LOW RANK APPROXIMATION BASED HYBRID PRECODING SCHEMES FOR MULTI-CARRIER SINGLE-USER MASSIVE MIMO SYSTEMS

Jianshu Zhang [1], *Ami Wiesel* [2], *Martin Haardt* [1]

[1]Communications Research Laboratory
Ilmenau University of Technology, Germany

[2] School of Computer Science and Engineering
The Hebrew University of Jerusalem, Israel

ABSTRACT

In this paper we study the hybrid precoding design problem for a frequency selective massive MIMO channel, e.g., the millimeter wave (mmWave) massive MIMO channel. In contrast to a traditional MIMO system, a hybrid analog-digital MIMO scheme is preferred for massive MIMO systems due to the high cost and power consumption of the radio frequency (RF) chains. The RF analog precoding is implemented using only phase shift networks, which impose constant modulus constraints on the RF precoding and decoding matrices. Moreover, there is just one common equivalent RF beamforming matrix for all subcarriers. The resulting sum rate maximization problem is non-convex and, therefore we resort to suboptimal solutions. Two methods are introduced, namely, the higher order SVD (HOSVD) based design and the sequential low rank unimodular approximation based design. The former approach exploits the truncated HOSVD of the equivalent channel while the latter approach approximates optimal unconstrained solutions by low rank unimodular approximations. Simulation results show that when the mmWave channel model is used, both approaches outperform the extension of the state of the art compressed sensing based algorithm to the multi-carrier case.

Index Terms— Massive MIMO, hybrid precoding, mmWave, higher order SVD, convex optimization.

I. INTRODUCTION

The massive MIMO technique, which uses orders of magnitude more antennas (e.g., 100 or more), can provide significant MIMO gains [1]. When combined with millimeter wave (mmWave) technology, it will not only gain from large chunks of underutilized spectrum in the mmWave band [2] but will also benefit from a significantly reduced form factor of the massive MIMO array [3]. Hence, mmWave massive MIMO communication is a potential technique for future wireless networks [4]. However, if a large number of RF chains is implemented to steer the massive number of antenna elements, the involved power consumption and the hardware cost are too high and therefore are impractical. To exploit the MIMO multiplexing gain under a reasonable cost, one promising solution is to deploy hybrid analog precoding schemes, realized using phase shifters or switches in the RF domain [5], and digital precoding schemes, implemented in the digital baseband domain as in conventional MIMO. If analog precoding is achieved using phase shifters only, the analog precoding matrix should have only constant modulus entries, which are stringent constraints that lead to significant challenges for the required signal processing [6]. Several papers have recently studied such hybrid precoding design problems, including [7], [8], [9], [10]. In [7] a compressed sensing based hybrid precoding is proposed to approximate the optimal unconstrained solution of a point-to-point MIMO system and is further refined in [8]. In [9] a low complexity codebook based hybrid precoding scheme is introduced for the multi-user mmWave downlink channel. A multi-user multi-carrier system is studied in [10]. However, the focus of [10] is on the required number of RF chains and phase shifters to achieve the performance of pure digital beamforming. Furthermore, the considered single stream beamforming solution is not general and cannot simply be used for the single user case. Hence, this motivates us to develop optimal hybrid precoding schemes for single user multi-carrier systems.

In this paper we study the hybrid precoder and decoder design for a single user multi-carrier massive MIMO system to achieve a MIMO multiplexing gain. The cyclic prefix OFDM (CP-OFDM) based multi-carrier modulation scheme is used. The RF precoding is implemented using only phase shifters, i.e., only constant modulus entries are allowed. Hence, equivalently we get the same phase shifts for all subcarriers. Due to these two constraints, the resulting sum rate maximization problem is non-convex and it might be intractable to find exact solutions. Thus, we resort to suboptimal solutions. First, if the constant modulus constraints are dropped, the truncated higher order singular value decomposition (HOSVD) provides common basis matrices for the subspaces spanned by all the channels, which are good candidates for the RF matrices. The optimal digital solutions are then obtained through the truncated SVD of the effective channels. Finally, under the constant modulus constraints approximate solutions are provided for the RF precoding and decoding matrices. Alternatively, without the constant modulus constraints hybrid precoding and decoding schemes can also be obtained via a low rank approximation of the optimal unconstrained solution. Such a low rank approximation is in general not available given the constant modulus constraints. Nevertheless, we derive a rank-N approximation of the optimal unconstrained solution by sequentially computing the best rank-1 unimodular approximations N times. This is inspired by the fact that we can obtain a local optimal solution of the best rank-1 unimodular approximation problem. Simulation results show that when the mmWave channel model is used, both proposed algorithms outperform the extension of the sparse precoding solution in [8] to the multi-carrier case. Moreover, they achieve almost the same multiplexing gain as the optimal unconstrained solution.

Notation: The operator $\|\cdot\|_{\mathrm{F}}$ denotes the Frobenius norm. The concatenation of matrices or tensors along the r-th dimension is denoted by $\sqcup_r$ ($r = 1, 2, 3$) [11]. The r-mode product between a tensor and a matrix is $\times_r$ [12]. The operation $|\cdot|$ computes the magnitude of a scalar or the determinant of a matrix, $(\cdot)^*$ denotes the complex conjugate operation, and $\angle(\boldsymbol{A})$ computes the phases of the matrix $\boldsymbol{A}$ element-wise.

