Hybrid Beamforming Design for Downlink MU-MIMO-OFDM Millimeter-Wave Systems

Sepideh Gherekhloo, Khaled Ardah, and Martin Haardt
Communications Research Laboratory (CRL), TU Ilmenau, Ilmenau, Germany
{sepideh.gherekhloo, khaled.ardah, martin.haardt}@tu-ilmenau.de

***Abstract*—Hybrid beamforming (HBF) structures have been recently proposed as one of the solutions to reduce cost and energy consumption of fully-digital multi-user MIMO systems in the millimeter-wave range. However, most of the initial references on HBF design assume frequency-flat single-carrier MIMO systems. Their extension to the frequency-selective multi-carrier MIMO-OFDM case is, in general, not a straightforward task, especially in partially-connected architectures. In this paper, we propose a new HBF design approach for multi-user massive MIMO-OFDM systems. Simulation results reveal that the proposed solution achieves a comparable performance to the benchmark methods in the fully-connected case, while significantly outperforming them in case of partially-connected architectures.**

***Index Terms*—Hybrid beamforming, millimeter-wave, multi-user MIMO-OFDM.**

I. Introduction

Massive MIMO and millimeter-wave communications are among key techniques in future 5G wireless networks to increase system capacity and achievable data rates [1]. However, a large number of radio-frequency (RF) chains render the conventional fully-digital architectures impractical, due to the associated high cost and energy consumption. Therefore, HBF structures are proposed to reduce the number of RF chains by dividing the beamforming matrix into a high-dimensional analog beamforming (ABF) matrix and a low-dimensional baseband beamforming (BBF) matrix [2], [3]. Initial works [4]–[11] focused on HBF design for MIMO systems assuming fully-connected ABF architectures, where each RF chain is connected to all antenna elements using a network of phase-shifters (PSs). To reduce the cost and power consumption of high resolution PSs, partially-connected ABF architectures are proposed [10], [12]–[14], where each RF chain is connected to only a subset of antennas.

The aforementioned references, however, assume frequency-flat single-carrier MIMO systems. Their extension to frequency-selective multi-carrier MIMO-OFDM systems is, in general, not a straightforward task, especially with the partially-connected architectures. Therefore, recent works [15]–[20] proposed HBF design methods considering MIMO-OFDM systems. For example, the authors in [19] consider a downlink multi-user MISO-OFDM system, where the BBF matrix is designed using the classical zero-forcing (ZF) method, while the ABF matrix is designed such that the approximated upper-bound of the system's spectral efficiency (SE) is maximized. However, the proposed solution in [19] is mostly applicable for fully-connected architectures, where its simple adaption to partially-connected architectures incurs a high performance lose. Differently, the authors in [20] consider a fully-connected downlink multi-user MIMO-OFDM system and propose a HBF design approach, where the ABF matrices are designed using a subspace matching approach by exploiting the tensor structure of the propagation channels, while the BBF matrices are designed afterward as a solution to a signal-to-leakage-plus-noise ratio optimization problem.

In this paper, we consider a frequency-selective downlink multi-user MIMO-OFDM system and propose a HBF design approach that is applicable to both fully-connected and partially-connected architectures. We assume that the transmit BBF matrices are designed using the classical ZF method, while the receive BBF matrices are designed using the MMSE approach [21]. Moreover, we derive an approximation of system's SE upper-bound, which is used as the objective function of the transmit ABF matrix optimization. To obtain a solution, we divide the problem into a series of convex sub-problems that are updated iteratively until convergence is obtained. Simulation results show that the proposed solution achieves a comparable performance to the benchmark methods for fully-connected architectures, while significantly outperforming them in the partially-connected case.

The authors gratefully acknowledge the support of the German Research Foundation (DFG) under contract no. HA 2239/14-1.

II. System Model and Problem Formulation

We consider a downlink multi-user MIMO-OFDM system, where a single base station (BS) communicates with K mobile stations (MSs). The BS is equipped with M_T antennas and $N_T \leq M_T$ RF chains, while each MS is equipped with M_R antennas, $N_R \leq M_R$ RF chains, and receives N_s data streams. Here, we assume that $N_T \overset{\text{def}}{=} KN_s$ and $N_R \overset{\text{def}}{=} N_s$. A cyclic-prefix (CP)-OFDM based multi-carrier modulation scheme with N subcarriers is applied to combat the multipath effect. Assuming that the CP length has the same length as the maximum excess delay of the channel, such that the intersymbol interference is avoided, the received signal on the nth subcarrier, $n \in \{1,\ldots,N\}$, at the kth MS, $k \in \{1,\ldots,K\}$, is given as

$$\hat{\boldsymbol{s}}_{k,n} = \boldsymbol{W}_{k,n}^{\mathrm{H}}\big(\boldsymbol{H}_{k,n}\boldsymbol{F}_{k,n}\boldsymbol{s}_{k,n} + \boldsymbol{H}_{k,n}\textstyle\sum_{j\neq k}\boldsymbol{F}_{j,n}\boldsymbol{s}_{j,n} + \boldsymbol{n}_{k,n}\big), \quad (1)$$

where $\boldsymbol{H}_{k,n} \in \mathbb{C}^{M_R \times M_T}$ is the MIMO frequency-domain channel, $\boldsymbol{F}_{k,n} \in \mathbb{C}^{M_T \times N_s}$ is the precoding matrix, $\boldsymbol{W}_{k,n} \in \mathbb{C}^{M_R \times N_s}$ is the decoding matrix, $\boldsymbol{s}_{k,n} \in \mathbb{C}^{N_s}$ is the data vector with $\mathbb{E}\{\boldsymbol{s}_{k,n}\boldsymbol{s}_{k,n}^H\} = \boldsymbol{I}_{N_s}$, and $\boldsymbol{n}_{k,n} \in \mathbb{C}^{M_R}$ is the additive white Gaussian noise with variance σ^2. We consider a HBF structure such that $\boldsymbol{F}_{k,n} \overset{\text{def}}{=} \boldsymbol{F}^{\mathrm{RF}}\boldsymbol{F}_{k,n}^{\mathrm{BB}}$ and $\boldsymbol{W}_{k,n} \overset{\text{def}}{=} \boldsymbol{W}_k^{\mathrm{RF}}\boldsymbol{W}_{k,n}^{\mathrm{BB}}$. Here,

