IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 22, 2019, accepted February 28, 2019, date of publication March 6, 2019, date of current version March 29, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2903166

Hybrid Precoding for Massive mmWave MIMO Systems

XIANRU LIU[1], XUEMING LI[2], SHU CAO[3], QINGYONG DENG[4,5], RONG RAN[6], (Member, IEEE), KIEN NGUYEN[7], (Senior Member, IEEE), AND PEI TINGRUI[3]

[1]School of automation, Central South University, Changsha 410086, China
[2]School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China
[3]Key Laboratory of Hunan Province for Internet of Things and Information Security, Xiangtan University, Xiangtan 411105, China
[4]School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China
[5]Key Laboratory of Hunan Province for Internet of Things and Information Security, Xiangtan University, Xiangtan 411105, China
[6]School of Electrical and Computer Engineering, Ajou University, Suwon 16499, South Korea
[7]Graduate School of Engineering, Chiba University, Chiba 263-8522, Japan

Corresponding author: Qingyong Deng (dengqingyong@xtu.edu.cn)

This work was supported in part by the Open Project Program of the Key Laboratory of Universal Wireless Communications under Grant 2016-KFKT-2016104, in part by the Ministry of Education, Beijing University of Posts and Telecommunications, in part by the Natural Science Foundation of China under Grant 61771414, Grant 61711540306, and Grant 61672447, and in part by the Natural Science Foundation of Hunan Province of China under Grant 2017JJ2249 and Grant 2018JJ1025.

ABSTRACT Due to high costs and power consumptions, fully digital baseband precoding schemes are usually prohibitive in millimeter-wave massive MIMO systems. Therefore, hybrid precoding strategies become promising solutions. In this paper, we present a novel real-time yet high-performance precoding strategy. Specifically, the eigenvectors corresponding to the larger eigenvalues of the right unitary matrix after singular value decomposition on an array response matrix are used to abstract the angle information of an analog precoding matrix. As the obtained eigenvectors correspond to the larger singular values, the major phase information of channels is captured. In this way, the iterative search process for obtaining the analog precoding vectors is avoided, and thus the hybrid precoding can be realized in parallel. To further improve its spectral-efficiency, we enlarge the resultant vector set by involving more relevant vectors in terms of their correlation values with the unconstrained optimal precoder, and a hybrid precoder is thus produced by using the vector set. The simulation results show that our proposed scheme achieves near the same performance as the orthogonal matching pursuit does, whereas it costs much fewer complexities than the OMP, and thus can be realized in parallel.

INDEX TERMS Millimeter wave communication, MIMO, wireless communication, hybrid precoding.

I. INTRODUCTION

As the capacity demand on wireless communication is increasing rapidly, the communication systems operating in microwave band are difficult to meet the quality-of-service (QoS) requirement. Hence, in the past several years, many advanced technologies such as relay cooperation, diversity multiplexing and cognitive radio are utilized to enhance spectral efficiency [1]–[5]. However, compared to the explosive increasing demand on system capacity, the obtained gain is too limited. Accordingly, exploring other radio frequency bands seems to be the essential solution to mitigate the crowded frequency bands.

To alleviate the shortage of frequency resource, millimeter wave (mmWave) is deemed as a potential radio frequency candidate. Due to its gigabit-per-second data rates [6]–[9], mmWave will be deployed in the future fifth generation (5G) wireless communication systems. Compared to the currently used frequency bands, the main characteristic of mmWave communication is its tremendous increase in carrier frequency. It implies that mmWave signals experience an orders-of-magnitude fading while propagating through the space. Therefore, it is difficult to ensure enough signal-to-noise ratio (SNR) for the conventional MIMO transceiver architecture. Fortunately, the substantially reduced wavelength enables the facility to put large antenna arrays in a much smaller space. Hence, mmWave systems can integrate massive multiple-input multiple-output (MIMO)

The associate editor coordinating the review of this manuscript and approving it for publication was Wei Xu.

transceiver elements to enhance signal gain and spectral efficiency. Consequently, massive mmWave MIMO architecture has been adopted as a standard in 5G communication systems [10]–[12], and becomes one of the research hotspots in the wireless communication domain.

For a massive mmWave MIMO system, to decrease the complexity of the receiver and cancel the interference among data streams, signals should be precoded before transmission. When fully digital baseband precoding is used, it is not difficult to precode multiple data streams to remove the inter-signal interference. However, an individual radio frequency (RF) chain is required for each antenna, which is impractical when the number of antennas is large. On the other hand, when a pure analog precoding mode is taken, RF chain can be shared among different antennas, and the cost in terms of hardware and power consumption can be reduced greatly. However, it is only applicable to the scenarios with single data stream, so it is not beneficial to improve spectral efficiency. Consequently, the methodology of combining the digital and analog modes to form a hybrid precoding scheme is introduced into massive mmWave MIMO systems. It not only determines precoders' structure, but also affects system performances [13]. Hence, a hybrid precoder which can achieve good performance has attracted a lot of research attentions.

In the past several years, much research are carried out around massive MIMO. In [14], the minimal requested number of RF chains and phase shifters in massive MIMO system is analyzed. For distributed massive MIMO systems, the benefits of beamforming training scheme are analyzed in [15], and its spectral efficiency with linear beamforming under pilot contamination is presented in [16]. For mixed-ADC/DAC multipair massive MIMO relaying systems, both exact and approximate closed-form expressions of the achievable rate are derived and a power scaling law is presented in [17]. For the similar system model, when two-way relaying with hardware impairments is considered, an approximation of the spectral efficiency with maximum ratio processing is derived in closed form in [18]. In [19], the design of directional hybrid precoding for multiuser mmwave communication systems with multiple eavesdroppers is investigated. To maximize the sum rate and the energy efficiency of massive MIMO systems, a codebook based hybrid precoding scheme with low complexity is proposed in [20].

In mmWave massive MIMO systems, the essence of hybrid precoding is to use two constrained matrices to approach the fully digital precoding matrix or inverted channel matrix. Accordingly, the intuitive solution of hybrid precoding is to decompose the fully digital precoding/channel matrix under given constraints. By directly decomposing the predesigned unconstrained digital precoder/combiner, a hybrid RF and baseband precoding/combining scheme for multi-stream transmission is proposed in [21]. In [22], an iterative precoding algorithm is proposed for mmWave systems, where all precoding operations are done in analog domain, and mmWave system can adapt the rank of its transmission in response to varying propagation conditions. In [23], a hybrid connected structure for hybrid beamforming is proposed, where precoding matrix optimization problem is decomposed into multiple precoding sub-matrix optimization problems and near-optimal hybrid digital and analog precoders are designed through factorizing the precoding sub-matrix for each sub-array. In [24], channel matrix is decomposed into an angle domain basis matrix and a gain matrix, which perfectly matches the structure of hybrid precoding. In [25], it is shown that the Gram matrix of frequency-selective channel can be decomposed into frequency-flat and frequency-selective components, which can be used for the analog and baseband precoders, respectively. All these research work are focused on decomposing the channel matrix of the mmWave large-scale MIMO systems through some mathematical methods. As the constraints of analog matrix should be considered during decomposition, the computational complexity is very high.

To decrease the complexity of hybrid precoder design, some low-complexity hybrid precoding schemes have been proposed. A new view of multi-user hybrid massive mMIMO systems from array signal processing perspective is also presented in [24], and a hybrid precoding and channel estimation strategies are proposed. In [26], an explicit analytical sum-rate expression for generic channel covariance-based beamforming scheme is derived, and a low-complexity joint statistical beamforming and user scheduling algorithm is proposed. In [27], exploiting the spatial structure of mmWave channels, precoding/combining is formulated as a sparse reconstruction problem, and a precoding algorithm is developed to accurately approximate optimal unconstrained precoder and combiner with low-cost RF hardware. Depending on the known array geometry and a low training and feedback, a hybrid analog/digital precoding algorithm is proposed in [28]. An angle-domain hybrid precoding is proposed in [29], and a novel channel tracking strategy is also presented. After the capacity optimization problem is decomposed into a series of subproblems that are easier to be handled by considering each antenna array one by one, a near-optimal iterative hybrid precoding scheme is proposed in [30] that is based on a more realistic subarray structure. An orthogonal matching and local search method is proposed to implement analog precoding in [31]. Based on the structure of mmWave channels, the design of a hybrid precoder is formulated as a sparsity constrained least squares problem, and a precoding algorithm is proposed to approximate the optimal unconstrained precoder in [32]. A hybrid precoding processor based on parallel-index-selection and orthogonal matching pursuit without matrix inversion is proposed in [33]. For multiuser massive MIMO systems, a hybrid block diagonalization scheme is proposed to approach the capacity performance of the traditional baseband digital method in [34]. By exploring the idea of turbo equalizer together with the tabu search algorithm, in [35], a novel hybrid precoding scheme is proposed to achieve near-optimal performance. In [36], after the relevance of directional precoding structures is established as a low-complexity and robust solution to meet the

demands of data rate, a simple class of directional schedulers is proposed based on the channel directional structure, and then the performance comparison between the single-user and multi-user scenarios is provided.

