Interference-Aware RZF Precoding for Multi Cell Downlink Systems
Axel Müller, Romain Couillet, Emil Björnson, Sebastian Wagner, Mérouane Debbah

To cite this version:

HAL Id: hal-01242434
https://hal.archives-ouvertes.fr/hal-01242434
Submitted on 12 Dec 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Abstract—Recently, a structure of an optimal linear precoder for multi cell downlink systems has been described in [1 Eq (3.33)]. Other references (e.g., [2], [3]) have used simplified versions of the precoder to obtain promising performance gains. These gains have been hypothesized to stem from the additional degrees of freedom that allow for interference mitigation through interference relegation to orthogonal subspaces. However, no conclusive or rigorous understanding has yet been developed.

In this paper, we build on an intuitive interference induction trade-off and the aforementioned precoding structure to propose an interference aware RZF (iaRZF) precoding scheme for multi cell downlink systems and we analyze its rate performance. Special emphasis is placed on the induced interference mitigation mechanism of iaRZF. 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 very poor CSI allows for significant sum-rate gains. Our obtained insights are then used to propose heuristic precoder parameters for arbitrary systems, whose effectiveness are shown in more involved system scenarios. Furthermore, calculation and implementation of these parameters does not require explicit inter base station cooperation.

Index Terms—Multi user MIMO, linear precoding, multi cell downlink, interference mitigation, random matrix theory.

A. Müller was with Intel Mobile Communications, Sophia Antipolis, France and with the Alcatel-Lucent Chair on Flexible Radio, Supélec, Gif-sur-Yvette, France. He is currently with the Mathematical and Algorithmic Sciences Lab, France Research Center, Huawei Technologies Co. Ltd., Boulogne-Billancourt, France (email: axel.mueller@huawei.com).

R. Couillet is with Laboratoire des Signaux et Systèmes (L2S, UMR CNRS 8506), CentraleSupelec - CNRS - Université Paris-Sud, Gif-sur-Yvette, France (email: romain.couillet@centralesupelec.fr).

E. Björnson was with the Alcatel-Lucent Chair on Flexible Radio, Supélec, Gif-sur-Yvette, France, and with the Department of Signal Processing, KTH Royal Institute of Technology, Stockholm, Sweden. He is currently with the Department of Electrical Engineering (ISY), Linköping University, Linköping, Sweden (email: emil.bjornson@liu.se).

S. Wagner is with Intel Mobile Communications, Sophia Antipolis, France (email: sebastian.wagner@intel.com).

M. Debbah is with Laboratoire des Signaux et Systèmes (L2S, UMR CNRS 8506), CentraleSupelec - CNRS - Université Paris-Sud, Gif-sur-Yvette, France, and with the Mathematical and Algorithmic Sciences Lab, France Research Center, Huawei Technologies Co. Ltd., Boulogne-Billancourt, France (email: merouane.debbah@centralesupelec.fr).

E. Björnson was 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).

Copyright (c) 2014 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

I. INTRODUCTION

The growth of data traffic and the number of user terminals (UTs) in cellular networks will likely persist for the foreseeable future [4]. In order to deal with the resulting demand, it is estimated [5] 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 [6]. Chapter 6.3.4. In recent times, this technique has been named the small cell approach [7], [8]. A large body of research indicates that interference still is a major limiting factor for capacity in multi cell scenarios [9], [10], especially in modern cellular networks that serve a multitude of users within 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 MIMO approach [11]. 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]. [12], [13]. 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 estimation errors and exploit the available CSI as efficiently as possible.

Arguably, the most successful and practically applicable precoding scheme used today is RZF precoding [14] (also known as minimum mean square error (MMSE) precoding, transmit Wiener filter, generalized eigenvalue-based beamformer, etc.; see [1 Remast 3.2]). Classical RZF precoders are only defined for single cell systems and thus do not take inter cell interference into account. Disregarding available information about inter cell interference is particularly detrimental in high density scenarios, where interference is a main performance limiting factor. Early multi cell extensions of the RZF scheme do not take the quality of CSI into account [15] and later ones either rely on heuristic distributed optimization algorithms or on inter cell cooperation [16] to determine the precoding vector. Thus, they offer limited insight into the precoder structure, i.e., into how the precoder works and how...
it could be improved.

An intuitive extension of the single cell RZF, with the goal of completely eliminating induced interference is to substitute the intra cell channel matrix $\mathbf{H}$ in the (qualitative, single antenna UTs) precoder formulation $\mathbf{F} = \mathbf{H}(\mathbf{H}^H + \xi \mathbf{I})^{-1}$ by a matrix $\tilde{\mathbf{H}}$, which is $\mathbf{H}$ projected onto the space orthogonal to the inter cell channel matrices, i.e., $\mathbf{F} = \tilde{\mathbf{H}}(\tilde{\mathbf{H}}^H + \xi \mathbf{I})^{-1}$. Hence induced interference can be completely removed at the cost of reduced signal power, if the CSI is perfectly known. However, it is immediately clear that this is a very harsh requirement, since the projection negatively affects the amount of signal energy received at the served UTs (unless $\mathbf{H} = \tilde{\mathbf{H}}$). Assuming the precoding objective is system wide sum-rate optimization, one realizes that single cell RZF is probably not optimal, since it reduces the rate in other cells due to induced interference. Thus, a trade-off between the two extremes is expected to be beneficial, especially when the channel matrices are estimated with dissimilar quality. In this paper we analyze the following class of precoders for multi-cell single antenna UT systems, which we will denote the following class of precoders for multi-cell single antenna UT systems, which we will denote $\text{iaRZF}$. This class allows for the desired trade-off, as will be shown later on:

$$F_m^l = \left( \sum_{k=1}^{L} \alpha_k^m \tilde{H}^m_k (\tilde{H}^m_k)^H + \xi_m \mathbf{I}_{N_m} \right)^{-1} \tilde{H}^m_k \nu_m^k. \quad (1)$$

Here $F_m^l$ is the linear precoder used by base station (BS) $m$ and $\tilde{H}^m_k$ denotes the imperfect estimate of the channel matrix from BS $m$ to the UTs in cell $l$. The factor $\xi_m$ is a regularization parameter and $\nu_m$ normalizes the precoder. Each channel matrix is assigned a factor $\alpha_k^m$, that can be interpreted as the importance placed on the respective estimated channel. It is easy to see how this structure can mimic single cell RZF under perfect CSI (choose $\alpha_k^m = 0$, $l \neq m$ and $\alpha_m^m = 1$). The weights $\alpha_k^m$ allow balancing signal power directed to the served users with interference induced to other cells. This can be used to optimize sum-rate performance in certain cases, as will be shown in Section II. In general the optimal weights are not known and the classical UL/DL duality approach (e.g., [11]) cannot be applied to find these weights, outside of power minimization settings or, if imperfect CSI is considered. We note that every BS can try to estimate the interference from other cells without explicit inter cell cooperation or communication, by means of blind or known pilot based schemes, though the CSI quality might be rather poor. Such estimation might be considered as implicit coordination. In [3] a simplified version of iaRZF was discussed, where a single subset of UT channels was weighted with respect to an estimated receive covariance matrix of all interfering channels. Hoydis et al. argued that “large [weights] make the precoding vectors more orthogonal to the interference subspace”, but they did not conclusively and rigorously show how or why this is achieved. The work in [3] builds upon results from [2, Theorem 6], which introduced the precoding structure offering a single common balance parameter weighting all inter cell interference. Differing from these works, we are more interested in increasing the sum-rate performance by giving the traditional RZF precorder even more additional degrees of freedom. Doing so enables it to separately take into account the interference induced to certain subspaces, associated with different cells. Increasing the degrees of freedom in the interference suppression was also shown to be effective in [12] for perfect CSI. Similar to the traditional RZF precoder our approach is empirical and based on several motivational aspects (see Section II). The iaRZF structure is also partially based on the work in [17] and [1, Eq (3.33)]. In the latter, one finds one of the most recent and general treatments of the multi cell RZF precoder, along with proof that the proposed structure is optimal w.r.t. many utility functions of practical interest (see also [18]).

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 poor 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 in which the number of transmit antennas and UTs are both large. Furthermore, new finite dimensional approaches for analyzing multi cell RZF precoding schemes are introduced and applied for limiting cases. 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 [20] for single cell systems and in [21] for multi cell systems, where only deterministic channel covariance matrices can be used in the analysis pertaining to the suppression of inter cell interference. Then, these extensions are used to optimize the sum-rate of the iaRZF precoding scheme in limiting cases. Insights gathered from this lead us to propose and motivate an appropriate heuristic scaling of the precoder weights w.r.t. various system parameters, that offers attractive sum-rate performance; also in non-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 $\mathbf{X}^*$ 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 $\mathbb{E}[]$. 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 $CN(\mathbf{x}, \mathbf{Q})$, where $\mathbf{x}$ 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 $\bot$ and $\perp$ to mean stochastic independence and dependence, respectively.

1Where $\xi$ represents some positive regularization parameter.
II. UNDERSTANDING iARZF

In order to intuitively understand and motivate the iARZF precoder, this section first analyzes its behavior and impact in a relatively simple system.