II. PROBLEM FORMULATION

We begin with a brief review of the standard point-to-point MIMO system where a multi-antenna base station (BS) transmits data to a multi-antenna user equipment (UE). The BS has M_{T} transmit antennas with $N_{\mathrm{T}}^{(\mathrm{RF})} = M_{\mathrm{T}}$ RF chains. The UE has M_{R} receive antennas and $N_{\mathrm{R}}^{(\mathrm{RF})} = M_{\mathrm{R}}$ RF chains. A CP-OFDM based multi-carrier modulation technique is applied to combat the multipath effect. The corresponding FFT size is N_{FFT}. Let

$\boldsymbol{s}[m] \in \mathbb{C}^{N_{\rm ss}}$ represent the transmitted symbol vector on the m-th subcarrier of the BS ($m \in \{1, \cdots, N_{\rm FFT}\}$), where $\boldsymbol{s}[m]$ has zero mean and $\mathbb{E}\{\boldsymbol{s}[m]\boldsymbol{s}^{\rm H}[m]\} = \boldsymbol{I}_{N_{\rm ss}}$. A digital precoding matrix $\boldsymbol{F}[m] \in \mathbb{C}^{M_{\rm T}\times N_{\rm ss}}$ is applied on a per-subcarrier basis, i.e., the transmitted frequency domain signal matrix $\boldsymbol{X}_{\rm BB} = [\boldsymbol{F}[1]\boldsymbol{s}[1] \quad \cdots \quad \boldsymbol{F}[N_{\rm FFT}]\boldsymbol{s}[N_{\rm FFT}]] \in \mathbb{C}^{N_{\rm T}^{\rm (RF)}\times N_{\rm FFT}}$. Afterwards, the precoded signal passes through the IFFT filter and a CP of length $N_{\rm CP}$ symbols is added. Finally, the average transmit power constraint has to be fulfilled such that $\mathbb{E}\{\|\boldsymbol{X}_{\rm BB}\|_{\rm F}^2\} \leq P_{\rm T}$.

We consider a frequency selective quasi-static block fading channel. Assume that $N_{\rm CP}$ symbols have the same length as the maximum excess delay of the channel such that the inter-symbol interference is avoided. After passing through the channel, first, the CP is removed from the received signal and by using the FFT filter the time domain signal is transformed into the frequency domain. Let $\boldsymbol{H}[m] \in \mathbb{C}^{M_{\rm R}\times M_{\rm T}}$ and $\boldsymbol{W}^{\rm H}[m] \in \mathbb{C}^{N_{\rm SS}\times M_{\rm R}}$ denote the discrete channel transfer function (CTF) and the digital decoding matrix on m-th subcarrier of the UE, respectively. The estimated received signal on the m-th subcarrier is obtained as

$$\hat{\boldsymbol{s}}[m] = \boldsymbol{W}^{\rm H}[m]\left(\boldsymbol{H}[m]\boldsymbol{F}[m]\boldsymbol{s}[m] + \boldsymbol{z}[m]\right), \tag{1}$$

where $\boldsymbol{z}[m]$ is the zero mean circularly symmetric complex Gaussian (ZMCSCG) noise with covariance matrix $\mathbb{E}\{\boldsymbol{z}[m]\boldsymbol{z}^{\rm H}[m]\} = \sigma_{\rm n}^2\boldsymbol{I}_{M_{\rm R}}$ for all m.

Assume that perfect channel knowledge is available at both the BS and the UE. The traditional joint MIMO precoding and decoding problem is to find matrices $\boldsymbol{F}[m]$ and $\boldsymbol{W}[m]$, $\forall m$, such that the achievable rate is maximized subject to the transmit power constraint at the BS, i.e.,

$$\begin{aligned}&\max_{\boldsymbol{F}[m],\boldsymbol{W}[m]\forall m} \quad \sum_{m=1}^{N_{\rm FFT}} \log_2\left|\boldsymbol{I}_{\rm ss} + \boldsymbol{R}_{\rm s}[m]\sigma_{\rm n}^2(\boldsymbol{W}^{\rm H}[m]\boldsymbol{W}[m])^{-1}\right|\\ &\text{s.t.} \quad \boldsymbol{R}_{\rm s}[m] = \boldsymbol{W}^{\rm H}[m]\boldsymbol{H}[m]\boldsymbol{F}[m]\cdot\Big(\boldsymbol{W}^{\rm H}[m]\boldsymbol{H}[m]\boldsymbol{F}[m]\Big)^{\rm H}\\ &\qquad \sum_{m=1}^{N_{\rm FFT}} \|\boldsymbol{F}[m]\|_{\rm F}^2 \leq P_{\rm T},\end{aligned} \tag{2}$$

This classical problem has a well known solution. The pairs of precoding and decoding matrices for each sub-carrier are easily obtained by the decoupled SVDs of each of the MIMO channels. Consider the m-th subcarrier and define the SVD of $\boldsymbol{H}[m]$ as

$$\boldsymbol{H}[m] = \boldsymbol{U}_{\rm s}[m]\boldsymbol{\Sigma}_{\rm s}[m]\boldsymbol{V}_{\rm s}^{\rm H}[m] + \boldsymbol{U}_{\rm n}[m]\boldsymbol{\Sigma}_{\rm n}[m]\boldsymbol{V}_{\rm n}^{\rm H}[m],$$

where $\boldsymbol{U}_{\rm s}[m] \in \mathbb{C}^{M_{\rm R}\times N_{\rm ss}}$ and $\boldsymbol{V}_{\rm s}[m] \in \mathbb{C}^{M_{\rm T}\times N_{\rm ss}}$ are the singular vectors corresponding to the dominant $N_{\rm ss}$ singular values. Then the optimal precoding and decoding matrices are

$$\boldsymbol{W}_{\rm opt}[m] = \boldsymbol{U}_{\rm s}[m] \qquad \boldsymbol{F}_{\rm opt}[m] = \boldsymbol{V}_{\rm s}[m], \tag{3}$$

To ensure optimality, these digital precoding and decoding matrices must also be loaded with powers according to the waterfilling algorithm [13]. For simplicity, hereinafter we will assume high SNR where uniform power allocation is almost optimal. To summarize this review, we emphasize that the core property of the solution is that it is based on *decoupled and unconstrained SVDs* for each subcarrier.

