Deterministic Annealing for Hybrid Beamforming Design in Multi-Cell MU-MIMO Systems

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

***Abstract*—This work deals with hybrid beamforming (HBF) for the MIMO Interfering Broadcast Channel (IBC), i.e. the Multi-Input Multi-Output (MIMO) Multi-User (MU) Multi-Cell downlink channel. HBF is a low complexity alternative to 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. We furthermore propose a deterministic annealing based approach to avoid issues of local optima that plague phase shifter constrained analog beamformers. Simulation results indicate that the proposed deterministic annealing based approach performs significantly better than state of the art Weighted Sum Mean Squared Error (WSMSE) or WSR based solutions. We also propose a closed form solution for the analog BF in case the number of RF chains equals or exceeds the total number of multipath components and the antenna array responses are phasors.**

***Index Terms*—Hybrid beamforming, massive MIMO, millimeter wave, weighted sum rate, deterministic annealing.**

I. INTRODUCTION

In this paper, Tx may denote transmit/transmitter/transmission and Rx may denote receive/receiver/reception. Hybrid beamforming is a two-stage architecture in which the beamformer (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 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 single user mmWave system.

Hybrid beamforming designs for multi-user systems can be found in [5]–[12]. In [5] and [6], the authors propose a two-stage hybrid precoding design. In the first stage, Mobile and Base Station (MS/BS) jointly select the best combination of Rx combiner and RF beamformer which maximizes the channel gain to that particular user, ignoring the effect of interference. The digital precoder is then chosen as the zero-forcing solution to the effective channel. In [8] the authors use Weighted Sum Rate (WSR) maximization as the target optimization criterion for the design of hybrid beamformers. However, they optimize the analog phasors using transmit power minimization criteria while the digital precoder is zero-forcing. In [13], we propose a Weighted Sum Mean Squared Error (WSMSE) based approach for the joint design of digital and analog beamformers for a multi-cell multi-user MIMO system. [11], [14] propose a hybrid BF design using sparse formulations and approximating the MMSE. In [14], orthogonal matching pursuit is used to select RF beamforming vectors from a set of candidate vectors and the digital BF is optimized by least-squares fitting of the analog-digital BF cascade to an all-digital solution.

The main issue with WSR/WSMSE optimization for a HBF hybrid design is the high non-convexity of the cost function. This implies that even if it is possible to show convergence to a local optimum [13], convergence to the global optimum cannot be guaranteed. To avoid the convergence to a local optimum, [15] proposed Deterministic Annealing (DA) for digital BF design in the MIMO interference channel.

A. *Contributions of this paper*

In this paper:

- We first propose a hybrid beamforming design based on the WSR criterion which is simplified using the minorization approach. The advantage compared to the WSMSE solution [13] is that the iterative algorithm converges faster (no ping-pong between Tx and Rx optimization, and direct power optimization).
- We derive conditions under which the HBF can attain the fully digital performance with sufficient number of RF chains.
- To overcome the issue of local optima, we propose a deterministic annealing approach for the design of the analog phasors.
- Numerical results suggest that the proposed alternating optimization based WSR maximizing algorithm performs better than state of the art solutions. Moreover, it is interesting to observe that the proposed DA based HBF design allows to narrow the gap to optimal fully digital solutions [16].

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

II. Multi-User MIMO System Model

In this paper we shall consider a multi-stream approach with d_k streams for user k. So, consider an Interfering BroadCast (IBC) (i.e. multi-cell MU downlink) system of C cells with a total of K users and N_t^c transmit antennas in cell c. User k is equipped with N_k antennas. $\mathbf{H}_{k,c}$ represents the $N_k \times N_t^c$ MIMO channel between user k and BS c and we define $\mathrm{E}\left[\mathbf{H}_{k,c}^H \mathbf{H}_{k,c}\right] = \mathbf{\Theta}_k^c$. User k receives

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

where $\mathbf{s}_k$, of size $d_k \times 1$, is the intended signal stream vector (all entries are white, unit variance). BS c serves $U_c = \sum\limits_{i:b_i=c} 1$ users. We are considering a noise whitened signal representation so that we get for the noise $\mathbf{v}_k \sim \mathcal{CN}(0, \mathbf{I}_{N_k})$. The analog beamformer $\mathbf{V}^c$ for base station c is of dimension $N_t^c \times M^c$ where M^c is the number of RF chains at BS c. The $M^c \times d_k$ digital beamformer is $\mathbf{G}_k$, where $\mathbf{G}_k = \left[\mathbf{g}_k^{(1)} \ldots \mathbf{g}_k^{(d_k)}\right]$ and $\mathbf{g}_k^{(s)}$ represents the beamformer for stream s of user k. The transmit power constraint at base station c can be written as $\mathrm{tr}\{\,\mathbf{V}^{c\,H}\mathbf{V}^c \sum_{i:b_i=c}^K \mathbf{G}_i\,\mathbf{G}_i^H\,\} \leq P_c$.

