An Upper Bound on the Capacity Loss Due to Imprecise Channel State Information for General Memoryless Fading Channels

Çağatay Candan, *Senior Member, IEEE*

***Abstract*—A remarkably simple upper bound on the capacity loss due to imprecise channel state information (CSI) is presented for single-input single-output (SISO) general memoryless fading channels, (Capacity Loss) $\leq \log(1 + \mathbf{var}(h - \widehat{h})\mathbf{SNR})$ where $\mathbf{var}(h - \widehat{h})$ represents the variance of channel estimation error, i.e. CSI inaccuracy. The bound extends earlier work on Rayleigh channels to general memoryless channels and confirms that the effect of CSI inaccuracies on the communication capacity is secondary to the receiver noise when the signal-to-noise ratio (SNR) is smaller than the reciprocal of the channel estimation error variance, SNR $< 1/\mathbf{var}(h - \widehat{h})$. An extension to the single-input multiple-output (SIMO) case is also provided.**

***Index Terms*—Ergodic capacity, capacity penalty, fading channels, channel state. Information.**

I. Introduction

THE impact of imprecise channel state information (CSI) on the capacity of single-input single-output (SISO) general memoryless fading channels with additive white Gaussian noise ($y = hx + n$) is considered. It is well known that the capacity of Rayleigh channels increases with $\log(\text{SNR})$ when CSI is perfectly known by the receiver, [1]. When CSI is not available, the capacity growth at high SNR is shown to be a doubly logarithmic function of SNR, $\log(\log(\text{SNR}))$, [2]. Recently, Lapidoth et al. have shown that the double logarithmic growth rate is not limited to the Rayleigh fading channels; but also applies to the channels with *noisy CSI* provided that the fading process and the side information on the fading process (noisy CSI) have finite second moments and differential entropy rates, [3, Theorem 4.2]. This important result indicates that for a fairly general class of fading processes, the capacity growth rate at high SNR is extremely slow when CSI is not *precisely* known. After the study of Lapidoth et al., Etkin et al. have examined the range of SNR values for which the double logarithmic growth rate is effective, [4]. Recently, the study of Etkin et al. has been compounded with the introduction of the fading number into the analysis, [5], [6]. In this letter, we revisit the impact of noisy CSI on the communication capacity and present a remarkably simple upper bound for the capacity loss due to imperfect CSI which is also known as the capacity penalty in the literature.

Lapidoth et al. have noted that the capacity loss of a flat fading channel (achieved by Gaussian codebooks and scaled nearest neighbor decoding) due to imperfect CSI can be considered as negligible, if the channel estimation error variance is much smaller than $1/\text{SNR}$, [7]. This important conclusion has been confirmed by Etkin et al. and extended to the time-varying channels having a Gauss-Markov fading structure in [4]. In this letter, we present the relation of (Capacity Penalty) $\leq \log(1 + \text{var}(h - \widehat{h})\text{SNR})$ immediately confirming and quantifying the earlier conclusions and extending the results for the Rayleigh channels to general memoryless fading channels. The presented bound, in spite of its simplicity, is also shown to be tighter than the Etkin bound, [4].

Manuscript received December 29, 2012. The associate editor coordinating the review of this letter and approving it for publication was T. Oechtering.

Ç. Candan is with the Department of Electrical and Electronics Engineering, Middle East Technical University (METU), Ankara, Turkey (e-mail: ccandan@metu.edu.tr).

Digital Object Identifier 10.1109/LCOMM.2013.060513.122930

II. Preliminaries

We consider a flat fading SISO channel with the system model of $y = hx + n$. Here, x and y represent the channel input and output, respectively. The random variable h is the fading coefficient with an arbitrary probability density function and the random variable n represents the zero-mean additive white Gaussian noise (AWGN) with the variance σ_n^2. It is assumed that the side information on h, which is shown by $\widehat{h}$, is available to the receiver. The main goal is to study the mutual information between channel input and output in the presence of side information, $I(x; y|\widehat{h})$.

