsymbol{C}_{\boldsymbol{x},\check{\boldsymbol{s}}|\hat{\boldsymbol{h}}}\right)^{-1}\right)\right]. \tag{10}$$

With the matrix inversion lemma,

$$\left(\boldsymbol{C}_{\check{\boldsymbol{s}}} - \boldsymbol{C}_{\check{\boldsymbol{s}},\boldsymbol{x}|\hat{\boldsymbol{h}}}\boldsymbol{C}_{\boldsymbol{x}|\hat{\boldsymbol{h}}}^{-1}\boldsymbol{C}_{\boldsymbol{x},\check{\boldsymbol{s}}|\hat{\boldsymbol{h}}}\right)^{-1} \triangleq \boldsymbol{C}_{\check{\boldsymbol{s}}}^{-1} + \boldsymbol{R}^{\mathrm{H}}\breve{\boldsymbol{H}}^{\mathrm{H}}\breve{\boldsymbol{C}}_{\boldsymbol{\eta}'}^{-1}\breve{\boldsymbol{H}}\boldsymbol{R},$$

where the matrices $\breve{\boldsymbol{H}}$ and $\breve{\boldsymbol{C}}_{\boldsymbol{\eta}'}$ are abbreviations defined via

$$\breve{\boldsymbol{H}}\boldsymbol{R} = \boldsymbol{C}_{\boldsymbol{x},\check{\boldsymbol{s}}|\hat{\boldsymbol{h}}}\boldsymbol{C}_{\check{\boldsymbol{s}}}^{-1} = \left(\hat{\boldsymbol{H}} + \mathrm{E}\left[\boldsymbol{\Omega}|\hat{\boldsymbol{H}}\right]\right)\boldsymbol{R} \tag{11a}$$

$$\begin{aligned}\breve{\boldsymbol{C}}_{\boldsymbol{\eta}'} &= \boldsymbol{C}_{\boldsymbol{x}|\hat{\boldsymbol{h}}} - \boldsymbol{C}_{\boldsymbol{x},\check{\boldsymbol{s}}|\hat{\boldsymbol{h}}}\boldsymbol{C}_{\check{\boldsymbol{s}}}^{-1}\boldsymbol{C}_{\check{\boldsymbol{s}},\boldsymbol{x}|\hat{\boldsymbol{h}}} \\ &= \boldsymbol{C}_{\boldsymbol{\eta}} + \boldsymbol{C}_{\boldsymbol{s}} * \boldsymbol{C}_{\boldsymbol{\omega}|\hat{\boldsymbol{h}}}\end{aligned} \tag{11b}$$

where '$*$' denotes a bilinear operator defined as

$$\boldsymbol{X} * \boldsymbol{Y} = (\mathbf{1}_N^{\mathrm{T}} \otimes \mathbf{I}_M)\big((\boldsymbol{X} \otimes \mathbf{1}_{M\times M}) \odot \boldsymbol{Y}\big)(\mathbf{1}_N \otimes \mathbf{I}_M).$$

wherein '$\odot$' stands for the Hadamard (entry-wise) product. From (11a), we can settle on the choice $\breve{\boldsymbol{H}} = \hat{\boldsymbol{H}} + \mathrm{E}\left[\boldsymbol{\Omega}|\hat{\boldsymbol{H}}\right]$. Assembling (8)–(11b), we obtain the lower bound

$$\underline{\mathcal{I}}_{\mathrm{G}}(\boldsymbol{s};\boldsymbol{x}|\hat{\boldsymbol{H}}) = \mathrm{E}_{\breve{\boldsymbol{H}}}\left[\log\det\left(\mathbf{I} + \breve{\boldsymbol{C}}_{\boldsymbol{\eta}'}^{-1}\breve{\boldsymbol{H}}\boldsymbol{C}_{\boldsymbol{s}}\breve{\boldsymbol{H}}^{\mathrm{H}}\right)\right]. \tag{12}$$

In the special case of MMSE channel estimation, we have by the orthogonality principle $\mathrm{E}\left[\boldsymbol{\Omega}|\hat{\boldsymbol{H}}\right] = 0$, so $\breve{\boldsymbol{H}} = \hat{\boldsymbol{H}}$ and $\breve{\boldsymbol{C}}_{\boldsymbol{\eta}'} = \boldsymbol{C}_{\boldsymbol{\eta}'}$, because $\boldsymbol{C}_{\boldsymbol{\omega}|\hat{\boldsymbol{h}}} = \boldsymbol{C}_{\boldsymbol{\omega}}$, whereby

$$\underline{\mathcal{I}}_{\mathrm{G}} = \mathrm{E}_{\hat{\boldsymbol{H}}}\left[\log\det\left(\mathbf{I} + \boldsymbol{C}_{\boldsymbol{\eta}'}^{-1}\hat{\boldsymbol{H}}\boldsymbol{C}_{\boldsymbol{s}}\hat{\boldsymbol{H}}^{\mathrm{H}}\right)\right]. \tag{13}$$

The latter lower bound is a generalization of that given in [1], [2]. This bound is valid for Gaussian input signals $\boldsymbol{s}$ and superior input distributions, including the unknown capacity-achieving distribution.

