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[†]

[*]ST-ERICSSON, 06904 Sophia-Antipolis, France, Email: {sebastian.wagner, romain.couillet}@stericsson.com
[†]EURECOM, 06904 Sophia-Antipolis, France, Email: {sebastian.wagner, dirk.slock}@eurecom.fr
[‡]SUPÉLEC, 91192 Gif sur Yvette, France, Email: {romain.couillet, merouane.debbah}@supelec.fr

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 $\beta \triangleq M/K$. 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 $(\cdot)^{\mathsf{H}}$, $\mathrm{tr}(\cdot)$ and $E[\cdot]$ denote conjugate transpose, trace and expectation, respectively. The $N \times N$ identity matrix is $\mathbf{I}_N$ 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 kth is

$$y_k = \mathbf{h}_k^{\mathsf{H}} \mathbf{x} + n_k, \quad k = 1, 2, \ldots, K, \qquad (1)$$

where $\mathbf{h}_k^{\mathsf{H}} \in \mathbb{C}^{1 \times M}$ is the channel of user k with independent and identical distributed (i.i.d.) complex Gaussian entries of zero mean and unit variance, $\mathbf{x} \in \mathbb{C}^{M \times 1}$ 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 σ^2. The transmit vector $\mathbf{x}$ is subject to an average power constraint

$$E[\|\mathbf{x}\|^2] \leq P. \qquad (2)$$

Therefore the (downlink) SNR ρ_{dl} at each user is $\rho_{dl} \triangleq P/\sigma^2$.

We assume a block-fading channel $\mathbf{H}^{\mathsf{H}} = [\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_K] \in \mathbb{C}^{M \times K}$, i.e. the channel $\mathbf{H}$ remains constant over a block of

T channel uses before changing independently from block to block. Due to the finite coherence interval T, only $\hat{\mathbf{H}}$, an imperfect estimate of the true channel matrix $\mathbf{H}$, is available at the BS. We model $\hat{\mathbf{H}}$ as

$$\hat{\mathbf{H}} = \sqrt{1 - \tau^2}\mathbf{H} + \tau\mathbf{Q}, \tag{3}$$

where $\mathbf{Q} \in \mathbb{C}^{K \times M}$ is the matrix of channel estimation errors containing i.i.d. entries of zero mean and unit variance, and $\tau \in [0, 1]$. The parameter τ reflects the amount of distortion in the channel estimate $\hat{\mathbf{H}}$. Furthermore, we suppose that $\mathbf{H}$ and $\mathbf{Q}$ are mutually independent as well as independent of the information symbols s_k and the noise terms n_k. A similar model for the imperfect CSIT has been used in [8]–[10].

Following [4]–[6], we constrain the user terminals to transmit at a peak power of P_{ul} per channel use. The UL noise is assumed to be white complex Gaussian with variance σ^2.[1]

III. TRANSMISSION SCHEME

We consider a TDD system where UL and DL share the *same* channel at different times. Therefore, it is possible for the BS to estimate the channel from known pilot signaling from the users. The channel coherence interval T is divided into T_t channel uses for UL training and $T - T_t$ channel uses for coherent transmission in the DL. Note that in order to coherently decode the information symbols, the users need to know their effective (precoded) channels. This is usually accomplished by a common training phase in the DL prior to the data transmission. As shown in [11], a minimal amount of training (at most one pilot symbol) is sufficient when data and pilots are processed jointly. Therefore, we assume that the users have perfect knowledge of their effective channels and we neglect the overhead associated with common training.

In a TDD system, the imperfections in the CSIT are caused by (i) channel estimation errors, (ii) imperfect channel reciprocity due to different hardware in the transmitter and receiver and (iii) the channel coherence interval T. In what follows we assume that the channel is perfectly reciprocal and we study the joint impact of (i) and (iii).

