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

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

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 systems 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 while their ratio remains bounded. We derive deterministic approximations of the empirical signal-to-interference plus noise ratio (SINR) at the receivers, which are tight as $M, K \to \infty$. In the course of this derivation, the per-user channel correlation model requires the development of a novel deterministic equivalent of the empirical Stieltjes transform of large dimensional random matrices with generalized variance profile. The deterministic SINR approximations enable us to solve various practical optimization problems. Under sum rate maximization, we derive (i) for RZF the optimal regularization parameter, (ii) for ZF the optimal number of users, (iii) for ZF and RZF the optimal power allocation scheme and (iv) 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 (MU) scenario, the inherent problems of SU-MIMO transmission can largely be overcome by exploiting multi-user 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]. However, the DPC precoder is non-linear and to this day too complex to be implemented efficiently in practical systems. 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.

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].

Although [9], [12], [13] assume perfect channel state information at the transmitter (CSIT) to determine theoretically optimal performance, this assumption is untenable in practice. It is indeed a particularly strong assumption, since the performance of all precoding strategies is crucially depending on the CSIT quality. 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 [14] for a recent survey.

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., we do not account for user scheduling, and consider ZF and regularized ZF (RZF) precoding under imperfect CSIT (modeled as a weighted sum of the true channel plus noise) as well as per-user channel correlation, i.e., the vector channel $\mathbf{h}_k \in \mathbb{C}^M$ of user k ($k = 1, \ldots, K$) satisfies $E[\mathbf{h}_k] = 0$ and $E[\mathbf{h}_k \mathbf{h}_k^{\mathsf{H}}] = \mathbf{\Theta}_k$. To obtain insights into the system behavior, we approximate the signal-to-interference plus noise ratio (SINR) by a deterministic quantity, where the novelty of this

S. Wagner is with ST-ERICSSON, Sophia-Antipolis

R. Couillet is with Supélec, Gif sur Yvette, France, e-mail:romain.couillet@supelec.fr.

S. Wagner and D.T.M. Slock are with EURECOM, Sophia-Antipolis, 06904, Route des Crêtes, B.P. 193, France, e-mail:{sebastian.wagner, dirk.slock}@eurecom.fr.

M. Debbah is with the Alcatel-Lucent Chair on Flexible Radio, SUPÉLEC, Gif sur Yvette, France e-mail: merouane.debbah@supelec.fr.

study lies in the *large* system approach. More precisely, we approximate the SINR γ_k of user k by a *deterministic equivalent* γ_k° such that $\gamma_k - \gamma_k^\circ \to 0$ almost surely, as the system dimensions M and K go jointly to infinity with bounded ratio $1 \leq \lim_{M,K\to\infty} \frac{M}{K} = \beta < \infty$. Hence, γ_k° becomes more accurate for increasing M, K. To derive γ_k°, we apply tools from the well-established field of large dimensional random matrix theory (RMT) [15], [16]. Previous work considered SINR approximations based on *bounds* on the average (with respect to the random channels $\mathbf{h}_k$) SINR. The deterministic equivalent γ_k° is not a bound but is a *tight* approximation, for asymptotically large M, K. Furthermore, the RMT tools allow us to consider advanced channel models like the per-user correlation model, which are usually extremely difficult to study exactly for finite dimensions. Interestingly, simulations suggest that γ_k° is very accurate even for small system dimension, e.g., $M = K = 16$. Currently, the 3GPP LTE-Advanced standard [17] already defines up to $M = 8$ transmit antennas further motivating the application of large system approximations to characterize the performance of wireless communication systems. Subsequently, we apply these SINR approximations to various practical optimization problems.

A. Related Literature

To the best of the authors' knowledge, Hochwald et al. [18] were the first to carry out a large system analysis with $M, K \to \infty$ and finite ratio for linear precoding under the notion of "channel hardening". In particular, they considered ZF precoding, called channel inversion (CI), for $M > K$ under perfect CSIT, and showed that the SINR for independent and identical distributed (i.i.d.) Gaussian channels converges to $\rho(\beta - 1)$, where ρ is the signal-to-noise ratio (SNR), independent of the applied power normalization strategy. They go on to derive the sum rate maximizing system loading $\beta^{\star\circ}$ for a fixed M. Their results are a special case of our analysis in Section III-B and Section V-A. The authors in [18] conclude by showing that for $\beta > 1$, ZF achieves a large fraction of the linear (with respect to K) sum rate growth. The work in [9] extends the analysis in [18] to the case $M = K$ and shows that the sum rate of ZF is constant in M as $M, K \to \infty$, i.e., the linear sum rate growth is lost. The authors in [9] counter this problem by introducing a regularization parameter α in the inverse of the channel matrix. Under the assumption of large M, K, perfect CSIT and for any rotationally-invariant channel distribution, [9] derives the regularization parameter $\alpha = \alpha^{\star\circ} = \frac{1}{\beta\rho}$ that maximizes the SINR. Note here that [9] does not apply the classic tools from large dimensional RMT to derive their results but rather find the solution by applying various expectations and approximations. In the present contribution, the RZF precoder of [9] is referred to as *channel distortion-unaware* RZF (RZF-CDU) precoder, since its design assumes perfect CSIT, although in practice, the available CSIT is erroneous or distorted. It has been observed in [9] that the RZF-CDU precoder is very similar to the transmit filter derived under the minimum mean square error (MMSE) criterion [19] and both become identical in the large M, K limit. Likewise, we will observe some similarities between RZF and MMSE filters when considering imperfect CSIT. The RZF precoder in [9] has been extended in [20] to account for channel quantization feedback under random vector quantization (RVQ). The authors in [20] do not apply tools from large RMT but use the same techniques as in [9] and obtain different results for the optimal regularization parameter and SINR compared to our results in Section VI.

The first work applying tools from large RMT to derive the asymptotic SINR under ZF and RZF precoding for correlated channels was [21]. However, in [21] the regularization parameter of the considered RZF precoder was set to fulfill the total average power constraint. Similar work [22] was published later, where the authors considered the RZF precoder in [9] and derived the asymptotic SINR for *uncorrelated* Gaussian channels. Moreover, they derived the asymptotically optimal regularization parameter $\alpha^{\star\circ} = \frac{1}{\beta\rho}$, already derived in [9], which is a special case of the result derived in Section IV. Another work [23], reproducing our results, noticed that the optimal regularization parameter in [9], [22] is independent of transmit correlation when the channel correlation is identical for all users.

In the large system limit and for channels with i.i.d. entries, the cross correlations between the user channels, and therefore the users' SINRs, are identical. It has been shown in [24] 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 [25] claimed that indeed the RZF precoder structure emerges as the optimal precoding solution for $M, K \to \infty$. This asymptotic optimality further motivates a detailed analysis of the RZF precoder for large system dimensions.

B. Contributions of the Present Work

In this paper, we provide a concise framework that directly extends and generalizes the results in [9], [18], [22], [23], [26] by accounting for per-user correlation and imperfect CSIT. Furthermore, we apply our SINR approximations to several limited-feedback scenarios that have been previously analyzed by applying bounds on the ergodic rate of finite dimensional systems. Our main contributions are summarized as follows:

- Motivated by the channel model, we derive a deterministic equivalent of the empirical Stieltjes transform of matrices with *generalized variance profile*, thereby extending the results in [27], [28].
- We propose deterministic equivalents for the SINR of ZF ($\beta > 1$) and RZF ($\beta \geq 1$) precoding under imperfect CSIT and channel with per-user correlation, i.e., deterministic approximations of the SINR, which are independent of the individual channel realizations, and (almost surely) exact as $M, K \to \infty$.
- Under imperfect CSIT and *common* correlation ($\mathbf{\Theta}_k = \mathbf{\Theta}\ \forall k$), we derive the sum rate maximizing RZF precoder called *channel-distortion aware* RZF (RZF-CDA) precoder.
- For ZF and RZF, under common correlation and *different* CSIT qualities, we derive the optimal power allocation scheme which is the solution of a water-filling algorithm.

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 maximizing 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 ρ, to exactly maintain an instantaneous per-user rate gap of $\log_2 b$ bits/s/Hz, almost surely as $M, K \to \infty$, 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/ZF: $B=(M-1)\log_2 \rho-(M-1)\log_2 2(b-1)$

  That is, the RZF-CDA precoder requires $(M-1)\log_2 \frac{b+1}{2}$ 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 ρ_{dl}, the sum rate maximizing amount of channel training scales as $\sqrt{T}$ and $1/\sqrt{\log(\rho_{dl})}$ for a fixed ρ_{dl} and T, respectively under both RZF-CDA and ZF precoding.

The remainder of the paper is organized as follows. Section II presents the transmission 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 users for ZF precoding and the optimal power allocation when the CSIT quality of the users is unequal. 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.

Most technical poofs are presented in the appendix. In these proofs, we apply several lemmas collected in Appendix VI.

Notation: In the following, boldface lower-case and upper-case characters denote vectors and matrices, respectively. The operators $(\cdot)^{\mathsf{H}}$, $\mathrm{tr}(\cdot)$ and $E[\cdot]$ denote conjugate transpose, trace and expectation, respectively. The $N\times N$ identity matrix is denoted $\mathbf{I}_N$, $\log(\cdot)$ is the natural logarithm and $\Im(z)$ is the imaginary part of $z\in\mathbb{C}$. $\|\mathbf{X}\|$ and $\lambda_{\min}(\mathbf{X})$ are the spectral radius and the minimum eigenvalue of the Hermitian matrix $\mathbf{X}$, respectively. The imaginary unit is denoted $\mathbf{i}$. The sets $\mathbb{R}^+$ and $\mathbb{C}^+$ are defined as $\{x : x>0\}$ and $\{x=r+\mathbf{i}v : r\in\mathbb{R}, v>0\}$. A random vector $\mathbf{x}\sim\mathcal{CN}(\mathbf{m},\mathbf{\Theta})$ is complex Gaussian distributed with mean vector $\mathbf{m}$ and covariance matrix $\mathbf{\Theta}$.

II. System Model

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

A. Transmission Model

Consider a MISO broadcast channel composed of a central transmitter equipped with M antennas and of K single-antenna non-cooperative receivers. We assume $M\geq K$, thus user scheduling is not taken into account. Furthermore, we suppose narrow-band transmission. The signal y_k received by user k at any time instant reads

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

where $\mathbf{h}_k\in\mathbb{C}^M$ is the random channel from the transmitter to user k, $\mathbf{x}\in\mathbb{C}^M$ is the transmit vector and the noise terms $n_k\sim\mathcal{CN}(0,\sigma^2)$ are independent. We assume that the channel $\mathbf{h}_k$ evolves according to a block-fading model, i.e., the channel is constant at every time instant but varies *independently* from one time instant to another.

The transmit vector $\mathbf{x}$ is a linear combination of the independent user symbols s_k and can be written as

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

where $\mathbf{g}_k\in\mathbb{C}^M$ and $p_k\geq 0$ are the precoding vector and the signal power of user k, respectively. Subsequently, we assume that user k has perfect knowledge of $\mathbf{h}_k$ and the effective channel $\mathbf{h}_k^{\mathsf{H}}\mathbf{g}_k$. In particular, an estimate of $\mathbf{h}_k^{\mathsf{H}}\mathbf{g}_k$ can be obtained through dedicated downlink training by precoding the pilots of user k by $\mathbf{g}_k$. The precoding vectors are normalized to satisfy the *average* total power constraint

$$E[\|\mathbf{x}\|^2]=\mathrm{tr}(\mathbf{P}\mathbf{G}^{\mathsf{H}}\mathbf{G})\leq P, \tag{1}$$

where $\mathbf{G}\triangleq[\mathbf{g}_1,\mathbf{g}_2,\ldots,\mathbf{g}_K]\in\mathbb{C}^{M\times K}$, $\mathbf{P}=\mathrm{diag}(p_1,\ldots,p_K)$ and $P>0$ is the total available transmit power.

Denote $\rho\triangleq P/\sigma^2$ the SNR. Under the assumption of Gaussian signaling, i.e., $s_k\sim\mathcal{CN}(0,1)$ and single-user decoding with perfect channel state information at the receivers, the SINR γ_k of user k is defined as [29]

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

The rate R_k of user k is given by

$$R_k=\log(1+\gamma_k) \tag{3}$$

and the ergodic sum rate is defined as

$$R_{\mathrm{sum}}=\sum_{k=1}^{K}E[R_k], \tag{4}$$

where the expectation is taken over the random channels $\mathbf{h}_k$.

B. Channel Model

Each user channel $\mathbf{h}_k$ is modeled as

$$\mathbf{h}_k=\sqrt{M}\mathbf{\Theta}_k^{1/2}\mathbf{z}_k, \tag{5}$$

where $\mathbf{\Theta}_k$ is the channel correlation matrix of user k and $\mathbf{z}_k$ has i.i.d. complex entries of zero mean and variance $1/M$. The channel transmit correlation matrices $\mathbf{\Theta}_k$ 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 $\mathbf{\Theta}_k$. Moreover, only an imperfect estimate $\hat{\mathbf{h}}_k$ of the true channel $\mathbf{h}_k$ is available at the transmitter which is modeled as [30]–[33]

$$\hat{\mathbf{h}}_k=\sqrt{M}\mathbf{\Theta}_k^{1/2}\left(\sqrt{1-\tau_k^2}\mathbf{z}_k+\tau_k\mathbf{q}_k\right)=\sqrt{M}\mathbf{\Theta}_k^{1/2}\hat{\mathbf{z}}_k, \tag{6}$$

where $\hat{\mathbf{z}}_k=\sqrt{1-\tau_k^2}\mathbf{z}_k+\tau_k\mathbf{q}_k$, $\mathbf{q}_k$ has i.i.d. entries of zero mean and variance $1/M$ independent of $\mathbf{z}_k$ and n_k. The

parameter $\tau_k \in [0,1]$ reflects the accuracy or quality of the channel estimate $\hat{\mathbf{h}}_k$, i.e., $\tau_k = 0$ corresponds to perfect CSIT, whereas for $\tau_k = 1$ the CSIT is completely uncorrelated to the true channel. The variation in the accuracy of the available CSIT $\hat{\mathbf{h}}_k$ between the different user channels $\mathbf{h}_k$ arises naturally. Firstly, there might be low mobility users and high mobility users with large or small channel coherence intervals, respectively. Therefore, the CSIT of the high mobility users will be outdated quickly and hence be very inaccurate. On the other hand, the CSIT of the low mobility users remains accurate since their channel does not change significantly from the time of the channel estimation until the time of precoding and coherent data transmission. Secondly, different CSIT qualities arise when the feedback rate varies among the users. For instance, if the CSIT is obtained from uplink training, the training length of each user could be different, leading to different channel estimation errors at the transmitter. Similarly, if the users feed back a quantized channel, they could use channel quantization codebooks of different sizes depending on their channel quality and the available uplink resources. However, for simplicity, we assume identical CSIT qualities $\tau_k = \tau \ \forall k$ for the optimization problems considered in Section VI and Section VII.

Remark 1: The model for imperfect CSIT in (6), is adequate for instance in a FDD system, where the channel $\mathbf{h}_k$ is *finely* quantized using a random codebook of i.i.d. vectors. Since the correlation matrices $\mathbf{\Theta}_k$ are known at both ends, user k solely quantizes the fast fading channel component $\mathbf{z}_k$ to the closest codebook vector $\hat{\mathbf{z}}_k$, which can be accurately approximated as $\hat{\mathbf{z}}_k = \sqrt{1-\tau_k^2}\mathbf{z}_k + \tau_k \mathbf{q}_k$. Subsequently, the user sends the codebook index back to the transmitter, where the estimated downlink channel is reconstructed by multiplying with $\sqrt{M}\mathbf{\Theta}_k^{1/2}$. For uncorrelated channels, this specific FDD system is studied in Section VI.

Define the compound estimated channel matrix $\hat{\mathbf{H}} \triangleq [\hat{\mathbf{h}}_1, \hat{\mathbf{h}}_2, \ldots, \hat{\mathbf{h}}_K]^{\mathsf{H}} \in \mathbb{C}^{K \times M}$. Therefore, the matrix $\frac{1}{M}\hat{\mathbf{H}}^{\mathsf{H}}\hat{\mathbf{H}}$ can be written as

$$\frac{1}{M}\hat{\mathbf{H}}^{\mathsf{H}}\hat{\mathbf{H}} = \sum_{k=1}^{K} \mathbf{\Theta}_k^{1/2}\hat{\mathbf{z}}_k\hat{\mathbf{z}}_k^{\mathsf{H}}\mathbf{\Theta}_k^{1/2}. \tag{7}$$

The per-user channel correlation model (also called generalized variance profile) is very general and encompasses various propagation environments. For instance, all channel coefficients $h_{k,i}$ of the vector channel $\mathbf{h}_k$ 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 $\mathbf{\Theta}_k = \mathrm{diag}(\sigma_{k,1}^2, \sigma_{k,2}^2, \ldots, \sigma_{k,M}^2)$, see [27], [28], [34]. Another possible scenario consists of an environment where all user channels have identical transmit correlation $\mathbf{\Theta}$, 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 $\mathbf{\Theta}_k = d_k\mathbf{\Theta}$. From a mathematical point of view, a homogeneous system with common user channel correlation $\mathbf{\Theta}_k = \mathbf{\Theta} \ \forall k$ 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 $\mathbf{\Theta}_k = \mathbf{I}_M \ \forall k$, in which case the approximated SINRs are given explicitly.

The model in (7) has never been considered in large dimensional RMT and therefore no results are available. The most general model studied, assumes a variance profile, first treated in [27] and extended in [28], which is a special case of the model in (7). Therefore, to be able to derive deterministic equivalents of the SINR, we need to extend the results in [27], [28] to account for the per-user correlation model in (7), which is done in the next section.

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 $\mathbf{\Theta}_k$. These results will be used in Sections IV-VII to solve practical optimization problems.

The following theorem extends the results in [27], [28], [35] by assuming a generalized variance profile. This theorem is required to cope with the channel model in (5) and forms the mathematical basis of the subsequent large system analysis of the MISO BC under RZF and ZF precoding.

Theorem 1: Let $\mathbf{B}_N = \mathbf{X}_N^{\mathsf{H}}\mathbf{X}_N + \mathbf{S}_N$ with $\mathbf{S}_N \in \mathbb{C}^{N\times N}$ Hermitian nonnegative definite and $\mathbf{X}_N \in \mathbb{C}^{n\times N}$ random. The ith column $\mathbf{x}_i$ of $\mathbf{X}_N^{\mathsf{H}}$ is $\mathbf{x}_i = \mathbf{\Psi}_i\mathbf{y}_i$, where the entries of $\mathbf{y}_i \in \mathbb{C}^{r_i}$ are i.i.d. of zero mean, variance $1/N$ and have eighth order moment of order $O\left(\frac{1}{N^4}\right)$. The matrices $\mathbf{\Psi}_i \in \mathbb{C}^{N\times r_i}$ are deterministic. Furthermore, let $\mathbf{\Theta}_i = \mathbf{\Psi}_i\mathbf{\Psi}_i^{\mathsf{H}} \in \mathbb{C}^{N\times N}$ and define $\mathbf{Q}_N \in \mathbb{C}^{N\times N}$ deterministic. Assume $\limsup_{N\to\infty}\sup_{1\le i\le n}\|\mathbf{\Theta}_i\| < \infty$ and let $\mathbf{Q}_N$ have uniformly bounded spectral norm (with respect to N). Define

$$m_{\mathbf{B}_N,\mathbf{Q}_N}(z) \triangleq \frac{1}{N}\mathrm{tr}\mathbf{Q}_N\left(\mathbf{B}_N - z\mathbf{I}_N\right)^{-1}. \tag{8}$$

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 $0 < \liminf_N \beta_N \le \limsup_N \beta_N < \infty$ and $0 < \liminf_N \beta_{N,i} \le \limsup_N \beta_{N,i} < \infty$, we have that

$$m_{\mathbf{B}_N,\mathbf{Q}_N}(z) - m^{\circ}_{\mathbf{B}_N,\mathbf{Q}_N}(z) \overset{N\to\infty}{\longrightarrow} 0, \tag{9}$$

almost surely, with $m^{\circ}_{\mathbf{B}_N,\mathbf{Q}_N}(z)$ given by

$$m^{\circ}_{\mathbf{B}_N,\mathbf{Q}_N}(z) = \frac{1}{N}\mathrm{tr}\mathbf{Q}_N\left(\frac{1}{N}\sum_{j=1}^{n}\frac{\mathbf{\Theta}_j}{1+e_{N,j}(z)} + \mathbf{S}_N - z\mathbf{I}_N\right)^{-1} \tag{10}$$

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

$$e_{N,i}(z) = \frac{1}{N}\mathrm{tr}\mathbf{\Theta}_i\left(\frac{1}{N}\sum_{j=1}^{n}\frac{\mathbf{\Theta}_j}{1+e_{N,j}(z)} + \mathbf{S}_N - z\mathbf{I}_N\right)^{-1} \tag{11}$$

which is the Stieltjes transform of a nonnegative finite measure on $\mathbb{R}^+$. Moreover, for $z<0$, the scalars $e_{N,1}(z), \ldots, e_{N,n}(z)$ are the unique nonnegative solutions to (11).

Note that (11) forms a system of n coupled equations, from which (10) is given explicitly.

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

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

$$e_{N,i}^{(k)}(z)=\frac{1}{N}\mathrm{tr}\boldsymbol{\Theta}_i\left(\frac{1}{N}\sum_{j=1}^{n}\frac{\boldsymbol{\Theta}_j}{1+e_{N,j}^{(k-1)}(z)}+\mathbf{S}_N-z\mathbf{I}_N\right)^{-1} \tag{12}$$

for $k>0$. Then, $\lim_{k\to\infty}e_{N,i}^{(k)}(z)=e_{N,i}(z)$ defined in (11) for $i\in\{1,2,\ldots,n\}$.

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

To derive a deterministic equivalent of the SINR under RZF and ZF precoding, we require the following assumptions on the correlation matrices $\boldsymbol{\Theta}_k$ and the power allocation matrix $\mathbf{P}$.

Assumption 1: All correlation matrices $\boldsymbol{\Theta}_k$ have uniformly bounded spectral norm on M, i.e.,

$$\limsup_{M,K\to\infty}\sup_{1\leq k\leq K}\|\boldsymbol{\Theta}_k\|<\infty. \tag{13}$$

Assumption 2: The power $p_{\max}=\max(p_1,\ldots,p_K)$ is of order $O(1/K)$, i.e.,

$$\|\mathbf{P}\|=O(1/K). \tag{14}$$

A. Regularized Zero-forcing Precoding

Consider the RZF precoding matrix

$$\mathbf{G}_{\mathrm{rzf}}=\xi\left(\hat{\mathbf{H}}^{\mathsf{H}}\hat{\mathbf{H}}+M\alpha\mathbf{I}_M\right)^{-1}\hat{\mathbf{H}}^{\mathsf{H}}, \tag{15}$$

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

From the total power constraint (1), we obtain ξ^2 as

$$\xi^2=\frac{P}{\mathrm{tr}\mathbf{P}\hat{\mathbf{H}}(\hat{\mathbf{H}}^{\mathsf{H}}\hat{\mathbf{H}}+M\alpha\mathbf{I}_M)^{-2}\hat{\mathbf{H}}^{\mathsf{H}}}=\frac{P}{\Psi},$$

where we defined $\Psi\triangleq\mathrm{tr}\mathbf{P}\hat{\mathbf{H}}(\hat{\mathbf{H}}^{\mathsf{H}}\hat{\mathbf{H}}+M\alpha\mathbf{I}_M)^{-2}\hat{\mathbf{H}}^{\mathsf{H}}$. Denoting $\hat{\mathbf{W}}\triangleq(\hat{\mathbf{H}}^{\mathsf{H}}\hat{\mathbf{H}}+M\alpha\mathbf{I}_M)^{-1}$, the SINR $\gamma_{k,\mathrm{rzf}}$ of user k in (2) under RZF precoding takes the form

$$\gamma_{k,\mathrm{rzf}}=\frac{p_k|\mathbf{h}_k^{\mathsf{H}}\hat{\mathbf{W}}\hat{\mathbf{h}}_k|^2}{\mathbf{h}_k^{\mathsf{H}}\hat{\mathbf{W}}\hat{\mathbf{H}}_{[k]}^{\mathsf{H}}\mathbf{P}_{[k]}\hat{\mathbf{H}}_{[k]}\hat{\mathbf{W}}\mathbf{h}_k+\frac{\Psi}{\rho}}, \tag{16}$$

where $\hat{\mathbf{H}}_{[k]}\triangleq[\hat{\mathbf{h}}_1,\ldots,\hat{\mathbf{h}}_{k-1},\hat{\mathbf{h}}_{k+1},\ldots,\hat{\mathbf{h}}_K]^{\mathsf{H}}\in\mathbb{C}^{K-1\times M}$ and $\mathbf{P}_{[k]}\triangleq\mathrm{diag}(p_1,\ldots,p_{k-1},p_{k+1},\ldots,p_K)$.

To derive a deterministic equivalent $\gamma_{k,\mathrm{rzf}}^{\circ}$ of the SINR $\gamma_{k,\mathrm{rzf}}$ defined in (16) such that $\gamma_{k,\mathrm{rzf}}-\gamma_{k,\mathrm{rzf}}^{\circ}\overset{M\to\infty}{\longrightarrow}0$, almost surely, we require the following assumption.

Assumption 3: The random matrix $\frac{1}{M}\hat{\mathbf{H}}^{\mathsf{H}}\hat{\mathbf{H}}$ has uniformly bounded spectral norm on M with probability one, i.e.,

$$\limsup_{M,K\to\infty}\left\|\frac{1}{M}\hat{\mathbf{H}}^{\mathsf{H}}\hat{\mathbf{H}}\right\|<\infty, \tag{17}$$

with probability one.

Remark 2: Assumption 3 holds true if $\sup_K|\{\boldsymbol{\Theta}_k:k=1,2,\ldots,K\}|<\infty$, where $|\mathcal{A}|$ denotes the cardinality of the set $\mathcal{A}$. That is, $\{\boldsymbol{\Theta}_k\}$ belongs to a *finite* family [36]. In particular, if $\boldsymbol{\Theta}_k=\boldsymbol{\Theta}$ $\forall k$, then Assumption 3 is satisfied, since $\frac{1}{M}\|\hat{\mathbf{H}}^{\mathsf{H}}\hat{\mathbf{H}}\|\leq\|\boldsymbol{\Theta}\|\|\hat{\mathbf{Z}}^{\mathsf{H}}\hat{\mathbf{Z}}\|$, where $\hat{\mathbf{Z}}=[\hat{\mathbf{z}}_1,\ldots,\hat{\mathbf{z}}_K]^{\mathsf{H}}$ and both $\|\boldsymbol{\Theta}\|$ and $\|\hat{\mathbf{Z}}^{\mathsf{H}}\hat{\mathbf{Z}}\|$ are uniformly bounded for all large M with probability one [37].

A deterministic equivalent $\gamma_{k,\mathrm{rzf}}^{\circ}$ of $\gamma_{k,\mathrm{rzf}}$ is provided in the following theorem.

