Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
February 2017, 24(1): 18–25
www.sciencedirect.com/science/journal/10058885
The Journal of China Universities of Posts and Telecommunications
http://jcupt.bupt.edu.cn

Low complexity asymptotically unitary algorithm for hybrid beamforming in mmWave communication systems

Li Xiaohui[1,2] (✉), Meng Meimei[1], Lin Yingchao[1], Hei Yongqiang[1]

1. State Key Laboratory of Integrated Service Networks, Xidian University, Xi'an 710071, China
2. Collaborative Innovation Center of Information Sensing and Understanding, Xidian University, Xi'an 710071, China

Abstract

In millimeter wave (mmWave) massive multiple-input multiple-output (MIMO) systems, because of the high hardware cost and high power consumption, the traditional fully digital beamforming (DBF) cannot be implemented easily. Meanwhile, analog beamforming which is implemented with phase shifters has high availability but suffers poor performance. Considering the advantages of above two, a potential solution is to design an appropriate hybrid analog and digital beamforming structure, where the available iterative optimization algorithm can get performance close to fully digital processing, but solving this sparse optimization problem faces with a high computational complexity. The key challenge of seeking out hybrid beamforming (HBF) matrices lies in leveraging the trade-off between the spectral efficiency performance and the computational complexity. In this paper, we propose an asymptotically unitary hybrid precoding (AUHP) algorithm based on antenna array response (AAR) properties to solve the HBF optimization problem. Firstly, we get the optimal orthogonal analog and digital beamforming matrices relying on the channel's path gain in absolute value by taking into account that the AAR matrices are asymptotically unitary. Then, an improved simultaneously orthogonal matching pursuit (SOMP) algorithm based on recursion is adopted to refine the hybrid combining. Numerical results demonstrate that our proposed AUHP algorithm enables a lower computational complexity with negligible spectral efficiency performance degradation.

Keywords massive MIMO, mmWave communication, hybrid beamforming, asymptotically unitary, low complexity

1 Introduction

Recently, the increasingly demanding requirement of high data rate transmission has promoted the research of new technology, which includes long-term evolution in unlicensed band (LTE-U), massive MIMO, mmWave and so on. Due to the research potential of the large unlicensed bandwidth, especially mmWave band (e.g. 60 GHz), mmWave wireless communication has drawn a great attention. Moreover, it has been regarded as a key technology in the fifth generation (5G) communication [1]. However, there is a biggest challenge of the mmWave communication that the high carrier frequency suffer a significant propagation attenuation. Fortunately, the short wavelength resulting from high frequency enables the transmitters and receivers to configure massive MIMO systems [2]. It is worth noting that the massive MIMO configurations are able to provide sufficient gains to compensate the serious path-loss by using beamforming.

Unfortunately, for massive MIMO systems which are equipped with hundreds of antennas at mmWave frequencies, the traditional fully DBF architecture is not practical due to the power consumption of the radio frequency (RF) chains and high hardware cost [3–4]. Meanwhile, analog beamforming which is implemented with phase shifters has high availability but suffers poor performance. Considering the advantages of above two, in this paper we consider a digital and analog hybrid beamforming architecture to reduce the high cost to ensure

Received date: 12-09-2016
Corresponding author: Li Xiaohui, E-mail: xhli@mail.xidian.edu.cn
DOI: 10.1016/S1005-8885(17)60183-3

the feasibility of mmWave communications which can reduce the number of RF chains by exploiting the limited scattering of mmWave communications [5]. Moreover, the HBF can achieve efficiency performance close to DBF [6].

In order to improve the feasibility of the mmWave communication, there have been designed different HBF architectures and proposed some HBF solutions to gain the digital and analog beamforming matrices in Refs. [7–10]. In Ref. [7], El Agach et al. proposed a suboptimal beam steering method only by simply considering the channel hardening phenomenon of massive MIMO systems, which can implement the transmit angle alignment base on channel state information (CSI) with low complexity. However, this strategy will suffers serious performance loss especially when the beam steering alone cannot capture the CSI at mmWave frequencies. In Ref. [9], this HBF design has been formulated into a sparse signal reconstruction issue which is widely used in image processing and then given a SOMP algorithm in detail. Although this iterative optimization algorithm can get performance close to fully digital processing, solving this sparse optimization problem faces with a high computational complexity. For the purpose to overcome the high computational complexity limitation, Rusu et al. in Ref. [9] propose a low complexity by exploiting the nature of semi-unitary, but this method needs to design an over complete matrix which is complex itself. Then, a modified SOMP algorithm is proposed in Ref. [10], which enable efficient parallel computation without sacrificing the spectral performance compared with the SOMP algorithm [7].

With the purpose of leveraging the trade-off between the spectral efficiency performance and the computational complexity, in this paper, we propose an AUHP algorithm to solve the HBF optimization problem. Firstly, taking into account that the AAR matrices are asymptotically unitary, we can get the optimal orthogonal analog and digital beamforming matrices relying on the channel's path gain in absolute value. Then, an improved SOMP algorithm based on recursion is adopted to refine the hybrid combining. This solution can achieve a great reduction of computational complexity with a negligible performance degradation. Moreover, numerical results show that the AUHP algorithm enables a negligible spectral efficiency performance degradation respect to the SOMP method in Ref. [8]. The main contributions of our work are as follows:

After exploiting the asymptotically unitary properties of AAR, an original algorithm is proposed to decrease the computational complexity compared to the SOMP by avoiding multi-iterations.

In order to further reduce the complexity, we adopt the compressive sampling matching pursuit (CoSaMP) algorithm, which identifies multiple atoms using a matched filter and combines it with the support-set AAR estimated in the previous iteration and which is superior to SOMP.

Building on the above, the simulation results show that the AUHP algorithm enables hybrid analog and digital beamforming structures to leverage the trade-off between the spectral efficiency performance and the computational complexity.

The rest of this paper is organized as follows: Sect. 2 describes the hybrid mmWave MIMO system model. Sect. 3 presents the hybrid analog and digital beamforming construction, and exhibits the AUHP algorithm in detail and gives the complexity analysis. Sect. 4 presents the simulation results and we conclude in Sect. 5.

