Transceiver Design in Multiuser MISO Systems with Imperfect Transmit CSI

Mario H. Castañeda, Israa Slim, Amine Mezghani and Josef A. Nossek
Institute for Circuit Theory and Signal Processing, Technische Universität München, Munich, Germany
Email: {maca,issl,amme,jono}@nws.ei.tum.de

***Abstract*—Multiple transmit antennas can be employed in a downlink channel to serve multiple users under the assumption of perfect *channel state information* (CSI) of all users at the base station. In practice, however, the channel knowledge available at the base station is not perfect. For instance, in case of a *time division duplex* (TDD) system, the transmit CSI is the estimated uplink channel. On the other hand, in a *frequency division duplex* (FDD) system, feedback from the users is required to obtain transmit CSI. In this case, we assume the transmit CSI consists of a quantized version of the normalized estimated channel vector and the statistics of the channel norm of each user. In this work we consider the transceiver design of the downlink of a multiuser MISO system based on the *minimum mean square error* (MMSE) criterion with imperfect CSI as discussed above.**

I. Introduction

It is well know that M transmit antennas at the BS can be employed in a downlink channel to serve K single-antenna users ($K \leq M$) under the assumption of perfect CSI. In practice, however, the channel knowledge available at the transmitter and receiver is not perfect. For instance, in a TDD system the BS can obtain an MMSE estimate of the uplink channels of the K users through K *dedicated* pilots, and due to the channel reciprocity the BS can employ the K estimated uplink channels as transmit CSI for the downlink. The quality of the transmit CSI for this case depends on the uplink SNR and training length [1].

In an FDD system, we assume the BS obtains knowledge of the downlink channel through limited feedback in the uplink consisting of B bits per user [2]. To this end, the K users first estimate their downlink channel with a *common* downlink pilot. Each of the K users then quantizes its estimated *channel direction information* (CDI), i.e. the normalized estimated channel vector, using the *random vector quantization* (RVQ) scheme [3] with B bits, which are then relayed back to the BS in the uplink. Hence, the BS has only access to a quantized version of the estimated CDI of each user and in addition to the statistics of the norm of the channel vector, i.e. the *channel magnitude information* (CMI), of each user as transmit CSI. The quality of the transmit CSI in this case depends on the downlink SNR, training length and on B. Note that although the feedback could be received erroneously at the BS, we will not consider the effect of the feedback errors in this work.

The optimum transmit strategy with respect to commonly employed figures of merit based on imperfect transmit CSI is unknown in general. However, there is some interesting work dealing with the problem of imperfect transmit CSI at the BS. The throughput degradation due to limited feedback with zero-forcing beamforming based on quantized CDI has been derived in [4]. The feedback strategy which provides quantized CDI considering feedback errors and feedback delay for zero-forcing beamforming has been optimized in [5]. The BER achieved under vector perturbation with limited feedback consisting of a quantized and estimated CDI has been treated in [6]. In this work we consider the transceiver design of the downlink of a multiuser MISO system based on the MMSE criteria with different types of imperfect transmit CSI. We consider a linear scheme instead of a non-linear technique due to simplicity and robustness to imperfect CSI of linear techniques. We treat the case $K \leq M$, i.e. we do not consider a scheduler. This papers is structured as follows. In Section II, the perfect CSI case is reviewed while the TDD case is treated in Section III. In Section IV we treat the theoretical case of perfect CDI in order to gain insight for the FDD case which is presented in Section V. Simulation results are discussed in Section VI, while Section VII concludes the paper.

II. Perfect CSI Case

Consider a *base station* (BS) equipped with M antennas in a single isolated cell transmitting to K i.i.d. single-antenna users with $K \leq M$. Denote the independent zero-mean transmitted signals with variance σ_s^2 for the K users as $\boldsymbol{s} \in \mathbb{C}^K$ with covariance matrix $\boldsymbol{R}_s = \sigma_s^2 \mathbf{1}_K$. Define also the channel from the BS to user k as $\boldsymbol{h}_k \in \mathbb{C}^M$ and the channel matrix $\boldsymbol{H} = [\boldsymbol{h}_1, \boldsymbol{h}_2, \cdots, \boldsymbol{h}_K]^{\mathrm{T}} \in \mathbb{C}^{K \times M}$. We assume the elements of $\boldsymbol{h}_k \ \forall k$ to be i.i.d. zero-mean unit-variance complex Gaussian random variables. With a linear precoder $\boldsymbol{P}$ at the BS and transmit power contraint $\operatorname{tr}\left(\boldsymbol{P}\boldsymbol{R}_s\boldsymbol{P}^{\mathrm{H}}\right) \leq P_{\mathrm{DL}}$, where P_{DL} is the available transmit power at the BS, the estimated signal at the K users is given by

$$\hat{\boldsymbol{s}} = g\left(\boldsymbol{H}\boldsymbol{P}\boldsymbol{s} + \boldsymbol{n}\right), \tag{1}$$

where g is a scalar receive filter applied at each user since there is no cooperation among the single-antenna users and $\boldsymbol{n} \in \mathbb{C}^K$ is the collection of the zero-mean AWGN experienced at each user with covariance matrix $\boldsymbol{R}_n = \sigma_n^2 \mathbf{1}_K$. Under the assumption of *perfect CSI*, i.e. $\boldsymbol{H}$ is known at the BS, g and $\boldsymbol{P}$ can be computed in order to minimize the system *mean square error* (MSE) $\mathrm{E}\left[\|\boldsymbol{s} - \hat{\boldsymbol{s}}\|_2^2\right]$ subject to the power constraint, i.e.

$$\{\boldsymbol{P}, g\} = \underset{\{\boldsymbol{P}, g\}}{\operatorname{argmin}} \ \mathrm{E}\left[\|\boldsymbol{s} - \hat{\boldsymbol{s}}_{\mathrm{e}}\|_2^2\right] \quad \text{s.t.} \quad \operatorname{tr}\left(\boldsymbol{P}\boldsymbol{R}_s\boldsymbol{P}^{\mathrm{H}}\right) \leq P_{\mathrm{DL}}, \tag{2}$$

where the expectation is taken over $\boldsymbol{s}$ and $\boldsymbol{n}$. The solution to (2) is [7]

$$g = \sqrt{\frac{\operatorname{tr}\left(\left(\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}+\xi\boldsymbol{1}_M\right)^{-2}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{R}_s\boldsymbol{H}\right)}{P_{\mathrm{DL}}}} \quad (3)$$

$$\boldsymbol{P} = \frac{1}{g}\left(\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}+\xi\boldsymbol{1}_M\right)^{-1}\boldsymbol{H}^{\mathrm{H}}, \quad (4)$$

where

$$\xi = \frac{\operatorname{tr}\left(\boldsymbol{R}_n\right)}{P_{\mathrm{DL}}} = \frac{K\sigma_n^2}{P_{\mathrm{DL}}}. \quad (5)$$

In addition, the MMSE with perfect CSI is

$$\mathrm{MMSE}_{\mathrm{p}} = \xi \operatorname{tr}\left(\left(\boldsymbol{H}\boldsymbol{H}^{\mathrm{H}}+\xi\boldsymbol{1}_M\right)^{-1}\boldsymbol{R}_s\right). \quad (6)$$

For $K < M$ and at high SNR, i.e. $\lambda_{\min}\left(\boldsymbol{H}\boldsymbol{H}^{\mathrm{H}}\right) \gg \xi$ with probabiliy 1, where $\lambda_{\min}(\boldsymbol{X})$ denotes the smallest eigenvalue of $\boldsymbol{X}$, we have

$$\mathrm{MMSE}_{\mathrm{p,high\text{-}SNR}} \approx \xi\,\sigma_s^2 \operatorname{tr}\left(\left(\boldsymbol{H}\boldsymbol{H}^{\mathrm{H}}\right)^{-1}\right), \quad (7)$$

and the average MMSE (over the channel realizations) is

$$\begin{aligned}\overline{\mathrm{MMSE}}_{\mathrm{p,high\text{-}SNR}} &\approx \xi\,\sigma_s^2\,\mathrm{E}\left[\operatorname{tr}\left(\left(\boldsymbol{H}\boldsymbol{H}^{\mathrm{H}}\right)^{-1}\right)\right] \\ &= \xi\,\sigma_s^2\,\frac{K}{M-K}, \end{aligned} \quad (8)$$

where in the second step we have used the result $\mathrm{E}\left[\operatorname{tr}\left(\left(\boldsymbol{H}\boldsymbol{H}^{\mathrm{H}}\right)^{-1}\right)\right] = \frac{K}{M-K}$ from equation (2.9) in [8].

III. TDD System: Estimated CSI

Consider now a TDD system where the BS estimates the users' uplink channels with the aid of K dedicated *orthogonal* pilot sequences in the uplink of length $T_{\mathrm{UL}} \geq K$[1]. We assume that the base station computes a *minimum mean square error* estimate of each user's uplink channel. In this case, the variance of the estimation error is given by [1]

$$\sigma_{e_{\mathrm{UL}}}^2 = \frac{1}{1+\frac{P_{\mathrm{UL}}}{\sigma_v^2}T_{\mathrm{UL}}}, \quad (9)$$

where P_{UL} is the transmit power at each user and σ_v^2 is the variance of the AWGN at each antenna of the BS in the uplink. Due to reciprocity in a TDD system, the uplink channel estimate serves as an estimate of the downlink channel. Let us denote the estimated downlink channel for user k as $\hat{\boldsymbol{h}}_k \in \mathbb{C}^M$ and $\hat{\boldsymbol{H}} = \left[\hat{\boldsymbol{h}}_1, \hat{\boldsymbol{h}}_2, \cdots, \hat{\boldsymbol{h}}_K\right]^{\mathrm{T}} \in \mathbb{C}^{K\times M}$ such that

$$\boldsymbol{H} = \hat{\boldsymbol{H}} + \boldsymbol{E}, \quad (10)$$

where $\boldsymbol{E} = \left[\boldsymbol{e}_1, \cdots, \boldsymbol{e}_K\right]^{\mathrm{T}} \in \mathbb{C}^{K\times M}$ is the estimation error matrix, where $\boldsymbol{e}_k$ is the estimation error vector when estimating the channel of user k. The elements of $\boldsymbol{E}$ are zero-mean complex Gaussian random variables with variance $\sigma_{e_{\mathrm{UL}}}^2$ while the elements of $\hat{\boldsymbol{H}}$ are zero-mean complex Gaussian random variables with variance $1-\sigma_{e_{\mathrm{UL}}}^2$. Recall that the MMSE estimate $\hat{\boldsymbol{h}}_k$ has the property that $\mathrm{E}\left[\hat{\boldsymbol{h}}_k^{\mathrm{H}}\boldsymbol{e}_k\right] = \boldsymbol{0}$ and in addition due to the independence assumption among the users, we have that $\mathrm{E}\left[\hat{\boldsymbol{h}}_k^{\mathrm{H}}\boldsymbol{e}_j\right] = \boldsymbol{0}\ \forall j,k$, such that

$$\mathrm{E}\left[\hat{\boldsymbol{H}}^{\mathrm{H}}\boldsymbol{E}\right] = \boldsymbol{0}. \quad (11)$$

