Dictionary-free Hybrid Precoders and Combiners for mmWave MIMO Systems

Roi Méndez-Rial[†], Cristian Rusu[†], Nuria González-Prelcic[†] and Robert W. Heath Jr.[‡]
[†] Universidade de Vigo, Vigo, Spain, Email: {roimr,crusu,nuria}@gts.uvigo.es
[‡] The University of Texas at Austin, Austin, TX 78712, Email: {rheath}@utexas.edu

Abstract—The high cost and power consumption of the radiofrequency chain and data converters at mmWave frequencies introduce hardware limitations into the design of MIMO precoders and combiners. MmWave hybrid precoding overcomes this limitation by dividing the spatial signal processing between the radio frequency and baseband domains. Analog networks of phase shifters have been proposed to implement the radio frequency precoders, since they achieve a good compromise between complexity and performance. In this paper, we propose a low complexity hybrid precoding design for the architecture based on phase shifters. The new method is a greedy algorithm based on the orthogonal matching pursuit algorithm, but replacing the costly correlation operations over a dictionary with the element-wise normalization of the first singular vector of the residual. The main advantage is that the design avoids any assumption on the antenna array geometry. Additionally, numerical results show the superiority of the proposed method in terms of achievable spectral efficiency over other previous solutions.

I. INTRODUCTION

Millimeter wave (mmWave) multiple-input multiple-output (MIMO) systems overcome future rising spectrum needs by enabling gigabit per second rates of communication for indoor and outdoor wireless systems [1]–[4]. The decrease of wavelength at mmWave makes possible large antenna arrays at receiver and transmitter. The cost and power consumption of the radio frequency (RF) components in the mmWave range, however, make it challenging to use one complete RF chain and one analog-to-digital (ADC) converter per antenna.

An immediate solution to overcome the limitation on the number of complete RF chains is to perform beamforming in analog using variable phase shifters [5]. This is the defacto approach in mmWave indoor communications in IEEE 802.11ad [6], and it has been introduced as an optional functionality in IEEE 802.15.3c [7]. The network of phase shifters steers the beam along the dominant propagation path of the channel using different strategies [8], [9], and supporting only single stream MIMO communications. Moreover, the availability of only quantized phase shifters limits the performance of the analog beamforming solutions. A different analog beamforming concept, beamspace MIMO [10], is based on a high-resolution discrete lens array. This avoids the limitations of the quantized phase shifters, but does not provide uniform performance across a broad range of angles.

Another approach to deal with a limited number of RF chains is to design hybrid precoders/combiners, which separate the spatial signal processing into an analog processing network and the baseband (BB) domain. This approach supports multi-stream communication. In [11], the precoder is designed using simple direct formulas approaching the waterfilling solution. No assumption on the array geometry is needed. The work considers dynamic power allocation with the number of data streams being dynamically optimized. Additionally, with a small number of data streams and large number of antennas, the power allocation benefits achieved are not significant versus low complexity equal power allocation solutions.

A hybrid solution especially designed for mmWave channels was recently proposed in [12], [13]. The method exploits the limited scattering nature of the mmWave channel and the presence of large antenna arrays. The design of the precoders is formulated as a sparse optimization problem with hardware constraints. It resembles the problem of sparse signal recovery via multiple measurement vector problem. The main limitation of this work is that it assumes known array geometries for both transmitter and receiver. Additionally, solving the sparse optimization problem still results in high complexity. The works in [11]–[13] assume perfect channel state information at the receiver. To overcome the high complexity limitation another technique for the precoder/combiner design has been proposed in [14]; a significant complexity reduction can be achieved, but the solution still depends on the array geometry.

In this paper, we propose a low-complexity hybrid precoding design using a fixed number of data streams and equal power allocation. The main advantage is that it avoids any assumption on the array geometry or channel structure. The method is a greedy algorithm based on orthogonal matching pursuit steps, as in [13], but replacing the correlation operations over a dictionary with the element-wise normalization of the first singular vector of the residual. Only a number of iterations equal to the difference between available RF chains and data streams is needed to complete the design. This results in a great reduction of the computational complexity with respect to the method in [13]. Numerical results show the superiority of the proposed solution in terms of achievable spectral efficiency.

II. SYSTEM MODEL

A single user mmWave MIMO system is shown in Fig-1. The transmitting BS is equipped with N_t antennas and L_t RF chains while the receiving MS with N_r antennas and L_r RF chains. N_s data streams are transmitted from the BS to the MS assuming $N_s \leq L_t \leq N_t$ and $N_s \leq L_r \leq N_r$.