A. Simple System

We start by examining a simple two cell downlink system, as depicted in Figure 1, which is a further simplification of the Wyner model \[22\], \[23\]. It features a two cell downlink system, with BSs serving single antenna users. For convenience we introduce the notations \( s \in \mathbb{C} \) and \( g \in \mathbb{C} \). We treat \( \varepsilon \) as an arbitrary interference channel gain/path-loss factor. The precoding matrix used at BS \( x \) is denoted \( \mathbf{F}_x \in \mathbb{C}^{N \times K} \). For the channel realizations we choose a simplified system.  

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

In this section we use the previously introduced iaRZF precoding scheme. Given our simple system the unnormalized precoder reads

\[
\mathbf{M}_x = \left( \alpha_x \mathbf{H}_x \mathbf{H}_x^H + \beta_x \mathbf{G}_x \mathbf{G}_x^H + \xi_x \mathbf{I} \right)^{-1} \mathbf{H}_x . \tag{2}
\]

One remarks that the regularization of the identity matrix can also be controlled by scaling \( \alpha \) and \( \beta \) at the same time, if \( \xi \) is fixed to an arbitrary value (e.g., 1). We usually keep \( \alpha \), \( \beta \) and \( \xi \) in our formulas to allow easy comparison to traditional RZF formulations. We assume the following normalization of the precoder:

\[
\mathbf{F}_x = \sqrt{K} \frac{\mathbf{M}_x}{\sqrt{\text{tr} \left( \mathbf{M}_x^H \mathbf{M}_x \right)}} \tag{3}
\]

### 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 section can be simply rewritten to fit the more traditional statistics of \( \mathbf{h}_k \sim \mathcal{CN}(0, \mathbf{I}_N) \) and \( g_{x,k} \sim \mathcal{CN}(0, \frac{1}{N} \mathbf{I}_N) \) for \( k = 1, \ldots, K \). The scaling factor \( 1/N \) is introduced for technical reasons and is further explained in Remark 1.

### B. Performance of Simple System

First, we compare the general performance of the proposed iaRZF scheme with classical approaches, i.e., single cell zero-forcing (ZF), maximum-ratio transmission (MRT) and RZF. The rate of UT\( x,k \) can be defined as

\[
\mathbf{r}_{x,k} = \text{log}_2 \left( 1 + \frac{\text{Sig}_{x,k}}{\text{Int}_{x,k} + \text{Int}_{x,k} + 1} \right)
\]

where \( \text{Sig}_{x,k} = \rho_x \mathbf{h}_x^H \mathbf{F}_{x,k} \mathbf{F}_{x,k}^H \mathbf{h}_{x,k} \), \( \text{Int}_{x,k} = \rho_x \mathbf{h}_x^H \mathbf{F}_{x,k} \mathbf{s}_{x,k} \mathbf{h}_{x,k} \), and \( \text{Int}_{x,k} = \rho_x g_{x,k} \mathbf{F}_{x,k} \mathbf{F}_{x,k}^H g_{x,k} \) denote the received signal power, received intra cell interference and received inter cell interference, respectively.

For comparison we used the following precoder formulations: \( \mathbf{M}_{x, \text{MRT}} = \mathbf{H}_x \), \( \mathbf{M}_{x, \text{ZF}} = \mathbf{H}_x (\mathbf{H}_x^H \mathbf{H}_x)^{-1} \), \( \mathbf{M}_{x, \text{RZF}} = \mathbf{H}_x (\mathbf{H}_x^H + \frac{\mathbf{C}_x}{N \rho_x})^{-1} \), where the regularization in \( \mathbf{M}_{x, \text{RZF}} \) is chosen according to \( 19 \), \( 20 \) and all precoders are normalized as in \( 3 \). The iaRZF weights have been chosen to be \( \alpha = \beta = N \rho_x \) and \( \xi = 1 \), to simplify comparison with traditional RZF precoding. The corresponding performance graphs, obtained by extensive Monte-Carlo (MC) simulations.

\[3\text{It can be shown, using results from Appendix C-A by taking } y_{x,k} = 1 \forall y, \text{ that this implies } \mathbf{F}_{x,k}^H \mathbf{F}_{x,k} \to 1, \text{ almost surely, under Assumption I} \]
in the simplified system (for \( N = 160, K = 40, \varepsilon = 0.7, \rho_1 = \rho_2 = \rho \)), are depicted in Figure 2.

We observe that iaRZF generally outperforms the other schemes. This is not surprising, as the classical 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 for \( \tau = 0.5 \)). Only for extremely bad CSI we observe that iaRZF wastes energy due to an improper choice of \( \alpha, \beta \). Thus, it performs worse than the other schemes, that do not take \( \tau \) into account for precoding. This problem can easily be circumvented by choosing proper weights that let \( \beta \to 0 \) for \( \tau \to 1 \); as will be elaborated later on.

### C. Properties of iaRZF for \( \alpha_x, \beta_x \to \infty \)

As has been briefly remarked by Hoydis et al. in [3], the iaRZF weights \( \alpha_x \) and \( \beta_x \) should, intuitively, allow to project the transmitted signal in subspaces orthogonal to the UTs’ ("own users") and UTs’ ("other users") channels, respectively. This behavior, in the limit cases of \( \alpha \) or \( \beta \to \infty \), is treated analytically in this subsection.

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

First, we introduce the notation \( \mathbf{P}_\mathcal{S} \) as a projection matrix on the space orthogonal to the column space of \( \mathbf{X} \) and we remind ourselves that \( \xi = 1 \) is still assumed. Following the path outlined in Appendix [A][B] and assuming \( \mathbf{H}_x^H \mathbf{H}_x \) invertible (true with probability 1), one finds for \( \alpha_x \to \infty \):

\[
\alpha_x \mathbf{M}_x \xrightarrow{\alpha_x \to \infty} \mathbf{H}_x \left( \mathbf{H}_x^H \mathbf{H}_x \right)^{-1}.
\]

Recall the received signal at the UTs of BS \( x \) in our simple model, due to (only) the intra cell users, is given as \( y_{\text{intra}} = \mathbf{H}_x^H \mathbf{F}_x s_x \). Thus,

\[
y_{\text{intra}} \xrightarrow{\text{iaRZF}} \mathbf{H}_x^H \mathbf{H}_x \left( \mathbf{H}_x^H \mathbf{H}_x \right)^{-1} s_x = \nu s_x
\]

where the precoder normalization leaves a scaling factor \( \nu \) that is independent of \( \alpha_x \). The Lemma 2 used here can be found in Appendix [A]. Hence, we see that for \( \alpha_x \to \infty \) and \( \beta_x \) bounded, the precoder acts similar to a traditional ZF precoder, i.e., the in-cell interference is completely suppressed in our system. The scaling factor includes the loss of energy caused by the alignment constraint, i.e., \( \nu \) decreases the more \( \text{span} \{ \mathbf{H}_x \} \) and \( \text{span} \{ \mathbf{G}_x \} \) intersect. It remains to mention that due to the iaRZF definition, exact ZF can only be achieved in the limit for \( N = K \), where \( \mathbf{H}_x \left( \mathbf{H}_x^H \mathbf{H}_x \right)^{-1} = \left( \mathbf{H}_x^H \mathbf{H}_x \right)^{-1} \mathbf{H}_x \); assuming the inverses exist.

Looking now at the limit \( \beta_x \to \infty \) and following Appendix [B-A], one arrives at

\[
\mathbf{M}_x \xrightarrow{\beta_x \to \infty} \left[ \mathbf{P}_G - \mathbf{P}_G \left( \alpha_x^{-1} \mathbf{I} + \mathbf{H}_x^H \mathbf{P}_G \mathbf{H}_x \right)^{-1} \mathbf{H}_x^H \mathbf{P}_G \right] \mathbf{H}_x = \mathbf{H} \left( \mathbf{I} + \alpha \mathbf{H}_x^H \mathbf{H}_x \right)^{-1}
\]

where we introduced \( \mathbf{H} = \mathbf{P}_G \mathbf{H}_x \) as the channel matrix \( \mathbf{H} \) projected on the space orthogonal to the channels of \( \mathbf{G} \). One remembers that the received signal due to inter cell interference in our simple model is given as

\[
y_{\text{inter}} = \mathbf{G}_x^H \mathbf{F}_x s_x
\]

which, via \( 5 \) in the case of \( \mathbf{M}_x \) and Lemma 2 directly gives \( y_{\text{inter}} = 0 \), i.e., we see that for \( \beta_x \to \infty \) the induced inter cell interference vanishes.

In \( 5, \) we finally see one of the main motivators for defining iaRZF, in the chosen form. Choosing \( \beta_x = 0 \) gives the standard single cell ZF solution; choosing \( \beta_x \to \infty \) gives an intuitively reasonable RZF precoder on projected channels that makes sure no interference is induced in the other cell. It stands to reason that a sum-rate optimal solution can be found as a trade-off between these two extremes, by balancing induced interference and received signal power.

