BOUNDS ON THE CAPACITY OF MISO CHANNELS WITH DIFFERENT TYPES OF IMPERFECT CSI

Mario Castañeda, Amine Mezghani and Josef A. Nossek

Lehrstuhl für Netzwerktheorie und Signalverarbeitung
Technische Universität München, Arcisstraße 21, 80333 Munich
{castaneda,mezghani,nossek}@nws.ei.tum.de

ABSTRACT

Consider the downlink of an isolated cell with a *base station* (BS) equipped with M multiple antennas and a single user with a single antenna. For this *multiple-input single-output* (MISO) channel, the capacity with perfect *channel state information* (CSI) the transmitter and receiver is known. In practice, however, it is actually more realistic to assume *imperfect* CSI at the transmitter. For instance, in a *frequency divsion duplex* (FDD) system with limited feedback, the available transmit CSI is *estimated, quantized, outdated* and affected by *erroneous* feedback. The quality of the transmit CSI then depends on the downlink training length, the number of feedback bits, the feedback error probability and on the rate of variation of the downlink channel. However, the capacity of the MISO channel with imperfect CSI in general is unknown. In this work, nonetheless, we compute bounds of the MISO channel capacity employing *coherent beamforming* at the BS with different types of imperfect CSI at the transmitter.

1. INTRODUCTION

Multiple transmit and receive antennas increase substantially the capacity of wireless communication links but under the assumption of perfect channel knowledge at the transmitter and receiver. In practice, however, the channel knowledge available at the transmitter and receiver is not perfect. For instance, the receiver can obtain a *minimum mean square error* (MMSE) estimate of the channel through training [1], and hence, we assume the available receive CSI to be *estimated*. In case of a *time division duplex* (TDD) system, the BS can employ the receive CSI from the uplink by making use of the channel reciprocity, such that the available transmit CSI is *estimated* and possibly outdated. Then, the quality of the transmit CSI depends on the training length employed in the uplink to estimate the channel and on the rate of variation of the channel.

In an FDD system, the available receive CSI in the uplink cannot be employed as transmit CSI in the downlink, since we assume the downlink and uplink channels to be uncorrelated. In this case, we assume the transmitter obtains partial knowledge of the downlink channel through *limited* feedback of B bits in the uplink [2]. To this end, the receiver in the downlink, i.e., the user, first estimates the downlink channel and then *quantizes* the downlink channel estimate with B bits, which are relayed back to the BS. For this, we employ the *random vector quantization* (RVQ) scheme [2], where we have at the transmitter (BS) and receiver (user) a codebook with 2^B random beamforming vectors i.i.d over the M-dimensional unit sphere. Assuming a faded- and noise-prone feedback link, i.e., the uplink, the B relayed bits could be received *erroneously* at the BS, which reduces the quality of the available CSI at the transmitter. Besides the previous issues, the transmit CSI is further degraded due to the delay incurred in the feedback process, i.e., the partial knowledge the BS has obtained at a given time slot can only be employed at a later time slot, when the channel could have changed. Hence, the transmit CSI is also *outdated* under the assumption of a time-varying channel. In this way, the transmit CSI available at the BS for the downlink transmission is *estimated, quantized, outdated* and affected by *erroneous* feedback. Such a characterization of the imperfect transmit CSI in FDD systems has also been assumed in [3], but has not been considered in general in the literature.

The capacity with imperfect CSI, such as the different types described above, is unknown in general and in such cases, one can instead recur to the computation of bounds on the capacity. In this work, we present bounds on the capacity of a single-user MISO channel with *coherent beamforming* at the BS in an FDD system with imperfect transmit CSI, which is estimated, quantized, outdated and affected by erroneous feedback, as a function of the downlink training length, the number of feedback bits B, the feedback error probability in the uplink and the rate of variation of the downlink channel. Additionally, we compute bounds on the capacity in a TDD system with estimated transmit CSI as a function of the uplink training length and of the rate of variation of the channel. As a reference we also include the capacities with perfect CSI and without CSI at the transmitter. To this end, this paper is organized as follows. In Section 2, the system model and the types of imperfect CSI are discussed. The capacity bounds are derived in Section 3. Section 4 presents numerical results and finally, Section 5 concludes the paper.

2. SYSTEM MODEL WITH IMPERFECT CSI

We assume that the downlink channel is constant for T symbols, which is the duration of a time slot. The first T_{DL} symbols are used to obtain an MMSE estimate of the downlink channel and the remaining $D = T - T_{\text{DL}}$ symbols are used to transmit information from the BS to the user. With coherent beamforming at the BS, the equivalent *single-input single-output* (SISO) downlink system of the D data symbols at time slot n is given by

$$\mathbf{y}[n] = \sqrt{P}\,\mathbf{w}_{\text{BS}}^{\text{H}}[n]\mathbf{h}[n]\mathbf{r}[n] + \mathbf{v}[n], \tag{1}$$

where $\mathbf{y}[n] \in \mathbb{C}^D$ are the received signals, $\mathbf{r}[n] \in \mathbb{C}^D$ are the transmit symbols with unit variance, $\mathbf{w}_{\text{BS}}[n] \in \mathbb{C}^M$ is the beamforming vector with unit norm, $\mathbf{h}[n] \in \mathbb{C}^M$ is the downlink MISO channel and $\mathbf{v}[n] \in \mathbb{C}^D$ is the *additive white Gaussian noise* (AWGN) with zero mean and variance σ_v^2 at time slot n. Additionally, P is the transmit power available at the BS. We assume Rayleigh fading, i.e. the elements of the channel vector $\mathbf{h}[n]$ are i.i.d. complex Gaussian random variables with zero mean and unit variance. Contrary to the common block fading assumption found in the literature, in this paper we consider *temporally correlated block fading*, i.e. $\mathbf{h}[n]$ is assumed to be constant for the coherence time of T symbols and is correlated with the channel in the previous time slot $\mathbf{h}[n-1]$ according to a first order Markov model:

$$\mathbf{h}[n] = \sqrt{\alpha}\,\mathbf{h}[n-1] + \sqrt{1-\alpha}\,\mathbf{g}[n-1], \tag{2}$$

where the elements of $\mathbf{g}[n-1] \in \mathbb{C}^M$ are i.i.d. zero-mean unit-variance complex Gaussian random variables and are uncorrelated with $\mathbf{h}[n-1]$ and $\sqrt{\alpha}$ is the correlation coefficent, with $\alpha \in [0,1]$. For instance, with a Jake's spectrum of the Rayleigh fading we would have that $\sqrt{\alpha} = \text{J}_0(2\pi T f_{\text{c}} s/c)$, where J_0 is the zeroth-order Bessel function of the first kind, f_{c} is the carrier frequency, c is the speed of light and s is the radial component of the velocity of the user, along a line from the BS to the user. Hence, α is a function of the velocity of the user. However, note that the autocorrelation function of the fading process with a Jake's spectrum cannot be modeled by a first order Markov model and it has solely been shown here to present a possible relation between α and the user's velocity. Also, we assume that α is unknown at the transmitter. In the following, we discuss different types of imperfect CSI in the context of the downlink of an FDD system and explain how the beamforming vector $\mathbf{w}_{\text{BS}}[n]$ is determined.

