Boriana Boiadjieva, Hussein Al-Shatri and Anja Klein, ”Hierarchical Beamforming in CRAN Using Random Matrix Theory,” in *Proc. of the IEEE International Conference on Communications - (IEEE ICC 2018)*, Kansas City, MO, USA, May 2018.

Hierarchical Beamforming in CRAN Using Random Matrix Theory

Boriana Boiadjieva, Hussein Al-Shatri and Anja Klein
Communications Engineering Lab
Technische Universität Darmstadt, Merckstrasse 25, 64283 Darmstadt, Germany
{b.boiadjieva, h.shatri, a.klein}@nt.tu-darmstadt.de

***Abstract*—In this paper, we propose a hierarchical beamforming approach for the downlink of a multicellular multiuser system with cloud radio access network (CRAN) which requires reduced fronthaul transmissions while benefiting from the centralized functionality of the cloud. We design the hierarchical beamforming as a concatenation of two beamformers. The first one is called inner beamformer and it is designed at the base station (BS) based on the instantaneous channel state information (CSI) of its own users. For this beamformer, the BSs apply regularized zero-forcing (RZF). The second one is called outer beamformer and it is defined at the cloud based on the global but only statistical CSI. All outer beamformers are designed together at the cloud to allow coordination among the BSs. This hierarchical approach provides a combination of distributed and centralized precoding and so it manages the intra-cell as well as the inter-cell interference. Because the cloud has only statistical channel knowledge, we apply random matrix theory to obtain deterministic approximations of the useful power, intra-cell and inter-cell interference power at every user. These approximations are closed-form expressions and allow the cloud to optimize the outer beamformers for diverse objectives. In this work, we propose a low complexity iterative outer beamformer design based on block diagonalization which maximizes the system sum rate. Simulations show that the deterministic approximations are tight and that the proposed hierarchical beamforming achieves a higher sum rate compared to the conventional distributed RZF.**

I. Introduction

Future mobile cellular networks will experience a huge growth in traffic while satisfying high requirements for latency and reliability. The forecast suggests that the mobile data traffic will grow at a compound annual growth rate of 46 percent between 2016 and 2021, and so it will reach 48.3 exabytes per month by 2021 [1]. Additionally, the Machine-to-Machine and Internet-of-Things devices will introduce the necessity for real-time processing of a huge amount of data and for the development of massive connectivity strategies requiring densified cells and large antenna arrays at the base stations (BSs). Therefore, the new promising architecture cloud radio access network (CRAN) has been proposed to provide centralization and coordination among the BSs. Through CRAN, the baseband processing will benefit from softwarization and flexible functional splits providing advanced coordination.

However, the coordination of densely deployed BSs, equipped with multiple antennas and serving many devices, is associated with a lot of signaling over the fronthaul links connecting the BSs with the cloud. Since the fronthaul is capacity constrained and time-delay constrained [2], the big amount of signaling might lead to unacceptable delays for the real-time processing. These fronthaul constraints become very critical for coordination techniques like the coordinated beamforming where the availability of channel state information (CSI) is a crucial factor defining the achievable network performance.

To address this challenge, different proposals can be found in the literature suggesting compressive CSI acquisition, fronthaul compression strategies or sparse precoding, see [3] and references therein. In [4] for instance, a new CSI acquisition has been proposed to reduce the CSI overhead and to design a stochastic beamforming. In [5] and [6], optimization algorithms for different fronthaul compression and beamforming strategies have been studied. In [7], we propose a coordinated hierarchical beamforming where the cloud has knowledge only of averaged channel link qualities and defines a transmission subspace size for every BS. To achieve high system performance while utilizing low complexity precoders, in this work, we let the cloud define not only the subspace sizes, but also the transmission subspaces for every BS.

In this paper, we propose a coordinated beamforming for the downlink of multicellular multiuser multiple input multiple output (MIMO) networks with CRAN which requires only statistical CSI. To achieve this, we design the precoder as concatenation of two beamformers. The first, so-called inner beamformer is designed at the BS based on the instantaneous CSI of its own users and can manage the intra-cell interference. The second, so-called outer beamformer is designed centrally together with all outer beamformers at the cloud. The outer beamformer adapts to the channel statistics, controls the inter-cell interference and defines the transmission subspace of the BS. Since the cloud has only statistical channel knowledge, which involves operations with large random matrices, we apply theorems from random matrix theory (RMT).

RMT provides deterministic approximations in systems with high dimensional random processes. Because the radio channel is often modeled as a large random matrix, theorems from RMT have already found application in diverse MIMO scenarios. For example, in [8] the authors consider a single cell scenario with a closed-form precoder and derive the deterministic equivalent of the signal to interference and noise ratio (SINR) at every user to solve various optimization problems like optimal number of users for zero-forcing precoding and the optimal regularization parameter for regularized zero-forcing (RZF). In [9], the authors define the SINR approximations at the users in a multicell network with RZF beamformers where all BSs serve all users simultaneously by exchanging data and signaling over backhaul links. Another related work is [10], where the precoder is split into inner and outer beamformers so that all inner beamformers perform RZF at the BSs while the outer beamformers are designed at the cloud according to an approximate optimization problem which maximizes a concave utility. To obtain the approximated problem, the authors assume that the inter-cell interference is zero and use deterministic approximations of the data rate for a single cell. In contrast, we consider a more general scenario where inter-cell interference is allowed and controlled at the cloud.

Further to previous works, we derive deterministic approximations of the useful power, intra-cell interference and inter-cell interference power terms at every user in the system, using only the second order statistics of the channels. The approximations are closed-form expressions and functions of the outer beamformers. Therefore, they enable outer beamforming optimization at the cloud for diverse objectives while no knowledge of instantaneous channels or inner beamformer realizations is needed. This results in a top-down one-shot hierarchical beamforming which requires only reduced fronthaul transmissions. More precisely, in our approach, the cloud first designs the outer beamformers centrally, sends each one of them to the corresponding BS and every BS simply concatenates the outer to its locally designed inner beamformer. Having this hierarchical structure, we propose an outer beamformer design based on iterative block diagonalization which requires only low computational complexity while maximizing the system sum rate.

The paper is organized as follows. In Section II, we introduce the system model and in Section III, the proposed hierarchical structure. Section IV presents the deterministic approximations of the power terms and Section V the iterative outer beamformer design. In Section VI, we show simulation results and in Section VII, we summarize the work.

Notations - We use lower case and upper case boldface letters to denote vectors and matrices, respectively. The ith entry of the vector $\mathbf{x}$ is denoted by $[\mathbf{x}]_i$ and the (i,j)th entry of the matrix $\mathbf{X}$ by $[\mathbf{X}]_{i,j}$. The operations $(\cdot)^{\mathrm{H}}$ and $\mathrm{tr}(\cdot)$ are Hermitian and trace of a matrix, respectively. An $N \times N$ diagonal matrix with entries of $\mathbf{x}$ is denoted by $\mathrm{diag}(\mathbf{x})$. $\mathbf{I}_N$ stands for the identity matrix of size $N \times N$. Euclidean norm of vector $\mathbf{x}$ and spectral norm of matrix $\mathbf{X}$ are denoted as $||\mathbf{x}||$ and $||\mathbf{X}||$, respectively. For a set $\mathcal{A}$, $|\mathcal{A}|$ denotes its cardinality.

II. System Model

We consider the downlink of a multicellular network with CRAN which coordinates L BSs. Every BS is equipped with M_l antennas and serves a set $\mathcal{K}_l$ of single-antenna users simultaneously, separating them spatially by a beamformer. Every cell has a set of users with size $K_l = |\mathcal{K}_l|$ such that $K_l \leq M_l$. The overall number of users in the system is $K = \sum_{l=1}^{L} K_l$. The index $l_k \in \{1, \dots, L\}$ denotes the serving BS of user k, with $k = 1, \dots, K$. The data symbol s_{k,l_k} to be sent to user k is available only at BS l_k and modeled as zero mean Gaussian process with variance one, i.e. $s_{k,l_k} \sim \mathcal{CN}(0,1)$.