978-1-7281-1946-5/20/$31.00 ©2020 IEEE

$\boldsymbol{F}^{\mathrm{RF}} \in \mathbb{C}^{M_T \times N_T}$ and $\boldsymbol{W}_k^{\mathrm{RF}} \in \mathbb{C}^{M_R \times N_R}$ are the ABF matrices, while $\boldsymbol{F}_{k,n}^{\mathrm{BB}} \in \mathbb{C}^{N_T \times N_s}$ and $\boldsymbol{W}_{k,n}^{\mathrm{BB}} \in \mathbb{C}^{N_R \times N_s}$ are the BBF matrices. We consider the two well-known ABF architectures [9], [10]: fully-connected (A1) and partially-connected (A2). In A1, each RF-chain is connected to all antennas using a networks of PSs. Therefore, $\boldsymbol{F}^{\mathrm{RF}} = [\boldsymbol{f}_1^{RF}, \ldots, \boldsymbol{f}_{N_T}^{\mathrm{RF}}]$, $\boldsymbol{f}_\ell^{\mathrm{RF}} \in \mathbb{C}^{M_T}$, $|[\boldsymbol{f}_\ell^{\mathrm{RF}}]_{[i]}| = \frac{1}{\sqrt{M_T}}$, $\boldsymbol{W}_k^{\mathrm{RF}} = [\boldsymbol{w}_{k,1}^{\mathrm{RF}}, \ldots, \boldsymbol{w}_{k,N_R}^{\mathrm{RF}}]$, $\boldsymbol{w}_{k,\ell}^{\mathrm{RF}} \in \mathbb{C}^{M_R}$, and $|[\boldsymbol{w}_{k,\ell}^{\mathrm{RF}}]_{[i]}| = \frac{1}{\sqrt{M_R}}$. In A2, each RF-chain is connected to a subset of the antennas using a network of PSs. Therefore, $\boldsymbol{F}^{\mathrm{RF}} = \mathrm{blkdiag}\{\boldsymbol{f}_1^{\mathrm{RF}}, \ldots, \boldsymbol{f}_{N_T}^{\mathrm{RF}}\}$, $\boldsymbol{f}_\ell^{\mathrm{RF}} \in \mathbb{C}^{M_T/N_T}$, $|[\boldsymbol{f}_\ell^{\mathrm{RF}}]_{[i]}| = \frac{1}{\sqrt{M_T/N_T}}$, $\boldsymbol{W}_k^{\mathrm{RF}} = \mathrm{blkdiag}\{\boldsymbol{w}_{k,1}^{\mathrm{RF}}, \ldots, \boldsymbol{w}_{k,N_R}^{\mathrm{RF}}\}$, $\boldsymbol{w}_{k,\ell}^{\mathrm{RF}} \in \mathbb{C}^{M_R/N_R}$, and $|[\boldsymbol{w}_{k,\ell}^{\mathrm{RF}}]_{[i]}| = \frac{1}{\sqrt{M_R/N_R}}$. The achievable SE at the kth MS on the nth subcarrier can be expressed as

$$\mathrm{SE}_{k,n} = \log|\boldsymbol{I}_{N_s} + \boldsymbol{\Phi}_{k,n}^{-1}\boldsymbol{W}_{k,n}^{\mathrm{H}}\boldsymbol{H}_{k,n}\boldsymbol{F}_{k,n}\boldsymbol{F}_{k,n}^{\mathrm{H}}\boldsymbol{H}_{k,n}^{\mathrm{H}}\boldsymbol{W}_{k,n}|, \quad (2)$$

where $\boldsymbol{\Phi}_{k,n} = \boldsymbol{W}_{k,n}^{\mathrm{H}}\big(\boldsymbol{H}_{k,n}\big(\sum_{j\neq k}\boldsymbol{F}_{j,n}\boldsymbol{F}_{j,n}^{\mathrm{H}}\big)\boldsymbol{H}_{k,n}^{\mathrm{H}} + \sigma^2\boldsymbol{I}_{M_R}\big)\boldsymbol{W}_{k,n}$ denotes the covariance matrix of the multi-user interference plus noise. Let $\boldsymbol{F}_n = [\boldsymbol{F}_{1,n}, \ldots, \boldsymbol{F}_{K,n}] \in \mathbb{C}^{M_T \times KN_s}$. Then, we enforce $\|\boldsymbol{F}_n\|_F^2 = P_n$, where $P_n \overset{\mathrm{def}}{=} \frac{P}{N}$ and P is the total transmit power.

In this paper, our goal is to design the HBF matrices to maximize the total SE of the system, i.e.,

$$\begin{aligned} \max_{\{\boldsymbol{F}_{k,n},\boldsymbol{W}_{k,n}\}} \quad & \mathrm{SE} = \sum_k\sum_n \mathrm{SE}_{k,n} \\ \text{s.t.} \quad & \|\boldsymbol{F}_n\|_F^2 = P_n, \forall n,\ \boldsymbol{F}^{\mathrm{RF}} \in \mathscr{F}, \text{ and } \boldsymbol{W}_k^{\mathrm{RF}} \in \mathscr{W}, \forall k, \end{aligned} \quad (3)$$

where $\mathscr{W}$ and $\mathscr{F}$ denote the sets of ABF matrices satisfying the constraints that are associated with each architecture as specified above. Note that (3) is a non-convex optimization problem, where the major difficulty comes from the joint optimization of the BBF and ABF matrices and the non-convex constraints in $\boldsymbol{W}_k^{\mathrm{RF}} \in \mathscr{W}, \forall k$, and $\boldsymbol{F}^{\mathrm{RF}} \in \mathscr{F}$. In the following, we relax problem (3) by decoupling the optimization of BBF and ABF matrices and treat them separately, which is a common practice in the literature, see [4], [13], [19], [20].