III. Minorization Approach

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

$$\begin{aligned}[\mathbf{V}\ \ \mathbf{G}] &= \arg\max_{\mathbf{V},\mathbf{G}} WSR(\mathbf{G},\mathbf{V})\\ &= \arg\max_{\mathbf{V},\mathbf{G}} \sum_{k=1}^K u_k \ln\det(\mathbf{R}_{\overline{k}}^{-1}\mathbf{R}_k),\end{aligned} \tag{2}$$

where the u_k are the rate weights, $\mathbf{G}$ represents the collection of digital BFs $\mathbf{G}_k$, $\mathbf{V}$ the collection of analog BFs $\mathbf{V}^{b_k}$. From [16], we can write,

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

where $\mathbf{R}_{\overline{k}}$ is the interference plus noise covariance matrix.With the definition of the Tx covariance matrices $\mathbf{Q}_i$, the power constraints can be written as,

$$\textstyle\sum_{k:b_k=c} \mathrm{tr}\{\mathbf{Q}_k\} \leq P_c\,. \tag{4}$$

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 the difference of convex functions (DC programming) approach as in [17] in which the WSR is written as the summation of a convex and a concave term. Consider the dependence of the WSR on $\mathbf{Q}_k$ alone:

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

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, consider the first order Taylor series expansion of $WSR_{\overline{k}}$ in $\mathbf{Q}_k$ around $\widehat{\mathbf{Q}}$ (i.e. all $\widehat{\mathbf{Q}}_i$).

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

Note that the linearized tangent expression for $WSR_{\overline{k}}$ constitutes a lower bound for it and hence the DC approach (in $\mathbf{Q}$) is also a minorization approach (in $\mathbf{Q}$ or $\mathbf{G}$). Now, dropping constant terms, reparameterizing the $\mathbf{Q}_k = \mathbf{G}_k\mathbf{G}_k^H$, performing this linearization for all users, and augmenting the WSR cost function with the Tx power constraints, we get the Lagrangian,

$$\begin{aligned}&WSR(\mathbf{G},\mathbf{V},\lambda) = \sum_{k=1}^K u_k \ln\det\left(\mathbf{I} + \mathbf{G}_k^H\mathbf{V}^{b_k\,H}\widehat{\mathbf{B}}_k\mathbf{V}^{b_k}\mathbf{G}_k\right)\\ &-\mathrm{tr}\left\{\mathbf{G}_k^H\mathbf{V}^{b_k\,H}\left(\widehat{\mathbf{A}}_k + \lambda_{b_k}\mathbf{I}\right)\mathbf{V}^{b_k}\mathbf{G}_k\right\} + \sum_{j=1}^C \lambda_j P_j,\end{aligned} \tag{7}$$

where $\widehat{\mathbf{B}}_k = \mathbf{H}_{k,b_k}^H\widehat{\mathbf{R}}_{\overline{k}}^{-1}\mathbf{H}_{k,b_k}$. In what follows, we shall optimize the WSR with perfect CSIT by alternating optimization between digital and analog beamformers.

A. Digital BF Design

The gradient w.r.t. $\mathbf{G}_k$ of (7) (which is still the same as that of (2)) leads to the solution as d_k dominant generalized eigenvectors

$$\mathbf{G}_k^{'} = \mathbf{V}_{1:d_k}\left(\mathbf{V}^{b_k\,H}\widehat{\mathbf{B}}_k\mathbf{V}^{b_k},\ \mathbf{V}^{b_k\,H}\left(\widehat{\mathbf{A}}_k + \lambda_{b_k}\mathbf{I}\right)\mathbf{V}^{b_k}\right), \tag{8}$$

with associated generalized eigenvalues $\Sigma_k = \Sigma_{1:d_k}(\mathbf{V}^{b_k\,H}\widehat{\mathbf{B}}_k\mathbf{V}^{b_k}, \mathbf{V}^{b_k\,H}(\widehat{\mathbf{A}}_k + \lambda_{b_k}\mathbf{I})\mathbf{V}^{b_k})$. Let $\mathbf{\Sigma}_k^{(1)} = \mathbf{G}_k^{'\,H}\mathbf{V}^{b_k\,H}\widehat{\mathbf{B}}_k\mathbf{V}^{b_k}\mathbf{G}_k^{'}$ and $\mathbf{\Sigma}_k^{(2)} = \mathbf{G}_k^{'\,H}\mathbf{V}^{b_k\,H}\widehat{\mathbf{A}}_k\mathbf{V}^{b_k}\mathbf{G}_k^{'}$. The advantage of formulation (7) is that it allows straightforward power adaptation: introducing stream powers in the diagonal matrices $\mathbf{P}_k \geq 0$ and substituting $\mathbf{G}_k = \mathbf{G}_k^{'}\,\mathbf{P}_k^{\frac{1}{2}}$ in (7) yields

