Sparse Hybrid Precoding and Combining in Millimeter Wave MIMO Systems

Aryan Kaushik*, **John Thompson***, **Mehrdad Yaghoobi***
*Institute for Digital Communications,
The University of Edinburgh, United Kingdom.
Email: *{A.Kaushik, J.S.Thompson, M.Yaghoobi-Vaighan}@ed.ac.uk

Abstract

Millimeter wave (mmWave) communication allows us to exploit a new spectrum band between 30 GHz to 300 GHz to meet the growing demands of capacity for fifth generation (5G) wireless communication systems. Multiple-input multiple-output (MIMO) antennas can be used to tackle higher path loss and attenuation at mmWave frequencies compared to microwave bands. Beamforming, called precoding at the transmitter, is performed digitally in conventional microwave frequency MIMO systems, but at mmWave frequencies the higher cost and power consumption of system components means that the system cannot implement one radio frequency (RF) chain per antenna. To enable spatial multiplexing, hybrid precoders using fewer RF chains than antennas emerge as cost-effective and power saving alternative for the transceiver architecture of mmWave MIMO systems. This paper demonstrates the hybrid precoder design with its spectral efficiency and energy efficiency characteristics, and we compare the performance with that of optimal digital precoding (with one RF chain per antenna) and simplified beam steering systems. It also includes two different algorithmic solutions to meet the optimization objective. The orthogonal matching pursuit (OMP) algorithm appears to provide high performance solution to the problem, whereas the gradient pursuit (GP) algorithm is proposed as a cost-effective and fast approximation solution that can still provide equally high performance.

I. Introduction

To advance the state of present wireless communication systems, researchers are primarily concerned about the evolution of fifth generation (5G) networks and even beyond. It is suggested that initial 5G standards may be introduced by 2020 [1]. Such advanced systems systems demand lower latency, lower infrastructure costs, ultra-high reliability, higher mobility, improved range, much higher throughput, and increased capacity of networks [2,3]. The main differences of 5G systems compared to fourth generation (4G) systems will be the use of much greater spectrum allocations, higher aggregate capacity, much higher bit rates, longer battery life, and higher reliability to support many simultaneous users in both licensed and unlicensed RF bands [4]. The emerging advanced consumer devices and developed communication systems have resulted in ever-increasing demands on bandwidth and capacity [5]. The current carrier frequency spectrum has been limited to the very crowded range between 700 MHz and 2.6 GHz leading to the worldwide need for more spectrum and higher capacity. In such scenario, millimeter Wave (mmWave) appears to be a promising technology for future wireless communication systems [4,5]. Utilizing the unused wireless spectrum at much higher frequencies makes mmWave technology different from existing wireless solutions. MmWave offers larger bandwidth channels resulting in much higher data rates, thus supporting much better internet-based access and higher connectivity [4]. MmWave spectrum is currently used for various applications such as satellite communication, radio applications, and backhaul networks. MmWave technology is already a very significant technology for wireless backhaul [6] along with the possibility of self-backhaul in cellular systems. However, mmWave cellular systems do hold certain challenges such as supporting directional communication, susceptibility to shadowing, intermittent connectivity, and processing power consumption by data converters [7].
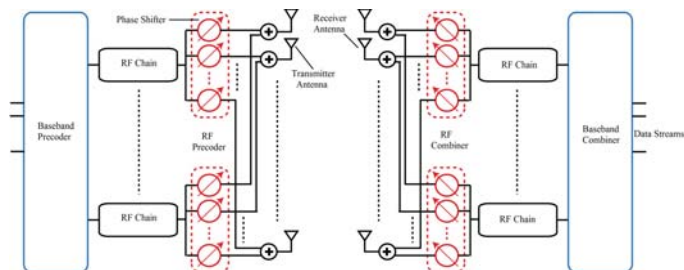

Fig. 1. Hardware block diagram of mmWave single-user fully-connected hybrid beamforming system.

MmWave technology fits very well with multiple-input multiple-output (MIMO) systems as the size of antenna arrays and associated electronics will reduce due to the shorter wavelengths [8]. MIMO technology has already been applied to commercial wireless local area networks and cellular systems at sub-6GHz frequencies. MIMO techniques at mmWave frequencies will be applied differently than at microwave frequencies due to changes in RF propagation and additional hardware constraints. Signal processing for mmWave MIMO systems is of critical importance. At lower frequencies, the signal processing actions are carried out at baseband leading to entirely digital signal processing solutions. While at higher frequencies, there are various hardware constraints making it difficult to have a separate radio frequency (RF) chain dedi-

cated to each antenna. Moreover, the practical implementations of system entities such as RF chains, power amplifiers, low noise amplifiers, and baseband connections are more difficult to construct at mmWave [9], and power consumption is a major issue as these entities become power hungry devices [10]. MmWave frequency systems will exploit polarization and spatial processing techniques such as very directional adaptive beamforming to improve the performance of the system. Deploying a large number of antennas results in high beamforming gain, forming directional beam patterns between transmitter and receiver, which further can assist in overcoming the higher path loss experienced at mmWave frequencies. One of the objectives of this paper is to focus on the sparse nature of the mmWave channel which allows us to use signal processing to enhance performance of mmWave systems towards ultimate performance limits.