Armed with this understanding, we now turn to the more challenging hybrid massive MIMO formulation. In contrast to the classical model, this formulation assumes that the precoding and decoding are achieved by hybrid analog and digital schemes. Moreover, the number of RF chains is assumed to be much smaller than the number of antenna elements, i.e., $M_{\rm T} \gg N_{\rm T}^{\rm (RF)}$ and $M_{\rm R} \gg N_{\rm R}^{\rm (RF)}$. We have $\min(N_{\rm T}^{\rm (RF)}, N_{\rm R}^{\rm (RF)}) \geq N_{\rm ss}$, where $N_{\rm ss}$ denotes the number of spatial streams. A CP-OFDM based technique is again used to combat the multipath effect. A digital precoding matrix $\boldsymbol{F}_{\rm BB}[m] \in \mathbb{C}^{N_{\rm T}^{\rm (RF)}\times N_{\rm ss}}$ is applied on a per-subcarrier basis. Afterwards, the precoded signal passes through the IFFT filter and a CP of length $N_{\rm CP}$ symbols is added, followed by an RF precoder $\boldsymbol{F}_{\rm RF} \in \mathbb{C}^{M_{\rm T}\times N_{\rm T}^{\rm (RF)}}$ using analog circuitry. We assume that the RF precoder is implemented using analog phase shifters. Hence, constant modulus constraints should be fulfilled for each element of $\boldsymbol{F}_{\rm RF}$, i.e., $|F_{{\rm RF},j,a}| = 1$ for all $j \in \{1, \cdots, M_{\rm T}\}$ and $a \in \{1, \cdots, N_{\rm T}^{\rm (RF)}\}$. The transmit power constraint is now expressed as $\mathbb{E}\{\|\boldsymbol{F}_{\rm RF}\boldsymbol{X}_{\rm BB}\|_{\rm F}^2\} \leq P_{\rm T}$. This practical formulation leads to two additional constraints on (2)

$$\{\boldsymbol{F}_{\rm RF}, \boldsymbol{W}_{\rm RF}\} \in \mathcal{F}^{\rm (RF)} \tag{4a}$$
$$\boldsymbol{W}[m] = \boldsymbol{W}_{RF}\boldsymbol{W}_{BB}[m], \quad \boldsymbol{F}[m] = \boldsymbol{F}_{RF}\boldsymbol{F}_{BB}[m], \tag{4b}$$

where $\mathcal{F}^{\rm (RF)}$ represents a general set of matrices or vectors with constant modulus entries. Thus, the hybrid precoding and decoding matrices are restricted in two challenging manners. First, the RF matrices are constant modulus matices. Second, the solutions across different subcarriers are now coupled via their joint RF solution [10]. The goal of this paper is to provide efficient solutions to problem (2) subject to the challenging constraints in (4). From a linear algebra perspective, we seek a joint matrix decomposition of multiple matrices where a joint basis consists of constant modulus elements. Our derivations will be based on two competing approaches: consecutive SVD via truncated HOSVD and concatenated SVD via sequential rank-1 decompositions.

III. Solution via HOSVD

In this section we propose an HOSVD based framework for hybrid analog-digital precoding. Indeed, the added subcarrier dimension, naturally suggests a higher order decomposition. Our method has two steps: First we consider the joint RF optimization via a truncated HOSVD. Then, we derive the decoupled digital matrices using standard SVDs over the effective channels on each subcarrier.

Define the tensor representation of the overall channel as

$$\boldsymbol{\mathcal{H}} = [\boldsymbol{H}[1]\sqcup_3 \quad \cdots \quad \sqcup_3\boldsymbol{H}[N_{\rm FFT}]] \in \mathbb{C}^{M_{\rm R}\times M_{\rm T}\times N_{\rm FFT}} \tag{5}$$

The HOSVD of $\boldsymbol{\mathcal{H}}$ is defined as

$$\boldsymbol{\mathcal{H}} = \boldsymbol{\mathcal{S}}_{M_{\rm R}\times M_{\rm T}\times N_{\rm FFT}} \times_1 \boldsymbol{U}_1 \times_2 \boldsymbol{U}_2 \times_3 \boldsymbol{U}_3 \tag{6}$$

where $\boldsymbol{U}_1 \in \mathbb{C}^{M_{\rm R}\times M_{\rm R}}$, $\boldsymbol{U}_2 \in \mathbb{C}^{M_{\rm T}\times M_{\rm T}}$, and $\boldsymbol{U}_3 \in \mathbb{C}^{N_{\rm FFT}\times N_{\rm FFT}}$ are unitary matrices. It is worth mentioning that the columns of $\boldsymbol{U}_1$ and $\boldsymbol{U}_2$ provide an orthonormal basis for the column space spanned by $\boldsymbol{H}_1 = [\boldsymbol{H}[1] \quad \cdots \quad \boldsymbol{H}[N_{\rm FFT}]] \in \mathbb{C}^{M_{\rm R}\times M_{\rm T}N_{\rm FFT}}$ and $\boldsymbol{H}_2 = \left[\boldsymbol{H}[1]^{\rm T} \quad \cdots \quad \boldsymbol{H}[N_{\rm FFT}]^{\rm T}\right] \in \mathbb{C}^{M_{\rm T}\times M_{\rm R}N_{\rm FFT}}$, respectively. Let $\boldsymbol{U}_{1,{\rm t}}$ and $\boldsymbol{U}_{2,{\rm t}}$ contain the first $N_{\rm R}^{\rm (RF)}$ and $N_{\rm T}^{\rm (RF)}$ columns of $\boldsymbol{U}_1$ and $\boldsymbol{U}_2$, respectively. Then we define the rank-$(N_{\rm R}^{\rm (RF)}, N_{\rm T}^{\rm (RF)}, N_{\rm FFT})$ truncated HOSVD as

$$\boldsymbol{\mathcal{H}}^{\rm trun} \approx \tilde{\boldsymbol{\mathcal{S}}}_{N_{\rm R}^{\rm (RF)}\times N_{\rm T}^{\rm (RF)}\times N_{\rm FFT}} \times_1 \boldsymbol{U}_{1,{\rm t}} \times_2 \boldsymbol{U}_{2,{\rm t}} \times_3 \boldsymbol{U}_3. \tag{7}$$

