

Two-Stage Hybrid Precoding Algorithm Based on Switch Network for Millimeter Wave MIMO Systems

Fulai Liu^{1, 2, *, †}, Xiaodong Kan^{1, 2, †}, Xiaoyu Bai², Ruiyan Du^{1, 2}, and Yanshuo Zhang^{1, 2}

Abstract—Owing to the hardware cost and power consumption limitation, hybrid precoding has been recently considered as an alternative to the fully digital precoding in millimeter wave (mmWave) large-scale multiple-input multiple-output (MIMO) systems. Although the number of radio frequency (RF) chains is reduced to a certain extent in the hybrid precoding structure, a great number of phase shifters are still needed. In this paper, we present a new hybrid precoding architecture based on switch network to decrease the power consumption of hybrid precoder by reducing the number of phase shifters greatly. The new hybrid precoding architecture consists of three parts, a digital precoder, an analog precoder, and a switch network, in which the switch network is used to offer a dynamic connection from phase shifters to antennas. Afterwards, a two-stage algorithm is proposed to determine each part of the hybrid precoding implementation. Specifically, the product of the analog precoding matrix and digital precoding matrix is viewed as a whole matrix firstly, thereby the original problem is simplified into a two-variable problem which is relatively easy to be solved. Then, the decomposition of the analog precoding matrix and digital precoding matrix is considered in the second stage. Simulation results show that the presented implementation can not only provide a better trade-off between hardware complexity and system performance, but also achieve higher energy efficiency with far fewer phase shifters than previous works.

1. INTRODUCTION

Millimeter wave (mmWave) communication systems, operating in the spectrum from 30 GHz to 300 GHz, have attracted extensive attention over the past years [1–3]. Multiple-input multiple-output (MIMO) is one of the promising techniques, which can exploit large-scale antenna elements at transceivers to achieve high beamforming gains and combat huge attenuation and penetration loss at mmWave frequencies [4]. However, increased power consumption and hardware cost make a fully digital precoder infeasible for large-scale mmWave MIMO systems [5]. To overcome this shortcoming, a hybrid precoding architecture, which only adopts a limited number of RF chains to connect a low-dimensional digital baseband precoder and a high-dimensional analog RF precoder, has recently received much consideration [6, 7].

In general, the hybrid precoding is categorized into fully-connected and partially-connected structures with phase shifters (FC-PSs and PC-PSs). Recently, several hybrid precoding algorithms have been presented for FC-PSs. The spatially sparse precoding algorithm in [7] reformulates the hybrid precoding problem as a sparse reconstruction problem and solves it by the orthogonal matching pursuit (OMP). Codebook-based hybrid precoding algorithm in [8] involves an iterative searching process in a predefined codebook to find the optimal hybrid precoding matrix. The works in [9–12] devise hybrid precoding algorithm by matrix decomposition and alternative minimization, respectively, and

Received 28 October 2018, Accepted 6 December 2018, Scheduled 29 December 2018

* Corresponding author: Fulai Liu (fulailiu@126.com).

¹ Engineer Optimization & Smart Antenna Institute, Northeastern University at Qinhuangdao, Qinhuangdao, China. ² School of Computer Science and Engineering, Northeastern University, Shenyang, China. [†] Fulai Liu and Xiaodong Kan have made the same contributions on this paper.

the objective of achieving spectral efficiency is close to that of fully digital solutions. The iterative column maximization algorithm and iterative coordinate descent algorithm for FC-PSs are studied in [13] and [14], respectively.

The hybrid precoding scheme based on FC-PSs enjoys fully beamforming gain, since each RF chain is connected to all antenna elements via phase shifters. However, the number of required phase shifters is as large as the product of the numbers of RF chains and antenna elements, which leads to excessive hardware cost and power consumption. To improve the hardware efficiency, the hybrid precoding scheme based on PC-PSs is one possible way to significantly reduce the number of phase shifters, in which each RF chain is only connected to a subset of antennas.

In [15, 16], a codebook-based design of hybrid precoders for PC-PSs is proposed for narrow-band and orthogonal frequency division multiplexing systems, respectively. The design complexity of these algorithms is low; however, the limited size of the codebook gives rise to an inevitable performance loss. In [17], a semidefinite relaxation (SDR_AltMin) algorithm is introduced by utilizing the idea of alternating minimization, which can provide substantial performance gains for PC-PSs. Based on realistic PC-PSs with low complexity, an iterative hybrid precoding algorithm is studied in [8], where successive interference cancellation is exploited to obtain the analog RF precoding matrix. In addition, to improve the system performance for PC-PSs, a greedy hybrid precoding algorithm and a modified K-means-based hybrid precoding algorithm are developed to dynamically optimize the sub-arrays in [18] and [19], respectively.

However, while the PC-PSs are studied substantially, there still exist an inevitable gap compared with the performance of FC-PSs [17]. To achieve the tradeoff among power consumption, hardware complexity and spectral efficiency of the hybrid precoder, in this paper, we focus on a novel hardware-efficient hybrid precoding architecture with switch network instead of phase shifters in mmWave MIMO systems. The main contributions of this paper can be summarized as follows.

- Firstly, a new hybrid architecture based on switch network is proposed in the analog processing stage. The objective of the proposed architecture is to further reduce power consumption of mmWave MIMO systems. Compared with the existing FC-PSs and PC-PSs, a dynamic switch network is added to connect the phase shifters and antenna elements in the proposed scheme, which can significantly reduce the number of phase shifters and achieve substantial hardware efficiency gain.
- Secondly, because it is challenging to jointly optimize the switch network matrix, digital precoding matrix, and RF precoding matrix, a two-stage alternating minimization algorithm is introduced to facilitate the solution of each part of the proposed scheme in this paper. In the first stage, the binary switch network matrix is derived analytically, and then, the decomposition of the digital precoding matrix and RF precoding matrix is considered in the second stage.
- Finally, we evaluate the performance of the proposed hybrid precoding architecture based on switch network in comparison with FC-PSs and PC-PSs. Simulation results show that the proposed hybrid precoding scheme based on switch network can yield reasonable reduction in the power consumption.

The rest of this paper is organized as follows. Section 2 briefly introduces the mmWave system model and proposed hybrid precoding implementation. The proposed hybrid precoding algorithm is developed in detail and followed by the problem formulation in Section 3. Section 4 presents several simulation results to verify the performance of the proposed algorithm. Finally, Section 5 provides a concluding remark to summarize the paper.