One of the simplest approaches to apply MIMO in mmWave systems is analog beamforming which can be implemented at both transmitter and receiver. This approach often connects antenna elements via phase shifters to a single RF chain which supports single stream communication only and does not provide spatial multiplexing gains. Hybrid beamforming can be implemented instead to enable spatial multiplexing and multi-user MIMO communication. Fig. 1 shows the basic structure of a mmWave single-user fully-connected hybrid beamforming system [11] with digital baseband precoding followed by constrained RF precoding implemented using RF phase shifters. The same number of phase shifters as antennas are connected to each RF chain which leads to a fully-connected architecture. Precoding generally refers to beamforming at the transmitter, which may be generalized to support multi-stream (or multi-layer) transmission. At the receiver end, signal combining techniques can be used. One may find the unique advantage associated with hybrid precoding is that, to approach the performance of unconstrained solutions, the digital precoder can correct analog limitations such as cancelling residual multi-stream interference. Although hybrid precoding currently makes compromise on power consumption and hardware complexity yet there is much scope to exploit energy and capacity efficient designs.

Reference [11] proposes a fully-connected hybrid precoder design which leads to a capacity efficient mmWave MIMO system. For an energy efficient design, [12] considers sub-connected architecture, where each RF chain is connected to only a subset of transmitter antennas requiring fewer phase shifters in comparison to the fully-connected architecture. This energy efficient hybrid precoding design is based on successive interference cancellation (SIC) providing near-optimal performance and proposing a low complexity algorithmic solution. Reference [13] considers both fully-connected and partially-connected structures to design a hybrid precoder. The fully-connected structure seems to outperform partially-connected structure in terms of capacity whereas the latter shows higher energy efficiency. In [14] an energy efficient optimization to design the hybrid precoder through the use of optimal number of RF chains is proposed. More generally [15] provides a overview on the relationship between energy efficiency and spectral efficiency for different configurations of a hybrid beamforming system.

This paper mainly exhibits spectral efficiency and energy efficiency characteristics of a hybrid precoder which are helpful in analyzing the throughput and energy variations with respect to the system parameters and the channel parameters. The simulation results are plotted with respect to signal-to-noise ratio (SNR) and the number of RF chains. The solution to the optimization problem implements orthogonal matching pursuit (OMP) at the transmitter and the receiver which appears to be a low complexity solution. Gradient Pursuit (GP) method is introduced as a novel solution to the optimization objective which has the same performance as OMP yet it is a cost-effective and fast approximation solution. The performance and run time comparisons between both the algorithmic solutions are performed and GP is implemented to plot the spectral efficiency and energy efficiency characteristics.

The following notations have been used throughout the paper: $\mathbf{A}$, $\mathbf{a}$, and a stand for a matrix, a vector, and a scalar, respectively; $\mathbf{A}^{(i)}$ represents the ith column of $\mathbf{A}$; transpose and conjugate transpose of $\mathbf{A}$ are denoted as $\mathbf{A}^T$ and $\mathbf{A}^*$, respectively; $||\mathbf{A}||_F$, tr$(\mathbf{A})$, and det $(\mathbf{A})$ represent the Frobenius norm, trace, and determinant of $\mathbf{A}$, respectively; $||\mathbf{a}||_p$ is the p-norm of $\mathbf{a}$; $[\mathbf{A}|\mathbf{B}]$ denotes horizontal concatenation; diag$(\mathbf{A})$ generates a vector by the diagonal elements of $\mathbf{A}$; $\mathbf{I}_N$ and $\mathbf{0}_{X \times Y}$ represent $N \times N$ identity matrix and $X \times Y$ all-zeros matrix, respectively; $\mathcal{CN}(\mathbf{a}; \mathbf{A})$ denotes a complex Gaussian vector having mean $\mathbf{a}$ and covariance matrix $\mathbf{A}$, and i.i.d. shows that the entries of that vector are independent and identically distributed. The expectation and real part of a complex variable are denoted as $\mathcal{E}[.]$ and $\Re\{.\}$, respectively.

II. System and Channel Models

This section presents the mmWave system model and channel model used in this paper.

A. System Model

Considering a single-user mmWave system with N_t antennas at the transmitter end, sending N_s data streams to N_r receiver antennas. N_t^{rf} and N_r^{rf} denote the number of RF chains at the transmitter with the limitation $N_s \leq N_t^{rf} \leq N_t$ and at the receiver with the limitation $N_s \leq N_r^{rf} \leq N_r$, respectively. In other words, in massive MIMO communication systems, based on the function of the RF chains and the hybrid precoding scheme, the number of RF chains is larger than or equal to the number of baseband data streams and smaller than or equal to number of the transmitter antennas. The matrices $\mathbf{F}_{bb}$ and $\mathbf{F}_{rf}$ denote the $N_t^{rf} \times N_s$ baseband precoder and the $N_t \times N_t^{rf}$ RF precoder, respectively. Similarly at the receiver end, the matrices $\mathbf{W}_{bb}$ and $\mathbf{W}_{rf}$ denote the $N_r^{rf} \times N_s$ baseband combiner and the $N_r \times N_r^{rf}$ RF combiner, respectively. Fig. 1 shows the system setup. The signal, $\mathbf{x} = \mathbf{F}_{rf}\mathbf{F}_{bb}\mathbf{s}$, is transmitted where $\mathbf{s}$ is the $N_s \times 1$ symbol vector such that $\mathcal{E}[\mathbf{ss}^*] = \frac{1}{N_s}\mathbf{I}_{N_s}$. All elements of $\mathbf{F}_{rf}$ and $\mathbf{W}_{rf}$ are constrained to have equal norm. The power constraint at the transmitter end is satisfied by $||\mathbf{F}_{rf}\mathbf{F}_{bb}||_F^2 = N_s$. Considering a narrowband block-fading propagation channel with $\mathbf{H}$ as $N_r \times N_t$ channel matrix, which is assumed to be known to both the transmitter and the receiver, a discrete-time model for the received signal is

$$\mathbf{y} = \sqrt{\rho}\mathbf{H}\mathbf{F}_{rf}\mathbf{F}_{bb}\mathbf{s} + \mathbf{n}, \qquad (1)$$

where $\mathbf{y}$ is the $N_r \times 1$ received vector, ρ is the average received power, and $\mathbf{n}$ is a noise vector with entries which are i.i.d. $\mathcal{CN}(0, \sigma_n^2)$. After combining processing, the processed received signal can be written as follows:

$$\tilde{\mathbf{y}} = \sqrt{\rho} \mathbf{W}_{bb}^* \mathbf{W}_{rf}^* \mathbf{H} \mathbf{F}_{rf} \mathbf{F}_{bb} \mathbf{s} + \mathbf{W}_{bb}^* \mathbf{W}_{rf}^* \mathbf{n}, \quad (2)$$

