Large System Analysis of Linear Precoding in Correlated MISO Broadcast Channels under Limited Feedback

Sebastian Wagner, Student Member, IEEE, Romain Couillet, Member, IEEE, Mérouane Debbah, Senior Member, IEEE, and Dirk T. M. Slock, Fellow, IEEE

Abstract—In this paper we study the sum rate performance of zero-forcing (ZF) and regularized ZF (RZF) precoding in large MISO broadcast channels under the assumptions of imperfect channel state information at the transmitter and per-user channel transmit correlation. Our analysis assumes that the number of transmit antennas $M$ and the number of single-antenna users $K$ are large and of the same order of magnitude. We derive deterministic approximations of the empirical signal-to-interference plus noise ratio (SINR) at the receivers, which are almost surely exact as $M, K \rightarrow \infty$. In course of this derivation, the assumed channel model requires the development of a novel deterministic equivalent of the empirical Stieltjes transform of the eigenvalue distribution of certain kinds of large dimensional random matrices. The deterministic SINR approximations enable us to solve various practical optimization problems. Under sum rate maximization, we derive (i) the optimal regularization term of the RZF precoder, (ii) for ZF the optimal number of active users and (iii) for ZF and RZF the optimal amount of feedback in large FDD/TDD multi-user systems. Numerical simulations suggest that the deterministic approximations are accurate even for small $M, K$.

Index Terms—Broadcast channel, random matrix theory, linear precoding, limited feedback, multi-user systems.

I. INTRODUCTION

The pioneering work in [1] and [2] revealed that the capacity of a point-to-point (single-user (SU)) multiple-input multiple-output (MIMO) channel can potentially increase linearly with the number of antennas. However, practical implementations quickly demonstrated that in most propagation environments the promised capacity gain of SU-MIMO is unachievable due to antenna correlation and line-of-sight components [3]. In a multi-user scenario, the inherent problems of SU-MIMO transmission can largely be overcome by exploiting multi-user (MU) diversity, i.e., sharing the spatial dimension not only between the antennas of a single receiver, but among multiple (non-cooperative) users. The underlying channel for MU-MIMO transmission is referred to as the MIMO broadcast channel (BC) or MU downlink channel. Although much more robust to channel correlation, the MIMO-BC suffers from inter-user interference at the receivers which can only be efficiently mitigated by appropriate (i.e., channel-aware) pre-processing at the transmitter.

It has been proved that dirty-paper coding (DPC) is a capacity achieving precoding strategy for the Gaussian MIMO-BC [4]–[8]. But the DPC precoder is non-linear and to this day too complex to be implemented efficiently in practical systems. However, it has been shown in [4], [9]–[11], that suboptimal linear precoders can achieve a large portion of the BC rate region while featuring low computational complexity. Thus, a lot of research has recently focused on linear precoding strategies.

A. Related Literature

In general, the rate maximizing linear precoder has no explicit form. Several iterative algorithms have been proposed in [12], [13], but no global convergence has been proved. Still, these iterative algorithms have a high computational complexity which motivates the use of further suboptimal linear transmit filters (i.e., precoders), by imposing more structure into the filter design. A straightforward technique is to precode by the inverse of the channel. This scheme is referred to as channel inversion or zero-forcing (ZF) [4].

In this contribution we focus on the multiple-input single-output (MISO) BC, where a central transmitter equipped with $M$ antennas communicates with $K$ single-antenna non-cooperative receivers. We assume $M \geq K$, i.e., no user scheduling is considered. Under this system setup, we carry out a large system analysis assuming that the number of transmit antennas $M$ as well as the number of users $K$ grow large while their ratio $\beta_M \equiv M/K$ remains bounded. In the following, the notation $M \rightarrow \infty$ implies that both $M$ and $K$ grow asymptotically large while $1 \leq \lim_{M \rightarrow \infty} \beta_M = \beta < \infty$.

A similar system has been studied in [14], [15]. In particular, it is shown in [14] that for $\beta > 1$, ZF achieves a large fraction of the linear (w.r.t. $K$) sum rate growth. The work in [9] extends the analysis in [14] to the case $\beta = 1$ and shows that the sum rate of ZF is constant in $M$ as $M \rightarrow \infty$. The authors in [9] counter this problem by introducing a regularization term in the inverse of the channel matrix. Under the assumption of large $M, K$ and for any rotationally-invariant channel distribution, [9] derives the regularization term that maximizes the signal-to-interference plus noise ratio (SINR). In this article, the regularized ZF (RZF) precoder of [9] is referred to as channel distortion-unaware RZF (RZF-CDU).
since its design assumes perfect channel state information at the transmitter (CSIT), although the available CSIT is erroneous or distorted. It has been observed that the RZF-CDA is very similar to the transmit filter derived under the minimum mean square error (MMSE) criterion [16] and both become identical in the large $M, K$ limit. Likewise, we will observe some similarities between ZF and MMSE filters when considering imperfect CSIT.

In the large system limit, i.e., $M \to \infty$, and for channels with independent and identically distributed (i.i.d.) entries, the cross correlations between the users’ channels, and therefore the users’ SINRs, are identical. It has been shown in [17] that for this symmetric case and equal noise variances, the SINR maximizing precoder is of closed form and coincides with the RZF precoder. Recently, the authors in [18] claimed that indeed the RZF precoder structure emerges as the optimal beamforming solution for $M \to \infty$. This asymptotic optimality motivates a detailed analysis of the RZF precoder for large system dimensions. The RZF-CDU has been analyzed for i.i.d. channels in a large single-cell setup in [19] and for a two-cell system in [18]. In [19] the authors derive the optimal regularization parameter using an asymptotic expression for the SINR. The expressions in [19] are a special case of the results derived in this paper. However, the approach and the tools for the derivations of the asymptotic SINR in [19] have been independently introduced earlier in the context of a different RZF precoder [20], where the regularization parameter is set to fulfill the total transmit power constraint. Moreover, the work in [21] extends the results in [19] by considering a common transmit correlation of the user channels. In particular, it is shown that the optimal regularization term derived in [9], [19] is independent of the transmit correlation.

Although it is legitimate that [9], [12], [13], [16] assume perfect CSIT to determine theoretically optimal performance, this assumption is untenable in practice. Also, it is a particularly strong assumption, since the performance of all precoding strategies is crucially depending on the CSIT. In practical systems, the transmitter has to acquire the channel state information (CSI) of the downlink channel by feedback signaling from the uplink. Since in practice the channel coherence time is finite, the information of the instantaneous channel state is inherently incomplete. For this reason, a lot of research has been carried out to understand the impact of imperfect CSIT on the system behavior, see [22] for a recent survey.

An information theoretic analysis of the impact of imperfect CSIT on the achievable rate of a ZF precoded MU-MISO downlink channel with $M = K$ has been carried out in [23]. Hereby, the author derives an upper bound on the ergodic per-user rate gap between perfect CSIT and imperfect CSIT under random vector quantization (RVQ) with $B$ feedback bits per user. Under finite-rate feedback, both [23] and [24] observe a sum-rate ceiling for high signal-to-noise ratios (SNR). In [23, Theorem 3] provides a formula for the minimum scaling of $B$ to maintain an average per-user rate gap of $\log_2 b$ bits/s/Hz and hence to achieve the full multiplexing gain of $K$. Although derived for ZF, the author claims that for all SNR, [23, Theorem 3] is more accurate for the RZF-CDU proposed in [9].

**B. Contributions of the Present Work**

This framework directly extends the analysis of ZF precoding in [14], [19], [21] by including per-user correlation and imperfect CSIT. More precisely, the vector channel $h_k \in \mathbb{C}^M$ from the transmitter to user $k$ is modeled as a non-line-of-sight fading channel with $E[|h_k|^2] = 0$ and $E[h_k h_k^H] = \Theta_k$, but only an estimate $\hat{h}_k$ is available at the transmitter. Furthermore, our approach allows for a unification and an extension of the RZF analysis in [9], [19], [21], [23]. In particular, we optimize the RZF-CDU proposed in [9], [19], [21] under imperfect CSIT, where the optimal regularization term is the solution to an implicit equation for common user correlation $\Theta_k = \Theta$ and has a explicit closed form for uncorrelated channels $\Theta_k = I_M$. This optimal RZF precoder is referred to as RZF channel distortion-aware (RZF-CDA). Moreover, for RZF and ZF, to maintain an instantaneous per-user rate offset of $\log_2 b$ bits/s/Hz as $M \to \infty$, almost surely, we derive the required scaling of the CSIT distortion (i.e. a measure of uncertainty in the CSIT) as a function of the SNR. Under RVQ, these results extend [23, Theorem 3].

Our main contributions can be summarized as follows:

- We propose deterministic equivalents for the SINR of ZF ($M > K$) and RZF ($M \geq K$) precoders, i.e., deterministic approximations of the SINR, which are independent of the individual channel realizations, and asymptotically (almost surely) exact as $M \to \infty$. Numerical results prove that these approximations are accurate even for finite $M$.

- Under imperfect CSIT and common correlation, we derive the sum rate maximizing regularization term.

For uncorrelated channels we obtain the following results:

- Under ZF precoding and imperfect CSIT, a closed-form approximate solution of the number of users $K$ that maximize the sum rate per transmit antenna for a fixed $M$.

- In large frequency-division duplex (FDD) systems, under RVQ, for $\beta = 1$ and high SNR $\rho$, to exactly maintain an instantaneous per-user rate gap of $\log_2 b$ bits/s/Hz, almost surely as $M \to \infty$, the number of feedback bits $B$ per user has to scale with

  - **RZF-CDA**: $B = (M-1) \log_2 \rho - (M-1) \log_2 (b^2-1)$

  - **RZF-CDU**: $B = (M-1) \log_2 \rho - (M-1) \log_2 2(b-1)$

  That is, the RZF-CDA precoder requires $(M-1) \log_2 b + 1/2$ bits less than RZF-CDU and ZF.

- In large time-division duplex (TDD) systems with channel coherence interval $T$, at high uplink SNR and downlink SNR $\rho_{dl}$, the sum rate maximizing amount of channel training scales as $\sqrt{T}$ and $1/\sqrt{\log(\rho_{dl})}$ for both RZF-CDA and ZF.

This paper focuses mainly on the applications of recent results in the field of large dimensional random matrix theory to problems in the MISO BC under linear precoding. Nevertheless, the per-user channel correlation model requires the extension of the results in [25]. This novel result is presented in this article and rigorously proved in the appendix.

The remainder of the paper is organized as follows. Section II presents the system model and channel model. In Section III,
we propose deterministic equivalents for the SINR of RZF and ZF precoding. In Section IV, we derive the sum rate maximizing regularization under RZF precoding. Section V studies the sum rate maximizing number of user for ZF precoding. Section VI analyses the optimal amount of feedback in a large FDD system. In Section VII, we study a large TDD system and derive the optimal amount of uplink channel training. Finally, in Section VIII, we summarize our results and conclude the paper.

The majority of the proofs is presented in the appendix. In these proofs, we apply several lemmas collected in Appendix VIII.

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 denoted $I_N$. $\log(\cdot)$ is the natural logarithm and $\Im(z)$ is the imaginary part of $z \in \mathbb{C}$. $|X|$ and $\lambda_{\min}(X)$ are the spectral radius and the minimum eigenvalue of $X$, respectively. The imaginary unit is denoted $i$. The sets $\mathbb{R}^+$ and $\mathbb{C}^+$ are defined as $\{x : x > 0\}$ and $\{x = r + iv : r \in \mathbb{R}, v > 0\}$.

II. SYSTEM MODEL

This section describes the transmission model as well as the underlying channel model.

A. Transmission Model

Consider the MISO broadcast channel composed of a central transmitter equipped with $M$ antennas and of $K$ single-antenna receivers. Assume narrow-band communication. The signal $y_k$ received by user $k$ at any time instant reads

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

where $h_k \in \mathbb{C}^M$ is the random channel from the transmitter to user $k$, $x \in \mathbb{C}^M$ is the transmit vector and the $n_k$ are independent circularly complex Gaussian noise terms with zero mean and variance $\sigma^2$.

The transmit vector $x$ is a linear combination of the independent user symbols $s_k \sim \mathcal{CN}(0, 1)$ and can be written as

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

where $g_k \in \mathbb{C}^M$ and $p_k \geq 0$ are the beamforming (BF) vector and the signal power of user $k$, respectively. The BF vectors are normalized to satisfy the average power constraint

$$E[\|x\|^2] = \text{tr}(PG^H G) \leq P,$$

where $G = [g_1, g_2, \ldots, g_K] \in \mathbb{C}^{M \times K}$, $P = \text{diag}(p_1, \ldots, p_K)$ and $P$ is the total available transmit power.

Denote $\rho \triangleq P/\sigma^2$ the signal-to-noise ratio (SNR). Under the assumption of Gaussian signalling and single-user decoding with perfect channel state information at the receivers, the SINR $\gamma_k$ of user $k$ takes the form

$$\gamma_k = \frac{p_k |h_k^H g_k|^2}{\sum_{j=1, j \neq k}^K p_j |h_j^H g_j|^2 + \sigma^2}.$$

The ergodic sum rate is defined as

$$R_{\text{sum}} = \sum_{k=1}^K E[\log(1 + \gamma_k)],$$

where the expectation is taken over the random channels $h_k$.

B. Channel Model

Each user channel $h_k$ is modeled as

$$h_k = \sqrt{M} \Theta_k^{1/2} z_k,$$

where $\Theta_k$ is the channel correlation matrix of user $k$ and $z_k$ has independent and identical distributed (i.i.d.) complex entries of zero mean and variance $1/M$. The channel transmit correlation matrices $\Theta_k$ are assumed to be slowly varying compared to the channel coherence time and thus are supposed to be perfectly known to the transmitter, whereas receiver $k$ has only knowledge about $\Theta_k$. Moreover, only an imperfect estimate $\hat{\Theta}_k$ of the true channel $\Theta_k$ is available at the transmitter which is modeled as $[26]–[29]$

$$\hat{\Theta}_k = \sqrt{M} \Theta_k^{1/2} \left(1 - \tau_k^2 z_k + \tau_k q_k\right) \triangleq \sqrt{M} \Theta_k^{1/2} z_k,$$

where $q_k$ has i.i.d. entries of zero mean and variance $1/M$ independent of $z_k$ and $n_k$. The parameter $\tau_k \in [0, 1]$ reflects the amount of uncertainty in the channel estimate $\hat{\Theta}_k$.

The per-user channel correlation model is very general and encompasses various propagation environments. For instance, all channel coefficients $h_{k,i}$ may have different variances $\sigma_{k,i}^2$ resulting from different attenuation of the signal while traveling to the receivers. This so called variance profile of the vector channel is obtained by setting $\Theta_k = \text{diag}(\sigma_{k,1}^2, \sigma_{k,2}^2, \ldots, \sigma_{k,M}^2)$, see [30]. Another possible scenario consists of an environment where all user channels experience identical transmit correlation $\Theta$, but where the users are heterogeneously scattered around the transmitter and hence experience different channel gains $d_k$. Such a setup can be modeled with $\Theta_k = d_k \Theta$. From a mathematical point of view, a homogeneous system with common user channel correlation $\Theta_k = \Theta$ is very attractive. In this case, the user channels are statistically equivalent and the deterministic SINR approximations can be computed by solving a single implicit equation instead of multiple systems of coupled implicit equations. A further simplification occurs when the channels are uncorrelated $\Theta_k = I_M$, in which case the approximated SINRs are given explicitly.

III. A DETERMINISTIC EQUIVALENT OF THE SINR

This section introduces deterministic approximations of the SINR under RZF and ZF precoding for various assumptions on the transmit correlation matrices $\Theta_k$ in form of theorems and corollaries. These results will be used in Sections IV-VII to solve practical optimization problems.

The following theorem extends the results in [25], [31] by assuming a correlated variance profile. This theorem is required for the channel model in (4) and forms the mathematical basis of the subsequent large system analysis of the MISO BC under RZF and ZF precoding.
Theorem 1: Let $B_N = X_N^H X_N + S_N$ with $S_N \in \mathbb{C}^{N \times N}$ Hermitian nonnegative definite and $X_N \in \mathbb{C}^{n \times N}$ random. The $i$th column of $X_N^H$ is $x_i = \Psi_j y_i$, where the entries of $y_i \in \mathbb{C}^n$ are i.i.d. of zero mean, variance $1/N$ and have eighth order moment of order $O(1/N^4)$. The matrices $\Psi_j \in \mathbb{C}^{N \times n}$ are deterministic. Furthermore, let $\Theta_i = \Psi_j \Psi_j^H \in \mathbb{C}^{N \times N}$ and define $Q_N \in \mathbb{C}^{N \times N}$ deterministic. Both $\Theta_i$ and $Q_N$ are assumed to have uniformly bounded spectral norm (with respect to $N$). Define

$$m_{B_N, Q_N}(z) \triangleq \frac{1}{N} \text{tr} Q_N (B_N - z I_N)^{-1}. \quad (6)$$

Then, for $z \in \mathbb{C} \setminus \mathbb{R}^+$, as $n, N$ grow large with ratios $\beta_{N,i} \triangleq N/r_i$ and $\beta_{N} \triangleq N/N$ such that $0 < \lim \inf_{N} \beta_{N} \leq \lim \sup_{N} \beta_{N} < \infty$ and $0 < \lim \inf_{N} \beta_{N,i} \leq \lim \sup_{N} \beta_{N,i} < \infty$, we have that

$$m_{B_N, Q_N}(z) - m_{B_N, Q_N} \xrightarrow{N \to \infty} 0, \quad (7)$$

almost surely, with $m_{B_N, Q_N}(z)$ given by

$$m_{B_N, Q_N}(z) = \frac{1}{N} \text{tr} Q_N \left( \frac{1}{N} \sum_{j=1}^{n} \frac{\Theta_j}{1 + e_{N,j}(z)} + S_N - z I_N \right)^{-1} \quad (8)$$

where $e_{N,1}(z), \ldots, e_{N,n}(z)$ form the unique solution of

$$e_{N,i}(z) = \frac{1}{N} \text{tr} \Theta_i \left( \frac{1}{N} \sum_{j=1}^{n} \frac{\Theta_j}{1 + e_{N,j}(z)} + S_N - z I_N \right)^{-1} \quad (9)$$

which is the Stieltjes transform of a nonnegative finite measure on $\mathbb{R}^+$. Moreover, for $z < 0$, the $e_{N,1}(z), \ldots, e_{N,n}(z)$ are the unique nonnegative solutions to (9). Note that (9) forms a system of $n$ coupled equations, from which (8) is given explicitly.

Proof: The proof of Theorem 1 is given in Appendix I. ■

Proposition 1 (Convergence of the Fixed Point Algorithm): Let $z \in \mathbb{C} \setminus \mathbb{R}^+$ and $\{e_{N,i}^{(k)}(z)\} (k \geq 0)$ be the sequence defined by $e_{N,i}^{(0)}(z) = -1/z$ and

$$e_{N,i}^{(k)}(z) = \frac{1}{N} \text{tr} \Theta_i \left( \frac{1}{N} \sum_{j=1}^{n} \frac{\Theta_j}{1 + e_{N,j}^{(k-1)}(z)} + S_N - z I_N \right)^{-1} \quad (10)$$

for $k \geq 0$. Then, $\lim_{k \to \infty} e_{N,i}^{(k)}(z) = e_{N,i}(z)$ defined in (9) for $i \in \{1, 2, \ldots, n\}$. ■

Proof: The proof of Proposition 1 is given in Appendix I-B and I-C.