Theorem 2: Let Assumptions 1, 2, and 3 hold true and let $\alpha>0$ and $\gamma_{k,\mathrm{rzf}}$ be the SINR of user k defined in (16). Then

$$\gamma_{k,\mathrm{rzf}}-\gamma_{k,\mathrm{rzf}}^{\circ}\overset{M\to\infty}{\longrightarrow}0, \tag{18}$$

almost surely, where $\gamma_{k,\mathrm{rzf}}^{\circ}$ is given by

$$\gamma_{k,\mathrm{rzf}}^{\circ}=\frac{p_k(1-\tau_k^2)\,(m_k^{\circ})^2}{\Upsilon_k^{\circ}(1-\tau_k^2[1-(1+m_k^{\circ})^2])+\frac{\Psi^{\circ}}{\rho}(1+m_k^{\circ})^2}, \tag{19}$$

with $m_k^{\circ}=e_k$, where the $e_1,\ldots,e_K$ form the unique positive solutions of

$$e_i=\frac{1}{M}\mathrm{tr}\boldsymbol{\Theta}_i\mathbf{T} \tag{20}$$

$$\mathbf{T}=\left(\frac{1}{M}\sum_{j=1}^{K}\frac{\boldsymbol{\Theta}_j}{1+e_j}+\alpha\mathbf{I}_M\right)^{-1} \tag{21}$$

and Ψ° and Υ_k° read

$$\Psi^{\circ}=\frac{1}{M}\sum_{j=1}^{K}\frac{p_je_j'}{(1+e_j)^2}, \tag{22}$$

$$\Upsilon_k^{\circ}=\frac{1}{M}\sum_{j=1,j\neq k}^{K}\frac{p_je_{j,k}'}{(1+e_j)^2}, \tag{23}$$

with $\mathbf{e}'=[e_1',\ldots,e_K']^{\mathsf{T}}$ and $\mathbf{e}_k'=[e_{1,k}',\ldots,e_{K,k}']^{\mathsf{T}}$ given by

$$\mathbf{e}'=(\mathbf{I}_K-\mathbf{J})^{-1}\mathbf{v}, \tag{24}$$

$$\mathbf{e}_k'=(\mathbf{I}_K-\mathbf{J})^{-1}\mathbf{v}_k, \tag{25}$$

where $\mathbf{J}$, $\mathbf{v}$ and $\mathbf{v}_k$ take the form

$$[\mathbf{J}]_{ij}=\frac{\frac{1}{M}\mathrm{tr}\boldsymbol{\Theta}_i\mathbf{T}\boldsymbol{\Theta}_j\mathbf{T}}{M(1+e_j)^2},$$

$$\mathbf{v}=\left[\frac{1}{M}\mathrm{tr}\boldsymbol{\Theta}_1\mathbf{T}^2,\ldots,\frac{1}{M}\mathrm{tr}\boldsymbol{\Theta}_K\mathbf{T}^2\right]^{\mathsf{T}},$$

$$\mathbf{v}_k=\left[\frac{1}{M}\mathrm{tr}\boldsymbol{\Theta}_1\mathbf{T}\boldsymbol{\Theta}_k\mathbf{T},\ldots,\frac{1}{M}\mathrm{tr}\boldsymbol{\Theta}_K\mathbf{T}\boldsymbol{\Theta}_k\mathbf{T}\right]^{\mathsf{T}}.$$

Proof: The proof of Theorem 2 is given in Appendix II. ∎

Corollary 1: Let Assumptions 1 and 2 hold true and let $\alpha>0$ and $\boldsymbol{\Theta}_k=\boldsymbol{\Theta}$ $\forall k$, then $\gamma_{k,\mathrm{rzf}}^{\circ}$ takes the form

$$\gamma_{k,\mathrm{rzf}}^{\circ}=\frac{\frac{p_k}{P/K}m^{\circ}(1-\tau_k^2)\left[e_{22}+\alpha\beta(1+m^{\circ})^2e_{12}\right]}{e_{22}(1-\frac{p_k}{P})\left[1-\tau_k^2(1-(1+m^{\circ})^2)\right]+\frac{e_{12}}{\rho}(1+m^{\circ})^2}, \tag{26}$$

where m° is the unique positive solution of

$$m^\circ = \frac{1}{M}\mathrm{tr}\boldsymbol{\Theta}\mathbf{T} \tag{27}$$

$$\mathbf{T} = \left(\frac{\boldsymbol{\Theta}/\beta}{1+m^\circ} + \alpha\mathbf{I}_M\right)^{-1} \tag{28}$$

and e_{ij} is given by

$$e_{ij} = \frac{1}{(1+m^\circ)^j}\frac{1}{M}\mathrm{tr}\boldsymbol{\Theta}^i\mathbf{T}^j. \tag{29}$$

Proof: Substituting $\boldsymbol{\Theta}_k = \boldsymbol{\Theta}\ \forall k$ into Theorem 2, we have $e_i = m_k^\circ = m^\circ$ given in (27), $e_i' = e' = [\beta(1+m^\circ)^2 e_{12}]/(\beta - e_{22})$ and $e_{i,k}' = \tilde{e}' = [\beta(1+m^\circ)^2 e_{22}]/(\beta - e_{22})$. Therefore, the terms Ψ° and Υ_k° become $(P/K)e_{12}/(\beta - e_{22})$ and $(P/K[1-p_k/P])e_{22}/(\beta - e_{22})$, respectively. Furthermore, m° can be written as

$$\begin{aligned} m^\circ &= \frac{1}{M}\mathrm{tr}\boldsymbol{\Theta}\mathbf{T}\left(\frac{\boldsymbol{\Theta}/\beta}{1+m^\circ} + \alpha\mathbf{I}_M\right)\mathbf{T} \\ &= \alpha(1+m^\circ)^2 e_{12} + \frac{1}{\beta}(1+m^\circ)e_{22}. \end{aligned} \tag{30}$$

Substituting these terms into (19) yields (26) which completes the proof. ∎

Note that under Assumption 2, the term $\frac{p_k}{P}$ in (26) can be omitted since the convergence in (18) still holds true. We will make use of this simplification when studying different applications of the SINR approximations.

Corollary 2: Let Assumption 2 hold true and let $\alpha > 0$ and $\boldsymbol{\Theta}_k = \mathbf{I}_M\ \forall k$, then $\gamma_{k,\mathrm{rzf}}^\circ$ takes the form

$$\gamma_{k,\mathrm{rzf}}^\circ = \frac{\frac{p_k}{P/K}m^\circ(1-\tau_k^2)\left[1+\alpha\beta(1+m^\circ)^2\right]}{(1-\frac{p_k}{P})\left[1-\tau_k^2(1-(1+m^\circ)^2)\right] + \frac{1}{\rho}(1+m^\circ)^2}, \tag{31}$$

where m° is given as

$$m^\circ = \frac{\beta - 1 - \beta\alpha + \sqrt{(\beta-1)^2 + 2(1+\beta)\alpha\beta + \alpha^2\beta^2}}{2\alpha\beta}. \tag{32}$$

Proof: Substituting $\boldsymbol{\Theta} = \mathbf{I}_M$ into Corollary 1, we have $e_{12} = e_{22}$ which yields (31). Moreover, (27) becomes a quadratic equation in m° with unique positive solution (32), which completes the proof. ∎

In particular, we will consider two different RZF precoders. The first RZF precoder is defined by $\alpha = \frac{1}{\beta\rho}$ and is referred to as RZF *channel distortion unaware* (RZF-CDU) precoder. Under imperfect CSIT the RZF-CDU precoder is mismatched to the true channel. The second RZF precoder is called RZF *channel distortion aware* (RZF-CDA) precoder and does account for imperfect CSIT. The optimal regularization parameter for the RZF-CDA precoder is derived in Section IV.

Moreover, there are two limiting cases of the RZF precoder corresponding to $\alpha \to \infty$ and $\alpha \to 0$. For $\alpha \to \infty$ the RZF precoder converges to the matched filter (MF) precoder $\mathbf{G}_{\mathrm{mf}} = \xi\hat{\mathbf{H}}^{\mathsf{H}}$. A deterministic equivalent $\gamma_{k,\mathrm{mf}}^\circ$ for the MF precoder can be derived by taking the limit $\gamma_{k,\mathrm{mf}}^\circ = \lim_{\alpha\to\infty}\gamma_{k,\mathrm{rzf}}^\circ$. However, since the performance of the MF precoder is rather poor and $\gamma_{k,\mathrm{mf}}^\circ$ does not involve Stieltjes transforms anymore, we will not discuss this precoding scheme in the present work. The reader is referred to [38] or [39] for a detailed large system analysis of the MF precoder. In the case of $\alpha \to 0$, the RZF precoder converges to the ZF precoder, which is discussed in the next section.

B. Zero-forcing Precoding

For $\alpha = 0$, the RZF precoding matrix in (15) reduces to the ZF precoding matrix $\mathbf{G}_{\mathrm{zf}}$ which reads

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

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

$$\xi^2 = \frac{P}{\mathrm{tr}\mathbf{P}(\hat{\mathbf{H}}\hat{\mathbf{H}}^{\mathsf{H}})^{-1}} = \frac{P}{\underline{\Psi}},$$

where $\underline{\Psi} \triangleq \mathrm{tr}\mathbf{P}(\hat{\mathbf{H}}\hat{\mathbf{H}}^{\mathsf{H}})^{-1}$. Defining $\underline{\hat{\mathbf{W}}} \triangleq \hat{\mathbf{H}}^{\mathsf{H}}(\hat{\mathbf{H}}\hat{\mathbf{H}}^{\mathsf{H}})^{-2}\hat{\mathbf{H}}$, the SINR $\gamma_{k,\mathrm{zf}}$ of user k in (2) under ZF precoding reads

$$\gamma_{k,\mathrm{zf}} = \frac{p_k|\mathbf{h}_k^{\mathsf{H}}\underline{\hat{\mathbf{W}}}\hat{\mathbf{h}}_k|^2}{\mathbf{h}_k^{\mathsf{H}}\underline{\hat{\mathbf{W}}}\hat{\mathbf{H}}_{[k]}^{\mathsf{H}}\mathbf{P}_{[k]}\hat{\mathbf{H}}_{[k]}\underline{\hat{\mathbf{W}}}\mathbf{h}_k + \frac{\underline{\Psi}}{\rho}}. \tag{33}$$

To obtain a deterministic equivalent of the SINR in (33), we need to ensure that the minimum eigenvalue of $\hat{\mathbf{H}}\hat{\mathbf{H}}^{\mathsf{H}}$ is bounded away from zero for all large M, almost surely. Therefore, the following assumption is required.

Assumption 4: There exists $\varepsilon > 0$ such that, for all large M, we have $\lambda_{\min}(\frac{1}{M}\hat{\mathbf{H}}\hat{\mathbf{H}}^{\mathsf{H}}) > \varepsilon$ with probability one.

Remark 3: If $\boldsymbol{\Theta}_k = \boldsymbol{\Theta}\ \forall k$ and $\lambda_{\min}(\boldsymbol{\Theta}) > \varepsilon > 0$ (i.e., in contrast to Theorem 2, $\boldsymbol{\Theta}$ must be invertible), for all M, then Assumption 4 holds true if $\beta > 1$. Indeed, for $\beta > 1$, from [37], there exists $\zeta > 0$ such that, for all large M, $\lambda_{\min}(\hat{\mathbf{Z}}\hat{\mathbf{Z}}^{\mathsf{H}}) > \zeta$, where $\hat{\mathbf{Z}} = [\hat{\mathbf{z}}_1, \ldots, \hat{\mathbf{z}}_K]^{\mathsf{H}}$, with probability one. Therefore, for all large M, $\lambda_{\min}(\frac{1}{M}\hat{\mathbf{H}}\hat{\mathbf{H}}^{\mathsf{H}}) \geq \lambda_{\min}(\hat{\mathbf{Z}}\hat{\mathbf{Z}}^{\mathsf{H}})\lambda_{\min}(\boldsymbol{\Theta}) > \zeta\varepsilon > 0$ almost surely.

Furthermore, we require the following assumption for the channel model with per-user correlation.

Assumption 5: Assume that $\underline{e}_i = \lim_{\alpha\to 0}\alpha e_i(\alpha)$ exists for all i and $\underline{e}_i > \varepsilon\ \forall i$ for some $\varepsilon > 0$, for all M.

Remark 4: Under these conditions, the $\underline{e}_1, \ldots, \underline{e}_K$ are the unique positive solutions of (36). In particular, Assumption 5 holds true if $\boldsymbol{\Theta}_k = \boldsymbol{\Theta}\ \forall k$, $\beta > 1$ and $\lambda_{\min}(\boldsymbol{\Theta}) > \varepsilon > 0$. This is detailed in the proof of Corollary 3.

Theorem 3: Let Assumptions 1, 2, 3, 4 and 5 hold true and let $\gamma_{k,\mathrm{zf}}$ be the SINR of user k under ZF precoding defined in (33). Then

$$\gamma_{k,\mathrm{zf}} - \gamma_{k,\mathrm{zf}}^\circ \overset{M\to\infty}{\longrightarrow} 0,$$

almost surely, where $\gamma_{k,\mathrm{zf}}^\circ$ is given by

$$\gamma_{k,\mathrm{zf}}^\circ = p_k\frac{1-\tau_k^2}{\tau_k^2\underline{\Upsilon}_k^\circ + \frac{\underline{\Psi}^\circ}{\rho}}, \tag{34}$$

where $\underline{\Psi}^\circ$ and $\underline{\Upsilon}_k^\circ$ read

$$\begin{aligned} \underline{\Psi}^\circ &= \frac{1}{M}\sum_{j=1}^K \frac{p_j}{\underline{e}_j}, \\ \underline{\Upsilon}_k^\circ &= \frac{1}{M}\sum_{j=1, j\neq k}^K p_j\frac{\underline{e}_{j,k}'}{\underline{e}_j^2}. \end{aligned} \tag{35}$$

The functions $\underline{e}_1, \ldots, \underline{e}_K$ form the unique positive solution of

$$\underline{e}_i = \frac{1}{M}\mathrm{tr}\boldsymbol{\Theta}_i\underline{\mathbf{T}} \tag{36}$$

$$\underline{\mathbf{T}} = \left(\frac{1}{M}\sum_{j=1}^{K}\frac{\boldsymbol{\Theta}_j}{\underline{e}_j} + \mathbf{I}_M\right)^{-1}. \tag{37}$$

Further, define $\underline{\mathbf{e}}'_k = [\underline{e}'_{1,k}, \ldots, \underline{e}'_{K,k}]^{\mathsf{T}}$, which is given as

$$\underline{\mathbf{e}}'_k = (\mathbf{I}_K - \underline{\mathbf{J}})^{-1}\,\underline{\mathbf{v}}_k, \tag{38}$$

where $\underline{\mathbf{J}}$ and $\underline{\mathbf{v}}_k$ take the form

$$[\underline{\mathbf{J}}]_{ij} = \frac{\frac{1}{M}\mathrm{tr}\boldsymbol{\Theta}_i\underline{\mathbf{T}}\boldsymbol{\Theta}_j\underline{\mathbf{T}}}{M\,\underline{e}_j^2},$$

$$\underline{\mathbf{v}}_k = \left[\frac{1}{M}\mathrm{tr}\boldsymbol{\Theta}_1\underline{\mathbf{T}}\boldsymbol{\Theta}_k\underline{\mathbf{T}}, \ldots, \frac{1}{M}\mathrm{tr}\boldsymbol{\Theta}_K\underline{\mathbf{T}}\boldsymbol{\Theta}_k\underline{\mathbf{T}}\right]^{\mathsf{T}}.$$

Proof: The proof of Theorem 3 is given in Appendix III. ∎

Corollary 3: Let Assumptions 1 and 2 hold true. Further, let $\beta > 1$, $\boldsymbol{\Theta}_k = \boldsymbol{\Theta}$ $\forall k$ with $\lambda_{\min}(\boldsymbol{\Theta}) > \varepsilon$, $\varepsilon > 0$, for all M, then Theorem 3 holds true and $\gamma^{\circ}_{k,\mathrm{zf}}$ takes the form

$$\gamma^{\circ}_{k,\mathrm{zf}} = \frac{p_k}{P/K}\frac{1-\tau_k^2}{\tau_k^2\underline{\Upsilon}^{\circ}\left[1-\frac{p_k}{P}\right] + \frac{\underline{\Psi}^{\circ}}{\rho}}$$

with

$$\underline{\Psi}^{\circ} = \frac{1}{\beta\underline{e}}, \tag{39}$$

$$\underline{\Upsilon}^{\circ} = \frac{\underline{e}_2/\underline{e}^2}{\beta - \underline{e}_2/\underline{e}^2}, \tag{40}$$

$$\underline{e}_2 = \frac{1}{M}\mathrm{tr}\boldsymbol{\Theta}^2\underline{\mathbf{T}}^2$$

where $\underline{e}$ is the unique positive solution of

$$\underline{e} = \frac{1}{M}\mathrm{tr}\boldsymbol{\Theta}\underline{\mathbf{T}}, \tag{41}$$

$$\underline{\mathbf{T}} = \left(\mathbf{I}_M + \frac{1}{\underline{e}\beta}\boldsymbol{\Theta}\right)^{-1}. \tag{42}$$

Proof: For $\boldsymbol{\Theta}_k = \boldsymbol{\Theta}$ $\forall k$, we obtain from (20)

$$\begin{aligned}\underline{e}_i &= \lim_{\alpha\to 0}\alpha e_i(\alpha) = \underline{e}\\ &= \lim_{\alpha\to 0}\left\{\frac{1}{M}\mathrm{tr}\boldsymbol{\Theta}\left(\frac{1}{\beta}\frac{\boldsymbol{\Theta}}{\alpha+\alpha e(\alpha)} + \mathbf{I}_M\right)^{-1}\right\}\\ &= \frac{1}{M}\mathrm{tr}\boldsymbol{\Theta}\left(\frac{\boldsymbol{\Theta}}{\beta\underline{e}} + \mathbf{I}_M\right)^{-1}.\end{aligned} \tag{43}$$

A lower bounded of (43) is given as $\underline{e} \geq \lambda_{\min}(\boldsymbol{\Theta})(1-1/\beta)$ which is uniformly bounded away from zero if $\boldsymbol{\Theta}$ is invertible and $\beta > 1$. Thus, under these conditions, Assumption 5 is satisfied. Moreover, the $\underline{e}'_{j,k}$ in (38) rewrite

$$\underline{e}'_{j,k} = \underline{e}' = \frac{\beta\underline{e}_2}{\beta - \frac{\underline{e}_2}{\underline{e}^2}}$$

and therefore,

$$\underline{\Upsilon}^{\circ}_k = \frac{\underline{e}_2/\underline{e}^2}{\beta - \frac{\underline{e}_2}{\underline{e}^2}}\frac{P}{K}\left[1-\frac{p_k}{P}\right].$$

Dividing $\underline{\Upsilon}^{\circ}_k$ by $\frac{P}{K}\left[1-\frac{p_k}{P}\right]$ and $\underline{\Psi}^{\circ} = \frac{P}{\underline{e}M}$ by P/K, we obtain $\underline{\Upsilon}^{\circ}$ given in (40) and $\underline{\Psi}^{\circ}$ given in (39), respectively, which completes the proof. ∎

Corollary 4: Let Assumption 2 hold true and let $\beta > 1$ and $\boldsymbol{\Theta}_k = \mathbf{I}_M$ $\forall k$, then $\gamma^{\circ}_{k,\mathrm{zf}}$ takes the explicit form

$$\gamma^{\circ}_{k,\mathrm{zf}} = \frac{p_k}{P/K}\frac{1-\tau_k^2}{\tau_k^2[1-\frac{p_k}{P}] + \frac{1}{\rho}}(\beta-1). \tag{44}$$

Proof: By substituting $\boldsymbol{\Theta} = \mathbf{I}_M$ into (41), $\underline{e}$ is explicitly given by $\underline{e} = (\beta-1)/\beta$. We further have $\frac{\underline{e}_2}{\underline{e}^2} = 1$ and $\underline{\Psi}^{\circ} = \underline{\Upsilon}^{\circ} = (\beta-1)^{-1}$. ∎

C. Rate Approximations

We are interested in the individual rates R_k of the users as well as the average system sum rate R_{sum}. Since the logarithm is a continuous function, by applying the continuous mapping theorem [40], it follows from the almost sure convergence $\gamma_k - \gamma^{\circ}_k \overset{M\to\infty}{\longrightarrow} 0$, that

$$R_k - R^{\circ}_k \overset{M\to\infty}{\longrightarrow} 0, \tag{45}$$

almost surely, where $R^{\circ}_k = \log(1+\gamma^{\circ}_k)$. An approximation $\hat{R}_{\mathrm{sum}}$ of the ergodic sum rate R_{sum} is obtained by replacing the instantaneous (i.e., without averaging over the channel distribution) SINR γ_k with its large system approximation γ°_k, i.e.,

$$\hat{R}_{\mathrm{sum}} = \sum_{k=1}^{K}\log\left(1+\gamma^{\circ}_k\right). \tag{46}$$

It follows that

$$\frac{1}{K}\left(R_{\mathrm{sum}} - \hat{R}_{\mathrm{sum}}\right) \overset{M\to\infty}{\longrightarrow} 0, \tag{47}$$

holds true almost surely.

Another quantity of interest is the rate gap between the achievable rate under perfect and imperfect CSIT. We define the rate gap ΔR_k of user k as

$$\Delta R_k \triangleq \bar{R}_k - R_k, \tag{48}$$

where $\bar{R}_k$ is the rate of user k under perfect CSIT, i.e., for $\tau_k^2 = 0$ $\forall k$. Then, from (45) it follows that a deterministic equivalent ΔR°_k of the rate gap of user k such that

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

almost surely, is given by

$$\Delta R^{\circ}_k = \bar{R}^{\circ}_k - R^{\circ}_k, \tag{49}$$

where $\bar{R}^{\circ}_k$ is a deterministic equivalent of the rate of user k under perfect CSIT.

Since we will require the per-user rate gaps for uncorrelated channels ($\boldsymbol{\Theta}_k = \mathbf{I}_M$ $\forall k$) in the limited feedback analysis in Sections VI and VII, we introduce hereafter ΔR°_k for RZF-CDU and ZF precoding.

Corollary 5 (RZF-CDU precoding): Let $\boldsymbol{\Theta}_k = \mathbf{I}_M$ $\forall k$, $p_k = P/K$ $\forall k$, $\tau_k^2 = \tau^2$ $\forall k$ and define $\Delta R_{k,\mathrm{rzf-cdu}}$ as the rate gap of user k under RZF-CDU precoding. Then a deterministic equivalent $\Delta R^{\circ}_{k,\mathrm{rzf-cdu}} = \Delta R^{\circ}_{\mathrm{rzf-cdu}}$ such that

$$\Delta R_{k,\mathrm{rzf-cdu}} - \Delta R^{\circ}_{\mathrm{rzf-cdu}} \overset{M\to\infty}{\longrightarrow} 0$$

almost surely, is given by

$$\Delta R^{\circ}_{\mathrm{rzf-cdu}} = \log\left(\frac{1+m^{\circ}}{1+\frac{m^{\circ}(1-\tau^2)\left[1+\frac{1}{\rho}(1+m^{\circ})^2\right]}{1-\tau^2+(1+m^{\circ})^2[\tau^2+\frac{1}{\rho}]}}\right),$$

where m° is given in (32).

Proof: With Corollary 2, compute $\Delta R^{\circ}_{\mathrm{rzf-cdu}}$ as defined in (49), where $\bar{R}^{\circ}_{\mathrm{rzf-cdu}} = \log(1+m^{\circ})$. ∎

Corollary 6 (ZF precoding): Let $\mathbf{\Theta}_k = \mathbf{I}_M \ \forall k$, $p_k = P/K \ \forall k$ and define $\Delta R_{k,\mathrm{zf}}$ to be the rate gap of user k under ZF precoding. Then

$$\Delta R_{k,\mathrm{zf}} - \Delta R^{\circ}_{k,\mathrm{zf}} \overset{M\to\infty}{\longrightarrow} 0$$

almost surely, with $\Delta R^{\circ}_{k,\mathrm{zf}}$ given by

$$\Delta R^{\circ}_{k,\mathrm{zf}} = \log\left(\frac{1+\rho(\beta-1)}{1+\rho\omega_k(\beta-1)}\right)$$

where ω_k is defined given by

$$\omega_k = \frac{1-\tau_k^2}{1+\tau_k^2\rho}. \tag{50}$$

Proof: Substitute the SINR from Corollary 4 into (49). ∎

Remark 5: In practice, one is often interested in the average system performance, e.g., the ergodic SINR $E[\gamma_k]$ or ergodic rate $E[R_k]$. Since the SINR γ_k is uniformly bounded on M for the considered precoding schemes, we can apply the dominated convergence theorem [40, Theorem 16.4] and obtain

$$E[\gamma_k] - \gamma_k^{\circ} \overset{M\to\infty}{\longrightarrow} 0,$$

where the expectation is taken over the probability space generating the sequence $\{\mathbf{H}(\omega),\ M \geq 1\}$ with $\mathbf{H} = [\mathbf{h}_1, \ldots, \mathbf{h}_K]^{\mathsf{H}} \in \mathbb{C}^{K\times M}$. The same holds true for the per-user rate R_k, i.e., $E[R_k] - R_k^{\circ} \overset{M\to\infty}{\longrightarrow} 0$.

D. Numerical Results

We validate Theorem 2 and Theorem 3 by comparing the ergodic sum rate (4), obtained by Monte-Carlo (MC) simulations of i.i.d. Rayleigh block-fading channels, to the large system approximation $\hat{R}_{\mathrm{sum}}$, for finite system dimensions and equal power allocation $\mathbf{P} = \frac{1}{K}\mathbf{I}_K$.

The correlation $\mathbf{\Theta}_k$ of the kth user channel is modeled as in [41] by assuming a diffuse two-dimensional field of isotropic scatterers around the receivers. The waves impinge the receiver k uniformly at an azimuth angle θ ranging from $\theta_{k,\min}$ to $\theta_{k,\max}$. Denoting d_{ij} the distance between transmit antenna i and j, the correlation is modeled as

$$[\mathbf{\Theta}_k]_{ij} = \frac{1}{\theta_{k,\max}-\theta_{k,\min}} \int_{\theta_{k,\min}}^{\theta_{k,\max}} e^{\,\mathbf{i}\frac{2\pi}{\lambda} d_{ij}\cos(\theta)} d\theta, \tag{51}$$

where λ 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 $\theta_{k,\min} = -\pi$ and $\theta_{k,\max} = \varphi_k - \pi$. Note that for small $\theta_{k,\max} - \theta_{k,\min}$ (in our example for small values of k), the corresponding signal of user k is highly correlated since the signal arrives from a very narrow angle. Thus, the correlation model (51) yields rank-deficient correlating matrices for some users. The transmitter is equipped with a uniform linear array (ULA). To ensure that $\|\mathbf{\Theta}_k\|$ 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.

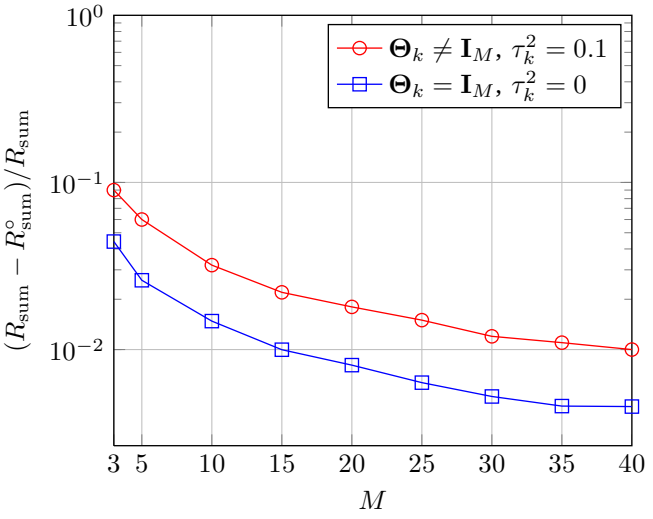

Fig. 1. RZF, $(R_{\mathrm{sum}} - R^{\circ}_{\mathrm{sum}})/R_{\mathrm{sum}}$ vs. M for a fixed SNR of $\rho = 10$ dB with $M = K$, $\alpha = 1/\rho$.

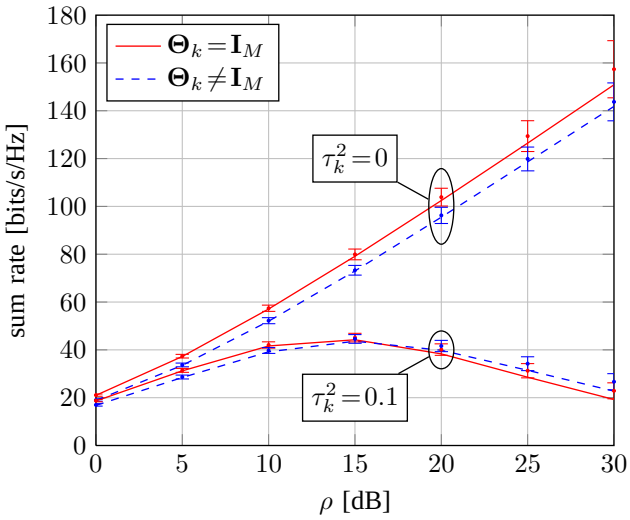

Fig. 2. RZF, sum rate vs. SNR with $M = K = 30$ and $\alpha = 1/\rho$, simulation results are indicated by circle marks with error bars indicating the standard deviation.

The simulation results presented in Figure 1 depict the absolute error of the sum rate approximation $\hat{R}_{\mathrm{sum}}$ compared to the ergodic sum rate R_{sum}, averaged over 10 000 independent channel realizations. The notation "$\mathbf{\Theta}_k \neq \mathbf{I}_M$" indicates that $\mathbf{\Theta}_k$ is modeled according to (51) with $d_{ij}/\lambda = 0.5$. From Figure 1, we observe that the approximated sum rate $\hat{R}_{\mathrm{sum}}$ becomes more accurate with increasing M.

Figures 2 and 3 compare the ergodic sum rate to the deterministic approximation (46) 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

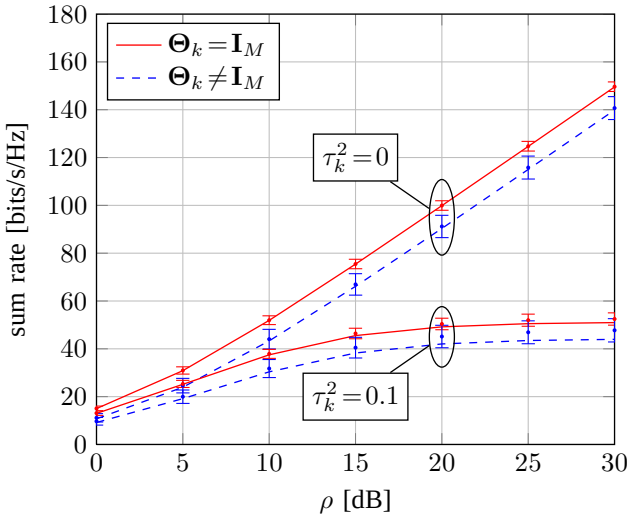

Fig. 3. ZF, sum rate vs. SNR with $M = 30$, $K = 15$, simulation results are indicated by circle marks with error bars indicating the standard deviation.

roughly 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 α does not account for τ_k^2 and thus the matrix $\hat{\mathbf{H}}^{\mathsf{H}}\hat{\mathbf{H}} + M\alpha\mathbf{I}_M$ 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 $\hat{\mathbf{H}}$ 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 thus the bounds in Proposition 12 (Appendix I-A) are 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 approximations in Theorems 2 and 3 are accurate even for small dimensions and can be applied to various optimization problems discussed in the sequel.

IV. Sum Rate Maximizing Regularization

The optimal regularization parameter $\alpha^{\star\circ}$ maximizing (46) is defined as

$$\alpha^{\star\circ} = \arg\max_{\alpha>0} \sum_{k=1}^{K} \log\left(1 + \gamma_{k,\mathrm{rzf}}^{\circ}\right). \tag{52}$$

In general, the optimization problem (52) is not convex in α and the solution has to be computed via a one-dimensional line search.