2.1. Estimated CSI

At each time slot, T_{DL} pilot symbols are employed to obtain an MMSE estimate $\hat{\mathbf{h}}$ of the downlink channel. Due to practical limitations, we assume an even power distribution, i.e. P, over the training and data payload phase. Based on the downlink channel estimate $\hat{\mathbf{h}}[n]$, the downlink channel $\mathbf{h}[n]$ can be written as

$$\mathbf{h}[n] = \hat{\mathbf{h}}[n] + \mathbf{e}[n], \tag{3}$$

where $\mathbf{e}[n]$ is the error vector whose elements are i.i.d. zero-mean complex Gaussian random variables with variance σ_e^2 given by [1]

$$\sigma_e^2 = \begin{cases} \dfrac{1+\rho(M-T_{\text{DL}})}{1+\rho M} & \text{for} \quad T_{\text{DL}} < M \\ \dfrac{1}{1+\rho T_{\text{DL}}} & \text{for} \quad T_{\text{DL}} \geq M \end{cases}, \tag{4}$$

where $\rho = \frac{P}{M\sigma_v^2}$. Additionally, the elements of the MMSE estimate $\hat{\mathbf{h}}$ are i.i.d. zero-mean Gaussian random variables with variance $(1-\sigma_e^2)$. Note that the quality of the estimated CSI depends on the training length T_{DL}. As the training length increases, i.e. $T_{\text{DL}} \to T$ the estimated CSI becomes closer to perfect CSI, i.e., $\sigma_e^2 \to 0$. Nonetheless, $D \to 0$ and there will be less symbols available to transmit data, thus reducing the downlink capacity. We then have a tradeoff between the training length and the data payload length as discussed in [1].

Remark 2.1: Estimation in TDD Systems. Consider a TDD system where the uplink utilizes the channel at time slot $n-1$ and the downlink makes use of the channel at time slot n. We assume that the BS does not know α and hence the BS employs the estimate of the uplink channel as transmit CSI for the downlink transmission. Denoting the uplink channel estimate at time slot $n-1$ as $\hat{\mathbf{h}}_{\text{UL}}[n-1]$, we have that the employed beamforming vector in the downlink is $\mathbf{w}_{\text{BS}}[n] = \frac{\hat{\mathbf{h}}_{\text{UL}}[n-1]}{\|\hat{\mathbf{h}}_{\text{UL}}[n-1]\|_2}$. However note that estimating the uplink *single-input multiple-output* (SIMO) channel is not exactly the same as estimating the MISO downlink channel. The uplink channels from the user to each antenna can be estimated independently at each antenna of the BS, since the M channels are orthogonal in space. This is contrast to the downlink, where at the single receiving antenna of the user we need to estimate M channels. Hence, the variance of the estimation error of $\hat{\mathbf{h}}_{\text{UL}}[n]$ and in turn of $\hat{\mathbf{h}}[n]$ in a TDD system is given by $\sigma_{e_{\text{TDD}}}^2 = \frac{1}{1+\frac{P_{\text{UL}}}{\sigma_u^2}T_{\text{UL}}}$, where $P_{\text{UL}}, \sigma_u^2$ and T_{UL} are the transmit power of the user in the uplink, the noise variance at the BS in the uplink and the uplink training length, respectively. Note, however, that only uplink resources are consumed for the estimation of the channel and in turn to obtain $\mathbf{w}_{\text{BS}}[n]$. Nevertheless when the $\mathbf{w}_{\text{BS}}[n]$ is employed in the downlink the channel could have already changed and hence the estimated CSI is also outdated.

2.2. Quantized CSI

The estimated channel is first available only at the receiver and in order to provide the transmitter with knowledge about the current channel in an FDD system, the estimated channel must be quantized and then relayed back to the BS with the limited feedback of B bits. For this, we employ the RVQ scheme, where we have at the transmitter (BS) and receiver (user) a codebook with 2^B random beamforming vectors $\mathbf{t}_j$, $j = 1, \ldots, 2^B$, i.i.d isotropically distributed over the M-dimensional unit sphere. For instance, the channel estimate

$\hat{\mathbf{h}}[n-1]$ is quantized by selecting the beamforming vector $\mathbf{w}[n]$ (to be used at time slot n) that best matches $\hat{\mathbf{h}}[n-1]$ [4]

$$\mathbf{w}[n] = \underset{\mathbf{t}_j}{\text{argmax}} \ |\mathbf{t}_j^{\text{H}}\hat{\mathbf{h}}[n-1]|^2. \tag{5}$$

Note that we have written the beamforming vector $\mathbf{w}$ without the subscript BS as $\mathbf{w}_{\text{BS}}$ shown in (1). The reason behind this will explained in the next subsection. The user feeds back the B bits representing the index of the beamforming vector $\mathbf{w}[n]$ in the codebook. The loss due to quantization depends on the number of feedback bits B. As B increases, the quantization error becomes smaller, but this comes at the expense of the consumption of uplink resources accompanied with a higher feedback error probability as it will be seen next.

2.3. Erroneous CSI

The feedback channel in an FDD system is the uplink, which is not error-free and so, the relayed CSI can be received erroneously. If at least one feedback symbol is in error, we assume there is a *total* loss of the feedback as in [3], such that the index received by the BS corresponds to a different beamforming vector than the one intended by the user. We further assume no error detection scheme for the feedback and hence, after erroneous feedback the beamforming vector applied by the BS, i.e., $\mathbf{w}_{\text{BS}}[n]$, would be different than the beamforming vector $\mathbf{w}[n]$, which was fedback by the user (c.f. (5)). In such a case, the beamforming vector applied by the BS would be completely *uncorrelated* with the actual downlink channel. Let us denote the probability of erroneous feedback, i.e., that $\mathbf{w}_{\text{BS}}[n] \neq \mathbf{w}[n]$, as p_ϵ. We do not discuss how p_ϵ is computed and simply assume it is given. For instance, assuming that the feedback bits are sent uncoded with QPSK symbols similar to [3] with a symbol error probability p_{S} in an unfaded uplink, we have that the feedback error probability is given as $p_\epsilon = 1-(1-p_{\text{S}})^{\frac{B}{2}}$. In this case, p_ϵ depends on the symbol error probability and on the number of feedback bits, i.e. on the quantization of the downlink channel estimate. A larger B would lead to a smaller quantization error, but also to a higher p_ϵ and hence, we have a tradeoff between the quantization loss and the feedback error probability. Furthermore, the probability p_{S} and of course p_ϵ depend on the uplink *signal to noise ratio* (SNR) per receive antenna and on the uplink training length [6].

2.4. Outdated CSI

Due to delay involved in the feedback process, the available CSI at the BS is outdated in a time-varying downlink channel. Between downlink channel estimates of successive time slots, a first order Markov model also holds similar to (2):

$$\hat{\mathbf{h}}[n] = \sqrt{\alpha'}\,\hat{\mathbf{h}}[n-1] + \sqrt{1-\alpha'}\,\hat{\mathbf{g}}[n-1], \tag{6}$$

where $\sqrt{\alpha'}$ is the correlation coefficient between channel estimates and the elements of $\hat{\mathbf{g}} \in \mathbb{C}^M$ are i.i.d. zero-mean complex Gaussian random variables with variance $1-\sigma_e^2$ and are uncorrelated with $\hat{\mathbf{h}}[n-1]$. To compute α', we need

