WILEY | Hindawi

Research Article

Hybrid Precoding Algorithm for Millimeter-Wave Massive MIMO Systems with Subconnection Structures

Xue Zhang[1] **and Feng Zhao** [2]

[1]*Key Laboratory of Cognitive Ratio and Information Processing, Guilin University of Electronic Technology, Guilin 541004, China*
[2]*Guangxi Colleges and Universities Key Laboratory of Complex System Optimization and Big Data Processing,*
 Yulin Normal University, Yulin 537000, China

Correspondence should be addressed to Feng Zhao; zhaofeng@guet.edu.cn

In mmWave massive MIMO systems, traditional digital precoding is difficult to be implemented because of the high cost and energy consumption of RF chains. Fortunately, the hybrid precoding which combines digital precoding and analog precoding not only solves this problem successfully, but also improves the performance of the system effectively. However, due to the constant mode constraint introduced by the phase shifter in the analog domain, it is difficult to solve the hybrid precoding directly. There is a solution which divides the total optimization problem into two stages to solve, that is, first fix the digital precoding matrix, solve the analog precoding matrix, and then optimize the digital precoding matrix according to the obtained analog precoding matrix. In this paper, a high energy-efficient hybrid precoding scheme is proposed for the subconnection structure. In the first stage, the optimization problem can be decomposed into a series of subproblems by means of the independent submatrix structure of the analog precoding matrix. When the optimized analog precoding matrix is obtained, the digital precoding matrix can be solved by the minimum mean error (MMSE). Finally, the digital precoding matrix is normalized to satisfy the constraint conditions. The simulation results demonstrate that the performance of the proposed algorithm is close to that of fully digital precoding based on subconnection structure and better than that of the existing algorithms. In addition, this paper presents the simulation analysis of the algorithm performance under imperfect channel state information. Simulation results show that when the estimation accuracy of channel state information is 0.8, the spectral efficiency of the proposed algorithm can already be maintained at a good level.

1. Introduction

With the rapid development of technology, the fifth generation mobile communication (5G) has attracted wide attention due to higher frequency, greater network capacity, and lower latency. It integrates many technologies, among which millimeter-wave (mmWave) communication and massive multiple-input multiple-output (MIMO) technology play important roles. Due to the rich spectrum resources available in the high frequency band, millimeter-wave communication can realize extremely high-speed and short distance communication with high gain, but there are also short transmission distance, poor penetration, and diffraction capability, vulnerability to climatic and environmental impacts [1], while massive MIMO systems, which use large arrays of antennas to communicate, can rapidly improve wireless data rates and system energy efficiency [2]. The shorter wavelength of mmWave signals enables large antenna arrays to be integrated into a smaller space, which implies that massive MIMO is feasible in wireless communications [3]. In addition, the combination of them can significantly improve user throughput, spectrum, and energy efficiency and increase the potential for mobile network capacity [4].

In millimeter-wave communication system, the cost of traditional digital precoding is too high to realize. Therefore, hybrid precoding, which combines digital precoding with analog precoding, is a better choice. It is usually based on two traditional structures: fully connected and subconnected.

Since the fully connected structure can approach the theoretical optimal spectrum efficiency, it has been extensively studied in the academic community. The hybrid precoding problem was reconstructed by the spatial sparsity of the mmWave channel in [5], and the orthogonal matching pursuit (OMP) algorithm is proposed to solve the problem. Based on the OMP algorithm, a greedy algorithm is proposed in [6]. The algorithm does not need to consider the geometry of the antenna array, which effectively reduces the computational complexity while achieving good performance. A threshold orthogonal matching pursuit (TOMP) algorithm with performance close to the optimal and higher than the OMP algorithm is proposed by setting an appropriate threshold [7]. Although the performance of the OMP algorithm is good, the complexity of the algorithm is relatively high. Therefore, a real-time and high performance precoding strategy based on singular value decomposition is studied in [8]. The performance of this scheme is similar to that of the OMP algorithm, and the complexity is much lower than that of the OMP algorithm. In [9], an alternative minimization (AltMin) algorithm is proposed by using the idea of manifold optimization, and the complexity of the algorithm is further reduced.

