SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •
Special Focus on Millimeter Wave Communications Techniques and Devices for 5G

August 2017, Vol. 60 080304:1–080304:10
doi: 10.1007/s11432-016-0640-5

Energy-efficient Butler-matrix-based hybrid beamforming for multiuser mmWave MIMO system

Jiahui LI, Limin XIAO, Xibin XU, Xin SU & Shidong ZHOU*

State Key Laboratory on Microwave and Digital Communications,
Tsinghua National Laboratory for Information Science and Technology,
Department of Electronic Engineering, Tsinghua University, Beijing 100084, *China*

Received September 27, 2016; accepted December 20, 2016; published online May 24, 2017

Abstract In order to reduce the cost and power consumption of radio frequency (RF) chains in a millimeter-wave (mmWave) multiple-input multiple-output (MIMO) system, hybrid analog/digital beamforming (HBF) can be utilized to reduce the number of RF chains. The HBF consists of an analog beamforming (ABF) stage and a digital beamforming (DBF) stage. The ABF is always realized by using phase shifters and the DBF is done in a low-dimensional digital domain. However, phase shifters have several drawbacks, such as high power consumption and inconsistency of insertion loss. In this paper, we propose an energy-efficient HBF structure to handle these problems, which utilizes the Butler phase shifting matrix in the ABF stage. With the Butler-matrix-based ABF, several fixed beam directions can be obtained, and the best beam directions of different Butler matrices can be chosen by using exhaustive search. To reduce the high complexity of exhaustive search, we further provide a low complexity HBF algorithm. Simulations under the conditions of perfect channel state information (CSI) and estimated CSI verify the effectiveness of our proposed Butler-matrix-based HBF structure and related algorithms.

Keywords Butler matrix, energy-efficient, hybrid beamforming, MIMO, mmWave, multiuser

Citation Li J H, Xiao L M, Xu X B, et al. Energy-efficient Butler-matrix-based hybrid beamforming for multiuser mmWave MIMO system. Sci China Inf Sci, 2017, 60(8): 080304, doi: 10.1007/s11432-016-0640-5

1 Introduction

MmWave frequency band has attracted much attention recently, since it can provide much more spectrum resource than the current cellular spectrum below 3 GHz to meet the extensive needs in the coming 5G era [1]. As proved by measurement campaigns shown in [2], the mmWave signals suffer from severe path loss, penetration loss, rain effect and atmospheric absorption. Therefore, a large number of antennas should be configured at the base station (BS) in a mmWave communication system, which can provide a high array gain to compensate for the severe signal attenuation. However, it is not cost-efficient and energy-efficient to equipped each antenna with an independent RF chain [3]. Therefore, HBF is proposed to reduce the number of RF chains and can provide good performance [4–7].

The HBF consists of an ABF stage and a DBF stage [4–7]. Take downlink transmission for example, the basic idea of HBF is to first process the transmitted signals in a low-dimensional digital domain, and

* Corresponding author (email: zhousd@mail.tsinghua.edu.cn)

then feed the post-processed signals through an analog phase shifting network, which is realized with phase shifters and is called the phase-shifter-based structure in the following part. Finally, the signals are sent out by antennas whose number is much larger than that of the RF chains.

There are mainly two different system architectures for HBF: the fully-connected architecture [4–6] and the sub-connected architecture [7]. In the fully-connected architecture, each RF chain is connected to all the antennas, making it too difficult to be realized due to the high hardware complexity and energy consumption caused by the huge phase shifting network [7]. While in the sub-connected architecture, each RF chain is only connected to the antennas in different sub-arrays, which makes it easier to be implemented in a real system. Therefore, we will focus on the more realistic sub-connected architecture. All the HBF systems considered in this paper are assumed to use the sub-connected architecture.

There exist many algorithms that can be utilized for the phase-shifter-based HBF structure, such as the SIC-based algorithm [7] and the beam steering algorithm [8], which can provide good system performance. However, as introduced in [9], there exist many problems on phase shifters. The phase shifters are partitioned into two categories: active phase shifters and passive phase shifters [9]. For the active phase shifters, power consumption is high and nonlinearity problems will result in a high noise figure. For the passive phase shifters, the dependency of the insertion loss on the phase shifting value will cause negative impact on system performance and it is hard to do calibration on this effect.

The Butler-matrix-based phase shifting structure can be utilized to solve these problems [10], which has low power consumption and low insertion loss. It is realized by using a Butler phase shifting matrix, which consists of a fixed phase shifting network that can achieve several fixed beam directions containing in a predefined codebook. An RF switch array is configured to choose the beam direction. Therefore, the Butler-matrix-based phase shifting structure is very suitable for the HBF system, but related analysis has not been considered yet.

In this paper, we propose a Butler-matrix-based HBF structure for multiuser mmWave MIMO systems and derive several related HBF algorithms. The ABF is done by using the Butler phase shifting matrix. Firstly we obtain the performance upper bound of the Butler-matrix-based HBF structure, which is achieved by using exhaustive search to find the best beam directions in the codebook for all the Butler matrices. However, the searching process is of high complexity, which is not practical. So we propose a low complexity HBF algorithm to reduce the complexity. Simulations indicate that, compared with the full digital system and the phase-shifter-based system [7, 8], the proposed structure not only has an acceptable rate performance, but also performs more energy-efficiently. Moreover, the effect of channel estimation error is also analyzed, which proves the effectiveness of our proposed algorithm.

The rest of this paper is organized as follows. The system model is introduced in Section 2. The related HBF algorithms for the proposed structure are presented and analyzed in Section 3. Simulation results are shown in Section 4, and at last the conclusion is drawn in Section 5.

Notations. Vectors are column vectors and denoted by lower case boldface and italic: $\boldsymbol{x}$. Matrices are upper case boldface: $\boldsymbol{A}$. $\boldsymbol{I}_N$ is the size-N identity matrix. $\mathbf{0}_{N\times M}$ is the $N \times M$ zero matrix. The transpose and Hermitian transpose are denoted by $(\cdot)^{\mathrm{T}}$ and $(\cdot)^{\mathrm{H}}$, respectively. The 2-norm of a vector $\boldsymbol{x}$ is denoted by $||\boldsymbol{x}||_2$, and the Frobenius-norm of a matrix $\boldsymbol{X}$ is denoted by $||\boldsymbol{X}||_{\mathrm{F}}$. $\mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma})$ stands for the circular symmetric complex Gaussian distribution with mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$. $\mathrm{E}[\cdot]$ is the expectation operator. $\mathrm{diag}\{a_1, \ldots, a_N\}$ and $\mathrm{blkdiag}\{\boldsymbol{A}_1, \ldots, \boldsymbol{A}_N\}$ are the operators to obtain a diagonal matrix and a block diagonal matrix, respectively. For a set $\mathcal{S}$, the number of elements in it is denoted by $|\mathcal{S}|$.