$$\begin{aligned}
\text{E}\Big[\hat{\mathbf{h}}[n]\hat{\mathbf{h}}^{\text{H}}[n-1]\Big] &= \text{E}\Big[\text{E}\Big[\hat{\mathbf{h}}[n]\hat{\mathbf{h}}^{\text{H}}[n-1] \mid \mathbf{h}[n],\mathbf{h}[n-1]\Big]\Big] \\
&= \text{E}\left[\text{E}\left[\hat{\mathbf{h}}[n]\,|\mathbf{h}[n]\right]\text{E}\left[\hat{\mathbf{h}}^{\text{H}}[n-1]\,|\mathbf{h}[n-1]\right]\right] \\
&= \text{E}\left[(1-\sigma_e^2)\mathbf{h}[n](1-\sigma_e^2)\mathbf{h}^{\text{H}}[n-1]\right] \\
&= \sqrt{\alpha}(1-\sigma_e^2)^2\mathbf{1}_M,
\end{aligned} \tag{7}$$

where $\mathbf{1}_M$ is an $M \times M$ identity matrix and the first step follows from the fact the estimates at different time slots conditioned on the actual channels are independent and the last step from (2). By definition, the correlation coefficient $\sqrt{\alpha'}$ is computed by dividing the covariance $\sqrt{\alpha}(1-\sigma_e^2)^2$ with the product of the standard deviation of the elements of $\hat{\mathbf{h}}[n]$, i.e. $\sqrt{1-\sigma_e^2}$, and the standard deviation of the elements of $\hat{\mathbf{h}}[n]$, i.e. $\sqrt{1-\sigma_e^2}$. Hence, we have that $\sqrt{\alpha'} = \sqrt{\alpha}(1-\sigma_e^2)$. Note that the correlation coefficient between the channel estimates depends not only on the correlation coefficient $\sqrt{\alpha}$ between the true channels but also on the estimation error. The reason for this can be explained with the following example. Even if the channels remain constant, i.e., $\alpha = 1$, such that $\mathbf{h}[n] = \mathbf{h}[n-1]$, this does not mean that $\hat{\mathbf{h}}[n] = \hat{\mathbf{h}}[n-1]$, since at each time slot there are different noise realizations!

3. CAPACITY BOUNDS WITH IMPERFECT CSI

The capacity with imperfect CSI is unknown and hence, we compute bounds on the MISO channel capacity with the discussed imperfect CSI. Let us first rewrite (1) using (3)

$$\begin{aligned}
\mathbf{y}[n] &= \sqrt{P}\,\mathbf{w}_{\text{BS}}^{\text{H}}[n]\left(\hat{\mathbf{h}}[n]+\mathbf{e}[n]\right)\mathbf{r}[n]+\mathbf{v}[n] \\
&= \sqrt{P}\,\mathbf{w}_{\text{BS}}^{\text{H}}[n]\hat{\mathbf{h}}[n]\mathbf{r}[n]+\mathbf{z}[n].
\end{aligned} \tag{8}$$

where $\mathbf{z}[n] = \sqrt{P}\,\mathbf{w}^{\text{H}}[n]\mathbf{e}[n]\mathbf{r}[n]+\mathbf{v}[n]$ is the effective noise. The elements of $\mathbf{z}[n]$ are not necessarily Gaussian distributed and can be correlated with the signal $\mathbf{r}[n]$[1]. The variance of the elements of $\mathbf{z}[n]$ is

$$\sigma_z^2 = \sigma_e^2 P + \sigma_v^2. \tag{9}$$

Recall that the first T_{DL} symbols in each time slot are used to compute a downlink channel estimate at the user, i.e., $\hat{\mathbf{h}}$. Therefore, the user knows the actual channel estimate $\hat{\mathbf{h}}[n]$ for the data detection in the last $T-T_{\text{DL}}$ symbols of the time slot. Additionally, at time slot n the user *expects* that the BS will be performing beamforming with $\mathbf{w}[n]$ (c.f. (5)), whose index was fedback from the user to the BS at time slot $n-1$. This will be the case if there was no feedback error and then $\mathbf{w}_{\text{BS}}[n] = \mathbf{w}[n]$, such that the user's expectation about the beamforming vector is correct. However, if the feedback was received with errors then $\mathbf{w}_{\text{BS}}[n] \neq \mathbf{w}[n]$ and the user's assumption about the beamforming vector is wrong! The user,

[1] If the $\mathbf{r}[n]$ are QPSK symbols, then $\mathbf{z}[n]$ is uncorrelated with $\mathbf{r}[n]$.

nevertheless, is not aware if a feedback error has occurred or not. But since the feedback error probability p_ϵ, for instance, depends only on the uplink SNR, which is a long term parameter, and on the uplink training length and B, which are fixed, we have that p_ϵ remains constant over many transmission blocks. Hence, even if the user is not aware of the feedback errors, it can be assumed that p_ϵ is known at the user.

3.1. Capacity Lower Bound with Imperfect CSI

In order to compute a lower bound of the capacity C with imperfect CSI we assume that transmitted symbols $\mathbf{r}[n]$ are Gaussian distributed, which is not necessary capacity achieving in this case. Let us denote the feedback error event by the binary random variable ϵ, where $\epsilon[n-1]=1$ in case of a feedback error at time slot $n-1$ and $\epsilon[n-1]=0$ otherwise. It is also known that replacing $\mathbf{z}[n]$ with a zero-mean Gaussian random variable with variance σ_z^2 which is independent of the signal $\mathbf{r}[n]$ minimizes the mutual information between the transmitted symbols $\mathbf{r}[n]$ and the received signal $\mathbf{y}[n]$ given in (8) [1]. A lower bound on the downlink capacity C_{lb} with imperfect CSI assuming the receiver knows $\hat{\mathbf{h}}[n]$ and $\mathbf{w}[n]$, (and not $\mathbf{w}_{\text{BS}}[n]$!) can be computed as follows

$$\begin{aligned}
T\cdot C &\overset{\text{(a)}}{\geq} I(\mathbf{r}[n];\mathbf{y}[n] \mid \hat{\mathbf{h}}[n],\mathbf{w}[n]) \\
&\overset{\text{(b)}}{=} I(\mathbf{r}[n];\mathbf{y}[n]|\hat{\mathbf{h}}[n],\mathbf{w}[n],\epsilon[n-1]) \\
&\quad +I(\epsilon[n-1];\mathbf{y}[n] \mid \hat{\mathbf{h}}[n],\mathbf{w}[n]) \\
&\quad -I(\epsilon[n-1];\mathbf{y}[n] \mid \hat{\mathbf{h}}[n],\mathbf{w}[n],\mathbf{r}[n]) \\
&\overset{\text{(c)}}{\geq} I(\mathbf{r}[n];\mathbf{y}[n] \mid \hat{\mathbf{h}}[n],\mathbf{w}[n],\epsilon[n-1]) \\
&\quad -I(\epsilon[n-1];\mathbf{y}[n] \mid \hat{\mathbf{h}}[n],\mathbf{w}[n],\mathbf{r}[n]) \\
&\overset{\text{(d)}}{=} I(\mathbf{r}[n];\mathbf{y}[n] \mid \hat{\mathbf{h}}[n],\mathbf{w}[n],\epsilon[n-1]) \\
&\quad -H(\epsilon[n-1] \mid \hat{\mathbf{h}}[n],\mathbf{w}[n],\mathbf{r}[n]) \\
&\quad +H(\epsilon[n-1] \mid \mathbf{y}[n],\hat{\mathbf{h}}[n],\mathbf{w}[n],\mathbf{r}[n]) \\
&\overset{\text{(e)}}{\geq} I(\mathbf{r}[n];\mathbf{y}[n] \mid \hat{\mathbf{h}}[n],\mathbf{w}[n],\epsilon[n-1]) \\
&\quad -H(\epsilon[n-1] \mid \hat{\mathbf{h}}[n],\mathbf{w}[n],\mathbf{r}[n]) \\
&\overset{\text{(f)}}{\geq} I(\mathbf{r}[n];\mathbf{y}[n]|\hat{\mathbf{h}}[n],\mathbf{w}[n],\epsilon[n-1])-H(\epsilon[n-1])
\end{aligned}$$