In the following, we confine ourselves to the case of common correlation $\mathbf{\Theta}_k = \mathbf{\Theta}$ $\forall k$, since for per-user correlation a common regularization parameter is not optimal anymore [12], [42]. Under common transmit correlation, we subsequently assume that the distortions τ_k^2 of the CSIT $\hat{\mathbf{h}}_k$ are identical for all users, since the users' channels are statistically equivalent. Under these conditions $\mathbf{P} = \frac{1}{K}\mathbf{I}_K$ maximizes (46) and the optimization problem (52) has the following solution.

Proposition 2: Let $\mathbf{\Theta}_k = \mathbf{\Theta}$, $0 \leq \tau_k = \tau < 1$ $\forall k$ and $\mathbf{P} = \frac{1}{K}\mathbf{I}_K$. The approximated SINR $\gamma_{k,\mathrm{rzf}}^{\circ}$ of user k under RZF precoding (equivalently, the approximated per-user rate and the sum rate) is maximized for a regularization parameter $\alpha \triangleq \alpha^{\star\circ}$, given as a positive solution to the fixed-point equation

$$\alpha^{\star\circ} = \frac{\left[1 + \nu(\alpha^{\star\circ}) + \tau^2\rho\frac{e_{22}(\alpha^{\star\circ})}{e_{12}(\alpha^{\star\circ})}\right]\frac{1}{\beta\rho}}{(1-\tau^2)[1+\nu(\alpha^{\star\circ})] + \tau^2\nu(\alpha^{\star\circ})[1+m^{\circ}(\alpha^{\star\circ})]^2} \tag{53}$$

where $m^{\circ}(\alpha)$ is defined in (27) and $\nu(\alpha)$ is given by

$$\nu(\alpha) = \frac{1}{(1+m^{\circ})e_{22}}\frac{e_{13}}{e_{12}}\left[\frac{e_{22}}{e_{12}} - \frac{e_{23}}{e_{13}}\right] \tag{54}$$

with e_{ij} defined in (29).

Proof: The proof is provided in Appendix IV. ■

Note that the solution in Proposition 2 assumes a *fixed* distortion τ^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 on the amount of channel training.

Under perfect CSIT ($\tau^2 = 0$), Proposition 2 simplifies to the well-known solution $\alpha^{\star\circ} = \frac{1}{\beta\rho}$, *independent* of $\mathbf{\Theta}$, which has previously been derived in [9], [22], [26]. As mentioned in [9], for *large* M the RZF-CDA precoder is identical to the MMSE precoder in [19], [43]. The authors in [26] showed that, under *perfect* CSIT, $\alpha^{\star\circ}$ is independent of the correlation $\mathbf{\Theta}$. However, for imperfect CSIT ($\tau^2 \neq 0$), the optimal regularization parameter (53) depends on the transmit correlation through $m^{\circ}(\alpha)$ and $e_{ij}(\alpha)$. For uncorrelated channels ($\mathbf{\Theta} = \mathbf{I}_M$), we have $e_{12} = e_{22}$ and $\nu(\alpha) = 0$ and therefore the explicit solution

$$\alpha^{\star\circ} = \left(\frac{1+\tau^2\rho}{1-\tau^2}\right)\frac{1}{\beta\rho}. \tag{55}$$

Note that in this case, it can be shown that $\alpha^{\star\circ}$ in (55) is the *unique* positive solution to (52).

For imperfect CSIT ($\tau^2 > 0$), the RZF-CDA precoder and the MMSE precoder with regularization parameter $\alpha_{\mathrm{MMSE}} = \tau^2\beta^{-1} + 1/(\beta\rho)$ [43] are not identical anymore, even in the large M, K limit. Unlike the case of perfect CSIT, $\alpha^{\star\circ}$ now depends on the correlation matrix $\mathbf{\Theta}$ through $m^{\circ}(\alpha^{\star\circ})$ and $e_{ij}(\alpha^{\star\circ})$. The impact of m° and e_{ij} on the sum rate of RZF-CDA precoding is evaluated through numerical simulations in Figure 5. Further note that since $m^{\circ}(\alpha)$ and e_{ij} are bounded from above under the conditions explained in Remark 6 below, at asymptotically high SNR the regularization parameter $\alpha^{\star\circ}$ in (53) converges to $\alpha_{\infty}^{\star\circ} \triangleq \lim_{\rho\to\infty}\alpha^{\star\circ}$, where $\alpha_{\infty}^{\star\circ}$ is a positive solution of

$$\alpha_{\infty}^{\star\circ} = \frac{\frac{\tau^2}{\beta}\frac{e_{22}(\alpha_{\infty}^{\star\circ})}{e_{12}(\alpha_{\infty}^{\star\circ})}}{(1-\tau^2)[1+\nu(\alpha_{\infty}^{\star\circ})] + \tau^2\nu(\alpha_{\infty}^{\star\circ})[1+m^{\circ}(\alpha_{\infty}^{\star\circ})]^2}. \tag{56}$$

For uncorrelated channels, the limit in (56) takes the form

$$\alpha_{\infty}^{\star\circ} = \frac{\tau^2}{(1-\tau^2)\beta}.$$

Thus, for asymptotically high SNR, RZF-CDA precoding is *not* the same as ZF precoding, since the regularization parameter $\alpha^{\star\circ}$ is non-zero due to the residual interference

caused by the imperfect CSIT. Similar observations have been made in [43] for the MMSE precoder.

Remark 6: Note that in (56) we apply the limit $\rho \to \infty$ on a result obtained from an SINR approximation which is almost surely exact as $M, K \to \infty$. This is correct if $\Psi = \mathrm{tr}\mathbf{P}\hat{\mathbf{H}}(\hat{\mathbf{H}}^{\mathsf{H}}\hat{\mathbf{H}} + M\alpha\mathbf{I}_M)^{-2}\hat{\mathbf{H}}^{\mathsf{H}}$ in (16) is bounded for asymptotically high SNR as $M, K \to \infty$. For $\tau^2 > 0$ it is clear that Ψ° is bounded since $\alpha^{\star\circ} > 0$ for all SNR. In the case where $\tau^2 = 0$, we have $\lim_{\rho\to\infty} \alpha^{\star\circ} = 0$ and thus for $\beta = 1$ the support of the limiting eigenvalue distribution of $\frac{1}{M}\hat{\mathbf{H}}\hat{\mathbf{H}}^{\mathsf{H}}$ includes zero resulting in an unbounded Ψ°. From Remark 3, for $\beta > 1$, $\mathbf{\Theta}_k = \mathbf{\Theta}\ \forall k$ and $\lambda_{\min}(\mathbf{\Theta}) > \varepsilon > 0$ there exists $\xi > 0$ such that $\lambda_{\min}(\frac{1}{M}\hat{\mathbf{H}}\hat{\mathbf{H}}^{\mathsf{H}}) > \xi$ for all large M. Thus, Ψ° is bounded. On the contrary, for $\mathbf{\Theta}_k \neq \mathbf{\Theta}_j$ $(k \neq j)$, $\beta > 1$ and $\lambda_{\min}(\mathbf{\Theta}_k) > \varepsilon > 0\ \forall k$, it has not been proved that $\lambda_{\min}(\frac{1}{M}\hat{\mathbf{H}}\hat{\mathbf{H}}^{\mathsf{H}}) > \xi$ and we have to evoke Assumption 4 to ensure that Ψ° is bounded. Thus, for $\tau^2 = 0$, the limit (56) is only well defined for $\beta > 1$. Further note that if Ψ° is bounded as $M, K \to \infty$ the limits $M, K \to \infty$ and $\rho \to \infty$ can be inverted without affecting the result.

For various special cases, substituting (53) into the deterministic equivalent of the SINR $\gamma^\circ_{k,\mathrm{rzf}}$ in (26) yields the following simplified expressions.

Corollary 7: Let Assumptions 1 and 2 hold true and let $\mathbf{\Theta}_k = \mathbf{\Theta}$, $\tau_k^2 = 0$, $p_k = P/K\ \forall k$, $\alpha^{\star\circ} = \frac{1}{\beta\rho}$ and $\gamma_{k,\mathrm{rzf-cda}}$ be the sum rate maximizing SINR of user k under RZF precoding. Then

$$\gamma_{k,\mathrm{rzf-cda}} - \gamma^\circ_{k,\mathrm{rzf-cda}} \overset{M\to\infty}{\longrightarrow} 0,$$

almost surely, where $\gamma^\circ_{k,\mathrm{rzf-cda}}$ is given by

$$\gamma^\circ_{k,\mathrm{rzf-cda}} \triangleq \gamma^\circ_{\mathrm{rzf-cda}} = m^\circ(-\alpha^{\star\circ}), \tag{57}$$

where $m^\circ(-\alpha^{\star\circ})$ is the unique positive solution to

$$m^\circ(-\alpha^{\star\circ}) = \frac{1}{M}\mathrm{tr}\mathbf{\Theta}\left(\frac{\mathbf{\Theta}/\beta}{1 + m^\circ(-\alpha^{\star\circ})} + \alpha^{\star\circ}\mathbf{I}_M\right)^{-1}.$$

Proof: Substituting $\alpha^{\star\circ} = \frac{1}{\beta\rho}$ into (26) together with $\tau^2 = 0$, we obtain (57) which completes the proof. ∎

For uncorrelated channels $\mathbf{\Theta}_k = \mathbf{I}_M\ \forall k$, the solution to (57) is explicit and summarized in the following corollary.

Corollary 8: Let $\mathbf{\Theta}_k = \mathbf{I}_M$, $\tau_k^2 = \tau^2$, $p_k = P/K\ \forall k$ and $\gamma_{k,\mathrm{rzf-cda}}$ be the sum rate maximizing SINR of user k under RZF precoding. Then $\gamma_{k,\mathrm{rzf-cda}} - \gamma^\circ_{k,\mathrm{rzf-cda}} \overset{M\to\infty}{\longrightarrow} 0$, almost surely, where $\gamma^\circ_{k,\mathrm{rzf-cda}}$ is given by

$$\gamma^\circ_{k,\mathrm{rzf-cda}} \triangleq \gamma^\circ_{\mathrm{rzf-cda}} = \frac{\omega}{2}\rho(\beta - 1) + \frac{\chi}{2} - \frac{1}{2}, \tag{58}$$

where $\omega \in [0, 1]$ and χ are given by

$$\omega = \frac{1 - \tau^2}{1 + \tau^2\rho}, \tag{59}$$

$$\chi(\omega) = \sqrt{(\beta - 1)^2\omega^2\rho^2 + 2(1 + \beta)\omega\rho + 1}. \tag{60}$$

Proof: Substituting $\mathbf{\Theta} = \mathbf{I}_M$ into Corollary 7 leads to a quadratic equation in $m^\circ(-\alpha^{\star\circ})$ for which the unique positive solution is given by (58), which completes the proof. ∎

A deterministic equivalent $\Delta R^\circ_{\mathrm{rzf-cda}}$ of the rate gap $\Delta R_{k,\mathrm{rzf-cda}}$ under RZF-CDA precoding is provided in the following corollary.

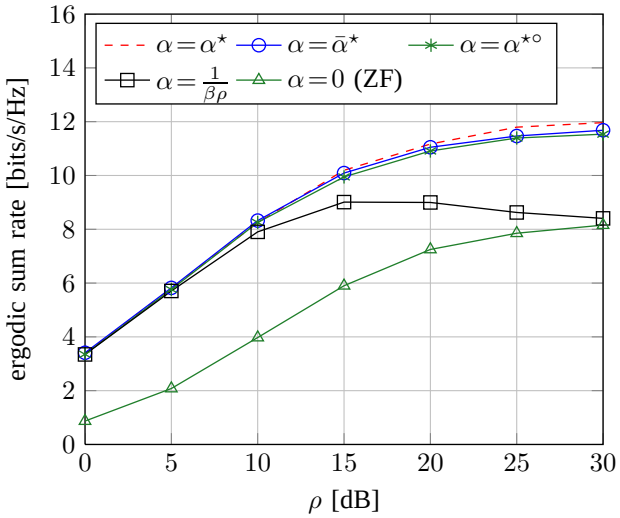

Fig. 4. RZF, ergodic sum rate vs. SNR with $M = K = 5$, $\mathbf{\Theta}_k = \mathbf{I}_M\ \forall k$, $\mathbf{P} = \frac{1}{K}\mathbf{I}_K$ and $\tau^2 = 0.1$.

Corollary 9 (RZF-CDA precoding): Let $\mathbf{\Theta}_k = \mathbf{I}_M\ \forall k$, $p_k = P/K\ \forall k$, $\tau_k^2 = \tau^2\ \forall k$ and define $\Delta R_{k,\mathrm{rzf-cda}}$ as the rate gap of user k under RZF-CDA precoding. Then,

$$\Delta R_{k,\mathrm{rzf-cda}} - \Delta R^\circ_{\mathrm{rzf-cda}} \overset{M\to\infty}{\longrightarrow} 0$$

almost surely, with

$$\Delta R^\circ_{\mathrm{rzf-cda}} = \log\left(\frac{1 + \rho(\beta - 1) + \chi(1)}{1 + \omega\rho(\beta - 1) + \chi(\omega)}\right),$$

where ω and χ are defined in (59) and (60), respectively.

Proof: With Corollary 8, compute $\Delta R^\circ_{\mathrm{rzf-cda}}$ as defined in (49). ∎

The impact of the regularization parameter on the ergodic sum rate is depicted in Figures 4 and 5.

In Figure 4, we compare the ergodic sum rate performance for different regularization parameters α with CSIT distortion $\tau_k^2 = \tau^2 = 0.1\ \forall k$. The upper bound $\alpha = \alpha^\star$ is obtained by optimizing α for every channel realization, whereas $\bar{\alpha}^\star$ maximizes the ergodic sum rate. It can be observed that both $\bar{\alpha}^\star$ and $\alpha^{\star\circ}$ perform close to the optimal $\alpha^\star$. Furthermore, if the channel quality τ^2 is unknown at the transmitter (and hence assumed to be equal to zero), the performance is decreasing as soon as τ^2 dominates (i.e. the inter-user interference limits the performance) the noise power σ^2 and approaches the sum rate of ZF precoding for high SNR. We conclude that (i) adapting the regularization parameter yields a significant performance increase and (ii) that the proposed RZF-CDA precoder with $\alpha^{\star\circ}$ performs close to optimal even for small system dimensions.

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

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

We compare two different RZF precoders: A first precoder coined RZF common correlation aware (RZF-CCA) that takes the channel correlation into account and computes α according

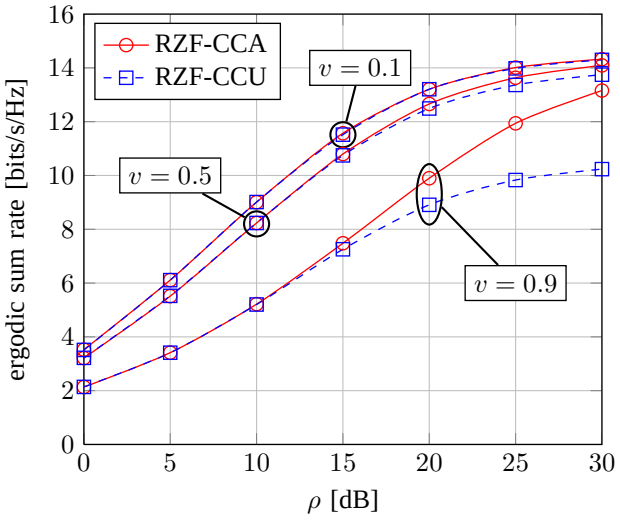

Fig. 5. RZF, ergodic sum rate vs. SNR with $M = K = 5$, $\mathbf{P} = \frac{1}{K}\mathbf{I}_K$ and $\tau^2 = 0.05$.

to (53), and a second precoder, called RZF common correlation unaware (RZF-CCU) that does not take $\mathbf{\Theta}$ into account and computes α as in (55). 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 precoders perform equally well at low SNR. Therefore, we conclude that it is beneficial to account for transmit correlation, especially in highly correlated channels. Further simulations (not provided here) suggest that the sum rate gain of RZF-CCA over RZF-CCU precoding is less pronounced for lower CSIT qualities (i.e., increasing τ^2), because in this case the impact of the CSIT quality τ^2 is more significant than the impact of $\mathbf{\Theta}$ on the sum rate.

V. Optimal Number of Users and Power Allocation

In this section, we address two problems: (i) the determination of the sum rate maximizing number of users per transmit antenna for a fixed M and (ii) the optimization of the power distribution among a *given* set of users with unequal CSIT qualities.

Consider problem (i). Intuitively, an optimal number of users $K^\star$ exists because serving more users creates more interference which in turn reduces the rates of the users. At some point the accumulated rate loss, due to the additional interference caused by scheduling another user, will outweigh the sum rate gain and hence the system sum rate will decrease. In particular, we consider a fair scenario where the SINR approximation of all users are equal. Here, the (approximated) optimal solution can be expressed under a closed form for ZF precoding.

In problem (ii), we optimize the power allocation matrix $\mathbf{P}$ for a given K. More precisely, we focus on common correlation $\mathbf{\Theta}_k = \mathbf{\Theta}\ \forall k$ with *different* CSIT qualities τ_k^2, since in this case the (approximated) optimal power distribution $\mathbf{P}^{\star\circ}$ is the solution of a classical water-filling algorithm.

A. Sum Rate Maximizing Number of Users

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

$$\beta^{\star\circ} = \arg\max_{\beta} \frac{1}{\beta}\frac{1}{K}\sum_{k=1}^{K}\log\left(1+\gamma_k^\circ\right), \tag{61}$$

where γ_k° denotes either $\gamma_{k,\mathrm{zf}}^\circ$ with $\beta > 1$ or $\gamma_{k,\mathrm{rzf}}^\circ$ with $\beta \geq 1$. In general (61) has to be solved by a one-dimensional line search. However, in case of ZF precoding and uncorrelated antennas, the optimization problem (61) has a closed-form solution given in the following proposition.

Proposition 3: Let $\mathbf{\Theta}_k = \mathbf{I}_M$, $\tau_k = \tau\ \forall k$ and $\mathbf{P} = \frac{P}{K}\mathbf{I}_K$, the sum rate maximizing system loading per transmit antenna $\beta^{\star\circ}$ is given by

$$\beta^{\star\circ} = \left(1-\frac{1}{a}\right)\left(1+\frac{1}{\mathcal{W}(x)}\right), \tag{62}$$

where $a = \frac{1-\tau^2}{\tau^2+\frac{1}{\rho}}$, $x = \frac{a-1}{e}$ and $\mathcal{W}(x)$ is the Lambert W-function defined as $z = \mathcal{W}(z)e^{\mathcal{W}(z)}$, $z \in \mathbb{C}$.

Proof: Substituting the SINR in Corollary 4 into (61) and differentiating along β leads to

$$\frac{a\beta}{1+a(\beta-1)} = \log\left(1+a(\beta-1)\right) \tag{63}$$

Denoting $w(\beta) = \frac{a-1}{a(\beta-1)+1}$, we can rewrite (63) as

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

Noticing that $w(\beta) = \mathcal{W}(x)$ and solving for β yields (62), which completes the proof. ■

For $\tau \in [0,1]$, $\beta > 1$ we have $w \geq -1$ and $x \geq -e^{-1}$. In this case $\mathcal{W}(x)$ is a well-defined function. If $\tau^2 = 0$, we obtain the results in [18], although in [18] they are not given in closed form. Note that for $\tau^2 = 0$, we have $\lim_{\rho\to\infty}\beta^{\star\circ} = 1$, i.e., the optimal system loading tends to one. Further note that only integer values of $M/\beta^{\star\circ}$ are meaningful in practice.

B. Power Optimization under Common Correlation

From Corollaries 1 and 3, the approximated sum rate (46) for both RZF and ZF precoding takes the form

$$\hat{R}_{\mathrm{sum}} = \sum_{k=1}^{K}\log\left[1+p_k\nu_k^\circ(\tau_k)\right], \tag{64}$$

with $\nu_k^\circ(\tau_k) = \gamma_k^\circ/p_k$, where the only dependence on user k stems from τ_k. The user powers $p_k^{\star\circ}$ that maximize (64), subject to $\sum_{k=1}^{K}p_k \leq P$, $p_k \geq 0$, are thus given by the classical water-filling solution [44]

$$p_k^{\star\circ} = \left[\mu - \frac{1}{\nu_k^\circ(\tau_k)}\right]^+, \tag{65}$$

where $[x]^+ \triangleq \max(0,x)$ and μ is the water level chosen to satisfy $\sum_{k=1}^{K}p_k = P$. For $\tau_k^2 = \tau^2\ forall k$, the optimal user powers (65) are all equal, i.e., $p_k^{\star\circ} = p^{\star\circ} = P/K$ and $\mathbf{P}^{\star\circ} \triangleq \mathrm{diag}(p_1^{\star\circ},\ldots,p_K^{\star\circ}) = \frac{P}{K}\mathbf{I}_K$. In this case though, it could still be beneficial to adapt the number of users as discussed in Section V-A.

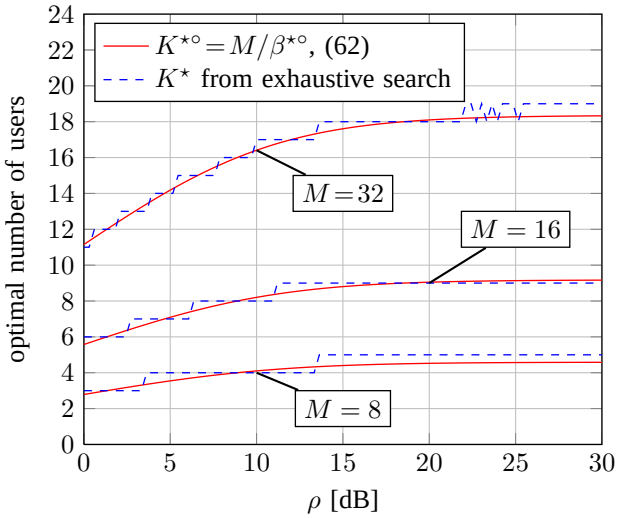

Fig. 6. ZF, sum rate maximizing number of users vs. SNR with $\mathbf{\Theta}_k = \mathbf{I}_M \ \forall k$, $\tau^2 = 0.1$ and $\mathbf{P} = \frac{1}{K}\mathbf{I}_K$.

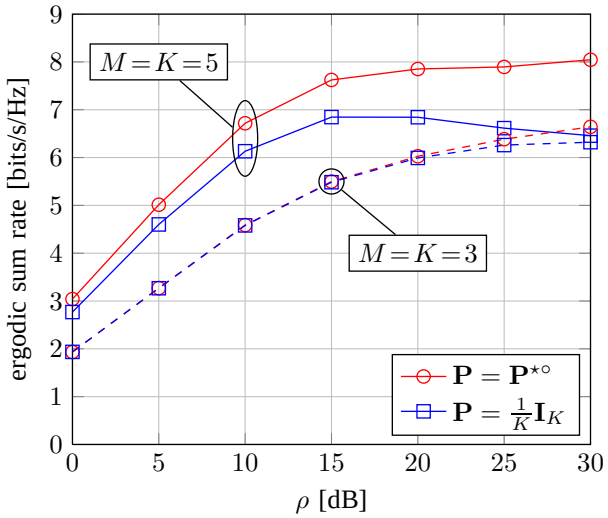

Fig. 8. RZF-CDU, R_{sum} vs. ρ with $\alpha = 1/\rho$, $\mathbf{\Theta}_k = \mathbf{I}_M \ \forall k$, $P = 1$ and $\tau_k^2 \in \mathcal{T}_1, \cup_{k=1}^3 \tau_k^2 = \mathcal{T}_1$ ($M = 5$) and $\tau_k^2 \in \mathcal{T}_2, \cup_{k=1}^3 \tau_k^2 = \mathcal{T}_2$ ($M = 3$).

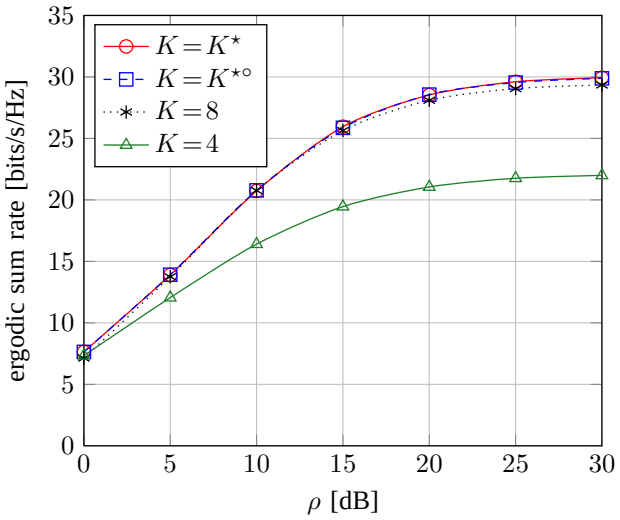

Fig. 7. ZF, R_{sum} vs. SNR with $M = 16$, $\mathbf{\Theta}_k = \mathbf{I}_M \ \forall k$, $\mathbf{P} = \frac{1}{K}\mathbf{I}_K$ and $\tau^2 = 0.1$.

C. Numerical Results

Figure 6 compares the optimal number of users $K^{\star\circ} = M/\beta^{\star\circ}$ in (62) to $K^{\star}$ obtained by choosing the $K \in \{1, 2, \ldots, M\}$ such that the ergodic sum rate is maximized, whereas Figure 7 depicts the impact of a suboptimal number of users on the ergodic sum rate of the system.

From Figure 6, it can be observed that (i) the approximated results $K^{\star\circ}$ do fit well with the simulation results even for small dimensions, (ii) $(K^{\star}, K^{\star\circ})$ increase with the SNR and (iii), for $\tau^2 \neq 0$, $(K^{\star}, K^{\star\circ})$ saturate for high SNR at a value lower than M. Therefore, under imperfect CSIT, it is not optimal anymore to serve the maximum number of users $K = M$ for asymptotically high SNR. Instead, depending on τ^2, a lower number of users $K < M$ should be served even at high SNR which implies a reduced multiplexing gain of the system. The impact of different numbers of users on the sum rate is depicted in Figure 7.

From Figure 7 we observe that (i) the approximate solution $K^{\star\circ}$ achieves most of the sum rate and (ii) adapting the number of users with the SNR is beneficial compared to a fixed K. Moreover, from Figure 6, we identify $K = 8$ as an 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 6 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 users must be adapted to the channel conditions and the approximate result $K^{\star\circ}$ is a good choice to determine the optimal number of users.

In Figure 8, under RZF-CDU precoding, we compare the ergodic sum rate performance with power allocation $\mathbf{P} = \mathbf{P}^{\star\circ}$ from (65) to equal power allocation $\mathbf{P} = \frac{1}{K}\mathbf{I}_K$. We consider a system with $M = K = 5$, where the CSIT qualities vary significantly among the users, i.e., $\tau_k^2 \in \mathcal{T}_1$ with $\mathcal{T}_1 = \{0.8, 0.3, 0.2, 0.1, 0.05\}$, $\cup_{k=1}^5 \tau_k^2 = \mathcal{T}_1$. We observe a significant gain over the whole SNR range when optimal power allocation is applied. In contrast, if the CSIT distortion of the users' channels with $M = K = 3$ does not differ considerably ($\tau_k^2 \in \mathcal{T}_2$, $\cup_{k=1}^3 \tau_k^2 = \mathcal{T}_2$ with $\mathcal{T}_2 = \{0.2, 0.15, 0.1\}$), we only observe a small gain at high SNR. For increasing SNR, the SINRs become increasingly distinct depending on the τ_k^2. Therefore, it might be optimal to turn off the users with lowest CSIT accuracy as the SNR increases, which explains why the sum rate gain is larger at high SNR than at low SNR. However, recall that the water-filling solution is optimal under Assumption 2 ($\|\mathbf{P}\| = O(1/K)$) and large M. We thus conclude that the optimal power allocation proposed in (65) achieves significant performance gains, especially at high SNR, when the quality of the available CSIT varies considerably among the users' channels.

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 codebook 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 an imperfect estimate of the true downlink channel. To obtain tractable expressions, we restrict the subsequent analysis to i.i.d. Gaussian channels $\mathbf{h}_k \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_M)\ \forall k$.

In the sequel, we follow the limited feedback analysis in [45], where each user's *channel direction* $\tilde{\mathbf{h}}_k \triangleq \frac{\mathbf{h}_k}{\|\mathbf{h}_k\|_2}$ is quantized using B bits which are subsequently fed back to the transmitter. Under Rayleigh fading, the channel $\mathbf{h}_k$ can be decomposed as $\mathbf{h}_k = \|\mathbf{h}_k\|_2 \cdot \tilde{\mathbf{h}}_k$, where we suppose that the channel magnitude $\|\mathbf{h}_k\|_2$ is perfectly known to the transmitter since it can be efficiently quantized with only a few bits [45]. Without loss of generality,[1] we assume random vector quantization (RVQ), where each user *independently* generates a random codebook $\mathcal{C}_k \triangleq \{\mathbf{w}_{ki}, \ldots, \mathbf{w}_{k2^B}\}$ containing 2^B vectors $\mathbf{w}_{ki} \in \mathbb{C}^M$ that are isotropically distributed on the M-dimensional unit sphere. Subsequently, user k quantizes its channel direction $\tilde{\mathbf{h}}_k$ to the closest $\mathbf{w}_{ki}$ according to

$$\hat{\tilde{\mathbf{h}}}_k = \arg \max_{\mathbf{w}_{ki} \in \mathcal{C}_k} \|\tilde{\mathbf{h}}_k^{\mathsf{H}} \mathbf{w}_{ki}\|.$$

Under RVQ, the quantized channel direction $\hat{\tilde{\mathbf{h}}}_k \in \mathcal{C}_k$ is isotropically distributed on the M-dimensional unit sphere due to the statistical properties of both, the random codebook $\mathcal{C}_k$ and the channels $\mathbf{h}_k$. Thus, for *fine quantization* with small errors, the entries of both $\tilde{\mathbf{h}}_k$ and $\hat{\mathbf{h}}_k = \|\mathbf{h}_k\|_2 \cdot \hat{\tilde{\mathbf{h}}}_k$ can be modeled with good approximation as i.i.d. Gaussian of zero mean and unit variance. The quantization error vector $\mathbf{e}_k$ can be approximated as $\mathbf{e}_k \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_M)$ [46] and we can write

$$\hat{\mathbf{h}}_k = \sqrt{1 - \tau_k^2}\mathbf{h}_k + \tau_k \mathbf{e}_k, \tag{66}$$

where τ_k^2 is the quantization error variance. The scaling in (66) is required to ensure that the elements of $\hat{\mathbf{h}}_k$ have unit variance. Therefore, the effect of imperfect CSIT under RVQ in (66) is captured by the channel model (6). For RVQ, the quantization error $\tau_k^2 \triangleq \|\tilde{\mathbf{h}}_k^{\mathsf{H}} \hat{\tilde{\mathbf{h}}}_k\|$ can be upper bounded as [45, Lemma 1]

$$\tau_k^2 < 2^{-\frac{B}{M-1}}. \tag{67}$$

The bound in (67) is tight for large B [45]. Moreover, since the quantization codebooks of the users are supposed to be of equal size, the resulting CSIT distortions can be assumed identical, i.e., $\tau_k^2 = \tau^2\ \forall k$. Under this assumption and equal power allocation, for large M, the SINR γ° is identical for all users and, hence, optimizing γ° is equivalent to optimizing the per-user rate $R^\circ = \log_2(1+\gamma^\circ)$ bits/s/Hz and the sum rate $\hat{R}_{\text{sum}} = KR^\circ$.