2 System model

The downlink transmission of a single-cell multiuser mmWave MIMO system with Butler-matrix-based HBF structure is considered in this paper. In the considered system, each RF chain is connected to an independent sub-array. Each sub-array is a uniform linear array (ULA) that has M antennas. The BS is equipped with N RF chains and thus has a total of $N_{\mathrm{t}} = MN$ antennas, transmitting N_{s} ($N_{\mathrm{s}} \leqslant N$)

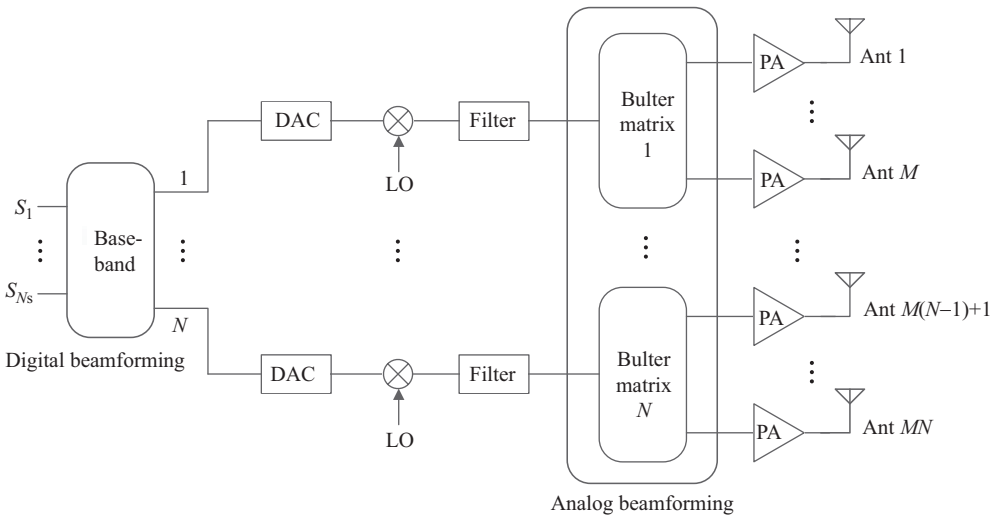

Figure 1 The Butler-matrix-based HBF structure for mmWave MIMO systems.

data streams at the same time. All the sub-arrays form a big ULA, as schematically shown in Figure 1. The BS is serving K users simultaneously, which are perfectly synchronized. The k-th ($k = 1, \ldots, K$) user has the same structure as the BS, where N_k RF chains are equipped and each RF chain is connected to a ULA with M_k antennas. A total of $N_{\mathrm{r}k} = M_k N_k$ antennas are configured for this user, receiving $N_{\mathrm{s}k}$ ($N_{\mathrm{s}k} \leqslant N_k, \sum_{k=1}^{K} N_{\mathrm{s}k} = N_{\mathrm{s}}$) data streams at the same time. Assume that the number of beam directions in the codebook is the same as that of the antennas in each sub-array, then we denote the codebooks of BS and users by $\mathcal{F}(|\mathcal{F}| = M)$ and $\mathcal{W}_k(|\mathcal{W}_k| = M_k)$, respectively. Transmission over flat-fading channels is assumed.

2.1 Channel model

To model the mmWave propagation environment, we consider using the clustered multipath channel model [11–13]. The narrowband fast fading channel matrix of the k-th user can be modeled as

$$\boldsymbol{H}_k = \sqrt{\frac{N_{\mathrm{t}} N_{\mathrm{r}k}}{N_{\mathrm{cl},k} N_{\mathrm{ray},k}}} \sum_{j=1}^{N_{\mathrm{cl},k}} \sum_{l=1}^{N_{\mathrm{ray},k}} \alpha_{jlk} \boldsymbol{a}_{N_{\mathrm{r}k}}(\theta^{\mathrm{r}}_{jlk}) \boldsymbol{a}_{N_{\mathrm{t}}}(\theta^{\mathrm{t}}_{jlk})^{\mathrm{H}}, \tag{1}$$

where $N_{\mathrm{cl},k}$ is the number of clusters, $N_{\mathrm{ray},k}$ is the number of propagation paths in each cluster, $\alpha_{jlk} \sim \mathcal{CN}(0, \delta^2_{jk})$ is the complex gain of the l-th path in the j-th cluster, $\theta^{\mathrm{r}}_{jlk}$ and $\theta^{\mathrm{t}}_{jlk}$ are the azimuth angle of arrival (AoA) and angle of departure (AoD) for this path, respectively. δ^2_{jk} represents the power of the j-th cluster, which satisfies $\frac{1}{N_{\mathrm{cl},k}} \sum_j \delta^2_{jk} = 1$. The AoAs $\theta^{\mathrm{r}}_{jlk}$ within the j-th cluster are assumed to be randomly distributed in according to a Laplacian distribution with mean cluster angle θ^{r}_{jk} and angular standard deviation $\delta_{\theta^{\mathrm{r}}}$. The AoDs $\theta^{\mathrm{t}}_{jlk}$ follow a similar distribution as the AoAs. As the ULA is utilized in the system, the steering vector $\boldsymbol{a}_N(\theta)$ is given by

$$\boldsymbol{a}_N(\theta) = \frac{1}{\sqrt{N}} \left[1, \mathrm{e}^{-\mathrm{j}2\pi\Delta\cos\theta}, \ldots, \mathrm{e}^{-\mathrm{j}2\pi(N-1)\Delta\cos\theta}\right]^{\mathrm{T}}, \tag{2}$$

where $\mathrm{j} = \sqrt{-1}$ and Δ is the antenna spacing normalized by carrier wavelength λ.