Using (7) the tensor based effective channel is obtained as

$$\boldsymbol{\mathcal{H}}^{\rm (eff)} = \tilde{\boldsymbol{\mathcal{S}}}_{N_{\rm R}^{\rm (RF)}\times N_{\rm T}^{\rm (RF)}\times N_{\rm FFT}} \times_1 (\boldsymbol{W}_{\rm RF}^{\rm H}\boldsymbol{U}_{1,{\rm t}}) \times_2 (\boldsymbol{F}_{\rm RF}^{\rm T}\boldsymbol{U}_{2,{\rm t}}) \times_3 \boldsymbol{U}_3.$$

Let us first consider the coupled solution ignoring the constant modulus constraints. Since $\boldsymbol{U}_1$ and $\boldsymbol{U}_2$ span a common column and row space of the column and row spaces of all the channels, it is reasonable to choose columns of $\boldsymbol{U}_1$ and $\boldsymbol{U}_2$ as the common RF matrices. Therefore, with only constraint (4b), the RF matrices are

$$\boldsymbol{W}_{\rm RF} = \boldsymbol{U}_{1,{\rm t}} \qquad \boldsymbol{F}_{\rm RF} = \boldsymbol{U}_{2,{\rm t}}^*. \tag{8}$$

Assuming high SNR, the optimal decoupled digital beamforming solutions on each subcarrier are given by

$$\boldsymbol{W}_{\mathrm{BB}}[m] = \tilde{\boldsymbol{U}}_{\mathrm{s}}[m] \qquad \boldsymbol{F}_{\mathrm{BB}}[m] = \tilde{\boldsymbol{V}}_{\mathrm{s}}[m], \tag{9}$$

where $\tilde{\boldsymbol{U}}_{\mathrm{s}}[m] \in \mathbb{C}^{N_{\mathrm{R}}^{(\mathrm{RF})} \times N_{\mathrm{ss}}}$ and $\tilde{\boldsymbol{V}}_{\mathrm{s}}[m] \in \mathbb{C}^{N_{\mathrm{T}}^{(\mathrm{RF})} \times N_{\mathrm{ss}}}$ consist of the dominant N_{ss} left and right singular vectors of the effective channel $\boldsymbol{U}_{1,\mathrm{t}} \boldsymbol{H}[m] \boldsymbol{U}_{2,\mathrm{t}}^{*}$ on the m-th subcarrier.

Implementation and complexity: Numerically $\boldsymbol{U}_1$ and $\boldsymbol{U}_2$ can be computed from the eigendecomposition of $\boldsymbol{H}_1 \boldsymbol{H}_1^{\mathrm{H}}$ and $\boldsymbol{H}_2 \boldsymbol{H}_2^{\mathrm{H}}$. If a cubic order complexity for the eigendecomposition is used, the truncated HOSVD based algorithm has a complexity of $\mathcal{O}(\max(M_{\mathrm{T}}^3, M_{\mathrm{T}}^2 M_{\mathrm{R}} N_{\mathrm{FFT}}))$ and $\mathcal{O}(\max(M_{\mathrm{R}}^3, M_{\mathrm{R}}^2 M_{\mathrm{T}} N_{\mathrm{FFT}}))$ for the computation of the precoding matrices and the decoding matrices, respectively.

Motivated by the coupled hybrid solution, we continue and add the constant modulus constraints in (4a). Unfortunately, these lead to a difficult non-convex optimization. We do not expect an optimal solution. Instead, we propose a simple natural heuristic

$$\boldsymbol{W}_{\mathrm{RF}} = e^{j\angle(\boldsymbol{U}_{1,\mathrm{t}})}, \boldsymbol{F}_{\mathrm{RF}} = e^{j\angle(\boldsymbol{U}_{2,\mathrm{t}}^{*})}. \tag{10}$$

IV. SOLUTION VIA CONCATENATED SVD

In this section we propose an alternative approach based on a single SVD with concatenated matrices. We begin with the classical optimal and unconstrained solution in (3) and try to approximate them within the feasible set defined by (4a) and (4b). As before, we first drop the constant modulus constraints. In what follows, we describe the approach with respect to the precoding matrices, and assume that an identical approach is applied on the decoding side as well. This leads to the following optimization

$$\min_{\boldsymbol{F}_{\mathrm{RF}}, \tilde{\boldsymbol{F}}_{\mathrm{BB}}[m]} \quad \sum_{m} \|\boldsymbol{F}_{\mathrm{opt}}[m] - \boldsymbol{F}_{\mathrm{RF}} \tilde{\boldsymbol{F}}_{\mathrm{BB}}[m]\|_{\mathrm{F}}^2 \tag{11}$$