The characteristics of subconnected structure make it achieve a better balance between performance and cost, which has attracted the attention of many scholars. Unlike the one-to-one correspondence between the RF chain and the antenna in the fully connected structure, each RF chain in the subconnected structure is only connected with part of the antenna. Therefore, research on hybrid precoding algorithms with subconnection structures has also become the focus of attention in recent years. For the subconnected structure, an AltMin algorithm is also developed with the help of semidefinite relaxation in [9]. Similarly, a new divide-and-conquer precoding scheme is proposed in [10]. The performance of this scheme is close to that of SDR-AltMin, which can save time effectively and is robust to potential saddle points with poor performance. In view of the hybrid precoding problem of energy-saving subconnection structure, a hybrid precoding scheme based on successive interference cancellation (SIC) is proposed in [11]. It does not require singular value decomposition and matrix inversion, and the computational complexity is much lower than traditional sparse reconstruction precoding algorithms. In [12], an optimization model depends on the minimum mean square error (MMSE) is established, and then, a low complexity hybrid precoding algorithm based on particle swarm ant colony optimization (PSACO) is proposed. The high performance hybrid precoding algorithm for subconnection structures is studied in [13], and the performance of the algorithm is better than that of the SIC algorithm when the channel state information is not perfect. A new subconnection structure is introduced in [14] to further reduce the power consumption of the system. According to this structure, an efficient hybrid precoding scheme based on service quality constraint is proposed. A general hybrid precoding algorithm is studied in [15], in which the supplementary matrix is introduced, and the regularization zero-forcing method is used to solve the problem.

Synthesizing the current research situation, most of the literature studies the ideal fully connection structure, but in practice, the structure has high complexity and energy consumption. Although the performance is considerable, the practical application is poor. Therefore, some scholars have gradually focused on the energy-saving and more practical subconnection structure, thus far, also achieved some results, but the performance of the hybrid precoding algorithm and the complexity of the algorithm still need to be further improved. Different from the traditional full-digital precoding, hybrid precoding can only control the phase of the data stream in the analog domain, but cannot adjust its amplitude. The constant modulus constraint of the analog precoding matrix makes the solution of the hybrid precoding more difficult.

Thus, in order to simplify the problem, we consider solving the analog precoding matrix and the digital precoding matrix, respectively, as in most references [11, 13]. The total achievable rate R of the system is expanded directly, and the optimization problem is decomposed into a series of subrate optimization by recursion in [11, 15]. Different from them, this paper starts with the equivalent problem of maximizing the achievable rate, the Euclidean distance derived in [5] to solve the optimization problem. We utilize the block diagonal form of the analog precoding and decompose the optimization problem into a series of optimization subproblems and solved in turn. The subproblem, which is each column of the analog precoding matrix, is derived from the corresponding Euclidean distance expansion. After the complete analog precoding matrix is obtained, the corresponding digital precoding matrix optimization problem is easy to solve. The simulation results also show that the algorithm can achieve considerable performance regardless of whether the acquired channel state information is perfect or not.

The rest of this paper is arranged as follows. Section 2 introduces the model of mmWave massive MIMO system and the channel model used. Section 3 describes the design process of the proposed algorithm in detail. In Section 4, the simulation results are given, and the performance of the proposed algorithm is analyzed. Finally, we conclude this paper in Section 5.

Notation: in this paper, $\mathbf{A}$, $\mathbf{a}$, and a denote a matrix, a vector, and a scalar, respectively; $(\mathbf{A})^T$, $(\mathbf{A})^H$, and $\|\mathbf{A}\|_F$ are the transpose, conjugate transpose, and Frobenius norm of $\mathbf{A}$, respectively; $\mathbf{A}^{-1}$ and $\mathbf{A}^\dagger$ represent the inverse and Moore-Penrose pseudo inverse of $\mathbf{A}$; $\mathrm{Tr}(\mathbf{A})$ indicates the trace of matrix $\mathbf{A}$; $\det(\mathbf{A})$ expresses the determinant of $\mathbf{A}$; $\mathrm{diag}(\mathbf{A})$ is the diagonal version of $\mathbf{A}$; $\mathrm{abs}(\mathbf{A})$ is the element-wise absolute value of $\mathbf{A}$; $\mathbf{A} \oslash \mathbf{B}$ is the Hadamard division between $\mathbf{A}$ and $\mathbf{B}$; $\mathscr{CN}(0, \sigma^2)$ is a complex Gaussian vector with mean 0 and covariance σ^2; $E(\cdot)$ denotes the expectation; $\mathbb{C}^{m \times n}$ denotes an $m \times n$ dimensional complex space; $\mathfrak{R}(a)$ is the real part of a.

2. System Model and Channel Model

2.1. System Model. We focus on subconnection architecture, which can avoid the use of a large number of phase shifters,
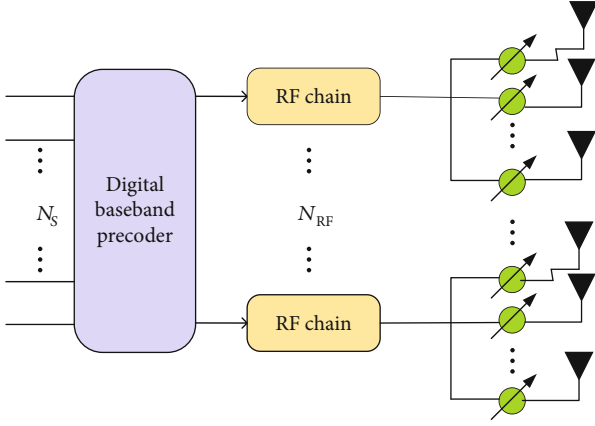
FIGURE 1: Subconnected architecture.