III. Baseband Beamforming Design

For given and fixed ABF matrices, the BS designs its transmit BBF matrices using the conventional linear ZF scheme to eliminate the multi-user interference. We assume that MS k applies the noise whitening filter $\boldsymbol{Q}_k = \big((\boldsymbol{W}_k^{\mathrm{RF}})^{\mathrm{H}}\boldsymbol{W}_k^{\mathrm{RF}}\big)^{-\frac{1}{2}}$ after applying its receive ABF. Let $\mathring{\boldsymbol{H}}_{k,n} \overset{\mathrm{def}}{=} \boldsymbol{Q}_k^H(\boldsymbol{W}_k^{\mathrm{RF}})^{\mathrm{H}}\boldsymbol{H}_{k,n} \in \mathbb{C}^{N_R \times M_T}$, $\mathring{\boldsymbol{H}}_n \overset{\mathrm{def}}{=} [\mathring{\boldsymbol{H}}_{1,n}^{\mathrm{T}}, \ldots, \mathring{\boldsymbol{H}}_{K,n}^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{C}^{KN_R \times M_T}$, and $\bar{\boldsymbol{H}}_n \overset{\mathrm{def}}{=} \mathring{\boldsymbol{H}}_n\boldsymbol{F}^{\mathrm{RF}} \in \mathbb{C}^{KN_R \times N_T}$. The transmit BBF matrix $\boldsymbol{F}_n^{\mathrm{BB}} \overset{\mathrm{def}}{=} [\boldsymbol{F}_{1,n}^{\mathrm{BB}}, \ldots, \boldsymbol{F}_{K,n}^{\mathrm{BB}}] \in \mathbb{C}^{N_T \times KN_R}$ for the nth subcarrier is given as

$$\boldsymbol{F}_n^{\mathrm{BB}} = \sqrt{P_n}\big(\bar{\boldsymbol{F}}_n^{\mathrm{BB}}/\|\boldsymbol{F}^{\mathrm{RF}}\bar{\boldsymbol{F}}_n^{\mathrm{BB}}\|_F\big), \quad (4)$$

where $\bar{\boldsymbol{F}}_n^{\mathrm{BB}} = \bar{\boldsymbol{H}}_n^{\mathrm{H}}(\bar{\boldsymbol{H}}_n\bar{\boldsymbol{H}}_n^{\mathrm{H}})^{-1}$ is the un-normalized ZF-based precoding matrix and $\boldsymbol{F}_{k,n}^{\mathrm{BB}} \in \mathbb{C}^{N_T \times N_s}$ is the kth sub-matrix of $\boldsymbol{F}_n^{\mathrm{BB}}$ related to the kth MS (recall that $N_R \overset{\mathrm{def}}{=} N_s$). On the other hand, the receive BBF matrix of the kth MS on the nth subcarrier is calculated using the MMSE scheme as $\boldsymbol{W}_{k,n}^{\mathrm{BB}} = \min_{\boldsymbol{W}_{k,n}^{\mathrm{BB}}} \mathbb{E}[\|\hat{\boldsymbol{s}}_{k,n} - \boldsymbol{s}_{k,n}\|^2]$, i.e., $\boldsymbol{W}_{k,n}^{\mathrm{BB}} \in \mathbb{C}^{N_R \times N_s}, \forall k,n$, is computed such that the mean-squared-error of the received signal is minimized, where its solution is given by [21], [22]

$$\boldsymbol{W}_{k,n}^{\mathrm{BB}} = \boldsymbol{\Omega}_{k,n}^{-1}(\boldsymbol{W}_k^{\mathrm{RF}})^{\mathrm{H}}\boldsymbol{H}_{k,n}\boldsymbol{F}_{k,n}, \quad (5)$$

where $\boldsymbol{\Omega}_{k,n} = \sum_j \boldsymbol{S}_{k,j,n}\boldsymbol{S}_{k,j,n}^{\mathrm{H}} + \sigma^2\boldsymbol{Q}_k^{\mathrm{H}}(\boldsymbol{W}_k^{\mathrm{RF}})^{\mathrm{H}}\boldsymbol{W}_k^{\mathrm{RF}}\boldsymbol{Q}_k$ and $\boldsymbol{S}_{k,j,n} = \boldsymbol{Q}_k^{\mathrm{H}}(\boldsymbol{W}_k^{\mathrm{RF}})^{\mathrm{H}}\boldsymbol{H}_{k,n}\boldsymbol{F}_{j,n}$. As $\boldsymbol{F}_{j,n}^{\mathrm{BB}}, \forall j$, are calculated as in (4), we have $\boldsymbol{\Omega}_{k,n} = \boldsymbol{S}_{k,k,n}\boldsymbol{S}_{k,k,n}^{\mathrm{H}} + \sigma^2\boldsymbol{I}_{N_R}$, where $\boldsymbol{S}_{k,j,n}\boldsymbol{S}_{k,j,n}^{\mathrm{H}} = \boldsymbol{0}, \forall j \neq k$, and $\sigma^2\boldsymbol{Q}_k^{\mathrm{H}}(\boldsymbol{W}_k^{\mathrm{RF}})^{\mathrm{H}}\boldsymbol{W}_k^{\mathrm{RF}}\boldsymbol{Q}_k = \sigma^2\boldsymbol{I}_{N_R}$.

IV. Analog Beamforming Design

Since the BBF matrices are designed using the ZF method, the multi-user interference can be neglected during the ABF design, where the problem can be formulated such that the capacity of each effective channel $\boldsymbol{H}_{k,n}^{\mathrm{eff}} \overset{\mathrm{def}}{=} (\boldsymbol{W}_k^{\mathrm{RF}})^{\mathrm{H}}\boldsymbol{H}_{k,n}\boldsymbol{F}^{\mathrm{RF}}$ is maximized. In the following, we decouple the ABF matrices design between $\boldsymbol{W}_k^{\mathrm{RF}}$ and $\boldsymbol{F}^{\mathrm{RF}}$ by treating them independently. For the kth MS, one approach to design $\boldsymbol{W}_k^{\mathrm{RF}}$ is

$$\max_{\boldsymbol{W}_k^{\mathrm{RF}} \in \mathscr{W}} \ \eta_k = \sum_n \log|\boldsymbol{I}_{N_R} + (\boldsymbol{W}_k^{\mathrm{RF}})^{\mathrm{H}}\boldsymbol{H}_{k,n}\boldsymbol{H}_{k,n}^{\mathrm{H}}\boldsymbol{W}_k^{\mathrm{RF}}|. \quad (6)$$