2) **Large-Scale Analysis:** In order to find an appropriate expression of the sum-rate that does not rely on random quantities, we anticipate results from Subsection [C]. There we find a deterministic limit to which the random values of SINR \( x \) almost surely converge, when \( N, K \to \infty \); assuming \( 0 < \varepsilon < \infty \). We can adapt the results from Theorem 1 to fit our current simplified model (by choosing the parameters of the generalized Theorem 1 as \( L = 2, K_x = K, N_x = N, \chi_x = 1, \tau_x = \tau, t_x = 2, \alpha_x = \alpha_x = \alpha_x = \beta_x, \xi = 1, \rho_x = \rho_x \), for \( x \in \{1, 2\} \)). Doing so ultimately results in the following performance indicators \( \text{Sig}_{x} \xrightarrow{\text{a.s.,}\; N,K \to \infty} \text{Sig}_{x} \) and

\[
\text{Int}_{x} \xrightarrow{\text{a.s.,}\; N,K \to \infty} \text{Int}_{x}, \;
\]

where

\[
\text{Sig}_{x} = P_x \left( \frac{\alpha_x \varepsilon e_x}{1 + \alpha_x \varepsilon e_x} \right) \left( \frac{\beta_x \varepsilon e_x}{1 + \beta_x \varepsilon e_x} \right)
\]

\[
\text{Int}_{x} = P_x \left( \frac{1}{1 + \alpha_x \varepsilon e_x} \right) + P_x e \left( \frac{1 + 2 \beta_x \varepsilon e_x}{1 + \beta_x \varepsilon e_x} \right)
\]

\[
\Delta \text{Int}_{x} \xrightarrow{\text{a.s.,}\; N,K \to \infty} \text{Int}_{x} \xrightarrow{\text{a.s.,}\; N,K \to \infty}
\]

\[
e_x = \left( 1 + \frac{\alpha_x \varepsilon e_x}{1 + \alpha_x \varepsilon e_x} + \frac{\beta_x \varepsilon e_x}{1 + \beta_x \varepsilon e_x} \right)^{-1}
\]

Fig. 2. Average user rate vs. transmit power to noise ratio (\( N = 160, K = 40, \varepsilon = 0.7, \rho_1 = \rho_2 = \rho \)).
where $e_x$ is the unique non negative solution to the fixed point equation (7). These expressions are precise in the large-scale regime ($N, K \to \infty$, $0 < K/N < \infty$) and good approximations for finite dimensions. As a consequence of the continuous mapping theorem (e.g., [21]) the above finally implies $\text{SINR}_x \xrightarrow{a.s.} \text{SINR}_x = \text{SINR}_x(\text{Int}_x + 1)^{-1}$.

After realizing that $0 < \lim \inf e_x < \lim \sup e_x < \infty$ for $K, N \to \infty$ (see Lemma[4], 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

$$\lim_{\alpha_x \to \infty} \text{Int}^{\text{BS}_x} = \lim_{\alpha_x \to \infty} P_x c \frac{1}{(1+\alpha_x e_x)^2} = 0$$

$$\lim_{\beta_x \to \infty} \text{Int}^{\text{BS}_x} = \lim_{\beta_x \to \infty} P_x c \frac{\varepsilon}{(1+\beta_x e_x)^2} = 0$$

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

The large system formulation can now also be used to look at the important practical case of mis-estimation of the channels to the other cell’s users. Employing again $0 < \lim \inf e_x < \lim \sup e_x < \infty$ and (6) leads to

$$\lim_{\alpha_x \to \infty} P_x c \frac{1}{(1+\alpha_x e_x)} = 0$$

$$\lim_{\beta_x \to \infty} P_x c \frac{\beta_x^2 + 2\beta_x e_x + \beta_x^{-1} + \varepsilon^2 \tau^2 e_x^2}{(\beta_x^{-1} + e_x)^2} = P_x c \tau^2 \varepsilon$$

i.e., for $\alpha_x \to \infty$ the intra cell interference still vanishes, under the assumption of perfect intra cell CSI. However, for $\beta_x \to \infty$ the induced inter cell interference converges to $P_x c \tau^2 \varepsilon$. Unsurprisingly we see that, due to imperfect CSI, the induced inter cell interference cannot be completely canceled any more.

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, we limit the model to the case $P_1 = P_2 = P$ and $\alpha_1 = \alpha_2 = \alpha$, $\beta_1 = \beta_2 = \beta$. Proceeding similar to the previous subsection, we obtain a formulation for the large-scale approximation of the (now equal) SINR values, when $\alpha \to \infty$. This is denoted $\text{SINR}^{\alpha \to \infty} = \text{SINR}^{\alpha \to \infty} \left(1 + \text{Int}^{\alpha \to \infty}\right)^{-1}$, where

$$\text{SINR}^{\alpha \to \infty} = P \left(1 - \frac{\varepsilon^2 c^2 e_x^2}{(1+\beta_x e_x)^2}\right)$$

$$\text{SINR}^{\alpha \to \infty} = P \tau^2 \frac{1 + 2\beta_x e_x}{(1+\beta_x e_x)^2}$$

and

$$e_{\tau} \equiv \text{SINR}^{\alpha \to \infty} = \left(1 + \frac{\varepsilon^2 c^2 e_x^2}{(1+\beta_x e_x)^2}\right)^{-1}$$

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

$$\frac{\partial \text{SINR}^{\alpha \to \infty}}{\partial \beta} = \frac{P \left(1 - \tau^2\right)}{P \tau^2 + 1} - 1. \quad (9)$$

In other words, in the perfect CSI case ($\tau = 0$), one should choose $\beta$ equal to the transmit power of the BSs. It also shows how one should scale $\beta$ in between the two obvious solutions, i.e., full weight on the interfering channel information for perfect CSI and no weight under random CSI ($\tau = 1$). We remark that the interference channel gain factor $\varepsilon$ is also implicitly included in the precoder. Thus for $\varepsilon \to 0$, we have $\beta \left(\text{tr} \mathbf{G}_1 \mathbf{G}_2\right)^2 \to 0$, while $\beta$ 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 $\beta \to \infty$. The SINR optimal weight for $\alpha$ can be found as (similar to Appendix B-C)

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

Thus, like in the perfect CSI case ($\tau = 0$), one should choose $\alpha$ equal to the transmit power of the BSs. The implications for other limit-cases are not so clear. We see that increasing the transmit power also increases the weight $\alpha$, up to the maximum value of $1/(c \tau^2)$. The weight reduces as the interference worsens, i.e., when $\tau^2, \varepsilon$ grow. This makes sense, as the precoder would give more importance on the interfering channel (by indirectly increasing $\beta$ via normalization). The weight is also reduced, if the cell performance is expected to be bad, i.e., $c$ approaches 1, which makes sense from a sum-rate optimization point of view.

Finally, we can easily calculate the SINR in the limit of both $\alpha$ and $\beta$ independently tending to infinity, as

$$\text{SINR}^{\alpha, \beta \to \infty} = \frac{P \left(1 - 2\varepsilon\right)}{P \tau^2 + 1} \quad (11)$$

The rationale behind all analyses in this section is that optimal weights in the limiting 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 (3) to find an “heuristic interference aware zero-forcing” precoder:

$$\mathbf{M}_{\perp}^{\alpha \perp F} = \mathbf{H}_x \left(\mathbf{H}_x^H \mathbf{H}_x\right)^{-1} - \mathbf{P}_H \mathbf{G}_x$$

$$\times \left(\frac{P \tau^2 + 1}{P \left(1 - \tau^2\right)} \mathbf{I} + \mathbf{G}_x \mathbf{P}_H \mathbf{G}_x\right)^{-1} \mathbf{G}_x^H \mathbf{H}_x \left(\mathbf{H}_x^H \mathbf{H}_x\right)^{-1} \quad (12)$$

To finish this section, we remark that doing these optimizations without the assumption of arbitrarily well estimated intra cell channels, is still an open problem.

4) Graphical Interpretation of the Results: Here we show the implications of the previous subsection on the system performance of our simple model. Of particular interest to us are comparisons of the heuristic scheme with, numerically obtained, 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 (α_{ls}^{opt} and β_{ls}^{opt}) are obtained using 4D grid search, thereby renouncing the α = 0, and β = 0 restrictions from before. Crucially, we see that the performance under (α_{ls}^{opt}, β_{ls}^{opt}) and (α_{ls}^{opt}, β_{ls}^{opt}) is virtually the same. The plot also contains the pair (α_{ls}^{opt}, 0), which corresponds to single cell RZF precoding. The weight α_{ls}^{opt} is again found by grid search. The performance is constant, as the precoder does not take the interfering channel into account. However, we see that the optimally weighted iaRZF is equal to single cell RZF, 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 α = α_{ls}^{opt} → ∞ for all curves and also give the familiar (α_{ls}^{opt}, β_{ls}^{opt} → ∞) curve as a benchmark. Furthermore, the actual value of β_{ls}^{opt} 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) βs 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 for the next section. There we 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 the more general type illustrated in Fig. 5. Each of the L cells consists of one BS associated with a number of single antenna UTs. In more detail, the lth BS is equipped with N_l transmit antennas and serves K_l UTs. We generally set N_l ≥ K_l in order to avoid scheduling complications. We assume transmission on a single narrowband carrier, full transmit-buffers, and universal frequency reuse among the cells.