For transmitted symbols following a Gaussian distribution, the achievable spectral efficiency can be expressed as follows:

$$R = \log_2 \det\{\mathbf{I}_{N_s} +$$
$$\frac{\rho}{N_s} \mathbf{R}_n^{-1} \mathbf{W}_{bb}^* \mathbf{W}_{rf}^* \mathbf{H} \mathbf{F}_{rf} \mathbf{F}_{bb} \mathbf{F}_{bb}^* \mathbf{F}_{rf}^* \mathbf{H}^* \mathbf{W}_{rf} \mathbf{W}_{bb}\}, \quad (3)$$

where $\mathbf{R}_n = \sigma_n^2 \mathbf{W}_{bb}^* \mathbf{W}_{rf}^* \mathbf{W}_{rf} \mathbf{W}_{bb}$ represents the noise covariance matrix after the combining processing.

B. Channel Model

The fading channel models used in traditional MIMO becomes inaccurate for mmWave channel modeling due to the high free-space path loss and large tightly-packed antenna arrays. So the mmWave propagation environment can be characterized by a narrowband clustered channel model, such as the Saleh-Valenzuela model [10]. For N_{cl} clusters and N_{ray} propagation paths each cluster, mmWave channel matrix can be depicted as follows:

$$\mathbf{H} = \sqrt{\frac{N_t N_r}{N_{cl} N_{ray}}} \sum_{i=1}^{N_{cl}} \sum_{l=1}^{N_{ray}} \alpha_{il} \mathbf{a}_r(\phi_{il}^r, \theta_{il}^r) \mathbf{a}_t(\phi_{il}^t, \theta_{il}^t)^*, \quad (4)$$

where α_{il} denotes the gain of lth ray in i^{th} cluster and it is assumed that α_{il} are i.i.d. $\mathcal{CN}(0, \sigma_{\alpha,i}^2)$, where $\sigma_{\alpha,i}^2$ is average power of the ith cluster such that $\sum_{i=1}^{N_{cl}} \sigma_{\alpha,i}^2 = \gamma$, γ being the normalization factor satisfying $\mathcal{E}[||\mathbf{H}||_F^2] = N_t N_r$, and $\gamma = \sqrt{\frac{N_t N_r}{N_{cl} N_{ray}}}$. Further, $\mathbf{a}_r(\phi_{il}^r, \theta_{il}^r)$ and $\mathbf{a}_t(\phi_{il}^t, \theta_{il}^t)$ represent the normalized receive and transmit array response vectors, where ϕ_{il}^t and θ_{il}^t are azimuth and elevation angles of departure, respectively, and ϕ_{il}^r and θ_{il}^r are azimuth and elevation angles of arrival, respectively. The antenna elements at the transmitter and the receiver can be modeled as ideal sectored elements [16] and then antenna element gains can be evaluated over the ideal sectors. In (4), the transmit and receive antenna element gains are considered unity over ideal sectors defined by $\phi_{il}^t \in [\phi_{min}^t, \phi_{max}^t]$ and $\theta_{il}^t \in [\theta_{min}^t, \theta_{max}^t]$; $\phi_{il}^r \in [\phi_{min}^r, \phi_{max}^r]$ and $\theta_{il}^r \in [\theta_{min}^r, \theta_{max}^r]$, respectively, and the gains are zero otherwise. This paper considers uniform linear array (ULA) antenna elements for simulations, where for a N_z-element ULA on z-axis, the array response vector can be expressed as follows [17]:

$$\mathbf{a}_z(\phi) = \frac{1}{\sqrt{N_z}} [e^{jm \frac{2\pi}{\lambda} d(\sin(\phi))}]^T, \quad (5)$$

where $0 \le m \le (N_z - 1)$ is a real integer counting through antennas, d is inter-element spacing, and λ is the signal wavelength. The array response vectors could also be computed considering a uniform planar array (UPA) of antenna elements in a two-dimensional plane [17].

III. Hybrid Precoder Design

It is usually difficult to find a global optimization solution for the joint optimization problem over transmitter and receiver precoders [18]. So, the design can be split into two sub-optimization problems, i.e, one focusing on designing $\mathbf{F}_{rf} \mathbf{F}_{bb}$ for the precoder and the other on designing $\mathbf{W}_{rf} \mathbf{W}_{bb}$ for the combiner. The mutual information obtained through Gaussian signaling over the channel is computed for the hybrid precoder $\mathbf{F}_{rf} \mathbf{F}_{bb}$, measuring the mutual dependence between the two matrices, as follows [11]:

$$\mathcal{I}(\mathbf{F}_{rf}, \mathbf{F}_{bb}) = \log_2 \det(\mathbf{I} + \frac{\rho}{N_s \sigma_n^2} \mathbf{H} \mathbf{F}_{rf} \mathbf{F}_{bb} \mathbf{F}_{bb}^* \mathbf{F}_{rf}^* \mathbf{H}^*) \ , \quad (6)$$