In the following, Theorem 1 will be applied to derive a deterministic equivalent of the SINR under RZF and ZF precoding.

A. Regularized Zero-forcing Precoding

Consider the RZF precoding matrix

$$G_{\text{rZF}} = \xi \left( \tilde{H}^H \tilde{H} + M \alpha I_M \right)^{-1} \tilde{H}^H, \quad (11)$$

where $\tilde{H} \triangleq [\hat{h}_1, \hat{h}_2, \ldots, \hat{h}_K]^H \in \mathbb{C}^{K \times M}$ is the channel estimate available at the transmitter, $\xi$ is a normalization scalar to fulfill the power constraint (1) and $\alpha > 0$ is the regularization term. Here, $\alpha$ is scaled by $M$ to ensure that $\alpha$ itself converges to a constant, as $M \to \infty$.

From the total power constraint (1), we obtain $\xi^2$ as

$$\xi^2 = \frac{P}{\text{tr} PH \tilde{H}^H \tilde{H} + M \alpha M} \triangleq \frac{P}{\Psi},$$

Denote $\Psi \triangleq (\tilde{H}^H \tilde{H} + M \alpha I_M)^{-1}$, the SINR (2) of user $k$ under RZF precoding takes the form

$$\gamma_{k, \text{rZF}} = \frac{p_k |h_k| \Psi}{h_k^H \Psi h_k + \Psi}, \quad (12)$$

where $h_k = [h_1, h_k, h_{k+1}, \ldots, h_K]^H \in \mathbb{C}^{K \times 1}$ and $P \triangleq \text{diag}(p_1, p_2, \ldots, p_K)$.

The following theorem provides a deterministic approximation of $\gamma_{k, \text{rZF}}$ defined in (12), which is almost surely exact as $M \to \infty$.

Theorem 2: Let $\gamma_{k, \text{rZF}}$ be the SINR of user $k$ defined in (12). Then

$$\gamma_{k, \text{rZF}} \xrightarrow{M \to \infty} 0,$$

almost surely, where $\gamma_{k, \text{rZF}}$ is given by

$$\gamma_{k, \text{rZF}} = \frac{p_k (1 - \tau_K^2)(m_k^2)}{\Upsilon_k^2 (1 - \tau_K^2)(1 + (1 + m_k^2)^2)} + \frac{\Psi}{\Psi^2 (1 + m_k^2)^2}, \quad (13)$$

where

$$m_k^2 = \frac{1}{M} \text{tr} \Theta_k V,$$

$$\Psi = \frac{1}{M} \sum_{j=1}^{K} \frac{p_j e_j}{(1 + e_j)^2},$$

$$\Upsilon_k^2 = \frac{1}{M} \sum_{j=1}^{K} \frac{p_j e_j'}{(1 + e_j)^2}.$$

Denoting $\Psi \triangleq (F + \alpha I_M)^{-1}$, three systems of $K$ coupled equations have to be solved. First, the $e_1, \ldots, e_K$ form the unique positive solutions of

$$e_i = \frac{1}{M} \text{tr} \Theta_i V,$$

$$F = \frac{1}{M} \sum_{j=1}^{K} \frac{\Theta_j e_j}{(1 + e_j)^2}.$$

Secondly, the $e'_1, \ldots, e'_K$ form the unique positive solutions of

$$e'_i = \frac{1}{M} \text{tr} \Theta_i V^2 (F' + I_M),$$

$$F' = \frac{1}{M} \sum_{j=1}^{K} \frac{\Theta_j e'_j}{(1 + e_j)^2}.$$

Finally, the $e'_{i,k}, \ldots, e'_{K,k}$ are the unique positive solutions of

$$e'_{i,k} = \frac{1}{M} \text{tr} \Theta_i V^2 (F'_k + \Theta_k),$$

$$F'_k = \frac{1}{M} \sum_{j=1}^{K} \frac{\Theta_j e'_{j,k}}{(1 + e_j)^2}.$$
Proof: The proof of Theorem 2 can be found in Appendix III.

**Corollary 1:** Let $\Theta_k = \Theta$, then $\gamma_k^\circ$ takes the form

$$
\gamma_k^\circ = \frac{p_k}{P/K} \frac{m^\circ(1 - \tau_k^2)}{e_2(1 - \tau_k^2) + \tau_k^2(1 + m^\circ)^2 e_2 + \frac{1}{\beta}(1 + m^\circ)^2 e_1},
$$

where $m^\circ$ is the unique positive solution of

$$
m^\circ = \frac{1}{M} \text{tr}\Theta \left( \frac{\Theta/\beta}{1 + m^\circ} + \alpha I_M \right)^{-1}
$$

and $e_1$ and $e_2$ are given by

$$
e_1 = \frac{1}{M} \text{tr}\Theta \left( \alpha(1 + m^\circ)I_M + \frac{1}{\beta} \Theta \right)^{-2}
$$

$$
e_2 = \frac{1}{M} \text{tr}\Theta^2 \left( \alpha(1 + m^\circ)I_M + \frac{1}{\beta} \Theta \right)^{-2}.
$$

Proof: Substituting $\Theta_k = \Theta$ into Theorem 2, we have

$$
e_i = \frac{m_k^\circ}{m^\circ} \text{given in (14)}, e'_i = \frac{[\beta(1 + m^\circ)^2 e_1]/(\beta - e_2) \text{ and } e'_k = \frac{[\beta(1 + m^\circ)^2 e_2]/(\beta - e_2)\text{.}}$

Therefore, the terms $\Psi^\circ$ and $\Psi_k^\circ$ become $(P/K)e_1/(\beta - e_2)$ and $(P/K)e_2/(\beta - e_2)$, respectively. Furthermore, $m^\circ$ can be written as

$$
m^\circ = \frac{1}{M} \text{tr}\Theta \left( \frac{\Theta/\beta}{1 + m^\circ} + \alpha I_M \right)\left( \frac{\Theta/\beta}{1 + m^\circ} + \alpha I_M \right)^{-1} \left( \frac{\Theta/\beta}{1 + m^\circ} + \alpha I_M \right)^{-1}
$$

$$
= \alpha(1 + m^\circ)^2 e_1 + \frac{1}{\beta}(1 + m^\circ)^2 e_2.
$$

Substituting these terms into (13) yields (14) which completes the proof.

**B. Zero-forcing Precoding**

For $\alpha = 0$, the RZF precoding matrix in (11) reduces to the ZF precoding matrix $G_{zf}$ which reads

$$
G_{zf} = \xi \bar{H}^H \left( \bar{H} \bar{H}^H \right)^{-1},
$$

where $\xi$ is a scaling factor to fulfill the power constraint (1) and is given by

$$
\xi^2 = \frac{P}{\text{tr}P(\bar{H} \bar{H}^H)^{-1}} \equiv P / \Psi.
$$

Defining $\tilde{W} \triangleq \bar{H}^H(\bar{H} \bar{H}^H)^{-2}\bar{H}$, the SINR (2) of user $k$ under ZF precoding reads

$$
\gamma_{k,zf} = \frac{p_k |h_k^H \tilde{W} u_{k,i}|^2}{h_k^H \tilde{W} h_k^H K_1 P_{k,i} \bar{H} \bar{h}_k W_k + \frac{\Psi}{\rho}}.
$$

To obtain a deterministic equivalent of the SINR (17), we need to ensure that the minimum eigenvalue of $\bar{H} \bar{H}^H$ is bounded away from zero for all large $M$, almost surely. Therefore, the following assumption is required.

**Assumption 1:** Let $\beta > 1$, there exists $\varepsilon > 0$ such that, for asymptotically large $M$, we have $\lambda_{\min}(\bar{H} \bar{H}^H) > \varepsilon$ with probability one.

**Remark 1:** With $\beta > 1$, we have $\lambda_{\min}(Z Z^H) > \zeta$ for $\zeta > 0$, where $Z = [z_1, \ldots, z_K]^H$. Let $\Theta_k = \Theta$ with $\lambda_{\min}(\Theta) < \varepsilon$ for some $\varepsilon > 0$. Then it follows that $\lambda_{\min}(\bar{H} \bar{H}^H) > \zeta$ and Assumption 1 holds true.

**Theorem 3:** Let Assumption 1 hold true and let $\gamma_{k,zf}$ be the SINR of user $k$ under ZF precoding defined in (17). Then

$$
\gamma_{k,zf} - \gamma_k^\circ \xrightarrow{M \to \infty} 0,
$$

almost surely, where $\gamma_k^\circ$ is given by

$$
\gamma_k^\circ = \frac{p_k}{\tau_k^2 + \frac{\Psi^\circ}{\rho}}
$$

with

$$
\Psi^\circ = \frac{1}{M} \sum_{j=1}^K \frac{p_j}{e_j},
$$

$$
\Gamma_k^\circ = \frac{1}{M} \sum_{j=1}^K \lambda_{k,i} \left( \frac{1 - \tau_k^2}{\beta} \right) \left( 1 + \frac{m_k^\circ}{\beta} \right),
$$

$$
m_{\Gamma,k}^\circ = \frac{1}{M} \sum_{j=1}^K \frac{u_{k,i}^H \Theta_k u_{k,i}}{e_j},
$$

$$
m_{\Gamma,k}^\circ = \frac{1}{M} \text{tr}D_{k,i} \tilde{W}^{-2}[\tilde{W} + \tilde{F}],
$$

where $\lambda_{k,i}$ the $i$th eigenvalue of $\Theta_k$ with corresponding eigenvector $u_{k,i}$ and $D_{k,i} = \text{diag}(u_{k,i}^H \Theta_k u_{k,i}, \ldots)$.

Denote $\tilde{V} \triangleq (\tilde{F} + \tilde{I}_M)^{-1}$, three systems of coupled equations have to be solved. First, the $\bar{e}_1, \ldots, \bar{e}_K$ form the unique solution of

$$
\bar{e}_i = \frac{1}{M} \text{tr}\Theta_i \tilde{V}
$$

$$
\tilde{F} = \frac{1}{M} \sum_{j=1}^K \Theta_j e_j.
$$

Secondly, the $\bar{e}_1, \ldots, \bar{e}_M$ form the unique solution of

$$
\bar{e}_i = \frac{1}{M} \text{tr}D_{i,k} \tilde{F}_i^{-2}[\tilde{F}_i + \tilde{P}],
$$

$$
\tilde{F} = \frac{1}{M} \sum_{j=1}^M D_{i,j} \bar{e}_j.
$$

Finally, the $\bar{e}_1, \ldots, \bar{e}_M$ are the unique solution of

$$
\bar{e}_i = \frac{1}{M} \text{tr}D_{i,k} \tilde{F}_i^{-2}[\tilde{F}_i + \tilde{P}],
$$

$$
\tilde{F} = \frac{1}{M} \sum_{j=1}^M D_{i,j} \bar{e}_j.
$$

Proof: The proof of Theorem 3 can be found in Appendix IV.

**Corollary 2:** Let $\beta > 1$ and $\Theta_k = \Theta$, then $\gamma_k^\circ$ takes the form

$$
\gamma_k^\circ = \frac{p_k}{P/K} \frac{1 - \tau_k^2}{\tau_k^2 + \frac{\Psi^\circ}{\rho}}
$$

with

\[ \hat{\Psi}^o = \frac{1}{\beta e}, \quad \text{(20)} \]
\[ \hat{\Upsilon}^o = \frac{e_2/e^2}{\beta - e_2/e^2}, \quad \text{(21)} \]
\[ \hat{e}_2 = \text{Tr} \Theta^2 \left( \text{I}_M + \frac{1}{\epsilon \beta} \Theta \right)^{-2} \]

where \( \hat{e} \) is the unique solution of

\[ \hat{e} = \text{Tr} \Theta \left( \text{I}_M + \frac{1}{\epsilon \beta} \Theta \right)^{-1}. \quad \text{(22)} \]

Proof: For \( \Theta_k = \Theta \) we have \( D_{k,i} = \lambda_i \text{I}_K \), where \( \lambda_i \) is the \( i \)th eigenvalue of \( \Theta \). We obtain \( \hat{e}_i = \hat{e} \) in (22), \( \hat{e}_i = \hat{e} = 1/(\beta \epsilon), \) \( e^*_i = \hat{e} = (P/K)(1/e^2)/[\beta - e_2/e^2] \) and \( \hat{e}_i = \lambda_i e^2 / \beta \). Dividing both \( \hat{\Psi}^o \) and \( \hat{\Upsilon}^o \) by \( \hat{e} \) we obtain (20) and (21), respectively.

Corollary 3: Let \( \Theta_k = \text{I}_M \), then \( \gamma_{k,\text{zf}}^o \) takes the explicit form

\[ \gamma_{k,\text{zf}}^o = \frac{p_k}{P/K} \frac{1 - \tau_k^2}{\tau_k + 1} (\beta - 1). \quad \text{(23)} \]

Proof of Corollary 3: By substituting \( \Theta = \text{I}_M \) into (22), \( \hat{e} \) is explicitly given by \( \hat{e} = (\beta - 1)/\beta \). We further have \( \hat{\Psi}^o = \hat{\Upsilon}^o = 1/(\beta - 1) \).

C. Approximation of the System Sum Rate

The objective function is an approximation \( R_{\text{sum}}^o \) of the ergodic sum rate (3), where the instantaneous SINR \( \gamma_k \) is replaced by its large system approximation \( \gamma_{k,\text{zf}}^o \) in Theorem 2 or \( \gamma_{k,\text{zf}}^o \) in Theorem 3, i.e.,

\[ R_{\text{sum}}^o = \sum_{k=1}^K \log (1 + \gamma_{k,\text{zf}}^o). \quad \text{(24)} \]

Since the logarithm is a continuous function, it follows from \( \gamma_k \to \gamma_{k,\text{zf}}^o \) as \( M \to \infty \), almost surely that

\[ \frac{1}{K} (R_{\text{sum}} - R_{\text{sum}}^o) \to 0, \quad \text{(25)} \]

almost surely by applying the continuous mapping theorem [33]. Other possible objective functions include maximizing the minimum SINR or minimizing the power for given SINR constraints.

D. Power Optimization for \( \Theta_k = \Theta \)

From Corollaries 1 and 2, the approximated sum rate (24) for both RZF and ZF precoding can be written in the form

\[ R_{\text{sum}}^o = \sum_{k=1}^K \log \left( 1 + p_k \nu_k^o (\gamma_k) \right), \quad \text{(26)} \]

with \( \nu_k^o = \gamma_{k,\text{zf}}^o / p_k \), where the only dependence on user \( k \) stems from \( \gamma_k \). The user powers \( p_k^o \) that maximize (26), subject to \( \sum_{k=1}^K p_k \leq P, \ p_k \geq 0 \), are thus given by the classical water-filling solution [34]

\[ p_k^o = \left[ \mu - \frac{1}{\nu_k^o} \right]^+, \quad \text{(27)} \]

where \( [x]^+ \triangleq \max(0,x) \) and \( \mu \) is the water level chosen to satisfy \( \sum_{k=1}^K p_k = P \).

It could be argued that as \( \Theta \) is common to all users, the distortion of the CSIT \( \tau_k^2 \) should be identical as well. But suppose for instance a TDD system, where the CSIT is obtained from uplink training by the users. The training length of each user could be different, leading to different channel estimation errors at the transmitter. One could also imagine that in a FDD system the users apply channel quantization codebooks of different size depending on their channel quality and the available uplink resources. Thus, at the transmitter the channel quantization error would vary among the users. However, for simplicity, we assume a common distortion \( \tau_k^2 = \tau^2 \) for the optimization problems considered in Section VI and Section VII. For \( \tau_k^2 = \tau^2 \), the optimal user powers (27) are all equal, i.e., \( p_k^o = p^* = P/K \) and \( P^o \triangleq \text{diag}(p_1^o, \ldots, p_K^o) = \frac{P}{K} \text{I}_K \).

E. Numerical Results

We validate Theorem 2 and Theorem 3 by comparing the ergodic SINR and ergodic sum rate (3), obtained by Monte-Carlo (MC) simulations of i.i.d. Rayleigh block-fading channels, to their respective approximations, for finite system dimensions and equal power allocation \( \text{P} = \text{I}_K \).

The correlation \( \Theta_k \) of the \( k \)th user channel is modeled as in [35] by assuming a diffuse two-dimensional field of isotropic scatterers around the receivers. The waves impinge the receiver \( k \) uniformly at an azimuth angle \( \theta \) ranging from \( \theta_k,\text{min} \) to \( \theta_k,\text{max} \). Denoting \( d_{ij} \) the distance between transmit antenna \( i \) and \( j \), the correlation is modeled as

\[ [\Theta_k]_{ij} = \frac{1}{\theta_k,\text{max} - \theta_k,\text{min}} \int_{\theta_k,\text{min}}^{\theta_k,\text{max}} e^{i 2\pi d_{ij} \cos(\theta)} d\theta, \quad \text{(28)} \]

where \( \lambda \) denotes the signal wavelength. The users are assumed to be distributed uniformly around the transmitter at an angle \( \varphi_k = 2\pi k/K \) and as a simple example we choose \( \theta_k,\text{min} = -\pi \) and \( \theta_k,\text{max} = \varphi_k - \pi \). Note that in this example only odd values of \( K \) yield different \( \Theta_k \). The transmitter is equipped with a uniform linear array (ULA). To ensure that \( \| \Theta_k \| \) is bounded as \( M \) grows large, we assume that the distance between adjacent antennas is independent of \( M \), i.e., the length of the ULA increases with \( M \).

The simulation results presented in Figure 1 validate Theorem 2 by comparing the deterministic equivalent of the SINR of user 1 \( \gamma_1,\text{ref} \) to the SINR \( \gamma_1,\text{ref} \) averaged over 10,000 independent channel realizations. Note that in general, Theorem 2 holds for instantaneous SINR, i.e., the deterministic equivalent \( \gamma_1,\text{ref} \) is almost surely exact to the SINR obtained from a single channel realization as \( M \to \infty \). However, since we are interested in the average behavior of the system, we chose to compare against ergodic SINR or sum rate\(^1\). The notation \( \Theta_k \neq \text{I}_M \) indicates that \( \Theta_k \) is modeled according to (28) with \( d_{ij}/\lambda = 0.5 \). From Figure 1, we observe, as expected, that \( \gamma_1,\text{ref} \) becomes more accurate with increasing \( M \).

\(^1\) Also note that with the dominated convergence theorem [33]

\[ \int_{\Omega} \gamma_k,\text{ref}(\omega) dP(\omega) - \gamma_k^o,\text{ref} \to 0 \quad \text{almost surely as} \ M \to \infty. \]
is not decreasing at high SNR, because the CSIT \( \hat{H} \) is much better conditioned. The optimal regularization is discussed in Section V. Further observe that in Figure 2 the deterministic approximation becomes less accurate for high SNR. The reason is that in the derivation of the approximated SINR we apply Theorem 1 in \( z = -\alpha = -1/\rho \) and the bounds in Appendix I-A, Proposition 7 are thus proportional to the sum rate \([\text{bits/s/Hz}]\) 

\[
E[\gamma_{1,\text{ref}}] - \gamma_{1,\text{ref}}^0
\]

\( \Theta_k \neq I_M, \tau_k^2 = 0.1 \)

\( \Theta_k = I_M, \tau_k^2 = 0 \)

Fig. 1. RZF, \( E[\gamma_{1,\text{ref}}] - \gamma_{1,\text{ref}}^0 \) vs. \( M \) with \( M = K, \alpha = 0.1, \rho = 10 \) dB and \( P = I_K \).