A. Uplink Training Phase

In our setup, the distortion τ^2 in the CSIT is solely caused by an imperfect channel estimation at the BS and is identical for all entries of $\mathbf{H}$. To acquire CSIT, each user transmits $T_t \geq K$ *orthogonal* pilot symbols over the UL channel to the BS. Subsequently the BS estimates all K channels simultaneously. At the BS, the signal $\mathbf{r}_k$ received from user k is given by

$$\mathbf{r}_k = \sqrt{T_t P_{ul}}\mathbf{h}_k + \mathbf{n}_k, \tag{4}$$

where we assumed perfect reciprocity of UL and DL channels. That is, the UL and DL channel coefficients are equal and the UL noise terms $\mathbf{n}_k = [n_1, n_2, \ldots, n_M]^{\mathsf{T}}$ are statistically equivalent to their respective DL analog. Subsequently the BS performs a minimum mean square error (MMSE) estimation

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

$$\hat{h}_{ij} = \sqrt{1 - \tau^2}h_{ij} + \tau q_{ij}, \tag{5}$$

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

$$\tau^2 = \frac{1}{1 + T_t \rho_{ul}}, \tag{6}$$

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

B. Data Transmission Phase

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

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

$$\mathbf{x} = \sum_{k=1}^{K} \mathbf{g}_k s_k, \tag{7}$$

where $\mathbf{g}_k \in \mathbb{C}^{M \times 1}$ 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 $\mathbf{G} \triangleq [\mathbf{g}_1, \mathbf{g}_2, \ldots, \mathbf{g}_K] \in \mathbb{C}^{M \times K}$, from (2) we obtain $\mathrm{tr}(\mathbf{G}\mathbf{G}^{\mathsf{H}}) \leq P$.

Under a sum power constraint, the optimal ZF precoder $\mathbf{G}_{\mathrm{zf}}$ [13] and RZF precoder $\mathbf{G}_{\mathrm{rzf}}$ take the form

$$\mathbf{G}_{\mathrm{zf}} = \xi\hat{\mathbf{H}}^{\mathsf{H}}\left(\hat{\mathbf{H}}\hat{\mathbf{H}}^{\mathsf{H}}\right)^{-1}, \tag{8}$$

$$\mathbf{G}_{\mathrm{rzf}} = \xi\left(\hat{\mathbf{H}}^{\mathsf{H}}\hat{\mathbf{H}} + M\alpha\mathbf{I}_M\right)^{-1}\hat{\mathbf{H}}^{\mathsf{H}}, \tag{9}$$

where the scaling factor ξ is set to fulfill the power constraint $\mathrm{tr}(\mathbf{G}\mathbf{G}^{\mathsf{H}}) \leq P$ and the regularization scalar $\alpha > 0$ in (9) is scaled by M to ensure that, as (K, M) grow large, both $\mathrm{tr}\hat{\mathbf{H}}^{\mathsf{H}}\hat{\mathbf{H}}$ and $\mathrm{tr}M\alpha\mathbf{I}_M$ 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

$$R_{\mathrm{sum}} = \frac{T - T_t}{T}\sum_{k=1}^{K} E_{\mathbf{H}}\log\left(1 + \gamma_k\right), \tag{10}$$

---

[1]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 $(T - T_t)/T$ appears since data is transmitted in the reduced interval $T - T_t$ and the SINR γ_k of user k reads

$$\gamma_k = \frac{|\mathbf{h}_k^\mathsf{H}\mathbf{g}_k|^2}{\sum_{j=1, j\neq k}^{K} |\mathbf{h}_k^\mathsf{H}\mathbf{g}_j|^2 + \sigma^2}. \tag{11}$$

B. Large System Approximation of the SINR

To gain valuable insight into the system behavior, several bounds of the sum rate (10) have been proposed in e.g. [4]–[6]. Here we utilize an approximation of the sum rate derived for large (K, M) but bounded ratio $\beta = M/K$ [7].