We consider the one-ring channel model [11] where the channel between user k and BS l, for $l = 1, \dots, L$ is

$$\mathbf{h}_{k,l} = \sqrt{a_{k,l}}\bar{\mathbf{\Theta}}_{k,l}^{1/2}\mathbf{z}_{k,l} = \mathbf{\Theta}_{k,l}^{1/2}\mathbf{z}_{k,l}, \tag{1}$$

where $a_{k,l}$ is the long-term path loss and $\bar{\mathbf{\Theta}}_{k,l} \in \mathbb{C}^{M_l \times M_l}$ is the correlation matrix of the channel between user k and BS l. $\mathbf{\Theta}_{k,l}$ defines the second order statistics of the channel which vary over large time scale. The vector $\mathbf{z}_{k,l} \in \mathbb{C}^{M_l \times 1}$ represents the fast fluctuations in $\mathbf{h}_{k,l}$ and varies few orders faster than the channel statistics. It is modeled as a random process with identically and independently distributed (i.i.d.) entries from zero mean Gaussian distribution with unit variance, i.e. $\mathbf{z}_{k,l} \sim \mathcal{CN}(0, \mathbf{I}_{M_l})$. We assume that the users are separated by at least few wavelengths, hence their channels are mutually independent. The resulting received signal at user k is

$$\begin{aligned} y_k = {} & \mathbf{h}_{k,l_k}^{\mathrm{H}}\sqrt{p_{k,l_k}}\mathbf{v}_{k,l_k}s_{k,l_k} + \sum_{i \in \mathcal{K}_{l_k}, i \neq k} \mathbf{h}_{k,l_k}^{\mathrm{H}}\sqrt{p_{i,l_k}}\mathbf{v}_{i,l_k}s_{i,l_k} \\ & + \sum_{\substack{l=1,\\ l \neq l_k}}^{L} \sum_{j \in \mathcal{K}_l} \mathbf{h}_{k,l}^{\mathrm{H}}\sqrt{p_{j,l}}\mathbf{v}_{j,l}s_{j,l} + n_k \end{aligned} \tag{2}$$

where $p_{k,l}$ is the power allocated at BS l for user k, $\mathbf{v}_{k,l} \in \mathbb{C}^{M_l \times 1}$ the beamforming vector at BS l for user k so that $\mathbf{V}_l = [\mathbf{v}_{i,l}]_{i \in \mathcal{K}_l} \in \mathbb{C}^{M_l \times K_l}$ is the beamformer at BS l. In (2), the four summands represent the useful signal, intra-cell interference, inter-cell interference and noise components, respectively. The noise at user k is modeled as zero mean white Gaussian process with variance $\sigma^2 = 1$.

III. Hierarchical Beamforming

We propose a hierarchical beamforming, which is a concatenation of inner beamformer $\mathbf{G}_l$ and outer beamformer $\mathbf{F}_l$, i.e. $\mathbf{V}_l = \mathbf{F}_l\mathbf{G}_l$ with $\mathbf{F}_l \in \mathbb{C}^{M_l \times M_l}$ and $\mathbf{G}_l = [\mathbf{g}_{i,l}]_{i \in \mathcal{K}_l} \in \mathbb{C}^{M_l \times K_l}$. The outer beamformer $\mathbf{F}_l$ is a centralized solution designed at the cloud, based on channel statistics and allowing coordination among the BSs. Its main objective is to define the transmission subspace in which every BS should transmit such that the overall system performance is improved compared to non-coordinated techniques. The inner beamformer $\mathbf{G}_l$, on the other hand, is designed at the BS dependent only on the users in the cell and based on the instantaneous CSI within the predefined transmission subspace. Having this hierarchical beamforming, the SINR at user k obeys the form

$$\gamma_k = \frac{S_k}{I_k^{ra} + I_k^{er} + \sigma^2} \tag{3}$$

$$S_k = p_{k,l_k}|\mathbf{h}_{k,l_k}^{\mathrm{H}}\mathbf{F}_{l_k}\mathbf{g}_{k,l_k}|^2, \tag{4}$$

$$I_k^{ra} = \sum_{i \in \mathcal{K}_{l_k}, i \neq k} p_{i,l_k}|\mathbf{h}_{k,l_k}^{\mathrm{H}}\mathbf{F}_{l_k}\mathbf{g}_{i,l_k}|^2, \tag{5}$$

$$I_k^{er} = \sum_{l=1, l \neq l_k}^{L} \sum_{j \in \mathcal{K}_l} p_{j,l}|\mathbf{h}_{k,l}^{\mathrm{H}}\mathbf{F}_l\mathbf{g}_{j,l}|^2, \tag{6}$$

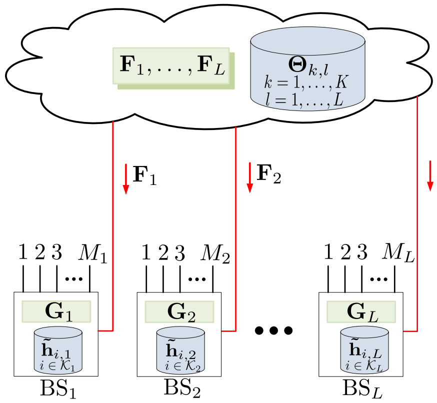


Fig. 1: Top-down one-shot hierarchical beamforming

where (4)-(6) are useful signal, intra-cell and inter-cell interference power terms, respectively. Furthermore, we can define the data rate at user k as $R_k = \log_2(1+\gamma_k)$ in bit/s/Hz and the overall system sum rate as $R_{sum} = \sum_{k=1}^{K} R_k$ in bit/s/Hz.

In general, the beamformers are coupled, meaning that to optimize the inner beamformer, we need information about the outer beamformer and vice versa. Such a design involves a lot of signaling over the fronthaul, which we aim to minimize. Therefore, we propose a top-down one-shot approach where the inner beamformer is of closed-form design. Having this approach, the cloud needs to know only the structure of the inner beamforming without any specific realization. We have applied the RZF [12], [13] so that

$$\mathbf{G}_l = \xi_l \bar{\mathbf{G}}_l = \xi_l \left(\tilde{\mathbf{H}}_l^{\mathrm{H}} \tilde{\mathbf{H}}_l + M_l \alpha_l \mathbf{I}_{M_l}\right)^{-1} \tilde{\mathbf{H}}_l^{\mathrm{H}} \tag{7}$$

where $\xi_l^2 = P_l / \mathrm{tr}(\mathbf{P}_l \bar{\mathbf{G}}_l^{\mathrm{H}} \bar{\mathbf{G}}_l)$ is a normalization parameter to fulfill the power constraint at BS l which has power budget P_l. $\mathbf{P}_l = \mathrm{diag}(\mathbf{p}_l) \in \mathbb{R}_+^{K_l \times K_l}$ is the power allocation matrix at BS l with $\mathbf{p}_l = [p_{i,l}]_{i\in\mathcal{K}_l} \in \mathbb{C}^{K_l \times 1}$. The matrix $\tilde{\mathbf{H}}_l$ is defined as $\tilde{\mathbf{H}}_l = [\tilde{\mathbf{h}}_{i,l}]_{i\in\mathcal{K}_l}^{\mathrm{H}} \in \mathbb{C}^{K_l \times M_l}$ where $\tilde{\mathbf{h}}_{k,l} = \mathbf{F}_l^{\mathrm{H}} \mathbf{h}_{k,l}$ is the effective channel between user k and BS l, i.e. the channel vector within the transmission subspace of the lth BS. The regularization parameter α_l controls the interference in the cell and it has been chosen so that it maximizes the SINR for single cell scenario with only local channel knowledge [14], [15]: $\alpha_l = (K_l \sigma^2)/(P_l M_l)$.

