Design of Single User Limited Feedback Systems

Mario H. Castañeda Garcia, *Student Member, IEEE*, Amine Mezghani, *Student Member, IEEE*, and Josef A. Nossek, *Fellow, IEEE*

***Abstract*—The available *channel state information* (CSI) in a limited feedback system like the *frequency division duplex* (FDD) downlink is not perfect since it is subject to estimation, quantization and feedback errors, and in addition, can be outdated. Despite the fact that the capacity of limited feedback systems is unknown in general, we derive a novel lower bound on the capacity of single user limited feedback systems with imperfect CSI. Based on this bound, we propose the design of an FDD system by finding the optimum training and number of feedback bits. To this end we also take the FDD uplink into account, since practical FDD systems represent *two-way* systems. We also provide closed-form approximations for the optimum downlink training, uplink training and number of feedback bits which basically maximize lower bounds on the FDD downlink and uplink capacity with imperfect CSI.**

***Index Terms*—Limited feedback, capacity bound, imperfect channel state information, beamforming.**

I. Introduction

THE capacity of wireless communication links increases by deploying multiple antennas under the assumption of perfect *channel state information* (CSI) [1]. In practice, however, the available CSI is far from perfect, as it is in the downlink of a single user *frequency division duplex* (FDD) system. Assuming multiple antennas at the *base station* (BS) and a single antenna at the user, the *imperfect* transmit CSI at the BS for the *multiple-input single-output* (MISO) downlink becomes available through *limited feedback* from the user [2], [3]. The available CSI is subject to estimation and quantization errors, and can be outdated and affected by feedback errors. Such imperfect CSI characterization has been addressed in [4] for the multiuser downlink with *zero-forcing* (ZF) beamforming. However, the quantization error in the single user case leads solely to an SNR offset [6] and not to a rate saturation with SNR [5], as in the multiuser case. Moreover, our focus is on the ergodic capacity with imperfect CSI rather than achievable rates.

Imperfect transmit CSI as described previously has not been treated generally for the single user case. In [6]–[14], the available CSI at the user before feedback is assumed to be perfect. Limited feedback with a noiseless feedback channel is assumed in [15], [16]. Since the downlink capacity with imperfect CSI is unknown, one can recur to the computation of bounds [15]–[17]. In contrast to the literature, in the first part of this work we present a *novel* lower bound on the ergodic capacity with Gaussian signalling for the single user MISO FDD downlink with imperfect transmit CSI. Although the lower bound is derived following our previous work [17], the new bound is tighter. The tightness of the lower bound is evaluated with an upper bound.

In the second part of this paper we address the design of single user limited feedback systems which we characterize as *two-way* FDD systems, with transmission in the downlink and the uplink. Due to the feedback, we have an inherent coupling and tradeoff between the FDD downlink and uplink rates, which has been considered in some literature such as [16], where the optimum scaling of the training and feedback length is determined. Furthermore, this tradeoff is characterized in [20] by means of an upper bound on the rate region which is composed of the downlink and uplink rates. The optimization of a single user two-way limited feedback system is presented in [21], [22] with respect to the time and power allocated for the feedback and data phase under the assumption of perfect receive CSI and error-free and instantaneous feedback.

Despite the coupling between the FDD uplink and downlink, the feedback represents a significant overhead only for short time slot lengths. Thus, we propose a system design by optimizing each link separately, taking certain aspects into account. In contrast to the previous literature, we derive *closed-form* approximations for the optimum uplink training, downlink training and number of feedback bits as a function of the system parameters. To this end, this paper is organized as follows. Sections II and III present the system model of the two-way FDD system and the effective downlink channel with limited feedback beamforming. The lower bound on the FDD downlink capacity is derived in Section IV, and the lower bound is evaluated in Section V. We discuss the two-way system optimization in Section VI. The optimum training and number of feedback bits for the design of single user limited feedback systems are derived in Section VII, and the results are validated in Section VIII. We conclude the paper in Section IX.

II. System Model

We consider an FDD two-way wireless communication system between a BS with M antennas and a single-antenna user.[1] We assume Rayleigh fading, i.e., all channels are i.i.d. complex

Manuscript received October 16, 2013; revised March 7, 2014 and June 29, 2014; accepted August 21, 2014. Date of publication September 8, 2014; date of current version October 8, 2014. The associate editor coordinating the review of this paper and approving it for publication was W. Gerstacker.

The authors are with the Institute for Circuit Theory and Signal Processing, Technische Universität München, 80333 Munich, Germany (e-mail: mario.castaneda@tum.de; amine.mezghani@tum.de; josef.a.nossek@tum.de).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TWC.2014.2354361

[1]This scenario could correspond to a multiuser cellular system, where a scheduler in a cell selects one user at a time to serve in each link. In this context, the intercell interference is considered part of the noise. Note that multiple users can be served with perfect CSI, but serving multiple users with imperfect CSI in the downlink leads to a rate saturation with SNR [5], which is not the case with a single user. In this case, serving only one user with beamforming achieves a higher sum rate.

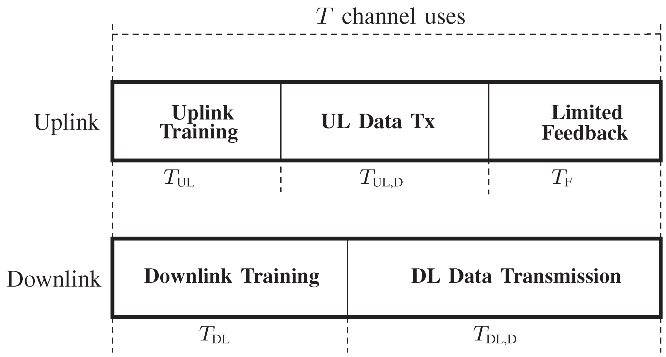

Fig. 1. Time slot structure in the FDD uplink and downlink.

Gaussian distributed with zero-mean and unit variance.[2] Thus, the downlink and uplink channel vectors are spatially uncorrelated. The time slot duration in the uplink and the downlink is T channel uses, where at each channel use one symbol is transmitted in each link. In addition, the uplink and downlink channels are assumed to be frequency flat, constant for one time slot, i.e., block fading, and are uncorrelated with one another.

In the downlink, the first T_{DL} channel uses are used to send a training sequence, which enables the user to estimate the downlink channel. The remaining $T_{\mathrm{DL,D}}$ symbols of a downlink time slot are reserved for transmitting data from the BS to the user, i.e., $T = T_{\mathrm{DL}} + T_{\mathrm{DL,D}}$.

Similarly, the first T_{UL} channel uses in the uplink time slot are allocated for transmitting a training sequence from the user to the BS, so the BS can estimate the uplink channel. In the next $T_{\mathrm{UL,D}}$ channel uses of the time slot, the user transmits data to the BS. The final T_{F} channel uses are reserved for the feedback of CSI from the user to the BS. Thus, in the uplink we have $T = T_{\mathrm{UL}} + T_{\mathrm{DL,D}} + T_{\mathrm{F}}$. An overview of the time slot structure is depicted in Fig. 1. The fed back CSI is available at the BS with a delay of one time slot, i.e., the CSI fed back during time slot $n-1$ in the uplink is available in the downlink at time slot n. Due to practical reasons, we do not consider an uneven power allocation among the different phases as in [20]–[23].

A. Uplink Channel

For the *uplink* (UL), we have a Gaussian *single-input multiple-output* (SIMO) channel with M receive antennas. With receive beamforming at the BS, the equivalent *single-input single-output* (SISO) system for the $T_{\mathrm{UL,D}}$ data symbols at time slot n is given by

$$\boldsymbol{y}_{\mathrm{UL}}[n] = \sqrt{P_{\mathrm{UL}}}\, \boldsymbol{w}_{\mathrm{UL}}^{\mathrm{H}}[n] \boldsymbol{h}_{\mathrm{UL}}[n] \boldsymbol{s}_{\mathrm{UL}}[n] + \boldsymbol{v}_{\mathrm{UL}}[n], \tag{1}$$

where at time slot n, $\boldsymbol{y}_{\mathrm{UL}}[n] \in \mathbb{C}^{T_{\mathrm{UL,D}}}$ is the signal after beamforming at the BS, $\boldsymbol{s}_{\mathrm{UL}}[n] \in \mathbb{C}^{T_{\mathrm{UL,D}}}$ are the transmit symbols with zero mean and unit variance, $\boldsymbol{w}_{\mathrm{UL}}[n] \in \mathbb{C}^{M}$ is the receive beamforming vector with unit norm, $\boldsymbol{h}_{\mathrm{UL}}[n] \in \mathbb{C}^{M}$ is the uplink SIMO channel and $\boldsymbol{v}_{\mathrm{UL}}[n] \in \mathbb{C}^{T_{\mathrm{UL,D}}}$ is the equivalent *additive white Gaussian noise* (AWGN), resulting after receive beamforming, with zero mean and variance σ_{n}^2, which is independent of the transmit signal. Furthermore, P_{UL} is the user's transmit power. With a scalar pilot sequence of length T_{UL}, the BS obtains a *minimum mean square error* (MMSE) estimate $\hat{\boldsymbol{h}}_{\mathrm{UL}}[n] \in \mathbb{C}^{M}$ of the uplink channel, such that

$$\boldsymbol{h}_{\mathrm{UL}}[n] = \hat{\boldsymbol{h}}_{\mathrm{UL}}[n] + \boldsymbol{e}_{\mathrm{UL}}[n], \tag{2}$$

where $\boldsymbol{e}_{\mathrm{UL}}[n] \in \mathbb{C}^{M}$ is the uplink estimation error. The MMSE estimate $\hat{\boldsymbol{h}}_{\mathrm{UL}}[n]$ is independent of $\boldsymbol{e}_{\mathrm{UL}}[n]$ and the elements of the $\hat{\boldsymbol{h}}_{\mathrm{UL}}[n]$ and $\boldsymbol{e}_{\mathrm{UL}}[n]$ are independent complex Gaussian random variables with zero mean and variance $1 - \sigma_{e_{\mathrm{UL}}}^2$ and $\sigma_{e_{\mathrm{UL}}}^2$, respectively, where the variance of the estimation error $\sigma_{e_{\mathrm{UL}}}^2$ is given by [23]

$$\sigma_{e_{\mathrm{UL}}}^2 = \frac{1}{1 + \frac{P_{\mathrm{UL}}}{\sigma_{\mathrm{n}}^2} T_{\mathrm{UL}}}. \tag{3}$$

The uplink data and the feedback can be detected at the BS with *maximum ratio combining* based on $\hat{\boldsymbol{h}}_{\mathrm{UL}}[n]$ such that the receive beamforming vector $\boldsymbol{w}_{\mathrm{UL}}[n]$ is $\boldsymbol{w}_{\mathrm{UL}}[n] = \hat{\boldsymbol{h}}_{\mathrm{UL}}[n] / \|\hat{\boldsymbol{h}}_{\mathrm{UL}}[n]\|_2$. Using (2), we can therefore rewrite (1) as

$$\begin{aligned} \boldsymbol{y}_{\mathrm{UL}}[n] = {} & \sqrt{P_{\mathrm{UL}}} \|\hat{\boldsymbol{h}}_{\mathrm{UL}}[n]\|_2\, \boldsymbol{s}_{\mathrm{UL}}[n] \\ & + \sqrt{P_{\mathrm{UL}}} \boldsymbol{w}_{\mathrm{UL}}^{\mathrm{H}}[n] \boldsymbol{e}_{\mathrm{UL}}[n] \boldsymbol{s}_{\mathrm{UL}}[n] + \boldsymbol{v}_{\mathrm{UL}}[n]. \end{aligned} \tag{4}$$

B. Downlink Channel

The *downlink* (DL) is Gaussian MISO channel with M transmit antennas. With transmit beamforming, the equivalent SISO downlink system for the $T_{\mathrm{DL,D}}$ data symbols at time slot n is

$$\boldsymbol{y}_{\mathrm{DL}}[n] = \sqrt{P_{\mathrm{DL}}}\, \boldsymbol{w}_{\mathrm{DL}}^{\mathrm{H}}[n] \boldsymbol{h}_{\mathrm{DL}}[n] \boldsymbol{s}_{\mathrm{DL}}[n] + \boldsymbol{v}_{\mathrm{DL}}[n], \tag{5}$$

where at time slot n, $\boldsymbol{y}_{\mathrm{DL}}[n] \in \mathbb{C}^{T_{\mathrm{DL,D}}}$ are the received signals, $\boldsymbol{s}_{\mathrm{DL}}[n] \in \mathbb{C}^{T_{\mathrm{DL,D}}}$ are the transmit symbols, $\boldsymbol{w}_{\mathrm{DL}}[n] \in \mathbb{C}^{M}$ is the transmit beamforming vector with unit norm, $\boldsymbol{h}_{\mathrm{DL}}[n] \in \mathbb{C}^{M}$ is the downlink MISO channel and $\boldsymbol{v}_{\mathrm{DL}}[n] \in \mathbb{C}^{T_{\mathrm{DL,D}}}$ is the AWGN with zero mean and variance σ_{n}^2, which is also independent of the transmit signal. P_{DL} is the transmit power available at the BS.