Figures 2 and 3 compare the ergodic sum rate to the deterministic approximation (24) under RZF and ZF precoding, respectively. The error bars indicate the standard deviation of the MC results. It can be observed that the approximation lies approximately within one standard deviation of the MC simulations. From Figure 2, under imperfect CSIT (\( \tau_k^2 = 0.1 \)), the sum rate is decreasing for high SNR, because the regularization parameter \( \alpha \) does not account for \( \tau_k^2 \) and thus the matrix \( \hat{H}^\dagger \hat{H} + \alpha I_M \) in the RZF precoder becomes ill-conditioned. Figure 3 shows that, for \( M > K \), the sum rate is not decreasing at high SNR, because the CSIT \( \hat{H} \) is much better conditioned. The optimal regularization is discussed in Section V. Further observe that in Figure 2 the deterministic approximation becomes less accurate for high SNR. The reason is that in the derivation of the approximated SINR we apply Theorem 1 in \( z = -\alpha = -1/\rho \) and the bounds in Appendix I-A, Proposition 7 are thus proportional to the

SNR. Therefore, to increase the accuracy of the approximated SINR, larger dimensions are required in the high SNR regime.

We conclude that the approximation in Theorem 2 is accurate even for finite dimensions and can be applied to various optimization problems discussed in the sequel.

In Figure 4, we compare the ergodic sum rate performance with optimal power allocation \( P = P^\ast \) in (27) to equal power allocation \( P = I_K \) under RZF-CDU precoding. In the system with \( M = K = 5 \), the quality of the CSIT varies significantly among the users, \( \tau_k^2 = \{0.8, 0.1, 0.3, 0.05, 0.2\} \) and we observe a significantly gain over the whole SNR range when optimal power allocation is applied. In contrast, the CSIT distortion of the users’ channels in the setup where \( M = K = 3 \) does not differ considerably (\( \tau_k^2 \neq \{0.1, 0.15, 0.2\} \)) and we observe only a small gain at high SNR. The gain at high SNR is larger compared to low SNR, because the interference is dominating in the high SNR regime and the CSIT distortion \( \tau_k^2 \) becomes the discriminating factor between the users’ SINRs which impacts the optimal power distribution. We conclude that the optimal power allocation proposed in (27) achieves significant performance gains, especially at high SNR, when the quality of the available CSIT varies considerably among the users’ channels.

IV. SUM RATE MAXIMIZING REGULARIZATION

The optimal regularization parameter \( \alpha^\ast \) maximizing (24) is defined as

\[
\alpha^\ast = \arg \max_{\alpha > 0} \sum_{k=1}^{K} \log \left( 1 + \gamma_{k,\text{ref}}^\alpha \right).
\]  

(29)

In general, the optimization problem (29) is not convex in \( \alpha \) and the solution needs to be computed via a one-dimensional line search.

In the following, we confine ourselves to the case of common correlation \( \Theta_k = \Theta \), since for per-user correlation a common regularization term is not optimal anymore [12].
Let \( \alpha = 1/\rho, \Theta_k = I_M, P = K \) and \( \tau_k^2 = \{0.8, 0.1, 0.3, 0.05, 0.2\} \) for \( M = 5 \) and \( \tau_k^2 = \{0.1, 0.15, 0.2\} \) for \( M = 3 \).

An improved deterministic approximation of the SINR of the optimal linear precoder in [12] under per-user correlation and weighted sum rate maximization has been derived in [36]. Since for common transmit correlation the users’ channels are statistically equivalent, it is reasonable to assume that the distortions \( \tau_k^2 \) of the CSIT \( \hat{H}_k \) are identical for all users. In this conditions, \( P = I_K \) maximizes (24) and the optimization problem (29) has the following solution.

**Proposition 2:** Let \( \Theta_k = \Theta, \tau_k = \tau \) and \( P = I_K \). The approximated SINR \( \gamma_k^\circ \) of user \( k \) under RZF precoding (equivalently, the approximated per-user rate and the sum rate) is maximized for a regularization term \( \alpha = \alpha^* \), given as a positive solution to the fixed-point equation

\[
\alpha^* = \frac{1 + \tau^2 \rho \eta(\alpha^*)}{(1 - \tau^2)\beta \rho},
\]

where \( \eta(\alpha) \) is given by

\[
\eta(\alpha) = \frac{e_2}{e_1} \frac{\text{tr} \Theta^2 (\alpha [1 + m^\circ(\alpha)] I_M + \Theta/\beta)^{-2}}{\text{tr} \Theta (\alpha [1 + m^\circ(\alpha)] I_M + \Theta/\beta)^{-2}}.
\]

with \( m^\circ \) defined in (14).

**Proof:** The proof is provided in Appendix V.

Note that the solution in Proposition 2 assumes a fixed distortion \( \tau^2 \). Later in Section VI the distortion becomes a function of the quantization codebook size and in Section VII it depends on the uplink SNR as well as the amount of channel training.

Under perfect CSIT (\( \tau^2 = 0 \)), Proposition 2 simplifies to the well-known solution \( \alpha^* = 1/(\beta \rho) \), which has previously been derived in [9], [19], [21]. As mentioned in [9], for large \( M \) the RZF precoder is identical to the MMSE precoder in [16], [37].

The authors in [21] showed that, under perfect CSIT, \( \alpha^* \) is independent of the correlation \( \Theta \). However, for imperfect CSIT (\( \tau^2 \neq 0 \)), the optimal regularization (30) depends on the transmit correlation through \( \eta(\alpha) \). For uncorrelated channels (\( \Theta = I_M \)), we have \( \eta(\alpha) = 1 \) and therefore the closed form solution

\[
\alpha^* = \frac{1 + \tau^2 \rho}{1 - \tau^2} \frac{1}{\beta \rho}.
\]

Moreover, for \( \tau^2 > 0 \), the RZF-CDA precoder and the MMSE precoder with \( \alpha_{\text{MMSE}} = \tau^2/\beta + 1/(\beta \rho) \) [37] are not identical anymore, even in the large \( M \) limit. Further note that, for \( \tau^2 > 0 \), at asymptotically high SNR the regularization term \( \alpha^* \) in (30) converges to

\[
\lim_{\rho \to \infty} \alpha^* = \frac{\tau^2}{1 - \tau^2} \frac{1}{\beta ^2}.
\]

Thus, for asymptotically high SNR, RZF-CDA is not zero-forcing (ZF) precoding, since the regularization parameter \( \alpha^* \) is non-zero due to the residual interference caused by the imperfect CSIT. Similar observations have been made in [37] for the MMSE precoder.

**Remark 2:** Note that in (33) we apply the limit \( \rho \to \infty \) on a result obtained from an SINR approximation which is almost surely exact as \( M \to \infty \). This is correct if \( \Psi \) in (12) is bounded as \( M \to \infty \). For \( \tau^2 > 0 \) it is clear that \( \Psi \rho \) is bounded since \( \alpha^* > 0 \). In the case where \( \tau^2 = 0 \), we have \( \lim_{\rho \to \infty} \alpha^* = 0 \) and thus for \( \beta < 1 \) the support of the limiting eigenvalue distribution of \( H^H H \) includes zero resulting in an unbounded \( \Psi \rho \). From Remark 1, it follows that \( \lambda_{\min}(H^HH) > \varepsilon \) for \( \varepsilon > 0 \), excluding the mass in zero, if \( \beta > 1 \). \( \Theta_k = \Theta \) and \( \lambda_{\min}(\Theta) > \zeta, \zeta > 0 \). Thus \( \Psi \rho \) is bounded. A similar result has not been proved for the model in Theorem 1 and we have to evoke Assumption 1 to ensure that \( \Psi \rho \) is bounded for \( \beta > 1 \). Thus, for \( \tau^2 = 0 \), the limit (33) is only defined for \( \beta > 1 \). Further note that if \( \Psi \rho \) is bounded as \( M \to \infty \) and \( \rho \to \infty \) can be inverted without affecting the result.

Substituting (30) into the deterministic equivalent for the SINR (13) yields the following simple expression.

**Corollary 4:** Let \( \Theta_k = \Theta, \tau_k = \tau \) and \( \gamma_{k,\text{rrz}}^\circ \) be the maximum SINR of user \( k \) under RZF precoding. Then \( \gamma_{k,\text{rrz}}^* - \gamma_{k,\text{rrz}}^\circ \overset{M \to \infty}{\sim} 0 \), almost surely, where \( \gamma_{k,\text{rrz}}^\circ \) is given by

\[
\gamma_{k,\text{rrz}}^\circ \triangleq \gamma_{k,\text{rrz}}^\circ = m^\circ(-\alpha^*)
\]

\[
= \frac{1}{M} \text{tr} \Theta \left( \frac{\Theta/\beta}{1 + m^\circ(-\alpha^*) + \alpha^* I_M} \right)^{-1}.
\]

**Proof:** Substituting (30) into (110) together with \( \eta = e_2/e_1 \), we obtain (34) which completes the proof.

For uncorrelated channels \( \Theta_k = I_M \), the solution to (34) is of closed form and summarized in the following corollary.

**Corollary 5:** Let \( \Theta_k = I_M, \tau_k = \tau \) and \( \gamma_{k,\text{rrz}}^* \) be the maximum SINR of user \( k \) under RZF precoding. Then \( \gamma_{k,\text{rrz}}^* - \gamma_{k,\text{rrz}}^\circ \overset{M \to \infty}{\sim} 0 \), almost surely, where \( \gamma_{k,\text{rrz}}^\circ \) is given by

\[
\gamma_{k,\text{rrz}}^\circ \triangleq \gamma_{k,\text{rrz}}^\circ = m^\circ(-\alpha^*) = \frac{\omega}{2}(\beta - 1) + \frac{\chi}{2} - \frac{1}{2}
\]

where \( \omega \in [0, 1] \) and \( \chi \) are given by

\[
\omega = \frac{1 - \tau^2}{1 + \tau^2 \rho},
\]

\[
\chi = \sqrt{(\beta - 1)^2 \omega^2 \rho^2 + 2(1 + \beta) \omega \rho + 1}.
\]
Proof: Substituting $\Theta = I_M$ into Corollary 4 leads to a quadratic equation in $\tilde{\alpha}^2(-\alpha^{*\circ})$ for which the only positive solution is given by (35), which completes the proof.

In the following the RZF precoder with sum rate maximizing regularization is referred to as RZF channel distortion aware (RZF-CCA) precoder.

The impact of the regularization term on the ergodic sum rate is depicted in Figures 5 and 6.

In Figure 5, we compare the ergodic sum rate performance for different regularization parameters $\alpha$ with CSIT distortion $\tau_k^2=0.1$. The upper bound $\alpha = \alpha^*$ is obtained by optimizing $\alpha$ for each channel realization, whereas $\alpha^*$ maximizes the ergodic sum rate. It can be observed that both $\alpha^*$ and $\alpha^{*\circ}$ perform close to the optimal $\alpha^*$. Furthermore, if the channel uncertainty $\tau_k^2$ is unknown at the transmitter (and hence assumed zero), the performance is decreasing as soon as $\tau_k^2$ dominates (i.e. the inter-user interference limits the performance) the noise power $\sigma^2$ and approaches the sum rate of ZF precoding for high SNR. We conclude that (i) adapting the regularization term yields a significant performance increase and (ii) that the proposed RZF-CDA with $\alpha^{*\circ}$ performs close to optimal even for small system dimensions.

In Figure 5, we simulate the impact of transmit correlation in the computation of $\alpha^{*\circ}$ on the sum rate. For this purpose, we use the standard exponential correlation model as done in [21], i.e.,

$$[\Theta]_{ij} = v^{|i-j|}.$$ 

We compare two different RZF precoder. A precoder coined RZF-CCA that takes the channel correlation into account and computes $\alpha$ according to (30). The second precoder, called RZF-CCU, does not take $\Theta$ into account and computes $\alpha$ as in (32). We observe that for high correlation, i.e., $v = 0.9$, the RZF-CCA precoder significantly outperforms the RZF-CCU precoder at medium to high SNR, whereas both precoder perform equally at low SNR. Therefore, we conclude that it is beneficial to account for transmit correlation, only in highly correlated channels. Further simulations suggest that this effect is less pronounced as the $\tau_k^2$ increase.

V. SUM RATE MAXIMIZING NUMBER OF ACTIVE USERS

Consider the problem of finding the optimal system loading $\beta^{*\circ}$ maximizing the approximated sum rate per transmit antenna for a fixed $M$, i.e.,

$$\beta^{*\circ} = \arg \max_{\beta} \frac{1}{\beta} \sum_{k=1}^K \log (1 + \gamma_k^2) ,$$

where $\gamma_k^2$ denotes either $\gamma_{k,x}^2$ with $\beta > 1$ or $\gamma_{k,tx}^2$ with $\beta \geq 1$. In general (38) has to be solved by a one-dimensional line search. However, in case of ZF precoding and uncorrelated antennas the optimization problem (38) has a closed-form solution given in the following proposition.

Proposition 3: Let $\Theta_k = I_M$, $P = I_K$ and $\tau_k = \tau$, the sum rate maximizing system loading per transmit antenna $\beta^{*\circ}$ is given by

$$\beta^{*\circ} = \left(1 - \frac{1}{a} \right) \left(1 + \frac{1}{W(x)} \right),$$

where $a = \frac{1-x}{x^{1+\frac{1}{\beta}}} \in \mathbb{R}$ and $W(x)$ is the Lambert W-function defined as $z = W(z)e^{W(z)}$, $z \in \mathbb{C}$.

Proof: Substituting the SINR in Corollary 3 into (38) and differentiating along $\beta$ leads to

$$a\beta \left(1 + a(\beta - 1) \right) = \log \left(1 + a(\beta - 1) \right)$$

Denoting $w(\beta) = \frac{a-1}{a(\beta-1)+1}$, we can rewrite (40) as

$$w(\beta)e^{w(\beta)} = x.$$ 

Noticing that $w(\beta) = W(x)$ and solving for $\beta$ yields (39), which completes the proof.

For $\tau \in [0, 1]$, $\beta > 1$ we have $w \geq -1$ and $x \in [-e^{-1}, \infty)$. In this case $W(x)$ is a single-valued function. If $\tau^2 = 0$, we obtain the results in [14], although in [14] they are not given in
closed form. Note that for $\tau^2 = 0$, we have $\lim_{\rho \to \infty} \beta^{\tau^2} = 1$, i.e., the optimal system loading tends to one. Further note that only rational values of $\beta$ are meaningful in practice.

Figure 7 compares the optimal number of active users $K^{\tau^2} = M/\beta^{\tau^2}$ in (39) to the optimal number of active users $K^\tau$ obtained by choosing the $K = 1, 2, \ldots, M$ maximizing the ergodic sum rate, whereas Figure 8 depicts the impact of a suboptimal number of active users on the ergodic sum rate of the system.

From Figure 7 it can be observed, that (i) the approximate results $K^{\tau^2}$ do fit well with the simulation results even for small dimensions, (ii) $(K^\tau, K^{\tau^2})$ increase with the SNR and (iii), for $\tau^2 \neq 0$, $(K^\tau, K^{\tau^2})$ saturate for high SNR at a value lower than $M$. The impact of the system loading on the sum rate is depicted in Figure 8.

From Figure 8 we observe that, (i) the approximate large $M$ solution $K^{\tau^2}$ achieves most of the sum rate and (ii) adapting the number of users is beneficial compared to a fixed $K$. Moreover, from Figure 7, we identify $K = 8$ as a optimal choice (for $M = 16$) for medium SNR and, as expected, the performance is optimal in the medium SNR regime and suboptimal at low and high SNR. From Figure 7 it is clear that $K = 4$ is highly suboptimal in the medium and high SNR range and we observe a significant loss in sum rate. Consequently, the number of active users must be adapted to the channel conditions and the approximate result $K^{\tau^2}$ is a good choice to determine the number of active users in the system.

VI. OPTIMAL FEEDBACK IN LARGE FDD MULTI-USER SYSTEMS

Consider a frequency-division duplex (FDD) system, where the users quantize their perfectly estimated channel vectors and send the quantization index back to the transmitter over an independent feedback channel of limited rate. The feedback channels are assumed to be error-free and of zero delay. The quantization codebooks are generated prior to transmission and are known to both transmitter and respective receiver. Due to the finite rate feedback link, imposing a finite codebook size, the transmitter has only access to a imperfect estimate of the true downlink channel.

In the sequel, we follow the limited feedback analysis in [23], where each user's channel direction $\tilde{h}_k \triangleq \frac{h_k}{\|h_k\|_2}$ is quantized using $B$ bits which are subsequently fed back to the transmitter. For analytical tractability we assume random vector quantization (RVQ), where each user independently generates a random codebook $C \triangleq \{w_{i_1}, \ldots, w_{i_{2^B}}\}$ containing $2^B$ vectors $w_i \in \mathbb{C}^M$ that are isotropically distributed on the $M$-dimensional unit sphere. Subsequently, user $k$ quantizes its channel direction $\tilde{h}_k$ to the closest $w_i$ according to

$$\hat{h}_k = \arg \max_{w_i \in C} \| \tilde{h}_k^H w_i \|.$$ 

Since we use RVQ, the quantization error is isotropically distributed in $\mathbb{C}^M$ and independent of the quantization vector $\hat{h}_k$. [23]. Therefore the channel direction $\hat{h}_k$ can be written as [23]

$$\hat{h}_k = \sqrt{1 - \tau_k^2} \tilde{h}_k + \tau_k \tilde{q},$$

(41)

where $\tilde{q}$ is independent of $\tau_k$ and is isotropically distributed in the nullspace of $\tilde{h}_k$. Thus, the effect of imperfect CSIT under RVQ (41) is captured by the channel model (5). The quantization error $\tau_k^2$ can be upper-bounded as [23, Lemma 1]

$$\tau_k^2 < 2^{-\frac{B}{\tau}}.$$ 

(42)

To obtain tractable expressions, we restrict the subsequent analysis to i.i.d. channels ($\Theta_k = I_M$). Moreover, since the quantization codebooks of the users are supposed to be of equal size, the resulting CSIT distortions are identical, i.e., $\tau_k^2 = \tau^2$. Under this assumption, for large $M$, the SINR $\gamma^0$ is identical for all users and hence optimizing the SINR is

\[\begin{align*}
K^{\tau^2} = M/\beta^{\tau^2}, \quad (39) \\
K^\tau \text{ from exhaustive search}
\end{align*}\]

Fig. 7. ZF, sum rate maximizing number of active users vs. SNR with $\Theta_k = I_M$, $\tau^2 = 0.1$ and $P = I_K$.

\[\begin{align*}
\text{K = K^{\tau^2}} \\
\text{K = K^*} \\
\text{K = 8} \\
\text{K = 4}
\end{align*}\]

Fig. 8. ZF, ergodic sum rate vs. SNR with $M = 16$, $\Theta_k = I_M$, $P = I_K$ and $\tau^2 = 0.1$.
In order to maintain a rate $R^o = \log_2(1 + \gamma^o)$ [bits/s/Hz] and the sum rate $R^o_{\text{sum}} = KR^o$. From $\gamma_k - \gamma^o k \to \infty$ almost surely, the continuous mapping theorem [33] implies that

$$R_k - R^o \overset{M \to \infty}{\longrightarrow} 0,$$

almost surely, where $R_k = \log_2(1 + \gamma_k)$ is the rate of user $k$. Let $\Delta R_k = R_k - R^o$, be the rate gap of user $k$ with $R_k$ the rate of user $k$ under perfect CSIT, i.e., for $\tau^2 = 0$. Then, a deterministic equivalent $\Delta R^o$ such that

$$\Delta R_k - \Delta R^o \overset{M \to \infty}{\longrightarrow} 0,$$

almost surely, is given by

$$\Delta R^o = \bar{R}^o - R^o,$$

where $\bar{R}^o$ is the per-user rate under perfect CSIT.