To obtain a solution of problem (6), we first rewrite its objective function in terms of its upper-bound as

$$\eta_k < \log|\boldsymbol{I}_{N_R} + (\boldsymbol{W}_k^{\mathrm{RF}})^{\mathrm{H}}\boldsymbol{A}_k\boldsymbol{W}_k^{\mathrm{RF}}| \quad (7)$$

which follows from the Jensen's inequality [16], [19] and $\boldsymbol{A}_k = \sum_n \boldsymbol{H}_{k,n}\boldsymbol{H}_{k,n}^{\mathrm{H}} \in \mathbb{C}^{M_R \times M_R}$. Therefore, (6) reduces to

$$\max_{\boldsymbol{W}_k^{\mathrm{RF}} \in \mathscr{W}} \quad \log|\boldsymbol{I}_{N_R} + (\boldsymbol{W}_k^{\mathrm{RF}})^{\mathrm{H}}\boldsymbol{A}_k\boldsymbol{W}_k^{\mathrm{RF}}|. \quad (8)$$

Problem (8) has been addressed recently in [13] considering an ABF architecture A2. Meanwhile, the authors in [14] proposed a unifying design approach for the flat-fading case that is applicable for both ABF architectures A1 and A2. In this paper, we assume that $\boldsymbol{W}_k^{\mathrm{RF}}, \forall k$, are updated using the proposed method in [14].

Assuming that $\boldsymbol{F}_{k,n}^{\mathrm{BB}}$ and $\boldsymbol{W}_{k,n}^{\mathrm{BB}}, \forall k,n$, are computed by (4) and (5), respectively. Then, (2) simplifies to [21]

$$\mathrm{SE}_{k,n} = \log|\boldsymbol{I}_{N_s} + \frac{1}{\sigma^2}(\boldsymbol{F}_{k,n}^{\mathrm{BB}})^{\mathrm{H}}(\boldsymbol{F}^{\mathrm{RF}})^{\mathrm{H}}\mathring{\boldsymbol{H}}_{k,n}^{\mathrm{H}}\mathring{\boldsymbol{H}}_{k,n}\boldsymbol{F}^{\mathrm{RF}}\boldsymbol{F}_{k,n}^{\mathrm{BB}}|. \quad (9)$$

Therefore, the objective function of (3) simplifies to

$$\mathrm{SE} = \sum_k\sum_n \mathrm{SE}_{k,n} = \sum_n \log|\boldsymbol{I}_{KN_s} + \frac{1}{\sigma^2}(\boldsymbol{F}_n^{\mathrm{BB}})^{\mathrm{H}}\bar{\boldsymbol{H}}_n^{\mathrm{H}}\bar{\boldsymbol{H}}_n\boldsymbol{F}_n^{\mathrm{BB}}|, \quad (10)$$

which can be upper-bounded as

$$\mathrm{SE} \leq KN_s \log\Big(1 + \frac{P_n}{\sigma^2(KN_R)^2}Tr[(\boldsymbol{F}^{\mathrm{RF}})^{\mathrm{H}}\boldsymbol{A}\boldsymbol{F}^{\mathrm{RF}}]\Big), \quad (11)$$

where $\boldsymbol{A} = \sum_n \mathring{\boldsymbol{H}}_n^{\mathrm{H}}\mathring{\boldsymbol{H}}_n$. The derivation of the upper-bound in (11) is provided in Appendix A. Therefore, $\boldsymbol{F}^{\mathrm{RF}}$ can be designed so that (11) is maximized, i.e.,

$$\max_{\boldsymbol{F}^{\mathrm{RF}} \in \mathbb{C}^{M_T \times N_T}} \ Tr\big[(\boldsymbol{F}^{\mathrm{RF}})^{\mathrm{H}}\boldsymbol{A}\boldsymbol{F}^{\mathrm{RF}}\big] \ \text{s.t.} \quad \boldsymbol{F}^{\mathrm{RF}} \in \mathscr{F}. \quad (12)$$

For architecture A1, a direct solution of (12) is to update $\boldsymbol{F}^{\mathrm{RF}}$ using the phase-angles of the dominant N_T eigenvectors $\boldsymbol{V}_{N_T} \in \mathbb{C}^{M_T \times N_T}$ of the estimated covariance matrix $\boldsymbol{A}$, i.e., [19]

$$\boldsymbol{F}^{\mathrm{RF}} = 1/\sqrt{M_T}\,e^{j(\boldsymbol{V}_{N_T})^{\angle}}, \quad (13)$$

where $(.)^{\angle}$ is a function that extracts the phase angles. For architecture A2, the authors in [19] simply update the ℓth block of $\boldsymbol{F}^{\mathrm{RF}}$ using the phase-angles of the dominant eigenvector of $\boldsymbol{A}_\ell$, where $\boldsymbol{A}_\ell, \ell \in \{1,\ldots,N_T\}$, is the channel covariance matrix of all antennas connected to the ℓth RF-chain. As will be shown latter, this simplified approach incurs a significant performance degradation. Next, we take another solution approach that is applicable to architectures A1 and A2.

We note that problem (12) can be equivalently written as

$$\max_{\{\boldsymbol{f}_\ell^{\mathrm{RF}}\}} \sum_\ell (\boldsymbol{f}_\ell^{\mathrm{RF}})^{\mathrm{H}} \boldsymbol{A} \boldsymbol{f}_\ell^{\mathrm{RF}}, \quad s.t. \quad \boldsymbol{f}_\ell^{\mathrm{RF}} \in \mathscr{F}, \forall \ell, \tag{14}$$

where $\boldsymbol{f}_\ell^{\mathrm{RF}} \in \mathbb{C}^{M_T}$ is the ℓth column of $\boldsymbol{F}^{\mathrm{RF}} \in \mathbb{C}^{M_T \times N_T}$. Clearly, problem (14) simplifies problem (12) by decoupling the optimization of $\boldsymbol{F}^{\mathrm{RF}}$ into N_T sub-problems. Although (14) is a non-convex problem, a solution of its ℓth sub-problem, i.e., $\boldsymbol{f}_\ell$ can be efficiently calculated using the proposed method in [9]. However, the problem formulation in (14) only considers maximizing the main diagonal elements of $(\boldsymbol{F}^{\mathrm{RF}})^{\mathrm{H}} \boldsymbol{A} \boldsymbol{F}^{\mathrm{RF}}$ but ignores the minimization of its off-diagonal entries, which correspond to the cross-correlation between the different columns of $\boldsymbol{F}^{\mathrm{RF}}$, i.e., $(\boldsymbol{f}_\ell^{\mathrm{RF}})^{\mathrm{H}} \boldsymbol{A} \boldsymbol{f}_k^{\mathrm{RF}}, \forall \ell \neq k$. Note that, using the proposed solution in (13), we are implicitly satisfying an additional set of contrarians of minimizing $|(\boldsymbol{f}_\ell^{\mathrm{RF}})^{\mathrm{H}} \boldsymbol{A} \boldsymbol{f}_k^{\mathrm{RF}}|^2, \forall \ell \neq k$, where we have $|(\boldsymbol{f}_\ell^{\mathrm{RF}})^{\mathrm{H}} \boldsymbol{A} \boldsymbol{f}_k^{\mathrm{RF}}|^2 = 0, \forall \ell \neq k$, if we simply set $\boldsymbol{F}^{\mathrm{RF}} = \boldsymbol{V}_{N_T}$. Therefore, we propose to reformulate problem (14) as