2.2 Signal model

2.2.1 *Downlink transmission*

The received signal vector at the k-th user is expressed as

$$\boldsymbol{y}_k = \frac{1}{\sqrt{\beta_k}} \boldsymbol{H}_k \boldsymbol{s} + \boldsymbol{n}_k, \tag{3}$$

where $\boldsymbol{H}_k \in \mathbb{C}^{N_{\mathrm{r}k} \times N_{\mathrm{t}}}$ is the downlink fast fading channel matrix with unit average power elements, β_k is the path loss between the BS and this user, and $\boldsymbol{n}_k \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I}_{N_{\mathrm{r}k}})$ is the normalized Gaussian noise vector at the receiver. $\boldsymbol{s}$ is the transmitted signal vector for all the users, given by

$$\boldsymbol{s} = \boldsymbol{F}_{\mathrm{RF}} \boldsymbol{F}_{\mathrm{BB}} \boldsymbol{x}, \tag{4}$$

where $\boldsymbol{F}_{\mathrm{RF}} \in \mathbb{C}^{N_{\mathrm{t}} \times N}$ is the analog precoding matrix, $\boldsymbol{F}_{\mathrm{BB}} = [\boldsymbol{F}_{\mathrm{BB},1}, \ldots, \boldsymbol{F}_{\mathrm{BB},K}] \in \mathbb{C}^{N \times N_{\mathrm{s}}}$ is the digital precoding matrix, and $\boldsymbol{x} = \left[\boldsymbol{x}_1^{\mathrm{T}}, \ldots, \boldsymbol{x}_K^{\mathrm{T}}\right]^{\mathrm{T}} \in \mathbb{C}^{N_{\mathrm{s}} \times 1}$ is the transmitted symbol vector. $\boldsymbol{F}_{\mathrm{BB},k} = \left[\boldsymbol{f}_{\mathrm{BB},k1}, \ldots, \boldsymbol{f}_{\mathrm{BB},kN_{\mathrm{s}k}}\right]$ is the digital precoding matrix for the k-th user. As the sub-connected architecture is assumed, we have

$$\boldsymbol{F}_{\mathrm{RF}} = \mathrm{blkdiag}\{\boldsymbol{f}_{\mathrm{RF},1}, \ldots, \boldsymbol{f}_{\mathrm{RF},N}\} = \begin{bmatrix} \boldsymbol{f}_{\mathrm{RF},1} & \boldsymbol{0}_{M\times 1} & \cdots & \boldsymbol{0}_{M\times 1} \\ \boldsymbol{0}_{M\times 1} & \boldsymbol{f}_{\mathrm{RF},2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \boldsymbol{0}_{M\times 1} \\ \boldsymbol{0}_{M\times 1} & \cdots & \boldsymbol{0}_{M\times 1} & \boldsymbol{f}_{\mathrm{RF},N} \end{bmatrix}, \tag{5}$$

with $\boldsymbol{f}_{\mathrm{RF},n} \in \mathbb{C}^{M\times 1}$ and $\boldsymbol{f}_{\mathrm{RF},n} \in \mathcal{F} (\forall n = 1, \ldots, N)$, where each element satisfies the constant-magnitude and normalized-power constraints, which can be described as $|\boldsymbol{f}_{\mathrm{RF},n}[m]| = 1/\sqrt{M} (\forall m = 1, \ldots, M)$. The digital precoding matrix has a normalized-power constraint that $||\boldsymbol{F}_{\mathrm{RF}} \boldsymbol{F}_{\mathrm{BB}}||_{\mathrm{F}}^2 = N_{\mathrm{s}}$. The transmitted symbol vector satisfies $\boldsymbol{x}_k \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{P}_k)$ with $\boldsymbol{P}_k = \mathrm{diag}\{p_{k1}, \ldots, p_{kN_{\mathrm{s}k}}\}$, which is the transmitting power matrix of the data streams corresponding to the k-th user.

To make it easy for performance evaluation, we assume equal power allocation among different data streams and that the path losses of different users are the same, which means that $\beta_k = \beta$ and $p_{kn} = p = \rho\beta$ $(\forall k = 1, \ldots, K, n = 1, \ldots, N_{\mathrm{s}k})$. Thus, the average received signal-to-noise ratio (SNR) per antenna and data stream is ρ. Assume that $\bar{\boldsymbol{x}} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I}_{N_{\mathrm{s}}})$ is the normalized transmitted symbol vector, then Eq. (3) can be written as

$$\boldsymbol{y}_k = \sqrt{\rho} \boldsymbol{H}_k \boldsymbol{F}_{\mathrm{RF}} \boldsymbol{F}_{\mathrm{BB}} \bar{\boldsymbol{x}} + \boldsymbol{n}_k. \tag{6}$$

To recover the transmitted symbols, the k-th user will do a combining with the received signal. The n-th recovered symbol is denoted by (7), where $\boldsymbol{w}_{kn} = \boldsymbol{W}_{\mathrm{RF},k} \boldsymbol{w}_{\mathrm{BB},kn} \in \mathbb{C}^{N_{\mathrm{r}k} \times 1}$ is the combining vector, which is one of the columns in the combining matrix $\boldsymbol{W}_k = \boldsymbol{W}_{\mathrm{RF},k} \boldsymbol{W}_{\mathrm{BB},k} \in \mathbb{C}^{N_{\mathrm{r}k} \times N_{\mathrm{s}k}}$. $\boldsymbol{W}_{\mathrm{RF},k} \in \mathbb{C}^{N_{\mathrm{r}k} \times N_k}$ is the analog combining matrix that satisfies $\boldsymbol{W}_{\mathrm{RF},k} = \mathrm{blkdiag}\{\boldsymbol{w}_{\mathrm{RF},k1}, \ldots, \boldsymbol{w}_{\mathrm{RF},kN_k}\}$ with $\boldsymbol{w}_{\mathrm{RF},kn} \in \mathbb{C}^{M_k \times 1}$ and $\boldsymbol{w}_{\mathrm{RF},kn} \in \mathcal{W}$ $(\forall n = 1, \ldots, N_k)$. The constant-magnitude and normalized-power constraints still exist, which means that $|\boldsymbol{w}_{\mathrm{RF},kn}[m]| = 1/\sqrt{M_k}$ $(\forall m = 1, \ldots, M_k)$,