Though the performance of the aforementioned hybrid precoding schemes is very close to the optimal unconstrained precoding, most of them require an iterative search. Hence, the real-time performance is not very good. In this paper, after analyzing the relationship between antenna response matrix and its right singular-value-decomposition (SVD) matrix, the most related vectors are selected out, which serves as analog precoder after each element is normalized. To further improve system spectral efficiency, the selected vector set is enlarged such that the most relevant base vector can be involved in term of the correlation with the fully digital precoding matrix. In this way, the iterative search manner is not required any longer, and then the hybrid precoding scheme can achieve high spectral efficiency with a parallel manner. The main contributions of this paper are summarized as follows.

- As a baseline of hybrid precoder design, SVD is adopted to obtain the phase information of mmWave large-scale MIMO channel. After SVD is carried out on the array response matrix, the right resultant matrix consists of a group orthogonal basis, which can be used to form array response vectors. Hence, the problem of designing hybrid precoding is converted into searching the most relevant basis vectors, which is reflected on the value of eigenvalues, and can be realized easily.
- Instead of taking advantage of the traditional optimization techniques used in hybrid precoding, such as OMP, OMP+LS, the proposed hybrid precoding scheme utilizes SVD on array response matrix, and the most relevant antenna response vectors are selected as analog precoding matrix directly after being normalized. In this way, the iterative search process, which is a major ingredient that leads high computational complexity, is avoided, and a hybrid precoder can be obtained directly. Consequently, hybrid precoding is promising to realize in parallel mode.
- Unlike the schemes in [30], [31], and [27], where the analog precoding vectors are obtained one by one through an iterative OMP, the proposed scheme can select out all the analog precoding vectors simultaneously. Consequently, the execution time should be short so that the massive mmWave MIMO system becomes more real-time.
- The accuracy of hybrid precoding relative to the OMP scheme can be controlled through setting the number of radio frequency chains. As we know, if all the right normalized eigenvectors corresponding to nonzero eigenvalues are selected out as analog precoding vectors, all the antenna response vectors can be reconstructed. Hence, the required number of radio frequency chains can be determined according to the requested accuracy.

The rest of the paper is organized as follows. In Section II, after the system model of mmWave large-scale MIMO is briefly introduced, the hybrid precoding problem is formulated. The proposed hybrid precoding scheme in parallel and its improved version are presented specifically in Section III. We evaluate the performance of our solution schemes in Section IV. Finally, Section V concludes the paper and presents our future research.

Notations: A is a matrix; a is a vector; a or A is a scalar; $A^{(i)}$ or $A_{(i)}$ is the i^{th} column of A; $(\bullet)'$, $(\bullet)^*$, and $(\bullet)^{-1}$ denote the transpose, conjugate transpose, and inversion, respectively; $E[\bullet]$ denotes expectation operation.

II. SYSTEM DESCRIPTION

In this section, we first describe our considered system model. Then, the characteristic of mmWave MIMO channel is analyzed.

A. SYSTEM MODEL

We consider a single-user mmWave system in which one base station with N_{BS} antennas transmits N_s data streams to a receiver with N_{MS} antennas simultaneously. To support the function of transferring multiple data streams in a parallel mode, the transmitter and receiver are equipped with multiple radio frequency (RF) chains. The number of RF chains are denoted as N_{RF} and M_{RF}, and they satisfy $N_s \leq N_{RF} \leq N_{BS}$ and $N_s \leq M_{RF} \leq N_{MS}$ constraints, respectively.

In this paper, we focus on the precoder design at the transmitter due to the reciprocities between the transmitting process and receiving process. The similar approach can be utilized in obtaining the combiner at the receiver, and is omitted for brevity. For the N_s transmitted data streams, before going through RF chains, they should be precoded in the baseband. Denote the baseband precoder, analog precoder and transmitted symbol vector as $\mathbf{F}_{BB}$, $\mathbf{F}_{RF}$, and $\mathbf{s}$, respectively, and their dimensions are $N_{RF} \times N_s$, $N_{BS} \times N_{RF}$ and $N_s \times 1$. Suppose the transmit power of $\mathbf{s}$ is normalized, i.e., $E[\mathbf{s}\mathbf{s}^*] = \frac{1}{N_s}\mathbf{I}_{N_s}$. After precoding in the baseband domain, the resultant vector given as $\mathbf{F}_{BB}\mathbf{s}$ is the input of N_{RF} RF chains for the up conversion. Next, an analog precoder $\mathbf{F}_{RF}$ is applied for adjusting phase/angle to maximize the system capacity or minimize the interference according to the system requirements. Therefore, the discrete-time transmitted signals is finally represented as $\mathbf{x} = \mathbf{F}_{RF}\mathbf{F}_{BB}\mathbf{s}$. As the precoder $\mathbf{F}_{RF}$ achieved by an analog phase shifter only adjusts the phases rather than the amplitudes of signals, $(\mathbf{F}_{RF}^{(i)}\mathbf{F}_{RF}^{(i)*})_{l,l} = N_t^{-1}$ should hold for any i, where $\mathbf{F}_{RF}^{(i)}$ denotes the i^{th} column vector of $\mathbf{F}_{RF}$ and $(.)_{l,l}$ denotes the l^{th} diagonal element of a matrix. Meanwhile, the transmitted signals should guarantee the power constraint, which results in $||\mathbf{F}_{RF}\mathbf{F}_{BB}||_F^2 = N_s$.

For narrowband block-fading propagation scenarios, the received signals at the receiver can be shown as (1), where $\mathbf{y}$ is the $N_{MS} \times 1$ received signal vector, $\mathbf{H}$ is the channel matrix with $N_{MS} \times N_{BS}$ dimensions, ξ is the average received power, and $\mathbf{n}$ is the additive white Gaussian noise (AWGN) vector

following i.i.d distribution CN(0, σ_n^2).

$$\mathbf{y} = \sqrt{\xi}\mathbf{H}\mathbf{F}_{RF}\mathbf{F}_{BB}\mathbf{s} + \mathbf{n}. \quad (1)$$

Without loss of generality, we assume that the channel matrix $\mathbf{H}$ is known at both the transmitter and receiver.[1] At the receiver(s), the received signals are processed in the similar way to recover the original transmitted data streams. Accordingly, the finally processed signals can be shown as (2), where $\mathbf{W}_{BB}$ and $\mathbf{W}_{RF}$ denote the baseband and analog combiners, respectively.

$$\hat{\mathbf{y}} = \sqrt{\xi}\mathbf{W}_{BB}^*\mathbf{W}_{RF}^*\mathbf{H}\mathbf{F}_{RF}\mathbf{F}_{BB}\mathbf{s} + \mathbf{W}_{BB}^*\mathbf{W}_{RF} * \mathbf{n}. \quad (2)$$

Similar to the precoder, the analog combiner $\mathbf{W}_{RF}$ at the receiver only provides the phase adjustments. So the elements of $\mathbf{W}_{RF}$ should have the same amplitudes, i.e. $(\mathbf{W}_{RF}^{(i)}\mathbf{W}_{RF}^{(i)*})_{l,l} = N_r^{-1}$. Given the system model (2), its spectral efficiency is described as (3) [27], where $R_n = \sigma_n^2\mathbf{W}_{BB}^*\mathbf{W}_{RF}^*\mathbf{W}_{RF}\mathbf{W}_{BB}$ is the noise covariance matrix after the received signals are combined at the receiver. Therefore, for mmWave massive MIMO systems, the task of hybrid precoding is to maximize R through designing $\mathbf{F}_{BB}$, $\mathbf{F}_{RF}$, $\mathbf{W}_{BB}$ and $\mathbf{W}_{RF}$ jointly.

