Hybrid Beamforming Design in Multi-Cell MU-MIMO Systems with Per-RF or Per-Antenna Power Constraints

Christo Kurisummoottil Thomas, Dirk Slock
EURECOM, Sophia-Antipolis, France, Email:{kurisumm,slock}@eurecom.fr

***Abstract*—This work deals with hybrid beamforming for the MIMO Interfering Broadcast Channel (IBC), i.e. the Multi-Input Multi-Output (MIMO) Multi-User (MU) Multi-Cell (MC) downlink (DL) channel. Hybrid beamforming (HBF) is a low complexity alternative for fully digital precoding in Massive MIMO systems. Hybrid architectures involve a combination of digital and analog processing that enables both beamforming and multiplexing gains. We consider BF design by maximizing the Weighted Sum Rate (WSR) for the case of Perfect Channel State Information at the Transmitter (CSIT). We optimize the WSR using minorization and alternating optimization, the result of which is observed to converge fast. The design is proposed for both fully and partially connected analog BF architectures. Moreover, we consider the BF design under realistic scenarios with per-RF (radio frequency) chain or per-antenna power constraints, leading to novel interference leakage aware water filling procedures. Simulation results illustrate the good WSR performance of the designs and the gains over naive constraint satisfaction approaches.**

***Index Terms*—Hybrid beamforming, massive MIMO, per RF power constraints, perfect CSIT.**

I. Introduction

In this paper, Tx may denote transmit/transmitter/ transmission, Rx may denote receive/receiver/reception, BF may denote beamforming/beamformer. Hybrid beamforming (HBF) is a two-stage architecture in which the BF is constructed by concatenation of a low-dimensional precoder (digital BF) and an analog BF, with the number of RF chains less than the number of antennas. This technique was first introduced in [1], with the analog precoder implemented using phase shifters. Hybrid precoding designs for single user (SU) systems can be found in [2]–[4]. The authors in [2] propose near-optimal solutions based on the formulation of sparse signal recovery for a SU mmWave system. In [5] the authors use Weighted Sum Rate (WSR) maximization as the target optimization criterion for the HBF design. However, they optimize the analog phasors using Tx power minimization criteria and a zero-forcing (ZF) solution for the digital precoder. In [6], we propose a Weighted Sum Mean Squared Error (WSMSE) based approach for the joint max WSR design of digital and analog beamformers for a multi-cell (MC) multi-user (MU) MIMO system. [7], [8] propose a HBF design using sparse formulations and approximating the minimum MSE (MMSE). In [7], orthogonal matching pursuit (OMP) based algorithms are used to successively select RF BF vectors from a set of candidate vectors and the corresponding digital BF are optimized by least squares fitting. [9] uses the decomposition of analog BF vectors and antenna array response vectors into Kronecker products of (unit modulus) factors. By exploiting that the inner product of Kronecker structured vectors is the product of factor inner products, different factors of the analog BF vectors are designed either for interference nulling or enhancing the signal power. While this interesting design proposes to exploit only multipath CSIT for the intercell interference, it is only applicable to Kronecker structured array responses and is suboptimal due to the ZF constraints and especially due to the number of phase variables (factors) that grows only logarithmically with the number of antennas.

In contrast to the conventional (sum-)power constraint (SPC) on the base station (BS), this paper considers a more realistic scenario with additionally per-RF or per-antenna power constraints (PRFPC/PAPC). In practice, each RF chain is equipped with a power amplifier and its linear range of the PA combined with Peak to Average Power Ratio (PAPR) considerations lead to a power constraint per power amplifier. Another scenario is the case of a distributed system where a central BS is connected via a high speed backbone network to remote antennas. Fully digital BF designs with PAPC can be found in [10]–[13]. [10] focuses on the design of BF vectors for a MISO system to minimize the per-antenna power while enforcing a set of SINR constraints for each user. ZF BF design with PAPC are discussed in [11], while [12] utilizes UL/DL duality of the sum MSE for the precoder design. Existing approaches for this problem are based on either interior point methods that do not favorably scale with the problem size or subgradient methods [14] that have very slow convergence rate.

A. Contributions of this paper

- We propose a novel HBF design (for both fully or partially connected structures) based on the WSR criterion which is simplified using minorization and alternating optimization. To the authors' best knowledge, this is the first paper to propose HBF design under the more realistic scenario of per-RF or per-antenna power constraints.
- We propose a novel analog phasors design using a deterministic annealing approach, leading to the only existing solution which avoids the big problem of local optima. Performance is significantly better than state of the art

solutions (compared e.g. to the alternating optimization for phasor design using the WSMSE method as in [6]).

- We propose a novel interference leakage aware water filling (ILA-WF) for the stream power optimization, even for just SPC, but also augmented with PRFPC or PAPC. We propose to solve the resulting convex Lagrange dual problem by alternating bisection but may other solutions can be considered. The ILA-WF allows automatic discovery of the sustainable number of streams per user in MIMO channels.
- The simulation results indicate that due to this power optimization approach and the avoidance of auxiliary Rx filters, the proposed algorithm can converge much faster compared to existing approaches.

Notation: In the following, boldface lower-case and upper-case characters denote vectors and matrices respectively. the operators $E[\cdot]$, tr$\{\cdot\}$, $(\cdot)^H$, $(\cdot)^T$ represents expectation, trace, conjugate transpose and transpose respectively. A circularly complex Gaussian random vector $\mathbf{x}$ with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Theta}$ is distributed as $\mathbf{x} \sim \mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Theta})$. $\mathbf{V}_{max}(\mathbf{A}, \mathbf{B})$ or $\mathbf{V}_{1:M}(\mathbf{A}, \mathbf{B})$ represents (normalized) dominant generalized eigenvector or the matrix formed by M (normalized) dominant generalized eigenvectors of $\mathbf{A}$ and $\mathbf{B}$. $vec(\mathbf{X})$ represents the vector obtained by stacking each of the columns of $\mathbf{X}$ and $unvec(\mathbf{X})$ represents the reverse operation. Further, $\mathbf{A} \succeq 0$ means that all the elements of the matrix $\mathbf{A}$ are ≥ 0.

