

Interference-Aware RZF Precoding for Multi-Cell Downlink Systems

Axel Müller^{§*}, Romain Couillet[‡], Emil Björnson^{§†}, Sebastian Wagner^{*}, and Mérouane Debbah[§]

^{*}Intel Mobile Communications, Sophia Antipolis, France [†]ACCESS Linnaeus Centre, Signal Processing Lab, Sweden. [‡]Department of Telecommunications and [§]Alcatel-Lucent Chair on Flexible Radio, Supélec, France
 {axel.mueller, romain.couillet, merouane.debbah}@supelec.fr, sebastian.wagner@intel.com, emil.bjornson@ee.kth.se

Abstract—Recently, the structure of the optimal linear precoder for multi-cell downlink systems has been described in [1]. Other references (e.g., [2]) have used simplified versions of the precoder to obtain promising performance gains. These gains have been hypothesized to stem from providing additional degrees of freedom that allow for interference mitigation through interference relegation to orthogonal subspaces. However, no conclusive or rigorous understanding has yet been proposed.

In this paper, we take an interference-aware adaption of the generally optimal precoding structure and analyze the rate performance in multi-cell scenarios. A special emphasis is placed on induced interference mitigation. For example, we will verify the intuitive expectation that the precoder structure can either completely remove induced inter-cell or intra-cell interference. We state new results from large-scale random matrix theory, that make it possible to give more intuitive and insightful explanations of the precoder behavior, also for cases involving imperfect channel state information (CSI). We remark especially that the interference-aware precoder makes use of all available information about interfering channels to improve performance. Even extremely bad CSI can be used to enhance the sum rate. Our obtained insights are then used to propose heuristic precoder parameters for arbitrary systems, whose effectiveness is shown in more involved system scenarios. Furthermore, determining these parameters does not require explicit inter base station cooperation. Using a simple heuristic version of the interference aware precoder, one finds that a sum rate performance, close to the optimally parameterized precoder one, can be achieved.

I. INTRODUCTION

The growth of data traffic and the number of user terminals (UTs) in cellular networks, will likely persist for the foreseeable future [3]. In order to deal with the resulting demand, it is estimated [4] that a thousand-fold increase in network capacity is required over the next 10 years. Given that the available spectral resources are severely limited, the majority of the wireless community sees massive network densification as the most realistic approach to solving most pressing issues. Also historically, shrinking cell size has been the single most successful technique in satisfying demand for network capacity [5, Chapter 6.3.4]. In recent times, this technique has been named the *small cell* approach [6], [7]. However, interference still is a major limiting factor for capacity in multi-cell scenarios [8], [9]. Also, The situation

is unlikely to improve, as modern cellular networks serve a multitude of users, using the same time/frequency resources. In general, we see a trend to using more and more antennas for interference mitigation, e.g. via the “massive” multiple-input multiple-output (MIMO) approach [10]. Here the number of transmit antennas surpasses the number of served UTs by an order of magnitude. Independent of this specific approach, the surplus antennas can be used to mitigate interference by using spatial precoding [1], [11], [12], [9]. The interference problem is generally compounded by the effect of imperfect knowledge concerning the channel state information (CSI). Such imperfections are unavoidable, as imperfect estimation algorithms, limited number of orthogonal pilot sequences, mobile UTs, delays, etc. can not be avoided in practice. Hence, one is interested in employing precoding schemes that are robust to CSI error and exploit the available CSI as efficiently as possible.

Arguably, the most successful and practically applicable precoding scheme used today is regularized zero forcing (RZF) precoding [13] (also known as MMSE precoding, transmit Wiener filter, generalized eigenvalue-based beamformer, etc.; see [1, Remark 3.2] for a comprehensive history of this precoding scheme). Classical RZF precoders are only defined for single cell systems and thus do not take inter-cell interference into account. This disregard of valuable information is particularly wasteful in high density scenarios, where high interference levels are performance limiting. It is, hence, advisable to look for RZF related precoding schemes that exploit any additional information about out-of-cell interference. Early multi-cell extensions of the RZF scheme do not take the quality of CSI into account [14] and later ones either rely on heuristic distributed optimization algorithms or on inter-cell cooperation [15] to determine the precoding vector. Thus, they offer limited insight into the precoder structure, how the precoder works and how it can be improved.

In [1, Eq (3.33)] we find the most recent and general treatment of the multi-cell RZF precoder, along with a proof that the proposed structure is optimal w.r.t. many utility functions of practical interest (see also [16]). The precoding structure in question is the following:

$$\mathbf{F}_m^m = \left(\sum_{l=1}^L \sum_{k=1}^{K_l} \alpha_{l,k}^m \mathbf{h}_{l,k}^m (\mathbf{h}_{l,k}^m)^H + \gamma_m \mathbf{I}_{N_m} \right)^{-1} \mathbf{H}_m^m \nu_m^{\frac{1}{2}}$$

E. Björnson is funded by the International Postdoc Grant 2012-228 from The Swedish Research Council. This research has been supported by the ERC Starting Grant 305123 MORE (Advanced Mathematical Tools for Complex Network Engineering).

which we will call generalized RZF (genRZF) from now on. Here the channel vectors from a base station (BS) m to UT k in cell l , is denoted $\mathbf{h}_{l,k}^m$ and the aggregated channel matrix from BS m to cell l is \mathbf{H}_l^m . The factor γ_m is a regularization parameter and precoder normalization is done via the variable ν_m . We notice that each channel is separately assigned a factor $\alpha_{l,k}^m$, that can be interpreted as the importance placed on the respective estimated channel. In [2] a largely simplified version of genRZF was discussed, where a certain set of UT channels was weighted with respect to an estimated receive covariance matrix of the interference channels. Hoydis et al. argued, that “large regularization parameters make the precoding vectors more orthogonal to the interference subspace”, but did not conclusively and rigorously show how or why this is achieved. In this paper we propose and analyze an intermediate class of RZF precoders, that we denote *interference-aware RZF* (iaRZF):

$$\mathbf{F}_m^m = \left(\sum_{l=1}^L \alpha_l^m \hat{\mathbf{H}}_l^m (\hat{\mathbf{H}}_l^m)^H + \gamma_m \mathbf{I}_{N_m} \right)^{-1} \hat{\mathbf{H}}_m^m \nu_m^{\frac{1}{2}} \quad (1)$$

where the weighting takes place with respect to each cell using α_l^m and $\hat{\mathbf{H}}_l^m$ is actually the imperfect estimate of each aggregated channel matrix. This structure achieves a middle ground between genRZF and the approach in [2]. It reduces the number of variables that need to be optimized, but still allows to manage interference induced to other cells in an multi-cell environment, by using excess antennas at the BSs. The weights α_l^m can be interpreted as a factor of importance placed on the respective estimated channels. It allows the balancing of signal power directed to the served users, with causing interference to other cells. This can be used to optimize sum rate performance, as will be shown in Section II. We note that estimation of the inter-cell interference can be considered as implicit coordination, but no inter-cell communication is necessary.

Building on our work in [17], this paper analyzes the proposed iaRZF scheme, showing that it can significantly improve sum-rate performance in high interference multi cellular scenarios. In particular, it is not necessary to have reliable estimations of interfering channels; even very bad CSI allow for significant gains. We facilitate intuitive understanding of the precoder through new methods of analysis in both finite and large dimensions. Special emphasis is placed on the induced interference mitigation mechanism of iaRZF. To obtain fundamental insights, we consider the large-system regime where the number of transmit antennas and UTs are both large. Our main contributions are as follows.

- We derive deterministic expressions for the asymptotic user rates, which also serve as accurate approximations in practical non-asymptotic regimes. Merely, the channel statistics are needed for calculation and implementation of our deterministic expressions.
- These novel expressions generalize the prior work in [18] for single-cell systems and in [19] for multi-cell systems where only deterministic statistical CSI is utilized for suppression of inter-cell interference.

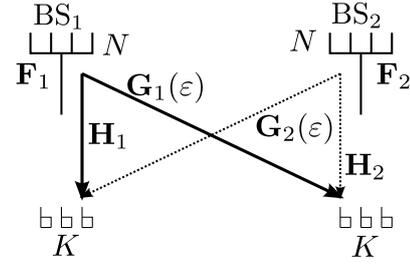


Fig. 1. Simple 2BSs Downlink System.

- These extensions are used to optimize the sum rate of the iaRZF precoding scheme in limiting cases.
- We propose and explain the appropriate heuristic scaling of the precoder weights w.r.t. various system parameters, that offers close to optimal sum rate performance; also in non-limit cases.
- Furthermore, new finite dimensional approaches to analyzing multi-cell RZF precoding schemes are introduced and applied for limiting cases.

The notation in this paper adheres to the following general rules. Boldface lower case is used for column vectors, and upper case for matrices. \mathbf{X}^T and \mathbf{X}^H denote the transpose, and conjugate transpose of \mathbf{X} , respectively, while $\text{tr}(\mathbf{X})$ is the matrix trace function. The expectation operator is denoted $\mathbb{E}[\cdot]$. The spectral norm of \mathbf{X} is denoted $\|\mathbf{X}\|_2$ and the euclidean norm of \mathbf{x} is denoted $\|\mathbf{x}\|_2$. Circularly symmetric complex Gaussian random vectors are denoted $\mathcal{CN}(\bar{\mathbf{x}}, \mathbf{Q})$, where $\bar{\mathbf{x}}$ is the mean and \mathbf{Q} is the covariance matrix. The set of all complex numbers is denoted by \mathbb{C} , with $\mathbb{C}^{N \times 1}$ and $\mathbb{C}^{N \times M}$ being the generalizations to vectors and matrices, respectively. The $M \times M$ identity matrix is written as \mathbf{I}_M , the zero vector of length M is denoted $\mathbf{0}_{M \times 1}$ and the zero matrix $\mathbf{0}_M$. Throughout this paper, superscripts generally refer to the origin (e.g., cell m) and subscripts generally denote the destination (e.g., cell l or UT k of cell l), when both information are needed. We employ \perp and $\not\perp$ to mean stochastic independence and dependence, respectively.

II. UNDERSTANDING IARZF

In order to intuitively understand and motivate the iaRZF precoder we first analyze its behavior and impact in a relatively simple system, which is introduced in the following subsection.

A. Simple System

We start by examining a simple downlink system depicted in Figure 1 that is a further simplification of the Wyner model [20], [21]. It features 2 BSs, BS₁ and BS₂, with N antennas each. Every BS serves one cell with K single antenna users. For convenience we introduce the notations $c = K/N$ and $\bar{x} = \text{mod}(x, 2) + 1$, $x \in \{1, 2\}$. In order to circumvent scheduling complications, we assume $N \geq K$. The aggregated channel matrix between BS _{x} and the affiliated users is denoted $\mathbf{H}_x = [\mathbf{h}_{x,1}, \dots, \mathbf{h}_{x,K}] \in \mathbb{C}^{N \times K}$ and the matrix pertaining to the

users of the other cell $\mathbf{G}_x(\varepsilon) = [\mathbf{g}_{x,1}, \dots, \mathbf{g}_{x,K}] \in \mathbb{C}^{N \times K}$, which is usually abbreviated as \mathbf{G}_x . We generally treat ε as an interference channel gain/path-loss factor. The precoding matrix used at BS_{*x*} is designated by $\mathbf{F}_x \in \mathbb{C}^{N \times K}$. For the channel realizations we choose a simple block-wise fast fading model, where $\mathbf{h}_{x,k} \sim \mathcal{CN}(0, \frac{1}{N} \mathbf{I}_N)$ and $\mathbf{g}_{x,k} \sim \mathcal{CN}(0, \varepsilon \frac{1}{N} \mathbf{I}_N)$ for $k = 1, \dots, K$.

Denoting $\mathbf{f}_{x,k}$ the k th column of \mathbf{F}_x , $\mathbf{F}_{x[k]}$ as \mathbf{F}_x with its k th column removed and $n_{x,k} \sim \mathcal{CN}(0, 1)$ the received additive Gaussian noise at UT_{*x,k*}, we define the received signal at UT_{*x,k*} as

$$y_{x,k} = \mathbf{h}_{x,k}^H \mathbf{f}_{x,k} s_{x,k} + \underbrace{\mathbf{h}_{x,k}^H \mathbf{F}_{x[k]} \mathbf{s}_{x[k]}}_{\text{intra-cell interference}} + \underbrace{\mathbf{g}_{\bar{x},k}^H \mathbf{F}_{\bar{x}} \mathbf{s}_{\bar{x}}}_{\text{inter-cell interference}} + n_{x,k} \quad (2)$$

where $\mathbf{s}_x \sim \mathcal{CN}(0, \rho_x \mathbf{I}_N)$ ¹ is the vector of transmitted Gaussian symbols. It defines the average per UT transmit power of BS_{*x*} as ρ_x (normalized w.r.t. noise). The notations $\mathbf{s}_{x[k]}$ and $s_{x,k}$ designate the transmit vector without symbol k and the transmit symbol of UT_{*x,k*}.

When calculating the precoder \mathbf{F}_x , we assume that the channel \mathbf{H}_x can be correctly estimated, however, we allow for mis-estimation of the ‘‘inter-cell interference channel’’ \mathbf{G}_x by adopting the generic Gauss-Markov formulation

$$\hat{\mathbf{G}}_x = \sqrt{1-\tau^2} \mathbf{G}_x + \tau \tilde{\mathbf{G}}_x.$$

Choosing $\tilde{\mathbf{g}}_{x,k} \sim \mathcal{CN}(0, \varepsilon \frac{1}{N} \mathbf{I}_N)$, we can vary the available CSI quality by adjusting $0 \leq \tau \leq 1$ appropriately.