Notations: we make the following notation throughout this paper: $\boldsymbol{A}$ is a matrix. $\boldsymbol{a}$ is a vector. a is a scalar. $(\bullet)^{\mathrm{T}}$ and $(\bullet)^{*}$ denote transpose and conjugate transpose respectively. $\boldsymbol{A}^{(i)}$ is the ith column of $\boldsymbol{A}$. $\|\bullet\|_{\mathrm{F}}$ is the Frobenius norm; $\mathrm{tr}(\bullet)$ is the trace. $r(\boldsymbol{A})$ is the rank of matrix $\boldsymbol{A}$. $[\boldsymbol{A}\,|\,\boldsymbol{B}]$ denotes horizontal concatenation. $\boldsymbol{I}_N$ is the $N\times N$ identity matrix; $\mathcal{CN}(\boldsymbol{a},\boldsymbol{A})$ is a complex Gaussian vector with mean $\boldsymbol{a}$ and covariance matrix $\boldsymbol{A}$. $\mathrm{E}[\bullet]$ denoted expectation.

2 Hybrid mmWave MIMO System model

2.1 System model

In this section, we present the hybrid mmWave communication system and give the mmWave channel model.

Firstly, a single user mmWave downlink system shown in Fig. 1 is considered. The base station (BS) equipped with N_{t} antennas transmits N_{S} data streams to the target mobile station (MS) equipped with N_{r} receive antennas. It is noting that the uplink is similar to this downlink, so we omit the uplink transmission progress for brevity. In this HBF architecture, the BS is equipped with

$N_{\mathrm{t}}^{\mathrm{RF}}$ transmit chains assuming that $N_{\mathrm{S}} \leqslant N_{\mathrm{t}}^{\mathrm{RF}} \leqslant N_{\mathrm{t}}$, and the MS is equipped with $N_{\mathrm{r}}^{\mathrm{RF}}$ receive chains assuming that $N_{\mathrm{S}} \leqslant N_{\mathrm{r}}^{\mathrm{RF}} \leqslant N_{\mathrm{r}}$, which ensure the multiple data steams mmWave communication. Thus, the BS apply a $N_{\mathrm{t}}^{\mathrm{RF}} \times N_{\mathrm{S}}$ baseband precoder $\boldsymbol{F}_{\mathrm{BB}}$ which has no constaint, followed by a $N_{\mathrm{t}} \times N_{\mathrm{t}}^{\mathrm{RF}}$ RF precoder $\boldsymbol{F}_{\mathrm{RF}}$ which elements need satisfy $\left(\boldsymbol{F}_{\mathrm{RF}}^{(i)} \boldsymbol{F}_{\mathrm{RF}}^{(i)*}\right)_{l,l} = 1/N_{\mathrm{t}}$.

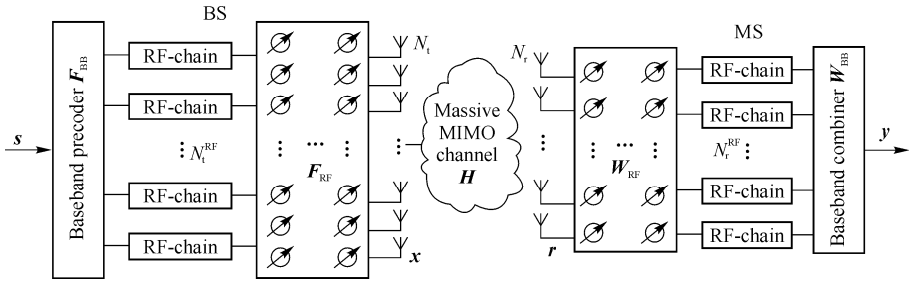

Fig. 1 Hardware block diagram of a mmWave single user system with HBF

In the BS, the discrete time transmitted data is presented as $\boldsymbol{x} = \boldsymbol{F}_{RF} \boldsymbol{F}_{\mathrm{BB}} \boldsymbol{s}$ where $\boldsymbol{s}$ is the data streams with that $\mathrm{E}\left[\boldsymbol{s}\boldsymbol{s}^*\right] = \left(1/N_{\mathrm{S}}\right) \boldsymbol{I}_{N_{\mathrm{S}}}$, and the total power constraint is $\left\|\boldsymbol{F}_{\mathrm{RF}} \boldsymbol{F}_{\mathrm{BB}}\right\|_{\mathrm{F}}^2 = N_{\mathrm{S}}$.

At the target MS, the received signal before detection is

$$\boldsymbol{r} = \sqrt{\rho} \boldsymbol{H} \boldsymbol{F}_{\mathrm{RF}} \boldsymbol{F}_{\mathrm{BB}} \boldsymbol{s} + \boldsymbol{n} \tag{1}$$

where $\boldsymbol{r}$ is the $N_{\mathrm{r}} \times 1$ receive vector, ρ denotes the average transmit power, $\boldsymbol{H}$ is the $N_{\mathrm{r}} \times N_{\mathrm{t}}$ channel matrix that satisfies $\mathrm{E}\left[\left\|\boldsymbol{H}\right\|_{\mathrm{F}}^2\right] = N_{\mathrm{t}} N_{\mathrm{r}}$, and $\boldsymbol{n}$ is the $N_{\mathrm{r}} \times 1$ noise follows i.i.d $\mathcal{CN}\left(0, \sigma_n^2\right)$ distribution. After the RF detection $\boldsymbol{W}_{\mathrm{RF}}$ and the baseband detection $\boldsymbol{W}_{\mathrm{BB}}$, we can get the original transmitted data streams denoted by $\boldsymbol{y}$ form the antenna received signal $\boldsymbol{r}$. The received data steams can be written as

$$\boldsymbol{y} = \sqrt{\rho} \boldsymbol{W}_{\mathrm{BB}}^* \boldsymbol{W}_{\mathrm{RF}}^* \boldsymbol{H} \boldsymbol{F}_{\mathrm{RF}} \boldsymbol{F}_{\mathrm{BB}} \boldsymbol{s} + \boldsymbol{W}_{\mathrm{BB}}^* \boldsymbol{W}_{\mathrm{RF}}^* \boldsymbol{n} \tag{2}$$

where $\boldsymbol{W}_{\mathrm{RF}}$ is the $N_{\mathrm{r}} \times N_{\mathrm{r}}^{\mathrm{RF}}$ matrix with unit amplitude limitation which elements need satisfy $\left(\boldsymbol{W}_{\mathrm{RF}}^{(i)} \boldsymbol{W}_{\mathrm{RF}}^{(i)*}\right)_{l,l} = 1/N_{\mathrm{r}}$ similar to the RF precoder, and $\boldsymbol{W}_{\mathrm{BB}}$ is the $N_{\mathrm{r}}^{\mathrm{RF}} \times N_{\mathrm{S}}$ baseband combining matrix without limitation.