Denote the precoder and scalar receiver based on $\hat{\boldsymbol{H}}$, i.e. with estimated CSI, as $\boldsymbol{P}_{\mathrm{e}}$ and g_{e}. In contrast to (1), the estimated signal at the users with estimated CSI can be written as

$$\begin{aligned}\hat{\boldsymbol{s}}_{\mathrm{e}} &= g_{\mathrm{e}}\left(\left(\hat{\boldsymbol{H}}+\boldsymbol{E}\right)\boldsymbol{P}_{\mathrm{e}}\boldsymbol{s}+\boldsymbol{n}\right) \\ &= g_{\mathrm{e}}\left(\hat{\boldsymbol{H}}\boldsymbol{P}_{\mathrm{e}}\boldsymbol{s}+\boldsymbol{n}_{\mathrm{e}}\right), \end{aligned} \quad (12)$$

where the effective noise now is

$$\boldsymbol{n}_{\mathrm{e}} = \boldsymbol{E}\boldsymbol{P}_{\mathrm{e}}\boldsymbol{s}+\boldsymbol{n}, \quad (13)$$

which is uncorrelated with the signal, i.e.

$$\begin{aligned}\mathrm{E}\left[\boldsymbol{s}\boldsymbol{n}_{\mathrm{e}}^{\mathrm{H}}\right] &= \mathrm{E}\left[\boldsymbol{s}\boldsymbol{s}^{\mathrm{H}}\boldsymbol{P}_{\mathrm{e}}^{\mathrm{H}}\boldsymbol{E}^{\mathrm{H}}\right]+\mathrm{E}\left[\boldsymbol{s}\boldsymbol{n}^{\mathrm{H}}\right] = \mathrm{E}\left[\boldsymbol{s}\boldsymbol{s}^{\mathrm{H}}\right]\boldsymbol{P}_{\mathrm{e}}^{\mathrm{H}}\mathrm{E}\left[\boldsymbol{E}^{\mathrm{H}}\right]+\boldsymbol{0} \\ &= \sigma_s^2\cdot\boldsymbol{1}_M\boldsymbol{P}_{\mathrm{e}}^{\mathrm{H}}\cdot\boldsymbol{0}+\boldsymbol{0} = \boldsymbol{0}, \end{aligned} \quad (14)$$

since the signal and the estimation error are independent. The covariance matrix of the effective noise $\boldsymbol{n}_{\mathrm{e}}$ is

$$\begin{aligned}\boldsymbol{R}_{n,\mathrm{e}} &= \mathrm{E}\left[\boldsymbol{n}_{\mathrm{e}}\boldsymbol{n}_{\mathrm{e}}^{\mathrm{H}}\right] \\ &\overset{(a)}{=} \mathrm{E}\left[\boldsymbol{E}\boldsymbol{P}_{\mathrm{e}}\boldsymbol{s}\boldsymbol{s}^{\mathrm{H}}\boldsymbol{P}_{\mathrm{e}}^{\mathrm{H}}\boldsymbol{E}^{\mathrm{H}}\right]+\mathrm{E}\left[\boldsymbol{n}\boldsymbol{n}^{\mathrm{H}}\right] \\ &\overset{(b)}{=} \mathrm{E}\left[\boldsymbol{E}\underbrace{\boldsymbol{P}_{\mathrm{e}}\boldsymbol{R}_s\boldsymbol{P}_{\mathrm{e}}^{\mathrm{H}}}_{\boldsymbol{A}}\boldsymbol{E}^{\mathrm{H}}\right]+\sigma_n^2\boldsymbol{1}_K \\ &\overset{(c)}{=} \sum_{i=1}^{M}\sum_{j=1}^{M}A_{i,j}\,\mathrm{E}\left[\boldsymbol{e}_{\mathrm{c},i}\,\boldsymbol{e}_{\mathrm{c},j}^{\mathrm{H}}\right]+\sigma_n^2\boldsymbol{1}_K \\ &\overset{(d)}{=} \sum_{i=1}^{M}A_{i,i}\,\mathrm{E}\left[\boldsymbol{e}_{\mathrm{c},i}\,\boldsymbol{e}_{\mathrm{c},i}^{\mathrm{H}}\right]+\sigma_n^2\boldsymbol{1}_K \\ &\overset{(e)}{=} \left(\sum_{i=1}^{M}A_{i,i}\right)\cdot\sigma_{e_{\mathrm{UL}}}^2\boldsymbol{1}_K+\sigma_n^2\boldsymbol{1}_K \\ &\overset{(f)}{=} \left(\operatorname{tr}\left(\boldsymbol{P}_{\mathrm{e}}\boldsymbol{R}_s\boldsymbol{P}_{\mathrm{e}}^{\mathrm{H}}\right)\cdot\sigma_{e_{\mathrm{UL}}}^2+\sigma_n^2\right)\boldsymbol{1}_K, \end{aligned} \quad (15)$$

where (a) follows from plugging (13) and from the fact that the signal and noise are uncorrelated, i.e. $\mathrm{E}\left[\boldsymbol{s}\boldsymbol{n}^{\mathrm{H}}\right] = \boldsymbol{0}$. Step (b) follows by taking the expectation over the noise and over the signal. For step (c) we took the substitution $\boldsymbol{A} = \boldsymbol{P}_{\mathrm{e}}\boldsymbol{R}_s\boldsymbol{P}_{\mathrm{e}}^{\mathrm{H}}$ and have denoted $A_{i,j}$ as the element in the i-th row and j-th column of $\boldsymbol{A}$ and we also denote $\boldsymbol{e}_{\mathrm{c},i}$ as the i-th column of $\boldsymbol{E}$. In (d) we make use of $\mathrm{E}\left[\boldsymbol{e}_{\mathrm{c},i}\,\boldsymbol{e}_{\mathrm{c},j}^{\mathrm{H}}\right] = \boldsymbol{0}$ for $i \neq j$ due to the independence among the users and channels, while in step (e) we make use of $\mathrm{E}\left[\boldsymbol{e}_{\mathrm{c},i}\,\boldsymbol{e}_{\mathrm{c},j}^{\mathrm{H}}\right] = \sigma_{e_{\mathrm{UL}}}^2\cdot\boldsymbol{1}_K$ due to the independence among the channels. For step (f) note that $\sum_{i=1}^{M}A_{i,i} = \operatorname{tr}(\boldsymbol{A}) = \operatorname{tr}\left(\boldsymbol{P}_{\mathrm{e}}\boldsymbol{R}_s\boldsymbol{P}_{\mathrm{e}}^{\mathrm{H}}\right)$.

Let us compute $\boldsymbol{P}_{\mathrm{e}}$ and g_{e} which minimize the MSE with estimated CSI, i.e.

$$\left\{\boldsymbol{P}_{\mathrm{e}}, g_{\mathrm{e}}\right\} = \underset{\{\boldsymbol{P}_{\mathrm{e}},g_{\mathrm{e}}\}}{\operatorname{argmin}}\ \mathrm{E}\left[\|\boldsymbol{s}-\hat{\boldsymbol{s}}_{\mathrm{e}}\|_2^2\right] \quad \text{s.t.} \quad \operatorname{tr}\left(\boldsymbol{P}_{\mathrm{e}}\boldsymbol{R}_s\boldsymbol{P}_{\mathrm{e}}^{\mathrm{H}}\right) \leq P_{\mathrm{DL}}, \quad (16)$$

[1]Note that due to the single-input multiple-output nature of the uplink, only one pilot symbol is really necessary in the noise-free case, to estimate each of the M channels from a user's antenna to the M receiving antennas at the base station. We assume, however, the base station estimates K uplink channel vectors simultaneously and hence the uplink training length T_{UL} must be of length at least K in order to have K orthogonal pilot sequences.

where the expectation is taken over $\boldsymbol{s}$, $\boldsymbol{n}$ and $\boldsymbol{E}$. The MSE is given as

$$\begin{aligned}\mathrm{E}\left[\|\boldsymbol{s}-\hat{\boldsymbol{s}}_{\mathrm{e}}\|_2^2\right] &= \mathrm{E}\left[\left\|\left(\mathbf{1}_K - g_{\mathrm{e}}\hat{\boldsymbol{H}}\boldsymbol{P}_{\mathrm{e}}\right)\boldsymbol{s} + g_{\mathrm{e}}\boldsymbol{n}_{\mathrm{e}}\right\|_2^2\right] \\ &= \mathrm{tr}\Big(\mathrm{E}\left[\left(\mathbf{1}_K - g_{\mathrm{e}}\hat{\boldsymbol{H}}\boldsymbol{P}_{\mathrm{e}}\right)\boldsymbol{s}\boldsymbol{s}^{\mathrm{H}}\left(\mathbf{1}_K - g_{\mathrm{e}}\hat{\boldsymbol{H}}\boldsymbol{P}_{\mathrm{e}}\right)^{\mathrm{H}}\right] \\ &\quad + |g_{\mathrm{e}}|^2\mathrm{E}\left[\boldsymbol{n}_{\mathrm{e}}\boldsymbol{n}_{\mathrm{e}}^{\mathrm{H}}\right]\Big) \\ &= \mathrm{tr}\left(\boldsymbol{R}_s - 2\mathrm{Re}\left(g_{\mathrm{e}}\hat{\boldsymbol{H}}\boldsymbol{P}_{\mathrm{e}}\boldsymbol{R}_s\right)\right) + \\ \mathrm{tr}\left(|g_{\mathrm{e}}|^2\hat{\boldsymbol{H}}\boldsymbol{P}_{\mathrm{e}}\boldsymbol{R}_s\boldsymbol{P}_{\mathrm{e}}^{\mathrm{H}}\hat{\boldsymbol{H}}^{\mathrm{H}}\right) &+ K\left(\mathrm{tr}\left(\boldsymbol{P}_{\mathrm{e}}\boldsymbol{R}_s\boldsymbol{P}_{\mathrm{e}}^{\mathrm{H}}\right)\sigma_{e_{\mathrm{UL}}}^2 + \sigma_n^2\right),\end{aligned}$$

where the third step follows from (14).

Note that $\boldsymbol{R}_{n,\mathrm{e}}$ depends on $\boldsymbol{P}_{\mathrm{e}}$ as can be seen from (15) and must be considered when solving (16), which results in

$$g_{\mathrm{e}} = \sqrt{\frac{\mathrm{tr}\left(\left(\hat{\boldsymbol{H}}^{\mathrm{H}}\hat{\boldsymbol{H}} + \xi_{\mathrm{e}}\mathbf{1}_M\right)^{-2}\hat{\boldsymbol{H}}^{\mathrm{H}}\boldsymbol{R}_s\hat{\boldsymbol{H}}\right)}{P_{\mathrm{DL}}}} \quad (17)$$

$$\boldsymbol{P}_{\mathrm{e}} = \frac{1}{g_{\mathrm{e}}}\left(\hat{\boldsymbol{H}}^{\mathrm{H}}\hat{\boldsymbol{H}} + \xi_{\mathrm{e}}\mathbf{1}_M\right)^{-1}\hat{\boldsymbol{H}}^{\mathrm{H}}, \quad (18)$$

with

$$\xi_{\mathrm{e}} = \frac{K\sigma_n^2}{P_{\mathrm{DL}}} + K\sigma_{e_{\mathrm{UL}}}^2 = \frac{\mathrm{tr}\left(\boldsymbol{R}_{n,\mathrm{e}}\right)}{P_{\mathrm{DL}}}, \quad (19)$$

where in the second equality we make use of the fact that the power constraint is active, i.e. $\mathrm{tr}\left(\boldsymbol{P}_{\mathrm{e}}\boldsymbol{R}_s\boldsymbol{P}_{\mathrm{e}}^{\mathrm{H}}\right) = P_{\mathrm{DL}}$, and so (15) becomes

$$\boldsymbol{R}_{n,\mathrm{e}} = \left(P_{\mathrm{DL}}\,\sigma_{e_{\mathrm{UL}}}^2 + \sigma_n^2\right)\mathbf{1}_K. \quad (20)$$

Substituting $\sigma_{e_{\mathrm{UL}}}^2$ from (9) in (19) gives us

$$\xi_{\mathrm{e}} = \xi + \frac{K}{1 + \frac{P_{\mathrm{UL}}}{\sigma_v^2}T_{\mathrm{UL}}}. \quad (21)$$