II. Multi-User MIMO System Model

In this paper we shall consider a multi-stream approach. So, consider a multi-cell multi-user downlink (MC MU DL) system with C cells and a total of K users. $\mathbf{H}_{k,b_i}$ represents the $N_k \times N_t^{b_i}$ MIMO channel between user k (associated to BS b_k) and BS b_i, with Tx side covariance matrix $\mathrm{E}\left[\mathbf{H}_{k,c}^H \mathbf{H}_{k,c}\right] = \boldsymbol{\Theta}_k^c$. User k receives

$$\mathbf{y}_k = \mathbf{H}_{k,b_k}\, \mathbf{V}^{b_k}\, \mathbf{G}_k\, \mathbf{s}_k + \sum_{i \neq k} \mathbf{H}_{k,b_i}\, \mathbf{V}^{b_i}\, \mathbf{G}_i\, \mathbf{s}_i + \mathbf{v}_k, \quad (1)$$

where $\mathbf{s}_k$ is the $d_k \times 1$ intended signal stream vector (entries are white, unit variance). We are considering a noise whitened signal representation so that we get for the noise $\mathbf{v}_k \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{N_k})$. The analog BF $\mathbf{V}^c$ for base station c is of dimension $N_t^c \times M^c$ where M^c is the number of RF chains. The $M^c \times d_k$ digital BF is $\mathbf{G}_k$. Note that the fully digital case can be recovered from our design by substituting $M^c = N_t^c$ and $\mathbf{V}^c = \mathbf{I}$.

There exist two types of phased arrays at mmWave frequencies: (i) passive phased arrays and (ii) active phased arrays [15]. Though passive phase shifters incur some power loss, they require only the same number of power amplifiers as RF units, leading to PRFPC considerations. Since there is a clear trend towards active systems, we also consider PAPC in section III-D. Although we do not model the power loss (which would complete the picture), simulations show that due to the reduced number of power constraints, passive systems with PRFPC have some power efficiency gain over active systems with PAPC.

III. WSR Maximization via Minorization and Alternating Optimization

Consider the optimization of the hybrid beamforming design using WSR maximization of the Multi-cell MU-MIMO system:

$$\max_{\mathbf{V},\mathbf{G}} WSR(\mathbf{V}, \mathbf{G}) = \max_{\mathbf{V},\mathbf{G}} \sum_{k=1}^{K} u_k \ln\det(\mathbf{R}_{\overline{k}}^{-1} \mathbf{R}_k), \quad (2)$$

where the u_k are the rate weights, $\mathbf{G}$ represents the collection of BFs $\mathbf{G}_k$, $\mathbf{V}$ the analog BFss $\mathbf{V}^c$. Also, we define

$$\begin{aligned} \mathbf{R}_{\overline{k}} &= \sum_{i=1, i\neq k}^{K} \mathbf{H}_{k,b_i} \mathbf{Q}_i \mathbf{H}_{k,b_i}^H + \mathbf{I}_{N_k}, \\ \mathbf{R}_k &= \sum_{i=1}^{K} \mathbf{H}_{k,b_i} \mathbf{Q}_i \mathbf{H}_{k,b_i}^H + \mathbf{I}_{N_k},\ \mathbf{Q}_k = \mathbf{V}^{b_k} \mathbf{G}_k \mathbf{G}_k^H \mathbf{V}^{b_k H} \end{aligned} \quad (3)$$

where ($\mathbf{R}_k$) $\mathbf{R}_{\overline{k}}$ is the Rx (signal plus) interference plus noise covariance matrix and $\mathbf{Q}_k$ is the Tx covariance matrix for user k. The per-RF power constraints (PRFPC) at BS c can be written as,

$$\sum_{k: b_k = c} \left[\mathbf{G}_k \mathbf{G}_k^H\right]_{i,i} \leq a_i^c,\ i = 1, ..., M^c, \quad (4)$$

where $\left[\mathbf{G}_k \mathbf{G}_k^H\right]_{i,i}$ represents the i^{th} diagonal element of $\mathbf{G}_k \mathbf{G}_k^H$. Further, also total Tx power constraints need to be satisfied, $\sum_{k: b_k = c} \mathrm{tr}\{Q_k\} \leq P^c$. The WSR problem is non-concave in the $\mathbf{Q}_k$ due to the interference terms. Therefore finding the global optimum is challenging. In order to render a feasible solution, we consider constructing a minorizer based on the difference of convex functions (DC programming) approach. Consider the dependence of WSR on $\mathbf{Q}_k$ alone.

$$\begin{aligned} WSR &= u_k \ln\det(\mathbf{R}_{\overline{k}}^{-1} \mathbf{R}_k) + WSR_{\overline{k}}, \\ WSR_{\overline{k}} &= \sum_{i=1, \neq k}^{K} u_i \ln\det(\mathbf{R}_{\overline{i}}^{-1} \mathbf{R}_i), \end{aligned} \quad (5)$$

where $\ln\det(\mathbf{R}_{\overline{k}}^{-1} \mathbf{R}_k)$ is concave in $\mathbf{Q}_k$ and $WSR_{\overline{k}}$ is convex in $\mathbf{Q}_k$. Since a linear function is simultaneously convex and concave, DC programming [16] introduces the first order Taylor series expansion of $WSR_{\overline{k}}$ in $\mathbf{Q}_k$ around $\widehat{\mathbf{Q}}$ (i.e. all $\widehat{\mathbf{Q}}_i$).

$$\begin{aligned} \underline{WSR}_{\overline{k}}(\mathbf{Q}_k, \widehat{\mathbf{Q}}) &= WSR_{\overline{k}}(\widehat{\mathbf{Q}}_k, \widehat{\mathbf{Q}}) - \mathrm{tr}\left\{(\mathbf{Q}_k - \widehat{\mathbf{Q}}_k) \widehat{\mathbf{A}}_k\right\}, \\ \widehat{\mathbf{A}}_k &= -\left.\frac{\partial WSR_{\overline{k}}(\mathbf{Q}_k, \widehat{\mathbf{Q}})}{\partial \mathbf{Q}_k}\right|_{\widehat{\mathbf{Q}}_k} = \sum_{i=1, \neq k}^{K} u_i \mathbf{H}_{i,b_k}^H (\widehat{\mathbf{R}}_{\overline{i}}^{-1} - \widehat{\mathbf{R}}_i^{-1}) \mathbf{H}_{i,b_k}. \end{aligned} \quad (6)$$

Note that the linearized tangent expression $\underline{WSR}_{\overline{k}}$ constitutes a (touching) lower bound for $WSR_{\overline{k}}$ via $-\mathrm{tr}\{\mathbf{R}^{-1} \boldsymbol{\Delta}\} \leq -\ln\det(\mathbf{R}^{-1}(\mathbf{R} + \boldsymbol{\Delta}))$ and $\mathbf{R}_k \geq \mathbf{R}_{\overline{k}}$. Hence the DC approach is also a minorization approach [17], regardless of the (re)parameterization of $\mathbf{Q}$. Now let $\widehat{\mathbf{B}}_k = \mathbf{H}_{k,b_k}^H \widehat{\mathbf{R}}_{\overline{k}}^{-1} \mathbf{H}_{k,b_k}$, $\boldsymbol{\Psi}_c = \mathrm{diag}\,(\Psi_{c,1}, ..., \Psi_{c,M^c})$ represents the Lagrange multipliers associated with the per-RF power constraints $\boldsymbol{\Phi}_c = \mathrm{diag}\,(a_1^c, ..., a_{M^c}^c)$. $\boldsymbol{\Psi}$ represents the set of all $\boldsymbol{\Psi}_c$ and $\boldsymbol{\Lambda} = \mathrm{diag}\,(\lambda_1, ..., \lambda_C)$. Then, dropping constant terms, reparameterizing the $\mathbf{Q}_k$ as in (3), performing this linearization for all users, and augmenting the WSR cost function with the Tx power constraints, we get the Lagrangian (7) which gets maximized alternatingly [17] between digital and analog BF