$$\begin{aligned}&WSR(\mathbf{P},\lambda) = \textstyle\sum_j^C \lambda_j P_j +\\ &\sum_{k=1}^K [u_k \ln\det(\mathbf{I} + \mathbf{P}_k\mathbf{\Sigma}_k^{(1)}) - \mathrm{tr}\{\mathbf{P}_k(\mathbf{\Sigma}_k^{(2)} + \lambda_{b_k}\mathbf{V}^{b_k\,H}\mathbf{V}^{b_k})\}],\end{aligned}$$

the optimization of which leads to the following interference leakage aware water filling (WF) (jointly for the $\mathbf{P}_k$ and λ_c)

$$\mathbf{P}_k = \left(u_k(\mathbf{\Sigma}_k^{(2)} + \lambda_{b_k}\mathbf{V}^{b_k\,H}\mathbf{V}^{b_k})^{-1} - \mathbf{\Sigma}_k^{-(1)}\right)^+, \tag{9}$$

where $(x)^+ = \max(0, x)$ is applied to all diagonal elements and the Lagrange multipliers are adjusted to satisfy the power constraints. This can be done by bisection and gets executed per BS. Given the digital BFs, we update the analog beamformers $\mathbf{V}^c$. First we consider the case in which the analog beamformer is unconstrained.

B. Design of Unconstrained Analog BF

To optimize $\mathbf{V}^c$, we set the gradient of (7) w.r.t. $\mathbf{V}^c$ equal to zero. Using the results $\nabla \ln\det \mathbf{X} = \mathrm{tr}(\mathbf{X}^{-1}\nabla\mathbf{X})$ and $\det(\mathbf{I}_M + \mathbf{X}\mathbf{Y}) = \det(\mathbf{I}_N + \mathbf{Y}\mathbf{X})$ from [18], we get

$$\sum_{k:b_k=c} \widehat{\mathbf{B}}_k \mathbf{V}^c \mathbf{G}_k \mathbf{G}_k^H \mathbf{W}_k - \sum_{k:b_k=c} (\widehat{\mathbf{A}}_k + \lambda_c \mathbf{I}) \mathbf{V}^c \mathbf{G}_k \mathbf{G}_k^H = 0,$$
$$\text{where } \mathbf{W}_k = u_k \left(\mathbf{I} + \mathbf{G}_k \mathbf{G}_k^H \mathbf{V}^{b_k\,H} \widehat{\mathbf{B}}_k \mathbf{V}^{b_k}\right)^{-1}. \tag{10}$$

Now using $vec(\mathbf{A}\mathbf{X}\mathbf{B}) = (\mathbf{B}^T \otimes \mathbf{A})vec(\mathbf{X})$ from [18], we get

$$\begin{aligned} &\mathbf{V}^c = unvec(\mathbf{V}_{max}(\mathbf{B}_c, \mathbf{A}_c)) \ \text{ with} \\ &\mathbf{B}_c = \sum_{k:b_k=c} \left((\mathbf{G}_k \mathbf{G}_k^H \mathbf{W}_k)^T \otimes \widehat{\mathbf{B}}_k\right), \\ &\mathbf{A}_c = \sum_{k:b_k=c} \left((\mathbf{G}_k \mathbf{G}_k^H)^T \otimes (\widehat{\mathbf{A}}_k + \lambda_c \mathbf{I})\right). \end{aligned} \tag{11}$$

The unconstrained BF derived here is used in Section V to design the deterministic annealing based analog phasors.

C. Design of Phase Shifter Constrained Analog Beamformer

Given the digital BFs, the phase shifter analog beamformer $\mathbf{V}^c$ can be found by performing alternating optimization elementwise. Accounting for the unit modulus constraints of the entries of $\mathbf{V}^c$ can be done by parameterizing as

$$\left|\mathbf{V}^c_{p,q}\right| = 1 \implies \mathbf{V}^c_{p,q} = e^{j\theta^c_{p,q}}. \tag{12}$$

Since the analog BF is common to all users in a cell c, from (7) we can write the WSR as a function of $\theta^c_{p,q}$ as

$$\begin{aligned} f\left(\theta^c_{p,q}\right) = \sum_{k:b_k=c} [u_k \ln\det(\mathbf{I} + \mathbf{C}^k_{p,q}\, e^{j\theta^c_{p,q}} + \mathbf{D}^k_{p,q}\, e^{-j\theta^c_{p,q}} \\ + \mathbf{T}^{k,1}_{\overline{p},\overline{q}}) - \mathrm{tr}(\mathbf{E}^k_{p,q}\, e^{j\theta^c_{p,q}} + \mathbf{F}^k_{p,q}\, e^{-j\theta^c_{p,q}} + \mathbf{T}^{k,2}_{\overline{p},\overline{q}})] + c_{\overline{p},\overline{q}}, \end{aligned} \tag{13}$$