Having closed-form solutions for the inner beamformers, the cloud is able to optimize the outer beamformers $\mathbf{F}_l$ for $\forall l$ without the knowledge of the realizations of $\mathbf{G}_l$. The optimized $\mathbf{F}_l$ is then transmitted to BS l and concatenated with $\mathbf{G}_l$, see Fig. 1. The challenge of optimizing the outer beamformers lies in the fact that the cloud has only statistical channel information, i.e., it does not know any realization of $\mathbf{h}_{k,l}$, but knows only the statistical CSI $\mathbf{\Theta}_{k,l}$ for $\forall k, l$. Therefore, to allow system analysis and design at the cloud based only on the available statistical knowledge, we apply RMT and derive deterministic equivalents of (4)-(6) which are closed-form expressions approximating the power terms.

IV. Deterministic Equivalents

The deterministic equivalents provide asymptotic expressions of functionals with large dimensional random matrices where the number of rows and columns increase to infinity while keeping their ratio constant [16]. The interpretation of these rows and columns for our model is number of users and number of antennas at BSs. Interestingly, the asymptotic expressions provide tight approximations of finite size systems and even for systems of very small dimensions.

Throughout this work, we denote the deterministic equivalent of a functional x as $\mathring{x}$ where $x - \mathring{x} \xrightarrow{a.s.} 0$. The notation "$\xrightarrow{a.s.}$" refers to almost sure convergence as $M_l, K_l \to \infty$ with ratio $\beta_l = M_l/K_l$ such that $0 < \liminf_{M_l,K_l} \beta_l \leq \limsup_{M_l,K_l} \beta_l < \infty$ for $l = 1, \dots, L$.

We approximate the power terms S_k, I_k^{ra} and I_k^{er} at user k for $k = 1, \dots, K$ by their deterministic equivalents $\mathring{S}_k$, $\mathring{I}_k^{ra}$ and $\mathring{I}_k^{er}$, respectively. Using these deterministic equivalents, we also define the approximations $\mathring{\gamma}_k$ of the SINR at every user, $\mathring{R}_k$ of the data rate at every user and $\mathring{R}_{sum}$ of the overall system sum rate.

To obtain $\mathring{S}_k$, $\mathring{I}_k^{ra}$ and $\mathring{I}_k^{er}$, we first define $\mathbf{e}_l$ and $\mathbf{T}_l$ for $l = \{1, \dots, L\}$ and $k = \{1, \dots, K\}$ which are from the unique solution of

$$\mathbf{e}_l = \left[\frac{1}{M_l} \mathrm{tr}(\mathbf{F}_l^{\mathrm{H}} \mathbf{\Theta}_{i,l} \mathbf{F}_l \mathbf{T}_l)\right]_{i\in\mathcal{K}_l} \in \mathbb{C}^{K_l \times 1}, \tag{8}$$

$$\mathbf{T}_l = \left(\frac{1}{M_l} \sum_{j\in\mathcal{K}_l} \frac{\mathbf{F}_l^{\mathrm{H}} \mathbf{\Theta}_{j,l} \mathbf{F}_l}{(1+[\mathbf{e}_l]_j)} + \alpha_l \mathbf{I}_{M_l}\right)^{-1}. \tag{9}$$

The expressions $\mathbf{e}_l$ and $\mathbf{T}_l$ can be defined using a fixed point algorithm which iteratively solves equations (8) and (9). Having $\mathbf{e}_l$ and $\mathbf{T}_l$, we just substitute them in the following set of equations:

$$\mathring{\Psi}_l = \frac{1}{M_l} \sum_{j\in\mathcal{K}_l} \frac{p_{j,l} [\mathbf{e}_l']_j}{(1+[\mathbf{e}_l]_j)^2}, \tag{10a}$$

$$\mathbf{e}_l' = \mathbf{D}_l \mathbf{v}_l, \tag{10b}$$

$$\mathbf{D}_l = (\mathbf{I}_{K_l} - \mathbf{J}_l)^{-1}, \tag{10c}$$

$$\mathbf{v}_l = \left[\frac{1}{M_l} \mathrm{tr}(\mathbf{F}_l^{\mathrm{H}} \mathbf{\Theta}_{t,l} \mathbf{F}_l \mathbf{T}_l^2)\right]_{t\in\mathcal{K}_l} \in \mathbb{C}^{K_l \times 1}, \tag{10d}$$

$$[\mathbf{J}_l]_{i,j} = \frac{\mathrm{tr}(\mathbf{F}_l^{\mathrm{H}} \mathbf{\Theta}_{i,l} \mathbf{F}_l \mathbf{T}_l \mathbf{F}_l^{\mathrm{H}} \mathbf{\Theta}_{j,l} \mathbf{F}_l \mathbf{T}_l)}{M_l^2 (1+[\mathbf{e}_l]_j)^2} \text{ for } i,j \in \mathcal{K}_l, \tag{10e}$$

$$\mathring{\Upsilon}_{k,l} = \begin{cases} \sum\limits_{j\in\mathcal{K}_l, j\neq k} \frac{p_{j,l} [\mathbf{c}_{k,l}']_j}{(1+[\mathbf{e}_l]_j)^2} & \text{for } l = l_k \\ \sum\limits_{j\in\mathcal{K}_l} \frac{p_{j,l} [\mathbf{c}_{k,l}']_j}{(1+[\mathbf{e}_l]_j)^2} & \text{otherwise} \end{cases} \tag{10f}$$

$$\mathbf{c}_{k,l}' = \mathbf{D}_l \mathbf{w}_{k,l}, \tag{10g}$$

$$\mathbf{w}_{k,l} = \left[\frac{1}{M_l} \mathrm{tr}(\mathbf{F}_l^{\mathrm{H}} \mathbf{\Theta}_{t,l} \mathbf{F}_l \mathbf{T}_l \mathbf{F}_l^{\mathrm{H}} \mathbf{\Theta}_{k,l} \mathbf{F}_l \mathbf{T}_l)\right]_{t\in\mathcal{K}_l} \in \mathbb{C}^{K_l \times 1} \tag{10h}$$

and so we obtain all required components to derive the deterministic equivalents of the power terms as

$$\mathring{S}_k = \frac{p_{k,l_k} P_{l_k} [\mathbf{e}_{l_k}]_k^2}{\mathring{\Psi}_{l_k}(1+[\mathbf{e}_{l_k}]_k)^2}, \tag{11}$$

$$\mathring{I}_k^{ra} = \frac{P_{l_k} \mathring{\Upsilon}_{k,l_k}}{M_{l_k} \mathring{\Psi}_{l_k}(1+[\mathbf{e}_{l_k}]_k)^2}, \tag{12}$$

$$\mathring{I}_k^{er} = \sum_{l=1, l\neq l_k}^{L} \frac{P_l}{M_l \mathring{\Psi}_l} \mathring{\Upsilon}_{k,l}. \tag{13}$$

Proof: see Appendix A.

From the definition of deterministic equivalents [16] and from the continuous mapping theorem [17], we can define the deterministic equivalent of the SINR γ_k as $\mathring{\gamma}_k = \mathring{S}_k/(\mathring{I}_k^{ra} + \mathring{I}_k^{er} + \sigma^2)$, the deterministic equivalent of data rate R_k as $\mathring{R}_k = \log_2(1+\mathring{\gamma}_k)$ such that $R_k - \mathring{R}_k \xrightarrow{a.s.} 0$ and the deterministic equivalent of the overall system sum rate R_{sum} as $\mathring{R}_{sum} = \sum_{k=1}^{K} \mathring{R}_k$ where $R_{sum} - \mathring{R}_{sum} \xrightarrow{a.s.} 0$.