The resulting MMSE with estimated CSI (averaged over the noise and the esimation error) reads similar to (6)

$$\mathrm{MMSE}_{\mathrm{e}} = \xi_{\mathrm{e}}\,\mathrm{tr}\left(\left(\hat{\boldsymbol{H}}\hat{\boldsymbol{H}}^{\mathrm{H}} + \xi_{\mathrm{e}}\mathbf{1}_M\right)^{-1}\boldsymbol{R}_s\right). \quad (22)$$

For $K < M$ and at high downlink SNR, i.e. $\lambda_{\min}\left(\hat{\boldsymbol{H}}\hat{\boldsymbol{H}}^{\mathrm{H}}\right) \gg \xi_{\mathrm{e}}$ with probability 1, we have

$$\begin{aligned}\mathrm{MMSE}_{\mathrm{e,high\text{-}SNR}} &\approx \xi_{\mathrm{e}}\,\sigma_s^2\,\mathrm{tr}\left(\left(\hat{\boldsymbol{H}}\hat{\boldsymbol{H}}^{\mathrm{H}}\right)^{-1}\right) \\ &= \left(\frac{K\sigma_n^2}{P_{\mathrm{DL}}} + K\,\sigma_{e_{\mathrm{UL}}}^2\right)\sigma_s^2\,\mathrm{tr}\left(\left(\hat{\boldsymbol{H}}\hat{\boldsymbol{H}}^{\mathrm{H}}\right)^{-1}\right),\end{aligned}$$

and averaging over the channel realizations

$$\begin{aligned}\overline{\mathrm{MMSE}}_{\mathrm{e,high\text{-}SNR}} &\overset{(a)}{\approx} \left(\frac{K\sigma_n^2}{P_{\mathrm{DL}}} + K\,\sigma_{e_{\mathrm{UL}}}^2\right)\sigma_s^2\,\frac{K}{(M-K)(1-\sigma_{e_{\mathrm{UL}}}^2)} \\ &\overset{(b)}{\geq} \xi\sigma_s^2\frac{K}{(M-K)} + \frac{K^2\,\sigma_{e_{\mathrm{UL}}}^2\sigma_s^2}{(M-K)(1-\sigma_{e_{\mathrm{UL}}}^2)} \\ &\overset{(c)}{=} \overline{\mathrm{MMSE}}_{\mathrm{p,high\text{-}SNR}} + \frac{K^2\,\sigma_s^2}{(M-K)\frac{P_{\mathrm{UL}}}{\sigma_v^2}T_{\mathrm{UL}}} \\ &\geq \overline{\mathrm{MMSE}}_{\mathrm{p,high\text{-}SNR}},\end{aligned} \quad (23)$$

where $\mathrm{E}\left[\mathrm{tr}\left((\hat{\boldsymbol{H}}\hat{\boldsymbol{H}}^{\mathrm{H}})^{-1}\right)\right] = \frac{K}{(M-K)(1-\sigma_{e_{\mathrm{UL}}}^2)}$ is used in step (a) which results from (2.9) in [8] and from the fact that the elements of $\hat{\boldsymbol{H}}$ have variance $(1-\sigma_{e_{\mathrm{UL}}}^2)$. Step (b) follows from $\sigma_{e_{\mathrm{UL}}}^2 \leq 1$ and in (c) we plugged in (8) and (9). Hence there is still a gap between the MMSE with perfect CSI and the MMSE with estimated CSI even at high SNR. At high SNR $\mathrm{MMSE}_{\mathrm{e}} \to \mathrm{MMSE}_{\mathrm{p}}$ only when $T_{\mathrm{UL}} \to \infty$ (which implies a consumption of uplink resources). However, for a finite T_{UL} there is a non-zero gap between the $\mathrm{MMSE}_{\mathrm{e}}$ and $\mathrm{MMSE}_{\mathrm{p}}$.

IV. Channel Direction Information (CDI)

In order to gain some insight for the FDD case, we first consider the theoretical case where the BS has only access to the perfect CDI of the users, i.e. $\frac{\boldsymbol{h}_k}{\|\boldsymbol{h}_k\|_2}\ \forall k$. Define $\boldsymbol{H}_{\mathrm{n}} = \left[\frac{\boldsymbol{h}_1}{\|\boldsymbol{h}_1\|_2}, \cdots, \frac{\boldsymbol{h}_K}{\|\boldsymbol{h}_K\|_2}\right]^{\mathrm{T}}$ such that $\boldsymbol{H} = \boldsymbol{B}\boldsymbol{H}_{\mathrm{n}}$ with $\boldsymbol{B} = \mathrm{diag}\left([\|\boldsymbol{h}_1\|_2, \cdots, \|\boldsymbol{h}_K\|_2]\right)^2$. Denote the precoder and the scalar receiver based on $\boldsymbol{H}_{\mathrm{n}}$, i.e. with perfect CDI, as $\boldsymbol{P}_{\mathrm{n}}'$ and g_{n}', respectively. The estimated signal at the users is

$$\hat{\boldsymbol{s}}_{\mathrm{n}}' = g_{\mathrm{n}}'\left(\boldsymbol{B}\boldsymbol{H}_{\mathrm{n}}\boldsymbol{P}_{\mathrm{n}}'\boldsymbol{s} + \boldsymbol{n}\right), \quad (24)$$

where g_{n}' and $\boldsymbol{P}_{\mathrm{n}}'$ are computed in order to minimize the MSE $\mathrm{E}\left[\|\boldsymbol{s} - \hat{\boldsymbol{s}}_{\mathrm{n}}'\|_2^2\right]$ subject to the power constraint, where the expectation is taken over $\boldsymbol{s}$, $\boldsymbol{n}$ and over the *channel magnitude information* (CMI) of the users, i.e. over $\boldsymbol{B}$, since the BS has only access to the CDI. In this case with perfect CDI and the statistics of the CMI we obtain

$$g_{\mathrm{n}}' = \sqrt{\frac{\mathrm{tr}\left(\left(\boldsymbol{H}_{\mathrm{n}}^{\mathrm{H}}\mathrm{E}\left[\boldsymbol{B}^2\right]\boldsymbol{H}_{\mathrm{n}} + \xi\mathbf{1}_M\right)^{-2}\boldsymbol{H}_{\mathrm{n}}^{\mathrm{H}}\mathrm{E}\left[\boldsymbol{B}\right]\boldsymbol{R}_s\mathrm{E}\left[\boldsymbol{B}\right]\boldsymbol{H}_{\mathrm{n}}\right)}{P_{\mathrm{DL}}}}$$

$$\boldsymbol{P}_{\mathrm{n}}' = \frac{1}{g_{\mathrm{n}}'}\left(\boldsymbol{H}_{\mathrm{n}}^{\mathrm{H}}\mathrm{E}\left[\boldsymbol{B}^2\right]\boldsymbol{H}_{\mathrm{n}} + \xi\mathbf{1}_M\right)^{-1}\boldsymbol{H}_{\mathrm{n}}^{\mathrm{H}}\,\mathrm{E}\left[\boldsymbol{B}\right],$$

and since the norm of each user's channel vector has the same chi-distribution, we have that

$$\mathrm{E}\left[\boldsymbol{B}\right] = \frac{\Gamma(M+\frac{1}{2})}{(M-1)!}\mathbf{1}_K \quad (25)$$

$$\mathrm{E}\left[\boldsymbol{B}^2\right] = M\,\mathbf{1}_K, \quad (26)$$

with $\Gamma(\bullet)$ as the Gamma function. The respective MMSE given the users' CDI $\boldsymbol{H}_{\mathrm{n}}$, i.e. averaging over the noise and the users' CMI $\boldsymbol{B}$, turns out to be

$$\begin{aligned}\mathrm{MMSE}_{\mathrm{n}}' = \sigma_s^2\mathrm{tr}\Bigg(&\left(\frac{\xi}{M}\mathbf{1}_K + \left(1 - \frac{\Gamma^2(M+\frac{1}{2})}{M!(M-1)!}\right)\boldsymbol{H}_{\mathrm{n}}\boldsymbol{H}_{\mathrm{n}}^{\mathrm{H}}\right) \times \\ &\left(\frac{\xi}{M}\mathbf{1}_K + \boldsymbol{H}_{\mathrm{n}}\boldsymbol{H}_{\mathrm{n}}^{\mathrm{H}}\right)^{-1}\Bigg).\end{aligned} \quad (27)$$

Note that with this approach, even in the noise-free case, i.e. $\xi = 0$, we have a residual MMSE. Hence, $\mathrm{MMSE}_{\mathrm{n}}'$ saturates at high SNR to the following residual error

$$\begin{aligned}\mathrm{MMSE}_{\mathrm{n},\xi=0}' &= \sigma_s^2\mathrm{tr}\left(\left(\left(1 - \frac{\Gamma^2(M+\frac{1}{2})}{M!(M-1)!}\right)\boldsymbol{H}_{\mathrm{n}}\boldsymbol{H}_{\mathrm{n}}^{\mathrm{H}}\right)\left(\boldsymbol{H}_{\mathrm{n}}\boldsymbol{H}_{\mathrm{n}}^{\mathrm{H}}\right)^{-1}\right) \\ &= \sigma_s^2\mathrm{tr}\left(\left(1 - \frac{\Gamma^2(M+\frac{1}{2})}{M!(M-1)!}\right)\cdot\mathbf{1}_K\right) \\ \overline{\mathrm{MMSE}}_{\mathrm{n},\xi=0}' &= \sigma_s^2 K\left(1 - \frac{\Gamma^2(M+\frac{1}{2})}{M!(M-1)!}\right) \approx\big|_{\substack{M=4\\K=3}}\ 0.181\cdot\sigma_s^2,\end{aligned}$$

[2]The operator $\mathrm{diag}(\boldsymbol{x})$ with $\boldsymbol{x} \in \mathbb{C}^K$ results in a $K \times K$ matrix with all zeros except on the diagonal, which contains the elements of $\boldsymbol{x}$. If $\boldsymbol{X} \in \mathbb{C}^{K\times K}$, $\mathrm{diag}(\boldsymbol{X})$ returns a vector with the elements of the diagonal of $\boldsymbol{X}$.

where the last equation represents also the average MMSE over the channel realizations, i.e. $\overline{\text{MMSE}}'_{\text{n},\xi=0}$, since $\text{MMSE}'_{\text{n},\xi=0}$ does not depend on the CDI $\boldsymbol{H}_{\text{n}}$.