The transmitter applies a hybrid precoder $\mathbf{F}_T$ to the symbol vector $\mathbf{s} \in \mathbb{C}^{N_s \times 1}$ with $\mathbb{E}[\mathbf{ss}^*] = \mathbf{I}$. The discrete-time transmitted signal is given by $\mathbf{x} = \mathbf{F}_T\mathbf{s}$. The signal is transmitted through a narrowband flat channel assuming
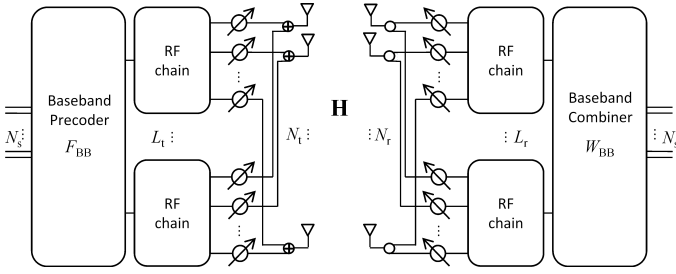
Fig. 1: Block diagram of the transmitter-receiver single user mmWave system architecture.

perfect synchronization to give the received signal

$$\mathbf{r} = \sqrt{\rho}\mathbf{H}\mathbf{F}_\mathrm{T}\mathbf{s} + \mathbf{n}, \tag{1}$$

where $\mathbf{H} \in \mathbb{C}^{N_\mathrm{r} \times N_\mathrm{t}}$ is the channel matrix such that $\mathbb{E}\left[\|\mathbf{H}\|_F^2\right] = N_\mathrm{t}N_\mathrm{r}$, ρ represents the average transmitted power per symbol and $\mathbf{n} \in \mathbb{C}^{N_\mathrm{r} \times 1}$ is the noise vector with $\mathcal{CN}(0, \sigma_n^2)$ entries. The MS applies a hybrid combiner $\mathbf{W}_\mathrm{T}$ to the signal. Assuming flat-fading and perfect synchronization, the discrete-time model for a single symbol period is

$$\mathbf{y} = \sqrt{\rho}\mathbf{W}_\mathrm{T}^*\mathbf{H}\mathbf{F}_\mathrm{T}\mathbf{s} + \mathbf{W}_\mathrm{T}^*\mathbf{n}. \tag{2}$$

The hybrid precoder $\mathbf{F}_\mathrm{T} = \mathbf{F}_\mathrm{RF}\mathbf{F}_\mathrm{BB}$ is composed of an RF precoder $\mathbf{F}_\mathrm{RF} \in \mathbb{C}^{N_\mathrm{t} \times L_\mathrm{t}}$ and a baseband precoder $\mathbf{F}_\mathrm{BB} \in \mathbb{C}^{L_\mathrm{t} \times N_\mathrm{s}}$. Equivalently, the hybrid combiner $\mathbf{W}_\mathrm{T} = \mathbf{W}_\mathrm{RF}\mathbf{F}_\mathrm{BB}$ is composed of an RF combiner $\mathbf{W}_\mathrm{RF} \in \mathbb{C}^{N_\mathrm{r} \times L_\mathrm{r}}$ and a baseband combiner $\mathbf{W}_\mathrm{BB} \in \mathbb{C}^{L_\mathrm{r} \times N_\mathrm{s}}$. The RF precoder and combiner are implemented in the analog domain. Therefore, the precoding and combining matrices $\mathbf{F}_\mathrm{RF}$ and $\mathbf{W}_\mathrm{RF}$ are subject to the specific hardware constraints. The RF precoder/combiner implemented with a network of variable phase shifters impose the constraint of unit norm entries in $\mathbf{F}_\mathrm{RF}, \mathbf{W}_\mathrm{RF}$.

III. PROBLEM FORMULATION

Assuming perfect channel state information at the receiver, we seek to design hybrid mmWave precoders and combiners, $\mathbf{F}_\mathrm{T} = \mathbf{F}_\mathrm{RF}\mathbf{F}_\mathrm{BB}$ and $\mathbf{W}_\mathrm{T} = \mathbf{W}_\mathrm{RF}\mathbf{W}_\mathrm{BB}$, to maximize the spectral efficiency [15]

$$\log_2 \left| \mathbf{I}_{N_\mathrm{s}} + \frac{\rho}{N_\mathrm{s}}\mathbf{R}_n^{-1}\mathbf{W}_\mathrm{T}^*\mathbf{H}\mathbf{F}_\mathrm{T}\mathbf{F}_\mathrm{T}^*\mathbf{H}^*\mathbf{W}_\mathrm{T} \right|, \tag{3}$$

where $\mathbf{R}_n$ is the noise covariance matrix after combining. We consider a total transmit power constraint given by $\|\mathbf{F}_\mathrm{T}\|_F^2 = N_\mathrm{s}$, and a fixed number of data streams $N_\mathrm{s} \leq \min(L_\mathrm{t}, L_\mathrm{r})$. We are interested in designs with equal power allocation and low computational complexity.

Given the singular value decomposition of the channel $\mathbf{H} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^*$, the optimal unconstrained hybrid precoder and combiner that achieve the capacity are

$$\mathbf{F}_\mathrm{opt} = \mathbf{V}_L\boldsymbol{\Gamma}\mathbf{A} \tag{4}$$
$$\mathbf{W}_\mathrm{opt} = \mathbf{B}\mathbf{U}_L. \tag{5}$$

Here $\mathbf{A}$ can be any $L \times L$ unitary matrix with $L = \min(L_\mathrm{t}, L_\mathrm{r})$, $\mathbf{B}$ any full-rank matrix of the same size and $\boldsymbol{\Gamma}$ is the diagonal matrix with the power allocation weights given by waterfilling [16]. Denote with $\mathbf{V}_L$ ($\mathbf{U}_L$) the matrix constructed by selecting

the first L columns of $\mathbf{V}$ ($\mathbf{U}$). Therefore, our goal can be formulated as the design of practical precoders and combiners at RF and BB that approximate these optimal solutions.