This problem can be solved by a single SVD over concatenated matrices in a least squares (LS) sense. Define $\bar{\boldsymbol{F}}_{\mathrm{opt}} = [\boldsymbol{F}_{\mathrm{opt}}[1] \quad \cdots \quad \boldsymbol{F}_{\mathrm{opt}}[N_{\mathrm{FFT}}]] \in \mathbb{C}^{M_{\mathrm{T}} \times N_{ss} N_{\mathrm{FFT}}}$ and $\bar{\boldsymbol{F}}_{\mathrm{BB}} = \left[\tilde{\boldsymbol{F}}_{\mathrm{BB}}[1] \quad \cdots \quad \tilde{\boldsymbol{F}}_{\mathrm{BB}}[N_{\mathrm{FFT}}]\right] \in \mathbb{C}^{N_{\mathrm{T}}^{(\mathrm{RF})} \times N_{ss} N_{\mathrm{FFT}}}$. This yields

$$\min_{\boldsymbol{F}_{\mathrm{RF}}, \bar{\boldsymbol{F}}_{\mathrm{BB}}} \quad \|\bar{\boldsymbol{F}}_{\mathrm{opt}} - \boldsymbol{F}_{\mathrm{RF}} \bar{\boldsymbol{F}}_{\mathrm{BB}}\|_{\mathrm{F}}^2 \tag{12}$$

Problem (12) is non-convex, yet it has a closed form solution via a truncated SVD

$$\boldsymbol{F}_{\mathrm{RF}} = \bar{\boldsymbol{U}}_{\mathrm{s}} \qquad \bar{\boldsymbol{F}}_{\mathrm{BB}} = \bar{\boldsymbol{V}}_{\mathrm{s}} \bar{\boldsymbol{\Sigma}}_{\mathrm{s}} \tag{13}$$

where $\bar{\boldsymbol{U}}_{\mathrm{s}} \in \mathbb{C}^{M_{\mathrm{T}} \times N_{\mathrm{T}}^{(\mathrm{RF})}}$ and $\bar{\boldsymbol{V}}_{\mathrm{s}} \in \mathbb{C}^{N_{\mathrm{ss}} N_{\mathrm{FFT}} \times N_{\mathrm{T}}^{(\mathrm{RF})}}$ contain the dominant $N_{\mathrm{T}}^{(\mathrm{RF})}$ left and right singular vectors of $\bar{\boldsymbol{F}}_{\mathrm{opt}}$. The matrix $\bar{\boldsymbol{\Sigma}}_{\mathrm{opt,s}} \in \mathbb{C}^{N_{\mathrm{T}}^{(\mathrm{RF})} \times N_{\mathrm{T}}^{(\mathrm{RF})}}$ is a diagonal matrix and its main diagonal elements are the dominant $N_{\mathrm{T}}^{(\mathrm{RF})}$ singular values of $\bar{\boldsymbol{F}}_{\mathrm{opt}}$. In practice, we also need to ensure that the transmit power constraint is satisfied. Therefore, the final $\boldsymbol{F}_{\mathrm{BB}} = [\boldsymbol{F}_{\mathrm{BB}}[1] \quad \cdots \quad \boldsymbol{F}_{\mathrm{BB}}[N_{\mathrm{FFT}}]]$ is defined as a scaled version of $\bar{\boldsymbol{F}}_{\mathrm{BB}}$ such that $\|\boldsymbol{F}_{\mathrm{RF}} \boldsymbol{F}_{\mathrm{BB}}\|_{\mathrm{F}}^2 = P_{\mathrm{T}}$.

Using the above intuition, we turn to the constant modulus constraint in (4a). More precisely, we need to solve

$$\min_{\boldsymbol{F}_{\mathrm{RF}}, \bar{\boldsymbol{F}}_{\mathrm{BB}}} \quad \|\bar{\boldsymbol{F}}_{\mathrm{opt}} - \boldsymbol{F}_{\mathrm{RF}} \bar{\boldsymbol{F}}_{\mathrm{BB}}\|_{\mathrm{F}}^2, \quad \text{s.t. } \boldsymbol{F}_{\mathrm{RF}} \in \mathcal{F}^{(\mathrm{RF})}, \tag{14}$$

in other words, a unimodular low rank approximation. This problem has a similar structure as its single carrier counterpart in [8]. As a benchmark, the spatially sparse precoding via orthogonal matching pursuit (OMP) algorithm in [8] can be extended to the multi-carrier case. Nevertheless, in the following we propose a novel approach from a matrix decomposition point of view.

Algorithm 1 Hybrid Precoding via Sequential Low Rank Unimodular Approximation (SeLoRUA)

Initialize: set $\boldsymbol{F}_{\mathrm{res}}^{(1)} = \bar{\boldsymbol{F}}_{\mathrm{opt}}$ and a tolerance factor ϵ.
Main step:
for $n = 1$ to $N_{\mathrm{T}}^{(\mathrm{RF})}$ **do**
  set an arbitrary unimodular $\boldsymbol{f}_n^{(0)}$.
  repeat
    $\boldsymbol{f}_n^{(p)} = e^{j\angle(\boldsymbol{F}_{\mathrm{res}}^{(n)} \boldsymbol{F}_{\mathrm{res}}^{(n)\mathrm{H}} \boldsymbol{f}_n^{(p-1)})}$.
  until $\|\boldsymbol{f}_n^{(p)} - \boldsymbol{f}_n^{(p-1)}\|_2 \leq \epsilon$
  return $\boldsymbol{F}_{\mathrm{RF}} = [\boldsymbol{F}_{\mathrm{RF}} | \boldsymbol{f}_n^{(p)}]$,
  $\bar{\boldsymbol{F}}_{\mathrm{BB}} = [\bar{\boldsymbol{F}}_{\mathrm{BB}}^{\mathrm{T}} | \boldsymbol{F}_{\mathrm{res}}^{(n)\mathrm{H}} \boldsymbol{f}_n^{(p)} / M_{\mathrm{T}}]^{\mathrm{T}}$
  update $\boldsymbol{F}_{\mathrm{res}}^{(n+1)} = \boldsymbol{F}_{\mathrm{res}}^{(n)} - \boldsymbol{f}_n^{(p)} \boldsymbol{f}_n^{(p)\mathrm{H}} \boldsymbol{F}_{\mathrm{res}}^{(n)} / M_{\mathrm{T}}$
end for
Output: $\boldsymbol{F}_{\mathrm{RF}}$, $\boldsymbol{F}_{\mathrm{BB}} = \bar{\boldsymbol{F}}_{\mathrm{BB}} \sqrt{P_{\mathrm{T}}} / \|\boldsymbol{F}_{\mathrm{RF}} \bar{\boldsymbol{F}}_{\mathrm{BB}}\|_{\mathrm{F}}$.