In this mmWave communication system, the achieved spectral efficiency can be calculated by

$$R_{\mathrm{sum}} = \mathrm{lb}\left(\left|\boldsymbol{I}_{N_{\mathrm{S}}} + \frac{\rho}{N_{\mathrm{S}}} \boldsymbol{R}_n^{-1} \boldsymbol{W}_{\mathrm{BB}}^* \boldsymbol{W}_{\mathrm{RF}}^* \boldsymbol{H} \boldsymbol{F}_{\mathrm{RF}} \boldsymbol{F}_{\mathrm{BB}} \cdot \boldsymbol{F}_{\mathrm{BB}}^* \boldsymbol{F}_{\mathrm{RF}}^* \boldsymbol{H}^* \boldsymbol{W}_{\mathrm{RF}} \boldsymbol{W}_{\mathrm{BB}}\right|\right) \tag{3}$$

where $\boldsymbol{R}_n = \sigma_n^2 \boldsymbol{W}_{\mathrm{BB}}^* \boldsymbol{W}_{\mathrm{RF}}^* \boldsymbol{W}_{\mathrm{RF}} \boldsymbol{W}_{\mathrm{BB}}$ is the noise covariance matrix after detection.

2.2 MmWave channel model

Refer to the mmWave communication channel model [11], we adopt the narrowband block fading channel model. In order to further simplify the model reasonably, we consider an uniform linear array (ULA) with L propagation paths and d inter antenna distance [12]. For the lth , $l = 1, 2, ..., L$ path, the AAR is denoted by

$$\boldsymbol{a}_{\mathrm{ULA}_y}\left(\phi_l\right) = \frac{1}{\sqrt{N_{\mathrm{t}}}}\left[1, \mathrm{e}^{\mathrm{j}\frac{2\pi}{\lambda} d \sin\phi_l}, ..., \mathrm{e}^{\mathrm{j}(N_{\mathrm{t}}-1)\frac{2\pi}{\lambda} d \sin\phi_l}\right]^{\mathrm{T}} \tag{4}$$

where λ is wavelength, ϕ_l represents the angles of arrival (AoA) at the BS or the angles of departure (AoD) at the MS in the azimuth plane implemented horizontal (2D) beamforming only, which can extensions to three-demensional (3D) beamforming architecture. With L scatter paths, the $N_{\mathrm{r}} \times N_{\mathrm{t}}$ mmWave downlink channel model $\boldsymbol{H}$ can be expressed as

$$\boldsymbol{H} = \sqrt{\frac{N_{\mathrm{t}} N_{\mathrm{r}}}{L}} \sum_{l=1}^{L} \alpha_l \boldsymbol{a}_{\mathrm{r}}\left(\phi_l^{\mathrm{r}}\right) \boldsymbol{a}_{\mathrm{t}}^*\left(\phi_l^{\mathrm{t}}\right) \tag{5}$$

where $\alpha_l \sim \mathcal{CN}(0,1)$ is the complex random gain of the lth path, $\boldsymbol{a}_{\mathrm{r}}\left(\phi_l^{\mathrm{r}}\right)$ and $\boldsymbol{a}_{\mathrm{t}}\left(\phi_l^{\mathrm{t}}\right)$ denote the AAR vectors at the MS and the BS, respectively.

3 AUHP algorithm

In this section, the design problem of HBF is given first and how many RF chains are needed to achieve the optimum performance is discussed. Then, in order to reduce the computational cost of the SOMP method, we propose the AUHP algorithm with low complexity by

exploiting that the AAR has asymptotically unitary properties. Moreover, the computational complexity analysis of three algorithms, including beam steering method, is presented in detail.

3.1 Design problem of HBF

There are four matrices should be solved during the coherence time of data stream transmission to achieve mmWave communication. In this paper, we design the hybrid precoding and combing matrices by maximizing the achievable spectral efficiency in Eq. (3) with low computational complexity.

Therefore, the optimal hybrid analog and digital beamforming are obtained by the following optimization problem,

$$\left.\begin{aligned} &\arg \max_{\boldsymbol{F}_{\mathrm{RF}},\boldsymbol{F}_{\mathrm{BB}},\boldsymbol{W}_{\mathrm{RF}},\boldsymbol{W}_{\mathrm{BB}}} R_{\mathrm{sum}} \\ &\text{s.t.} \\ &\quad \mathrm{Tr}\left(\boldsymbol{F}_{\mathrm{RF}}\boldsymbol{F}_{\mathrm{BB}}\boldsymbol{W}_{\mathrm{RF}}^{*}\boldsymbol{W}_{\mathrm{BB}}^{*}\right)\leqslant\rho \\ &\quad \left|\boldsymbol{F}_{\mathrm{RF}}(i,\mathrm{j})\right|^{2}=1;\quad \forall i,j \\ &\quad \left|\boldsymbol{W}_{\mathrm{RF}}(i,\mathrm{j})\right|^{2}=1;\quad \forall i,j \end{aligned}\right\} \tag{6}$$

However, maximizing the spectral efficiency involves a joint optimization of four matrices variables $\left(\boldsymbol{F}_{\mathrm{RF}},\boldsymbol{F}_{\mathrm{BB}},\boldsymbol{W}_{\mathrm{RF}},\boldsymbol{W}_{\mathrm{BB}}\right)$ with non-convex constrains for $\boldsymbol{F}_{\mathrm{RF}}$ and $\boldsymbol{W}_{\mathrm{RF}}$. As a result, finding the global optimal solution is complex, which will damage the realizability. In this paper, we come up with an algorithm to solve the optimization problems Eq. (6) achieving high complexity reduction.

Before proposing the HBF solution, an interesting question, that is, how many RF chains are needed to achieve the optimum performance, should be discussed.

Lemma 1 In order to get a full-rank hybrid precoding matrix, it is necessary that $N_{\mathrm{t}}^{\mathrm{RF}}>N_{\mathrm{S}}$.

Proof Note that $r\left(\boldsymbol{F}_{\mathrm{RF}}\boldsymbol{F}_{\mathrm{BB}}\right)\leqslant N_{\mathrm{t}}^{\mathrm{RF}}$, so as to implement a rank N_{S}, $N_{\mathrm{t}}^{\mathrm{RF}}\geqslant N_{\mathrm{S}}$ is needed.

Lemma 2 For $N_{\mathrm{S}}>1$, the optimal precoder can be realized using the structure $N_{\mathrm{t}}^{\mathrm{RF}}\geqslant 2N_{\mathrm{S}}$.

Proof This is presented in Ref. [13].

3.2 The proposed AUHP algorithm

To simplify the transceiver design, we can temporarily decouple the problem. Firstly, we design the analog precoder matrix $\boldsymbol{F}_{\mathrm{RF}}$ and the baseband digital precoder matrix $\boldsymbol{F}_{\mathrm{BB}}$ jointly. Then we obtain the analog combiner $\boldsymbol{W}_{\mathrm{RF}}$ and the digital combiner $\boldsymbol{W}_{\mathrm{BB}}$.