We consider *temporally correlated block fading*, i.e., $\boldsymbol{h}_{\mathrm{DL}}[n]$ is assumed to be constant for the coherence time of T symbols and is correlated with the channel in the previous time slot $\boldsymbol{h}_{\mathrm{DL}}[n-1]$ according to a first order Markov model as in [4]:

$$\boldsymbol{h}_{\mathrm{DL}}[n] = \sqrt{\alpha}\, \boldsymbol{h}_{\mathrm{DL}}[n-1] + \sqrt{1-\alpha}\, \boldsymbol{g}[n-1]. \tag{6}$$

The elements of $\boldsymbol{g}[n-1] \in \mathbb{C}^{M}$ are i.i.d. zero-mean unit-variance complex Gaussian random variables and are uncorrelated with $\boldsymbol{h}_{\mathrm{DL}}[n-1]$, and $\sqrt{\alpha}$ is the correlation coefficient, with $0 \leq \alpha < 1$. Furthermore, α is unknown at the transmitter, i.e., we assume no channel prediction. With M T_{DL}-dimensional pilot vectors transmitted from the M antennas of the BS [23],

[2]This holds, for instance, in a rich scattering environment and a uniform linear array with isotropic radiators separated by half the signal wavelength at the BS, so the channels are spatially uncorrelated [18] and there is no antenna mutual coupling [19].

the user obtains an MMSE estimate $\hat{\boldsymbol{h}}_{\mathrm{DL}}[n] \in \mathbb{C}^M$, such that the downlink channel is

$$\boldsymbol{h}_{\mathrm{DL}}[n] = \hat{\boldsymbol{h}}_{\mathrm{DL}}[n] + \boldsymbol{e}_{\mathrm{DL}}[n], \tag{7}$$

where $\boldsymbol{e}_{\mathrm{DL}}[n]$ is the error vector which is independent of $\hat{\boldsymbol{h}}_{\mathrm{DL}}[n]$. The elements of the $\hat{\boldsymbol{h}}_{\mathrm{DL}}$ and $\boldsymbol{e}_{\mathrm{DL}}$ are independent complex Gaussian random variables with zero mean and variance $1-\sigma^2_{e_{\mathrm{DL}}}$ and $\sigma^2_{e_{\mathrm{DL}}}$, respectively, where the variance of the estimation error $\sigma^2_{e_{\mathrm{DL}}}$ for $T_{\mathrm{DL}} \geq M$ is given by [23]

$$\sigma^2_{e_{\mathrm{DL}}} = \frac{1}{1+\frac{P_{\mathrm{DL}}}{M\sigma_{\mathrm{n}}^2}T_{\mathrm{DL}}}. \tag{8}$$

Afterward, the downlink channel estimate is quantized with B bits. For the BS to perform transmit beamforming in the downlink, the user needs to quantize only the *channel direction information* (CDI) of the estimated channel, which is a unit-norm vector. For this, we employ the *random vector quantization* (RVQ) scheme [15], where we have at the transmitter (BS) and receiver (user) a codebook $\mathcal{C}$ with 2^B random beamforming vectors $\boldsymbol{t}_i$, $i = 1, \ldots, 2^B$, i.i.d. isotropically distributed over the M-th dimensional unit sphere.[3] The CDI of the downlink channel $\hat{\boldsymbol{h}}_{\mathrm{DL}}[n-1]$, for instance, is quantized by selecting the beamforming vector $\boldsymbol{w}_{\mathrm{FB}}[n]$ (to be used at time slot n due to the feedback delay) that best matches the current channel estimate:

$$\boldsymbol{w}_{\mathrm{FB}}[n] = \underset{\boldsymbol{t}_i \in \mathcal{C}}{\arg\max} \; \left|\boldsymbol{t}_i^{\mathrm{H}}\hat{\boldsymbol{h}}_{\mathrm{DL}}[n-1]\right|^2. \tag{9}$$

After quantizing its estimated CDI, the user feeds back to the BS the B bits representing the index of the beamforming vector $\boldsymbol{w}_{\mathrm{FB}}[n]$. We assume the feedback is transmitted *uncoded* and employs $T_{\mathrm{F}} = B/2$ QPSK symbols in the uplink during the limited feedback phase (see Fig. 1). We employ QPSK symbols for the feedback, since higher modulation schemes suffer from higher symbol error probabilities. The feedback is sent *uncoded* due to the fact that codewords span many time slots, which implies the feedback would be even more outdated after decoding.

With a faded- and noise-prone uplink (feedback link), the feedback can be received erroneously at the BS. The feedback is sent without error detection and hence, we assume the BS to be unaware of a feedback error.[4] As in [4], we assume no optimized mapping of the B quantization bits in the feedback message, such that a bit error leads to a total feedback loss. Thus, beamforming based on an erroneously received message is equivalent to random beamforming, i.e., no antenna gain can be exploited. This is why the fed back beamforming vector $\boldsymbol{w}_{\mathrm{FB}}[n]$ has a different subscript than the beamforming vector applied at the BS $\boldsymbol{w}_{\mathrm{DL}}[n]$ (see (5)); they are only equal in case of error-free reception of the feedback message.

[3]We employ RVQ not just due to mathematical tractability, but also because it is asymptotically optimal in terms of capacity, when the number of antennas and number of feedback bits tend to infinity with a fixed ratio [6], [7]. In addition, RVQ is also able to perform very well for low number of transmit antennas [2], [6] compared to other optimized codebooks.

[4]Although we assume the BS to be unaware of feedback errors, the BS could be aware, to some extent, of a feedback error based on the correlation of received beamforming vectors with previously received vectors. For instance, the BS could employ the last correctly received beamforming vector instead of a beamforming vector given by an erroneously received feedback message.

C. Feedback Error Probability

The BS can receive the feedback message employing maximum ratio combining which is based on the uplink channel estimate. As stated in [23]–[25], the worst effect the estimation error can have is to behave as independent AWGN with variance $P_{\mathrm{UL}}\sigma^2_{e_{\mathrm{UL}}}$, such that we approximate the SNR during the feedback phase, for instance, at time slot $n-1$ with $\overline{\gamma}_{\mathrm{F}}[n-1] = (P_{\mathrm{UL}}\|\hat{\boldsymbol{h}}_{\mathrm{UL}}[n-1]\|_2^2)/(\sigma_{\mathrm{n}}^2 + P_{\mathrm{UL}}\sigma^2_{e_{\mathrm{UL}}})$, where the numerator results from performing maximum ratio combining with the uplink channel estimate $\hat{\boldsymbol{h}}_{\mathrm{UL}}[n-1]$. With this SNR approximation, we compute the average bit error probability with QPSK and maximum ratio combining with $\hat{\boldsymbol{h}}_{\mathrm{UL}}[n-1]$ by using [26, (7.20)]:

$$\mathrm{E}\left[p_{\mathrm{b}}\left(\overline{\gamma}_{\mathrm{F}}[n-1]\right)\right] = \left(\frac{1-\Lambda}{2}\right)^M \sum_{m=0}^{M-1}\begin{bmatrix} M-1+m \\ m \end{bmatrix}\left(\frac{1+\Lambda}{2}\right)^m, \tag{10}$$

where the expectation is taken over the channel estimate $\hat{\boldsymbol{h}}_{\mathrm{UL}}$ and $\Lambda = \sqrt{(\lambda_{\mathrm{UL}}/2)/(1+(\lambda_{\mathrm{UL}}/2))}$, where λ_{UL}

$$\lambda_{\mathrm{UL}} = \frac{P_{\mathrm{UL}}\left(1-\sigma^2_{e_{\mathrm{UL}}}\right)^2}{\sigma_{\mathrm{n}}^2 + P_{\mathrm{UL}}\sigma_{e_{\mathrm{UL}}}}. \tag{11}$$

Despite assuming the worst case noise for the estimation error, simulation results indicate that $\mathrm{E}[p_{\mathrm{b}}(\overline{\gamma}_{\mathrm{F}}[n-1])]$ is indeed a good approximation of the actual feedback error probability.

Since one symbol error leads to total feedback loss, the average feedback error probability is

$$p_\epsilon = \mathrm{E}\left[1-(1-p_{\mathrm{b}})^B\right] \leq 1-(1-\mathrm{E}\left[p_{\mathrm{b}}\right])^B, \tag{12}$$

where the expectation is taken over the channel estimate $\hat{\boldsymbol{h}}_{\mathrm{UL}}$. Since the upper bound resulting from Jensen's inequality is very tight for small values of p_{b}, we employ it as an approximation for the average feedback error probability.

III. Downlink Effective Channel

In *general*, let us define the effective downlink channel $h_{\mathrm{DL}}[n] = \boldsymbol{w}_{\mathrm{DL}}^{\mathrm{H}}[n]\boldsymbol{h}_{\mathrm{DL}}[n] \in \mathbb{C}$, such that (5) can be rewritten as

$$\boldsymbol{y}_{\mathrm{DL}}[n] = \sqrt{P_{\mathrm{DL}}}\; h_{\mathrm{DL}}[n]\boldsymbol{s}_{\mathrm{DL}}[n] + \boldsymbol{v}_{\mathrm{DL}}[n]. \tag{13}$$

Since they have different distributions, we make a distinction between the effective channel after erroneous and after correct feedback, and denote them as $h_{\mathrm{DL,ef}}[n]$ and $h_{\mathrm{DL,cf}}[n]$, respectively. Without feedback errors (correct feedback), $\boldsymbol{w}_{\mathrm{DL}}[n] = \boldsymbol{w}_{\mathrm{FB}}[n]$ and the equivalent channel denoted by $h_{\mathrm{DL,cf}}[n] = \boldsymbol{w}_{\mathrm{FB}}^{\mathrm{H}}[n]\boldsymbol{h}_{\mathrm{DL}}[n]$ is given by

$$\begin{aligned} h_{\mathrm{DL,cf}}[n] &= \boldsymbol{w}_{\mathrm{FB}}^{\mathrm{H}}[n]\left(\sqrt{\alpha}\hat{\boldsymbol{h}}_{\mathrm{DL}}[n-1] + \sqrt{\alpha}\boldsymbol{e}_{\mathrm{DL}}[n-1] + \sqrt{1-\alpha}\boldsymbol{g}[n-1]\right) \\ &= \sqrt{\alpha}\frac{\boldsymbol{w}_{\mathrm{FB}}^{\mathrm{H}}[n]\hat{\boldsymbol{h}}_{\mathrm{DL}}[n-1]}{\|\hat{\boldsymbol{h}}_{\mathrm{DL}}[n-1]\|_2}\|\hat{\boldsymbol{h}}_{\mathrm{DL}}[n-1]\|_2 + \sqrt{\alpha}e_{\mathrm{DL}} + \sqrt{1-\alpha}g, \end{aligned} \tag{14}$$

where the first line follows from (6), with (7) at time slot $n-1$. In the last line we made the following substitutions $e_{\mathrm{DL}} = \boldsymbol{w}_{\mathrm{FB}}^{\mathrm{H}}[n]\boldsymbol{e}_{\mathrm{DL}}[n-1]$ and $g = \boldsymbol{w}_{\mathrm{FB}}^{\mathrm{H}}[n]\boldsymbol{g}[n-1]$. Since $\boldsymbol{w}_{\mathrm{FB}}[n]$ has unit norm and is independent of $\boldsymbol{e}_{\mathrm{DL}}[n-1]$ and $\boldsymbol{g}[n-1]$, e_{DL} and g are complex Gaussian with zero mean and variance $\sigma_{e_{\mathrm{DL}}}^2$ and 1, respectively. Let us now define

$$\nu = \frac{\left|\boldsymbol{w}_{\mathrm{FB}}^{\mathrm{H}}[n]\hat{\boldsymbol{h}}_{\mathrm{DL}}[n-1]\right|^2}{\left\|\hat{\boldsymbol{h}}_{\mathrm{DL}}[n-1]\right\|_2^2}. \tag{15}$$

By recalling (9), notice that $\nu \in [0,1]$ is the maximum of 2^B square absolute values consisting of the inner product of two uniformly distributed unit vectors.

Using (14) and with $\mathrm{Re}\{z\}$ as the real part of z, $|h_{\mathrm{DL,cf}}[n]|^2$ can be expressed as

$$\begin{aligned}|h_{\mathrm{DL,cf}}[n]|^2 =& \alpha\nu\left\|\hat{\boldsymbol{h}}_{\mathrm{DL}}[n-1]\right\|_2^2 + 2\sqrt{\alpha(1-\alpha)}\,\mathrm{Re}\left\{g\, e_{\mathrm{DL}}^*\right\}\\ &+ 2\sqrt{\alpha(1-\alpha)}\,\mathrm{Re}\left\{\boldsymbol{w}_{\mathrm{FB}}^{\mathrm{H}}[n]\hat{\boldsymbol{h}}_{\mathrm{DL}}[n-1]\, g^*\right\} + (1-\alpha)\,|g|^2\\ &+ 2\alpha\,\mathrm{Re}\left\{\boldsymbol{w}_{\mathrm{FB}}^{\mathrm{H}}[n]\hat{\boldsymbol{h}}_{\mathrm{DL}}[n-1]\, e_{\mathrm{DL}}^*\right\} + \alpha\,|e_{\mathrm{DL}}|^2. \end{aligned} \tag{16}$$

Since g and e_{DL} have zero mean, are independent of $\hat{\boldsymbol{h}}_{\mathrm{DL}}[n-1]$ and in addition, ν and $\hat{\boldsymbol{h}}_{\mathrm{DL}}[n-1]$ are also independent [12], [27], the expected value of $|h_{\mathrm{DL,cf}}[n]|^2$ is

$$\begin{aligned}\mathrm{E}\left[|h_{\mathrm{DL,cf}}[n]|^2\right] &= \alpha\mathrm{E}\left[\left\|\hat{\boldsymbol{h}}_{\mathrm{DL}}[n-1]\right\|_2^2\right]\mathrm{E}[\nu] + (1-\alpha) + \alpha\,\sigma_{e_{\mathrm{DL}}}^2\\ &= \alpha\left(M\,\mathrm{E}[\nu]-1\right)\left(1-\sigma_{e_{\mathrm{DL}}}^2\right) + 1,\end{aligned} \tag{17}$$

which also results from $\|\hat{\boldsymbol{h}}_{\mathrm{DL}}[n-1]\|_2^2$ being chi-squared distributed with $2M$ degrees of freedom, each with variance $(1-\sigma_{e_{\mathrm{DL}}}^2)/2$. Define $\mu_\nu = \mathrm{E}[\nu]$, which is given by [13]

$$\mu_\nu = \mathrm{E}[\nu] = 1 - 2^B\mathrm{Beta}\left(2^B, \frac{M}{M-1}\right) \tag{18}$$

$$\approx 1 - 2^{-\frac{B}{M-1}}, \tag{19}$$

where $\mathrm{Beta}(a,b)$ is the Beta function [28] and the approximation results from [5].