In the following, under RVQ, we will derive the necessary scaling of the distortion $\tau^2$ to ensure that

$$\Delta R^o - \log_2 b \overset{M \to \infty}{\longrightarrow} 0,$$

almost surely with $b \geq 1$. That is, a constant rate gap of $\log_2 b$ is maintained exactly as $M \to \infty$. A constant rate gap ensures that the full multiplexing gain of $K$ is achieved. Thus, the proposed scaling also guarantees a larger but constant rate gap to the optimal DPC solution. The choice of a rate offset $\log_2 b$ is motivated by mere mathematical convenience to avoid terms of the form $2^b$. Further note that $\log_2 b$ can be easily translated into a power offset at high SNR (see [23] and references therein). That is, a sum rate offset of $M \log_2 b$ corresponds to a power offset of $3 \log_2 b$ dB.

With this strategy we closely follow [23]. In particular we compare the scaling of $\tau^2$ under RZF-CD-A, RZF-CDU and ZF ($\beta > 1$) precoding to the upper bound given for ZF ($\beta = 1$) precoding in [23, Theorem 3]. For the sake of comparison we restate [23, Theorem 3].

**Theorem 4:** [23, Theorem 3]. In order to maintain a rate offset no larger than $\log_2 b$ (per user) between zero-forcing with perfect CSIT and with finite-rate feedback (i.e., $\Delta R(\rho) \leq \log_2 b \forall \rho$), it is sufficient to scale the number of feedback bits per mobile according to

$$B^o = (M - 1) \log_2 \rho - (M - 1) \log_2 (b - 1) \approx \frac{M - 1}{3} \rho \log_2 b - (M - 1) \log_2 (b - 1).$$

It is also mentioned that the result in [23, Theorem 3] holds true for RZF-CDU precoding for high SNR, since ZF and RZF-CDU converge for high SNR. Furthermore, it is claimed, corroborated by simulation results, that [23, Theorem 3] is true under RZF-CDU precoding for all SNR.

In order to correctly interpret the subsequent results, it is important to understand the differences between our approach and the approach in [23]. The scaling given in [23, Theorem 3] is a strict upper bound on the ergodic per-user rate gap $E_{\text{H}}[\Delta R_k]$ for all SNR and all $M = K$. In contrast, our approach yields a necessary scaling of $\tau^2$ that maintains a given instantaneous target rate gap $\log_2 b$ exactly as $M \to \infty$. Therefore, our results are not upper bounds for finite $M$, i.e., we cannot guarantee that $\Delta R_k < \log_2 b$ for finite dimensions. But since for asymptotically large $M$, the rate gap is maintained exactly and we apply an upper bound on the CSIT distortion under RVQ (42), it follows that our results become indeed upper bounds for large $M$. Simulations reveal that under the derived scaling of $\tau^2$, the per-user rate gap is very close to $\log_2 b$ even for small dimension, e.g. $M = 10$.

Concerning the ergodic and instantaneous per user rate gap, the reader is reminded that our results hold also for ergodic per-user rates as a consequence of the dominated convergence theorem.

Consequently, a comparison of the results in [23] to our solutions is meaningful especially for larger values of $M$, where our results become upper bounds.

### A. Channel Distortion Aware Regularized Zero-forcing Precoding

The rate gap per user under RZF-CDA precoding is given in the following theorem.

**Theorem 5:** Let $\Theta_k = I_M$ and define $\Delta R_{k,\text{ref}}$ as the difference between the per-user rate of RZF-CDA precoding under perfect CSIT and imperfect CSIT. Then a deterministic equivalent $\Delta R^o_{\text{ref}}$ such that

$$\Delta R_{k,\text{ref}} - \Delta R^o_{\text{ref}} \overset{M \to \infty}{\longrightarrow} 0$$

almost surely, is given by

$$\Delta R^o_{\text{ref}} = \log_2 \left( \frac{1 + g(1, \beta)}{1 + g(\omega, \beta)} \right) \quad \text{[bits/s/Hz]},$$

where $\omega$ is given in (36) and

$$g(x, \beta) = x \rho (\beta - 1) + \sqrt{(1 - \beta)^2 x^2 \rho^2 + 2(1 + \beta) x \rho + 1}.$$

**Proof:** With Corollary 5, compute $\Delta R^o_{\text{ref}}$ as defined in (43).

Following the work in [23], we extend [23, Theorem 3] to RZF-CDA precoding for large $M$ in the following theorem.

**Theorem 6:** Let $\Theta_k = I_M$. Then the CSIT distortion $\tau^2$, such that the rate gap $\Delta R_{k,\text{ref}}$ of user $k$ between RZF-CDA precoding with perfect CSIT and imperfect CSIT satisfies

$$\Delta R_{k,\text{ref}} - \log_2 b \overset{M \to \infty}{\longrightarrow} 0$$

almost surely, has to scale as

$$\tau^2 = \frac{\phi^*_\text{ref}(\rho, b)}{\rho},$$

$$\phi^*_\text{ref}(\rho, b) = \rho \frac{[(1 + \beta)b + w(\beta - 1)] - \frac{1}{2\rho}(w^2 - b^2)}{1 + \beta b + w(\beta - 1) + \frac{1}{2\rho}(w^2 - b^2)},$$

$$w(\rho, b) = 1 - b + g(1, \beta).$$

**Proof:** Set $\Delta R^o_{\text{ref}} = \log_2 b$ and solve for $\tau^2$.

**Corollary 6:** In the conditions of Theorem 6 with $\beta = 1$, the distortion $\tau^2$ has to scale with

$$\tau^2 = \frac{1 + 4 \rho - \frac{w^2}{2\rho} - \frac{1}{3} + \frac{w^2}{2\rho^2}}{\rho}.$$
If the SNR grows to infinity, the term $\phi_{\text{rf}}^\circ(\rho, b)$ in (45) converges to the following limits,

$$\lim_{\rho \to \infty} \phi_{\text{rf}}^\circ(\rho, b) = \begin{cases} b^2 - 1 & \text{if } \beta = 1 \\ b - 1 & \text{if } \beta > 1. \end{cases}$$

(46)

Details can be found in Appendix VI-A.

For a direct comparison of Theorem 6 to [23, Theorem 3], we use the upper bound on the quantization distortion (42), i.e., $\tau^2 = 2 - \frac{M}{\rho}$, where $B_{\text{rf}}^\circ$ is the number of feedback bits per user under RZF-CDA precoding. Thus, (44) can be rewritten as

$$B_{\text{rf}}^\circ = (M - 1) \log_2 \rho - (M - 1) \log_2 \phi_{\text{rf}}^\circ(\rho, b).$$

(47)

### B. Channel Distortion Unaware Regularized Zero-forcing Precoding

Although the RZF-CDU precoder is suboptimal under imperfect CSIT, the results are useful to compare to the work in [23]. For the sake of comparison to [23, Theorem 3], we state the following theorem.

**Theorem 7:** Let $\Theta_k = I_M$. Then the CSIT distortion $\tau^2$, such that the rate gap $\Delta R_{k,\text{rf}}$ with $\alpha = 1/(\beta \rho)$ of user $k$ between RZF-CDU precoding with perfect CSIT and imperfect CSIT satisfies

$$\Delta R_{k,\text{rf}} - \log_2 b \xrightarrow{M \to \infty} 0$$

almost surely, has to scale as

$$\tau^2 = \phi_{\text{rf}}^\circ(\rho, b),$$

$$\phi_{\text{rf}}^\circ(\rho, b) = \frac{(b - 1 - \gamma^\ast)[\rho + \tilde{m}] + b m^\circ[\rho + \tilde{m}]}{(b - 1 - \gamma^\ast)[1 - \tilde{m}] + b m^\circ[1 + \frac{1}{\rho} \tilde{m}]}$$

with $\tilde{m} = (1 + m^{\gamma^\ast})^2$ and the approximated SINR under prefect CSIT $\gamma^\ast = [m^\circ(1 + \tilde{m}/\rho)]/[1 + \tilde{m}/\rho]$.

**Proof:** With Corollary 1, for $\alpha = 1/(\beta \rho)$, compute the rate gap per user $\log_2[1/(1 + \tau_{\text{rf}}^2)]$, equate it to $\log_2 b$ and solve for $\tau^2$. □

Applying the upper bound on the CSIT distortion under RVQ (42) with $B_{\text{rf}}^\circ$ bits per user, we obtain

$$B_{\text{rf}}^\circ = (M - 1) \log_2 \rho - (M - 1) \log_2 \phi_{\text{rf}}^\circ(\rho, b).$$

(48)

For asymptotically high SNR, $\phi_{\text{rf}}^\circ(\rho, b)$ converges to the following limits,

$$\lim_{\rho \to \infty} \phi_{\text{rf}}^\circ(\rho, b) = \begin{cases} 2(b - 1) & \text{if } \beta = 1 \\ b - 1 & \text{if } \beta > 1. \end{cases}$$

(49)

Details are provided in Appendix VI-B.

### C. Zero-forcing Precoding

The following results are only valid for $\beta > 1$ and thus, they can not be directly compared to [23, Theorem 3] which are derived under the assumption $M = K$.

**Theorem 8:** Let $\beta > 1$, $\Theta_k = I_M$ and define $\Delta R_{k,\text{zf}}$ to be the difference of the per-user rate under ZF precoding of perfect CSIT and imperfect CSIT. Then

$$\Delta R_{k,\text{zf}} - \Delta R_{\text{zf}}^\circ \xrightarrow{M \to \infty} 0$$

almost surely, with $\Delta R_{\text{zf}}^\circ$ given by

$$\Delta R_{\text{zf}}^\circ = \log_2 \left( \frac{1 + p(\beta - 1)}{1 + (1 + \tau^2)(\beta - 1)/\tau^2} \right) \text{[bits/s/Hz].}$$

**Proof:** Substitute the SINR from Corollary 3 into (43). □

**Theorem 9:** Let $\beta > 1$ and $\Theta_k = I_M$, to maintain a rate offset $\Delta R_{k,\text{zf}}$ such that

$$\Delta R_{k,\text{zf}} - \log_2 b \xrightarrow{M \to \infty} 0$$

almost surely, the distortion $\tau^2$ has to scale according to

$$\tau^2 = \phi_{\text{zf}}^\circ(\rho, b),$$

$$\phi_{\text{zf}}^\circ(\rho, b) = \frac{(b - 1)[1 + \rho(\beta - 1)]}{1 - b + (\beta - 1)[\rho + b]}.$$  

(50)

**Proof:** Set $\Delta R_{\text{zf}}^\circ = \log_2 b$ and solve for $\tau^2$. □

For asymptotically high SNR, $\phi_{\text{zf}}^\circ(\rho, b)$ in (50) converges to

$$\lim_{\rho \to \infty} \phi_{\text{zf}}^\circ(\rho, b) = b - 1.$$  

Under RVQ with $B_{\text{zf}}^\circ$ per user, we have

$$B_{\text{zf}}^\circ = (M - 1) \log_2 \rho - (M - 1) \log_2 \phi_{\text{zf}}^\circ(\rho, b).$$

(51)

For $\beta = 1$, the optimal scaling for high SNR is identical to RZF-CDU precoding (49) since both precoders are equivalent at high SNR.

### D. Discussion and Numerical Results

At this point we can draw the following conclusions. The optimal scaling of the CSIT distortion is lower for $\beta = 1$ compared to $\beta > 1$. For $\beta = 1$ the optimal scaling of the feedback bits $B_{\text{rf}}^\circ$, $B_{\text{zf}}^\circ$ and $B$ for ZF in [23, Theorem 3] are different, even in the high SNR limit. In fact, for large $M$, under RZF-CDU and ZF precoding, the upper bound in [23, Theorem 3] is too pessimistic in the scaling of the feedback bits. From (48) and (49), a more accurate choice is

$$B_{\text{rf}}^\circ = (M - 1) \log_2 \rho - (M - 1) \log_2 \phi_{\text{rf}}(\rho, b),$$

(52)

i.e., $M - 1$ bits less than proposed in [23, Theorem 3]. However, recall that (52) becomes an upper bound for large $M$ and a rate gap of at least $\log_2 b$ bits/s/Hz cannot be guaranteed for small values of $M$. Moreover, for high SNR, $\beta = 1$ and large $M$, to maintain a rate offset of $\log_2 b$, the RZF-CDA precoder requires $(M - 1) \log_2 (\frac{b}{b+1})$ bits less than the RZFCDU and ZF precoder and $(M - 1) \log_2 (b+1)$ bits less than the scaling proposed in [23, Theorem 3].

In contrast, for $\beta > 1$ and high SNR, we have $B_{\text{rf}}^\circ = B_{\text{zf}}^\circ$. Intuitively, the reason is that, for $\beta > 1$, the channel matrix is well conditioned and the RZF and ZF perform very closely. Therefore, both schemes are equally sensitive to imperfect CSIT and thus the scaling of $\tau^2$ is the same for high SNR.

Notice that our model comprises a generic distortion of the CSIT. That is, the distortion can be a combination of different additional factors, e.g. channel estimation at the receivers,
channel mismatch due to feedback delay or feedback errors. Moreover, we consider i.i.d. block-fading channels, which can be seen as a worst case scenario in terms of feedback overhead. It is possible to exploit channel correlation in time, frequency and space to refine the CSIT or to reduce the amount of feedback.

Figures 9 and 10 depict the ergodic sum rate of RZF precoding under RVQ and the corresponding number of feedback bits per user $B$, respectively. To avoid an infinitely high regularization term $\alpha^{\ast}$, the minimum number of feedback bits is set to one.

From Figure 9, we observe that (i) the desired sum rate offset of 10 bits/s/Hz is approximately maintained over the given SNR range when $B$ is chosen according to (47) and (48) under RZF-CDA and RZF-CDU precoding, respectively, (ii) given an equal number of feedback bits (47), the RZF-CDA precoder achieves a significantly higher sum rate compared to RZF-CDU for medium and high SNR, e.g. about 2.5 bits/s/Hz at 20 dB and (iii) to maintain a sum rate offset of $M$ bits/s/Hz, the proposed a feedback scaling of $B = \frac{M-1}{3} \rho_{dB}$ [23] is very pessimistic, since the sum rate offset is about 6 bits/s/Hz.

We conclude that the proposed RZF-CDA precoder significantly increases the sum rate for a given feedback rate or equivalently significantly reduces the amount of feedback given a target rate. Moreover, the scaling of the number of feedback bits under RZF-CDU precoding proposed in (47) appears to be more accurate than the scaling proposed in [23, Theorem 3].

VII. OPTIMAL TRAINING IN LARGE TDD MULTI-USER SYSTEMS

Consider a time division duplex (TDD) system where uplink (UL) and downlink (DL) share the same channel at different times. Therefore, the transmitter estimates the channel from known pilot signaling of the receivers. The channel coherence interval $T$, i.e. the amount of channel uses for which the channel is approximately constant, is divided into $T_i$ channel uses for UL training and $T - T_i$ 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 [38], 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) assuming $\Theta_k = I_M$.

A. Uplink Training Phase

In our setup, the distortion $\tau^2$ of the CSIT is solely caused by an imperfect channel estimation at the transmitter and is identical for all entries of $H$. To acquire CSIT, each user transmits $T_i \geq K$ orthogonal pilot symbols over the UL channel to the transmitter. Subsequently the transmitter estimates all $K$ channels simultaneously. At the transmitter, the signal $r_k$ received from user $k$ is given by

$$r_k = \sqrt{T_i P_{ul}} h_k + n_k,$$

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 $n_k = [n_1, n_2, \ldots, n_M]^T$ are statistically equivalent to their respective DL analog. Subsequently the transmitter performs a MMSE estimation of each channel coefficient $\hat{h}_{ij}$ $(i = 1, \ldots, K, \ j = 1, \ldots, M)$. Due to the orthogonality property of the MMSE estimation [39], the estimates $\hat{h}_{ij}$ of $h_{ij}$ and the corresponding estimation errors $\hat{h}_{ij}$ 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 [40]

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

(53)

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

**B. Optimization of Channel Training**

We focus on equal power allocation among the users, i.e., $P = I_K$. Since $T_i$ channel uses have already been consumed to train the transmitter over the user channels, there remains an interval of length $T - T_i$ for DL data transmission.

To compute the training length $T_i$ that maximizes the sum rate approximation (24), we substitute $\gamma^o_{k,ul}$ from Corollary 3 into (24) and the approximated sum rate $R_{\text{sum}}^{o,xf}$ under ZF precoding takes the form

$$R_{\text{sum}}^{o,xf} = K \left(1 - \frac{T_i}{T}\right) \log \left(1 + \frac{1 - \tau^2}{\tau^2 + \frac{1}{\rho_{dl}}} (\beta - 1)\right),$$

(54)

where $\rho_{dl} \triangleq \frac{P}{\sigma^2}$. Similarly, for RZF-CDA the approximated sum rate $R_{\text{sum}}^{o,rf}$ reads

$$R_{\text{sum}}^{o,rf} = K \left(1 - \frac{T_i}{T}\right) \log \left(1 + \frac{T_i \rho_{ul} (\beta - 1)}{1 + T_i \rho_{ul} + \frac{1}{\rho_{dl}}}\right),$$

(55)

where $\gamma^o_{r,f}$ is given in Corollary 5.

Substituting (53) into (54) and (55), we obtain

$$R_{\text{sum}}^{o,xf} = K \left(1 - \frac{T_i}{T}\right) \log \left(1 + \frac{T_i \rho_{ul} (\beta - 1)}{1 + T_i \rho_{ul} + \frac{1}{\rho_{dl}}}\right),$$

(56)

$$R_{\text{sum}}^{o,rf} = K \left(1 - \frac{T_i}{T}\right) \log \left(\frac{1}{2} + \frac{1}{2} \rho_{dl} (\beta - 1) + \frac{d(w)}{2}\right),$$

(57)

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

(58)

$$w = \frac{T_i \rho_{ul} \rho_{dl}}{1 + T_i \rho_{ul} + \rho_{dl}}.$$

For $\beta > 1$ under ZF precoding and $\beta \geq 1$ for RZF-CDA precoding, it is easy to verify that the functions $R_{\text{sum}}^{o,xf}$ and $R_{\text{sum}}^{o,rf}$ are strictly concave in $T_{i,xf}$ and $T_{i,rf}$ in the interval $[K, T]$, respectively, where $K$ is the minimum amount of training required, due to the orthogonality constraint of the pilot sequences. Therefore, we can apply standard convex optimization algorithms [41] to evaluate

$$T_{i,xf}^* = \arg \max_{K \leq T_{i,xf} \leq T} R_{\text{sum}}^{o,xf},$$

(59)

$$T_{i,rf}^* = \arg \max_{K \leq T_{i,rf} \leq T} R_{\text{sum}}^{o,rf}.$$  

(60)

In the following we derive approximate closed-form solutions to (59) and (60) for high SNR. We distinguish two cases, (i) the UL and DL SNR vary with finite ratio $c \triangleq \rho_{dl}/\rho_{ul}$ and (ii) $\rho_{ul}$ varies while $\rho_{dl}$ remains finite. In contrast to case (i), the system in case (ii) is interference-limited due to the finite transmit power of the users.