The lth BS transmits a data symbol vector s_l = [s_{l,1}, ..., s_{l,K_l}]^T intended for its K_l uniquely associated UTs. This BS uses the linear precoding matrix F_{l} \in \mathbb{C}^{N_l \times K_l}, where the columns f_{l,k} \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} ∈ \mathbb{C} at the kth UT in cell l is

\[
y_{l,k} = \sqrt{\gamma_{l,k}} h_{l,k}^H f_{l,k} s_{l,k} + \sum_{k' \neq k} \sqrt{\gamma_{l,k}} h_{l,k'}^H f_{l,k'} s_{l,k'} + \sum_{m \neq l} \sqrt{\gamma_{m,k}} h_{m,k}^H F_{m} s_{m} + n_{l,k}
\]

where n_{l,k} \sim \mathcal{CN}(0,1) is an additive noise term. The transmission symbols are chosen from a Gaussian codebook, i.e., s_{l,k} \sim \mathcal{CN}(0,1). We assume block-wise small scale Rayleigh fading, thus the channel vectors are modeled as h_{l,k} \sim \mathcal{CN}(0, \frac{1}{N_l} I_{N_l}). The path-loss and other large-scale fading effects are incorporated in the \gamma_{l,k} factors. The scaling factor \frac{1}{N_l} 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. Our setting here assumes that interference from other cells dominates w.r.t. other types of rate limitations, such as pilot contamination, which is thus not accounted for in our system model. According to recent works, this assumption is sensible if one considers practical ranges of antennas (on the order of 100s) in systems
with optimized pilot-reuse \textsuperscript{26}, and also when using non-ideal hardware \textsuperscript{27}, Figure 14).

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 \textsuperscript{21} and using channel reciprocity or a frequency-division duplex protocol, where temporal correlation is exploited as in \textsuperscript{23}. 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, or user mobility. To model imperfect CSI without making explicit assumptions on the acquisition process, we employ the generic Gauss-Markov formulation (see, e.g., \textsuperscript{20}, \textsuperscript{29}, \textsuperscript{30}) and we define the estimated channel vectors \( \hat{h}_{l,k}^\gamma \) to be

\[
\hat{h}_{l,k}^\gamma = \sqrt{\lambda_{l,k}^\gamma} \left[ \sqrt{1 - \tau_{l,k}^\gamma} \hat{h}_{l,k}^\tau + \tau_{l,k}^\gamma \hat{h}_{l,k}^\nu \right]
\]

(11)

where \( \hat{h}_{l,k}^\gamma \sim \mathcal{CN}(0, \frac{1}{N_m} I_{N_m}) \) 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 \( \tau_{l,k}^m \in [0, 1] \); a small value for \( \tau_{l,k}^m \) implies a good estimate. Furthermore, we remark that these choices imply \( \hat{h}_{l,k}^\gamma \sim \mathcal{CN}(0, \lambda_{l,k}^\gamma I_{N_m}) \). For convenience later on, we define the aggregated estimated channel matrices as \( \hat{H}_l^m = [\hat{h}_{l,1}^m, \ldots, \hat{h}_{l,K_l}^m] \in \mathbb{C}^{N_m \times K_l} \).

C. iaRZF and Power Constraints

Following the promising results observed in Section \textsuperscript{11}, we continue our analysis of the iaRZF precoding matrices \( F_m^l \), \( m = 1, \ldots, L \), introduced in \textsuperscript{11}. For some derivations, it will turn out to be useful to restate this precoder as

\[
F_m^l = \left( \alpha_m^l \hat{H}_m^l (\hat{H}_m^l)^H + Z^m + \xi_m I_{N_m} \right)^{-1} \hat{H}_m^l \nu_m
\]

where \( Z^m = \sum_{i \neq l} \alpha_i^m \hat{H}_i^m (\hat{H}_i^m)^H \). The \( \alpha_i^m \) can be considered as weights pertaining to the importance one wishes to attribute to the respective estimated channel. We remark that the regularization parameter \( \xi_m \) is usually chosen to be the number of users over the total transmit power \textsuperscript{19} in classical RZF. The factors \( \nu_m \) are used to fulfill the average per UT transmit power constraint \( P_m \textsuperscript{4} \) pertaining to BS \( m \): \( \frac{1}{K_m} \text{tr} \left( F_m^l (F_m^l)^H \right) = P_m \).

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 expected (w.r.t. the transmitted symbols \( s_i^\ell \)) received signal power at the \( k \)th UT of cell \( l \), i.e., \( UT_{l,k} \), is

\[
\mathbb{E}[s_{l,k}^\ell] = h_{l,k}^\ell H_{l,k}^\ell (F_{l,k}^\ell)^H h_{l,k}^\ell + \tau_{l,k}^\ell H_{l,k}^\ell (F_{l,k}^\ell)^H h_{l,k}^\ell \]  

(13)

Similarly, the interference power is

\[
\mathbb{E}[\tau_{l,k}^\ell H_{l,k}^\ell (F_{l,k}^\ell)^H h_{l,k}^\ell]
\]

\[
\tau_{l,k}^\ell H_{l,k}^\ell (F_{l,k}^\ell)^H h_{l,k}^\ell \]  

(14)

where

\[
F_{l,k}^\ell \]  

(15)

and \( H_{l,k}^\ell \) with its \( k \)th column removed. Hence, the SINR at \( UT_{l,k} \) can be expressed as

\[
\text{SINR}_{l,k} = \frac{\mathbb{E}[s_{l,k}^\ell] + \mathbb{E}[\tau_{l,k}^\ell H_{l,k}^\ell (F_{l,k}^\ell)^H h_{l,k}^\ell]}{\mathbb{E}[\tau_{l,k}^\ell H_{l,k}^\ell (F_{l,k}^\ell)^H h_{l,k}^\ell]}
\]

(16)

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 ergodic sum-rate is expressed as

\[
R_{\text{sum}} = \mathbb{E} \sum_{l,k} r_{l,k} = \mathbb{E} \sum_{l,k} \log(1 + \text{SINR}_{l,k})
\]

(17)

where SINRs are random quantities defined by the system model.

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 \textsuperscript{IV-B}). In certain special cases, optimizations of such approximations w.r.t. many performance measures, can be carried out analytically (see for example \textsuperscript{20}). 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 (16). DEs are preferable to standard limit calculations, as they are precise in the limit case, they are also defined for finite dimensions and they provably approach the random quantity for increasing dimensions. Introducing the ratio \( c_i = K_i / N_i \), we make the following technical assumption in order to obtain a DE.

**Assumption A-1.** \( N_i, K_i \to \infty \), such that for all \( i \) we have

\[
0 < \lim \inf c_i \leq \lim \sup c_i < \infty.
\]

This asymptotic regime is denoted \( N \to \infty \) for brevity.
In other words, we require for $N_i$ and $K_i$ to grow large at the same speed. By extending the analytical approach in [20] and [21] to the SINR expression in (16), we obtain a DE of the SINR, which is denoted $\text{SINR}_{l,k}$ in the following.

**Theorem 1** (Deterministic Equivalent of the SINR). Under Assumptions A1 we have

$$\text{SINR}_{l,k} - \text{SINR}_{l,k} \xrightarrow{a.s.} 0$$

Here

$$\text{SINR}_{l,k} = \frac{\text{SINR}_{l,k}(\text{Int}_{l,k} + 1)}{\text{SINR}_{l,k}(\text{Int}_{l,k} + 1)}^{-1}$$

with

$$\text{SINR}_{l,k}(\text{Int}_{l,k} + 1) = m(l) \left(1 + \alpha_m \chi_{l,k}^2 (\nu^2) (y_{l,k}^2)ight)^2$$

and

$$\text{Int}_{l,k} = \sum_{m=1}^{L} \nu_m \left(1 + 2\epsilon_m c_m + \alpha_m \chi_{l,k}^2 (\nu^2) \right) \nu_m g_m (y_{l,k}^2)$$

given $\nu_m = \alpha_m \chi_{l,k}^2 (\nu^2)$. The parameter $\nu_m$, the abbreviations $g_m$ and $y_{l,k}^2$, as well as the corresponding fixed-point equation $\epsilon_m$ and $\epsilon'_m$ are given in the following.

First, we define $\epsilon_m$ to be the unique positive solution of the fixed-point equation

$$\epsilon_m = \left(\xi_m + \frac{1}{N_m} \sum_{j=1}^{K_m} \alpha_m \chi_{m,j}^2 (\nu^2) + \frac{1}{N_m} \sum_{k=1}^{K_i} \alpha_i \chi_{l,k}^2 (\nu^2) \right)^{-1}$$

where $\nu_{l,k}^2 = \left(1 + \alpha_l \chi_{l,k}^2 (\nu^2) \right)^{-1}$. We also have $\nu_m = \frac{P_m K_m}{(N_m g_m)}$ with

$$g_m = \frac{1}{N_m} \sum_{j=1}^{K_m} \chi_{m,j}^2 (\nu^2)$$

and $\epsilon'_m$ can be found directly, once $\epsilon_m$ is known:

$$\epsilon'_m = \left[\frac{1}{N_m} \sum_{j=1}^{K_m} \chi_{m,j}^2 (\nu^2) \right]^{-1}$$

$$+ \frac{1}{N_m} \sum_{j=1}^{K_m} \chi_{m,j}^2 (\nu^2) \left(\nu_m g_m^2 - \epsilon_m^2 \right)$$

**Proof.** See Appendix C.

By employing dominated convergence arguments and the continuous mapping theorem (e.g., [22]), we see that Theorem 1 implies, for each UT $(l,k)$,

$$r_{l,k} = \log_2 \left(1 + \text{SINR}_{l,k} \right) \xrightarrow{a.s.} 0$$

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 Paragraph II-C3) “limit-optimal” iaRZF precoder weights. Furthermore, we provide simulations that corroborate the viability of the proposed precoder, also in systems that are substantially different to the idealized system used in Section II.