While designing hybrid precoders and combiners for mmWave MIMO systems, we are very much concerned about hardware complexity, spectral efficiency, and energy consumption for baseband processing and analog processing entities such as analog-to-digital converters (ADCs), digital-to-analog converters (DACs), RF chains, phase shifters, and power amplifiers. Sparing use of these entities can lead the system to operate in a very energy efficient manner. For instance, as the number of RF chains increase, more energy would get consumed leading to a decrease in energy efficiency. Measuring the energy efficiency characteristics with respect to the number of RF chains, as shown in Section IV, is quite helpful to design a energy efficient hybrid beamforming system. Meanwhile, the hybrid precoder optimization problem can be formulated as follows:

$$(\mathbf{F}_{rf}^{opt}, \mathbf{F}_{bb}^{opt}) = \max_{\mathbf{F}_{rf}, \mathbf{F}_{bb}} \mathcal{I}(\mathbf{F}_{rf}, \mathbf{F}_{bb}),$$
$$\text{s.t.} \ \ \mathbf{F}_{rf} \in \mathcal{F}_{rf}, \quad (7)$$
$$||\mathbf{F}_{rf} \mathbf{F}_{bb}||_F^2 = N_s,$$

where $\mathcal{F}_{rf}$ denotes the set of $N_t \times N_t^{rf}$ matrices having elements of constant magnitude. For such a non-convex constraint, it is difficult to yield general solutions to the problem. So in order to design the near-optimal hybrid precoder, certain assumptions and approximations can be exploited as in [11] to simplify the above problem. Equation (7) can be transformed in terms of the Euclidean distance between $\mathbf{F}_{rf} \mathbf{F}_{bb}$ and the channel's optimal fully digital precoder $\mathbf{F}_{opt}$. The hybrid precoder $\mathbf{F}_{rf} \mathbf{F}_{bb}$ can be located in a constrained space to be as close as possible to the optimal matrix $\mathbf{F}_{opt}$ in the unconstrained space. So the Euclidean distance $||\mathbf{F}_{opt} - \mathbf{F}_{rf} \mathbf{F}_{bb}||_F$ should be as small as possible for maximum throughput. We compute the channel's singular value decomposition (SVD) as $\mathbf{H} = \mathbf{U_H} \Lambda_\mathbf{H} \mathbf{V_H^*}$, where $\mathbf{U_H} \in \mathcal{C}^{N_r \times N_r}$ and $\mathbf{V_H} \in \mathcal{C}^{N_t \times N_t}$ are unitary matrices, and $\Lambda_\mathbf{H} \in \Re^{N_r \times N_t}$ is a rectangular matrix of singular values in decreasing order whose diagonal elements are non-negative real numbers and whose non-diagonal elements are zero. The optimal matrix $\mathbf{F}_{opt}$ is comprised of the first N_s columns of $\mathbf{V_H}$. As the array response vectors $\mathbf{a}_t(\phi_{il}^t, \theta_{il}^t)$ are constant-magnitude phase-only vectors and $\mathcal{F}_{rf}$ denotes the set of $N_t \times N_t^{rf}$ matrices having elements of constant magnitude, we can restrict $\mathcal{F}_{rf}$ to be a set of basis vectors $\mathbf{a}_t(\phi_{il}^t, \theta_{il}^t)$ in order to find the best low dimensional representation of the optimal matrix $\mathbf{F}_{opt}$. So the hybrid precoder optimization

problem can further be stated as follows:

$$(\mathbf{F}_{rf}^{opt}, \mathbf{F}_{bb}^{opt}) = \min_{\mathbf{F}_{rf},\mathbf{F}_{bb}} ||\mathbf{F}_{opt} - \mathbf{F}_{rf}\mathbf{F}_{bb}||_F,$$
$$\text{s.t. } \mathbf{F}_{rf}^{(i)} \in \{\mathbf{a}_t(\phi_{il}^t, \theta_{il}^t), \forall i, l\}, \quad (8)$$
$$||\mathbf{F}_{rf}\mathbf{F}_{bb}||_F^2 = N_s,$$

One may note here that the constraint on $\mathbf{F}_{rf}^{(i)}$ may be added into the optimization given (8) to obtain the problem as follows:

$$\tilde{\mathbf{F}}_{bb}^{opt} = \min_{\tilde{\mathbf{F}}_{bb}} ||\mathbf{F}_{opt} - \mathbf{A}_t\tilde{\mathbf{F}}_{bb}||_F,$$
$$\text{s.t. } ||\text{diag}(\tilde{\mathbf{F}}_{bb}\tilde{\mathbf{F}}_{bb}^*)||_0 = N_t^{rf}, \quad (9)$$
$$||\mathbf{A}_t\tilde{\mathbf{F}}_{bb}||_F^2 = N_s,$$

where $\mathbf{A}_t$ is an $N_t \times N_{cl}N_{ray}$ matrix consisting of array response vectors and $\tilde{\mathbf{F}}_{bb}$ is an $N_{cl}N_{ray} \times N_s$ matrix. The matrices $\mathbf{A}_t$ and $\tilde{\mathbf{F}}_{bb}$ help to obtain $\mathbf{F}_{rf}^{opt}$ and $\mathbf{F}_{bb}^{opt}$ as the N_t^{rf} non-zero rows of $\tilde{\mathbf{F}}_{bb}$ will give us the baseband precoder matrix $\mathbf{F}_{bb}^{opt}$ and the corresponding N_t^{rf} columns of $\mathbf{A}_t$ will provide the RF precoder matrix $\mathbf{F}_{rf}^{opt}$. Equation (9) basically reformulates (8) into a sparsity constrained reconstruction problem with one variable. The problem can now be addressed as a sparse approximation problem [19], and orthogonal matching pursuit (OMP) [20] can be used as an algorithmic solution to this problem. The receiver side follows a problem definition, optimization objective, and the same algorithmic solution can be used with minimal changes. As the hybrid combiner design has a similar mathematical formulation except for the extra transmitter power constraint at the transmitter, this paper mainly focuses on hybrid precoder design and the hybrid combiner design has been omitted. One may note here that by assuming the hybrid precoders $\mathbf{F}_{rf}\mathbf{F}_{bb}$ to be fixed, the hybrid combiners $\mathbf{W}_{rf}\mathbf{W}_{bb}$ can be designed in order to minimize the mean-squared-error (MSE) between the transmitted and processed received signals by using the linear minimum mean-square error (MMSE) receiver.

