Optimal Training in Large TDD Multi-user Downlink Systems under Zero-forcing and Regularized Zero-forcing Precoding

Sebastian Wagner†, Romain Couillet‡, Mérouane Debbah†, and Dirk T. M. Slock†

Abstract—This paper considers a large multi-user time-division duplex (TDD) system, where the base station (BS) acquires channel state information via pilot signaling from the users. In the downlink the BS employs zero-forcing (ZF) and regularized zero-forcing (RZF) precoding. We derive the optimal sum rate maximizing amount of channel training using sum rate approximations from the large system analysis of MISO downlink channels under (R)ZF precoding. Moreover, in the regime of high signal-to-noise ratio (SNR), we derive approximate solutions of the optimal amount of training for both schemes that are of closed-form. By comparing the two schemes, we find that RZF requires less training than ZF, but the training interval of both schemes is equal for asymptotically high SNR. Furthermore, simulations are carried out which demonstrate the accuracy of our approximate solutions.

I. INTRODUCTION

It is well known that the capacity region of the multiple-input single-output (MISO) broadcast channel is achieved by dirty-paper coding (DPC) [1]. Although optimal, DPC has a high computational complexity. Therefore, attention has been drawn away from DPC toward more practical schemes such as zero-forcing (ZF) or regularized ZF (RZF) precoding. However, the achievable sum rate under all precoding strategies is heavily dependent on the channel state information available at the transmitter (CSIT). In practical systems, the channel is approximately constant over a finite amount of channel uses and thus, the CSIT is always imperfect. Consequently, it is important to study the impact of limited channel coherence interval on the achievable system throughput.

In this contribution, we focus on time-division duplex (TDD) multi-user systems, where uplink (UL) and downlink (DL) channels are assumed perfectly reciprocal. In a TDD system, CSIT is usually obtained from pilot signaling in the UL and is subsequently used to pre-process the data streams by applying ZF or RZF precoding. Since channel acquisition and data transmission are carried out within one channel coherence interval, there exists a non-trivial trade-off in the allocation of resources between channel training and data transmission.

Previous work has primarily focused on describing this trade-off under ZF precoding for an equal number of users $K$ and transmit antennas $M$ [2]–[6], using bounds on the sum-rate gap between perfect CSIT and imperfect CSIT. In this work we use a different approach. Instead of bounds on the sum rate for all $(K, M)$ we provide deterministic approximations that are asymptotically accurate as $(K, M)$ grow large at a constant rate $\beta \approx M/K$. These approximations are referred to as deterministic equivalents. More precisely, we make use of deterministic equivalents of the signal-to-interference plus noise ratio (SINR) under ZF precoding for $M > K$ and RZF precoding for $M \geq K$ [7] to characterize the optimal trade-off between channel training and data transmission. Moreover, we derive closed-form expressions for the sum rate maximizing amount of training in the high signal-to-noise ratio (SNR) regime.

The contributions of this paper are (i) the application of large system approximations to compute the optimal (sum rate maximizing) amount of channel training, (ii) the comparison of the optimal channel training under ZF precoding ($\beta > 1$) and RZF precoding ($\beta \geq 1$), and (iii) novel closed-form solutions of the optimal training for both schemes at high SNR.

Notation: In the following boldface lower-case and upper-case characters denote vectors and matrices, respectively. The operators $(\cdot)^H$, $\text{tr}(\cdot)$ and $E[\cdot]$ denote conjugate transpose, trace and expectation, respectively. The $N \times N$ identity matrix is $I_N$ and $\log(\cdot)$ is the logarithm to base $e$.

II. SYSTEM MODEL

Consider a MISO downlink channel where an $M$-antenna base station (BS) transmits to $K$ single-antenna mobile users. At a given time instant the signal $y_k$ received by user $k$th is

$$y_k = h_k^H x + n_k, \quad k = 1, 2, \ldots, K,$$

where $h_k^H \in \mathbb{C}^{1 \times M}$ is the channel of user $k$ with independent and identical distributed (i.i.d.) complex Gaussian entries of zero mean and unit variance, $x \in \mathbb{C}^{M \times 1}$ is the transmit vector and the $n_1, n_2, \ldots, n_K$ are i.i.d. complex Gaussian additive noise terms with zero mean and variance $\sigma^2$. The transmit vector $x$ is subject to an average power constraint

$$E[\|x\|^2] \leq P.$$  

Therefore the (downlink) SNR $\rho_d$ at each user is $\rho_d \triangleq P/\sigma^2$.