$$\begin{aligned} \hat{x}_{kn} &= \boldsymbol{w}_{kn}^{\mathrm{H}} (\sqrt{\rho} \boldsymbol{H}_k \boldsymbol{F}_{\mathrm{RF}} \boldsymbol{F}_{\mathrm{BB}} \bar{\boldsymbol{x}} + \boldsymbol{n}_k) \\ &= \underbrace{\sqrt{\rho} \boldsymbol{w}_{kn}^{\mathrm{H}} \boldsymbol{H}_k \boldsymbol{F}_{\mathrm{RF}} \boldsymbol{f}_{\mathrm{BB},kn} \bar{x}_{kn}}_{\text{Desired signal}} + \underbrace{\sqrt{\rho} \boldsymbol{w}_{kn}^{\mathrm{H}} \boldsymbol{H}_k \sum_{m \neq n} \boldsymbol{F}_{\mathrm{RF}} \boldsymbol{f}_{\mathrm{BB},km} \bar{x}_{km}}_{\text{Inter-stream interference}} \\ &\quad + \underbrace{\sqrt{\rho} \boldsymbol{w}_{kn}^{\mathrm{H}} \boldsymbol{H}_k \sum_{i \neq k} \boldsymbol{F}_{\mathrm{RF}} \boldsymbol{F}_{\mathrm{BB},i} \bar{\boldsymbol{x}}_i}_{\text{Inter-user interference}} + \underbrace{\boldsymbol{w}_{kn}^{\mathrm{H}} \boldsymbol{n}_k}_{\text{Noise}}. \end{aligned} \tag{7}$$

2.2.2 *Channel estimation*

To acquire CSI at the BS and each user, one possible way is to do beam sweeping: all sub-arrays on the BS side send orthogonal pilot signals with all possible beams, and each time when the BS sub-arrays fix their beamforming vectors, the users will receive the pilot signals with all sub-arrays and all possible beams. Suppose the orthogonal pilot signals are known by each user, then the equivalent downlink channel response matrix (the orthogonal pilot signals have been removed) between the j-th $(j = 1, 2, \ldots, N)$ RF chain at the BS and the i-th $(i = 1, 2, \ldots, N_k)$ RF chain at user k can be denoted as

$$\boldsymbol{Y}_{ij} = \boldsymbol{A}_{M_k}^{\mathrm{H}} (\boldsymbol{H}_{k,ij} \boldsymbol{A}_M + \boldsymbol{N}_{k,ij}), \tag{8}$$

where $\boldsymbol{A}_{M_k} \in \mathbb{C}^{M_k \times M_k}$ and $\boldsymbol{A}_M \in \mathbb{C}^{M \times M}$ are the codebook matrices of the k-th user and the BS, respectively. $\boldsymbol{N}_{k,ij} \in \mathbb{C}^{M_k \times M}$ is the normalized Gaussian noise matrix at the receiver and $\boldsymbol{H}_{k,ij} \in \mathbb{C}^{M_k \times M}$ is the channel matrix from the i-th sub-array at the BS to the j-th sub-array at the k-th user. As the codebook matrices are invertible, the channel matrix can be estimated by

$$\hat{\boldsymbol{H}}_{k,ij} = (\boldsymbol{A}_{M_k}^{\mathrm{H}})^{-1} \boldsymbol{Y}_{ij} \boldsymbol{A}_M^{-1} = \boldsymbol{H}_{k,ij} + \boldsymbol{N}_{k,ij} \boldsymbol{A}_M^{-1}. \tag{9}$$

Then, the estimation of the k-th user's channel matrix can be obtained with

$$\hat{\boldsymbol{H}}_k = \begin{bmatrix} \hat{\boldsymbol{H}}_{k,11} & \cdots & \hat{\boldsymbol{H}}_{k,1N} \\ \vdots & \ddots & \vdots \\ \hat{\boldsymbol{H}}_{k,N_k 1} & \cdots & \hat{\boldsymbol{H}}_{k,N_k N} \end{bmatrix} = \boldsymbol{H}_k + \Delta \boldsymbol{H}_k. \tag{10}$$

The estimated CSI will be fed back to the BS by each user.

2.2.3 *Achievable downlink spectral efficiency*

Here we assume that each user knows its own estimated CSI and the related beamforming matrix/vector, then the ergodic achievable downlink sum spectral efficiency (SE) of the system can be expressed as

$$\eta_{\mathrm{SE}} = \mathrm{E}_{\boldsymbol{H}} \left[\sum_{k=1}^{K} \sum_{n=1}^{N_{sk}} \log_2(1 + \gamma_{kn}) \right], \tag{11}$$

with γ_{kn} the signal-to-interference-plus-noise ratio (SINR) of the n-th data stream of the k-th user, which can be calculated according to (7) and is given by (12). $\boldsymbol{H} = [\boldsymbol{H}_1^{\mathrm{H}}, \ldots, \boldsymbol{H}_K^{\mathrm{H}}]^{\mathrm{H}}$ is the fast fading channel matrix of this system,

$$\gamma_{kn} = (|\boldsymbol{w}_{\mathrm{BB},kn}^{\mathrm{H}} \boldsymbol{W}_{\mathrm{RF},k}^{\mathrm{H}} \hat{\boldsymbol{H}}_k \boldsymbol{F}_{\mathrm{RF}} \boldsymbol{f}_{\mathrm{BB},kn}|^2) \Big/ \Bigg(\mathrm{E} \Bigg[\sum_{(i,m) \neq (k,n)} |\boldsymbol{w}_{\mathrm{BB},kn}^{\mathrm{H}} \boldsymbol{W}_{\mathrm{RF},k}^{\mathrm{H}} \boldsymbol{H}_k \boldsymbol{F}_{\mathrm{RF}} \boldsymbol{f}_{\mathrm{BB},im}|^2 + |\boldsymbol{w}_{\mathrm{BB},kn}^{\mathrm{H}} \boldsymbol{W}_{\mathrm{RF},k}^{\mathrm{H}} \Delta \boldsymbol{H}_k \boldsymbol{F}_{\mathrm{RF}} \boldsymbol{f}_{\mathrm{BB},kn}|^2 \Big| \hat{\boldsymbol{H}}_k \Bigg] + \frac{1}{\rho} \|\boldsymbol{W}_{\mathrm{RF},k} \boldsymbol{w}_{\mathrm{BB},kn}\|_2^2 \Bigg). \tag{12}$$

2.3 Power consumption model

We consider the power consumption of the BS, which can be easily extended to the user side when uplink transmission is considered. In each Butler matrix, the power consumption is caused by the RF switch that chooses the beam direction, whose power can be modeled as $P_{\mathrm{SW}} = 5$ mW [9].