In this section we choose the precoding to be the previously introduced iaRZF, the unnormalized form of which the simple system reads

$$\mathbf{M}_x = (\alpha_x \mathbf{H}_x \mathbf{H}_x^H + \beta_x \mathbf{G}_x \mathbf{G}_x^H + \gamma_x \mathbf{I})^{-1} \mathbf{H}_x. \quad (3)$$

One remarks that the normalization of the identity matrix can also be controlled by only scaling α_x and β_x at the same time and fixing γ_x to an arbitrary value (e.g., 1). We still keep all three variables to facilitate easy adaption for applications that are closer to traditional RZF (set $\alpha, \beta = 1$) or closer to the general precoder (set $\gamma = 1$). We assume the following normalization of the precoder:

$$\mathbf{F}_x = \sqrt{K} \frac{\mathbf{M}_x}{\sqrt{\text{tr}(\mathbf{M}_x^H \mathbf{M}_x)}} \quad (4)$$

i.e., it is assured that the sum energy of the precoder $\text{tr}(\mathbf{F}_x^H \mathbf{F}_x)$ is K ².

Remark 1 (Channel Scaling $1/N$). *The statistics of the channel matrices in this section incorporate the factor $1/N$, which simplifies comparisons with the later, more general, large-scale results (see Section III). This can also be interpreted, as transferring a scaling of the transmit power into the channel itself. The precoder formulations presented in the current*

¹We remark that ρ_x is of order 1.

²It can be shown, using results from Appendix C-A, that this implies $\|\mathbf{f}_{x,k}\|_2^2 \rightarrow 1$, almost surely, under Assumption 1 for the given simplified system.

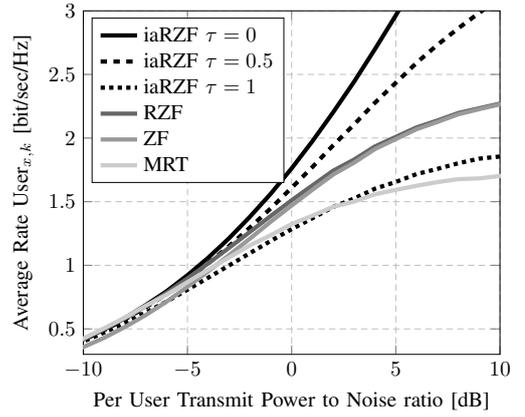


Fig. 2. Average user rate vs. transmit power to noise ratio ($N = 160$, $K = 40$, $\varepsilon = 0.7$, $\rho_1 = \rho_2 = \rho$).

section can be simply rewritten to fit the more traditional statistics of $\mathbf{h}_k \sim \mathcal{CN}(0, \mathbf{I}_N)$ and $\mathbf{g}_k \sim \mathcal{CN}(0, \varepsilon \mathbf{I}_N)$, by using

$$\tilde{\mathbf{M}}_x = (\alpha_x \mathbf{H}_x \mathbf{H}_x^H + \beta_x \mathbf{G}_x \mathbf{G}_x^H + N\gamma_x \mathbf{I})^{-1} \mathbf{H}_x$$

instead of \mathbf{M} . This equation shows that, under the chosen model, the regularization implicitly scales with N . However, one can either chose γ or α, β appropriately, to achieve any scaling.

B. Performance of Simple System

First, we compare the general performance of the proposed iaRZF scheme with classical approaches, i.e., non-cooperative zero-forcing (ZF), maximum-ratio transmission (MRT) and RZF. The rate of UT_{*x,k*} can be defined as

$$\text{Rate}_{x,k} = \log_2 \left(1 + \frac{\text{Sig}_{x,k}}{\text{Int}_{x,k}^a + \text{Int}_{x,k}^r + 1} \right)$$

where $\text{Sig}_{x,k} = \rho_x \mathbf{h}_{x,k}^H \mathbf{f}_{x,k} \mathbf{f}_{x,k}^H \mathbf{h}_{x,k}$, $\text{Int}_{x,k}^a = \rho_x \mathbf{h}_{x,k}^H \mathbf{F}_{x[k]} \mathbf{F}_{x[k]}^H \mathbf{h}_{x,k}$ and $\text{Int}_{x,k}^r = \rho_{\bar{x}} \mathbf{g}_{\bar{x},k}^H \mathbf{F}_{\bar{x}} \mathbf{F}_{\bar{x}}^H \mathbf{g}_{\bar{x},k}$ denote the received signal power, received intra-cell interference and received inter-cell interference, respectively.

For comparison we used the following (pre-normalization) precoders: $\mathbf{M}_x^{\text{MRT}} = \mathbf{H}_x$, $\mathbf{M}_x^{\text{ZF}} = \mathbf{H}_x (\mathbf{H}_x^H \mathbf{H}_x)^{-1}$, $\mathbf{M}_x^{\text{RZF}} = \mathbf{H}_x (\mathbf{H}_x^H \mathbf{H}_x + \frac{K}{N\rho_x} \mathbf{I})^{-1}$, where the regularization in $\mathbf{M}_x^{\text{RZF}}$ is chosen according to [16], [18]. The iaRZF weights have been chosen to be $\alpha = \beta = N\rho_x$ and $\gamma = 1$, hence simplifying comparison with RZF precoding. The corresponding performance graphs, obtained by extensive Monte-Carlo simulations, can be found in Figure 2.

We observe that iaRZF largely outperforms the other schemes. This is not surprising, as the non-cooperative schemes do not take information about the interfered UTs into account. What is surprising, however, is the gain in performance even for very bad channel estimates (see curve $\tau = 0.5$). Only for extremely bad CSI we observe that iaRZF wastes energy due to non-optimized choice of α, β . Thus, it performs worse than the other schemes, that do not take τ into account for precoding. This problem can easily be circumvented by choosing proper weights that let $\beta \rightarrow 0$ for $\tau \rightarrow 1$; as will be shown later on.

C. iaRZF for $\alpha_x, \beta_x \rightarrow \infty$

As has been briefly remarked by Hoydis et al. in [2], the iaRZF weights α_x and β_x should, intuitively, allow to project the transmitted signal to subspaces orthogonal to the UT_x 's ("own users") and $\text{UT}_{\bar{x}}$'s ("other users") channels, respectively. This behavior, in the limit cases of α_x or $\beta_x \rightarrow \infty$, is analyzed in this subsection.

1) *Finite Dimensional Analysis*: Limiting ourselves to finite dimensional methods and to the perfect CSI case ($\tau = 0$), we can already obtain the following insights.

First, we introduce the notation $\mathbf{P}_{\mathbf{X}}^\perp$ as a projection matrix on the space orthogonal to the column space of \mathbf{X} . Following the path outlined in Appendix B-A, one finds for the limit $\alpha_x \rightarrow \infty$ and assuming $\mathbf{H}_x^H \mathbf{H}_x$ invertible (true with probability 1):

$$\alpha_x \mathbf{M}_x \xrightarrow{\alpha_x \rightarrow \infty} \mathbf{H}_x (\mathbf{H}_x^H \mathbf{H}_x)^{-1} - \mathbf{P}_{\mathbf{H}_x}^\perp \mathbf{G}_x (\beta_x^{-1} \mathbf{I} + \mathbf{G}_x^H \mathbf{P}_{\mathbf{H}_x}^\perp \mathbf{G}_x)^{-1} \mathbf{G}_x^H \mathbf{H}_x (\mathbf{H}_x^H \mathbf{H}_x)^{-1} \quad (5)$$

Recall that the received signal at the UTs of BS_x in our simple model, due to (only) the intra-cell users, is given as³

$$\mathbf{y}_x^{\text{intra}} = \mathbf{H}_x^H \mathbf{F}_x \mathbf{s}_x \stackrel{\text{Lem. 2}}{=} k \mathbf{H}_x^H \mathbf{H}_x (\mathbf{H}_x^H \mathbf{H}_x)^{-1} \mathbf{s}_x = k \mathbf{s}_x$$

where the normalization leaves a scaling factor k that is independent of α_x . The Lemma 2 used here can be found in Appendix A. Thus, we see that for $\alpha_x \rightarrow \infty$ and β_x bounded, the precoder acts similar to a traditional ZF precoder. Thus, the intra-cell interference is completely suppressed in our system.

Looking at the limit $\beta_x \rightarrow \infty$, one arrives at

$$\mathbf{M}_x \xrightarrow{\beta_x \rightarrow \infty} \left[\mathbf{P}_{\mathbf{G}_x}^\perp - \mathbf{P}_{\mathbf{G}_x}^\perp \mathbf{H}_x (\alpha_x^{-1} \mathbf{I} + \mathbf{H}_x^H \mathbf{P}_{\mathbf{G}_x}^\perp \mathbf{H}_x)^{-1} \mathbf{H}_x^H \mathbf{P}_{\mathbf{G}_x}^\perp \right] \mathbf{H}_x.$$

One remembers that the received signal due to inter-cell interference in our simple model is given as

$$\mathbf{y}_x^{\text{inter}} = \mathbf{G}_{\bar{x}}^H \mathbf{F}_{\bar{x}} \mathbf{s}_{\bar{x}}$$

which from the above gives $\mathbf{y}_x^{\text{inter}} = 0$. Hence, using Lemma 2, we see that for $\beta_x \rightarrow \infty$ and α_x bounded, the induced inter-cell interference vanishes.

2) *Large-Scale Analysis*: We want to be able to study the impact of all system parameters on the average rate performance in more detail. Many insights on this matter are hidden by the inherent randomness of the SINRs. In order to find an expression of the sum rate that does not rely on random quantities, we anticipate results from Subsection III-E. There we find a deterministic limit to which the random values of SINR_x almost surely converge, when $N, K \rightarrow \infty$, assuming $0 < c < \infty$. This will also serve to motivate, how those later results are advantageous to intuitively and easily analyze more general system models pertaining to iaRZF formulations. We can adapt the results from Theorem 1 to fit our the current simplified model, by choosing $L = 2, K_x = K, N_x = N, \chi_x^x = 1, \chi_{\bar{x}}^x = \varepsilon, \tau_{\bar{x}}^x = \tau, \tau_x^x = 0, \alpha_x^x = \alpha_x, \alpha_{\bar{x}}^x = \beta_x, \gamma = 1, P_x = \rho_x$, for $x \in \{1, 2\}$. Doing so ultimately results in

³This includes the signal, as well as, the intra-cell interference part.

the following performance indicators $\text{Sig}_x \xrightarrow{\text{a.s.}} \overline{\text{Sig}}_x$ and $\text{Int}_x \xrightarrow{\text{a.s.}} \overline{\text{Int}}_x$, where

$$\begin{aligned} \overline{\text{Sig}}_x &= P_x \left(1 - \frac{c\alpha_x^2 e_x^2}{(1 + \alpha_x e_x)^2} - \frac{c\beta_x^2 \varepsilon^2 e_x^2}{(1 + \beta_x \varepsilon e_x)^2} \right) \\ \overline{\text{Int}}_x &= P_x c \underbrace{\frac{1}{(1 + \alpha_x e_x)^2}}_{\text{from BS } x} + P_{\bar{x}} c \varepsilon \underbrace{\frac{1 + 2\beta_{\bar{x}} \varepsilon \tau^2 e_{\bar{x}} + \beta_{\bar{x}}^2 \varepsilon^2 \tau^2 e_{\bar{x}}^2}{(1 + \beta_{\bar{x}} \varepsilon e_{\bar{x}})^2}}_{\text{from BS } \bar{x}} \end{aligned} \quad (6)$$

$$\Delta \overline{\text{Int}}_x^{\text{BS}x} + \overline{\text{Int}}_x^{\text{BS}\bar{x}}$$

$$e_x = \left(1 + \frac{c\alpha_x}{1 + \alpha_x e_x} + \frac{c\beta_x \varepsilon}{1 + \beta_x \varepsilon e_x} \right)^{-1} \quad (7)$$

where e_x is the unique non-negative solution to the fixed point equation (7). These expressions are precise in the large-scale regime ($N, K \rightarrow \infty, 0 < K/N < \infty$) and good approximations for finite dimensions. As a consequence of the continuous mapping theorem the above finally implies $\text{SINR}_x \xrightarrow{\text{a.s.}} \overline{\text{SINR}}_x = \overline{\text{Sig}}_x (\overline{\text{Int}}_x + 1)^{-1}$.

After realizing that $0 < \liminf e_x < \limsup e_x < \infty$ for $K, N \rightarrow \infty$ (see Lemma 6), the large-scale formulations give the insights we already obtained from the finite dimensional analysis (see previous subsection). Slightly simplifying (6) to reflect the perfect CSI case ($\tau = 0$), one obtains

$$\begin{aligned} \lim_{\alpha_x \rightarrow \infty} \overline{\text{Int}}_x^{\text{BS}x} &= \lim_{\alpha_x \rightarrow \infty} P_x c \frac{1}{(1 + \alpha_x e_x)^2} = 0 \\ \lim_{\beta_x \rightarrow \infty} \overline{\text{Int}}_x^{\text{BS}\bar{x}} &= \lim_{\beta_x \rightarrow \infty} P_{\bar{x}} c \frac{\varepsilon}{(1 + \beta_x \varepsilon e_x)^2} = 0 \end{aligned}$$

i.e., for $\alpha_x \rightarrow \infty$ the intra-cell interference vanishes and for $\beta_x \rightarrow \infty$ the induced inter-cell interference vanishes. Hence, at this point we have re-obtained the results from the previous subsection, which only used on finite dimensional techniques.