Algorithm 1: Hybrid Precoder Design through Orthogonal Matching Pursuit (OMP) [20]

Require: $\mathbf{F}_{opt}$
1: $\mathbf{F}_{rf} = \emptyset$
2: $\mathbf{F}_{res} = \mathbf{F}_{opt}$
3: **for** $i \leq N_t^{rf}$
4: $\quad \mathbf{\Psi} = \mathbf{A}_t^*\mathbf{F}_{res}$
5: $\quad k = \arg\max_{l=1,\ldots,N_{cl}N_{ray}}(\mathbf{\Psi}\mathbf{\Psi}^*)_{l,l}$
6: $\quad \mathbf{F}_{rf} = \left[\mathbf{F}_{rf} \mid \mathbf{A}_t^{(k)}\right]$
7: $\quad \mathbf{F}_{bb} = (\mathbf{F}_{rf}^*\mathbf{F}_{rf})^{-1}\mathbf{F}_{rf}^*\mathbf{F}_{opt}$
8: $\quad \mathbf{F}_{res} = \frac{\mathbf{F}_{opt}-\mathbf{F}_{rf}\mathbf{F}_{bb}}{||\mathbf{F}_{opt}-\mathbf{F}_{rf}\mathbf{F}_{bb}||_F}$
9: **end for**
10: $\mathbf{F}_{bb} = \sqrt{N_s}\frac{\mathbf{F}_{bb}}{||\mathbf{F}_{rf}\mathbf{F}_{bb}||_F}$
11: return $\mathbf{F}_{rf}, \mathbf{F}_{bb}$

Algorithm 1 starts by finding the array response vector $\mathbf{a}_t(\phi_{il}^t, \theta_{il}^t)$ along which the optimal precoder has the maximum projection, and then concatenates that selected column vector into the RF precoder $\mathbf{F}_{rf}$ as shown in step 6. It then continues to find least squares solution to the baseband precoder $\mathbf{F}_{bb}$,

and then the residual precoding matrix $\mathbf{F}_{res}$ is computed in order to remove the contribution of the selected vector. Then the algorithm continues to find the column along which $\mathbf{F}_{res}$ has the largest projection until all RF chains have been used. The transmit power constraint is satisfied at step 10, which is applicable for a general case of $N_s \geq 1$.

To develop fast approximate OMP algorithms that require less storage, [21] proposes improvements to greedy strategies using directional pursuit methods, and discusses optimization schemes on the basis of gradient, conjugate gradient, and approximate conjugate gradient approaches. The gradient pursuit (GP) method is introduced as a novel solution to the optimization objective exhibiting the same performance as OMP, cheaper cost consumption, and faster processing time. Unlike OMP where optimum signal approximation is achieved on all the selected atoms, GP makes use of a single gradient direction for the approximation avoiding the need to consider all the atoms and hence leading to reduced computation time. The computation time is considerably less for large MIMO configurations when implementing GP, as shown in section IV. Algorithm 2 starts in the same way as Algorithm 1. There is a index set which is updated at each iteration as shown in step 6 which is used to generate baseband precoder matrix $\mathbf{F}_{bb}$. The gradient direction, as mentioned in step 8, is computed at each iteration and the step-size is determined explicitly making use of the gradient direction, as shown in step 10. Finally the RF precoder matrix $\mathbf{F}_{rf}$ and the baseband precoder matrix $\mathbf{F}_{bb}$ are obtained at the end of the algorithm. The transmit power constraint is satisfied at step 14.

Algorithm 2: Hybrid Precoder Design through Gradient Pursuit (GP) [21]

Require: $\mathbf{F}_{opt}$
1: $\mathbf{F}_{rf} = \emptyset$, $\Gamma = \emptyset$
2: $\mathbf{F}_{res} = \mathbf{F}_{opt}$, $\mathbf{F}_{bb} = 0$
3: **for** $i \leq N_t^{rf}$
4: $\quad \mathbf{\Psi} = \mathbf{A}_t^*\mathbf{F}_{res}$
5: $\quad k = \arg\max_{l=1,\ldots,N_{cl}N_{ray}}(\mathbf{\Psi}\mathbf{\Psi}^*)_{l,l}$
6: $\quad \Gamma = \Gamma \cup k$
7: $\quad \mathbf{F}_{rf} = \left[\mathbf{F}_{rf} \mid \mathbf{A}_t^{(k)}\right]$
8: $\quad \mathbf{D} = \mathbf{F}_{rf}^*\mathbf{F}_{res}$
9: $\quad \mathbf{C} = \mathbf{F}_{rf}\mathbf{D}$
10: $\quad g = \frac{\text{tr}\{\mathbf{F}_{res}^*\mathbf{C}\}}{||\mathbf{C}||_F^2}$
11: $\quad \mathbf{F}_{bb}|_\Gamma = \mathbf{F}_{bb}|_\Gamma - g\mathbf{D}$
12: $\quad \mathbf{F}_{res} = \mathbf{F}_{res} - g\mathbf{C}$
13: **end for**
14: $\mathbf{F}_{bb} = \sqrt{N_s}\frac{\mathbf{F}_{bb}}{||\mathbf{F}_{rf}\mathbf{F}_{bb}||_F}$
15: return $\mathbf{F}_{rf}, \mathbf{F}_{bb}$