Each antenna is connected to a power amplifier, whose power consumption can be modeled as $P_{\mathrm{PA}} = P_{\mathrm{t}}/\eta_{\mathrm{PA}}$, where P_{t} and η_{PA} are the average transmitting power of each antenna and the efficiency of the power amplifier, respectively. η_{PA} can be set as 0.375 [14]. The RF chain contains a mixer, a local oscillator, a filter and a base-band amplifier. The power consumed by each RF chain can be expressed as

$$P_{\mathrm{RF}} = P_{\mathrm{mixer}} + P_{\mathrm{LO}} + P_{\mathrm{filter}} + P_{\mathrm{BB\ amp}}. \tag{13}$$

The reference values for them are given in [3]: $P_{\mathrm{mixer}} = 19$ mW, $P_{\mathrm{LO}} = 5$ mW, $P_{\mathrm{filter}} = 14$ mW, $P_{\mathrm{BB\ amp}} = 5$ mW. Thus the total power consumption for each RF chain is $P_{\mathrm{RF}} = 43$ mW. The power consumption of each DAC is $P_{\mathrm{DAC}} = 255$ mW, and it is approximated that the power consumption of the base-band processing is $P_{\mathrm{BB}} = P_{\mathrm{DAC}}$ [9].

Therefore, the total power consumption of the Butler-matrix-based HBF structure can be given by

$$P_{\mathrm{Butler}} = N(P_{\mathrm{DAC}} + P_{\mathrm{RF}} + P_{\mathrm{SW}}) + MNP_{\mathrm{PA}} + P_{\mathrm{BB}}. \tag{14}$$

As a comparison, the power consumption of the phase-shifter-based HBF structure is

$$P_{\mathrm{Phase}} = N(P_{\mathrm{DAC}} + P_{\mathrm{RF}}) + MN(P_{\mathrm{PA}} + P_{\mathrm{PS}}) + P_{\mathrm{BB}}, \tag{15}$$

where $P_{\mathrm{PS}} = 30$ mW is the power consumption of each phase shifter [9]. And the power consumption of the full digital structure is

$$P_{\mathrm{Digital}} = MN(P_{\mathrm{DAC}} + P_{\mathrm{RF}} + P_{\mathrm{PA}}) + P_{\mathrm{BB}}. \tag{16}$$

3 HBF algorithm design

In this section, the HBF algorithms for the Butler-matrix-based HBF structure are considered. Here we assume that the estimated CSI is known at the BS and the related users. Our goal is to maximize the instantaneous SE with the constraints shown in Subsection 2.2,

$$\bar{\eta}_{\mathrm{SE}} = \sum_{k=1}^{K} \sum_{n=1}^{N_{\mathrm{s}k}} \log_2(1+\gamma_{kn}). \tag{17}$$

Since it is hard to solve the joint optimization problem directly, we would like to design ABF and DBF separately.

3.1 The exhaustive searching HBF algorithm

The exhaustive searching HBF algorithm for the Butler-matrix-based system can be utilized to obtain the performance upper bound. As shown in Algorithm 1, in the ABF stage, exhaustive searching through the codebooks of all the sub-arrays is done to find the best beam directions for the BS and each user, respectively. The achievable rate of the effective channel, which is the channel processed after analog precoding and combining, is utilized as the selection criterion. In the DBF stage, the digital block diagonalization (BD) algorithm [15] is processed with the low-dimensional effective channel.

Algorithm 1 The proposed HBF algorithm for the Butler-matrix-based system

Input: BS Butler matrix codebook $\mathcal{F}(|\mathcal{F}| = M)$, user k Butler matrix codebook $\mathcal{W}_k$ $(|\mathcal{W}_k| = M_k)$, and the estimated downlink channel matrix $\hat{\boldsymbol{H}}_k$ $(k = 1, \ldots, K)$.

Output: The analog precoding/combining matrix $\boldsymbol{F}_{\mathrm{RF}}/\boldsymbol{W}_{\mathrm{RF},k}$, the digital precoding/combining matrix $\boldsymbol{F}_{\mathrm{BB}}/\boldsymbol{W}_{\mathrm{BB},k}$.

The ABF stage:

Precoder:

The BS selects $\boldsymbol{f}^*_{\mathrm{RF},n}(n = 1, \ldots, N)$ that solve the following problem:

$$\{\boldsymbol{f}^*_{\mathrm{RF},n}\} = \underset{\substack{\forall \boldsymbol{f}_{\mathrm{RF},n} \in \mathcal{F} \\ (n=1,\ldots,N)}}{\operatorname{argmax}} \log_2 \det \left[\boldsymbol{I}_N + \rho \boldsymbol{F}^{\mathrm{H}}_{\mathrm{RF}} \hat{\boldsymbol{H}}^{\mathrm{H}} \hat{\boldsymbol{H}} \boldsymbol{F}_{\mathrm{RF}} \right].$$

The BS sets $\boldsymbol{F}_{\mathrm{RF}} = \mathrm{blkdiag}\{\boldsymbol{f}^*_{\mathrm{RF},1}, \ldots, \boldsymbol{f}^*_{\mathrm{RF},N}\}$.

Combiner:

The k-th $(k = 1, \ldots, K)$ user selects $\boldsymbol{w}^*_{\mathrm{RF},ki}(i = 1, \ldots, N_k)$ that solve the following problem:

$$\{\boldsymbol{w}^*_{\mathrm{RF},ki}\} = \underset{\substack{\forall \boldsymbol{w}_{\mathrm{RF},ki} \in \mathcal{W}_k \\ (i=1,\ldots,N_k)}}{\operatorname{argmax}} \log_2 \det \left[\boldsymbol{I}_{N_k} + \rho \boldsymbol{W}^{\mathrm{H}}_{\mathrm{RF},k} \hat{\boldsymbol{H}}_k \hat{\boldsymbol{H}}^{\mathrm{H}}_k \boldsymbol{W}_{\mathrm{RF},k} \right].$$

User k sets $\boldsymbol{W}_{\mathrm{RF},k} = \mathrm{blkdiag}\{\boldsymbol{w}^*_{\mathrm{RF},k1}, \ldots, \boldsymbol{w}^*_{\mathrm{RF},kN_k}\}$.