In this context, a first design strategy with equal power allocation and fixed N_s was proposed in [13]. The approach is to first abstract the receiver operation, assuming a perfect maximum likelihood receiver, and focus on the precoder design. Therefore, they jointly seek for $\mathbf{F}_\mathrm{RF}$ and $\mathbf{F}_\mathrm{BB}$ that maximize the mutual information

$$\arg\max_{\mathbf{F}_\mathrm{RF},\mathbf{F}_\mathrm{BB}} \log_2 \left| \mathbf{I} + \frac{\rho}{N_\mathrm{s}\sigma_n^2}\mathbf{H}\mathbf{F}_\mathrm{RF}\mathbf{F}_\mathrm{BB}\mathbf{F}_\mathrm{BB}^*\mathbf{F}_\mathrm{RF}^*\mathbf{H}^* \right|$$
$$\text{s.t.} \quad \mathbf{F}_\mathrm{RF} \in \mathcal{F}_\mathrm{RF}$$
$$\|\mathbf{F}_\mathrm{RF}\mathbf{F}_\mathrm{BB}\|_F^2 = N_\mathrm{s}, \tag{6}$$

where $\mathcal{F}_\mathrm{RF}$ stands for the set of feasible RF implementable precoders with phase shifters, i.e. precoders with constant norm entries. Since (6) does not have a general solution when constraints are imposed on $\mathbf{F}_\mathrm{RF}$, some approximations and assumptions are used. Further constraining the set of feasible RF precoders $\mathcal{F}_\mathrm{RF}$ to the set of array response vectors $\mathbf{A}_\mathrm{t} = [\mathbf{a}_t(\phi_1)...\mathbf{a}_t(\phi_N)]$, with N the angular resolution,

$$\arg\min_{\mathbf{A}_\mathrm{t},\mathbf{F}_\mathrm{BB}} \|\mathbf{F}_\mathrm{opt} - \mathbf{A}_\mathrm{t}\mathbf{F}_\mathrm{BB}\|_F$$
$$\text{s.t.} \quad \|\text{diag}(\mathbf{F}_\mathrm{BB}\mathbf{F}_\mathrm{BB}^*)\|_0 = L_\mathrm{t}$$
$$\|\mathbf{A}_\mathrm{t}\mathbf{F}_\mathrm{BB}\|_F^2 = N_\mathrm{s}, \tag{7}$$

which can be solved with a variant of simultaneous orthogonal matching pursuit (SOMP). An equivalent problem is solved to design the combiner.

This solution is specifically designed for equal power allocation and fixed number of data streams. When $N_\mathrm{s} < L_\mathrm{t}$, it benefits from the extra RF chains and seeks for the best linear combination that approximates the optimal precoder. The main drawback is the use of a dictionary of array steering vectors and the costly correlation operations. To build the dictionary, some assumptions on the channel structure and array geometry have to be made. The accuracy of the approximation depends on the resolution considered to design the dictionary. Additionally, by further constraining the set of feasible RF precoders to the set of array steering vectors, the method does not exploit all the available degrees of freedom in the RF domain.

Our final goal will be finding an alternative dictionary-free solution to (6) which does not depend on the array geometry and exhibiting at the same time a low computational complexity.

IV. HOW MANY RF CHAINS ARE NEEDED?

An interesting question before designing the precoder/combiner is how many RF chains are needed to achieve the performance of the optimum unconstrained precoder with a hybrid design with phase shifters constraints for a given receiving antenna array size. The case where $N_\mathrm{s} = 1$ was analyzed in [11]. Any vector $\mathbf{f}_\mathrm{opt} \in \mathbb{C}^{N_\mathrm{t} \times 1}$ can be written as

$$\mathbf{f}_\mathrm{opt} = \mathbf{F}_\mathrm{RF}\mathbf{f}_\mathrm{BB}, \tag{8}$$

with $\mathbf{F}_\mathrm{RF} \in \mathcal{F}_\mathrm{RF}^{N_\mathrm{t} \times L_\mathrm{t}}$ and $\mathbf{f}_\mathrm{BB} \in \mathbb{C}^{L_\mathrm{t} \times 1}$ if and only if $L_\mathrm{t} \geq 2$. See [11] for the complete proof. The solutions are not unique, a possible design with direct formulas is provided in [11].