Given that the CSI matrix $\boldsymbol{H}$ is known and can be expressed in its singular value decomposition (SVD) form as $\boldsymbol{H}=\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{*}$, where $\boldsymbol{U}$ and $\boldsymbol{V}$, the left and right singular vectors of $\boldsymbol{H}$, are $N_{\mathrm{r}}\times r(\boldsymbol{H})$ and $N_{\mathrm{t}}\times r(\boldsymbol{H})$ unitary matrices, and $\boldsymbol{\Sigma}$ is the $r(\boldsymbol{H})\times r(\boldsymbol{H})$ diagonal matrix of singular values arranged in decreasing order.

For the case of N_{S} data streams, the optimal fully digital precoder and combiner can be simply gained, i.e., $\boldsymbol{F}_{\mathrm{opt}}=\boldsymbol{V}(:,1:N_{\mathrm{S}})$ and $\boldsymbol{W}_{\mathrm{opt}}=\boldsymbol{U}(:,1:N_{\mathrm{S}})$. Unfortunately, the analog progressing is subject to many constraints. Therefore, the hybrid precoder (combiner) $\boldsymbol{F}_{\mathrm{RF}}\boldsymbol{F}_{\mathrm{BB}}$ $\left(\boldsymbol{W}_{\mathrm{RF}}\boldsymbol{W}_{\mathrm{BB}}\right)$ can be just made sufficiently 'close to' the optimal result $\boldsymbol{F}_{\mathrm{opt}}$ $\left(\boldsymbol{W}_{\mathrm{opt}}\right)$.

Assuming a perfect maximum likelihood (ML) receiver, the approach in Ref. [8] is first to focus on the hybrid precoders, i.e., $\boldsymbol{F}_{\mathrm{RF}}$ and $\boldsymbol{F}_{\mathrm{BB}}$. The optimal HBF problem is converted into solving the following optimization problem

$$\left.\begin{aligned} &\arg\max\ \mathrm{lb}\left|\boldsymbol{I}+\frac{\rho}{N_{\mathrm{S}}\sigma_{n}^{2}}\boldsymbol{H}\boldsymbol{F}_{\mathrm{RF}}\boldsymbol{F}_{\mathrm{BB}}\boldsymbol{F}_{\mathrm{BB}}^{*}\boldsymbol{F}_{\mathrm{RF}}^{*}\boldsymbol{H}^{*}\right| \\ &\text{s.t.} \\ &\quad \boldsymbol{F}_{\mathrm{RF}}\in\boldsymbol{\Gamma}_{\mathrm{RF}} \\ &\quad \left\|\boldsymbol{F}_{\mathrm{RF}}\boldsymbol{F}_{\mathrm{BB}}\right\|_{\mathrm{F}}^{2}=N_{\mathrm{S}} \end{aligned}\right\} \tag{7}$$

Since amplitude constrains are imposed on $\boldsymbol{F}_{\mathrm{RF}}$, the problem Eq. (7) cannot have a traditional solution. In Ref. [8], some reasonable approximations and assumptions are given and proposed. The set of feasible RF implementable precoders with phase shifters $\boldsymbol{\Gamma}_{\mathrm{RF}}$ to the set of array response $\boldsymbol{A}_{\mathrm{t}}=\left[\boldsymbol{a}_{\mathrm{t}}\left(\phi_{1}^{t}\right),\boldsymbol{a}_{\mathrm{t}}\left(\phi_{2}^{\mathrm{t}}\right),...,\boldsymbol{a}_{\mathrm{t}}\left(\phi_{L}^{\mathrm{t}}\right)\right]$, with L is the angular resolution. Detailed theoretical derivation shows that the near optimal HBF matrices can be solve by minimizing the Frobenius norm with respect to the fully digital unconstrained result, which can be formulated as

$$\left.\begin{aligned} &\arg\min_{\boldsymbol{F}_{\mathrm{RF}},\boldsymbol{F}_{\mathrm{BB}}}\left\|\boldsymbol{F}_{\mathrm{opt}}-\boldsymbol{A}_{\mathrm{t}}\boldsymbol{F}_{\mathrm{BB}}\right\|_{\mathrm{F}} \\ &\text{s.t.} \\ &\quad \left\|\mathrm{diag}\left(\boldsymbol{F}_{\mathrm{BB}}\boldsymbol{F}_{\mathrm{BB}}^{*}\right)\right\|_{0}=N_{\mathrm{t}}^{\mathrm{RF}} \\ &\quad \left\|\boldsymbol{A}_{\mathrm{t}}\boldsymbol{F}_{\mathrm{BB}}\right\|_{\mathrm{F}}^{2}=N_{\mathrm{S}} \end{aligned}\right\} \tag{8}$$

This near optimal problem can be solved with the

SOMP algorithm which general used in sparse signal recovery [14]. An equivalent problem applies to solve the combiner matrices. The simulation results and theoretical analysis indicate that the proposed algorithm can accurately approximate optimal fully digital unconstrained beamforming even when the transceiver hardware constrains are considered. However, solving the sparse optimization problem using SOMP still results in high complexity due to the iterative selection process and the process of solving the pseudo-inverse matrix in each iteration, which makes the SOMP is not suitable for the realization of mmWave communication.

To overcome the limitation of high complexity, using the following Lemma 3, we propose the AUHP algorithm with low complexity to solving the optimization problem by exploiting that the AAR $\boldsymbol{A}_{\mathrm{t}}$ has asymptotically unitary properties.

Lemma 3 For an ULA system with azimuth AoA drawn independently from a continuous distribution, the massive MIMO systems equip BS antenna arrays with an order of magnitude more elements, i.e., a hundred antennas or more. Therefore, the BS AAR vectors $\boldsymbol{a}_{\mathrm{t}}\left(\phi_l^{\mathrm{t}}\right)$ are asymptotically orthogonal, i.e., $\boldsymbol{a}_{\mathrm{t}}\left(\phi_k^{\mathrm{t}}\right) \perp \boldsymbol{a}_{\mathrm{t}}\left(\phi_l^{\mathrm{t}}\right), k \neq l$ as the number of the BS antennas N_{t} tends to infinity and the number of paths is $L = o(N)$ in the channel.

Proof This is presented in the Appendix A.