where $c_{\overline{p},\overline{q}}$ are terms that are independent of $\theta^c_{p,q}$. The steps leading to these expressions are derived in Appendix A. Setting the derivative of (13) w.r.t. $\theta^c_{p,q}$ to zero we get

$$\begin{aligned} e^{j\theta^c_{p,q}} \sum_{k:b_k=c} \mathrm{tr}\{\widetilde{\mathbf{W}}_k\, \mathbf{C}^k_{p,q} - \mathbf{E}^k_{p,q}\} = \\ e^{-j\theta^c_{p,q}} \sum_{k:b_k=c} \mathrm{tr}\{\widetilde{\mathbf{W}}_k\, \mathbf{D}^k_{p,q} - \mathbf{F}^k_{p,q}\} \end{aligned} \tag{14}$$

where $\widetilde{\mathbf{W}}_k = u_k \left(\mathbf{I} + \mathbf{G}_k^H \mathbf{V}^{b_k\,H} \widehat{\mathbf{B}}_k \mathbf{V}^{b_k} \mathbf{G}_k\right)^{-1}$.

This leads to two extrema for $\theta^c_{p,q}$ of which the best one needs to be chosen:

$$\begin{aligned} &\theta^c_{p,q} = \arg\max_{\theta^{c\,1}_{p,q},\theta^{c\,2}_{p,q}} f\left(\theta^c_{p,q}\right), \ \theta^{c\,1}_{p,q} = -\frac{\angle a}{2}, \\ &\theta^{c\,2}_{p,q} = \pi - \frac{\angle a}{2}, a = \frac{\sum_{k:b_k=c} \mathrm{tr}\{\widetilde{\mathbf{W}}_k\, \mathbf{C}^k_{p,q} - \mathbf{E}^k_{p,q}\}}{\sum_{k:b_k=c} \mathrm{tr}\{\widetilde{\mathbf{W}}_k\, \mathbf{D}^k_{p,q} - \mathbf{F}^k_{p,q}\}}. \end{aligned} \tag{15}$$

Alternating WSR maximization between digital and analog BF now leads to Algorithm 1.

Algorithm 1 Hybrid BF Design via Alternating Minorizer

Given: $P_c, \mathbf{H}_{k,c}, u_k \,\forall k, c$.
Initialization: $\mathbf{V}^c = e^{j\angle \mathbf{V}_{1:M^c}(\sum_{k:b_k=c}\mathbf{\Theta}^c_k, \sum_{i:b_i\neq c}\mathbf{\Theta}^c_i)}$,
The $\mathbf{G}_k$ are taken as the ZF precoders for the effective channels $\mathbf{H}_{k,b_k}\mathbf{V}^{b_k}$ with uniform powers.
Iteration (j) :

1) Compute $\widehat{\mathbf{B}}_k, \widehat{\mathbf{A}}_k, \,\forall k$ from (6), (7).
2) Update $\mathbf{G}_k^{'(j)}$, $\forall k$, from (8).
3) Update $\mathbf{P}_k$ and λ_c, $\forall k, c$ from (9).
4) Update $(\mathbf{V}^c_{p,q})^{(j)}$, $\forall c, \forall (p, q)$, from (15) (phasor constrained) or from (11) (unconstrained).
5) Check for convergence of the WSR: if not go to step 1).

IV. Hybrid Beamformer Capabilities

In this section we analyze to what extent a hybrid BF can achieve the same performance as a fully digital BF. In particular we shall see that this is possible for a sufficient number of RF chains and with the antenna array responses being phasors. Consider a specular or pathwise channel model with say L multi-paths per link. For notational simplicity we shall consider a uniform L and $N_k = N_r, \forall k$. Let the antenna array response for BS c be $\mathbf{h}^c_t(\phi)$ for Angle of Departure (AoD) ϕ. We assume that all entries of $\mathbf{h}^c_t(\phi)$ have the same magnitude. Then the collective $N_t \times L$ multipath Tx array response $\mathbf{H}_{t,k}$ for the downlink channel of user k is

$$\mathbf{H}^c_{t,k} = \left[\mathbf{h}^c_t\left(\phi_{k,1}\right)\ \mathbf{h}^c_t\left(\phi_{k,2}\right)\ \ldots\ \mathbf{h}^c_t\left(\phi_{k,L}\right)\right]^*, \tag{16}$$

and the concatenated antenna array response matrix to all users can be written as, $\overline{\mathbf{H}}^c_t = \left[\mathbf{H}^c_{t,1}\ \mathbf{H}^c_{t,2}\ \ldots\ \mathbf{H}^c_{t,K}\right]$, of dimension $N_t \times N_p$, where we denote the total number of paths $N_p = LK$. Similarly we define $\overline{\mathbf{H}}^c_r$ and $\overline{\mathbf{A}}^c$ for the concatenated Rx antenna array responses and complex path amplitudes. $\overline{\mathbf{A}}^c$ is a $N_p \times N_p$ block diagonal matrix with blocks of size $L \times L$ and $\overline{\mathbf{H}}^c_r$ is a $KN_r \times N_p$ block diagonal matrix with blocks of size $N_r \times L$. Finally, we can write the $KN_r \times N_t$ MIMO channel from BS c to all a users as $\mathbf{H}^{c\,H} = \overline{\mathbf{H}}^c_t\, \overline{\mathbf{A}}^{c\,H}\, \overline{\mathbf{H}}^{c\,H}_r$.