To motivate the capacity bound discussed in the next section, we present a concrete communication system example for which the bound is applicable. The same set-up is also utilized in the numerical results section to compare the proposed bound with the alternatives.

The exemplary channel is considered to be block-fading. Stated differently, it is assumed that once the fading coefficient h is selected from the associated density, the coefficient remains constant for a number of symbols, [1].

The side information $\widehat{h}$ on the fading coefficient can be generated through a pilot-assisted scheme as in [8], [9]. Here, it is assumed that the value of the channel coefficient h is estimated via the transmission of a pilot and the side information $\widehat{h}$ is produced through the framework of linear minimum mean square error (LMMSE) estimation:

$$\widehat{h}_{\text{LMMSE}} = \frac{E\{h^2\}x_p^*}{E\{h^2\}|x_p|^2 + \sigma_n^2} y$$
$$\text{var}(h - \widehat{h}_{\text{LMMSE}}) = \frac{E\{h^2\}}{E\{h^2\}\text{SNR}_\text{T} + 1}. \quad (1)$$

Here, x_p denotes the pilot symbol, y denotes the observed channel output ($y = hx_p + n$) and $\text{SNR}_\text{T} = |x_p|^2/E\{n^2\}$ is the SNR during the transmission of the pilot. It should be noted that the SNR utilized in the training phase can be set higher than the data transmission SNR (operational SNR) to improve the quality of CSI.

We would like to note that the bound described in this paper has no dependence on the mechanism generating the side

1089-7798/13\$31.00 © 2013 IEEE

information $\widehat{h}$. Here, the pilot-assisted scheme is presented due to the availability of simple closed form relations for the error variance. Any other mechanism, including a genie assisted one, can also be utilized to generate $\widehat{h}$. Readers further interested in the pilot assisted scheme can also examine [10] where the impact of noisy CSI on the zero-outage capacity is studied.

III. Upper Bound On Capacity Loss

Considering the system model $y = hx + n$, the mutual information between x and (y, h) given $\widehat{h}$, $I(x; y, h|\widehat{h})$, can be written in two different ways using the chain rule:

$$\begin{aligned} I(x; y, h|\widehat{h}) &\overset{(2a)}{=} I(x; y|\widehat{h}) + I(x; h|\widehat{h}, y) \\ &\overset{(2b)}{=} I(x; h|\widehat{h}) + I(x; y|\widehat{h}, h). \end{aligned} \tag{2}$$

Since the input of the channel, which is x, is independent from other random variables; $I(x; h|\widehat{h})$, appearing in line (2b), is $I(x; h|\widehat{h}) = H(h|\widehat{h}) - H(h|\widehat{h}, x) = 0$. In addition, since the channel output y given h is independent from $\widehat{h}$; we have $I(x; y|\widehat{h}, h) = I(x; y|h)$. These two facts simplify line (2b) to $I(x; y|h)$, which is the mutual information under precise CSI. By combining lines (2a) and (2b), we get the following relation:

$$I(x; y|\widehat{h}) = I(x; y|h) - I(x; h|\widehat{h}, y). \tag{3}$$

It should be clear that $I(x; y|h)$ and $I(x; y|\widehat{h})$ are the mutual information with precise and imprecise CSI, respectively. The term $I(x; h|\widehat{h}, y)$ in (3) shows the loss in the mutual information due to imprecise CSI. Our main goal is to find an upper bound for the mentioned penalty term. The penalty term can be further processed as follows:

$$\begin{aligned} I(x; h|\widehat{h}, y) &\overset{(4a)}{=} H(h|\widehat{h}, y) - H(h|\widehat{h}, y, x) \\ &\overset{(4b)}{\leq} H(h|\widehat{h}) - H(h|\widehat{h}, y, x) \\ &\overset{(4c)}{=} H(h|\widehat{h}, x) - H(h|\widehat{h}, y, x) \\ &\overset{(4d)}{=} I(y; h|\widehat{h}, x). \end{aligned} \tag{4}$$

Here, line (4b) follows from the fact that conditioning does not increase the entropy. Line (4c) follows from the fact that the input distribution is independent from other random variables, and the last line results from the definition of mutual entropy.