$$\max_{\{\boldsymbol{f}_\ell^{\mathrm{RF}}\}} \sum_\ell \left((\boldsymbol{f}_\ell^{\mathrm{RF}})^{\mathrm{H}} \boldsymbol{A} \boldsymbol{f}_\ell^{\mathrm{RF}} - \Upsilon_\ell \right) \text{ s.t. } \boldsymbol{f}_\ell^{\mathrm{RF}} \in \mathscr{F}, \forall \ell, \tag{15}$$

where $\Upsilon_\ell = \sum_{k \neq \ell} |(\boldsymbol{f}_k^{\mathrm{RF}})^{\mathrm{H}} \boldsymbol{A} \boldsymbol{f}_\ell^{\mathrm{RF}}|^2$, which represents the total *interference-leakage* caused by $\boldsymbol{f}_\ell^{\mathrm{RF}}$. We propose to solve (15) in an alternating fashion as following. Let $\boldsymbol{a}_\ell \stackrel{\text{def}}{=} \boldsymbol{A} \boldsymbol{f}_\ell^{\mathrm{RF}} \in \mathbb{C}^{M_T}, \ell \in \{1,\ldots,N_T\}$. Then, (15) can be rewritten as

$$\max_{\{\boldsymbol{f}_\ell^{\mathrm{RF}}\}} \sum_\ell \left(\boldsymbol{a}_\ell^{\mathrm{H}} \boldsymbol{f}_\ell^{\mathrm{RF}} - \sum_{k \neq \ell} |\boldsymbol{a}_k^{\mathrm{H}} \boldsymbol{f}_\ell^{\mathrm{RF}}|^2 \right) \text{ s.t. } \boldsymbol{f}_\ell^{\mathrm{RF}} \in \mathscr{F}, \forall \ell. \tag{16}$$

Problem (16) is non-convex, due to the constant modulus constrains. Note that when $\boldsymbol{f}_\ell^{\mathrm{RF}} \in \mathscr{F}$, then $\|\boldsymbol{f}_\ell^{\mathrm{RF}}\|^2 = 1$. Therefore, we propose to relax (16) and write it as

$$\max_{\{\boldsymbol{f}_\ell^{\mathrm{RF}}\}} \sum_\ell \left(\boldsymbol{a}_\ell^{\mathrm{H}} \boldsymbol{f}_\ell^{\mathrm{RF}} - \sum_{k \neq \ell} |\boldsymbol{a}_k^{\mathrm{H}} \boldsymbol{f}_\ell^{\mathrm{RF}}|^2 \right) \text{ s.t. } \|\boldsymbol{f}_\ell^{\mathrm{RF}}\|^2 = 1, \forall \ell, \tag{17}$$

which is a convex optimization problem. The optimal solution to $\boldsymbol{f}_\ell^{\mathrm{RF}}$ satisfying the Karush-Kuhn-Tucker (KKT) conditions is given by *Proposition 4.1*.

Proposition 4.1: The optimal solution to problem (17) w.r.t. $\boldsymbol{f}_\ell^{\mathrm{RF}}$ is given as

$$\boldsymbol{f}_{\ell,\mathrm{opt}}^{\mathrm{RF}} = \mu_\ell \left(\textstyle\sum_{k \neq \ell} \boldsymbol{a}_k \boldsymbol{a}_k^{\mathrm{H}} + \boldsymbol{I}_{M_T} \right)^{-1} \boldsymbol{a}_\ell, \tag{18}$$

where $\mu_\ell = 1/\| \left(\sum_{k \neq \ell} \boldsymbol{a}_k \boldsymbol{a}_k^{\mathrm{H}} + \boldsymbol{I}_{M_T} \right)^{-1} \boldsymbol{a}_\ell \|$. Please refer to Appendix B. Algorithm 1 summarizes the proposed approach for updating $\boldsymbol{F}^{\mathrm{RF}}$, where in step 5, we enforce the constant modulus constraints of $\boldsymbol{f}_{\ell,\mathrm{opt}}^{\mathrm{RF}} \in \mathscr{F}$. The symbol $\odot$ refers to an element-wise product and $\boldsymbol{\phi}_\ell \in \mathbb{B}^{M_T}$ is a binary vector. For architecture A1, $X_T \stackrel{\text{def}}{=} M_T$ and $\boldsymbol{\phi}_\ell \stackrel{\text{def}}{=} \boldsymbol{1}_{M_T}, \forall \ell$, i.e., an all ones vector. For architecture A2, $X_T \stackrel{\text{def}}{=} M_T / N_T$ and $\boldsymbol{\phi}_\ell$ contains ones at the indices $[(\ell-1)\frac{M_T}{N_T} + 1, \ldots, \ell \frac{M_T}{N_T}]$ and zeros elsewhere.

Remark 1: From our simulation results, we notice that Algorithm 1 always converges to a point where the cost function stops increasing or decreasing, although not monotonically. We leave the convergence proof to a future work, which can be carried out by interpreting problem (17) from a game theory view-point [23], [24], where the proof can be accomplished by showing its convergence to a Nash equilibrium point .