C. Upper Bound

We expand the mutual information as

$$\mathcal{I}(\boldsymbol{x},\boldsymbol{s}|\hat{\boldsymbol{H}}) = \mathcal{H}(\boldsymbol{x}|\hat{\boldsymbol{H}}) - \mathcal{H}(\boldsymbol{x}|\boldsymbol{s},\hat{\boldsymbol{H}}) \tag{14}$$

and upper bound the entropy of the output $\boldsymbol{x}|\hat{\boldsymbol{H}}$ by the Gaussian entropy of covariance $\mathrm{cov}(\boldsymbol{x}|\hat{\boldsymbol{H}})$ [1]:

$$\mathcal{H}(\boldsymbol{x}|\hat{\boldsymbol{H}}) \leq \mathrm{E}_{\hat{\boldsymbol{H}}}\left[\log\det\left(\pi\mathrm{e}\ \mathrm{cov}(\boldsymbol{x}|\hat{\boldsymbol{H}})\right)\right].$$

On the other hand, the subtrahend in (14) equals $\mathcal{H}(\boldsymbol{\eta}'|\boldsymbol{s},\hat{\boldsymbol{H}})$ because $\boldsymbol{\eta}' = \boldsymbol{x} - \hat{\boldsymbol{H}}\boldsymbol{s}$ and a constant added to a random variable does not change its entropy. Since $\boldsymbol{\eta}'|\boldsymbol{s},\hat{\boldsymbol{H}}$ is Gaussian, said subtrahend can be expressed with the classic $\log\det$-formula of Gaussian entropy, i.e.,

$$\mathcal{H}(\boldsymbol{\eta}'|\boldsymbol{s},\hat{\boldsymbol{H}}) = \mathrm{E}_{\boldsymbol{s}}\left[\log\det\left(\pi\mathrm{e}\left(\boldsymbol{s}\boldsymbol{s}^{\mathrm{H}} * \boldsymbol{C}_{\boldsymbol{\omega}|\hat{\boldsymbol{h}}} + \boldsymbol{C}_{\boldsymbol{\eta}}\right)\right)\right]$$

Assembling all the results above, one can show that the resulting upper bound can be represented as

$$\overline{\mathcal{I}} = \underline{\mathcal{I}}_{\mathrm{G}} + \mathit{\Delta} \tag{15}$$

where the non-negative bound gap $\mathit{\Delta}$ reads as

$$\mathit{\Delta} = \overline{\mathcal{I}} - \underline{\mathcal{I}}_{\mathrm{G}} = \mathrm{E}_{\boldsymbol{s}}\left[\log\frac{\det(\boldsymbol{C}_{\boldsymbol{\eta}} + \boldsymbol{C}_{\boldsymbol{s}} * \boldsymbol{C}_{\boldsymbol{\omega}|\hat{\boldsymbol{h}}})}{\det(\boldsymbol{C}_{\boldsymbol{\eta}} + \boldsymbol{s}\boldsymbol{s}^{\mathrm{H}} * \boldsymbol{C}_{\boldsymbol{\omega}|\hat{\boldsymbol{h}}})}\right]. \tag{16}$$

Note that in contrast to $\underline{\mathcal{I}}_{\mathrm{G}}$, for which we needed to assume Gaussian input symbols, the upper bound $\overline{\mathcal{I}}$ as given in (15) still depends on the distribution of $\boldsymbol{s}$. This means that it is not restricted to any particular input distribution. It may be used, for instance, as an upper bound for the Gaussian distribution (we then write $\overline{\mathcal{I}}_{\mathrm{G}}$ for notational consistence), or for the unknown capacity-achieving distribution.