The term $I(y; h|\widehat{h}, x)$ in (4) upper bounds the term $I(x; h|\widehat{h}, y)$, which is the term related with the capacity penalty. This term can be interpreted as the reduction in the uncertainty of h given $\widehat{h}$ after the observation of channel output with the knowledge of channel input. This quantity can be interpreted as the information revelation about h in addition to the side-information $\widehat{h}$ by the observation of channel output for a known input. Combining equations (3) and (4), we reach the following bound on the mutual information:

$$I(x; y|\widehat{h}) \geq I(x; y|h) - I(y; h|\widehat{h}, x). \tag{5}$$

We note that the bound given above is valid for any distribution on x.

The next step is to maximize the loss term $I(y; h|\widehat{h}, x)$ in (5). The loss term given for a fixed value of $\widehat{h} = \widehat{h}_f$ and $x = x_f$, $I(y; h|\widehat{h} = \widehat{h}_f, x = x_f)$, can be maximized by noting that the channel $y = hx + n$ becomes the standard AWGN channel for the estimation of h given $\widehat{h}_f$ and x_f upon the observation of y. It is well known that the mutual information is maximized if the random variable h given $\widehat{h}$ is Gaussian distributed, that is

$$I(y; h|\widehat{h} = \widehat{h}_f, x = x_f) \leq \log\left(1 + \frac{\sigma^2_{h|\widehat{h}_f}|x_f|^2}{\sigma^2_n}\right). \tag{6}$$

The mutual information $I(y; h|\widehat{h}, x)$ appearing in (5) can be calculated by averaging $I(y; h|\widehat{h} = \widehat{h}_f, x = x_f)$ over the joint distribution of $\widehat{h}$ and x. We present an upper bound for $I(y; h|\widehat{h}, x)$ in (7) (presented on the next page). In the derivation of (7), the bound in (6) and Jensen's inequality, $E\{\log(x)\} \leq \log(E\{x\})$, is used and in the rightmost term, the ratio of σ^2_x/σ^2_n is denoted as SNR.

If the density $f^*_x(x)$ maximizes the mutual information for the perfect CSI case, $C^{\text{Precise}}_{\text{CSI}} = \max_{f_x(x)} I(x; y|h) = I(x; y|h)\mid_{f_x(x)=f^*_x(x)}$; by substituting $f^*_x(x)$ for the input density in (5) and then using (7), we get the following:

$$\begin{aligned} I(x; y|\widehat{h})\mid_{f_x(x)=f^*_x(x)} &\geq C^{\text{Precise}}_{\text{CSI}} - I(y; h|\widehat{h}, x)\mid_{f_x(x)=f^*_x(x)} \\ &\geq C^{\text{Precise}}_{\text{CSI}} - \log\left(1 + E_{\widehat{h}}\left\{\sigma^2_{h|\widehat{h}}\right\}\text{SNR}\right). \end{aligned}$$

Since $C^{\text{Imprecise}}_{\text{CSI}} \geq I(x; y|\widehat{h})\mid_{f_x(x)=f^*_x(x)}$, where $C^{\text{Imprecise}}_{\text{CSI}} = \max_{f_x(x)} I(x; y|\widehat{h})$; we immediately have

$$C^{\text{Imprecise}}_{\text{CSI}} \geq C^{\text{Precise}}_{\text{CSI}} - \log\left(1 + E_{\widehat{h}}\left\{\sigma^2_{h|\widehat{h}}\right\}\text{SNR}\right). \tag{8}$$