We assume a block-fading channel $H = [h_1, h_2, \ldots, h_K] \in \mathbb{C}^{M \times K}$, i.e. the channel $H$ remains constant over a block of
of each channel coefficient $h_{ij}$ ($i = 1, \ldots, K, j = 1, \ldots, M$). Due to the orthogonality property of the MMSE estimation [12], the estimates $\hat{h}_{ij}$ of $h_{ij}$ and the corresponding estimation errors $\tilde{h}_{ij}$ are i.i.d. complex Gaussian distributed and we can write

$$\hat{h}_{ij} = \sqrt{1 - \tau^2} h_{ij} + \tau q_{ij},$$

where both $h_{ij}$ and $q_{ij}$ have zero mean unit variance. The variance $\tau^2$ of the estimation error $\hat{h}_{ij}$ is given by [4]

$$\tau^2 = \frac{1}{1 + T_i \rho_{ul}},$$

where we define the uplink SNR $\rho_{ul}$ as $\rho_{ul} = P_{ul}/\sigma^2$.

B. Data Transmission Phase

We focus on equal power allocation among the users. The information symbols $s_k$ of the $k$th user are i.i.d. complex Gaussian $s_k \sim \mathcal{CN}(0,1)$. Since $T_i$ channel uses have already been consumed to train the BS about the user channels, there remains an interval of length $T - T_i$ for DL data transmission.

Prior to the transmission, the symbols $s_k$ are linearly precoded to form the transmit signal

$$x = \sum_{k=1}^{K} g_k s_k,$$

where $g_k \in \mathbb{C}^{M \times 1}$ is the beamforming (BF) vector of user $k$. The BF vectors are not of unit norm but set to fulfill the power constraint (2), i.e. denoting $G \triangleq [g_1, g_2, \ldots, g_K] \in \mathbb{C}^{M \times K}$, from (2) we obtain $\text{tr}(GG^H) \leq P$.

Under a sum power constraint, the optimal ZF precoder $G_{zf}$ [13] and RZF precoder $G_{rzf}$ take the form

$$G_{zf} = \xi \hat{H}^H (\hat{H} \hat{H}^H)^{-1},$$

$$G_{rzf} = \xi \left( \hat{H}^H \hat{H} + M \alpha I_M \right)^{-1} \hat{H}^H,$$

where the scaling factor $\xi$ is set to fulfill the power constraint $\text{tr}(GG^H) \leq P$ and the regularization scalar $\alpha > 0$ in (9) is scaled by $M$ to ensure that, as $(K, M)$ grow large, both $\text{tr}\hat{H}^H \hat{H}$ and $\text{tr}M \alpha I_M$ grow with the same order of magnitude.

IV. PERFORMANCE METRIC AND LARGE SYSTEM APPROXIMATION

In this section we introduce the ergodic sum rate as our performance measure and briefly review a large system approximation of the SINR of ZF and RZF precoding.

A. Ergodic Sum Rate

We measure the performance of the system in terms of ergodic sum rate $R_{sum}$, defined as

$$R_{sum} = \frac{T - T_i}{T} \sum_{k=1}^{K} E_H \log (1 + \gamma_k),$$

The assumption of equal noise power in UL and DL might not hold in practice, since the sources of noise are different for UL and DL and it is likely that they have different variances.
where the factor \((T - T_i)/T\) appears since data is transmitted in the reduced interval \(T - T_i\) and the SINR \(\gamma_k\) of user \(k\) reads
\[
\gamma_k = \frac{\left| h_{k,k}^T g_k \right|^2}{\sum_{j=1, j \neq k}^K \left| h_{k,j}^T g_j \right|^2 + \sigma^2}.
\]