Algorithm 1 Proposed Analog Beamforming Design

Input: $\boldsymbol{A} = \sum_n \mathring{\boldsymbol{H}}_n^{\mathrm{H}} \mathring{\boldsymbol{H}}_n \in \mathbb{C}^{M_T \times M_T}$ and $\boldsymbol{F}^{\mathrm{RF}}[0] \in \mathscr{F}$

1: Compute $\boldsymbol{a}_\ell[0] \stackrel{\text{def}}{=} \boldsymbol{A} \boldsymbol{f}_\ell^{\mathrm{RF}}[0] \in \mathbb{C}^{M_T}, \forall \ell$, and set $n = 1$
2: **while** not converged **do**
3: **for** $\ell = 1$ to N_T **do**
4: Compute $\boldsymbol{f}_{\ell,\mathrm{opt}}$ using (18)
5: Set $\boldsymbol{f}_\ell^{\mathrm{RF}}[n] = 1/\sqrt{X_T} e^{j(\boldsymbol{f}_{\ell,\mathrm{opt}} \odot \boldsymbol{\phi}_\ell)^{\angle}}$
6: Set $\boldsymbol{a}_\ell[n] \stackrel{\text{def}}{=} \boldsymbol{A} \boldsymbol{f}_\ell^{\mathrm{RF}}[n] \in \mathbb{C}^{M_T}$
7: **end for**
8: **end while**

Output: $\boldsymbol{F}^{\mathrm{RF}}[n]$

V. Numerical Results

We assume that each channel matrix is generated using the model shown in [15], where we fix the number of channel paths $L = 8$. We define the signal-to-noise ratio (SNR) as $\rho = \frac{P}{\sigma^2}$ dB and assume $K = 4$, $M_R = 8$, and $N_s = 2$. For comparison, we show simulation results of the proposed HBF methods in [19], [20][1], [13][2], and the fully-digital ZF-based approach[3].

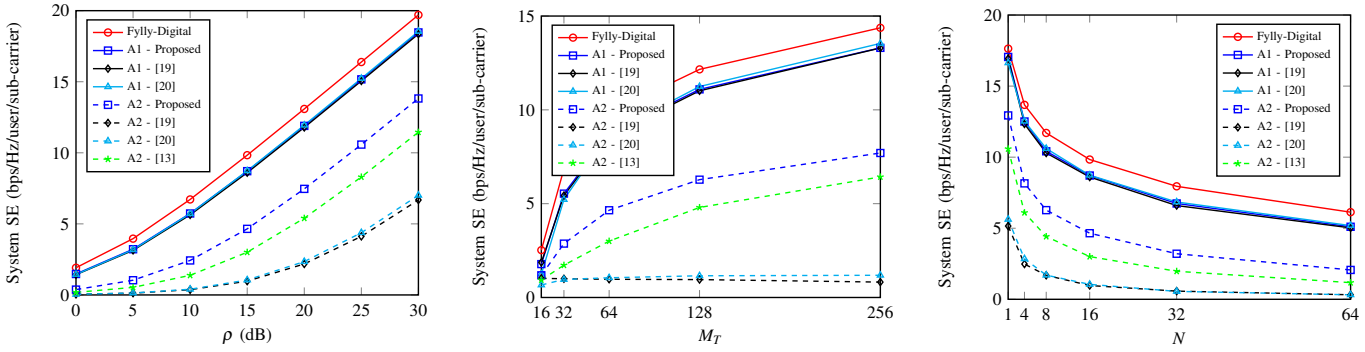
In Fig. 1, we show the system SE, normalized by the number of sub-carriers N and the number of MSs K, *versus* SNR in Fig. 1a, *versus* M_T in Fig. 1b, and *versus* N in Fig. 1c.

From Fig. 1, we can see that all the HBF algorithms using ABF architecture A1 (HBF-A1) achieve a comparable performance to each other with some performance gap compared to the fully-digital architecture. As expected, the SE with the ABF architecture A2 (HBF-A2) incurs a large performance gap compared to the HBF-A1 and the fully-digital counterparts. Nonetheless, we can see that the proposed HBF-A2 approach outperforms the other HBF-A2 methods in [13], [19], [20]. The proposed HBF-A2 methods in [19], [20] simplify the ABF design by completely decoupling the optimization between the ABF blocks. Differently, the proposed HBF-A2 method in [13] partially decouples the optimization between the ABF blocks by sequentially updating them, where in each step, the previously designed blocks are taken into account when designing the current block. On the other hand,

[1] In [20], the authors consider the fully-connected architecture A1. Analogous to [19], the proposed solution can be simply extended to the partially-connected architecture A2 by updating the ℓth block of the ABF matrix considering the unfolding of the channel-tensor of all antennas connected to the ℓth RF-chain.

[2] In [13], the authors consider the partially-connected architecture A2. Its extension to the fully-connected architecture A1 coincide with the proposed method in [19], hence, its corresponding simulation results are not shown.

[3] To have a fair comparison, we assumed that the fully-digital ZF transmit precoding matrix is calculated using (4), while assuming that $\boldsymbol{F}^{\mathrm{RF}} \stackrel{\text{def}}{=} \boldsymbol{I}_{M_T}$ and $\boldsymbol{W}_k^{\mathrm{RF}} \stackrel{\text{def}}{=} \boldsymbol{U}_k \in \mathbb{C}^{M_R \times N_T}$, where $\boldsymbol{U}_k$ consists of the N_R dominant eigenvectors of the channel covariance matrix $\boldsymbol{A}_k$ given by (7).



(a) $M_T = 64$ and $N = 16$ (b) $\rho = 15$ dB and $N = 16$ (c) $\rho = 15$ dB and $M_T = 64$

Fig. 1: System SE (bps/Hz/user/sub-carrier) vs. SNR, vs. number of transmit antennas M_T, and vs. number of sub-carriers N.

the proposed method in this paper updates the ABF matrix sequentially and iteratively with the aim of finding a good balance between maximizing the received signal in each block and minimizing the interference leakage to the other blocks.

Moreover, we can see from Fig. 1c that the SE per sub-carrier per user decreases when the number of sub-carriers N increases. This can be explained by the following two reasons. The first reason is due to the assumed power allocation approach (recall that $P_n \overset{\text{def}}{=} P/N$). Therefore, when N increases, the power allocation per sub-carrier P_n decreases. The second reason, and more importantly, is that the capacity of every equivalent channel $\boldsymbol{H}_{k,n}^{\text{eff}}$ decreases when N increases. This is due to the fact that the common transmit ABF matrix $\boldsymbol{F}^{\text{RF}}$ is computed from $\boldsymbol{A}$ that is given by (11). Therefore, when N increases, $\boldsymbol{F}^{\text{RF}}$ will span a larger subspace that is formed by N covariance matrices, which results in a capacity reduction of each $\boldsymbol{H}_{k,n}^{\text{eff}}, \forall n,k$. Note that the un-normalized SE increases with the increasing N, which can be checked by multiplying each point in Fig. 1c by $K = 4$ and the respective value of N.