1) Case 1: finite ratio $\rho_{dl}/\rho_{ul}$: We derive approximate, but explicit, solutions for the optimal training intervals $T_{i,xf}^*, T_{i,rf}^*$ in the high SNR regime and derive their limiting values for asymptotically low SNR.

a) High SNR Regime: An approximated closed form solution to (59) and (60) is summarized in the following proposition.

**Proposition 4:** Let $\rho_{dl}, \rho_{ul}$ be large with $c = \rho_{dl}/\rho_{ul}$ finite. Then, an approximation of the sum rate maximizing amount of channel training $T_{i,xf}^*$ and $T_{i,rf}^*$ under ZF and RZF-CDA precoding is given by

$$T_{i,xf}^* = \max \left[ c \left(1 + \frac{T_i}{T} \right) \log \left(1 + \frac{c}{2} \frac{\rho_{ul}}{\rho_{dl}}\right), \right]$$

(61)

$$T_{i,rf}^* = \begin{cases} \max \left[ \frac{c}{2} \sqrt{1 + 2 \frac{T_i}{T} \frac{\rho_{ul}}{\rho_{dl}} - \frac{c}{2}}, K \right] & \text{if } \beta = 1, \\ \max \left[ \frac{c}{2} \sqrt{1 + 2 \frac{T_i}{T} \frac{\rho_{ul}}{\rho_{dl}} - \frac{c}{2}}, K \right] & \text{if } \beta > 1, \end{cases}$$

(62)

where $\tilde{R}_{xf} = \log(1 + \rho_{dl} (\beta - 1))$ and $\tilde{R}_{rf} = \frac{1}{2} + \frac{1}{2} \rho_{dl} (\beta - 1) + \frac{d(w)}{2}$.

**Proof:** The proof is presented in Appendix VII.

Thus, the optimal training intervals scale as $T_{i,xf}^* \sim \sqrt{T}$ and $T_{i,rf}^* \sim 1/\sqrt{\log(\rho_{dl})}$. Under ZF precoding the same scaling has been reported in [42]-[44].

From (61) and (62) it is clear that, as $\rho_{dl} \to \infty$, both $T_{i,xf}^*, T_{i,rf}^*$ tend to $K$, the minimum amount of training, since $\tilde{R}_{xf}^*, \tilde{R}_{rf}^* \to \infty$.

Moreover, for $\beta > 1$, $\tilde{R}_{xf}^* > \tilde{R}_{rf}^*$ with equality if $\rho_{dl} \to \infty$. Therefore, RZF-CDA 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-CDA requires less training compared to the scenario where $\beta > 1$.

b) Low SNR Regime: For asymptotically low SNR $\rho_{dl}, \rho_{ul} \to 0$ with constant ratio $c = \rho_{dl}/\rho_{ul}$ the optimal amount of training is given in the subsequent proposition.

**Proposition 5:** Let $\rho_{dl}, \rho_{ul} \to 0$ with constant ratio $c = \rho_{dl}/\rho_{ul}$. Then the sum rate maximizing amount of channel training $T_{i,xf}^*$ and $T_{i,rf}^*$ under ZF and RZF-CDA precoding converges to

$$\lim_{\rho_{ul} \to \infty} T_{i,xf}^* = \lim_{\rho_{ul} \to \infty} T_{i,rf}^* = \frac{T}{2}.$$  

(63)

**Proof:** Applying $\log(1 + x) = x + O(x^2)$ and $\rho_{ul} = \rho_{dl}/c$, equations (56) and (57) take the form

$$R_{\text{sum}}^{o,xf} = K \left(1 - \frac{T_i}{T}\right) \frac{T_i \rho_{ul} (\beta - 1)}{c} \rho_{dl}^2 + O(\rho_{dl}^2),$$

(64)

$$R_{\text{sum}}^{o,rf} = K \left(1 - \frac{T_i}{T}\right) \frac{T_i \rho_{ul} (\beta - 1)}{c} \rho_{dl}^2 + O(\rho_{dl}^2).$$

(65)

Maximizing equations (64) and (65) with respect to $T_{i,xf}$ and $T_{i,rf}$, respectively, yields (63).

For ZF precoding, the limit has also been reported in [45].

2) Case 2: $\rho_{dl} \gg \rho_{ul}$ with finite $\rho_{ul}$: This scenario models a high capacity DL channel where the primary sum rate loss stems from the inaccurate CSIT estimate due to a limited rate UL signalling caused e.g. by a finite transmit power of the users. Thus the system becomes interference limited and the...
Setting the derivative of (67) with respect to $T$ rate maximizing amount of channel training given in the following proposition. 

**Proposition 6:** Let $\rho_{dl} \to \infty$ with finite $\rho_{ul}$. Then the sum rate maximizing amount of channel training $T_{t,xf}^{\circ}$ is given by

$$T_{t,xf}^{\circ} = \frac{1}{\rho_{ul}(\beta - 1)} \left( \frac{a}{W(ae)} - 1 \right),$$

(66)

where $W(z)$ is the Lambert W-function, defined as the unique solution to $z = W(z)e^{W(z)}$, $z \in \mathbb{R}$.

**Proof:** For ZF precoding and $\rho_{dl} \to \infty$, the sum rate (56) can be approximated as

$$P_{\text{sum}} \approx K \left( 1 - \frac{T_{t,xf}^{\circ}}{T} \right) \log (1 + T_{t,xf}^{\circ} \rho_{ul}(\beta - 1)).$$

(67)

Setting the derivative of (67) with respect to $T_{t,xf}$ to zero, yields

$$\log(a/\omega(T_{t,xf})) = \omega(T_{t,xf}) - 1,$$

(68)

where $a \triangleq \rho_{ul}T(\beta - 1) + 1$ and $\omega(T_{t,xf}) = (Ta)/[T + T_{t,xf}(a - 1)]$. Equation (68) can be written as

$$\omega(T_{t,xf})e^{\omega(T_{t,xf})} = ae.$$

Notice that $\omega(T_{t,xf}) = W(ae)$. Thus, solving $\omega(T_{t,xf}) = W(ae)$ for $T_{t,xf}$ yields (66).

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

For RZF-CDA, no accurate closed-form solution to (60) can be obtained.

**C. Numerical Results**

In Figure 11, we compare the approximated optimal training intervals $T_{t,xf}^{\circ}$, $T_{t,ref}^{\circ}$ to $T_{t,xf}^{\circ}$, $T_{t,ref}^{\circ}$ computed via exhaustive search and averaged over 1,000 independent channel realizations. The regularization term $\alpha$ is computed using the large system approximation $\alpha^{\circ}$ in (32). Figure 11 shows that the approximate solutions $T_{t,xf}^{\circ}$, $T_{t,ref}^{\circ}$ become very accurate for $K = 16$. Moreover, it can be observed that the approximations in (61) and (62) match very well. Further notice that for $\beta = 2$, ZF and RZF need approximately the same amount of training, as predicted by equations (61) and (62).

Figure 12 depicts the optimal relative amount of training $T_{t,xf}^{\circ}/T$ for ZF and RZF precoding. We observe that $T_{t,xf}^{\circ}/T$ 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 transmission. Furthermore, $T_{t,xf}^{\circ}/T$ saturates at $K/T$ due to the orthogonality constraint on the pilot sequences. Furthermore, as expected from (61) and (62), we observe that the optimal amount of training is less for RZF than for ZF precoding. Moreover, the relative amount of training $T_{t,xf}^{\circ}/T$ 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.
Finally, we demonstrate in Appendix I-C that $\rho$ is used for all SNR. We conclude that our approximation in (66) achieves very good performance and can therefore be utilized to compute $T_{t,sf}$ very efficiently.

VIII. CONCLUSION

In this paper we derived deterministic equivalents of the SINR of ZF and RZF precoding by applying novel results from large dimensional random matrix theory. These approximations are valid for any SNR and shown to be very accurate even for finite system dimensions. Therefore, they provide useful tools for many applications, such as the sum rate maximizing number of users in a cell or the optimal ratio between channel training and data transmission in large TDD systems. In particular, we proposed a RZF precoder which takes the information about the channel transmit correlation and the CSIT mismatch into account. Among others, we find that, given a target user rate, the improved precoder design can significantly decrease the amount of necessary feedback rate. Although not yet practical, large numbers of transmit antennas are expected to be widely deployed in the future, at which point the results derived in this paper will prove all the more valuable to system designers.

APPENDIX I

PROOF OF THEOREM 1

The proof is structured as follows: In Appendix I-A, we prove the convergence of (7). Subsequently in Appendix I-B we show that the sequence $\{e_{N,i}^{(k)}(z)\}$ defined by (10) converges to $e_{N,i}(9)$ as $k \rightarrow \infty$ if properly initialized. Finally, we demonstrate in Appendix I-C that $e_{N,i}$ satisfies $|m_{B_N, \Theta_i} - e_{N,i}| \overset{N \rightarrow \infty}{\longrightarrow} 0$, almost surely.

A. Proof of Convergence

The objective is to approximate the random variable $m_{B_N, Q_N}(z)$ by a deterministic functional $1/\sqrt{N} \text{tr}D^{-1}$ such that

$$Q_N(B_N - zI_N)^{-1} = D^{-1} \left[ D - (X_N^H X_N + S_N - zI_N)Q_N^{-1} \right] Q_N(B_N - zI_N)^{-1},$$

(70)

Since $D$ is deterministic, it can account for the deterministic matrices $S_N$, $Q_N$ and $zI_N$ inside the brackets in (70). Therefore we set

$$D = (R + S_N - zI_N)Q_N^{-1},$$

(71)

where $R$ is to be determined later, and obtain

$$Q_N(B_N - zI_N)^{-1} = D^{-1}R(B_N - zI_N)^{-1} - D^{-1}X_N^H X_N(B_N - zI_N)^{-1}.$$

Consider the term $D^{-1}X_N^H X_N(B_N - zI_N)^{-1}$. Taking the trace, together with $X_N^H X_N = \sum_{i=1}^n \Psi_i y_i^H \Psi_i^H$, we have

$$\frac{1}{N} \text{tr}D^{-1} X_N^H X_N(B_N - zI_N)^{-1} = \frac{1}{N} \sum_{i=1}^n y_i^H \Psi_i^H \Psi_i^H(B_N - zI_N)^{-1} = \frac{1}{N} \sum_{i=1}^n y_i^H \Psi_i^H \Psi_i^H(B_N - zI_N)^{-1}D^{-1} \Psi_i y_i.$$

Denoting $B[i] = B_N - \Psi_i y_i^H \Psi_i^H$ and applying Lemma 2, we obtain

$$\frac{1}{N} \text{tr}Q_N(B_N - zI_N)^{-1} - \frac{1}{N} \text{tr}D^{-1} = \frac{1}{N} \sum_{i=1}^n y_i^H \Psi_i^H(B[i] - zI_N)^{-1} D^{-1} \Psi_i y_i.$$

Therefore, the left-hand side of (69) takes the form

$$\frac{1}{N} \text{tr}Q_N(B_N - zI_N)^{-1} - \frac{1}{N} \text{tr}D^{-1} = \frac{1}{N} \sum_{i=1}^n y_i^H \Psi_i^H(B[i] - zI_N)^{-1} D^{-1} \Psi_i y_i.$$

The choice of an appropriate value for $R$, such that (69) is satisfied, requires some intuition. From Lemma 4 we know that

$$y_i^H \Psi_i^H(B[i] - zI_N)^{-1} \Psi_i y_i - \frac{1}{N} \text{tr} \Theta_i(B[i] - zI_N)^{-1} N \rightarrow \infty 0,$$

almost surely. Then, from Lemma 8, we surely have

$$\frac{1}{N} \text{tr} \Theta_i(B[i] - zI_N)^{-1} - \frac{1}{N} \text{tr} \Theta_i(B_N - zI_N)^{-1} N \rightarrow \infty 0.$$ 

From the previous arguments, $R$ will be chosen as

$$R = \frac{1}{N} \sum_{i=1}^n \frac{\Theta_i}{1 + \frac{1}{N} \text{tr} \Theta_i(B[i] - zI_N)^{-1}}.$$ 

(73)

Denote $m_{B_N, \Theta_i}(z) \triangleq \frac{1}{N} \text{tr} \Theta_i(B_N - zI_N)^{-1}$. Substituting $Q_N = \Theta_i$ into (71), we define a deterministic quantity $e_{N,i}(z)$ as

$$e_{N,i}(z) = \frac{1}{N} \text{tr} \Theta_i \left( \frac{1}{N} \sum_{j=1}^n \frac{\Theta_j}{1 + e_{N,j}(z)} + S_N - zI_N \right)^{-1}.$$ 

The term $e_{N,i}(z)$ is shown later in Appendix I-C to satisfy $m_{B_N, \Theta_i}(z) - e_{N,i}(z) \overset{N \rightarrow \infty}{\longrightarrow} 0$, almost surely. The remainder of this subsection proves (69) for the specific choice of $R$ in
Let the following upper bounds be well defined and let the entries of $y_i$ have eighth order moment of order $O\left(\frac{1}{N^2}\right)$. Then the $p$th order moments $E\left[|d_{i}^{(1)}|^p\right]$, $(l=1, 2, 3, 4)$ can be bounded as

$$E\left[|d_{i}^{(1)}|^p\right] \leq 2^{p-1}\left(\frac{\beta T^3 Q |z|^3}{(3\beta)^7}\right)^p \frac{1}{N^p} \left(\frac{C_p^{(1)}}{N^{p/2}} + 1\right)$$

$$E\left[|d_{i}^{(2)}|^p\right] \leq \frac{|z|^4}{(3\beta)^4} C_p^{(2)} P^{(N/2)}$$

$$E\left[|d_{i}^{(3)}|^p\right] \leq \left(\frac{|z|^T Q}{N (3\beta)^3}\right)^p \left[ 1 + \frac{\beta T^2 |z|^2}{(3\beta)^4} \right]$$

$$E\left[|d_{i}^{(4)}|^p\right] \leq 2^{p-1}\left(\frac{\beta T^3 Q |z|^3}{(3\beta)^7}\right)^p \frac{1}{N^p} \left(\frac{C_p^{(4)}}{N^{p/2}} + \frac{T^p}{(3\beta)^3} P^{(N/2)}\right),$$

where the $C_p^{(i)}$, $i \in \{1, 2, 4\}$ are constants depending only on $p$.

**Proof:** The proof is based on various common inequalities. Applying Lemma 9, $|d_{i}^{(1)}|$ can be upper-bounded as

$$|d_{i}^{(1)}| \leq |z| \left|\sum_{z}^{\beta} \left|y_i^{(1)}\right| (y_i^{(1)} - z) \left[D_{[-i]}^{(1)} - D_{[-i]}^{(1)}\right] y_i^{(1)}\right|.$$ 

We further bound $|d_{i}^{(1)}|$ by applying Lemmas 10 and 12 with the fact that $\|B_{[-i]} - zI\|^{-1} \leq \frac{1}{3\beta}$. Together with (76) we have

$$|d_{i}^{(1)}| \leq \frac{|z| T}{(3\beta)^2} \|y_i\|_2 \|D_{[-i]}^{(1)} - D_{[-i]}^{(1)}\|.$$ 

Similarly, with Lemma 1, it can be shown that $\|D_{[-i]}^{(1)} - D_{[-i]}^{(1)}\| \leq \frac{\beta T^2 |z|^3}{N (3\beta)^2}$ and thus

$$|d_{i}^{(1)}| \leq \frac{\beta T^3 Q |z|^3}{N (3\beta)^7} \|y_i\|_2.$$ 

The $p$th order moment of $|d_{i}^{(1)}|$ thus satisfies

$$E\left[|d_{i}^{(1)}|^p\right] \leq \left[\frac{\beta T^3 Q |z|^3}{(3\beta)^7}\right]^p \frac{1}{N^p} E\left[|y_i^{(1)} y_i - 1|^p\right].$$

Applying the inequality $|x + y| \leq 2^{p-1}(|x|^p + |y|^p)$ yields

$$E\left[|d_{i}^{(1)}|^p\right] \leq 2^{p-1}\left(\frac{\beta T^3 Q |z|^3}{(3\beta)^7}\right)^p \left[ E\left[|y_i^{(1)} y_i - 1|^p\right] + 1\right].$$

If the moments $E[|d_{i}^{(1)}|^4]$ and $E[|d_{i}^{(1)}|^{2p}]$ exist and are bounded, we apply Lemma 3 and obtain (78). For the sake of brevity, we omit the derivations of the remaining moments $E[|d_{i}^{(1)}|^p], l = \{2, 3, 4\}$, since the techniques are similar to the previous procedure.

From Proposition 7, we conclude that all $E[|d_{i}^{(1)}|^p]$ are summable if $p = 2 + \varepsilon, \varepsilon > 0$. Therefore, since the upper bounds in Proposition 7 are independent of $N$, $E[|w_N|^p]$, $p = 2 + \varepsilon$ is summable and the Borel-Cantelli Lemma thus implies that $w_N \xrightarrow{N \to \infty} 0$, almost surely.
We now prove the existence and uniqueness of a solution to (9).

B. Proof of Convergence of the Fixed Point Equation

In this section we consider the fixed point equation (9). We first prove that, properly initialized, the sequence \( \{e_{N,i}^{(k)}\}, (k = 1, 2, \ldots) \), converges to a limit \( e_{N,i} \) as \( k \to \infty \). Subsequently, we show that this limit \( e_{N,i} \) satisfies \( |m_{B_N, \Theta, -e_{N,i}}| \to 0 \) almost surely.

**Proposition 8:** Let \( z \in \mathbb{C}^+ \) and \( \{e_{N,i}^{(k)}(z)\} (k \geq 0) \) be the sequence defined by (10). If \( \{e_{N,i}^{(0)}(z)\} \) is a Stieltjes transform, then all \( \{e_{N,i}^{(k)}(z)\} (k > 0) \) are Stieltjes transforms as well.