Throughout this paper, bold upper-case letters and bold lower-case letters are used for matrix and vector, respectively. For example, \mathbf{a} and \mathbf{A} stand for a column vector and a matrix, respectively. $\mathbf{A}_{i,j}$ is the element (i, j) th of \mathbf{A} . The transpose, conjugate transpose and Moore-Penrose pseudo-inverse of \mathbf{A} are represented by \mathbf{A}^T , \mathbf{A}^H and \mathbf{A}^\dagger . $|\mathbf{A}|$ and $\|\mathbf{A}\|_F$ denote the determinant and Frobenius norm of \mathbf{A} . $\text{tr}(\mathbf{A})$ and $\text{vec}(\mathbf{A})$ indicate the trace and vectorization of \mathbf{A} . $\|\mathbf{a}\|_2$ denotes the 2-norm of \mathbf{a} . \mathbf{I}_N stands for the $N \times N$ identity matrix. Expectation is introduced by $\mathbb{E}[\cdot]$, and the real part of a complex variable is represented by $\Re\{\cdot\}$. $\mathcal{CN}(\mu, \sigma^2)$ is the complex Gaussian distribution with mean μ and variance σ^2 , and $\mathcal{U}(a, b)$ represents the uniform distribution between a and b .

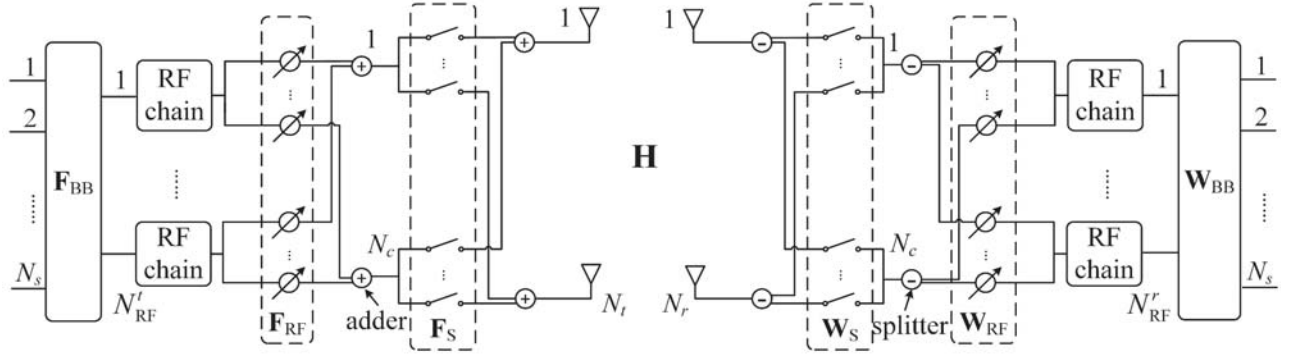


Figure 1. A single-user mmWave MIMO system with hybrid precoder and combiner implementation.

2. SYSTEM MODEL

Consider a single-user mmWave MIMO system model as shown in Fig. 1, in which N_s independent data streams are sent and collected from N_t antenna elements at the transmitter and N_{RF}^t RF chains to N_r antenna elements at the receiver and N_{RF}^r RF chains. In this model, only a small number of RF chains are available, i.e., $N_s \leq N_{\text{RF}}^t \leq N_t$ and $N_s \leq N_{\text{RF}}^r \leq N_r$. In the proposed implementation, it can be seen that the signal from each RF chain is transmitted by N_c available phase shifters, where $N_c \ll N_t$, which can improve the energy efficiency effectively. Compared with high-resolution phase shifters, the switch network only has binary on-off states, thus, the implementation of an adaptive switch network is much easier than phase shifters [20]. The transmitted signal vector \mathbf{x} can be given by $\mathbf{x} = \mathbf{F}_S \mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}} \mathbf{s}$ [21], where $\mathbf{F}_S \in \mathbb{C}^{N_t \times N_c}$, $\mathbf{F}_{\text{RF}} \in \mathbb{C}^{N_c \times N_{\text{RF}}^t}$ and $\mathbf{F}_{\text{BB}} \in \mathbb{C}^{N_{\text{RF}}^t \times N_s}$ stand for the binary switch network matrix, analog RF precoding matrix and digital baseband precoding matrix, respectively. $\mathbf{s} \in \mathbb{C}^{N_s \times 1}$ is the symbol vector such that $\mathbb{E}[\mathbf{s}\mathbf{s}^H] = \frac{1}{N_s} \mathbf{I}_{N_s}$. Since the analog RF precoder is implemented using phase shifters, which can only adjust the phases of the signals, all entries of \mathbf{F}_{RF} are subjected to constant modulus constraint. To reflect that, the constraint can be given by $|(\mathbf{F}_{\text{RF}})_{i,j}| = 1$, $i = 1, 2, \dots, N_c$, $j = 1, 2, \dots, N_{\text{RF}}^t$. The normalized total transmit power constraint is given by $\|\mathbf{F}_S \mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}}\|_F^2 = N_s$. For simplicity, a narrow-band block-fading propagation channel model is considered, and the received signal after decoding processing is expressed as [21]

$$\mathbf{y} = \sqrt{\rho} \mathbf{W}_{\text{BB}}^H \mathbf{W}_{\text{RF}}^H \mathbf{W}_S^H \mathbf{H} \mathbf{F}_S \mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}} \mathbf{s} + \mathbf{W}_{\text{BB}}^H \mathbf{W}_{\text{RF}}^H \mathbf{W}_S^H \mathbf{n}, \quad (1)$$

where ρ represents the average received power, and \mathbf{n} is the vector of i.i.d. $\mathcal{CN}(0, \sigma_n^2)$ noise, in which σ_n^2 stands for the noise power. \mathbf{H} denotes the $N_r \times N_t$ channel matrix, $\mathbf{W}_S \in \mathbb{C}^{N_r \times N_c}$, $\mathbf{W}_{\text{RF}} \in \mathbb{C}^{N_c \times N_{\text{RF}}^r}$ and $\mathbf{W}_{\text{BB}} \in \mathbb{C}^{N_{\text{RF}}^r \times N_s}$ stand for the switch matrix, RF combining matrix and baseband combining matrix, respectively. The RF combiner is also subjected to the constant modulus constraint, i.e., $|(\mathbf{W}_{\text{RF}})_{i,j}| = 1$, $i = 1, 2, \dots, N_c$, $j = 1, 2, \dots, N_{\text{RF}}^r$.