The large system formulation can also be used to judge the impact of the practically very important case of mis-estimation of the channels to the other cell's users. Remembering again $0 < \liminf e_x < \limsup e_x < \infty$ and (6) leads to

$$\begin{aligned} \lim_{\alpha_x \rightarrow \infty} P_x c \frac{1}{(1 + \alpha_x e_x)^2} &= 0 \\ \lim_{\beta_x \rightarrow \infty} P_{\bar{x}} c \frac{(\beta_x^{-2} + 2\varepsilon \tau^2 e_x \beta_x^{-1} + \varepsilon^2 \tau^2 e_x^2) \varepsilon}{(\beta_x^{-1} + \varepsilon e_x)^2} &= P_{\bar{x}} c \tau^2 \varepsilon \end{aligned}$$

i.e., for $\alpha_x \rightarrow \infty$ the intra-cell interference still vanishes, but for $\beta_x \rightarrow \infty$ the induced inter-cell interference converges to $P_{\bar{x}} c \tau^2 \varepsilon$. Hence we see that the induced inter-cell interference cannot be completely canceled any more, due to imperfect CSI. The impact of this is directly proportional to the transmit power, distance/gain, number of excessive antennas ($N - K$) and CSI quality obtained by the interfering BS.

3) *Large Scale Optimization*: One advantage of the large-scale approximation, is the possibility to find asymptotically optimal weights for the limit behavior of iaRZF. However, to keep the calculations within reasonable effort, one needs to limit the model to $P_1 = P_2 = P$. In this case the symmetry of the system entails $\alpha_1 = \alpha_2 = \alpha$ and $\beta_1 = \beta_2 = \beta$. Employing the steps from the previous subsection, we obtain a complete formulation for the large-scale approximation of

the (now equal) SINR values, when $\alpha \rightarrow \infty$. This is denoted $\overline{\text{SINR}}^{\alpha \rightarrow \infty} = \overline{\text{Sig}}^{\alpha \rightarrow \infty} \left(1 + \overline{\text{Int}}^{\alpha \rightarrow \infty}\right)^{-1}$ where

$$\begin{aligned} \overline{\text{Sig}}^{\alpha \rightarrow \infty} &= P \left(1 - c - \frac{c\beta^2 \varepsilon^2 e^2}{(1 + \beta \varepsilon e)^2}\right) \\ \overline{\text{Int}}^{\alpha \rightarrow \infty} &= P c \varepsilon \frac{1 + 2\beta \varepsilon \tau^2 e + \beta^2 \varepsilon^2 \tau^2 e^2}{(1 + \beta \varepsilon e)^2} \end{aligned}$$

and

$$e \stackrel{\Delta}{=} e^{\alpha \rightarrow \infty} = \left(1 + \frac{c}{e} + \frac{c\beta \varepsilon}{1 + \beta \varepsilon e}\right)^{-1}. \quad (8)$$

The optimal values of the weight β in limit case $\alpha \rightarrow \infty$ can be found by solving $\partial \overline{\text{SINR}}^{\alpha \rightarrow \infty} / \partial \beta = 0$. This leads (see Appendix B-C) to

$$\beta_{opt}^{\alpha \rightarrow \infty} = \frac{P(1 - \tau^2)}{P c \varepsilon \tau^2 + 1}. \quad (9)$$

This states, that in the perfect CSI case ($\tau = 0$), one should chose β equal to the transmit power of the BSs. It also shows how one should scale β in between the two obvious solutions, i.e., full weight on the interfering channel information for perfect CSI and no weight (disregard all information on the interfering channel) for random CSI ($\tau = 1$). We remark that the interference channel gain factor ε is also implicitly included in the precoder. Thus for $\varepsilon \rightarrow 0$, we have $\beta \|\mathbf{G}_x^H \mathbf{G}_x\| \rightarrow \mathbf{0}_K$, while β remains bounded. Hence no energy is wasted to precode for non-existent interference, as one would expect.

The same large-scale optimization can also be carried out for the limit of $\beta \rightarrow \infty$. The SINR optimal weight for α can be found as (similar to Appendix B-C)

$$\alpha_{opt}^{\beta \rightarrow \infty} = \frac{P}{P c \varepsilon \tau^2 + 1} = \frac{1}{c \varepsilon \tau^2 + 1/P}. \quad (10)$$

The result states, analog to the previous outcome, that in the perfect CSI case ($\tau = 0$), one should chose α equal to the transmit power of the BSs. However, unlike for $\beta_{opt}^{\alpha \rightarrow \infty}$, the implications for other limit-cases are not so clear. We see that increasing the transmit power also increases the weight α , up to the maximum value of $1/(c\varepsilon\tau^2)$. The weight reduces as the interference worsens, i.e., when τ^2 , c or ε grow. This makes sense, as the precoder would give more importance on the interfering channel (by indirectly increasing β via normalization).

Finally, we can easily calculate the SINR in the limit of both α and β independently tending to infinity:

$$\overline{\text{SINR}}^{\alpha, \beta \rightarrow \infty} = \frac{P(1 - 2c)}{P c \varepsilon \tau^2 + 1}.$$

The rationale behind all analyses in this section is, that optimal weights in the limit case often make for good heuristic approximations in more general cases. For instance, one can re-introduce the weights, found under the large-scale assumption, into the finite dimensional limit formulations. Particularly interesting for this approach is combining (9) with (5) to achieve a new structure, which could be considered a heuristic

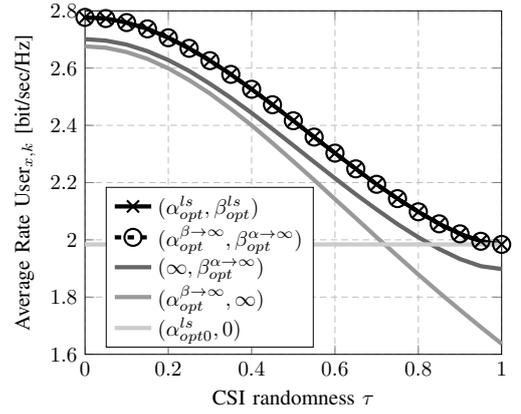


Fig. 3. Average user rate vs. CSI quality for adaptive precoder weights ($N = 160$, $K = 40$, $\varepsilon = 0.7$, $P = 10\text{dB}$).

iaZF precoder:

$$\begin{aligned} \mathbf{M}_x^{iaZF} &= \mathbf{H}_x (\mathbf{H}_x^H \mathbf{H}_x)^{-1} - \mathbf{P}_{\mathbf{H}_x}^\perp \mathbf{G}_x \\ &\times \left(\frac{P c \varepsilon \tau^2 + 1}{P(1 - \tau^2)} \mathbf{I} + \mathbf{G}_x^H \mathbf{P}_{\mathbf{H}_x}^\perp \mathbf{G}_x \right)^{-1} \mathbf{G}_x^H \mathbf{H}_x (\mathbf{H}_x^H \mathbf{H}_x)^{-1}. \end{aligned}$$

4) *Graphical Interpretation of the Results:* We will now proceed to show and compare the influence of the results from the previous subsection on the system performance of our simple model. Particularly interesting here are comparisons to numerically found, sum rate optimal weights.

In Figure 3 we analyze the average UT rate with respect to CSI randomness (τ), for different sets of precoder weights (α, β) , that (mostly) adapt to the available CSI quality. The values $(\alpha_{opt}^{ls}, \beta_{opt}^{ls})$ are obtained using 2D line search. Crucially, we see that the performance under $(\alpha_{opt}^{ls}, \beta_{opt}^{ls})$ and $(\alpha_{opt}^{\beta \rightarrow \infty}, \beta_{opt}^{\alpha \rightarrow \infty})$ is practically the same (the curves actually are the same within plotting precision).

The plot also contains the pair $(\alpha_{opt}^{ls}, 0)$, which corresponds to MMSE precoding. The weight α_{opt}^{ls} is again found by line search, hence we name the curve “optimal” (w.r.t sum rate) MMSE precoding. The performance is constant, as the precoder does not take the interfering channel (i.e., τ) into account. However, we see that the optimally weighted iaRZF reduced back to MMSE, when the channel estimation is purely random.

In Figure 4 we illustrate the effect of (sub-optimally, but conveniently) choosing a constant value for β . We set $\alpha = \alpha_{opt}^{\beta \rightarrow \infty}$ for all curves and also give the familiar $(\alpha_{opt}^{\beta \rightarrow \infty}, \beta_{opt}^{\alpha \rightarrow \infty})$ curve, as a benchmark. Furthermore, the actual value of β_{opt}^{ls} is given on a second axis to illustrate how one would need to adapt β for optimal average rate performance. Overall one observes that a constant value for β is (unsurprisingly) only acceptable for a limited region of the CSI quality spectrum. Small values of β fit well for large τ , middle values fit well for small τ . Overly large (or small) β s do not reach optimal performance in any region.

The encouraging performance of iaRZF using the optimal weights derived under limit assumptions, paired with the promise of simple and intuitive insights, provides motivation

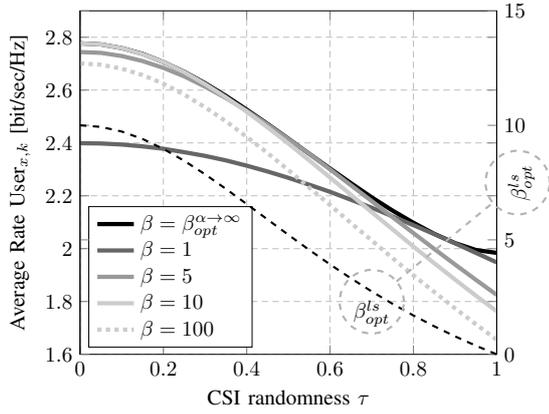


Fig. 4. Average user rate vs. CSI quality for constant precoder weights ($N = 160$, $K = 40$, $\varepsilon = 0.7$, $\alpha = \alpha_{opt}^{\beta \rightarrow \infty}$, $P = 10\text{dB}$).

for the next section, where we will apply the iaRZF scheme to a more general system.

III. GENERAL SYSTEM FOR IARZF ANALYSIS

A. System Model

In the following, we analyze cellular downlink multi-user MIMO systems of a more general type. We look at L cells, each consisting of one BS associated with a number of single antenna UTs. In more detail, the l th BS is equipped with N_l transmit antennas and serves K_l UTs. We generally set $N_l \geq K_l$ in order to avoid scheduling complications. We assume transmission on a single narrow-band carrier, full transmit buffers, and universal frequency reuse among the cells.

The l th BS transmits a data symbol vector $\mathbf{x}_l = [x_{l,1}, \dots, x_{l,K_l}]^T$ intended for its K_l uniquely associated UTs. This BS uses the linear precoding matrix $\mathbf{F}_l^l \in \mathbb{C}^{N_l \times K_l}$, where the columns $\mathbf{f}_{l,k}^l \in \mathbb{C}^{N_l}$ constitute the precoding vectors for each UT. We note that BSs do not directly interact with each other and users from other cells are explicitly not served. Thus, the received signal $y_{l,k} \in \mathbb{C}$ at the k th UT in cell l is

$$y_{l,k} = \sqrt{\chi_{l,k}^l} (\mathbf{h}_{l,k}^l)^H \mathbf{f}_{l,k}^l x_{l,k} + \sum_{k' \neq k} \sqrt{\chi_{l,k}^l} (\mathbf{h}_{l,k}^l)^H \mathbf{f}_{l,k'}^l x_{l,k'} + \sum_{m \neq l} \sqrt{\chi_{l,k}^m} (\mathbf{h}_{l,k}^m)^H \mathbf{F}_m^m \mathbf{x}_m + n_{l,k} \quad (11)$$

where $n_{l,k} \sim \mathcal{CN}(0,1)$ an additive noise term. The transmission symbols are chosen from a Gaussian codebook, i.e., $x_{l,k} \sim \mathcal{CN}(0,1)$. We assume block-wise small scale Rayleigh fading, thus the channel vectors are modeled as $\mathbf{h}_{l,k}^m \sim \mathcal{CN}(\mathbf{0}, \frac{1}{N_m} \mathbf{I}_{N_m})$. The path-loss and other large-scale fading effects are incorporated in the $\chi_{l,k}^m$ factors. The scaling factor $\frac{1}{N_m}$ in the fading variances is of technical nature and utilized in the asymptotic analysis. It can be canceled for a given arbitrarily sized system by modifying the transmission power accordingly; similar to Remark 1.

B. Imperfect Channel State Information

The UTs are assumed to perfectly estimate the respective channels to their serving BS, which enables coherent reception. This is reasonable, even for moderately fast traveling

users, if proper downlink reference signals are alternated with data symbols. Generally, downlink CSI can be obtained using either a time-division duplex protocol where the BS acquires channel knowledge from uplink pilot signaling [19] or a frequency-division duplex protocol, where temporal correlation is exploited as in [22]. In both cases, the transmitter usually has imperfect knowledge of the instantaneous channel realizations, e.g., due to imperfect pilot-based channel estimation, delays in the acquisition protocols, and user mobility. To model imperfect CSI without making explicit assumptions on the acquisition protocol, we employ the generic Gauss-Markov formulation (see e.g. [18], [23], [24]) and we define the estimated channel vectors $\hat{\mathbf{h}}_{l,k}^m \in \mathbb{C}^{N_m}$ to be

$$\hat{\mathbf{h}}_{l,k}^m = \sqrt{\chi_{l,k}^m} \left[\sqrt{(1 - (\tau_l^m)^2)} \mathbf{h}_{l,k}^m + \tau_l^m \tilde{\mathbf{h}}_{l,k}^m \right] \quad (12)$$

where $\tilde{\mathbf{h}}_{l,k}^m \sim \mathcal{CN}(0, \frac{1}{N_m} \mathbf{I}_{N_m})$ is the normalized independent estimation error. Using this formulation, we can set the accuracy of the channel acquisition between the UTs of cell l and the BS of cell m by selecting $\tau_l^m \in [0, 1]$; a small value for τ_l^m implies a good estimate. Furthermore, we remark that these choices imply $\hat{\mathbf{h}}_{l,k}^m \sim \mathcal{CN}(0, \chi_{l,k}^m \frac{1}{N_m} \mathbf{I}_{N_m})$. For convenience later on, we define the aggregated estimated channel matrices as $\hat{\mathbf{H}}_l^m = [\hat{\mathbf{h}}_{l,1}^m, \dots, \hat{\mathbf{h}}_{l,K_l}^m] \in \mathbb{C}^{N_m \times K_l}$.