**A. Heuristic Generalization of Optimal Weights**

Following from the encouraging sum-rate performance results in Paragraph II-C3, it seems promising to intuitively generalize the heuristic weights to systems with arbitrary many BSs, transmit powers, CSI randomness and user/antenna ratios. Adapting the structures obtained in [20] and (10), we define the general heuristic precoder weights as

$$\tilde{a}_b^\epsilon = \frac{P_a (1 - (\tau_b^a)^2)}{P_a c_b^\epsilon (\tau_b^a)^2 + 1}$$

Here, we introduced the new notation $\epsilon^a_b$, which is the average gain factor between BS $a$ and the UTs of cell $b$, i.e., $\epsilon^a_b = \frac{1}{K_b} \sum_k \chi_{b,k}$. One can intuitively understand by remembering that $\tilde{a}_b^\epsilon$ represents the “importance” of the associated channels (from BSs to UTs in cell $b$). The numerator increases the weight, i.e., makes orthogonality a priority, if the interfering BS uses a large transmit power ($P_a$). Poor channel estimates reduce importance again. The intuitive reason for having $\epsilon^a_b$ in the denominator is not immediately evident, since one would expect to place lower importance on UTs that are very far away. This behavior becomes clear over the realization, that the estimated channels in our model are not normalized (see (11)). Thus, the approximate effective weight of the precoder with respect to a normalized channel is $w^a_b = \tilde{a}_b^\epsilon w^a_b$. Hence, for $\epsilon^a_b \to 0$, we have $w^a_b \to 0$, i.e., no importance is placed on very weak channels. Using the same deliberation for $\epsilon^a_b \to \infty$, we have $w^a_b \to \infty$. Thus no energy is wasted on far away interferers/weak channels, even if one has perfect CSI of those channels.

We remind ourselves that in order to arrive at (20), we assumed $\xi = 1$. Furthermore, systems serving one cluster of closely located UTs per BS, reproduce the initial simplified system closely and, thus, should respond particularly well to the heuristic weights.

**B. Performance**

In order to verify viability of 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 denoted by $\tau_0^1 = \tau_0^2 = \tau_0$ and inter cell wise by $\tau_1^1 = \tau_1^2 = \tau_0$. Given the new more general 2 BS model with randomly distributed UTs for each realization, one has to consider 4 different channel weights ($\alpha_1^a$, $\alpha_2^a$, $\alpha_1^b$, and $\alpha_2^b$) in the whole system. The transmit power to noise ratio (per UT) at each BS is taken to be equal, i.e., $P_1 = P_2 = P$. For this system we obtain the average UT rate performance, shown in Figure 7. The markers denote results of MC simulations that randomize over UT placement scenarios and channel realizations, when the precoding weights are numerically optimal ($\alpha_{opt}^a$, $\alpha_{opt}^b$) per cell parameters. Case (c) refers to the weighted sum of the two cases (a) and (b), which is virtually the same. We also see that even when diverging from the simple system ($\tau_0 = 0$, case a) by choosing $\tau_a = 0.1$, i.e., case b, the heuristic weights still perform practically the same as exhaustive numerical optimization. The same holds true for the more extreme case c, where we chose to modify the numbers of BS antennas, UTs and the CSI quality on a per cell basis ($N_1 = 160$, $N_2 = 60$, $K_1 = 40$, $K_2 = 20$, $P_e = P$, $\tau_0 = 0.1$, $\tau_0 = 0.5$).

Also, we look at a more complex system of 4 BSs (see Figure 8). The BSs, of height 0.1 units, are placed on the corners of a square with edge length 1 units. The UT distribution and pathloss are chosen as before. Figure 8 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 $\tau_a$ for each BS to the adherent UTs, $\tau_b$ for each BS to UTs of BSs 1 unit away and $\tau_c$ for each BS to UTs of BSs $\sqrt{2}$ units away. It is, thus, reasonable to choose $\tau_a < \tau_b < \tau_c$. In the graph we compare the heuristic weights ($\tilde{\alpha}_{ab}^a$) with various other weighting approaches. Round markers stem from Monte-Carlo simulations of the performance pertaining to the heuristic weights DES and confirm the viability of the large scale approximation. The benchmark numeric result in this figure is obtained from optimizing the 16 precoder weights via extensive numerical search, using $\tilde{\alpha}_{ab}^a$ as a starting point. Of practical interest is the performance of the simplified heuristics $\alpha_{ab}^a = P_a(1 - (\tau_{ab}^a)^2)$. This choice means that no interference is taken into account, i.e., $\alpha_{ab}^a = 0$. We observe that most of the viability of the heuristic method comes from this part. Only at low SNR, where interference is the dominant problem, and very high SNR does the $P_a(1 - (\tau_{ab}^a)^2)$ approach become noticeably suboptimal. The constant weight approach ($\alpha_{ab}^a = 0.1$) behaves like in Section II in that it is only a good match for a limited part of the curve. For comparison purposes, we also compare with standard non-cooperative RZF, as defined in Subsection II-B.

In general, employing $\tilde{\alpha}_{ab}^a$ is most advantageous in high interference scenarios, as would be expected due to the “interference aware” conception of the precoder. 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 $\tilde{\alpha}_{ab}^a$. We also note that better CSI widens the gap between the $\tilde{\alpha}_{ab}^a$ and $\alpha_{ab}^a = P_a(1 - (\tau_{ab}^a)^2)$ weighted iaRZF approaches.
V. CONCLUSION

In this paper, we analyzed a linear precoder structure for multi cell systems, based on an intuitive interference induction trade-off and recent results on multi cell RZF, 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 mitigation. The effectiveness of these large-scale results has been demonstrated in practical finite dimensional systems.

Generally one denotes \( P_X \) as the projection matrix onto the column space of \( X \) and \( P_X^\perp \) as the projection matrix onto the orthogonal space of the column space of \( X \).

**Definition 1** (Notation of Resolvents). Given the notations from Section III we define resolvent matrices of \( H_n^a \) as:

\[
Q_a \triangleq \left( \alpha_a \hat{H}_n^a (\hat{H}_n^a)^H + Z_n + \xi_a I_N \right)^{-1}.
\]

We will also use of the following modified versions

\[
Q_{a[bc]} \triangleq \left( \alpha_a \hat{H}_{n[bc]}^a (\hat{H}_{n[bc]}^a)^H + Z_n + \xi_a I_N \right)^{-1},
\]

\[
Q_{a[b]} \triangleq \left( \alpha_a \hat{H}_{n[b]}^a (\hat{H}_{n[b]}^a)^H + Z_n + \xi_a I_N \right)^{-1},
\]

**Lemma 3** (Matrix Inversion Lemma [32 Lemma 2.2]). Let \( A \) be an \( M \times M \) invertible matrix and \( x \in \mathbb{C}^M, c \in \mathbb{C} \) for which \( A + cxx^H \) is invertible. Then, as an application of (21), we have

\[
x^H (A + cxx^H)^{-1} = \frac{x^H A^{-1}}{1 + cxx^H A^{-1} x}.
\]

For the previously defined resolvent matrices, we have in particular

\[
Q_{a[b]} = \frac{Q_{a[bc]} h_{a[b]}^H}{1 + \alpha_a^2 (h_{a[b]}^H)^2}.
\]

**Lemma 4** (Convergence of Quadratic Forms [33]). Let \( x_M = [X_1, \ldots, X_M]^T \) be an \( M \times 1 \) vector where the \( X_n \) are i.i.d. Gaussian random variables with unit variance. Let \( A_M \) be an \( M \times M \) matrix independent of \( X_M \). If in addition \( \limsup_M \|A\|_2 < \infty \) then we have that

\[
\frac{1}{M} x_M^H A_M x_M - \frac{1}{M} \text{tr} (A_M) \xrightarrow{a.s.} 0.
\]

**Corollary 1.** Let \( A_M \) be as in Lemma 4 i.e., \( \limsup_M \|A\|_2 < \infty \), and \( x_M, y_M \) be random, mutually independent with complex Gaussian entries of zero mean and variance 1. Then we have

\[
\frac{1}{M} y_M^H A_M x_M \xrightarrow{a.s.} 0.
\]

**Lemma 5.** [Rank-One Perturbation Lemma [24 Lemma 14.3]] Let \( Q_a \) and \( Q_{a[bc]} \) be the resolvent matrices as defined in Definition 1. Then, for any matrix \( A \) we have:

\[
\text{tr} \left[ A (Q_a - Q_{a[bc]}) \right] \leq \frac{1}{\xi_a} \|A\|_2.
\]

**APPENDIX B**

**SIMPLE SYSTEM LIMIT BEHAVIOR PROOFS**

In this section, we provide the proofs pertaining to the limit behavior of the simple system in Section II.