as shown in Figure 1. Consider a single-user mmWave massive MIMO system as shown in Figure 2. The transmitter and the receiver are equipped with N_t and N_r antennas, respectively [16]. N_s data streams in the baseband are first precoded by a digital precoder $\mathbf{D}$ and then precoded by an analog precoder after passing through the corresponding RF chain. After that, each data stream is transmitted by a subantenna array with only M antennas associated with the corresponding RF chain. The number of RF chains at the transmitter and the receiver is N_{RF}^t and N_{RF}^r, respectively. To achieve multiple data transmission, the following constraints must be satisfied: $N_s \leq N_{\mathrm{RF}}^t \leq N_t$ and $N_s \leq N_{\mathrm{RF}}^r \leq N_t$.

The hybrid precoder is composed of a digital baseband precoder $\mathbf{D}$ and an analog precoder $\mathbf{A}$, and their dimensions are $N_{\mathrm{RF}}^t \times N_s$ and $N_t \times N_{\mathrm{RF}}^t$. Assuming that the initial signal is $\mathbf{s}$, the transmission signal can be written as $\mathbf{x} = \mathbf{ADs}$, where $\mathbf{s}$ is the $N_s \times 1$ symbol vector and satisfies $E[\mathbf{ss}^T] = 1/N_s \mathbf{I}_{N_s}$. $\|\mathbf{AD}\|_F^2 = N_s$ is the normalized power constraint of the system [17]. The received signal vector $\mathbf{y} = [y_1, y_2, \cdots, y_K]^T$ of the system can be expressed as

$$\mathbf{y} = \sqrt{\rho}\mathbf{Hx} + \mathbf{n} = \sqrt{\rho}\mathbf{HADs} + \mathbf{n}, \qquad (1)$$

where ρ is the average received power; $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$ denotes the channel matrix, the baseband transmission signal vector can be expressed as $\mathbf{s} = [s_1, s_2, \cdots, s_N]^T$, and $\mathbf{A}$ and $\mathbf{D}$ represent the analog precoding matrix and the digital precoding matrix, respectively. $\mathbf{n}$ denotes the noise vector with independently and identically distributed i.i.d. $\mathscr{CN}(0, \sigma^2)$ entries. $\mathbf{F} = \mathbf{AD}$ presents the hybrid precoding matrix of size $N_t \times N_r$, which satisfies $\|\mathbf{F}\|_F \leq N_s$ to meet the total transmit power constraint.

In subconnection architecture, each RF chain is connected to M ($M = N_t/N_{\mathrm{RF}}$) antenna via M phase shifter [18]. Therefore, the digital precoding matrix $\mathbf{D}$ in this architecture is the diagonal matrix, $\mathbf{D} = \mathrm{diag}\,[\bar{d}_1, \bar{d}_2, \cdots, \bar{d}_N]$, where $\bar{d}_n \in \mathbb{R}, n = 1, 2, \cdots, N$. The corresponding analog precoding matrix is a block diagonal matrix, which is represented as

$$\mathbf{A} = \begin{bmatrix} \bar{\mathbf{a}}_1 & 0 & \cdots & 0 \\ 0 & \bar{\mathbf{a}}_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \bar{\mathbf{a}}_N \end{bmatrix}_{NM \times N}. \qquad (2)$$

$\bar{\mathbf{a}}_n \in \mathbb{C}^{M \times 1}$ can be regarded as an analog weighting vector. Due to the constant modulus constraint, each element in the $\bar{\mathbf{a}}_n$ has the same amplitude and different phases [18].

2.2. Channel Model. Unlike traditional low frequency channels, the propagation characteristics of millimeter-wave channels no longer subject to Rayleigh fading. Due to the high path loss in free space, the spatial selectivity of millimeter-wave propagation is limited [19]. In addition, large-scale dense antenna arrays will result in obvious antenna correlation in millimeter-wave channels [20]. Therefore, the traditional channel model is not suitable for the millimeter-wave channel in the massive MIMO system. For simplicity, this paper adopts the geometric Saleh-Valenzuela model in most literature [20, 21]. The channel matrix is given by

$$\mathbf{H} = \sqrt{\frac{N_t N_r}{N_{\mathrm{cl}} N_{\mathrm{ray}}}} \sum_{i=1}^{N_{\mathrm{cl}}} \sum_{k=1}^{N_{\mathrm{ray}}} \alpha_{ik} \mathbf{a}_r(\phi_{ik}^r, \theta_{ik}^r) \mathbf{a}_t(\phi_{ik}^t, \theta_{ik}^t)^H, \qquad (3)$$