Nevertheless, each user k has not made use of the instantaneous norm $\|h_k\|_2$, i.e. of the CMI, which is known at each user. The saturation of the MMSE can be avoided if each user k multiplies its received signal with $\frac{1}{\|\boldsymbol{h}_k\|_2}$ besides the scalar filter g_{n}, such that the resulting channel $\boldsymbol{B}^{-1}\boldsymbol{H} = \boldsymbol{H}_{\text{n}}$ is inherently known at the BS. The multiplication with $\frac{1}{\|\boldsymbol{h}_k\|_2}$ represents a normalization with each user's instantaneous CMI. Let us denote the precoder and the scalar receiver with perfect CDI and with the multiplication of $\frac{1}{\|\boldsymbol{h}_k\|_2}$ at each user k as $\boldsymbol{P}_{\text{n}}$ and g_{n}, respectively. Now the estimated signal is given as

$$\begin{aligned}\hat{\boldsymbol{s}}_{\text{n}} &= g_{\text{n}}\boldsymbol{B}^{-1}\left(\boldsymbol{B}\boldsymbol{H}_{\text{n}}\boldsymbol{P}_{\text{n}}\boldsymbol{s}+\boldsymbol{n}\right)\\ &= g_{\text{n}}\left(\boldsymbol{H}_{\text{n}}\boldsymbol{P}_{\text{n}}\boldsymbol{s}+\boldsymbol{B}^{-1}\boldsymbol{n}\right),\end{aligned} \tag{28}$$

since $\boldsymbol{B}^{-1} = \text{diag}\left(\left[\frac{1}{\|\boldsymbol{h}_1\|_2},\cdots,\frac{1}{\|\boldsymbol{h}_K\|_2}\right]\right)$. Now the overall scalar receiver for each user k is $\frac{g_{\text{k}}}{\|\boldsymbol{h}_k\|_2}$ which is based on the CDI and CMI, while $\boldsymbol{P}_{\text{n}}$ will only be based on the CDI and on the statistics of the CMI since we have assumed that the BS has only acces to the CDI. The effective noise $\boldsymbol{B}^{-1}\boldsymbol{n}$ has covariance matrix $\boldsymbol{R}_{n,\text{n}} = \sigma_n^2\text{E}\left[\boldsymbol{B}^{-2}\right] = \frac{\sigma_n^2}{M-1}\boldsymbol{1}_K$ because

$$\text{E}\left[\boldsymbol{B}^{-2}\right] = \frac{1}{M-1}\boldsymbol{1}_K, \tag{29}$$

since the diagonal of $\boldsymbol{B}^{-2}$ contains the inverse of i.i.d chi-squared random variables with $2M$ degrees of freedom and variance $\frac{1}{2}$. $\boldsymbol{P}_{\text{n}}$ and g_{n} are computed in order to minimize the MSE $\text{E}\left[\|\boldsymbol{s}-\hat{\boldsymbol{s}}_{\text{n}}\|_2^2\right]$ subject to the power constraint, where the expectation is taken over $\boldsymbol{s}$ and the effective noise $\boldsymbol{B}^{-1}\boldsymbol{n}$, by which we also average over the CMI of the users but in a different manner as before. The effective noise in this case is still uncorrelated with the normalized channel since the CMI and the CDI are independent [3]. The solution to the optimization problem with perfect CDI and the normalization with the CMI at each user k results in

$$g_{\text{n}} = \sqrt{\frac{\text{tr}\left(\left(\boldsymbol{H}_{\text{n}}^{\text{H}}\boldsymbol{H}_{\text{n}}+\xi_{\text{n}}\boldsymbol{1}_M\right)^{-2}\boldsymbol{H}_{\text{n}}^{\text{H}}\boldsymbol{R}_s\boldsymbol{H}_{\text{n}}\right)}{P_{\text{DL}}}} \tag{30}$$

$$\boldsymbol{P}_{\text{n}} = \frac{1}{g_{\text{n}}}\left(\boldsymbol{H}_{\text{n}}^{\text{H}}\boldsymbol{H}_{\text{n}}+\xi_{\text{n}}\boldsymbol{1}_M\right)^{-1}\boldsymbol{H}_{\text{n}}^{\text{H}}, \tag{31}$$

where

$$\xi_{\text{n}} = \frac{\text{tr}\left(\boldsymbol{R}_{n,\text{n}}\right)}{P_{\text{DL}}} = \frac{K\sigma_n^2}{P_{\text{DL}}(M-1)} = \frac{\xi}{M-1}. \tag{32}$$

The respective MMSE given the users' CDI $\boldsymbol{H}_{\text{n}}$, i.e. averaging over the noise and the users' CMI $\boldsymbol{B}$, results in

$$\begin{aligned}\text{MMSE}_{\text{n}} &= \xi_{\text{n}}\sigma_s^2\,\text{tr}\left(\left(\boldsymbol{H}_{\text{n}}\boldsymbol{H}_{\text{n}}^{\text{H}}+\xi_{\text{n}}\boldsymbol{1}_M\right)^{-1}\right)\\ &= \xi\sigma_s^2\,\text{tr}\left(\left((M-1)\boldsymbol{H}_{\text{n}}\boldsymbol{H}_{\text{n}}^{\text{H}}+\xi\boldsymbol{1}_M\right)^{-1}\right).\end{aligned} \tag{33}$$

Contrary to (27), (33) does not saturate at high SNR since each user makes use of its known CMI. For $K < M$ and at high SNR, i.e. $\lambda_{\min}\left(\boldsymbol{H}_{\text{n}}\boldsymbol{H}_{\text{n}}^{\text{H}}\right) \gg \xi_{\text{n}}$ with probability 1, (33) becomes

$$\text{MMSE}_{\text{n,high-SNR}} = \frac{\xi}{M-1}\sigma_s^2\,\text{tr}\left(\left(\boldsymbol{H}_{\text{n}}\boldsymbol{H}_{\text{n}}^{\text{H}}\right)^{-1}\right), \tag{34}$$

and averaging this MMSE over the CDI's we get

$$\overline{\text{MMSE}}_{\text{n,high-SNR}} = \xi\,\sigma_s^2\,\frac{K}{M-K} = \overline{\text{MMSE}}_{\text{p,high-SNR}}, \tag{35}$$

where we have used $\text{E}\left[\text{tr}\left(\left(\boldsymbol{H}_{\text{n}}\boldsymbol{H}_{\text{n}}^{\text{H}}\right)^{-1}\right)\right] = \frac{K(M-1)}{M-K}$ which is derived in the following:

$$\begin{aligned}\text{E}\left[\text{tr}\left(\left(\boldsymbol{H}\boldsymbol{H}^{\text{H}}\right)^{-1}\right)\right] &= \text{E}\left[\text{tr}\left(\left(\boldsymbol{B}\boldsymbol{H}_{\text{n}}\boldsymbol{H}_{\text{n}}^{\text{H}}\boldsymbol{B}\right)^{-1}\right)\right]\\ &\overset{(a)}{=} \text{E}\left[\text{tr}\left(\boldsymbol{B}^{-2}\left(\boldsymbol{H}_{\text{n}}\boldsymbol{H}_{\text{n}}^{\text{H}}\right)^{-1}\right)\right]\\ &\overset{(b)}{=} \text{tr}\left(\text{E}\left[\boldsymbol{B}^{-2}\right]\text{E}\left[\left(\boldsymbol{H}_{\text{n}}\boldsymbol{H}_{\text{n}}^{\text{H}}\right)^{-1}\right]\right)\\ &\overset{(c)}{=} \text{tr}\left(\frac{1}{M-1}\boldsymbol{1}_K\,\text{E}\left[\left(\boldsymbol{H}_{\text{n}}\boldsymbol{H}_{\text{n}}^{\text{H}}\right)^{-1}\right]\right)\\ &= \frac{\text{E}\left[\text{tr}\left(\left(\boldsymbol{H}_{\text{n}}\boldsymbol{H}_{\text{n}}^{\text{H}}\right)^{-1}\right)\right]}{M-1} \overset{(d)}{=} \frac{K}{M-K},\end{aligned}$$

where (a) follows from the property of the trace $\text{tr}(\boldsymbol{X}\boldsymbol{Y}) = \text{tr}(\boldsymbol{Y}\boldsymbol{X})$, (b) follows from the fact that the CDI and CMI are independent [9] and (c) follows from (29). Step (d) is due to the result $\text{E}\left[\text{tr}\left(\left(\boldsymbol{H}\boldsymbol{H}^{\text{H}}\right)^{-1}\right)\right] = \frac{K}{M-K}$ from equation (2.9) in [8]. From the last equality we can conclude that

$$\text{E}\left[\text{tr}\left(\left(\boldsymbol{H}_{\text{n}}\boldsymbol{H}_{\text{n}}^{\text{H}}\right)^{-1}\right)\right] = \frac{K(M-1)}{M-K}. \tag{36}$$

$\overline{\text{MMSE}}_{\text{n,high-SNR}}$ from (35) has been averaged over the CDI's and the CMI's, since in (33) we have averaged over the effective noise $\boldsymbol{B}^{-1}\boldsymbol{n}$, which implies averaging also over the CMI. Observe also that $\overline{\text{MMSE}}_{\text{n,high-SNR}} = \overline{\text{MMSE}}_{\text{p,high-SNR}}$, i.e. the average MMSE with perfect CDI and the statistics of the CMI converges to the average MMSE with perfect CDI and perfect CMI at high SNR. Recall that $K < M$ and that no user selection is performed at the base station.

V. FDD System: Quantized CDI

In an FDD system the estimate of the uplink channels cannot be employed as transmit CSI in the downlink, so we assume the BS obtains transmit CSI through limited feedback of B bits from each user, similarly as it is done for the single-user case in [10] and for the multi-user case in [5]. As explained in the following, the BS obtains a quantized version of an estimate of the CDI of each user, which together with the statistics of the users' CMI comprise the available transmit CSI at the BS. To this end, each user k obtains a minimum mean square error estimate of its downlink channel using a common pilot of length $T_{\text{DL}} \geq M$ which is emitted from the base station. Since the transmit power at the base station is allocated over M antennas we have that the error variance is given as [1]

$$\sigma_{e_{\text{DL}}}^2 = \frac{1}{1+\frac{P_{\text{DL}}}{M\sigma_n^2}T_{\text{DL}}}, \tag{37}$$

where P_{DL} is the transmit power at the BS and σ_n^2 is the variance of the AWGN at the users. Denote the estimated channel of user k as $\hat{\boldsymbol{h}}_k \in \mathbb{C}^M$ and $\hat{\boldsymbol{H}} = \left[\hat{\boldsymbol{h}}_1,\cdots,\hat{\boldsymbol{h}}_K\right]^{\text{T}} \in \mathbb{C}^{K\times M}$. The channel from all users can be written as

$$\boldsymbol{H} = \hat{\boldsymbol{H}} + \boldsymbol{E}, \tag{38}$$

where $\boldsymbol{E} = [\boldsymbol{e}_1, \cdots, \boldsymbol{e}_K]^{\mathrm{T}} \in \mathbb{C}^{K\times M}$ is the estimation error matrix, where $\boldsymbol{e}_k$ is the estimation error vector when user k estimates its channel. The elements of $\boldsymbol{E}$ are zero-mean complex Gaussian random variables with variance $\sigma^2_{e_{\mathrm{DL}}}$ while the elements of $\hat{\boldsymbol{H}}$ are zero-mean complex Gaussian random variables with variance $(1-\sigma^2_{e_{\mathrm{DL}}})$. Here we have abused the notation since we have employed the same notation for the estimated channel and error that arises in the TDD system described in Section III. The only difference is the variance of the estimation error, which in a TDD system it is given in (9) while in an FDD system it is given in (37). As in the TDD case, the MMSE estimate $\hat{\boldsymbol{h}}_k$ has the property that $\mathrm{E}\left[\hat{\boldsymbol{h}}_k^{\mathrm{H}}\boldsymbol{e}_k\right] = \boldsymbol{0}$ and furthermore due to the independence assumption among the users, we have that $\mathrm{E}\left[\hat{\boldsymbol{h}}_k^{\mathrm{H}}\boldsymbol{e}_j\right] = \boldsymbol{0}\ \forall j, k$, such that

$$\mathrm{E}\left[\hat{\boldsymbol{H}}^{\mathrm{H}}\boldsymbol{E}\right] = \boldsymbol{0}. \tag{39}$$