[1]The derived scaling results hold for *any* quantization codebook [45].

In the following, in particular under RVQ, we will derive the necessary scaling of the distortion τ^2 to ensure that

$$\Delta R_k - \log_2 b \overset{M\to\infty}{\longrightarrow} 0,$$

almost surely, where ΔR_k is defined in (48) and $b \geq 1$. That is, a constant rate gap of $\log_2 b$ is maintained *exactly* as $M, K \to \infty$. 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 with perfect CSIT. The choice of a rate offset $\log_2 b$ is motivated by mere mathematical convenience to avoid terms of the form 2^b and to be compliant with [45].

With this strategy we closely follow [45]. In [45, Theorem 1], the author derived an upper bound of the *ergodic* per-user gap $\Delta\tilde{R}_{\text{zf}}$ for ZF precoding with $M = K$ and *unit norm* precoding vectors under RVQ, which is given by

$$\Delta\tilde{R}_{\text{zf}} < \log_2\left(1 + \rho \cdot 2^{-\frac{B}{M-1}}\right). \tag{68}$$

We cannot directly compare the deterministic equivalents to the upper bound in (68) for two reasons, (i) under ZF precoding and $M = K$, a deterministic equivalent for the per-user rate gap does not exist and (ii) [45] considers unit norm precoding vectors, whereas in this paper we only impose a total power constraint (1). Concerning (i), at high SNR, we can use the deterministic equivalent for RZF-CDU precoding given in Corollary 5 as a good approximation for ZF precoding, since for high SNR the rates of RZF-CDU and ZF precoding converge. Regarding (ii), deriving a deterministic equivalent of the SINR under linear precoding with a unit norm power constraint on the precoding vectors is difficult, since it introduces an additional non-trivial dependence on the channel. However, it is useful to compare the accuracy of the upper bound in (68) and the deterministic equivalent $\Delta R_{k,\text{rzf}-\text{cdu}}^\circ$ in Corollary 5 at high SNR.

Figure 9, depicts the per-user rate gap as a function of the feedback bits B per user under ZF precoding at a SNR of 25 dB. We simulated the ergodic per-user rate gap $\Delta\tilde{R}_{\text{zf}}$ and $E[\Delta R_{k,\text{zf}}]$ of ZF precoding with unit norm precoding vectors and total power constraint, respectively. We compare the numerical results to the upper bound (68) and to the deterministic equivalent $\Delta R_{k,\text{rzf}-\text{cdu}}^\circ$ for $M = K = 5$ and $M = K = 10$. For both system dimensions $\Delta\tilde{R}_{\text{zf}}$ and $E[\Delta R_{k,\text{zf}}]$ are close, suggesting that our results derived under the total power constraint may be good approximations for the case of unit norm precoding vectors as well. As mentioned in [45], the accuracy of the upper bound increases with increasing B but the deterministic equivalent $\Delta R_{k,\text{rzf}-\text{cdu}}^\circ$ appears to be more accurate for both $M = K = 5$ and $M = K = 10$. In fact, for $M = K = 10$, $\Delta R_{k,\text{rzf}-\text{cdu}}^\circ$ approximates the per-user rate gap significantly more accurately than the upper bound (68) for the given SNR. We conclude that the proposed deterministic equivalent $\Delta R_{k,\text{rzf}-\text{cdu}}^\circ$ is sufficiently accurate and can be used to derive scaling laws for the optimal feedback rate.

In the following, we compare the scaling of τ^2 under RZF-CDA, RZF-CDU and ZF ($M > K$) precoding to the upper

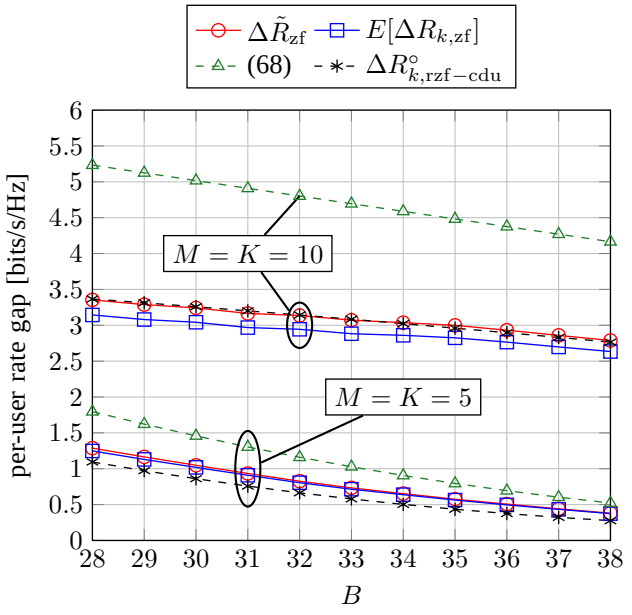

Fig. 9. ZF, per-user rate gap vs. number of bits per user with $\rho = 25$ dB, $\mathbf{\Theta}_k = \mathbf{I}_M \ \forall k$.

bound given for ZF ($M = K$) precoding in [45, Theorem 3]. For the sake of comparison, we restate [45, Theorem 3].

Theorem 4: [45, 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

$$\begin{aligned} B_{\mathrm{zf}} &= (M-1)\log_2 \rho - (M-1)\log_2(b-1) \\ &\approx \frac{M-1}{3}\rho_{\mathrm{dB}} - (M-1)\log_2(b-1). \end{aligned}$$

where $\rho_{\mathrm{dB}} = 10\log_{10}\rho$. It is also mentioned that the result in [45, Theorem 3] holds true for RZF-CDU precoding for high SNR, since ZF and RZF-CDU precoding converge for asymptotically high SNR. Furthermore, it is claimed, corroborated by simulation results, that [45, 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 [45]. The scaling given in [45, Theorem 3] is a strict upper bound on the *ergodic* per-user rate gap $E_{\mathbf{H}}[\Delta R_k]$ for all SNR and all $M = K$ under a unit norm constraint on the precoding vectors. In contrast, our approach yields a necessary scaling of τ^2 that maintains a given *instantaneous* target rate gap $\log_2 b$ *exactly* as $M, K \to \infty$ under a total power constraint. Therefore, our results are *not* upper bounds for small M, i.e., we cannot guarantee that $\Delta R_k < \log_2 b$ for small 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 (67), it follows that our results become indeed upper bounds for large M. Simulations reveal that under the derived scaling of τ^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, see Remark 5.

Consequently, a comparison of the results in [45] to our solutions is meaningful, especially for larger values of M where our results become upper bounds.

In the following section, we apply the deterministic equivalents of the per-user rate gap under RZF-CDA, RZF-CDU and ZF precoding provided in Corollaries 9, 5 and 6, respectively, to derive scaling laws for the amount of feedback necessary to achieve full multiplexing gain.

A. Channel Distortion Aware Regularized Zero-forcing Precoding

Proposition 4: Let $\mathbf{\Theta}_k = \mathbf{I}_M \ \forall k$. Then the CSIT distortion τ^2, such that the rate gap $\Delta R_{k,\mathrm{rzf-cda}}$ of user k between RZF-CDA precoding with perfect CSIT and imperfect CSIT satisfies

$$\Delta R_{k,\mathrm{rzf-cda}} - \log_2 b \overset{M\to\infty}{\longrightarrow} 0$$

almost surely, has to scale as

$$\tau^2 = \frac{\phi^{\circ}_{\mathrm{rzf-cda}}(\rho, b)}{\rho}, \tag{69}$$

$$\phi^{\circ}_{\mathrm{rzf-cda}}(\rho, b) = \frac{\rho\left[(1+\beta)b + \delta(\beta-1)\right] - \frac{1}{2b}(\delta^2 - b^2)}{(1+\beta)b + \delta(\beta-1) + \frac{1}{2b}(\delta^2 - b^2)}, \tag{70}$$

$$\delta = 1 - b + \chi(1) + \rho(\beta - 1),$$

where χ is defined in (60). With $\beta = 1$, the distortion τ^2 has to scale as

$$\tau^2 = \frac{1 + 4\rho - \frac{\delta^2}{b^2}}{3 + \frac{\delta^2}{b^2}} \frac{1}{\rho}.$$

Proof: Set $\Delta\bar{R}_{\mathrm{rzf-cda}}$ given in Corollary 9 equal to $\log_2 b$ and solve for τ^2. ∎

Although the proposed scaling of τ^2 in (69) converges to zero for asymptotically high SNR, we can approximate the term $\phi^{\circ}_{\mathrm{rzf-cda}}(\rho, b)$ in the high SNR regime.

Proposition 5: For asymptotically high SNR, the term $\phi^{\circ}_{\mathrm{rzf-cda}}(\rho, b)$ defined in (70) converges to the following limits,

$$\lim_{\rho\to\infty} \phi^{\circ}_{\mathrm{rzf-cda}}(\rho, b) = \begin{cases} b^2 - 1 & \text{if } \beta = 1 \\ b - 1 & \text{if } \beta > 1. \end{cases} \tag{71}$$

Proof: For $\beta = 1$ observe that δ scales as $2\sqrt{\rho}$. Thus, for $\rho \to \infty$, (70) converges to $b^2 - 1$. If $\beta > 1$, the term δ takes the form

$$\delta = 1 - b + (\beta - 1)\rho + |1 - \beta|\rho\,(1 + o(1)) \overset{\rho\to\infty}{\longrightarrow} 2\rho(\beta - 1) + 1 - b.$$

Therefore, for $\rho \to \infty$, (70) converges to $b - 1$, which completes the proof. ∎

Remark 7: Note that $\lim_{\rho\to\infty} \frac{\phi^{\circ}_{\mathrm{rzf-cda}}(\rho,b)}{\rho} = 0$ and thus, we require $\beta > 1$ to ensure that the limit $\rho \to \infty$ of the deterministic equivalent is well defined, see Remark 6. However, for finite SNR with the approximation in Proposition 5, we have $\tau^2 > 0$ and the scaling result holds true.

To compare Proposition 4 to [45, Theorem 3], we use the upper bound on the quantization distortion (67), i.e., $\tau^2 = 2^{-\frac{B^{\circ}_{\mathrm{rzf-cda}}}{M-1}}$, where $B^{\circ}_{\mathrm{rzf-cda}}$ is the number of feedback bits per user under RZF-CDA precoding. Thus, (69) can be rewritten as

$$B^{\circ}_{\mathrm{rzf-cda}} = (M-1)\log_2\rho - (M-1)\log_2\phi^{\circ}_{\mathrm{rzf-cda}}(\rho,b). \quad (72)$$

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 [45].

Proposition 6: Let $\mathbf{\Theta}_k = \mathbf{I}_M \ \forall k$. Then the CSIT distortion τ^2, such that the rate gap $\Delta R_{k,\mathrm{rzf-cdu}}$ with $\alpha = 1/(\beta\rho)$ of user k between RZF-CDU precoding with perfect CSIT and imperfect CSIT satisfies

$$\Delta R_{k,\mathrm{rzf-cdu}} - \log_2 b \overset{M\to\infty}{\longrightarrow} 0$$

almost surely, has to scale as

$$\tau^2 = \frac{\phi^{\circ}_{\mathrm{rzf-cdu}}(\rho,b)}{\rho},$$
$$\phi^{\circ}_{\mathrm{rzf-cdu}}(\rho,b) = \frac{(b-1)(1+m^{\circ})(\rho+\bar{m}^{\circ})}{(b-1-m^{\circ})[1-\bar{m}^{\circ}] + bm^{\circ}[1+\frac{1}{\rho}\bar{m}^{\circ}]},$$

where m° is defined in (32) and $\bar{m}^{\circ} \triangleq (1+m^{\circ})^2$.

Proof: Set $\Delta R_{k,\mathrm{rzf-cdu}}$ from Corollary 5 equal to $\log_2 b$ and solve for τ^2. ∎

An approximation of the term $\phi^{\circ}_{\mathrm{rzf-cdu}}(\rho,b)$ at high SNR is given in the following proposition.

Proposition 7: For asymptotically high SNR, $\phi^{\circ}_{\mathrm{rzf-cdu}}(\rho,b)$ converges to the following limits,

$$\lim_{\rho\to\infty}\phi^{\circ}_{\mathrm{rzf-cdu}}(\rho,b) = \begin{cases} 2(b-1) & \text{if } \beta = 1 \\ b-1 & \text{if } \beta > 1. \end{cases} \quad (73)$$

Proof of Proposition 7: For $\beta = 1$ and ρ large, m° scales as $\sqrt{\rho}$. Therefore, $\lim_{\rho\to\infty}\phi^{\circ}_{\mathrm{rzf-cdu}}(\rho,b) = 2(b-1)$. If $\beta > 1$, for large ρ, the term m° scales as $\rho(\beta-1)$. With this approximation we obtain $\lim_{\rho\to\infty}\phi^{\circ}_{\mathrm{rzf-cdu}}(\rho,b) = b-1$, which completes the proof. ∎

Applying the upper bound on the CSIT distortion under RVQ (67) with $B^{\circ}_{\mathrm{rzf-cdu}}$ bits per user, we obtain

$$B^{\circ}_{\mathrm{rzf-cdu}} = (M-1)\log_2\rho - (M-1)\log_2\phi^{\circ}_{\mathrm{rzf-cdu}}(\rho,b). \quad (74)$$

C. Zero-forcing Precoding

The following results are only valid for $\beta > 1$ and thus, they cannot be compared to [45, Theorem 3] which are derived under the assumption $M = K$. However, for high SNR the results for the RZF-CDU precoder are a good approximation for the ZF precoder as well, even for $\beta = 1$.

Corollary 10: Let $\beta > 1$ and $\mathbf{\Theta}_k = \mathbf{I}_M \ \forall k$. To maintain a rate offset $\Delta R_{k,\mathrm{zf}}$ such that

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

almost surely, the distortion τ^2 has to scale according to

$$\tau^2 = \frac{\phi^{\circ}_{\mathrm{zf}}(\rho,b)}{\rho},$$
$$\phi^{\circ}_{\mathrm{zf}}(\rho,b) = \frac{(b-1)[1+\rho(\beta-1)]}{1-b+(\beta-1)[\rho+b]}. \quad (75)$$

Proof: From Corollary 6, set $\Delta R^{\circ}_{\mathrm{zf}} = \log_2 b$ and solve for τ^2. ∎

Proposition 8: For asymptotically high SNR, $\phi^{\circ}_{\mathrm{zf}}(\rho,b)$ in (75) converges to

$$\lim_{\rho\to\infty}\phi^{\circ}_{\mathrm{zf}}(\rho,b) = b-1. \quad (76)$$

Proof: From (75), the result is immediate. ∎

Under RVQ with B°_{zf} feedback bits per user, we have

$$B^{\circ}_{\mathrm{zf}} = (M-1)\log_2\rho - (M-1)\log_2\phi^{\circ}_{\mathrm{zf}}(\rho,b). \quad (77)$$

D. Discussion and Numerical Results

At this point, we can draw the following conclusions. The optimal scaling of the CSIT distortion τ^2 is lower for $\beta = 1$ compared to $\beta > 1$. For $\beta = 1$, the optimal scaling of the feedback bits $B^{\circ}_{\mathrm{rzf-cda}}$, $B^{\circ}_{\mathrm{rzf-cdu}}$ and B for ZF in [45, Theorem 3] are different, even at high SNR. In fact, for large M, under RZF-CDU precoding and ZF precoding, the upper bound in [45, Theorem 3] appears to be too pessimistic in the scaling of the feedback bits. From (74) and (73), a more accurate choice may be

$$B^{\circ}_{\mathrm{rzf-cdu}} = (M-1)\log_2\rho - (M-1)\log_2(2(b-1)), \quad (78)$$

i.e., $M-1$ bits less than proposed in [45, Theorem 3]. However, recall that (78) 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 $(M-1)\log_2(\frac{b+1}{2})$ bits *less* than the RZF-CDU and ZF precoder and $(M-1)\log_2(b+1)$ bits *less* than the scaling proposed in [45, Theorem 3].

In contrast, for $\beta > 1$ and high SNR, we have $B^{\circ}_{\mathrm{rzf-cda}} = B^{\circ}_{\mathrm{rzf-cdu}} = B^{\circ}_{\mathrm{zf}}$. Intuitively, the reason is that, for $\beta > 1$, the channel matrix is well conditioned and the RZF and ZF precoders perform similarly. Therefore, both schemes are equally sensitive to imperfect CSIT and thus the scaling of τ^2 is the same for high SNR.

Note 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 (see [47]) as long as they can be modeled as additive noise (6). 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 10 and 11 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 parameter $\alpha^{\star\circ}$, the minimum number of feedback bits is set to one.

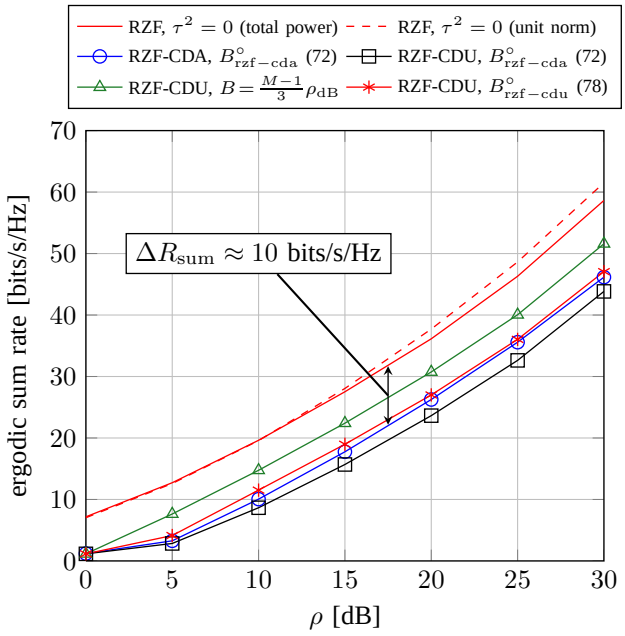

Fig. 10. RZF, ergodic sum rate vs. SNR under RZF precoding and RVQ with B feedback bits per user, where B is chosen to maintain a sum rate offset of $K \log_2 b = 10$, $\mathbf{\Theta}_k = \mathbf{I}_M \ \forall k$ and $M = K = 10$.

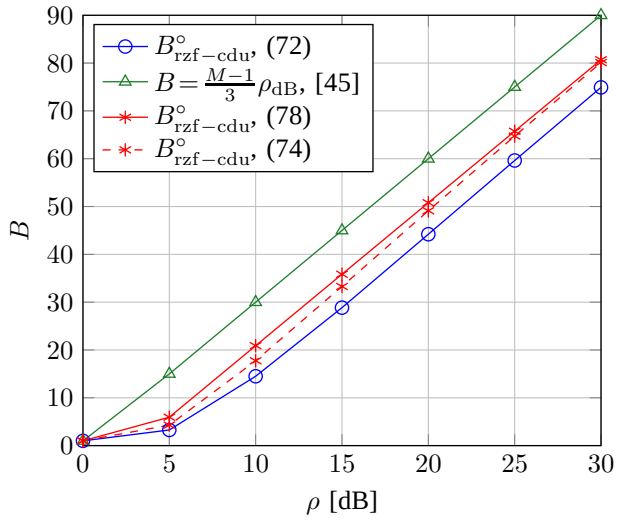

Fig. 11. RZF, B feedback bits per user vs. SNR, with B to maintain a sum rate offset of $K \log_2 b = 10$ and $\mathbf{\Theta}_k = \mathbf{I}_M \ \forall k$, $M = K = 10$.

In Figure 10, we plot the ergodic sum rate for RZF precoding under perfect CSIT with total power constraint (red solid lines) and unit norm constraint on the precoding vectors (red dashed line). We observe, that the sum rate under unit norm constraint is slightly larger at high SNR, suggesting that our scaling results for RZF precoding derived under a total power constraint become inaccurate under the unit norm constraint at high SNR. Hence, one has to be cautious when comparing the scaling in [45, Theorem 3] directly to the scaling derived with the large system approximations at high SNR. From Figure 10, we further 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 (72) and the high SNR approximation in (78) under RZF-CDA and RZF-CDU precoding, respectively, (ii) given an equal number of feedback bits (72), 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 K bits/s/Hz, the proposed feedback scaling of $B = \frac{M-1}{3}\rho_{\mathrm{dB}}$ for unit norm precoding vectors [45] is very pessimistic, since the sum rate offset to RZF with total power constraint and unit norm constraint is about 6 bits/s/Hz and 7 bits/s/Hz at 20 dB, respectively.

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 [45, Theorem 3] appears to be less accurate under a total power constraint than our large system approximation in (72).

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_t channel uses for UL training and $T - T_t$ channel uses for coherent transmission in the DL. Note that in order to coherently decode the information symbols, the users need to know their effective (precoded) channels. This is usually accomplished by a dedicated training phase (using precoded pilots) in the DL prior to the data transmission. As shown in [48], 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 the DL training.

In the considered TDD system, the imperfections in the CSIT are caused by (i) channel estimation errors in the UL, (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) for uncorrelated channels ($\mathbf{\Theta}_k = \mathbf{I}_M \ \forall k$).

A. Uplink Training Phase

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

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

where we assumed perfect reciprocity of UL and DL channels and P_{ul} is the average available transmit power at the receivers. That is, the UL and DL channel coefficients are equal

and the UL noise $\mathbf{n}_k = [n_1, n_2, \ldots, n_M]^{\mathsf{T}}$ is assumed identical for all users and statistically equivalent to its DL analog. Subsequently, the transmitter performs an MMSE estimation of each channel coefficient $h_{ij} \sim \mathcal{CN}(0,1)$ ($i = 1, \ldots, K$, $j = 1, \ldots, M$). Due to the orthogonality property of the MMSE estimation [49], the estimates $\hat{h}_{ij}$ of h_{ij} and the corresponding estimation errors $\tilde{h}_{ij} = h_{ij} - \hat{h}_{ij}$ are uncorrelated and i.i.d. complex Gaussian distributed. Hence, we can write

$$\hat{h}_{ij} = h_{ij} + \tilde{h}_{ij},$$

where h_{ij} and $\tilde{h}_{ij}$ are independent with zero mean and variance $1 - \tau^2$ and τ^2, respectively. The variance τ^2 of the estimation error $\tilde{h}_{ij}$ is given by [47]

$$\tau^2 = \frac{1}{1 + T_t \rho_{ul}}, \tag{79}$$

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

B. Optimization of Channel Training

We focus on equal power allocation among the users, i.e., $p_k = P/K \ \forall k$, because it is optimal for large M and $\tau_k^2 = \tau^2 \ \forall k$, see Section V-B. Since T_t channel uses have already been consumed to train the transmitter about the user channels, there remains an interval of length $T - T_t$ for DL data transmission and thus we have the pre-log factor $1 - T_t/T$. The net sum rate approximation reads

$$\hat{R}_{\text{sum}} = K \left(1 - \frac{T_t}{T}\right) \log\left(1 + \gamma_k^{\circ}\right). \tag{80}$$

To compute the training length T_t that maximizes the net sum rate approximation (80), we substitute $\gamma_{k,\text{zf}}^{\circ}$ from Corollary 4 into (80) and the approximated net sum rate $\hat{R}_{\text{sum}}^{\text{zf}}$ under ZF precoding takes the form

$$\hat{R}_{\text{sum}}^{\text{zf}} = K \left(1 - \frac{T_{t,\text{zf}}}{T}\right) \log\left(1 + \frac{1 - \tau^2}{\tau^2 + \frac{1}{\rho_{dl}}}(\beta - 1)\right), \tag{81}$$

where $\rho_{dl} \triangleq P/\sigma^2$. Similarly, for RZF-CDA precoding the approximated net sum rate $\hat{R}_{\text{sum}}^{\text{rzf}}$ reads

$$\hat{R}_{\text{sum}}^{\text{rzf}} = K \left(1 - \frac{T_{t,\text{rzf}}}{T}\right) \log\left(1 + \gamma_{\text{rzf}}^{\circ}\right), \tag{82}$$

where $\gamma_{\text{rzf}}^{\circ}$ is given in Corollary 8.

Substituting (79) into (81) and (82), we obtain

$$\hat{R}_{\text{sum}}^{\text{zf}} = K \left(1 - \frac{T_{t,\text{zf}}}{T}\right) \log\left(1 + \frac{T_{t,\text{zf}} \rho_{ul} (\beta - 1)}{1 + T_{t,\text{zf}} \frac{\rho_{ul}}{\rho_{dl}} + \frac{1}{\rho_{dl}}}\right), \tag{83}$$

$$\hat{R}_{\text{sum}}^{\text{rzf}} = K \left(1 - \frac{T_{t,\text{rzf}}}{T}\right) \log\left(\frac{1}{2} + \frac{1}{2}\omega \rho_{dl} (\beta - 1) + \frac{\chi(\omega)}{2}\right), \tag{84}$$

$$\chi(\omega) = \sqrt{(\beta - 1)^2 \omega^2 \rho_{dl}^2 + 2\omega \rho_{dl} (1 + \beta) + 1}, \tag{85}$$

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

For $\beta > 1$ under ZF precoding and $\beta \geq 1$ for RZF-CDA precoding, it is easy to verify that the functions $\hat{R}_{\text{sum}}^{\text{zf}}$ and $\hat{R}_{\text{sum}}^{\text{rzf}}$ are strictly concave in $T_{t,\text{zf}}$ and $T_{t,\text{rzf}}$ in the interval $[K, T]$, respectively, where K is the minimum amount of training required, due to the orthogonality constraint of the pilot sequences. Therefore, we can apply standard convex optimization algorithms [50] to evaluate

$$T_{t,\text{zf}}^{\star\circ} = \underset{K \leq T_{t,\text{zf}} \leq T}{\arg\max}\ \hat{R}_{\text{sum}}^{\text{zf}}, \tag{86}$$

$$T_{t,\text{rzf}}^{\star\circ} = \underset{K \leq T_{t,\text{rzf}} \leq T}{\arg\max}\ \hat{R}_{\text{sum}}^{\text{rzf}}. \tag{87}$$

In the following, we derive approximate explicit solutions to (86) and (87) 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) ρ_{dl} varies, while ρ_{ul} 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 ρ_{dl}/ρ_{ul}: We derive approximate, but explicit, solutions for the optimal training intervals $T_{t,\text{zf}}^{\star\circ}, T_{t,\text{rzf}}^{\star\circ}$ in the high SNR regime and derive their limiting values for asymptotically low SNR.

a) High SNR Regime: An approximate closed form solution to (86) and (87) is summarized in the following proposition.

Proposition 9: Let ρ_{dl}, ρ_{ul} be large with $c = \rho_{dl}/\rho_{ul}$ constant. Then, an approximation of the sum rate maximizing amount of channel training $T_{t,\text{zf}}^{\star\circ}$ and $T_{t,\text{rzf}}^{\star\circ}$ under ZF and RZF-CDA precoding is given by

$$T_{t,\text{zf}}^{\star\circ} = \max\left[\frac{c}{2}\sqrt{1 + 2\frac{2T + c}{c\bar{R}_{\text{zf}}^{\circ}}} - \frac{c}{2}, K\right], \tag{88}$$

$$T_{t,\text{rzf}}^{\star\circ} = \begin{cases} \max\left[\frac{c}{2}\sqrt{1 + \frac{2T + c}{c\bar{R}_{\text{rzf}}^{\circ}}} - \frac{c}{2}, K\right] & \text{if } \beta = 1, \\ \max\left[\frac{c}{2}\sqrt{1 + 2\frac{2T + c}{c\bar{R}_{\text{rzf}}^{\circ}}} - \frac{c}{2}, K\right] & \text{if } \beta > 1, \end{cases} \tag{89}$$

where $\bar{R}_{\text{zf}}^{\circ} = \log(1 + \rho_{dl}(\beta - 1))$ and $\bar{R}_{\text{rzf}}^{\circ} = \log(\frac{1}{2} + \frac{1}{2}\rho_{dl}(\beta - 1) + \frac{\chi(1)}{2})$.

Proof: The proof is presented in Appendix V. ∎

Thus, for a fixed DL SNR ρ_{dl}, the optimal training intervals scale as $T_{t,\text{zf}}^{\star\circ}, T_{t,\text{rzf}}^{\star\circ} \sim \sqrt{T}$. Likewise, for a constant T, the optimal training intervals scale as $T_{t,\text{zf}}^{\star\circ}, T_{t,\text{rzf}}^{\star\circ} \sim 1/\sqrt{\log(\rho_{dl})}$. Under ZF precoding the same scaling has been reported in [51]–[53]. From this scaling it is clear that, as $\rho_{dl} \to \infty$, $T_t^{\star\circ}$ tends to K, the minimum amount of training.