$$R = \log_2(|I_{N_s} + \frac{\xi}{N_s}R_n^{-1}\mathbf{W}_{BB}^*\mathbf{W}_{RF}^*\mathbf{H}\mathbf{F}_{RF}\mathbf{F}_{BB} \times \mathbf{F}_{BB}^*\mathbf{F}_{RF}^*\mathbf{H}^*\mathbf{W}_{RF}\mathbf{W}_{BB}|). \quad (3)$$

B. CLUSTER-BASED CHANNEL MODEL

Due to the characteristics of severe path-loss and high level of correlation among antennas, the traditional channel model is not applicable for mmWave large-scale MIMO transceiver array. Consequently, a narrow-band clustered channel representation based on the extended Saleh-Valenzuela model is usually adopted [38]–[40], where the channel matrix $\mathbf{H}$ is assumed to be a sum of the contributions of N_{cl} different clusters, each of which has N_{ray} propagation paths. Accordingly, the channel model of mmWave massive MIMO systems can be abstracted as (4), where γ is a normalization factor, α_{il} is the complex gain of the l^{th} propagation path in the i^{th} cluster. Functions $\Lambda_r(\bullet)$ and $\Lambda_t(\bullet)$ represent the gains of transmit and receive antenna, respectively, that correspond to different departure and arrival angles. $\mathbf{a}_r(\bullet)$ and $\mathbf{a}_t(\bullet)$ denote the normalized respective receive and transmit array response vectors when azimuth and elevation angles vary, and they are independent [27]. Moreover, ϕ_{il}^r (θ_{il}^r) and ϕ_{il}^t (θ_{il}^t) present the arrival and departure azimuth (elevation) angles, respectively, of the l^{th} ray in the i^{th} cluster.

$$\mathbf{H} = \gamma\sum_{i,l}\alpha_{i,l}\Lambda_r(\phi_{il}^r,\theta_{il}^t)\Lambda_t(\phi_{il}^r,\theta_{il}^t)\mathbf{a}_r(\phi_{il}^r,\theta_{il}^t)\mathbf{a}_t(\phi_{il}^r,\theta_{il}^t)^*. \quad (4)$$

Note that when uniform planar array structure is considered, both $\mathbf{a}_r(\bullet)$ and $\mathbf{a}_t(\bullet)$ can be expressed as the format given as (5), where W and H are the numbers of antenna elements, $k = \frac{2\pi}{\lambda}$ (λ is the corresponding wavelength) and d is the inter-element spacing between two neighbor antennas, $N = WH$ is the size of the antenna array. For simplicity, we use $\mathbf{A}_t$ and $\mathbf{A}_r$ to represent the matrices consisting of all array response vectors $\mathbf{a}_t(\bullet)$ and $\mathbf{a}_r(\bullet)$ respectively. These matrix representations will be used in the sequel.

$$\mathbf{a}_{UPA}(\phi,\theta) = \frac{1}{\sqrt{N}}[1,\cdots,e^{kd(msin(\phi)sin(\theta)+ncos(\theta))}, \cdots,e^{kd((W-1)sin(\phi)sin(\theta)+(H-1)cos(\theta))}]^T. \quad (5)$$

III. HYBRID PRECODING STRATEGY

In this section, after analyzing the relation between the channel matrix and antenna array response vectors, the problem of hybrid precoding is formulated. Then, a novel low-complexity hybrid precoding strategy and its improved version are proposed.

A. PROBLEM FORMULATION

To maximize R in (3), $\mathbf{F}_{BB}$, $\mathbf{F}_{RF}$, $\mathbf{W}_{BB}$ and $\mathbf{W}_{RF}$ should be optimal. Nevertheless, due to the non-convex constraints of $\mathbf{F}_{RF}$ and $\mathbf{W}_{RF}$, it is extremely difficult to tackle the optimization problem maximizing R via $\mathbf{F}_{RF}$, $\mathbf{W}_{BB}$ and $\mathbf{W}_{RF}$. Consequently, the problem of designing analog/digital precoders and combiners is usually decoupled into two independent sub-problems. One is designing precoders for the transmitter, another is designing combiners for the receiver, and both of them follow the similar approach to obtain solutions. At the transmitter, for data streams $\mathbf{s}$, the achievable mutual information of the mmWave massive MIMO system can be shown as (6).

$$I(\mathbf{F}_{RF},\mathbf{F}_{BB}) = \log_2(|\mathbf{I} + \frac{\xi}{N_S\sigma_n^2}\mathbf{H}\mathbf{F}_{RF}\mathbf{F}_{BB}\mathbf{F}_{BB}^*\mathbf{F}_{RF}^*\mathbf{H}^*|). \quad (6)$$

To maximize $I(\mathbf{F}_{RF},\mathbf{F}_{BB})$ in (6), the classical methods usually consist of two stages. First, considering the entirety of $\mathbf{F}_{RF}\mathbf{F}_{BB}$ and omitting all hardware limitations, the mutual information can be maximized by $\mathbf{F}_{opt} = \mathbf{V}(:,1:N_s)$, where the SVD of $\mathbf{H}$ is $\mathbf{U}\Sigma\mathbf{V}^*$. $\mathbf{F}_{opt}$ is the fully optimal baseband precoding matrix, which consists of the N_s column vectors in $\mathbf{V}$ that correspond to the highest singular values in Σ. Second, considering the hardware constraints, $\mathbf{F}_{RF}$ and $\mathbf{F}_{BB}$ are designed using some optimization strategies to approach $\mathbf{F}_{opt}$. The commonly used method of designing $\mathbf{F}_{RF}$ is to select the most relevant array response vectors. As a result, we formulate the optimization problem as [27],

$$\begin{aligned}&\arg\min_{\mathbf{A}_t,\hat{\mathbf{F}}_{BB}}\{||\mathbf{F}_{opt}-\mathbf{A}_t\hat{\mathbf{F}}_{BB}||_F\}\\ &\text{subject to } ||\text{diag}(\hat{\mathbf{F}}_{BB}\hat{\mathbf{F}}_{BB}^*)||_0 = N_{RF}\\ &\qquad ||\mathbf{A}_t\hat{\mathbf{F}}_{BB}||_F^2 = N_s. \end{aligned} \quad (7)$$

where $\mathbf{A}_t$ is the antenna response matrix with the dimension of $N_{BS}\times N_{cl}N_{ray}$ (N_{cl} and N_{ray} are the numbers of scattering cluster and propagation paths per cluster, respectively, in clustered channel model, which will be described briefly in the

[1]For practical mmWave massive MIMO systems, the channel state information (CSI) can be obtained through channel estimation technology at receiver, and feedback to transmitter.

next subsection), and $||\bullet||_0$ denotes pseudo norm returning the number of non-zero elements. Implicitly, by searching the most suitable antenna response vectors in $\mathbf{A}_t$, the constraints of the analog precoder can be neglected.

Once the analog precoder $\mathbf{F}_{RF}\mathbf{F}_{BB}$ has been designed, the combiner $\mathbf{W}_{BB}\mathbf{W}_{RF}$ at the receiver can be obtained via the similar approach. Let us assume $\mathbf{W}_{opt} = \mathbf{W}_{BB}\mathbf{W}_{RF}$ and relax the hardware limitations, the optimal $\mathbf{W}_{opt}$ can be achieved by minimizing the mean squared error (MMSE) between the transmitted and processed received signals [37]. Then, the problem of designing hybrid combiners can be formulated as (8), where $\mathbf{y}$ is the received signal vector, and Ω_{RF} is a constant-gain phase-only entries with $N_{MS} \times M_{RF}$ size.

$$\begin{aligned} &\arg \min_{\mathbf{W}_{RF},\mathbf{W}_{BB}} \{||E[\mathbf{y}\mathbf{y}^*]^{1/2}\mathbf{W}_{opt} - \mathbf{W}_{RF}\mathbf{W}_{BB}||_F\} \\ &\text{subject to } \mathbf{W}_{RF} \in \Omega_{RF}. \end{aligned} \tag{8}$$

Note that the process to solve (8) and (6) at the transmitter and receiver are similar, and hence we only present the strategy to design hybrid precoder in this paper, and the method to optimize the hybrid combiners is omitted for the sake of brevity.