C. iaRZF and Power Constraints

Following the promising results observed in Section II, we continue our analysis of the iaRZF precoding matrices \mathbf{F}_m^m , $m = 1, \dots, L$, introduced in (1). For some derivations, it will turn out to be useful to restate this precoder as

$$\mathbf{F}_m^m = \left(\alpha_m^m \hat{\mathbf{H}}_m^m (\hat{\mathbf{H}}_m^m)^H + \mathbf{Z}^m + \gamma_m \mathbf{I}_{N_m} \right)^{-1} \hat{\mathbf{H}}_m^m \nu_m^{\frac{1}{2}}$$

where $\mathbf{Z}^m = \sum_{l \neq m} \alpha_l^m \hat{\mathbf{H}}_l^m (\hat{\mathbf{H}}_l^m)^H$. The α_l^m can be considered as weights pertaining to the importance one wishes to attribute to the respective estimated channel. We remark, that the regularization parameter γ_m is usually chosen to be the number of users over the total transmit power [16] in classical RZF. The factors ν_m are used to fulfill the average per UT transmit power constraint P_m^4 , pertaining to BS m :

$$\frac{1}{K_m} \text{tr} [\mathbf{F}_m^m (\mathbf{F}_m^m)^H] = P_m. \quad (13)$$

D. Performance Measure

Most performance measures in cellular systems are functions of the SINRs at each UT; e.g., (weighted) sum rate and outage probability. Under the treated system model, the received signal power (in expectation to the transmitted symbols $x_{l,k}^{(l)}$) at the k th UT of cell l , i.e., UT $_{l,k}$, is

$$\text{Sig}_{l,k}^{(l)} = \chi_{l,k}^l (\mathbf{h}_{l,k}^l)^H \mathbf{f}_{l,k}^l (\mathbf{f}_{l,k}^l)^H \mathbf{h}_{l,k}^l \quad (14)$$

⁴We remark that choosing P_m of order 1 will assure proper scaling of all terms of the SINR in the following (see (17)).

Similarly, the interference power is

$$\text{Int}_{l,k}^{(l)} = \sum_{m \neq l} \chi_{l,k}^m (\mathbf{h}_{l,k}^m)^H \mathbf{F}_m^m (\mathbf{F}_m^m)^H \mathbf{h}_{l,k}^m + \chi_{l,k}^l (\mathbf{h}_{l,k}^l)^H \mathbf{F}_{l[k]}^l (\mathbf{F}_{l[k]}^l)^H \mathbf{h}_{l,k}^l \quad (15)$$

where

$$\mathbf{F}_{l[k]}^l = \left(\alpha_l^l \hat{\mathbf{H}}_l^l (\hat{\mathbf{H}}_l^l)^H + \mathbf{Z}^l + \gamma_l \mathbf{I}_{N_l} \right)^{-1} \hat{\mathbf{H}}_{l[k]}^l \nu_l^{\frac{1}{2}} \quad (16)$$

and $\hat{\mathbf{H}}_{l[k]}^l$ is $\hat{\mathbf{H}}_l^l$ with its k th column removed. Hence, the SINR at UT $_{l,k}$ can be expressed as

$$\text{SINR}_{l,k} = \text{Sig}_{l,k}^{(l)} (\text{Int}_{l,k} + 1)^{-1}. \quad (17)$$

In the following, we focus on the sum rate, which is a commonly used performance measure utilizing the SINR values and straightforward to interpret. Under the assumption that interference is treated as noise, the sum rate expressed as

$$R_{\text{sum}} = \sum_{l,k} R_{l,k} = \sum_{l,k} \log(1 + \text{SINR}_{l,k})$$

where SINRs are random quantities defined by the system model. This randomness obscures the influence of the system parameters on sum rate performance.

E. Deterministic Equivalent of the SINR

In order to obtain tractable and insightful expressions of the system performance, we propose a large scale approximation. This allows us to state the sum rate expression in a deterministic and compact form that can readily be interpreted and optimized. Also, the large system approximations are accurate in both massive MIMO systems and conventional small-scale MIMO of tractable size, as will be evidenced later via simulations (see Subsection IV-B). In certain special cases, optimizations of such approximations w.r.t. many performance measures, can be carried out analytically (see for example [18]). In almost all cases, optimizations can be done numerically. We will derive a deterministic equivalent (DE) of the SINR values that allows for a large scale approximation of the sum rate expression in (17). DEs are preferable to standard limit calculations, as they are precise in the limit case, are also defined for finite dimensions and provably approach the random quantity for increasing dimensions (see, e.g., [25] and [26] for more information). The DE is based on the following technical assumption. Introducing the ratio $c_i = K_i/N_i$, we make the following assumption.

Assumption A-1. $N_i, K_i \rightarrow \infty$, such that for all i we have

$$0 < \liminf c_i \leq \limsup c_i < \infty.$$

This asymptotic regime is denoted $N \rightarrow \infty$ for brevity.

Thus, we require for N_i and K_i to grow large at the same speed. By extending the analytical approach in [18] and [19] to the SINR expression in (17), we obtain a DE of the SINR, which is denoted $\overline{\text{SINR}}_{l,k}$ in the following.

Theorem 1 (Deterministic Equivalent of the SINR). *Under A-1, we have*

$$\text{SINR}_{l,k} - \overline{\text{SINR}}_{l,k} \xrightarrow[N \rightarrow \infty]{\text{a.s.}} 0.$$

Here $\overline{\text{SINR}}_{l,k} = \overline{\text{Sig}}_{l,k}^{(l)} (\overline{\text{Int}}_{l,k} + 1)^{-1}$ with

$$\begin{aligned} \overline{\text{Sig}}_{l,k}^{(l)} &= \bar{\nu}_l (\chi_{l,k}^l)^2 e_{(l)}^2 (1 - (\tau_l^l)^2) (y_{l,k}^l)^2 \\ \overline{\text{Int}}_{l,k} &= \sum_{m=1}^L \bar{\nu}_m \left(1 + 2x_{l,k}^m e_{(m)} + \alpha_l^m \chi_{l,k}^m x_{l,k}^m e_{(m)}^2 \right) \chi_{l,k}^m g_{(m)} (y_{l,k}^m)^2 \end{aligned}$$

given $x_{l,k}^m = \alpha_l^m \chi_{l,k}^m (\tau_l^m)^2$. The parameter $\bar{\nu}_m$, the abbreviations $g_{(m)}$ and $y_{l,k}^m$, as well as the corresponding fixed-point equation $e_{(m)}$ and $e'_{(m)}$ are given in the following.

First, we define $e_{(m)}$ to be the unique positive solution of the fixed-point equation

$$e_{(m)} = \left(\gamma_m + \frac{1}{N_m} \sum_{j=1}^{K_m} \alpha_l^m \chi_{m,j}^m y_{m,j}^m + \frac{1}{N_m} \sum_{l \neq m} \sum_{k=1}^{K_l} \alpha_l^m \chi_{l,k}^m y_{l,k}^m \right)^{-1} \quad (18)$$

where $y_{l,k}^m = \left(1 + \alpha_l^m \chi_{l,k}^m e_{(m)} \right)^{-1}$. We also have $\bar{\nu}_m = P_m K_m / (N_m g_{(m)})$ with

$$g_{(m)} = -\frac{1}{N_m} \sum_{j=1}^{K_m} \chi_{m,j}^m e'_{(m)} (y_{m,k}^m)^2$$

and $e'_{(m)}$ can be found directly, once $e_{(m)}$ is known:

$$e'_{(m)} = \left[\frac{1}{N_m} \sum_{j=1}^{K_m} (\alpha_l^m)^2 (\chi_{m,j}^m)^2 (y_{m,j}^m)^2 + \frac{1}{N_m} \sum_{l \neq m} \sum_{k=1}^{K_l} (\alpha_l^m)^2 (\chi_{l,k}^m)^2 (y_{l,k}^m)^2 - e_{(m)}^{-2} \right]^{-1}. \quad (19)$$

Proof: See Appendix C. \blacksquare

By employing dominated convergence arguments and the continuous mapping theorem (e.g., [26]), we see that Theorem 1 implies, for each UT (l, k) ,

$$R_{l,k} - \log_2(1 + \overline{\text{SINR}}_{l,k}) \xrightarrow[N \rightarrow \infty]{\text{a.s.}} 0. \quad (20)$$

These results have already been used in Section II and will also serve as the basis in the following.

IV. NUMERICAL RESULTS

In this section we will, first, introduce a heuristic generalization of the previously found (see Subsection II-C3) ‘‘limit-optimal’’ iaRZF precoder weights. Furthermore, we provide simulations that corroborate the viability of the proposed precoder, even in systems that are substantially different to the idealized system used in Section II.

A. Heuristic Generalization of Optimal Weights

Subsection II-C3 resulted in some optimal iaRZF precoder weights for the case of 2 BSs and under various assumptions, most prominently that the respective other weight is infinitely large. We have already observed in Subsection II-C4 that these precoder weights, also offer virtually optimal performance, when they are applied in the non-limit weight case. Now it is

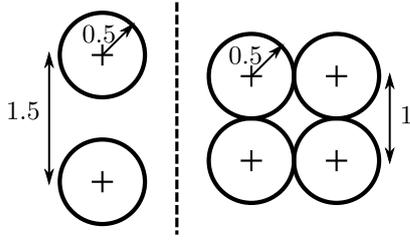


Fig. 5. Geometry 2 BS and 4 BS Downlink Models.

natural to go one step further and to intuitively generalize the heuristic weights to systems with arbitrary many BSs, transmit powers, CSI randomness and user/antenna ratios. Following the insights and the structures discovered before (see (9) and (10)), we define the general heuristic precoder weights as

$$\tilde{\alpha}_b^a = \frac{P_a(1 - (\tau_b^a)^2)}{P_b c_b \varepsilon_b^a (\tau_b^a)^2 + 1}. \quad (21)$$

Here we introduced the new notation ε_b^a , which we take to be the average gain factor between BS a and the UTs of cell b . Thus, it can be calculated as $\varepsilon_b^a = \frac{1}{K_b} \sum_k \chi_{b,k}^a$. One can intuitively understand (21) by remembering that α_b^a should be proportional to the “importance” of the associated channels (from BS a to UTs b). Hence, the numerator places more importance on BSs with more transmit power and less importance on badly estimated channels. The denominator deemphasizes the importance of induced interference, when the receiving cell features large transmit power and many “extra” antennas. Also, bad channel estimates reduce importance again; analogously to the numerator. The intuitive reason for having ε_b^a in the denominator becomes clear once one realizes that the estimated channels in our model are not normalized (see (12)). Thus, the approximate effective weight of the precoder with respect to a normalized channel is $w_b^a = \tilde{\alpha}_b^a \varepsilon_b^a$. Hence, for $\varepsilon_b^a \rightarrow 0$, we have $w_b^a \rightarrow 0$, i.e., no importance is placed on very weak channels. Using the same deliberation, we notice that for $\varepsilon_b^a \rightarrow \infty$ we have w_b^a tending to some constant value and for $\tau_b^a \rightarrow 0$ we have $w_b^a \rightarrow P_a \varepsilon_b^a$. Especially the last observation is important in order to see why no energy is wasted on far away interferers/weak channels, even if one has perfect CSI of those channels.

B. Performance

In order to verify the heuristic approach, we introduce two models (see Figure 5). In the first one, two BSs are distanced 1.5 units, have a height of 0.1 units and use 160 antennas each. Around each BS, 40 single antenna UTs of height 0, are randomly (uniformly) distributed within a radius of 1 unit. Hence, one obtains clear non-overlapping clusters that are closely related to the Wyner-like simplified model in Section II. The pathloss between each BS and all UTs is defined as the inverse of the distance to the power of 2.8. The quality of CSI estimation between a BS and its associated UTs is defined by $\tau_1^1 = \tau_2^2 = \tau_a$ and inter-cell wise by $\tau_2^1 = \tau_1^2 = \tau_b$. Due to the symmetry we can assume that

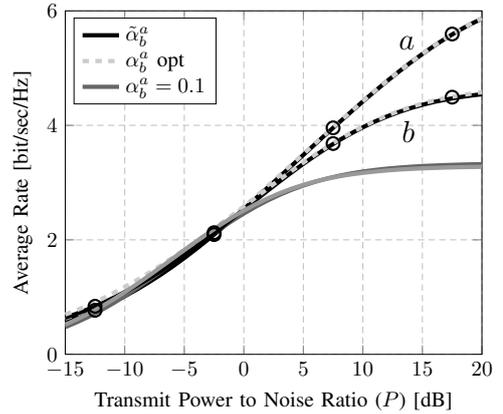


Fig. 6. 2 BSs: Average rate vs. transmit power to noise ratio ($N_x = 160$, $K_x = 40$, $P_X = P$, $(\tau_a, \tau_b) \in \{(0, 0.4), (0.1, 0.5)\}$, i.e., case a and b).

the chosen channel weighting pertaining to intra-cell channels are the same for both BSs and will be denoted $\alpha_1^1 = \alpha_2^2 = \alpha$. Similarly, the inter-cell weights will be denoted $\alpha_1^2 = \alpha_2^1 = \beta$. The transmit power to noise ratio (per UT) at each BS is taken equal, i.e., $P_1 = P_2 = P$. For this system we obtain the average UT rate performance, shown in Figure 6. The markers denote results of Monte-Carlo simulations that randomize over UT placement scenarios and channel realizations, when the precoding weights are chosen as in (21). The main point of this graph is to compare the performance under heuristic weights and numerically optimal weights, found via 2D line search. We observe that the performance of both approaches is virtually the same. Furthermore, one sees that constant weights exhibit the same problems as in Section II. Interesting is also the observation that, when one diverges prominently from the simple system ($\tau_a = 0$), by choosing $\tau_a = 0.1$, the heuristic weights still perform practically the same as exhaustive numerical optimization.