Assume that perfect channel state information (CSI) is known at both the transmitter and receiver [21]. When Gaussian symbols are transmitted over the channel, and the achievable spectral efficiency can be given by

$$R = \log_2 \left(\left| \mathbf{I}_{N_s} + \frac{\rho}{\sigma_n^2 N_s} (\mathbf{W}_S \mathbf{W}_{\text{RF}} \mathbf{W}_{\text{BB}})^H \mathbf{H} \mathbf{F}_S \mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}} \mathbf{F}_{\text{BB}}^H \mathbf{F}_{\text{RF}}^H \mathbf{F}_S^H \mathbf{H}^H (\mathbf{W}_S \mathbf{W}_{\text{RF}} \mathbf{W}_{\text{BB}}) \right| \right). \quad (2)$$

Based on the Saleh-Valenzuela model [7], the mmWave channel matrix \mathbf{H} is assumed to be a sum of N_{cl} clusters, each of which consists of N_{ray} rays. Let $L = N_{\text{cl}} N_{\text{ray}}$ be the total number of propagation paths, and the channel matrix \mathbf{H} can be expressed as

$$\mathbf{H} = \sqrt{\frac{N_t N_r}{L}} \sum_{i=1}^{N_{\text{cl}}} \sum_{l=1}^{N_{\text{ray}}} \alpha_{il} \mathbf{a}_r(\phi_{il}^r, \theta_{il}^r) \mathbf{a}_t^H(\phi_{il}^t, \theta_{il}^t), \quad (3)$$

where α_{il} denotes the complex path gain of the l th ray in the i th cluster; $\mathbf{a}_r(\phi_{il}^r, \theta_{il}^r)$ and $\mathbf{a}_t(\phi_{il}^t, \theta_{il}^t)$ represent array response vectors of the receiver and transmitter, respectively, where ϕ_{il}^r , θ_{il}^r , ϕ_{il}^t and θ_{il}^t

are azimuth and elevation angles of arrival and departure (AoAs and AoDs), respectively. The array response vector for an $M \times N$ uniform planar array (UPA) is defined as [22]

$$\mathbf{a}(\phi_{il}, \theta_{il}) = \frac{1}{\sqrt{MN}} \left[1 \quad \dots \quad e^{j\frac{2\pi}{\lambda}d(p \sin \phi_{il} \sin \theta_{il} + q \cos \theta_{il})} \quad \dots \quad e^{j\frac{2\pi}{\lambda}d((M-1) \sin \phi_{il} \sin \theta_{il} + (N-1) \cos \theta_{il})} \right]^T, \quad (4)$$

where $0 \leq p \leq M-1$ and $0 \leq q \leq N-1$ are the antenna indices in the 2D plane, and d and λ denote the antenna spacing and signal wavelength, respectively.

Then, the hybrid precoding and combining problem can be viewed as the two approximation problems as follows [7]

$$\begin{aligned} \min_{\mathbf{F}_S, \mathbf{F}_{\text{RF}}, \mathbf{F}_{\text{BB}}} \quad & \|\mathbf{F}_{\text{opt}} - \mathbf{F}_S \mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}}\|_F^2 \\ \text{s.t.} \quad & (\mathbf{F}_S)_{m,n} \in \{0, 1\}, \quad m = 1, 2, \dots, N_t, \quad n = 1, 2, \dots, N_c, \\ & |(\mathbf{F}_{\text{RF}})_{i,j}| = 1, \quad i = 1, 2, \dots, N_c, \quad j = 1, 2, \dots, N_{\text{RF}}^t, \\ & \|\mathbf{F}_S \mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}}\|_F^2 = N_s, \end{aligned} \quad (5)$$

$$\begin{aligned} \min_{\mathbf{W}_S, \mathbf{W}_{\text{RF}}, \mathbf{W}_{\text{BB}}} \quad & \|\mathbf{W}_{\text{opt}} - \mathbf{W}_S \mathbf{W}_{\text{RF}} \mathbf{W}_{\text{BB}}\|_F^2 \\ \text{s.t.} \quad & (\mathbf{W}_S)_{m,n} \in \{0, 1\}, \quad m = 1, 2, \dots, N_r, \quad n = 1, 2, \dots, N_c, \\ & |(\mathbf{W}_{\text{RF}})_{i,j}| = 1, \quad i = 1, 2, \dots, N_c, \quad j = 1, 2, \dots, N_{\text{RF}}^r. \end{aligned} \quad (6)$$

Due to similar mathematical formulations to Eqs. (5) and (6), the remaining of this paper will mainly focus on the precoding problem in Eq. (5), and the combining problem can be tackled via a similar approach. According to the MIMO theory, the $N_t \times N_s$ optimal fully digital precoding matrix can be given by $\mathbf{F}_{\text{opt}} = \mathbf{V}_{:,1:N_s}$, where $\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$ is the singular value decomposition (SVD) of channel matrix \mathbf{H} .

However, to maximize spectral efficiency requires a joint optimization over \mathbf{F}_S , \mathbf{F}_{RF} and \mathbf{F}_{BB} , which is intractable due to the binary constraint of \mathbf{F}_S and the unit modulus constraint of \mathbf{F}_{RF} . To solve these issues, the problem in Eq. (5) will be analyzed by a two-stage algorithm as follows.

3. ALGORITHM FORMULATION

In this section, an effective two-stage hybrid precoding algorithm is described in detail to tackle the problem in Eq. (5). Specifically, to derive the solution of binary switch network matrix \mathbf{F}_S , the product of \mathbf{F}_{RF} and \mathbf{F}_{BB} is viewed as a whole matrix \mathbf{A} firstly, and then, the decomposition problem of \mathbf{A} into \mathbf{F}_{RF} and \mathbf{F}_{BB} is considered. Therefore, the problem in Eq. (5) can be regarded as the combination of subproblems $\mathcal{P}1$ and $\mathcal{P}2$ as follows

$$\begin{aligned} \mathcal{P}1 : \quad & \left\{ \mathbf{F}_S^{\text{opt}}, \mathbf{A}^{\text{opt}} \right\} = \arg \min_{\mathbf{F}_S, \mathbf{A}} \|\mathbf{F}_{\text{opt}} - \mathbf{F}_S \mathbf{A}\|_F^2 \\ & \text{s.t.} \quad (\mathbf{F}_S)_{m,n} \in \{0, 1\} \\ & \quad \|\mathbf{F}_S \mathbf{A}\|_F^2 = N_s, \end{aligned} \quad (7)$$

$$\begin{aligned} \mathcal{P}2 : \quad & \left\{ \mathbf{F}_{\text{RF}}^{\text{opt}}, \mathbf{F}_{\text{BB}}^{\text{opt}} \right\} = \arg \min_{\mathbf{F}_{\text{RF}}, \mathbf{F}_{\text{BB}}} \|\mathbf{A}^{\text{opt}} - \mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}}\|_F^2 \\ & \text{s.t.} \quad |(\mathbf{F}_{\text{RF}})_{i,j}| = 1 \\ & \quad \|\mathbf{F}_S^{\text{opt}} \mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}}\|_F^2 = N_s. \end{aligned} \quad (8)$$