$$C_{\text{lb}}=\frac{1}{T}I(\mathbf{r}[n];\mathbf{y}[n]|\hat{\mathbf{h}}[n],\mathbf{w}[n],\epsilon[n-1])-\frac{1}{T}h_{\text{b}}(p_\epsilon) \tag{10}$$

where $I(p;q \mid m)$ denotes the mutual information between p and q given m and $H(p \mid q)$ denotes the conditional entropy of p given q. Inequality (a) follows by assuming that $\mathbf{r}[n]$ and $\mathbf{z}[n]$ are Gaussian distributed as explained above, step (b) from the chain rule of the mutual information, inequality (c) from the non-negativity of the mutual information, step (d) derives from the definition of the mutual information, inequality (e) from the non-negativity of the entropy and step (f) from the fact that conditioning reduces the entropy: $H(\epsilon[n-1] \mid \hat{\mathbf{h}}[n],\mathbf{w}[n],\mathbf{r}[n]) \leq H(\epsilon[n-1])$. Note that $H(\epsilon[n-1])=h_{\text{b}}(p_\epsilon)$, which is the binary entropy function with probability p_e for the error event $\epsilon[n-1]=1$, i.e.,

$$h_{\text{b}}(p_\epsilon)=-p_\epsilon\log_2 p_\epsilon-(1-p_\epsilon)\log_2(1-p_\epsilon). \tag{11}$$

The first term in (10) represents the capacity when the user is aware of the feedback error events but does not know the beamforming vector $\mathbf{w}_{\text{BS}}[n]$ actually applied by the BS after a feedback error which can be written as

$$\begin{aligned}
&I(\mathbf{r}[n];\mathbf{y}[n] \mid \hat{\mathbf{h}}[n],\mathbf{w}[n],\epsilon[n-1]) = \\
&\qquad (1-p_\epsilon)\, I\left(\mathbf{r}[n];\mathbf{y}[n] \mid \hat{\mathbf{h}}[n],\mathbf{w}_{\text{BS}}[n]{=}\mathbf{w}[n]\right) \\
&\qquad p_\epsilon\, I\left(\mathbf{r}[n];\mathbf{y}[n] \mid \hat{\mathbf{h}}[n],\mathbf{w}_{\text{BS}}[n]{\neq}\mathbf{w}[n]\right).
\end{aligned} \tag{12}$$

$I\left(\mathbf{r}[n];\mathbf{y}[n] \mid \hat{\mathbf{h}}[n],\mathbf{w}_{\text{BS}}[n]{=}\mathbf{w}[n]\right)$ corresponds to the mutual information assuming perfect feedback and that the effective estimated channel after beamforming, i.e., $\mathbf{w}_{\text{BS}}^{\text{H}}[n]\hat{\mathbf{h}}[n]$, is also perfectly known. The second term corresponds to the mutual information of a noncoherent SISO channel with coherence length D and *unknown* effective channel $h[n] = \mathbf{w}^{\text{H}}[n]\mathbf{h}[n]$. As explained in the previous section, after a feedback error, the applied beamforming vector $\mathbf{w}_{\text{BS}}[n]$ is totally uncorrelated with the channel and hence, it can be considered as if the beamforming vector were random. In this case the *unknown* channel $h[n]$ is a zero-mean unit-variance complex Gaussian random variable with $\text{E}\left[|h[n]|^2\right] = 1$. We can lower bound the second term using the following

$$\begin{aligned}
I\left(\mathbf{r}[n];\mathbf{y}[n] \mid \hat{\mathbf{h}}[n],\mathbf{w}_{\text{BS}}[n]{\neq}\mathbf{w}[n]\right) &= I(\mathbf{r}[n];\mathbf{y}[n]) \\
&= I(\mathbf{r}[n];\mathbf{y}[n] \mid h[n]) + I(h[n];\mathbf{y}[n]) \\
&\quad -I(h[n];\mathbf{y}[n] \mid \mathbf{r}[n]) \\
&\geq I(\mathbf{r}[n];\mathbf{y}[n] \mid h[n]) - I(h[n];\mathbf{y}[n] \mid \mathbf{r}[n]),
\end{aligned} \tag{13}$$

where the second step follows from the chain rule for the mutual information and the last step follows from the non-negativity of the mutual information. The first term in the expression above can be computed in closed form [7]

$$\begin{aligned}
I(\mathbf{r}[n];\mathbf{y}[n] \mid h[n]) &= D\ \text{E}\left[\log_2\left(1+\frac{P}{\sigma_v^2}|h[n]|^2\right)\right] \\
&= D\log_2(\text{e})\ \text{e}^{\sigma_v^2/P}\text{E}_1(\sigma_v^2/P),
\end{aligned} \tag{14}$$

where $\text{E}_k(z)$ is the generalized exponential integral defined as $\text{E}_k(z)=\int_1^\infty \frac{\text{e}^{-zt}}{t^k}\text{d}t$. The second term in (13) is the mutual information between $h[n]$ and $\mathbf{y}[n]=h[n]\mathbf{x}[n]+\mathbf{v}[n]$ given $\mathbf{r}[n]$, which can be lower bounded as

$$\begin{aligned}
I(h[n];\mathbf{y}[n] \mid \mathbf{r}[n]) &= \text{E}\left[\log_2\left(\left|\mathbf{1}_D+\frac{P}{\sigma_v^2}\mathbf{r}[n]\mathbf{r}^{\text{H}}[n]\right|\right)\right] \\
&= \text{E}\left[\log_2\left(1+\frac{P}{\sigma_v^2}\|\mathbf{r}[n]\|_2^2\right)\right] \\
&= \log_2(\text{e})\ \text{e}^{\sigma_v^2/P}\sum_{k=1}^{D}\text{E}_k(\sigma_v^2/P),
\end{aligned} \tag{15}$$

where the expectation is taken over $\mathbf{r}[n]$ and the last step follows similarly to the last step in (14).

3.2. Capacity Lower Bound with Perfect Feedback

Now we focus on the second term of (12), i.e. the mutual information under perfect feedback with known $\mathbf{w}_{\text{BS}}^{\text{H}}[n]\hat{\mathbf{h}}[n]$

$$I\left(\mathbf{r}[n];\mathbf{y}[n]|\hat{\mathbf{h}}[n],\mathbf{w}_{\text{BS}}[n]{=}\mathbf{w}[n]\right){=}I\left(\mathbf{r}[n];\mathbf{y}[n]|\mathbf{w}_{\text{BS}}[n]\hat{\mathbf{h}}[n]\right)$$
$$= \text{E}[\log_2(1{+}\frac{P}{\sigma_v^2}|\mathbf{w}_{\text{BS}}[n]\hat{\mathbf{h}}[n]|^2)], \quad (16)$$

with a lower bound given by the following theorem.

Theorem 3.1. *A capacity lower bound with perfect feedback and estimated, quantized and outdated transmit CSI reads as*

$$\frac{D}{T}\left(1-\frac{1}{\sqrt{2\pi}}\frac{\sigma_\eta}{\mu_\eta}\right)\log_2\left(1+\frac{(\alpha'(ME[\nu]-1)+1)(1-\sigma_e^2)}{\sigma_e^2+\frac{\sigma_v^2}{P}}\right),$$

Proof. Let us first define

$$\nu = \frac{|\mathbf{w}^{\text{H}}[n]\hat{\mathbf{h}}[n-1]|^2}{\|\hat{\mathbf{h}}[n-1]\|_2^2} \quad (17)$$
$$\eta = |\mathbf{w}^{\text{H}}[n]\hat{\mathbf{h}}[n]|^2, \quad (18)$$

and present the following useful result from [8].