$$\begin{aligned}\mathcal{L}(\mathbf{V},\mathbf{G},\boldsymbol{\Lambda},\boldsymbol{\Psi}) = \sum_{c=1}^{C}\lambda_c P^c + \sum_{c=1}^{C}\text{tr}\{\boldsymbol{\Psi}_c\boldsymbol{\Phi}_c\} + \\ \sum_{k=1}^{K} u_k \ln\det\left(\mathbf{I}+\mathbf{G}_k^H\mathbf{V}^{b_k H}\widehat{\mathbf{B}}_k\mathbf{V}^{b_k}\mathbf{G}_k\right) \\ -\text{tr}\left\{\mathbf{G}_k^H\left(\mathbf{V}^{b_k H}\left(\widehat{\mathbf{A}}_k+\lambda_{b_k}\mathbf{I}\right)\mathbf{V}^{b_k}+\boldsymbol{\Psi}_{b_k}\right)\mathbf{G}_k\right\}\,.\end{aligned} \tag{7}$$

A. Digital BF Design

Maximizing (7) w.r.t. $\mathbf{G}_k$ leads to the KKT conditions

$$\begin{aligned}\mathbf{V}^{b_k H}\widehat{\mathbf{B}}_k\mathbf{V}^{b_k}\mathbf{G}_k = \left(\mathbf{V}^{b_k H}\left(\widehat{\mathbf{A}}_k+\lambda_{b_k}\mathbf{I}\right)\mathbf{V}^{b_k}+\boldsymbol{\Psi}_{b_k}\right)\mathbf{G}_k \\ \times\ \tfrac{1}{u_k}(\mathbf{I}+\mathbf{G}_k^H\mathbf{V}^{b_k H}\widehat{\mathbf{B}}_k\mathbf{V}^{b_k}\mathbf{G}_k)\end{aligned} \tag{8}$$

with solution d_k dominant generalized eigenvectors (g.e.v.)

$$\mathbf{G}_k^{'} = \mathbf{V}_{1:d_k}\left(\mathbf{V}^{b_k H}\widehat{\mathbf{B}}_k\mathbf{V}^{b_k}, \mathbf{V}^{b_k H}\left(\widehat{\mathbf{A}}_k+\lambda_{b_k}\mathbf{I}\right)\mathbf{V}^{b_k}+\boldsymbol{\Psi}_{b_k}\right) \tag{9}$$

with eigenvalues $\boldsymbol{\Sigma}_k = \frac{1}{u_k}(\mathbf{I}+\mathbf{G}_k^H\mathbf{V}^{b_k H}\widehat{\mathbf{B}}_k\mathbf{V}^{b_k}\mathbf{G}_k)$. The gradient in (8), which would be the same with $\underline{WSR}$ replaced by WSR, leads to g.e.v. conditions whereas maximizing $\mathcal{L}$ in (7) leads to select the dominant g.e.v. Let $\mathbf{S}_k = \mathbf{G}_k^{'H}\mathbf{V}^{b_k H}\widehat{\mathbf{B}}_k\mathbf{V}^{b_k}\mathbf{G}_k'$, $\mathbf{W}_k = \mathbf{G}_k^{'H}\mathbf{V}^{b_k H}\widehat{\mathbf{A}}_k\mathbf{V}^{b_k}\mathbf{G}_k'$, and $\mathbf{T}_k(\lambda_{b_k},\boldsymbol{\Psi}_{b_k}) = \mathbf{W}_k + \mathbf{G}_k^{'H}(\lambda_{b_k}\mathbf{V}^{b_k H}\mathbf{V}^{b_k}+\boldsymbol{\Psi}_{b_k})\mathbf{G}_k'$. Note that g.e.v. diagonalize $\mathbf{S}_k$, $\mathbf{T}_k(\lambda_{b_k}^{(j-1)},\boldsymbol{\Psi}_{b_k}^{(j-1)})$ and $\boldsymbol{\Sigma}_k$ (see further for iteration index (j)). As g.e.v. are normalized, the stream powers $\mathbf{P}_k \geq 0$ (diagonal) need to be optimized separately. But this is straightforward from (7): substituting $\mathbf{G}_k = \mathbf{G}_k'\mathbf{P}_k^{\frac{1}{2}}$ in (7) yields,

$$\begin{aligned}\mathcal{L}(\mathbf{V},\mathbf{G}',\mathbf{P},\boldsymbol{\Lambda},\boldsymbol{\Psi}) = \textstyle\sum_{c=1}^{C}(\lambda_c P^c + \text{tr}\{\boldsymbol{\Psi}_c\boldsymbol{\Phi}_c\}) + \\ \textstyle\sum_{k=1}^{K}[u_k\ln\det(\mathbf{I}+\mathbf{S}_k\mathbf{P}_k) - \text{tr}\{\mathbf{T}_k(\lambda_{b_k},\boldsymbol{\Psi}_{b_k})\,\mathbf{P}_k\}].\end{aligned} \tag{10}$$

B. Optimization of Power Variables: $\mathbf{P},\boldsymbol{\Lambda},\boldsymbol{\Psi}$

The optimization of (10) w.r.t. $\mathbf{P}_k$ leads to the following interference leakage aware water filling (ILA-WF)

$$\begin{aligned}\left(u_k(\mathbf{W}_k+\mathbf{G}_k^{'H}(\lambda_{b_k}\mathbf{V}^{b_k H}\mathbf{V}^{b_k}+\boldsymbol{\Psi}_{b_k})\mathbf{G}_k')^{-1} - \mathbf{S}_k^{-1}\right)^{+} \\ = \mathbf{P}_k^o(\lambda_{b_k},\boldsymbol{\Psi}_{b_k}) = \left(u_k\,\mathbf{T}_k^{-1}(\lambda_{b_k},\boldsymbol{\Psi}_{b_k}) - \mathbf{S}_k^{-1}\right)^{+}\end{aligned} \tag{11}$$

where, $(\mathbf{X})^{+}$ denotes the positive semi-definite part of Hermitian $\mathbf{X}$. We substitute the optimized power distribution $\mathbf{P}_k^o(\lambda_{b_k},\boldsymbol{\Psi}_{b_k})$ in (10) yielding the Lagrange dual function