The rightmost term in (8), $\log(1 + E_{\widehat{h}}\{\sigma^2_{h|\widehat{h}}\}\text{SNR})$, is the desired capacity penalty term. The penalty term can be further simplified, at the peril of making the bound looser, via the conditional variance formula, $\text{var}(y) = E\{\text{var}(y|x)\} + \text{var}(E\{y|x\})$:

$$\begin{aligned} E_{\widehat{h}}\{\sigma^2_{h|\widehat{h}}\} &\overset{(9a)}{=} \text{var}(h) - \text{var}\left(E\{h|\widehat{h}\}\right) \\ &\overset{(9b)}{=} \text{var}\left(h - \widehat{\widehat{h}}_{\text{MMSE}}(\widehat{h})\right) \\ &\overset{(9c)}{\leq} \text{var}\,(h - \widehat{h}). \end{aligned} \tag{9}$$

In line (9a), $\widehat{\widehat{h}}_{\text{MMSE}}(\widehat{h}) = E\{h|\widehat{h}\}$ is the minimum mean square error (MMSE) estimate of h given the side-information $\widehat{h}$. Line (9b) follows from the orthogonality property of MMSE estimators. Line (9c) follows from the fact that $\widehat{h}$ is an estimate of h without any optimality properties and can not do better than $\widehat{\widehat{h}}_{\text{MMSE}}(\widehat{h})$ in the mean square sense.

Combining (8) and (9), we get the following lower bound on the channel capacity

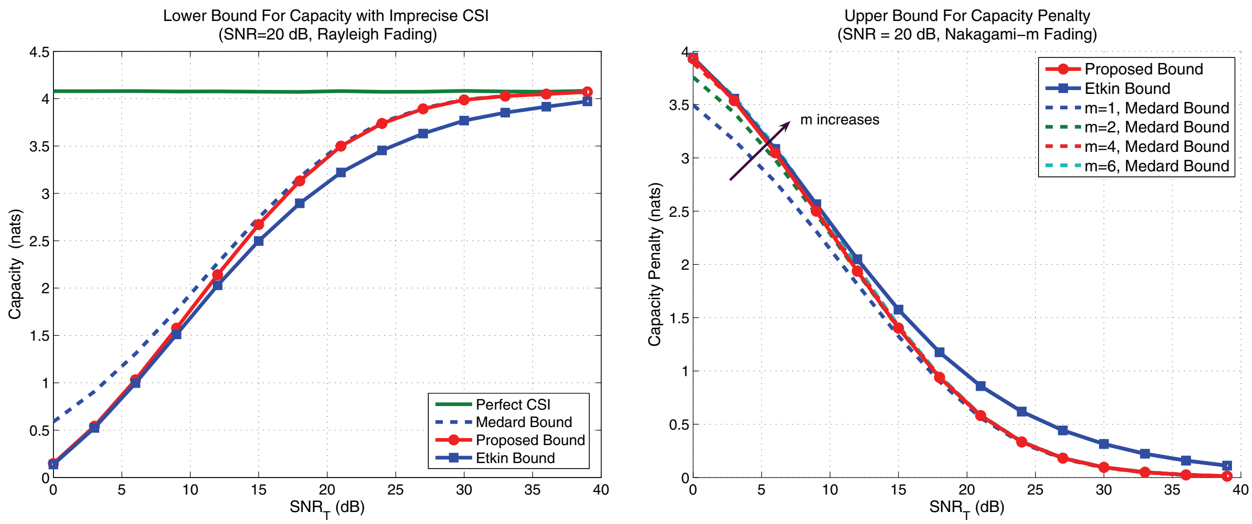
$$C^{\text{Imprecise}}_{\text{CSI}} \geq C^{\text{Precise}}_{\text{CSI}} - \log\left(1 + \text{var}(h - \widehat{h})\text{SNR}\right). \tag{10}$$

The proposed upper bound for the capacity penalty can also be expressed as follows:

$$C^{\text{Proposed}}_{\text{Penalty}} \leq \log\left(1 + \text{var}(h - \widehat{h})\text{SNR}\right). \tag{11}$$