Theorem 3.2. *For a concave function $\Omega(\eta)$, we have that*

$$E[\Omega(\eta)] \geq \left(1-\frac{d_\eta}{2\mu_\eta}\right)\Omega(\mu_\eta),$$

where $\eta \in [0,\infty[$ and μ_η and $d_\eta = E[|\eta-\mu_\eta|]$ are the expected value and the absolute deviation of η, respectively.

From (16) note that $\Omega(\eta) = \log_2(1+\eta)$ and to employ Theorem 3.2, we first need the expected value of η which is

$$\begin{aligned}
\mu_\eta &= \text{E}\left[|\mathbf{w}^{\text{H}}[n]\hat{\mathbf{h}}[n]|^2\right] \\
&= \alpha'\text{E}\left[|\mathbf{w}^{\text{H}}[n]\hat{\mathbf{h}}[n-1]|^2\right] + (1-\alpha')\text{E}\left[|\mathbf{w}^{\text{H}}[n]\hat{\mathbf{g}}[n-1]|^2\right] \\
&= \alpha'\text{E}\left[\|\hat{\mathbf{h}}[n-1]\|_2^2\frac{|\mathbf{w}^{\text{H}}\hat{\mathbf{h}}[n-1]|^2}{\|\hat{\mathbf{h}}[n-1]\|_2^2}\right]+(1-\alpha')(1-\sigma_e^2) \\
&= \alpha'\text{E}\left[\|\hat{\mathbf{h}}[n-1]\|_2^2\right]\text{E}\left[\frac{|\mathbf{w}^{\text{H}}\hat{\mathbf{h}}[n-1]|^2}{\|\hat{\mathbf{h}}[n-1]\|_2^2}\right]+(1-\alpha')(1-\sigma_e^2) \\
&= (\alpha' M\,\text{E}[\nu]+1-\alpha')(1-\sigma_e^2),
\end{aligned} \quad (19)$$

where the second step makes use of (6), the third steps follows since $\mathbf{w}$ has unit norm and the variance of the elements of $\hat{\mathbf{g}}[n-1]$ is $(1-\sigma_e^2)$, the fourth step derives from the fact that $\|\hat{\mathbf{h}}[n-1]\|_2^2$ and $\nu = |\mathbf{w}^{\text{H}}[n]\hat{\mathbf{h}}[n-1]|^2/\|\hat{\mathbf{h}}[n-1]\|_2^2$ are independent [5]. Now we need $\text{E}[\nu]$, which can be shown is given by [2]

$$\text{E}[\nu] = 1-2^B\text{Beta}\left(2^B, M/(M-1)\right), \quad (20)$$

where Beta$(\bullet,\bullet)$ is the Beta function and recall that B are the quantization bits. Let now characterize the variance of η

$$\sigma_\eta^2 = \text{E}\left[(\eta-\mu_\eta)^2\right] = \text{E}\left[\eta^2\right]-\mu_\eta^2, \quad (21)$$

for which we need the second moment of η, i.e.

$$\begin{aligned}
\text{E}\left[\eta^2\right] &= \alpha'^2\text{E}\left[|\mathbf{w}^{\text{H}}[n]\hat{\mathbf{h}}[n-1]|^4\right]+(1-\alpha')^2\,\text{E}\left[|\mathbf{w}^{\text{H}}[n]\hat{\mathbf{g}}[n-1]|^4\right] \\
&\quad + 4\alpha'(1-\alpha')\,\text{E}\left[|\mathbf{w}^{\text{H}}[n]\hat{\mathbf{h}}[n-1]|^2\right]\text{E}\left[|\mathbf{w}^{\text{H}}[n]\hat{\mathbf{g}}[n-1]|^2\right] \\
&= \alpha'^2\text{E}\left[\|\hat{\mathbf{h}}[n-1]\|_2^4\right]\text{E}\left[\nu^2\right]+(1-\alpha')^2\,\text{E}\left[|\mathbf{w}^{\text{H}}[n]\hat{\mathbf{h}}[n-1]|^4\right] \\
&\quad + 4\alpha'(1-\alpha')\,\text{E}\left[\|\hat{\mathbf{h}}[n-1]\|_2^2\right]\text{E}[\nu]\,\text{E}\left[|\mathbf{w}^{\text{H}}[n]\hat{\mathbf{h}}[n-1]|^2\right] \\
&= \alpha'^2(1-\sigma_e^2)^2M(M+1)\text{E}\left[\nu^2\right]+2(1-\alpha')^2(1-\sigma_e^2)^2 \\
&\quad + 4\alpha'(1-\alpha')(1-\sigma_e^2)^2M\text{E}[\nu],
\end{aligned} \quad (22)$$

where the second step follows by replacing $|\mathbf{w}^{\text{H}}[n]\,\hat{\mathbf{h}}[n-1]|^2$ for $\|\hat{\mathbf{h}}[n-1]\|_2^2\ \nu$ and then using the fact that $\|\hat{\mathbf{h}}[n-1]\|_2^2$ and ν are independent. The fourth step is derived from noting that $\text{E}\left[\|\hat{\mathbf{h}}\|_2^4\right] = M(M+1)(1-\sigma_e^2)^2$, i.e. the second moment of the chi-square distributed random variable $\|\hat{\mathbf{h}}\|_2^2$ with $2M$ degrees of freedom, each degree of freedom with variance $(1-\sigma_e^2)/2$. To evaluate $\text{E}\left[\eta^2\right]$ in the above expression, we still need $\text{E}\left[\nu^2\right]$, which can be shown is given by

$$\text{E}\left[\nu^2\right]{=}1-2^{B+1}\text{Beta}\left(2^B,\frac{M}{M{-}1}\right){+}2^B\text{Beta}\left(2^B,\frac{M{+}1}{M{-}1}\right). \quad (23)$$

Using (19)-(23) we can calculate the ratio $\frac{\sigma_\eta}{\mu_\eta} =$

$$\frac{\sqrt{\alpha'^2(M(M{+}1)\text{E}[\nu^2]{-}M^2\text{E}[\nu]^2)+(1{-}\alpha')^2{+}2\alpha'(1{-}\alpha')M\text{E}[\nu]}}{\alpha'(M\text{E}[\nu]{-}1){+}1}. \quad (24)$$

To compute a lower bound of (16) with Theorem 3.2, we need d_η, but exact evaluation of d_η appears intractable in general. However, we can find an upper bound for d_η based on σ_η. To this end let us compute μ_η, d_η and σ_η by setting $B\to\infty$ (no quantization, i.e. $\nu\to 1$) and $\alpha'=1$ (no outdating)

$$\begin{aligned}
\mu_\eta|_{\substack{B\to\infty\\ \alpha'=1}} &= (1-\sigma_e^2)M \\
d_\eta\,|_{\substack{B\to\infty\\ \alpha'=1}} &= \frac{2(1-\sigma_e^2)M^M}{(M-1)!\ \text{e}^M} \\
\sigma_\eta^2\,|_{\substack{B\to\infty\\ \alpha'=1}} &= (1-\sigma_e^2)^2M,
\end{aligned}$$

which we can use to obtain

$$\frac{d_\eta}{\mu_\eta}\bigg|_{\substack{B\to\infty\\ \alpha'=1}} = \frac{2M^M}{M!\ \text{e}^M} \leq \frac{2}{\sqrt{2\pi M}} \quad (25)$$

$$\frac{2}{\sqrt{2\pi}}\ \frac{\sigma_\eta}{\mu_\eta}\bigg|_{\substack{B\to\infty\\ \alpha=1}} = \frac{2}{\sqrt{2\pi M}}, \quad (26)$$

where the inequality follows from Stirling's approximation. From (25) and (26) we can deduce that for the general case

$$\frac{d_\eta}{\mu_\eta} \leq \frac{2}{\sqrt{2\pi}}\frac{\sigma_\eta}{\mu_\eta}. \quad (27)$$

Substituting (24) in (27) and then plugging this with (19) in the result of Theorem 3.2 multiplied with the ratio of data symbols to total number of symbols concludes the proof. □

We are ready now to give C_{lb} from (10).