Since $\boldsymbol{F}_{\mathrm{opt}}$ is a column unitary matrix, i.e., $\boldsymbol{F}_{\mathrm{opt}}$ has orthonormal columns but is not square. Therefore, its 'close to' matrix need to obey the approximate equation $\left(\boldsymbol{F}_{\mathrm{RF}}\boldsymbol{F}_{\mathrm{BB}}\right)^{*}\boldsymbol{F}_{\mathrm{RF}}\boldsymbol{F}_{\mathrm{BB}} \approx \boldsymbol{I}_{N_{\mathrm{S}}}$, and thus $\boldsymbol{F}_{\mathrm{RF}}{}^{*}\boldsymbol{F}_{\mathrm{RF}}$ and $\boldsymbol{F}_{\mathrm{BB}}{}^{*}\boldsymbol{F}_{\mathrm{BB}}$ need to be closely approximate unitary matrices, i.e., $\boldsymbol{F}_{\mathrm{RF}}{}^{*}\boldsymbol{F}_{\mathrm{RF}} \approx \boldsymbol{I}_{N_{\mathrm{t}}^{\mathrm{RF}}}$ and $\boldsymbol{F}_{\mathrm{BB}}{}^{*}\boldsymbol{F}_{\mathrm{BB}} \approx \boldsymbol{I}_{N_{\mathrm{S}}}$. By the above argument, we have known that $\boldsymbol{F}_{\mathrm{RF}} = \left[\boldsymbol{a}_{\mathrm{t}}\left(\phi_1^{t}\right), \boldsymbol{a}_{\mathrm{t}}\left(\phi_2^{\mathrm{t}}\right), \ldots, \boldsymbol{a}_{\mathrm{t}}\left(\phi_{N_{\mathrm{t}}^{\mathrm{RF}}}^{\mathrm{t}}\right)\right]$, this assumes $|\alpha_i| \geqslant |\alpha_j|$, $\forall i < j$, which indicates the RF precoder columns can be resolved in an iteration.

Next, we consider the design of the base band digital precoder assuming that the RF precoder has been known. Therefore, $\boldsymbol{H}_{\mathrm{eff}} = \boldsymbol{H}\boldsymbol{F}_{\mathrm{RF}}$ can be seen as an effective CSI. Then, we can find the closed-form solution for the optimal digital precoder $\boldsymbol{F}_{\mathrm{BB}}$ by maximizing the spectral efficiency as follows:

$$\left.\begin{aligned} &\arg\max_{\boldsymbol{F}_{\mathrm{BB}}} \ \mathrm{lb}\left|\boldsymbol{I}_{N_{\mathrm{r}}} + \frac{1}{\sigma_n^2}\boldsymbol{H}_{\mathrm{eff}}\boldsymbol{F}_{\mathrm{BB}}\boldsymbol{F}_{\mathrm{BB}}^{*}\boldsymbol{H}_{\mathrm{eff}}^{*}\right| \\ &\text{s.t.} \\ &\quad \mathrm{tr}\left(\boldsymbol{F}_{\mathrm{RF}}^{*}\boldsymbol{F}_{\mathrm{RF}}\boldsymbol{F}_{\mathrm{BB}}\boldsymbol{F}_{\mathrm{BB}}^{*}\right) \leqslant \rho \end{aligned}\right\} \tag{9}$$

If we denote $\boldsymbol{H}_{\mathrm{e}} = \boldsymbol{H}_{\mathrm{eff}}\left(\boldsymbol{F}_{\mathrm{RF}}^{*}\boldsymbol{F}_{\mathrm{RF}}\right)^{1/2}$, the problem Eq. (7) should have an evident solution, i.e., $\boldsymbol{F}_{\mathrm{BB}} = \left(\boldsymbol{F}_{\mathrm{RF}}^{*}\boldsymbol{F}_{\mathrm{RF}}\right)^{1/2}$. $\boldsymbol{U}_{\mathrm{e}}\boldsymbol{\Sigma}_{\mathrm{e}}$, where $\boldsymbol{U}_{\mathrm{e}}$ is the set of eigenvectors corresponding to the N_{S} largest eigenvalues of $\boldsymbol{H}_{\mathrm{e}}^{*}\boldsymbol{H}_{\mathrm{e}}$, and $\boldsymbol{\Sigma}_{\mathrm{e}}$ is the diagonal matrix of powers allocated by water-filling.

Algorithm 1 The proposed AUHP algorithm

Require: $\boldsymbol{H} = \boldsymbol{A}_{\mathrm{r}}\mathrm{diag}(z)\boldsymbol{A}_{\mathrm{t}}^{*}$

Output: the HBF matrix $\boldsymbol{F}_{\mathrm{RF}}$ and $\boldsymbol{F}_{\mathrm{BB}}$.

1: [vec,pos]=sort(abs(z), 'descend')

2: $\boldsymbol{F}_{\mathrm{RF}} = \boldsymbol{A}_{\mathrm{t}}\left(:, \mathrm{pos}\left(1: N_{\mathrm{t}}^{\mathrm{RF}}\right)\right)$

3: $\boldsymbol{H}_{\mathrm{eff}} = \boldsymbol{H}\boldsymbol{F}_{\mathrm{RF}}$

4: $\boldsymbol{H}_{\mathrm{e}} = \boldsymbol{H}_{\mathrm{eff}}\left(\boldsymbol{F}_{\mathrm{RF}}^{*}\boldsymbol{F}_{\mathrm{RF}}\right)^{1/2}$

5: $\boldsymbol{F}_{\mathrm{BB}} = \left(\boldsymbol{F}_{\mathrm{RF}}^{*}\boldsymbol{F}_{\mathrm{RF}}\right)^{1/2}\boldsymbol{U}_{\mathrm{e}}\boldsymbol{\Sigma}_{\mathrm{e}}$

6: Return $\boldsymbol{F}_{\mathrm{RF}}$ and $\boldsymbol{F}_{\mathrm{BB}}$

By avoiding $N_{\mathrm{t}}^{\mathrm{RF}}$ iterations using asymptotically orthogonal AAR matrix $\boldsymbol{A}_{\mathrm{t}}$, the AUHP method enables to decrease the computational complexity compared to the SOMP.