Finally, we look a more complex system of 4 BSs (see Figure 5). The BSs, of height 0.1 units, are placed on the corners of a square with edge length 1 units. The UTs are of height 0 and are distributed uniformly in a disc of radius 0.5 units around the corresponding BS. The pathloss is calculated as the inverse of distance to the power of 2.8. Figure 7 shows the performance of the 4 BS system, assuming that each BS has 160 antennas with a power constraint of P per UT and serves 40 UTs. We assume that the CSI randomness is overwhelmingly determined by inter-BS distance, i.e., we have τ_a for each BS to the adherent UTs, τ_b for each BS to UTs of BSs 1 unit away and τ_c for each BS to UTs of BSs $\sqrt{2}$ units away. It is, thus, reasonable to chose $\tau_a < \tau_b < \tau_c$. In the graph we compare the heuristic weights with various other weighting approaches. Round markers stem from a Monte-Carlo simulation of the performance pertaining to the heuristic weights, in order to confirm the applicability of our DEs. The benchmark “numeric” result in this figure is obtained from optimizing the 8 precoder weights via extensive numerical search, using $\tilde{\alpha}_b^a$ as a starting point. The observed performance is always better than the heuristic approach, which is not surprising, as the randomly positioned and non-clustered structure of UTs is

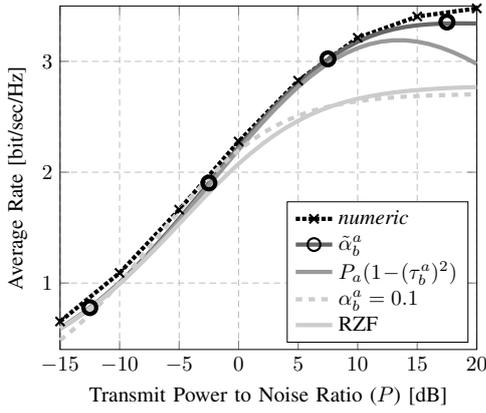


Fig. 7. 4 BSs: Average rate vs. transmit power to noise ratio ($N_x = 160$, $K_x = 40$, $P_x = P$, $(\tau_a, \tau_b, \tau_c) = (0.1, 0.3, 0.4)$).

taking the scenario very far away from the original simplified system of Section II. More interesting is the performance of taking $\alpha_b^a = P_a(1 - (\tau_b^a)^2)$. This configuration conforms to not taking any interference into account, i.e., $\varepsilon_b^a = 0$. We observe that most of the gains of the heuristic method come from this part; only at very high powers, where interference is the dominant problem, the $P_a(1 - (\tau_b^a)^2)$ approach is noticeably suboptimal. Similarly, choosing $\alpha_b^a = (1 - (\tau_b^a)^2)$ performs well at middle and high transmit SNR, but loses efficacy at low SNR. The constant weight approach behaves like in Section II, in that it is only a good match for a limited part of the curve. However, given the “mis-matched” general scenario, we see that it can also outperform the heuristic weights. For comparison purposes, we also compare with standard non-cooperative RZF, as defined in Subsection II-B.

In general, employing $\tilde{\alpha}_b^a$ is most advantageous in high interference scenarios, as would be expected due to the “interference aware” conception of the precoder. The figure generally implies that the heuristic approach is close to the numerical optimum, however we can not be sure that numeric optimization finds the true optimum. Carrying out the same simulations for different levels of CSI randomness, one observes that the gain of using the heuristic variant of iaRZF is substantial as long as the estimations of the interfering channels are not too bad. For extremely bad CSI, standard non-cooperative RZF can outperform iaRZF with $\tilde{\alpha}_b^a$. We also note that better CSI widens the gap between the $\tilde{\alpha}_b^a$ and $\alpha_b^a = P_a(1 - (\tau_b^a)^2)$ weighted iaRZF versions.

V. CONCLUSION

In this paper, we analyzed a variant of the generally optimal linear precoding structure for multi-cell system, denoted iaRZF. It was shown that the relegation of interference into orthogonal subspaces by iaRZF can be explained rigorously and intuitively, even without assuming large scale systems. For example, one can indeed observe that the precoder can either completely get rid of inter-cell or intra-cell interference (assuming perfect channel knowledge).

Stating and proving new results from large-scale random matrix theory, allowed us to give more conclusive and intuitive

insights into the behavior of the precoder, especially with respect to imperfect CSI knowledge and induced interference mitigation. The effectiveness of these large-scale results has been demonstrated in practical finite dimensional systems. Most importantly, we concluded that iaRZF can use all available (also very bad) interference channel knowledge to obtain significant performance gains, while not requiring explicit inter base station cooperation.

Moreover, it is possible to analytically optimize the iaRZF precoder weights in certain limit scenarios using our large-scale results. Insights from this were used to propose a heuristic generalization of the limit optimal iaRZF weighting for arbitrary systems. The efficacy of the heuristic iaRZF approach has been demonstrated by achieving a sum-rate close to the numerically optimally weighted iaRZF, for a wide range of general and practical systems. The effectiveness of our heuristic approach has been intuitively explained by mainly balancing the importance of available knowledge about various channel and system variables.

APPENDIX A

USEFUL NOTATION AND LEMMAS

Lemma 1 (Common Matrix Identities). *Let \mathbf{A} , \mathbf{B} be complex invertible matrices and \mathbf{C} a rectangular complex matrix, all of proper size. We restate the following, well known, relationships:*

Woodbury Identity:

$$(\mathbf{A} + \mathbf{C}\mathbf{B}\mathbf{C}^H)^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{C}(\mathbf{B}^{-1} + \mathbf{C}^H\mathbf{A}^{-1}\mathbf{C})^{-1}\mathbf{C}^H\mathbf{A}^{-1}. \quad (22)$$

Searl Identity:

$$(\mathbf{I} + \mathbf{A}\mathbf{B})^{-1}\mathbf{A} = \mathbf{A}(\mathbf{I} + \mathbf{B}\mathbf{A})^{-1}. \quad (23)$$

Resolvent Identity:

$$\mathbf{A}^{-1} + \mathbf{B}^{-1} = -\mathbf{A}^{-1}(\mathbf{A} - \mathbf{B})\mathbf{B}^{-1}. \quad (24)$$

Lemma 2 (Unitary Projection Matrices). *Let \mathbf{X} be an $N \times K$ complex matrix, where $N \geq K$ and $\text{rank}(\mathbf{X}) = K$. We define $\mathbf{P}_\mathbf{X} = \mathbf{X}(\mathbf{X}^H\mathbf{X})^{-1}\mathbf{X}^H$ and $\mathbf{P}_\mathbf{X}^\perp = \mathbf{I} - \mathbf{P}_\mathbf{X}$. It follows (see e.g., [27, Chapter 5.13])*

$$\begin{aligned} \mathbf{P} &= \mathbf{P}^2 \Leftrightarrow \mathbf{P} = \mathbf{P}^H \\ \mathbf{P}_\mathbf{X}^\perp \mathbf{X} &= \mathbf{0} \Leftrightarrow \mathbf{X}^H \mathbf{P}_\mathbf{X}^\perp = \mathbf{0}. \end{aligned}$$

Generally one denotes $\mathbf{P}_\mathbf{X}$ as the projection matrix onto the column space of \mathbf{X} and $\mathbf{P}_\mathbf{X}^\perp$ as the projection matrix onto the orthogonal space of the column space of \mathbf{X} .

Definition 1 (Notation of Resolvents). *Given the notations from Section III, we define the resolvent matrix of $\hat{\mathbf{H}}_a^a$ as*

$$\mathbf{Q}_a \triangleq \left(\alpha_a^a \hat{\mathbf{H}}_a^a (\hat{\mathbf{H}}_a^a)^H + \mathbf{Z}^a + \gamma_a \mathbf{I}_{N_a} \right)^{-1}$$

and we will also make use of the following modified versions

$$\begin{aligned} \mathbf{Q}_{a[bc]} &\triangleq \left(\alpha_a^a \hat{\mathbf{H}}_a^a (\hat{\mathbf{H}}_a^a)^H + \mathbf{Z}^a - \alpha_b^a \hat{\mathbf{h}}_{b,c}^a (\hat{\mathbf{h}}_{b,c}^a)^H + \gamma_a \mathbf{I}_{N_a} \right)^{-1} \\ \mathbf{Q}_{a[b]} &\triangleq \left(\alpha_a^a \hat{\mathbf{H}}_a^a (\hat{\mathbf{H}}_a^a)^H + \mathbf{Z}^a - \alpha_a^a \hat{\mathbf{h}}_{a,b}^a (\hat{\mathbf{h}}_{a,b}^a)^H + \gamma_a \mathbf{I}_{N_a} \right)^{-1} \\ &= \left(\alpha_a^a \hat{\mathbf{H}}_{a[b]}^a (\hat{\mathbf{H}}_{a[b]}^a)^H + \mathbf{Z}^a + \gamma_a \mathbf{I}_{N_a} \right)^{-1}. \end{aligned}$$

Lemma 3 (Matrix Inversion Lemma [28, Lemma 2.2]). *Let \mathbf{A} be an $M \times M$ invertible matrix and $\mathbf{x} \in \mathbb{C}^M$, $c \in \mathbb{C}$ for which $\mathbf{A} + c\mathbf{x}\mathbf{x}^H$ is invertible. Then, as an application of (22), we have*

$$\mathbf{x}^H (\mathbf{A} + c\mathbf{x}\mathbf{x}^H)^{-1} = \frac{\mathbf{x}^H \mathbf{A}^{-1}}{1 + c\mathbf{x}^H \mathbf{A}^{-1} \mathbf{x}}.$$

For the previously defined resolvent matrices, we have in particular

$$\mathbf{Q}_a \hat{\mathbf{h}}_{a,b}^a = \frac{\mathbf{Q}_{a[b]} \hat{\mathbf{h}}_{a,b}^a}{1 + \alpha_a^a (\hat{\mathbf{h}}_{a,b}^a)^H \mathbf{Q}_{a[b]} \hat{\mathbf{h}}_{a,b}^a}.$$

Lemma 4 (Convergence of Quadratic Forms [29]). *Let $\mathbf{x}_M = [X_1, \dots, X_M]^T$ be an $M \times 1$ vector where the X_n are i.i.d. Gaussian complex random variables with unit variance. Let \mathbf{A}_M be an $M \times M$ matrix independent of \mathbf{x}_M . If in addition $\limsup_M \|\mathbf{A}\|_2 < \infty$ then we have that*

$$\frac{1}{M} \mathbf{x}^H \mathbf{A}_M \mathbf{x} - \frac{1}{M} \text{tr}(\mathbf{A}_M) \xrightarrow[M \rightarrow +\infty]{a.s.} 0.$$

Corollary 1. *Let \mathbf{A}_M be as in Lemma 4, i.e., $\limsup_M \|\mathbf{A}\|_2 < \infty$, and $\mathbf{x}_M, \mathbf{y}_M$ be random, mutually independent with complex Gaussian entries of zero mean and variance 1. Then, for any $p \geq 2$ we have*

$$\frac{1}{M} \mathbf{y}_M^H \mathbf{A}_M \mathbf{x}_M \xrightarrow[M, K \rightarrow +\infty]{a.s.} 0.$$

Lemma 5. [Rank-One Perturbation Lemma [26, Lemma 14.3]] *Let \mathbf{Q}_a and $\mathbf{Q}_{a[b]}$ be the resolvent matrices as defined in Definition 1. Then, for any matrix \mathbf{A} we have:*

$$\text{tr} [\mathbf{A} (\mathbf{Q}_a - \mathbf{Q}_{a[b]})] \leq \frac{1}{\gamma_a} \|\mathbf{A}\|_2.$$

APPENDIX B

SIMPLE SYSTEM LIMIT BEHAVIOR PROOFS

A. Finite Dimensions

In order to simplify the notation we will not explicitly state the index x in the following, unless needed, hence the normalized precoder \mathbf{F} for each of the two cells is $\mathbf{F} = \sqrt{KM}/\sqrt{\text{tr} \mathbf{M}} \mathbf{M}^H$ for $\mathbf{M} = (\alpha \mathbf{H}\mathbf{H}^H + \beta \mathbf{G}\mathbf{G}^H + \gamma \mathbf{I})^{-1} \mathbf{H}$.