**Proof:** Suppose (10) is initialized by \( e_{N,i}^{(0)}(z) = -1/z \), which is the Stieltjes transform of a function with a single mass in zero. We demonstrate that at all subsequent iterations \( k > 0 \), the corresponding \( e_{N,i}^{(k)}(z) \) are Stieltjes transforms for all \( N \). For ease of notation we omit the dependence on \( z \), the \( e_{N,i}^{(k+1)} \) are given by

\[
e_{N,i}^{(k+1)} = \frac{1}{N} \text{tr} \Theta_i \left( \frac{1}{N} \sum_{j=1}^{n} e_{N,j}^{(k)} \Theta_j + S_N - zI_N \right)^{-1}
\]

where \( e_{N,j}^{(k)} = 1/(1 + e_{N,j}^{(k)}) \). In (79), multiplying \( A_k \) from the right by \((A_k^H)^{-1}A_k^H\), we obtain

\[
e_{N,i}^{(k+1)} = \frac{1}{N} \text{tr} A_k^H \Theta_i A_k \left[ \frac{1}{N} \sum_{j=1}^{n} e_{N,j}^{(k)} \Theta_j \right] + v_i^{(k)},
\]

where \( v_i^{(k)} = \frac{1}{N} \text{tr} A_k^H \Theta_i A_k [S_N - zI_N] \). Denoting \( r_i^{(k)} \triangleq \frac{1}{N} \left[ \frac{1}{N} \text{tr} A_k^H \Theta_i A_k \Theta_i \right], \frac{1}{N} \text{tr} A_k^H \Theta_i A_k \Theta_i \right] \) and \( c_{N,i}^{(k)} \triangleq \left[ c_{N,i}^{(n)}, \ldots, c_{N,i}^{(1)} \right]^T \), (80) takes the form

\[
e_{N,i}^{(k+1)} = r_i^{(k)} c_{N,i}^{(k)} + v_i^{(k)}.
\]

Since the \( \Theta_i \) are uniformly bounded w.r.t. \( N \), we have \( r_i^{(k)}, v_i^{(k)} > 0 \). To show that \( e_{N,i}^{(k+1)} \) are Stieltjes transforms of a nonnegative finite measure, the following three conditions must be verified [25, Proposition 2.2]: (i) \( e_{N,i}^{(k+1)}(z) \in \mathbb{C}^+ \), (ii) \( z e_{N,i}^{(k+1)}(z) \in \mathbb{C}^+ \) and (iii) \( \lim_{y \to +\infty} -i y e_{N,i}^{(k+1)}(iy) < \infty \). From (81) it is easy to verify that all three conditions are met, which completes the proof.

We are now in a position to show that any sequence \( \{e_{N,i}^{(k)}(z)\} \), \( k > 0 \) converges to a limit \( e_{N,i}(z) \) as \( k \to \infty \).

**Proposition 9:** Any sequence \( \{e_{N,i}^{(k)}(z)\}, (k > 0) \) defined by (10) converges to \( e_{N,i}(z) \) as \( k \to \infty \) if \( e_{N,i}^{(0)}(z) \) is a Stieltjes transform.

**Proof:** Let \( e_{N,i}^{(k)}(z) = \frac{1}{N} \text{tr} \Theta_k A_k^{(k-1)} \) and \( e_{N,i}^{(k+1)}(z) = \frac{1}{N} \text{tr} \Theta_k A_k^{(k)} \), where

\[
A_k^{(k-1)} = \left[ \frac{1}{N} \sum_{j=1}^{n} \frac{\Theta_j}{1 + e_{N,j}^{(k-1)}(z)} + S_N - zI_N \right]^{-1}
\]

\[
A_k^{(k)} = \left[ \frac{1}{N} \sum_{j=1}^{n} \frac{\Theta_j}{1 + e_{N,j}^{(k)}(z)} + S_N - zI_N \right]^{-1}
\]

Applying Lemma 1, the difference \( |e_{N,i}^{(k)}(z) - e_{N,i}^{(k+1)}(z)| \) is

\[
|e_{N,i}^{(k)}(z) - e_{N,i}^{(k+1)}(z)| = \frac{1}{N} \text{tr} A_k^{(k+1)} \Theta_k A_k^{(k)} \left[ \frac{1}{N} \sum_{j=1}^{n} \frac{e_{N,j}^{(k)} - e_{N,j}^{(k-1)}}{1 + e_{N,j}^{(k)}(z)} \right] \left[ \frac{1}{N} \sum_{j=1}^{n} \frac{e_{N,j}^{(k-1)}}{1 + e_{N,j}^{(k-1)}(z)} \right]^{-1}
\]

(82)

With Lemmas 9, 11 and 12, (82) can be bounded as

\[
|e_{N,i}^{(k)}(z) - e_{N,i}^{(k+1)}(z)| \leq C \sup_{1 \leq i \leq n} |e_{N,i}^{(k)} - e_{N,i}^{(k-1)}|,
\]

(83)

where \( C = \frac{\beta \sqrt{\tau}}{(\delta_{\max})^2} \). Clearly, the sequence \( \{e_{N,i}^{(k)}(z)\} \) converges to a limit \( e_{N,i}(z) \) for \( z \) restricted to the set \( \{ z \in \mathbb{C}^+: C < 1 \} \). Proposition 8 shows that all \( \{e_{N,i}^{(k)}\} \), including the limit \( e_{N,i} \), are Stieltjes transforms and therefore analytic and bounded in closed subsets of \( \mathbb{C}^+ \). Since \( \{e_{N,i}^{(k)}(z)\} \) for \( z \in \mathbb{C}^+: C < 1 \) is at least countable and has a cluster point, Vitali’s convergence theorem ensures that the sequence \( \{e_{N,i}^{(k)}\} \) must converge for all \( z \in \mathbb{C}\setminus\mathbb{R}^+ \).

It is easy to verify, that the previous holds also true for \( z \in \mathbb{C}^- \).

For \( z < 0 \), the existence of a unique solution to (9) as well as the convergence of (10) from any real initial point can be proved within the framework of standard interference functions [46]. The strategy is as follows. Let \( e_{N}(z) = [e_{N,1}(z), e_{N,2}(z), \ldots, e_{N,N}(z)]^T \in \mathbb{R}^N \) and \( f(e_{N}) = [f_1(e_{N}), f_2(e_{N}), \ldots, f_N(e_{N})]^T \in \mathbb{R}^n \), where

\[
f_i(e_{N}) = \frac{1}{N} \text{tr} \Theta_i \left( \frac{1}{N} \sum_{j=1}^{n} \frac{\Theta_j}{1 + e_{N,j}(z)} + S_N - zI_N \right)^{-1}
\]

Theorems 1 and 2 in [46] prove that, if \( f(e_{N}) \) is a feasible standard interference function, then (10) converges to a unique solution \( e_{N} \) with all nonnegative entries for any initial point \( e_{N}^{(0)}, \ldots, e_{N,n}^{(0)} \). The proof that \( f(e_{N}) \) is feasible as well as a standard interference function is straightforward and details are omitted in this correspondence.

The uniqueness of \( e_{N} \), whose entries are Stieltjes transforms of nonnegative finite measures, ensures the functional uniqueness of \( e_{N,1}(z), \ldots, e_{N,N}(z) \) as Stieltjes transform solution to (9) for \( z \in \mathbb{C}\setminus\mathbb{R}^+ \). This completes the proof of uniqueness.

The following section proves that \( e_{N,i}(z) = \lim_{k \to \infty} e_{N,i}^{(k)}(z) \) satisfies \( |m_{B_N, \Theta, -e_{N,i}(z)}| \to 0 \) almost surely.
C. Proof of Convergence of the Deterministic Equivalent

In Section I-A we showed that \( \frac{1}{N} \text{tr} Q_N (B_N - zI_N)^{-1} = \frac{1}{N} \text{tr} Q_N (B_N - zI_N)^{-1} - \frac{1}{N} \text{tr} Q_N (R + S_N - zI_N)^{N \to \infty} 0 \), almost surely. Furthermore in Section I-B we proved that the sequence defined by (9) converges to a limit \( e_{N,i}(z) \). It remains to prove that

\[
\begin{align*}
    m_{B_N, \Theta_i} - e_{N,i} & = \frac{1}{N} \text{tr} \Theta_i (B_N - zI_N)^{-1} \\
    & - \frac{1}{N} \text{tr} \frac{1}{N} \sum_{j=1}^{n} \frac{\Theta_j}{1 + e_{N,j}(z)} + S_N - zI_N \to 0, \quad \text{N \to \infty}.
\end{align*}
\]

Almost surely. Denote \( w_{N,i} \equiv w_{\Theta_i} \). Applying Lemma 1, (84) can be written as

\[
\begin{align*}
    m_{B_N, \Theta_i} - e_{N,i} &= w_{N,i} + \frac{1}{N} \text{tr} \Theta_i (A + S_N - zI_N)^{-1} - e_{N,i}(z) \\
    &= w_{N,i} + \frac{1}{N} \text{tr} \Theta_i (A + S_N - zI_N)^{-1} |A - B| (B + S_N - zI_N)^{-1},
\end{align*}
\]

where \( A \equiv \frac{1}{N} \sum_{j=1}^{n} \frac{\Theta_j}{1 + e_{N,j}(z)} \) and \( B \equiv \frac{1}{N} \sum_{j=1}^{n} \frac{\Theta_j}{1 + e_{N,j}(z)} \). Applying Lemmas 9 and 11, \( |m_{B_N, \Theta_i} - e_{N,i}| \) can be bounded as

\[
|m_{B_N, \Theta_i} - e_{N,i}| \leq |w_{N,i}| + \left\| \Theta_i \right\| \left\| A^{-1} \right\| \left\| B^{-1} \right\| \times \left\| \frac{1}{N} \sum_{j=1}^{n} \Theta_j \frac{|m_{B_N, \Theta_j} - e_{N,j}|}{(1 + e_{N,j})(1 + e_{N,j})} \right\|.
\]

(85)

Similar to (83), with Lemma 12, (85) can be further bounded as

\[
|m_{B_N, \Theta_i} - e_{N,i}| \leq |w_{N,i}| + C \sup_{1 \leq i \leq n} |m_{B_N, \Theta_i} - e_{N,i}|.
\]

Taking the supremum over all \( i = 1, \ldots, n \), we obtain

\[
\sup_{1 \leq i \leq n} |m_{B_N, \Theta_i} - e_{N,i}| \leq \sup_{1 \leq i \leq n} |w_{N,i}|.
\]

(86)

From (86), on the set \( \{ z \in \mathbb{C}^+ : 0 < C < 1 \} \), it suffices to show that \( \sup_{1 \leq i \leq n} |w_{N,i}| \) goes to zero sufficiently fast. For any \( \varepsilon > 0 \) we have

\[
P \left( \sup_{1 \leq i \leq n} |w_{N,i}| > \varepsilon \right) \leq P \left( \sum_{i=1}^{n} |w_{N,i}| > \varepsilon \right) \leq \sum_{i=1}^{n} P \left( |w_{N,i}| > \varepsilon \right) = \sum_{i=1}^{n} P \left( |w_{N,i}|^p > \varepsilon^p \right).
\]

(87)

Applying Markov's inequality, (87) can be further bounded as

\[
P \left( \sup_{1 \leq i \leq n} |w_{N,i}| \geq \varepsilon \right) \leq \frac{1}{\varepsilon^p} \sum_{i=1}^{n} E \left[ |w_{N,i}|^p \right].
\]

For all \( n \) and \( p = 4 + \varepsilon \) with \( \varepsilon > 0 \), the term \( \sum_{i=1}^{n} E \left[ |w_{N,i}|^p \right] \) is summable and we can apply the Borel-Cantelli Lemma which implies \( \sup_{1 \leq i \leq n} w_{N,i} \to 0 \), almost surely.

On \( \{ z \in \mathbb{C}^+ : 0 < C < 1 \} \), the \( e_{N,i}(z) \) are summable and have a cluster point. Furthermore, Proposition 8 assures that the \( e_{N,i}(z) \) are Stieltjes transforms and hence uniformly bounded on every closed set in \( \mathbb{C} \setminus \mathbb{R}^+ \). Therefore, Vitali's convergence theorem applies, and extends the convergence region of (84) to \( z \in \mathbb{C} \setminus \mathbb{R}^+ \).

APPENDIX II

DETERMINISTIC EQUIVALENT OF THE EMPIRICAL
STIETLJES TRANSFORM OF THE EIGENVALUES OF
\( B_n = X_N X_N^H + \tilde{S}_n \)

The following theorem is required in the proof of a deterministic equivalent of the SINR under ZF precoding in Appendix IV.

Theorem 10: Let \( \tilde{B}_n = X_N X_N^H + \tilde{S}_n \) with \( \tilde{S}_n \in \mathbb{C}^{n \times n} \) Hermitian nonnegative definite and \( X_N \in \mathbb{C}^{n \times N} \) random. The \( i \)th column \( x_i \) of \( X_N^H \) is \( x_i = \Psi_i \gamma_i \), where the entries of \( y_i \) are i.i.d. zero-mean, variance \( \frac{1}{N} \) and have eighth order moment of order \( O \left( \frac{1}{N^2} \right) \). The matrices \( \Psi_i \in \mathbb{C}^{N \times r_i} \) are deterministic. Furthermore, let \( \Theta_i = \Psi_i \Psi_i^H \in \mathbb{C}^{N \times N} \) and define \( Q_n \in \mathbb{C}^{n \times n} \) deterministic. Both \( \Theta_i \) and \( Q_n \) are assumed to have uniformly bounded spectral norm (with respect to \( N \)). Define

\[
m_{B_n, Q_n}(z) = \frac{1}{n} \text{tr} \tilde{Q}_n \left( \tilde{B}_n - zI_n \right)^{-1}.
\]

(88)

Then, for \( z \in \mathbb{C} \setminus \mathbb{R}^+ \), as \( n, N \) grow large with ratios \( \beta_i \equiv N/r_i \) and \( \beta N/n \) such that \( 0 < \text{lim inf}_N N \beta(N) \leq \text{lim sup}_N \beta(N) < \infty \) and \( 0 < \text{lim inf}_N N \beta_n \leq \text{lim sup}_N \beta_n, \), we have

\[
m_{B_n, Q_n}(z) - m_{B_n, Q_n}(z) \to 0, \quad \text{N \to \infty}.
\]

(89)

almost surely, with \( m_{B_n, Q_n}(z) \) given by

\[
m_{B_n, Q_n}(z) = \frac{1}{n} \text{tr} \tilde{Q}_n \left( \frac{1}{N} \sum_{j=1}^{N} \frac{V_j}{1 + e_{N,j}(z)} + \tilde{S}_n - zI_n \right)^{-1}
\]

(90)

where \( V_i = \text{diag}(u_{i1}^H \Theta_1 u_{i1}, u_{i2}^H \Theta_2 u_{i2}, \ldots, u_{in}^H \Theta_n u_{in}) \) with \( u_{ij} \) the \( i \)th eigenvector of \( \Theta_1 \). The \( e_{N,1}(z), \ldots, e_{N,N}(z) \) form the unique solution of

\[
e_{N,i}(z) = \frac{1}{n} \text{tr} V_i \left( \frac{1}{N} \sum_{j=1}^{N} \frac{V_j}{1 + e_{N,j}(z)} + \tilde{S}_n - zI_n \right)^{-1}
\]

(91)

which is the Stieltjes transform of a nonnegative finite measure on \( \mathbb{R}^+ \). Moreover, for \( z < 0 \), the \( e_{N,1}(z), \ldots, e_{N,N}(z) \) are the unique nonnegative solutions to (91). Furthermore, under Assumption 1, the theorem extends to \( z = 0 \).

Proof: The proof follows the same lines as the proof of Theorem 1 in Appendix I. We find a functional \( \frac{1}{n} \text{tr} D^{-1} \) such that

\[
\frac{1}{n} \text{tr} Q_n \left( \tilde{B}_n - zI_n \right)^{-1} - \frac{1}{n} \text{tr} D^{-1} N \to \infty 0
\]

(92)

almost surely. Applying Lemma 1 and setting \( D = (R + S_n - zI_n) Q_n^{-1} \) we obtain

\[
Q_n \left( \tilde{B}_n - zI_n \right)^{-1} - D^{-1} = D^{-1} R \left( \tilde{B}_n - zI_n \right)^{-1} - D^{-1} X_N X_N^H \left( \tilde{B}_n - zI_n \right)^{-1}.
\]

(93)
Instead of working with $X_N$, we will work with $X_N \triangleq X_N U_1$, where $U_1 = [u_{11}, \ldots, u_{1M}]$ is the eigenvector basis of $\Theta_1$. In order to apply Lemma 4, we need to characterize the $i$th column $\tilde{x}_i$ of $X_N$, which reads

$$
\tilde{x}_i = \left[ y_1^H \Psi, y_2^H \Psi, \ldots, y_n^H \Psi \right]^T u_{i1}, \ldots, u_{iM}
$$

where $u_{i1}$ is the $i$th eigenvector of $\Theta_1$, $V_i \triangleq \text{diag}(u_{i1}^H \Theta_1 u_{i1}, u_{i1}^H \Theta_1 u_{i2}, \ldots, u_{i1}^H \Theta_1 u_{iM})$ and $y_i \in \mathbb{C}^n$ has i.i.d. entries of zero mean and variance $1/N$. Consider the term $D^{-1}X_N X_N^H (B_N - z I_N)^{-1}$. Taking the trace, together with $X_N X_N = \sum_{i=1}^N V_i^{1/2} \tilde{y}_i^H \tilde{y}_i V_i^{1/2}$, we have

$$
\frac{1}{n} \text{tr} D^{-1}X_N X_N^H (B_N - z I_N)^{-1} = \frac{1}{n} \sum_{i=1}^N \tilde{y}_i^H V_i^{1/2} (B_N - z I_N)^{-1} D^{-1} V_i^{1/2} \tilde{y}_i
$$

Denoting $\tilde{B}_i = B_N - V_i^{1/2} \tilde{y}_i^H V_i^{1/2}$ and applying the Lemma 2 yields

$$
\frac{1}{n} \text{tr} \tilde{Q}_n (B_N - z I_N)^{-1} - \frac{1}{n} \text{tr} D^{-1} - \frac{1}{n} \text{tr} \tilde{B}_i (B_N - z I_N)^{-1} - \frac{1}{n} \text{tr} \tilde{Q}_n (B_N - z I_N)^{-1} - \frac{1}{n} \text{tr} \tilde{B}_i (B_N - z I_N)^{-1} = \frac{1}{n} \sum_{i=1}^N \tilde{y}_i^H V_i^{1/2} (B_N - z I_N)^{-1} D^{-1} V_i^{1/2} \tilde{y}_i
$$

With the same arguments as in Appendix I, we choose $\tilde{R}$ as

$$
\tilde{R} = \frac{1}{N} \sum_{i=1}^N \frac{V_i}{1 + \frac{1}{n} \text{tr} V_i (B_N - z I_N)^{-1}} V_i
$$

A deterministic equivalent $\tilde{c}_{N,i}(z)$ of $\frac{n}{N} \tilde{m}_{B_n} V_i(z) = \frac{1}{n} \tilde{c}_{N,i}(z)$ such that $\frac{n}{N} \tilde{m}_{B_n} V_i(z) - \tilde{c}_{N,i}(z) \xrightarrow{N \to \infty} 0$ almost surely is given by (91). The remainder of the proof is equivalent to Appendix I since $V_i$ has uniformly bounded spectral norm w.r.t. $N$.

Under Assumption 1, we have that $\lambda_{\min}(B_n) > \varepsilon$ for some $\varepsilon > 0$ and hence $z = 0$ is valid.

### Appendix III

#### Proof of Theorem 2

The strategy is as follows: The SINR $\gamma_{k,ref}$ in (12) consists of three terms, (i) the signal power $|h_k^H \hat{W} h_k|^2$, (ii) the interference power $h_k^H \hat{W} h_k$ and (iii) the term $\tilde{L}$ of the power normalization. For each of these three terms we will subsequently derive a deterministic equivalent which together constitute the final expression for $\gamma_{k,ref}$.