Figure 1 represents the three mutual information bounds $\underline{\mathcal{I}}_{\mathrm{G}}$, $\overline{\mathcal{I}}_{\mathrm{G}}$, and the perfect CSI upper bound $\overline{\mathcal{I}}_{\mathrm{p}}$, which is used e.g. in [15], and whose general definition is

$$\overline{\mathcal{I}}_{\mathrm{p}} = \mathrm{E}_{\boldsymbol{H}}\left[\log\det\left(\mathbf{I} + \boldsymbol{C}_{\boldsymbol{\eta}}^{-1}\boldsymbol{H}\boldsymbol{C}_{\boldsymbol{s}}\boldsymbol{H}^{\mathrm{H}}\right)\right]. \tag{17}$$

The plot of Figure 1 was generated for a SISO channel with a channel variance $\sigma_h^2 = 1$, no temporal noise correlation (i.e., $\boldsymbol{C}_{\boldsymbol{\eta},\mathrm{time}} = \mathbf{I}$) and equal transmit power σ_s^2 for both the training and data transmit phases. In the considered SISO case, all three bounds have closed-form expressions. The aspect of Figure 1, however, is similar for higher-dimensional cases. The closed-form expressions allow for the calculation of the following two limits [cf. (31) and (33)]:

$$\mathit{\Delta}_{\text{high SNR}} = \log(2) - \varsigma(1) \approx 0.0968 \ll \gamma \approx 0.5772$$

where the function $x \mapsto \varsigma(x)$ is defined as $\varsigma(x) = \mathrm{e}^x\,\mathrm{E}_1(x)$, with the exponential integral $\mathrm{E}_1(x) = \int_x^\infty \mathrm{e}^{-t}/t\,\mathrm{d}t$, and

$$\lim_{\mathrm{SNR}\to\infty}\left[\overline{\mathcal{I}}_{\mathrm{p}} - \underline{\mathcal{I}}_{\mathrm{G}}\right] = \log(2) \approx 0.6931.$$

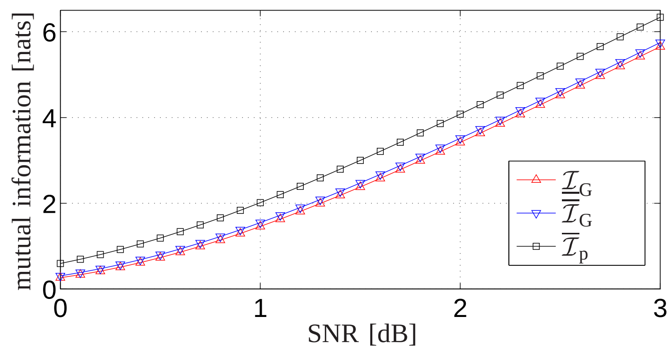

FIG. 1. Lower bound $\underline{\mathcal{I}}_{\mathrm{G}}$, upper bound $\overline{\mathcal{I}}_{\mathrm{G}}$ and perfect CSI upper bound $\overline{\mathcal{I}}_{\mathrm{p}}$ as functions of the SNR

IV. OPTIMAL CHANNEL ESTIMATION

It is difficult to show anything about the optimal structure of the estimator just by differentiating the mutual information bounds as given in (12) and (15) with respect to the channel estimator $\boldsymbol{G}$. So, for the purpose of studying how the bounds behave as a function of the channel estimator, a more viable approach consists in resorting to a parametrization of the channel estimator $\boldsymbol{G}$, which reads as

$$\boldsymbol{G} = \boldsymbol{C}_{\hat{\boldsymbol{h}}}^{1/2}\boldsymbol{F}^{\mathrm{H}}(\boldsymbol{S}\boldsymbol{C}_{\boldsymbol{h}}\boldsymbol{S}^{\mathrm{H}} + \boldsymbol{C}_{\boldsymbol{\eta}_{\mathrm{T}}})^{-1/2}, \tag{18}$$

where $\boldsymbol{F} \in \mathbb{C}^{MT\times MN}$ is a tall subunitary matrix, i.e., $\boldsymbol{F}^{\mathrm{H}}\boldsymbol{F} = \mathbf{I}$. The three matrix factors in (18) can be interpreted as three consecutive stages, which successively

- decorrelate and whiten the estimator's input,
- perform a subunitary rank-reduction,
- recorrelate the estimator's output with the covariance $\boldsymbol{C}_{\hat{\boldsymbol{h}}}$.

For sufficient estimators, the subunitary rank-reduction (second stage) reads as

$$\boldsymbol{F}_{\mathrm{opt}}^{\mathrm{H}} = \left(\boldsymbol{C}_{\boldsymbol{h}}\boldsymbol{S}^{\mathrm{H}}(\boldsymbol{S}\boldsymbol{C}_{\boldsymbol{h}}\boldsymbol{S}^{\mathrm{H}} + \boldsymbol{C}_{\boldsymbol{\eta}_{\mathrm{T}}})^{-1}\boldsymbol{S}\boldsymbol{C}_{\boldsymbol{h}}\right)^{1/2} \cdot \\ \cdot\, \boldsymbol{C}_{\boldsymbol{h}}\boldsymbol{S}^{\mathrm{H}}(\boldsymbol{S}\boldsymbol{C}_{\boldsymbol{h}}\boldsymbol{S}^{\mathrm{H}} + \boldsymbol{C}_{\boldsymbol{\eta}_{\mathrm{T}}})^{-1/2}. \tag{19}$$