Note that the column vectors of the fully digital precoding matrix $\mathbf{F}_{opt}$ are orthogonal [27]. However, they are not necessarily complete orthogonal bases for the channel's row space since some eigenvectors corresponding to non-zero eigenvalues are not selected out due to the constraint in the number column vectors (only N_S column vectors are taken out). On the other hand, from (4), it can be seen that the linear combination of all array response vectors $\mathbf{a}_t(\phi_{il}^t, \theta_{il}^t)$ forms the channel matrix $\mathbf{H}$. Hence, it can be concluded that the linear combination of $a_t(\phi_{il}^t, \theta_{il}^t)$ can be used to present F_{opt} accurately. Therefore, we can argue that hybrid precoder design is to find two smaller matrices $\mathbf{F}_{RF}$ and $\mathbf{F}_{BB}$ such that $\mathbf{F}_{opt} = \mathbf{F}_{RF}F_{BB}$. Moreover, $\mathbf{F}_{BB}$ should be unit-amplitude which implies that the column vectors of $\mathbf{F}_{BB}$ is able to present $\mathbf{F}_{opt}$. Design (7) can be equally converted into the following problem

$$\begin{aligned} (\mathbf{F}_{RF}^{opt}, \mathbf{F}_{BB}^{opt}) = \arg \min &||\mathbf{F}_{opt} - \mathbf{F}_{RF}\mathbf{F}_{BB}||_F, \\ &\text{subject to } \mathbf{F}_{RF}^{(i)} \in \{\mathbf{a}_t(\phi_{il}^t, \theta_{il}^t), \forall i, l\}, \\ &||\mathbf{F}_{RF}\mathbf{F}_{BB}||_F^2 = N_s. \end{aligned} \tag{9}$$

B. SOLUTION STRATEGY

To solve the optimization problem in (9), its essence is to select out the N_{RF} most relevant antenna response vectors from $\{\mathbf{a}_t(\phi_{il}^t, \theta_{il}^t), \forall i, l\}$, which is denoted as $\mathbf{A}_t$ for notational simplicity. In the existing hybrid precoding schemes, optimal matching strategies are commonly used, such as [27] and [31]. Though the performance of these hybrid precoding schemes is very close to that of the fully digital baseband precoding scheme, its computational complexity is exponentially increasing with the number of transceiver antennas, which leads to huge latency. In fact, for the antenna response matrix $\mathbf{A}_t$, if an orthogonal basis can be used to present it, the corresponding phase information is obtained, and an approximation analog precoder can be constructed. In this way, hybrid precoding can be realized without using any iterative search process, and computational complexity can be reduced greatly.

For antenna response matrix $\mathbf{A}_t'$, when it is decomposed using SVD, it can be shown as $\mathbf{A}_t' = \mathbf{U}\Sigma\mathbf{V}^*$, where the dimensions of $\mathbf{A}_t'$, $\mathbf{U}$, Σ and $\mathbf{V}$ are $N_{cl}N_{ray} \times N_{BS}$, $N_{cl}N_{ray} \times N_{cl}N_{ray}$, $N_{cl}N_{ray} \times N_{BS}$ and $N_{BS} \times N_{BS}$, respectively. To see more clear relationship between $\mathbf{A}_t'$ and $\mathbf{V}$, let us assume $\boldsymbol{\beta} = \mathbf{U}\Sigma$. Then, let $\boldsymbol{\beta}_i$, $\mathbf{A}_i$ and $\mathbf{V}_i$ denote the i^{th} row/row/column of $\boldsymbol{\beta}$, $\mathbf{A}_t'$ and $\mathbf{V}$, respectively, for the sake of convenience. Consequently, the following relation can be obtained.

$$\mathbf{A}_i = [\boldsymbol{\beta}_i V_1^* \ \boldsymbol{\beta}_i \mathbf{V}_2^*, \cdots, \boldsymbol{\beta}_i \mathbf{V}_{N_{BS}}^*] = \boldsymbol{\beta}_i[\mathbf{V}_1^* \ \mathbf{V}_2^*, \cdots, \mathbf{V}_{N_{BS}}^*]. \tag{10}$$

It implies that the orthogonal basis $\mathbf{V}_i^* \in \forall i$ can be used to express $\mathbf{A}_i$. For (10), as $\boldsymbol{\beta}_i$ can be removed into the part of digital precoder, the remaining task of the hybrid precoding is to select out the most relevant orthogonal vectors from $\mathbf{V}$. Therefore, searching the most relevant vectors in $\mathbf{A}_t$ successively is equivalent to obtaining the most suitable orthogonal basis which can be realized in the parallel model.

For $\mathbf{V}_i^*$, though its norm is one, the norms of its element is not necessary one. Hence, its elements should be normalized. let $\mathbf{V}_i^* = [\mathbf{V}_{1i}, \mathbf{V}_{2i}, \cdots, \mathbf{V}_{N_{BS}i}]$. Then, it can be rewritten as

$$\begin{aligned} &[\mathbf{V}_{1i}, \mathbf{V}_{2i}, \cdots, \mathbf{V}_{N_{BS}i}] \\ &\quad = [\frac{\mathbf{V}_{1i}}{|\mathbf{V}_{1i}|}, \frac{\mathbf{V}_{2i}}{|\mathbf{V}_{2i}|}, \cdots, \frac{\mathbf{V}_{N_{BS}i}}{|\mathbf{V}_{N_{BS}i}|}] \\ &\qquad \times \begin{bmatrix} |\mathbf{V}_{1i}| & \times 0 & \times \cdots, & \times 0 \\ 0 & \times |\mathbf{V}_{2i}| & \times \cdots, & \times 0 \\ \cdots & \times \cdots & \times \cdots, & \times \cdots \\ 0 & \times 0 & \times \cdots, & \times |\mathbf{V}_{N_{BS}i}| \end{bmatrix}, \end{aligned} \tag{11}$$

in which $[\frac{\mathbf{V}_{1i}}{|\mathbf{V}_{1i}|}, \frac{\mathbf{V}_{2i}}{|\mathbf{V}_{2i}|}, \cdots, \frac{\mathbf{V}_{N_{BS}i}}{|\mathbf{V}_{N_{BS}i}|}]$ can be realized with the analog phase shifter, and

$$\begin{bmatrix} |\mathbf{V}_{1i}| & 0 & \cdots, & 0 \\ 0 & |\mathbf{V}_{2i}| & \cdots, & 0 \\ \cdots & \cdots & \cdots, & \cdots \\ 0 & 0 & \cdots, & |\mathbf{V}_{N_{BS}i}| \end{bmatrix}$$

can be removed from the part of baseband precoder. From the above analysis, if all the vectors corresponding to non-zero eigenvalues are taken out for the construction of analog phase shifter, the performance of the hybrid precoding in this way is equal to that of the OMP scheme. Obviously, as the number of RF chains is limited,[2] the construction of $\mathbf{A}_t$ is probably not perfect. Consequently, its performance is degraded compared to the optimal iterative search scheme.

Based on the above analysis, the proposed hybrid precoding scheme can be summarized as *Algorithm 1*[3].

[2]It can be seen that the number of RF chains is equal to the number of the selected vectors from $\mathbf{A}_t$.

[3]norm($\bullet$) function means normalizing each elements of the matrix.