Theorem 1. *For a multi-cell MU MIMO system with $M \geq N_p$ and phasor antenna responses, to achieve optimal all-digital precoding performance, the analog beamformer can be chosen as the Tx side concatenated antenna array response.*

Proof: From [16] or [15, eq. (13)], the optimal all-digital beamformer is of the form

$$\begin{aligned} &(\mathbf{H}^{c\,H}\mathbf{D}^c_1\mathbf{H}^c + \lambda_c\mathbf{I})^{-1}\mathbf{H}^{c\,H}\mathbf{D}^c_2 \\ &= \mathbf{H}^{c\,H}\mathbf{B}^c = \overline{\mathbf{H}}^c_t\ \overline{\mathbf{A}}^{c\,H}\ \overline{\mathbf{H}}^{c\,H}_r\ \mathbf{B}^c \end{aligned} \tag{17}$$
$$\text{where } \mathbf{B}^c = \left(\lambda_c\mathbf{I} + \mathbf{D}^c_1\mathbf{H}^c\mathbf{H}^{c\,H}\right)^{-1}\mathbf{D}^c_2,$$

$\mathbf{D}^c_1$, $\mathbf{D}^c_2$ are block diagonal matrices and we used the identity $(\mathbf{I} + \mathbf{X}\mathbf{Y})^{-1}\mathbf{X} = \mathbf{X}(\mathbf{I} + \mathbf{Y}\mathbf{X})^{-1}$. Under the Theorem assumptions we can then separate the BFs as

$$\mathbf{V}^c = \overline{\mathbf{H}}^c_t,\ \mathbf{G}^c = \overline{\mathbf{A}}^{c\,H}\overline{\mathbf{H}}^{c\,H}_r\mathbf{B}^c. \tag{18}$$

Hence $\mathbf{V}$ depends only on the Tx antenna array responses. $\square$ Note that whereas the digital BF $\mathbf{G}$ in (18) is a function of the instantaneous CSIT, the analog BF $\mathbf{V}^c$ is only a function of AoDs, hence only of the slow fading channel components. This explains why the outdated CSIT based update for $\mathbf{V}$ in a mixed time scale scenario in [13] has a performance close to that of an instantaneous CSIT update based $\mathbf{V}$. Also, the theorem above motivates us to use the concatenated antenna array response matrix as the initialization of the analog BF for the Algorithm 1, when the number of RF chains M is greater than N_p or even when it is not, by taking the M strongest paths.

V. Deterministic Annealing for Global Convergence

In this section, we analyze how to improve the performance of the alternating optimization algorithm proposed (Algorithm 1) in the scenario in which the number of specular paths across all users exceed the number of RF chains. In the previous sections we considered the hybrid beamforming design using the WSR cost function which is a non-convex function. Due to which the algorithm will converge to different local optima depending on the initialization. So we consider here one approach called deterministic annealing (DA) to avoid the problem of local optima. In DA, we use a temperature parameter to track the global optimum with a homotopy method starting from a convex problem. Starting with a high temperature, where we know the optimal solution, we slowly decrease the temperature to reach the desired solution. If at high temperature we know the global optimum value, then if the temperature variations are slow, at the next value the global optimum will have the previous solution in its region of attraction. For the analog beamforming design using phasors, numerical results show that it converges to a local optimum and that it is very sensitive to the initialization used. In DA, we start from the optimal unconstrained $\mathbf{V}$ (note that HBF with factored digital and analog BFs has its own convexity issues that can be resolved with a separate DA strategy as in [15]). Then the gradual forcing of the amplitude of the unconstrained $\mathbf{V}$ entries to 1 allows to approach the global optimum. Here the amplitude relaxation parameter of each $\mathbf{V}$ entry is related to the temperature parameter. Note that in resulting Algorithm 2, d is some constant smaller than 1, say 0.9. The number of iterations required is a number of time constants of $e^{d\,t}$.

Algorithm 2 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 1 converge first.

1) Scale $\forall\,(i,j) : |\mathbf{V}^c_{i,j}| \leftarrow e^{d\ln|\mathbf{V}^c_{i,j}|}$.
2) Reoptimize all $\theta^c_{i,j}$ and all digital BFs using Algorithm 1.
3) Update stream powers and Lagrange multipliers.
4) Go to 1) for a number of iterations.
5) Finally redo 2)-3) a last time with all $|\mathbf{V}^c_{i,j}| = 1$ in 1).