A first try at (14) suggests to fix $\boldsymbol{F}_{\mathrm{RF}}$ and solve for $\bar{\boldsymbol{F}}_{\mathrm{BB}}$ in closed form using the Moore-Penrose pseudo inverse. This leads to a difficult minimization of a complicated function of $\boldsymbol{F}_{\mathrm{RF}}$ subject to constant modulus constraints. It does not seem to have a simple solution. However, it does suggest a different approach based on deflation, i.e., the procedure of constructing SVD sequentially [14]. Supposing that $\boldsymbol{F}_{\mathrm{RF}}$ and $\bar{\boldsymbol{F}}_{\mathrm{BB}}^{\mathrm{H}}$ are vectors (i.e., $N_{\mathrm{T}}^{(\mathrm{RF})} = 1$) and are denoted by $\boldsymbol{f}_1$ and $\boldsymbol{b}_1$, respectively, then this yields

$$\min_{\boldsymbol{f}_1, \boldsymbol{b}_1} \|\bar{\boldsymbol{F}}_{\mathrm{opt}} - \boldsymbol{f}_1 \boldsymbol{b}_1^{\mathrm{H}}\|_{\mathrm{F}}^2, \quad \text{s.t. } \boldsymbol{f}_1 \in \mathcal{F}^{(\mathrm{RF})}. \tag{15}$$

Solving for $\boldsymbol{b}_1$ yields

$$\boldsymbol{b}_1 = \bar{\boldsymbol{F}}_{\mathrm{opt}}^{\mathrm{H}} \cdot \boldsymbol{f}_1 / M_{\mathrm{T}}. \tag{16}$$

Inserting (16) into the objective function of (15), we get

$$\|\bar{\boldsymbol{F}}_{\mathrm{opt}} - \boldsymbol{f}_1 \boldsymbol{b}_1^{\mathrm{H}}\|_{\mathrm{F}}^2 = \|\bar{\boldsymbol{F}}_{\mathrm{opt}}\|_{\mathrm{F}}^2 - \boldsymbol{f}_1^{\mathrm{H}} \bar{\boldsymbol{F}}_{\mathrm{opt}} \bar{\boldsymbol{F}}_{\mathrm{opt}}^{\mathrm{H}} \boldsymbol{f}_1 / M_{\mathrm{T}}. \tag{17}$$

This leads to

$$\max_{\boldsymbol{f}_1} \ \boldsymbol{f}_1^{\mathrm{H}} \bar{\boldsymbol{F}}_{\mathrm{opt}} \bar{\boldsymbol{F}}_{\mathrm{opt}}^{\mathrm{H}} \boldsymbol{f}_1 / M_{\mathrm{T}} \quad \text{s.t.} \quad \boldsymbol{f}_1 \in \mathcal{F}^{(\mathrm{RF})}. \tag{18}$$

The resulting problem is expressed as

$$\max_{\boldsymbol{f}_1} \{\boldsymbol{f}_1^{\mathrm{H}} \boldsymbol{A} \boldsymbol{f}_1\}, \quad \text{s.t. } \boldsymbol{f}_1 \in \mathcal{F}^{(\mathrm{RF})}, \tag{19}$$

where $\boldsymbol{A} = \bar{\boldsymbol{F}}_{\mathrm{opt}} \bar{\boldsymbol{F}}_{\mathrm{opt}}^{\mathrm{H}}$. Problem (19) is in general non-convex and it is termed as the unimodular quadratic programming (UQP) problem in [15]. A local optimal solution to (19) can be obtained by a power method like iteration, i.e., in the p-th iteration we compute

$$\boldsymbol{f}_1^{(p)} = e^{j\angle(\boldsymbol{A} \cdot \boldsymbol{f}_1^{(p-1)})}, \tag{20}$$

where it is proven in [15] that this power method like iteration converges to a local optimum of (19).

Thus, we have reduced the vector case (or rank one case) of (14) to a standard UQP with an efficient power iteration like algorithmic solution. In the more general case when $\boldsymbol{F}_{\mathrm{RF}}$ and $\bar{\boldsymbol{F}}_{\mathrm{BB}}$ are matrices, we continue by deflation. Specifically, we can achieve a low rank unimodular approximation of $\bar{\boldsymbol{F}}_{\mathrm{opt}}$ by sequentially computing its rank-1 unimodular approximations $N_{\mathrm{T}}^{(\mathrm{RF})}$ times. Let $\boldsymbol{f}_n \in \mathbb{C}^{M_{\mathrm{T}}}$ and $\boldsymbol{b}_n \in \mathbb{C}^{N_{\mathrm{ss}} N_{\mathrm{FFT}}}$ denotes the n-th column of $\boldsymbol{F}_{\mathrm{RF}}$ and $\bar{\boldsymbol{F}}_{\mathrm{BB}}^{\mathrm{H}}$, respectively. To obtain $\boldsymbol{f}_n$ and $\boldsymbol{b}_n$, we calculate the best rank-1 unimodular approximation of the following matrix

$$\boldsymbol{F}_{\mathrm{res}}^{(n)} = \boldsymbol{F}_{\mathrm{res}}^{(n-1)} - \boldsymbol{f}_{n-1} \boldsymbol{b}_{n-1}^{\mathrm{H}} = \left(\boldsymbol{I}_{M_{\mathrm{T}}} - \frac{\boldsymbol{f}_{n-1} \boldsymbol{f}_{n-1}^{\mathrm{H}}}{M_{\mathrm{T}}}\right) \boldsymbol{F}_{\mathrm{res}}^{(n-1)}.$$

Finally, our proposed hybrid precoding via sequential low rank unimodular approximation (SeLoRUA) is summarized in Algorithm 1.