Similarly, for RZF the approximated sum rate reads
\[
R_{\text{sum}}^\circ = K \left( 1 - \frac{T_i,\text{zf}}{T} \right) \log \left( 1 + \frac{T_i,\text{zf} \rho_{ul} (\beta - 1) + d(w)}{1 + T_i,\text{zf} \rho_{ul} + T_i,\text{zf} \rho_{ul} + d(w)} \right),
\]

\[
R_{\text{sum}}^{\circ,\text{zf}} = K \left( 1 - \frac{T_i,\text{zf}}{T} \right) \log \left( 1 + \frac{T_i,\text{zf} \rho_{ul} (\beta - 1) + d(w)}{1 + T_i,\text{zf} \rho_{ul} + d(w)} \right),
\]

\[
R_{\text{sum}}^{\circ,\text{rzf}} = K \left( 1 - \frac{T_i,\text{rzf}}{T} \right) \log \left( 1 + \frac{T_i,\text{rzf} \rho_{ul} (\beta - 1) + d(w)}{1 + T_i,\text{rzf} \rho_{ul} + d(w)} \right).
\]

\[
R_{\text{sum}}^{\circ,\text{rzf}} = K \left( 1 - \frac{T_i,\text{rzf}}{T} \right) \log \left( 1 + \frac{T_i,\text{rzf} \rho_{ul} (\beta - 1) + d(w)}{1 + T_i,\text{rzf} \rho_{ul} + d(w)} \right).
\]

\[
R_{\text{sum}}^{\circ,\text{rzf}} = K \left( 1 - \frac{T_i,\text{rzf}}{T} \right) \log \left( 1 + \frac{T_i,\text{rzf} \rho_{ul} (\beta - 1) + d(w)}{1 + T_i,\text{rzf} \rho_{ul} + d(w)} \right).
\]

\[
R_{\text{sum}}^{\circ,\text{rzf}} = K \left( 1 - \frac{T_i,\text{rzf}}{T} \right) \log \left( 1 + \frac{T_i,\text{rzf} \rho_{ul} (\beta - 1) + d(w)}{1 + T_i,\text{rzf} \rho_{ul} + d(w)} \right).
\]

\[
R_{\text{sum}}^{\circ,\text{rzf}} = K \left( 1 - \frac{T_i,\text{rzf}}{T} \right) \log \left( 1 + \frac{T_i,\text{rzf} \rho_{ul} (\beta - 1) + d(w)}{1 + T_i,\text{rzf} \rho_{ul} + d(w)} \right).
\]

\[
R_{\text{sum}}^{\circ,\text{rzf}} = K \left( 1 - \frac{T_i,\text{rzf}}{T} \right) \log \left( 1 + \frac{T_i,\text{rzf} \rho_{ul} (\beta - 1) + d(w)}{1 + T_i,\text{rzf} \rho_{ul} + d(w)} \right).
\]

\[
R_{\text{sum}}^{\circ,\text{rzf}} = K \left( 1 - \frac{T_i,\text{rzf}}{T} \right) \log \left( 1 + \frac{T_i,\text{rzf} \rho_{ul} (\beta - 1) + d(w)}{1 + T_i,\text{rzf} \rho_{ul} + d(w)} \right).
\]

\[
R_{\text{sum}}^{\circ,\text{rzf}} = K \left( 1 - \frac{T_i,\text{rzf}}{T} \right) \log \left( 1 + \frac{T_i,\text{rzf} \rho_{ul} (\beta - 1) + d(w)}{1 + T_i,\text{rzf} \rho_{ul} + d(w)} \right).
\]

\[
R_{\text{sum}}^{\circ,\text{rzf}} = K \left( 1 - \frac{T_i,\text{rzf}}{T} \right) \log \left( 1 + \frac{T_i,\text{rzf} \rho_{ul} (\beta - 1) + d(w)}{1 + T_i,\text{rzf} \rho_{ul} + d(w)} \right).
\]

\[
R_{\text{sum}}^{\circ,\text{rzf}} = K \left( 1 - \frac{T_i,\text{rzf}}{T} \right) \log \left( 1 + \frac{T_i,\text{rzf} \rho_{ul} (\beta - 1) + d(w)}{1 + T_i,\text{rzf} \rho_{ul} + d(w)} \right).
\]

\[
R_{\text{sum}}^{\circ,\text{rzf}} = K \left( 1 - \frac{T_i,\text{rzf}}{T} \right) \log \left( 1 + \frac{T_i,\text{rzf} \rho_{ul} (\beta - 1) + d(w)}{1 + T_i,\text{rzf} \rho_{ul} + d(w)} \right).
\]