For deriving the downlink capacity bound we require the variance of ν:

$$\sigma_\nu^2 = \mathrm{E}\left[(\nu - \mathrm{E}[\nu])^2\right] = \mathrm{E}[\nu^2] - \mathrm{E}^2[\nu]. \tag{20}$$

With the pdf $f_\nu(\nu)$ given in [13], $\mathrm{E}[\nu^2]$ can be obtained as

$$\begin{aligned}\mathrm{E}[\nu^2] &= \int_0^1 \nu^2 f_\nu(\nu)\,\mathrm{d}\nu\\ &= 2^B(M-1)\int_0^1 \nu^2(1-\nu)^{M-2}\left(1-(1-\nu)^{M-1}\right)^{2^B-1}\mathrm{d}\nu\\ &= 2^B\int_0^1\left(1-(1-x)^{\frac{1}{M-1}}\right)^2 x^{2^B-1}\,\mathrm{d}x\\ &= 1 - 2\cdot 2^B\mathrm{Beta}\left(2^B, \frac{M}{M-1}\right) + 2^B\mathrm{Beta}\left(2^B, \frac{M+1}{M-1}\right),\end{aligned} \tag{21}$$

where in the third line we employ the substitution $x = 1-(1-\nu)^M - 1$. The last line results from evaluating the square and from the definition of the Beta function.

A. Transmit Beamforming After Erroneous Feedback

After a feedback error at time slot $n-1$, there is a total feedback loss and the applied beamforming vector $\boldsymbol{w}_{\mathrm{DL}}[n] \neq \boldsymbol{w}_{\mathrm{FB}}[n]$ is random with respect to $\boldsymbol{h}_{\mathrm{DL}}[n]$. Thus, the effective channel after erroneous feedback is a zero-mean unit-variance complex Gaussian random variable, i.e.,

$$h_{\mathrm{DL,ef}}[n] = \boldsymbol{w}_{\mathrm{DL}}^{\mathrm{H}}[n]\boldsymbol{h}_{\mathrm{DL}}[n] \sim \mathcal{CN}(0,1). \tag{22}$$

Hence, the feedback is correct ($\boldsymbol{w}_{\mathrm{DL}}[n] = \boldsymbol{w}_{\mathrm{FB}}[n]$) with probability $1-p_\epsilon$ and the effective channel is (14). On the other hand, the feedback is erroneous with probability p_ϵ and the effective channel is (22). In the following, let the binary random variable θ indicate a feedback error event: $\theta[n-1]=1$ in case of a feedback error at time slot $n-1$ and $\theta[n-1]=0$ otherwise.

IV. Downlink Capacity Bounds

A lower bound on the ergodic downlink capacity C_{DL} assuming Gaussian signalling (i.e., we do not optimize the input distribution with the imperfect CSI) is provided by the next theorem.

Theorem 4.1: A lower bound on the downlink capacity C_{DL} of a limited feedback system taking into account estimation, quantization and feedback errors and outdating is given by

$$\begin{aligned}&C_{\mathrm{DL}}\\ &\geq \frac{(1-p_\epsilon)\log_2(\mathrm{e})}{T}\\ &\times\left(T_{\mathrm{DL,D}}\left(1-\frac{\sigma_\nu}{2\mu_\nu}\right)\mathrm{e}^{1/\chi}\sum_{k=1}^{M}\mathrm{E}_k\left(\frac{1}{\chi}\right) - \mathrm{e}^{\frac{1}{\xi}}\sum_{k=1}^{T_{\mathrm{DL,D}}}\mathrm{E}_k\left(\frac{1}{\xi}\right)\right)\\ &-\frac{H_{\mathrm{b}}(p_\epsilon)}{T}\\ &+\frac{p_\epsilon}{T}\log_2(\mathrm{e})\mathrm{e}^{\frac{\sigma_{\mathrm{n}}^2}{P_{\mathrm{DL}}}}\left(T_{\mathrm{DL,D}}\mathrm{E}_1\left(\frac{\sigma_{\mathrm{n}}^2}{P_{\mathrm{DL}}}\right) - \sum_{k=1}^{T_{\mathrm{DL,D}}}\mathrm{E}_k\left(\frac{\sigma_{\mathrm{n}}^2}{P_{\mathrm{DL}}}\right)\right)\\ &\triangleq C_{\mathrm{DL,lb}},\end{aligned} \tag{23}$$

where $\mathrm{E}_k(z) = \int_1^\infty(\mathrm{e}^{-zt}/t^k)\,\mathrm{d}t$ is the generalized exponential integral [28, (5.1.4)], $H_{\mathrm{b}}(\bullet)$ is the binary entropy function, $\chi = (P_{\mathrm{DL}}\alpha\mu_\nu(1-\sigma^2 e_{\mathrm{DL}}))/\sigma_{\mathrm{n}}^2$ and $\xi = (P_{\mathrm{DL}}\mathrm{E}[|h_{\mathrm{DL,cf}}[n]|^2])/\sigma_{\mathrm{n}}^2$.

Proof: The proof is given in Appendix A. □

An upper bound to the downlink capacity with Gaussian signalling is derived based on [25]:

$$\begin{aligned}C_{\mathrm{DL}} &\overset{(a)}{\leq} \frac{1}{T}I\left(\boldsymbol{s}_{\mathrm{DL}}[n];\boldsymbol{y}_{\mathrm{DL}}[n]\mid h_{\mathrm{DL}}[n],\theta[n-1]\right)\\ &\overset{(b)}{=} \frac{1}{T}(1-p_\epsilon)\,I\left(\boldsymbol{s}_{\mathrm{DL}}[n];\boldsymbol{y}_{\mathrm{DL}}[n]\mid h_{\mathrm{DL,cf}}[n]\right)\\ &\quad+\frac{1}{T}p_\epsilon I\left(\boldsymbol{s}_{\mathrm{DL}}[n];\boldsymbol{y}_{\mathrm{DL}}[n]\mid h_{\mathrm{DL,ef}}[n]\right)\end{aligned}$$

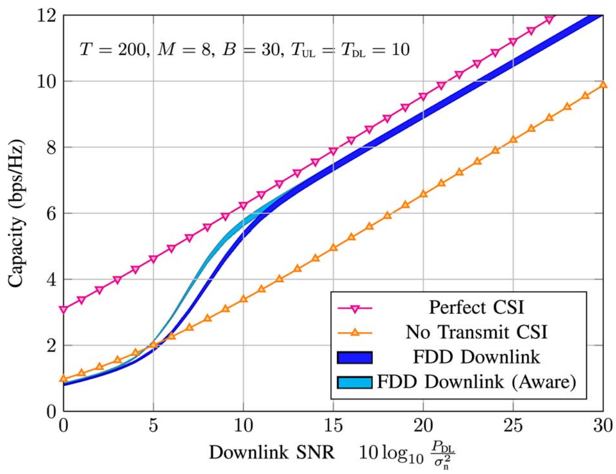

Fig. 2. Downlink capacity vs. SNR.

$$\begin{aligned}
&\overset{(c)}{=} \frac{T_{\mathrm{DL,D}}}{T}(1-p_\epsilon)\mathrm{E}_{h_{\mathrm{DL,cf}}}\left[\log_2\left(1+\frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2}|h_{\mathrm{DL,cf}}[n]|^2\right)\right] \\
&\quad + \frac{T_{\mathrm{DL,D}}}{T}p_\epsilon \log_2(\mathrm{e})\mathrm{e}^{\frac{\sigma_{\mathrm{n}}^2}{P_{\mathrm{DL}}}}\mathrm{E}_1\left(\frac{\sigma_{\mathrm{n}}^2}{P_{\mathrm{DL}}}\right) \\
&\overset{(d)}{\leq} \frac{T_{\mathrm{DL,D}}}{T}(1-p_\epsilon)\log_2\left(1+\frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2}(\alpha(M\mathrm{E}[\nu]-1)(1-\sigma_e^2)+1)\right) \\
&\quad + \frac{T_{\mathrm{DL,D}}}{T}p_\epsilon \log_2(\mathrm{e})\mathrm{e}^{\frac{\sigma_{\mathrm{n}}^2}{P_{\mathrm{DL}}}}\mathrm{E}_1\left(\frac{\sigma_{\mathrm{n}}^2}{P_{\mathrm{DL}}}\right) \\
&\triangleq C_{\mathrm{DL,ub}}, \qquad (24)
\end{aligned}$$

where step (a) follows by assuming a genie provides the effective downlink channel $h_{\mathrm{DL}}[n]$ and $\theta[n-1]$ to the user, such that the user is aware of feedback errors. Step (b) results from the effective channel with correct and erroneous feedback. Given the effective downlink channel, either $h_{\mathrm{DL,cf}}[n]$ or $h_{\mathrm{DL,ef}}[n]$, Gaussian signalling is optimum, which has been employed in step (c), where we also used (58). In step (d), the first term results from Jensen's inequality and (17).

V. Performance Analysis

To assess the lower bound, consider the following scenario: $T = 200$, $M = 8$, $B = 30$, $T_{\mathrm{UL}} = T_{\mathrm{DL}} = 10$, and $\alpha = 0.9994$.[5] We also set $P_{\mathrm{UL}} = P_{\mathrm{DL}}/M$, i.e., the user has $1/M$-th the available power of the BS, to introduce a scaling of the transmit power with the number of antennas. With the assumption $T_{\mathrm{UL}} = T_{\mathrm{DL}}$, the estimation error (c.f. (3) and (8)) in the uplink and downlink are the same for a given power setting. Although we focus on the downlink, the uplink parameters determine the feedback error probability. For the described scenario, Fig. 2 depicts the downlink capacity bounds (23) and (24) as a function of the downlink SNR. The actual capacity with imperfect CSI lies in the shaded region enclosed by the bounds. The tightness of the lower bound (23) can clearly be observed. As references we show the capacity with perfect CSI and an upper bound on the capacity without transmit CSI [1]. Furthermore, we note the slightly increasing loss with SNR with respect to the perfect CSI case, resulting from the training overhead $(T - T_{\mathrm{DL}})/T < 1$.

For small SNRs the downlink capacity, when the BS is unaware of feedback errors, can be smaller than the capacity without transmit CSI. For such SNRs, the downlink capacity is dominated by the capacity with erroneous feedback, since the occurrence of feedback errors is not negligible. For a downlink SNR of 4 dB, the SNR in the uplink is approximately -5 dB (recall $M = 8$ and $P_{\mathrm{UL}} = P_{\mathrm{DL}}/M$). With uplink SNRs less than -5 dB and $B = 30$, the feedback error probability p_ϵ is larger than 92%! In such cases, the performance in the downlink is equivalent to having only one transmit antenna. As the uplink SNR increases, the feedback error probability decreases such that for uplink SNRs larger than 1 dB, i.e., downlink SNRs larger than 10 dB with $M = 8$ and $P_{\mathrm{UL}} = P_{\mathrm{DL}}/M$, the feedback error probability p_ϵ is already less than 12%.

Since we assume the BS is unaware of feedback errors, after erroneous feedback the BS sends only one stream with a random beamforming vector instead of sending M independent streams with equally shared power, which is the optimum strategy when the BS has no transmit CSI [1]. One way to increase the performance of the downlink in such cases would be to enable the BS to be aware of feedback errors,[6] so the BS can transmit with the previous beamforming vector if this was received error-free. The capacity bounds for this case are also depicted in the plot with the legend (Aware). For moderate to high SNR, there is practically no gain of being aware of feedback errors, since the feedback error probability is negligible. However, even if the BS is aware of the feedback errors (as described before) for very low SNRs, the optimum transmit strategy for the no transmit CSI case could be slightly better. Besides enabling the BS to become aware of feedback errors, another straightforward approach to counter the feedback errors is to increase the uplink SNR in order to reduce the feedback error probability.

We now plot the FDD capacity bounds as a function of the number of feedback bits B in Fig. 3 for $T = 500$, $M = 8$, $T_{\mathrm{UL}} = T_{\mathrm{DL}} = 10$, and $\alpha = 0.9994$. The downlink SNR is 10 dB and consequently with $P_{\mathrm{UL}} = P_{\mathrm{DL}}/M$ the UL SNR ≈ 4 dB. The tradeoff between the quantization error (number of feedback bits) and the feedback error probability in the downlink is visible. The maximum for both capacity bounds is achieved with around 22–26 bits. Hence, we posit the optimum number of feedback bits for the actual FDD downlink capacity with imperfect CSI to be around 22–26 bits. In addition, as the number of feedback bits increase, the reduction in the quantization error ($\mu_\nu \to 1$) cannot compensate for the increase in the transmission errors in the feedback link, which consequently leads to a decrease of the downlink capacity.

[5] Assuming a Jake's spectrum for the Rayleigh fading we would have $\sqrt{\alpha} = \mathrm{J}_0(2\pi t_{\mathrm{s}} f_{\mathrm{c}} s/c)$, where $\mathrm{J}_0(\bullet)$ is the zeroth-order Bessel function of the first kind, such that the value of α results from assuming a time slot duration $t_{\mathrm{s}} = 1$ ms, a carrier frequency $f_{\mathrm{c}} = 2$ GHz and a user's velocity along the line from the BS to the user of $s = 3$ km/h.