Extension to SIMO Channels: We assume a flat fading system with a single transmitter and $R \geq 1$ receiving antennas. The system model can be written as $\mathbf{y} = \mathbf{h}x + \mathbf{n}$ where $\mathbf{y}$ is

$$I(y;h|\widehat{h},x) \le E_{x,\widehat{h}}\left\{\log\left(1+\frac{\sigma^2_{h|\widehat{h}}|x|^2}{\sigma_n^2}\right)\right\} \le \log\left(1+\frac{E_{\widehat{h}}\left\{\sigma^2_{h|\widehat{h}}\right\}\sigma_x^2}{\sigma_n^2}\right) = \log\left(1+E_{\widehat{h}}\left\{\sigma^2_{h|\widehat{h}}\right\}\mathrm{SNR}\right) \quad (7)$$



(a) Lower Bound for Capacity (b) Upper Bound for Capacity Penalty

Fig. 1. Figure 1(a) compares the capacity lower bounds for the Rayleigh channel. Figure 1(b) shows the capacity penalty upper bounds for Nakagami-m channels.

$R \times 1$ column vector denoting the output of each antenna at the receiver. The elements of vector $\mathbf{h}$ denotes the path gain between the transmitter and antennas, $\mathbf{n}$ is a zero mean complex Gaussian vector with covariance matrix $\mathbf{C_n}$. The estimate for path gain vector is denoted with $\widehat{\mathbf{h}}$.

The equations (2) to (5) exactly apply to the SIMO system with the replacement of $y \to \mathbf{y}$, $h \to \mathbf{h}$ and $\widehat{h} \to \widehat{\mathbf{h}}$. The loss term becomes $I(x;\mathbf{y}|\mathbf{h}) - I(\mathbf{y};\mathbf{h}|\widehat{\mathbf{h}},x)$. Similar to the scalar case, the loss term is maximized with the Gaussian distributed $\mathbf{h}$ given $\widehat{\mathbf{h}}$. The maximum value attained for a fixed $\widehat{\mathbf{h}}$ and x is $\log\left(\left|\mathbf{C_n}^{-1}\mathbf{C}_{\mathbf{h}|\widehat{\mathbf{h}}}|x|^2+\mathbf{I}\right|\right)$, where $\mathbf{I}$ is the identity matrix. By averaging this result over $\widehat{\mathbf{h}}$ and x, as in (7), we can get the desired result. As in the scalar case, the Jensen's inequality can be applied to upper bound the result of averaging, since $\log(\det(\cdot))$ is a concave function, [11, p.74]. Once Jensen's inequality is applied, the loss term can be expressed as $\log\left(\left|\mathbf{C_n}^{-1}E\{\mathbf{C}_{\mathbf{h}|\widehat{\mathbf{h}}}\}\sigma_x^2+\mathbf{I}\right|\right)$ in complete analogy with (7). Finally, it is possible to show that $E\{\mathbf{C}_{\mathbf{h}|\widehat{\mathbf{h}}}\} \le \mathrm{Cov}(\mathbf{h}-\widehat{\mathbf{h}})$ by adapting the steps given in (9) for scalar random variables to the random vectors and we reach the following bound for the SIMO system:

$$C_{\mathrm{CSI}}^{\mathrm{Imprecise}} \ge C_{\mathrm{CSI}}^{\mathrm{Precise}} - \log\left(\left|\mathbf{I}+\mathbf{C_n}^{-1}\mathrm{Cov}(\mathbf{h}-\widehat{\mathbf{h}})\sigma_x^2\right|\right). \quad (12)$$

Note that when the receiver noise is white, $\mathbf{C_n} = \sigma_n^2\mathbf{I}$, the loss term becomes $\log\left(\left|\mathbf{I}+\mathrm{Cov}(\mathbf{h}-\widehat{\mathbf{h}})\mathrm{SNR}\right|\right)$, where $\mathrm{SNR} = \sigma_x^2/\sigma_n^2$ is the SNR at each antenna. The result for the SISO case is immediately retrieved by setting $R = 1$ in the last expression.