VI. Simulation Results

Simulations to validate the performance of the proposed hybrid BF algorithms are presented for a single cell system with K single antenna users. The pathwise channel model $\mathbf{h}_k$ for user k can be written as $\mathbf{h}_k = \sum_{i=1}^{L} \alpha_{k,i}\,\mathbf{h}_t(\phi_{k,i})$, where $\alpha_{k,i}$ are the complex path gains which are assumed to be Gaussian with variance distributed according to an exponential profile. In the Uniform Linear Array (ULA) $\mathbf{h}_t\,(\phi_{k,i})$, the AoD $\phi_{k,i}$ are assumed to be uniformly distributed in the interval $[0^o, 30^o]$. Furthermore, we consider the case in which the number of RF chains $M < LK$ (with local optima issues). Notations used in the figure: CoCSIT refers to covariance CSIT and EV refers to dominant eigen vectors of the sum of the channel covariance matrices of all users. We compare the performance of the proposed algorithms with the optimal fully digital BF [16] (referred to as "Optimal Fully Digital [Christensen et al]"), approximate WSR based hybrid design [8] (referred to as "Approximate WSR [Sohrabi, Wei Yu]", WSMSE based alternating optimization [13] (referred to as "WSMSE HBF") and the covariance CSIT based scheme [19] (referred to as "V CoCSIT and G R-ZF [S.Park et al]"). "V Random Initialization" refers to the case when Algorithm 1 starts with random phases for the analog BF.

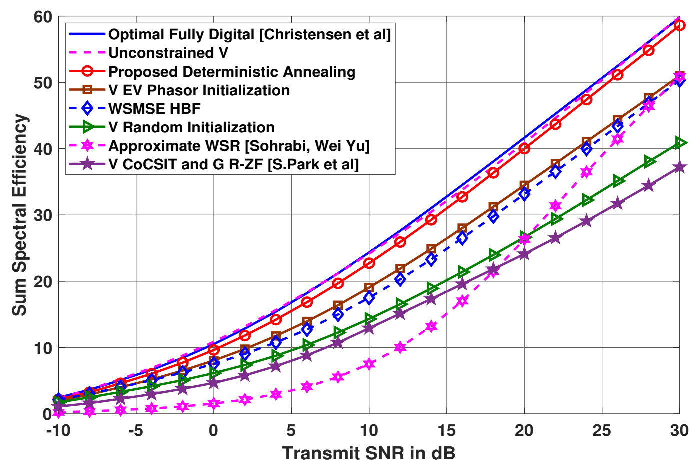

Fig. 1. Sum Rate comparisons for, $N_t = 32, M = 16, K = 8, C = 1, L = 4$ paths.

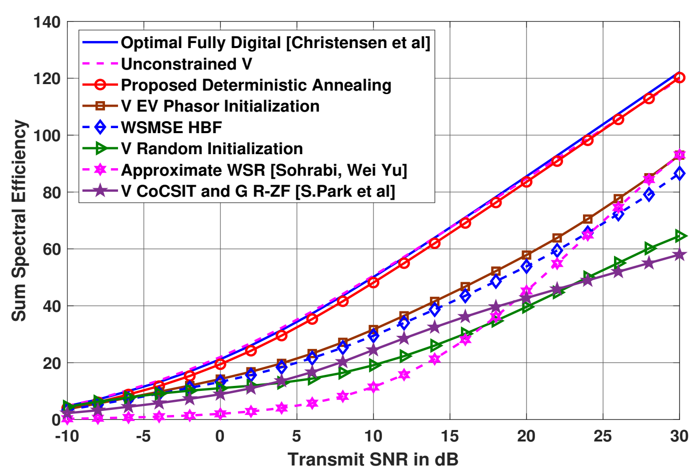

Fig. 2. Sum Rate comparisons for, $N_t = 64, M = 16, K = 16, C = 1, L = 2$ paths.

It is evident from the figures that the DA based approach (Algorithm 2) performs significantly better than just alternating optimization (Algorithm 1) and also the state of the art methods.

VII. Conclusion

In this paper, we derived and presented an optimal beamforming algorithm for the hybrid beamforming scenario in a Multi-cell MU-MIMO system. An iterative algorithm is obtained which jointly optimizes both analog and digital beamformers. In order to solve the issue of local optima in the non-concave WSR we introduced deterministic annealing for the analog phasors design. Simulation results indicate that the resulting global optimum is much better than typical local optima and that the thus optimized HBF performance can be very close to the optimal fully digital performance.