For the general case with $N_s \geq 1$ and $\mathbf{F}_{\text{opt}} \in \mathbb{C}^{N_t \times N_s}$, the aim is to find the minimum number of L_t for what there exist $\mathbf{F}_{\text{RF}} \in \mathcal{F}_{\text{RF}}^{N_t \times L_t}$ and $\mathbf{F}_{\text{BB}} \in \mathbb{C}^{L_t \times N_s}$ providing

$$\mathbf{F}_{\text{opt}} = \mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}. \tag{9}$$

A sufficient condition is that $L_t \geq 2N_s$. This result is a direct consequence of the single stream result. Since each column of $\mathbf{F}_{\text{opt}}$ can be decomposed as a linear combination of two columns from the RF precoder, we can write

$$\begin{aligned} \mathbf{F}_{\text{opt}} &= [\mathbf{f}_{\text{opt}}^1 \ldots \mathbf{f}_{\text{opt}}^{N_s}] \\ &= [\mathbf{F}_{\text{RF}}^1 \ldots \mathbf{F}_{\text{RF}}^{N_s}] \begin{bmatrix} \mathbf{f}_{\text{BB}}^1 & & 0 \\ & \ddots & \\ 0 & & \mathbf{f}_{\text{BB}}^{N_s} \end{bmatrix}, \end{aligned} \tag{10}$$

where $\mathbf{F}_{\text{RF}}^i \in \mathcal{F}_{\text{RF}}^{N_t \times 2}$ and $\mathbf{f}_{\text{BB}}^i \in \mathbb{C}^{2 \times 1}$ for $i = 1 \ldots N_s$. Equation (10) is equivalent to $\mathbf{f}_{\text{opt}}^i = \mathbf{F}_{\text{RF}}^i \mathbf{f}_{\text{BB}}^i$ for $i = 1 \ldots N_s$, for which we know that there exists a solution. The same formulas as for the single data stream case can be applied to design the hybrid precoder.

Notice that for the case $L_t \geq 2N_s$, the solution to the Frobenius norm based optimization problem is exactly the optimum solution that maximizes the spectral efficiency. We focus on designing near optimal hybrid precoders when the available number of RF chains is $L_t < 2N_s$.

V. DECOUPLED HYBRID PRECODING

The first simplest idea to reduce the complexity of the joint analog/baseband optimization problem is to decouple problem (3) into the analog and digital domains. The analog+baseband decoupling strategy is stated as follows:

- Analog: Find the optimum $\mathbf{F}_{\text{RF}}$ and $\mathbf{W}_{\text{RF}}$ that maximize the spectral efficiency of the system

$$\hat{\mathbf{y}} = \mathbf{W}_{\text{RF}}^* \mathbf{H} \mathbf{F}_{\text{RF}} \hat{\mathbf{s}} + \mathbf{n}. \tag{11}$$

- Baseband: Fixed $\mathbf{F}_{\text{RF}}$ and $\mathbf{W}_{\text{RF}}$, find $\mathbf{F}_{\text{BB}}$ and $\mathbf{W}_{\text{BB}}$ that maximizes the spectral efficiency of the equivalent system

$$\mathbf{y} = \mathbf{W}_{\text{BB}}^* \hat{\mathbf{H}} \mathbf{F}_{\text{BB}} \mathbf{s} + \mathbf{W}_{\text{RF}}^* \mathbf{n}. \tag{12}$$

The equivalent channel is $\hat{\mathbf{H}} = \mathbf{K}_w^{-1/2} \mathbf{W}_{\text{RF}}^* \mathbf{H} \mathbf{F}_{\text{RF}}$, with $\mathbf{K}_w = (\mathbf{W}_{\text{RF}}^* \mathbf{W}_{\text{RF}})$ the noise covariance matrix.

Finding the optimum feasible RF precoder/combiner that maximize the spectral efficiency of (11),

$$\log_2 \left| \mathbf{I}_{L_r} + \frac{\rho}{L_r} \mathbf{R}_n^{-1} \mathbf{W}_{\text{RF}}^* \mathbf{H} \mathbf{F}_{\text{RF}} \mathbf{F}_{\text{RF}}^* \mathbf{H}^* \mathbf{W}_{\text{RF}} \right|,$$

requires solving an intractable optimization problem. Following the discussion in [13], near optimal RF precoders can be found by minimizing the Frobenius norm with respect to the optimum unconstrained solution. Given the singular value decomposition (SVD) of the channel, $\mathbf{H} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^*$, the optimum precoder, $\mathbf{F}_{\text{RF}}^{\text{opt}} = \mathbf{V}_{L_t}$, and combiner, $\mathbf{W}_{\text{opt}}^{\text{RF}} = \mathbf{U}_{L_r}$, are given by the first L_t, L_r singular vectors of $\mathbf{U}$ and $\mathbf{V}$. The idea is to solve the following optimization problem

$$\begin{aligned} \underset{\mathbf{F}_{\text{RF}}}{\arg\min} & \|\mathbf{F}_{\text{RF}}^{\text{opt}} - \mathbf{F}_{\text{RF}}\|_F \\ \text{s.t.} \quad & \mathbf{F}_{\text{RF}} \in \mathcal{F}_{\text{RF}}. \end{aligned} \tag{13}$$