[6] We do not discuss how the BS could achieve this, but one possible naive approach would be for the BS to compare each fed back beamforming vector with previously received beamforming vectors. Assuming that the feedback error probability is not so large, the BS could determine to some extent whether the current feedback message was received with error if the corresponding beamforming vector is not correlated with the beamforming vectors from previously received feedback messages.

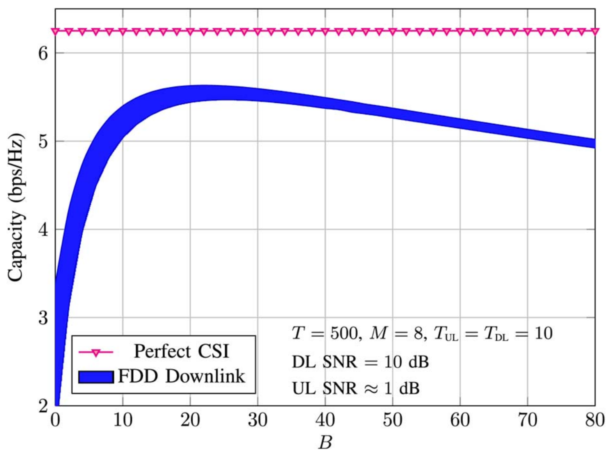

Fig. 3. Downlink capacity vs. number of feedback bits B.

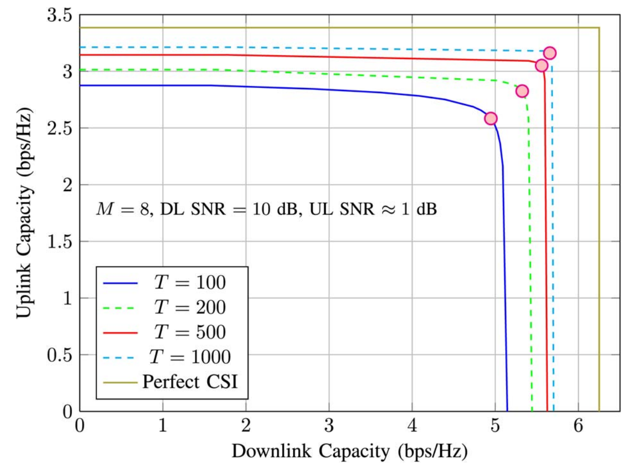

Fig. 4. Achievable uplink-downlink rate region.

Based on the results, we assert the tightness of the bounds mainly depends on the value of the ratio B/M. The bounds are looser for smaller values of B/M and tighter for larger values of B/M. This is a result of the quantization error scaling with $2^{-B/(M-1)}$ (see (19)). In addition, since Jensen's inequality becomes tighter with increasing number of antennas M, the upper bound (24) becomes tighter with increasing M. This is due to the fact that the magnitude square of the effective downlink channel after correct feedback becomes more deterministic.

VI. Two-Way System

To properly characterize the FDD system, we take into account the FDD uplink as well as the FDD downlink. To this end, a lower bound on the ergodic capacity C_{UL} of the uplink (given only an estimate of the uplink channel) can be obtained by assuming the term from the estimation error, i.e., $\boldsymbol{w}_{\text{UL}}^{\text{H}}[n]\boldsymbol{e}_{\text{UL}}[n]\boldsymbol{s}_{\text{UL}}[n]$ in (4) behaves as worst case noise [23]–[25]:

$$C_{\text{UL}} \geq \frac{T-T_{\text{UL}}-T_{\text{F}}}{T}\mathrm{E}_{\hat{\boldsymbol{h}}_{\text{UL}}}\left[\log_2\left(1+\frac{P_{\text{UL}}\|\hat{\boldsymbol{h}}_{\text{UL}}\|_2^2}{P_{\text{UL}}\sigma_{e_{\text{UL}}}^2+\sigma_{\text{n}}^2}\right)\right] \quad (25)$$

$$= \frac{T-T_{\text{UL}}-T_{\text{F}}}{T}\log_2(\mathrm{e})\mathrm{e}^{1/\lambda_{\text{UL}}}\sum_{k=1}^{M}\mathrm{E}_k\left(\frac{1}{\lambda_{\text{UL}}}\right)$$

$$\triangleq C_{\text{UL,lb}}, \quad (26)$$

which follow from [23, (17)] and [31, (8.40)] with [28, (5.1.45)]. Recall λ_{UL} is given in (11).

For the FDD system design, i.e., for determining the optimum values of the training length and number of feedback bits, we propose using the lower bounds of the uplink and downlink which represent achievable rates in each link. Hence, we have a bi-objective optimization problem

$$\max_{T_{\text{UL}},T_{\text{DL}},B}\ (C_{\text{UL,lb}}, C_{\text{DL,lb}}) \quad \text{s.t.} \quad 0 \leq T_{\text{DL}} \leq T,$$

$$0 \leq T_{\text{UL}} + \frac{B}{2} \leq T, \quad (27)$$

where the bounds are given in (26) and (23). The solution to the optimization (27) gives a set of Pareto optimal points. In case the rate region formed by uplink and downlink achievable rate is concave, we can obtain the Pareto optimal points from the single-objective optimization [32]:

$$\max_{T_{\text{UL}},T_{\text{DL}},B}\ \beta C_{\text{UL,lb}} + (1-\beta)C_{\text{DL,lb}} \quad \text{s.t.} \quad 0 \leq T_{\text{DL}} \leq T,$$

$$0 \leq T_{\text{UL}} + \frac{B}{2} \leq T, \quad (28)$$

for $\beta \in [0,1]$. We can solve the previous optimization numerically for $\beta \in [0,1]$ in order to obtain pairs of achievable rates in the uplink and the downlink [33], [34]. The set of Pareto optimal points for the optimization (27) is shown in Fig. 4 for different time slot lengths T. For the figure we have additionally assumed $M = 8$, $\alpha = 1$ (no outdating) and as in the previous plot, a downlink SNR = 10 dB and uplink SNR ≈ 1 dB. From the figure, we observe that the rate region formed by the uplink and downlink achievable rate is indeed concave.

The coupling between the FDD uplink and downlink due to the feedback overhead is only significant for short time slot lengths. For instance, the maximum uplink achievable rate is obtained by setting $B = 0$, i.e., no uplink resources are employed to aid the downlink. However, this does not imply that the downlink rate is zero, since the downlink can still achieve a positive rate corresponding to that with no CSI or that of random beamforming with a single user. On the other hand, the maximum downlink achievable rate is obtained by using the uplink solely as a feedback link which leads to a zero rate for the uplink. For the case $T = 1000$, the optimum resource allocation which maximizes solely the downlink achievable rate is $B = 30$, $T_{\text{DL}} = 13$, and $T_{\text{UL}} = 985$. Although there are plenty of resources to have a larger B, we still have the tradeoff due to the feedback error probability shown in Fig. 3. For example, increasing the number of feedback bits from $B = 30$ to $B = 60$ increases the CDI quantization quality μ_ν slightly (from 0.9520 to 0.9975), which in turn increases the term in the capacity corresponding to correct feedback. This increase in B, nevertheless, almost doubles the feedback error probability from 8.72% to 16.68% (c.f. (12)), such that $B = 30$ is the optimum choice! The rest of the time slot is employed for uplink training to obtain the smallest possible uplink estimation error.

For the system design of an FDD downlink/uplink (two-way system), there is no single figure of merit as in a one-way system. Similar to [27], we consider the sum rate of the achievable rate on both links as an appropriate figure of merit for optimizing the FDD system

$$\{T_{\mathrm{UL}}^\star, T_{\mathrm{DL}}^\star, B^\star\} = \underset{T_{\mathrm{UL}}, T_{\mathrm{DL}}, B}{\arg\max} \quad C_{\mathrm{UL,lb}} + C_{\mathrm{DL,lb}}$$

$$\text{subject to} \quad 0 \le T_{\mathrm{DL}} \le T, \quad 0 \le T_{\mathrm{UL}} + \frac{B}{2} \le T, \tag{29}$$

where $C_{\mathrm{UL,lb}}$ and $C_{\mathrm{DL,lb}}$ are given in (26) and (23). This operating point is indicated with a circle for different T in Fig. 4. Due to the fact that the coupling between the two links is only significant for small values of T, for large T there is practically no capacity loss for each link when optimizing each link separately, as compared to optimizing the links jointly through the uplink-downlink sum rate optimization. Lastly, we note that the gap in the uplink with respect to the perfect CSI case is smaller than in the downlink due to the fact that the quality of the CSI in the downlink is worse, i.e., in addition to estimation errors, it is subject to quantization and feedback errors. Although the capacity with perfect CSI of both links is equal for the same SNR, the depicted downlink capacity with perfect CSI is higher because we assume $P_{\mathrm{UL}} = P_{\mathrm{DL}}/M$.

VII. System Design

Based on the previous discussion, we propose optimizing the uplink and downlink separately to determine the optimum training and number of feedback bits. As will be seen, the optimization of the training lengths can be easily decoupled. When optimizing the number of feedback bits, however, we have to consider two aspects of the feedback: feedback errors and the overhead.

A. Uplink Training Length

Although the optimization of the training length has been addressed in [23], no closed form expression has been derived there. To optimize the uplink training length, we find the uplink training length which maximizes the uplink achievable rate by ignoring the feedback overhead in the uplink. Hence, from (26) and with $T_{\mathrm{F}} = 0$, we can bound $C_{\mathrm{UL,lb}}$

$$\frac{T - T_{\mathrm{UL}}}{T} \log_2(\mathrm{e})\, \mathrm{e}^{\frac{1}{\lambda_{\mathrm{UL}}}} \sum_{k=1}^{M} \mathrm{E}_k\left(\frac{1}{\lambda_{\mathrm{UL}}}\right) \ge \frac{T - T_{\mathrm{UL}}}{T} \log_2(\mathrm{e})\, \mathrm{e}^{\frac{1}{\lambda_{\mathrm{UL}}}} \mathrm{E}_1\left(\frac{1}{\lambda_{\mathrm{UL}}}\right). \tag{30}$$

where λ_{UL} given in (11) depends on T_{UL}. The inequality results from $\mathrm{e}^{1/\lambda_{\mathrm{UL}}}\mathrm{E}_k(1/\lambda_{\mathrm{UL}}) \ge 0$. Also, it turns out the uplink training length T_{UL} which maximizes each of the summands $\mathrm{e}^{1/\lambda_{\mathrm{UL}}}\mathrm{E}_k(1/\lambda_{\mathrm{UL}})$ (which are *concave* functions of T_{UL}) is approximately the same. For simplicity, we focus solely on the first summand $k = 1$. With [28, (5.1.20)], we can bound this summand as

$$\frac{T - T_{\mathrm{UL}}}{T} \mathrm{e}^{\frac{1}{\lambda_{\mathrm{UL}}}} \mathrm{E}_1\left(\frac{1}{\lambda_{\mathrm{UL}}}\right) \ge \frac{T - T_{\mathrm{UL}}}{T} \log(1 + \lambda_{\mathrm{UL}}). \tag{31}$$

The uplink training length T_{UL} which maximizes $\mathrm{e}^{1/\lambda_{\mathrm{UL}}}\mathrm{E}_1(1/\lambda_{\mathrm{UL}})$ turns out to be approximately the same as the T_{UL} which maximizes $\log(1 + \lambda_{\mathrm{UL}})$. Hence, we posit that maximizing $C_{\mathrm{UL,lb}}$ over T_{UL} is equivalent to maximizing

$$(T - T_{\mathrm{UL}}) \log\left(\frac{1 + \rho_{\mathrm{UL}} + T_{\mathrm{UL}}\left(\rho_{\mathrm{UL}}^2 + \rho_{\mathrm{UL}}\right)}{1 + \rho_{\mathrm{UL}} + T_{\mathrm{UL}}\rho_{\mathrm{UL}}}\right), \tag{32}$$

which follows by substituting λ_{UL} from (11) in (31) with $\sigma_{e_{\mathrm{UL}}}^2$ from (3) and the UL SNR $\rho_{\mathrm{UL}} = P_{\mathrm{UL}}/\sigma_{\mathrm{n}}^2$. Since (32) is concave with respect to T_{UL}, we take the derivative of (32) with respect to T_{UL} and set it to zero to find the optimum uplink training length. Setting the derivative to zero gives us

$$-\log\left(1 + \frac{\rho_{\mathrm{UL}}^2 T_{\mathrm{UL}}}{1 + \rho_{\mathrm{UL}} + \rho_{\mathrm{UL}} T_{\mathrm{UL}}}\right) + \frac{\rho_{\mathrm{UL}}^2 (T - T_{\mathrm{UL}})}{1 + \rho_{\mathrm{UL}} + \left(\rho_{\mathrm{UL}}^2 + 2\rho_{\mathrm{UL}}\right) T_{\mathrm{UL}} + \rho_{\mathrm{UL}}^2 T_{\mathrm{UL}}^2} = 0. \tag{33}$$

With $\rho_{\mathrm{UL}} T_{\mathrm{UL}} \gg 1 + \rho_{\mathrm{UL}}$, which holds since usually $\rho_{\mathrm{UL}} T_{\mathrm{UL}} < 1$, we have $\log(1 + (\rho_{\mathrm{UL}}^2 T_{\mathrm{UL}}/(1 + \rho_{\mathrm{UL}} + \rho_{\mathrm{UL}} T_{\mathrm{UL}}))) \approx \log(1 + \rho_{\mathrm{UL}})$, such that the first term in (33) does not depend on T_{UL}. Employing this approximation in (33), we can find an approximate solution to (33) by solving:

$$T_{\mathrm{UL}}^2 + \left(\left(1 + \frac{2}{\rho_{\mathrm{UL}}}\right) + \frac{1}{\log(1 + \rho_{\mathrm{UL}})}\right) T_{\mathrm{UL}} + \left(\frac{1}{\rho_{\mathrm{UL}}^2} + \frac{1}{\rho_{\mathrm{UL}}}\right) - \frac{T}{\log(1 + \rho_{\mathrm{UL}})} = 0,$$

whose only positive solution represents an approximation for the optimum UL training length:

$$\begin{aligned} T_{\mathrm{UL}}^\star &\approx \sqrt{\frac{T}{\log(1+\rho_{\mathrm{UL}})} + \frac{\frac{1}{2} + \frac{1}{\rho_{\mathrm{UL}}}}{\log(1+\rho_{\mathrm{UL}})} + \frac{1}{4\log^2(1+\rho_{\mathrm{UL}})} + \frac{1}{4}} - \left(\frac{1}{2} + \frac{1}{\rho_{\mathrm{UL}}} + \frac{1}{2\log(1+\rho_{\mathrm{UL}})}\right) \\ &\approx \sqrt{\frac{T}{\log\left(1 + \frac{P_{\mathrm{UL}}}{\sigma_{\mathrm{n}}^2}\right)}} - \left(\frac{1}{2} + \frac{1}{\frac{P_{\mathrm{UL}}}{\sigma_{\mathrm{n}}^2}} + \frac{1}{2\log\left(1 + \frac{P_{\mathrm{UL}}}{\sigma_{\mathrm{n}}^2}\right)}\right), \end{aligned} \tag{34}$$

which is independent of M, scales $\mathcal{O}(\sqrt{T})$ and is inversely proportional to the UL SNR $\rho_{\mathrm{UL}} = P_{\mathrm{UL}}/\sigma_{\mathrm{n}}^2$.