**A. Deterministic equivalent for $\Psi$**

The term $\Psi = tr \hat{P} \hat{H}^H \hat{H} + M \alpha I_M$ can be written as

$$
\Psi = \sum_{k=1}^K \frac{1}{M} \text{tr} \frac{h_k^H}{1 + \frac{1}{n} \text{tr} h_k^H} \hat{H}^H \hat{H} + M \alpha I_M
$$

where $C_{[k]} \triangleq \Gamma_{[k]} + \alpha I_M$ with $\Gamma_{[k]} \triangleq \frac{1}{M} \hat{H}_k^H \hat{H}_k$ and in (a) we applied Lemma 2 twice together with (5). For $M$ large, we apply Lemma 4 and obtain

$$
\Psi - \frac{1}{M} \sum_{k=1}^K \frac{1}{M} m_{[k]} \Theta_{[k]} \left( \frac{\tilde{c}_{[k]}^o}{\tilde{c}_{[k]}^o} \right)^2 M \to \infty 0
$$

as well. Where $\Psi = \frac{1}{M} \sum_{k=1}^K m_{[k]} \Theta_{[k]} \left( \frac{\tilde{c}_{[k]}^o}{\tilde{c}_{[k]}^o} \right)^2 M \to \infty 0$, almost surely, where in (b) we applied Lemma 6, the definition (6) and denoted $m_{[k]} \Theta_{[k]} \left( \frac{\tilde{c}_{[k]}^o}{\tilde{c}_{[k]}^o} \right)$ the derivative of $m_{[k]} \Theta_{[k]} \left( \frac{\tilde{c}_{[k]}^o}{\tilde{c}_{[k]}^o} \right)$ along $z$ at $z = -\alpha$. To obtain a deterministic equivalent $m_{[k]} \Theta_{[k]} \left( \frac{\tilde{c}_{[k]}^o}{\tilde{c}_{[k]}^o} \right)$ and $m_{[k]} \Theta_{[k]} \left( \frac{\tilde{c}_{[k]}^o}{\tilde{c}_{[k]}^o} \right)$, respectively, we apply Theorem 1. With the notations of Theorem 2, we have $m_{[k]} \Theta_{[k]} \left( \frac{\tilde{c}_{[k]}^o}{\tilde{c}_{[k]}^o} \right) = \frac{1}{M} \text{tr} \hat{H}_k^H \hat{H}_k$ and therefore

$$
\Psi - \frac{1}{M} \sum_{k=1}^K m_{[k]} \Theta_{[k]} \left( \frac{\tilde{c}_{[k]}^o}{\tilde{c}_{[k]}^o} \right)^2 = \frac{1}{M} \sum_{k=1}^K \frac{1}{M} m_{[k]} \Theta_{[k]} \left( \frac{\tilde{c}_{[k]}^o}{\tilde{c}_{[k]}^o} \right) M \to \infty 0,
$$

satisfies $\Psi - \Psi^o M \to \infty 0$, almost surely.

**B. Deterministic equivalent for $h_k^H \hat{W}_k$**

Similar to the derivations in (98) and (99), we have

$$
\frac{n}{N} \tilde{m}_{B_n} V_i(z) = \frac{1}{n} \tilde{c}_{N,i}(z) \xrightarrow{N \to \infty} 0
$$

A deterministic equivalent $\tilde{c}_{N,i}(z)$ of $\frac{n}{N} \tilde{m}_{B_n} V_i(z) = \frac{1}{n} \tilde{c}_{N,i}(z)$ such that $\frac{n}{N} \tilde{m}_{B_n} V_i(z) - \tilde{c}_{N,i}(z) \xrightarrow{N \to \infty} 0$ almost surely is given by (91). Since $\Psi$ and $z_k$ are independent, we apply Lemma 5 and obtain

$$
\Psi - \frac{1}{M} \sum_{k=1}^K \frac{1}{M} m_{[k]} \Theta_{[k]} \left( \frac{\tilde{c}_{[k]}^o}{\tilde{c}_{[k]}^o} \right) M \to \infty 0,
$$

almost surely.

**C. Deterministic equivalent of $h_k^H \hat{W}_k^H \hat{P}_k^H \hat{H}_k^H \hat{W}_k$**

With (4) and $C \triangleq \Gamma + \alpha I_M$, $\Gamma \triangleq \frac{1}{M} \hat{H}_k^H \hat{H}_k$, we have

$$
\frac{n}{N} \tilde{m}_{B_n} V_i(z) = \frac{1}{n} \tilde{c}_{N,i}(z) \xrightarrow{N \to \infty} 0
$$

A deterministic equivalent $\tilde{c}_{N,i}(z)$ of $\frac{n}{N} \tilde{m}_{B_n} V_i(z) = \frac{1}{n} \tilde{c}_{N,i}(z)$ such that $\frac{n}{N} \tilde{m}_{B_n} V_i(z) - \tilde{c}_{N,i}(z) \xrightarrow{N \to \infty} 0$ almost surely is given by (91). Since $\Psi$ and $z_k$ are independent, we apply Lemma 5 and obtain

$$
\Psi - \frac{1}{M} \sum_{k=1}^K \frac{1}{M} m_{[k]} \Theta_{[k]} \left( \frac{\tilde{c}_{[k]}^o}{\tilde{c}_{[k]}^o} \right) M \to \infty 0,
$$

almost surely.
where we denoted $A_k \triangleq \Theta_k^{1/2} C^{-1} \Theta_k^{1/2}$ and $B_k \triangleq \Theta_k^{1/2} C^{-1} H[k] P[k] H[k]^{H} C^{-1} \Theta_k^{1/2}$. Noting that $c_0 + c_1 = 1$ and $c_0 c_1 - c_2 = 0$, we apply Lemma 7 to each of the four quadratic forms in (102) and obtain

$$z_k^H A_k z_k - \frac{u(1 + c_1 u)}{1 + u} M \to 0,$$

$$z_k^H A_k q_k - \frac{-c_2 u^2}{1 + u} M \to 0,$$

$$z_k^H B_k z_k - \frac{u'(1 + c_1 u)}{1 + u} M \to 0,$$

$$e_k^H B_k z_k - \frac{-c_2 u u'}{1 + u} M \to 0,$$

almost surely, where $u = \frac{1}{M} \text{tr} \Theta_k C[k]^{-1}$ and $u' = \frac{1}{M} \text{tr} P[k] H[k] C[k]^{-1} \Theta_k C[k]^{-1} H[k]^{H}$. Substituting the random terms in (102) by their respective deterministic equivalents yields

$$h_k^H \tilde{\mathbf{W}} H[k] P[k] \tilde{\mathbf{H}} H[k]^{H} \tilde{\mathbf{W}} k - \left[ \frac{1}{M} \frac{u'(1 + c_1 u)}{1 + u} \right] M \to 0$$

$$\frac{1}{M} \frac{c_0 (1 + c_1 u)^2 + c_1 c_2 u^2 - 2 c_2 u (1 + c_1 u) u'}{(1 + u)^2} M \to 0$$

(103)

almost surely. The second term in brackets of (103) reduces to $\frac{1}{M} \frac{1 - \tau^2}{(1 + u)^2} u'$ and we obtain

$$h_k^H \tilde{\mathbf{W}} H[k] P[k] \tilde{\mathbf{H}} H[k]^{H} \tilde{\mathbf{W}} k - \left[ \frac{1}{M} \frac{1 - \tau^2}{(1 + u)^2} u' \right] M \to 0$$

(104)

almost surely. From Lemma 6 we have

$$u - m \gamma_k, \Theta_k^{(-\alpha)} M \to 0,$$

$$u' - \gamma_k, \Theta_k^{(-\alpha)} M \to 0,$$

almost surely, where $m \gamma_k, \Theta_k^{(-\alpha)} = \frac{1}{M} \text{tr} \Theta_k C^{-1}$ and $\gamma_k = \frac{1}{M^2} \text{tr} \Theta_k C^{-1} H[k]^{H} C^{-1} H[k]$. Therefore, (104) becomes

$$h_k^H \tilde{\mathbf{W}} H[k] P[k] \tilde{\mathbf{H}} H[k]^{H} \tilde{\mathbf{W}} k - \left( \frac{1 - \tau^2}{(1 + u)^2} u' \right) M \to 0$$

(105)

almost surely. Defining $\tilde{\mathbf{C}} \triangleq \Theta_k^{1/2} \Gamma \Theta_k^{1/2} + \alpha \Theta_k^{1/2}$, we rewrite $\gamma_k$ as

$$\gamma_k = \frac{1}{M} \sum_{j=1}^{K} p_j \hat{z}_j^H \hat{\Theta}_j \hat{z}_j - \gamma_k, \Theta_k^{(-\alpha)} M \to 0$$

Applying Lemmas 2, 4 and 6, we obtain almost surely

$$\gamma_k - \frac{1}{M} \sum_{j=1}^{K} p_j \frac{1}{M} \text{tr} \Theta_k^{1/2} \Theta_j \Theta_k^{1/2} C^{-2} \Theta_k^{1/2} \Theta_j \hat{z}_j \hat{z}_j$$

$$\Rightarrow \gamma_k - \frac{1}{M} \sum_{j=1}^{K} p_j m'_{C, \Theta_k^{(-1/2)} \Theta_k^{(-1/2)}}(0) \to 0$$

where $m'_{C, \Theta_k^{(-1/2)} \Theta_k^{(-1/2)}}(0) = d \frac{1}{M} \text{tr} \Theta_j (\Gamma + \alpha I_M - z \Theta_k)^{-1}$ at $z = 0$. Finally, we apply Theorem 1 and obtain $\gamma_k$ such that $\gamma_k, \Theta_k^{(-\alpha)} M \to 0$, almost surely, as

$$\gamma_k = \frac{1}{M} \sum_{j=1}^{K} p_j \frac{m_{C, \Theta_k^{(-1/2)} \Theta_k^{(-1/2)}}(0)}{1 + c_j ^2}$$

which completes the proof.

APPENDIX IV

PROOF OF THEOREM 3

The SINR $\gamma_k, \alpha$ for ZF of user $k$ is derived as

$$\gamma_k, \alpha = \lim_{\alpha \to 0} \gamma_k, \alpha.$$

Denote $\mathbf{R} \triangleq \tilde{\mathbf{H}}^{H} \tilde{\mathbf{H}} + M \alpha \mathbf{I}_M$, the terms in the SINR of ZRF that depend on $\alpha$ are

$$m_k = \text{tr} \Theta_k \mathbf{R}^{-1},$$

$$\Psi = \text{tr} \tilde{\mathbf{H}}^{H} \tilde{\mathbf{H}} \mathbf{R}^{-2},$$

$$\gamma_k = \text{tr} \mathbf{R}^{-1} \Theta_k \mathbf{R}^{-1} \tilde{\mathbf{H}}^{H} \tilde{\mathbf{H}}.$$

Since $\tilde{\mathbf{H}}^{H} \tilde{\mathbf{H}}$ is singular for $M > K$, we apply the matrix inversion lemma and subsequently take the limit $\alpha \to 0$ of $\Psi$ and $\gamma_k$. We obtain

$$\tilde{\Psi} \triangleq \lim_{\alpha \to 0} \Psi = \text{tr} \tilde{\mathbf{H}}^{H} \tilde{\mathbf{H}} \mathbf{R}^{-1},$$

$$\tilde{\gamma}_k \triangleq \lim_{\alpha \to 0} \gamma_k = \text{tr} \Theta_k \mathbf{R}^{-1} \tilde{\mathbf{H}}^{H} \mathbf{R}^{-1} \tilde{\mathbf{H}}^{H}.$$

At this point, we require Assumption 1 to ensure that the minimum eigenvalue of $\tilde{\mathbf{H}}^{H} \tilde{\mathbf{H}}$ is bounded away from zero for all large $M$, almost surely.

Since $m_k$ grows with $\alpha$ as $O(1/\alpha)$ we have

$$\gamma_k, \alpha = \lim_{\alpha \to 0} \gamma_k, \alpha.$$

We proceed by deriving deterministic equivalents $\tilde{\Psi}$ and $\tilde{\gamma}_k$ for $\Psi$ and $\gamma_k$, respectively.

Define $\Gamma \triangleq \frac{1}{M} \tilde{\mathbf{H}}^{H} \tilde{\mathbf{H}}$, the term $\tilde{\Psi}$ takes the form

$$\tilde{\Psi} = \frac{K}{M} \left[ \frac{1}{K} \text{tr} \right] = \frac{K}{M} m_{\Gamma, \mathbf{P}}(0).$$

A deterministic equivalent $\tilde{\Psi}$ such that $\Psi - \tilde{\Psi} M \to 0$ almost surely, is given by

$$\tilde{\Psi} = \frac{K}{M} m_{\Gamma, \mathbf{P}}(0).$$

The Stieltjes transform $m_{\Gamma, \mathbf{P}}(z)$ is obtained from $m_{\Gamma}(z)$ by the following relation [47, Lemma 3.1]

$$m_{\Gamma}(z) = \frac{M}{K} m_{\Gamma, \mathbf{P}}(z) + \frac{1}{z} \frac{1}{K} - 1$$

$$= \frac{1}{K} \left[ - \text{tr} \mathbf{V} + \text{tr} I_M \right] - \frac{1}{z}$$

(106)
where \( \tilde{V} \triangleq ( \tilde{F} + I_M )^{-1}, \tilde{F} \triangleq \frac{1}{M} \sum_{j=1}^{K} \Theta_j \) and we defined \( \tilde{e}_i(z) \triangleq -z e_i(z) \), which reads
\[
\tilde{e}_i(z) = \frac{1}{M} \text{tr} \Theta_i \tilde{V}.
\]

Applying the Lemma 1 to (106) yields
\[
m^\circ_{\tilde{F}, \tilde{P}}(z) = \frac{1}{z} K \left[ \frac{1}{z} \frac{1}{M} \text{tr} \left( \frac{1}{M} \sum_{j=1}^{K} \Theta_j \tilde{V} \right) - \frac{1}{z} \right] = \frac{1}{z} K \frac{1}{z} \frac{1}{M} \sum_{j=1}^{K} \frac{1}{\tilde{e}_j(z) - \frac{1}{z}} = \frac{1}{K} \sum_{j=1}^{K} \frac{1}{\tilde{e}_j(z) - z}.
\]

(107)

To obtain \( m^\circ_{\tilde{F}, \tilde{P}}(0) \), observe that \( \text{tr} \tilde{P} ( \tilde{H}^H \tilde{H}^{-1} ) = \text{tr} ( \tilde{H}^H \tilde{H}^{-1} ) \)
where the columns \( \tilde{h}_k \) of \( \tilde{H}^H \) are correlated as \( E[ \tilde{h}_k \tilde{h}_k^H ] = \Theta_k / p_k \triangleq \Theta_k \). Substituting \( \tilde{\Theta}_k \) by \( \Theta_k / p_k \) in (107) yields
\[
m^\circ_{\tilde{F}, \tilde{P}}(0) = \frac{1}{K} \sum_{j=1}^{K} \frac{p_j}{\tilde{e}_j(0)}.
\]

Therefore, denoting \( \tilde{e}_j \triangleq \tilde{e}_j(0) \), we obtain
\[
\tilde{\Psi}^\circ = \frac{1}{M} \sum_{j=1}^{K} \frac{p_j}{\tilde{e}_j}.
\]

In order to find \( \tilde{\Theta}_k^\circ \), we first write \( \tilde{\Theta}_k = U_k \text{diag}(\lambda_1, \ldots, \lambda_M) U_k^H \), where \( U_k \) is a unitary matrix. Denote \( \tilde{H}^H \triangleq H U_k \). The \( i \)-th column \( \tilde{h}_i^H \) of \( \tilde{H}^H \) reads
\[
\tilde{h}_i^H = \sqrt{M} \begin{bmatrix} \tilde{h}_1^H \Theta_1^{1/2} \\ \tilde{h}_2^H \Theta_2^{1/2} \\ \vdots \\ \tilde{h}_M^H \Theta_M^{1/2} \end{bmatrix} u_{k,i},
\]
where \( u_{k,i} \) is the \( i \)-th column of \( U_k \). The elements \( \tilde{h}_j, j = 1, \ldots, K, \) of \( \tilde{h}_i^H \) are independent with zero mean and variance \( E[ \tilde{h}_i^H \tilde{h}_j^H ] = u_{k,i}^H \Theta_j u_{k,i} \). Applying Lemma 2 twice, (105) takes the form
\[
\tilde{T}_k = \sum_{i=1}^{M} \lambda_{k,i} \tilde{h}_i^H \tilde{H}^H \tilde{H}^{-1} \tilde{P} \tilde{H}^H \tilde{H}^{-1} \tilde{h}_i^H = \sum_{i=1}^{M} \tilde{h}_i^H \tilde{H}^H \tilde{H}^{-1} \tilde{P} \tilde{H}^H \tilde{H}^{-1} \tilde{h}_i^H.
\]

Denoting \( \tilde{D}_{k,i} \triangleq \text{diag}(u_{k,i}^H \Theta_j u_{k,i}, \ldots, u_{k,i}^H \Theta_K u_{k,i}) \) and applying Lemma 4 and Lemma 6, we obtain
\[
\tilde{T}_k - \frac{1}{M} \sum_{i=1}^{M} \lambda_{k,i} \frac{1}{1 + \frac{1}{1 + M} \text{tr} \tilde{D}_{k,i} \tilde{P}^{-1}} M \rightarrow \infty 0
\]
\[
\Leftrightarrow \tilde{T}_k - \frac{1}{M} \sum_{i=1}^{M} \lambda_{k,i} \frac{1}{1 + M} \text{tr} \tilde{D}_{k,i} \tilde{P}^{-1} M \rightarrow \infty 0,
\]

almost surely, where we defined \( \tilde{T}_k \triangleq \frac{1}{M} \tilde{H}^H \tilde{H} \) and \( \tilde{D}_{k,i} \triangleq \text{diag}(u_{k,i}^H \Theta_j u_{k,i}, \ldots, u_{k,i}^H \Theta_K u_{k,i}) \). A deterministic equivalent \( m^\circ_{\tilde{T}_k, \tilde{P}}(0) = \frac{1}{M} \text{tr} \tilde{D}_{k,i} \tilde{P}^{-1} \) such that \( m^\circ_{\tilde{T}_k, \tilde{P}}(0) = m^\circ_{\tilde{T}_k, \tilde{P}}(0) M \rightarrow \infty 0 \) almost surely, is given by
\[
m^\circ_{\tilde{T}_k, \tilde{P}}(0) = \frac{1}{K} \sum_{j=1}^{K} \frac{u_{k,i}^H \Theta_j u_{k,i}}{\tilde{e}_j(0)}.
\]

Note that \( m^\circ_{\tilde{T}_k, \tilde{P}}(z) = \frac{1}{M} \text{tr} \tilde{D}_{k,i} \tilde{P}(\tilde{P}^{-1} - \tilde{P}^{-1} z P) \) and thus a deterministic equivalent \( m^\circ_{\tilde{T}_k, \tilde{P}}(0) = m^\circ_{\tilde{T}_k, \tilde{P}}(0) = \frac{d}{dz} \frac{1}{M} \text{tr} \tilde{D}_{k,i} \tilde{P}(\tilde{P}^{-1} - \tilde{P}^{-1} z P) \mid_{z=0} \) is obtained by applying Theorem 10 and is given by (19).