Theorem IV.1. *Any linear bijective operation applied on the channel estimate $\hat{\boldsymbol{h}}$ does not modify the value of the mutual information bounds, as long as the estimator $\boldsymbol{G}$ is full-rank.*

It can be shown, in fact, that by inserting the parametrized expression (18) of the channel estimator $\boldsymbol{G}$ into the expressions of $\underline{\mathcal{I}}_{\mathrm{G}}$ and of $\mathit{\Delta}$, the latter result to be independent of the last estimator stage $\boldsymbol{C}_{\hat{\boldsymbol{h}}}^{1/2}$ (recorrelator). From this observation, we can directly infer Theorem IV.1.

Theorem IV.2. *All sufficient channel estimators are equivalently optimal with regard to maximizing both mutual information bounds.*

Due to space limitations, we omit the proof. Note that the Theorems IV.1 and IV.2 are closely linked. Indeed, the invariance property IV.1 implies that all sufficient estimators are equivalent, because they all have the form $\boldsymbol{X}^{-1}\boldsymbol{S}^{\mathrm{H}}\boldsymbol{C}_{\boldsymbol{\eta}_{\mathrm{T}}}^{-1}$ with an arbitrary last stage $\boldsymbol{X}^{-1}$. On the other hand, said invariance property IV.1 holds not only for sufficient estimators, but also for any non-sufficient full-rank estimator $\boldsymbol{G}$.

Corollary IV.1. *The lower bound expression* (13)*, which is valid for an MMSE channel estimator, is also valid for any other sufficient channel estimator. As another consequence, the worst-case noise interpretation given in [2] of said lower bound* (13) *is more generally applicable to any sufficient channel estimator.*

The properties described by the above Theorems IV.1 and IV.2 are similar to certain properties of the *true* mutual information $\mathcal{I}(\boldsymbol{s};\boldsymbol{x}|\hat{\boldsymbol{h}})$. In fact, the invariance property IV.1 is analogous to the invariance of $\mathcal{I}(\boldsymbol{s};\boldsymbol{x}|\hat{\boldsymbol{h}})$ against invertible operations performed on $\hat{\boldsymbol{h}}$. In fact, such an operation corresponds to a change of variables (variable substitution in the integral representation of the mutual information) which does not affect the value of $\mathcal{I}(\boldsymbol{s};\boldsymbol{x}|\hat{\boldsymbol{h}})$. As to the optimality of sufficient estimation IV.2, it also holds for the true mutual information: it has been proven in [3] that the pilot receive symbols $\boldsymbol{x}_{\mathrm{T}}$ are as informative as the conditional mean estimation $\hat{\boldsymbol{h}}$, i.e.,

$$\mathcal{I}(\boldsymbol{s};\boldsymbol{x}|\hat{\boldsymbol{h}}) = \mathcal{I}(\boldsymbol{s};\boldsymbol{x}|\boldsymbol{x}_{\mathrm{T}}). \tag{20}$$

Hence, there can exist no better estimator than the conditional mean estimator. Combining this result with the aforementioned invariance property, we can infer that all sufficient channel estimators are optimal with regard to maximizing the true mutual information.

V. On the Bound Gap

In order to analyze the accuracy and tightness of the lower and upper bound pair, we are interested in knowing the highest possible values attained by their difference Δ as given in (16). That is, we search for the supremum of Δ over all admissible triples $(\boldsymbol{C}_{\boldsymbol{\omega}|\hat{\boldsymbol{h}}}, \boldsymbol{C}_{\boldsymbol{\eta}}, \boldsymbol{C}_{\boldsymbol{s}})$. The individual constraints imposed on these three matrices are that $\boldsymbol{C}_{\boldsymbol{s}} \in \mathbb{C}^{N\times N}$ is positive semidefinite with $\mathrm{tr}(\boldsymbol{C}_{\boldsymbol{s}}) \leq E_{\mathrm{tx}}$, and that $\boldsymbol{C}_{\boldsymbol{\omega}|\hat{\boldsymbol{h}}} \in \mathbb{C}^{MN\times MN}$ and $\boldsymbol{C}_{\boldsymbol{\eta}} \in \mathbb{C}^{M\times M}$ are positive definite. Additionally, $\boldsymbol{C}_{\boldsymbol{\omega}|\hat{\boldsymbol{h}}}$ and $\boldsymbol{C}_{\boldsymbol{\eta}}$ are interdependent via the relation