After estimating its channel, user k normalizes its estimated channel $\hat{\boldsymbol{h}}_k$ in order to obtain $\hat{\boldsymbol{h}}_{k,\mathrm{n}} = \frac{\hat{\boldsymbol{h}}_k}{\|\hat{\boldsymbol{h}}_k\|_2}$, which represents an estimate of the CDI with $\|\hat{\boldsymbol{h}}_k\|_2$ as an estimate of the CMI. Each user k then quantizes its estimated CDI, i.e. the *normalized* channel $\hat{\boldsymbol{h}}_{k,\mathrm{n}}$, with B bits, which are then relayed back to the BS. We assume the feedback link to be error-free and orthogonal among the users. For the quantization, we consider the *random vector quantization* (RVQ) [3] as follows

$$\hat{\boldsymbol{h}}_{k,\mathrm{q}} = \underset{\boldsymbol{t}_{k,j}\in\mathcal{C}_k}{\operatorname{argmax}}\ |\boldsymbol{t}_{k,j}^{\mathrm{H}}\hat{\boldsymbol{h}}_{k,\mathrm{n}}|, \tag{40}$$

where each user k has a different codebook $\mathcal{C}_k$ consisting of 2^B unit-norm random beamforming vectors $\boldsymbol{t}_{k,j} \in \mathcal{C}_k$, which is also available at the BS. We assume a different codebook for each user, since otherwise there exists a non-zero probability that two or more users feed back the same channel vector. Denote $\hat{\boldsymbol{H}}_{\mathrm{q}} = \left[\hat{\boldsymbol{h}}_{1,\mathrm{q}}, \cdots, \hat{\boldsymbol{h}}_{K,\mathrm{q}}\right]^{\mathrm{T}} \in \mathbb{C}^{K\times M}$. After *error-free* feedback of the B bits from each user, the BS would have access to $\hat{\boldsymbol{H}}_{\mathrm{q}}$ as transmit CSI. Note that $\hat{\boldsymbol{h}}_{k,\mathrm{q}}$ represents a quantized version of the estimated *channel direction information* of user k. Let us denote $c_k = \hat{\boldsymbol{h}}_{k,\mathrm{q}}^{\mathrm{H}}\hat{\boldsymbol{h}}_{k,\mathrm{n}} \in \mathbb{C}$ with $|c_k| \le 1$ since

$$|c_k| = \|\hat{\boldsymbol{h}}_{k,\mathrm{q}}\|_2\|\hat{\boldsymbol{h}}_{k,\mathrm{n}}\|_2\cos\theta_k = \cos\theta_k, \tag{41}$$

since $\hat{\boldsymbol{h}}_{k,\mathrm{q}}$ and $\hat{\boldsymbol{h}}_{k,\mathrm{n}}$ have unit norm and where θ_k is the angle between $\hat{\boldsymbol{h}}_{k,\mathrm{q}}$ and $\hat{\boldsymbol{h}}_{k,\mathrm{n}}$. In addition we point out that $\arg\ (c_k) \neq 0$, which has to be considered in the design of the receiver. With $c_k = \hat{\boldsymbol{h}}_{k,\mathrm{q}}^{\mathrm{H}}\hat{\boldsymbol{h}}_{k,\mathrm{n}}$, we can express the normalized estimated channel in terms of the quantized channel as

$$\hat{\boldsymbol{h}}_{k,\mathrm{n}} = c_k\hat{\boldsymbol{h}}_{k,\mathrm{q}} + \boldsymbol{e}_{k,\mathrm{q}}, \tag{42}$$

such that $\hat{\boldsymbol{h}}_{k,\mathrm{q}}$ and the quantization error $\boldsymbol{e}_{k,\mathrm{q}}$ are orthogonal:

$$\hat{\boldsymbol{h}}_{k,\mathrm{q}}^{\mathrm{H}}\boldsymbol{e}_{k,\mathrm{q}} = 0, \tag{43}$$

$$\operatorname{diag}\left(\hat{\boldsymbol{H}}_{\mathrm{q}}^{\mathrm{H}}\boldsymbol{E}_{\mathrm{q}}\right) = \boldsymbol{0}. \tag{44}$$

This means that $\boldsymbol{e}_{\mathrm{q},k}$ lies in the nullspace of $\hat{\boldsymbol{h}}_{k,\mathrm{q}}^{\mathrm{H}}$ and is uniformly distributed in the $M-1$ dimensional nullspace. In addition, let us compute the norm of $\boldsymbol{e}_{k,\mathrm{q}}$:

$$\begin{aligned} 1 = \|\hat{\boldsymbol{h}}_{k,\mathrm{n}}\|_2^2 &= \|c_k\hat{\boldsymbol{h}}_{k,\mathrm{q}} + \boldsymbol{e}_{k,\mathrm{q}}\|_2^2 \\ &\overset{(a)}{=} |c_k|^2\|\hat{\boldsymbol{h}}_{k,\mathrm{q}}\|_2^2 + \|\boldsymbol{e}_{k,\mathrm{q}}\|_2^2 \\ &\overset{(b)}{=} \cos^2\theta_k + \|\boldsymbol{e}_{k,\mathrm{q}}\|_2^2 \\ \|\boldsymbol{e}_{k,\mathrm{q}}\|_2^2 &\overset{(c)}{=} 1-\cos^2\theta_k = \sin^2\theta_k, \end{aligned} \tag{45}$$

where (a) follows from (43), while (b) and (c) follow from the fact that $\|\hat{\boldsymbol{h}}_{k,\mathrm{q}}\|_2^2 = 1$ and $\|\hat{\boldsymbol{h}}_{k,\mathrm{n}}\|_2^2 = 1$. Note that the quantization error of user k decreases as θ_k becomes smaller. Denote $\hat{\boldsymbol{B}} = \operatorname{diag}\left(\left[\|\hat{\boldsymbol{h}}_1\|_2, \cdots, \|\hat{\boldsymbol{h}}_K\|_2\right]\right)$, $\boldsymbol{C} = \operatorname{diag}\left([c_1, \cdots, c_K]\right)$ and $\boldsymbol{E}_{\mathrm{q}} = [\boldsymbol{e}_{1,\mathrm{q}}, \cdots, \boldsymbol{e}_{K,\mathrm{q}}]^{\mathrm{T}}$. Using (42) and collecting the estimated CDI of all users, we have

$$\hat{\boldsymbol{H}}_{\mathrm{n}} = \boldsymbol{C}_k\hat{\boldsymbol{H}}_{\mathrm{q}} + \boldsymbol{E}_{\mathrm{q}}. \tag{46}$$

The channel $\boldsymbol{H}$ from (38) is rewritten as

$$\begin{aligned} \boldsymbol{H} &= \hat{\boldsymbol{H}} + \boldsymbol{E} = \hat{\boldsymbol{B}}\hat{\boldsymbol{H}}_{\mathrm{n}} + \boldsymbol{E} \\ &= \hat{\boldsymbol{B}}\left(\boldsymbol{C}\hat{\boldsymbol{H}}_{\mathrm{q}} + \boldsymbol{E}_q\right) + \boldsymbol{E} \\ &= \hat{\boldsymbol{B}}\boldsymbol{C}\hat{\boldsymbol{H}}_{\mathrm{q}} + \hat{\boldsymbol{B}}\boldsymbol{E}_q + \boldsymbol{E}. \end{aligned} \tag{47}$$

A. Transceiver Optimization with Quantized CDI

Motivated by the result from the perfect CDI case (c.f. Section IV), each user k multiplies its received signal with $\frac{1}{c_k\|\hat{\boldsymbol{h}}_k\|_2}$ besides the scalar filter g_{q} such that the estimated signal of the K users is given by

$$\hat{\boldsymbol{s}}_{\mathrm{q}} = g_{\mathrm{q}}\ \boldsymbol{C}^{-1}\hat{\boldsymbol{B}}^{-1}\left(\boldsymbol{H}\boldsymbol{P}\boldsymbol{s} + \boldsymbol{n}\right) = g_{\mathrm{q}}\left(\hat{\boldsymbol{H}}_{\mathrm{q}}\boldsymbol{P}_{\mathrm{q}}\boldsymbol{s} + \boldsymbol{n}_{\mathrm{q}}\right), \tag{48}$$

where the effective noise

$$\boldsymbol{n}_{\mathrm{q}} = \boldsymbol{C}^{-1}\boldsymbol{E}_q\boldsymbol{P}_{\mathrm{q}}\boldsymbol{s} + \boldsymbol{C}^{-1}\hat{\boldsymbol{B}}^{-1}\boldsymbol{E}\boldsymbol{P}_{\mathrm{q}}\boldsymbol{s} + \boldsymbol{C}^{-1}\hat{\boldsymbol{B}}^{-1}\boldsymbol{n}. \tag{49}$$

Note that the scalar receiver at user k is actually $\frac{g_{\mathrm{q}}}{c_k\|\hat{\boldsymbol{h}}_k\|_2}$ which is based not only on the estimated CDI but also on the estimated CMI and c_k. We compute now $\boldsymbol{P}_{\mathrm{q}}$ and g_{q} which minimize the MSE based on $\hat{\boldsymbol{H}}_{\mathrm{q}}$ and the statistics of the estimated CMI, i.e. $\hat{\boldsymbol{B}}$, are known at the BS.

$$\{\boldsymbol{P}_{\mathrm{q}}, g_{\mathrm{q}}\} = \underset{\{\boldsymbol{P}_{\mathrm{q}}, g_{\mathrm{q}}\}}{\operatorname{argmin}}\ \mathrm{E}\left[\|\boldsymbol{s} - \hat{\boldsymbol{s}}_{\mathrm{q}}\|_2^2\right]\ \text{ s.t. } \operatorname{tr}\left(\boldsymbol{P}_{\mathrm{q}}\boldsymbol{R}_s\boldsymbol{P}_{\mathrm{q}}^{\mathrm{H}}\right) \le P_{\mathrm{DL}}, \tag{50}$$

where the expected value is taken over $\boldsymbol{s}$, $\boldsymbol{n}$, $\boldsymbol{E}$, $\hat{\boldsymbol{B}}$, $\boldsymbol{C}$ and $\boldsymbol{E}_{\mathrm{q}}$ since $\hat{\boldsymbol{H}}_{\mathrm{q}}$ is known at the BS. Let us first calculate $\mathrm{E}\left[\|\boldsymbol{s} - \hat{\boldsymbol{s}}_{\mathrm{q}}\|_2^2\right]$, i.e.