We define a *deterministic equivalent* of the SINR γ_k as any γ_k°, such that

$$\gamma_k - \gamma_k^\circ \overset{M\to\infty}{\longrightarrow} 0, \tag{12}$$

almost surely. Thus, γ_k° is an approximation of γ_k for all SNR and its accuracy increases as (K, M) grow large. We use γ_k° to approximate the sum rate as

$$R_{\text{sum}}^\circ = \frac{T - T_t}{T} \sum_{k=1}^{K} \log\left(1 + \gamma_k^\circ\right). \tag{13}$$

In [7], we present numerical results showing that R_{sum}° is an accurate approximation of R_{sum} even for not so large (K, M). Now, we briefly review the results for γ_k° under ZF and RZF precoding.

1) Zero-forcing Precoding: In [7, Corollary 2], for $\beta > 1$, the SINR γ_{zf} of ZF precoding can be approximated by

$$\gamma_{\text{zf}}^\circ = \frac{1 - \tau^2}{\tau^2 + \frac{1}{\rho_{dl}}}(\beta - 1). \tag{14}$$

2) Regularized Zero-forcing Precoding: In [7, Corollary 1], for $\beta \geq 1$, the SINR γ_{rzf} of RZF precoding can be approximated by

$$\gamma_{\text{rzf}}^\circ = \frac{\beta(1 - \alpha^{\star\circ}) - 1 + d(\alpha^{\star\circ}, \beta)}{2\alpha^{\star\circ}\beta}, \quad \text{with} \tag{15}$$

$$d(\alpha^{\star\circ}, \beta) = \sqrt{\beta^2(\alpha^{\star\circ})^2 + 2\alpha^{\star\circ}\beta(1 + \beta) + (1 - \beta)^2}, \tag{16}$$

$$\alpha^{\star\circ} = \left(\frac{1 + \tau^2\rho_{dl}}{1 - \tau^2}\right)\frac{1}{\beta\rho_{dl}}. \tag{17}$$

Notice that both γ_{zf}° and $\gamma_{\text{rzf}}^\circ$ are independent of the particular user k.

V. Optimization of Channel Training

The approximation (13) is very accurate and can be used to optimize the training length T_t. By substituting γ_{zf}° into (13), the approximated sum rate $R_{\text{sum}}^{\circ,\text{zf}}$ of ZF precoding takes the form

$$R_{\text{sum}}^{\circ,\text{zf}} = K\left(1 - \frac{T_{t,\text{zf}}}{T}\right)\log\left(1 + \frac{1 - \tau^2}{\tau^2 + \frac{1}{\rho_{dl}}}(\beta - 1)\right). \tag{18}$$

Similarly, for RZF the approximated sum rate $R_{\text{sum}}^{\circ,\text{rzf}}$ is

$$R_{\text{sum}}^{\circ,\text{rzf}} = K\left(1 - \frac{T_{t,\text{rzf}}}{T}\right)\log\left(1 + \gamma_{\text{rzf}}^\circ\right), \tag{19}$$

where $\gamma_{\text{rzf}}^\circ$ is given in (15).

The distortion τ^2 in the CSIT is solely caused by an imperfect channel estimation at the BS and is identical for all entries of $\mathbf{H}$. Substituting (6) into (18) and (19), we obtain

$$R_{\text{sum}}^{\circ,\text{zf}} = K\left(1 - \frac{T_{t,\text{zf}}}{T}\right)\log\left(1 + \frac{T_{t,\text{zf}}\rho_{ul}(\beta - 1)}{1 + T_{t,\text{zf}}\frac{\rho_{ul}}{\rho_{dl}} + \frac{1}{\rho_{dl}}}\right), \tag{20}$$

$$R_{\text{sum}}^{\circ,\text{rzf}} = K\left(1 - \frac{T_{t,\text{rzf}}}{T}\right)\log\left(\frac{1}{2} + \frac{1}{2}w\rho_{dl}(\beta - 1) + \frac{d(w)}{2}\right), \tag{21}$$

$$d(w) = \sqrt{(1 - \beta)^2 w^2 \rho_{dl}^2 + 2w\rho_{dl}(1 + \beta) + 1}, \tag{22}$$

$$w = \frac{T_{t,\text{rzf}}\rho_{ul}}{1 + T_{t,\text{rzf}}\rho_{ul} + \rho_{dl}}. \tag{23}$$