where N_{cl} is the number of scattering clusters and each cluster contributes N_{ray} propagation paths. α_{ik} is the gain of the kth ray in the ith scattering cluster, and it is subject to i.i.d. $\mathscr{CN}(0, \sigma_{\alpha,i}^2)$, where $\sigma_{\alpha,i}^2$ indicates the power of the ith clusters [22]. $\mathbf{a}_r(\phi_{ik}^r, \theta_{ik}^r)$ and $\mathbf{a}_t(\phi_{ik}^t, \theta_{ik}^t)$ represent the normalized antenna array response vectors at the receiver and transmitter, respectively. $\phi_{ik}^r(\theta_{ik}^r)$ and $\phi_{ik}^t(\theta_{ik}^t)$ are the azimuth (elevation) angles of the kth ray in the $i$thscattering cluster on the receiving and transmitting sides, respectively. The mean cluster angles $\phi_{ik}^r(\theta_{ik}^r)$ and $\phi_{ik}^t(\theta_{ik}^t)$ are subject to uniform-random distribution in the range of $[0, 2\pi)$ [23]. Within the ith cluster, the arrival and departure angles of each ray follow Laplace distribution with mean cluster angles $\phi_{ik}^r(\theta_{ik}^r)$ and $\phi_{ik}^t(\theta_{ik}^t)$ as the standard deviation.

The antenna array response vector depends on the structure of the antenna array at the transmitter and receiver. There are two types of antenna array structures: uniform linear array (ULA) and uniform planar array (UPA). For millimeter-wave massive MIMO systems, a uniform planar array is conducive to the packaging of antennas at the transceiver, and the realization of beam-forming will be more flexible. On the other hand, as the UPA array can bring better energy efficiency and spectrum efficiency, we mainly consider the UPA. In the case of a uniform planar array with horizontal W antenna elements and vertical H antenna elements, the array response vector is given by [24]
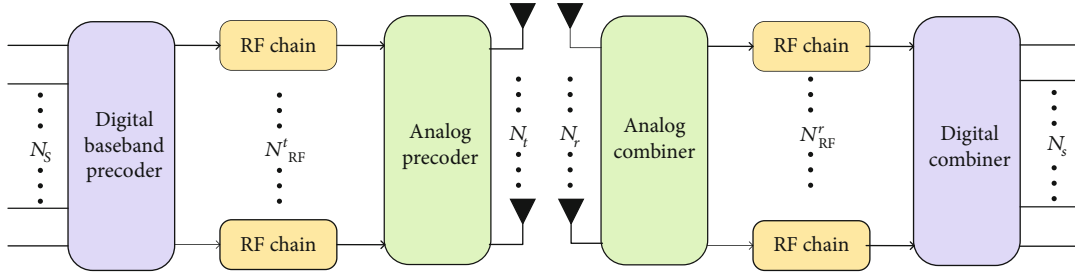
FIGURE 2: Single-user mmWave massive MIMO system.

$$\mathbf{a}_{UPA}(\phi,\theta) = \frac{1}{\sqrt{W \times H}}$$
$$\cdot \left[1,\cdots,e^{j(2\pi/\lambda)d(m\,\sin\,\phi\,\sin\,\theta+n\,\cos\,\theta)},\cdots,e^{j(2\pi/\lambda)d((W-1)\,\sin\,\phi\,\sin\,\theta+(H-1)\,\cos\,\theta)}\right]^T, \tag{4}$$

where λ is the wavelength of the carrier signal, d is the distance between adjacent antennas, and m and n satisfy $0 \le m < W$ and $0 \le n < H$, respectively.

3. Hybrid Precoding Algorithm Design

In this paper, an approximate optimal hybrid precoding algorithm is proposed for the subconnection structure. Because the joint optimization problem can be temporarily decoupled [5], this paper mainly discusses the design of the hybrid precoding algorithm at the transmitter of the system.

The fundamental goal of the hybrid precoding algorithm is to maximize the spectral efficiency R by designing the optimal hybrid precoder. The spectral efficiency of the system is expressed as [5] follows:

$$R = \log_2\left(\left|\mathbf{I}_{N_s} + \frac{\rho}{N_s\sigma^2}\mathbf{H}\mathbf{F}\mathbf{F}^H\mathbf{H}^H\right|\right), \tag{5}$$

where $\mathbf{I}_{N_s}$ denotes a $N_s \times N_s$ unit matrix and $\mathbf{F}$ presents the hybrid precoding matrix with block diagonal form. Thus, the hybrid precoder optimization problem in this paper can be expressed as

$$\left(\mathbf{A}^{\mathrm{opt}}, \mathbf{D}^{opt}\right) = \underset{\mathbf{A},\mathbf{D}}{\arg\max}\, R,$$
$$s.t.\,\mathbf{A} \in \mathscr{A}, \tag{6}$$
$$\|\mathbf{A}\mathbf{D}\|_F^2 = N_s.$$