Theorem 3.3. *A lower bound on the capacity with imperfect transmit CSI which is estimated, quantized, outdated and affected by feedback errors is given by*

$$\begin{aligned} C_{lb} &= \frac{D}{T}\, p_\epsilon \log_2(\mathrm{e})\, \mathrm{e}^{\frac{\sigma_v^2}{P}} \mathrm{E}_1\left(\frac{\sigma_v^2}{P}\right) \\ &+ \frac{D}{T}(1-p_\epsilon)\left(1-\frac{1}{\sqrt{2\pi}}\frac{\sigma_\eta}{\mu_\eta}\right)\times \\ &\quad \log_2\left(1+\frac{(\alpha'(ME[\nu]-1)+1)(1-\sigma_e^2)}{\sigma_e^2+\frac{\sigma_v^2}{P}}\right) \\ &- \frac{1}{T}\left(p_\epsilon \log_2(\mathrm{e})\, \mathrm{e}^{\frac{\sigma_v^2}{P}} \sum_{k=1}^{D} \mathrm{E}_k\left(\frac{\sigma_v^2}{P}\right) + h_b(p_\epsilon)\right). \end{aligned}$$

Proof. The result follows by plugging (11), (14), (15) and the result from Theorem 3.1 in (10) with $\frac{\sigma_\eta}{\mu_\eta}$ given by (24). □

3.3. Capacity Upper Bound with Imperfect CSI

We now draw our attention to computing an upper bound on the mutual information between $\mathbf{r}[n]$ and $\mathbf{y}[n]$ in given in (8).

Theorem 3.4. *An upper bound on the capacity under beamforming of the MISO channel with imperfect CSI in an FDD system is given by*

$$\begin{aligned} C_{ub} &= \frac{D}{T}(1-p_\epsilon)\log_2\left(1+\frac{P}{\sigma_v^2}(\alpha'(ME[\nu]-1)(1-\sigma_e^2)+1)\right) \\ &\quad + \frac{D}{T}\, p_\epsilon \log_2(\mathrm{e})\mathrm{e}^{\frac{\sigma_v^2}{P}} \mathrm{E}_1\left(\frac{\sigma_v^2}{P}\right). \end{aligned}$$

Proof. This upper bound can be derived by assuming the existence of a genie which provides extra information to the user namely, the exact channel vector $\mathbf{h}[n]$ and the actual beamforming vector applied by the BS $\mathbf{w}_{\text{BS}}[n]$, such that

$$\begin{aligned} C &\overset{\text{(a)}}{\leq} \frac{D}{T}\mathrm{E}\left[\log_2\left(1+\frac{P}{\sigma_v^2}\|\mathbf{w}_{\text{BS}}^{\text{H}}[n]\mathbf{h}[n]\|_2^2\right)\right] \\ &\overset{\text{(b)}}{=} \frac{D}{T}(1-p_\epsilon)\,\mathrm{E}\left[\log_2\left(1+\frac{P}{\sigma_v^2}\|\mathbf{w}^{\text{H}}[n]\mathbf{h}[n]\|_2^2\right)\right] \\ &\quad + \frac{D}{T}\, p_\epsilon\, \mathrm{E}\left[\log_2\left(1+\frac{P}{\sigma_v^2}|h[n]|^2\right)\right] \\ &\overset{\text{(c)}}{\leq} \frac{D}{T}(1-p_\epsilon)\log_2\left(1+\frac{P}{\sigma_v^2}\mathrm{E}\left[\|\mathbf{w}^{\text{H}}[n]\hat{\mathbf{h}}[n]+\mathbf{w}^{\text{H}}[n]\mathbf{e}[n]\|_2^2\right]\right) \\ &\quad + \frac{D}{T}\, p_\epsilon\, \log_2(\mathrm{e})\, \mathrm{e}^{\frac{\sigma_v^2}{P}}\, \mathrm{E}_1\left(\frac{\sigma_v^2}{P}\right) \\ &\overset{\text{(d)}}{=} \frac{D}{T}(1-p_\epsilon)\log_2\left(1+\frac{P}{\sigma_v^2}(\mu_\eta+\sigma_e^2)\right) \\ &\quad + \frac{D}{T}\, p_\epsilon\, \log_2(\mathrm{e})\, \mathrm{e}^{\frac{\sigma_v^2}{P}}\, \mathrm{E}_1(\sigma_v^2/P). \end{aligned}$$

where (a) follows from knowing $\mathbf{w}_{\text{BS}}^{\text{H}}$ and $\mathbf{h}[n]$ at the user and (b) from the fact that $\mathbf{w}_{\text{BS}}[n] = \mathbf{w}[n]$ in case of no feedback errors and in case of a feedback error we have $h[n] = \mathbf{w}_{\text{BS}}^{\text{H}}\mathbf{h}[n]$ which is Gaussian distributed. In Step (c) the first term results from Jensen's inequality and (3) and the second term from [7]. Step (d) arises from (18) and then the result of the theorem follows by plugging (19). □

3.4. Capacity Bounds in a TDD System

As discussed in Remark 2.1, in a TDD system the transmit CSI is estimated and outdated. The capacity bounds for this case can be derived based on the results of Theorems 3.3 and 3.4 by neglecting the effects of the quantization and of the feedback errors. First we present a lower bound on the downlink capacity C_{TDD} in a TDD system.

Theorem 3.5. *A lower bound on the downlink capacity of a TDD system with estimated and outdated CSI is*

$$\begin{aligned} C_{TDD,lb} &= \left(1-\frac{1}{\sqrt{2\pi}}\frac{\sigma_{\eta,TDD}}{\mu_{\eta,TDD}}\right)\log_2\left(1+\frac{P(\alpha'(M-1)+1)}{\sigma_v^2}\right) \\ &\quad -\frac{1}{T}\log_2(\mathrm{e})\, \mathrm{e}^{\frac{\sigma_v^2}{P(\alpha'(M-1)+1)}} \sum_{k=1}^{T} \mathrm{E}_k\left(\frac{\sigma_v^2}{P(\alpha'(M-1)+1)}\right). \end{aligned}$$

Proof. The proof follows by taking the result from Theorem 3.3 without quantization, with perfect feedback and without downlink resources employed for training, i.e. $T_{\text{DL}} = 0$ and $\sigma_e^2 = 0$. Additionally, as explained in Remark 2.1 the variance of the estimation error is different in an TDD system, i.e. $\sigma_{e_{\text{TDD}}}^2 = \frac{1}{1+\frac{P_{\text{UL}}}{\sigma_u^2}T_{\text{UL}}}$. Plugging $\mathrm{E}[\nu] = 1$ (no quantization), $p_\epsilon = 0$ (no feedback errors) and $D = T$ (no downlink training) in (24) and Theorem 3.3 gives us the result of the theorem where the substracted term represents the rate penalty for not knowing the effective channel $\mathbf{w}_{\text{BS}}^{\text{H}}[n]\mathbf{h}[n]$ at the user. With $\mathrm{E}[\nu] = 1$ and $p_\epsilon = 0$ in (24) we have that

$$\frac{\sigma_{\eta,\text{TDD}}}{\mu_{\eta,\text{TDD}}} = \frac{\sqrt{\alpha'(2-\alpha')(M-1)+1}}{\alpha'(M-1)+1}. \tag{28}$$

□

An upper bound on the downlink capacity under beamforming in a TDD system can be derived by making use of Theorem 3.4.