Therefore, we can approximate diverse parameters by deterministic expressions without knowing any actual channel realization. The only required information to determine the approximations is the statistical CSI $\mathbf{\Theta}_{k,l}$. Additionally, all deterministic equivalents are derived as functions of the outer beamformers $\mathbf{F}_l$ for $\forall l$ which allows the cloud to perform centralized optimization with respect to the matrices $\mathbf{F}_l$.

V. Outer Beamformer Design

The outer beamformers $\mathbf{F}_l$ for $l = 1, \ldots, L$ are designed centrally at the cloud and based only on the statistical CSI. Our objective is to maximize the system sum rate while satisfying a total power constraint at each BS:

$$\begin{aligned} &\underset{\mathbf{F}_1,\ldots,\mathbf{F}_L}{\arg\max}\, \mathring{R}_{sum} \\ \text{s.t. :}&\mathrm{tr}(\mathbf{P}_l \mathbf{G}_l^{\mathrm{H}} \mathbf{G}_l) \le P_l \qquad \text{for } l = 1, \ldots, L. \end{aligned} \tag{14}$$

This maximization problem involves a non-convex optimization and its solution is non-tractable. Therefore, we propose an iterative approach which is based on block diagonalization [18].

A. Iterative Block Diagonalization

We propose an iterative algorithm where in every iteration, the transmission subspace of BS b for $b \in \{1, \ldots, L\}$ is selected such that the system sum rate $\mathring{R}_{sum}$ is maximized given all selected subspaces in the previous iteration. The algorithm converges when the cloud cannot find new transmission subspaces which further improve the sum rate.

To select the transmission subspace for BS b, we define $\mathbf{B}_b^s = \left[\mathbf{F}_b^{\mathrm{H}} \mathbf{\Theta}_{i,b} \mathbf{F}_b\right]_{i \in \mathcal{K}_b}$ of size $M_b \times K_b M_b$ and $\mathbf{B}_b^i = \left[\mathbf{F}_b^{\mathrm{H}} \mathbf{\Theta}_{j,b} \mathbf{F}_b\right]_{j \in \{\mathcal{K}_l : l=1,\ldots,L \text{ and } l \neq b\}}$ of size $M_b \times M_b \sum_{l=1, l\neq b}^{L} K_l$. As next step, we perform singular value decomposition (SVD) on the matrix $\mathbf{B}_b^i$ and denote its left singular vectors by $\mathbf{E}_b$. We choose an orthogonal basis $\mathbf{E}_b^0$ which is a selection of the vectors $\mathbf{E}_b$ corresponding to the weakest N_b^i singular values. Then, we project the "serving" transmission space of BS b onto this orthogonal basis and denote the projection by $\mathbf{M}_b$, i.e. $\mathbf{M}_b = (\mathbf{E}_b^0)^{\mathrm{H}} \mathbf{B}_b^s$. Using SVD, we define the N_b^s dominant modes and collect them in the matrix $\mathbf{M}_b^1$. The resulting outer beamformer for BS b is $\mathbf{F}_b = \mathbf{E}_b^0 \mathbf{M}_b^1$ and has dimensions $M_b \times N_b^s$. This outer beamformer defines the transmission subspace which is orthogonal or nearly orthogonal to the subspace of all users to which BS b produces interference. At the same time, only the strongest dimensions, occupied by the transmission subspace of served users, are taken into account.

The choice of N_b^i and N_b^s plays an important role in the resulting data rate. Therefore, we do not set them constant, but let the cloud find the values of N_b^i and N_b^s which result in highest sum rate $\mathring{R}_{sum}$ for every iteration.

VI. Simulations Results

In this section, we present simulation results showing the accuracy of the derived deterministic approximations and the average system performance. We execute simulations for $L = 3$ BSs under different conditions by changing diverse parameters such as number M_l of antennas, number K_l of users, their ratio $\beta_l = M_l/K_l$, scattering environment for strong and weak channel correlation and signal to noise ratio (SNR), i.e. $\rho = P_l/\sigma^2$ for $\forall l$. To obtain an average performance, we consider N_f frames, where in every frame the second order channel statistics are constant, and N_r channel realizations in each frame.

A. General Setup

The correlation $\bar{\mathbf{\Theta}}_{k,l}$ is modeled using the discrete uniform distribution [19] and assumes that there are $N_{k,l}$ scatterers evenly spaced within angular spread $\Delta_{k,l}$ around the kth user with angle of arrival $\phi_{k,l}$ with respect to the lth BS. We assume that the BSs are equipped with a uniform linear array with distance $d = 0.5\lambda$ between two neighbouring antenna elements where λ is the carrier wavelength. The (m,n)th element of $\bar{\mathbf{\Theta}}_{k,l}$ is given by

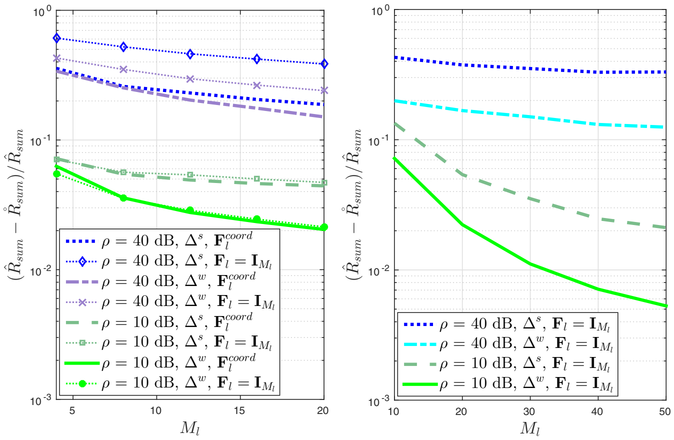
$$[\bar{\mathbf{\Theta}}_{k,l}]_{m,n} = \frac{1}{N_{k,l}} \sum_{i=1}^{N_{k,l}} e^{-j2\pi \frac{d}{\lambda}(m-n)\cos(\theta_{k,l,i})} \tag{15}$$

where $\theta_{k,l,i} = \theta_{k,l}^{\min} + \Delta\theta_{k,l}(i-1)$ with $\theta_{k,l}^{\min} = \phi_{k,l} - \frac{\Delta_{k,l}}{2}$ and $\Delta\theta_{k,l} = \Delta_{k,l}/(N_{k,l}-1)$ for $i = 1, \ldots, N_{k,l}$.

In the simulations below, we have distinguished between strong Δ^s and weak Δ^w channel correlation scenarios. To simulate them, we let the angular spread $\Delta_{k,l}$ of scatterers around a user be a random variable with uniform distribution where for the weak correlation we have $\Delta^w \sim \mathcal{U}(1°, 60°)$ and for the strong correlation $\Delta^s \sim \mathcal{U}(1°, 5°)$ with $N_{k,l} = \Delta_{k,l}$.

For all simulations, we assume that the power is equally distributed between the users, i.e. $\mathbf{P}_l = (P_l/K_l)\mathbf{I}_{K_l}$ and that all users are uniformly distributed in hexagonal cells, with radius $R_{cell} = 50$ m. To model the long term path loss, we consider the distance $D_{k,l}$ between user k and BS l, the reference distance $D_0 = 10$ m and the path loss exponent $\alpha_{loss} = 3$ such that $a_{k,l} = (D_{k,l}/D_0)^{-\alpha_{loss}}$.

For the proposed iterative block diagonalization, at the beginning of the optimization, the initial outer beamformers are



(a) $\beta_l = 2$ (b) $\beta_l = 10$

Fig. 2: Relative sum rate difference

set to be identity matrices, i.e. $\mathbf{F}_l = \mathbf{I}_{M_l}$ for $l = 1, \ldots, L$ and the cloud selects N_l^i and N_l^s to maximize the overall system sum rate such that $K_l \leq N_l^i \leq M_l$ and $K_l \leq N_l^s \leq N_l^i$.