Hybrid precoding introduces phase shifters in the analog domain to process the transmitted signal. Since the phase shifter can only adjust the phase of the signal and cannot adjust the amplitude of the signal, the nonzero terms in the analog precoding matrix need to have the same modulus [18]. $\mathscr{A}$ is used to represent the set of $\mathbf{A}$ matrices that satisfy the constant modulus constraint. Obviously, problem (6) belongs to nonconvex optimization problem due to the existence of constant modulus constraint. Although the constraints make it difficult to obtain the exact solution directly, the approximate optimal mixed precoding matrix

can be obtained by solving the approximation of (6). As described in [5], problem (6) can be further equated to problem (7):

$$\left(\mathbf{A}^{\mathrm{opt}}, \mathbf{D}^{\mathrm{opt}}\right) = \underset{\mathbf{A},\mathbf{D}}{\arg\min}\, \left\|\mathbf{F}^{opt} - \mathbf{A}\mathbf{D}\right\|_F^2,$$
$$s.t.\,\mathbf{A} \in \mathscr{A}, \tag{7}$$
$$\|\mathbf{A}\mathbf{D}\|_F^2 = N_s.$$

The Euclidean distance is used to make the optimal mixed precoding as close as possible to the unconstrained optimal digital precoding. It is shown in reference [5] that the objective function in the minimization problem (7) can maximize the spectral efficiency. The optimal mixed precoding matrix in the problem can be obtained by singular value decomposition of the channel matrix. The channel matrix singular value decomposition is represented by $\mathbf{H} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^H$, and the unconstrained optimal precoding matrix $\mathbf{F}^{\mathrm{opt}}$ consists of the first N_s columns of the right singular matrix.

Problem (7) is fundamentally aimed at obtaining approximately optimal analog precoding matrix $\mathbf{A}$ and digital precoding matrix $\mathbf{D}$. The constant mode constraint in (7) makes it difficult to solve the problem directly. Thanks to the special block diagonal form of the hybrid precoding matrix; it can be seen that the different subarrays are not correlated. This allows us to decompose the overall optimization problem into a series of suboptimization problems, each of which can be represented by $\bar{\mathbf{f}}_n = \bar{d}_n\bar{\mathbf{a}}_n$.

$$\mathbf{F} = \mathbf{A}\mathbf{D} = \mathrm{diag}\left[\bar{\mathbf{a}}_1, \bar{\mathbf{a}}_2,\cdots,\bar{\mathbf{a}}_N\right] \times \mathrm{diag}\left[\bar{d}_1, \bar{d}_2,\cdots\bar{d}_N\right]$$
$$= \mathrm{diag}\left[\bar{\mathbf{f}}_1, \bar{\mathbf{f}}_2,\cdots,\bar{\mathbf{f}}_N\right] = \begin{bmatrix} \bar{\mathbf{f}}_1 & 0 & \cdots & 0 \\ 0 & \bar{\mathbf{f}}_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \bar{\mathbf{f}}_N \end{bmatrix}, \tag{8}$$

where $\mathbf{f}_n$ is the n column of the hybrid precoding matrix $\mathbf{F}$ and $\bar{\mathbf{f}}_n$ is the nonzero element in the n column, which is the nonzero element matrix of column matrix of $\mathbf{F}$. The concrete structure of $\mathbf{f}_n$ is shown as

$$\mathbf{f}_n = \left[\mathbf{0}_{1\times M(n-1)}, \bar{\mathbf{f}}_n, \mathbf{0}_{1\times M(N-n)}\right]^T. \tag{9}$$

It is obvious from formula (8) that the design of hybrid precoder $\mathbf{F}$ can be converted into optimizing analog precoder $\mathbf{A}$ and digital precoder $\mathbf{D}$ separately, or the problem is equivalent to optimizing each independent subantenna array $\bar{\mathbf{f}}_n$. Therefore, we assume that the digital precoding matrix $\mathbf{D}$ is fixed and first optimize the analog precoding matrix $\mathbf{A}$. The optimization problem is expressed as

$$\mathbf{A}^{\text{opt}} = \arg \min_{\mathbf{A}} \left\| \mathbf{F}^{\text{opt}} - \mathbf{AD} \right\|_F^2,$$

$$s.t. \mathbf{A} \in \mathscr{A}, \tag{10}$$

$$\left\| \mathbf{AD} \right\|_F^2 = N_s.$$

Since the hybrid precoding matrix of subconnected structure is a special block-diagonal form, the total optimization problem is decomposed into a series of independent subarray optimization problems. The power constraint in (10) can be scaled to meet the constraint after the analog precoder and the digital precoder are optimized. With the help of (8), the subarray optimization problem is described as