$$\begin{aligned}g(\boldsymbol{\Lambda},\boldsymbol{\Psi}) = \mathcal{L}(\mathbf{V},\mathbf{G}',\mathbf{P}^o(\boldsymbol{\Lambda},\boldsymbol{\Psi}),\boldsymbol{\Lambda},\boldsymbol{\Psi}) = \sum_{c=1}^{C} g_c(\lambda_c,\boldsymbol{\Psi}_c) \\ g_c(\lambda_c,\boldsymbol{\Psi}_c) = \lambda_c P^c + \text{tr}\{\boldsymbol{\Psi}_c\boldsymbol{\Phi}_c\} \\ +\textstyle\sum_{k:b_k=c}[u_k\ln\det(\mathbf{I}+\mathbf{S}_k\mathbf{P}_k^o) - \text{tr}\{\mathbf{T}_k(\lambda_{b_k},\boldsymbol{\Psi}_{b_k})\,\mathbf{P}_k^o\}]\end{aligned} \tag{12}$$

where we omitted the dependence of $g()$ on $\mathbf{V},\mathbf{G}'$, which are currently fixed in the alternating optimization process, as we maximize over $\mathbf{P}$. $\boldsymbol{\Lambda},\boldsymbol{\Psi}$ should be chosen such that $g(\boldsymbol{\Lambda},\boldsymbol{\Psi})$ is finite. Further, the non-negativity of $\boldsymbol{\Lambda}$ and $\boldsymbol{\Psi}$ imposes constraints on the dual objective function. Formally, the Lagrangian dual problem per cell can be stated as follows:

$$\min_{\lambda_c,\boldsymbol{\Psi}_c} g_c(\lambda_c,\boldsymbol{\Psi}_c) \text{ subject to } \lambda_c \geq 0, \boldsymbol{\Psi}_c \succeq 0\,,\ \forall c\,. \tag{13}$$

Since the dual function $g_c(\lambda_c,\boldsymbol{\Psi}_c)$ is the pointwise supremum of a family of functions of $\lambda_c,\boldsymbol{\Psi}_c$, it is convex [18] and the globally optimal value $\lambda_c,\boldsymbol{\Psi}_c$ can be found by a multitude of convex optimization techniques. We propose to use the alternating bisection method as in Algorithm 1. This requires to specify search ranges. We can take the lower bounds $(\underline{\lambda}_c,\underline{\Psi}_{c,i}) = (0,0)$. The upper bounds are obtained by finding the largest value over users such that the strongest mode of that user loses power with the corresponding power constraint being the only active one: $\overline{\lambda}_c = \max\limits_{k:b_k=c}(u_k\mathbf{S}_k - \mathbf{W}_k)_{1,1}/(\mathbf{G}_k^{'H}\mathbf{V}^{cH}\mathbf{V}^c\mathbf{G}_k')_{1,1}$ and $\overline{\Psi}_{c,i} = \max\limits_{k:b_k=c}(u_k\mathbf{S}_k - \mathbf{W}_k)_{1,1}/|(\mathbf{G}_k')_{i,1}|^2$. To simplify the description of the method in Algorithm 1, we introduce $\Psi_{c,0} = \lambda_c$. Also, $\boldsymbol{\Psi}_{c,\bar{i}}$ denotes all components of $\boldsymbol{\Psi}_c$ except for $\Psi_{c,i}$ and we take some liberty in ordering arguments of $g_c()$. The complexity could be reduced by reducing the bisection search ranges in consecutive sweeps of overall alternating optimization sweeps.

Algorithm 1 Alternating bisection for Lagrange multipliers

Initialization: $\underline{\Psi}_{c,i} = 0, \overline{\Psi}_{c,i}\,, \forall c, i.$
for $c = 1, ..., C$
  Repeat until convergence
    for $i = 0, 1, ..., M^c$
      $\Psi_{c,i} = (\underline{\Psi}_{c,i} + \overline{\Psi}_{c,i})/2$
      if $g_c(\boldsymbol{\Psi}_{c,\bar{i}}, \underline{\Psi}_{c,i}) < g_c(\boldsymbol{\Psi}_{c,\bar{i}}, \overline{\Psi}_{c,i})$, $\overline{\Psi}_{c,i} = \Psi_{c,i}$,
      else $\underline{\Psi}_{c,i} = \Psi_{c,i}$
    end for
end for

With the optimized λ_{b_k} and $\boldsymbol{\Psi}_{b_k}$, $\mathbf{P}_k^o(\lambda_{b_k},\boldsymbol{\Psi}_{b_k})$ is no longer diagonal. So consider its eigen decomposition $\mathbf{P}_k^o = \mathbf{U}_k\mathbf{P}_k\mathbf{U}_k^H$ leading to the new diagonal $\mathbf{P}_k$ and absorb the unitary $\mathbf{U}_k$: $\mathbf{G}_k' \leftarrow \mathbf{G}_k'\mathbf{U}_k$. Note that the minorization approach, which avoids introducing Rxs, can at every BF update allow to introduce an arbitrary number of streams per user by determining multiple dominant generalized eigenvectors, and then let the ILA-WF operation decide how many streams can actually be sustained. Given the digital BFs and the Lagrange multipliers, the analog BF $\mathbf{V}^c$ can be found by alternating optimization.

C. Design of Unconstrained Analog BF

At first we consider the case in which the analog BF is unconstrained. Hence the resulting design would also be applicable to more general two-stage BF design [19] in which the outer BF stage ($\mathbf{V}^c$) is in common to all users in a cell.

1) Fully Connected Case: To optimize $\mathbf{V}^c$, we equate the gradient of (7) w.r.t. $\mathbf{V}^c$ to zero. Using $\partial\ln\det\mathbf{X} = \text{tr}(\mathbf{X}^{-1}\partial\mathbf{X})$ and $\det(\mathbf{I}_M+\mathbf{AB}) = \det(\mathbf{I}_N+\mathbf{BA})$ from [20], we get

$$\begin{aligned}\sum_{k:b_k=c}\left(\widehat{\mathbf{B}}_k\mathbf{V}^c\mathbf{G}_k\boldsymbol{\zeta}_k\mathbf{G}_k^H - (\widehat{\mathbf{A}}_k+\lambda_c\mathbf{I})\mathbf{V}^c\mathbf{G}_k\mathbf{G}_k^H\right) = 0, \\ \text{with } \boldsymbol{\zeta}_k = u_k\left(\mathbf{I}+\mathbf{G}_k^H\mathbf{V}^{b_k H}\widehat{\mathbf{B}}_k\mathbf{V}^{b_k}\mathbf{G}_k\right)^{-1}.\end{aligned} \tag{14}$$