3.1. An Upper Bound for the Objective

Let's focus on the subproblem $\mathcal{P}1$ firstly. Note that the columns of the optimal fully digital precoding matrix \mathbf{F}_{opt} are mutually orthogonal in order to mitigate the inter-stream interference [17, 23]. Inspired by it, we impose a similar constraint on matrix \mathbf{A} , i.e.,

$$\mathbf{A}^H \mathbf{A} = \gamma^2 \mathbf{B}^H \mathbf{B} = \gamma^2 \mathbf{I}_{N_s} \quad (9)$$

where $\mathbf{A} = \gamma\mathbf{B}$, γ represents a scaling factor, and \mathbf{B} stands for a semi-unitary matrix with the same dimension as \mathbf{A} . Thus, the objective function of $\mathcal{P}1$ can be rewritten as

$$\begin{aligned} \|\mathbf{F}_{\text{opt}} - \mathbf{F}_S \mathbf{A}\|_F^2 &= \text{tr}(\mathbf{F}_{\text{opt}}^H \mathbf{F}_{\text{opt}}) - \text{tr}(\mathbf{F}_{\text{opt}}^H \mathbf{F}_S \mathbf{A}) - \text{tr}(\mathbf{A}^H \mathbf{F}_S^H \mathbf{F}_{\text{opt}}) + \text{tr}(\mathbf{A}^H \mathbf{F}_S^H \mathbf{F}_S \mathbf{A}) \\ &= \|\mathbf{F}_{\text{opt}}\|_F^2 - 2\gamma \Re\{\text{tr}(\mathbf{B} \mathbf{F}_{\text{opt}}^H \mathbf{F}_S)\} + \gamma^2 \|\mathbf{F}_S \mathbf{B}\|_F^2. \end{aligned} \quad (10)$$

Unfortunately, it is still intractable to directly optimize Eq. (10). To overcome this issue, the Frobenius norm $\|\mathbf{F}_S \mathbf{B}\|_F^2$ is upper bounded by

$$\|\mathbf{F}_S \mathbf{B}\|_F^2 = \text{tr}(\mathbf{B}^H \mathbf{F}_S^H \mathbf{F}_S \mathbf{B}) = \text{tr} \left(\begin{bmatrix} \mathbf{I}_{N_s} & \\ & \mathbf{0} \end{bmatrix} \mathbf{K}^H \mathbf{F}_S^H \mathbf{F}_S \mathbf{K} \right) < \text{tr}(\mathbf{K}^H \mathbf{F}_S^H \mathbf{F}_S \mathbf{K}) = \|\mathbf{F}_S\|_F^2, \quad (11)$$

where $\mathbf{B} \mathbf{B}^H = \mathbf{K} \begin{bmatrix} \mathbf{I}_{N_s} & \\ & \mathbf{0} \end{bmatrix} \mathbf{K}^H$ is the SVD of $\mathbf{B} \mathbf{B}^H$. Hence, we can formulate the objective function as

$$\|\mathbf{F}_{\text{opt}}\|_F^2 - 2\gamma \Re\{\text{tr}(\mathbf{B} \mathbf{F}_{\text{opt}}^H \mathbf{F}_S)\} + \gamma^2 \|\mathbf{F}_S\|_F^2. \quad (12)$$

3.2. Alternating Minimization

By adopting the upper bound given by Eq. (12) as the new objective function, the subproblem $\mathcal{P}1$ can be further simplified as [23]

$$\begin{aligned} \min_{\gamma, \mathbf{F}_S, \mathbf{B}} \quad & \gamma^2 \|\mathbf{F}_S\|_F^2 - 2\gamma \Re\{\text{tr}(\mathbf{B} \mathbf{F}_{\text{opt}}^H \mathbf{F}_S)\} \\ \text{s.t.} \quad & (\mathbf{F}_S)_{m,n} \in \{0, 1\} \\ & \mathbf{B}^H \mathbf{B} = \mathbf{I}_{N_s}. \end{aligned} \quad (13)$$

When regarding \mathbf{F}_S and γ as being fixed, the problem in Eq. (13) can be recast as

$$\max_{\mathbf{B}} \quad \gamma \Re\{\text{tr}(\mathbf{B} \mathbf{F}_{\text{opt}}^H \mathbf{F}_S)\} \quad \text{s.t.} \quad \mathbf{B}^H \mathbf{B} = \mathbf{I}_{N_s}. \quad (14)$$

Obviously, the objective function in Eq. (14) only has one optimization variable \mathbf{B} , which can be reformulated by the dual norm [24], i.e.,

$$\gamma \Re\{\text{tr}(\mathbf{B} \mathbf{F}_{\text{opt}}^H \mathbf{F}_S)\} \leq |\text{tr}(\gamma \mathbf{B} \mathbf{F}_{\text{opt}}^H \mathbf{F}_S)| \stackrel{(a)}{\leq} \|\mathbf{B}\|_{\infty} \|\gamma \mathbf{F}_{\text{opt}}^H \mathbf{F}_S\|_1 = \|\gamma \mathbf{F}_{\text{opt}}^H \mathbf{F}_S\|_1 = \sum_{i=1}^{N_s} \sigma_i, \quad (15)$$

where (a) obeys the Hölder's inequality, and $\|\cdot\|_1$ and $\|\cdot\|_{\infty}$ represent the one and infinite Schatten norms, respectively. The equality holds only when

$$\mathbf{B} = \mathbf{V}_1 \mathbf{U}^H, \quad (16)$$

where $\gamma \mathbf{F}_{\text{opt}}^H \mathbf{F}_S = \mathbf{U} \mathbf{\Sigma} \mathbf{V}_1^H$ follows the SVD, in which $\mathbf{\Sigma}$ denotes a diagonal matrix with non-zero singular values $\sigma_1, \dots, \sigma_{N_s}$.

With fixed \mathbf{B} , we add a constant term $\|\Re\{\mathbf{F}_{\text{opt}} \mathbf{B}^H\}\|_F^2$ to the objective function in Eq. (13), and the optimization problem can be recast as

$$\min_{\gamma, \mathbf{F}_S} \quad \|\Re\{\mathbf{F}_{\text{opt}} \mathbf{B}^H\} - \gamma \mathbf{F}_S\|_F^2 \quad \text{s.t.} \quad (\mathbf{F}_S)_{m,n} \in \{0, 1\}. \quad (17)$$

Considering the binary constraint of the switch network matrix \mathbf{F}_S , it is easy to know that $(\mathbf{F}_S)_{m,n} = 1$ if the corresponding (m, n) th entry in $\Re\{\mathbf{F}_{\text{opt}} \mathbf{B}^H\}$ is closer to γ than 0, and $(\mathbf{F}_S)_{m,n} = 0$ otherwise. Thus, the problem in Eq. (17) can be equivalently considered by [23]

$$\begin{aligned} \min_{\gamma, \mathbf{s}} \quad & \|\mathbf{x} - \gamma \mathbf{s}\|_2^2 \\ \text{s.t.} \quad & \mathbf{s} \in \{0, 1\}, \end{aligned} \quad (18)$$

where $\mathbf{x} = \text{vec}\{\Re\{\mathbf{F}_{\text{opt}} \mathbf{B}^H\}\}$, and $\mathbf{s} = [s_1, s_2, \dots, s_k]^T = \text{vec}\{\mathbf{F}_S\}$, in which $k = N_t N_c$.