$$\boldsymbol{C}_{\boldsymbol{\omega}|\hat{\boldsymbol{h}}} = \left(\boldsymbol{C}_{\boldsymbol{h}}^{-1} + \boldsymbol{S}^{\mathrm{H}}\boldsymbol{G}^{\mathrm{H}}(\boldsymbol{G}\boldsymbol{C}_{\boldsymbol{\eta}_{\mathrm{T}}}\boldsymbol{G}^{\mathrm{H}})^{-1}\boldsymbol{G}\boldsymbol{S}\right)^{-1} \tag{21}$$

with $\boldsymbol{C}_{\boldsymbol{\eta}_{\mathrm{T}}} = \boldsymbol{C}_{\boldsymbol{\eta},\mathrm{time}}^{\mathrm{T}} \otimes \boldsymbol{C}_{\boldsymbol{\eta}}$. Let us call $\mathcal{D}_{\mathrm{indep.}}$ the set of admissible triples $(\boldsymbol{C}_{\boldsymbol{\omega}|\hat{\boldsymbol{h}}}, \boldsymbol{C}_{\boldsymbol{\eta}}, \boldsymbol{C}_{\boldsymbol{s}})$ *without* this interdependence constraint, and $\mathcal{D}_{\mathrm{interdep.}}$ the set of these triples fulfilling said interdependence constraint. Let us call

$$\Delta_{\mathrm{sup,interdep.}} = \sup_{\mathcal{D}_{\mathrm{interdep.}}} \Delta \qquad \Delta_{\mathrm{sup,indep.}} = \sup_{\mathcal{D}_{\mathrm{indep.}}} \Delta.$$

We obviously have $\mathcal{D}_{\mathrm{interdep.}} \subset \mathcal{D}_{\mathrm{indep.}}$, and thus

$$\Delta_{\mathrm{sup,interdep.}} \leq \Delta_{\mathrm{sup,indep.}}. \tag{22}$$

In the next two subsections, we successively investigate the suprema $\Delta_{\mathrm{sup,indep.}}$ and $\Delta_{\mathrm{sup,interdep.}}$.

A. Independent Noise and Estimation Error Covariance

If $\boldsymbol{C}_{\boldsymbol{\eta}}$ and $\boldsymbol{C}_{\boldsymbol{\omega}|\hat{\boldsymbol{h}}}$ can be chosen freely, as was implicitly assumed in [1], it is easy to show that the supremum of Δ over all positive definite $\boldsymbol{C}_{\boldsymbol{\eta}}$ is attained when $\boldsymbol{C}_{\boldsymbol{\eta}} \to \boldsymbol{0}$, i.e.,

$$\Delta_{\mathrm{sup,indep.}} = \sup_{(\boldsymbol{C}_{\boldsymbol{\omega}|\hat{\boldsymbol{h}}},\boldsymbol{C}_{\boldsymbol{s}})} \left[\log \frac{\det(\boldsymbol{C}_{\boldsymbol{s}} * \boldsymbol{C}_{\boldsymbol{\omega}|\hat{\boldsymbol{h}}})}{\det(\boldsymbol{s}\boldsymbol{s}^{\mathrm{H}} * \boldsymbol{C}_{\boldsymbol{\omega}|\hat{\boldsymbol{h}}})}\right]. \tag{23}$$

The remaining two matrix variables $\boldsymbol{C}_{\boldsymbol{s}}$ and $\boldsymbol{C}_{\boldsymbol{\omega}|\hat{\boldsymbol{h}}}$ can be merged into one by introducing a white complex Gaussian $\boldsymbol{z} \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{0}, \mathbf{I})$ such that $\boldsymbol{s} = \boldsymbol{C}_{\boldsymbol{s}}^{1/2}\boldsymbol{z}$. Then, due to a property of the '$*$' operator, we have

$$\boldsymbol{s}\boldsymbol{s}^{\mathrm{H}} * \boldsymbol{C}_{\boldsymbol{\omega}|\hat{\boldsymbol{h}}} = \boldsymbol{z}\boldsymbol{z}^{\mathrm{H}} * \boldsymbol{C} \tag{24}$$

with a positive definite $\boldsymbol{C} \in \mathbb{C}^{MN\times MN}$ defined as

$$\boldsymbol{C} = \left(\boldsymbol{C}_{\boldsymbol{s}}^{\mathrm{T}/2} \otimes \mathbf{I}\right)\boldsymbol{C}_{\boldsymbol{\omega}|\hat{\boldsymbol{h}}}\left(\boldsymbol{C}_{\boldsymbol{s}}^{*/2} \otimes \mathbf{I}\right). \tag{25}$$