Moreover, for $\beta > 1$, $\bar{R}_{\text{rzf}}^{\circ} \geq \bar{R}_{\text{zf}}^{\circ}$ 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 10: Let $\rho_{dl}, \rho_{ul} \to 0$ with constant ratio $c = \rho_{dl}/\rho_{ul}$ and $T \geq 2K$. Then, the sum rate maximizing amount of channel training $T_{t,\text{zf}}^{\star\circ}$ and $T_{t,\text{rzf}}^{\star\circ}$ under ZF and RZF-CDA precoding converges to

$$\lim_{\rho_{dl} \to 0} T_{t,\text{zf}}^{\star\circ} = \lim_{\rho_{dl} \to 0} T_{t,\text{rzf}}^{\star\circ} = \frac{T}{2}. \tag{90}$$

Proof: Applying $\log(1+x) = x+O(x^2)$ and $\rho_{ul} = \rho_{dl}/c$, equations (83) and (84) take the form

$$\hat{R}_{\text{sum}}^{\text{zf}} = K\left(1-\frac{T_{t,\text{zf}}}{T}\right)\frac{T_{t,\text{zf}}(\beta-1)}{c}\rho_{dl}^2 + O(\rho_{dl}^4), \quad (91)$$

$$\hat{R}_{\text{sum}}^{\text{rzf}} = K\left(1-\frac{T_{t,\text{rzf}}}{T}\right)\frac{T_{t,\text{rzf}}\beta}{c}\rho_{dl}^2 + O(\rho_{dl}^4). \quad (92)$$

Maximizing equations (91) and (92) with respect to $T_{t,\text{zf}}$ and $T_{t,\text{rzf}}$, respectively, yields (90). Since, by definition, we assume orthogonal pilot sequences, hence $T_t \geq K$, the result (90) implies that $T \geq 2K$, which completes the proof. ■

For ZF precoding, the limit has also been reported in [54].

2) Case 2: $\rho_{dl} \gg \rho_{ul}$ with finite ρ_{ul}: This scenario models a high capacity DL channel where the primary sum rate loss stems from the inaccurate CSIT estimate due to limited-rate UL signaling caused, e.g., by a finite transmit power of the users. Thus, the system becomes interference-limited and the optimal amount of channel training under ZF precoding is given in the following proposition.

Proposition 11: Let $\rho_{dl} \to \infty$ and ρ_{ul} finite. Then the (approximated) sum rate maximizing amount of channel training $T_{t,\text{zf}}^{\star\circ}$ is given by

$$T_{t,\text{zf}}^{\star\circ} = \frac{1}{\rho_{ul}(\beta-1)}\left(\frac{a}{\mathcal{W}(ae)}-1\right), \quad (93)$$

where $\mathcal{W}(z)$ is the Lambert W-function.

Proof: For ZF precoding and $\rho_{dl}\to\infty$, the sum rate (83) can be approximated as

$$\hat{R}_{\text{sum}}^{\text{zf}} \approx K\left(1-\frac{T_{t,\text{zf}}}{T}\right)\log\left(1+T_{t,\text{zf}}\rho_{ul}(\beta-1)\right). \quad (94)$$

Setting the derivative of (94) with respect to $T_{t,\text{zf}}$ to zero, yields

$$\log(a/\omega(T_{t,\text{zf}})) = \omega(T_{t,\text{zf}}) - 1, \quad (95)$$

where $a \triangleq \rho_{ul}T(\beta-1)+1$ and $\omega(T_{t,\text{zf}}) \triangleq (Ta)/[T+T_{t,\text{zf}}(a-1)]$. Equation (95) can be written as

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

Notice that $\omega(T_{t,\text{zf}}) = \mathcal{W}(ae)$. Thus, solving $\omega(T_{t,\text{zf}}) = \mathcal{W}(ae)$ for $T_{t,\text{zf}}$ yields (93). ■

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

For RZF-CDA precoding, no accurate closed-form solution to (87) has yet been found.

C. Numerical Results

In Figure 12, we compare the approximated optimal training intervals $T_{t,\text{zf}}^{\star\circ}, T_{t,\text{rzf}}^{\star\circ}$ to $T_{t,\text{zf}}^{\star}, T_{t,\text{rzf}}^{\star}$ computed via exhaustive search and averaged over 1 000 independent channel realizations. The regularization parameter α is computed using the large system approximation $\alpha^{\star\circ}$ in (55). Figure 12 shows that the approximate solutions $T_{t,\text{zf}}^{\star\circ}, T_{t,\text{rzf}}^{\star\circ}$ become very accurate for $K = 16$. Moreover, it can be observed that the approximations in (88) and (89) match very well. Further note that for $\frac{M}{K} = 2$, ZF and RZF-CDA need approximately the same amount of training, as predicted by equations (88) and (89).

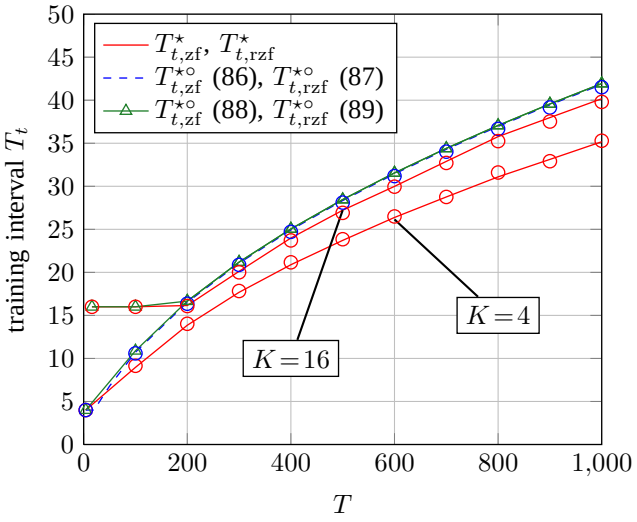

Fig. 12. ZF and RZF-CDA, optimal amount of training with $\frac{M}{K} = 2$, $\rho_{dl} = 20$ dB, $\rho_{ul} = 10$ dB, $\mathbf{\Theta}_k = \mathbf{I}_M$ $\forall k$, RZF is indicated by circle marks.

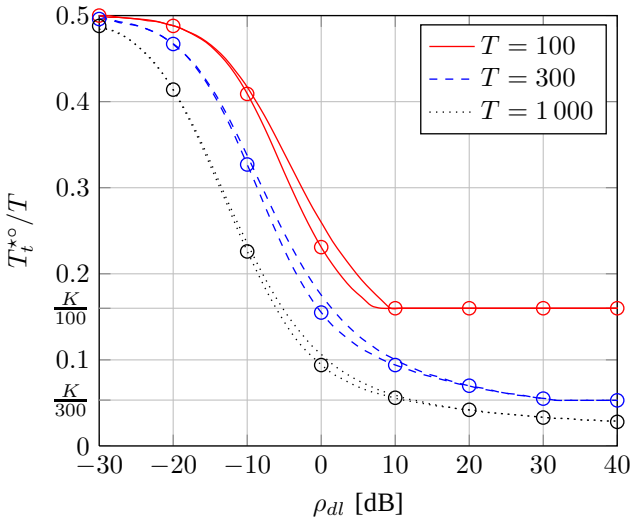

Fig. 13. ZF and RZF-CDA, optimal relative amount of training $T_t^{\star\circ}/T$ vs. ρ_{dl} with $M = 32$, $K = 16$, $\rho_{dl}/\rho_{ul} = 10$, $\mathbf{\Theta}_k = \mathbf{I}_M$ $\forall k$, RZF is indicated by circle marks.

Figure 13 depicts the optimal relative amount of training $T_t^{\star\circ}/T$ for ZF and RZF-CDA precoding. We observe that $T_t^{\star\circ}/T$ decreases with increasing SNR as $1/\sqrt{\log(\rho_{dl})}$. That is, for increasing SNR, the estimation becomes more accurate and resources for channel training are reallocated to data transmission. Furthermore, $T_t^{\star\circ}/T$ saturates at K/T due to the orthogonality constraint on the pilot sequences. As expected from (88) and (89), we observe that the optimal amount of training is less for RZF-CDA than for ZF precoding. Moreover, the relative amount of training $T_t^{\star\circ}/T$ for both ZF and RZF-CDA converges at low SNR to $1/2$ and at high SNR to the minimum amount of training K, as predicted by the theoretical analysis.

Figure 14 shows the ergodic sum rate under ZF precoding with fixed UL SNR $\rho_{ul} = 5$ dB for various training intervals. We observe (i) no significant difference in the performance

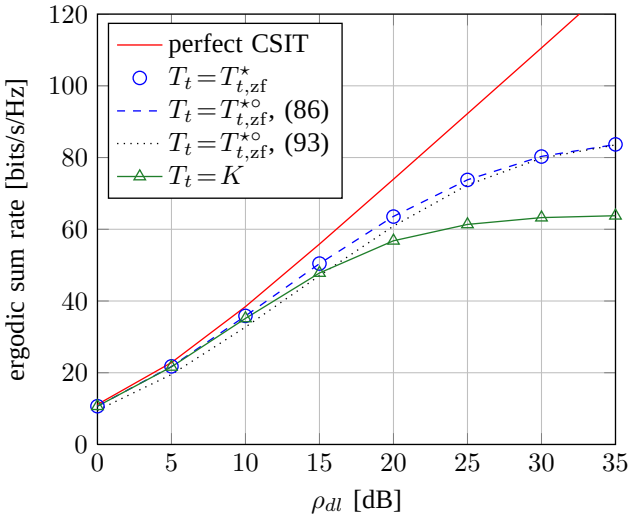

Fig. 14. ZF, ergodic sum rate vs. downlink SNR with $M = 32$, $K = 16$, $\rho_{ul} = 5$ dB, $\mathbf{\Theta}_k = \mathbf{I}_M \ \forall k$ and $T = 1\,000$.

of the schemes employing either optimal training $T^{\star}_{t,\mathrm{zf}}$, computed via exhaustive search, or $T^{\star\circ}_{t,\mathrm{zf}}$ obtained from a convex optimization of the large system approximation (83), (ii) a small performance loss at low and medium SNR of the (high-SNR) approximation of $T^{\star\circ}_{t,\mathrm{zf}}$ in (93) and (iii) a significant performance loss if the minimum training interval $T_{t,\mathrm{zf}} = K$ is used for all SNR. We conclude that our approximation in (93) achieves very good performance and can therefore be utilized to compute $T_{t,\mathrm{zf}}$ very efficiently.

VIII. Conclusion

In this paper, we presented a consistent framework for the study of ZF and RZF precoding schemes based on the theory of large dimensional random matrices. The tools from RMT allowed us to consider a very realistic channel model accounting for per-user channel correlation as well as individual channel gains for each link. The system performance under this general type of channel is extremely difficult to study for finite dimensions but becomes feasible by assuming large system dimensions. Simulation results indicated that these approximations are very accurate even for small system dimensions and reveal the deterministic dependence of the system performance on several important system parameters, such as the transmit correlation, signal powers, SNR, and CSIT quality. Applied to practical optimization problems, the deterministic approximations lead to important insights into the system behavior, which are consistent with previous results, but go further and extend them to more realistic channel models and other linear precoding techniques. Furthermore, the proposed channel-independent performance approximations can be used to simulate the system behavior without having to carry out extensive Monte Carlo simulations.

Appendix I
Proof of Theorem 1

The proof is structured as follows: In Appendix I-A, we prove that $m_{\mathbf{B}_N,\mathbf{Q}_N}(z) - \frac{1}{N}\mathrm{tr}\mathbf{D}^{-1} \overset{N\to\infty}{\longrightarrow} 0$ almost surely, where $\mathbf{D}$ is an auxiliary random variable involving the terms $m_{\mathbf{B}_N,\mathbf{\Theta}_i}(z)$. Appendix I-B shows that the sequence $\{e^{(k)}_{N,i}(z)\}$ defined by (12) converges to $e_{N,i}$ (11) as $k \to \infty$, if properly initialized. Finally, in Appendix I-C we demonstrate that $e_{N,i}$ satisfies $|m_{\mathbf{B}_N,\mathbf{\Theta}_i} - e_{N,i}| \overset{N\to\infty}{\longrightarrow} 0$, almost surely.

A. Convergence to an Auxiliary Variable

The objective is to approximate the random variable $m_{\mathbf{B}_N,\mathbf{Q}_N}(z)$ by an appropriate functional $\frac{1}{N}\mathrm{tr}\mathbf{D}^{-1}$ such that

$$\frac{1}{N}\mathrm{tr}\mathbf{Q}_N\left(\mathbf{B}_N - z\mathbf{I}_N\right)^{-1} - \frac{1}{N}\mathrm{tr}\mathbf{D}^{-1} \overset{N\to\infty}{\longrightarrow} 0, \tag{96}$$

almost surely. Take $z \in \mathbb{C}^+$. From (96) we proceed by applying Lemma 2 and obtain

$$\mathbf{Q}_N\left(\mathbf{B}_N - z\mathbf{I}_N\right)^{-1} - \mathbf{D}^{-1} = \mathbf{D}^{-1}\left[\mathbf{D} - \left(\mathbf{X}_N^{\mathsf{H}}\mathbf{X}_N + \mathbf{S}_N - z\mathbf{I}_N\right)\mathbf{Q}_N^{-1}\right]\mathbf{Q}_N\left(\mathbf{B}_N - z\mathbf{I}_N\right)^{-1}. \tag{97}$$

We choose $\mathbf{D}$ as

$$\mathbf{D} = \left(\mathbf{R} + \mathbf{S}_N - z\mathbf{I}_N\right)\mathbf{Q}_N^{-1}, \tag{98}$$

where $\mathbf{R}$ is to be determined later, and obtain

$$\begin{aligned}&\mathbf{Q}_N\left(\mathbf{B}_N - z\mathbf{I}_N\right)^{-1} - \mathbf{D}^{-1}\\ &= \mathbf{D}^{-1}\mathbf{R}\left(\mathbf{B}_N - z\mathbf{I}_N\right)^{-1} - \mathbf{D}^{-1}\mathbf{X}_N^{\mathsf{H}}\mathbf{X}_N\left(\mathbf{B}_N - z\mathbf{I}_N\right)^{-1}.\end{aligned}$$

Consider the term $\mathbf{D}^{-1}\mathbf{X}_N^{\mathsf{H}}\mathbf{X}_N\left(\mathbf{B}_N - z\mathbf{I}_N\right)^{-1}$. Taking the trace, together with $\mathbf{X}_N^{\mathsf{H}}\mathbf{X}_N = \sum_{i=1}^n \mathbf{\Psi}_i\mathbf{y}_i\mathbf{y}_i^{\mathsf{H}}\mathbf{\Psi}_i^{\mathsf{H}}$, we have

$$\begin{aligned}&\frac{1}{N}\mathrm{tr}\mathbf{D}^{-1}\mathbf{X}_N^{\mathsf{H}}\mathbf{X}_N\left(\mathbf{B}_N - z\mathbf{I}_N\right)^{-1}\\ &= \frac{1}{N}\mathrm{tr}\mathbf{D}^{-1}\sum_{i=1}^n \mathbf{\Psi}_i\mathbf{y}_i\mathbf{y}_i^{\mathsf{H}}\mathbf{\Psi}_i^{\mathsf{H}}\left(\mathbf{B}_N - z\mathbf{I}_N\right)^{-1}\\ &= \frac{1}{N}\sum_{i=1}^n \mathbf{y}_i^{\mathsf{H}}\mathbf{\Psi}_i^{\mathsf{H}}\left(\mathbf{B}_N - z\mathbf{I}_N\right)^{-1}\mathbf{D}^{-1}\mathbf{\Psi}_i\mathbf{y}_i.\end{aligned}$$

Denoting $\mathbf{B}_{[i]} = \mathbf{B}_N - \mathbf{\Psi}_i\mathbf{y}_i\mathbf{y}_i^{\mathsf{H}}\mathbf{\Psi}_i^{\mathsf{H}}$ and applying Lemma 1, we obtain

$$\begin{aligned}&\frac{1}{N}\mathrm{tr}\mathbf{D}^{-1}\mathbf{X}_N^{\mathsf{H}}\mathbf{X}_N\left(\mathbf{B}_N - z\mathbf{I}_N\right)^{-1}\\ &= \frac{1}{N}\sum_{i=1}^n \frac{\mathbf{y}_i^{\mathsf{H}}\mathbf{\Psi}_i^{\mathsf{H}}\left(\mathbf{B}_{[i]} - z\mathbf{I}_N\right)^{-1}\mathbf{D}^{-1}\mathbf{\Psi}_i\mathbf{y}_i}{1 + \mathbf{y}_i^{\mathsf{H}}\mathbf{\Psi}_i^{\mathsf{H}}\left(\mathbf{B}_{[i]} - z\mathbf{I}_N\right)^{-1}\mathbf{\Psi}_i\mathbf{y}_i}.\end{aligned}$$

Therefore, the left-hand side of (96) takes the form

$$\begin{aligned}&\frac{1}{N}\mathrm{tr}\mathbf{Q}_N\left(\mathbf{B}_N - z\mathbf{I}_N\right)^{-1} - \frac{1}{N}\mathrm{tr}\mathbf{D}^{-1}\\ &= \frac{1}{N}\mathrm{tr}\mathbf{D}^{-1}\mathbf{R}\left(\mathbf{B}_N - z\mathbf{I}_N\right)^{-1}\\ &\quad - \frac{1}{N}\sum_{i=1}^n \frac{\mathbf{y}_i^{\mathsf{H}}\mathbf{\Psi}_i^{\mathsf{H}}\left(\mathbf{B}_{[i]} - z\mathbf{I}_N\right)^{-1}\mathbf{D}^{-1}\mathbf{\Psi}_i\mathbf{y}_i}{1 + \mathbf{y}_i^{\mathsf{H}}\mathbf{\Psi}_i^{\mathsf{H}}\left(\mathbf{B}_{[i]} - z\mathbf{I}_N\right)^{-1}\mathbf{\Psi}_i\mathbf{y}_i}.\end{aligned} \tag{99}$$

The choice of an appropriate value for $\mathbf{R}$, such that (96) is satisfied, requires some intuition. From Lemma 4 we know that $\mathbf{y}_i^{\mathsf{H}}\mathbf{\Psi}_i^{\mathsf{H}}\left(\mathbf{B}_{[i]} - z\mathbf{I}_N\right)^{-1}\mathbf{\Psi}_i\mathbf{y}_i - \frac{1}{N}\mathrm{tr}\mathbf{\Theta}_i\left(\mathbf{B}_{[i]} - z\mathbf{I}_N\right)^{-1} \overset{N\to\infty}{\longrightarrow} 0$, almost surely. Then, from Lemma 8, we surely have

$$\frac{1}{N}\mathrm{tr}\mathbf{\Theta}_i\left(\mathbf{B}_{[i]} - z\mathbf{I}_N\right)^{-1} - \frac{1}{N}\mathrm{tr}\mathbf{\Theta}_i\left(\mathbf{B}_N - z\mathbf{I}_N\right)^{-1} \overset{N\to\infty}{\longrightarrow} 0.$$

From the previous arguments, $\mathbf{R}$ will be chosen as

$$\mathbf{R} = \frac{1}{N}\sum_{i=1}^{n} \frac{\mathbf{\Theta}_i}{1+\frac{1}{N}\mathrm{tr}\mathbf{\Theta}_i\left(\mathbf{B}_N - z\mathbf{I}_N\right)^{-1}}. \tag{100}$$

Note that $\mathbf{R}$ is random since it depends on $\mathbf{B}_N$. The remainder of this subsection proves (96) for the specific choice of $\mathbf{R}$ in (100). Substituting (100) into (99) we obtain

$$\begin{aligned} w_N \triangleq w_{\mathbf{Q}_N} \triangleq \frac{1}{N}\mathrm{tr}\mathbf{Q}_N\left(\mathbf{B}_N - z\mathbf{I}_N\right)^{-1} - \frac{1}{N}\mathrm{tr}\mathbf{D}^{-1} \quad (101)\\ = \frac{1}{N}\sum_{i=1}^{n} \frac{\frac{1}{N}\mathrm{tr}\mathbf{\Theta}_i\left(\mathbf{B}_N - z\mathbf{I}_N\right)^{-1}\mathbf{D}^{-1}}{1+\frac{1}{N}\mathrm{tr}\mathbf{\Theta}_i\left(\mathbf{B}_N - z\mathbf{I}_N\right)^{-1}} \\ - \frac{1}{N}\sum_{i=1}^{n} \frac{\mathbf{y}_i^{\mathsf{H}}\mathbf{\Psi}_i^{\mathsf{H}}\left(\mathbf{B}_{[i]} - z\mathbf{I}_N\right)^{-1}\mathbf{D}^{-1}\mathbf{\Psi}_i\mathbf{y}_i}{1+\mathbf{y}_i^{\mathsf{H}}\mathbf{\Psi}_i^{\mathsf{H}}\left(\mathbf{B}_{[i]} - z\mathbf{I}_N\right)^{-1}\mathbf{\Psi}_i\mathbf{y}_i}. \quad (102)\end{aligned}$$

In order to prove that $w_N \overset{N\to\infty}{\longrightarrow} 0$, almost surely, we divide the left-hand side of (102) into $4n$ terms, i.e.,

$$w_N = \frac{1}{N}\sum_{i=1}^{n}\left[d_i^{(1)} + d_i^{(2)} + d_i^{(3)} + d_i^{(4)}\right]. \tag{103}$$

It is then easier to show that each $d_i^{(l)}$, $(l=1,2,3,4)$, converges to zero, sufficiently fast, as $N\to\infty$, which will imply $w_N \overset{N\to\infty}{\longrightarrow} 0$, almost surely. The $d_i^{(l)}$ are chosen as

$$\begin{aligned} d_i^{(1)} &= \frac{\mathbf{y}_i^{\mathsf{H}}\mathbf{\Psi}_i^{\mathsf{H}}\left(\mathbf{B}_{[i]} - z\mathbf{I}_N\right)^{-1}\mathbf{D}_{[i]}^{-1}\mathbf{\Psi}_i\mathbf{y}_i}{1+\mathbf{y}_i^{\mathsf{H}}\mathbf{\Psi}_i^{\mathsf{H}}\left(\mathbf{B}_{[i]} - z\mathbf{I}_N\right)^{-1}\mathbf{\Psi}_i\mathbf{y}_i} \\ &\quad - \frac{\mathbf{y}_i^{\mathsf{H}}\mathbf{\Psi}_i^{\mathsf{H}}\left(\mathbf{B}_{[i]} - z\mathbf{I}_N\right)^{-1}\mathbf{D}^{-1}\mathbf{\Psi}_i\mathbf{y}_i}{1+\mathbf{y}_i^{\mathsf{H}}\mathbf{\Psi}_i^{\mathsf{H}}\left(\mathbf{B}_{[i]} - z\mathbf{I}_N\right)^{-1}\mathbf{\Psi}_i\mathbf{y}_i} \\ d_i^{(2)} &= \frac{\frac{1}{N}\mathrm{tr}\mathbf{\Theta}_i\left(\mathbf{B}_{[i]} - z\mathbf{I}_N\right)^{-1}\mathbf{D}_{[i]}^{-1}}{1+\mathbf{y}_i^{\mathsf{H}}\mathbf{\Psi}_i^{\mathsf{H}}\left(\mathbf{B}_{[i]} - z\mathbf{I}_N\right)^{-1}\mathbf{\Psi}_i\mathbf{y}_i} \\ &\quad - \frac{\mathbf{y}_i^{\mathsf{H}}\mathbf{\Psi}_i^{\mathsf{H}}\left(\mathbf{B}_{[i]} - z\mathbf{I}_N\right)^{-1}\mathbf{D}_{[i]}^{-1}\mathbf{\Psi}_i\mathbf{y}_i}{1+\mathbf{y}_i^{\mathsf{H}}\mathbf{\Psi}_i^{\mathsf{H}}\left(\mathbf{B}_{[i]} - z\mathbf{I}_N\right)^{-1}\mathbf{\Psi}_i\mathbf{y}_i} \\ d_i^{(3)} &= \frac{\frac{1}{N}\mathrm{tr}\mathbf{\Theta}_i\left(\mathbf{B}_N - z\mathbf{I}_N\right)^{-1}\mathbf{D}^{-1}}{1+\mathbf{y}_i^{\mathsf{H}}\mathbf{\Psi}_i^{\mathsf{H}}\left(\mathbf{B}_{[i]} - z\mathbf{I}_N\right)^{-1}\mathbf{\Psi}_i\mathbf{y}_i} \\ &\quad - \frac{\frac{1}{N}\mathrm{tr}\mathbf{\Theta}_i\left(\mathbf{B}_{[i]} - z\mathbf{I}_N\right)^{-1}\mathbf{D}_{[i]}^{-1}}{1+\mathbf{y}_i^{\mathsf{H}}\mathbf{\Psi}_i^{\mathsf{H}}\left(\mathbf{B}_{[i]} - z\mathbf{I}_N\right)^{-1}\mathbf{\Psi}_i\mathbf{y}_i} \\ d_i^{(4)} &= \frac{\frac{1}{N}\mathrm{tr}\mathbf{\Theta}_i\left(\mathbf{B}_N - z\mathbf{I}_N\right)^{-1}\mathbf{D}^{-1}}{1+\frac{1}{N}\mathrm{tr}\mathbf{\Theta}_i\left(\mathbf{B}_N - z\mathbf{I}_N\right)^{-1}} \\ &\quad - \frac{\frac{1}{N}\mathrm{tr}\mathbf{\Theta}_i\left(\mathbf{B}_N - z\mathbf{I}_N\right)^{-1}\mathbf{D}^{-1}}{1+\mathbf{y}_i^{\mathsf{H}}\mathbf{\Psi}_i^{\mathsf{H}}\left(\mathbf{B}_{[i]} - z\mathbf{I}_N\right)^{-1}\mathbf{\Psi}_i\mathbf{y}_i}, \end{aligned}$$

where we defined

$$\mathbf{D}_{[i]}^{-1} = \mathbf{Q}_N\left(\frac{1}{N}\sum_{i=1}^{n}\frac{\mathbf{\Theta}_i}{1+m_{\mathbf{B}_{[i]},\mathbf{\Theta}_i}(z)} - z\mathbf{I}_N + \mathbf{S}_N\right)^{-1},$$

where $m_{\mathbf{B}_{[i]},\mathbf{\Theta}_i}(z) = \frac{1}{N}\mathrm{tr}\mathbf{\Theta}_i\left(\mathbf{B}_{[i]} - z\mathbf{I}_N\right)^{-1}$.

In the course of the development of the proof, we require the existence of moments of order p of w_N in (103), i.e., $E\left[|w_N|^p\right] \neq 0$, for some integer p. First we bound (103) as $E[|w_N|^p] \leq E[(\sum_{i=1}^{4n}\tilde{d}_i)^p]$. The application of Hölder's inequality yields

$$E\left[|w_N|^p\right] \leq \left(\frac{4}{\beta}\right)^{p-1}\frac{1}{N}\sum_{i=1}^{n}\sum_{l=1}^{4}E\left[|d_i^{(l)}|^p\right].$$

Furthermore, for some $T, Q<\infty$, we can uniformly bound $\mathbf{\Theta}_i$ and $\mathbf{Q}_N$ as

$$\limsup_{N\to\infty}\sup_{1\leq i\leq n}\|\mathbf{\Theta}_i\| \leq T \tag{104}$$

$$\limsup_{N\to\infty}\|\mathbf{Q}_N\| \leq Q. \tag{105}$$

Proposition 12: Let the following upper bounds be well defined and let the entries of $\mathbf{y}_i$ have eighth order moment of order $O\left(\frac{1}{N^4}\right)$. Then the pth order moments $E\left[|d_i^{(l)}|^p\right]$, $(l=1,2,3,4)$ can be bounded as

$$\begin{aligned} E\left[|d_i^{(1)}|^p\right] &\leq 2^{p-1}\left(\frac{\beta T^3 Q|z|^3}{(\Im z)^7}\right)^p\frac{1}{N^p}\left(\frac{C_p^{(1)}}{N^{p/2}}+1\right) \quad (106)\\ E\left[|d_i^{(2)}|^p\right] &\leq \frac{|z|^4}{(\Im z)^4}\frac{C_p^{(2)}}{N^{p/2}}, \\ E\left[|d_i^{(3)}|^p\right] &\leq \left(\frac{|z|TQ}{N(\Im z)^3}\right)^p\left[1+\frac{\beta T^2|z|^2}{(\Im z)^4}\right]^p, \\ E\left[|d_i^{(4)}|^p\right] &\leq 2^{p-1}\left(\frac{TQ|z|^2}{(\Im z)^4}\right)^p\left[\frac{C_p^{(4)}}{N^{p/2}}+\frac{T^p}{N^p(\Im z)^p}\right], \end{aligned}$$

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

$$|d_i^{(1)}| \leq \frac{|z|}{\Im z}\left|\mathbf{y}_i^{\mathsf{H}}\mathbf{\Psi}_i^{\mathsf{H}}\left(\mathbf{B}_{[i]} - z\mathbf{I}_N\right)^{-1}\left[\mathbf{D}_{[i]}^{-1} - \mathbf{D}^{-1}\right]\mathbf{\Psi}_i\mathbf{y}_i\right|.$$

We further bound $|d_i^{(1)}|$ by applying Lemmas 10 and 12 with the fact that $\|(\mathbf{B}_{[i]} - z\mathbf{I}_N)^{-1}\| \leq \frac{1}{\Im z}$. Together with (104) we have

$$|d_i^{(1)}| \leq \frac{|z|T}{(\Im z)^2}\|\mathbf{y}_i\|_2^2\|\mathbf{D}_{[i]}^{-1} - \mathbf{D}^{-1}\|.$$

Similarly, with Lemma 2, it can be shown that $\|\mathbf{D}_{[i]}^{-1} - \mathbf{D}^{-1}\| \leq \frac{\beta T^2 Q|z|^2}{N(\Im z)^5}$ and thus

$$|d_i^{(1)}| \leq \frac{\beta T^3 Q|z|^3}{N(\Im z)^7}\|\mathbf{y}_i\|_2^2.$$

The pth order moment of $|d_i^{(1)}|$ thus satisfies

$$E\left[|d_i^{(1)}|^p\right] \leq \left[\frac{\beta T^3 Q|z|^3}{(\Im z)^7}\right]^p\frac{1}{N^p}E\left[\left|\mathbf{y}_i^{\mathsf{H}}\mathbf{y}_i\right|^p\right].$$

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

$$E\left[|d_i^{(1)}|^p\right] \leq 2^{p-1}\left(\frac{\beta T^3 Q|z|^3}{N(\Im z)^7}\right)^p\left(E\left[\left|\mathbf{y}_i^{\mathsf{H}}\mathbf{y}_i - 1\right|^p\right] + 1\right).$$

If the moments $E[|d_i^{(1)}|^4]$ and $E[|d_i^{(1)}|^{2p}]$ exist and are bounded, we can apply Lemma 3 and obtain (106). For the sake of brevity, we omit the derivations of the remaining

moments $E[|d_i^{(l)}|^p]$, $l=\{2,3,4\}$, since the techniques are similar to the previous procedure. ∎

From Proposition 12, we conclude that all $E[|d_i^{(l)}|^p]$ are summable if $p=2+\varepsilon$, $\varepsilon>0$. Therefore, $E[|w_N|^p]$ is summable for $p=2+\varepsilon$ and hence the Borel-Cantelli Lemma [40] implies that $w_N \overset{N\to\infty}{\longrightarrow} 0$, almost surely. Note that with the same approach, the convergence region can be extended to $z\in\mathbb{C}\setminus\mathbb{R}^+$.