Sort all the entries of \mathbf{x} in ascending order, i.e., $\tilde{\mathbf{x}} = [\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_k]^T$, in which $\tilde{x}_1 \leq \tilde{x}_2 \leq \dots \leq \tilde{x}_k$. Therefore, the objective function in Eq. (17) can be rewritten as

$$f(\gamma) = \|\tilde{\mathbf{x}} - \gamma \mathbf{s}\|_2^2 = \begin{cases} \sum_{j=1}^i (\tilde{x}_j - \gamma)^2 + \sum_{j=i+1}^k \tilde{x}_j^2, & \gamma < 0 \quad \text{and} \quad \frac{\gamma}{2} \in \mathcal{I}_i \\ \sum_{j=1}^i \tilde{x}_j^2 + \sum_{j=i+1}^k (\tilde{x}_j - \gamma)^2, & \gamma > 0 \quad \text{and} \quad \frac{\gamma}{2} \in \mathcal{I}_i \end{cases}, \quad (19)$$

$$= \begin{cases} i\gamma^2 - 2\sum_{j=1}^i \tilde{x}_j\gamma + \sum_{j=1}^i \tilde{x}_j^2, & \gamma < 0 \quad \text{and} \quad \gamma \in \mathcal{R}_i \\ (k-i)\gamma^2 - 2\sum_{j=i+1}^k \tilde{x}_j\gamma + \sum_{j=1}^k \tilde{x}_j^2, & \gamma > 0 \quad \text{and} \quad \gamma \in \mathcal{R}_i, \end{cases}$$

where all the entries split the line into $k+1$ intervals $\{\mathcal{I}_i\}_{i=1}^k$, such that $\mathcal{I}_i = [\tilde{x}_i, \tilde{x}_{i+1}]$. Basically, $f(\gamma)$ is viewed as a quadratic function within each interval $\mathcal{R}_i = [2\tilde{x}_i, 2\tilde{x}_{i+1}]$. Hence, the optimal solution in Eq. (17) is given by

$$\gamma^{\text{opt}} = \arg \min_{\{\tilde{x}_i, \tilde{x}_i\}_{i=1}^k} \{f(2\tilde{x}_i), f(\tilde{x}_i)\}, \quad (20)$$

$$\mathbf{F}_S^{\text{opt}} = \begin{cases} \mathbb{1}\{\Re\{\mathbf{F}_{\text{opt}}\mathbf{B}^H\} > \frac{\gamma}{2}\mathbf{1}_{\mathbf{N}_t \times \mathbf{N}_c}\}, & \gamma > 0 \\ \mathbb{1}\{\Re\{\mathbf{F}_{\text{opt}}\mathbf{B}^H\} < \frac{\gamma}{2}\mathbf{1}_{\mathbf{N}_t \times \mathbf{N}_c}\}, & \gamma < 0, \end{cases} \quad (21)$$

where $\mathbb{1}(\cdot)$ stands for the indicator function, and $\mathbf{1}_{\mathbf{m} \times \mathbf{n}}$ represents an $m \times n$ matrix with all matrix elements equal to one. In addition, \tilde{x}_i is the i th smallest entry in \mathbf{x} , and

$$\bar{x}_i = \begin{cases} \frac{\sum_{j=1}^i \tilde{x}_j}{i}, & \bar{x}_i < 0 \quad \text{and} \quad \bar{x}_i \in \mathcal{R}_i \\ \frac{\sum_{j=i+1}^k \tilde{x}_j}{k-i}, & \bar{x}_i > 0 \quad \text{and} \quad \bar{x}_i \in \mathcal{R}_i \\ +\infty, & \text{otherwise.} \end{cases} \quad (22)$$

This means that γ^{opt} can only be obtained either at the endpoint of the intervals $\{\mathcal{R}_i\}_{i=1}^k$, i.e., $\{2\tilde{x}_i\}_{i=1}^k$, or at the vertexes of the parabolas, i.e., $\{\tilde{x}_i\}_{i=1}^k$, if they fall into the intervals. In other words, γ^{opt} can only be given by comparing the values of the objective function of all the endpoints and vertexes, as indicated in Eq. (20). And the optimal switch matrix $\mathbf{F}_S^{\text{opt}}$ can be obtained by a closed-form solution of Eq. (21), thereby the superiority and benefits of the surrogate objective function given by Eq. (13) are verified.

Now, the subproblem $\mathcal{P}1$ has been solved, and $\mathbf{F}_S^{\text{opt}}$ has been given. Afterwards, let's focus on the subproblem $\mathcal{P}2$, which is an optimization problem of \mathbf{F}_{RF} and \mathbf{F}_{BB} . This kind of problem has been extensively studied, such as the OMP-based sparse precoding algorithm [7] and PE-AltMin algorithm [17]. In this paper, we chose the PE-AltMin algorithm [17] to solve as it can achieve better performance with relatively low computational cost. Based on the above discussion, with the closed-form solutions derived in Eqs. (16), (20) and (21), the proposed two-stage algorithm based on switch network in mmWave systems can be summarized and presented with pseudo-codes in Table 1.

4. SIMULATION RESULTS

In this section, the simulation results are obtained to evaluate the presented algorithm for single-user mmWave MIMO communication systems as shown in Fig. 1. The UPAs of the transmitter

Table 1. Pseudo-code of proposed hybrid precoding algorithm.

Proposed Two-Stage Hybrid Precoding Algorithm
Input: \mathbf{F}_{opt}
Initializations:
$\mathbf{F}_{\text{RF}}^{(0)}$ and $\mathbf{F}_{\text{BB}}^{(0)}$ are constructed with random initial values randomly, $t = 0$
repeat
1: Fix $\mathbf{B}^{(t)}$, optimize $\gamma^{(t)}$ and $\mathbf{F}_{\text{S}}^{(t)}$ according to (20) and (21), respectively;
2: Fix $\gamma^{(t)}$ and $\mathbf{F}_{\text{S}}^{(t)}$, update $\mathbf{B}^{(t)}$ with (16);
3: $\mathbf{A}^{(t)} = \gamma^{(t)}\mathbf{B}^{(t)}$;
4: Fix $\mathbf{F}_{\text{RF}}^{(t)}$, compute the SVD: $\mathbf{A}^{(t)H}\mathbf{F}_{\text{RF}}^{(t)} = \mathbf{U}_1^{(t)}\mathbf{S}^{(t)}\mathbf{V}_2^{(t)H}$;
5: $\mathbf{F}_{\text{DD}}^{(t)} = \mathbf{V}_2^{(t)}\mathbf{U}_1^{(t)H}$;
6: Fix $\mathbf{F}_{\text{DD}}^{(t)}$, and $\arg\{\mathbf{F}_{\text{RF}}^{(t+1)}\} = \arg(\mathbf{A}^{(t)}\mathbf{F}_{\text{DD}}^{(t)H})$;
7: $t = t + 1$;
until convergence;
9: $\mathbf{F}_{\text{BB}} = \frac{\sqrt{N_s}}{\ \mathbf{F}_{\text{RF}}\mathbf{F}_{\text{DD}}\ _F}\mathbf{F}_{\text{DD}}$.
Output: $\mathbf{F}_{\text{S}}, \mathbf{F}_{\text{RF}}, \mathbf{F}_{\text{BB}}$

and receiver are equipped with 12×12 and 4×4 antenna elements, respectively. Without loss of generality, it is assumed that the same numbers of RF chains are set at the transmitter and receiver, i.e., $N_{\text{RF}}^t = N_{\text{RF}}^r = N_{\text{RF}}$.