1) $\beta \rightarrow \infty$: For the limit when $\beta \rightarrow \infty$ we use (22) with $\mathbf{A} = \beta \mathbf{G}\mathbf{G}^H + \gamma \mathbf{I}$ and $\mathbf{C}\mathbf{B}\mathbf{C}^H = \mathbf{H}\alpha \mathbf{I}\mathbf{H}^H$ to reformulate the matrix \mathbf{M}

$$\begin{aligned} \mathbf{M} &= (\alpha \mathbf{H}\mathbf{H}^H + \beta \mathbf{G}\mathbf{G}^H + \gamma \mathbf{I})^{-1} \mathbf{H} \\ &= \left[\mathbf{Q}_G - \mathbf{Q}_G \mathbf{H} (\alpha^{-1} \mathbf{I} + \mathbf{H}^H \mathbf{Q}_G \mathbf{H})^{-1} \mathbf{H}^H \mathbf{Q}_G \right] \mathbf{H} \end{aligned}$$

where

$$\begin{aligned} \mathbf{Q}_G &= (\beta \mathbf{G}\mathbf{G}^H + \gamma \mathbf{I})^{-1} \\ &\stackrel{(22)}{=} \gamma^{-1} \mathbf{I} - \gamma^{-1} \mathbf{G} \left(\frac{\gamma}{\beta} \mathbf{I} + \mathbf{G}^H \mathbf{G} \right)^{-1} \mathbf{G}^H. \end{aligned}$$

We now let $\beta \rightarrow \infty$, assuming $\mathbf{G}^H \mathbf{G}$ is invertible (which true with probability 1) and γ bounded. In this regime, we remember Lemma 2, and rewrite $\mathbf{Q}_G = \gamma^{-1} \mathbf{P}_G^\perp$. One finally arrives at

$$\mathbf{M} \xrightarrow{\beta \rightarrow \infty} \left[\gamma^{-1} \mathbf{P}_G^\perp - \gamma^{-2} \mathbf{P}_G^\perp \mathbf{H} (\alpha^{-1} \mathbf{I} + \gamma^{-1} \mathbf{H}^H \mathbf{P}_G^\perp \mathbf{H})^{-1} \mathbf{H}^H \mathbf{P}_G^\perp \right] \mathbf{H}.$$

2) $\alpha \rightarrow \infty$: Introducing the abbreviations $\mathbf{Q}_H = (\mathbf{H}\mathbf{H}^H + \frac{\gamma}{\alpha} \mathbf{I})^{-1}$ and $\bar{\mathbf{Q}}_H = (\mathbf{H}^H \mathbf{H} + \frac{\gamma}{\alpha} \mathbf{I})^{-1}$, we can rewrite the matrix \mathbf{M} as follows.

$$\begin{aligned} \alpha \mathbf{M} &= \left(\mathbf{H}\mathbf{H}^H + \frac{\beta}{\alpha} \mathbf{G}\mathbf{G}^H + \frac{\gamma}{\alpha} \mathbf{I} \right)^{-1} \mathbf{H} \\ &\stackrel{(22)}{=} \left[\mathbf{Q}_H - \mathbf{Q}_H \mathbf{G} \left(\frac{\alpha}{\beta} \mathbf{I} + \mathbf{G}^H \mathbf{Q}_H \mathbf{G} \right)^{-1} \mathbf{G}^H \mathbf{Q}_H \right] \mathbf{H} \\ &\stackrel{(23)}{=} \mathbf{H} \bar{\mathbf{Q}}_H - \mathbf{Q}_H \mathbf{G} \left(\frac{\alpha}{\beta} \mathbf{I} + \mathbf{G}^H \mathbf{Q}_H \mathbf{G} \right)^{-1} \mathbf{G}^H \mathbf{H} \bar{\mathbf{Q}}_H. \end{aligned}$$

Applying (24) to the expression $(\mathbf{H}\mathbf{H}^H + \frac{\gamma}{\alpha} \mathbf{I})^{-1} + (-\frac{\gamma}{\alpha} \mathbf{I})^{-1}$, one eventually finds the relationship $\mathbf{Q}_H = \alpha \gamma^{-1} (\mathbf{I} - \mathbf{H} \mathbf{Q}_H \mathbf{H}^H)$. Hence,

$$\begin{aligned} \alpha \mathbf{M} &= \mathbf{H} \bar{\mathbf{Q}}_H - \gamma^{-1} (\mathbf{I} - \mathbf{H} \bar{\mathbf{Q}}_H \mathbf{H}^H) \\ &\mathbf{G} \left[\frac{1}{\beta} \mathbf{I} + \gamma^{-1} \mathbf{G}^H (\mathbf{I} - \mathbf{H} \bar{\mathbf{Q}}_H \mathbf{H}^H) \mathbf{G} \right]^{-1} \mathbf{G}^H \mathbf{H} \bar{\mathbf{Q}}_H. \end{aligned}$$

Now, taking the limit of $\alpha \rightarrow \infty$, assuming $\mathbf{H}^H \mathbf{H}$ invertible (true with probability 1), and recognizing $\mathbf{P}_H^\perp = \mathbf{I} - \mathbf{H} (\mathbf{H}^H \mathbf{H})^{-1} \mathbf{H}^H$ we arrive at

$$\begin{aligned} \alpha \mathbf{M} &\xrightarrow{\alpha \rightarrow \infty} \mathbf{H} (\mathbf{H}^H \mathbf{H})^{-1} - \gamma^{-1} \left[\mathbf{I} - \mathbf{H} (\mathbf{H}^H \mathbf{H})^{-1} \mathbf{H}^H \right] \mathbf{G} \\ &\quad \left\{ \beta^{-1} \mathbf{I} + \gamma^{-1} \mathbf{G}^H \left[\mathbf{I} - \mathbf{H} (\mathbf{H}^H \mathbf{H})^{-1} \mathbf{H}^H \right] \mathbf{G} \right\}^{-1} \\ &\quad \mathbf{G}^H \mathbf{H} (\mathbf{H}^H \mathbf{H})^{-1} \\ &= \mathbf{H} (\mathbf{H}^H \mathbf{H})^{-1} \\ &\quad - \gamma^{-1} \mathbf{P}_H^\perp \mathbf{G} \left\{ \beta^{-1} \mathbf{I} + \gamma^{-1} \mathbf{G}^H \mathbf{P}_H^\perp \mathbf{G} \right\}^{-1} \mathbf{G}^H \mathbf{H} (\mathbf{H}^H \mathbf{H})^{-1}. \end{aligned}$$

B. Large-Scale Approximation

We remind ourselves, that for perfect and imperfect CSI the resulting fixed point equations are equivalent:

$$e = \left(1 + \frac{c}{\alpha^{-1} + e} + \frac{c\varepsilon}{\beta^{-1} + \varepsilon e} \right)^{-1}. \quad (25)$$

Lemma 6 (e is Bounded). *For either $\alpha \rightarrow \infty$ and β, ε bounded or $\beta \rightarrow \infty$ and α, ε bounded, we have*

$$0 < \liminf e < \limsup e < \infty.$$

Proof: 1) $e < \infty$ when α or $\beta \rightarrow \infty$.

This follows immediately from contradiction, when one takes $e \rightarrow \infty$ in (25).

2) e positive when α or $\beta \rightarrow \infty$.

We take either $\alpha \rightarrow \infty$ and β, ε bounded or $\beta \rightarrow \infty$ and α, ε bounded. For the case $\alpha \rightarrow \infty$, we first denote $\gamma = \alpha e$. Now we assume γ to be bounded for $\alpha \rightarrow \infty$

$$\gamma = \lim_{\alpha \rightarrow \infty} \left(\frac{1}{\alpha} + \frac{c}{1 + \gamma} + \frac{c\beta\varepsilon}{\alpha + \beta\varepsilon\gamma} \right)^{-1} = \left(\frac{c}{1 + \gamma} \right)^{-1}$$

thus implying $\gamma = \frac{1}{c-1} < 0$, as $c < 1$. Case 1 directly contradicts the assumption and case 2 is contradicting, as e can not be negative for positive values of α, β, c and ε . Thus, γ is not bounded for $\alpha \rightarrow \infty$, hence e can neither be zero nor negative. For the case of $\beta \rightarrow \infty$, we denote $\gamma = \beta e$ and proceed analogously. ■

C. Large-Scale Optimization $\alpha \rightarrow \infty$

Continuing from Appendix B-B, we see that in the limit $\alpha \rightarrow \infty$ the large-scale approximation of the SINR values, pertaining to the users of each cell, i.e., $\overline{\text{SINR}}^{\alpha \rightarrow \infty}$, is indeed as stated in Subsection II-C3.

Differentiating $\overline{\text{SINR}}^{\alpha \rightarrow \infty}$ w.r.t. β , while taking into account that e is an abbreviation for $e^{\alpha \rightarrow \infty}(\beta)$ leads us to

$$\frac{\partial \overline{\text{SINR}}^{\alpha \rightarrow \infty}}{\partial \beta} = -2Pc\varepsilon^2 [e + \beta e'] \quad (26)$$

$$\times \frac{t_1}{[P(c\beta^2 e^2 \varepsilon^3 \tau^2 + 2c\beta e \varepsilon^2 \tau^2 + c\varepsilon) + \beta^2 e^2 \varepsilon^2 + 2\beta e \varepsilon + 1]^2}$$

where we used e' as shorthand for $\frac{\partial e^{\alpha \rightarrow \infty}(\beta)}{\partial \beta}$ and

$$t_1 = P[c - 1 - \beta \varepsilon e + 2\beta c \varepsilon e] + \beta e + \beta^2 e^2 \varepsilon - P_{\bar{x}} \tau^2 [c - 1 - \beta \varepsilon e + \beta c \varepsilon e - \beta^2 c e^2 \varepsilon^2]$$

Realizing that the denominator of (26) can not become zero, we have two possible solutions for $\frac{\partial \overline{\text{SINR}}^{\alpha \rightarrow \infty}}{\partial \beta} = 0$. In Lemma 7 we show that $e + \beta e' > 0$, hence we only need to deal with the term t_1 . We remember from (8) that $c - 1 - \beta \varepsilon e + 2\beta c \varepsilon e + e + \beta \varepsilon e^2 = 0$. Thus,

$$P[c - 1 - \beta \varepsilon e + 2\beta c \varepsilon e] + \beta e + \beta^2 e^2 \varepsilon = -Pe - P\beta \varepsilon e^2 + \beta e + \beta^2 \varepsilon e^2.$$

Hence,

$$t_1 = (\varepsilon e^2 + P\tau^2 c e^2 \varepsilon^2) \left(\beta - \frac{P(1 - \tau^2)}{Pc\varepsilon\tau^2 + 1} \right) \left(\beta + \frac{1}{\varepsilon e} \right).$$

Given that only the middle term can become zero, we find β_{opt} to be as stated in (9). The physical interpretation of the SINR guarantees this point to be the maximum.

Lemma 7. *Given the notation and definitions from Appendix B-C, $e + \beta e' > 0$.*

Proof Sketch: From [28] we know that an object of the form

$$m(z) = [-z + c \int \frac{t}{1 + tm(z)} d\nu(t)]^{-1}$$

where ν is a non-negative finite measure, is a so-called Stieltjes transform of a measure ν , defined $\forall z \notin \text{Supp}(\nu)$. Adapting (25) by re-naming $\tilde{e} \triangleq \beta \varepsilon e$ we see that it is indeed a valid Stieltjes transform for an appropriately chosen measure. Finally, one recognizes $\beta e' + e$ as the derivative of a Stieltjes transform, which is always positive. ■

APPENDIX C

PROOF OF THEOREM 1

A. Power Normalization Term

We start by finding a DE of the term ν_m , which will turn out to be a frequently reoccurring object, throughout this Section. From (13), we see that the power normalization term ν_m is defined by the relationship

$$\frac{P_m K_m}{\nu_m N_m} = \frac{1}{N_m} \text{tr} \left[\hat{\mathbf{H}}_m^m (\hat{\mathbf{H}}_m^m)^H \mathbf{Q}_m^2 \right]$$

$$= \frac{\partial}{\partial \gamma_m} \left\{ \frac{1}{\alpha_m^m N_m} \text{tr} \left[(\mathbf{Z}^m + \gamma_m \mathbf{I}_{N_m}) \mathbf{Q}_m \right] \right\} \quad (27)$$

where we used the general identities $\frac{\partial}{\partial y} \left\{ -\text{tr} \left[\mathbf{A} (\mathbf{A} + \mathbf{B} + y\mathbf{I})^{-1} \right] \right\} = \text{tr} \left[\mathbf{A} (\mathbf{A} + \mathbf{B} + y\mathbf{I})^{-2} \right]$ and $\mathbf{A} (\mathbf{A} + \mathbf{B} + y\mathbf{I})^{-1} = \mathbf{I} - (\mathbf{B} + y\mathbf{I}) (\mathbf{A} + \mathbf{B} + y\mathbf{I})^{-1}$. The goal now is to find a deterministic object \bar{X}_m that satisfies

$$\frac{1}{N_m} \text{tr} \left[\hat{\mathbf{H}}_m^m (\hat{\mathbf{H}}_m^m)^H \mathbf{Q}_m^2 \right] - \bar{X}_m \xrightarrow{\text{a.s.}} 0$$

for the regime defined in A-1.