Since the trace of $\boldsymbol{C}_{\boldsymbol{s}}$ is upper-bounded, but the trace of $\boldsymbol{C}_{\boldsymbol{\omega}|\hat{\boldsymbol{h}}}$ may take any value, the trace constraint on $\boldsymbol{C}_{\boldsymbol{s}}$ can be disregarded for finding the supremum (23). With (25), the expression (23) now reads as

$$\Delta_{\mathrm{sup,indep.}} = \sup_{\boldsymbol{C}\succ\boldsymbol{0}} \left[\log \frac{\det(\mathbf{I} * \boldsymbol{C})}{\det(\boldsymbol{z}\boldsymbol{z}^{\mathrm{H}} * \boldsymbol{C})}\right]. \tag{26}$$

Due to the linearity of the '$*$' operator, we see in (26) that $\Delta_{\mathrm{sup,indep.}}$ is independent of any norm of $\boldsymbol{C}$. Therefore, $\Delta_{\mathrm{sup,indep.}}$ is also expressible as a maximum:

$$\Delta_{\mathrm{sup,indep.}} = \max_{\boldsymbol{C}\succeq\boldsymbol{0}} \left[\log \frac{\det(\mathbf{I} * \boldsymbol{C})}{\det(\boldsymbol{z}\boldsymbol{z}^{\mathrm{H}} * \boldsymbol{C})}\right]. \tag{27}$$

For a channel $\boldsymbol{H}$ with i.i.d. Gaussian unit-variance entries and a $\boldsymbol{C}_{\boldsymbol{\omega}|\hat{\boldsymbol{h}}}$ having the structure of a scaled identity matrix, a proof was given in [1] that this supremum (27) equals

$$\Delta_{\mathrm{sup,indep.}} = M\gamma \tag{28}$$

with the Euler-Mascheroni constant $\gamma = 0.57721\,56649\ldots$ However, in [1], this specific structure of $\boldsymbol{C}_{\boldsymbol{\omega}|\hat{\boldsymbol{h}}}$ is assumed *a priori*, which is a somewhat unrealistic assumption. In order to be fulfilled, it would require very specific assumptions on the system parameters. However, by following a different approach from that in [1], it can be shown that (28) holds generally for MISO ($M = 1$) and SIMO channels, without any further assumptions. The proof of the equality (28) can even be extended to MIMO channels if the following three assumptions[2] are met:

- $\boldsymbol{H}$ is a Kronecker channel with only transmit correlation, i.e., $\boldsymbol{h} \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{0}, \boldsymbol{C}_{\mathrm{Tx}}^{\mathrm{T}} \otimes \mathbf{I})$,
- The additive noise $\boldsymbol{\eta}$ is spatially white, i.e., $\boldsymbol{C}_{\boldsymbol{\eta}} = \sigma_{\boldsymbol{\eta}}^2\mathbf{I}$,
- The receive antennas are non-cooperative during the channel estimation phase, which signifies that $\boldsymbol{G} = \boldsymbol{G}_{\mathrm{r}}^{\mathrm{T}} \otimes \mathbf{I}$ with $\boldsymbol{G}_{\mathrm{r}} \in \mathbb{C}^{T\times N}$.

Simulations for the most general case of a MIMO channel with arbitrary channel correlations suggest that the equality

[2]With these three assumptions, one can show that the matrix $\boldsymbol{C}$ from (25) adopts the form of a Kronecker product, which facilitates the search for a maximum as in (27).

(28) holds generally. When expressed as a maximum as in (27), it is attained when $\boldsymbol{C} = \boldsymbol{e}_j \boldsymbol{e}_j^{\mathrm{T}} \otimes \boldsymbol{C}_0$ with any positive definite $\boldsymbol{C}_0 \in \mathbb{C}^{M\times M}$ and any $1 \leq j \leq N$.

B. Interdependent Noise and Estimation Error Covariance

In constrast to the previous section, we now take into account the interdependency between $\boldsymbol{C}_{\boldsymbol{\eta}}$ and $\boldsymbol{C}_{\boldsymbol{\omega}|\hat{\boldsymbol{h}}}$ via (21). We write $\boldsymbol{C}_{\boldsymbol{\eta}} = \alpha\, \boldsymbol{C}_{\boldsymbol{\eta},0}$ with a factor $\alpha > 0$ and a unit-trace positive definite $\boldsymbol{C}_{\boldsymbol{\eta},0}$. The factor α thus represents the noise power, i.e., $\alpha = \sigma_{\boldsymbol{\eta}}^2$, while the structure of spatial noise correlations is captured in the normalized matrix $\boldsymbol{C}_{\boldsymbol{\eta},0}$. Accordingly, we write $\boldsymbol{C}_{\boldsymbol{\eta}_{\mathrm{T}}} = \alpha\, \boldsymbol{C}_{\boldsymbol{\eta},\mathrm{time}}^{\mathrm{T}} \otimes \boldsymbol{C}_{\boldsymbol{\eta},0}$ for the training noise. It can be shown that, if we define the low- and high-SNR limits of $\varDelta$ respectively as