$$\bar{\mathbf{a}}_n^{\text{opt}} = \arg \min_{\bar{\mathbf{a}}_n} \left\| \mathbf{F}_n^{\text{opt}} - \bar{\mathbf{a}}_n \bar{d}_n \right\|_F^2,$$

$$s.t. \left| (\bar{\mathbf{a}}_n)_m \right| = 1. \tag{11}$$

As can be seen from the objective function (11), the optimization problem of the analog precoding matrix $\mathbf{A}^{\text{opt}}$ is further decomposed into optimization of nonzero submatrix $\bar{\mathbf{a}}_n$. The objective function of the (11) suboptimization problem is expanded as

$$\begin{aligned}
\left\| \mathbf{F}_n^{\text{opt}} - \bar{\mathbf{a}}_n \bar{d}_n \right\|_F^2 &= \text{Tr}\left(\left(\mathbf{F}_n^{\text{opt}} - \bar{\mathbf{a}}_n \bar{d}_n \right)^H \left(\mathbf{F}_n^{\text{opt}} - \bar{\mathbf{a}}_n \bar{d}_n \right) \right) \\
&= \text{Tr}\left(\left(\mathbf{F}_n^{\text{opt}} \right)^H \left(\mathbf{F}_n^{\text{opt}} \right) \right) - 2\Re \left(\text{Tr}\left(\bar{d}_n^H \bar{\mathbf{a}}_n^H \mathbf{F}_n^{\text{opt}} \right) \right) \\
&\quad + \text{Tr}\left(\bar{d}_n^H \bar{\mathbf{a}}_n^H \bar{\mathbf{a}}_n \bar{d}_n \right) = \text{Tr}\left(\left(\mathbf{F}_n^{\text{opt}} \right)^H \left(\mathbf{F}_n^{\text{opt}} \right) \right) \\
&\quad - 2\Re \left(\text{Tr}\left(\bar{\mathbf{a}}_n^H \mathbf{F}_n^{\text{opt}} \bar{d}_n^H \right) \right) + \text{Tr}\left(\bar{d}_n^H \bar{\mathbf{a}}_n^H \bar{\mathbf{a}}_n \bar{d}_n \right) \\
&= \text{Tr}\left(\left(\mathbf{F}_n^{\text{opt}} \right)^H \left(\mathbf{F}_n^{\text{opt}} \right) \right) - 2\Re \left(\bar{d}_n \text{Tr}\left(\bar{\mathbf{a}}_n^H \mathbf{F}_n^{\text{opt}} \right) \right) + M \bar{d}_n^2.
\end{aligned} \tag{12}$$

$\text{Tr}(AB) = \text{Tr}(BA)$ and $\bar{\mathbf{a}}_n^H \bar{\mathbf{a}}_n = M$ are used in the expansion of the equation (12). Because all the terms in $\bar{\mathbf{a}}_n$ have the same amplitude 1, $\bar{\mathbf{a}}_n^H \bar{\mathbf{a}}_n = M$ is established apparently.

From (12), we can see that the expansion result is made up of three parts, in which the first term $\text{Tr}((\mathbf{F}_n^{\text{opt}})^H (\mathbf{F}_n^{\text{opt}}))$ and the last term $M \bar{d}_n^2$ are fixed, and the goal of the problem is converted to maximize the real part of $\bar{d}_n \text{Tr}(\bar{\mathbf{a}}_n^H \mathbf{F}_n^{\text{opt}})$, that is, to maximize the real part of $\bar{\mathbf{a}}_n^H \mathbf{F}_n^{\text{opt}}$. Under constant modulus constraint, when the phase of each term of the $\bar{\mathbf{a}}_n$ is equal to the phase of $\mathbf{F}_n^{\text{opt}}$ corresponding term, the $\bar{\mathbf{a}}_n^H \mathbf{F}_n^{\text{opt}}$ contains only the real part and reaches the maximum. Therefore, the optimal solution of $\bar{\mathbf{a}}_n$ can be obtained by preserving the phase of the term in $\mathbf{F}_n^{\text{opt}}$ and normalizing it.

$$\bar{\mathbf{a}}_n = \mathbf{F}_n^{\text{opt}} \oslash \text{abs}\left(\mathbf{F}_n^{\text{opt}} \right), \tag{13}$$

If $n < N_{\text{RF}}^t$, then set $n = n + 1$ and continue to solve the next subproblem in a similar way. By solving $\bar{\mathbf{a}}_n$ successively, the complete optimized analog precoding matrix $\mathbf{A} = \text{diag}[\bar{\mathbf{a}}_1, \bar{\mathbf{a}}_2, \cdots, \bar{\mathbf{a}}_N]$ can be obtained. When the analog precoding matrix is determined, the corresponding optimal digital precoding matrix $\mathbf{D}$ can be easily obtained by the minimum mean square error (MMSE).

$$\mathbf{D} = \mathbf{A}^\dagger \mathbf{F}^{\text{opt}}. \tag{14}$$

Finally, the digital precoding matrix is normalized to meet the previously ignored power constraints.

$$\mathbf{D} = \frac{\sqrt{N_s}}{\left\| \mathbf{AD} \right\|_F} \mathbf{D}. \tag{15}$$

The complete process of hybrid precoding design is shown in Algorithm 1.