This problem is easier to solve and has analytical solution. In fact, it is equivalent to minimizing the square error of each entry $\|[\mathbf{F}_{\text{RF}}^{\text{opt}}]_{i,k} - [\mathbf{F}_{\text{RF}}]_{i,k}\|_2^2$. Since $\left[\mathbf{F}_{\text{RF}}^{\text{opt}}\right]_{i,k} = \beta_{i,k} e^{j\psi_{i,k}}$, the RF solution to (13) is given by the simplest selection $[\mathbf{F}_{\text{RF}}]_{i,k} = e^{j\phi_{i,k}}$ with $\phi_{i,k} = \psi_{i,k}$. $\mathbf{F}_{\text{RF}}$ can be computed efficiently as the element-wise normalization of $\mathbf{F}_{\text{RF}} = \mathbf{F}_{\text{RF}}^{\text{opt}} \oslash |\mathbf{F}_{\text{RF}}^{\text{opt}}|$. Equivalently, $\mathbf{W}_{\text{RF}} = \mathbf{W}_{\text{RF}}^{\text{opt}} \oslash |\mathbf{W}_{\text{RF}}^{\text{opt}}|$.

The solution to the baseband problem is known. The unitary optimal precoder/combiner that maximize the mutual information is given by the singular value decomposition of the equivalent channel $\hat{\mathbf{H}} = \hat{\mathbf{U}}\hat{\boldsymbol{\Sigma}}\hat{\mathbf{V}}^*$. The digital precoder and combiner are $\mathbf{F}_{\text{BB}} = \hat{\mathbf{V}}_{N_s}$ and $\mathbf{W}_{\text{BB}}^* = \hat{\mathbf{U}}_{N_s}\mathbf{K}_w^{-1/2}$, where $\hat{\mathbf{V}}_{N_s}$ and $\hat{\mathbf{U}}_{N_s}$ are the first N_s eigenvectors of $\hat{\mathbf{U}}$ and $\hat{\mathbf{V}}$.

VI. GREEDY HYBRID PRECODING

The second proposed algorithm is a greedy method that finds a near optimal solution to the problem of minimizing $\|\mathbf{F}_{\text{opt}} - \mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}\|_F$ with the hardware constraints. The algorithm starts by initializing the first N_s columns of $\mathbf{F}_{\text{RF}}$ with the element-wise normalization of $\mathbf{F}_{\text{opt}}$. When $L_t = N_s$, this selection minimizes the Frobenius norm $\|\mathbf{F}_{\text{opt}} - \mathbf{F}_{\text{RF}}\|_F$. Then, $\mathbf{F}_{\text{BB}}$ is computed with regular least squares $\mathbf{F}_{\text{BB}} = \mathbf{F}_{\text{RF}} \setminus \mathbf{F}_{\text{opt}}$. When $N_s \leq L_t$ the algorithm follows a greedy strategy to complete the $N_t \times L_t$ $\mathbf{F}_{\text{RF}}$ after the initialization. In each step, the objective is to add a column to $\mathbf{F}_{\text{RF}}$ which leads to the highest reduction of the residual $\mathbf{R}^{(k)} = \mathbf{F}_{\text{opt}} - \mathbf{F}_{\text{RF}}^{(k)}\mathbf{F}_{\text{BB}}^{(k)}$. In $L_t - N_s$ steps the process is completed. After the initialization, the algorithms basically follows the steps of orthogonal matching pursuit, replacing the costly correlation operation by the computation of the main singular vector of the residual followed by its element-wise normalization. The Greedy Hybrid Precoding (GHP) is described in detail in Algorithm 1.

Algorithm 1 – Greedy Hybrid Precoding (GHP) .

Initialization: Perform the singular value decomposition of the channel $\mathbf{H} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^*$ and build the optimum unconstrained precoder $\mathbf{F}_{\text{opt}}$ using the first N_s columns of $\mathbf{V}$.
Main steps:

1) Initialize $\mathbf{F}_{\text{RF}}^{(0)}$ with the element-wise normalization $\mathbf{F}_{\text{RF}}^{(0)} = \mathbf{F}_{\text{opt}} \oslash |\mathbf{F}_{\text{opt}}|$.
2) For $k = 0 \ldots L_t - N_s - 1$:
   a) Update $\mathbf{F}_{\text{BB}}^{(k)} = \mathbf{F}_{\text{RF}}^{(k)} \setminus \mathbf{F}_{\text{opt}}$.
   b) Compute the residual $\mathbf{R}^{(k)} = \mathbf{F}_{\text{opt}} - \mathbf{F}_{\text{RF}}^{(k)}\mathbf{F}_{\text{BB}}^{(k)}$.
   c) Compute the first singular vector $\mathbf{u}_1$ of the singular value decomposition $\mathbf{R}^{(k)} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^*$.
   d) Append the element-wise normalization of $\mathbf{u}_1$ as the new unital column $\mathbf{F}_{\text{RF}}^{(k+1)} = \begin{bmatrix} \mathbf{F}_{\text{RF}}^{(k)} & \mathbf{u}_1 \oslash |\mathbf{u}_1| \end{bmatrix}$.