B. Downlink Training Length

To find the optimum downlink training length, we simplify the analysis by considering solely (56), i.e., the correct feedback component of the downlink achievable rate under the assumption of no feedback errors, no quantization and no outdating, i.e., by setting $\alpha = 1$, $\mu_\nu = 1$ and $\sigma_\nu = 0$. Considering only (56) with the previous assumptions, we can write

$$\frac{T - T_{\mathrm{DL}}}{T} \log_2(\mathrm{e}) \, \mathrm{e}^{\frac{1}{\chi}} \sum_{k=1}^{M} \mathrm{E}_k \left(\frac{1}{\chi}\right) \geq \frac{T - T_{\mathrm{DL}}}{T} \log_2(\mathrm{e}) \, \mathrm{e}^{\frac{1}{\chi}} \mathrm{E}_1 \left(\frac{1}{\chi}\right), \quad (35)$$

where the inequality follows as (30) in the uplink. In addition, χ which is defined in (57), with the mentioned assumptions results in

$$\chi \overset{\substack{\mu_\nu = 1 \\ \alpha = 1}}{=} \frac{P_{\mathrm{DL}}(1 - \sigma_{e_{\mathrm{DL}}}^2)}{\sigma_{\mathrm{n}}^2}. \quad (36)$$

Similarly as in the uplink, maximizing (35) over T_{DL} is approximately equivalent to maximizing

$$(T - T_{\mathrm{DL}}) \log \left(1 + \frac{\rho_{\mathrm{DL}} T_{\mathrm{DL}} \frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2}}{1 + \rho_{\mathrm{DL}} T_{\mathrm{DL}}}\right), \quad (37)$$

which follows from $\rho_{\mathrm{DL}} = P_{\mathrm{DL}}/M\sigma_{\mathrm{n}}^2$ and (31) and by substituting χ from (36) with $\sigma_{e_{\mathrm{DL}}}^2$ from (8).

We can obtain an approximate solution which maximizes (35) by taking the derivative of (37) with respect to T_{DL} and setting it to zero. After taking the derivative and approximating $\log(1 + (\rho_{\mathrm{DL}} T_{\mathrm{DL}}(P_{\mathrm{DL}}/\sigma_{\mathrm{n}}^2)/(1 + \rho_{\mathrm{DL}} T_{\mathrm{DL}}))) \approx \log(1 + (P_{\mathrm{DL}}/\sigma_{\mathrm{n}}^2))$ we obtain the following quadratic equation in T_{DL}

$$\rho_{\mathrm{DL}}^2 \left(1 + \frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2}\right) T_{\mathrm{DL}}^2 + \rho_{\mathrm{DL}} \left(2 + \frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2} + \frac{\frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2}}{\log\left(1 + \frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2}\right)}\right) T_{\mathrm{DL}} + 1 - \frac{\rho_{\mathrm{DL}} \frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2}}{\log\left(1 + \frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2}\right)} T = 0,$$

which can be rewritten as

$$T_{\mathrm{DL}}^2 + \frac{1}{1 + \frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2}} \left(\frac{2}{\frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2}} + 1 + \frac{1}{\log\left(1 + \frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2}\right)}\right) M T_{\mathrm{DL}} + \frac{M^2}{\left(\frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2}\right)^2 \left(1 + \frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2}\right)} - \frac{T\, M}{\left(1 + \frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2}\right) \log\left(1 + \frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2}\right)} = 0. \quad (38)$$

To solve the previous quadratic equation, we consider the fact that

$$\frac{1}{1 + \frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2}} \left(\frac{2}{\frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2}} + 1 + \frac{1}{\log\left(1 + \frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2}\right)}\right) \ll 1 \quad (39)$$

$$\frac{1}{\left(\frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2}\right)^2 \left(1 + \frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2}\right)} \ll 1, \quad (40)$$

which hold for practical SNRs. With $T_{\mathrm{DL}} \geq M$ (assumed in (8) for $\sigma_{e_{\mathrm{DL}}}^2$), (39) and (40), we can neglect the second and third term in (38) and approximate the optimum downlink training length

$$T_{\mathrm{DL}}^\star \approx \max \left(M, \sqrt{\frac{TM}{\left(1 + \frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2}\right) \log\left(1 + \frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2}\right)}}\right), \quad (41)$$

where we employ the $\max$ operator, since the positive root could be less than M but we assumed $T_{\mathrm{DL}} \geq M$ for $\sigma_{e_{\mathrm{DL}}}^2$ in (8). In contrast to the uplink, the downlink training does depend on M.

C. Number of Feedback Bits

To find an approximation of the optimum $B^\star$, we ignore the effects of the estimation and outdating, i.e., $\sigma_{e_{\mathrm{DL}}}^2 = 0$ and $\alpha = 1$. In addition, we focus on the correct feedback component of the downlink achievable rate with the correct feedback probability. Although the feedback overhead is in the uplink, we incorporate it in the downlink as in [15], [16], to perform the optimization only in the downlink. We also set $\sigma_\nu = 0$, and with $T_{\mathrm{F}} = B/2$, we assume $T_{\mathrm{DL,D}} = T - (B/2)$ instead of $T_{\mathrm{DL,D}} = T - T_{\mathrm{DL}}$. With these assumptions, based on the first term in (23), we have

$$\frac{T - \frac{B}{2}}{T} (1 - p_\epsilon) \log_2(\mathrm{e}) \, \mathrm{e}^{\frac{1}{\chi}} \sum_{k=1}^{M} \mathrm{E}_k \left(\frac{1}{\chi}\right) \geq \frac{T - \frac{B}{2}}{T} (1 - p_\epsilon) \log_2(\mathrm{e}) \, \mathrm{e}^{\frac{1}{\chi}} \mathrm{E}_1 \left(\frac{1}{\chi}\right), \quad (42)$$

where χ from (57) is given in this case by

$$\chi \overset{\substack{\sigma_{e_{\mathrm{DL}}}^2 = 0 \\ \alpha = 1}}{=} \frac{P_{\mathrm{DL}} \mu_\nu}{\sigma_{\mathrm{n}}^2} \approx \frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2} \left(1 - 2^{-\frac{B}{M-1}}\right),$$

and where the approximation follows from (19). The expression (42) considers the three consequences of the feedback: feedback overhead, feedback errors, and the quantization error.

Following the procedure employed before, maximizing (42) is equivalent to maximizing

$$\left(T - \frac{B}{2}\right) (1 - \overline{p_b})^B \log \left(1 + \frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2} \left(1 - 2^{-\frac{B}{M-1}}\right)\right), \quad (43)$$

which follows from (42) using (31) with χ given above instead of λ_{UL} and approximating p_ϵ with (12), where we have replaced $\mathrm{E}[p_{\mathrm{b}}(\overline{\gamma}_{\mathrm{F}}[n])]$ with $\overline{p_{\mathrm{b}}}$. We recall the bit error probability defined in (10) depends on the uplink estimation error variance, uplink SNR and number of antennas M at the BS. We compute this bit error probability assuming the optimum training length given in (34). By taking the derivative of (43) with respect to B and setting it to zero we obtain

$$\begin{aligned}&\left(\frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2}\left(1-2^{-\frac{B}{M-1}}\right)\right)\\&\times\left(\log\left(1+\frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2}\right)+\log\left(1-\frac{\frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2}}{1+\frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2}}2^{-\frac{B}{M-1}}\right)\right)\\&\times\left(-\frac{1}{2}+\left(T-\frac{B}{2}\right)\log\left(1-\overline{p_{\mathrm{b}}}\right)\right)+\log 2\frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2}\frac{T-\frac{B}{2}}{M-1}\\&+\log 2\frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2}\frac{T-\frac{B}{2}}{M-1}\left(1-2^{-\frac{B}{M-1}}\right)=0,\end{aligned}$$

where after employing the approximation

$$\log\left(1-\frac{\frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2}}{1+\frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2}}2^{-\frac{B}{M-1}}\right)\approx-\frac{\frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2}}{1+\frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2}}2^{-\frac{B}{M-1}},$$

that is valid for $(P_{\mathrm{DL}}/\sigma_{\mathrm{n}}^2)/(1+(P_{\mathrm{DL}}/\sigma_{\mathrm{n}}^2))2^{-(B/(M-1))}\ll 1$, which holds for practical values of B, we can write

$$\begin{aligned}&\left(\frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2}\left(1-2^{-\frac{B}{M-1}}\right)\right)\left(\log\left(1+\frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2}\right)\right.\\&\left.-\frac{\frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2}}{1+\frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2}}2^{-\frac{B}{M-1}}+\left(\frac{\frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2}}{1+\frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2}}-\frac{\frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2}}{1+\frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2}}\right)\right)\\&\times\left(-\frac{1}{2}+\left(T-\frac{B}{2}\right)\log\left(1-\overline{p_{\mathrm{b}}}\right)\right)\\&+\log 2\frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2}\frac{T-\frac{B}{2}}{M-1}+\log 2\frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2}\frac{T-\frac{B}{2}}{M-1}\left(1-2^{-\frac{B}{M-1}}\right)=0.\end{aligned}$$

Furthermore, with

$$x=\frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2}\left(1-2^{-\frac{B}{M-1}}\right), \tag{44}$$

we can rewrite the previous equation to obtain the following quadratic equation in x

$$\begin{aligned}&x^2+\left(1-\frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2}+\left(1+\frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2}\right)\log\left(1+\frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2}\right)\right.\\&\qquad\left.-\frac{\log 2\left(1+\frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2}\right)}{(M-1)\left(\log(1-\overline{p_{\mathrm{b}}})-\frac{1}{2T-B)}\right)}\right)x\\&\quad-\frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2}+\left(1+\frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2}\right)\log\left(1+\frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2}\right)\\&\quad-\frac{\log 2\left(1+\frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2}\right)}{(M-1)\left(\log(1-\overline{p_{\mathrm{b}}})-\frac{1}{2T-B}\right)}=0.\end{aligned}$$

By approximating $1/(2T-B)\approx 1/(2T)$ in the denominator of the last term, we have a quadratic equation in x which is a function of B. Solving this equation shows us that the only positive root,[7] which at the end leads to a positive value for the number of feedback bits B, reads as (see equation at the bottom of the page). With

$$N=\left(\frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2}+1\right)\left(\log\left(\frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2}+1\right)+1-\frac{\frac{\log 2}{M-1}}{\log(1-\overline{p_{\mathrm{b}}})-\frac{1}{2T}}\right) \tag{46}$$

$$d=-4\left(\frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2}+1\right)^2\log\left(\frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2}+1\right), \tag{47}$$

we can simplify the term with the square root in (45) as:

$$\sqrt{N^2+d}\approx N+\frac{d}{2N}, \tag{48}$$

which follows from the first-order Taylor series of the square root function. Hence, by employing the approximation (48) with (46) and (47), we simplify the solution (45) as follows

$$x\approx\frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2}-\frac{\left(\frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2}+1\right)\log\left(\frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2}+1\right)}{\log\left(\frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2}+1\right)+1-\frac{\log 2}{(M-1)\left(\log(1-\overline{p_{\mathrm{b}}})-\frac{1}{2T}\right)}},$$

[7]With $x=(P_{\mathrm{DL}}/\sigma_{\mathrm{n}}^2)(1-2^{-(B/(M-1))})$ we have that $B=-(M-1)\log_2(1-(x/(P_{\mathrm{DL}}/\sigma_{\mathrm{n}}^2)))$, such that $x\le 0$ leads to $B\le 0$.

$$\begin{aligned}x=&-\frac{1}{2}\left(1-\frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2}+\left(1+\frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2}\right)\log\left(1+\frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2}\right)-\frac{\log 2\left(1+\frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2}\right)}{M-1\left(\log(1-\overline{p_{\mathrm{b}}})-\frac{1}{2T}\right)}\right)\\&+\frac{1+\frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2}}{2}\times\sqrt{\left(\log\left(1+\frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2}\right)+1-\frac{\frac{\log 2}{(M-1)}}{\log(1-\overline{p_{\mathrm{b}}})-\frac{1}{2T}}\right)^2-4\log\left(1+\frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2}\right)}\end{aligned} \tag{45}$$

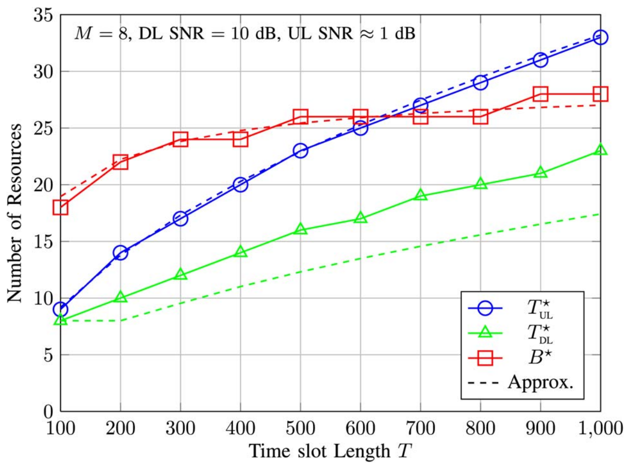

Fig. 5. Optimum resource allocation vs. T.