We have $\mathbf{W}_{\text{RF}}^H \mathbf{W}_{\text{RF}} \propto \mathbf{I}$ on the receiving side, while the digital combining matrix satisfies $\mathbf{W}_{BB}^H \mathbf{W}_{BB} \propto \mathbf{I}$. If $\mathbf{H}_{\text{eff}} = \mathbf{HAD}$ is considered as an equivalent matrix, the spectral efficiency can be given by [13]

$$\begin{aligned}
R &= \log_2 \det \left(I_{N_s} + \frac{\rho}{N_s \sigma_n^2} (W_{\text{RF}} W_{BB})^H H_{\text{eff}} H_{\text{eff}}^H W_{\text{RF}} W_{BB} \right) \\
&= \log_2 \det \left(I_{N_s} + \frac{\rho}{N_s \sigma_n^2} H_{\text{eff}}^H W_{\text{RF}} W_{BB} (W_{\text{RF}} W_{BB})^H H_{\text{eff}} \right).
\end{aligned} \tag{16}$$

The spectrum efficiency of the receiver and the mutual information of the transmitter have similar structures, so the analog receiver and the digital receiver can be designed in the same way as the transmitter.

The algorithm proposed in this paper is to iteratively solve a single column vector. As mentioned in [11], the expected performance of the algorithm can be achieved when the number of the iterations is set to 5. The HPD-PS algorithm in [13] requires $5N_s$ iterations, and the magnitude of the major complexity of each iteration is $\mathcal{O}(N_t^2 / N_{\text{RF}}^{t\,2})$, while the complexity of each iteration of the SIC algorithm in [11] is $\mathcal{O}(N_t^2 / N_s^2)$. The main complexity of each iteration of the algorithm presented in this paper is about $\mathcal{O}(2N_t N_s / N_{\text{RF}}^t)$. It can be seen that compared with other subconnection structure hybrid precoding algorithms, the computational complexity of the algorithm in this paper is not high.

In addition, for mmWave communications, one nature is broadband. Therefore, the case of extending the algorithm to wideband communication is briefly discussed here. The same as in narrowband systems, the extra orthogonality constraint on the digital precoder also has been negligible impact on spectral efficiency in mmWave OFDM systems [9]. That is to say, the main difference between the hybrid precoding schemes of wideband and narrowband systems is the design of analog precoding. In mmWave MIMO-OFDM systems, the signals of all subcarriers usually need to share an analog

```
Input: H
1:    [∼ , ∼ ,V_d] = svd(H) ;
2:    F^{opt} = (V_d)_{:,1:Ns} ;
3:    for l ≤ L do
4:        for n ≤ N_{RF} do
5:            ā_n = F_n^{opt} ⊘ abs(F_n^{opt})
6:        end for
7:    end for
8:    A = diag [ā_1, ā_2,···,ā_N]
9:    Compute D according to (16);
10:   Normalize D with (17);
11:   return A, D
```

ALGORITHM 1: Spectrum Efficient Hybrid Precoding Design Based on Subconnected Architecture.

precoder. Thus, the solution of analog precoding can be extended to wideband system by carrier aggregation and other related operations.

4. Simulation Results

In this section, the performance of the proposed algorithm is simulated and compared with the hybrid precoding algorithm based on continuous interference cancellation in [11] and the HPD-PS algorithm in [13]. In the result diagram, for reference, the performance of full-digital precoding and full-analog precoding for subconnection structures is given.

   In this simulation, the default parameters are set as follows. The channel matrix is generated from the channel description above. We model the propagation environment as a $N_{cl} = 8$ cluster environment with $N_{ray} = 10$ rays per cluster [24]. The angular spread of each cluster is 10 degrees. The carrier frequency is set as 28 GHz [11]. The UPA array is used for the antenna of transmitter and receiver, and the distance between the adjacent antennas is $d = \lambda/2$. We assumed that the AODs follow the uniform distribution within $[-\pi/6, \pi/6]$, while the AOAs follow the uniform distribution within $[-\pi, \pi]$ [22]. Moreover, the maximum number of iterations is set to 5.