We now prove the existence and uniqueness of a solution to (11).

B. Proof of Convergence of the Fixed Point Equation

In this section we consider the fixed point equation (11). We first prove that, properly initialized, the sequence $\{e_{N,i}^{(k)}\}$, $(k=1,2,\dots)$, converges to a limit $e_{N,i}$ as $k\to\infty$. Subsequently, we show that this limit $e_{N,i}$ satisfies $|m_{\mathbf{B}_N,\mathbf{\Theta}_i}-e_{N,i}| \overset{N\to\infty}{\longrightarrow} 0$, almost surely.

Proposition 13: Let $z\in\mathbb{C}^+$ and $\{e_{N,i}^{(k)}(z)\}$ $(k\geq 0)$ be the sequence defined by (12). 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 (12) 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

$$e_{N,i}^{(k+1)} = \frac{1}{N}\mathrm{tr}\mathbf{\Theta}_i\left(\frac{1}{N}\sum_{j=1}^{n} c_{N,j}^{(k)}\mathbf{\Theta}_j+\mathbf{S}_N-z\mathbf{I}_N\right)^{-1} \triangleq \frac{1}{N}\mathrm{tr}\mathbf{\Theta}_i\mathbf{A}_k, \tag{107}$$

where $c_{N,j}^{(k)}=1/(1+e_{N,j}^{(k)})$. In (107), multiplying $\mathbf{A}_k$ from the right by $(\mathbf{A}_k^{\mathsf{H}})^{-1}\mathbf{A}_k^{\mathsf{H}}$, we obtain

$$e_{N,i}^{(k+1)} = \frac{1}{N}\mathrm{tr}\mathbf{A}_k^{\mathsf{H}}\mathbf{\Theta}_i\mathbf{A}_k\left[\frac{1}{N}\sum_{j=1}^{n} c_{N,j}^{*,(k)}\mathbf{\Theta}_j\right]+v_i^{(k)}, \tag{108}$$

where $v_i^{(k)}=\frac{1}{N}\mathrm{tr}\mathbf{A}_k^{\mathsf{H}}\mathbf{\Theta}_i\mathbf{A}_k[\mathbf{S}_N-z^*\mathbf{I}_N]$. Denoting $\mathbf{r}_i^{(k)} \triangleq \frac{1}{N}[\frac{1}{N}\mathrm{tr}\mathbf{A}_k^{\mathsf{H}}\mathbf{\Theta}_i\mathbf{A}_k\mathbf{\Theta}_1,\dots,\frac{1}{N}\mathrm{tr}\mathbf{A}_k^{\mathsf{H}}\mathbf{\Theta}_i\mathbf{A}_k\mathbf{\Theta}_n]^{\mathsf{T}}$ and $\mathbf{c}_N^{(k)} \triangleq [c_{N,1}^{(k)},\dots,c_{N,n}^{(k)}]^{\mathsf{T}}$, (108) takes the form

$$e_{N,i}^{(k+1)} = \mathbf{r}_i^{\mathsf{T},(k)}\mathbf{c}_N^{\mathsf{H},(k)}+v_i^{(k)}. \tag{109}$$

Since the $\mathbf{\Theta}_i$ are uniformly bounded w.r.t. N, we have $\mathbf{r}_i^{(k)}, v_i^{(k)}>0$. To show that $e_{N,i}^{(k+1)}$ are Stieltjes transforms of a nonnegative finite measure, the following three conditions must be verified [28, Proposition 2.2]: For $z\in\mathbb{C}^+$ (i) $e_{N,i}^{(k+1)}(z)\in\mathbb{C}^+$, (ii) $ze_{N,i}^{(k+1)}(z)\in\mathbb{C}^+$ and (iii) $\lim_{y\to+\infty}-\mathbf{i}ye_{N,i}^{(k+1)}(\mathbf{i}y)<\infty$. From (109) 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 14: Any sequence $\{e_{N,i}^{(k)}(z)\}$, $(k>0)$ defined by (12) converges to a Stieltjes transform, denoted $e_{N,i}(z)$ as $k\to\infty$ if $e_{N,i}^{(0)}(z)$ is a Stieltjes transform.

Proof: Let $e_{N,i}^{(k)}(z)=\frac{1}{N}\mathrm{tr}\mathbf{\Theta}_i\mathbf{A}^{(k-1)}$ and $e_{N,i}^{(k+1)}(z)=\frac{1}{N}\mathrm{tr}\mathbf{\Theta}_i\mathbf{A}^{(k)}$, where

$$\mathbf{A}^{(k-1)}=\left(\frac{1}{N}\sum_{j=1}^{n}\frac{\mathbf{\Theta}_j}{1+e_{N,j}^{(k-1)}(z)}+\mathbf{S}_N-z\mathbf{I}_N\right)^{-1},$$

$$\mathbf{A}^{(k)}=\left(\frac{1}{N}\sum_{j=1}^{n}\frac{\mathbf{\Theta}_j}{1+e_{N,j}^{(k)}(z)}+\mathbf{S}_N-z\mathbf{I}_N\right)^{-1}.$$

Applying Lemma 2, the difference $|e_{N,i}^{(k)}(z)-e_{N,i}^{(k+1)}(z)|$ is

$$|e_{N,i}^{(k)}-e_{N,i}^{(k+1)}| = \left|\frac{1}{N}\mathrm{tr}\mathbf{A}^{(k+1)}\mathbf{\Theta}_i\mathbf{A}^{(k)}\left[\frac{1}{N}\sum_{j=1}^{n}\mathbf{\Theta}_j\frac{e_{N,j}^{(k)}-e_{N,j}^{(k-1)}}{\left[1+e_{N,j}^{(k)}\right]\left[1+e_{N,j}^{(k-1)}\right]}\right]\right| \tag{110}$$

With Lemmas 9, 11 and 12, (110) can be bounded as

$$|e_{N,i}^{(k)}-e_{N,i}^{(k+1)}|\leq C\sup_{1\leq i\leq n}|e_{N,i}^{(k)}-e_{N,i}^{(k-1)}|, \tag{111}$$

where $C=\frac{\beta T^2|z|^2}{(\Im z)^4}$. Clearly, the sequence $\{e_{N,i}^{(k)}\}$ converges to a limit $e_{N,i}$ for z restricted to the set $\{z\in\mathbb{C}^+ : C<1\}$. Proposition 13 shows that all $\{e_{N,i}^{(k)}\}$ are uniformly bounded Stieltjes transforms and therefore their limit is analytic. 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 [15, Theorem 3.11] ensures that the sequence $\{e_{N,i}^{(k)}\}$ must converge for all $z\in\mathbb{C}\setminus\mathbb{R}^+$ and their limit is $e_{N,i}(z)$.

It is straightforward to verify, that the previous holds also true for $z\in\mathbb{C}^-$. ∎

Remark 8: For $z<0$, the existence of a unique solution to (11) as well as the convergence of (12) from any real initial point can be proved within the framework of *standard interference functions* [55]. The strategy is as follows. Let $\bar{\mathbf{e}}_N \triangleq \bar{\mathbf{e}}_N(z)=[\bar{e}_{N,1}(z),\bar{e}_{N,2}(z),\dots,\bar{e}_{N,n}(z)]^{\mathsf{T}}\in\mathbb{R}^n$ and $\mathbf{f}(\bar{\mathbf{e}}_N)=[f_1(\bar{\mathbf{e}}_N),f_2(\bar{\mathbf{e}}_N),\dots,f_n(\bar{\mathbf{e}}_N)]^{\mathsf{T}}\in\mathbb{R}^n$, where

$$f_i(\bar{\mathbf{e}}_N)=\frac{1}{N}\mathrm{tr}\mathbf{\Theta}_i\left(\frac{1}{N}\sum_{j=1}^{n}\frac{\mathbf{\Theta}_j}{1+\bar{e}_{N,j}(z)}+\mathbf{S}_N-z\mathbf{I}_N\right)^{-1}.$$

Theorems 1 and 2 in [55] prove that, if $\mathbf{f}(\bar{\mathbf{e}}_N)$ is a feasible standard interference function, then (12) converges to a unique solution $\mathbf{e}_N$ with all nonnegative entries for any initial point $e_{N,i}^{(0)},\dots,e_{N,n}^{(0)}$. The proof that $\mathbf{f}(\bar{\mathbf{e}}_N)$ is feasible as well as a standard interference function is straightforward and details are omitted in this correspondence.

The uniqueness of $\mathbf{e}_N$, whose entries are Stieltjes transforms of nonnegative finite measures, ensures the functional uniqueness of $e_{N,i}(z),\dots,e_{N,n}(z)$ as a Stieltjes transform solution to (11) for $z\in\mathbb{C}\setminus\mathbb{R}^+$. This completes the proof of uniqueness.

Denote $m_{\mathbf{B}_N,\mathbf{\Theta}_i}(z) \triangleq \frac{1}{N}\mathrm{tr}\mathbf{\Theta}_i\left(\mathbf{B}_N-z\mathbf{I}_N\right)^{-1}$. In the following section, we prove that $e_{N,i}(z)=\lim_{k\to\infty}e_{N,i}^{(k)}(z)$ satisfies $|m_{\mathbf{B}_N,\mathbf{\Theta}_i}(z)-e_{N,i}(z)| \overset{N\to\infty}{\longrightarrow} 0$, almost surely.

C. Proof of Convergence of the Deterministic Equivalent

In Section I-A we showed that $w_N = \frac{1}{N}\mathrm{tr}\mathbf{Q}_N\left(\mathbf{B}_N - z\mathbf{I}_N\right)^{-1} - \frac{1}{N}\mathrm{tr}\mathbf{Q}_N\left(\mathbf{R}+\mathbf{S}_N - z\mathbf{I}_N\right)^{-1} \overset{N\to\infty}{\longrightarrow} 0$, almost surely. Furthermore, in Section I-B we proved that the sequence defined by (11) converges to a limit $e_{N,i}$. It remains to prove that

$$m_{\mathbf{B}_N,\mathbf{\Theta}_i} - e_{N,i} = \frac{1}{N}\mathrm{tr}\mathbf{\Theta}_i\left(\mathbf{B}_N - z\mathbf{I}_N\right)^{-1} - \frac{1}{N}\mathrm{tr}\mathbf{\Theta}_i\left(\frac{1}{N}\sum_{j=1}^{n}\frac{\mathbf{\Theta}_j}{1+e_{N,j}(z)} + \mathbf{S}_N - z\mathbf{I}_N\right)^{-1} \overset{N\to\infty}{\longrightarrow} 0, \tag{112}$$

almost surely. Denote $w_{N,i} \triangleq w_{\mathbf{\Theta}_i}$ with $w_{\mathbf{\Theta}_i}$ defined in (101). Applying Lemma 2, (112) can be written as

$$\begin{aligned} &m_{\mathbf{B}_N,\mathbf{\Theta}_i} - e_{N,i} \\ &= w_{N,i} + \frac{1}{N}\mathrm{tr}\mathbf{\Theta}_i\left(\mathbf{A}+\mathbf{S}_N - z\mathbf{I}_N\right)^{-1} - e_{N,i}(z) \\ &= w_{N,i} - \frac{1}{N}\mathrm{tr}\mathbf{\Theta}_i\bar{\mathbf{A}}^{-1}\left[\mathbf{A}-\mathbf{B}\right]\bar{\mathbf{B}}^{-1}, \end{aligned}$$

where $\bar{\mathbf{A}} \triangleq \mathbf{A}+\mathbf{S}_N - z\mathbf{I}_N$, $\mathbf{A} \triangleq \frac{1}{N}\sum_{l=1}^{n}\frac{\mathbf{\Theta}_l}{1+\frac{1}{N}\mathrm{tr}\mathbf{\Theta}_l(\mathbf{B}_N - z\mathbf{I}_N)^{-1}}$ and $\bar{\mathbf{B}} \triangleq \mathbf{B}+\mathbf{S}_N - z\mathbf{I}_N$, $\mathbf{B} \triangleq \frac{1}{N}\sum_{j=1}^{n}\frac{\mathbf{\Theta}_j}{1+e_{N,j}}$. Applying Lemmas 9 and 11, $|m_{\mathbf{B}_N,\mathbf{\Theta}_i} - e_{N,i}|$ can be bounded as

$$|m_{\mathbf{B}_N,\mathbf{\Theta}_i} - e_{N,i}| \le |w_{N,i}| + \|\mathbf{\Theta}_i\|\|\bar{\mathbf{A}}^{-1}\|\|\bar{\mathbf{B}}^{-1}\| \times \left\|\frac{1}{N}\sum_{j=1}^{n}\mathbf{\Theta}_j\frac{|m_{\mathbf{B}_N,\mathbf{\Theta}_j} - e_{N,j}|}{(1+m_{\mathbf{B}_N,\mathbf{\Theta}_j})(1+e_{N,j})}\right\|. \tag{113}$$

Similar to (111), with Lemma 12, (113) can be further bounded as

$$|m_{\mathbf{B}_N,\mathbf{\Theta}_i} - e_{N,i}| \le |w_{N,i}| + C\sup_{1\le i\le n}|m_{\mathbf{B}_N,\mathbf{\Theta}_i} - e_{N,i}|,$$

where $C = \frac{\beta T^2|z|^2}{(\Im z)^4}$. Taking the supremum over all $i=1,\ldots,n$, we obtain

$$\sup_{1\le i\le n}|m_{\mathbf{B}_N,\mathbf{\Theta}_i} - e_{N,i}|\left[1-C\right] \le \sup_{1\le i\le n}|w_{N,i}|. \tag{114}$$

From (114), on the set $\{z\in\mathbb{C}^+ : 0<C<1\} \neq \emptyset$, it suffices to show that $\sup_{1\le i\le n}|w_{N,i}|$ goes to zero sufficiently fast. For any $\varepsilon>0$ we have

$$\begin{aligned} P\left(\sup_{1\le i\le n}|w_{N,i}| > \varepsilon\right) &\le \sum_{i=1}^{n}P\left(|w_{N,i}| > \varepsilon\right) \\ &= \sum_{i=1}^{n}P\left(|w_{N,i}|^p > \varepsilon^p\right). \end{aligned} \tag{115}$$

Applying Markov's inequality, (115) can be further bounded as

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

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 $\sup_{1\le i\le n}w_{N,i} \overset{N\to\infty}{\longrightarrow} 0$, almost surely.

On $\{z\in\mathbb{C}^+ : 0<C<1\}$, the $e_{N,i}(z)$ are summable and have a cluster point. Furthermore, Proposition 13 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 [15, Theorem 3.11] applies, and extends the convergence region of (112) to $z\in\mathbb{C}\setminus\mathbb{R}^+$.

Since (112) holds true, the following convergence holds almost surely

$$\frac{1}{N}\mathrm{tr}\mathbf{D}^{-1} - \frac{1}{N}\mathrm{tr}\mathbf{Q}_N\left(\frac{1}{N}\sum_{i=1}^{n}\frac{\mathbf{\Theta}_i}{1+e_{N,i}} + \mathbf{S}_N - z\mathbf{I}_N\right)^{-1} \overset{N\to\infty}{\longrightarrow} 0. \tag{116}$$

The convergence in (116) implies the convergence in (9), which completes the proof.

Appendix II
Proof of Theorem 2

The strategy is as follows: The SINR $\gamma_{k,\mathrm{rzf}}$ in (16) consists of three terms, (i) the scaled signal power $|\mathbf{h}_k^{\mathsf{H}}\hat{\mathbf{W}}\hat{\mathbf{h}}_k|^2$: (ii) the scaled interference power $\mathbf{h}_k^{\mathsf{H}}\hat{\mathbf{W}}\hat{\mathbf{H}}_{[k]}^{\mathsf{H}}\mathbf{P}_{[k]}\hat{\mathbf{H}}_{[k]}\hat{\mathbf{W}}\mathbf{h}_k$ (both scaled by ξ^{-2}) and (iii) the term Ψ of the power normalization. For each of these three terms we will subsequently derive a deterministic equivalent which together constitute the final expression for $\gamma_{k,\mathrm{rzf}}^{\circ}$.

A. Deterministic equivalent for Ψ

The term $\Psi = \mathrm{tr}\mathbf{P}\hat{\mathbf{H}}(\hat{\mathbf{H}}^{\mathsf{H}}\hat{\mathbf{H}} + M\alpha\mathbf{I}_M)^{-2}\hat{\mathbf{H}}^{\mathsf{H}}$ can be written as

$$\Psi = \sum_{k=1}^{K}p_k\hat{\mathbf{h}}_k^{\mathsf{H}}\left(\hat{\mathbf{H}}^{\mathsf{H}}\hat{\mathbf{H}} + M\alpha\mathbf{I}_M\right)^{-2}\hat{\mathbf{h}}_k \tag{117}$$

$$\overset{(a)}{=} \frac{1}{M}\sum_{k=1}^{K}p_k\frac{\hat{\mathbf{z}}_k^{\mathsf{H}}\mathbf{\Theta}_k^{1/2}\mathbf{C}_{[k]}^{-2}\mathbf{\Theta}_k^{1/2}\hat{\mathbf{z}}_k}{\left(1+\hat{\mathbf{z}}_k^{\mathsf{H}}\mathbf{\Theta}_k^{1/2}\mathbf{C}_{[k]}^{-1}\mathbf{\Theta}_k^{1/2}\hat{\mathbf{z}}_k\right)^2}, \tag{118}$$

where $\mathbf{C}_{[k]} \triangleq \mathbf{\Gamma}_{[k]} + \alpha\mathbf{I}_M$ with $\mathbf{\Gamma}_{[k]} \triangleq \frac{1}{M}\hat{\mathbf{H}}_{[k]}^{\mathsf{H}}\hat{\mathbf{H}}_{[k]}$ and in (a) we applied Lemma 1 twice together with (6). For M large and under Assumptions 1, we apply Lemma 4 and obtain

$$\Psi - \frac{1}{M}\sum_{k=1}^{K}p_k\frac{\frac{1}{M}\mathrm{tr}\mathbf{\Theta}_k\mathbf{C}_{[k]}^{-2}}{\left(1+\frac{1}{M}\mathrm{tr}\mathbf{\Theta}_k\mathbf{C}_{[k]}^{-1}\right)^2} \overset{M\to\infty}{\longrightarrow} 0$$

$$\overset{(b)}{\Leftrightarrow}\ \Psi - \frac{1}{M}\sum_{k=1}^{K}p_k\frac{m'_{\mathbf{\Gamma},\mathbf{\Theta}_k}(-\alpha)}{\left(1+m_{\mathbf{\Gamma},\mathbf{\Theta}_k}(-\alpha)\right)^2} \overset{M\to\infty}{\longrightarrow} 0,$$

almost surely, where in (b) we applied Lemma 6, the definition (8) and denoted $m'_{\mathbf{\Gamma},\mathbf{\Theta}_k}(-\alpha)$ the derivative of $m_{\mathbf{\Gamma},\mathbf{\Theta}_k}(z)$ along z at $z=-\alpha$. Applying Theorem 1 to $m_{\mathbf{\Gamma},\mathbf{\Theta}_k}(z)$, we obtain

$$m_{\mathbf{\Gamma},\mathbf{\Theta}_k}(-\alpha) - \frac{1}{M}\mathrm{tr}\mathbf{\Theta}_k\mathbf{T} \overset{M\to\infty}{\longrightarrow} 0,$$

$$m'_{\mathbf{\Gamma},\mathbf{\Theta}_k}(-\alpha) - \frac{1}{M}\mathrm{tr}\mathbf{\Theta}_k\mathbf{T}' \overset{M\to\infty}{\longrightarrow} 0,$$

almost surely, where $\mathbf{T}$ is defined in (21) and $\mathbf{T}'$ is given by

$$\mathbf{T}' = \mathbf{T}\left[\frac{1}{M}\sum_{j=1}^{K}\frac{\boldsymbol{\Theta}_j e_j'}{(1+e_j)^2} + \mathbf{I}_M\right]\mathbf{T}. \tag{119}$$

Define $\mathbf{e}' = [e_1', \ldots, e_K']^\mathsf{T}$ with $e_i' = \frac{1}{M}\mathrm{tr}\boldsymbol{\Theta}_i\mathbf{T}'$. The system of K equations formed by the e_i' takes the form $\mathbf{e}' = \mathbf{J}\mathbf{e}' + \mathbf{v}$ and the explicit solution $\mathbf{e}'$ is given in (24). Substituting $m_{\boldsymbol{\Gamma},\boldsymbol{\Theta}_k}(-\alpha)$ and $m'_{\boldsymbol{\Gamma},\boldsymbol{\Theta}_k}(-\alpha)$ by their respective deterministic equivalents e_k and e_k', we obtain Ψ° in (22) such that $\Psi - \Psi^\circ \overset{M\to\infty}{\longrightarrow} 0$, almost surely.

B. Deterministic equivalent for $\mathbf{h}_k^\mathsf{H}\hat{\mathbf{W}}\hat{\mathbf{h}}_k$

Similar to the derivations in (117) and (118), we have

$$\mathbf{h}_k^\mathsf{H}\hat{\mathbf{W}}\hat{\mathbf{h}}_k = \frac{\mathbf{z}_k^\mathsf{H}\boldsymbol{\Theta}_k^{1/2}\mathbf{C}_{[k]}^{-1}\boldsymbol{\Theta}_k^{1/2}\hat{\mathbf{z}}_k}{1+\hat{\mathbf{z}}_k^\mathsf{H}\boldsymbol{\Theta}_k^{1/2}\mathbf{C}_{[k]}^{-1}\boldsymbol{\Theta}_k^{1/2}\hat{\mathbf{z}}_k}$$
$$= \frac{\sqrt{1-\tau_k^2}\mathbf{z}_k^\mathsf{H}\boldsymbol{\Theta}_k^{1/2}\mathbf{C}_{[k]}^{-1}\boldsymbol{\Theta}_k^{1/2}\mathbf{z}_k}{1+\hat{\mathbf{z}}_k^\mathsf{H}\boldsymbol{\Theta}_k^{1/2}\mathbf{C}_{[k]}^{-1}\boldsymbol{\Theta}_k^{1/2}\hat{\mathbf{z}}_k} + \frac{\tau_k\mathbf{z}_k^\mathsf{H}\boldsymbol{\Theta}_k^{1/2}\mathbf{C}_{[k]}^{-1}\boldsymbol{\Theta}_k^{1/2}\mathbf{q}_k}{1+\hat{\mathbf{z}}_k^\mathsf{H}\boldsymbol{\Theta}_k^{1/2}\mathbf{C}_{[k]}^{-1}\boldsymbol{\Theta}_k^{1/2}\hat{\mathbf{z}}_k}.$$

Since $\mathbf{q}_k$ and $\mathbf{z}_k$ are independent, we apply Lemma 5 together with Lemma 4 and 6 and obtain

$$\mathbf{h}_k^\mathsf{H}\hat{\mathbf{W}}\hat{\mathbf{h}}_k - \sqrt{1-\tau_k^2}\frac{m_k^\circ}{1+m_k^\circ} \overset{M\to\infty}{\longrightarrow} 0,$$

almost surely.

C. Deterministic equivalent of $\mathbf{h}_k^\mathsf{H}\hat{\mathbf{W}}\hat{\mathbf{H}}_{[k]}^\mathsf{H}\mathbf{P}_{[k]}\hat{\mathbf{H}}_{[k]}\hat{\mathbf{W}}\mathbf{h}_k$

With (5) and $\mathbf{C} \triangleq \boldsymbol{\Gamma} + \alpha\mathbf{I}_M$, $\boldsymbol{\Gamma} \triangleq \frac{1}{M}\hat{\mathbf{H}}^\mathsf{H}\hat{\mathbf{H}}$, we have

$$\mathbf{h}_k^\mathsf{H}\hat{\mathbf{W}}\hat{\mathbf{H}}_{[k]}^\mathsf{H}\mathbf{P}_{[k]}\hat{\mathbf{H}}_{[k]}\hat{\mathbf{W}}\mathbf{h}_k$$
$$= \frac{1}{M}\mathbf{z}_k^\mathsf{H}\boldsymbol{\Theta}_k^{1/2}\mathbf{C}^{-1}\hat{\mathbf{H}}_{[k]}^\mathsf{H}\mathbf{P}_{[k]}\hat{\mathbf{H}}_{[k]}\mathbf{C}^{-1}\boldsymbol{\Theta}_k^{1/2}\mathbf{z}_k \tag{120}$$
$$= \frac{1}{M}\mathbf{z}_k^\mathsf{H}\boldsymbol{\Theta}_k^{1/2}\mathbf{C}_{[k]}^{-1}\hat{\mathbf{H}}_{[k]}^\mathsf{H}\mathbf{P}_{[k]}\hat{\mathbf{H}}_{[k]}\mathbf{C}^{-1}\boldsymbol{\Theta}_k^{1/2}\mathbf{z}_k +$$
$$\frac{1}{M}\mathbf{z}_k^\mathsf{H}\boldsymbol{\Theta}_k^{1/2}\left[\mathbf{C}^{-1}-\mathbf{C}_{[k]}^{-1}\right]\hat{\mathbf{H}}_{[k]}^\mathsf{H}\mathbf{P}_{[k]}\hat{\mathbf{H}}_{[k]}\mathbf{C}^{-1}\boldsymbol{\Theta}_k^{1/2}\mathbf{z}_k. \tag{121}$$