B. Accuracy

In this subsection, we examine how tight the deterministic approximations are. We compare the average sum rate $\mathring{R}_{sum}$ obtained by the deterministic equivalents with the ergodic sum rate $\hat{R}_{sum}$ obtained by the conventional Monte Carlo approach. We consider both with ($\mathbf{F}_l^{coord.}$) and without ($\mathbf{F}_l = \mathbf{I}_{M_l}$) outer beamforming designs while $\mathbf{G}_l$ utilizes RZF.

In Fig. 2, we present the relative sum rate difference between $\mathring{R}_{sum}$ and $\hat{R}_{sum}$ by changing M_l and K_l such that their ratio β_l is constant. We investigate the performance for $\rho = 10$ dB, $\rho = 40$ dB and for both strong and weak correlation. For $\beta_l = 2$ we have $N_f = 500$ and $N_r = 5000$ and for $\beta_l = 10$, $N_f = 250$ and $N_r = 2500$. The results show that the difference between the two approximation methods is small even for system scenarios of small numbers of antennas and users as well as that applying outer beamformer, which suppresses the inter-cell interference, improves the estimation. We observed that the deterministic approximations reach time savings of orders of magnitude. From the figures, we conclude that the difference of the two approximation methods becomes smaller when:

- the SNR is weaker
- the channel correlation is weaker
- the numbers of antennas and users are larger

The first two observations, can be explained by the rank-one perturbation lemma (Lemma 2.1 in [20]) evaluated at α_l, as applied in our derivations, from which we know that for higher $\alpha_l = 1/(\rho\beta_l)$ and for lower spectral norm of the correlation matrix, we obtain a tighter approximation. The last observation is explained also by the RMT from which we know that the closer we are to the large limit of dimensions approaching infinity, the tighter the empirical and the limiting expressions are. Therefore, for large M_l and K_l, the two approximation methods converge to the same asymptotic result.

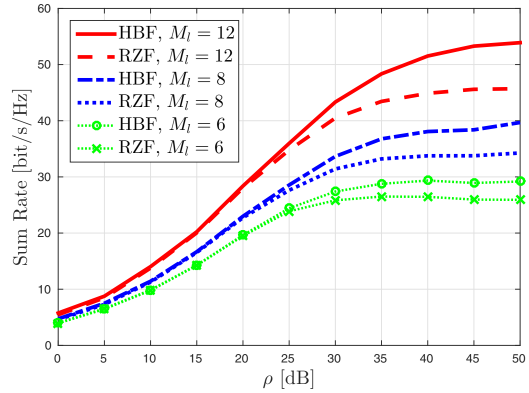


Fig. 3: System performance for weakly correlated channels

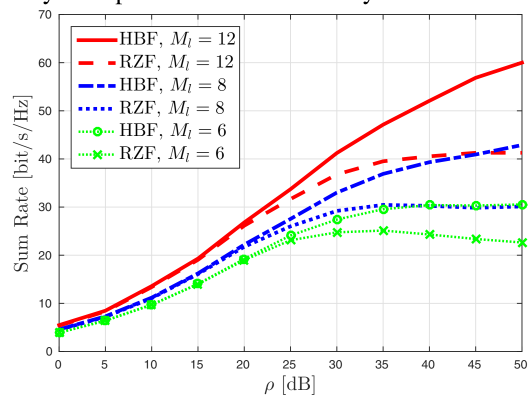


Fig. 4: System performance for strongly correlated channels

C. Average System Performance

In this subsection, we present the average system performance of the proposed hierarchical beamforming with iterative block diagonalization designing the outer beamformers.

In Fig. 3 and 4, we have the sum rate of the proposed hierarchical beamforming (HBF) compared to the conventional distributed RZF where every BS applies RZF to the channel matrix of its own users. The simulation setup is for $M_l = 6$, $M_l = 8$ and $M_l = 12$, $K_l = 4$, with weakly or strongly correlated channels, and with N_f and N_r both equal to 1000. From the simulation results, we conclude that managing the inter-cell interference through outer beamforming is beneficial even for very small scenarios of only three cells where we have only small inter-cell interference. The slight drop in RZF for some scenarios is due to the choice of α_l which can be optimized by taking into account that the system is multicellular and so to improve the system performance. From the simulations, we summarized the following observations:

- Comparing HBF and RZF, we observe that in the high SNR regime, having inter-cell interference management through outer beamformer becomes essential.
- The more antennas we have, the higher the diversity is and so the gain of the outer beamforming becomes higher.

In general, the more inter-cell interference is present in the system, the more crucial becomes its control and, hence, it is more beneficial to apply an outer beamformer.

VII. Conclusions

In this paper, we proposed a coordinated hierarchical beamforming for the downlink of a multicellular multiuser system with CRAN which requires only reduced transmissions over the fronthaul. Applying RMT, we derive deterministic approximations of the received useful, intra-cell and inter-cell interference power terms at every user using only the available statistics at the cloud. These approximations provide us with reliable performance estimation even for small system dimensions as well as with impressive time savings compared to traditional Monte Carlo approach. Additionally, the approximations are closed-form expressions and, thus, enable the cloud to perform optimization for diverse objectives with respect to the outer beamformers. In this work, we proposed a low complexity iterative approach for the outer beamformer design which maximizes the system sum rate. Simulations demonstrate that our proposed hierarchical beamformer achieves higher sum rate compared to the non-coordinated RZF scheme.

Acknowledgment

This work has been performed in the context of DFG funded CRC 1053 MAKI.

Appendix A

Here, we provide the proof of the deterministic equivalents (11)-(13). To derive them, we make the following two assumptions. The first one is that the random process $\bar{\mathbf{z}}_{k,l} = \mathbf{z}_{k,l}/\sqrt{M_l}$ is zero mean i.i.d. with variance $1/M_l$ and its eighth order moment is of order $1/M_l^4$. The second assumption is that the matrix $\tilde{\mathbf{\Theta}}_{k,l} = \mathbf{F}_l^{\mathrm{H}}\mathbf{\Theta}_{k,l}\mathbf{F}_l$ is deterministic with uniformly bounded spectral norm, i.e., $\limsup_{M_l\to\infty}\sup_{1\le k\le K}||\tilde{\mathbf{\Theta}}_{k,l}|| < \infty$ for $\forall l$.

We define $\tilde{\mathbf{C}}_l = \tilde{\mathbf{\Gamma}}_l + \alpha_l\mathbf{I}_{M_l}$ with $\tilde{\mathbf{\Gamma}}_l = \frac{1}{M_l}\tilde{\mathbf{H}}_l^{\mathrm{H}}\tilde{\mathbf{H}}_l$ and $\tilde{\mathbf{C}}_{[p],l} = \tilde{\mathbf{\Gamma}}_{[p],l} + \alpha_l\mathbf{I}_{M_l}$ with $\tilde{\mathbf{\Gamma}}_{[p],l} = \frac{1}{M_l}\tilde{\mathbf{H}}_{[p],l}^{\mathrm{H}}\tilde{\mathbf{H}}_{[p],l}$ where $\tilde{\mathbf{H}}_{[p],l}$ is the matrix $\tilde{\mathbf{H}}_l$ without the pth row as well as the following lemmas and theorems:

L1: Matrix inversion lemma, Eq. (2.2) in [21]
L2: Derivative of matrix inverse, Eq. (59) in [22]
L3: Trace lemma, Lemma 14.2 in [16]
L4: Rank-one perturbation lemma, Lemma 2.1 in [20]
T1: Theorem 1 in [8]
T2: Dominated convergence theorem (Theorem 16.4 in [23])