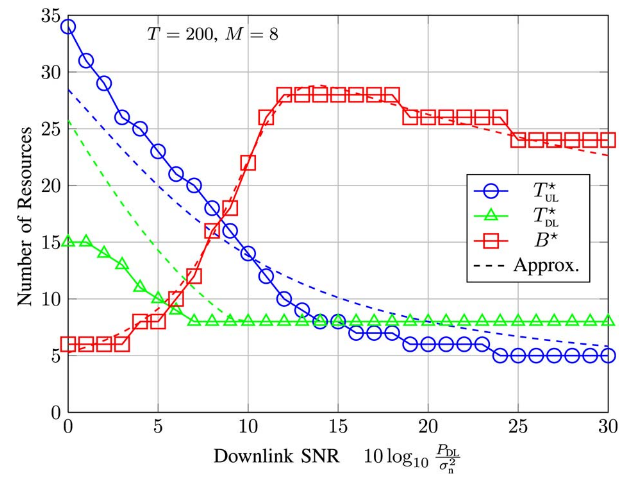

Fig. 6. Optimum resource allocation vs. SNR.

such that with (44), we obtain an approximation for the optimum number of feedback bits:

$$B^\star = (M-1)\log_2\left(\frac{1}{1+\frac{\sigma_{\text{n}}^2}{P_{\text{DL}}}} + \frac{1}{\left(1+\frac{\sigma_{\text{n}}^2}{P_{\text{DL}}}\right)\log\left(\frac{P_{\text{DL}}}{\sigma_{\text{n}}^2}+1\right)} - \frac{\frac{\log 2}{M-1}}{\left(1+\frac{\sigma_{\text{n}}^2}{P_{\text{DL}}}\right)\log\left(\frac{P_{\text{DL}}}{\sigma_{\text{n}}^2}+1\right)\left(\log(1-\overline{p_{\text{b}}})-\frac{1}{2T}\right)}\right), \quad (49)$$

which scales linearly with M and increases quite slowly with the time slot length T, as will be observed in the simulation results. The scaling with $P_{\text{DL}}/\sigma_{\text{n}}^2$ depends on the SNR region.

VIII. Simulation Results and Discussion

Although several approximations were made to obtain the optimum uplink training (34), downlink training length (41) and number of feedback bits (49), the expressions are very close to the optimum values computed numerically, i.e., by solving (29) via an exhaustive search. To this end, we plot the optimum resource allocation obtained numerically along with the derived approximations as a function of different system parameters. For the optimum B obtained numerically, we assume only even values, since we consider QPSK feedback symbols. In Fig. 5, the optimum values are plotted as a function of the time slot length T for $M = 8$, $\alpha = 1$, $P_{\text{UL}} = P_{\text{DL}}/M$ and DL SNR $= 10$ dB. The results confirm the scaling of the training lengths with $\sqrt{T}$ derived in the approximations and also in [35]. Note that with $P_{\text{UL}} = P_{\text{DL}}/M$, $\sigma^2_{e_{\text{DL}}} = \sigma^2_{e_{\text{UL}}}$ for $T_{\text{DL}} = T_{\text{UL}}$, with $T_{\text{DL}} \geq M$. Since the optimum uplink training length is larger than the optimum downlink training length for each T, we see the uplink channel needs to be estimated better than the downlink channel. Recall that for the lower bound on the uplink channel capacity we have considered the estimation error as additive noise.

From Fig. 5, we also observe the optimum number of feedback bits increases very slowly with T. Although [21] suggests the feedback load should increase with $\log_2 T$, this could be due to the fact that the feedback error probability was not considered. There is a point after which increasing B leads solely to a slight reduction of the quantization error which does not compensate the increase in the feedback error probability. Under the assumed setting, as $T \to \infty$ the optimum feedback load actually converges to a constant

$$B^\star \approx (M-1)\log_2\left(\frac{-\frac{1}{M-1}}{\left(1+\frac{\sigma_{\text{n}}^2}{P_{\text{DL}}}\right)\log\left(\frac{P_{\text{DL}}}{\sigma_{\text{n}}^2}+1\right)\log(1-\overline{p_{\text{b}}})}\right) \approx 28$$

feedback bits, which follows by setting $1/(2T) = 0$ in (49), since the third term inside the logarithm dominates the first two with the assumed setting. In general, for a given M and SNR, B converges to a constant as $T \to \infty$ and does not increase unboundedly.

We now depict in Fig. 6 the numerical and approximated optimum values for T_{UL}, T_{DL} and B as a function of the downlink SNR $P_{\text{DL}}/\sigma_{\text{n}}^2$ in dB with the uplink SNR given by $P_{\text{UL}}/\sigma_{\text{n}}^2 = P_{\text{DL}}/(M\sigma_{\text{n}}^2)$. We further assume $T = 200$, $M = 8$, and $\alpha = 1$. We can see the training lengths scale inversely proportional to the SNR until converging to the minimum value ($M = 8$ for the downlink and 1 for the uplink at higher SNRs), since the channel can be estimated well enough with the minimum training length at high SNRs. As for the optimum B, we observe an interesting behavior. It first increases swiftly with the SNR and after reaching a maximum value, decreases slowly at high SNRs. For low downlink SNRs, which implies even lower uplink SNRs due to $P_{\text{UL}}/\sigma_{\text{n}}^2 = P_{\text{DL}}/(M\sigma_{\text{n}}^2)$, the optimum number of feedback bits is small, since the bit error probability in the feedback link is very large. As the SNR increases and the feedback link becomes more reliable, the optimum B also increases until the quantization error is small enough, such that any further decreaser with additional bits does not compensate the loss due to feedback overhead and increase of feedback errors. As the SNR increases further, the optimum B decreases very slowly as indicated by the approximation given in (49). Hence, for a fixed T, the downlink channel vector can be quantized more coarsely as the SNR increases due to the increase of the available power.

Fig. 7 depicts the optimum resource allocation as a function of the number M of antennas at the BS. We assume $T = 200$,

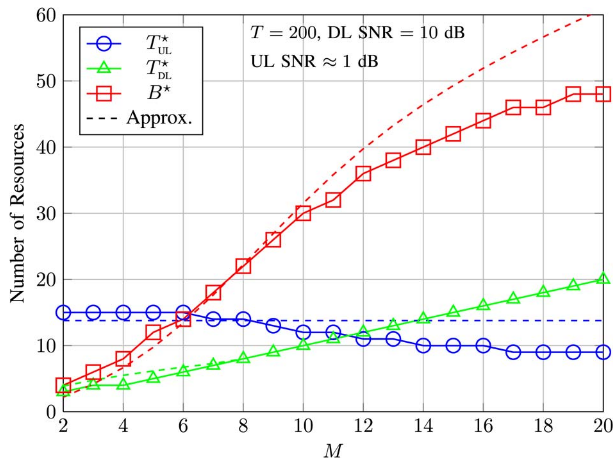

Fig. 7. Optimum resource allocation vs. number of antennas M.

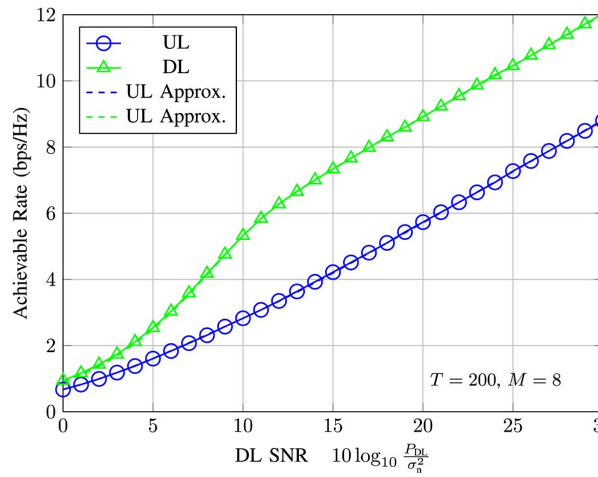

Fig. 9. Optimized rates vs. SNR.

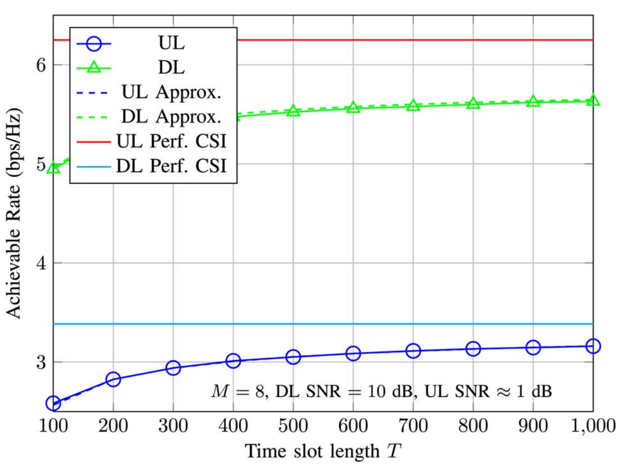

Fig. 8. Optimized rates vs. time slot length T.

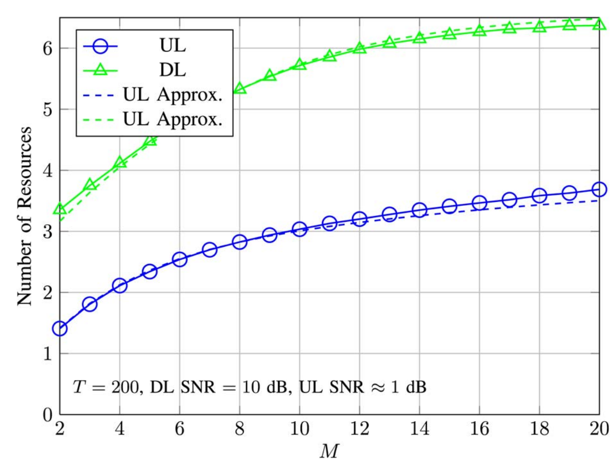

Fig. 10. Optimized rates vs. number of antennas M.

$\alpha = 1$, $P_{\mathrm{UL}} = P_{\mathrm{DL}}/M$ and $10\log_{10}(P_{\mathrm{DL}}/\sigma_{\mathrm{n}}^2) = 10$ dB. As derived in (34), we observe the uplink training length is almost independent of M. For large M, the approximation given in (49) does not seem very good. However, simulation results indicate that the optimum number of feedback bits is not so sensitive for large values of M.

We now compare the uplink and downlink achievable rates resulting from the resource allocation obtained numerically from (29) and with the approximations (34), (41), and (49). We plot the lower bounds (26) and (23) for both cases as a function of T, downlink SNR and M in Figs. 8–10, respectively. When the parameter is not in the abscissa of the plot, we assume $T = 200$, $M = 8$, $\alpha = 1$, $10\log_{10}(P_{\mathrm{DL}}/\sigma_{\mathrm{n}}^2) = 10$ dB. As before, we set $P_{\mathrm{UL}}/\sigma_{\mathrm{n}}^2 = P_{\mathrm{DL}}/(M\sigma_{\mathrm{n}}^2)$, which leads to higher rates in the downlink for a given DL-UL SNR setting. With this relationship, the UL rate is parameterized based on the DL SNR in Fig. 9. From the results, we observe the rates with the optimum resource allocation obtained numerically or with the approximations are basically the same. Although as shown in Figs. 5–7, the approximations for the optimum T_{UL}, T_{DL} and B do not always exactly match the optimum values obtained numerically, the approximated values are close enough to the optimum operating point.

A. Further Considerations

Channel Outdating: We have yet to consider the effect of the outdating of the downlink channel. It turns out the training lengths and the feedback load are basically independent of the channel variation for realistic scenarios. As $\alpha \to 0$, feedback becomes meaningless, and hence, the optimum downlink training length and number of feedback bits should be set to zero. For practical values of α, the resource allocation is independent of the channel variations since the discussed channel estimation and feedback scheme do not consider channel prediction or make use of previous channel estimates.

Multiple Antennas at the User: The system design can be generalized to multiple antennas at the user by performing antenna combining in the downlink and transmit beamforming in the uplink. In the downlink, the user can combine the receive signal over its antennas to produce an equivalent downlink channel vector from the actual MIMO channel. Antenna combining can be performed, for instance, such that the quantization error of the equivalent downlink channel is reduced [37]. The effective downlink channel can afterward be quantized with RVQ and fed back as discussed in this work. Moreover, to enable transmit beamforming in the uplink, the user requires feedback from the BS of its preferred quantized CDI, similarly as previously discussed for the downlink. This implies feedback takes place in both links as presented in [20]. In this scenario, the derived optimum uplink training is no longer valid, since it assumes a single transmit antenna at the user. However, we can employ the derived approximation of the optimum downlink training length for the optimum training length in both links, since it is based on beamforming with multiple antennas. Similarly, the derived optimum number of feedback bits can be employed to determine the optimum feedback load in both the downlink and uplink.