Algorithm 1 Low-Complexity Hybrid Precoding Algorithm

Obtain F_{opt} through SVD on $\mathbf{H}$.
$\mathbf{A}_t' = \mathbf{U}\Sigma\mathbf{V}^*$
$\mathbf{F}_{RF} = [\mathbf{V}_1, \mathbf{V}_2, \cdots, \mathbf{V}_{N_{RF}}]^*$.
$\mathbf{F}_{RF} = norm(\mathbf{F}_{RF})$
$\mathbf{F}_{BB} = (\mathbf{F}_{RF}^*\mathbf{F}_{RF})^{-1}\mathbf{F}_{RF}^*\mathbf{F}_{opt}$
return $\mathbf{F}_{RF}, \mathbf{F}_{BB}$

Algorithm 2 The Improved Low-Complexity Hybrid Precoding Process

Require : $\mathbf{F}_{opt}$.
$\mathbf{A}_t' = \mathbf{U}\Sigma\mathbf{V}^*$.
$\tilde{\mathbf{F}}_{RF} = [\mathbf{V}_1, \mathbf{V}_2, \cdots, \mathbf{V}_n]^*$.
$\Psi = \tilde{\mathbf{F}}_{RF}'\mathbf{F}_{opt}$.
$\Omega = \mathrm{diag}(\Psi\Psi^*)$.
$[Y, I] = \mathrm{sort}(\Omega, 'descend')$.
$\hat{\mathbf{F}}_{RF} = \tilde{\mathbf{F}}_{RF}(:, I(1 : N_{RF}))$.
$F_{RF} = \mathrm{norm}(\hat{F_{RF}})$.
$\mathbf{F}_{BB} = (\mathbf{F}_{RF}^*\mathbf{F}_{RF})^{-1}\mathbf{F}_{RF}^*\mathbf{F}_{opt}$.
return $\mathbf{F}_{RF}, \mathbf{F}_{BB}$.

For mmWave massive MIMO transceiver, as $N_{RF} << N_{BS}$ and $N_{RF} << N_{cl}N_{ray}$, it is incomplete that the N_{RF} most relevant orthogonal basis vectors are used to approximate the antenna response matrix $\mathbf{A}_t$. Furthermore, in the OMP hybrid precoding schemes, just the N_{RF} relevant antenna response vectors in $\mathbf{A}_t$ are selected out as the hybrid precoder. Consequently, there is a gap between the performance of *Algorithm 1* and that of the fully digital baseband precoding scheme. As a compromise, an enlarged candidate orthogonal basis set is considered next, which means more orthogonal basis vectors are selected according to the correlation with the fully digital baseband precoder $\mathbf{F}_{opt}$. This idea is detailedly demonstrated in *Algorithm 2*, where n is the parameter used to set the scale of the candidate orthogonal basis vectors and $N_{RF} \leq n \leq N_{BS}$.

C. REAL-TIME ANALYSIS

To see the computational complexity of the proposed hybrid precoding scheme, we compare it with the classical hybrid precoding schemes in [27] and [31]. From *Algorithm 1* and *Algorithm 2*, it can be seen that most of the computing operations are concentrated on the SVD, sorting and matrix inversion (but just be required once). The traditional hybrid precoding schemes, such as OMP and OMP+LS, usually consist of an iterative search process, in which sorting, matrix inversion and normalization operations are repeated for each iteration, and are time-consuming. Consequently, the proposed hybrid precoding is more real-time than the traditional schemes.

IV. PERFORMANCE EVALUATION

In this section, numerical simulations are presented to show the performance of our proposed hybrid precoding schemes.

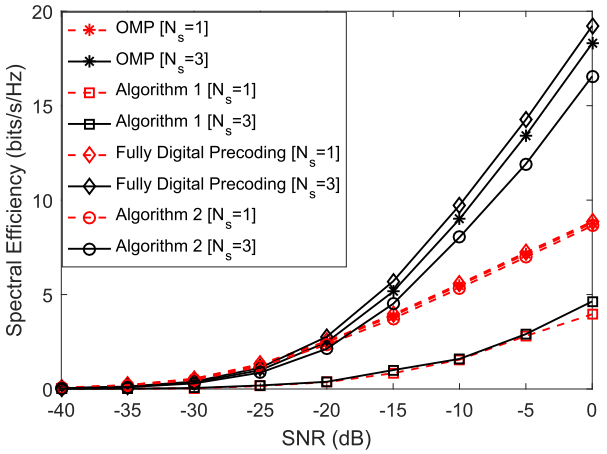


FIGURE 1. The spectral efficiency of four precoding schemes with the varying SNR, where $N_S = 1$ and $N_S = 3$ are set under the conditions $N_{BS} = 64$, $N_{MS} = 16$ and $N_{RF} = M_{RF} = 4$, respectively.

In the simulation, the cluster-based channel model is considered, in which $N_{cl} = 8$ and there are $N_{ray} = 10$ rays per cluster. Both the arrival and departure azimuth and elevation angles of each cluster follows the Laplacian distribution. Similar to [27] and [41], the spread angle in azimuth and elevation angles are also set as 20° and 60°, respectively. Moreover, it is also assumed that all clusters have the equal transmit power. In practical mmWave massive MIMO systems, the base stations are usually more capable to configure larger massive transceiver antenna array compared to the mobile terminals. Hence, in our simulation, we let the number of antennas at the base station be larger than that of the mobile user, and the scenarios with $N_{BS} = 64$, $N_{MS} = 16$ and $N_{BS} = 256$, $N_{MS} = 64$ are considered.

Fig. 1 and Fig. 2 show the spectral efficiency with the varying SNR under the condition that $N_{BS} = 64$, $N_{MS} = 16$ for $N_{RF} = M_{RF} = 4$ and $N_{RF} = M_{RF} = 8$, respectively. For the purpose of comparison, the simulation results of the fully digital precoding and OMP schemes are presented as well. In order to identify the impact of the number of data streams on the system spectral efficiency, we also consider $N_s = 1$ and 3 cases.

From Fig. 1 and Fig. 2, we can observe that the spectral efficiency achieved by *Algorithm 1* and *Algorithm 2* approach to that of the OMP scheme with respect to the number of RF chains. For example, in Fig. 2, with $N_{RF} = M_{RF} = 8$, the spectral performance of *Algorithm 2* is almost as same as that of the OMP and *Fully Digital Precoding* schemes. That is because, the more RF chains are available, the more basis vectors can be used to generate the more accurate representative matrix of the antenna response matrix $\mathbf{A}_t$.

It is not surprise that *Algorithm 2* always outperforms *Algorithm 1*. But we notice that their performance gap decreases as the number of RF chain increases. The essential of our proposed schemes is how to select the relevant basis vectors to construct a new matrix which can accurately represent the antenna response matrix $\mathbf{A}_t$. In *Algorithm 1*,

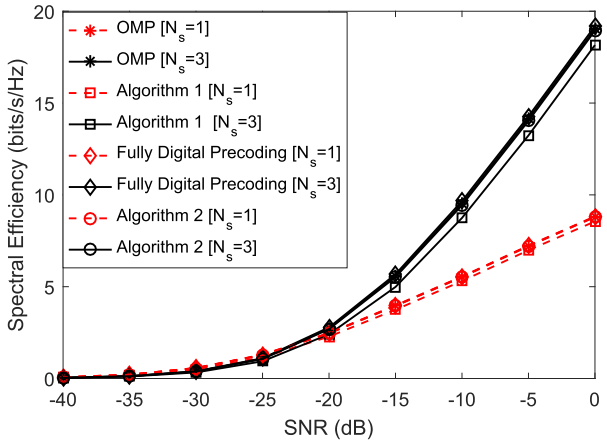


FIGURE 2. The spectral efficiency of four precoding schemes with the varying SNR, where $N_S = 1$ and $N_S = 3$ are set under the conditions $N_{BS} = 64$, $N_{MS} = 16$ and $N_{RF} = M_{RF} = 8$, respectively.