For $\beta > 1$ under ZF precoding and $\beta \geq 1$ for RZF precoding, it is easy to verify that the functions $R_{\text{sum}}^{\circ,\text{zf}}$ and $R_{\text{sum}}^{\circ,\text{rzf}}$ are strictly concave in $T_{t,\text{zf}}$ and $T_{t,\text{rzf}}$ in the interval $[K, T]$, respectively, where K is the minimum amount of training required, due to the orthogonality of the pilot sequences. Therefore we can apply standard convex optimization algorithms [14] to evaluate

$$T_{t,\text{zf}}^{\star\circ} = \underset{K \leq T_{t,\text{zf}} \leq T}{\arg\max} R_{\text{sum}}^{\circ,\text{zf}}, \tag{24}$$

$$T_{t,\text{rzf}}^{\star\circ} = \underset{K \leq T_{t,\text{rzf}} \leq T}{\arg\max} R_{\text{sum}}^{\circ,\text{rzf}}. \tag{25}$$

In the following we derive approximate closed-form solutions to (24) and (25) for high SNR. We distinguish two cases, (i) $\rho_{dl}, \rho_{ul} \to \infty$, $c \triangleq \rho_{dl}/\rho_{ul}$ finite and (ii) $\rho_{dl} \to \infty$, ρ_{ul} finite. In contrast to case (i), the system in case (ii) is interference-limited due to the finite transmit power of the users.

A. Case 1: $\rho_{dl}, \rho_{ul} \to \infty$ with finite ratio ρ_{dl}/ρ_{ul}

We derive approximate, but explicit, solutions for the optimal training intervals $T_{t,\text{zf}}^{\star\circ}, T_{t,\text{rzf}}^{\star\circ}$ in the high SNR regime and derive their limiting values for asymptotically low SNR.

1) High SNR Regime: The sum rate R_{sum}° can be written as a function of the per-user rate under perfect CSIT $\bar{R}^\circ$ and the per-user rate gap ΔR° as

$$R_{\text{sum}}^\circ = K\left(1 - \frac{T_t}{T}\right)\left[\bar{R}^\circ - \Delta R^\circ\right], \tag{26}$$

where for ZF and RZF we have $\bar{R}_{\text{zf}}^\circ = \log(1 + \rho_{dl}(\beta - 1))$ and $\bar{R}_{\text{rzf}}^\circ = \frac{1}{2} + \frac{1}{2}\rho_{dl}(\beta - 1) + \frac{d(1)}{2}$, respectively, and

$$\Delta R_{\text{zf}}^\circ = \log\left(\frac{(\beta - 1)(\rho_{dl} + 1)}{1 + \frac{1}{\rho_{dl}} + T_{t,\text{zf}}[\frac{1}{c} + \rho_{ul}(\beta - 1)]}\right) \tag{27}$$

$$\Delta R_{\text{rzf}}^\circ = \log\left(\frac{1 + \rho_{dl}(\beta - 1) + d(1)}{1 + w\rho_{dl}(\beta - 1) + d(w)}\right). \tag{28}$$