Following the approach in [8], we derive separately the deterministic equivalent of every power term. We begin with the normalization parameter $\xi_l^2 = P_l/\mathrm{tr}(\mathbf{P}_l\bar{\mathbf{G}}_l^{\mathrm{H}}\bar{\mathbf{G}}_l) = P_l/\Psi_l$. With **L1** and **L3**, we obtain

$$\Psi_l = \sum_{j\in\mathcal{K}_l}\frac{p_{j,l}}{M_l^2}\tilde{\mathbf{h}}_{j,l}^{\mathrm{H}}\tilde{\mathbf{C}}_l^{-2}\tilde{\mathbf{h}}_{j,l} \xrightarrow{a.s.} \sum_{j\in\mathcal{K}_l}\frac{p_{j,l}\mathrm{tr}(\tilde{\mathbf{\Theta}}_{j,l}\tilde{\mathbf{C}}_{[j],l}^{-2})}{(M_l + \mathrm{tr}(\tilde{\mathbf{\Theta}}_{j,l}\tilde{\mathbf{C}}_{[j],l}^{-1}))^2}.$$

According to **T1** and $m_{\mathbf{Y},\mathbf{A}}(z) \triangleq \mathrm{tr}\big(\mathbf{A}\big(\mathbf{Y} - z\mathbf{I}_N\big)^{-1}\big)/N$ the Stieltjes transform of random matrices at point z with $\mathbf{Y} = \mathbf{X} + \mathbf{B}$ where $\mathbf{A}$, $\mathbf{B}$ deterministic and $\mathbf{X}$ random and of size $N\times N$, we have $m_{\tilde{\mathbf{\Gamma}}_l,\tilde{\mathbf{\Theta}}_{j,l}}(-\alpha_l) = \mathrm{tr}(\tilde{\mathbf{\Theta}}_{j,l}(\tilde{\mathbf{\Gamma}}_l{+}\alpha_l\mathbf{I}_{M_l})^{-1})/M_l$ and $\mathring{m}_{\tilde{\mathbf{\Gamma}}_l,\tilde{\mathbf{\Theta}}_{j,l}}(-\alpha_l) = \mathrm{tr}(\tilde{\mathbf{\Theta}}_{j,l}\mathbf{T}_l)/M_l = [\mathbf{e}_l]_j$ for $j\in\mathcal{K}_l$ with $\mathbf{T}_l$ and $[\mathbf{e}_l]_j$ as in (9) and (8), respectively. Additionally, $m'_{\tilde{\mathbf{\Gamma}}_l,\tilde{\mathbf{\Theta}}_{j,l}}(-\alpha_l) = \mathrm{tr}\big(\tilde{\mathbf{\Theta}}_{j,l}(\tilde{\mathbf{\Gamma}}_l + \alpha_l\mathbf{I}_l)^{-2}\big)/M_l$ and from **T2**, we have $\mathring{m}'_{\tilde{\mathbf{\Gamma}}_l,\tilde{\mathbf{\Theta}}_{\mathbf{j},\mathbf{l}}}(-\alpha_l) = \mathrm{tr}(\tilde{\mathbf{\Theta}}_{j,l}\mathbf{T}'_l)/M_l = [\mathbf{e}'_l]_j$ where the derivative is with respect to $z = -\alpha_l$. Using **L2** on $\mathbf{T}'_l$, we obtain the derivative of $\mathbf{e}_l$ in a convenient matrix form $\mathbf{e}'_l = (\mathbf{I}_{M_l} - \mathbf{J}_l)^{-1}\mathbf{v}_l$ with $\mathbf{J}_l$ and $\mathbf{v}_l$ as in (10e) and (10d).

Therefore, after applying **L4** and the theorems form above, we obtain $\Psi_l - \mathring{\Psi}_l \xrightarrow{a.s.} 0$ for $\mathring{\Psi}_l = \sum_{j\in\mathcal{K}_l}[p_{j,l}[\mathbf{e}'_l]_j]/[M_l(1+[\mathbf{e}_l]_j)^2]$ as in (10a) and the deterministic equivalent of the normalization parameter is equal to $P_l/\mathring{\Psi}_l$. □

Having $S_k = \xi_{l_k}^2 p_{k,l_k}\bar{S}_k$, we apply **L1** on $\bar{S}_k$ and obtain

$$\bar{S}_k = \left|\frac{\bar{\mathbf{z}}_{k,l_k}^{\mathrm{H}}\tilde{\mathbf{\Theta}}_{k,l_k}^{\mathrm{H}/2}\tilde{\mathbf{C}}_{[k],l_k}^{-1}\tilde{\mathbf{\Theta}}_{k,l_k}^{1/2}\bar{\mathbf{z}}_{k,l_k}}{1+\bar{\mathbf{z}}_{k,l_k}^{\mathrm{H}}\tilde{\mathbf{\Theta}}_{k,l_k}^{\mathrm{H}/2}\tilde{\mathbf{C}}_{[k],l_k}^{-1}\tilde{\mathbf{\Theta}}_{k,l_k}^{1/2}\bar{\mathbf{z}}_{k,l_k}}\right|^2.$$

Note that we decompose $\tilde{\mathbf{\Theta}}_{k,l_k} = \tilde{\mathbf{\Theta}}_{k,l_k}^{1/2}\tilde{\mathbf{\Theta}}_{k,l_k}^{\mathrm{H}/2}$ where $\tilde{\mathbf{\Theta}}_{k,l}^{1/2} = \mathbf{F}_l^{\mathrm{H}}\mathbf{\Theta}_{k,l}^{1/2}$ and $\tilde{\mathbf{\Theta}}_{k,l}^{\mathrm{H}/2} = \mathbf{\Theta}_{k,l}^{H/2}\mathbf{F}_l$ and that the correlation matrix enjoys the symmetric property $\mathbf{\Theta}_{k,l}^{1/2} = \mathbf{\Theta}_{k,l}^{\mathrm{H}/2}$. For the term in the absolute value, we perform manipulations with **L3**, **L4** and **T1** analogical to the approach above and we obtain

$$\mathring{\bar{S}}_k = \frac{\mathring{m}^2_{\tilde{\mathbf{\Gamma}}_{l_k},\tilde{\mathbf{\Theta}}_{k,l_k}}(-\alpha_{l_k})}{\big(1+\mathring{m}_{\tilde{\mathbf{\Gamma}}_{l_k},\tilde{\mathbf{\Theta}}_{k,l_k}}(-\alpha_{l_k})\big)^2} = \frac{[\mathbf{e}_{l_k}]_k^2}{(1+[\mathbf{e}_{l_k}]_k)^2}$$

which is the deterministic equivalent of the useful power without allocated power and normalization parameter. □

Having $I_k^{ra} = \xi_{l_k}^2\bar{I_k^{ra}}$, we apply **L1** on $\bar{I_k^{ra}}$ and obtain

$$\frac{\bar{\mathbf{z}}_{k,l_k}^{\mathrm{H}}\tilde{\mathbf{\Theta}}_{k,l_k}^{\mathrm{H}/2}\tilde{\mathbf{C}}_{[k],l_k}^{-1}\tilde{\mathbf{H}}_{[k],l_k}^{\mathrm{H}}\mathbf{P}_{[k],l_k}\tilde{\mathbf{H}}_{[k],l_k}\tilde{\mathbf{C}}_{[k],l_k}^{-1}\tilde{\mathbf{\Theta}}_{k,l_k}^{1/2}\bar{\mathbf{z}}_{k,l_k}}{M_{l_k}\big(1+\bar{\mathbf{z}}_{k,l_k}^{\mathrm{H}}\tilde{\mathbf{\Theta}}_{k,l_k}^{\mathrm{H}/2}\tilde{\mathbf{C}}_{[k],l_k}^{-1}\tilde{\mathbf{\Theta}}_{k,l_k}^{1/2}\bar{\mathbf{z}}_{k,l_k}\big)^2} \tag{A1}$$