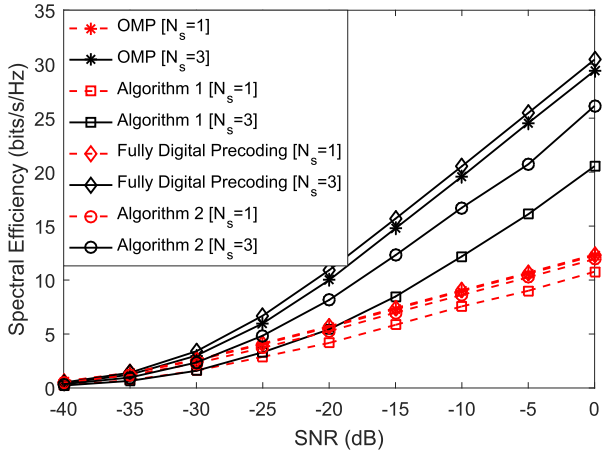


FIGURE 3. The spectral efficiency of four precoding schemes with the varying SNR, where $N_S = 1$ and $N_S = 3$ are set under the conditions $N_{BS} = 256$, $N_{MS} = 64$ and $N_{RF} = M_{RF} = 4$, respectively.

the number of basis vectors of $\mathbf{A}_t$, which is the same as that of RF chains, is used to produce an analogy precoder. Since the channel matrix $\mathbf{H}$ consists of vectors of $\mathbf{A}_t$, the optimal analogy precoder should be heavily related to $\mathbf{A}_t$. But the number of RF chain limits the number of basis vector candidates which we can select to construct the representative matrix, there is not guarantee that the selected vectors can contribute enough to produce a good precoder. However, in *Algorithm 2*, a larger number of eigenvectors are allowed to be correlated with the optimal precoder, so more vector candidates can be used to determine the most relevant basis vectors. Therefore, a better precoder is possible to be produced compared to *Algorithm 1*, which is confirmed by simulation results. Furthermore, when the number of RF chains increases, the probability of selecting the most relevant basis vectors becomes larger, and hence the spectral efficiency gap is reduced.

To show the impact of antenna array size on the spectral efficiency, the simulation results for the cases $N_{BS} = 256$ and $N_{MS} = 64$ are presented in Fig. 3 and Fig. 4 for

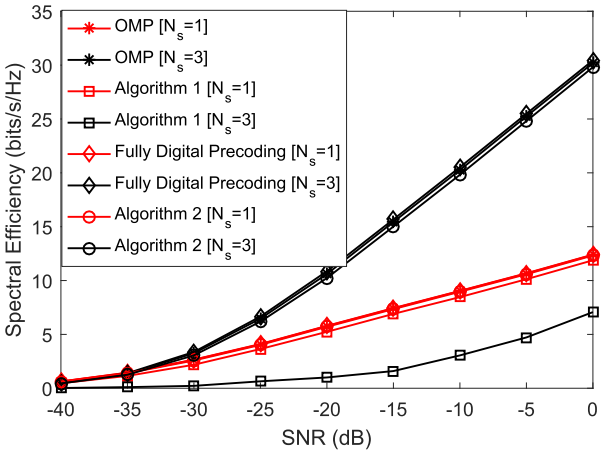


FIGURE 4. The spectral efficiency of four precoding schemes with the varying SNR, where $N_S = 1$ and $N_S = 3$ are set under the conditions $N_{BS} = 256$, $N_{MS} = 64$ and $N_{RF} = M_{RF} = 8$, respectively.

$N_{RF} = 4$ and $N_{RF} = 8$, respectively. In Fig. 3, we can observe that the performance gap between the proposed *Algorithm 2* and the OMP gets bigger compared to Fig. 1 due to the increased number of antennas. It is because that when the size of the antenna array increases, the number of transmission paths increase. Consequently, the components that construct $\mathbf{A}_t$ becomes more complex, and it is more difficult to construct $\mathbf{A}_t$ by using the orthogonal vectors. However, when we increase the number of RF chains, as illustrated in Fig. 4, *Algorithm 2* approaches the OMP for both $N_s = 1$ and 3 cases, while costs much less complexities.

In Fig. 5, we consider three cases that $N_s = 1$, 2 and 3 under the conditions that $N_{BS} = 64$, $N_{MS} = 16$ and $N_{RF} = M_{RF} = 8$. From simulation results, we can observe that the spectral efficiencies of the proposed two algorithms increase in term of the number of data streams, which follows the general behaviors of MIMO systems. In Fig. 6, we preserve $N_{RF} = M_{RF} = 8$ and only increase the numbers of transmit and receive antennas up to 256 and 64, respectively. The spectral efficiency of the proposed *Algorithm 2* increases as the number of data streams increases. But the proposed *Algorithm 1* with $N_s = 3$ does not achieve the best performance. That is because the precoder via *Algorithm 1* with the degree of freedom of $N_{RF} = M_{RF} = 8$ is not robust enough to deal with additional interference caused by the increased data streams.

Therefore, in order to learn the impact of the number of data streams on the spectral efficiency thoroughly, we illustrate the spectral efficiencies with respect to the number of data streams when SNR is given as $0dB$ in Fig. 7. From simulation results, we can observe that the fully digital precoding scheme and OMP follow the general behaviors of MIMO systems that the spectral efficiency near linearly increases as the number of data streams at the middle or high SNR regimes. Our proposed *Algorithm 2* behaves the same as the OMP when N_s is less than 6, but starts to degrade when N_s is over 6, and the proposed *Algorithm 1* degrades from $N_s = 2$. It is not

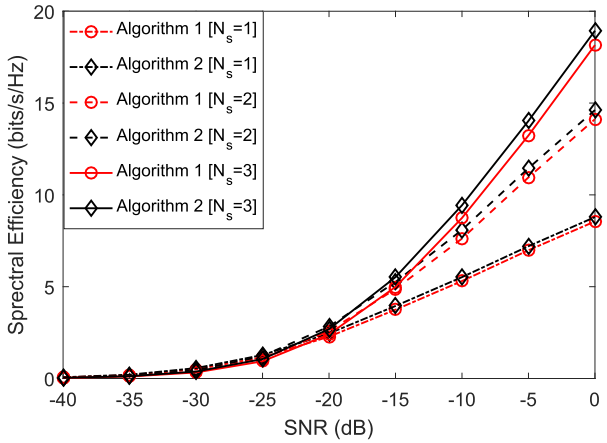


FIGURE 5. The spectral efficiency of the proposed hybrid precoding schemes with the varying SNR, where $N_s = 1$, $N_s = 2$ and $N_s = 3$ are set under the conditions $N_{BS} = 64$, $N_{MS} = 16$ and $N_{RF} = M_{RF} = 8$, respectively.

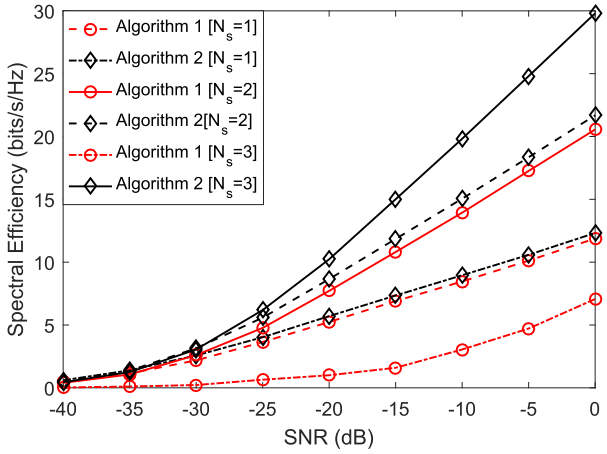


FIGURE 6. The spectral efficiency of the proposed hybrid precoding schemes with the varying SNR, where $N_s = 1$, $N_s = 2$ and $N_s = 3$ are set under the conditions $N_{BS} = 256$, $N_{MS} = 64$ and $N_{RF} = M_{RF} = 8$, respectively.

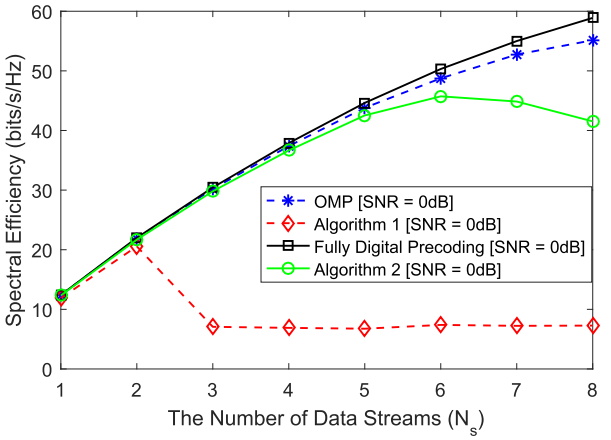


FIGURE 7. The spectral efficiency of four precoding schemes with the increasing N_s, where SNR = 0dB is set under the conditions $N_{BS} = 256$, $N_{MS} = 64$ and $N_{RF} = M_{RF} = 8$, respectively.