3) Final update $\mathbf{F}_{\text{BB}} = \mathbf{F}_{\text{RF}} \setminus \mathbf{F}_{\text{opt}}$.
4) Normalization $\mathbf{F}_{\text{BB}} = \sqrt{N_s} \frac{\mathbf{F}_{\text{BB}}}{\|\mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}\|_F}$

Regarding the computational complexity of GHP, the main differences with respect to an SOMP based algorithm as [13], are the steps involved in the computation of $\mathbf{F}_{\text{RF}}$, steps (c) and (d) of GHP. Step (c) computes the main singular vector of the current residual. The main singular vector can be well approximated by a few, in the order of N_s, iterations of the power method. Step (d) is just an element-wise normalization

stage that enforces the unit magnitude constraint. The order of complexity of steps (c) and (d) are $\mathcal{O}(N_t N_s^2)$ and $\mathcal{O}(N_t)$ respectively. These steps substitute the correlation operation over a dictionary and the selection of the atom that has the largest total correlation with the current residual. Computing the correlation takes $\mathcal{O}(N_t^2 N_s)$ operations for a resolution $\mathcal{O}(N_t)$, while the atom selection step requires $\mathcal{O}(N_t^2)$. The previous analysis holds for one iteration of both algorithms. Notice that for $N_t \gg N_s$, GHP leads to a great complexity reduction, since the number of operations depends only linearly on N_t. Additionally, GHP will run for $L_t - N_s$ iterations, less than the L_t iterations of SOMP. Overall, we do expect GHP to perform faster than SOMP in any circumstances. Furthermore, solving least squares problems with $\mathbf{F}_{RF}$ that increases in size, as in step (a), can be done efficiently by avoiding a full Cholesky factorization of $\mathbf{F}_{RF}^* \mathbf{F}_{RF}$ at each step and just performing an update of the factorization computed in the previous iteration.

Table I summarizes for each method the computational complexity of each step. The first singular value decomposition of the channel and the final normalization step are not included since they are common to all the strategies. Notice that the decoupled strategy provides the hybrid precoder and combiner with only one run of the algorithm, while GHP and SOMP have to be applied twice, one for each.

TABLE I: Computational complexity

Decoupled	
Operation	Complexity
$\mathbf{F}_{RF} = \mathbf{V}_{L_t} \oslash \|\mathbf{V}_{L_t}\|$	$\mathcal{O}(N_t L_t)$
$\mathbf{W}_{RF} = \mathbf{W}_{RF} \oslash \|\mathbf{W}_{RF}\|$	$\mathcal{O}(N_t L_t)$
$\mathbf{K}_w = (\mathbf{W}_{RF}^* \mathbf{W}_{RF})$	$\mathcal{O}(N_t L_t^2)$
$\mathbf{K}_w^{-1/2} \hat{\mathbf{H}} = \mathbf{W}_{RF}^* \mathbf{H} \mathbf{F}_{RF}$	$\mathcal{O}(N_t^2 L_t)$
$\text{SVD}(\hat{\mathbf{H}})$	$\mathcal{O}(L_t^3)$
Overall	$\mathcal{O}(N_t^2 L_t)$
GHP	
Operation	Complexity
$\mathbf{F}_{RF} = \mathbf{V}_{L_t} \oslash \|\mathbf{V}_{L_t}\|$	$\mathcal{O}(N_t L_t)$
$(L_t - N_s) \times \mathbf{F}_{BB} = \mathbf{F}_{RF} \setminus \mathbf{F}_{opt}$	$\mathcal{O}((L_t - N_s) N_t L_t^2)$
$(L_t - N_s) \times \mathbf{R} = \mathbf{F}_{opt} - \mathbf{F}_{RF} \mathbf{F}_{BB}$	$\mathcal{O}((L_t - N_s) N_t L_t N_s)$
$(L_t - N_s) \times$ Rank 1 decomp. of $\mathbf{R}$	$\mathcal{O}((L_t - N_s) N_t N_s^2)$
$(L_t - N_s) \times \mathbf{u}_1 \oslash \|\mathbf{u}_1\|$	$\mathcal{O}((L_t - N_s) N_t)$
$\mathbf{F}_{BB} = \mathbf{F}_{RF} \setminus \mathbf{F}_{opt}$	$\mathcal{O}(N_t L_t^2)$
Overall	$\mathcal{O}(N_t L_t^2 N_s)$
SOMP	
Operation	Complexity
$(L_t) \times$ Correlation $\mathbf{A}^* \mathbf{R}$	$\mathcal{O}(N_t^2 L_t N_s)$
$(L_t) \times$ Atom selection	$\mathcal{O}(N_t^2 L_t)$
$(L_t) \times \mathbf{F}_{BB} = \mathbf{F}_{RF} \setminus \mathbf{F}_{opt}$	$\mathcal{O}(N_t L_t^3)$
$(L_t) \times \mathbf{R} = \mathbf{F}_{opt} - \mathbf{F}_{RF} \mathbf{F}_{BB}$	$\mathcal{O}(N_t L_t^2 N_s)$
$(L_t) \times \mathbf{R} = \mathbf{R}/\|\mathbf{R}\|_F$	$\mathcal{O}(N_t L_t)$
Overall	$\mathcal{O}(N_t^2 L_t N_s)$