\[
R_{\text{sum}}^{\circ,\text{rzf}} = K \left( 1 - \frac{T_i,\text{rzf}}{T} \right) \log \left( 1 + \frac{T_i,\text{rzf} \rho_{ul} (\beta - 1) + d(w)}{1 + T_i,\text{rzf} \rho_{ul} + d(w)} \right).
\]
Denoting $\psi \triangleq 1 + \frac{1}{\rho_{ul}} + T_{t,x}T_{1,c} + \rho_{ul}(\beta - 1)$, the derivatives take the form
\[
\frac{\partial R_{\text{sum}}^{\rho}}{\partial T_{t,x}} = -\frac{K}{T}(R_{\text{st}} - \Delta R_{\text{st}}^o) + \frac{K}{\psi^3 + (\beta - 1)(\rho_{ul} + 1)\psi},
\]
for $T_{t,x}$, and
\[
\frac{\partial R_{\text{sum}}^{\rho}}{\partial T_{t,\text{rxf}}} = -\frac{K}{T}(R_{\text{rxf}} - \Delta R_{\text{rxf}}^o)
\]
for $T_{t,\text{rxf}}$, respectively, where $w' = \partial w/\partial T_{t,\text{rxf}} = (1/\rho_{ul} + c)/(T_{t,\text{rxf}} + 1/\rho_{ul} + c)$ and $d' = \partial d/\partial T_{t,\text{rxf}} = [(\beta - 1)^2 w'/\rho_{ul} + w'/\rho_{ul} (1 + \beta)/d]$. In (29) and (30) the per-user rate-gap $\Delta R_{t,x}$ and $\Delta R_{t,\text{rxf}}$ can be neglected since $\Delta R_{t,x}^o \ll R_{t,x}$ and $\Delta R_{t,\text{rxf}}^o \ll R_{t,\text{rxf}}$, respectively. For $\rho_{ul} \to \infty$ and $c = \rho_{ul}/\rho_{ul}$ finite, solving (29) and (30) for $T_{t,x}$ and $T_{t,\text{rxf}}$, respectively, we obtain
\[
T_{t,x}^{\rho} = \max \left[ \frac{c}{2} \sqrt{1 + 2\frac{2T + c}{\rho_{ul}} - \frac{c}{2}}, K \right],
\]
and
\[
T_{t,\text{rxf}}^{\rho} = \left\{ \begin{array}{ll}
\max \left[ \frac{c}{2} \sqrt{1 + 2\frac{T + c}{\rho_{ul}} - \frac{c}{2}}, K \right] & \text{if } \beta = 1,
\max \left[ \frac{c}{2} \sqrt{1 + 2\frac{T + c}{\rho_{ul}} - \frac{c}{2}}, K \right] & \text{if } \beta > 1.
\end{array} \right.
\]
Thus, the optimal training intervals scale as $T_{t,x}^{\rho} \sim \sqrt{\psi}$ and $T_{t,\text{rxf}}^{\rho} \sim 1/\sqrt{\log(\rho_{ul})}$, under ZF precoding the same scaling has been reported in [2], [3], [6].

From (31) and (32) it is clear that as $\rho_{ul} \to \infty$ both $T_{t,x}^{\rho}, T_{t,\text{rxf}}^{\rho}$ tend to $K$, the minimum amount of training.

Moreover, for $\beta > 1$, $R_{t,x}^o \geq R_{t,\text{rxf}}^o$ with equality if $\rho_{ul} \to \infty$. Therefore, RZF requires less training than ZF, but the training interval of both schemes is equal for asymptotically high SNR. In case of full system loading ($\beta = 1$), RZF requires less training compared to the scenario where $\beta > 1$.

2) Low SNR Regime: For asymptotically low SNR $\rho_{dl}, \rho_{ul} \to 0$ with constant ratio $c = \rho_{dl}/\rho_{ul}$, applying a second order Taylor expansion around $\rho_{ul} = 0$, equations (20) and (21) take the form
\[
R_{\text{sum}}^{\rho} = K \left( 1 - \frac{T_{t,x}}{T} \right) T_{t,x}(\beta - 1) + \rho_{ul} + o(1),
\]
and
\[
R_{\text{sum}}^{\rho} = K \left( 1 - \frac{T_{t,\text{rxf}}}{T} \right) T_{t,\text{rxf}}(\beta - 1) + \rho_{ul} + o(1).
\]
Maximizing equations (33) and (36) with respect to $T_{t,x}$ and $T_{t,\text{rxf}}$, respectively, yields $T_{t,x}^{\rho} = T_{t,\text{rxf}}^{\rho} = T/2$. For ZF precoding, the limit has also been reported in [5].