Spatially Correlated Channels: The rank of the correlation matrix is usually smaller than the number of BS antennas in this case. The downlink channel vector can thus be described based on the eigenbasis by an effective channel of reduced dimension, i.e., the rank of the correlation matrix. The presented framework can be applied in this case by considering the estimation and quantization of the effective downlink channel vector of reduced dimension. The derived approximations for the optimum downlink training length and feedback load are valid by considering the reduced dimension instead of the number of BS antennas. To this end the BS needs to know the downlink channel eigenbasis, which can be obtained via long-term feedback.

Multiple Users: the system design can be generalized with multiple users by performing ZF beamforming in the downlink based on the CDI feedback of the users [5]. If the number of users is larger than the number of BS antennas, user selection can be performed. Furthermore, MMSE *receive* beamforming can be applied in the uplink based on the users' estimated uplink channels. For this scenario, the derived approximation for the downlink training length is valid since it is based on a common downlink pilot. In the uplink, however, dedicated training is required for the BS to estimate the uplink channel vector of the users. This implies the optimum uplink training length depends on the number of users, similarly as the optimum downlink training length depends on the number of BS antennas. Furthermore, the optimum feedback load should take into account that serving multiple users in the downlink based on quantized CDI leads to a rate saturation due to the unknown multiuser interference resulting from the quantization [5].

IX. Conclusion

We derived a tight lower bound on the ergodic capacity of single user limited feedback systems. To this end, we assume the available CSI at the BS is affected by estimation, quantization, feedback errors and outdating. Although there is no single figure of merit in two-way systems, we address the design of the FDD uplink and downlink by maximizing the sum of achievable rates in both links by considering the derived downlink bound as well as a bound on the uplink. For the FDD system design, we provide closed-form approximations for the optimum downlink training, uplink training and number of feedback bits. Although we have assumed the RVQ scheme in this work, the derived results are still useful for other quantization schemes (codebooks) since RVQ is not only asymptotically optimal, but also performs quite well for low number of antennas.

Appendix

A. Proof of Theorem 4.1

Proof: A lower bound to the downlink capacity C_{DL} can be obtained as follows.

$$
\begin{aligned}
C_{\mathrm{DL}} &= \frac{1}{T}\left(I\left(\boldsymbol{s}[n];\boldsymbol{y}[n]\right) + I\left(\left(\boldsymbol{s}[n], h_{\mathrm{DL}}[n]\right);\boldsymbol{y}[n]\right)\right.\\
&\qquad \left. - I\left(\left(\boldsymbol{s}[n], h_{\mathrm{DL}}[n]\right);\boldsymbol{y}[n]\right)\right)\\
&\overset{\text{(a)}}{=} \frac{1}{T}\left(I\left(h_{\mathrm{DL}}[n];\boldsymbol{y}[n]\right) + I\left(\boldsymbol{s}[n];\boldsymbol{y}[n] \mid h_{\mathrm{DL}}[n]\right)\right.\\
&\qquad \left. - I\left(h_{\mathrm{DL}}[n];\boldsymbol{y}[n] \mid \boldsymbol{s}[n]\right)\right)\\
&\overset{\text{(b)}}{\geq} \frac{1}{T}\left(I\left(\boldsymbol{s}[n];\boldsymbol{y}[n] \mid h_{\mathrm{DL}}[n]\right) - I\left(h_{\mathrm{DL}}[n];\boldsymbol{y}[n] \mid \boldsymbol{s}[n]\right)\right.\\
&\qquad + I\left(\left(\boldsymbol{s}[n], \theta[n-1]\right);\boldsymbol{y}[n] \mid h_{\mathrm{DL}}[n]\right)\\
&\qquad - I\left(\left(\boldsymbol{s}[n], \theta[n-1]\right);\boldsymbol{y}[n] \mid h_{\mathrm{DL}}[n]\right)\\
&\qquad - I\left(\left(h_{\mathrm{DL}}[n], \theta[n-1]\right);\boldsymbol{y}[n] \mid \boldsymbol{s}[n]\right)\\
&\qquad \left. + I\left(\left(h_{\mathrm{DL}}[n], \theta[n-1]\right);\boldsymbol{y}[n] \mid \boldsymbol{s}[n]\right)\right)\\
&\overset{\text{(c)}}{=} \frac{1}{T}\left(I\left(\theta[n-1];\boldsymbol{y}[n] \mid h_{\mathrm{DL}}[n]\right)\right.\\
&\qquad + I\left(\boldsymbol{s}[n];\boldsymbol{y}[n] \mid h_{\mathrm{DL}}[n], \theta[n-1]\right)\\
&\qquad - I\left(\theta[n-1];\boldsymbol{y}[n] \mid \boldsymbol{s}[n]\right)\\
&\qquad \left. - I\left(h_{\mathrm{DL}}[n];\boldsymbol{y}[n] \mid \boldsymbol{s}[n], \theta[n-1]\right)\right)\\
&\overset{\text{(s)}}{\geq} \frac{1}{T}\left(I\left(\boldsymbol{s}[n];\boldsymbol{y}[n] \mid h_{\mathrm{DL}}[n], \theta[n-1]\right)\right.\\
&\qquad - I\left(h_{\mathrm{DL}}[n];\boldsymbol{y}[n] \mid \boldsymbol{s}[n], \theta[n-1]\right)\\
&\qquad \left. - I(\theta[n-1];\boldsymbol{y}[n] \mid \boldsymbol{s}[n])\right)\\
&\overset{\text{(e)}}{=} \frac{1}{T}\left(I\left(\boldsymbol{s}[n];\boldsymbol{y}[n] \mid h_{\mathrm{DL}}[n], \theta[n-1]\right)\right.\\
&\qquad - I(h_{\mathrm{DL}}[n];\boldsymbol{y}[n] \mid \boldsymbol{s}[n],\\
&\qquad \left. - H\left(\theta[n-1] \mid \boldsymbol{s}[n]\right) + H\left(\theta[n-1] \mid \boldsymbol{s}[n], \boldsymbol{y}[n]\right)\right)\\
&\overset{\text{(f)}}{\geq=} \frac{1}{T}\left(I\left(\boldsymbol{s}[n];\boldsymbol{y}[n] \mid h_{\mathrm{DL}}[n], \theta[n-1]\right)\right.\\
&\qquad \left. - I(h_{\mathrm{DL}}[n];\boldsymbol{y}[n] \mid \boldsymbol{s}[n], \theta[n-1]) - H_{\mathrm{b}}(p_\epsilon)\right), \quad (50)
\end{aligned}
$$

where the effective downlink channel $h_{\mathrm{DL}}[n] = h_{\mathrm{DL,cf}}[n]$ after correct feedback, or $h_{\mathrm{DL}}[n] = h_{\mathrm{DL,cf}}[n]$ after erroneous feedback and $I(B;C)$ represents the mutual information between B and C. The first line in the previous derivation results from considering the mutual information between the transmit signal $\boldsymbol{s}_{\mathrm{DL}}[n]$ and the receive signal $\boldsymbol{y}_{\mathrm{DL}}[n]$ (after dropping the subscript DL to simplify notation) in one time slot consisting of T channel uses. Step (a) follows from applying the two definitions of the chain rule $I(A,B;C) = I(B;C) + I(A;C \mid B) = I(A;C) + I(B;C \mid A)$ on the two identical terms. Inequality (b) results from the non-negativity of the mutual information. Step (c) arises similarly to (a) from the chain rule on both pairs of identical terms, while step (d) follows from the non-negativity of the mutual information. Step (e) results from the definition of the mutual information with $H(A \mid B)$ as the conditional entropy[8] of A given B. Step (f) arises from the non-negativity of the entropy and from $H(\theta[n-1] \mid \boldsymbol{s}[n]) \geq H(\theta[n-1])$ which is the binary entropy function $H_{\mathrm{b}}(p_\epsilon)$, with probability p_e for the error event $\theta[n-1] = 1$.

[8]We employ H, which usually represents the entropy of discrete random variables, to denote the entropy of continuous random variables to avoid confusion in the notation with the effective downlink channel denoted by the lowercase letter h.

The first two terms in (50) can be expressed as

$$\frac{1}{T}I(\boldsymbol{s}[n];\boldsymbol{y}[n]\mid h_{\mathrm{DL}}[n],\theta[n-1]) =\frac{1-p_\epsilon}{T}I(\boldsymbol{s}[n];\boldsymbol{y}[n]\,|\,h_{\mathrm{DL,cf}}[n])+\frac{p_\epsilon}{T}I(\boldsymbol{s}[n];\boldsymbol{y}[n]\mid h_{\mathrm{DL,ef}}[n]) \tag{51}$$

$$\frac{1}{T}I(h_{\mathrm{DL}}[n];\boldsymbol{y}[n]\,|\,\boldsymbol{s}[n],\theta[n-1]) =\frac{1-p_\epsilon}{T}I(h_{\mathrm{DL,cf}}[n];\boldsymbol{y}[n]\,|\,\boldsymbol{s}[n])+\frac{p_\epsilon}{T}I(h_{\mathrm{DL,ef}}[n];\boldsymbol{y}[n]\,|\,\boldsymbol{s}[n]). \tag{52}$$

The first term in (51) is a scaled version of the mutual information given the effective channel after correct feedback. This mutual information can be lower bounded as

$$\frac{1}{T}I\left(\boldsymbol{s}[n];\boldsymbol{y}[n]\mid h_{\mathrm{DL,cf}}[n]\right)$$

$$\overset{(a)}{=}\frac{T_{\mathrm{DL,D}}}{T}\mathrm{E}\left[\log_2\left(1+\frac{P_{\mathrm{DL}}|h_{\mathrm{DL,cf}}[n]|^2}{\sigma_{\mathrm{n}}^2}\right)\right] \tag{53}$$

$$\overset{(b)}{\geq}\frac{T_{\mathrm{DL,D}}}{T}\mathrm{E}_{\nu,\|\hat{\boldsymbol{h}}_{\mathrm{DL}}\|_2^2}\left[\log_2\left(1+\frac{P_{\mathrm{DL}}\alpha\nu\|\hat{\boldsymbol{h}}_{\mathrm{DL}}[n-1]\|_2^2}{\sigma_{\mathrm{n}}^2}\right)\right] \tag{54}$$

$$\overset{(c)}{\geq}\frac{T_{\mathrm{DL,D}}}{T}\left(1-\frac{\sigma_\nu}{2\mu_\nu}\right)\mathrm{E}_{\|\hat{\boldsymbol{h}}_{\mathrm{DL}}\|_2^2}\left[\log_2\left(1+\frac{P_{\mathrm{DL}}\alpha\mu_\nu\|\hat{\boldsymbol{h}}_{\mathrm{DL}}\|_2^2}{\sigma_{\mathrm{n}}^2}\right)\right] \tag{55}$$

$$\overset{(d)}{=}\frac{T_{\mathrm{DL,D}}}{T}\left(1-\frac{\sigma_\nu}{2\mu_\nu}\right)\log_2(\mathrm{e})\mathrm{e}^{\frac{1}{\chi}}\sum_{k=1}^{M}\mathrm{E}_k\left(\frac{1}{\chi}\right), \tag{56}$$

where $\mathrm{E}_k(z)$ is the generalized exponential integral [28, (5.1.4)]. Step (a) results from assuming Gaussian signalling (recall that we do not optimize the input distribution). The inequality in step (b) is proven in Appendix B. For step (c), we employ the fact that ν and $\|\hat{\boldsymbol{h}}_{\mathrm{DL}}[n-1]\|_2^2$ are independent and we make use of a lower bound on concave functions given in [30, (15)] where the absolute deviation $d_\nu \leq \sigma_\nu$. Finally, we find a closed form expression for the lower bound given in (55) by making use of the result from [31, (8.40)] with [28, (5.1.45)], which holds when the elements of $\hat{\boldsymbol{h}}_{\mathrm{DL}}[n-1]$ are independent complex Gaussian random variables. For the last result (56), we have also employed the following substitution

$$\chi=\frac{P_{\mathrm{DL}}\alpha\mu_\nu(1-\sigma_{e_{\mathrm{DL}}}^2)}{\sigma_{\mathrm{n}}^2}. \tag{57}$$

The second term in (51) consists of the mutual information given the effective channel $h_{\mathrm{DL,ef}}[n]$ after erroneous feedback which is distributed according to (22) and hence, is given by [31, (8.34)]

$$\frac{1}{T}I\left(\boldsymbol{s}[n];\boldsymbol{y}[n]\mid h_{\mathrm{DL,ef}}[n]\right)=\frac{T_{\mathrm{DL,D}}}{T}\log_2(\mathrm{e})\,\mathrm{e}^{\frac{\sigma_{\mathrm{n}}^2}{P_{\mathrm{DL}}}}\mathrm{E}_1\left(\frac{\sigma_{\mathrm{n}}^2}{P_{\mathrm{DL}}}\right). \tag{58}$$

The second term in (50), i.e., (52), can be viewed as a penalty for not really knowing the effective downlink channel $h_{\mathrm{DL}}[n]$. The first term in (52) can be *upper* bounded using

$$\frac{1}{T}I(h_{\mathrm{DL,cf}}[n];\boldsymbol{y}[n]\mid\boldsymbol{s}[n]) \leq\frac{1}{T}\log_2(\mathrm{e})\,\mathrm{e}^{\frac{\sigma_{\mathrm{n}}^2}{P_{\mathrm{DL}}\mathrm{E}\left[|h_{\mathrm{DL,cf}}[n]|^2\right]}}\sum_{k=1}^{T_{\mathrm{DL,D}}}\mathrm{E}_k\left(\frac{\sigma_{\mathrm{n}}^2}{P_{\mathrm{DL}}\mathrm{E}\left[|h_{\mathrm{DL,cf}}[n]|^2\right]}\right), \tag{59}$$

which results by assuming $h_{\mathrm{DL,cf}}[n]$ to be Gaussian distributed, which maximizes the entropy given the variance of the effective channel, where the variance is upper bounded by $\mathrm{E}[|h_{\mathrm{DL,cf}}[n]|^2]$, which is defined in (17). In (59), we have also employed [31, (8.40)] with [28, (5.1.45)].