with $\mathbf{P}_{[p],l}$ being $\mathbf{P}_l$ without the pth row and column. We apply **L3** and **L4** to the numerator of (A1) and obtain

$$\begin{aligned}\Upsilon_{k,l_k} &= \mathrm{tr}\big(\tilde{\mathbf{\Theta}}_{k,l_k}\tilde{\mathbf{C}}_{[k],l_k}^{-1}\tilde{\mathbf{H}}_{[k],l_k}^{\mathrm{H}}\mathbf{P}_{[k],l_k}\tilde{\mathbf{H}}_{[k],l_k}\tilde{\mathbf{C}}_{[k],l_k}^{-1}\big)/M_{l_k}\\ &\xrightarrow{a.s.} \mathrm{tr}\big(\mathbf{P}_{[k],l_k}\tilde{\mathbf{H}}_{[k],l_k}\tilde{\mathbf{C}}_{l_k}^{-1}\tilde{\mathbf{\Theta}}_{k,l_k}\tilde{\mathbf{C}}_{l_k}^{-1}\tilde{\mathbf{H}}_{[k],l_k}^{\mathrm{H}}\big)/M_{l_k}\\ &= \sum_{\substack{j\in\mathcal{K}_{l_k}\\ j\neq k}} p_{j,l_k}\underbrace{\bar{\mathbf{z}}_{j,l_k}^{\mathrm{H}}\tilde{\mathbf{\Theta}}_{j,l_k}^{\mathrm{H}/2}\tilde{\mathbf{C}}_{l_k}^{-1}\tilde{\mathbf{\Theta}}_{k,l_k}\tilde{\mathbf{C}}_{l_k}^{-1}\tilde{\mathbf{\Theta}}_{j,l_k}^{1/2}\bar{\mathbf{z}}_{j,l_k}}_{\upsilon_1}.\end{aligned}$$

Because $\bar{\mathbf{z}}_{\mathbf{j},\mathbf{l_k}}$ is independent of $\tilde{\mathbf{C}}_{[\mathbf{j}],\mathbf{l_k}}$, after manipulations with **L1**, **L3** and **L4** on υ_1, we achieve the convergence

$$\upsilon_1 - \frac{\mathrm{tr}\big(\tilde{\mathbf{\Theta}}_{j,l_k}\tilde{\mathbf{C}}_{l_k}^{-1}\tilde{\mathbf{\Theta}}_{k,l_k}\tilde{\mathbf{C}}_{l_k}^{-1}\big)/M_{l_k}}{\big(1+\mathrm{tr}\big(\tilde{\mathbf{\Theta}}_{j,l_k}\tilde{\mathbf{C}}_{l_k}^{-1}\big)/M_{l_k}\big)^2} \xrightarrow{a.s.} 0. \tag{A2}$$

The denominator is of already known form. Therefore, let us consider the numerator in the last expression (A2) which equals $m'_{\tilde{\mathbf{\Gamma}}_{l_k}-z\tilde{\mathbf{\Theta}}_{k,l_k},\tilde{\mathbf{\Theta}}_{j,l_k}}(-\alpha_{l_k})\big|_{z=0}$. From

$$\mathring{m}_{\tilde{\mathbf{\Gamma}}_{l_k}-z\tilde{\mathbf{\Theta}}_{k,l_k},\tilde{\mathbf{\Theta}}_{j,l_k}}(-\alpha_{l_k}) = \mathrm{tr}(\tilde{\mathbf{\Theta}}_{j,l_k}\mathbf{Q}_{k,l_k})/M_{l_k}$$

with

$$\mathbf{Q}_{k,l_k} = \left(\frac{1}{M_{l_k}}\sum_{j\in\mathcal{K}_{l_k}}\frac{\tilde{\mathbf{\Theta}}_{j,l_k}}{1+[\mathbf{c}_{k,l_k}]_j} - z\tilde{\mathbf{\Theta}}_{k,l_k} + \alpha_{l_k}\mathbf{I}_{M_{l_k}}\right)^{-1},$$

$[\mathbf{c}_{k,l_k}]_i = \mathrm{tr}(\tilde{\mathbf{\Theta}}_{i,l_k}\mathbf{Q}_{k,l_k})/M_{l_k}$

and after **T2**, we obtain the deterministic equivalent of the numerator in (A2)

$$\mathring{m}'_{\tilde{\mathbf{\Gamma}}_{l_k}-z\tilde{\mathbf{\Theta}}_{k,l_k},\tilde{\mathbf{\Theta}}_{j,l_k}}(-\alpha_{l_k}) = \mathrm{tr}(\tilde{\mathbf{\Theta}}_{j,l_k}\mathbf{Q}'_{k,l_k})/M_{l_k}$$

for $j \in (\mathcal{K}_{l_k}, j \neq k)$ and with **L2**, the derivative expression $\mathbf{Q}'_{k,l_k}$ becomes

$$\mathbf{Q}'_{k,l_k} = \mathbf{Q}_{k,l_k}\left(\frac{1}{M_{l_k}}\sum_{j\in\mathcal{K}_{l_k}}\frac{\tilde{\mathbf{\Theta}}_{j,l_k}[\mathbf{c}'_{k,l_k}]_j}{(1+[\mathbf{c}_{k,l_k}]_j)^2}+\tilde{\mathbf{\Theta}}_{k,l_k}\right)\mathbf{Q}_{k,l_k}$$

Evaluating at $z = 0$, we obtain the terms $\mathbf{Q}_{k,l_k} = \mathbf{T}_{l_k}$, $[\mathbf{c}_{k,l_k}]_j = [\mathbf{e}_{l_k}]_j$ and

$$\mathbf{Q}'_{k,l_k} = \left(\mathbf{T}_{l_k}\sum_{j\in\mathcal{K}_{l_k}}\frac{\tilde{\mathbf{\Theta}}_{j,l_k}[\mathbf{c}'_{k,l_k}]_j}{M_{l_k}(1+[\mathbf{e}_{l_k}]_j)^2}\mathbf{T}_{l_k}+\mathbf{T}_{l_k}\tilde{\mathbf{\Theta}}_{k,l_k}\mathbf{T}_{l_k}\right).$$

Hence, $\mathring{\upsilon}_1$ is described in a convenient matrix form by $\mathbf{c}'_{k,l}$ and $\mathbf{w}_{k,l}$ as defined in (10g) and (10h) for $l = l_k$.

Consequently, we obtain the deterministic equivalent of Υ_{k,l_k} which is required for the deterministic equivalent of I_k^{ra}

$$\mathring{\Upsilon}_{k,l_k} = \sum_{j\in\mathcal{K}_{l_k},j\neq k}\frac{p_{j,l_k}[\mathbf{c}'_{k,l_k}]_j}{(1+[\mathbf{e}_{l_k}]_j)^2}. \qquad \square$$

The derivations of the deterministic equivalents for I_k^{er} are similar to those for I_k^{ra}. One should take into account the statistical independence between the random process $\bar{\mathbf{z}}_{k,l}$ for $k \in \mathcal{K}_{l_k}$ with the inner beamformer $\mathbf{G}_l$ for $\forall l \neq l_k$.