To do this, we apply [18, Theorem 1] to (27), where we set the respective variables to be $\Psi_i = \chi_{m,i}^m \mathbf{I}$, $\mathbf{Q}_N = \mathbf{Z}^m + \gamma_m \mathbf{I}_{N_m}$, $\mathbf{B}_N = \alpha_m^m \hat{\mathbf{H}}_m^m (\hat{\mathbf{H}}_m^m)^H + \mathbf{Z}^m$ and $z = -\gamma_m$. Thus, we find the (partially deterministic) quantity

$$\bar{X}_m = \frac{\partial}{\partial \gamma_m} \frac{1}{\alpha_m^m N_m} \text{tr} \left[(\mathbf{Z}^m + \gamma_m \mathbf{I}_{N_m}) \cdot \left(\frac{1}{N_m} \sum_{j=1}^{K_m} \frac{\alpha_m^m \chi_{m,j}^m \mathbf{I}_{N_m}}{1 + e_{(m)}^j} + \mathbf{Z}^m + \gamma_m \mathbf{I}_{N_m} \right)^{-1} \right]$$

where $e_{(m)}^j = \alpha_m^m \chi_{m,i}^m e_{(m)}$ and

$$e_{(m)} = \frac{1}{N_m} \text{tr} \left(\frac{1}{N_m} \sum_{j=1}^{K_m} \frac{\alpha_m^m \chi_{m,j}^m \mathbf{I}_{N_m}}{1 + \alpha_m^m \chi_{m,j}^m e_{(m)}} + \mathbf{Z}^m + \gamma_m \mathbf{I}_{N_m} \right)^{-1}.$$

Remark 2. *In order to reuse the results from this section later on, it will turn out to be useful to realize the following relationship involving $e_{(m)}$.*

$$\frac{1}{N_m} \text{tr} \mathbf{Q}_m - e_{(m)} \xrightarrow{\text{a.s.}} 0. \quad (28)$$

This can be quickly verified by [18, Theorem 1], when choosing $\Psi_i = \chi_{m,i}^m \mathbf{I}$, $\mathbf{Q}_N = \mathbf{I}$, $\mathbf{B}_N = \alpha_m^m \hat{\mathbf{H}}_m^m (\hat{\mathbf{H}}_m^m)^H + \mathbf{Z}^m$ and $z = -\gamma_m$.

One notices, that the fixed-point equation $e_{(m)}$ contains the term \mathbf{Z}^m , which is not deterministic. Thus, our found objects are not yet DEs. In order to resolve this situation we condition \mathbf{Z}^m to be fixed for now. Under this assumption we now find the DE of $e_{(m)}$. To do this, it is necessary to realize that $e_{(m)}$ contains another Stieltjes transform:

$$e_{(m)} = \frac{1}{N_m} \text{tr} \left[(\mathbf{Z}^m + \beta_m \mathbf{I}_{N_m})^{-1} \right]$$

where

$$\beta_m = \frac{1}{N_m} \sum_{j=1}^{K_m} \frac{\alpha_m^m \chi_{m,j}^m}{1 + \alpha_m^m \chi_{m,j}^m e_{(m)}} + \gamma_m. \quad (29)$$

The solution becomes immediate once we rephrase \mathbf{Z}^m as

$$\mathbf{Z}^m = \sum_{l \neq m} \sum_{k=1}^{K_l} \alpha_l^m \hat{\mathbf{h}}_{l,k}^m (\hat{\mathbf{h}}_{l,k}^m)^H = \check{\mathbf{H}}_{[m]}^m \mathbf{A}_{[m]}^m \left(\check{\mathbf{H}}_{[m]}^m \right)^H$$

where $\check{\mathbf{H}}_{[m]}^m \in \mathbb{C}^{N_m \times K_{[m]}}$, with $K_{[m]} = \sum_{l \neq m} K_l$, is the aggregated matrix of the vectors $\check{\mathbf{h}}_{l,k}^m \sim \mathcal{CN}(0, \frac{1}{N_m} \mathbf{I}_{N_m})$, $\forall l \neq m$ and

$$\mathbf{A}_{[m]}^m = \text{diag} \left[\alpha_1^m \chi_{1,1}^m, \dots, \alpha_1^m \chi_{1,K_1}^m, \alpha_2^m \chi_{2,1}^m, \dots, \alpha_2^m \chi_{2,K_2}^m, \dots, \alpha_{m-1}^m \chi_{m-1,1}^m, \dots, \alpha_{m-1}^m \chi_{m-1,K_{m-1}}^m, \alpha_{m+1}^m \chi_{m+1,1}^m, \dots, \alpha_B^m \chi_{B,K_B}^m \right]$$

Now, one can directly apply [28] or [26][Th 3.13, Eq 3.23] with $\mathbf{T} = \mathbf{A}_{[m]}^m$ and $\mathbf{X} = (\tilde{\mathbf{H}}_{[m]}^m)^H$. Being careful with the notation ($\mathbf{X}\mathbf{T}\mathbf{X}^H$ instead of $(\tilde{\mathbf{H}}_{[m]}^m)^H \mathbf{A}_{[m]}^m \tilde{\mathbf{H}}_{[m]}^m$), we arrive at:

$$e_{(m)} = \frac{1}{N_m} \text{tr} \left\{ \left[\tilde{\mathbf{H}}_{[m]}^m \mathbf{A}_{[m]}^m (\tilde{\mathbf{H}}_{[m]}^m)^H + \beta_m \mathbf{I}_{N_m} \right]^{-1} \right\}$$

where

$$e_{(m)} - \frac{1}{N_m} \left[\beta_m + \frac{1}{N_m} \sum_{l \neq m} \sum_k \frac{\alpha_l^m \chi_{l,k}^m}{1 + \alpha_l^m \chi_{l,k}^m e_{(m)}} \right] \xrightarrow{\text{a.s.}} 0.$$

Here we used Remark 2 and β_m is given in (29)

Now, combining the intermediate results, using again Remark 2 and the relationship $\text{tr} \mathbf{A} (\mathbf{A} + x\mathbf{I})^{-1} = \text{tr} \mathbf{I} - x \text{tr} (\mathbf{A} + x\mathbf{I})^{-1}$ with $\mathbf{A} = \mathbf{Z}^m + \gamma_m \mathbf{I}_{N_m}$, we arrive at

$$\bar{X}_m = -\frac{1}{\alpha_m^m N_m} \sum_{j=1}^{K_m} \frac{\alpha_m^m \chi_{m,j}^m e'_{(m)}}{(1 + \alpha_m^m \chi_{m,j}^m e_{(m)})^2}$$

where $e'_{(m)}$ is shorthand for $\partial/\partial\gamma_m e_{(m)}$ and can be found (by tedious calculus) to be as stated in (19), which concludes this part of the proof.

B. Signal Power Term

The important part of the signal power term (14) (w.r.t. this proof) is $(\mathbf{h}_{l,k}^l)^H \mathbf{Q}_l \hat{\mathbf{h}}_{l,k}^l$. We will now derive a DE of this quantity. Before proceeding, we remind ourselves that our chosen model of the estimated channel (12) entails the following relationships: $\mathbf{h}_{l,k}^l \perp \tilde{\mathbf{h}}_{l,k}^l$, $\hat{\mathbf{h}}_{l,k}^l \not\perp \mathbf{h}_{l,k}^l$, $\hat{\mathbf{h}}_{l,k}^l \not\perp \tilde{\mathbf{h}}_{l,k}^l$, $\mathbf{Q}_{l[k]} \perp \hat{\mathbf{h}}_{l,k}^l$, $\mathbf{Q}_{l[k]} \perp \mathbf{h}_{l,k}^l$. Also, formulations containing $\hat{\mathbf{h}}_{l,k}^l$ can often be split into two terms comprising $\mathbf{h}_{l,k}^l$ and $\tilde{\mathbf{h}}_{l,k}^l$. Hence, the application of Lemmas 3, 4, 5 and Corollary 1, in the following is well justified. Employing (28) one sees

$$(\mathbf{h}_{l,k}^l)^H \mathbf{Q}_l \hat{\mathbf{h}}_{l,k}^l - \frac{\sqrt{\chi_{l,k}^l} \sqrt{(1 - (\tau_l^l)^2)} e_{(l)}}{1 + \alpha_l^l \chi_{l,k}^l e_{(l)}} \xrightarrow{\text{a.s.}} 0.$$

Finally, applying this result to the complete formulation (14), we arrive at the familiar term from Theorem 1.

C. Preparation for Interference Terms

In this subsection we derive the deterministic equivalents of the two terms $(\mathbf{h}_{l,k}^l)^H \mathbf{B} \mathbf{Q}_l \mathbf{h}_{l,k}^l$ and $(\mathbf{h}_{l,k}^l)^H \mathbf{B} \mathbf{Q}_l \tilde{\mathbf{h}}_{l,k}^l$, where $\mathbf{B} \in \mathbb{C}^{N_l \times N_l}$ has uniformly bounded spectral norm w.r.t. N_l and is independent of $\mathbf{h}_{l,k}^l$ and $\tilde{\mathbf{h}}_{l,k}^l$. The following approach is based on and slightly generalizes [18, Lemma 7]. First, we realize that

$$\mathbf{Q}_a^{-1} - \mathbf{Q}_{a[bc]}^{-1} = c_0 \mathbf{h}_{b,c}^a (\mathbf{h}_{b,c}^a)^H + c_2 \mathbf{h}_{b,c}^a (\tilde{\mathbf{h}}_{b,c}^a)^H + c_2 \tilde{\mathbf{h}}_{b,c}^a (\mathbf{h}_{b,c}^a)^H + c_1 \tilde{\mathbf{h}}_{b,c}^a (\tilde{\mathbf{h}}_{b,c}^a)^H \quad (30)$$

where $c_0 = \alpha_b^a \chi_{b,c}^a (1 - (\tau_b^a)^2)$, $c_1 = \alpha_b^a \chi_{b,c}^a (\tau_b^a)^2$ and $c_2 = \alpha_b^a \chi_{b,c}^a \sqrt{(1 - (\tau_b^a)^2)} \tau_b^a$. We omitted designating the dependencies of c on a and b , as this is always clear from the context. To ease the exposition, we also introduce the following abbreviations $Y_1 \triangleq (\hat{\mathbf{h}}_{l,k}^l)^H \mathbf{Q}_{l[k]} \hat{\mathbf{h}}_{l,k}^l$, $Y_4 \triangleq (\mathbf{h}_{l,k}^l)^H \mathbf{B} \mathbf{Q}_l \mathbf{h}_{l,k}^l$, $Y_2 \triangleq (\mathbf{h}_{l,k}^l)^H \mathbf{Q}_l \tilde{\mathbf{h}}_{l,k}^l$,

$Y_5 \triangleq (\tilde{\mathbf{h}}_{l,k}^l)^H \mathbf{Q}_{l[k]} \tilde{\mathbf{h}}_{l,k}^l$, $Y_3 \triangleq (\mathbf{h}_{l,k}^l)^H \mathbf{B} \mathbf{Q}_l \mathbf{h}_{l,k}^l$ and $Y_6 \triangleq (\mathbf{h}_{l,k}^l)^H \mathbf{Q}_l \mathbf{h}_{l,k}^l$. Finally, we begin with the term $(\mathbf{h}_{l,k}^l)^H \mathbf{B} \mathbf{Q}_l \tilde{\mathbf{h}}_{l,k}^l$:

$$\begin{aligned} & (\mathbf{h}_{l,k}^l)^H \mathbf{B} \mathbf{Q}_l \tilde{\mathbf{h}}_{l,k}^l - (\mathbf{h}_{l,k}^l)^H \mathbf{B} \mathbf{Q}_{l[k]} \tilde{\mathbf{h}}_{l,k}^l \\ & \stackrel{(24)}{=} -(\mathbf{h}_{l,k}^l)^H \mathbf{B} \mathbf{Q}_l (\mathbf{Q}_l^{-1} - \mathbf{Q}_{l[k]}^{-1}) \mathbf{Q}_{l[k]} \tilde{\mathbf{h}}_{l,k}^l \end{aligned}$$

and, using (30), we find

$$(\mathbf{h}_{l,k}^l)^H \mathbf{B} \mathbf{Q}_l \tilde{\mathbf{h}}_{l,k}^l = \frac{Y_3 - (\mathbf{h}_{l,k}^l)^H \mathbf{B} \mathbf{Q}_l \mathbf{h}_{l,k}^l (c_0 Y_2 + c_2 Y_5)}{1 + c_2 Y_2 + c_1 Y_5} \quad (31)$$

Similarly, for the term $(\mathbf{h}_{l,k}^l)^H \mathbf{B} \mathbf{Q}_l \mathbf{h}_{l,k}^l$ we arrive at

$$\begin{aligned} & (\mathbf{h}_{l,k}^l)^H \mathbf{B} \mathbf{Q}_l \mathbf{h}_{l,k}^l (1 + c_0 Y_6 + c_2 Y_1) \\ & = Y_4 - (\mathbf{h}_{l,k}^l)^H \mathbf{B} \mathbf{Q}_l \tilde{\mathbf{h}}_{l,k}^l (c_2 Y_5 + c_1 Y_1). \end{aligned} \quad (32)$$

Now, applying (31) to (32), one arrives at

$$\begin{aligned} & (\mathbf{h}_{l,k}^l)^H \mathbf{B} \mathbf{Q}_l \mathbf{h}_{l,k}^l \left[(1 + c_0 Y_6 + c_2 Y_1) - \frac{(c_0 Y_2 + c_2 Y_5)(c_2 Y_6 + c_1 Y_1)}{1 + c_2 Y_2 + c_1 Y_5} \right] \\ & = Y_4 - \frac{(\mathbf{h}_{l,k}^l)^H \mathbf{B} \mathbf{Q}_l \tilde{\mathbf{h}}_{l,k}^l (c_2 Y_6 + c_1 Y_1)}{1 + c_2 Y_2 + c_1 Y_5}. \end{aligned} \quad (33)$$