Due to some limitations such as size of the mobile terminals, the number of receive antennas N_{r} does not meet the condition so that the matrix $\boldsymbol{A}_{\mathrm{r}}$ cannot be approximate to unitary matrix. Therefore, using the proposed AUHP method in MS can seriously hinder overall system performance. In order to avoid this serious impact, we adopt an improved iterative hybrid beamforming algorithm in MS. The sparse signal reconstruction algorithm via SOMP allows the HBF systems to approach their unconstrained theoretical limits on spectral efficiency.

$$\left.\begin{aligned} &\arg\min_{\boldsymbol{W}_{\mathrm{RF}}\boldsymbol{F}_{\mathrm{BB}}} \left\|\boldsymbol{R}_{yy}^{\frac{1}{2}}\left(\boldsymbol{W}_{\mathrm{MMSE}} - \boldsymbol{W}_{\mathrm{RF}}\boldsymbol{W}_{\mathrm{BB}}\right)\right\|_2^2 \\ &\text{s.t.} \\ &\quad \boldsymbol{W}_{\mathrm{RF}}(:, i); \quad \forall i \in \left\{\boldsymbol{a}_{\mathrm{r}}\left(\phi_l^{\mathrm{r}}\right),\ 1 \leqslant l \leqslant L\right\} \end{aligned}\right\} \tag{10}$$

where $\boldsymbol{R}_{yy} = \left(\rho/N_{\mathrm{S}}\right)\boldsymbol{H}\boldsymbol{F}_{\mathrm{RF}}\boldsymbol{F}_{\mathrm{BB}}\boldsymbol{F}_{\mathrm{BB}}^{*}\boldsymbol{F}_{\mathrm{RF}}^{*}\boldsymbol{H}^{*} + \sigma_n^2\boldsymbol{I}_{N_{\mathrm{r}}}$, and $\boldsymbol{W}_{\mathrm{MMSE}}$ is the optimal minimum mean square error

(MMSE) combiner.

Considering that the SOMP method needs $N_{\mathrm{r}}^{\mathrm{RF}}$ iterations and needs to compute the pseudo-inverse matrix at each iteration, which leads to high computational complexity. In order to guarantee the real time mmWave communication synchronization, we need to seek a low complexity algorithm with fewer iteration times. Therefore, we adopt the CoSaMP algorithm, which is on the basis of the SOMP algorithm but superior to the SOMP method. The improved CoSaMP scheme combines several improvement ideas from the existing literature in the field of image processing to accelerate the SOMP algorithm [15]. Meanwhile, results provide strong guarantee on performance.

In each iteration, the CoSaMP algorithm choose multiple alternative solutions using a matched filter per time. Then the set of obtained results should be combined with the support set $\boldsymbol{A}_{\mathrm{r}}$ that is estimated in the previous iteration to update the solution set until meet the end condition. Using least-squares scheme in the process of updating, this method can choose a new $N_{\mathrm{r}}^{\mathrm{RF}}$ (the number of RF chains) dimensional subspace from $\boldsymbol{A}_{\mathrm{r}}$ by reducing the beamforming error of the hybrid structure. Empirically we observed that the matched filter will weaken the performance of least-squares and hence deteriorate the system performance of the CoSaMP when the number of elements selected at a time is too large. Experiments show that choosing a fixed number of elements $2N_{\mathrm{r}}^{\mathrm{RF}}$ is appropriate. In this section, the specific CoSaMP algorithm steps are no longer described in detail.

3.3 Computational complexity analysis

In this section, we next want to analyze and compare the computational complexities for the three involved algorithms, namely the proposed AUHP algorithm, the SOMP algorithm in Ref. [8], and the beam steering algorithm in Ref. [7]. Considering that all three methods above need to compute the SVD of the CSI $\boldsymbol{H}$ to obtain the optimum fully digital reference unconstrained precoder $\boldsymbol{F}_{\mathrm{opt}}$ and combiner $\boldsymbol{W}_{\mathrm{opt}}$. So we compare their computational complexity after we have done the SVD.

Moreover, the distinction in computational complexity between the above three mainly displays in how to obtain the analog precoder $\boldsymbol{F}_{\mathrm{RF}}$ and the baseband digital precoder $\boldsymbol{F}_{\mathrm{BB}}$. The complexity of the proposed AUHP scheme to obtain $\boldsymbol{F}_{\mathrm{RF}}=\boldsymbol{A}_{\mathrm{t}}\left(:,\mathrm{pos}\left(1:N_{\mathrm{t}}^{\mathrm{RF}}\right)\right)$ is $O\left(N_{\mathrm{t}}N_{\mathrm{t}}^{\mathrm{RF}}\right)$, and the complexity to obtain the baseband digital precoder $\boldsymbol{F}_{\mathrm{BB}}$ by water-filling strategy is $O\left(N_{\mathrm{t}}^{2}N_{\mathrm{t}}^{\mathrm{RF}}\right)$. Therefore, the overall complexity of the AUHP method in transmitter is $O\left(N_{\mathrm{t}}^{2}N_{\mathrm{t}}^{\mathrm{RF}}\right)$. In Table 1, we lists their computational complexities of these three algorithms in detail. Then we could see that the beam steering algorithm in Ref. [7] has the lower computational complexity but suffer a poor performance especially the number of transceiver antennas is small. The SOMP method allows the HBF system to approach their optimal fully digital unconstrained performance on spectral efficiency, while its computational complexity is the highest. It is worth noting that our proposed AUHP algorithm reduces the computational complexity of the SOMP method and enables highly spectral efficiency simultaneously.

Table 1 Computational complexity

Algorithm	Computational complexity
Transmiter: beam steering	$O\left(N_{\mathrm{t}}^{2}N_{\mathrm{t}}^{\mathrm{RF}}\right)$
Receiver: beam steering	$O\left(N_{\mathrm{r}}^{2}N_{\mathrm{r}}^{\mathrm{RF}}\right)$
Transmiter: SOMP	$O\left(N_{\mathrm{t}}^{2}N_{\mathrm{t}}^{\mathrm{RF}}N_{\mathrm{S}}\right)$
Receiver: MMSE	$O\left(N_{\mathrm{r}}^{2}N_{\mathrm{r}}^{\mathrm{RF}}N_{\mathrm{S}}\right)$
Transmiter: AUHP	$O\left(N_{\mathrm{t}}N_{\mathrm{t}}^{\mathrm{RF}}\right)$
Receiver: CoSaMP	$O\left(N_{\mathrm{r}}N_{\mathrm{r}}^{\mathrm{RF}}\,\mathrm{lb}\,N_{\mathrm{r}}\right)$

4 Performance analysis

In this section, some simulation results are presented to show the performance of the proposed AUHP algorithm in comparison with the existing two methods presented above under mmWave communication channel model. In order to describe the propagation environment mentioned in Sect. 2, we establish a mmWave geometric channel model with N_{cl} clusters and N_{ray} propagation paths per pair of transceiver antennas. Moreover, assuming that the total transmission power constrained is same and the three algorithms adopt equal power allocation scheme. Then, $\eta=\rho/\sigma_{n}^{2}$ denotes the signal to noise ratio (SNR) of the received signal and the inter-antenna spacing is set to half a wavelength.

Firstly, the spectral efficiencies achieved by the optimal fully DBF and the HBF using beam steering method for different $N_{\mathrm{t}}\times N_{\mathrm{r}}$ with different SNR values are presented in Fig. 2. This implies that the loss due to beam steering

scheme increases as antenna array size decreases, which guides the thought of our proposed AUHP method.