**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 \( \mathbf{F} = \sqrt{K_M} / \sqrt{\text{tr} \mathbf{MM}} \) for \( M = (\alpha H^H + \beta G^H + \xi I)^{-1} \).
1) $\beta \to \infty$: For the limit when $\beta \to \infty$ we use \ref{21} with $A = \beta GG^H + \xi I$ and $CBC^H = H\alpha HH^H$ to reformulate the matrix $M$

$$M = (\alpha HH^H + \beta GG^H + \xi I)^{-1} H = \left[ QG - QG H (\alpha^{-1} I + H^H QG H)^{-1} H^H QG \right] H$$

where

$$QG = (\beta GG^H + \xi I)^{-1} \xi^{-1} I - \xi^{-1} G \left( \frac{\xi}{\beta} I + G^H G \right)^{-1} G^H.$$ 

We now let $\beta \to \infty$, assuming $G^H G$ is invertible (which is true with probability 1) and $\xi$ bounded. In this regime, we remember Lemma \ref{2} and rewrite $QG = \xi^{-1} P_G$. One finally arrives at

$$M \xrightarrow{\beta \to \infty} \left[ \xi^{-1} P_G^H - \xi^{-2} P_G^H (\alpha^{-1} I + \xi^{-1} H^H P_G H)^{-1} H^H P_G^H \right] H.$$

Relying further on properties of projection matrices ($P_G^H = P_H^H P_G$, $(P_G^H)^H = P_G$) and introducing the matrix $H = P_G H$, as the channel matrix $H$ projected on the space orthogonal to the channels of $G$, we get

$$M \xrightarrow{\beta \to \infty} \left[ \xi^{-1} P_G^H - \xi^{-2} P_G^H (\alpha^{-1} I + \xi^{-1} H^H P_G H)^{-1} H^H P_G^H \right] H = H \left( \xi I + \alpha H^H H \right)^{-1}.$$

2) $\alpha \to \infty$: Introducing the abbreviations $Q_H = (H^H + \frac{\xi}{\alpha} I)^{-1}$ and $Q_H = (H^H H + \frac{\xi}{\alpha} I)^{-1}$, we can rewrite the matrix $M$ as follows.

$$\alpha M = \left( HH^H + \frac{\beta}{\alpha} GG^H + \frac{\xi}{\alpha} I \right)^{-1} H$$

$$\xrightarrow{\alpha \to \infty} \left[ QH - QH G \left( \frac{\alpha}{\beta} I + G^H QH G \right)^{-1} G^H QH \right] H$$

$$\xrightarrow{\alpha \to \infty} H QH - QH G \left( \frac{\alpha}{\beta} I + G^H QH G \right)^{-1} G^H QH.$$

Applying \ref{23} to the expression $\left( HH^H + \frac{\xi}{\alpha} I \right)^{-1} + \left( -\frac{\xi}{\alpha} I \right)^{-1}$, one eventually finds the relationship $QH = \alpha \xi^{-1} (I - H QH H^H)$. Hence,

$$\alpha M \xrightarrow{\alpha \to \infty} H QH - \xi^{-1} (I - H QH H^H) \times G \left[ \frac{1}{\beta} I + \xi^{-1} G^H (I - H QH H^H) G \right]^{-1} G^H QH.$$

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

$$\alpha M \xrightarrow{\alpha \to \infty} (H^H H)^{-1} - \xi^{-1} (I - H (H^H H)^{-1} H^H) G$$

$$\xrightarrow{\beta \to \infty} \left[ \beta^{-1} I + \xi^{-1} G^H \left( I - H (H^H H)^{-1} H^H \right) G \right]^{-1}$$

$$\xrightarrow{\alpha \to \infty} H (H^H H)^{-1}$$

$$\xrightarrow{\beta \to \infty} \left( \beta^{-1} I + \xi^{-1} G^H P_H G \right)^{-1} G^H (H^H H)^{-1}.$$

B. Large-Scale Approximation

In this subsection, we primarily show that the fixed point equation $e$ is bounded in the sense of $0 < \lim \text{inf} e < \lim \text{sup} e < \infty$. This knowledge simplifies the limit calculations in Subsection \ref{1C} to simple operations. We remind ourselves, that for perfect and imperfect CSI the resulting fixed point equations are equivalent:

$$e = \left( 1 + \frac{c}{\alpha^2 + \epsilon^2} + \frac{c \beta e}{\alpha + \beta \epsilon e} \right)^{-1}$$

with $e$ where we abbreviated $\epsilon_\alpha$ with $e$ for notational convenience.

**Lemma 6** ($e$ is Bounded). For either $\alpha \to \infty$ and $\beta, \epsilon$ bounded or $\beta \to \infty$ and $\alpha, \epsilon$ bounded, we have

$$0 < \lim \text{inf} e < \lim \text{sup} e < \infty.$$

**Proof.** 1) $e < \infty$ when $\alpha$ or $\beta \to \infty$, follows immediately from contradiction, when one takes $e \to \infty$ in \ref{24}.

2) To see that $e$ positive when $\alpha$ or $\beta \to \infty$, we take either $\alpha \to \infty$ and $\beta, \epsilon$ bounded or $\beta \to \infty$ and $\alpha, \epsilon$ bounded. For the case $\alpha \to \infty$, we first denote $v = \epsilon \alpha$ and we look at

$$v = \left( \frac{1}{\alpha + \epsilon} + \frac{c \beta \epsilon}{\alpha + \beta \epsilon \epsilon} \right)^{-1}.$$

Now assume $v$ to be bounded for $\alpha \to \infty$

$$v = \lim_{\alpha \to \infty} \left( \frac{1}{\alpha + \epsilon} + \frac{c \beta \epsilon}{\alpha + \beta \epsilon \epsilon} \right)^{-1} = \left( \frac{1}{\alpha + \epsilon} \right)^{-1}$$

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

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

Continuing from Appendix \ref{B} we see that in the limit $\alpha \to \infty$ the large-scale approximation of the SINR values pertaining to the users of each cell, i.e., $\text{SINR} \xrightarrow{\alpha \to \infty}$, is indeed as stated in Paragraph \ref{1C}.

Differentiating $\text{SINR}$ w.r.t. $\beta$, while taking into account that $e$ is an abbreviation for $\epsilon_\beta \to \infty$ leads us to

$$\frac{\partial \text{SINR}}{\partial \beta} = -2 P e e^2 [e + \beta e^2]$$

$$\times \frac{t_1}{[P (e \beta^2 e^2 e^4 + 2 \epsilon \beta e^2 + 2 \epsilon^2 e^2 + 2 \beta e^2 + 1)^2}.$$
where we used $e'$ as shorthand for $\frac{\partial e^{-\beta t}}{\partial t}$ and 
\[ t_1 = \frac{P[1 - 2e\beta e^{\beta}] + \beta e^{\beta} e^2}{P - P e^{2\beta e^{\beta}}} \]

Realizing that the denominator of (25) cannot become zero, we have two possible solutions for $\beta$ \text{SINR} - r^{\infty} / \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 $c(1 - \beta e^{\beta} + 2\beta e^2 e^2)$, and similarly 
\[ P[1 - 2e\beta e^{\beta}] + \beta e^{\beta} e^2 = -Pe - P e^{2e^{2\beta}}. \]

Hence, 
\[ t_1 = \left( e^{2e^2} + P^2 e^{2e^2} \right) \left( \beta - \frac{P(1 - \tau^2)}{Pe^{2\tau^2}} \right) \left( \beta + \frac{1}{e} \right) \]

Given that the only middle term can become zero, we find $\beta_{\text{opt}}$ to be 
\[ \beta_{\text{opt}} = \frac{P(1 - \tau^2)}{Pe^{2\tau^2} + 1} \] (26)

as stated in (9). The physical interpretation of the SINR guarantees this point to be the maximum. We used the assumption $e + \beta e' > 0$ to arrive at the previous result. This claim is proved by the following lemma.

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

**Proof Sketch.** From (32) we know that an object of the form 
\[ m(z) = \left[ -z + c \int \frac{t}{1 + tm(z)} \ dt \right]^{-1} \]

where $\nu$ is a non negative finite measure, is a so-called Stieltjes transform of a measure $\nu$, defined $\forall z \notin \text{Supp}(\nu)$. Adapting (24) by re-naming $e = \beta e \nu$, 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. \hfill \Box

**APPENDIX C**

**PROOF OF THEOREM 1**

The objective of this section is to find a DE for the SINR term (19). A broad outline of the required steps is as follows. In the beginning of the proof we condition that $Z^m$ is fixed to some realization and we follow the steps given in Appendix II for the power normalization $\nu_m$. Invoking Theorem 1 we obtain the fundamental equations for $e_m$. We then, allow $Z^m$ to be random and apply Theorem 3.13 to obtain (17). Invoking Tonelli’s theorem, it is admissible to apply the two theorems one after the other, as $Z^m$ is a bounded sequence with probability one. The DEs of all required terms are found by following Appendix II again. This is true for the terms from Subsection II-D as well. However here the interference terms ask for a slightly more generalized version of (20 Lemma 7).