Denoting $\psi \triangleq 1 + \frac{1}{\rho_{dl}} + T_{t,\mathrm{zf}}[\frac{1}{c} + \rho_{ul}(\beta - 1)]$, the derivatives take the form

$$\frac{\partial R_{\mathrm{sum}}^{\circ,\mathrm{zf}}}{\partial T_{t,\mathrm{zf}}} = -\frac{K}{T}(\bar{R}_{\mathrm{zf}}^{\circ} - \Delta R_{\mathrm{zf}}^{\circ}) + K\left(1 - \frac{T_{t,\mathrm{zf}}}{T}\right)$$
$$\times \frac{(\beta - 1)(\rho_{dl} + 1)[\frac{1}{c} + \rho_{ul}(\beta - 1)]}{\psi^2 + (\beta - 1)(\rho_{dl} + 1)\psi}, \quad (29)$$

$$\frac{\partial R_{\mathrm{sum}}^{\circ,\mathrm{rzf}}}{\partial T_{t,\mathrm{rzf}}} = -\frac{K}{T}(\bar{R}_{\mathrm{rzf}}^{\circ} - \Delta R_{\mathrm{rzf}}^{\circ})$$
$$+ K\left(1 - \frac{T_{t,\mathrm{rzf}}}{T}\right)\frac{w'\rho_{dl}(\beta - 1) + d'}{1 + w\rho_{dl}(\beta - 1) + d}, \quad (30)$$

where $w' = \partial w/\partial T_{t,\mathrm{rzf}} = (1/\rho_{ul} + c)/(T_{t,\mathrm{rzf}} + 1/\rho_{ul} + c)^2$ and $d' = \partial d/\partial T_{t,\mathrm{rzf}} = [(\beta - 1)^2 ww'\rho_{dl}^2 + w'\rho_{dl}(1 + \beta) + 1]/d$. In (29) and (30) the per-user rate-gap $\Delta R_{\mathrm{zf}}^{\circ}$ and $\Delta R_{\mathrm{rzf}}^{\circ}$ can be neglected since $\Delta R_{\mathrm{zf}}^{\circ} \ll \bar{R}_{\mathrm{zf}}^{\circ}$ and $\Delta R_{\mathrm{rzf}}^{\circ} \ll \bar{R}_{\mathrm{rzf}}^{\circ}$, respectively. For $\rho_{dl}, \rho_{ul} \to \infty$ and $c = \rho_{dl}/\rho_{ul}$ finite, solving (29) and (30) for $T_{t,\mathrm{zf}}$ and $T_{t,\mathrm{rzf}}$, respectively, we obtain

$$T_{t,\mathrm{zf}}^{\star\circ} = \max\left[\frac{c}{2}\sqrt{1 + 2\frac{2T + c}{c\bar{R}_{\mathrm{zf}}^{\circ}}} - \frac{c}{2}, K\right], \quad (31)$$

$$T_{t,\mathrm{rzf}}^{\star\circ} = \begin{cases} \max\left[\frac{c}{2}\sqrt{1 + \frac{2T + c}{c\bar{R}_{\mathrm{rzf}}^{\circ}}} - \frac{c}{2}, K\right] & \text{if } \beta = 1, \\ \max\left[\frac{c}{2}\sqrt{1 + 2\frac{2T + c}{c\bar{R}_{\mathrm{rzf}}^{\circ}}} - \frac{c}{2}, K\right] & \text{if } \beta > 1. \end{cases} \quad (32)$$

Thus, the optimal training intervals scale as $T_{t,\mathrm{zf}}^{\star\circ}, T_{t,\mathrm{rzf}}^{\star\circ} \sim \sqrt{T}$ and $T_{t,\mathrm{zf}}^{\star\circ}, T_{t,\mathrm{rzf}}^{\star\circ} \sim 1/\sqrt{\log(\rho_{dl})}$. Under ZF precoding the same scaling has been reported in [2], [3], [6].

From (31) and (32) it is clear that as $\rho_{dl} \to \infty$ both $T_{t,\mathrm{zf}}^{\star\circ}, T_{t,\mathrm{rzf}}^{\star\circ}$ tend to K, the minimum amount of training.