Theorem 3.6. *An upper bound on the downlink capacity of a TDD system with estimated CSI is*

$$C_{TDD,ub} = \log_2\left(1+\frac{P}{\sigma_v^2}((M-1)(1-\sigma_{e_{TDD}}^2)+1)\right).$$

Proof. The proof makes use of Theorem 3.4, which corresponds to the capacity upper bound in an FDD system and hence the result needs to be modified for the TDD case. For the TDD case, we have no quantization and hence $\mathrm{E}[\nu] = 1$ (c.f. (17)). Additionally, we have no outdating and no feedback and therefore, $\alpha' = 1$ and $p_\epsilon = 0$. The final result follows by replacing $\mathrm{E}[\nu] = 1$, $\alpha' = 1$ and $p_\epsilon = 0$ in Theorem 3.4 and by setting $D = T$. □

3.5. Capacity with Perfect CSI

For comparison we present the capacity of the MISO downlink channel with perfect CSI at the transmitter and receiver. The beamforming vector applied by the BS in the downlink is $\mathbf{w}_{\text{BS}}[n] = \frac{\mathbf{h}[n]}{\|\mathbf{h}[n]\|_2}$ such that (1) can be rewritten as

$$\mathbf{y}[n] = \sqrt{P}\,\|\mathbf{h}[n]\|_2 \mathbf{r}[n] + \mathbf{v}[n]. \tag{29}$$

The capacity $C_{\text{Perf. CSI}}$ with perfect CSI can be computed in closed form and is given by [7]

$$C_{\text{Perf. CSI}} = \log_2(\mathrm{e})\, \mathrm{e}^{\frac{\sigma_v^2}{P}} \sum_{k=1}^{M} \mathrm{E}_k\left(\frac{\sigma_v^2}{P}\right). \tag{30}$$

This capacity is similar to the result from Theorem 3.6 except that in (30) the factor $\frac{T-T_{\text{UL}}}{T} = 1$ and $\sigma^2_{e_{\text{TDD}}} = 0$ since there is no estimation loss as the user knows the channel perfectly.

3.6. Capacity without Transmit CSI but with Receive CSI

Also for comparison we present the capacity when the BS has no CSI but the receiver can estimate the downlink channel through training. Note that this is not the same capacity with non-coherent detection like in (13). Without CSI at the transmitter the use of multiple antennas are transparent to the user in a single-user MISO channel. If the BS either applies a random beam $\mathbf{w}_{\text{BS}}[n]$ or equivalently transmits with only one antenna to the user, the equivalent SISO channel $h[n]$ is a scalar which is Gaussian distributed. Assuming the receiver can estimate the scalar channel $h[n]$ with T_{DL} symbols, such that the channel estimate is $\hat{h}[n]$ we have that

$$\begin{aligned}\mathbf{y}[n] &= \sqrt{P}\,\hat{h}[n]\mathbf{r}[n] + \sqrt{P}\,e'[n]\mathbf{r}[n] + \mathbf{v}[n] \\ &= \sqrt{P}\,\hat{h}[n]\mathbf{r}[n] + \mathbf{z}'[n]. \end{aligned}\tag{31}$$

where $e'[n]$ is the estimation error with variance given exactly like the variance of the estimation error in a TDD system (Remark 2.1), i.e. $\sigma^2_{e'} = \frac{1}{1+\frac{P}{\sigma_v^2}T_{\text{DL}}}$, and where the effective noise $\mathbf{z}'[n] = \sqrt{P}\;e'[n]\mathbf{r}[n] + \mathbf{v}[n]$ with variance $\sigma_v^2 + \sigma^2_{e'}$. The capacity without CSI C_{NoTxCSI} at the transmitter is unknown and so we compute lower and upper bounds on the capacity of (31). We can compute a lower bound by first assuming that $\mathbf{r}[n]$ and $\mathbf{z}'[n]$ are Gaussian distributed as has been explained before, such that

$$\begin{aligned} C_{\text{NoTxCSI}} \geq\; & \frac{D}{T}\,\mathrm{E}\left[\log_2\left(1+\frac{P}{\sigma_v^2+\sqrt{P}\sigma^2_{e'}}|h[n]|^2\right)\right] \\ \geq\; & \frac{D}{T}\left(1-\frac{1}{\mathrm{e}}\right)\log_2\left(1+\frac{1-\sigma^2_{e'}}{\frac{\sigma_v^2}{P}+\sigma^2_{e'}}\right) = C_{\text{NoTxCSI,lb}}, \end{aligned}\tag{32}$$

where the second step follows by making use of Theorem 3.2 where $\eta = |h[n]|^2$ is an exponential distributed random variable with expected value given by $1-\sigma^2_{e'}$ and absolute deviation given by $\frac{2}{\mathrm{e}}\left(1-\sigma^2_{e'}\right)$. An upper bound can be computed by making use of Theorem 3.6 and setting $M = 1$

$$C_{\text{NoTxCSI,ub}} = \frac{D}{T}\log_2\left(1+\frac{P}{\sigma_v^2}\right).$$

4. NUMERICAL RESULTS AND DISCUSSION

In this section we present some results of the bounds computed in the previous section. Fig. 1 depicts the bounds as a function of $\frac{P}{\sigma_v^2}$ in dB for $M = 10$, $T = 500$, $T_{\text{DL}} = 100$, $B = 20$, $\alpha = 0.9$, $p_\epsilon = 0$, $\sigma_u^2 = 1$, $P_{\text{UL}} = P/M$ and $T_{\text{UL}} \in \{10, 100\}$ (the last 3 parameters have an impact on the TDD case). One can see the impact of the different types and degrees of the CSI on the capacity. The best case with imperfect CSI consists of having just estimated and outdated CSI at the transmitter, i.e. the TDD case. Nevertheless, in the TDD system, the performance of the downlink is directly related to the quality of the channel estimate in the uplink, i.e., T_{UL} and P_{UL}/σ_u, while the FDD downlink performance is influenced by the number of feedback bits and their reliability. For this reason, the FDD system could achieve higher downlink rates than the TDD system in some cases.

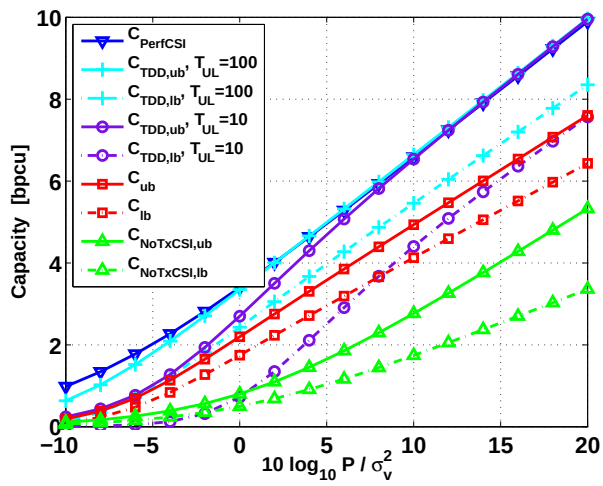

Fig. 1. Capacity Bounds vrs $\frac{P}{\sigma_v^2}$ for the Different Scenarios.