The mmWave propagation channel consists of totally $L = N_{\text{cl}}N_{\text{ray}} = 40$ paths, which are divided into 5 clusters \mathcal{C}_i , $i = 1, 2, \dots, 5$, and each contains 8 rays. When the signal leaves the transmitter, the average azimuth/elevation AoD of each cluster, i.e., $\phi_{\mathcal{C}_i}^t = \frac{1}{8} \sum_{l \in \mathcal{C}_i} \phi_l^t$, $\theta_{\mathcal{C}_i}^t = \frac{1}{8} \sum_{l \in \mathcal{C}_i} \theta_l^t$ ($i = 1, 2, \dots, 5$), is uniformly distributed within $(0, 2\pi)$. The azimuth/elevation AoD of rays in each cluster follows Laplace distribution, i.e., $\phi_{l, l \in \mathcal{C}_i}^t \sim \mathcal{L}(\mu_{\phi, \mathcal{C}_i}^t, b_{\phi, \mathcal{C}_i}^t)$, $\theta_{l, l \in \mathcal{C}_i}^t \sim \mathcal{L}(\mu_{\theta, \mathcal{C}_i}^t, b_{\theta, \mathcal{C}_i}^t)$, where $\mu_{\phi, \mathcal{C}_i}^t = \phi_{\mathcal{C}_i}^t$, $\mu_{\theta, \mathcal{C}_i}^t = \theta_{\mathcal{C}_i}^t$ are the location parameters, and $b_{\phi, \mathcal{C}_i}^t = b_{\theta, \mathcal{C}_i}^t = 10^\circ$ are the scale parameters. The statistic properties of azimuth/elevation AoA are the same as the azimuth/elevation AoD. The complex gain of each path α_l follows the standard complex normal distribution $\mathcal{CN}(0, 1)$. All algorithms are used to solve for the whole hybrid precoding in this section. It is assumed that the perfect CSI is known at both the transmitter and receiver instantaneously. All of the simulation results are averaged over 1000 random channel realizations.

Figure 2 shows the spectral efficiencies achieved by the several precoding algorithms when SNR = 0 dB, $N_s = N_{\text{RF}} = 4$, and $N_c = 30$. As can be seen from Fig. 2, with the increase of SNR, the performance of all algorithms gradually improves. Obviously, although the number of phase shifters is small, i.e., $N_{\text{PS}} = N_c N_{\text{RF}} = 120$, the proposed hybrid precoding architecture achieves higher spectral efficiency than the OMP-based sparse precoding algorithm for FC-PSs with $N_{\text{PS}} = N_t N_{\text{RF}} = 576$ [7] and the SDR_AltMin algorithm for PC-PSs with $N_{\text{PS}} = N_t = 144$ [17], while incurring small loss in the system performance compared to the optimal fully digital precoder. Besides, the proposed algorithm far outperforms the antenna selection structure A3 based on switch network (SW-AS) presented in [20].

Figure 3 plots the spectral efficiency and energy efficiency comparison against the number of N_c with SNR = 0 dB, $N_s = N_{\text{RF}} = 4$. When taking energy efficiency into consideration, the energy efficiency η can be defined as

$$\eta = \frac{R}{P} = \frac{R}{P_t + P_{\text{total}}}, \quad (23)$$

where the unit of η is bits/Hz/J, and $P_t = \|\mathbf{F}_{\text{S}}\mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}\|_F^2 = N_s$ is the transmitted power. Then, the total power consumption P_{total} is given by $P_{\text{total}} = N_{\text{RF}}P_{\text{RF}} + N_{\text{PS}}P_{\text{PS}} + N_{\text{SW}}P_{\text{SW}}$, in which P_{RF} ,

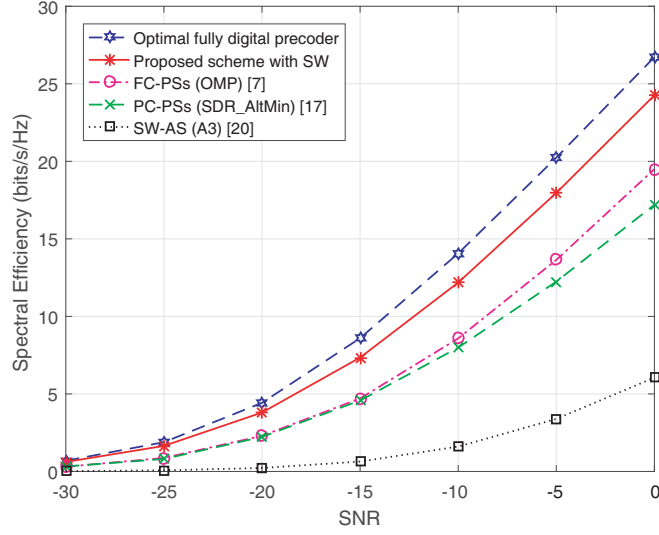


Figure 2. Spectral efficiency for various SNR.