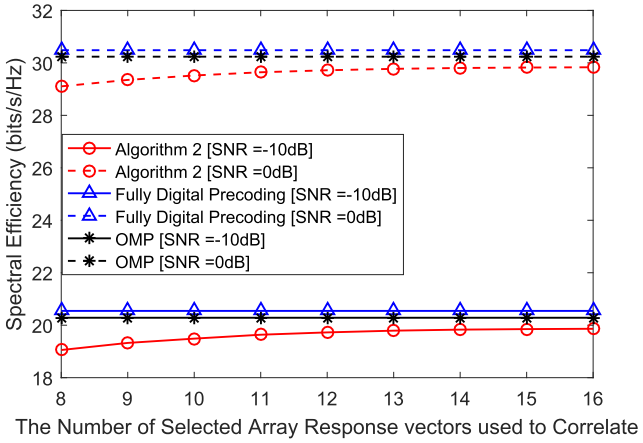


FIGURE 8. The spectral efficiency with the increasing number of selected array response vectors to correlate with the fully digital precoding matrix, where SNR = $-10dB$ and SNR = 0dB are set under the conditions $N_{BS} = 256$, $N_{MS} = 64$ and $N_{RF} = M_{RF} = 8$, respectively.

difficult to understand that given $\mathbf{A}_t$, the selected orthogonal vectors which are relevant to F_{opt}, become the loose representatives of the vectors of F_{opt} when the number of data stream increases. Therefore, the interference cancellation capability of the precoder is degraded, which results in the performance loss. The issue can be addressed by increasing the number of candidate vectors used to generate the precoder (i.e., *Algorithm 2*), but the complexity is increased as well. There arises a tradeoff between the interference cancellation capability of the precoder via *Algorithm 2* and the number of basis vectors used to generate the representative matrix. Namely, a tradeoff between performance and complexity. Fortunately, since $N_{RF} \geq 2\,N_s$ is required to approach the optimal according to the results given in [14], our proposed *Algorithm 2* is implementable for the most reasonable scenarios.

Finally, we illustrate the impact of the number of candidate vectors used in producing precoders in Fig. 8, where we still consider $N_{BS} = 256$, $N_{MS} = 64$, $N_{RF} = M_{RF} = 8$ and $N_s = 3$. From simulation results, we can see that when the number of candidate vectors is greater than 12, the spectral efficiencies achieved by *Algorithm 2* approach to a constant. For $N_{RF} = M_{RF} = 8$, only four additional vector candidates are added to the selection set and obtain near-optimal performance, which causes negligible complexity increments compared to *Algorithm 1*.

V. CONCLUSION

In this paper, we propose a novel hybrid precoding scheme for a mmWave massive MIMO system with the cluster-based channel model. Unlike the existing precoding schemes, which generally require an iterative and successive search processes to achieve the analog phase shifters, our proposed scheme exploits the correlation values between the eigenvectors of the array response matrix and the digital optimal precoder in selecting the most relevant basis vectors which are used to reconstruct the representative matrix of the array

response matrix. Therefore, the iterative search process is no longer required and the proposed scheme can significantly reduce the complexity and be realized in parallel. Meanwhile, the proposed scheme achieves a comparable performance as OMP.

REFERENCES

[1] L. Mohjazi, M. Dianati, G. K. Karagiannidis, S. Muhaidat, and M. Al-Qutayri, "RF-powered cognitive radio networks: Technical challenges and limitations," *IEEE Commun. Mag.*, vol. 53, no. 4, pp. 94–100, Apr. 2015.

[2] S. Wang, R. Ruby, V. C. M. Leung, and Z. Yao, "A low-complexity power allocation strategy to minimize sum-source-power for multi-user single-AF-relay networks," *IEEE Trans. Commun.*, vol. 64, no. 8, pp. 3275–3283, Aug. 2016.

[3] J. Mietzner, R. Schober, L. Lampe, W. H. Gerstacker, and P. A. Hoeher, "Multiple-antenna techniques for wireless communications—A comprehensive literature survey," *IEEE Commun. Surveys Tuts.*, vol. 11, no. 2, pp. 87–105, 2nd Quart., 2009.

[4] Q. Deng *et al.*, "Dynamic spectrum sharing for hybrid access in OFDMA-based cognitive femtocell networks," *IEEE Trans. Veh. Tech.*, vol. 67, no. 11, pp. 10830–10840, Nov. 2018.

[5] S. Wang, R. Ruby, V. C. M. Leung, Z. Yao, X. Liu, and Z. Li, "Sum-power minimization problem in multisource single-AF-relay networks: A new revisit to study the optimality," *IEEE Trans. Veh. Technol.*, vol. 66, no. 11, pp. 9958–9971, Nov. 2017.

[6] P. B. Papazian, G. A. Hufford, R. J. Achatz, and R. Hoffman, "Study of the local multipoint distribution service radio channel," *IEEE Trans. Broadcast.*, vol. 43, no. 2, pp. 175–184, Jun. 1997.

[7] T. S. Rappaport, G. R. MacCartney, M. K. Samimi, and S. Sun, "Wideband millimeter-wave propagation measurements and channel models for future wireless communication system design," *IEEE Trans. Commun.*, vol. 63, no. 9, pp. 3029–3056, Sep. 2015.

[8] R. C. Daniels and R. W. Heath, Jr., "60 GHz wireless communications: Emerging requirements and design recommendations," *IEEE Veh. Technol. Mag.*, vol. 2, no. 3, pp. 41–50, Sep. 2007.

[9] J. G. Andrews, T. Bai, M. N. Kulkarni, A. Alkhateeb, A. K. Gupta, and R. W. Heath, Jr., "Modeling and analyzing millimeter wave cellular systems," *IEEE Trans. Commun.*, vol. 65, no. 1, pp. 403–430, Jan. 2017.

[10] A. I. Sulyman, A. T. Nassar, M. K. Samimi, G. R. MacCartney, Jr., T. S. Rappaport, and A. Alsanie, "Radio propagation path loss models for 5G cellular networks in the 28 GHz and 38 GHz millimeter-wave bands," *IEEE Commun. Mag.*, vol. 52, no. 9, pp. 78–86, Sep. 2014.

[11] M. Elkashlan, T. Q. Duong, and H.-H. Chen, "Millimeter-wave communications for 5G: Fundamentals: Part I [guest editorial]," *IEEE Commun. Mag.*, vol. 52, no. 9, pp. 52–54, Sep. 2014.

[12] J. G. Andrews *et al.*, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

[13] J. Zhang, L. Dai, X. Li, Y. Liu, and L. Hanzo, "On low-resolution ADCs in practical 5G millimeter-wave massive MIMO systems," *IEEE Commun. Mag.*, vol. 56, no. 7, pp. 205–211, Jul. 2018.

[14] T. E. Bogale, L. B. Le, A. Haghighat, and L. Vandendorpe, "On the number of RF chains and phase shifters, and scheduling design with hybrid analog-digital beamforming," *IEEE Trans. Wireless Commun.*, vol. 15, no. 5, pp. 3311–3326, May 2016.

[15] J. Li, D. Wang, P. Zhu, and X. You, "Benefits of beamforming training scheme in distributed large-scale MIMO systems," *IEEE Access*, vol. 6, pp. 7432–7444, 2018.

[16] J. Li, D. Wang, P. Zhu, J. Wang, and X. You, "Downlink spectral efficiency of distributed massive MIMO systems with linear beamforming under pilot contamination," *IEEE Trans. Veh. Technol.*, vol. 67, no. 2, pp. 1130–1145, Feb. 2018.

[17] J. Zhang, L. Dai, Z. He, B. Ai, and O. A. Dobre, "Mixed-ADC/DAC multipair massive MIMO relaying systems: Performance analysis and power optimization," *IEEE Trans. Commun.*, vol. 67, no. 1, pp. 140–153, Jan. 2019.

[18] J. Zhang, X. Xue, E. Björnson, B. Ai, and S. Jin, "Spectral efficiency of multipair massive MIMO two-way relaying with hardware impairments," *IEEE Wireless Commun. Lett.*, vol. 7, no. 1, pp. 14–17, Feb. 2018.