For the fully connected hybrid precoder design, it is quite interesting to observe the energy performance. Reference [15] suggests that energy efficiency ε can be defined as the ratio between spectral efficiency R and total power consumption P_{tot} as shown in (10). The total power consumption is the sum of power consumed for transmission, and baseband processing

and analog processing entities.

$$\varepsilon = \frac{R}{P_{tot}}$$

$$= \frac{R}{P_{cp} + N_t^{rf} P_{rf} + N_{ps}(P_{ps} + P_{pa})} bits/Hz/J, \quad (10)$$

where $N_{ps}, P_{cp}, P_{rf}, P_{ps}$, and P_{pa} represent the number of phase shifters, the common power of transmitter, the power per RF chain, the power per phase shifter, and the power per power amplifier. The energy consumed by the RF chains is a major concern leading to high value of P_{rf} with substantial increase in each RF chain. In a fully-connected hybrid precoder structure, one can consider that N_{ps} is equal to $N_t^{rf} N_t$ [12,13].

IV. Simulation Results

This section demonstrates the spectral efficiency and energy efficiency characteristics of the hybrid precoder design. For observation, there are 10 rays for each cluster and there are 8 clusters in total, i.e., $N_{ray} = 10$ and $N_{cl} = 8$. The average power of each cluster is unity, i.e., $\sigma_{\alpha,i} = 1$. The azimuth and elevation angles of departure and arrival are computed on the basis of a Laplacian distribution with uniformly distributed mean angles within the range of $60°$ to $120°$ in the azimuth domain, and $80°$ to $100°$ in the elevation domain. The angle spread which is the standard deviation of the Laplacian distribution of the angles is set to be $7.5°$. The antenna elements in the ULA are spaced by half wavelength distance. The symbol vector $\mathbf{s}$ is generated using quadrature amplitude modulation (QAM) scheme. The signal-to-noise ratio (SNR) is determined as $\frac{\rho}{\sigma_n^2}$ for the plots. All the simulation results are averaged over 5000 random channel realizations.
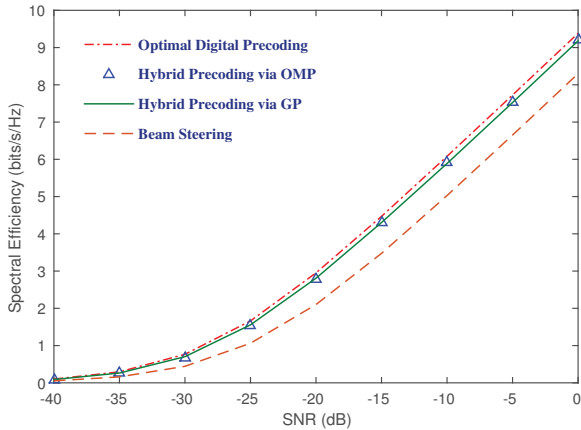
and at the receiver ends. Hybrid precoding implements 4 RF chains both at the transmitter and the receiver, i.e., $N_t^{rf} = N_r^{rf} = 4$. Both OMP and GP algorithmic solutions have been implemented for the hybrid precoder design. It can be observed that hybrid precoding performs slightly worse than optimal digital precoding but it is clearly better than beam steering. Moreover, the hybrid precoder using GP shows the same performance characteristics as that for OMP. GP provides a fast approximation solution as it requires less run time than OMP, which provides us a novel cost-effective solution to design the hybrid precoders.
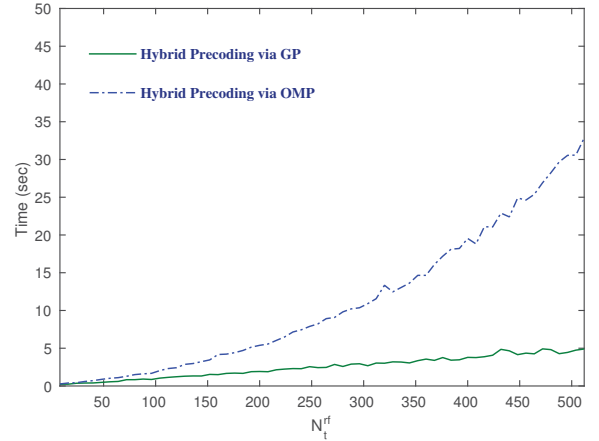

Fig. 3. Time evaluation with respect to number of RF chains for OMP and GP for 512×512 mmWave system with $N_{cl}=12$, $N_{ray} = 20$, $N_s = 8$ and SNR = -25 dB
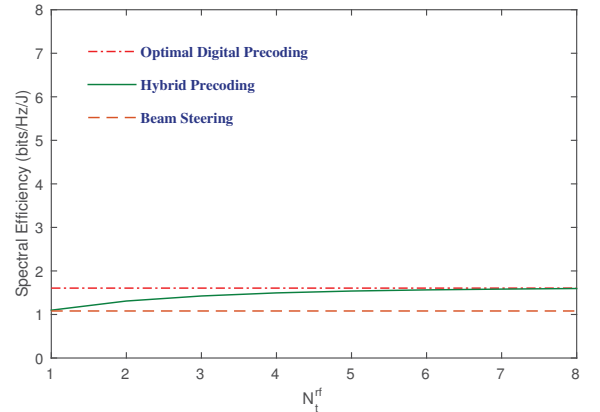

Fig. 2. Spectral efficiency for several precoding solutions for 64×16 fully-connected mmWave system with $N_s = 1$, $N_{cl} = 8$, and $N_{ray} = 10$.

Fig. 4. Spectral efficiency for several fully-connected precoder designs while SNR = -25 dB.