The DBF stage:

The digital block diagonalization (BD) algorithm [15] is utilized with the low-dimensional effective channel matrix $\bar{\boldsymbol{H}} = [\hat{\boldsymbol{H}}^{\mathrm{H}}_1 \boldsymbol{W}_{\mathrm{RF},1}, \ldots, \hat{\boldsymbol{H}}^{\mathrm{H}}_K \boldsymbol{W}_{\mathrm{RF},K}]^{\mathrm{H}} \boldsymbol{F}_{\mathrm{RF}}$, then we can obtain the digital precoding/combining matrix $(\boldsymbol{F}_{\mathrm{BB}}/\boldsymbol{W}_{\mathrm{BB},k})$ (To meet the transmission power constraint, set $\boldsymbol{F}_{\mathrm{BB}} = \frac{\sqrt{N_{\mathrm{s}}}\boldsymbol{F}_{\mathrm{BB}}}{||\boldsymbol{F}_{\mathrm{RF}}\boldsymbol{F}_{\mathrm{BB}}||_{\mathrm{F}}}$).

The searching complexity of the ABF stage is $\mathcal{O}(M^N + \sum_{k=1}^{K} M_k^{N_k})$. At each searching time, a series of matrix multiplications are done to obtain the achievable rate. It can be seen that when the number of antennas or RF chains increases, the searching complexity will increase much faster, which is hard to be accepted.

3.2 The proposed low complexity solution

In order to reduce the searching complexity, we propose a low complexity algorithm, which is different from Algorithm 1 in the ABF stage. As described in Algorithm 2, the exhaustive searching process to find the best ABF vectors is replaced with a sequential searching process. By doing so, the searching complexity of the ABF stage is reduced to $\mathcal{O}(MN + \sum_{k=1}^{K} M_k N_k)$, which increases linearly with the number of antennas or RF chains.

Algorithm 2 The **ABF stage** of the low complexity HBF algorithm for the Butler-matrix-based system

Precoder:

The BS selects $\boldsymbol{f}_{\mathrm{RF},n}^{*}(n=1,\ldots,N)$ sequentially that solve the following problem:

for $n=1$ to N **do**

$$\boldsymbol{f}_{\mathrm{RF},n}^{*}=\underset{\forall \boldsymbol{f}_{\mathrm{RF},n}\in\mathcal{F}}{\operatorname{argmax}}\log_2\det\left[\boldsymbol{I}_n+\rho\bar{\boldsymbol{F}}_{\mathrm{RF},n}^{\mathrm{H}}\hat{\boldsymbol{H}}^{\mathrm{H}}\hat{\boldsymbol{H}}\bar{\boldsymbol{F}}_{\mathrm{RF},n}\right],$$

where $\bar{\boldsymbol{F}}_{\mathrm{RF},n}=\begin{bmatrix}\mathrm{blkdiag}\{\boldsymbol{f}_{\mathrm{RF},1}^{*},\ldots,\boldsymbol{f}_{\mathrm{RF},n-1}^{*},\boldsymbol{f}_{\mathrm{RF},n}\}\\ \boldsymbol{0}_{M(N-n)\times n}\end{bmatrix}$.

end for

The BS sets $\boldsymbol{F}_{\mathrm{RF}}=\mathrm{blkdiag}\{\boldsymbol{f}_{\mathrm{RF},1}^{*},\ldots,\boldsymbol{f}_{\mathrm{RF},N}^{*}\}$.

Combiner:

The k-th $(k=1,\ldots,K)$ user selects $\boldsymbol{w}_{\mathrm{RF},ki}^{*}(i=1,\ldots,N_k)$ sequentially that solve the following problem:

for $i=1$ to N_k **do**

$$\boldsymbol{w}_{\mathrm{RF},ki}^{*}=\underset{\forall \boldsymbol{w}_{\mathrm{RF},ki}\in\mathcal{W}_k}{\operatorname{argmax}}\log_2\det\left[\boldsymbol{I}_i+\rho\bar{\boldsymbol{W}}_{\mathrm{RF},ki}^{\mathrm{H}}\hat{\boldsymbol{H}}_k\hat{\boldsymbol{H}}_k^{\mathrm{H}}\bar{\boldsymbol{W}}_{\mathrm{RF},ki}\right],$$

where $\bar{\boldsymbol{W}}_{\mathrm{RF},ki}=\begin{bmatrix}\mathrm{blkdiag}\{\boldsymbol{w}_{\mathrm{RF},k1}^{*},\ldots,\boldsymbol{w}_{\mathrm{RF},k(i-1)}^{*},\boldsymbol{w}_{\mathrm{RF},ki}\}\\ \boldsymbol{0}_{M_k(N_k-i)\times i}\end{bmatrix}$.

end for

User k sets $\boldsymbol{W}_{\mathrm{RF},k}=\mathrm{blkdiag}\{\boldsymbol{w}_{\mathrm{RF},k1}^{*},\ldots,\boldsymbol{w}_{\mathrm{RF},kN_k}^{*}\}$.

4 Simulation results

In this section, we evaluate the performance of the proposed Butler-matrix-based HBF structure and its related HBF algorithms. The sub-connected full digital system and the phase-shifter-based HBF system with algorithms given by [7,8] are utilized as comparisons. To make a fair comparison, the digital stage of [7,8] is replaced by the BD algorithm. Both SE and energy efficiency (EE) are considered.

Consider the downlink of a single-cell multiuser mmWave system with ULAs that have half-wavelength antenna spacing. The BS is equipped with $N_{\mathrm{t}}=32$ transmit antennas and $N=4$ RF chains. Each RF chain is connected to a sub-array with $M=8$ antennas. There are $K=2$ users served simultaneously by the BS. Each user has $N_{\mathrm{r}k}=8$ antennas and $N_k=2$ RF chains. Each RF chain of users is connected to a sub-array with $M_k=4$ antennas. The transmitting data stream number is assumed to be equal to the number of RF chains.