VII. SIMULATION

We consider the narrow-band clustered channel model in [13] with $N_{cl} = 4$ clusters and $N_{ray} = 8$ propagation paths per cluster. We assume all clusters are of equal power satisfying the normalization constraint $\mathbb{E}[\|\mathbf{H}\|_F^2] = N_t N_r$. The angles of departure and arrival are normal randomly distributed with mean cluster angle uniformly randomly distributed in $[0, 2\pi]$. The angle spread is set to 7.5. The same total power constraint
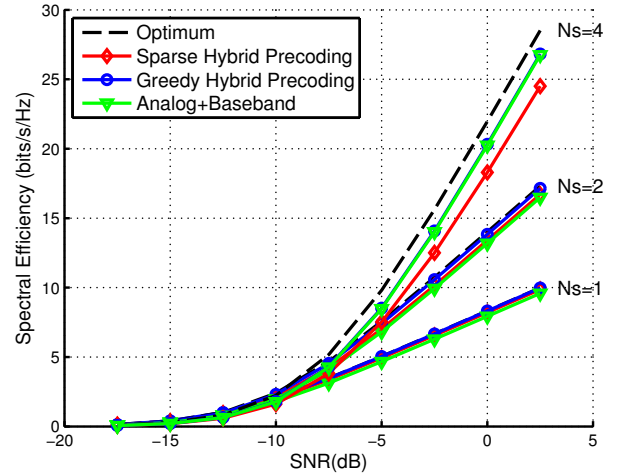

Fig. 2: Achievable spectral efficiency. ULA system $N_t = N_r = 32$ antennas and $L_t = L_r = 4$ RF chains. $N_s \in \{1, 2, 4\}$ data streams are considered.

is fixed for all precoders with equal power allocation per stream and SNR $= \frac{\rho}{\sigma_n^2}$.

Fig. 2 shows the spectral efficiency achieved by the proposed hybrid precoders GHP and analog+digital, together with the sparse hybrid precoder [13] and the optimum unconstrained solution given by the SVD of the channel for different SNR values. Both transmitter and receiver are assumed to have ULAs with $N_t = N_r = 32$ antennas and $L_t = L_r = 4$ RF chains with which they transmit $N_s \in \{1, 2, 4\}$ data streams. All the hybrid precoders achieve spectral efficiencies close to those achieved by the optimum unconstrained solution, within a small gap increasing for higher N_s. The GHP overcomes the analog+digital and the sparse hybrid precoder for any number of data streams. The analog+digital nearly overlaps with GHP when the number of streams equals the number of RF chains and performs slightly worse for $N_s = 1$ or 2.

Fig. 3 shows the spectral efficiency achieved by the hybrid precoders in terms of the number of RF chains. The same ULA system with $N_t = N_r = 32$ is considered with equal number of RF chains in transmission and reception ($L_t = L_r$) varying from 1 to 10. The number of streams equals the number of RF chains $N_s = L_t$. The SNR is fixed to 0 dB. We see that the GHP and the analog+digital precoders nearly overlap, both with a non-negligible improvement with respect to the sparse hybrid precoder. The gap between the spectral efficiency achieved by the hybrid precoders and the unconstrained optimal solution increases with the number of RF chains. To explore the performance when the number of RF chains is greater than the number of data streams, Fig.-4 plots the spectral efficiency for the same set up with the difference that now the number of data streams is fixed $N_s \in \{1, 2, 4\}$. We see that there is a slight improvement in the achievable spectral efficiency when the number of RF chains increases over the number of data streams with GHP and the sparse hybrid precoder. It means that the singular vectors of the channel are well approximated as a linear combination of unital vectors with GHP and as a linear combination of steering vectors with the sparse hybrid precoder. This gain is more noticeable for
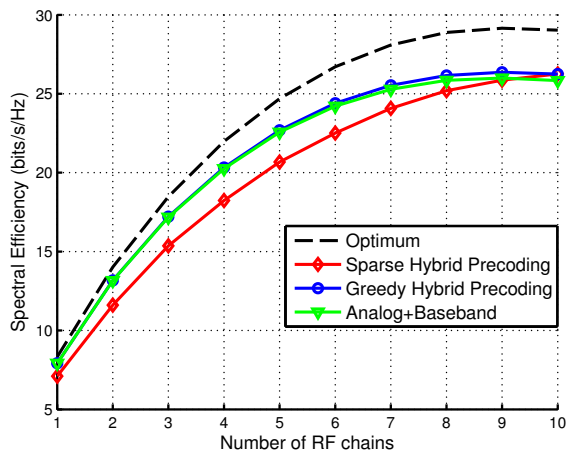
Fig. 3: Spectral Efficiency as a function of the number of RF chains. ULA system $N_t = N_r = 32$ antennas and $L_t = L_r = 4$ RF chains. The number of data streams equal the number of RF chains $N_s = L_t$.
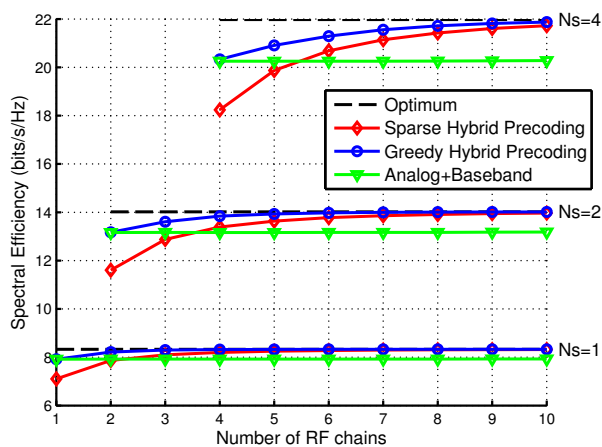