Fig. 2 shows the spectral efficiency versus SNR plot for several precoding solutions. For a single-user 64×16 mmWave system with a single stream being transmitted and received, the parameters are set in such a way that the hybrid precoder $\mathbf{F}_{rf}\mathbf{F}_{bb}$ can be made sufficiently close to the optimal precoder $\mathbf{F}_{opt}$. The optimal digital precoder uses N_t RF chains at the transmitter and N_r RF chains at the receiver, while beam steering [22] uses only a single RF chain both at the transmitter

The run time for GP is less than that of OMP for both small and large MIMO configurations. Fig. 3 shows the run time characteristics with respect to the number of RF chains for both GP and OMP for a large 512×512 mmWave system with $N_{cl} = 12$, $N_{ray} = 20$, $N_s = 8$, and SNR = -25dB. The time difference between both the algorithmic solutions is considerable which shows that GP is a better practical solution and more efficient than OMP to design a hybrid precoder. As GP has the same performance but less run time, the rest of

Common power of transmitter	$P_{cp} = 10$ W
Power per RF chain	$P_{rf} = 100$ mW
Power per phase shifter	$P_{ps} = 10$ mW
Power per power amplifier	$P_{pa} = 300$ mW

TABLE I.   SIMULATION PARAMETERS FOR THE POWER MODEL [10].

the plots in this paper make use of GP as the algorithmic solution to find the optimum precoder. Fig. 4 plots the spectral efficiency characteristics of the hybrid precoder, the optimal digital precoder, and beam steering system with respect to the number of RF chains at a SNR of -25 dB. It can be observed from Fig. 4 that the spectral efficiency of the hybrid precoder increases gradually and starts approximating the performance of the optimal digital precoder. It also clearly outperforms the beam steering approach in terms of spectral efficiency with increase in number of RF chains for a certain SNR (such as -25 dB).
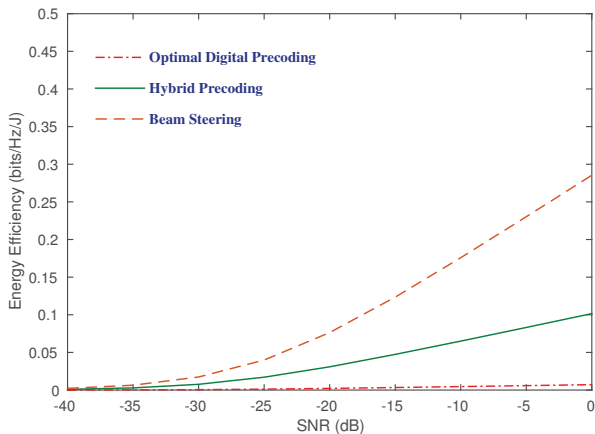

Fig. 5.   Energy efficiency for several precoding solutions for 64×16 fully-connected mmWave system with $N_s = 1$, $N_{cl} = 8$, and $N_{ray} = 10$.

Fig. 5 shows the energy efficiency versus SNR plot for several precoding solutions. To illustrate the achievable energy efficiency of different precoding solutions, the parameters in (10) are set as as shown in Table I and the other required parameters are same as used to obtain Fig. 2. The energy efficiency performance of the hybrid precoder clearly appears to outperform the optimal digital precoder as the SNR increases. However, the beam steering approach performs better in terms of energy efficiency as only one RF chain is being used in that system which reduces the energy consumption considerably. As N_{ps} is scaled linearly with N_t^{rf} and N_t, the energy consumption will significantly increase with respect to N_t^{rf}. For the same reason, beam steering outperforms hybrid precoding and optimal digital precoding as number of RF chains increases for a certain SNR (such as -25 dB) as shown in Fig. 6. The hybrid precoding performs exactly the same as beam steering in terms of energy efficiency with use of a single RF chain. One should note that, in order to achieve a significant spectral efficiency gain while accepting an increase
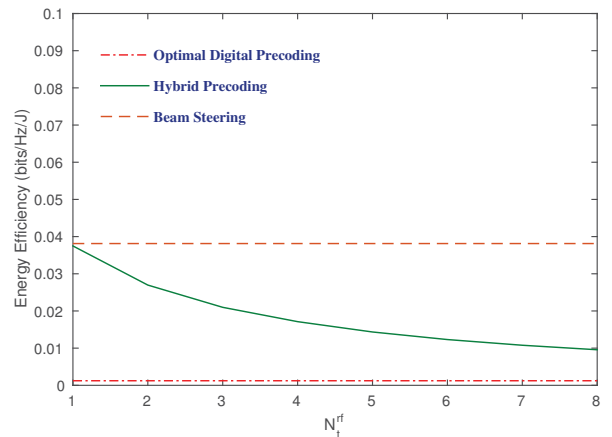

Fig. 6.   Energy efficiency for several fully-connected precoder designs while SNR = -25 dB.

in the energy consumption, the hybrid precoder solution might be a better approach to follow. For instance, to obtain a gain of 1 bits/s/Hz over the beam steering approach, the hybrid precoder will exhibit 0.11 bits/Hz/J less energy efficiency than beam steering at SNR = -10 dB as observed from Fig. 2 and Fig. 5.