$$\varDelta_{\text{low SNR}} = \lim_{\alpha\to\infty} \varDelta \qquad \varDelta_{\text{high SNR}} = \lim_{\alpha\to 0} \varDelta, \tag{29}$$

then the low-SNR limit is for any type of channel estimation

$$\varDelta_{\text{low SNR}} = 0, \tag{30}$$

whereas the high-SNR limit for *sufficient* channel estimation is

$$\varDelta_{\text{high SNR}} = M \mathrm{E}_{\boldsymbol{s}} \left[\log \frac{1 + \mathrm{tr}\big(\boldsymbol{C}_{\boldsymbol{s}} (\boldsymbol{T}\boldsymbol{C}_{\boldsymbol{\eta},\mathrm{time}}^{-1} \boldsymbol{T}^{\mathrm{H}})^{-1}\big)}{1 + \boldsymbol{s}^{\mathrm{H}} (\boldsymbol{T}\boldsymbol{C}_{\boldsymbol{\eta},\mathrm{time}}^{-1} \boldsymbol{T}^{\mathrm{H}})^{-1} \boldsymbol{s}} \right]. \tag{31}$$

It can be shown that for all admissible $\boldsymbol{C}_{\boldsymbol{s}}$, this high-SNR limit is not larger than a constant which is itself strictly smaller than $M\gamma$. Additionally, we claim that $\varDelta$ grows monotonically with the SNR (in the sense that it decreases monotonically with α), taking values from 0 for the low-SNR limit (30) up to the high-SNR limit (31). This monotonicity can be proven for SIMO channels. A proof for MISO/MIMO channels seems more difficult, but simulations seem to corroborate this conjecture. Therefore,

$$\varDelta_{\text{sup,interdep.}} = \max_{\substack{\boldsymbol{C}_{\boldsymbol{s}} \succeq \boldsymbol{0} \\ \mathrm{tr}(\boldsymbol{C}_{\boldsymbol{s}}) \leq E_{\mathrm{tx}}}} \varDelta_{\text{high SNR}}, \tag{32}$$

and we have, all in all,

$$\varDelta_{\text{sup,interdep.}} < \varDelta_{\text{sup,indep.}} = M\gamma. \tag{33}$$

VI. Optimal Training

It can be shown that for a SISO channel, the lower bound for an MMSE estimator (13) acquires the form $\underline{\mathcal{I}}_{\mathrm{G}} = \varsigma(\sigma_{\boldsymbol{\eta}'}^2 / \sigma_{\boldsymbol{s}}^2 \sigma_{\hat{\boldsymbol{h}}}^2)$, where $\varsigma(x) = \mathrm{e}^x\, \mathrm{E}_1(x)$, as in Section III-C. The pilot symbol matrix $\boldsymbol{T}$ reduces to the row vector $\boldsymbol{t}^{\mathrm{T}}$, the training noise matrix $\boldsymbol{C}_{\boldsymbol{\eta}_{\mathrm{T}}}$ reduces to $\sigma_\eta^2 \boldsymbol{C}_{\boldsymbol{\eta},\mathrm{time}}$. The problem of finding the optimal training sequence $\boldsymbol{t}^{\mathrm{T}}$ which maximizes the mutual information lower bound can be stated as

$$\boldsymbol{t}_{\mathrm{opt}} = \operatorname*{argmax}_{\|\boldsymbol{t}\|_2^2 \leq 1} \left[\varsigma\left(\frac{\sigma_{\boldsymbol{\eta}'}^2}{\sigma_{\boldsymbol{s}}^2 \sigma_{\hat{\boldsymbol{h}}}^2} \right) \right]. \tag{34}$$