Appendix A
Derivation of phasors in $\mathbf{V}$

Adding the phase shifter constraint, we identify the dependence of (7) on a single element $\mathbf{V}^c_{p,q}$. We simplify each of the quadratic terms in the expression for WSR. First let us consider each element (r,s) of the matrix $\mathbf{G}_k^H \mathbf{V}^{c\,H} \widehat{\mathbf{B}}_k \mathbf{V}^c \mathbf{G}_k$ (for $k : b_k = c$),

$$\begin{aligned}&\mathbf{g}_k^{(r)\,H}\mathbf{V}^{c\,H}\widehat{\mathbf{B}}_k\mathbf{V}^c\mathbf{g}_k^{(s)} = ((\mathbf{V}^c\mathbf{g}_k^{(r)})_p)^H(\widehat{\mathbf{B}}_k)_{p,p}(\mathbf{V}^c\mathbf{g}_k^{(s)})_p\\ &+((\mathbf{V}^c\mathbf{g}_k^{(r)})_{\overline{p}})^H(\widehat{\mathbf{B}}_k)_{\overline{p},p}(V^c\mathbf{g}_k^{(s)})_p+\\ &((\mathbf{V}^c\mathbf{g}_k^{(r)})_p)^H(\widehat{\mathbf{B}}_k)_{p,\overline{p}}(\mathbf{V}^c\mathbf{g}_k^{(s)})_{\overline{p}}\\ &+((\mathbf{V}^c\mathbf{g}_k^{(r)})_{\overline{p}})^H(\widehat{\mathbf{B}}_k)_{\overline{p},\overline{p}}(\mathbf{V}^c\mathbf{g}_k^{(s)})_{\overline{p}},\end{aligned} \tag{19}$$

where $(\mathbf{x})_p$ represents the p^{th} element of vector $\mathbf{x}$, $(\mathbf{x})_{\overline{p}}$ represents all other elements, $(\mathbf{B})_{p,p}$ represents element (p,p) of matrix $\mathbf{B}$, $(\mathbf{B})_{\overline{p},p}$ represents all elements in column p except for row p, etc. Note that $\left(\mathbf{V}^c\mathbf{g}_k^{(r)}\right)_{\overline{p}}$ does not contain $\mathbf{V}^c_{p,q}$. The p^{th} term of $\mathbf{V}^c\mathbf{g}_k^{(r)}$ can be written in terms of $\mathbf{V}^c_{p,q}$ as :

$$\left(\mathbf{V}^c\mathbf{g}_k^{(r)}\right)_p = \mathbf{V}^c_{p,q}\,\mathbf{g}_{k,q}^{(r)} + \mathbf{V}^c_{p,\overline{q}}\,\mathbf{g}_{k,\overline{q}}^{(r)}, \tag{20}$$

where $\mathbf{V}^c_{p,l}$ represents element (p,l) element of $\mathbf{V}^c$ and $\mathbf{g}^r_{k,q}$ represents the q^{th} element of $\mathbf{g}_k^{(r)}$, $\mathbf{g}_{k,\overline{q}}^{(r)}$ represents all other elements, etc. Now substituting $\mathbf{V}^c_{p,q} = e^{j\theta^c_{p,q}}$, (19) can be written as :

$$\begin{aligned}&\mathbf{g}_k^{(r)\,H}\mathbf{V}^{c\,H}\widehat{\mathbf{B}}_k\mathbf{V}^c\mathbf{g}_k^{(s)} = (\mathbf{V}^{c\,H}_{p,\overline{q}}\mathbf{g}_{k,\overline{q}}^{(r)})^H(\widehat{\mathbf{B}}_k)_{p,p}e^{j\theta^c_{p,q}}\mathbf{g}_{k,q}^{(s)}+\\ &(\mathbf{V}^c_{p,\overline{q}}\mathbf{g}_{k,\overline{q}}^{(s)})(\widehat{\mathbf{B}}_k)_{p,p}e^{-j\theta^c_{p,q}}\mathbf{g}_{k,q}^{(r)\,H}\\ &+((\mathbf{V}^c\mathbf{g}_k^{(r)})_{\overline{p}})^H(\widehat{\mathbf{B}}_k)_{\overline{p},p}e^{j\theta^c_{p,q}}\mathbf{g}_{k,q}^{(s)}\\ &+(\widehat{\mathbf{B}}_k)_{p,\overline{p}}(\mathbf{V}^c\mathbf{g}_k^{(s)})_{\overline{p}}e^{-j\theta^c_{p,q}}\mathbf{g}_{k,q}^{(r)\,H} + \text{“terms”}.\end{aligned} \tag{21}$$