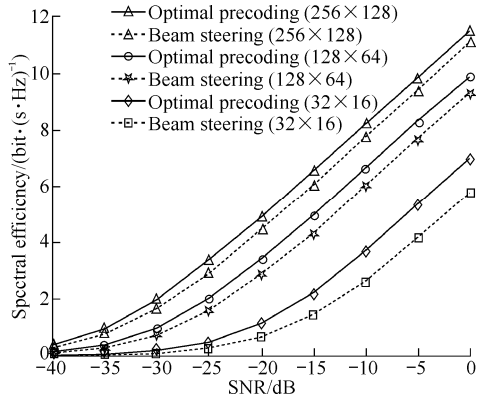

Fig. 2 Spectral efficiencies achieved by the unconstrained SVD-based fully DBF

We next want to simulate and compare the spectral efficiencies achieved by these four algorithms contains the optimal unconstrained beamforming. Fig. 3 presents the spectral efficiencies achieved by various beamforming methods for $N_t \times N_r = 64 \times 16$ mmWave communication systems with linear antenna arrays at the transmitter and receiver. The propagation channel model is a clusters environment with $N_{cl} = 8$, $N_{ray} = 10$. Then assuming that the HBF equipped with $N_t^{RF} = N_r^{RF} = 4$ RF chains.

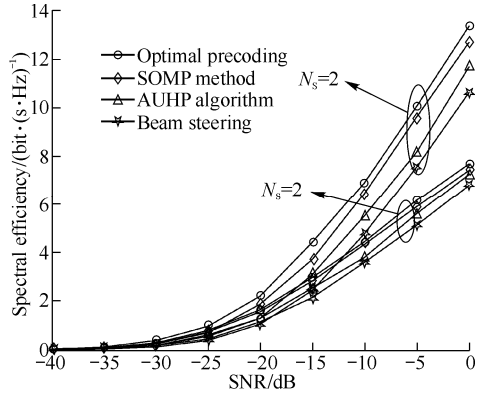

Fig. 3 Spectral efficiencies achieved by the four various beamforming methods for $N_t \times N_r = 64 \times 16$ mmWave systems

Fig. 4 shows the spectral efficiencies achieved by various beamforming schemes for $N_t \times N_r = 128 \times 32$ mmWave systems with linear antenna arrays at the transmitter and receiver.

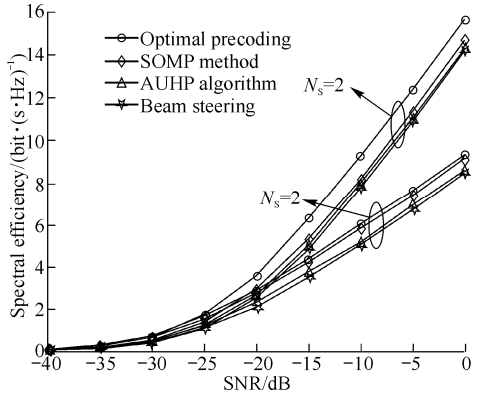

Fig. 4 Spectral efficiencies achieved by four various beamforming methods for $N_t \times N_r = 128 \times 32$ mmWave systems

Figs. 3 and 4 can explain that the proposed AUHP algorithm achieves spectral efficiencies which are essentially equal to those achieved by the SOMP solution in the case $N_S = 1$ and within a small performance loss in the case $N_S = 2$. These show that our proposed AUHP solution can accurately approximate the near-optimal (86%) performance of the iterative approximation optimal algorithm. Moreover, when compared to beam steering scheme in Ref. [7], there is a non-negligible performance improvement.

In order to analyze and explore spectral efficiencies in mmWave communication systems with larger scale antenna arrays, Fig. 5 shows the spectral efficiencies achieved by the four various beamforming methods for $N_t \times N_r = 256 \times 64$ mmWave systems with $N_t^{RF} = N_r^{RF} =$ 6 chains. Fig. 5 can illustrate that these three hybrid analog and DBF algorithm can get similar spectral efficiencies in both $N_S = 1$ and $N_S = 2$ cases when the transmitter and receiver have large enough antenna scale.

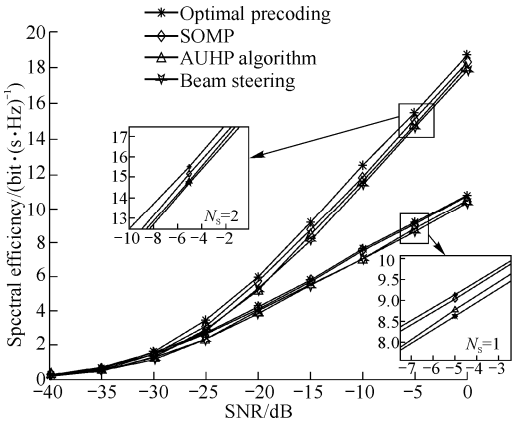

Fig. 5 Spectral efficiency achieved by various beamforming methods for $N_t \times N_r = 256 \times 64$ mmWave systems

5 Conclusions

In this paper, we consider hybrid analog and DBF architectures in mmWave communication systems and adopt a realistic mmWave sparse channel model. In this paper, in order to overcome the high hardware cost and power consumption and enhance the feasibility of mmWave communication, we propose an AUHP algorithm by exploiting the AAR properties. Numerical results and computational analysis can demonstrate that the AUHP algorithm enables a lower computational complexity with negligible spectral efficiency performance degradation, namely the proposed AUHP method can leverage the trade-off between the spectral efficiency performance and the computational complexity.

Appendix A Proof of Lemma 3

The N_{t} element ULAs take the form

$$\boldsymbol{a}_{\mathrm{t}}\left(\phi_l^{\mathrm{t}}\right)=\frac{1}{\sqrt{N_{\mathrm{t}}}}\left[1,\mathrm{e}^{\mathrm{j}\frac{2\pi}{\lambda}d\sin\phi_l^{\mathrm{t}}},...,\mathrm{e}^{\mathrm{j}(N_{\mathrm{t}}-1)\frac{2\pi}{\lambda}d\sin\phi_l^{\mathrm{t}}}\right]^{\mathrm{T}} \tag{A.1}$$

We exploit the dot products, in absolute value, between any two such distinct vectors $\boldsymbol{a}_{\mathrm{t}}(\phi_k^{\mathrm{t}})$ and $\boldsymbol{a}_{\mathrm{t}}(\phi_l^{\mathrm{t}})$ ($l\neq k$).