Now with $vec(\mathbf{AXB}) = (\mathbf{B}^T\otimes\mathbf{A})vec(\mathbf{X})$ [20], we get

$$\mathbf{V}^c = unvec(\mathbf{V}_{max}(\mathbf{B}_c,\mathbf{A}_c)),\ \text{with} \tag{15}$$

$$\begin{aligned}\mathbf{B}_c = \sum_{k:b_k=c}\left((\mathbf{G}_k\boldsymbol{\zeta}_k\mathbf{G}_k^H)^T\otimes\widehat{\mathbf{B}}_k\right), \\ \mathbf{A}_c = \sum_{k:b_k=c}\left((\mathbf{G}_k\mathbf{G}_k^H)^T\otimes(\widehat{\mathbf{A}}_k+\lambda_c\mathbf{I})\right)\,.\end{aligned} \tag{16}$$

2) Partially Connected Case: As in [21], in a partially connected phase shifting network, each RF chain is connected to a subset of antennas. Assuming each RF chain is connected to $L_t^c = N_t^c/M^c$ antennas, the analog precoder matrix can be written as a block diagonal matrix

$$\mathbf{V}^c = \begin{bmatrix} \mathbf{v}_1^c & 0 & \dots & 0 \\ 0 & \mathbf{v}_2^c & & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{v}_{M^c}^c \end{bmatrix} \tag{17}$$

where $\mathbf{v}_i^c \in \mathbb{C}^{L_t^c \times 1}$ (with unit magnitude elements in the phasor case). The advantage of a partially connected structure is that we need only N_t^c phase shifters. But at the cost of degradation in performance compared to a fully connected structure where there is more phase control. We define $\widetilde{\mathbf{B}}_{c,k}$ as the $N_t^c M^c \times N_t^c$ matrix obtained by concatenating the following subsets of columns: $(i-1)N_t^c + 1 : (i-1)N_t^c + L_t^c, i = 1, ..., M^c$ of $(\mathbf{G}_k \boldsymbol{\zeta}_k \mathbf{G}_k^H)^T \otimes \widehat{\mathbf{B}}_k$. We define $\widetilde{\mathbf{A}}_{c,i}$ similarly and let $\widetilde{\mathbf{V}}^c = [\mathbf{v}_1^{c\,T}, \mathbf{v}_2^{c\,T}, ..., \mathbf{v}_M^{c\,T}]^T$. Then optimizing (7) w.r.t. $\widetilde{\mathbf{V}}^c$ yields

$$\widetilde{\mathbf{V}}^c = \mathbf{V}_{max}(\sum_{k:b_k=c} \widetilde{\mathbf{B}}_{c,k}, \sum_{k:b_k=c} \widetilde{\mathbf{A}}_{c,k}). \tag{18}$$

Alternating WSR maximization between digital BF and an unconstrained analog BF now leads to Algorithm 2.

Algorithm 2 Hybrid BF Design via Alternating Minorizer

Given: $P^c, \boldsymbol{\Phi}_c, \mathbf{H}_{k,c}, u_k, \forall k, c.$

Initialization: $(\mathbf{V}^c)^{(0)} = \mathbf{V}_{1:M^c}(\sum_{k:b_k=c} \boldsymbol{\Theta}_k^c, \sum_{i:b_i \neq c} \boldsymbol{\Theta}_i^c)$, The $\mathbf{G}_k^{(0)}$ are taken as the ZF precoders for the effective channels $\mathbf{H}_{k,b_k}\mathbf{V}^{b_k}$ with uniform powers (from SPC).

Iteration (j) :

1) Compute $\widehat{\mathbf{Q}}_k^{(j)}, \widehat{\mathbf{B}}_k^{(j)}, \widehat{\mathbf{A}}_k^{(j)}, \forall k$ from (3), (6), (7).
2) Update $\mathbf{G}_k^{\prime(j)}$, $\forall k$, from (9).
3) Update $\lambda_c^{(j)}, \boldsymbol{\Psi}_c^{(j)}$ $\forall c$ using Algorithm 1 and thus $\mathbf{P}_k^{(j)}$ $\forall k$, from (11).
4) Update $(\mathbf{V}^c)^{(j)}$, $\forall c$, from (15) for fully connected case or from (18) for partially connected case.
5) Check for convergence of the WSR: if not go to step 1).

D. Hybrid Beamforming Design with Per-Antenna Power Constraints

Per-antenna power constraints for HBF can be written as

$$\sum_{k:b_k=c} \left[\mathbf{V}^c \mathbf{G}_k \mathbf{G}_k^H \mathbf{V}^{c\,H}\right]_{i,i} \leq a_i^c, \; i = 1, ..., N_t^c. \tag{19}$$

Substituting the above modified power constraints, WSR alternating maximization through minorization leads to the following expressions for the BFs (fully connected case)

$$\begin{aligned} &\mathbf{G}_k' = \mathbf{V}_{1:d_k}\Big(\mathbf{V}^{b_k H}\widehat{\mathbf{B}}_k\mathbf{V}^{b_k}, \mathbf{V}^{b_k H}\Big(\widehat{\mathbf{A}}_k + \lambda_{b_k}\mathbf{I} + \boldsymbol{\Psi}_{b_k}'\Big)\mathbf{V}^{b_k}\Big) \\ &\mathbf{V}^c = unvec(\mathbf{V}_{max}(\mathbf{B}_c, \mathbf{A}_c')), \\ &\text{where } \mathbf{A}_c' = \sum_{k:b_k=c} \Big((\mathbf{G}_k\mathbf{G}_k^H)^T \otimes (\widehat{\mathbf{A}}_k + \lambda_c\mathbf{I} + \boldsymbol{\Psi}_{b_k}')\Big), \\ &\text{and } \boldsymbol{\Psi}_c' = \text{diag}\,(\Psi_{c,1}, ..., \Psi_{c,N_t^c}). \end{aligned} \tag{20}$$

The ILA-WF can be modified similarly. Note that as for PRFPC, the maximum number of power constraints that can be satisfied with equality is the number of streams (stream powers).

E. Algorithm Convergence

The convergence proof of [22] does not apply directly because the power constraints here are not separable in the BF variables. The ingredients required are minorization [17], alternating or cyclic optimization [17] (also called block coordinate descent), Lagrange dual function [18], saddle-point interpretation [18] and KKT conditions [18]. For the WSR cost function $WSR(\mathbf{Q})$ in (5) we construct the minorizer as in (6), (7) leading to

$$\begin{aligned} &WSR(\mathbf{Q}) \geq \underline{WSR}(\mathbf{Q}, \widehat{\mathbf{Q}}) = \\ &\sum_{k=1}^{K} [u_k \ln\det(\mathbf{I} + \widehat{\mathbf{B}}_k\mathbf{Q}_k) - \text{tr}\{\widehat{\mathbf{A}}_k(\mathbf{Q}_k - \widehat{\mathbf{Q}}_k)\}] \end{aligned} \tag{21}$$