A. Power Normalization Term

We start by finding a DE of the term $\nu_m$, which will turn out to be a frequently reoccurring object throughout this Section. From [22], we see that the power normalization term $\nu_m$ is defined by the relationship

\[ \frac{P_m K_m}{\nu_m N_m} - \frac{1}{N_m} \text{tr} \left[ \hat{H}_m^m \hat{H}_{m,m}^m Q_m^2 \right] - \frac{e}{\nu_m} \] \[ = \frac{\partial}{\partial \xi_m} \left( \frac{1}{\alpha_m^{\nu_m}} \text{tr} \left[ (Z^m + \xi_m I_{N_m}) Q_m \right] \right) \] (27)

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

\[ \frac{1}{N_m} \text{tr} \left[ \hat{H}_m^m \hat{H}_{m,m}^m Q_m^2 \right] - \tilde{X}_m \xrightarrow{a.s.} N \to \infty 0 \]

for the regime defined in A-1.

To do this, we apply [20 Theorem 1] to (27), where we set the respective variables to be $Y_m = \chi_{m,m}^m$, $Q_N = Z^m + \xi_m I_{N_m}$, $B_N = \alpha_m^m \hat{H}_m^m \hat{H}_{m,m}^m + Z^m$ and $z = -\xi_m$. Thus, we find the (partially deterministic) quantity

\[ \tilde{X}_m = \frac{1}{N_m} \text{tr} \left[ (Z^m + \xi_m I_{N_m}) \right] \left( \frac{1}{N_m} \sum_{j=1}^{K_m} \alpha_m^{\nu_m} \chi_{m,m}^m I_{N_m} + Z^m + \xi_m I_{N_m} \right)^{-1} \]

where $e_m^m = \alpha_m^{\nu_m} \chi_{m,m}^m e_m$ and

\[ e_m = \frac{1}{N_m} \text{tr} \left( 1 + \alpha_m^{\nu_m} \chi_{m,m}^m e_m + Z^m + \xi_m I_{N_m} \right)^{-1} \]

**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 $e_m$.

\[ \frac{1}{N_m} \text{tr} Q_m - e_m \xrightarrow{a.s.} N \to \infty 0. \] (28)

This can be quickly verified by using [20 Theorem 1], via choosing $R_m = \chi_{m,m}^m I$, $D_N = I$, $B_N = \alpha_m^m \hat{H}_m^m \hat{H}_{m,m}^m + Z^m$ and $z = -\xi_m$.

One notices, that the fixed-point equation $e_m$ contains the term $Z^m$, which is not deterministic. Thus, the derived objects are not yet DEs. In order to resolve this we need to condition $Z^m$ 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:

\[ e_m = \frac{1}{N_m} \text{tr} \left[ (Z^m + \beta_m I_{N_m})^{-1} \right] \]

where

\[ \beta_m = \frac{1}{N_m} \sum_{j=1}^{K_m} \alpha_m^{\nu_m} \chi_{m,m}^m e_m + \xi_m. \] (29)
The solution becomes immediate once we rephrase $Z^m$ as

$$Z^m = \sum_{l \neq m}^{K_l} \alpha_m h_{l,k}(h_{l,k})^H = H_{[m]}^m A_{[m]}^m (H_{[m]}^m)^H$$

where $H_{[m]}^m \in \mathbb{C}^{N_m \times K_m}$, with $K_m = \sum_{l \neq m} K_l$, is the aggregated matrix of the vectors $h_{l,k}^m \in \mathcal{CN}(0, \frac{1}{N_m} I_{N_m})$, \forall $l \neq m$ and

$$A_{[m]}^m = \text{diag}\left[\alpha_m^1 \chi_{l,k_1}, \ldots, \alpha_m^L \chi_{l,k_L}\right]$$

i.e., a diagonal matrix with the terms pertaining to $\alpha_m^m$ removed.

One can directly apply (32) or (33) (Theorem 3.13, Eq 3.23) with $T = A_{[m]}^m$ and $X = (H_{[m]}^m)^H$. Being careful with the notation $(XTX)^H$ instead of $(H_{[m]}^m)^H A_{[m]}^m H_{[m]}^m$), we arrive at:

$$e_m = \frac{1}{N_m} \text{tr} \left\{ \left[ H_{[m]}^m A_{[m]}^m (H_{[m]}^m)^H + \beta_m I_{N_m} \right]^{-1} \right\}$$

where

$$e_m = \frac{1}{N_m} \left[ \frac{1}{N_m} \sum_{l \neq m}^{K_l} \alpha_m^m \chi_{l,k,l}^m \frac{\beta_m}{1 + \alpha_m^m \chi_{l,k,l}^m} \right]^{-1} \xrightarrow{\text{a.s.}} 0 \quad N \to \infty .$$

Here we used Remark 2 and $\beta_m$ is given in (29).

Combining the intermediate results, using Remark 2 and the relationship $\text{tr} A (A + x I)^{-1} = \text{tr} 1 - x \text{tr} (A + x I)^{-1}$ with $A = Z^m + \xi_{m} I_{N_m}$, we arrive at

$$\bar{X}_m = -\frac{1}{\alpha_m^m N_m} \sum_{l \neq m}^{K_l} \frac{\alpha_m^m \chi_{l,k,l}^m e_m}{(1 + \alpha_m^m \chi_{l,k,l}^m e_m)^2}$$

where $e_m$ is shorthand for $\partial / \partial \xi_m e_m$ and can found (by prolonged calculus) to be as stated in (18), which concludes this part of the proof.

B. Signal Power Term

The important part of finding the DE of the signal power term (13) to find a DE of $(h_{l,k}^l)^H Q_{l,k} h_{l,k}^l$, which will now first be done. Before proceeding, we remind ourselves that our chosen model of the estimated channel (11) entails the following relationships: $h_{l,k}^l \perp h_{l,k}^k, h_{l,k}^l \perp h_{l,k}^l, Q_{l,k} \perp h_{l,k}^l, Q_{l,k} \perp h_{l,k}^l$. Also, formulations containing $h_{l,k}^l$ can often be split into two terms comprising $h_{l,k}^l$ and $h_{l,k}^l$. Hence, the application of Lemmas 3 [4] 5 and Corollary 1 in the following is well justified. Employing (29) one sees

$$(h_{l,k}^l)^H Q_{l,k} h_{l,k}^l \leq \frac{\chi_{l,k}^l \sqrt{(1 - \tau_l^2)^2 e_l (1)}}{1 + \alpha_l^l \chi_{l,k}^l e_l (1)} \xrightarrow{\text{a.s.}} 0 \quad N \to \infty .$$

Finally, applying this result to the complete formulation (13), we arrive at the familiar term from Theorem 1:

$$\sum_{l,k} (i) \chi_{l,k}^l \sqrt{(1 - \tau_l^2)^2 e_l (1)} \left( 1 - \tau_l^2 \right) \left( f_{l,k}^l \right)^2 .$$

C. Preparation for Interference Terms

In this subsection we derive the deterministic equivalents of the two terms $(h_{l,k}^l)^H B Q_{l,k} h_{l,k}^l$ and $(h_{l,k}^l)^H B Q_{l,k} h_{l,k}^l$ where $B \in \mathbb{C}^{N_l \times N_l}$ has uniformly bounded spectral norm w.r.t. $N_l$ and is independent of $h_{l,k}^l$ and $h_{l,k}^l$. The following approach is based on and slightly generalizes [20] Lemma 7. First, it is helpful to realize an implication of our resolvent notation (Definition 1) and channel estimation model (11):

$$Q_{a}^{-1} - Q_{a}^{-1} = c_0 h_{b,c}^a (h_{b,c}^a)^H + c_0 b_{b,c}^a (h_{b,c}^a)^H + c_0 h_{b,c}^a (h_{b,c}^a)^H + c_0 b_{b,c}^a (h_{b,c}^a)^H .$$

$$\left( h_{l,k}^l \right)^H B Q_{l,k} h_{l,k}^l \leq \left( h_{l,k}^l \right)^H B Q_{l,k} h_{l,k}^l \leq \left( h_{l,k}^l \right)^H B Q_{l,k} h_{l,k}^l .

Finally, we begin with the term $(h_{l,k}^l)^H B Q_{l,k} h_{l,k}^l$:

$$\left( h_{l,k}^l \right)^H B Q_{l,k} h_{l,k}^l \leq \left( h_{l,k}^l \right)^H B Q_{l,k} h_{l,k}^l \leq \left( h_{l,k}^l \right)^H B Q_{l,k} h_{l,k}^l .

and, using (30), we find

$$h_{l,k}^l \right)^H B Q_{l,k} h_{l,k}^l = \frac{Y_3 - \left( h_{l,k}^l \right)^H B Q_{l,k} h_{l,k}^l (c_0 Y_2 + c_2 Y_5)}{1 + c_2 Y_2 + c_1 Y_5} .$$

Similarly, for the term $(h_{l,k}^l)^H B Q_{l,k} h_{l,k}$ we arrive at

$$\left( h_{l,k}^l \right)^H B Q_{l,k} h_{l,k}^l (1 + c_0 Y_6 + c_2 Y_1) \leq Y_4 - \left( h_{l,k}^l \right)^H B Q_{l,k} h_{l,k}^l (c_2 Y_5 + c_1 Y_1) .