For $I_k^{er} = \sum_{\forall l\neq l_k}\xi_l^2 I_{k,l}^{\bar{er}}/M_l$ with $I_{k,l}^{\bar{er}}$ the scaled inter-cell interference from BS l at user k, we apply **L3** on $I_{k,l}^{\bar{er}}$

$$\underbrace{\bar{\mathbf{z}}_{k,l}^{\mathrm{H}}\tilde{\mathbf{\Theta}}_{k,l}^{\mathrm{H}/2}\tilde{\mathbf{C}}_l^{-1}\tilde{\mathbf{H}}_l^{\mathrm{H}}\mathbf{P}_l\tilde{\mathbf{H}}_l\tilde{\mathbf{C}}_l^{-1}\tilde{\mathbf{\Theta}}_{k,l}^{1/2}\bar{\mathbf{z}}_{k,l}-\mathrm{tr}(\tilde{\mathbf{\Theta}}_{k,l}\tilde{\mathbf{C}}_l^{-1}\tilde{\mathbf{H}}_l^{\mathrm{H}}\mathbf{P}_l\tilde{\mathbf{H}}_l\tilde{\mathbf{C}}_l^{-1})/M_l}_{\mathbf{\Lambda}_{k,l}} \xrightarrow{a.s.} 0.$$

$$\mathbf{\Lambda}_{k,l} = \sum_{j\in\mathcal{K}_l}p_{j,l}\underbrace{\bar{\mathbf{z}}_{j,l}^{\mathrm{H}}\tilde{\mathbf{\Theta}}_{j,l}^{\mathrm{H}/2}\tilde{\mathbf{C}}_l^{-1}\tilde{\mathbf{\Theta}}_{k,l}\tilde{\mathbf{C}}_l^{-1}\tilde{\mathbf{\Theta}}_{j,l}^{1/2}\bar{\mathbf{z}}_{j,l}}_{\upsilon_2}$$

and with **L1**, **L3** and **L4** consequently, we achieve

$$\upsilon_2 - \frac{\mathrm{tr}\big(\tilde{\mathbf{\Theta}}_{j,l}\tilde{\mathbf{C}}_l^{-1}\tilde{\mathbf{\Theta}}_{k,l}\tilde{\mathbf{C}}_l^{-1}\big)/M_l}{\big(1+\mathrm{tr}\big(\tilde{\mathbf{\Theta}}_{j,l}\tilde{\mathbf{C}}_l^{-1}\big)/M_l\big)^2} \xrightarrow{a.s.} 0.$$

Applying **T1** and **T2**, similar to the analysis above, we obtain the deterministic equivalent of υ_2

$$\mathring{\upsilon}_2 = \frac{\mathring{m}'_{\tilde{\mathbf{\Gamma}}_l-z\tilde{\mathbf{\Theta}}_{k,l},\tilde{\mathbf{\Theta}}_{j,l}}(-\alpha_l)\big|_{z=0}}{(1+\mathring{m}_{\tilde{\mathbf{\Gamma}}_l,\tilde{\mathbf{\Theta}}_{j,l}}(-\alpha_l))^2}$$

Hence, we describe $\mathring{\mathbf{\Lambda}}_{k,l}$ as $\mathring{\Upsilon}_{k,l}$ in (10f) for $\forall l \neq l_k$

$$\mathring{\mathbf{\Lambda}}_{k,l} = \mathring{\Upsilon}_{k,l} = \sum_{j\in\mathcal{K}_l}\frac{p_{j,l}[\mathbf{c}'_{k,l}]_j}{(1+[\mathbf{e}_l]_j)^2} \text{ for } \forall l \neq l_k. \qquad \square$$

Note that for a single cell scenario without outer beamformer, i.e. $L = 1$ and $\mathbf{F}_1 = \mathbf{I}_{M_1}$, there is no inter-cell interference, the BS simply performs RZF for its own users and the deterministic equivalents boil down to these in [8].

References

[1] Cisco, “Cisco visual networking index: Forecast and methodology, 2016 - 2021,” *White paper Cisco public*, June 2017.

[2] M. Peng, C. Wang, V. Lau, and H. Poor, “Fronthaul-constrained cloud radio access networks: Insights and challenges,” *IEEE Wireless Communications*, vol. 22, no. 2, pp. 152–160, Apr. 2015.

[3] M. Peng, Y. Sun, X. Li, Z. Mao, and C. Wang, “Recent advances in cloud radio access networks: System architectures, key techniques, and open issues,” *IEEE Communications Surveys & Tutorials*, vol. 18, no. 3, pp. 2282–2308, 2016.

[4] Y. Shi, J. Zhang, and K. B. Letaief, “CSI overhead reduction with stochastic beamforming for cloud radio access networks,” *IEEE International Conference on Communications*, pp. 5154–5159, 2014.

[5] Y. Zhou and W. Yu, “Fronthaul compression and transmit beamforming optimization for multi-antenna uplink C-RAN,” *IEEE Transactions on Signal Processing*, vol. 64, no. 16, pp. 4138–4151, Aug. 2016.

[6] J. Kang, O. Simeone, J. Kang, and S. Shamai, “Fronthaul compression and precoding design for C-RANs over ergodic fading channels,” *IEEE Transactions on Vehicular Technology*, vol. 65, no. 7, pp. 5022–5032, Jul. 2016.

[7] H. Al-Shatri and A. Klein, “Hierarchical beamforming for downlink cloud radio access networks,” *14th International Symposium on Wireless Communication Systems*, 2017.

[8] S. Wagner, R. Couillet, M. Debbah, and D. Slock, “Large system analysis of linear precoding in correlated MISO broadcast channels under limited feedback,” *IEEE Transactions on Information Theory*, vol. 58, no. 7, pp. 4509–4537, Jul. 2012.

[9] J. Zhang, C. Wen, S. Jin, X. Gao, and K. Wong, “Large system analysis of cooperative multi-cell downlink transmission via regularized channel inversion with imperfect CSIT,” *IEEE Transactions on Wireless Communications*, vol. 12, no. 10, pp. 4801–4813, Oct. 2013.

[10] A. Liu and V. K. N. Lau, “Hierarchical interference mitigation for massive MIMO cellular networks,” *IEEE Transactions on Signal Processing*, vol. 62, no. 18, pp. 4786–4797, Sept. 2014.

[11] D. Shiu, G. J. Foschini, M. J. Gans, and J. M. Kahn, “Fading correlation and its effect on the capacity of multielement antenna systems,” *IEEE Transactions on Communications*, vol. 48, no. 3, pp. 502–513, Mar. 2000.

[12] M. Joham, K. Kusume, M. H. Gzara, W. Utschick, and J. A. Nossek, “Transmit wiener filter for the downlink of TDD DS-CDMA systems,” *7th IEEE ISSSTA*, vol. 1, no. 1, pp. 9–13, Sept. 2002.

[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 Transactions on Communications*, vol. 53, no. 1, pp. 195–202, Jan. 2005.

[14] V. K. Nguyen and J. S. Evans, “Multiuser transmit beamforming via regularized channel inversion: A large system analysis,” *IEEE Global Telecommunications Conference*, pp. 1–4, Dec. 2008.

[15] R. Muharar and J. Evans, “Downlink beamforming with transmit-side channel correlation: A large system analysis,” in *IEEE International Conference on Communications*, Jun. 2011, pp. 1–5.

[16] R. Couillet and M. Debbah, *Random Matrix Methods for Wireless Communications*. Cambridge University Press, 2011.

[17] A. W. V. der Vaart, *Asymptotic Statistics*. Cambridge University Press, 1998.

[18] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, “Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels,” *IEEE Transactions on Signal Processing*, vol. 52, no. 2, pp. 461–471, Feb. 2004.

[19] R. Ertel, P. Cardieri, K. Sowerby, T. Rappaport, and J. Reed, “Overview of spatial channel models for anfenna array communication systems,” *IEEE Personal Communications*, vol. 5, no. 1, pp. 10–22, Feb 1998.

[20] Z. Bai and J. Silverstein, “On the signal-to-interference-ratio of CDMA systems in wireless communications,” *The Annals of Applied Probability*, vol. 17, pp. 81–101, Feb. 2007.

[21] 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.

[22] K. B. Petersen and M. S. Pedersen, *The Matrix Cookbook*. Technical University of Denmark, Nov. 2012.

[23] P. Billingsley, *Probability and Measure*, 3rd ed. John Wiley & Sons, Inc., 1995.