Moreover, for $\beta > 1$, $\bar{R}_{\mathrm{rzf}}^{\circ} \geq \bar{R}_{\mathrm{zf}}^{\circ}$ with equality if $\rho_{dl} \to \infty$. 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 $\rho_{dl} = 0$, equations (20) and (21) take the form

$$R_{\mathrm{sum}}^{\circ,\mathrm{zf}} = K\left(1 - \frac{T_{t,\mathrm{zf}}}{T}\right)\frac{T_{t,\mathrm{zf}}(\beta - 1)}{c}\rho_{dl}^2 + o(1), \quad (33)$$

$$R_{\mathrm{sum}}^{\circ,\mathrm{rzf}} = K\left(1 - \frac{T_{t,\mathrm{rzf}}}{T}\right)\frac{T_{t,\mathrm{rzf}}\beta}{c}\rho_{dl}^2 + o(1). \quad (34)$$

Maximizing equations (33) and (36) with respect to $T_{t,\mathrm{zf}}$ and $T_{t,\mathrm{rzf}}$, respectively, yields $T_{t,\mathrm{zf}}^{\star\circ} = T_{t,\mathrm{rzf}}^{\star\circ} = T/2$. For ZF precoding, the limit has also been reported in [5].

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

For ZF precoding and $\rho_{dl} \to \infty$, the sum rate (20) can be approximated as

$$R_{\mathrm{sum}}^{\circ,\mathrm{zf}} \approx K\left(1 - \frac{T_{t,\mathrm{zf}}}{T}\right)\log\left(1 + T_{t,\mathrm{zf}}\rho_{ul}(\beta - 1)\right). \quad (35)$$

Setting the derivative of (35) with respect to $T_{t,\mathrm{zf}}$ to zero, yields

$$\log(a/\omega(T_{t,\mathrm{zf}})) = \omega(T_{t,\mathrm{zf}}) - 1, \quad (36)$$
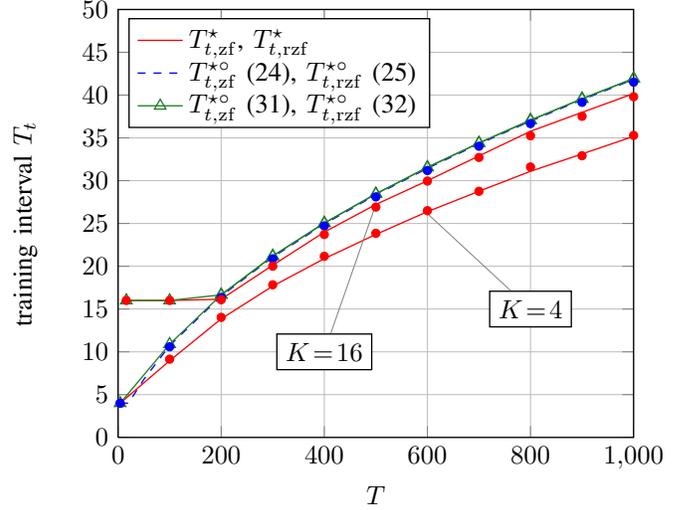
Fig. 1. ZF and RZF, optimal amount of training with $\beta = 2$, $\rho_{dl} = 20$ dB, $\rho_{ul} = 10$ dB, RZF is indicated by circle marks.

where $a \triangleq \rho_{ul}T(\beta - 1) + 1$ and $\omega(T_{t,\mathrm{zf}}) \triangleq (Ta)/[T + T_{t,\mathrm{zf}}(a - 1)]$. Equation (36) can be written as

$$\omega(T_{t,\mathrm{zf}})e^{\omega(T_{t,\mathrm{zf}})} = ae. \quad (37)$$

Notice that $\omega(T_{t,\mathrm{zf}}) = \mathcal{W}(ae)$, where $\mathcal{W}(z)$ is the Lambert W-function, defined as the unique solution to $z = \mathcal{W}(z)e^{\mathcal{W}(z)}$, $z \in \mathbb{R}$. Solving $\omega(T_{t,\mathrm{zf}}) = \mathcal{W}(ae)$ for $T_{t,\mathrm{zf}}$ yields

$$T_{t,\mathrm{zf}}^{\star\circ} = \frac{1}{\rho_{ul}(\beta - 1)}\left(\frac{a}{\mathcal{W}(ae)} - 1\right). \quad (38)$$

For asymptotically low ρ_{ul} we obtain $\lim_{\rho_{ul} \to 0} T_{t,\mathrm{zf}}^{\star\circ} = T/2$.