where $\underline{WSR}(\widehat{\mathbf{Q}}, \widehat{\mathbf{Q}}) = WSR(\widehat{\mathbf{Q}})$. The minorizer, which is concave in $\mathbf{Q}$, still has the same gradient as $WSR(\widehat{\mathbf{Q}})$ and hence KKT conditions are not affected. Now reparameterizing $\mathbf{Q}$ in terms of $\mathbf{P}, \mathbf{G}', \mathbf{V}$ as in (3) and adding the power constraints to the minorizer, we get the Lagrangian (10). Every alternating update of $\mathcal{L}$ w.r.t. $\mathbf{V}$, $\mathbf{G}'$, or $(\mathbf{P}, \boldsymbol{\Lambda}, \boldsymbol{\Psi})$ leads to an increase of the WSR, ensuring convergence (within each of these 3 parameter goups, we further alternate between each user or BS). For the KKT conditions, at the convergence point, the gradients of $\mathcal{L}$ w.r.t. $\mathbf{V}$ or $\mathbf{G}'$ correspond to the gradients of the Lagrangian of the original WSR. For fixed $\mathbf{V}$ and $\mathbf{G}'$, $\mathcal{L}$ is concave in $\mathbf{P}$, hence we have strong duality for the saddle point $\max_{\mathbf{P}} \min_{\Lambda,\boldsymbol{\Psi}} \mathcal{L}$. Also, at the convergence point the solution to $\min_{\Lambda,\boldsymbol{\Psi}} \mathcal{L}(\mathbf{V}^o, \mathbf{G}'^o, \mathbf{P}^o, \boldsymbol{\Lambda}, \boldsymbol{\Psi})$ satisfies the gradient KKT condition for $\mathbf{P}$ and the complementary slackness conditions for $c = 1, \dots, C$

$$\begin{aligned} &\lambda_c^o\,(P^c - \textstyle\sum_{k:b_k=c} \text{tr}\{\mathbf{V}^{co}\mathbf{G}_k'^o\mathbf{P}_k^o\mathbf{G}_k'^{o\,H}\mathbf{V}^{co\,H}\} = 0, \\ &\text{tr}\{\boldsymbol{\Psi}_c^o(\boldsymbol{\Phi}_c^o - \textstyle\sum_{k:b_k=c} \mathbf{G}_k'^o\mathbf{P}_k^o\mathbf{G}_k'^{o\,H})\} = 0 \end{aligned} \tag{22}$$

where all individual factors in the products are nonnegative (and for $\boldsymbol{\Psi}_c^o$, the sum of nonnegative terms being zero implies all the terms being zero).

In the proposed approach, $g(\boldsymbol{\Lambda}, \boldsymbol{\Psi}|\mathbf{V}, \mathbf{G}') = \max_{\mathbf{P}} \mathcal{L}(\mathbf{V}, \mathbf{G}', \mathbf{P}, \boldsymbol{\Lambda}, \boldsymbol{\Psi})$. In contrast, in [16], Lagrangian duality and alternating optimization are interchanged with dual function $g(\boldsymbol{\Lambda}) = \max_{\mathbf{V},\mathbf{G}',\mathbf{P}} \mathcal{L}(\mathbf{V}, \mathbf{G}', \mathbf{P}, \boldsymbol{\Lambda})$ (no PRFPC or PAPC), leading to more complex iterations and a power optimization that is further away from classical water filling.

IV. Deterministic Annealing for Phase Shifter Constrained Analog BF

In this section, we propose an algorithm modification to design the analog BF for a phase shifter constrained case. Accounting of the unit modulus constraints of the entries of $\mathbf{V}^c$ can be done by parameterizing as

$$\left|\mathbf{V}_{m,n}^c\right| = 1 \;\Rightarrow\; \mathbf{V}_{m,n}^c = e^{j\,\theta_{m,n}^c}. \tag{23}$$

One can push the alternating optimization paradigm further to the level of each phasor by representing the WSR as a function of each phasor element $f(\theta_{m,n}^c)$ and optimizing it.

A similar such approach can be found in [6], where the optimization is w.r.t. the WSMSE. However, a major drawback of such designs is the strong non-convexity of the cost function with many local optima. Hence depending on the initialization the algorithm will converge to different local optima. So we consider here one approach called deterministic annealing (DA) (or homotopy method) to avoid the problem of local optima. In DA, we use a homotopy parameter to gradually move the problem from a simpler one (with known global optimum) to the final complex problem, solving the problem for every consecutive homotopy parameter instance, initialized by the previous solution. If the homotopy parameter varies slowly, the global optimum of the previous problem instance will be in the region of attraction of the next global optimum. For the analog beamforming design using phasors, we start from the optimal unconstrained $\mathbf{V}^c$ which has close to all digital performance. Then with the gradual forcing of the amplitude of the unconstrained $\mathbf{V}^c$ to 1 we approach the global optimum for the phasor case. The steps for the DA method are given in Algorithm 3. Note that in Algorithm 3, b is some constant value less than 1, say 0.9. Then in n iterations, the coefficient amplitudes are modified by $|\mathbf{V}^c_{i,j}|^{b^n}$. As n varies from 0 to ∞, the homotopy parameter $t_n = b^n$ varies from 1 to 0 and $|\mathbf{V}^c_{i,j}|^{t_n}$ varies from $|\mathbf{V}^c_{i,j}|$ to 1. We get

$$(\mathbf{\Lambda}^{(n)}, \mathbf{\Psi}^{(n)}, \boldsymbol{\theta}^{(n)}, \mathbf{G}'^{(n)}, \mathbf{P}^{(n)}) = \arg\min_{\mathbf{\Lambda},\mathbf{\Psi}} \max_{\boldsymbol{\theta},\mathbf{G}',\mathbf{P}} \mathcal{L}(|\mathbf{V}|^{t_n}, \boldsymbol{\theta}, \mathbf{G}', \mathbf{P}, \mathbf{\Lambda}, \mathbf{\Psi})$$

initialized by $(\mathbf{\Lambda}^{(n-1)}, \mathbf{\Psi}^{(n-1)}, \boldsymbol{\theta}^{(n-1)}, \mathbf{G}'^{(n-1)}, \mathbf{P}^{(n-1)})$.

Algorithm 3 Deterministic Annealing for Analog Beamformer

Let $\mathbf{V}^c_{i,j} = |\mathbf{V}^c_{i,j}|e^{j\theta^c_{i,j}}$. Let the unconstrained $\mathbf{V}^c$ design (joint $\mathbf{V}^c$ and all $\mathbf{G}_k$ using Algorithm 2) converge first.

1) Scale $\forall c, (i,j) : |\mathbf{V}^c_{i,j}| \leftarrow |\mathbf{V}^c_{i,j}|^b$.
2) Reoptimize all $\theta^c_{i,j}$.
3) Update $\mathbf{G}'$, $\mathbf{P}$, $\mathbf{\Lambda}$, $\mathbf{\Psi}$ using Algorithm 2.
4) Go to 1) for a number of iterations.
5) Redo 2)-3) a last time with all $|\mathbf{V}^c_{i,j}| = 1$.