Comments: The suggested relation for the capacity penalty is closely related to the one given by Etkin et al. for Rayleigh fading MIMO channels evolving according to the first order Gauss-Markov process, [4, Eq. 8]. It can be noted that the expression for the capacity penalty given by Etkin et al. is significantly more complicated than the proposed one. Yet, as shown in the numerical results section, the presented bound captures the behaviour of the Etkin bound and also presents some improvements over this bound. It should be noted that the presented bound is most useful for the regime 1 operation described in [4]. More specifically, the presented bound is useful when SNR is much smaller than the double exponential of the fading number χ, i.e. $\mathrm{SNR} \ll \exp(\exp(\chi))$, [5, Sec. V]. The presented bound is not sophisticated enough to capture the behaviour in other regimes where the capacity is dominated either by the fading number or the term $\log(\log(\mathrm{SNR}))$, [6].

IV. Numerical Results

We consider a flat fading channel $y = \sqrt{\mathrm{SNR}}\,hx + n$ with an arbitrary density for the fading coefficient h. The noise n is circularly symmetric complex Gaussian distributed with zero mean and unit variance. Without any loss of generality, $E\{|h|^2\} = E\{|x|^2\}$ is assumed to be unity. In this section, $\mathrm{SNR_T}$ represents the SNR in the training phase of the pilot-assisted system described in the preliminaries section.

We note that when LMMSE is utilized to produce the estimate $\widehat{h}$ as in (1), $\mathrm{var}(h - \widehat{h}_{\mathrm{LMMSE}})$ becomes simply $1/(\mathrm{SNR_T}+1)$. Then, the capacity penalty due to imprecise knowledge of h can be immediately written as follows:

$$C_{\mathrm{Penalty}}^{\mathrm{Proposed}} \le \log\left(1+\frac{\mathrm{SNR}}{\mathrm{SNR_T}+1}\right). \quad (13)$$

Here, SNR and $\mathrm{SNR_T}$ refer to the signal-to-noise ratio for the data transmission and training, respectively. It should be noted that the LMMSE estimator depends only on the second order moments of the fading coefficient, i.e. has no dependence on its distribution; hence, the presented bound for the capacity loss is valid for all fading distributions.

When the bound proposed by Etkin et al. is adapted to the examined problem by setting $\epsilon = 1/(\text{SNR}_\text{T}+1)$ and $n_r = n_t = 1$ in [4, Eq. 8], we have the following result:

$$C^{\text{Etkin}}_{\text{Penalty}} \leq \log\left(\Upsilon + \sqrt{\Upsilon^2 - \text{SNR}_\text{T}/(\text{SNR}_\text{T}+1)}\right) \qquad (14)$$

where $2\Upsilon = 1 + \frac{\text{SNR}}{\text{SNR}_\text{T}+1} + \frac{\text{SNR}_\text{T}}{\text{SNR}_\text{T}+1}$. It can be easily verified that $C^{\text{Proposed}}_{\text{Penalty}} \leq C^{\text{Etkin}}_{\text{Penalty}}$. As $\text{SNR}_\text{T} \rightarrow \infty$, the Etkin bound approaches the proposed bound.

A related bound for the capacity with imprecise CSI is given by Medard in [12, Eq. 46]:

$$C^{\text{Imprecise}}_{\text{CSI,Medard}} \geq E_{\widehat{h}}\left\{\log\left(1 + \frac{\widehat{h}\,\text{SNR}}{\text{var}(h-\widehat{h})\text{SNR}+1}\right)\right\}. \qquad (15)$$

It should be noted that the Medard bound is a lower bound for the capacity. The penalty term is not immediately available for this bound. Different from other bounds, the Medard bound depends on the fading distribution, requires Monte-Carlo integration for its calculation.