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

$$\left( h_{l,k}^l \right)^H B Q_{l,k} h_{l,k}^l \left( 1 + c_0 Y_6 + c_2 Y_1 \right) = Y_4 - \left( h_{l,k}^l \right)^H B Q_{l,k} h_{l,k}^l (c_2 Y_5 + c_1 Y_1) .

(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:

$$\left( h_{l,k}^l \right)^H B Q_{l,k} h_{l,k}^l \left( 1 + c_0 Y_6 + c_2 Y_1 \right) = Y_4 - \left( h_{l,k}^l \right)^H B Q_{l,k} h_{l,k}^l (c_2 Y_5 + c_1 Y_1) .

where the values for $u'$, $c_0$ and $c_2$ are not yet of concern. Thus, (33) finally leads to

$$\left( h_{l,k}^l \right)^H B Q_{l,k} h_{l,k}^l \left( 1 + c_0 Y_6 + c_2 Y_1 \right) = Y_4 - \left( h_{l,k}^l \right)^H B Q_{l,k} h_{l,k}^l (c_2 Y_5 + c_1 Y_1) .

\left( h_{l,k}^l \right)^H B Q_{l,k} h_{l,k}^l \left( 1 + c_0 Y_6 + c_2 Y_1 \right) = Y_4 - \left( h_{l,k}^l \right)^H B Q_{l,k} h_{l,k}^l (c_2 Y_5 + c_1 Y_1) .$$
and we finally find the expression we were looking for

$$\left(h_{l,k}^h\right)^{\text{th}}BQ_{\text{h}_l^l} = \frac{u'(1+c_1u_1)}{1+c_1u_1+c_0u_2+(c_0c_1-c^2_2)u_1u_2} \xrightarrow{\text{a.s.}} 0. \quad (34)$$

In order to find the second original term \(\left(h_{l,k}^h\right)^{\text{th}}BQ_{\text{h}_l^l}\), we reform and plug (32) into (31) and follow analogously the path we took to arrive at (34). We finally find

$$\left(h_{l,k}^h\right)^{\text{th}}BQ_{\text{h}_l^l} = \frac{-c_2u_1u'}{1+c_1u_1+c_0u_2+(c_0c_1-c^2_2)u_1u_2} \xrightarrow{\text{a.s.}} 0. \quad (35)$$

### D. Interference Power Terms

Having obtained the preparation results in Appendix C, we can now continue to find the DEs for different parts of the interference power term. From (13) we arrive at

$$\text{Int}_{t}^{(l)} = \sum_{m \neq l} \nu_{m} x_{l,k}^{m} \left(h_{l,k}^h\right)^{\text{th}}Q_{\text{m},\text{h}_l^l} \left(\hat{H}_{l,k}^{\text{h}}\right)^{\text{th}}Q_{l,k} \hat{h}_{l,k}^{h} \xrightarrow{\text{Part A}} \text{Part B} \quad (36)$$

We start by treating (36). Part B first. Employing the relationships ABD = ACD + A (B - C) and (25) one finds

- **Part B** = \(\left(h_{l,k}^h\right)^{\text{th}}Q_{l,k} \hat{H}_{l,k}^{\text{h}} \left(\hat{H}_{l,k}^{\text{h}}\right)^{\text{th}}Q_{l,k} \hat{h}_{l,k}^{h}\)

Using the relationship (30) pertaining to \(Q_{l}^{-1} - Q_{l}^{-1}\), we can split Part B as

**Part B** = \(X_1 - c_0X_3X_1 - c_2X_3X_2 - c_2X_4X_1 - c_1X_4X_2\).

Where we have found and abbreviated the 4 quadratic forms,

- **X_1** = \(\left(h_{l,k}^h\right)^{\text{th}}Q_{l,k} \hat{H}_{l,k}^{\text{h}} \left(\hat{H}_{l,k}^{\text{h}}\right)^{\text{th}}Q_{l,k} \hat{h}_{l,k}^{h}\)

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

\(u_1 = \left(h_{l,k}^h\right)^{\text{th}}Q_{\text{h}_l^l} \hat{h}_{l,k}^{h} \Rightarrow u_1 - c_1 \xrightarrow{\text{a.s.}} 0\).

Analogously,

\(u_2 - c_1 \xrightarrow{\text{a.s.}} 0\).

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

\(\left(h_{l,k}^h\right)^{\text{th}}Q_{\text{h}_l^l} \hat{H}_{l,k}^{\text{h}} \left(\hat{H}_{l,k}^{\text{h}}\right)^{\text{th}}Q_{\text{h}_l^l} \hat{h}_{l,k}^{h} \Rightarrow u' = g_1 \xrightarrow{\text{a.s.}} 0\).

where the last step makes have use of the results in Appendix C-A. Also, we remind ourselves that we have

\(c_0 = \frac{\alpha_l c_l (1 - (\tau^l_{1})^2)}{2}, c_1 = \frac{\alpha_l c_l (\tau^l_{1})^2}{2}\) and \(c_2 = \frac{\alpha_l c_l (1 - (\tau^l_{1})^2)^2}{2},\) hence \(c_0 + c_1 = c_1\) and \(c_0c_1 - c_2 = 0\). So, finally, we have

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

and similarly

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

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

\(X_3 = u' \frac{(1+c_1u)}{1+(c_1+c_0)u} \xrightarrow{\text{a.s.}} 0\)

\(X_4 = -u' \frac{(1+c_1u)}{1+(c_1+c_0)u} \xrightarrow{\text{a.s.}} 0\).

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

\(\text{Part B} = gi \frac{1 - (\tau^l_{1})^2}{\left(1 + \alpha_l c_l \epsilon_1\right)^2} + g(\tau^l_{1})^2\).

The next step is to derive the DE of (56) Part \(\text{A}_m\), i.e., \(\text{Part A}_m\). Fortunately, the sum obliges \(m \neq l\) and, thus, the same derivation like for Part B applies. Hence, we arrive at

\(\text{Part A}_m = \frac{1 - (\tau^l_{m})^2}{\left(1 + \alpha_l c_l \epsilon_1 m\right)^2} + g_m(\tau^l_{m})^2\).

Combing Part B and the sum of Part \(\text{A}_m\) with our original expression of the interference power, we arrive at the familiar expression from Theorem 1.


Sebastian Wagner received the M.Sc. degree in Electrical Engineering from the Technische Universität Dresden, Germany in 2008 and the Ph.D. degree from TELECOM ParisTech (EURECOM), Sophia Antipolis, France in 2011. From 2008 to 2011, he was an research engineer at ST-ERICSSON, Sophia Antipolis, France, where he worked toward the Ph.D. degree. In 2012 he joined EURECOM as a post-doctoral fellow to work on the OpenAirInterface project. Since 2013 he is with Intel Mobile Communications developing and implementing algorithms for LTE products. His current research interests include signal processing for multiple antenna systems and the application of random matrix theory to wireless communications.

Mérouane Debbah entered the Ecole Normale Supérieure de Cachan (France) in 1996 where he received his M.Sc and Ph.D. degrees respectively. He worked for Motorola Labs (Saclay, France) from 1999-2002 and the Vienna Research Center for Telecommunications (Vienna, Austria) until 2003. From 2003 to 2007, he joined the Mobile Communications department of the Institut Eurecom (Sophia Antipolis, France) as an Assistant Professor. Since 2007, he is a Full Professor at Supélec (Gif-sur-Yvette, France). From 2007 to 2014, he was director of the Alcatel-Lucent Chair on Flexible Radio. Since 2014, he is Vice-President of the Huawei France R&D center and director of the Mathematical and Algorithmic Sciences Lab. His research interests are in information theory, signal processing and wireless communications. He is an Associate Editor in Chief of the journal Random Matrix Theory and Applications and was an associate and senior area editor for IEEE Transactions on Signal Processing respectively in 2011-2013 and 2013-2014. Mérouane Debbah is a recipient of the ERC grant MORE (Advanced Mathematical Tools for Complex Network Engineering). He is an IEEE Fellow, a WWRF Fellow and a member of the academic senate of Paris-Saclay. He is the recipient of the Mario Boella award in 2005, the 2007 IEEE GLOBECOM best paper award, the Wi-Opt 2009 best paper award, the 2010 Newcom++ best paper award, the WUN CogCom Best Paper 2012 and 2013 Award, the 2014 WCNC best paper award as well as the Valuetools 2007, Valuetools 2008, CrownCom2009, Valuetools 2012 and SAM 2014 best student paper awards. In 2011, he received the IEEE Glavieux Prize Award and in 2012, the Qualcomm Innovation Prize Award.