Similar to Appendix C-B, we notice that Y_1 , Y_2 and Y_3 converge almost surely to 0 in the large system limit: $Y_1, Y_2, Y_3 \xrightarrow{\text{a.s.}} 0$. We also foresee that

$$Y_4 - u' \xrightarrow{\text{a.s.}} 0, \quad Y_5 - u_1 \xrightarrow{\text{a.s.}} 0, \quad Y_6 - u_2 \xrightarrow{\text{a.s.}} 0$$

where the values for u' , u_1 and u_2 are not yet of concern. Thus, (33) finally leads to

$$(\mathbf{h}_{l,k}^l)^H \mathbf{B} \mathbf{Q}_l \mathbf{h}_{l,k}^l - \frac{u' (1 + c_1 u_1)}{1 + c_1 u_1 + c_0 u_2 + (c_0 c_1 - c_2^2) u_1 u_2} \xrightarrow{\text{a.s.}} 0. \quad (34)$$

In order to find the second original term $(\mathbf{h}_{l,k}^l)^H \mathbf{B} \mathbf{Q}_l \tilde{\mathbf{h}}_{l,k}^l$, we reform and plug (32) into (31) and follow analogously the path we took to arrive at (34). We finally find

$$(\mathbf{h}_{l,k}^l)^H \mathbf{B} \mathbf{Q}_l \tilde{\mathbf{h}}_{l,k}^l - \frac{-c_2 u_1 u'}{1 + c_1 u_1 + c_0 u_2 + (c_0 c_1 - c_2^2) u_1 u_2} \xrightarrow{\text{a.s.}} 0. \quad (35)$$

D. Interference Power Terms

Having obtained the preparation results in Appendix C-C we can now continue to find the DEs for different parts of the interference power term. From (15) we arrive at

$$\begin{aligned} \text{Int}_{l,k}^{(l)} &= \sum_{m \neq l} \nu_m \chi_{l,k}^m \underbrace{(\mathbf{h}_{l,k}^l)^H \mathbf{Q}_m \hat{\mathbf{H}}_m^m (\hat{\mathbf{H}}_m^m)^H \mathbf{Q}_m \mathbf{h}_{l,k}^m}_{\text{Part A}_m} \\ & \quad + \nu_l \chi_{l,k}^l \underbrace{(\mathbf{h}_{l,k}^l)^H \mathbf{Q}_l \hat{\mathbf{H}}_{l[k]}^l (\hat{\mathbf{H}}_{l[k]}^l)^H \mathbf{Q}_l \mathbf{h}_{l,k}^l}_{\text{Part B}}. \end{aligned} \quad (36)$$

We start by treating (36) Part B first. Employing the relationships $\mathbf{A}\mathbf{B}\mathbf{C} = \mathbf{A}\mathbf{C}\mathbf{D} + \mathbf{A}(\mathbf{B} - \mathbf{C})\mathbf{D}$ and (24) one finds

$$\begin{aligned} \text{Part B} &= (\mathbf{h}_{l,k}^l)^H \mathbf{Q}_{l[k]} \hat{\mathbf{H}}_{l[k]}^l (\hat{\mathbf{H}}_{l[k]}^l)^H \mathbf{Q}_l \hat{\mathbf{H}}_{l[k]}^l \mathbf{h}_{l,k}^l \\ & \quad - (\mathbf{h}_{l,k}^l)^H \mathbf{Q}_l \left[\mathbf{Q}_l^{-1} - \mathbf{Q}_{l[k]}^{-1} \right] \mathbf{Q}_{l[k]} \hat{\mathbf{H}}_{l[k]}^l (\hat{\mathbf{H}}_{l[k]}^l)^H \mathbf{Q}_l \mathbf{h}_{l,k}^l. \end{aligned}$$

We rewrite Part B as

$$\text{Part B} = X_1 - c_0 X_3 X_1 - c_2 X_3 X_2 - c_2 X_4 X_1 - c_1 X_4 X_2.$$

Where we have found and abbreviated the 4 quadratic forms, $X_1 = (\mathbf{h}_{l,k}^l)^H \mathbf{Q}_{l[k]} \hat{\mathbf{H}}_{l[k]}^l (\hat{\mathbf{H}}_{l[k]}^l)^H \mathbf{Q}_l \mathbf{h}_{l,k}^l$, $X_2 = (\hat{\mathbf{h}}_{l,k}^l)^H \mathbf{Q}_{l[k]} \hat{\mathbf{H}}_{l[k]}^l (\hat{\mathbf{H}}_{l[k]}^l)^H \mathbf{Q}_l \mathbf{h}_{l,k}^l$, $X_3 = (\mathbf{h}_{l,k}^l)^H \mathbf{Q}_l \mathbf{h}_{l,k}^l$ and $X_4 = (\mathbf{h}_{l,k}^l)^H \mathbf{Q}_l \mathbf{h}_{l,k}^l$.

To find the deterministic equivalents for X_1 and X_2 , we can use (34) and (35), respectively, where $\mathbf{B} = \mathbf{Q}_{l[k]} \hat{\mathbf{H}}_{l[k]}^l (\hat{\mathbf{H}}_{l[k]}^l)^H$. The respective variables u_1 , u_2 and u' for this choice of \mathbf{B} are found (using the same standard techniques as in Appendix C-B) to be

$$u_1 = (\tilde{\mathbf{h}}_{l,k}^l)^H \mathbf{Q}_{l[k]} \tilde{\mathbf{h}}_{l,k}^l \Rightarrow u_1 - e(l) \xrightarrow{\text{a.s.}} 0.$$

Analogously, $u_1 - e(l) \xrightarrow{\text{a.s.}} 0$. Hence, we see that u_1 and u_2 converge to the same value and we will abbreviate them henceforth as u . For the still missing term u' we arrive at

$$u' = (\mathbf{h}_{l,k}^l)^H \mathbf{Q}_{l[k]} \hat{\mathbf{H}}_{l[k]}^l (\hat{\mathbf{H}}_{l[k]}^l)^H \mathbf{Q}_l \mathbf{h}_{l,k}^l \Rightarrow u' - g(l) \xrightarrow{\text{a.s.}} 0$$

where the last step makes use of the results in Appendix C-A. Also, we remind ourselves that we have $c_0 = \alpha_l^l \chi_{l,k}^l (1 - (\tau_l^l)^2)$, $c_1 = \alpha_l^l \chi_{l,k}^l (\tau_l^l)^2$ and $c_2 = \alpha_l^l \chi_{l,k}^l \sqrt{(1 - (\tau_l^l)^2) \tau_l^l}$, hence $c_0 + c_1 = \alpha_l^l \chi_{l,k}^l$ and $c_0 c_1 - c_2^2 = 0$. So, finally, we have

$$X_1 - \frac{u'(1+c_1u)}{1+(c_1+c_0)u} \xrightarrow{\text{a.s.}} 0 \text{ and } X_2 - \frac{-c_2uu'}{1+(c_1+c_0)u} \xrightarrow{\text{a.s.}} 0.$$

To find the DEs for X_3 and X_4 , we can again use (34) and (35), respectively. This time $\mathbf{B} = \mathbf{I}$ and hence the variables simplify to $u' = u_1 = u_2 \triangleq u$, where $u - e(l) \xrightarrow{\text{a.s.}} 0$. Thus,

$$X_3 - \frac{u(1+c_1u)}{1+(c_1+c_0)u} \xrightarrow{\text{a.s.}} 0 \text{ and } X_4 - \frac{-c_2u^2}{1+(c_1+c_0)u} \xrightarrow{\text{a.s.}} 0.$$

Combining all results after further simplifications, we can express the DE of Part B, i.e., $\overline{\text{Part B}}$, as

$$\overline{\text{Part B}} = g(l) \frac{1 - (\tau_l^l)^2}{\left(1 + \alpha_l^l \chi_{l,k}^l e(l)\right)^2} + g(l) (\tau_l^l)^2.$$

The next step is to derive the DE of (36) Part A_m , i.e., $\overline{\text{Part } A_m}$. Fortunately, the sum obliges $m \neq l$ and, thus, the same derivation like for Part B applies. Hence, we arrive at

$$\overline{\text{Part } A_m} = g(m) \frac{1 - (\tau_l^m)^2}{\left(1 + \alpha_l^m \chi_{l,k}^m e(m)\right)^2} + g(m) (\tau_l^m)^2.$$

Combining Part B and the sum of Part A_m with our original expression of the interference power, we arrive at the expression in Theorem 1.

REFERENCES

- [1] E. Björnson and E. Jorswieck, "Optimal resource allocation in coordinated multi-cell systems," *Foundations and Trends in Communications and Information Theory*, vol. 9, no. 2-3, pp. 113–381, 2013.
- [2] J. Hoydis, K. Hosseini, S. ten Brink, and M. Debbah, "Making Smart Use of Excess Antennas: Massive MIMO, Small Cells, and TDD," *Bell Labs Technical Journal*, vol. 18, no. 2, pp. 5–21, 2013.
- [3] Cisco, "Cisco visual networking index: Global mobile data traffic forecast update, 2012–2017," *White Paper*, 2013.
- [4] G. Americas, "Meeting the 1000x Challenge: The Need for Spectrum, Technology and Policy Innovation," Tech. Rep., October 2013.
- [5] W. Webb, *Wireless Communications: The Future*. John Wiley & Sons, 2007.
- [6] J. G. Andrews, H. Claussen, M. Dohler, S. Rangan, and M. C. Reed, "Femtocells: Past, present, and future," *Selected Areas in Communications, IEEE Journal on*, vol. 30, no. 3, pp. 497–508, 2012.
- [7] J. Hoydis, M. Kobayashi, and M. Debbah, "Green small-cell networks," *IEEE Veh. Technol. Mag.*, vol. 6, no. 1, pp. 37–43, Mar. 2011.
- [8] H. Dai, A. F. Molisch, and H. V. Poor, "Downlink capacity of interference-limited MIMO systems with joint detection," *Wireless Communications, IEEE Transactions on*, vol. 3, no. 2, pp. 442–453, 2004.
- [9] D. Gesbert, S. Hanly, H. Huang, S. Shamai, O. Simeone, and W. Yu, "Multi-cell MIMO cooperative networks: A new look at interference," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1380–1408, 2010.
- [10] T. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, 2010.
- [11] R. Irmer, H. Droste, P. Marsch, M. Grieger, G. Fettweis, S. Brueck, H.-P. Mayer, L. Thiele, and V. Jungnickel, "Coordinated Multipoint: Concepts, Performance, and Field Trial Results," *IEEE Communications Magazine*, vol. 49, no. 2, pp. 102–111, 2011.
- [12] D. Gesbert, M. Kountouris, R. Heath, C.-B. Chae, and T. Sälzer, "Shifting the MIMO paradigm," *IEEE Signal Process. Mag.*, vol. 24, no. 5, pp. 36–46, 2007.
- [13] C. Peel, B. Hochwald, and A. 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, 2005.
- [14] R. Zakhour and D. Gesbert, "Distributed multicell-MISO precoding using the layered virtual SINR framework," *IEEE Trans. Wireless Commun.*, vol. 9, no. 8, pp. 2444–2448, 2010.
- [15] E. Björnson and B. Ottersten, "On the Principles of Multicell Precoding with Centralized and Distributed Cooperation," in *Proc. WCSP*, 2009.
- [16] E. Björnson, M. Bengtsson, and B. Ottersten, "Optimal Multiuser Transmit Beamforming: A Difficult Problem with a Simple Solution Structure," *IEEE Signal Processing Magazine*, 2014.
- [17] A. Müller, E. Björnson, R. Couillet, and M. Debbah, "Analysis and Management of Heterogeneous User Mobility in Large-scale Downlink Systems," in *Proceedings of the Asilomar Conference on Signals, Systems, and Computers*, 2013, pp. 1–5.
- [18] S. Wagner, R. Couillet, M. Debbah, and D. Slock, "Large system analysis of linear precoding in MISO broadcast channels with limited feedback," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4509–4537, 2012.
- [19] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?" *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, Feb. 2013.
- [20] A. Wyner, "Shannon-theoretic approach to a Gaussian cellular multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 40, no. 6, pp. 1713–1727, 1994.
- [21] J. Xu, J. Zhang, and J. Andrews, "On the accuracy of the Wyner model in cellular networks," *IEEE Trans. Wireless Commun.*, vol. 10, no. 9, pp. 3098–3109, 2011.
- [22] J. Choi, D. Love, and P. Bidigare, "Downlink training techniques for FDD massive MIMO systems: Open-loop and closed-loop training with memory," Sept. 2013, submitted, arXiv:1309.7712.
- [23] C. Wang and R. Murch, "Adaptive downlink multi-user MIMO wireless systems for correlated channels with imperfect CSI," *IEEE Trans. Wireless Commun.*, vol. 5, no. 9, pp. 2435–2436, Sept. 2006.
- [24] B. Nosrat-Makouei, J. Andrews, and R. Heath, "MIMO interference alignment over correlated channels with imperfect CSI," *IEEE Trans. Signal Process.*, vol. 59, no. 6, pp. 2783–2794, Jun. 2011.
- [25] 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, 2007.
- [26] R. Couillet and M. Debbah, *Random Matrix Methods for Wireless Communications*. Cambridge University Press, 2011.
- [27] C. D. Meyer, *Matrix Analysis and Applied Linear Algebra*. Siam, 2000, vol. 2, <http://www.matrixanalysis.com/DownloadChapters.html>.
- [28] 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, 1995.
- [29] 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.