$$\begin{aligned} \mathrm{E}\left[\|\boldsymbol{s} - \hat{\boldsymbol{s}}_{\mathrm{q}}\|_2^2\right] &= \mathrm{E}\left[\left\|\left(\boldsymbol{1}_K - g_{\mathrm{q}}\hat{\boldsymbol{H}}_{\mathrm{q}}\boldsymbol{P}_{\mathrm{q}}\right)\boldsymbol{s} + g_{\mathrm{q}}\boldsymbol{n}_{\mathrm{q}}\right\|_2^2\right] \\ &= \operatorname{tr}\left(\mathrm{E}\left[\left(\boldsymbol{1}_K - g_{\mathrm{q}}\hat{\boldsymbol{H}}_{\mathrm{q}}\boldsymbol{P}_{\mathrm{q}}\right)\boldsymbol{s}\boldsymbol{s}^{\mathrm{H}}\left(\boldsymbol{1}_K - g_{\mathrm{q}}\hat{\boldsymbol{H}}_{\mathrm{q}}\boldsymbol{P}_{\mathrm{q}}\right)^{\mathrm{H}}\right]\right) \\ &\quad + \operatorname{tr}\left(|g_{\mathrm{q}}|^2\mathrm{E}\left[\boldsymbol{n}_{\mathrm{q}}\boldsymbol{n}_{\mathrm{q}}^{\mathrm{H}}\right]\right) \\ &= \operatorname{tr}\Big(\boldsymbol{R}_s - 2\mathrm{Re}\left(g_{\mathrm{q}}\hat{\boldsymbol{H}}_{\mathrm{q}}\boldsymbol{P}_{\mathrm{q}}\boldsymbol{R}_s\right) + \\ &\quad |g_{\mathrm{q}}|^2\hat{\boldsymbol{H}}_{\mathrm{q}}\boldsymbol{P}_{\mathrm{q}}\boldsymbol{R}_s\boldsymbol{P}_{\mathrm{q}}^{\mathrm{H}}\hat{\boldsymbol{H}}_{\mathrm{q}}^{\mathrm{H}} + |g_{\mathrm{q}}|^2\boldsymbol{R}_{n,\mathrm{q}}\Big), \end{aligned} \tag{51}$$

where $\boldsymbol{R}_{n,\mathrm{q}}$ is the covariance matrix of the effective noise and the cross terms with $\mathrm{E}\left[\boldsymbol{s}_{\mathrm{q}}\boldsymbol{n}_{\mathrm{q}}^{\mathrm{H}}\right] = \boldsymbol{0}$ (similarly to (14)) due to the fact that the estimation error $\boldsymbol{E}$, quantization error $\boldsymbol{E}_{\mathrm{q}}$ and noise $\boldsymbol{n}$ are independent with the signal and in addition have

zero mean, i.e. $\mathrm{E}[\boldsymbol{E}]=\mathbf{0}$, $\mathrm{E}[\boldsymbol{E}_{\mathrm{q}}]=\mathbf{0}$ and $\mathrm{E}[\boldsymbol{n}]=\mathbf{0}$. For the optimization we need to compute $\mathrm{tr}(\boldsymbol{R}_{n,\mathrm{q}})=\mathrm{tr}\left(\mathrm{E}\left[\boldsymbol{n}_{\mathrm{q}}\boldsymbol{n}_{\mathrm{q}}^{\mathrm{H}}\right)\right]$:

$$\begin{aligned}\mathrm{tr}(\boldsymbol{R}_{n,\mathrm{q}})=&\ \mathrm{tr}\left(\mathrm{E}\left[\boldsymbol{C}^{-1}\boldsymbol{E}_q\boldsymbol{P}_{\mathrm{q}}\boldsymbol{s}\boldsymbol{s}^{\mathrm{H}}\boldsymbol{P}_{\mathrm{q}}^{\mathrm{H}}\boldsymbol{E}_q^{\mathrm{H}}\boldsymbol{C}^{-1,\mathrm{H}}\right]\right)+\\ &\mathrm{tr}\left(\mathrm{E}\left[\boldsymbol{C}^{-1}\hat{\boldsymbol{B}}^{-1}\boldsymbol{E}\boldsymbol{P}_{\mathrm{q}}\boldsymbol{s}\boldsymbol{s}^{\mathrm{H}}\boldsymbol{P}_{\mathrm{q}}^{\mathrm{H}}\boldsymbol{E}^{\mathrm{H}}\hat{\boldsymbol{B}}^{-1}\boldsymbol{C}^{-1,\mathrm{H}}\right]\right)+\\ &\mathrm{tr}\left(\mathrm{E}\left[\boldsymbol{C}^{-1}\hat{\boldsymbol{B}}^{-1}\boldsymbol{n}\boldsymbol{n}^{\mathrm{H}}\hat{\boldsymbol{B}}^{-1}\boldsymbol{C}^{-1,\mathrm{H}}\right]\right),\end{aligned}\tag{52}$$

where the other cross terms are $\mathbf{0}$ because the signal and the noise are uncorrelated and because the estimation error $\boldsymbol{E}$ is independent of the quantization error $\boldsymbol{E}_{\mathrm{q}}$ and has zero mean. We proceed to compute the three terms in (52).

Let us first focus on the second and third term, which are easier to compute. The second term can be written as

$$\begin{aligned}&\mathrm{tr}\left(\mathrm{E}\left[\boldsymbol{C}^{-1}\hat{\boldsymbol{B}}^{-1}\boldsymbol{E}\boldsymbol{P}_{\mathrm{q}}\boldsymbol{s}\boldsymbol{s}^{\mathrm{H}}\boldsymbol{P}_{\mathrm{q}}^{\mathrm{H}}\boldsymbol{E}^{\mathrm{H}}\hat{\boldsymbol{B}}^{-1}\boldsymbol{C}^{-1,\mathrm{H}}\right]\right)\overset{(a)}{=}\\ &\mathrm{tr}\left(\mathrm{E}\left[\boldsymbol{C}_{\mathrm{m}}^{-2}\hat{\boldsymbol{B}}^{-1}\boldsymbol{E}\boldsymbol{P}_{\mathrm{q}}\boldsymbol{R}_s\boldsymbol{P}_{\mathrm{q}}^{\mathrm{H}}\boldsymbol{E}^{\mathrm{H}}\hat{\boldsymbol{B}}^{-1}\right]\right)\overset{(b)}{=}\\ &\mathrm{E}\left[\cos^{-2}\theta_k\right]\mathrm{tr}\left(\mathrm{E}\left[\hat{\boldsymbol{B}}^{-2}\right]\mathrm{E}\left[\boldsymbol{E}\boldsymbol{P}_{\mathrm{q}}\boldsymbol{R}_s\boldsymbol{P}_{\mathrm{q}}^{\mathrm{H}}\boldsymbol{E}^{\mathrm{H}}\right]\right)\overset{(c)}{=}\\ &\frac{\mathrm{E}\left[\cos^{-2}\theta_k\right]}{(M-1)(1-\sigma_{e_{\mathrm{DL}}}^2)}\mathrm{tr}\left(\mathrm{E}\left[\boldsymbol{E}\boldsymbol{P}_{\mathrm{q}}\boldsymbol{R}_s\boldsymbol{P}_{\mathrm{q}}^{\mathrm{H}}\boldsymbol{E}^{\mathrm{H}}\right]\right)\overset{(d)}{=}\\ &\frac{\mathrm{E}\left[\cos^{-2}\theta_k\right]}{(M-1)(1-\sigma_{e_{\mathrm{DL}}}^2)}\mathrm{tr}\left(\mathrm{tr}\left(\boldsymbol{P}_{\mathrm{q}}\boldsymbol{R}_s\boldsymbol{P}_{\mathrm{q}}^{\mathrm{H}}\right)\sigma_{e_{\mathrm{DL}}}^2\mathbf{1}_K\right)=\\ &\frac{\mathrm{E}\left[\cos^{-2}\theta_k\right]}{(M-1)(1-\sigma_{e_{\mathrm{DL}}}^2)}K\,\sigma_{e_{\mathrm{DL}}}^2\,\mathrm{tr}\left(\boldsymbol{P}_{\mathrm{q}}\boldsymbol{R}_s\boldsymbol{P}_{\mathrm{q}}^{\mathrm{H}}\right),\end{aligned}\tag{53}$$

where in (a) we took the expectation over $\boldsymbol{s}$, applied the property of the trace $\mathrm{tr}(\boldsymbol{XY})=\mathrm{tr}(\boldsymbol{XY})$ and have substituted

$$\begin{aligned}\boldsymbol{C}_{\mathrm{m}}^{-2}&=\boldsymbol{C}^{-1,\mathrm{H}}\boldsymbol{C}^{-1}=\mathrm{diag}\left(\left[\frac{1}{|c_1|^2},\cdots,\frac{1}{|c_K|^2}\right]\right)\\ &=\mathrm{diag}\left(\left[\cos^{-2}\theta_1,\cdots,\cos^{-2}\theta_K\right]\right),\end{aligned}\tag{54}$$

where we used (41). In step (b) we have taken the expectation over $\boldsymbol{C}_{\mathrm{m}}^{-2}$ since it is independent of the other random variables and we point that the angles $\theta_1,\cdots,\theta_K$ all have the same distribution, i.e. of θ_k. In step (b) we have also applied the beforementioned property of the trace and use that $\hat{\boldsymbol{B}}$ is independent of $\boldsymbol{E}$. Step (c) follows from (29) with the fact that the estimated channel has variance $(1-\sigma_{e_{\mathrm{DL}}}^2)$. Step (d) can be derived similary to (15) and taking into account that the estimation error variance is $\sigma_{e_{\mathrm{DL}}}^2$.

The third term in (52) is

$$\begin{aligned}&\mathrm{tr}\left(\mathrm{E}\left[\boldsymbol{C}^{-1}\hat{\boldsymbol{B}}^{-1}\boldsymbol{n}\boldsymbol{n}^{\mathrm{H}}\hat{\boldsymbol{B}}^{-1}\boldsymbol{C}^{-1,\mathrm{H}}\right]\right)=\mathrm{tr}\left(\mathrm{E}\left[\boldsymbol{C}_{\mathrm{m}}^{-2}\hat{\boldsymbol{B}}^{-1}\boldsymbol{R}_n\hat{\boldsymbol{B}}^{-1}\right]\right)\\ &=\mathrm{E}\left[\cos^{-2}\right]\mathrm{tr}\left(\mathrm{E}\left[\hat{\boldsymbol{B}}^{-2}\right]\boldsymbol{R}_n\right)=\frac{\mathrm{E}\left[\cos^{-2}\theta_k\right]}{(M-1)(1-\sigma_{e_{\mathrm{DL}}}^2)}K\sigma_{\mathrm{n}}^2.\end{aligned}\tag{55}$$

B. Noise due to the Quantization

We now compute the first term in (52), where the expectation is taken over $\boldsymbol{C}$, $\boldsymbol{E}_q$ and $\boldsymbol{s}$. We can write

$$\begin{aligned}&\mathrm{tr}\left(\mathrm{E}\left[\boldsymbol{C}^{-1}\boldsymbol{E}_q\boldsymbol{P}_{\mathrm{q}}\boldsymbol{s}\boldsymbol{s}^{\mathrm{H}}\boldsymbol{P}_{\mathrm{q}}^{\mathrm{H}}\boldsymbol{E}_q^{\mathrm{H}}\boldsymbol{C}^{-1,\mathrm{H}}\right]\right)=\\ &\mathrm{tr}\left(\mathrm{E}\left[\boldsymbol{E}_q^{\mathrm{H}}\boldsymbol{C}_{\mathrm{m}}^{-2}\boldsymbol{E}_q\right]\boldsymbol{P}_{\mathrm{q}}\boldsymbol{R}_s\boldsymbol{P}_{\mathrm{q}}^{\mathrm{H}}\right)=\mathrm{tr}\left(\mathrm{E}\left[\boldsymbol{E}_q^{',\mathrm{H}}\boldsymbol{E}_{\mathrm{q}}^{'}\right]\boldsymbol{P}_{\mathrm{q}}\boldsymbol{R}_s\boldsymbol{P}_{\mathrm{q}}^{\mathrm{H}}\right),\end{aligned}\tag{56}$$

where in the first step we have taken the expectation over $\boldsymbol{s}$, applied the beforementioned property of the trace and used (54). The last equality follows from the substitution

$$\boldsymbol{E}_{\mathrm{q}}^{'}=\boldsymbol{C}_{\mathrm{m}}^{-1}\boldsymbol{E}_q=\left[\frac{\boldsymbol{e}_{\mathrm{q},1}}{|c_1|},\cdots,\frac{\boldsymbol{e}_{\mathrm{q},K}}{|c_K|},\right]^{\mathrm{T}}.\tag{57}$$