B. Case 2: $\rho_{dl} \to \infty$ with finite $\rho_{ul}$

For ZF precoding and $\rho_{dl} \to \infty$, the sum rate (20) can be approximated as
\[
R_{\text{sum}}^{\rho} \approx K \left( 1 - \frac{T_{t,x}}{T} \right) \log(1 + T_{t,x}/T_{t,x}^R_{\text{ul}}(\beta - 1)).
\]
Setting the derivative of (35) with respect to $T_{t,x}$ to zero, yields
\[
\log(a/\omega(T_{t,x})) = \omega(T_{t,x}) - 1,
\]
where $a \triangleq \rho_{dl}(\beta - 1) + 1$ and $\omega(T_{t,x}) \triangleq T(T_{t,x}/T + T_{t,x}^R(\alpha - 1)/T)$. Equation (36) can be written as
\[
\omega(T_{t,x})e^{\omega(T_{t,x})} = ae.
\]
Notice that $\omega(T_{t,x}) = W(ae)$, where $W(z)$ is the Lambert W-function, defined as the unique solution to $z = W(z)e^{W(z)}$, $z \in \mathbb{R}$. Solving $\omega(T_{t,x}) = W(ae)$ for $T_{t,x}$ yields
\[
T_{t,x}^{\rho} = \frac{1}{\rho_{ul}(-1)} \left( \frac{a}{W(ae)} - 1 \right).
\]
For asymptotically low $\rho_{al}$ we obtain $\lim_{\rho_{al} \to 0} T_{t,x}^{\rho} = T/2$. For RZF, no accurate closed-form solution to $T_{t,\text{rxf}}$ can be obtained.

VI. NUMERICAL RESULTS

In this section we present simulation results that verify our theoretical derivations.

The accuracy of the SINR approximations (14) and (15) of ZF and RZF, respectively, has been established in [7] and needs no further justification. In Figure 1, we compare the approximated optimal training intervals $T_{t,x}^{\rho}, T_{t,\text{rxf}}^{\rho}$ to $T_{t,x}, T_{t,\text{rxf}}$ computed via exhaustive search and averaged over 1,000 independent channel realizations. The regularization term $\alpha$ is computed using the large system approximation $\alpha^{\rho}$ in (17). Figure 1 shows that the approximate solutions $T_{t,x}^{\rho}, T_{t,\text{rxf}}^{\rho}$ are starting to become very accurate for $K = 16$. Moreover, it can be observed that the approximations in (31) and (32) match very well. Further notice that for $\beta = 2$, ZF and RZF need approximately the same amount of training, as predicted by equations (31) and (32).

Figure 2 depicts the optimal relative amount of training $T_{t,x}^{\rho}/T$ for ZF and RZF precoding. We observe that $T_{t,x}^{\rho}/T$ decreases with increasing SNR as $O(1/\sqrt{\log(\rho_{al})})$. That is, for increasing SNR the estimation becomes more accurate and resources for channel training are reallocated to data.
transmission. Furthermore, $T_{t}^{\circ}/T$ saturates at $K/T$ due to the orthogonality constraint on the pilot sequences. Furthermore, as expected from (31) and (32), we observe that the optimal amount of training is less for RZF than for ZF precoding. Moreover, the relative amount of training $T_{t}^{\circ}/T$ for both ZF and RZF converges at low SNR to $1/2$ and at high SNR to the minimum amount of training $K$, as predicted by the theoretical analysis.

Figure 3 shows the ergodic sum rate under ZF precoding with fixed UL SNR $\rho_{ul} = 5$ dB and $T = 1000$. We observe (i) no significant difference in the performance of the schemes employing either optimal training $T_{t, zf}^{*}$ computed via exhaustive search or $T_{t, zf}$ obtained from a convex optimization of the large system approximation (20), (ii) a small performance loss at low and medium SNR of the (high-SNR) approximation of $T_{t, zf}^{*}$ in (38) and (iii) a significant performance loss if the minimum training interval $K$ is used for all SNR. We conclude that our approximation in (38) achieves very good performance and can therefore be utilized to compute $T_{t, zf}$ very efficiently.

VII. CONCLUSION

This paper analyzed the optimal trade-off between channel training and data transmission for a given channel coherence interval in a TDD multi-user system under ZF and RZF precoding in the downlink. Based on a large system approximation of the sum rate we derive novel closed-form approximations of the optimal training interval that maximizes the system sum rate. The results help the system designer to easily determine the optimal amount of pilot signaling as a function of the basic system parameters.

REFERENCES