For RZF, no accurate closed-form solution to $T_{t,\mathrm{rzf}}$ 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 $T_{t,\mathrm{zf}}^{\star\circ}, T_{t,\mathrm{rzf}}^{\star\circ}$ to $T_{t,\mathrm{zf}}^{\star}, T_{t,\mathrm{rzf}}^{\star}$ computed via exhaustive search and averaged over 1,000 independent channel realizations. The regularization term α is computed using the large system approximation $\alpha^{\star\circ}$ in (17). Figure 1 shows that the approximate solutions $T_{t,\mathrm{zf}}^{\star\circ}, T_{t,\mathrm{rzf}}^{\star\circ}$ 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 $T_t^{\star\circ}/T$ for ZF and RZF precoding. We observe that $T_t^{\star\circ}/T$ decreases with increasing SNR as $O(1/\sqrt{\log(\rho_{dl})})$. That is, for increasing SNR the estimation becomes more accurate and resources for channel training are reallocated to data
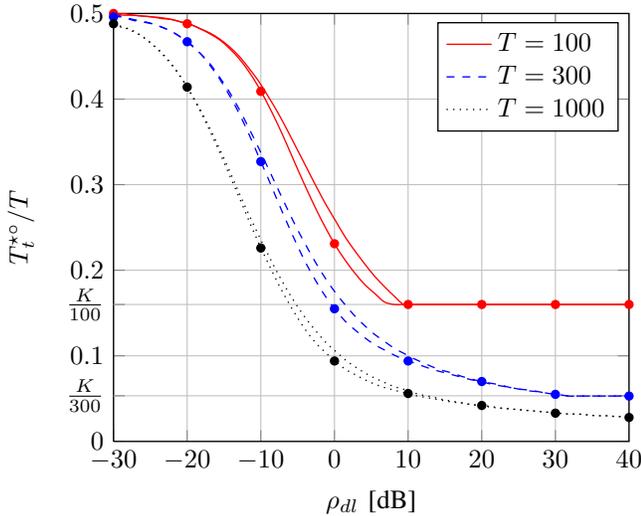
Fig. 2. ZF and RZF, optimal relative amount of training $T_t^{\star\circ}/T$ vs. ρ_{dl} with $M = 32$, $K = 16$, $\rho_{dl}/\rho_{ul} = 10$, RZF is indicated by circle marks.
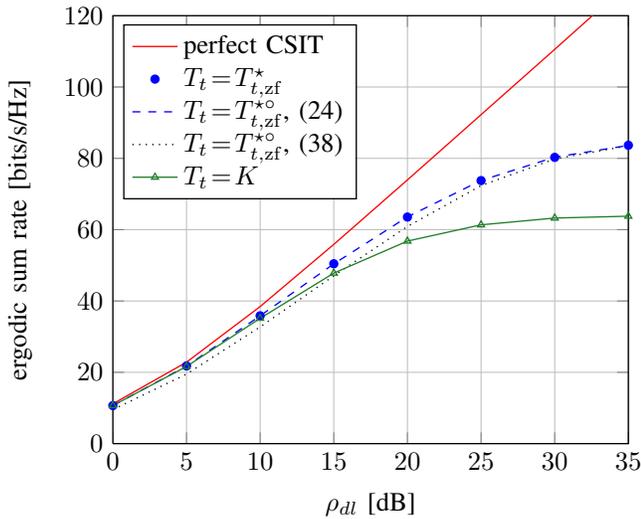
Fig. 3. ZF, ergodic sum rate vs. downlink SNR with $M = 32$, $\beta = 2$, $\rho_{ul} = 5$ dB and $T = 1000$.