Here “terms” denote the terms which are independent of $\mathbf{V}^c_{p,q}$. Define the following matrices $\mathbf{C}_k^{p,q}$ and $\mathbf{D}_k^{p,q}$ whose entries are,

$$\begin{aligned}(\mathbf{D}_k^{p,q})_{r,s} &= (\mathbf{V}^{c\,H}_{p,\overline{q}}\mathbf{g}_{k,\overline{q}}^{(r)})^H(\widehat{\mathbf{B}}_k)_{p,p}\mathbf{g}_{k,q}^{(s)}\\ &\quad+((\mathbf{V}^c\mathbf{g}_k^{(r)})_{\overline{p}})^H(\widehat{\mathbf{B}}_k)_{\overline{p},p}\mathbf{g}_{k,q}^{(s)},\\ (\mathbf{C}_k^{p,q})_{r,s} &= (\mathbf{V}^c_{p,\overline{q}}\mathbf{g}_{k,\overline{q}}^{(s)})(\widehat{\mathbf{B}}_k)_{p,p}\mathbf{g}_{k,q}^{(r)\,H}\\ &\quad+(\widehat{\mathbf{B}}_k)_{p,\overline{p}}(\mathbf{V}^c\mathbf{g}_k^{(s)})_{\overline{p}}\mathbf{g}_{k,q}^{(r)\,H}.\end{aligned} \tag{22}$$

Then we can rewrite $\mathbf{G}_k^H\mathbf{V}^{c\,H}\widehat{\mathbf{B}}_k\mathbf{V}^c\mathbf{G}_k$ as

$$\mathbf{G}_k^H\mathbf{V}^{c\,H}\widehat{\mathbf{B}}_k\mathbf{V}^c\mathbf{G}_k = \mathbf{D}_k^{p,q}e^{j\theta^c_{p,q}} + \mathbf{C}_k^{p,q}e^{-j\theta^c_{p,q}} + \mathbf{T}^{k,1}_{\overline{p},\overline{q}}. \tag{23}$$

Similarly we can write,

$$\mathbf{G}_k^H\mathbf{V}^{c\,H}(\widehat{\mathbf{A}}_k+\lambda_c\mathbf{I})\mathbf{V}^c\mathbf{G}_k = \mathbf{E}_k^{p,q}e^{j\theta^c_{p,q}} + \mathbf{F}_k^{p,q}e^{-j\theta^c_{p,q}} + \mathbf{T}^{k,2}_{\overline{p},\overline{q}}. \tag{24}$$

Here $\mathbf{T}^{k,1}_{\overline{p},\overline{q}}, \mathbf{T}^{k,2}_{\overline{p},\overline{q}}$ are matrices with terms independent of $\theta^c_{p,q}$.

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, pp. 4091–4103, 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, pp. 1499–1513, 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, pp. 831–846, October 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] A. Alkhateeb, R. W. H. Jr., and G. Leus, “Achievable rates of multi-user millimeter wave systems with hybrid precoding,” in *Proc. IEEE Int.Conf. Commun*, London, UK, June 2015, pp. 1232–1237.

[6] Q. Shi, M. Razaviyayn, Z. Q. Luo, and C. He, “Limited feedback hybrid precoding for multi-user millimeter wave systems,” *IEEE Trans. on Wireless Commun*, vol. 14, no. 11, pp. 6481–6494, November 2015.

[7] 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 analog-digital beamforming,” *IEEE Trans. on Wireless Commun.*, vol. 15, no. 5, May 2016.

[8] 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, pp. 501–513, April 2016.

[9] R. Mndez-Rial, N. Gonzlez-Prelcic, and R. W. Heath, “Adaptive hybrid precoding and combining in mmWave multiuser MIMO systems based on compressed covariance estimation,” in *IEEE 6th Intl. Wkshp. on Comput. Adv. in Multi-Sensor Adapt. Process. (CAMSAP)*, Cancun, Mexico, December 2015.

[10] D. H. N. Nguyen, L. B. Le, and T.Le-Ngoc, “Hybrid MMSE precoding for mmWave multiuser MIMO systems,” in *IEEE Intl. Conf. on Commun.*, Kuala Lumpur, Malaysia, May 2016.

[11] 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, pp. 19 167–19 181, April 2017.

[12] L. Liang, Y. Dai, W. Xu, and X. Dong, “How to approach zero-forcing under RF chain limitations in large mmWave multiuser systems ?” in *Proc. IEEE/CIC Intl. Conf. Commun, China*, October 2014.

[13] C. K. Thomas and D. Slock, “Mixed time scale weighted sum rate maximization for hybrid beamforming in multi-cell MU-MIMO systems,” in *Globecom Wkshps.*, Singapore, December 2017.

[14] M. Kim and Y. H. Lee, “MSE-based hybrid rf/baseband processing for millimeter-wave communication systems in MIMO interference channels,” *IEEE Trans. on Vehicular Technology*, vol. 64, no. 6, pp. 2714–2720, June 2015.

[15] 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.

[16] 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.*, December 2008.

[17] S. J. Kim and G. B. Giannakis, “Optimal resource allocation for MIMO ad-hoc cognitive radio networks,” in *IEEE Trans. on Inf. Theory*, vol. 57, May 2011, pp. 3117–3131.

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

[19] 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.