$$\left|\left\langle\boldsymbol{a}_{\mathrm{t}}(\phi_k^{\mathrm{t}}),\boldsymbol{a}_{\mathrm{t}}(\phi_l^{\mathrm{t}})\right\rangle\right|=\left|\boldsymbol{a}_{\mathrm{t}}(\phi_k^{\mathrm{t}})^{*}\boldsymbol{a}_{\mathrm{t}}(\phi_l^{\mathrm{t}})\right|=$$

$$\frac{1}{N_{\mathrm{t}}}\left|1+\mathrm{e}^{\mathrm{j}\frac{2\pi}{\lambda}d\left(\sin\phi_l^{\mathrm{t}}-\sin\phi_k^{\mathrm{t}}\right)}+...+\mathrm{e}^{\mathrm{j}2(N_{\mathrm{t}}-1)\frac{2\pi}{\lambda}\left(\sin\phi_l^{\mathrm{t}}-\sin\phi_k^{\mathrm{t}}\right)}\right|=$$

$$\frac{1}{N_{\mathrm{t}}}\left|\frac{1-\mathrm{e}^{\mathrm{j}\frac{2\pi}{\lambda}d\left(\sin\phi_l^{\mathrm{t}}-\sin\phi_k^{\mathrm{t}}\right)N_{\mathrm{t}}}}{1-\mathrm{e}^{\mathrm{j}\frac{2\pi}{\lambda}d\left(\sin\phi_l^{\mathrm{t}}-\sin\phi_k^{\mathrm{t}}\right)}}\right|=\frac{1}{N_{\mathrm{t}}}\left|\Sigma_{N_{\mathrm{t}}}\right| \tag{A.2}$$

where $\Sigma_{N_{\mathrm{t}}}$ is the sum of preceding N_{t} items. Since ϕ_k^{t} and ϕ_l^{t} are chosen independently from a continuous distribution, each term $\left\langle\boldsymbol{a}_{\mathrm{t}}(\phi_k^{\mathrm{t}}),\boldsymbol{a}_{\mathrm{t}}(\phi_l^{\mathrm{t}})\right\rangle$ is a geometric sequences with ratio $\mathrm{e}^{\mathrm{j}\frac{2\pi}{\lambda}d\left(\sin\phi_l^{\mathrm{t}}-\sin\phi_k^{\mathrm{t}}\right)}$ except the first one. And we have $\left(\sin\phi_l^{\mathrm{t}}-\sin\phi_k^{\mathrm{t}}\right)\neq 0$ with probability one, so the ratio $\mathrm{e}^{\mathrm{j}\frac{2\pi}{\lambda}d\left(\sin\phi_l^{\mathrm{t}}-\sin\phi_k^{\mathrm{t}}\right)}<1$. Thus, when N_{t} tends to infinity and the number of paths $L=o(N)$ in the channel, the $\Sigma_{N_{\mathrm{t}}}$ converges to 0. That is to say the array response vectors are orthogonal, i.e., $\boldsymbol{a}_{\mathrm{t}}\left(\phi_k^{\mathrm{t}}\right)\perp\boldsymbol{a}_{\mathrm{t}}\left(\phi_l^{\mathrm{t}}\right)$, $k\neq l$.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (61201134), State Key Science and Research Project (MJ-2014-S-37), and the 111 Project (B08038).

References

1. 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
2. Gao Z, Dai L L, Mi D, et al. Mmwave massive-MIMO-based wireless backhaul for the 5G ultra-dense network. IEEE Wireless Communications, 2015, 22(5): 13–21
3. Ding L, Liu R, Jiang B, et al. Limited feedback unitary precoding using improved euclidean distance metrics for spatial multiplexing systems. Proceedings of the 2010 International Conference on Wireless Communications and Signal Processing (WCSP'10), Oct 21–23, 2010, Suzhou, China. Piscataway, NJ, USA: IEEE, 2010: 6p
4. Shi X N, Siriteanu C, Yoshizawa S, et al. MIMO precoding performance for correlated and estimated Rician fading. Proceedings of the 11th International Symposium on Communications and Information Technologies (ISCIT'11), Oct 12–14, 2011, Hangzhou, China. Piscataway, NJ, USA: IEEE, 2011: 48–52
5. Han S, I C L, Xu Z K, et al. Large-scale antenna systems with hybird analog and digital beamforming for millimeter wave 5G. IEEE Communications Magazine, 2015, 53(1): 186–194
6. Jian A, Zhang J A, Huang X J, et al. Massive hybrid antenna array for millimeter-wave cellular communications. IEEE Wireless Communications, 2015, 22(1): 79–87
7. El Ayach C, Health R W, Abu-Surra S, et al. The capacity optimality of beam steering in large millimeter wave MIMO systems. Proceedings of the IEEE 13th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC'12), Jun 17–20, 2012, Cesme, Turkey. Piscataway, NJ, USA: IEEE, 2012: 100–104
8. El Ayach O, Rajagopal S, Abu-Surra S, et al. Spatially sparse precoding in millimeter wave MIMO systems. IEEE Transactions on Wireless Communications, 2014, 13(3): 1499–1513
9. Rusu C, Mendez-Rial R, Gonzalez-Prelcic N, et al. Low complexity hybird sparse precoding and combining in millimeter wave MIMO systems. Proceedings of the 2015 IEEE International Conference on Communications (ICC'15), Jun 8–12, 2015, London, UK. Piscataway, NJ, USA: IEEE, 2015: 1340–1345
10. Lee Y Y, Wang C H, Huang Y H. A hybrid RF/baseband precoding processor based on parallel-index-selection matrix-inversion-bypass simultaneous orthogonal matching pursuit for millimeter wave MIMO systems. IEEE Transactions on Signal Processing, 2015, 63(2): 305–317
11. Saleh A A M, Valenzuela R A. A statistical model for indoor multipath propagation. IEEE Journal on Selected Areas in Communications, 1987, 5(2): 128–137
12. Forenza A, Love D J, Health R W Jr. Simplified spatial correlation models for clustered MIMO channels with different array configurations. IEEE Transactions on Vehicular Technology, 2007, 56(4): 1924–1934
13. Zhang X Y, Molisch A F, Kung S Y. Variable-phase-shift-based RF-baseband codesign for MIMO antenna selection. IEEE Transactions on Signal Processing, 2005, 53(11): 4091–4103
14. Tropp J A, Anna C. Gilbert A C. Signal recovery from random measurements via orthogonal matching pursuit. IEEE Transactions on Information Theory, 2007, 53(12): 4656–4666
15. Zhang L. Image adaptive reconstruction based on compressive sensing via CoSaMP. Proceedings of the IEEE 2nd International Conference on Information Science and Control Engineering (ICISCE'15), Apr 24–26, 2015, Shanghai, China. Piscataway, NJ, USA: IEEE, 2015: 760–763

(Editor: Wang Xuying)