[19] Y. Huang, J. Zhang, and M. Xiao, "Constant envelope hybrid precoding for directional millimeter-wave communications," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 4, pp. 845–859, Apr. 2018.

[20] S. He, J. Wang, Y. Huang, B. Ottersten, and W. Hong, "Codebook-based hybrid precoding for millimeter wave multiuser systems," *IEEE Trans. Signal Process.*, vol. 65, no. 20, pp. 5289–5304, Oct. 2017.

[21] W. Ni, X. Dong, and W. S. Lu, "Near-optimal hybrid processing for massive MIMO systems via matrix decomposition," *IEEE Trans. Signal Process.*, vol. 65, no. 15, pp. 3922–3933, Aug. 2017.

[22] O. E. Ayach, R. W. Heath, Jr., S. Rajagopal, and Z. Pi, "Multimode precoding in millimeter wave MIMO transmitters with multiple antenna sub-arrays," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Atlanta, GA, USA, Dec. 2013, pp. 3476–3480.

[23] D. Zhang, Y. Wang, X. Li, and W. Xiang, "Hybridly connected structure for hybrid beamforming in mmWave massive MIMO systems," *IEEE Trans. Commun.*, vol. 66, no. 2, pp. 662–674, Feb. 2018.

[24] H. Lin, F. Gao, S. Jin, and G. Y. Li, "A new view of multi-user hybrid massive MIMO: Non-orthogonal angle division multiple access," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2268–2280, Oct. 2017.

[25] W. M. Chan, T. Kim, H. Ghauch, and M. Bengtsson, "Subspace estimation and hybrid precoding for wideband millimeter-wave MIMO systems," in *Proc. 50th Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, USA, Nov. 2016, pp. 286–290.

[26] C. Zhang, Y. Huang, Y. Jing, S. Jin, and L. Yang, "Sum-rate analysis for massive MIMO downlink with joint statistical beamforming and user scheduling," *IEEE Trans. Wireless Commun.*, vol. 16, no. 4, pp. 2181–2194, Apr. 2017.

[27] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, Jr., "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.

[28] A. Alkhateeb, R. W. Heath, Jr., and G. Leus, "Achievable rates of multi-user millimeter wave systems with hybrid precoding," in *Proc. IEEE Int. Conf. Commun.*, London, U.K., Jun. 2015, pp. 1232–1237.

[29] J. Zhao, F. Gao, W. Jia, S. Zhang, S. Jin, and H. Lin, "Angle domain hybrid precoding and channel tracking for millimeter wave massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 10, pp. 6868–6880, Oct. 2017.

[30] L. Dai, X. Gao, J. Quan, S. Han, and C.-L. I, "Near-optimal hybrid analog and digital precoding for downlink mmWave massive MIMO systems," in *Proc. IEEE Int. Conf. Commun.*, London, U.K., Jun. 2015, pp. 1334–1339.

[31] C. Rusu, R. Méndez-Rial, N. González-Prelcic, and R. W. Heath, Jr., "Low complexity hybrid sparse precoding and combining in millimeter wave MIMO systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, London, U.K., Jun. 2015, pp. 1340–1345.

[32] O. E. Ayach, R. W. Heath, Jr., S. Abu-Surra, S. Rajagopal, and Z. Pi, "Low complexity precoding for large millimeter wave MIMO systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Canada, U.K., Jun. 2012, pp. 3724–3729.

[33] Y.-Y. Lee, C.-H. Wang, and Y.-H. Huang, "A hybrid RF/baseband precoding processor based on parallel-index-selection matrix-inversion-bypass simultaneous orthogonal matching pursuit for millimeter wave MIMO systems," *IEEE Trans. Signal Process.*, vol. 63, no. 2, pp. 305–317, Jan. 2015.

[34] W. Ni and X. Dong, "Hybrid block diagonalization for massive multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 64, no. 1, pp. 201–211, Jan. 2016.

[35] X. Gao, L. Dai, C. Yuen, and Z. Wang, "Turbo-like beamforming based on tabu search algorithm for millimeter-wave massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 65, no. 7, pp. 5731–5737, Jul. 2016.

[36] V. Raghavan, S. Subramanian, J. Cezanne, A. Sampath, O. H. Koymen, and J. Li, "Single-user versus multi-user precoding for millimeter wave MIMO systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 6, pp. 1387–1401, Jun. 2017.

[37] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*. Upper Saddle River, NJ, USA: Prentice-Hall, 2000.

[38] H. Xu, V. Kukshya, and T. S. Rappaport, "Spatial and temporal characteristics of 60-GHz indoor channels," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 3, pp. 620–630, Apr. 2002.

[39] A. M. Sayeed, "Deconstructing multiantenna fading channels," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2563–2579, Oct. 2002.

[40] V. Raghavan and A. M. Sayeed, "Sublinear capacity scaling laws for sparse MIMO channels," *IEEE Trans. Inf. Theory*, vol. 57, no. 1, pp. 345–364, Jan. 2011.

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

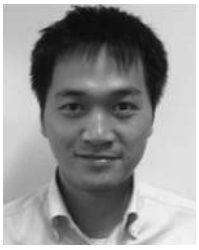
XIANRU LIU received the master's and Ph.D. degrees in control science and engineering from Central South University, Changsha, China, in 2006 and 2011, respectively, where she is currently an Instructor. Her research interests include computer vision and image understanding, machine learning, wireless cooperative communication, and cognitive networks.

XUEMING LI received the B.E. degree in electronics engineering from the University of Science and Technology of China, in 1992, and the Ph.D. degree in electronics engineering from the Beijing University of Posts and Telecommunications (BUPT), in 1997. From 1997 to 1999, he was a Postdoctoral Researcher with the Institute of Information Science, Beijing Jiaotong University. He has been with BUPT, since 1999. In 2002, he was a Guest Lecturer with Karlsruhe University, Germany. His current research interests include digital image processing, video coding, and multimedia telecommunication. He has undertaken many state and enterprise Research and Development projects. He has authored three books and over 50 papers in the field of multimedia information processing and transmission. He is currently a Senior Member of the Chinese Institute of Electrics, and a Senior Member of the China Society of Image and Graphics.

SHU CAO received the bachelor's degree in electronic science and technology from the Hunan Institute of Engineering, China, in 2016. Her research interests include millimeter-wave communication, massive MIMO, and machine learning.

QINGYONG DENG received the B.Eng. degree in electrical information engineering, and the M.Eng. degree in signal and information processing from Xiangtan University, in 2004 and 2009, respectively. He is currently pursuing the Ph.D. degree with the Beijing University of Posts and Telecommunications. Since 2004, he has been a Lecturer with the College of Information Engineering, Xiangtan University. His current researches focus on cognitive networks, machine-to-machine communication, smart grid, and wireless communication.

RONG RAN (M'05) received the Ph.D. degree from the Department of Electrical and Electronic Engineering, Yonsei University, in 2009. In 2009, she joined ETRI, Daejeong, and has worked on the IEEE 802. 16m standardization. She was a Research Associate with The Hong Kong University of Science and Technologies, in 2010. Since 2014, she has been a Faculty of the Electrical and Computer Engineering Department, Ajou University. Her current research interests include 5G wireless communications, machine/deep learning, and sparse signal processing. She is a member of the KICS.



KIEN NGUYEN (S'08–M'12–SM'16) received the B.E. degree in electronics and telecommunication from the Hanoi University of Science and Technology, Vietnam, and the Ph.D. degree in informatics from the Graduate University for Advanced Studies, Hayama, Japan, in 2004 and 2012, respectively. He joined the National Institute of Information Communications Technology, Yokosuka, Japan, in 2014, as a Researcher. His research interests include novel, software-based, evolvable networking technologies for the next generation of mobile networks and the Internet of Things. He is a member of the IEICE.

PEI TINGRUI has majored in signal and information processing and received the Ph.D. degree from the Beijing University of Posts and Telecommunications, in 2004. He visited Japan as a Researcher, from 2006 to 2007. He is currently a Professor and a Doctoral Supervisor. He focuses on the research of wireless sensor networks, ad hoc, mobile communication networks, and social computing.