Finally, substituting \( \frac{1}{M} \text{tr} \tilde{D}_{k,i} \tilde{P}^{-1} \) by \( \frac{1}{M} \text{tr} \tilde{D}_{k,i} \tilde{P}^{-2} \) by their respective deterministic equivalents \( m^\circ_{\tilde{T}_k, \tilde{P}}(0) \) and \( m^\circ_{\tilde{T}_k, \tilde{P}}(0) \), we obtain \( \tilde{T}_k^\circ \) given in (18), such that \( \tilde{T}_k^\circ - \tilde{T}_k^\circ \rightarrow M \rightarrow \infty 0 \), almost surely, which completes the proof.

APPENDIX V

PROOF OF PROPOSITION 2

The proof follows the lines of [21] with adaptations to account for imperfect CSIT. From Corollary 1, the SINR \( \gamma^\circ_{\text{trf}} \) can be written under the form
\[
\gamma^\circ_{\text{trf}} = \rho \beta m^\circ(1 - \tau^2) \Gamma,
\]
where
\[
\Gamma = \frac{\beta e_2 + \alpha (1 + m^\circ) e_1}{\rho e_2(1 - \tau^2) + \tau^2 \rho(1 + m^\circ) e_2 + (1 + m^\circ)^2 e_1}
\]
with \( e_1 \) and \( e_2 \) defined in (15) and (16), respectively. Taking the derivative along \( \alpha \), we obtain
\[
\frac{\partial \gamma^\circ_{\text{trf}}}{\partial \alpha} = \rho \beta m^\circ(1 - \tau^2) \Gamma \left[ \frac{m^\circ}{m^\circ + \Gamma'} \right]
\]
\[
= \rho \beta m^\circ(1 - \tau^2) \Gamma \times \left[ \frac{2(1 + m^\circ) m^\circ e_1 + \alpha (1 + m^\circ)^2 e_1 + \beta e_2}{\beta e_2 + \alpha (1 + m^\circ)^2 e_1} \right]
\]
\[
- \left[ \frac{1 - \tau^2 + \tau^2 (1 + m^\circ)^2 \rho e_2 + 2 \tau^2 \rho (1 + m^\circ) m^\circ e_2}{1 - \tau^2 + \tau^2 (1 + m^\circ)^2 \rho e_2 + (1 + m^\circ)^2 e_1} \right]
\]
\[
+ \left[ \frac{2(1 + m^\circ) m^\circ e_1 + (1 + m^\circ)^2 e_1}{1 - \tau^2 + \tau^2 (1 + m^\circ)^2 \rho e_2 + (1 + m^\circ)^2 e_1} \right],
\]

(108)

where we used the fact that
\[
m^\circ = \frac{(1 + m^\circ)^2 e_1}{\beta e_2 + \alpha (1 + m^\circ)^2 e_1}.
\]
Denoting $\chi \triangleq (1+m^o)^2 e_1$, $\psi \triangleq 2(1+m^o)m^o e_1 + (1+m^o)^2 e'_1$ and $\phi \triangleq 1 - \tau^2 + \tau^2 (1+m^o)^2$, (108) takes the form
\[
\frac{\partial \gamma_{\text{ref}}}{\partial \alpha} = \rho \beta m^o (1-\tau^2) \Gamma \\
\times \left[ \frac{\left(1 + \alpha \psi - \rho \phi e_1 + \chi + 2\tau^2 \rho (1+m^o) m^o e_2 \right)}{\phi e_1 + \chi} \right] \\
= \rho^2 \beta m^o (1-\tau^2) \Gamma \left[ \frac{\left( \alpha - \frac{1}{2\beta \phi} \right) (e_2 \psi - e'_2 \chi)}{\phi} \right] \\
- 2\tau^2 (1+m^o) m^o e_2 \left( e_2 \psi - e'_2 \chi \right),
\]

where $Z = (\frac{1}{2} e_2 + \alpha \chi)(\rho \phi e_1 + \chi)$. From [21] we have that $e_2 \psi - e'_2 \chi = 2(1+m^o) m^o e_2$ and thus
\[
\frac{\partial \gamma_{\text{ref}}}{\partial \alpha} = \frac{2\rho^2 \beta m^o (1-\tau^2) (1+m^o) m^o e_2 \Gamma}{Z} \\
\times \left[ \frac{\left( \alpha - \frac{1}{2\beta \phi} \right) T (e_2 \psi - e'_2 \chi)}{\phi} \right]. \quad (109)
\]

Denoting $\Omega \triangleq 2\rho^2 \beta m^o (1-\tau^2) (1+m^o) m^o e_2 \Gamma / Z$ and rewriting the term in brackets in (109), we obtain
\[
\frac{\partial \gamma_{\text{ref}}}{\partial \alpha} = \Omega \left[ \alpha - \frac{1 + \tau^2 \rho \eta(\alpha)}{(1-\tau^2) \rho \beta} \right] = 0,
\]

where $\eta(\alpha)$ is given in (31). For $\rho \in (0, \infty)$, the term $\Omega > 0$ and therefore the optimal regularization parameter $\alpha^*$ is given by (30). The fixed-point equation (30) has always a solution since the right-hand side of (30) is bounded from below and from above. More precisely, we have $\lim_{\alpha \to -\infty} \eta(\alpha) = \text{tr} \Theta^2 / \text{tr} \Theta$ and $\lim_{\alpha \to 0} \eta(\alpha) = M / \text{tr} \Theta^{-1} > 0$, which completes the proof.

**APPENDIX VI**

**DETAILS ON HIGH SNR LIMITS IN SECTION VI**

**A. Channel Distortion Aware Regularized Zero-forcing Precoding**

For $\beta = 1$ observe that $g(1, \beta)$ scales as $2\sqrt{\rho}$. Thus, for $\rho \to \infty$, (45) converges to $b^2 - 1$.

If $\beta > 1$, the term $g(1, \beta)$ takes the form
\[
g(1, \beta) = (\beta - 1)\rho + \sqrt{1 - \beta^2} \rho (1 + o(1)) \to \infty 2\rho (\beta - 1).
\]

Therefore, for $\rho \to \infty$, (45) converges to $b - 1$.

**B. Channel Distortion Unaware Regularized Zero-forcing Precoding**

For $\beta = 1$ and $\rho$ large, both terms $m^o$ and $\gamma^*$ scale as $\sqrt{\rho}$, whereas $\bar{m}$ scales as $\rho$. Therefore, $\lim_{\rho \to \infty} \phi_{\text{ref}}(\rho, b) = 2(b - 1)$. If $\beta > 1$, for large $\rho$, the terms $m^o$ and $\gamma^*$ scale as $\rho (\beta - 1)$ and $\bar{m}$ scales as $\rho^2 (\beta - 1)^2$. Note that $\gamma^*$ converges to the SINR of ZF precoding (23), for $\tau^2 = 0$ and $\rho \to \infty$. With this approximation we obtain $\lim_{\rho \to \infty} \phi_{\text{ref}}(\rho, b) = b - 1$.

**APPENDIX VII**

**PROOF OF PROPOSITION 4**

The sum rate $R_{\text{sum}}^o$ can be written as a function of the per-user rate under perfect CSIT $\bar{R}^o$ and the per-user rate gap $\Delta R^o$ as
\[
R_{\text{sum}}^o = K \left( 1 - \frac{T_i}{T} \right) [\bar{R}^o - \Delta R^o],
\]
where for ZF and RZF-CDA we have $R_{\text{zf}}^o = \log(1 + \rho_d(\beta - 1))$ and $R_{\text{zf}}^o = \frac{1}{2} + \frac{1}{2} \rho_d(\beta - 1) + \frac{d(1)}{T}$, respectively, and
\[
\Delta R_{\text{zf}}^o = \log \left( \frac{(\beta - 1)(\rho_d + 1)}{1 + \frac{1}{\rho_d} + T_{\text{zf}}[\bar{R}^o + \rho_d(\beta - 1)]} \right)
\]
\[
\Delta R_{\text{zf}}^o = \log \left( \frac{1 + \rho_d(\beta - 1) + d(1)}{1 + w_d(\beta - 1) + d(w)} \right),
\]

where $d(w)$ is defined in (58). Denoting $\psi \triangleq 1 + \frac{1}{\rho_d} + T_{\text{zf}}[1 + \rho_u(\beta - 1)]$, the derivatives take the form
\[
\frac{\partial R_{\text{sum}}^o}{\partial T_{\text{zf}}} = -\frac{K}{T} \Delta R_{\text{zf}}^o + K \left( 1 - \frac{T_{\text{zf}}}{T} \right) \left( \frac{(\beta - 1)(\rho_d + 1) + \rho_u(\beta - 1)}{\psi^2(\beta - 1)} \right)
\]
\[
\frac{\partial R_{\text{sum}}^o}{\partial T_{\text{zf}}} = -\frac{K}{T} \Delta R_{\text{zf}}^o + K \left( 1 - \frac{T_{\text{zf}}}{T} \right) \frac{w_d(\beta - 1) + d'}{1 + w_d(\beta - 1) + d'}, \quad (111)
\]

where $w' = \partial w / \partial T_{\text{zf}} = (1/\rho_u + c)/T_{\text{zf}} + (1/\rho_u + c)^2$ and $d' = \partial d / \partial T_{\text{zf}} = (1/\beta - 1)^2 w_d(\beta - 1) + (1/\beta - 1)/d$. In (110) and (111) the per-user rate-gap $\Delta R_{\text{zf}}^o$ and $\Delta R_{\text{zf}}^o$ can be neglected since $\Delta R_{\text{zf}}^o \ll \bar{R}^o$ and $\Delta R_{\text{zf}}^o \ll \bar{R}^o$, respectively. For $\rho_d, \rho_u \to \infty$ and $c = \rho_d / \rho_u$ finite, solving (110) and (111) for $T_{\text{zf}}$ and $T_{\text{zf}}$, respectively, yields (61) and (62), respectively, which completes the proof.

**APPENDIX VIII**

**IMPORTANT LEMMAS**

**Lemma 1** (Resolvent Identity): Let $U$ and $V$ be two invertible complex matrices of size $N \times N$. Then
\[
U^{-1} - V^{-1} = -U^{-1}(U - V)V^{-1}.
\]

**Lemma 2** (Matrix Inversion Lemma): [31, Lemma 2.2] Let $U$ be an $N \times N$ invertible matrix and $x \in \mathbb{C}^N$, $c \in \mathbb{C}$ for which $U + cx^H$ is invertible. Then
\[
x^H (U + cx^H)^{-1} = \frac{x^H U^{-1} - x^H}{1 + cx^H U^{-1} x}.
\]

**Lemma 3**: [48, Lemma B.26] Let $A \in \mathbb{C}^{N \times N}$ be a deterministic matrix and $x \in \mathbb{C}^N$ have i.i.d. complex entries of zero mean, variance $1/N$ and bounded $l^1$-th order moment $\|x\| \leq \nu_l$. Then for any $p \geq 1$
\[
\|Ax\|_p \leq \frac{C_p}{N^{p/2}} \left( \frac{1}{N} \text{tr} A A^H \right)^{p/2} \left[ \nu_1^{p/2} + \nu_2 p \right], \quad (112)
\]
where $C_p$ is a constant solely depending on $p$. 
Lemma 4: [47, Lemma 14.2] Let $A_1, A_2, \ldots, A_N \in \mathbb{C}^{N \times N}$ be a series of random matrices generated by the probability space $(\Omega, \mathcal{F}, P)$ such that, for $\omega \in A \subset \Omega$, with $P(A) = 1$, $\|A_N(\omega)\| < K(\omega) < \infty$, uniformly on $N$. Let $x_1, x_2, \ldots$, with $x_N \in \mathbb{C}^N$, be random vectors of i.i.d. entries with zero mean, variance $1/N$ and eighth order moment of order $O(1/N^4)$, independent of $A_N$. Then
\[
x_N^H A_N x_N - \frac{1}{N} \text{tr} A_N \xrightarrow{N \to \infty} 0,
\]
almost surely.

Proof: The proof unfolds from a direct application of the Tonelli theorem, [33, Theorem 18.3]. Denoting $(X, Y, P_X)$ the probability space that generates the series $x_1, x_2, \ldots$, we have that for every $\omega \in A$ (i.e., for every realization $A_1(\omega), A_2(\omega), \ldots,$), the trace lemma, [47, Theorem 3.4], holds true. From [33, Theorem 18.3], the space $B$ of couples $(x, y) \in Y \triangleq X \times \Omega$ for which the trace lemma holds, satisfies
\[
\int_Y 1_B(x, y) dP_Y(x, y) = \int_X \int_X 1_B(x, y) dP_X(x) dP_\Omega(\omega).
\]
If $\omega \in A$, then $1_B(x, y) = 1$ on a subset of $X$ of probability one. The inner integral therefore equals one whenever $\omega \in A$. As for the outer integral, since $P(A) = 1$, it also equals one, and the result is proved.

Lemma 5: Let $A_N$ be as in Lemma 4 and $x_N, y_N \in \mathbb{C}^N$ be random, mutually independent with standard i.i.d. entries of zero mean, variance $1/N$ and eighth order moment of order $O(1/N^4)$, independent of $A_N$.
\[
y_N^H A_N x_N \xrightarrow{N \to \infty} 0,
\]
amost surely.

Proof: Remark that $E |y_N^H A_N x_N|^4 < c/N^2$ for some constant $c > 0$ independent of $N$. The result then unfolds from the Markov inequality the Borel-Cantelli Lemma [33] and the Tonelli theorem [33, Theorem 18.3].

Lemma 6: [47, Lemma 14.3] Let $A_1, A_2, \ldots, A_N \in \mathbb{C}^{N \times N}$, be deterministic with uniformly bounded spectral norm and $B_1, B_2, \ldots, B_N \in \mathbb{C}^{N \times N}$, be random Hermitian, with eigenvalues $\lambda_{1}^{B_N} \leq \cdots \leq \lambda_{N}^{B_N}$ such that, with probability one, there exist $\varepsilon > 0$ for which $\lambda_{i}^{B_N} > \varepsilon$ for all large $N$. Then for $v \in \mathbb{C}^N$
\[
\frac{1}{N} \text{tr} A_N B_N^{-1} - 1 - \frac{1}{N} \text{tr} A_N (B_N + vv^H)^{-1} \xrightarrow{N \to \infty} 0
\]
amost surely, where $B_N^{-1}$ and $(B_N + vv^H)^{-1}$ exist with probability one.

Proof: The proof unfolds similarly as above, with some particular care to be taken. For $\omega \in B$, the smallest eigenvalue of $B_N(\omega)$ is uniformly greater than $\varepsilon(\omega)$. Therefore, with $B_N(\omega)$ and $B_N(\omega) + vv^H$ invertible and, taking $z = -\varepsilon(\omega)/2$, we can write
\[
\frac{1}{N} \text{tr} A_N B_N^{-1}(\omega)
\]
\[
= \frac{1}{N} \text{tr} A_N \left( B_N(\omega) - \frac{\varepsilon(\omega)}{2} I_N + \frac{\varepsilon(\omega)}{2} I_N \right)^{-1}
\]
and
\[
\frac{1}{N} \text{tr} A_N \left( B_N(\omega) + vv^H \right)^{-1}
\]
\[
= \frac{1}{N} \text{tr} A_N \left( \left[ B_N(\omega) + \frac{\varepsilon(\omega)}{2} I_N + \frac{\varepsilon(\omega)}{2} I_N \right]^{-1}
\]
Under these notations, $B_N(\omega) - \varepsilon(\omega)/2I_N$ and $B_N(\omega) + vv^H - \varepsilon(\omega)/2I_N$ are still nonnegative definite for all $N$. Therefore, the rank-1 perturbation lemma, [49, Lemma 2.1], can be applied for this $\omega$. But then, from the Tonelli theorem again, in the space that generates the couples $((x_1, x_2, \ldots), (B_1, B_2, \ldots))$, the subspace where the rank-1 perturbation lemma applies has probability one, which is what needed to be proved.

Lemma 7: Let $U, V \in \mathbb{C}^{N \times N}$ be invertible and of uniformly bounded spectral norm. Let $x, y \in \mathbb{C}^N$ have i.i.d. complex entries of zero mean, variance $1/N$ and finite 8th order moment and be mutually independent as well as independent of $U, V$. Define $c_0, c_1, c_2 \in \mathbb{R}^+$ such that $c_0c_1 - c_2^2 \geq 0$ and let $u \xrightarrow{N \to \infty} 0$ in $U$ and $u' \xrightarrow{N \to \infty} 0$ in $V$. Then we have
\[
x_N^H U (V + c_0xx^H + c_1yy^H + c_2xy^H + c_2yx^H)^{-1} x
\]
\[
- \frac{u'(1 + c_1u)}{(c_0c_1 - c_2^2)u^2 + (c_0 + c_1)u + 1} \xrightarrow{N \to \infty} 0
\]
amost surely. Furthermore
\[
x_N^H U (V + c_0xx^H + c_1yy^H + c_2xy^H + c_2yx^H)^{-1} y
\]
\[
- \frac{-c_2uu'}{(c_0c_1 - c_2^2)u^2 + (c_0 + c_1)u + 1} \xrightarrow{N \to \infty} 0
\]
amost surely.

Proof: Denote $V = (A + c_0xx^H + c_1yy^H + c_2xy^H + c_2yx^H)^{-1}$. Now $x_N^H U x$ can be resolved using Lemma 1
\[
x_N^H U x - x_N^H U A^{-1} x = x_N^H U (V^{-1} - A) A^{-1} x
\]
\[
= -x_N^H U (c_0xx^H + c_1yy^H + c_2xy^H + c_2yx^H) A^{-1} x.
\]

Equation (113) can be rewritten as
\[
x_N^H U x =
\]
\[
x_N^H U (A + c_0xx^H + c_1yy^H + c_2xy^H + c_2yx^H)^{-1}
\]
\[
= 1 + c_0xx^H A^{-1} x + c_2yx^H A^{-1} x
\]
Similarly to (113), we apply Lemma 1 to $x_N^H U y$. Thus, we obtain an expression involving the terms $x_N^H U A^{-1} x$, $y_N^H A^{-1} y$, $x_N^H U A^{-1} x$ and $y_N^H A^{-1} x$. To complete the proof we apply Lemma 4 and Lemma 5, with $u = \frac{1}{N} \text{tr} A^{-1}$ and $u' = \frac{1}{N} \text{tr} U^{-1}$ and we have
\[
x_N^H U y - \frac{u'(1 + c_1u)}{(c_0c_1 - c_2^2)u^2 + (c_0 + c_1)u + 1} \xrightarrow{N \to \infty} 0,
\]
amost surely. Similarly we have
\[
x_N^H U y - \frac{-c_2uu'}{(c_0c_1 - c_2^2)u^2 + (c_0 + c_1)u + 1} \xrightarrow{N \to \infty} 0,
\]
amost surely. Note that as $c_0, c_1, c_2 \in \mathbb{R}^+$ and $c_0c_1 \geq c_2^2$, equations (114) and (115) hold since then $(c_0c_1 - c_2^2)u^2 +