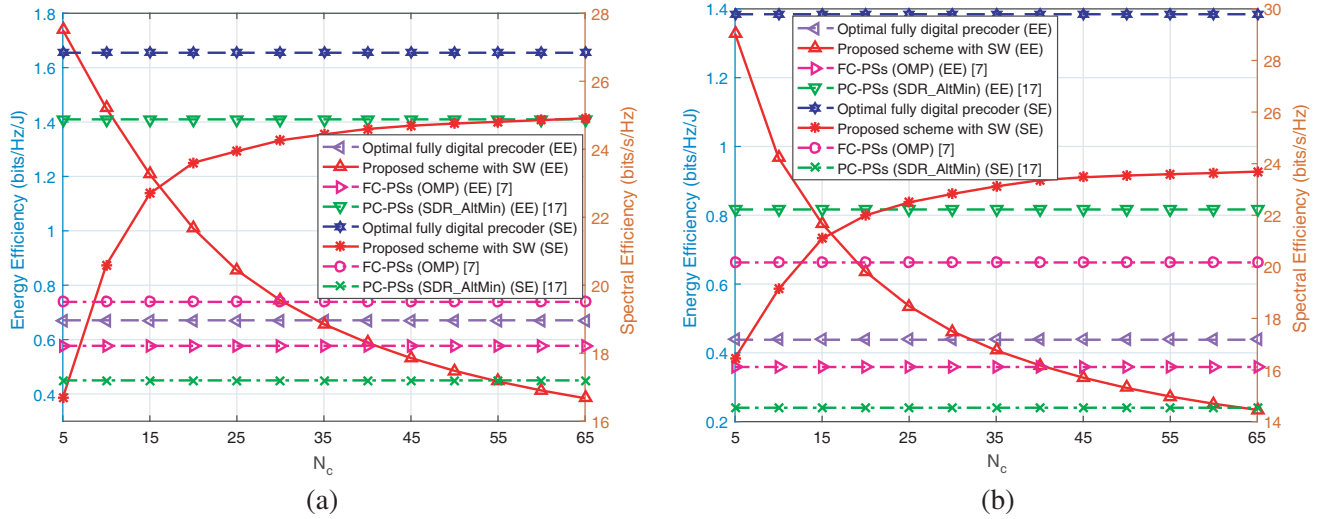


Figure 3. Performance comparison for various N_c . (a) $N_t = 12 \times 12$, (b) $N_t = 16 \times 16$.

P_{PS} and P_{SW} denote the power consumed by RF chains, phase shifters and switches, respectively, and N_{RF} , N_{PS} , and N_{SW} stand for the numbers of required RF chains, phase shifters, and switches in the transmitter, respectively. In this paper, we use the practical values $P_{RF} = 250$ mW [8], $P_{PS} = 50$ mW [23] and $P_{SW} = 5$ mW [23]. It can be demonstrated that the proposed hybrid precoding scheme based on switch network provides substantial energy efficiency gain over the optimal fully digital precoder, while only imposes an acceptable spectral efficiency loss. It shows that the energy efficiency decreases with increasing number of N_c for the proposed hybrid precoding implementation. The reason is that the diversity of the RF chains cannot compensate the increasing power consumption of the phase shifters and switches. Also, the proposed precoding scheme can achieve higher energy efficiency than the FC-PSs [7] when $N_c < 40$ in Fig. 3(a) and $N_c < 37$ in Fig. 3(b). For spectral efficiency, the performance given by the proposed algorithm increases with the increase of N_c , which almost saturates when $N_c > 45$.

It is noteworthy that $N_c = 8$ in the proposed scheme is enough to achieve a comparable spectral efficiency as the FC-PSs in Fig. 3(a). Hence, we compare the power consumption of the proposed scheme with the FC-PSs [7] and PC-PSs [17] when $N_c = 8$ in Table 2. Explicitly, to achieve the equivalent

Table 2. Power consumption of hybrid precoder for different hybrid precoding algorithms.

	RF chain		Phase shifter		Switch		Total Power P_{total}
	Number N_{RF}^t	Power P_{RF}	Number N_{PS}	Power P_{PS}	Number N_{SW}	Power P_{SW}	
FC-PSs (OMP) [7]	4	250 mW	576	50 mW	0	5 mW	29.8 W
FC-PSs (SDR_AltMin) [17]	4	250 mW	144	50 mW	0	5 mW	8.2 W
Proposed scheme with SW	4	250 mW	32	50 mW	1152	5 mW	8.36 W

spectral efficiency, the power consumption of FC-PSs is almost 3 times more than the proposed scheme. However, for approximately equal power consumption when $N_c = 8$, the spectral efficiency between the proposed scheme and PC-PSs has great gap. As can be seen from Fig. 3(b), the proposed scheme can achieve an approximately spectral efficiency as the FC-PSs when $N_c = 13$. The total power consumption of the PC-PSs is $P_{total} = 52.2$ W, and the total power consumption of the proposed scheme is $P_{total} = 20.24$ W, which demonstrates the attractive benefits of the proposed algorithm in this paper in terms of energy efficiency, even if the antenna array is large enough. For Fig. 3(b), the proposed scheme is enough to achieve a comparable spectral efficiency as the FC-PSs when $N_c = 13$.

Figure 4 compares the performance of different hybrid precoding algorithms for various N_{RF} when $N_s = N_{RF}$. The simulation parameters are the same as those in Fig. 2. It can be observed that for the single-stream communication system ($N_{RF} = 1$), the difference between the proposed scheme and optimal fully digital precoder is small, and the proposed hybrid precoding scheme outperforms the FC-PSs in [7] and the PC-PSs in [17]. However, for multiple-streams scenario ($N_{RF} > 1$), the proposed algorithm can consistently offer about 5 bits/s/Hz and 8 bits/s/Hz performance gain compared with the OMP-based sparse precoding algorithm for FC-PSs and the SDR_AltMin algorithm for PC-PSs, respectively. In addition, the SW-AS for A3 in [20] suffers from serious performance loss. Based on the above discussion, we verify the superiority of the proposed scheme under various system settings.

Figure 5 illustrates the spectral efficiencies achieved by different algorithms as the number of propagation paths L when SNR = 0 dB, $N_s = N_{RF} = 4$. It is observed that the spectral efficiency gap between the proposed algorithm and optimal fully digital precoder is small, and this gap is less than 2 bits/s/Hz. In addition, the proposed algorithm always outperforms the other comparative algorithms

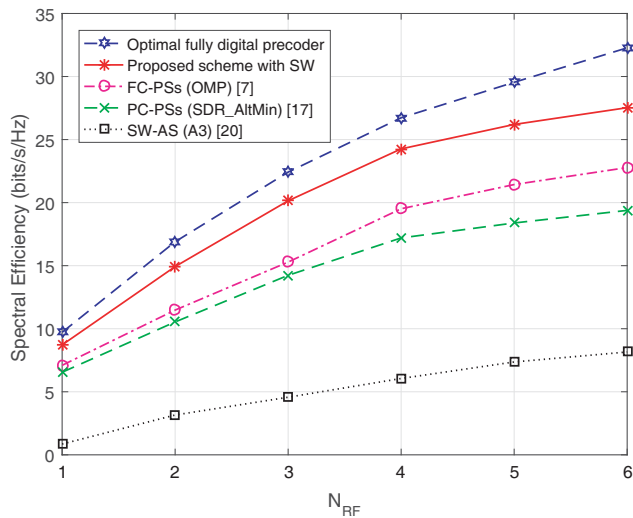


Figure 4. Spectral efficiency for various N_{RF} .