VI. Conclusions

In this paper, a new HBF design method is proposed for multi-user massive MIMO-OFDM millimeter-wave systems, which updates the ABF matrix sequentially and iteratively with the aim of finding a good balance between maximizing the received signal in each block and minimizing the interference leakage to the other blocks. Simulation results reveal that the proposed method has a comparable performance with the benchmark methods in the fully-connected HBF architectures, while it significantly outperforming them in the partially-connected case.

Appendix A
Derivation of Equation (11)

Substituting $\boldsymbol{F}_{k,n}^{\text{BB}}$ given by (4) into (10), then we can write

$$\text{SE} = KN_s \sum_n \log\left(1 + \frac{P_n}{\sigma^2 \|\boldsymbol{F}^{\text{RF}} \bar{\boldsymbol{F}}_n^{\text{BB}}\|_{\text{F}}^2}\right), \tag{19}$$

where we have used the fact that $(\bar{\boldsymbol{F}}_n^{\text{BB}})^{\text{H}} \bar{\boldsymbol{H}}_n^{\text{H}} \bar{\boldsymbol{H}}_n \bar{\boldsymbol{F}}_n^{\text{BB}} = \boldsymbol{I}_{KN_s}$. Then, after substituting $\bar{\boldsymbol{F}}_n^{\text{BB}} = \bar{\boldsymbol{H}}_n^{\text{H}} (\bar{\boldsymbol{H}}_n \bar{\boldsymbol{H}}_n^{\text{H}})^{-1}$, the term $\zeta \overset{\text{def}}{=} \|\boldsymbol{F}^{\text{RF}} \bar{\boldsymbol{F}}_n^{\text{BB}}\|_F^2$ can be written as

$$\begin{aligned} \zeta &= Tr[(\bar{\boldsymbol{H}}_n \bar{\boldsymbol{H}}_n^{\text{H}})^{-1} \bar{\boldsymbol{H}}_n (\boldsymbol{F}^{\text{RF}})^H \boldsymbol{F}^{\text{RF}} \bar{\boldsymbol{H}}_n^{\text{H}} (\bar{\boldsymbol{H}}_n \bar{\boldsymbol{H}}_n^{\text{H}})^{-1}] \\ &\approx Tr[(\bar{\boldsymbol{H}}_n \bar{\boldsymbol{H}}_n^{\text{H}})^{-1}] \end{aligned} \tag{20}$$

where the approximation follows by noting that $(\boldsymbol{F}^{\text{RF}})^{\text{H}} \boldsymbol{F}^{\text{RF}}$ can be approximated as $(\boldsymbol{F}^{\text{RF}})^{\text{H}} \boldsymbol{F}^{\text{RF}} \approx \boldsymbol{I}_{N_T}$, especially when M_T is large [7]. Using (20), we can approximate (19) as

$$\begin{aligned} \text{SE} &\approx KN_s \sum_n \log\Big(1 + \frac{P_n}{\sigma^2 Tr[(\mathring{\boldsymbol{H}}_n \boldsymbol{F}^{\text{RF}} (\boldsymbol{F}^{\text{RF}})^{\text{H}} \mathring{\boldsymbol{H}}_n^{\text{H}})^{-1}]}\Big), \\ &\overset{(a)}{\leq} KN_s \log\Big(1 + \frac{P_n}{\sigma^2} \sum_n \frac{1}{Tr[(\mathring{\boldsymbol{H}}_n \boldsymbol{F}^{\text{RF}} (\boldsymbol{F}^{\text{RF}})^{\text{H}} \mathring{\boldsymbol{H}}_n^{\text{H}})^{-1}]}\Big) \qquad (21) \\ &\overset{(b)}{\leq} KN_s \log\Big(1 + \frac{P_n}{\sigma^2 (KN_R)^2} \sum_n Tr[\mathring{\boldsymbol{H}}_n \boldsymbol{F}^{\text{RF}} (\boldsymbol{F}^{\text{RF}})^{\text{H}} \mathring{\boldsymbol{H}}_n^{\text{H}}]\Big) \\ &= KN_s \log\Big(1 + \frac{P_n}{\sigma^2 (KN_R)^2} Tr[(\boldsymbol{F}^{\text{RF}})^{\text{H}} \boldsymbol{A} \boldsymbol{F}^{\text{RF}}]\Big), \qquad (22) \end{aligned}$$

where $\boldsymbol{A} = \sum_n \mathring{\boldsymbol{H}}_n^{\text{H}} \mathring{\boldsymbol{H}}_n$, (a) follows from Jensen's inequality, and (b) follows from using the fact that $Tr[\boldsymbol{C}^{-1}] \geq n^2 / Tr[\boldsymbol{C}]$ for any positive definite matrix $\boldsymbol{C} \in \mathbb{C}^{n \times n}$.

Appendix B
Proof of Proposition 4.1

We define the Lagrangian function of (17) as

$$L(\boldsymbol{f}_\ell, \lambda_\ell) = -\sum_\ell \big(\boldsymbol{a}_\ell^{\text{H}} \boldsymbol{f}_\ell - \sum_{k \neq j} |\boldsymbol{a}_k^{\text{H}} \boldsymbol{f}_\ell|^2\big) + \lambda_\ell (\|\boldsymbol{f}_\ell\|^2 - 1), \tag{23}$$

where λ_ℓ is the Lagrangian multiplier associated with problem (17). Taking the gradient of $L(\boldsymbol{f}_\ell, \lambda_\ell)$ w.r.t. $\boldsymbol{f}_\ell$ as

$$\frac{\partial L(\boldsymbol{f}_\ell, \lambda_\ell)}{\partial \boldsymbol{f}_\ell} = -\boldsymbol{a}_\ell + \sum_{k \neq \ell} \boldsymbol{a}_k \boldsymbol{a}_k^{\text{H}} \boldsymbol{f}_\ell + \lambda_\ell \boldsymbol{f}_\ell. \tag{24}$$

Equating (24) to zero and solving for $\boldsymbol{f}_\ell$, we have $\boldsymbol{f}_{\ell,\text{opt}} = \big(\sum_{k \neq \ell} \boldsymbol{a}_k \boldsymbol{a}_k^{\text{H}} + \lambda_\ell \boldsymbol{I}_{M_T}\big)^{-1} \boldsymbol{a}_\ell$. Since $\|\boldsymbol{f}_\ell\|^2 = 1$, we can simply update $\boldsymbol{f}_\ell$ as $\boldsymbol{f}_{\ell,\text{opt}} = \mu_\ell \big(\sum_{k \neq \ell} \boldsymbol{a}_k \boldsymbol{a}_k^H + \boldsymbol{I}\big)^{-1} \boldsymbol{a}_\ell$, where $\mu_\ell = 1 / \|\big(\sum_{k \neq \ell} \boldsymbol{a}_k \boldsymbol{a}_k^{\text{H}} + \boldsymbol{I}_{M_T}\big)^{-1} \boldsymbol{a}_\ell\|$.