V.   Conclusion

This paper is focused on evaluating the spectral efficiency and energy efficiency characteristics of a hybrid precoder which help in designing capacity and energy efficient hybrid mmWave communication systems. The spectral efficiency and energy efficiency characteristics of a hybrid precoder are compared with that of optimal digital precoding (with one RF chain per antenna) and simplified beam steering systems. It can be observed that the hybrid precoder design provides near-optimal spectral efficiency, and outperforms the optimal digital precoder significantly in terms of energy efficiency. While compared to the conventional beam steering approach, the hybrid precoder shows notable performance gain in terms of spectral efficiency. However, beam steering outperforms hybrid precoding in terms of energy efficiency with respect to SNR and number of RF chains. The gradient pursuit (GP) method is introduced as a novel algorithmic solution to the optimization objective. The orthogonal matching pursuit (OMP) algorithm appears to provide high performance solution to the problem, whereas the GP algorithm is proposed as a cost-effective and fast approximation solution. GP shows the same performance as OMP but it requires less run time for both small and large MIMO configurations. This research work will be extended to design an energy efficient hybrid precoder with a fully-connected architecture through optimizing the baseband precoder and RF precoder matrices along with optimizing the number of RF chains, and compare the energy performance of the fully optimized hybrid precoder to the hybrid precoder before optimization, the optimal digital precoder, and the simplified beam steering system.

References

[1] Li, X., Gani, A., Salleh, R., and Zakaria, O.: 'The future of mobile wireless communication networks', IEEE Intern. Conf. Commun. Software and Networks, Macau, Feb. 2009, pp. 554-557.

[2] NGMN 5G White Paper, v.1.0, Feb. 2015, pp. 1-125.

[3] 2020 Networld White Paper for Research Beyond 5G, v.1.0, Oct. 2015, pp. 1-43.

[4] Rappaport, T. S., Shu, S., Mayzus, R., Hang, Z., Azar, Y., Wang, K., Wong, G. N., Schulz, J. K., Samimi, M., and Gutierrez: 'Millimeter wave mobile communications for 5G cellular: It will work!', IEEE Access, 2013, 1, pp. 335-349.

[5] Pi, Z., and Khan, F.: 'An introduction to millimeter-wave mobile broadband systems', IEEE Commun. Mag., 2011, 49, (6), pp. 101-107.

[6] Hur, S., Kim, T., Love, D. J., Krogmeier, J. V., Thomas, T. A., and Ghosh, A.: 'Millimeter wave beamforming for wireless backhaul and access in small cell networks', IEEE Trans. Commun., 2013, 61, (10), pp, 4391-4403, .

[7] Rangan, S., Rappaport, T. S., and Erkip, E.: 'Millimeter-wave cellular wireless networks: potentials and challenges', Proc. IEEE, 2014, 102, (3), pp. 366-385.

[8] Biglarbegian, B., Fakharzadeh, M., Busuioc, D., Nezhad-Ahmadi, M.-R., and Safavi-Naeini, S.: 'Optimized microstrip antenna arrays for emerging millimeter-wave wireless applications', IEEE Trans. Antenn. and Propag., 2011, 59, (5), pp. 1742-1747.

[9] Zhang, J. A., Huang, X., Dyadyuk, V., and Guo, Y. J.: 'Massive hybrid antenna array for millimeter-wave cellular communications', IEEE Wireless Commun., 2015, 22, (1), pp. 79-87,

[10] Rappaport, T. S., Heath, R. W., Daniels, R. C., and Murdock, J. N.: 'Millimeter wave wireless communications' (Prentice-Hall, Sept. 2014).

[11] Ayach, O. E., Rajagopal, S., Abu-Surra, S., Pi, Z., and Heath, R. W.: 'Spatially sparse precoding in millimeter wave MIMO systems', IEEE Trans. Wireless Commun., 2014, 13, (3), pp. 1499-1513.

[12] Gao, X., Dai, L., Han, S., I, C.-L., and Heath, R. W.: 'Energy-efficient hybrid analog and digital precoding for mmWave MIMO systems with large antenna arrays', IEEE J. Sel. Areas Commun., 2016, 34, (4), pp. 1-12.

[13] Yu, X., Shen, J. C., Zhang, J., and Letaief, K. B.: 'Alternating minimization algorithms for hybrid precoding in millimeter wave MIMO systems', IEEE J. Sel. Topics Signal Process., 2016, 10, (3), pp. 485-500.

[14] Zi, R., Ge, X., Thompson, J., Wang, C.-X., Wang, H., and Han, T.: 'Energy efficiency optimization of 5G radio frequency chain systems', IEEE J. Sel. Areas Commun., 2016, 34, (4), pp. 1-16.

[15] Han, S., I, C.-L., Xu, Z., and Rowell, C.: 'Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G', IEEE Commun. Mag., 2015, 53, (1), pp. 186-194.

[16] Singh, S., Mudumbai, R., and Madhow, U.: 'Interference analysis for highly directional 60-GHz mesh networks: The case for rethinking medium access control', IEEE/ACM Trans. Netw., 2011, 19, (5), pp. 1513-1527.

[17] Balanis, C.: 'Antenna Theory' (Wiley, 1997).

[18] Palomar, D. P., Cioffi, J. M., and Lagunas, M. A.: 'Joint Tx-Rx beamforming design for multicarrier MIMO channels: a unified framework for convex optimization', IEEE Trans. Signal Process., 2003, 51, (9), pp. 2381-2401.

[19] Tropp, J. A., Gilbert, A. C., and Strauss, M. J.: 'Algorithms for simultaneous sparse approximation-part I: greedy pursuit", Signal Process., 2006, 86, (3), pp. 572-588.

[20] Tropp, J., and Gilbert, A.: 'Signal recovery from random measurements via orthogonal matching pursuit', IEEE Trans. Info. Theory, 2007, 53, (12), pp. 4655-4666.

[21] Blumensath, T., and Davies, M. E.: 'Gradient pursuits', IEEE Trans. on Signal Process., 2008, 56, (6), pp. 2370-2382.

[22] Ayach, O. E., Heath, R. W., Abu-Surra, S., Rajagopal, S., and Pi, Z.: 'The capacity optimality of beam steering in large millimeter wave MIMO systems', in Proc. 2012 13th IEEE Int. Work. Signal Process. Advances Wireless Commun., Cesme, June 2012, pp. 100-104.