Substituting $\mathbf{C}^{-1}-\mathbf{C}_{[k]}^{-1} = -\mathbf{C}^{-1}(\mathbf{C}-\mathbf{C}_{[k]})\mathbf{C}_{[k]}^{-1}$ with $\mathbf{C}-\mathbf{C}_{[k]} = \boldsymbol{\Theta}_k^{1/2}(c_0\mathbf{z}_k\mathbf{z}_k^\mathsf{H} + c_1\mathbf{q}_k\mathbf{q}_k^\mathsf{H} + c_2\mathbf{z}_k\mathbf{q}_k^\mathsf{H} + c_2\mathbf{q}_k\mathbf{z}_k^\mathsf{H})\boldsymbol{\Theta}_k^{1/2}$, where $c_0 \triangleq 1-\tau_k^2$, $c_1 \triangleq \tau_k^2$ and $c_2 \triangleq \tau_k\sqrt{1-\tau_k^2}$ into (121), we obtain a sum of five terms

$$\mathbf{h}_k^\mathsf{H}\hat{\mathbf{W}}\hat{\mathbf{H}}_{[k]}^\mathsf{H}\mathbf{P}_{[k]}\hat{\mathbf{H}}_{[k]}\hat{\mathbf{W}}\mathbf{h}_k = \frac{1}{M}\mathbf{z}_k^\mathsf{H}\mathbf{B}_k\mathbf{z}_k$$
$$-\frac{c_0}{M}\mathbf{z}_k^\mathsf{H}\mathbf{A}_k\mathbf{z}_k\mathbf{z}_k^\mathsf{H}\mathbf{B}_k\mathbf{z}_k - \frac{c_1}{M}\mathbf{z}_k^\mathsf{H}\mathbf{A}_k\mathbf{q}_k\mathbf{q}_k^\mathsf{H}\mathbf{B}_k\mathbf{z}_k$$
$$-\frac{c_2}{M}\mathbf{z}_k^\mathsf{H}\mathbf{A}_k\mathbf{z}_k\mathbf{q}_k^\mathsf{H}\mathbf{B}_k\mathbf{z}_k - \frac{c_2}{M}\mathbf{z}_k^\mathsf{H}\mathbf{A}_k\mathbf{q}_k\mathbf{z}_k^\mathsf{H}\mathbf{B}_k\mathbf{z}_k, \tag{122}$$

where we denoted $\mathbf{A}_k \triangleq \boldsymbol{\Theta}_k^{1/2}\mathbf{C}^{-1}\boldsymbol{\Theta}_k^{1/2}$ and $\mathbf{B}_k \triangleq \boldsymbol{\Theta}_k^{1/2}\mathbf{C}_{[k]}^{-1}\hat{\mathbf{H}}_{[k]}^\mathsf{H}\mathbf{P}_{[k]}\hat{\mathbf{H}}_{[k]}\mathbf{C}^{-1}\boldsymbol{\Theta}_k^{1/2}$. Noting that $c_0+c_1=1$ and $c_0c_1-c_2^2=0$, we apply Lemma 7 to each of the four quadratic forms in (122). Under Assumption 1, we obtain

$$\mathbf{z}_k^\mathsf{H}\mathbf{A}_k\mathbf{z}_k - \frac{u(1+c_1u)}{1+u} \overset{M\to\infty}{\longrightarrow} 0,$$
$$\mathbf{z}_k^\mathsf{H}\mathbf{A}_k\mathbf{q}_k - \frac{-c_2u^2}{1+u} \overset{M\to\infty}{\longrightarrow} 0,$$

almost surely, where $u = \frac{1}{M}\mathrm{tr}\boldsymbol{\Theta}_k\mathbf{C}_{[k]}^{-1}$. Moreover, under Assumptions 1, 3 and $\|\mathbf{P}\| < \infty$ uniformly on M, we have

$$\mathbf{z}_k^\mathsf{H}\mathbf{B}_k\mathbf{z}_k - \frac{u'(1+c_1u)}{1+u} \overset{M\to\infty}{\longrightarrow} 0,$$
$$\mathbf{q}_k^\mathsf{H}\mathbf{B}_k\mathbf{z}_k - \frac{-c_2uu'}{1+u} \overset{M\to\infty}{\longrightarrow} 0,$$

almost surely, where $u' = \frac{1}{M}\mathrm{tr}\mathbf{P}_{[k]}\hat{\mathbf{H}}_{[k]}\mathbf{C}_{[k]}^{-1}\boldsymbol{\Theta}_k\mathbf{C}_{[k]}^{-1}\hat{\mathbf{H}}_{[k]}^\mathsf{H}$. Substituting the random terms in (122) by their respective deterministic equivalents yields

$$\mathbf{h}_k^\mathsf{H}\hat{\mathbf{W}}\hat{\mathbf{H}}_{[k]}^\mathsf{H}\mathbf{P}_{[k]}\hat{\mathbf{H}}_{[k]}\hat{\mathbf{W}}\mathbf{h}_k - \left[\frac{1}{M}\frac{u'(1+c_1u)}{1+u}\right.$$
$$\left.-\frac{1}{M}\frac{c_0(1+c_1u)^2 - c_1c_2^2u^2 - 2c_2^2u}{(1+u)^2}uu'\right] \overset{M\to\infty}{\longrightarrow} 0, \tag{123}$$

almost surely. The second term in brackets of (123) reduces to $\frac{1}{M}\frac{1-\tau_k^2}{(1+u)^2}uu'$ and we obtain

$$\mathbf{h}_k^\mathsf{H}\hat{\mathbf{W}}\hat{\mathbf{H}}_{[k]}^\mathsf{H}\mathbf{P}_{[k]}\hat{\mathbf{H}}_{[k]}\hat{\mathbf{W}}\mathbf{h}_k -$$
$$\frac{1}{M}\frac{1-\tau_k^2\left[1-(1+u)^2\right]}{(1+u)^2}u' \overset{M\to\infty}{\longrightarrow} 0, \tag{124}$$

almost surely. From Lemma 6 we have

$$u - m_{\boldsymbol{\Gamma},\boldsymbol{\Theta}_k}(-\alpha) \overset{M\to\infty}{\longrightarrow} 0,$$
$$\frac{1}{M}u' - \Upsilon_k \overset{M\to\infty}{\longrightarrow} 0,$$

almost surely, where $m_{\boldsymbol{\Gamma},\boldsymbol{\Theta}_k}(-\alpha) = \frac{1}{M}\mathrm{tr}\boldsymbol{\Theta}_k\mathbf{C}^{-1}$ and $\Upsilon_k = \frac{1}{M^2}\mathrm{tr}\mathbf{P}_{[k]}\hat{\mathbf{H}}_{[k]}\mathbf{C}^{-1}\boldsymbol{\Theta}_k\mathbf{C}^{-1}\hat{\mathbf{H}}_{[k]}^\mathsf{H}$. Therefore, (124) becomes

$$\mathbf{h}_k^\mathsf{H}\hat{\mathbf{W}}\hat{\mathbf{H}}_{[k]}^\mathsf{H}\mathbf{P}_{[k]}\hat{\mathbf{H}}_{[k]}\hat{\mathbf{W}}\mathbf{h}_k -$$
$$\frac{\Upsilon_k\left[1-\tau_k^2\left(1-(1+m_{\boldsymbol{\Gamma},\boldsymbol{\Theta}_k}(-\alpha))^2\right)\right]}{(1+m_{\boldsymbol{\Gamma},\boldsymbol{\Theta}_k}(-\alpha))^2} \overset{M\to\infty}{\longrightarrow} 0,$$

almost surely. We rewrite Υ_k as

$$\Upsilon_k = \frac{1}{M}\sum_{j=1,j\neq k}^{K} p_j\hat{\mathbf{z}}_j^\mathsf{H}\boldsymbol{\Theta}_j^{1/2}\mathbf{C}^{-1}\boldsymbol{\Theta}_k\mathbf{C}^{-1}\boldsymbol{\Theta}_j^{1/2}\hat{\mathbf{z}}_j.$$

Applying Lemmas 1, 4 and 6, we obtain almost surely

$$\Upsilon_k - \frac{1}{M}\sum_{j=1,j\neq k}^{K} p_j\frac{\frac{1}{M}\mathrm{tr}\boldsymbol{\Theta}_j\mathbf{C}^{-1}\boldsymbol{\Theta}_k\mathbf{C}^{-1}}{\left[1+\frac{1}{M}\mathrm{tr}\boldsymbol{\Theta}_j\left(\boldsymbol{\Gamma}+\alpha\mathbf{I}_M\right)^{-1}\right]^2} \overset{M\to\infty}{\longrightarrow} 0.$$

A deterministic equivalent e_i of $m_{\boldsymbol{\Gamma},\boldsymbol{\Theta}_i}(-\alpha) = \frac{1}{M}\mathrm{tr}\boldsymbol{\Theta}_i\left(\boldsymbol{\Gamma}+\alpha\mathbf{I}_M\right)^{-1}$ such that $m_{\boldsymbol{\Gamma},\boldsymbol{\Theta}_i}(-\alpha) - e_i \overset{M\to\infty}{\longrightarrow} 0$, almost surely is given in (20). To derive a deterministic equivalent for $\frac{1}{M}\mathrm{tr}\boldsymbol{\Theta}_j\mathbf{C}^{-1}\boldsymbol{\Theta}_k\mathbf{C}^{-1}$, we can assume the $\boldsymbol{\Theta}_k$ invertible because the result is also a deterministic equivalent for non-invertible matrices $\boldsymbol{\Theta}_k$, which is proved in [39, Theorem 4]. Define $\bar{\mathbf{C}} \triangleq \boldsymbol{\Theta}_k^{-1/2}\boldsymbol{\Gamma}\boldsymbol{\Theta}_k^{-1/2} + \alpha\boldsymbol{\Theta}_k^{-1}$, we have

$$\frac{1}{M}\mathrm{tr}\boldsymbol{\Theta}_j\mathbf{C}^{-1}\boldsymbol{\Theta}_k\mathbf{C}^{-1} = \frac{1}{M}\mathrm{tr}\boldsymbol{\Theta}_k^{-1/2}\boldsymbol{\Theta}_j\boldsymbol{\Theta}_k^{-1/2}\bar{\mathbf{C}}^{-2}$$
$$= \frac{d}{dz}\frac{1}{M}\mathrm{tr}\boldsymbol{\Theta}_j(\boldsymbol{\Gamma}+\alpha\mathbf{I}_M - z\boldsymbol{\Theta}_k)^{-1}.$$

Denote $m_{\boldsymbol{\Gamma}-z\boldsymbol{\Theta}_k,\boldsymbol{\Theta}_j}(-\alpha) = \frac{1}{M}\mathrm{tr}\boldsymbol{\Theta}_j(\boldsymbol{\Gamma}+\alpha\mathbf{I}_M - z\boldsymbol{\Theta}_k)^{-1}$. Applying Theorem 1, we obtain $m_{\boldsymbol{\Gamma}-z\boldsymbol{\Theta}_k,\boldsymbol{\Theta}_j}(-\alpha) -$

$\frac{1}{M}\mathrm{tr}\boldsymbol{\Theta}_j\mathbf{T}_k(z) \overset{M\to\infty}{\longrightarrow} 0$, almost surely, where $\mathbf{T}_k(z)$ is given by

$$\mathbf{T}_k(z) = \left(\frac{1}{M}\sum_{j=1}^{K}\frac{\boldsymbol{\Theta}_j}{1+e_{j,k}(z)} + \alpha\mathbf{I}_M - z\boldsymbol{\Theta}_k\right)^{-1}, \quad (125)$$

where $e_{i,k}(z) = \frac{1}{M}\mathrm{tr}\boldsymbol{\Theta}_i\mathbf{T}_k(z)$. By differentiating along z, we have

$$m'_{\boldsymbol{\Gamma}-z\boldsymbol{\Theta}_k,\boldsymbol{\Theta}_j}(-\alpha) - \frac{1}{M}\mathrm{tr}\boldsymbol{\Theta}_j\mathbf{T}'_k(z) \overset{M\to\infty}{\longrightarrow} 0, \quad (126)$$

almost surely, where $\mathbf{T}'_k(z) = \frac{d}{dz}\mathbf{T}_k(z)$ is given by

$$\mathbf{T}'_k(z) = \mathbf{T}_k(z)\left[\frac{1}{M}\sum_{j=1}^{K}\frac{\boldsymbol{\Theta}_j e'_{j,k}(z)}{(1+e_{j,k}(z))^2} + \boldsymbol{\Theta}_k\right]\mathbf{T}_k(z).$$

Setting $z=0$, we have $e_i = e_{i,k}(0) = \frac{1}{M}\mathrm{tr}\boldsymbol{\Theta}_i\mathbf{T}$ with $\mathbf{T} = \mathbf{T}_k(0)$ defined in (21) and the $e'_{1,k},\ldots,e'_{K,k}$ are the unique positive solutions of $e'_{i,k} = \frac{1}{M}\boldsymbol{\Theta}_i\mathbf{T}'_k(0)$. Define $\mathbf{e}'_k = [e'_{1,k},\ldots,e'_{K,k}]^{\mathsf{T}}$ and $\mathbf{J}$ and $\mathbf{v}_k$ as

$$[\mathbf{J}]_{ij} = \frac{\frac{1}{M}\mathrm{tr}\boldsymbol{\Theta}_i\mathbf{T}\boldsymbol{\Theta}_j\mathbf{T}}{M(1+e_j)^2}, \quad (127)$$

$$\mathbf{v}_k = \left[\frac{1}{M}\mathrm{tr}\boldsymbol{\Theta}_1\mathbf{T}\boldsymbol{\Theta}_k\mathbf{T},\ldots,\frac{1}{M}\mathrm{tr}\boldsymbol{\Theta}_K\mathbf{T}\boldsymbol{\Theta}_k\mathbf{T}\right]^{\mathsf{T}}. \quad (128)$$

Therefore, $\mathbf{e}'_k$ is given explicitly as

$$\mathbf{e}'_k = (\mathbf{I}_K - \mathbf{J})^{-1}\mathbf{v}_k. \quad (129)$$

Note that $\mathbf{I}_K - \mathbf{J}$ is always invertible since $\mathbf{e}'_k$ is a unique positive solution. Finally, substituting $m_{\boldsymbol{\Gamma},\boldsymbol{\Theta}_j}(-\alpha)$ and $\frac{1}{M}\mathrm{tr}\boldsymbol{\Theta}_j\mathbf{C}^{-1}\boldsymbol{\Theta}_k\mathbf{C}^{-1}$ by their respective deterministic equivalents e_j and $e'_{j,k}$, we obtain Υ°_k in (23) such that $\Upsilon_k - \Upsilon^\circ_k \overset{M\to\infty}{\longrightarrow} 0$, almost surely.

If all available transmit power is allocated to a single user (i.e., $p_k = P$), both Ψ° and Υ°_k are of order $O(1/M)$ and hence $\gamma^\circ_{k,\mathrm{rzf}}$ grows unbounded with M. Therefore, we require Assumption 2 to ensure that the convergence in (18) holds true, which completes the proof.

Appendix III
Proof of Theorem 3

We bound $|\gamma_{k,\mathrm{zf}} - \gamma^\circ_{k,\mathrm{zf}}|$ by adding and subtracting $\gamma_{k,\mathrm{rzf}}(\alpha)$ and $\gamma^\circ_{k,\mathrm{rzf}}(\alpha)$ and applying the triangle inequality. We obtain

$$|\gamma_{k,\mathrm{zf}} - \gamma^\circ_{k,\mathrm{zf}}| \le |\gamma_{k,\mathrm{zf}} - \gamma_{k,\mathrm{rzf}}(\alpha)| + |\gamma_{k,\mathrm{rzf}}(\alpha) - \gamma^\circ_{k,\mathrm{rzf}}(\alpha)| + |\gamma^\circ_{k,\mathrm{rzf}}(\alpha) - \gamma^\circ_{k,\mathrm{zf}}|. \quad (130)$$

To show that $|\gamma_{k,\mathrm{zf}} - \gamma^\circ_{k,\mathrm{zf}}| \to 0$ almost surely as $M, K \to \infty$, take $\varepsilon > 0$ arbitrarily small. For $\alpha > 0$ small enough, we will demonstrate that $|\gamma_{k,\mathrm{zf}} - \gamma_{k,\mathrm{rzf}}(\alpha)| < \frac{\varepsilon}{3}$ almost surely and $|\gamma^\circ_{k,\mathrm{rzf}}(\alpha) - \gamma^\circ_{k,\mathrm{zf}}| < \frac{\varepsilon}{3}$ independently of M and K. Furthermore, we show that for M, K large enough, $|\gamma_{k,\mathrm{rzf}}(\alpha) - \gamma^\circ_{k,\mathrm{rzf}}(\alpha)| < \frac{\varepsilon}{3}$ almost surely, from which we conclude that (130) can be made as small as desired.

In order to prove that $|\gamma_{k,\mathrm{zf}} - \gamma_{k,\mathrm{rzf}}(\alpha)| < \frac{\varepsilon}{3}$ for α small enough, it suffices to study the matrices $\hat{\mathbf{W}} = (\hat{\mathbf{H}}^{\mathsf{H}}\hat{\mathbf{H}} + M\alpha\mathbf{I}_M)^{-1}$ and $\underline{\hat{\mathbf{W}}} = \hat{\mathbf{H}}^{\mathsf{H}}(\hat{\mathbf{H}}\hat{\mathbf{H}}^{\mathsf{H}})^{-2}\hat{\mathbf{H}}$ in the SINR of RZF precoding (16) and ZF precoding (33). Applying the matrix inversion lemma, $\hat{\mathbf{W}}$ takes the form

$$\hat{\mathbf{W}} = \hat{\mathbf{H}}^{\mathsf{H}}(\hat{\mathbf{H}}\hat{\mathbf{H}}^{\mathsf{H}} + M\alpha\mathbf{I}_K)^{-2}\hat{\mathbf{H}} + M\alpha(\hat{\mathbf{H}}^{\mathsf{H}}\hat{\mathbf{H}} + M\alpha\mathbf{I}_M)^{-2}.$$

Under Assumption 4, $\lambda_{\min}(\hat{\mathbf{H}}\hat{\mathbf{H}}^{\mathsf{H}}) > \varepsilon > 0$ and, since $\lambda_{\max}(\hat{\mathbf{H}}\hat{\mathbf{H}}^{\mathsf{H}})$ is almost surely bounded for all large M, K, for any continuous functional $f(\hat{\mathbf{W}})$ we have $|f(\hat{\mathbf{W}}) - f(\underline{\hat{\mathbf{W}}})| \overset{\alpha\to 0}{\longrightarrow} 0$ with probability one. Therefore, $|\gamma_{k,\mathrm{zf}} - \gamma_{k,\mathrm{rzf}}(\alpha)| \overset{\alpha\to 0}{\longrightarrow} 0$ uniformly on M, K almost surely.

From Theorem 2, we have immediately that for any $\alpha > 0$, $|\gamma_{k,\mathrm{rzf}}(\alpha) - \gamma^\circ_{k,\mathrm{rzf}}(\alpha)| \overset{M\to\infty}{\longrightarrow} 0$ almost surely.

In order to prove $|\gamma^\circ_{k,\mathrm{rzf}}(\alpha) - \gamma^\circ_{k,\mathrm{zf}}| < \frac{\varepsilon}{3}$ for α small enough, uniformly on M, rewrite $\gamma^\circ_{k,\mathrm{rzf}}(\alpha)$ as

$$\gamma^\circ_{k,\mathrm{rzf}}(\alpha) = \frac{p_k(1-\tau_k^2)(\alpha e_k)^2}{\Upsilon^\circ_k(\alpha^2 - \tau_k^2[\alpha^2 - (\alpha+\alpha e_k)^2]) + \frac{\Psi^\circ}{\rho}(\alpha+\alpha e_k)^2}. \quad (131)$$

To show that $\gamma^\circ_{k,\mathrm{zf}} = \lim_{\alpha\to 0}\gamma^\circ_{k,\mathrm{rzf}}(\alpha)$, we need to verify that the limit $\alpha \to 0$ of both numerator and denominator in (131) exists and that the denominator is uniformly bounded away from zero. Define $\underline{e}_i = \lim_{\alpha\to 0}\alpha e_i(\alpha)$. Under Assumption 5, all $\underline{e}_i$ exist and are strictly positive. Since $\alpha e_i(\alpha)$ is holomorphic for $\alpha > 0$, and is bounded away from zero in a neighborhood of zero, by continuity extension in $\alpha = 0$, we obtain the limit $\alpha \to 0$ as

$$\underline{e}_i = \lim_{\alpha\to 0}\left\{\frac{1}{M}\mathrm{tr}\boldsymbol{\Theta}_i\left(\frac{1}{M}\sum_{j=1}^{K}\frac{\boldsymbol{\Theta}_j}{\alpha+\alpha e_j(\alpha)} + \mathbf{I}_M\right)^{-1}\right\} = \frac{1}{M}\mathrm{tr}\boldsymbol{\Theta}_i\underline{\mathbf{T}}, \quad (132)$$

where $\underline{\mathbf{T}}$ is given in (37). It is easy to verify that $\underline{e} \triangleq \sup_i \underline{e}_i$ is uniformly bounded on M. We have

$$|\underline{e}| \le \sup_i \|\boldsymbol{\Theta}_i\|. \quad (133)$$

Define $\underline{\mathbf{e}} \triangleq [\underline{e}_1,\ldots,\underline{e}_K]^{\mathsf{T}}$, $f_i : \underline{\mathbf{e}} \mapsto \frac{1}{M}\mathrm{tr}\boldsymbol{\Theta}_i\underline{\mathbf{T}}(\underline{\mathbf{e}})$ and $\mathbf{f}(\underline{\mathbf{e}}) = [f_1(\underline{\mathbf{e}}),\ldots,f_K(\underline{\mathbf{e}})]^{\mathsf{T}}$. Under Assumption 5, there exists a fixed point $\mathbf{f}(\underline{\mathbf{e}}^*) = \underline{\mathbf{e}}^*$, where $\underline{\mathbf{e}}^* \triangleq [\underline{e}_1^*,\ldots,\underline{e}_K^*]^{\mathsf{T}}$ with $\underline{e}_i^* > 0\ \forall i$. In this case, we can extend the results in [55][2] and show that the iterative fixed point algorithm defined by $\underline{\mathbf{e}}^{(n+1)} = \mathbf{f}(\underline{\mathbf{e}}^{(n)})$, $(n \ge 0)$, converges to the unique positive solution $\underline{\mathbf{e}}^*$ for any initial point $\underline{\mathbf{e}}^{(0)}$, $\underline{e}_i^{(0)} > 0\ \forall i$.

Furthermore, we need to show that both $\underline{\Upsilon}^\circ_k = \lim_{\alpha\to 0}\Upsilon^\circ_k$ and $\underline{\Psi}^\circ = \lim_{\alpha\to 0}\Psi^\circ$ exist and are uniformly bounded on M. Observe that

$$\lim_{\alpha\to 0}\alpha^2 e'_i = \underline{e}_i \quad (134)$$

and we obtain

$$\underline{\Psi}^\circ = \lim_{\alpha\to 0}\frac{1}{M}\sum_{j=1}^{K}p_j\frac{\alpha^2 e'_j}{(\alpha+\alpha e_j)^2} = \frac{1}{M}\sum_{j=1}^{K}\frac{p_j}{\underline{e}_j}. \quad (135)$$

[2]Since $\mathbf{f}(\underline{\mathbf{e}})$ can be extended by continuity in zero, where it satisfies $\mathbf{f}(0) = 0$, the positivity property of $\mathbf{f}(\underline{\mathbf{e}})$, defined in [55], does not hold. We precisely need to show that $\underline{\mathbf{e}}^{(n+1)} = \mathbf{f}(\underline{\mathbf{e}}^{(n)})$ can not converge to the fixed point 0, which unfolds from Assumption 5 with similar arguments as in [55].

Therefore, $0 < \underline{\Psi}^{\circ} < \infty$ for all $\underline{e}_i > 0$. Similarly, define $\underline{e}'_{j,k} = \lim_{\alpha\to 0} \alpha^2 e'_{j,k}$ given in (38) and thus

$$\underline{\Upsilon}_k^{\circ} = \lim_{\alpha\to 0} \frac{1}{M} \sum_{j=1, j\neq k}^{K} p_j \frac{\alpha^2 e'_{j,k}}{(\alpha + \alpha e_j)^2} = \frac{1}{M} \sum_{j=1, j\neq k}^{K} p_j \frac{\underline{e}'_{j,k}}{\underline{e}_j^2},$$

satisfying $0 < \underline{\Upsilon}_k^{\circ} < \infty$ for all $\underline{e}_i > 0$. To fulfill the constraints $\underline{e}_i > 0$, we have to evoke Assumption 4. The limit $\gamma_{k,\mathrm{zf}}^{\circ} = \lim_{\alpha\to 0} \gamma_{k,\mathrm{rzf}}^{\circ}(\alpha)$ is given by (34), which completes the proof.

Appendix IV
Proof of Proposition 2

The proof is inspired by [26] with adaptations to account for imperfect CSIT. From Corollary 1 with $p_k = P/K\ \forall k$ and $\tau_k = \tau\ \forall k$, for large M, K, the SINR $\gamma_{\mathrm{rzf}}^{\circ}$ takes the form

$$\gamma_{\mathrm{rzf}}^{\circ} = \rho\beta m^{\circ}(1-\tau^2)\Gamma,$$

where

$$\Gamma = \frac{\frac{1}{\beta} e_{22} + \alpha(1+m^{\circ})^2 e_{12}}{\rho e_{22}(1-\tau^2) + \tau^2\rho(1+m^{\circ})^2 e_{22} + (1+m^{\circ})^2 e_{12}}$$

with m° and e_{ij} defined in (27) and (29), respectively. Taking the derivative along α, we obtain

$$\frac{\partial \gamma_{\mathrm{rzf}}^{\circ}}{\partial \alpha} = \rho\beta m^{\circ}(1-\tau^2)\Gamma \left[\frac{m'^{\circ}}{m^{\circ}} + \frac{\Gamma'}{\Gamma}\right], \tag{136}$$

where

$$m'^{\circ} = -\frac{(1+m^{\circ})^2 e_{12}}{1 - \frac{e_{22}}{\beta}}. \tag{137}$$

and thus, together with (30), we have

$$\frac{m'^{\circ}}{m^{\circ}} = -\frac{(1+m^{\circ})^2 e_{12}}{\frac{1}{\beta} e_{22} + \alpha(1+m^{\circ})^2 e_{12}}.$$

Therefore, (136) becomes

$$\begin{aligned}\frac{\partial \gamma_{\mathrm{rzf}}^{\circ}}{\partial \alpha} &= \rho\beta m^{\circ}(1-\tau^2)\Gamma \\ &\times \left[\frac{2\alpha(1+m^{\circ})m'^{\circ}e_{12} + \alpha(1+m^{\circ})^2 e'_{12} + \frac{1}{\beta}e'_{22}}{\frac{1}{\beta}e_{22} + \alpha(1+m^{\circ})^2 e_{12}}\right. \\ &- \frac{[1-\tau^2+\tau^2(1+m^{\circ})^2]\rho e'_{22} + 2\tau^2\rho(1+m^{\circ})m'^{\circ}e_{22}}{[1-\tau^2+\tau^2(1+m^{\circ})^2]\rho e_{22} + (1+m^{\circ})^2 e_{12}} \\ &\left. - \frac{2(1+m^{\circ})m'^{\circ}e_{12} + (1+m^{\circ})^2 e'_{12}}{[1-\tau^2+\tau^2(1+m^{\circ})^2]\rho e_{22} + (1+m^{\circ})^2 e_{12}}\right].\end{aligned} \tag{138}$$

Denoting $\chi \triangleq (1+m^{\circ})^2 e_{12}$, $\psi \triangleq 2(1+m^{\circ})m'^{\circ}e_{12} + (1+m^{\circ})^2 e'_{12}$ and $\phi \triangleq 1-\tau^2+\tau^2(1+m^{\circ})^2$, (138) takes the form

$$\begin{aligned}\frac{\partial \gamma_{\mathrm{rzf}}^{\circ}}{\partial \alpha} &= \rho\beta m^{\circ}(1-\tau^2)\Gamma \\ &\times \left[\frac{\frac{1}{\beta}e'_{22} + \alpha\psi}{\frac{1}{\beta}e_{22} + \alpha\chi} - \frac{\rho\phi e'_{22} + \psi + 2\tau^2\rho(1+m^{\circ})m'^{\circ}e_{22}}{\rho\phi e_{22} + \chi}\right] \\ &= \frac{\phi\rho^2\beta m^{\circ}(1-\tau^2)\Gamma}{Z}\left[\left(\alpha - \frac{1}{\beta\rho\phi}\right)(e_{22}\psi - e'_{22}\chi)\right. \\ &\left. - \frac{2\tau^2(1+m^{\circ})m'^{\circ}e_{22}[\frac{e_{22}}{\beta} + \alpha\chi]}{\phi}\right],\end{aligned}$$

where $Z = (\frac{1}{\beta}e_{22} + \alpha\chi)(\rho\phi e_{22} + \chi)$. Denoting

$$\begin{aligned}\Omega &\triangleq \frac{2\phi\rho^2\beta m^{\circ}(1-\tau^2)(1+m^{\circ})m'^{\circ}e_{12}e_{22}\Gamma}{Z} \\ \nu &\triangleq \frac{(1+m^{\circ})^2[e'_{12}e_{22} - e_{12}e'_{22}]}{2(1+m^{\circ})m'^{\circ}e_{12}e_{22}},\end{aligned} \tag{139}$$

we obtain

$$\frac{\partial \gamma_{\mathrm{rzf}}^{\circ}}{\partial \alpha} = \Omega\left[\left(\alpha - \frac{1}{\beta\rho\phi}\right)(1+\nu) - \frac{\tau^2[\frac{e_{22}}{\beta} + \alpha\chi]}{\phi e_{12}}\right]. \tag{140}$$

Rewriting the term in brackets in (140), we have

$$\frac{\partial \gamma_{\mathrm{rzf}}^{\circ}}{\partial \alpha} = \Omega\left[\alpha - \frac{[1+\nu+\tau^2\rho\frac{e_{22}}{e_{12}}]\frac{1}{\beta\rho}}{(1-\tau^2)(1+\nu) + \tau^2\nu(1+m^{\circ})^2}\right] = 0.$$

Since $\Omega \neq 0$ for $\rho > 0$ and $\tau^2 < 1$, the optimal regularization parameter $\alpha^{\star\circ}$ is given by (53). Substituting (137) into (139), the term ν takes the form

$$\nu = \frac{1 - \frac{e_{22}}{\beta}}{2(1+m^{\circ})e_{12}} \frac{e'_{12}}{e_{22}} \left[\frac{e'_{22}}{e'_{12}} - \frac{e_{22}}{e_{12}}\right]. \tag{141}$$

With (30) and (137), we obtain $e'_{12} = \frac{-2e_{13}}{1-e_{22}/\beta}$ and $e'_{22} = \frac{-2e_{23}}{1-e_{22}/\beta}$. Substituting these terms into (141) yields (54), which completes the proof.