The tradeoff between training and data payload can be observed in Fig. 2, where the bounds are shown as a function of T_{DL}. Only the bounds for the FDD case, i.e. C_{ub} and C_{lb}, and the lower bound for the case without CSI at the transmitter are functions of T_{DL}. The parameters are the same as before except that $\frac{P}{\sigma_v^2} = 6$ dB and $T_{\text{UL}} = 100$. Obviously, if $M \ll T$, then it is optimum to spent more than $T_{\text{DL}} = M$ symbols in order to reduce the estimation error.

Fig. 3 presents the FDD bounds as function of the feedback bits B. Now, we assume a feedback error probability given by $p_\epsilon = 1-(1-p_{\text{s}})^{\frac{B}{2}}$ with $p_{\text{s}} = 0.005$. This corresponds, for instance, to an uncoded QPSK (unfaded) transmission link with symbol error probability p_{s}. Again, the other parameters are the same as given in the beginning of the section except that $M = 4$, $\frac{P}{\sigma_v^2} = 6$ dB and $T_{\text{UL}} = 100$. One can observe an optimum number of feedback bits B. The reason for this is that, larger B decreases the quantization loss while leading to an increased feedback error probability p_ϵ.

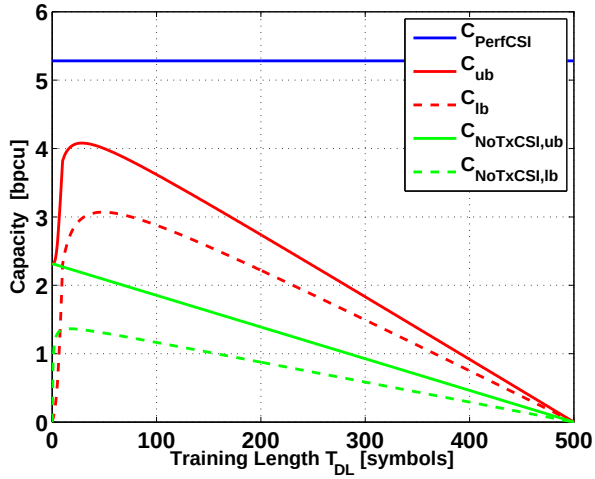

Fig. 2. Capacity Bounds vrs Training Length T_{DL}

The results are given for different α. Interestingly, the optimal values of B obtained, each for the lower and upper bounds of the capacity, are nearly the same, and thus, they probably maximize its exact value. Fig. 4 shows the bounds as a func-

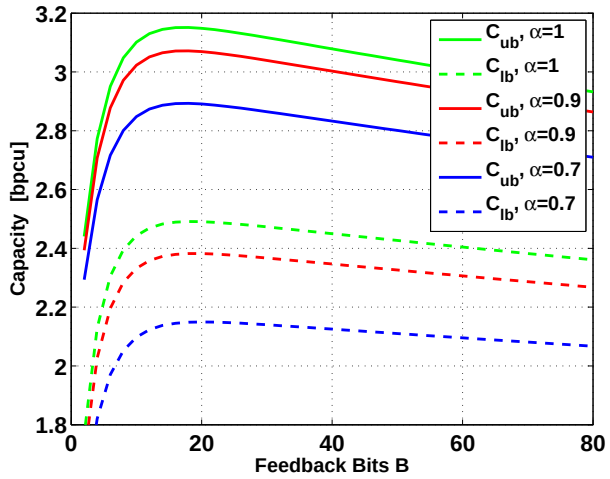

Fig. 3. FDD System Capacity Bounds vrs B

tion of M for fixed B. The parameters are the same as in the previous figure, except that $B = 20$. As it can be observed from the upper as well as the lower bounds, as the number of antennas M approaches the number of feedback bits B or the training length T_{DL}, the increase in FDD capacity becomes marginal. Thus, in terms of FDD capacity, there is barely any benefit from having M beyond T_{DL} or B.

5. CONCLUSIONS

We developed a general MISO downlink framework that considers the inherent tradeoff between training, feedback and data in downlink communications. Based on it, we derived lower and upper bounds on the downlink sum capacity for various communication scenarios and settings, taking into account estimated, quantized and delayed CSI and erroneous feedback. Numerical results shows that, even under this conditions, the feedback of imperfect CSI is still beneficial. The capacity bounds may have interesting applications in the analysis and design of wireless systems, such as the optimization of resource allocation between data, training and feedback.

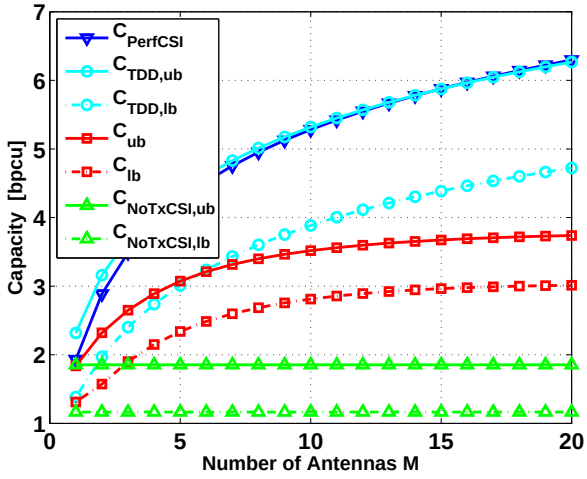

Fig. 4. Capacity Bounds vrs Number of Antennas M

6. REFERENCES

[1] B. Hassibi and B. M. Hochwald, "How much training is needed in a multiple-antenna wireless link?", *IEEE Trans. Inform. Theory*, vol. 49, pp. 951-964, Apr. 2003.

[2] C. K. Au-Yeung and D. J. Love, "On the performance of random vector quantization limited feedback beamforming in a MISO System", *IEEE Trans. Wireless Commun.*, vol. 6, Feb. 2007.

[3] G. Caire, N. Jindal, M. Kobayashi and N. Ravindran, "Multiuser MIMO Downlink Made Practical: Achievable Rates with Simple Channel State Estimation and Feedback Schemes," *Arxiv preprint cs.IT/0710.2642*.

[4] W. Santipach and M. L. Honig, "Capacity of Beamforming with Limited Training and Feedback", *Proceedings IEEE ISIT*, pp. 376-380, Jul. 2006.

[5] J. C. Roh and B. D. Rao, "Transmit beamforming in multiple-antenna systems with finite rate feedback: A VQ-based approach", *IEEE Trans. Inform. Theory*, vol. 52, no. 3, pp. 1101-1112, Mar. 2006.

[6] M. Castaneda, A. Mezghani, and J. A. Nossek, "Optimal Resource Allocation in Downlink/Uplink of Single-User MISO/SIMO FDD Systems with Limited Feedback", *submitted to the IEEE SPAWC* 2009.

[7] Mohamed-Slim Alouini, "Adaptive and Diversity Techniques for Wireless Digital Communications over Fading Channels," *Phd Thesis*, p. 128.

[8] A. Ben-Tal and E. Hochman, "More bounds on the expectation of a convex function of a random variable", *Journal of Applied Probability*, vol. 9, pp. 803 812, 1972.