The computational complexity of the SeLoRUA algorithm is dominated by the power method like iteration. In each iteration, two consecutive matrix-vector multiplications are computed, which accounts for a complexity of $\mathcal{O}(M_{\mathrm{T}} N_{\mathrm{ss}} N_{\mathrm{FFT}})$. The SeLoRUA algorithm can also be directly applied at the receiver side to obtain $\boldsymbol{W}_{\mathrm{BB}} = [\boldsymbol{W}_{\mathrm{BB}}[1] \quad \cdots \quad \boldsymbol{W}_{\mathrm{BB}}[N_{\mathrm{FFT}}]]$ and $\boldsymbol{W}_{\mathrm{RF}}$. Note that there is no guaranteed orthogonality for the obtained columns of $\boldsymbol{F}_{\mathrm{RF}}$ via the SeLoRUA algorithm. Hence, this sequential rank-$N_{\mathrm{T}}^{(\mathrm{RF})}$ unimodular approximation is not equivalent to the best rank-$N_{\mathrm{T}}^{(\mathrm{RF})}$ unimodular approximation of $\bar{\boldsymbol{F}}_{\mathrm{opt}}$ in the LS sense. However, numerical results show that the SeLoRUA algorithm provides a sufficiently good approximation accuracy for our applications.

V. SIMULATION RESULTS

The proposed algorithms are evaluated using Monte-Carlo simulations. The maximum allowable power P_{T} is fixed to unity. The SNR is thus defined as $\mathrm{SNR} = 1/(N_{\mathrm{FFT}}\sigma_n^2)$. We set $N_{\mathrm{FFT}} = 128$. However, for computational simplicity only 8 out of 128 subcarriers are used for data transmission. Other than the proposed truncated HOSVD based algorithm (denoted as ”HOSVD”) in Section III and the sequential low rank unimodular approximation (denoted as ”SeLoRUA”) in Section IV, two benchmark algorithms are compared, i.e., the optimal unconstrained solution in (3) (denoted as ”Optimal Unconstrained”) and the sparse precoding based design in [8] extended to the multi-carrier scenario (denoted as ”MC-OMP”). The tolerance factor is $\epsilon = 10^{-4}$ when implementing the SeLoRUA algorithm and we use 7 quantization bits when implementing the MC-OMP algorithm [8]. All the simulation results are obtained by averaging over 1000 channel realizations.

The simulated channel is a mmWave channel with a geometric channel model of $L = 7$ scatterers. Each scatterer contributes to a single propagation path between the BS and the UE. The discrete CTF of the UE on the m-th subcarrier is modeled as [16]

$$\boldsymbol{H}[m] = \frac{1}{\sqrt{L}} \sum_{\ell=1}^{L} \alpha_\ell \boldsymbol{a}_{\mathrm{R},\ell}(\theta_{\mathrm{R},\ell}) \boldsymbol{a}_{\mathrm{T},\ell}^{\mathrm{H}}(\theta_{\mathrm{T},\ell}) e^{-j2\pi \frac{\ell \cdot m}{N_{\mathrm{FFT}}}}, \quad (21)$$

where α_ℓ is the random complex gain of the ℓ-th path, with zero-mean and $\mathbb{E}\{|\alpha_\ell|^2\} = 1$. The variables $\{\theta_{\mathrm{R},\ell}, \theta_{\mathrm{T},\ell}\} \in [0, 2\pi)$ denote the angle of arrival and the angle of departure of the ℓ-th path, respectively. Finally, $\boldsymbol{a}_{\mathrm{R},\ell}(\cdot)$ and $\boldsymbol{a}_{\mathrm{T},\ell}(\cdot)$ are the array steering vectors of the BS and the UEs, respectively. As in [7], a uniform linear array (ULA) geometry is used at both ends. The inter-element spacing of the ULA is equal to half of the wavelength.

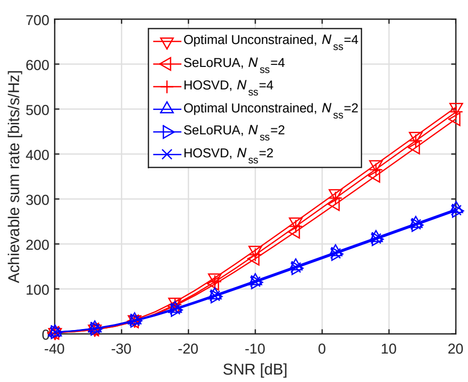

Fig. 1. Achievable sum rate vs. SNR for $M_{\mathrm{T}} = 256$, $M_{\mathrm{R}} = 64$, $N_{\mathrm{T}}^{(\mathrm{RF})} = N_{\mathrm{R}}^{(\mathrm{RF})} = 5$, if the constant modulus constraints are not taken into account.