transmission. Furthermore, $T_t^{\star\circ}/T$ 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 $T_t^{\star\circ}/T$ 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 for various training intervals. We observe (i) no significant difference in the performance of the schemes employing either optimal training $T_{t,\mathrm{zf}}^{\star}$ computed via exhaustive search or $T_{t,\mathrm{zf}}^{\star\circ}$ 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 $T_{t,\mathrm{zf}}^{\star\circ}$ 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 $T_{t,\mathrm{zf}}$ 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

[1] H. Weingarten, Y. Steinberg, and S. Shamai, "The Capacity Region of the Gaussian Multiple-Input Multiple-Output Broadcast Channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 3936–3964, Sep. 2006.

[2] M. Kobayashi, G. Caire, and N. Jindal, "How Much Training and Feedback are Needed in MIMO Broadcast Channels?" in *Proc. IEEE International Symposium on Information Theory (ISIT'08)*, Toronto, Canada, Jul. 2008, pp. 2663–2667.

[3] M. Kobayashi, N. Jindal, and G. Caire, "Optimized Training and Feedback for MIMO Downlink Channels," in *Proc. IEEE Information Theory Workshop on Networking and Information Theory (ITW'09)*, Volos, Greece, Jun. 2009, pp. 226–230.

[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] J. Jose, A. Ashikhmin, P. Whiting, and S. Vishwanath, "Linear Precoding for Multi-User Mulitple Antenna TDD Systems," *IEEE Trans. Inf. Theory*, submitted for publication. [Online]. Available: http://arxiv.org/abs/0812.0621

[6] U. Salim and D. Slock, "How much Feedback is Required for TDD Multi-Antenna Broadcast Channels with User Selection?" *EURASIP Journal on Advances in Signal Processing*, 2010, to appear.

[7] S. Wagner, R. Couillet, M. Debbah, and D. T. M. Slock, "Large System Analysis of Linear Precoding in MISO Broadcast Channels with Limited Feedback," *IEEE Trans. Inf. Theory*, submitted for publication. [Online]. Available: http://arxiv.org/abs/0906.3682

[8] A. D. Dabbagh and D. J. Love, "Multiple Antenna MMSE Based Downlink Precoding with Quantized Feedback or Channel Mismatch," *IEEE Trans. Commun.*, vol. 56, no. 11, pp. 1859–1868, Nov. 2008.

[9] T. Yoo and A. Goldsmith, "Capacity and Power Allocation for Fading MIMO Channels with Channel Estimation Error," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 2203–2214, May 2006.

[10] A. A. Hutter, E. de Carvalho, and J. M. Cioffi, "On the Impact of Channel Estimation for Multiple Antenna Diversity reception in Mobile OFDM Systems," in *Proc. IEEE Conference Record of the Asilomar Conference on Signals, Systems, and Computers*, vol. 2, Pacific Grove, CA, Nov. 2000, pp. 1820–1824.

[11] N. Jindal, A. Lozano, and T. L. Marzetta, "What is the value of joint processing of pilots and data in block-fading channels," in *Proc. IEEE International Symposium on Information Theory (ISIT'09)*, Seoul, South Korea, Jun. 2009, pp. 2189–2193.

[12] H. V. Poor, *An Introduction to Signal Detection and Estimation*, 2nd ed. Springer, 1994.

[13] A. Wiesel, Y. C. Eldar, and S. Shamai, "Zero-Forcing Precoding and Generalized Inverses," *IEEE Trans. Signal Process.*, vol. 56, no. 9, pp. 4409–4418, Sep. 2008.

[14] S. Boyd and L. Vandenberghe, *Convex Optimization*, 6th ed. New York, USA: Cambridge University Press, 2008.