REFERENCES

[1] A. L. Swindlehurst, E. Ayanoglu, P. Heydari, and F. Capolino, "Millimeter-wave massive MIMO: the next wireless revolution?" *IEEE Commun. Mag.*, vol. 52, no. 9, pp. 56–62, Sep. 2014.

[2] R. W. Heath, N. Gonzlez-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 436–453, Apr. 2016.

[3] T. E. Bogale, L. B. Le, A. Haghighat, and L. Vandendorpe, "On the number of RF chains and phase shifters, and scheduling design with hybrid analogdigital beamforming," *IEEE Trans. Wireless Commun.*, vol. 15, no. 5, pp. 3311–3326, May 2016.

[4] O. E. Ayach, R. W. Heath, S. Abu-Surra, S. Rajagopal, and Z. Pi, "Low complexity precoding for large millimeter wave MIMO systems," in *Proc. IEEE International Conference on Communications (ICC)*, Jun. 2012, pp. 3724–3729.

[5] L. Liang, W. Xu, and X. Dong, "Low-complexity hybrid precoding in massive multiuser MIMO systems," *IEEE Commun. Lett.*, vol. 3, no. 6, pp. 653–656, Dec. 2014.

[6] W. Ni and X. Dong, "Hybrid block diagonalization for massive multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 64, no. 1, pp. 201–211, Jan. 2016.

[7] F. Sohrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale antenna arrays," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 501–513, Apr. 2016.

[8] X. Wu, D. Liu, and F. Yin, "Hybrid beamforming for multi-user massive MIMO systems," *IEEE Trans. Commun.*, vol. 66, no. 9, pp. 3879–3891, Sep. 2018.

[9] K. Ardah, G. Fodor, Y. C. B. Silva, W. C. Freitas, and F. R. P. Cavalcanti, "A unifying design of hybrid beamforming architectures employing phase shifters or switches," *IEEE Trans. Veh. Technol.*, vol. 67, no. 11, pp. 11 243–11 247, Nov. 2018.

[10] R. Mndez-Rial, C. Rusu, N. Gonzlez-Prelcic, A. Alkhateeb, and R. W. Heath, "Hybrid MIMO architectures for millimeter wave communications: Phase shifters or switches?" *IEEE Access*, vol. 4, pp. 247–267, 2016.

[11] J. Zhang, M. Haardt, I. Soloveychik, and A. Wiesel, "A channel matching based hybrid analog-digital strategy for massive multi-user MIMO downlink systems," in *Proc. IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, Jul. 2016, pp. 1–5.

[12] D. Zhang, Y. Wang, X. Li, and W. Xiang, "Hybridly connected structure for hybrid beamforming in mmwave massive MIMO systems," *IEEE Trans. Commun.*, vol. 66, no. 2, pp. 662–674, Feb. 2018.

[13] X. Gao, L. Dai, S. Han, C. I, and R. W. Heath, "Energy-efficient hybrid analog and digital precoding for mmwave MIMO systems with large antenna arrays," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 998–1009, Apr. 2016.

[14] K. Ardah, G. Fodor, Y. C. B. Silva, W. Cruz, and A. Almeida, "Hybrid analog-digital beamforming design for SE and EE maximization in massive MIMO networks," *IEEE Trans. Veh. Technol.*, pp. 1–1, 2019.

[15] J. Zhang, A. Wiesel, and M. Haardt, "Low rank approximation based hybrid precoding schemes for multi-carrier single-user massive MIMO systems," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Mar. 2016, pp. 3281–3285.

[16] D. Zhu, B. Li, and P. Liang, "A novel hybrid beamforming algorithm with unified analog beamforming by subspace construction based on partial CSI for massive MIMO-OFDM systems," *IEEE Trans. Commun.*, vol. 65, no. 2, pp. 594–607, Feb 2017.

[17] Y. Kwon, J. Chung, and Y. Sung, "Hybrid beamformer design for mmwave wideband multi-user MIMO-OFDM systems : (invited paper)," in *Proc. IEEE 18th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Jul. 2017, pp. 1–5.

[18] F. Sohrabi and W. Yu, "Hybrid analog and digital beamforming for mmwave OFDM large-scale antenna arrays," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 7, pp. 1432–1443, Jul. 2017.

[19] Y. Liu and J. Wang, "Low-complexity OFDM-based hybrid precoding for multiuser massive MIMO systems," *IEEE Wireless Commun. Lett.*, pp. 1–1, 2019.

[20] D. Zhang, Y. Wang, X. Li, and W. Xiang, "Hybrid beamforming for downlink multiuser millimetre wave MIMO-OFDM systems," *IET Communications*, vol. 13, no. 11, pp. 1557–1564, 2019.

[21] S. S. Christensen, R. Agarwal, E. De Carvalho, and J. M. Cioffi, "Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 4792–4799, Dec. 2008.

[22] K. Ardah, Y. C. B. Silva, and F. R. P. Cavalcanti, "Decentralized linear transceiver design in multicell MIMO broadcast channels," *Journal of Communication and Information Systems (JCIS)*, vol. 32, no. 1, pp. 102–115, Oct. 2017.

[23] J. Ni and H. Xiao, "Game theoretic approach for joint transmit beamforming and power control in cognitive radio MIMO broadcast channels," *EURASIP Journal on Wireless Communications and Networking*, vol. 2016, no. 1, p. 98, Apr. 2016.

[24] B. Basutli and S. Lambotharan, "Game-theoretic beamforming techniques for multiuser multi-cell networks under mixed quality of service constraints," *IET Signal Processing*, vol. 11, no. 5, pp. 631–639, 2017.