The channels are generated according to the clustered multipath channel model in (1) with $N_{\mathrm{cl},k}=6$ scattering clusters, each having $N_{\mathrm{ray},k}=5$ propagation paths. The cluster power is generated randomly from a uniform random variable in $[0,1]$ and then the total cluster power corresponding to each user is normalized. The azimuth AoA and AoD within a cluster are assumed to be Laplacian-distributed with angular standard deviation $\delta_{\theta^{\mathrm{r}}}=\delta_{\theta^{\mathrm{t}}}=5^{\circ}$ [11]. The mean cluster angles are assumed to be uniformly distributed within $[0,360^{\circ}]$. The codebooks consist of steering vectors chosen from the DFT matrices. Therefore, we have $\boldsymbol{a}_M\ (\arccos(\frac{2m}{M}-1))(m=1,\ldots,M)$ in $\mathcal{F}$ and $\boldsymbol{a}_{M_k}\ (\arccos(\frac{2n}{M_k}-1))(n=1,\ldots,M_k)$ in $\mathcal{W}_k$. The sizes of codebook for the BS and users are eight and four, respectively.

Two different conditions of CSI are considered in this part. The first one is that perfect CSI is known by the BS and users, which means that there is no channel estimation error. The second one is that CSI is estimated by the BS and users, so channel estimation error exists. Here we assume that the SNR of channel estimation is the same as that of data transmission.

In the simulations, the multipath angles of each user and their related path average powers are generated independently for 100 times. Each time when given these parameters, the complex propagation gains are generated independently for 1000 times to get the SINR of each data stream. Then the ergodic SE can be obtained.

4.1 Spectral efficiency performance evaluation

The ergodic achievable downlink sum SE given in (11) is utilized to validate the proposed structure and related algorithms.

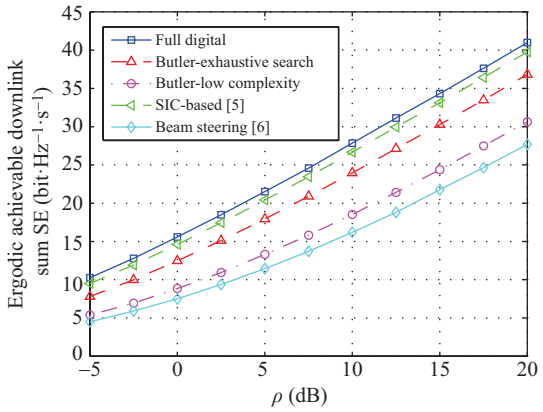

Figure 2 (Color online) The ergodic achievable downlink sum SE vs. ρ under the condition of perfect CSI: $N_{\rm s} = N = 4, M = 8, N_{\rm t} = 32, K = 2, N_{sk} = N_k = 2, M_k = 4, N_{rk} = 8$.

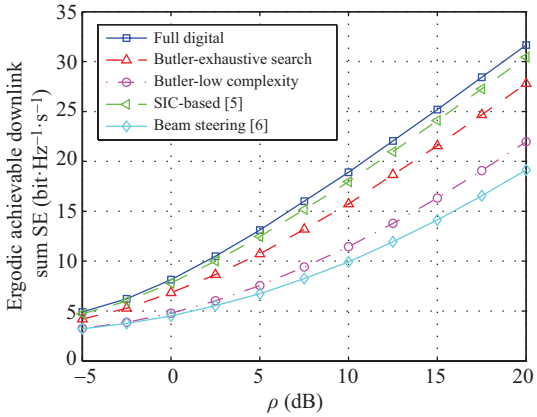

Figure 3 (Color online) The ergodic achievable downlink sum SE vs. ρ under the condition of estimated CSI: $N_{\rm s} = N = 4, M = 8, N_{\rm t} = 32, K = 2, N_{sk} = N_k = 2, M_k = 4, N_{rk} = 8$.

As shown in Figure 2, under the condition of perfect CSI, we can see that the Butler-matrix-based system with exhaustive search can achieve ninety percent of the sum SE achieved by the phase-shifter-based system with SIC-based algorithm [7] and eighty five percent of that achieved by the full digital system. Moreover, it outperforms the phase-shifter-based system with beam steering algorithm [8] a lot. Using the low complexity algorithm, the performance of the Butler-matrix-based system deteriorates nearly twenty percent, but it still performs better than the phase-shifter-based system with beam steering algorithm [8].

Then, as shown in Figure 3, when considering channel estimation error, the relative performance does not change much, which means that channel estimation error does not have additional negative effect on the Butler-matrix-based system, compared with the full digital system and the phase-shifter-based system.

4.2 Energy efficiency performance evaluation

According to the power consumption model given in Subsection 2.3, we can obtain EE with [14],

$$\eta_{\rm EE} = \frac{W\eta_{\rm SE}}{P_{\rm total}} \quad ({\rm bit}\cdot{\rm J}^{-1}), \tag{18}$$

where W is the bandwidth of the system that is set as 200 MHz, $P_{\rm total}$ is calculated with (14)–(16) for the Butler-matrix-based system, the phase-shifter-based system and the full digital system, respectively. Assume that the average path loss is 100 dB. The noise power spectral density is -174 dBm·Hz^{-1}. Each time when given a ρ, we can calculate the average transmitting power of antenna at the BS side and get the power consumption of the power amplifier. Then EE can be calculated with (18).

In Figure 4, under the condition of perfect CSI, it can be seen that as SE increases, the trend of EE is similar for all the systems. When SE is low, EE increases with SE. After reaching a peak point, EE decreases with SE. It is obvious that the HBF system performs more energy-efficient than the full digital system. Considering the HBF systems, we can see that the Butler-matrix-based system with exhaustive search achieves a better EE performance than the phase-shifter-based system with both SIC-based algorithm [7] and beam steering algorithm [8]. The peak point obtained by the Butler-matrix-based system has an improvement of twenty seven percent compared with the phase-shifter-based system with SIC-based algorithm. When utilizing the low complexity algorithm in the Butler-matrix-based system, the EE performance is worse compared with that of the exhaustive search algorithm. It is due to the fact that the exhaustive search algorithm can achieve a better SE performance when they have the same SNR. But the Butler-matrix-based system can still have a better EE performance when SE is low, and its peak point is at a similar level as that of the phase-shifter-based system with SIC-based algorithm.