However, even though the DA ensure that the phasors design (for a given unconstrained analog BF) converges to the global optimum, the overall alternating WSR optimization algorithm of digital BFs and analog precoders can still converge to a local optimum depending on the initialization. To ensure convergence to the overall global optimum, another layer of DA can be added by solving the WSR problem at increasing SNR values starting from near 0 to the desired SNR [23].

V. Simulation Results

Simulations are presented for a single cell with K single antenna users. Assuming a Uniform Linear Array (ULA), the MISO channel for user k can be written as the pathwise model $\mathbf{h}_k = \sum_{i=1}^{L} \alpha_i \mathbf{a}_t(\phi_i)$, where α_i is the complex path gain which is assumed to be complex Gaussian with exponentially distributed variance profile. ϕ_i corresponds to the Angle of Departure (AoD) which is assumed to be uniformly distributed in the interval $[0, 30]$.Notations used in the figure: CoCSIT refers to covariance CSIT , iCSIT refers to perfect CSIT. SPC, PAPC, PRFPC refer to sum power, per-antenna power or per-RF power constraints. Firstly, in Fig. 1, we consider the case of only SPC. We compare the performance of 3 proposed algorithms with the optimal fully digital BF [24], approximate WSR based hybrid design [5], WSR HBF via a WSMSE based design [6] and a CoCSIT based scheme [25]. Except the curve denoted as "Proposed Partially Connected Analog BF", all others are for fully connected analog BF.

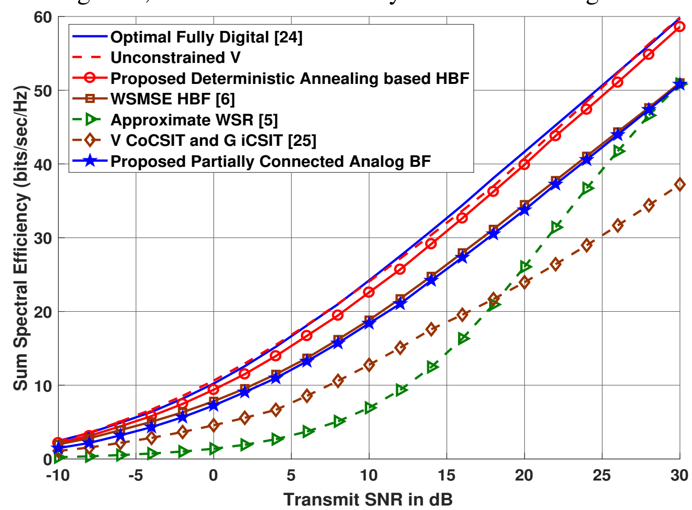

Fig. 1: Sum rates, $N_t = 32, M = 16, K = 8, C = 1, L = 4$.

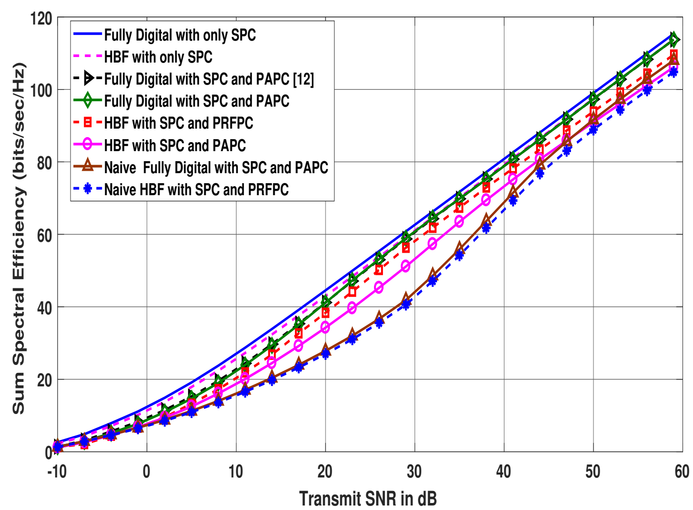

Fig. 2: Sum rates, $N_t = 32, M = 16, K = 8, C = 1, L = 4$.

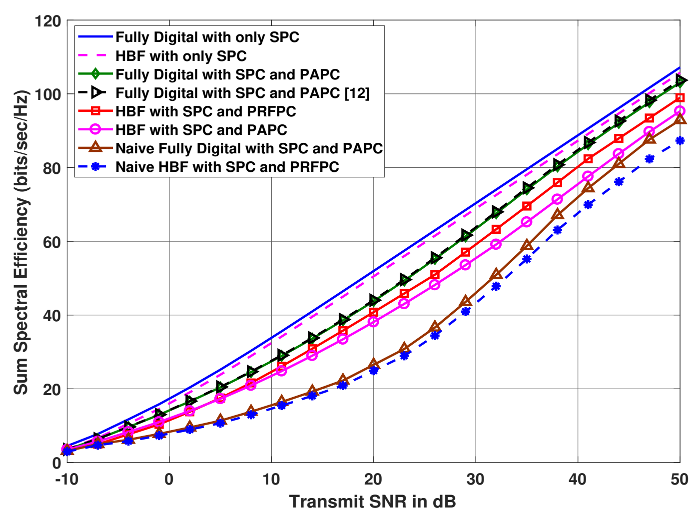

Fig. 3: Sum rates, $N_t = 64, M = 16, K = 8, C = 1, L = 4$.

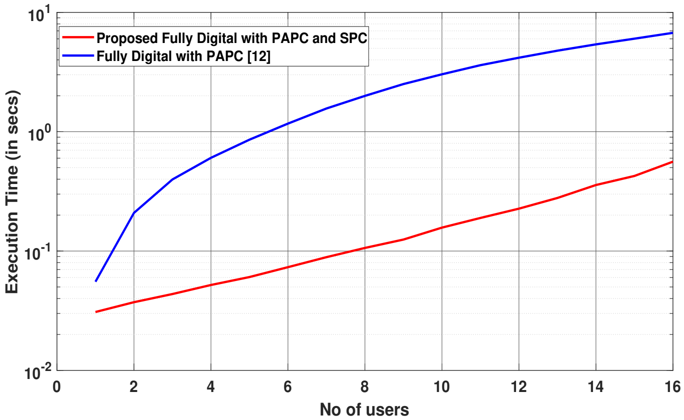

Fig. 4: Execution time comparison.

It is clear that the proposed unconstrained HBF solution has the same performance as the fully digital solution. With phase shifter constrained analog precoder, the proposed DA based design narrows the gap to the fully digital performance and performs much better than state of the art solutions such as WSMSE which suffer from the issue of local optima. In Fig. 2, we compare our fully digital and HBF designs based on SPC and/or PAPC and/or PRFPC. Imposing PAPC or PRFPC in addition to the SPC degrades the sum rate but less for PRFPC as there are fewer constraints. Our digital SPC+PAPC designs performs identically to that in [12]. The optimized designs for PAPC or PRFPC outperform naive designs in which the SPC BF is scaled down to satisfy the PAPC or PRFPC constraints, esp. at intermediate SNR. Fig. 3 confirms those observations for a larger number of BS antennas.