Appendix V
Proof of Proposition 9

The sum rate $\hat{R}_{\mathrm{sum}}$ can be written as a function of the per-user rate under perfect CSIT $\bar{R}^{\circ}$ and the per-user rate gap ΔR° as

$$\hat{R}_{\mathrm{sum}} = K\left(1 - \frac{T_t}{T}\right)\left[\bar{R}^{\circ} - \Delta R^{\circ}\right],$$

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

$$\begin{aligned}\Delta R_{\mathrm{zf}}^{\circ} &= \log\left(\frac{(\beta-1)(\rho_{dl}+1)}{1 + \frac{1}{\rho_{dl}} + T_{t,\mathrm{zf}}[\frac{1}{c} + \rho_{ul}(\beta-1)]}\right), \\ \Delta R_{\mathrm{rzf}}^{\circ} &= \log\left(\frac{1 + \rho_{dl}(\beta-1) + \chi(1)}{1 + \omega\rho_{dl}(\beta-1) + \chi(\omega)}\right),\end{aligned}$$

where $\chi(\omega)$ is defined in (85). Denoting $\psi \triangleq 1 + \frac{1}{\rho_{dl}} + T_{t,\mathrm{zf}}[\frac{1}{c} + \rho_{ul}(\beta-1)]$, the derivatives take the form

$$\begin{aligned}\frac{\partial \hat{R}_{\mathrm{sum}}^{\mathrm{zf}}}{\partial T_{t,\mathrm{zf}}} &= -\frac{K}{T}(\bar{R}_{\mathrm{zf}}^{\circ} - \Delta R_{\mathrm{zf}}^{\circ}) + K\left(1 - \frac{T_{t,\mathrm{zf}}}{T}\right) \\ &\times \frac{(\beta-1)(\rho_{dl}+1)[\frac{1}{c} + \rho_{ul}(\beta-1)]}{\psi^2 + (\beta-1)(\rho_{dl}+1)\psi},\end{aligned} \tag{142}$$

$$\begin{aligned}\frac{\partial \hat{R}_{\mathrm{sum}}^{\mathrm{rzf}}}{\partial T_{t,\mathrm{rzf}}} &= -\frac{K}{T}(\bar{R}_{\mathrm{rzf}}^{\circ} - \Delta R_{\mathrm{rzf}}^{\circ}) \\ &+ K\left(1 - \frac{T_{t,\mathrm{rzf}}}{T}\right)\frac{\omega'\rho_{dl}(\beta-1) + \chi'}{1 + \omega\rho_{dl}(\beta-1) + \chi},\end{aligned} \tag{143}$$

where $\omega' = \partial\omega/\partial T_{t,\mathrm{rzf}} = (1/\rho_{ul} + c)/(T_{t,\mathrm{rzf}} + 1/\rho_{ul} + c)^2$ and $\chi' = \partial\chi/\partial T_{t,\mathrm{rzf}} = [(\beta-1)^2\omega\omega'\rho_{dl}^2 + \omega'\rho_{dl}(1+\beta) + 1]/\chi$. In (142) and (143) the per-user rate-gap $\Delta R_{\mathrm{zf}}^{\circ}$ and $\Delta R_{\mathrm{rzf}}^{\circ}$ can be neglected, since at high SNR $\Delta R_{\mathrm{zf}}^{\circ} \ll \bar{R}_{\mathrm{zf}}^{\circ}$ and $\Delta R_{\mathrm{rzf}}^{\circ} \ll \bar{R}_{\mathrm{rzf}}^{\circ}$, respectively. Treating $\bar{R}_{\mathrm{zf}}^{\circ}, \bar{R}_{\mathrm{rzf}}^{\circ}$ as constant, for $\rho_{dl}, \rho_{ul} \to \infty$

and $c=\rho_{dl}/\rho_{ul}$ finite, solving (142) and (143) for $T_{t,\mathrm{zf}}$ and $T_{t,\mathrm{rzf}}$, respectively, yields (88) and (89), respectively, which completes the proof.

APPENDIX VI
IMPORTANT LEMMAS

Lemma 1 (Matrix Inversion Lemma): [35, Lemma 2.2] Let $\mathbf{U}$ be an $N\times N$ invertible matrix and $\mathbf{x}\in\mathbb{C}^N$, $c\in\mathbb{C}$ for which $\mathbf{U}+c\mathbf{x}\mathbf{x}^{\mathsf{H}}$ is invertible. Then

$$\mathbf{x}^{\mathsf{H}}\left(\mathbf{U}+c\mathbf{x}\mathbf{x}^{\mathsf{H}}\right)^{-1}=\frac{\mathbf{x}^{\mathsf{H}}\mathbf{U}^{-1}}{1+c\mathbf{x}^{\mathsf{H}}\mathbf{U}^{-1}\mathbf{x}}.$$

Lemma 2 (Resolvent Identity): Let $\mathbf{U}$ and $\mathbf{V}$ be two invertible complex matrices of size $N\times N$. Then

$$\mathbf{U}^{-1}-\mathbf{V}^{-1}=-\mathbf{U}^{-1}(\mathbf{U}-\mathbf{V})\mathbf{V}^{-1}.$$

Lemma 3: [56, Lemma B.26] Let $\mathbf{A}\in\mathbb{C}^{N\times N}$ be a deterministic matrix and $\mathbf{x}\in\mathbb{C}^N$ have i.i.d. complex entries of zero mean, variance $1/N$ and bounded lth order moment $E|x_i|^l\leq\nu_l$. Then for any $p\geq 1$

$$E\left|\mathbf{x}^{\mathsf{H}}\mathbf{A}\mathbf{x}-\frac{1}{N}\mathrm{tr}\mathbf{A}\right|^p\leq\frac{C_p}{N^{p/2}}\left(\frac{1}{N}\mathrm{tr}\mathbf{A}\mathbf{A}^{\mathsf{H}}\right)^{p/2}\left[\nu_4^{p/2}+\nu_{2p}\right], \tag{144}$$

where C_p is a constant solely depending on p.

Lemma 4: [15, Lemma 14.2] Let $\mathbf{A}_1,\mathbf{A}_2,\ldots$, with $\mathbf{A}_N\in\mathbb{C}^{N\times N}$, be a series of random matrices generated by the probability space $(\Omega,\mathcal{F},P)$ such that, for $\omega\in A\subset\Omega$, with $P(A)=1$, $\|\mathbf{A}_N(\omega)\|<K(\omega)<\infty$, uniformly on N. Let $\mathbf{x}_1,\mathbf{x}_2,\ldots$, with $\mathbf{x}_N\in\mathbb{C}^N$, 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 $\mathbf{A}_N$. Then

$$\mathbf{x}_N^{\mathsf{H}}\mathbf{A}_N\mathbf{x}_N-\frac{1}{N}\mathrm{tr}\mathbf{A}_N\overset{N\to\infty}{\longrightarrow}0,$$

almost surely.

Proof: The proof unfolds from a direct application of the Tonelli theorem, [40, Theorem 18.3]. Denoting $(X,\mathcal{X},P_X)$ the probability space that generates the series $\mathbf{x}_1,\mathbf{x}_2,\ldots$, we have that for every $\omega\in A$ (i.e., for every realization $\mathbf{A}_1(\omega),\mathbf{A}_2(\omega),\ldots$), the trace lemma, [15, Theorem 3.4], holds true. From [40, Theorem 18.3], the space B of couples $(x,\omega)\in Y\triangleq X\times\Omega$ for which the trace lemma holds, satisfies

$$\int_Y 1_B(x,\omega)dP_Y(x,\omega)=\int_\Omega\int_X 1_B(x,\omega)dP_X(x)dP_\Omega(\omega).$$

If $\omega\in A$, then $1_B(x,\omega)=1$ on a subset of X of probability one. Therefore, the inner integral 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 $\mathbf{A}_N$ be as in Lemma 4 and $\mathbf{x}_N,\mathbf{y}_N\in\mathbb{C}^N$ 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 $\mathbf{A}_N$.

$$\mathbf{y}_N^{\mathsf{H}}\mathbf{A}_N\mathbf{x}_N\overset{N\to\infty}{\longrightarrow}0,$$

almost surely.

Proof: Remark that $E\left[|\mathbf{y}_N^{\mathsf{H}}\mathbf{A}_N\mathbf{x}_N|^4\right]<c/N^2$ for some constant $c>0$ independent of N. The result then unfolds from the Markov inequality the Borel-Cantelli Lemma [40] and the Tonelli Theorem [40, Theorem 18.3]. ∎

Lemma 6: [15, Lemma 14.3] Let $\mathbf{A}_1,\mathbf{A}_2,\ldots$, with $\mathbf{A}_N\in\mathbb{C}^{N\times N}$, be deterministic with uniformly bounded spectral norm and $\mathbf{B}_1,\mathbf{B}_2,\ldots$, with $\mathbf{B}_N\in\mathbb{C}^{N\times N}$, be random Hermitian, with eigenvalues $\lambda_1^{\mathbf{B}_N}\leq\ldots\leq\lambda_N^{\mathbf{B}_N}$ such that, with probability one, there exist $\varepsilon>0$ for which $\lambda_1^{\mathbf{B}_N}>\varepsilon$ for all large N. Then for $\mathbf{v}\in\mathbb{C}^N$

$$\frac{1}{N}\mathrm{tr}\mathbf{A}_N\mathbf{B}_N^{-1}-\frac{1}{N}\mathrm{tr}\mathbf{A}_N(\mathbf{B}_N+\mathbf{v}\mathbf{v}^{\mathsf{H}})^{-1}\overset{N\to\infty}{\longrightarrow}0$$

almost surely, where $\mathbf{B}_N^{-1}$ and $(\mathbf{B}_N+\mathbf{v}\mathbf{v}^{\mathsf{H}})^{-1}$ 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 $\mathbf{B}_N(\omega)$ is uniformly greater than $\varepsilon(\omega)$. Therefore, with $\mathbf{B}_N(\omega)$ and $\mathbf{B}_N(\omega)+\mathbf{v}\mathbf{v}^{\mathsf{H}}$ invertible and, taking $z=-\varepsilon(\omega)/2$, we can write

$$\begin{aligned}&\frac{1}{N}\mathrm{tr}\mathbf{A}_N\mathbf{B}_N^{-1}(\omega)\\&=\frac{1}{N}\mathrm{tr}\mathbf{A}_N\left(\left[\mathbf{B}_N(\omega)-\frac{\varepsilon(\omega)}{2}\mathbf{I}_N\right]+\frac{\varepsilon(\omega)}{2}\mathbf{I}_N\right)^{-1}\end{aligned}$$

and

$$\begin{aligned}&\frac{1}{N}\mathrm{tr}\mathbf{A}_N\left(\mathbf{B}_N(\omega)+\mathbf{v}\mathbf{v}^{\mathsf{H}}\right)^{-1}\\&=\frac{1}{N}\mathrm{tr}\mathbf{A}_N\left(\left[\mathbf{B}_N(\omega)+\mathbf{v}\mathbf{v}^{\mathsf{H}}-\frac{\varepsilon(\omega)}{2}\mathbf{I}_N\right]+\frac{\varepsilon(\omega)}{2}\mathbf{I}_N\right)^{-1}.\end{aligned}$$

Under these notations, $\mathbf{B}_N(\omega)-\frac{\varepsilon(\omega)}{2}\mathbf{I}_N$ and $\mathbf{B}_N(\omega)+\mathbf{v}\mathbf{v}^{\mathsf{H}}-\frac{\varepsilon(\omega)}{2}\mathbf{I}_N$ are still nonnegative definite for all N. Therefore, the rank-1 perturbation lemma, [57, Lemma 2.1], can be applied for this ω. But then, from the Tonelli theorem again, in the space that generates the couples $((\mathbf{x}_1,\mathbf{x}_2,\ldots),(\mathbf{B}_1,\mathbf{B}_2,\ldots))$, the subspace where the rank-1 perturbation lemma applies has probability one, which completes the proof. ∎

Lemma 7: Let $\mathbf{U},\mathbf{V},\mathbf{\Theta}\in\mathbb{C}^{N\times N}$ be of uniformly bounded spectral norm with respect to N and let $\mathbf{V}$ be invertible. Further, define $\mathbf{x}\triangleq\mathbf{\Theta}^{1/2}\mathbf{z}$ and $\mathbf{y}\triangleq\mathbf{\Theta}^{1/2}\mathbf{q}$ where $\mathbf{z},\mathbf{q}\in\mathbb{C}^N$ 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 $\mathbf{U},\mathbf{V}$. Define $c_0,c_1,c_2\in\mathbb{R}^+$ such that $c_0c_1-c_2^2\geq 0$ and let $u\triangleq\frac{1}{N}\mathrm{tr}\mathbf{\Theta}\mathbf{V}^{-1}$ and $u'\triangleq\frac{1}{N}\mathrm{tr}\mathbf{\Theta}\mathbf{U}\mathbf{V}^{-1}$. Then we have

$$\begin{aligned}&\mathbf{x}^{\mathsf{H}}\mathbf{U}\left(\mathbf{V}+c_0\mathbf{x}\mathbf{x}^{\mathsf{H}}+c_1\mathbf{y}\mathbf{y}^{\mathsf{H}}+c_2\mathbf{x}\mathbf{y}^{\mathsf{H}}+c_2\mathbf{y}\mathbf{x}^{\mathsf{H}}\right)^{-1}\mathbf{x}\\&-\frac{u'(1+c_1u)}{(c_0c_1-c_2^2)u^2+(c_0+c_1)u+1}\overset{N\to\infty}{\longrightarrow}0,\end{aligned}$$

almost surely. Furthermore,

$$\begin{aligned}&\mathbf{x}^{\mathsf{H}}\mathbf{U}\left(\mathbf{V}+c_0\mathbf{x}\mathbf{x}^{\mathsf{H}}+c_1\mathbf{y}\mathbf{y}^{\mathsf{H}}+c_2\mathbf{x}\mathbf{y}^{\mathsf{H}}+c_2\mathbf{y}\mathbf{x}^{\mathsf{H}}\right)^{-1}\mathbf{y}\\&-\frac{-c_2uu'}{(c_0c_1-c_2^2)u^2+(c_0+c_1)u+1}\overset{N\to\infty}{\longrightarrow}0,\end{aligned}$$

almost surely.

Proof: Denote $\mathbf{V} = (\mathbf{A} + c_0\mathbf{x}\mathbf{x}^{\mathsf{H}} + c_1\mathbf{y}\mathbf{y}^{\mathsf{H}} + c_2\mathbf{x}\mathbf{y}^{\mathsf{H}} + c_2\mathbf{y}\mathbf{x}^{\mathsf{H}})^{-1}$. Now $\mathbf{x}^{\mathsf{H}}\mathbf{U}\mathbf{V}\mathbf{x}$ can be resolved using Lemma 2

$$\begin{aligned}\mathbf{x}^{\mathsf{H}}\mathbf{U}\mathbf{V}\mathbf{x} - \mathbf{x}^{\mathsf{H}}\mathbf{U}\mathbf{A}^{-1}\mathbf{x} &= \mathbf{x}^{\mathsf{H}}\mathbf{U}\mathbf{V}\left(\mathbf{V}^{-1} - \mathbf{A}\right)\mathbf{A}^{-1}\mathbf{x} \\ &= -\mathbf{x}^{\mathsf{H}}\mathbf{U}\mathbf{V}(c_0\mathbf{x}\mathbf{x}^{\mathsf{H}} + c_1\mathbf{y}\mathbf{y}^{\mathsf{H}} + c_2\mathbf{x}\mathbf{y}^{\mathsf{H}} + c_2\mathbf{y}\mathbf{x}^{\mathsf{H}})\mathbf{A}^{-1}\mathbf{x}.\end{aligned} \tag{145}$$

Rewrite (145) as

$$\mathbf{x}^{\mathsf{H}}\mathbf{U}\mathbf{V}\mathbf{x} = \frac{\mathbf{x}^{\mathsf{H}}\mathbf{U}\mathbf{A}^{-1}\mathbf{x} - \mathbf{x}^{\mathsf{H}}\mathbf{U}\mathbf{V}\mathbf{y}(c_1\mathbf{y}^{\mathsf{H}}\mathbf{A}^{-1}\mathbf{x} + c_2\mathbf{x}^{\mathsf{H}}\mathbf{A}^{-1}\mathbf{x})}{1 + c_0\mathbf{x}^{\mathsf{H}}\mathbf{A}^{-1}\mathbf{x} + c_2\mathbf{y}^{\mathsf{H}}\mathbf{A}^{-1}\mathbf{x}}.$$

Similarly to (145), we apply Lemma 2 to $\mathbf{x}^{\mathsf{H}}\mathbf{U}\mathbf{V}\mathbf{y}$. Thus, we obtain an expression involving the terms $\mathbf{x}^{\mathsf{H}}\mathbf{U}\mathbf{A}^{-1}\mathbf{x}$, $\mathbf{y}^{\mathsf{H}}\mathbf{A}^{-1}\mathbf{y}$, $\mathbf{x}^{\mathsf{H}}\mathbf{U}\mathbf{A}^{-1}\mathbf{y}$ and $\mathbf{y}^{\mathsf{H}}\mathbf{A}^{-1}\mathbf{x}$. To complete the proof, we apply Lemma 4 and Lemma 5, with $u = \frac{1}{N}\mathrm{tr}\mathbf{\Theta}\mathbf{A}^{-1}$ and $u' = \frac{1}{N}\mathrm{tr}\mathbf{\Theta}\mathbf{U}\mathbf{A}^{-1}$ and obtain

$$\mathbf{x}^{\mathsf{H}}\mathbf{U}\mathbf{V}\mathbf{x} - \frac{u'(1 + c_1 u)}{(c_0 c_1 - c_2^2)u^2 + (c_0 + c_1)u + 1} \overset{N\to\infty}{\longrightarrow} 0, \tag{146}$$

almost surely. Similarly we have

$$\mathbf{x}^{\mathsf{H}}\mathbf{U}\mathbf{V}\mathbf{y} - \frac{-c_2 u u'}{(c_0 c_1 - c_2^2)u^2 + (c_0 + c_1)u + 1} \overset{N\to\infty}{\longrightarrow} 0, \tag{147}$$

almost surely. Note that as $c_0, c_1, c_2 \in \mathbb{R}^+$ and $c_0 c_1 \geq c_2^2$, the convergence in (146) and (147) still holds since $(c_0 c_1 - c_2^2)u^2 + (c_0 + c_1)u + 1$ is bounded away from zero, which completes the proof. ■

Lemma 8: [57, Lemma 2.1] Let $\zeta > 0$, $\mathbf{B}, \mathbf{A} \in \mathbb{C}^{N\times N}$ with $\mathbf{B}$ Hermitian nonnegative definite, $\tau \in \mathbb{R}$ and $\mathbf{q} \in \mathbb{C}^N$. Then

$$\left|\mathrm{tr}\mathbf{A}\left[(\mathbf{B} + \zeta\mathbf{I}_N)^{-1} - (\mathbf{B} + \tau\mathbf{q}\mathbf{q}^{\mathsf{H}} + \zeta\mathbf{I}_N)^{-1}\right]\right| \leq \frac{\|\mathbf{A}\|}{\zeta}.$$

Lemma 9: [15, Corollary 2.2] Let $z \in \mathbb{C}^+$, $t > 0$, $\mathbf{q} \in \mathbb{C}^N$ and $\mathbf{B} \in \mathbb{C}^{N\times N}$ Hermitian nonnegative definite. Then

$$\left|\frac{1}{1 + t\mathbf{q}^{\mathsf{H}}\left(\mathbf{B} + z\mathbf{I}_N\right)^{-1}\mathbf{q}}\right| \leq \frac{|z|}{\Im z}.$$

Lemma 10: Let $\mathbf{q} \in \mathbb{C}^N$ and $\mathbf{A} \in \mathbb{C}^{N\times N}$ Hermitian nonnegative definite, then

$$\mathbf{q}^{\mathsf{H}}\mathbf{A}\mathbf{q} \leq \|\mathbf{A}\|\|\mathbf{q}\|_2^2.$$

Lemma 11: Let $\mathbf{A} \in \mathbb{C}^{N\times N}$ be Hermitian nonnegative-definite, then

$$\frac{1}{N}\mathrm{tr}\mathbf{A} \leq \|\mathbf{A}\|.$$

Lemma 12: Let $\mathbf{A}, \mathbf{B} \in \mathbb{C}^{N\times N}$ Hermitian nonnegative-definite, then

$$\|\mathbf{A}\mathbf{B}\| \leq \|\mathbf{A}\|\|\mathbf{B}\|.$$

ACKNOWLEDGEMENTS

This research has been partially supported by EU NoE Newcom++ and ANR project SESAME.

REFERENCES

[1] G. J. Foschini and M. J. Gans, "On Limits of Wireless Communications in a Fading Environment when Using Multiple Antennas," *Wireless Personal Communications*, vol. 6, no. 3, pp. 311–335, Mar. 1998.

[2] E. Telatar, "Capacity of Multi-antenna Gaussian Channels," *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–595, Feb. 1999.

[3] D. Gesbert, M. Kountouris, R. W. Heath, Jr., C. B. Chae, and T. Sälzer, "From Single User to Multiuser Communications: Shifting the MIMO Paradigm," *IEEE Signal Process. Mag.*, vol. 24, no. 5, pp. 36–46, May 2007.

[4] G. Caire and S. Shamai, "On the Achievable Throughput of a Multi-antenna Gaussian Broadcast Channel," *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1691–1706, Jul. 2003.

[5] P. Viswanath and D. N. C. Tse, "Sum Capacity of the Vector Gaussian Broadcast Channel and Uplink-Downlink Duality," *IEEE Trans. Inf. Theory*, vol. 49, no. 8, pp. 1912–1921, Aug. 2003.

[6] W. Yu and J. M. Cioffi, "Sum Capacity of Gaussian Vector Broadcast Channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 9, pp. 1875–1892, Sep. 2004.

[7] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, Achievable Rates, and Sum-Rate Capacity of Gaussian MIMO Broadcast Channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2658–2668, Oct. 2003.

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

[9] C. B. Peel, B. M. Hochwald, and A. L. Swindlehurst, "A Vector-Perturbation Technique for Near-Capacity Multiantenna Multiuser Communication–Part I: Channel Inversion and Regularization," *IEEE Trans. Commun.*, vol. 53, no. 1, pp. 195–202, Jan. 2005.

[10] T. Yoo and A. Goldsmith, "On the Optimality of Multiantenna Broadcast Scheduling Using Zero-Forcing Beamforming," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 3, pp. 528–541, Mar. 2006.

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

[12] S. Christensen, R. Agarwal, and J. M. Cioffi, "Weighted Sum-Rate Maximization using Weighted MMSE for MIMO-BC Beamforming Design," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 4792–4799, Dec. 2008.

[13] S. Shi, M. Schubert, and H. Boche, "Rate Optimization for Multiuser MIMO Systems With Linear Processing," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 4020–4030, Aug. 2008.

[14] D. J. Love, R. W. Heath, Jr., V. K. N. Lau, D. Gesbert, B. D. Rao, and M. Andrews, "An Overview of Limited Feedback in Wireless Communication Systems," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 8, pp. 1341–1365, Oct. 2008.

[15] R. Couillet and M. Debbah, *Random Matrix Methods for Wireless Communications*, 1st ed. Cambridge, UK: Cambridge University Press, 2011.

[16] A. M. Tulino and S. Verdú, *Random Matrix Theory and Wireless Communications*. Delft, Netherlands: Now Publishers Inc., 2004.

[17] Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA), "Further advancements for E-UTRA physical layer aspects (Release 9)," 3GPP TR 36.814 V9.0.0, Tech. Rep., Mar. 2010.

[18] B. Hochwald and S. Vishwanath, "Space-Time Multiple Access: Linear Growth in the Sum Rate," in *Proc. IEEE Annual Allerton Conference on Communication, Control, and Computing*, Monticello, Illinois, Oct. 2002, pp. 387–396.

[19] M. Joham, K. Kusume, M. H. Gzara, W. Utschick, and J. A. Nossek, "Transmit Wiener Filter for the Downlink of TDD DS-CDMA Systems," in *Proc. IEEE International Symposium on Spread Spectrum Techniques and Applications (ISSSTA'02)*, vol. 1, Prague, Czech Republic, Sep. 2002, pp. 9–13.

[20] D. Hwang, B. Clercks, and G. Kim, "Regularized Channel Inversion with Quantized Feedback in Down-link Multiuser Channels," *IEEE Trans. Wireless Commun.*, vol. 8, no. 12, pp. 5785–5789, Dec. 2009.

[21] R. Couillet, S. Wagner, M. Debbah, and A. Silva, "The Space Frontier: Physical Limits of Multiple Antenna Information Transfer," in *Proceedings of the 3rd ACM International Conference on Performance Evaluation Methodologies and Tools (VALUETOOLS'08)*, no. 84, Athens, Greece, Oct. 2008.

[22] V. K. Nguyen and J. S. Evans, "Multiuser Transmit Beamforming via Regularized Channel Inversion: A Large System Analysis," in *Proc.*

IEEE Global Communications Conference (GC'08), New Orleans, LO, Dec. 2008, pp. 1–4.

[23] R. Muharar and J. Evans, "Downlink beamforming with transmit-side channel correlation: A large system analysis," in *Proc. IEEE International Conference on Communications (ICC'11)*, Kyoto, Japan, Jun. 2011.

[24] A. Wiesel, Y. C. Eldar, and S. Shamai, "Linear Precoding via Conic Optimization for Fixed MIMO Receivers," *IEEE Trans. Signal Process.*, vol. 54, no. 1, pp. 161–176, Jan. 2006.

[25] R. Zakhour and S. V. Hanly, "Base station cooperation on the downlink: Large system analysis," *IEEE Trans. Inf. Theory*, Jun. 2010. [Online]. Available: http://arxiv.org/abs/1006.3360

[26] R. Muharar and J. Evans, "Downlink Beamforming with Transmit-side Channel Correlation: A Large System Analysis," *unpublished*, 2009. [Online]. Available: http://www.cubinlab.ee.unimelb.edu.au/ rmuharar/doc/techRepCorr.pdf

[27] V. L. Girko, *Theory of Stochastic Canonical Equations*, 1st ed. Boston, MA: Dordrecht, 2001.

[28] W. Hachem, P. Loubaton, and J. Najim, "Deterministic Equivalents for Certain Functionals of Large Random Matrices," *Annals of Applied Probability*, vol. 17, no. 3, pp. 875–930, Jun. 2007.

[29] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, NJ: Wiley & Sons, 2006.

[30] B. Nosrat-Makouei, J. G. Andrews, and R. W. Heath, Jr., "MIMO Interference Alignment Over Correlated Channels With Imperfect CSIT," *IEEE Trans. Signal Process.*, vol. 59, no. 6, pp. 2783–2794, Jun. 2011.

[31] M. Ding and S. D. Blostein, "MIMO Minimum Total MSE Transceiver Design With Imperfect CSI at Both Ends," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 1141–1150, Mar. 2009.

[32] C. Wang and R. D. Murch, "Adaptive Downlink Multi-User MIMO Wireless Systems for Correlated Channels with Imperfect CSI," *IEEE Trans. Wireless Commun.*, vol. 5, no. 9, pp. 2453–2446, Sep. 2006.

[33] T. Yoo and A. Goldsmith, "MIMO Capacity with Channel Uncertainty: Does Feedback Help?" in *Proc. IEEE Global Communications Conference (GC'04)*, Dallas, USA, Dec. 2004, pp. 96–100.

[34] W. Hachem, P. Loubaton, and J. Najim, "A CLT For Information-theoretic Statistics of Gram Random Matrices with a Given Variance Profile," *Annals of Applied Probability*, vol. 18, no. 6, pp. 2071–2130, Dec. 2009.

[35] J. W. Silverstein and Z. D. Bai, "On the Empirical Distribution of Eigenvalues of a Class of Large Dimensional Random Matrices," *Journal of Multivariate Analysis*, vol. 54, no. 2, pp. 175–192, Aug. 1995.

[36] R. Couillet, M. Debbah, and J. W. Silverstein, "A Deterministic Equivalent for the Capacity Analysis of Correlated Multi-user MIMO Channels," *IEEE Trans. Inf. Theory*, submitted for publication. [Online]. Available: http://arxiv.org/abs/0906.3667v3

[37] Z. D. Bai and J. W. Silverstein, "No Eigenvalues Outside the Support of the Limiting Spectral Distribution of Large Dimensional Sample Covariance Matrices," *Annals of Probability*, vol. 26, no. 1, pp. 316–345, Jan. 1998.

[38] S. Wagner, "MU-MIMO Transmission and Reception Techniques for the Next Generation of Cellular Wireless Standards (LTE-A)," Ph.D. dissertation, TELECOM ParisTech (EURECOM), Oct. 2011.

[39] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO: How many antennas do we need?" in *Proc. IEEE Annual Allerton Conference on Communication, Control, and Computing (Allerton'11)*, Urbana-Champaing, Illinois, US, Sep. 2011.

[40] P. Billingsley, *Probability and Measure*, 3rd ed. Hoboken, NJ: John Wiley & Sons, Inc., 1995.

[41] W. C. Jakes and D. C. Cox, *Microwave Mobile Communications*. Hoboken, NJ: Wiley-IEEE Press, 1994.

[42] S. Wagner and D. T. M. Slock, "Weighted Sum Rate Maximization of Correlated MISO Broadcast Channels under Linear Precoding: A Large System Analysis," in *Proc. IEEE International Workshop on Signal Processing Advances for Wireless Communications (SPAWC'11)*, San Francisco, USA, Jun. 2011.

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

[44] D. P. Palomar and J. R. Fonollosa, "Practical Algorithms for a Family of Waterfilling Solutions," *IEEE Trans. Signal Process.*, vol. 53, no. 2, pp. 686–695, Feb. 2005.

[45] N. Jindal, "MIMO Broadcast Channels With Finite-Rate Feedback," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 5045–5060, Nov. 2006.

[46] D. Marco and D. L. Neuhoff, "The validity of the additive noise model for uniform scalar quantizers," *IEEE Trans. Inf. Theory*, vol. 51, no. 5, pp. 1739–1755, May 2005.

[47] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, "Multiuser MIMO Achievable Rates With Downlink Training and Channel State Feedback," *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 2845–2866, Jun. 2010.

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

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

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

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

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

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

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

[55] R. D. Yates, "A Framework for Uplink Power Control in Cellular Radio Systems," *IEEE J. Sel. Areas Commun.*, vol. 13, no. 7, pp. 1341–1347, Sep. 1995.

[56] Z. Bai and J. W. Silverstein, *Spectral Analysis of Large Dimensional Random Matrices*, 2nd ed. Spring Street, NY: Springer, 2010.

[57] Z. D. Bai and J. W. Silverstein, "On the Signal-to-Interference Ratio of CDMA Systems in Wireless Communications," *Annals of Applied Probability*, vol. 17, no. 1, pp. 81–101, Feb. 2007.