Figure 1(a) compares the bounds for the Rayleigh fading channel (or equivalently Nakagami-m channel with $m = 1$) at SNR = 20 dB as the training SNR_T varies from 0 to 40 dB. The capacity for perfect CSI is also provided. It can be noted from Figure 1(a) that the proposed bound presents some improvements over the Etkin bound for all SNR_T values. The bound is poorer than the Medard bound for $\text{SNR}_\text{T} < \text{SNR}$; but it is almost identical to the Medard bound for $\text{SNR}_\text{T} > \text{SNR}$.

Figure 1(b) repeats the same comparison for Nakagami-m fading channels for $m = \{1, 2, 4, 6\}$. Different from Figure 1(a), the bounds for the capacity penalty are presented. As noted before, the Medard bound (dashed lines) depends on the fading distribution and the other bounds have no dependency on the fading distribution. It can noted from Figure 1(b) that the proposed bound is identical to the Medard bound for higher order Nakagami channels and consistently better than Etkin bound. (Similar results are received for Rician channels.)

It is indeed surprising that even though both the capacity with the precise CSI and the Medard bound is difficult to analytically express, their difference (capacity penalty) is well characterized with the proposed relation, at least for the standard fading distributions. In addition, the suggested bound is shown to present some improvements over the Etkin bound. Yet we consider the main contribution of the present work as the revelation of a remarkably simple relation for the capacity penalty.

V. Conclusions

A simple bound for the capacity loss due to the imperfections in CSI is derived. The bound is in accord with the conclusions of Lapidoth et al. and Etkin et al. stating that the capacity is limited by the receiver noise (not by the channel estimation errors) when the channel estimation error variance is smaller than the reciprocal of SNR, [4], [7]. To the best of our knowledge, the proposed bound is the simplest one confirming the stated conclusion and in spite of its simplicity, the bound is shown to be tighter than the Etkin bound and comparable with the Medard bound which does not lend itself to analysis.

References

[1] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.

[2] G. Taricco and M. Elia, "Capacity of fading channel with no side information," *Electron. Lett.*, vol. 33, no. 16, pp. 1368–1370, July 1997.

[3] 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, no. 10, pp. 2426–2467, Oct. 2003.

[4] R. Etkin and D. Tse, "Degrees of freedom in some underspread MIMO fading channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1576–1608, Apr. 2006.

[5] T. Koch and A. Lapidoth, "The fading number and degrees of freedom in non-coherent MIMO fading channels: a peace pipe," in *Proc. 2005 IEEE Int. Symp. Inf. Theory*, pp. 661–665.

[6] S. Moser, "The fading number of multiple-input multiple-output fading channels with memory," *IEEE Trans. Inf. Theory*, vol. 55, no. 6, pp. 2716–2755, June 2009.

[7] A. Lapidoth and S. Shamai, "Fading channels: how perfect need "perfect side information" be?" *IEEE Trans. Inf. Theory*, vol. 48, no. 5, pp. 1118–1134, May 2002.

[8] P. Dayal, M. Brehler, and M. Varanasi, "Leveraging coherent space-time codes for noncoherent communication via training," *IEEE Trans. Inf. Theory*, vol. 50, no. 9, pp. 2058–2080, Sept. 2004.

[9] A. Asyhari, T. Koch, and A. Guillen i Fabregas, "Nearest neighbour decoding and pilot-aided channel estimation in stationary Gaussian flat-fading channels," in *Proc. 2011 IEEE Int. Symp. Inf. Theory*, pp. 2786–2790.

[10] C. Candan, "Capacity of zero-outage scheme under imprecise channel state information," *IEEE Commun. Lett.*, vol. 17, no. 1, pp. 127–130, Jan. 2013.

[11] S. Boyd, *Convex Optimization*. Cambridge University Press, 2004.

[12] M. Medard, "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.