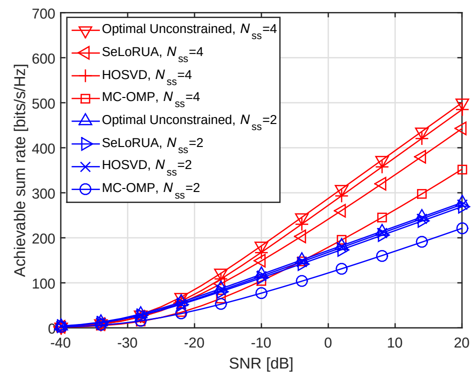

Fig. 2. Achievable sum rate vs. SNR for $M_{\mathrm{T}} = 256$, $M_{\mathrm{R}} = 64$, $N_{\mathrm{T}}^{(\mathrm{RF})} = N_{\mathrm{R}}^{(\mathrm{RF})} = 5$.

Fig. 1 and Fig. 2 compare the performance of the proposed algorithms without and with the constant modulus constraints, respectively. Fig. 1 illustrates that under the considered simulation settings the performance loss due to the coupling in the RF is negligible especially when the number of spatial streams is small. As can be seen from Fig. 2, both the truncated HOSVD based algorithm and the SeLoRUA algorithm suffer from the additional constant modulus constraints. But they all outperform the compressed sensing based algorithm. Compared to the optimal unconstrained solution, they provide almost the same multiplexing gain. The gap between the optimal unconstrained solution and the proposed methods reduces as N_{ss} decreases.

VI. CONCLUSION

We have studied the hybrid analog-digital precoding design for maximizing the sum rate of a single user multi-carrier massive MIMO system. Analog precoding is identical for all subcarriers and is realized using phase shifters. Thus the resulting optimization problem is non-convex. We have proposed two efficient algorithms to solve it. The first algorithm is based on a two-step consecutive low rank approximation, where in the first step the truncated HOSVD of the effective channel is used to compute the RF matrices. In the second step, the optimal digital precoder and decoder are provided by the truncated SVDs of the equivalent channels on a per-subcarrier basis. Thereby, there is no residual inter-stream interference, which makes this design more suitable for interference sensitive cases. The second algorithm approximates the optimal unconstrained solution by sequentially computing its best rank-1 unimodular approximation. The residual interference is non-zero after using the proposed SeLoRUA algorithm. However, the resulting approximation accuracy it provides is sufficient for our application. Simulation results show that both algorithms outperform an extension of the state of the art sparse precoding algorithm [8] to the multi-carrier case and achieve the same multiplexing gain as the optimal solution.

ACKNOWLEDGMENTS

The authors acknowledge the financial support by the Carl-Zeiss-Foundation (http://carl-zeiss-stiftung.de/).

VII. REFERENCES

[1] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, “Scaling up MIMO: opportunities and challenges with very large arrays,” *Signal Processing Magazine*, vol. 30, no. 1, 2013.

[2] W. Roh, J.-Y. Seol, J. Park, B. Lee, J. Lee, Y. Kim, J. Cho, K. Cheun, and F. Aryanfar, "Millimeter-wave beamforming as an enabling technology for 5G cellular communications: Theoretical feasibility and prototype results," *IEEE Communications Magazine*, pp. 106–113, Feb. 2014.

[3] A. L. Swindlehurst, E. Ayanoglu, P. Heydari, and F. Capolino, "Millimeter-wave massive MIMO: The next wireless revolution?," *IEEE Communications Magazine*, pp. 56–62, Sept. 2014.

[4] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter-wave cellular wireless networks: Potentials and challenges," *Proceedings of the IEEE*, vol. 102, pp. 366–385, Mar. 2014.

[5] R. Mendez-Rial, C. Rusu, A. Alkhateeb, N. G. Prelcic, and R. W. Heath, Jr., "Channel estimation and hybrid combining for mmWave: phase shifters or switches," in *Proc. of Information Theory and Applications Workshop (ITA 2015)*, UCSD, USA, Feb. 2015.

[6] V. Venkateswaran and A. van der Veen, "Analog beamforming in MIMO communications with phase shift networks and online channel estimation," *IEEE Transactions on Signal Processing*, vol. 58, pp. 4131–4143, Aug. 2010.

[7] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, Jr., "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 13, pp. 1499–1513, Mar. 2014.

[8] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, Jr., "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, pp. 831–846, Oct. 2014.

[9] A. Alkhateeb, G. Leus, and R. W. Heath, Jr., "Achievable rates of multi-user millimeter wave systems with hybrid precoding," in *Proc. IEEE International Conference on Communications (ICC 2015)*, London, UK, June 2015.

[10] 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," *submitted to IEEE Transactions on Wireless Communications, available at http://http://arxiv.org/abs/1410.2609*, pp. 1–30, 2015.

[11] M. Haardt, F. Roemer, and G. Del Galdo, "Higher-order SVD based subspace estimation to improve the parameter estimation accuracy in multi-dimensional harmonic retrieval problems," *IEEE Transactions on Signal Processing*, vol. 56, pp. 3198–3213, July 2008.

[12] L. de Lathauwer, B. de Moor, and J. Vanderwalle, "A multilinear singular value decomposition," *SIAM J. Matrix Anal. Appl.*, vol. 21, no. 4, 2000.

[13] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*, Cambridge University Press, 2003.

[14] G. H. Golub and C. F. Van Loan, *Matrix Computations*, John Hopkins University Press, 1996.

[15] M. Soltanalian and P. Stoica, "Designing unimodular codes via quadratic optimization," *IEEE Transactions on Signal Processing*, vol. 62, pp. 1221–1234, Mar. 2014.

[16] W. U. Bajwa, J. Haupt, A. M. Sayeed, and R. Nowak, "Compressed channel sensing: A new approach to estimating sparse multipath channels," *Proceedings of the IEEE*, vol. 98, pp. 1058–1076, June 2010.