By employing [31, (8.40)] with [28, (5.1.45)] we can also compute the second term in (52)

$$\frac{1}{T}I(h_{\mathrm{DL,ef}}[n];\boldsymbol{y}[n]\mid\boldsymbol{s}[n])=\frac{1}{T}\log_2(\mathrm{e})\,\mathrm{e}^{\frac{\sigma_{\mathrm{n}}^2}{P_{\mathrm{DL}}}}\sum_{k=1}^{T_{\mathrm{DL,D}}}\mathrm{E}_k\left(\frac{\sigma_{\mathrm{n}}^2}{P_{\mathrm{DL}}}\right), \tag{60}$$

since $\boldsymbol{s}[n]$ is Gaussian distributed with unit variance. The lower bound on the downlink capacity C_{DL} results by plugging (56) and (58) in (51), and (59) and (60) in (52), and afterward, substituting the resulting expressions for (51) and (52) in (50). We also employed $T_{\mathrm{DL,D}}=T-T_{\mathrm{DL}}$. □

B. Proof of the Inequality (54)

The effective channel $h_{\mathrm{DL,cf}}[n]=\boldsymbol{w}_{\mathrm{FB}}^{\mathrm{H}}[n]\boldsymbol{h}_{\mathrm{DL}}[n]$ in (14) can be written as $h_{\mathrm{DL,cf}}[n]=y+z$, with $y=\sqrt{\alpha}(\boldsymbol{w}_{\mathrm{FB}}^{\mathrm{H}}[n]\hat{\boldsymbol{h}}_{\mathrm{DL}}[n-1]/\|\hat{\boldsymbol{h}}_{\mathrm{DL}}[n-1]\|_2)\|\hat{\boldsymbol{h}}_{\mathrm{DL}}[n-1]\|_2$ and $z=\sqrt{\alpha}e_{\mathrm{DL}}+\sqrt{1-\alpha}g$. Based on the discussion after (14), z is then a zero-mean complex Gaussian random variables with variance $\sigma_z^2=\alpha\sigma_{e_{\mathrm{DL}}}^2+(1-\alpha)$.

The expected value in (53) can thus be expressed as

$$\mathrm{E}_{y,z}\left[\log_2\left(1+\frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2}|y+z|^2\right)\right]$$

$$=\mathrm{E}_y\left[\mathrm{E}_z\left[\log_2\left(1+\frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2}|y+z|^2\right)\middle|\,y\right]\right] \tag{61}$$

$$\geq\mathrm{E}_y\left[\log_2\left(1+\frac{P_{\mathrm{DL}}}{\sigma_{\mathrm{n}}^2}|y|^2\right)\right], \tag{62}$$

where the second line is a direct consequence of $\mathrm{E}_z[\log_2(1+(P_{\mathrm{DL}}/\sigma_{\mathrm{n}}^2)|y+z|^2)\mid y]\geq\log_2(1+(P_{\mathrm{DL}}/\sigma_{\mathrm{n}}^2)|y|^2)$, which results from $R=\mathrm{E}_z[\log_2(1+(P_{\mathrm{DL}}/\sigma_{\mathrm{n}}^2)|y+z|^2)\mid y]$ being equal to $\log_2(1+(P_{\mathrm{DL}}/\sigma_{\mathrm{n}}^2)|y|^2)$ for $\sigma_z=0$ along with the fact that R is an non-decreasing function of σ_z. This can be seen by taking the derivative of R with respect to σ_z and observing that $\partial R/\partial\sigma_z\geq 0$ for $\sigma_z\geq 0$, or equivalently with z' as a zero mean *unit* variance complex Gaussian random variable, that $\mathrm{E}_{z'}[(\mathrm{Re}\{yz'\}+\sigma_z|z'|^2)/(1+(P_{\mathrm{DL}}/\sigma_{\mathrm{n}}^2))|y+\sigma_z z'|^2)\,|\,y]\geq 0$ for $\sigma_z\geq 0$, which results from the fact that z' has zero mean and has a symmetric distribution. Finally, inequality (54) results from (62) and with ν defined in (15) such that $|y|^2=\alpha\nu\|\hat{\boldsymbol{h}}_{\mathrm{DL}}[n-1]\|_2^2$.

References

[1] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, Dec. 1999.

[2] D. J. Love, R. W. Heath, Jr., W. Santipach, and M. L. Honig, "What is the value of limited feedback for MIMO channels?" *IEEE Commun. Mag.*, vol. 42, no. 10, pp. 54–59, Oct. 2004.

[3] D. J. Love *et al.*, "An overview of limited feedback in wireless communication systems," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 8, pp. 1341–1365, Oct. 2008.

[4] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, "Multiuser MIMO achievable rates with downlink training and channel state feedback," *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 2845–2866, Jun. 2010.

[5] N. Jindal, "MIMO broadcast channels with finite-rate feedback," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 5045–5060, Nov. 2006.

[6] W. Santipach and M. L. Honig, "Asymptotic capacity of beamforming with limited feedback," in *Proc. IEEE ISIT*, Chicago, IL, USA, Jun./Jul. 2004, p. 289.

[7] W. Santipach and M. L. Honig, "Capacity of multi-antenna fading channel with a quantized precoding matrix," *IEEE Trans. Inf. Theory*, vol. 55, no. 3, pp. 1218–1234, Mar. 2009.

[8] R. S. Blum, "MIMO with limited feedback of channel state information," in *Proc. ICASSP*, Apr. 2003, vol. 4, pp. 89–92.

[9] D. J. Love and R. W. Heath, Jr., "Grassmannian beamforming for multiple-input multiple-output wireless systems," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2735–2745, Oct. 2003.

[10] K. K. Mukkavilli, A. Sabharwal, E. Erkip, and B. Aazhang, "On beamforming with finite rate feedback in multiple antenna systems," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2562–2579, Oct. 2003.

[11] V. K. N. Lau, Y. Liu, and T.-A. Chen, "On the design of MIMO block-fading channels with feedback-link capacity constraint," *IEEE Trans. Commun.*, vol. 52, no. 1, pp. 62–70, Jan. 2004.

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

[13] 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, no. 2, pp. 458–462, Feb. 2007.

[14] S. A. Jafar and S. Srinivasa, "On the optimality of beamforming with quantized feedback," *IEEE Trans. Commun.*, vol. 55, no. 12, pp. 2288–2302, Dec. 2007.

[15] W. Santipach and M. L. Honig, "Capacity of beamforming with limited training and feedback," in *Proc. IEEE ISIT*, Jul. 2006, pp. 376–380.

[16] W. Santipach and M. L. Honig, "Optimization of training and feedback overhead over block fading channels," *IEEE Trans. Inf. Theory*, vol. 56, no. 12, pp. 6103–6115, Dec. 2010.

[17] M. Castañeda, A. Mezghani, and J. A. Nossek, "Bounds on the capacity of MISO channels with different types of imperfect CSI," in *Proc. Int. ITG WSA*, Feb. 2009, pp. 1–8.

[18] W. C. Jakes, Ed., *Microwave Mobile Communications*. New York, NY, USA: Wiley, 1974.

[19] M. T. Ivrlač and J. A. Nossek, "Toward a circuit theory of communication," *IEEE Trans. Circuits Syst. I, Reg. Paper*, vol. 57, no. 7, pp. 1663–1683, Jul. 2010.

[20] C. K. Au-Yeung and D. J. Love, "Design and analysis of two-way limited feedback beamforming systems," in *Proc. IEEE Asilomar Conf. Signals, Syst. Comput.*, Nov. 2007, pp. 1946–1950.

[21] D. J. Love and C. K. Au-Yeung, "On resource allocation in two-way limited feedback beamforming systems," in *Proc. Inf. Theory Appl. Workshop*, Jan./Feb. 2008, pp. 188–192, UCSD.

[22] D. J. Love and C. K. Au-Yeung, "Optimization and tradeoff analysis of two-way limited feedback beamforming systems," *IEEE Trans. Wireless Commun.*, vol. 8, no. 5, pp. 2570–2579, May 2009.

[23] B. Hassibi and B. M. Hochwald, "How much training is needed in a multiple-antenna wireless link?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–964, Apr. 2003.

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

[25] A. Lapidoth and S. Shamai, "Fading channels: How perfect need 'Perfect Side Information' be?" *IEEE Trans. Inform. Theory*, vol. 48, no. 5, pp. 1118–114, May 2002.

[26] A. Goldsmith, *Wireless Communications*. New York, NY, USA: Cambridge Univ. Press, 2005.

[27] M. T. Ivrlač and J. A. Nossek, "On the problem of bandwidth partitioning in FDD block-fading single-user MISO/SIMO systems," in *Proc. EURASIP J. Adv. Signal Process.*, Jan. 2008, vol. 2008, pp. 1–13.

[28] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions With Formulas, Graphs and Mathematical Tables*. New York, NY, USA: Dover, 1972.

[29] J. Hadar and W. R. Russell, "Rules for ordering uncertain prospects," *Am. Econ. Rev.*, vol. 59, no. 1, pp. 25–34, Mar. 1969.

[30] A. Ben-Tal and E. Hochman, "More bounds on the expectation of a convex function of a random variable," *J. Appl. Probability*, vol. 9, no. 4, pp. 803–812, Dec. 1972.

[31] M. Alouini, "Adaptive and diversity techniques for wireless digital communications over fading channels," Ph.D. dissertation, California Inst. Technol., Pasadena, CA, USA.

[32] K. M. Miettinen, *Nonlinear Multiobjective Optimization*. Norwell, MA, USA: Kluwer, 1999.

[33] M. Castañeda, A. Mezghani, and J. A. Nossek, "Optimal resource allocation in the downlink/uplink of single-user MISO/SIMO FDD systems with limited feedback," in *Proc. IEEE 10th Workshop SPAWC*, Jun. 2009, pp. 354–358.

[34] M. Castañeda, A. Chathoth, and J. A. Nossek, "Achievable rates in the SIMO-Uplink/MISO-Downlink of an FDD system with imperfect CSI," in *Proc. ISWCS*, Sep. 2009, pp. 121–125.

[35] M. Kobayashi, G. Caire, and N. Jindal, "Optimized training and feedback for MIMO downlink channels," in *Proc. IEEE Inf. Theory Workshop*, Jun. 2009, pp. 226–230.

[36] A. Papoulis and S. U. Pillai, *Probability, Random Variables, Stochastic Processes*. New York, NY, USA: McGraw-Hill, 2002.

[37] N. Jindal, "Antenna combining for the MIMO downlink channel," *IEEE Trans. Wireless Commun.*, vol. 7, no. 10, pp. 3834–3844, Oct. 2008.

Mario H. Castañeda Garcia (S'04) was born in Tegucigalpa, Honduras. He received the B.S. degree in electrical engineering from the Universidad Nacional Autónoma de Honduras, Tegucigalpa, Honduras, in 2001 and the M.Sc. degree in communications engineering from the Technische Universität München (TUM), Munich, Germany, in 2004. From 2005 to 2014 he worked as a research and teaching assistant at the Institute for Circuit Theory and Signal Processing at TUM. In 2014 he joined Huawei's European Research Center in Munich. His research interests comprise signal processing for multiuser wireless communication systems.

Amine Mezghani (S'08) was born in Sfax, Tunisia. He received the "Diplome d'Ingénieur" degree from Ecole Centrale de Paris, France, and the Dipl.-Ing. degree in electrical engineering from the Technische Universität München (TUM), Munich, Germany, both in 2006. Since 2006, he has been working as research assistant at the Institute for Circuit Theory and Signal Processing, TUM. His research interests focus on the study of wireless communications, information theory and signal processing under low-precision analog-to-digital converters.

Josef A. Nossek (S'72–M'74–SM'81–F'93) received the Dipl.-Ing. and Dr.Techn. degrees in electrical engineering from Vienna University of Technology, Vienna, Austria, in 1974 and 1980, respectively. He joined Siemens AG, Munich, Germany, in 1974, where he was engaged in the design of both passive and active filters for communication systems. From 1987 to 1989, he was the Head of the Radio Systems Design Department, where he was instrumental in introducing high-speed VLSI signal processing into digital microwave radio. Since April 1989, he has been a Professor of circuit theory and design with the Technische Universität München (TUM), Munich, Germany. Dr. Nossek was the President Elect, President, and Past President of the IEEE Circuits and Systems Society in 2001, 2002, and 2003, respectively. He was the President of Verband der Elektrotechnik, Elektronik und Informationstechnik e.V. (VDE) in 2007 and 2008. He was the recipient of the ITG Best Paper Award in 1988, the Mannesmann Mobilfunk (currently Vodafone) Innovations Award in 1998, and the Award for Excellence in Teaching from the Bavarian Ministry for Science, Research and Art in 1998. From the IEEE Circuits and Systems Society, he received the Golden Jubilee Medal for "Outstanding Contributions to the Society" in 1999 and the Education Award in 2008. He was the recipient of the "Bundesverdienstkreuz am Bande" in 2008. In 2011 he received the IEEE Guillemin-Cauer Best Paper Award for his paper "Toward a Circuit Theory of Communication."