It can be shown that

$$\boldsymbol{t}_{\mathrm{opt}} = \operatorname*{argmax}_{\|\boldsymbol{t}\|_2^2 \leq 1} \left[\boldsymbol{t}^{\mathrm{H}} \boldsymbol{C}_{\boldsymbol{\eta}_{\mathrm{T}}}^{-1} \boldsymbol{t} \right], \tag{35}$$

that is, $\boldsymbol{t}_{\mathrm{opt}}$ is the eigenvector of $\boldsymbol{C}_{\boldsymbol{\eta}_{\mathrm{T}}}$ corresponding to the smallest eigenvalue of $\boldsymbol{C}_{\boldsymbol{\eta}_{\mathrm{T}}}$, i.e., of $\boldsymbol{C}_{\boldsymbol{\eta},\mathrm{time}}$. In [16], based on findings from [2], the pilot symbols are chosen so as to minimize the channel estimation error variance. As a consequence of this and several other system assumptions, the optimal $\boldsymbol{T}$ is found via a waterfilling algorithm and the authors show that $\boldsymbol{T}$ is tall (in constrast to our model, which assumes that $\boldsymbol{T}$ is broad). However, the above result (35) suggests that the approach from [16] of minimizing the channel estimation MSE may be suboptimal in terms of maximizing $\underline{\mathcal{I}}_{\mathrm{G}}$.

Références

[1] T. Yoo and A. Goldsmith, “Capacity and Power Allocation for Fading MIMO Channels With Channel Estimation Error,” *IEEE Transactions on Information Theory*, vol. 52, no. 5, pp. 2203–2214, May 2006.

[2] B. Hassibi and B. M. Hochwald, “How Much Training is Needed in Multiple-Antenna Wireless Links?” *IEEE Transactions on Information Theory*, vol. 49, no. 4, pp. 951–963, April 2003.

[3] J. Baltersee, G. Fock, and H. Meyr, “Achievable Rate of MIMO Channels with Data-Aided Channel Estimation and Perfect Interleaving,” *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 12, pp. 2358–2368, Dec 2001.

[4] I. E. Telatar, “Capacity of Multi-antenna Gaussian Channels,” *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–595, November/December 1999.

[5] G. J. Foschini and M. J. Gans, “On Limits of Wireless Communications in a Fading Environment When Using Multiple Antennas,” *Wireless Personal Communications*, vol. 6, no. 3, pp. 311–335, March 1998.

[6] W. Yu, W. Rhee, S. Boyd, and J. M. Cioffi, “Iterative Water-Filling for Gaussian Vector Multiple-Access Channels,” *IEEE Transactions on Information Theory*, vol. 50, no. 1, pp. 145–152, January 2004.

[7] P. Viswanath and D. N. C. Tse, “Sum Capacity of the Vector Gaussian Broadcast Channel and Uplink-Downlink Duality,” *IEEE Transactions on Information Theory*, vol. 49, no. 8, pp. 1912–1921, August 2003.

[8] S. Vishwanath, N. Jindal, and A. Goldsmith, “Duality, Achievable Rates, and Sum-Rate Capacity of Gaussian MIMO Broadcast Channels,” *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2658–2668, October 2003.

[9] M. Schubert and H. Boche, “Iterative Multiuser Uplink and Downlink Beamforming Under SINR Constraints,” *IEEE Transactions on Signal Processing*, vol. 53, no. 7, pp. 2324–2334, July 2005.

[10] R. Hunger and M. Joham, “A General Rate Duality of the MIMO Multiple Access Channel and the MIMO Broadcast Channel,” in *Proc. 51st IEEE Global Communications Conference (IEEE Globecom 2008)*, November 2008.

[11] H. Weingarten, Y. Steinberg, and S. Shamai (Shitz), “The Capacity Region of the Gaussian Multiple-Input Multiple-Output Broadcast Channel,” *IEEE Transactions on Information Theory*, vol. 52, no. 9, pp. 3936–3964, September 2006.

[12] N. Jindal, W. Rhee, S. Vishwanath, S. A. Jafar, and A. Goldsmith, “Sum Power Iterative Water-Filling for Multi-Antenna Gaussian Broadcast Channels,” *IEEE Transactions on Information Theory*, vol. 51, no. 4, pp. 1570–1580, April 2005.

[13] R. Hunger, D. A. Schmidt, and W. Utschick, “Sum-Capacity and MMSE for the MIMO Broadcast Channel without Eigenvalue Decompositions,” in *Proc. ISIT 2007*, June 2007.

[14] S. Jafar and A. Goldsmith, “Multiple-Antenna Capacity in Correlated Rayleigh Fading with Channel Covariance Information,” *IEEE Transactions on Wireless Communications*, vol. 4, no. 3, pp. 990–997, May 2005.

[15] M. Medard, “The Effect upon Channel Capacity in Wireless Communications of Perfect and Imperfect Knowledge of the Channel,” *IEEE Transactions on Information Theory*, vol. 46, no. 3, pp. 933–946, May 2000.

[16] A. Soysal and S. Ulukus, “Optimum Power Allocation for Single-User MIMO and Multi-User MIMO-MAC with Partial CSI,” *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 7, pp. 1402–1412, September 2007.