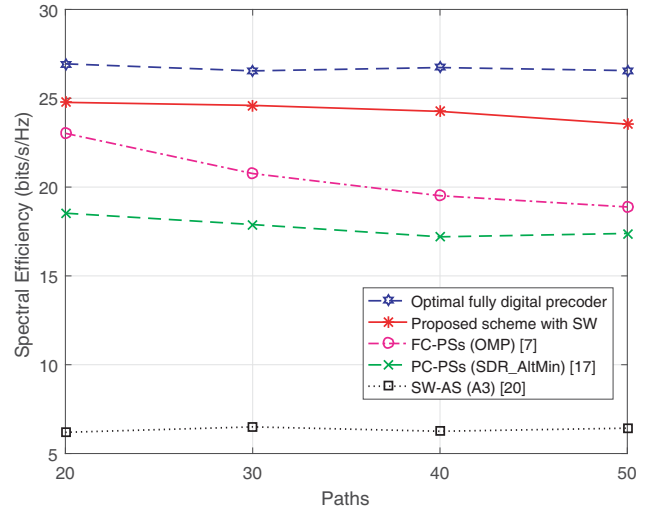


Figure 5. Spectral efficiency for various paths.

and offers a steady gain over the whole range of L , while the OMP-based algorithm [7] experiences serious performance degradation. Especially when $L \geq 40$, up to about 5 bits/s/Hz performance gain can be provided by the proposed algorithm. In a nutshell, the proposed algorithm is able to consistently offer a significant performance gain whether the channel is sparse or relatively rich scattered.

5. CONCLUSION

In this paper, a hardware-efficient architecture for hybrid precoding is considered for single-user mmWave MIMO communication systems. The new implementation introduces a switch network to dynamically connect the phase shifters and antennas, which can significantly reduce the power consumption of the hybrid precoder. Then a two-stage hybrid precoding algorithm is proposed to determine the digital precoding matrix, RF precoding matrix and switch network matrix. Simulation results show that the presented algorithm can not only improve the spectral efficiency effectively of mmWave MIMO communication systems but also achieve higher energy efficiency with much fewer phase shifters than the existing works.

ACKNOWLEDGMENT

This work was supported by the Natural Science Foundation of Hebei Province (Grant No. F2016501139), the Fundamental Research Funds for the Central Universities (Grant No. N172302002 and No. N162304002), and the National Natural Science Foundation of China (Grant No. 61501102).

REFERENCES

1. Han, S., C.-L. I, Z. Xu, and C. Rowell, "Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G," *IEEE Communications Magazine*, Vol. 53, No. 1, 186–194, 2015.
2. Kutty, S. and D. Sen, "Beamforming for millimeter wave communications: An inclusive survey," *IEEE Communications Surveys & Tutorials*, Vol. 18, No. 2, 949–973, 2016.
3. Bai, X., F. Liu, and R. Du, "An alternating iterative hybrid beamforming method for millimeter wave large-scale antenna arrays," *2017 Progress In Electromagnetics Research Symposium — Fall (PIERS — FALL)*, 2769–2776, Singapore, Nov. 19–22, 2017.
4. Heath, R. W., N. González-Prelcic, Jr., S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE Journal of Selected Topics in Signal Processing*, Vol. 10, No. 3, 436–452, 2016.
5. Rajashekar, R. and L. Hanzo, "Iterative matrix decomposition aided block diagonalization for mm-wave multiuser MIMO systems," *IEEE Transactions on Wireless Communications*, Vol. 16, No. 3, 1372–1384, 2017.
6. Liu, F., R. Du, X. Kan, and X. Wang, "W-LS-IR algorithm for hybrid precoding in wideband millimeter wave MIMO systems," *Progress in Electromagnetics Research M*, Vol. 72, 187–195, 2018.
7. Ayach, O. E., S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Transactions Wireless Communications*, Vol. 13, No. 3, 1499–1513, 2014.
8. Gao, X., L. Dai, S. Han, C.-L. I, and R. W. Heath, "Energy-efficient hybrid analog and digital precoding for mmwave MIMO systems with large antenna arrays," *IEEE Journal on Selected Areas in Communications*, Vol. 34, No. 4, 998–1009, 2016.
9. Alkhateeb, A., O. E. Ayach, G. Leus, and R. W. Heath, "Hybrid precoding for millimeter wave cellular systems with partial channel knowledge," *IEEE Information Theory and Applications Workshop*, 1–5, 2013.
10. Ni, W. and X. Dong, "Hybrid block diagonalization for massive multiuser MIMO systems," *IEEE Transactions on Communications*, Vol. 64, No. 1, 201–211, 2016.

11. Méndez-Rial, R., C. Rusu, N. González-Prelcic, and R. W. Heath, "Dictionary-free hybrid precoders and combiners for mmwave MIMO systems," *IEEE International Workshop on Signal Processing Advances in Wireless Communications*, 151–155, 2016.
12. Chen, C. E., "An iterative hybrid transceiver design algorithm for millimeter wave MIMO systems," *IEEE Wireless Communications Letters*, Vol. 4, No. 3, 285–288, 2015.
13. Sohrabi, F. and Y. Wei, "Hybrid digital and analog beamforming design for large-scale antenna arrays," *IEEE Journal on Selected Topics in Signal Processing*, Vol. 10, No. 3, 501–513, 2016.
14. Sohrabi, F. and Y. Wei, "Hybrid digital and analog beamforming design for large-scale MIMO systems," *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2929–2933, 2015.
15. Singh, J. and S. Ramakrishna, "On the feasibility of codebook-based beamforming in millimeter wave systems with multiple antenna arrays," *IEEE Transactions on Wireless Communications*, Vol. 14, No. 5, 2670–2683, 2015.
16. Kim, C., T. Kim, and J.-Y. Seol, "Multi-beam transmission diversity with hybrid beamforming for MIMO-OFDM systems," *IEEE Globecom Workshops*, 61–65, 2013.
17. Yu, X., J. C. Shen, J. Zhang, and K. B. Letaief, "Alternating minimization algorithms for hybrid precoding in millimeter wave MIMO systems," *IEEE Journal of Selected Topics in Signal Processing*, Vol. 10, No. 3, 485–500, 2016.
18. Park, S., A. Alkhateeb, and R. W. Heath, "Dynamic subarrays for hybrid precoding in wideband mmWave MIMO systems," *IEEE Transactions on Wireless Communications*, Vol. 16, No. 5, 2907–2920, 2017.
19. Yu, X., J. Zhang, and K. B. Letaief, "Partially-connected hybrid precoding in mm-Wave systems with dynamic phase shifter networks," *IEEE International Workshop on Signal Processing Advances in Wireless Communications*, 129–133, 2017.
20. Méndez-Rial, R., C. Rusu, N. González-Prelcic, and R. W. Heath, "Hybrid MIMO architectures for millimeter wave communications: Phase Shifters or Switches?," *IEEE Access*, Vol. 4, 247–267, 2015.
21. Alkhateeb, A., O. E. Ayach, G. Leus, and R. W. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE Journal of Selected Topics in Signal Processing*, Vol. 8, No. 5, 831–846, 2017.
22. Balanis, C., *Antenna Theory*, Wiley, 1997.
23. Yu, X., J. Zhang, and K. B. Letaief, "A hardware-efficient analog network structure for hybrid precoding in millimeter wave systems," *IEEE Journal of Selected Topics in Signal Processing*, Vol. 12, No. 2, 282–297, 2018.
24. Horn, R. A. and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, U.K., 2012.