In order to compute the expectation $\mathrm{E}\left[\boldsymbol{E}_q^{',\mathrm{H}}\boldsymbol{E}_{\mathrm{q}}^{'}\right]$ in (56), we have to recall that $\boldsymbol{E}_{\mathrm{q}}$ (and also $\boldsymbol{E}_{\mathrm{q}}^{'}$) is dependent on $\hat{\boldsymbol{H}}_{\mathrm{q}}$ as can be seen from (44). This means that $\frac{\boldsymbol{e}_{\mathrm{q},k}}{c_k|}$ lies in the nullspace of $\hat{\boldsymbol{h}}_{k,\mathrm{q}}^{\mathrm{H}}$, for which we denote a unitary basis as $\boldsymbol{U}_k\in\mathbb{C}^{M\times M-1}$, such that $\boldsymbol{h}_k^{\mathrm{H}}\boldsymbol{U}_k=\mathbf{0}$. Hence, we can write

$$\frac{\boldsymbol{e}_{\mathrm{q},k}}{|c_k|}=\boldsymbol{U}_k\boldsymbol{x}_k,\tag{58}$$

where $\boldsymbol{x}_k\in\mathbb{C}^{M-1}$ is a random vector uniformly distributed over the $M-1$ dimensional space, with i.i.d elements. Using (41), (45) and $\boldsymbol{U}_k^{\mathrm{H}}\boldsymbol{U}_k=\mathbf{1}_{M-1}$ we can show that

$$\begin{aligned}\left\|\frac{\boldsymbol{e}_{\mathrm{q},k}}{|c_k|}\right\|_2^2&=\|\boldsymbol{U}_k\boldsymbol{x}_k\|_2^2=\boldsymbol{x}_k^{\mathrm{H}}\boldsymbol{U}_k^{\mathrm{H}}\boldsymbol{U}_k\boldsymbol{x}_k\\ &=\|\boldsymbol{x}_k\|_2^2=\frac{\|\boldsymbol{e}_{\mathrm{q},k}\|_2^2}{|c_k|^2}\\ \|\boldsymbol{x}_k\|_2^2&=\frac{\sin^2\theta_k}{\cos^2\theta_k}=\tan^2\theta_k,\end{aligned}\tag{59}$$

and since the elements of $\boldsymbol{x}_k=[x_{k,1},\cdots,x_{k,M-1}]^{\mathrm{T}}$ are i.i.d. we have that

$$\begin{aligned}\mathrm{E}\left[\|\boldsymbol{x}_k\|_2^2\right]&=\mathrm{E}\left[\sum_i^{M-1}|x_{k,i}|^2\right]=\sum_i^{M-1}\mathrm{E}\left[|x_{k,i}|^2\right]\\ &=(M-1)\,\mathrm{E}\left[|x_{k,i}|^2\right]\overset{(a)}{=}\mathrm{E}\left[\tan^2\theta_k\right]\\ \mathrm{E}\left[|x_{k,i}|^2\right]&=\frac{1}{M-1}\mathrm{E}\left[\tan^2\theta_k\right],\end{aligned}\tag{60}$$

where (a) follows from (59). We are finally ready to compute $\mathrm{E}\left[\boldsymbol{E}_q^{',\mathrm{H}}\boldsymbol{E}_{\mathrm{q}}^{'}\right]$

$$\begin{aligned}\mathrm{E}\left[\boldsymbol{E}_q^{',\mathrm{H}}\boldsymbol{E}_{\mathrm{q}}^{'}\right]&=\mathrm{E}\left[\sum_k^K\left(\frac{\boldsymbol{e}_{\mathrm{q},k}^*}{|c_k|}\right)\left(\frac{\boldsymbol{e}_{\mathrm{q},k}^{\mathrm{T}}}{|c_k|}\right)\right]\\ &=\sum_k^K\mathrm{E}\left[\left(\frac{\boldsymbol{e}_{\mathrm{q},k}^*}{|c_k|}\right)\left(\frac{\boldsymbol{e}_{\mathrm{q},k}^{\mathrm{T}}}{|c_k|}\right)\right]\\ &\overset{(a)}{=}\sum_k^K\boldsymbol{U}_k^*\mathrm{E}\left[\boldsymbol{x}_k^*\boldsymbol{x}_k^{\mathrm{T}}\right]\boldsymbol{U}_k^{\mathrm{T}}\\ &\overset{(b)}{=}\sum_k^K\boldsymbol{U}_k^*\cdot\mathrm{E}\left[|x_{k,i}|^2\right]\mathbf{1}_{M-1}\cdot\boldsymbol{U}_k^{\mathrm{T}}\\ &\overset{(c)}{=}\sum_k^K\frac{\mathrm{E}\left[\tan^2\theta_k\right]}{M-1}\boldsymbol{U}_k^*\boldsymbol{U}_k^{\mathrm{T}}\\ &\overset{(d)}{=}\frac{\mathrm{E}\left[\tan^2\theta_k\right]}{M-1}\sum_k^K\left(\mathbf{1}_M-\hat{\boldsymbol{h}}_{\mathrm{q},k}^*\hat{\boldsymbol{h}}_{\mathrm{q},k}^{\mathrm{T}}\right)\\ &=\frac{\mathrm{E}\left[\tan^2\theta_k\right]}{M-1}\left(K\mathbf{1}_M-\hat{\boldsymbol{H}}_{\mathrm{q}}^{\mathrm{H}}\hat{\boldsymbol{H}}_{\mathrm{q}}\right),\end{aligned}\tag{61}$$

where (a) follows from (58), (b) from the i.i.d. property of the elements of $\boldsymbol{x}_k$ and (c) follows from (60) and from the fact that all users have the same codebook size, i.e. they have the same statistics. For step (d) note that $\boldsymbol{P}_{\mathcal{N},k}=\boldsymbol{U}_k\boldsymbol{U}_k^{\mathrm{H}}$ is the projector

onto the nullspace of $\boldsymbol{h}_k$ which is $\boldsymbol{P}_{\mathcal{N},k} = \mathbf{1}_M - \hat{\boldsymbol{h}}_{\mathrm{q},k}\hat{\boldsymbol{h}}_{\mathrm{q},k}^{\mathrm{H}}$ since $\hat{\boldsymbol{h}}_{\mathrm{q},k}$ has unit norm. Step (d) follows by taking the conjugate of $\boldsymbol{P}_{\mathcal{N},k}$, i.e. $\boldsymbol{P}_{\mathcal{N},k}^* = \left(\boldsymbol{U}_k\boldsymbol{U}_k^{\mathrm{H}}\right)^* = \left(\mathbf{1}_M - \hat{\boldsymbol{h}}_{\mathrm{q},k}^*\hat{\boldsymbol{h}}_{\mathrm{q},k}^{\mathrm{T}}\right)$.

Plugging (61) in (56) gives us that the first term in (52) is

$$\mathrm{tr}\left(\mathrm{E}\left[\boldsymbol{C}^{-1}\boldsymbol{E}_q\boldsymbol{P}_{\mathrm{q}}\boldsymbol{s}\boldsymbol{s}^{\mathrm{H}}\boldsymbol{P}_{\mathrm{q}}^{\mathrm{H}}\boldsymbol{E}_q^{\mathrm{H}}\boldsymbol{C}^{-1,\mathrm{H}}\right]\right) = \frac{\mathrm{E}\left[\tan^2\theta_k\right]}{M-1}\mathrm{tr}\left(\left(K\mathbf{1}_M - \hat{\boldsymbol{H}}_{\mathrm{q}}^{\mathrm{H}}\hat{\boldsymbol{H}}_{\mathrm{q}}\right)\boldsymbol{P}_{\mathrm{q}}\boldsymbol{R}_s\boldsymbol{P}_{\mathrm{q}}^{\mathrm{H}}\right). \quad (62)$$

$\mathrm{tr}\left(\boldsymbol{R}_{n,\mathrm{q}}\right)$ from (52) using (62), (53) and (55) is given as

$$\mathrm{tr}\left(\boldsymbol{R}_{n,\mathrm{q}}\right) = \frac{\mathrm{E}\left[\tan^2\theta_k\right]}{M-1}\mathrm{tr}\left(\left(K\mathbf{1}_M - \hat{\boldsymbol{H}}_{\mathrm{q}}^{\mathrm{H}}\hat{\boldsymbol{H}}_{\mathrm{q}}\right)\boldsymbol{P}_{\mathrm{q}}\boldsymbol{R}_s\boldsymbol{P}_{\mathrm{q}}^{\mathrm{H}}\right) + \frac{\mathrm{E}\left[\cos^{-2}\theta_k\right]}{(M-1)(1-\sigma_{e_{\mathrm{DL}}}^2)}K\left(\mathrm{tr}\left(\boldsymbol{P}_{\mathrm{q}}\boldsymbol{R}_s\boldsymbol{P}_{\mathrm{q}}^{\mathrm{H}}\right)\sigma_{e_{\mathrm{DL}}}^2 + \sigma_{\mathrm{n}}^2\right). \quad (63)$$

C. MMSE with Quantized CDI

Note that $\mathrm{tr}\left(\boldsymbol{R}_{n,\mathrm{q}}\right)$ depends on $\boldsymbol{P}_{\mathrm{q}}$ as can be seen from (63). Considering this, the solution to (50) using (51) results in

$$g_{\mathrm{q}} = \sqrt{\frac{\mathrm{tr}\left(\left((1-\kappa)\hat{\boldsymbol{H}}_{\mathrm{q}}^{\mathrm{H}}\hat{\boldsymbol{H}}_{\mathrm{q}} + \xi_{\mathrm{q}}\mathbf{1}_M\right)^{-2}\hat{\boldsymbol{H}}_{\mathrm{q}}^{\mathrm{H}}\boldsymbol{R}_s\hat{\boldsymbol{H}}_{\mathrm{q}}\right)}{P_{\mathrm{DL}}}} \quad (64)$$

$$\boldsymbol{P}_{\mathrm{q}} = \frac{1}{g_{\mathrm{q}}}\left((1-\kappa)\hat{\boldsymbol{H}}_{\mathrm{q}}^{\mathrm{H}}\hat{\boldsymbol{H}}_{\mathrm{q}} + \xi_{\mathrm{q}}\mathbf{1}_M\right)^{-1}\hat{\boldsymbol{H}}_{\mathrm{q}}^{\mathrm{H}}, \quad (65)$$

where

$$\kappa = \frac{\mathrm{E}\left[\tan^2\theta_k\right]}{M-1} \quad (66)$$

$$\xi_{\mathrm{q}} = K\frac{\mathrm{E}\left[\tan^2\theta_k\right]}{M-1} + K\frac{\mathrm{E}\left[\cos^{-2}\theta_k\right]}{(M-1)(1-\sigma_{e_{\mathrm{DL}}}^2)}\left(\sigma_{e_{\mathrm{DL}}}^2 + \frac{\sigma_n^2}{P_{\mathrm{DL}}}\right), \quad (67)$$

where we can use the approximation $\mathrm{E}\left[\cos^{-2}\theta_k\right] \approx \frac{1}{\mathrm{E}[\cos^2\theta_k]}$ and where $\mathrm{E}\left[\cos^2\theta_k\right] = 1 - \mathrm{E}\left[\sin^2\theta_k\right]$, with [4]

$$\mathrm{E}\left[\sin^2\theta_k\right] = 2^B\,\mathrm{Beta}\left(2^B, \frac{M}{M-1}\right), \quad (68)$$

and $\mathrm{E}\left[\tan^2\theta_k\right] \approx \frac{\mathrm{E}[\sin^2\theta_k]}{\mathrm{E}[\cos^2\theta_k]}$. Both approximations represent lower bounds for very small B but become quite tight as B increases. When solving (50) we point out that the power constraint is fulfilled with equality, i.e. $\mathrm{tr}\left(\boldsymbol{P}_{\mathrm{q}}\boldsymbol{R}_s\boldsymbol{P}_{\mathrm{q}}^{\mathrm{H}}\right) = P_{\mathrm{DL}}$. Note that in this case $\xi_{\mathrm{q}} \neq \frac{\mathrm{tr}(\boldsymbol{R}_{n,\mathrm{q}})}{P_{\mathrm{DL}}}$, since $\mathrm{tr}\left(\boldsymbol{R}_{n,\mathrm{q}}\right)$ depends on $\hat{\boldsymbol{H}}_{\mathrm{q}}$ as can be seen in (63). Let us recall that the overall receiver in this case is $\frac{g_{\mathrm{q}}}{c_k\|\hat{\boldsymbol{h}}_k\|_2}$, which makes use of the estimated CDI and CMI. The resulting MMSE given the quantized estimated CDI $\hat{\boldsymbol{H}}_{\mathrm{q}}$, i.e. averaging over $\boldsymbol{n}$, $\boldsymbol{E}$, $\boldsymbol{E}_{\mathrm{q}}$, $\hat{\boldsymbol{B}}$, and $\boldsymbol{C}$, results in

$$\mathrm{MMSE}_{\mathrm{q}} = \sigma_s^2\mathrm{tr}\left(\left(\xi_{\mathrm{q}}\mathbf{1}_M - \kappa\hat{\boldsymbol{H}}_{\mathrm{q}}\hat{\boldsymbol{H}}_{\mathrm{q}}^{\mathrm{H}}\right)\left((1-\kappa)\hat{\boldsymbol{H}}_{\mathrm{q}}\hat{\boldsymbol{H}}_{\mathrm{q}}^{\mathrm{H}} + \xi_{\mathrm{q}}\mathbf{1}_M\right)^{-2} \times \left((1-2\kappa)\hat{\boldsymbol{H}}_{\mathrm{q}}\hat{\boldsymbol{H}}_{\mathrm{q}}^{\mathrm{H}} + \xi_{\mathrm{q}}\mathbf{1}_M\right)\right). \quad (69)$$