Fig. 4: Spectral Efficiency as function of the number of RF chains. ULA system $N_t = N_r = 32$ and $L_t = L_r = 4$. A fixed number of data streams $N_s \in \{1, 2, 4\}$ is considered.

high N_s, with GHP performing better than the sparse hybrid precoder in all the cases. In this experiment, GHP needs a number of RF chains equal to twice the number of streams to achieve the optimal performance. On the other hand, the decoupled analog+digital solution doesnt benefit from the extra RF chains. Therefore, the applicability of the analog+baseband precoder is limited to the case when $N_s = L_t$. We note that although an increment on the number of RF chains over the number of data stream represents an improvement with the hybrid precoders, this gain is relatively small. In practice, there will be a trade off between this marginal gain and other consideration such as the high power consumption and cost per RF chain.

VIII. CONCLUSION

In this paper we developed a low complexity hybrid precoding design for mmWave MIMO systems including an analog preprocessing network of phase shifters. The main advantage of the method is the avoidance of any assumption on the array geometry and the use of a dictionary with the associated costly correlation operations. This results in a lower overall complexity. Simulation results show that the achievable spectral efficiencies are close to the unconstrained solution and higher than other popular approaches.

REFERENCES

[1] T. Rappaport, R. W. Heath Jr., T. Daniels, and J. Murdock, *Millimeter wave wireless communications*. Prentice Hall, 2014.

[2] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 101–107, 2011.

[3] 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 Commun. Mag.*, vol. 52, no. 2, pp. 106–113, 2014.

[4] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. Wong, J. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, 2013.

[5] A. Hajimiri, H. Hashemi, A. Natarajan, and A. Komijani, "Integrated Phased Array Systems in Silicon," *Proc. IEEE*, vol. 93, no. 9, pp. 1637–1655, Sep. 2005.

[6] "Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications. Amendment 3: Enhancements for Very High Throughput in the 60 GHz Band," *IEEE Std. 802.11 ad*, 2012.

[7] T. Baykas, C.-S. Sum, Z. Lan, J. Wang, M. Rahman, H. Harada, and S. Kato, "Ieee 802.15.3c: the first ieee wireless standard for data rates over 1 gb/s," *Communications Magazine, IEEE*, vol. 49, no. 7, pp. 114–121, July 2011.

[8] J. Wang, Z. Lan, C.-W. Pyo, T. Baykas, C.-S. Sum, M. Rahman, J. Gao, R. Funada, F. Kojima, H. Harada, and S. Kato, "Beam codebook based beamforming protocol for multi-gbps millimeter-wave wpan systems," *Selected Areas in Communications, IEEE Journal on*, vol. 27, no. 8, pp. 1390–1399, October 2009.

[9] S. Hur, T. Kim, D. Love, J. Krogmeier, T. Thomas, and A. Ghosh, "Millimeter wave beamforming for wireless backhaul and access in small cell networks," *Communications, IEEE Transactions on*, vol. 61, no. 10, pp. 4391–4403, October 2013.

[10] J. Brady, N. Behdad, and A. Sayeed, "Beamspace mimo for millimeter-wave communications: System architecture, modeling, analysis, and measurements," *Antennas and Propagation, IEEE Transactions on*, vol. 61, no. 7, pp. 3814–3827, July 2013.

[11] X. Zhang, A. F. Molisch, and S. Kung, "Variable-phase-shift-based RF-baseband codesign for MIMO antenna selection," *IEEE Trans. Signal Process.*, vol. 53, no. 11, pp. 4091–4103, Nov. 2005.

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

[13] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, pp. 1499–1513, 2014.

[14] C. Rusu, R. Méndez-Rial, N. González-Prelcic, and R. W. Heath Jr., "Low Complexity Hybrid Sparse Precoding and Combining in Millimeter Wave MIMO Systems," in *IEEE International Conference on Communications*, 2015.

[15] A. Goldsmith and S. A. Jafar, "Capacity limits of MIMO channels," *Sel. Areas Commun.*, 2003.

[16] E. Telatar, "Capacity of Multi-antenna Gaussian Channels," *European transactions on telecommunications*, 1999.