In Fig. 4, for the fully digital PAPC, we compare the execution time in Matlab for the proposed solution to that of the geometric programming (GP) approach in [12] for the power allocation (which is solved using interior point methods). The digital BF computation has similar complexity ($\mathcal{O}(N_t^3)$) between SMSE in [12] and the proposed solution. The complexity $(N_t+1)x$ of the alternating bisection is linear in the number of power constraints, where x represents the complexity associated with the evaluation of $g(\mathbf{\Lambda}, \mathbf{\Psi})$). GP has a worst case polynomial time complexity. Faster convergence of the minorization approach compared to the SMSE solution and the reduced complexity of the alternating bisection vs the GP lead to a much shorter execution time for the proposed algorithm as shown in Fig. 4.

VI. Conclusions

In this paper, we presented a WSR maximizing algorithm for digital or HBF, with unconstrained amplitude or phasor analog BF, fully or partially connected, in a Multi-User Multi-Cell MIMO system. We considered for the first time the more realistic scenario of per-RF or per-antenna power constraints for a HBF system. Convergence of the alternating minorization approach was shown and adding deterministic annealing allowed to attain the global optimum.

References

[1] X. Zhang, A. Molisch, and S. Kung, "Variable-phase-shift-based RF-baseband codesign for MIMO antenna selection," *IEEE Trans. on Sig. Process.*, vol. 53, no. 11, 2005.

[2] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. on Wireless Commun.*, vol. 13, no. 3, March 2014.

[3] A. Alkhateeb, O. E. Ayach, G. Leus, and R. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. for Sel. Topics in Sig Process.*, vol. 8, no. 5, Oct. 2014.

[4] C. Rusu, R. Mndez-Rial, N. Gonzlez-Prelcic, and R. W. H. Jr., "Low complexity hybrid sparse precoding and combining in millimeter wave MIMO systems," in *Proc. IEEE Int. Conf. Commun.*, June 2015.

[5] F. Sohrabi and W. Yu, "Hybrid digital and analog beamforming design for large scale antenna arrays," *IEEE J. for Sel. Topics in Sig. Process.*, vol. 10, no. 3, April 2016.

[6] C. K. Thomas and D. Slock, "Mixed Time Scale Weighted Sum Rate Maximization for Hybrid Beamforming in Multi-Cell MU-MIMO Systems," in *Globecom ET5GBWMN Wkshp*, Singapore, Dec. 2017.

[7] M. Kim and Y. Lee, "MSE-Based Hybrid RF/Baseband Processing for Millimeter-Wave Communication Systems in MIMO Interference Channels," *IEEE Trans. Vehicular Technology*, June 2015.

[8] D. H. N. Nguyen, L. B. Le, T. Le-Ngoc, and R. W. Heath, "Hybrid MMSE precoding and combining designs for mmWave multiuser systems," *IEEE Access*, vol. 5, April 2017.

[9] G. Zhu, K. Huang, V. K. N. Lau, B. Xia, X. Li, and S. Zhang, "Hybrid beamforming via the kronecker decomposition for the millimeter-wave massive MIMO systems," *IEEE J. on Sel. Areas of Commun.*, vol. 35, no. 9, Sep. 2017.

[10] W. Yu and T. Lan, "Transmitter optimization for the multi-antenna downlink with per-antenna power constraints," *IEEE Trans. Signal Proc.*, vol. 55, no. 6, June 2007.

[11] K. Karakayali, R. Yates, G. Foschini, and R. Valenzuela, "Optimum zero-forcing beamforming with per-antenna power constraints," in *Proc. IEEE Int'l Symp. on Info. Theo. (ISIT)*, Nice, France, 2017.

[12] T. E. Bogale and L. Vandendorpe, "Sum MSE optimization for downlink multiuser mimo systems with per antenna power constraint: Downlink-uplink duality approach," in *IEEE 22nd Intl Symp. on PIMRC*, 2011.

[13] T. M. Pham, R. Farrell, J. Dooley, E. Dutkiewicz, D. N. Nguyen, and L.-N. Tran, "Efficient zero-forcing precoder design for weighted sum-rate maximization with per-antenna power constraint," *IEEE Trans. on Vehic. Techn.*, N. 2017.

[14] R. Zhang, "Cooperative multi-cell block diagonalization with per-basestation power constraints," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, Dec. 2010.

[15] J. S. Herd and M. D. Conway, "The Evolution to Modern Phased Array Architectures," *Proc. IEEE*, Mar. 2016.

[16] S. J. Kim and G. B. Giannakis, "Optimal Resource Allocation for MIMO Ad-hoc Cognitive Radio Networks," *IEEE Trans. Inf. Theory*, May 2011.

[17] P. Stoica and Y. Selén, "Cyclic Minimizers, Majorization Techniques, and the Expectation-Maximization Algorithm: A Refresher," *IEEE Sig. Proc. Mag.*, Jan. 2004.

[18] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge Univ. Press, 2004.

[19] J. Chen and V. K. N. Lau, "Two-Tier Precoding for FDD Multi-Cell Massive MIMO Time-Varying Interference Networks," *IEEE J. Sel. Areas. Comm.*, June 2014.

[20] K. B. Petersen and M. S. Pedersen, "The matrix cookbook," in *URL http://www2.imm.dtu.dk/pubdb/p.php?3274.*, Nov. 2011.

[21] R. Mendez-Rial, C. Rusu, N. Gonzalez-Prelcic, A. Alkhateeb, and R. Heath, "Hybrid MIMO architectures for millimeter wave communications: Phase shifters or switches ?" *IEEE Access*, Jan. 2016.

[22] Q. Shi, M. Razaviyayn, Z. Q. Luo, and C. He, "An iteratively weighted mmse approach to distributed sum-utility maximization for a MIMO interfering broadcast channel," *IEEE Trans. on Sig. Process.*, vol. 59, no. 9, Sep. 2011.

[23] F. Negro, I. Ghauri, and D. T. M. Slock, "Deterministic annealing design and analysis of the noisy MIMO interference channel," in *Proc. IEEE Inf. Theo. and Applic. Workshop (ITA), San Diego, CA, USA*, 2011.

[24] S. S. Christensen, R. Agarwal, E. de Carvalho, and J. Cioffi, "Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design," *IEEE Trans. on Wireless Commun.*, Dec. 2008.

[25] S. Park, J.Park, A. Yazdan, and R. W. Heath, "Exploiting spatial channel covariance for hybrid precoding in Massive MIMO systems," *IEEE Trans. on Sig. Process.*, vol. 65, no. 14, July 2017.