Computing an expression for the MMSE with quantized CDI at high SNR's is not tractable as in the previous cases. However, due to the quantization error we point out that $\mathrm{MMSE}_{\mathrm{q}}$ will saturate at high SNR. $\mathrm{MMSE}_{\mathrm{q}}$ approaches $\mathrm{MMSE}_{\mathrm{n}}$ only when $T_{\mathrm{DL}} \to \infty$ and $B \to \infty$. At high SNR, the quantization error $\frac{\mathrm{E}[\tan^2\theta_k]}{M-1}$ dominates the estimation error $\frac{\mathrm{E}[\cos^{-2}\theta_k]}{(M-1)(1-\sigma_{e_{\mathrm{DL}}}^2)}\sigma_{e_{\mathrm{DL}}}^2$ in (67). Restricting the quantization error to be smaller than the estimation error in order to have a decreasing quantization plus estimation error with increasing SNR ξ^{-1}, we have that

$$\frac{\mathrm{E}\left[\tan^2\theta_k\right]}{M-1} < \frac{\mathrm{E}\left[\cos^{-2}\theta_k\right]}{(M-1)}\frac{\sigma_{e_{\mathrm{DL}}}^2}{(1-\sigma_{e_{\mathrm{DL}}}^2)},$$

which can be very well approximated as

$$\frac{\mathrm{E}\left[\sin^2\theta_k\right]}{\mathrm{E}\left[\cos^2\theta_k\right](M-1)} < \frac{1}{\mathrm{E}\left[\cos^2\theta_k\right](M-1)}\frac{M\,\xi}{K\,T_{\mathrm{DL}}}, \quad (70)$$

by using the mentioned approximations for $\mathrm{E}\left[\tan^2\theta_k\right]$ and $\mathrm{E}\left[\cos^{-2}\theta_k\right]$ and by noting that $\frac{\sigma_{e_{\mathrm{DL}}}^2}{(1-\sigma_{e_{\mathrm{DL}}}^2)} = \frac{M}{K\,\xi\,T_{\mathrm{DL}}}$ when plugging in (37). Using the approximation $\mathrm{E}\left[\sin^2\theta_k\right] \approx 2^{-\frac{B}{M-1}}$ from [4] in (70) to solve for B, we get that the scaling for B as a function of ξ^{-1} should be

$$B > (M-1)\left(\log_2\left(\xi^{-1}\right) + \log_2\left(\frac{K\,T_{\mathrm{DL}}}{M}\right)\right), \quad (71)$$

i.e. the number of feedback bits should scale with ξ^{-1} in dB. A similar result is derived in [4] in order to maintain a constant rate gap to the sum rate with zero-forcing and perfect CSI.

VI. Simulation Results

We now present some results for the transceiver design with the different types of transmit CSI. In Fig. 1 the average BER as a function of $\xi^{-1} = \frac{P_{\mathrm{DL}}}{\mathrm{tr}(\boldsymbol{R}_n)}$ in dB for each of the discussed types of transmit CSI is depicted. ξ^{-1} represents the average user SNR. We have assumed QPSK symbols, $M = 4$, $K = 4$, $T_{\mathrm{UL}} = 4, 12$ (uplink training length in the TDD case (estimated CSI)) and $T_{\mathrm{DL}} = 4, 12$ (downlink training length in the FDD case (quantized CDI)). For the TDD case we assume that the uplink power scales with the downlink power as $P_{\mathrm{UL}} = \frac{P_{\mathrm{DL}}}{M}$ and that we have the same noise variance in the uplink and downlink, i.e $\sigma_v^2 = \sigma_n^2$. In addition, the performance of the FDD case is shown for different values of B. The theoretical case with perfect CDI described by (28) is shown also as reference. One can observe a performance gap between the BER with perfect CSI and the BER with estimated CSI. As discussed for the MMSE, the estimation error decreases with increasing SNR, but the gap is still present at high SNR. However, as T_{UL} increases, the gap to the perfect CSI reduces as expected.

In addition, we observe how the BER saturates for the different values of B in the FDD case. In Fig. 2, we consider the same scenario as before but with the average MMSE per user as figure of merit and in addition for the FDD case we assume that the number of feedback bits scale with the SNR ξ^{-1} as $B = (M-1)\left(\log_2\left(\xi^{-1}\right)\right)$ based on (71), which with $M = 4$ means that $B = \xi_{\mathrm{dB}}^{-1}$. In this case, we can observe how the MMSE decreases with almost the same slope as the other curves. A similar behaviour is also observed for the average BER. However, we point out we have assumed error-free feedback and this result will change once feedback errors are considered. Nevertheless, assuming that the transmit

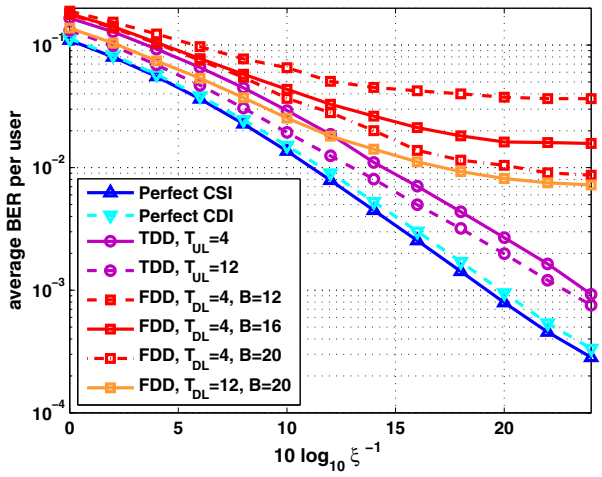

Fig. 1. BER vs. $10\log_{10}\xi^{-1}$ for different types of transmit CSI.

power of the users in the feedback link (uplink) scales with the downlink power, the effect of the feedback errors will be more pronounced at low to moderate SNR.

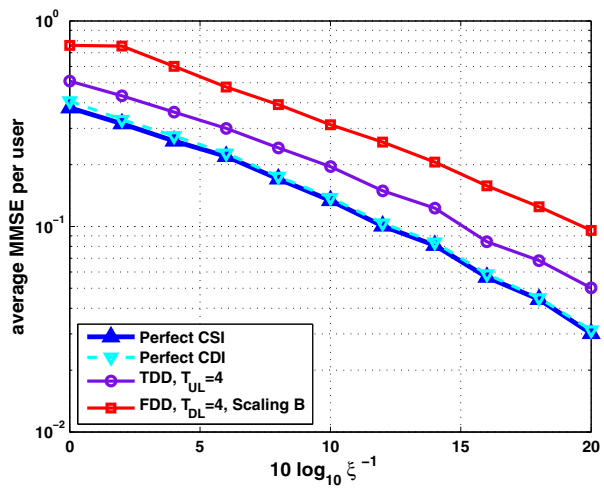

Fig. 2. BER vs. $10\log_{10}\xi^{-1}$ for different types of transmit CSI.

VII. Conclusion and Future Work

As can be observed from the plots and from the discussed analysis, imperfect transmit CSI in practical systems is still beneficial. Specifically for the FDD case, we have considered the feedback of a quantized version of the users' estimated CDI, which is used with the statistics of the users' CMI to compute a precoder at the BS. On the other hand, the scalar receiver at each user depends on the estimated CMI and on g_{q}. However, we have assumed the same g_{q} for all users. With QPSK, g_{q} is irrelevant, but for higher modulation scheme we will consider a scalar receiver which is only based on the local information availabe at each user.

In this work we have considered the case $K \leq M$. For $K > M$ we point out that in a TDD system the BS would have access to an estimate of all the users and could easily decide which users to served based on a given scheduler and the estimated CSI. Under our assumptions, the main drawback in this case is the increase of the training overhead in the uplink. We have assumed that $T_{\mathrm{UL}} \geq K$ in order to have orthogonal pilot sequences among the users such that the BS can estimate each user channel without interference from the other users. Nevertheless, this comes at the expense of uplink resources, a fact that has to be considered when optimizing T_{UL}. The tradeoff between the training length and the interference-contaminated estimated CSI is a subject of future work.

In a FDD system, dealing with the case $K > M$ is more elaborate. First we state that the performance with the transmit CSI in an FDD system is bounded by the performance with the theoretical case of perfect CDI. This upper bound is achieved as the training length and the number of feedback bits go to infinity. However, we have seen that the average performance with perfect CDI is not very far from the perfect CSI case, but we point out that this is under the assumption of users with the same average CMI, i.e. pathloss, and for the case $K \leq M$, i.e. no scheduling is involved. In the case of $K > M$, a scheduler at the BS could make use of the users' quantized-estimated CDI and CMI. We are also interested in analyzing the loss incurred when scheduling $K > M$ users just based on perfect CDI. It is clear, nonetheless, that in order to exploit multiuser diversity the BS would need also access to the users' CMI. What the users should feedback and how should it be done in an FDD system with $K > M$ is also a matter of future work. Finally, the optimization of multiuser systems considering the overhead due to feedback and training will also be pursued.

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] 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] 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, Feb. 2007.

[4] N. Jindal, "MIMO Broadcast Channels with Finite-Rate Feedback", *IEEE Transactions on Information Theory*, Vol. 52, No. 11, Nov. 2006.

[5] 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*, 2008.

[6] D. J. Ryan, I. B. Collings, I. V. L. Clarkson and R. W. Heath Jr., "Performance of Vector Perturbation Multiuser MIMO Systems with Limited Feedback", *IEEE Transactions on Communications*, Vol. 57, No. 9, pp. 2633-2644, Sep. 2009.

[7] M. Joham, M., W. Utschick, J. A. Nossek, "Linear Transmit Processing in MIMO Communications Systems", *IEEE Transactions on Signal Processing*, Vol. 53, No. 8, Aug. 2005, pp. 2700 - 2712.

[8] A.M. Tulino, S. Verdu, *Foundations and Trends™ in Communications and Information Theory*, Vol. 1, No. 1, pp. 1-182, 2004.

[9] 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.

[10] M. Castañeda, A. Mezghani and J. A. Nossek, "Bounds on the Capacity of MISO Channels with Different Types of Imperfect CSI", *ITG Workshop in Smart Antennas (WSA)*, Feb. 2009.