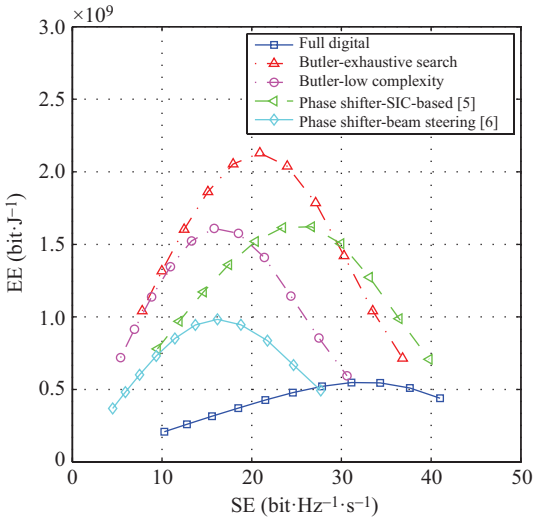

Figure 4 (Color online) EE vs. SE under the condition of perfect CSI: $N_{\rm s} = N = 4, M = 8, N_{\rm t} = 32, K = 2, N_{sk} = N_k = 2, M_k = 4, N_{{\rm r}k} = 8$.

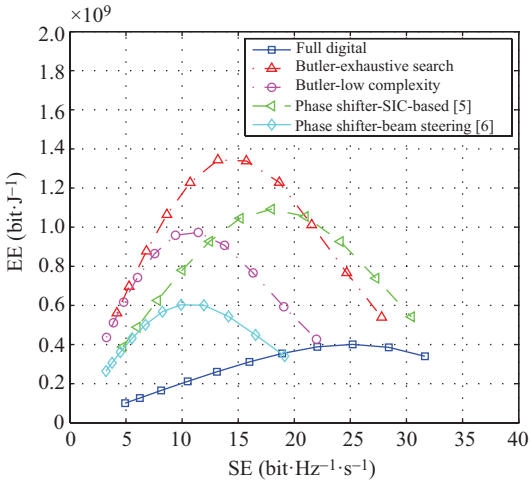

Figure 5 (Color online) EE vs. SE under the condition of estimated CSI: $N_{\rm s} = N = 4, M = 8, N_{\rm t} = 32, K = 2, N_{sk} = N_k = 2, M_k = 4, N_{{\rm r}k} = 8$.

Considering channel estimation error, as shown in Figure 5, the Butler-matrix-based system with exhaustive search still performs the best. And with the low complexity algorithm, the EE performance of the Butler-matrix-based system deteriorates faster than the phase-shifter-based system with SIC-based algorithm, but when SE is low, it still outperforms the phase-shifter-based system a lot.

5 Conclusion

A Butler-matrix-based HBF structure is proposed in this paper. We provide two HBF algorithms for the proposed HBF structure: one is proposed to find the best beam directions under this structure, which obtains the performance upper bound; the other one is proposed to reduce the searching complexity, which is more practical to be implemented. Simulation results indicate that, for both the perfect CSI and the estimated CSI conditions, the Butler-matrix-based HBF structure with the related algorithms can achieve an acceptable SE performance compared with the phase-shifter-based HBF structure, and a better EE performance can be obtained, which verifies the effectiveness of the proposed HBF structure.

Acknowledgements This work was supported by National Basic Research Program of China (973 Program) (Grant No. 2012CB316002), National Natural Science Foundation of China (Grant No. 61201192), National High Technology Research and Development Program of China (863 Program) (Grant No. 2015AA01A701), Science Fund for Creative Research Groups of NSFC (Grant No. 61321061), International Science and Technology Cooperation Program (Grant No. 2012DFG12010), National S&T Major Project (Grant No. 2015ZX03002002), Key Grant Project of Chinese Ministry of Education (Grant No. 313005), Tsinghua University Initiative Scientific Research (Grant No. 2015Z02-3), Open Research Fund of National Mobile Communications Research Laboratory, Southeast University (Grant No. 2012D02), and Tsinghua-Qualcomm Joint Research Program.

Conflict of interest The authors declare that they have no conflict of interest.

References

1 Pi Z, Khan F. An introduction to millimeter-wave mobile broadband systems. IEEE Commun Mag, 2011, 49: 101–107
2 Rappaport T S, Sun S, Mayzus R, et al. Millimeter wave mobile communications for 5G cellular: it will work! IEEE Access, 2013, 1: 335–349
3 Rangan S, Rappaport T S, Erkip E, et al. Energy efficient methods for millimeter wave picocellular systems. In: Proceedings of IEEE Communications Theory Workshop, Phuket, 2013. 1–25

4 Sohrabi F, Yu W. Hybrid digital and analog beamforming design for large-scale antenna arrays. IEEE J Sel Topics Signal Process, 2016, 10: 501–513

5 Li J H, Xiao L M, Xu X B, et al. Robust and low complexity hybrid beamforming for uplink multiuser mmWave MIMO systems. IEEE Commun Lett, 2016, 20: 1140–1143

6 Zi R, Ge X H, Thompson J, et al. Energy efficiency optimization of 5G radio frequency chain systems. IEEE J Sel Areas Commun, 2016, 34: 758–771

7 Gao X Y, Dai L L, Han S F, et al. Energy-efficient hybrid analog and digital precoding for mmWave MIMO systems with large antenna arrays. IEEE J Sel Areas Commun, 2016, 34: 998–1009

8 Alkhateeb A, Leus G, Heath R W. Limited fedback hybrid precoding for multi-user millimeter wave systems. IEEE Trans Wirel Commun, 2015, 14: 6481–6494

9 Méndez-Rial R, Rusu C, González-Prelcic N, et al. Hybrid MIMO architectures for millimeter wave communications: phase shifters or switches? IEEE Access, 2016, 4: 247–267

10 Chin T Y, Chang S F, Wu J C, et al. A 25-GHz compact low-power phased-array receiver with continuous beam steering in CMOS technology. IEEE J Solid-State Circ, 2010, 45: 2273–2282

11 Forenza A, Love D, Heath R W. Simplified spatial correlation models for clustered MIMO channels with different array configurations. IEEE Trans Veh Tech, 2007, 56: 1924–1934

12 El Ayach O, Heath R W, Rajagopal S, et al. Multimode precoding in millimeter wave MIMO transmitters with multiple antenna sub-arrays. In: Proceedings of IEEE Global Communication Conference (GLOBECOM), Atlanta, 2013. 3476–3480

13 Alkhateeb A, Heath R W. Frequency selective hybrid precoding for limited feedback millimeter wave systems. IEEE Trans Commun, 2016, 64: 1801–1818

14 Han S, Chih-Lin I, Xu Z K, et al. Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G. IEEE Commun Mag, 2015, 53: 186–194

15 Stankovic V, Haardt M. Generalized design of multi-user MIMO precoding matrices. IEEE Trans Wirel Commun, 2008, 7: 953–961