4.1. Performance under Perfect CSI. We first consider the ideal state, that is, the millimeter-wave massive MIMO communication scenario in which the CSI can be obtained. Figure 3 shows a comparison of spectral efficiency in a millimeter-wave MIMO system with $NM \times N_r = 64 \times 16$, and the number of RF chain is 8. Obviously, the subconnection structure algorithm proposed in this paper is better than the SIC hybrid precoding algorithm in the whole analog signal-to-noise ratio range. In addition, the spectral efficiency of the proposed algorithm is close to that of the optimal unconstrained full-digital precoding algorithm with subconnected structure.

   Considering that the continuous interference cancellation algorithm was proposed in 2016, Figure 4 further shows that when the antenna size of the transmitter is $N_t = 144$ and the receiver is $N_r = 16$, $N = 4$, the proposed algorithm is compared with the spectrum efficiency of the hybrid precoding algorithm based on successive interference cancellation in [11] and the HPD-PS algorithm in [13]. It has been shown
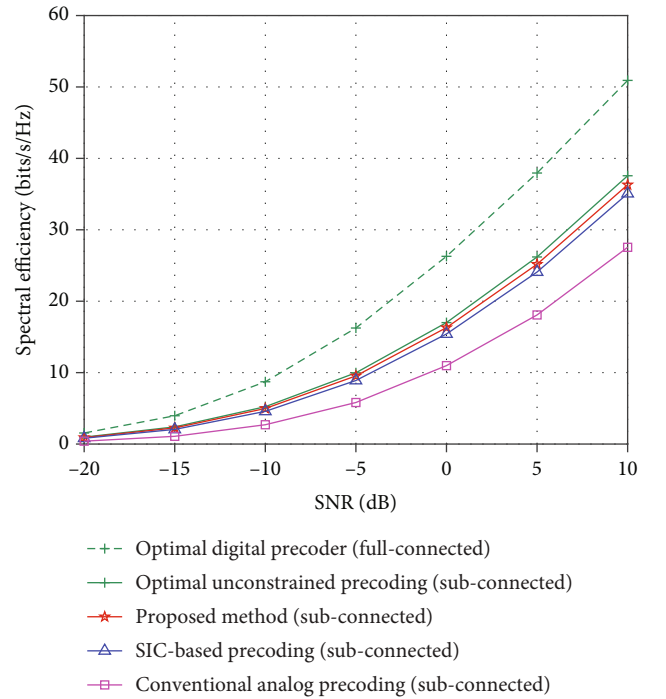

FIGURE 3: Spectral efficiency comparison of different algorithms, when $NM \times N_r = 64 \times 16 (N = 8)$.

in [13] that the performance of HPD-PS algorithm is better than that of SIC hybrid precoding algorithm, and the result can be seen in Figure 4. In addition, the hybrid precoding algorithm in this paper is much closer to the unconstrained full-digital precoding algorithm than the HPD-PS algorithm.

   There is a gap of 5 dB between the optimal fully connected method and the optimal unconstrained precoding for subconnected structure. However, the subconnection structure is much lower in cost and hardware complexity than the full-connection structure. Taking the system model of this article as an example, the fully connected structure requires at least $N^2 M$ phase shifters, while the subconnected structure only needs NM phase shifters.

4.2. Impact of Imperfect CSI. However, in the actual communication process, due to the lack of cooperation between the
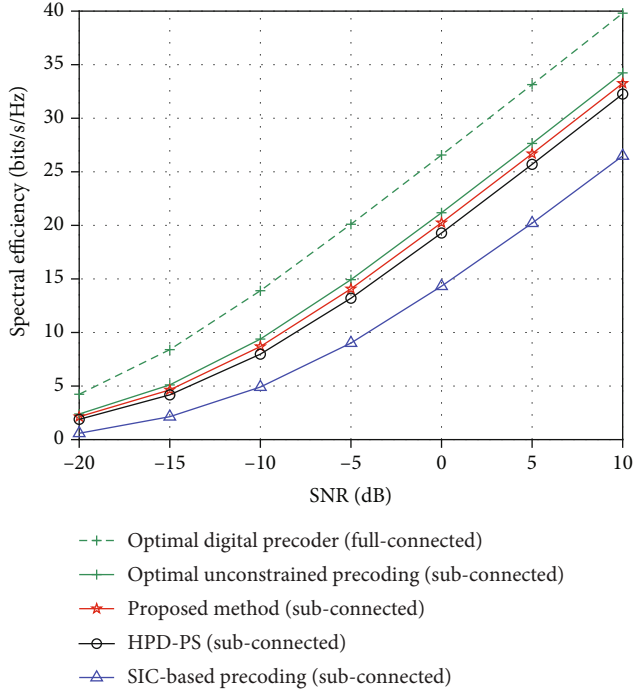
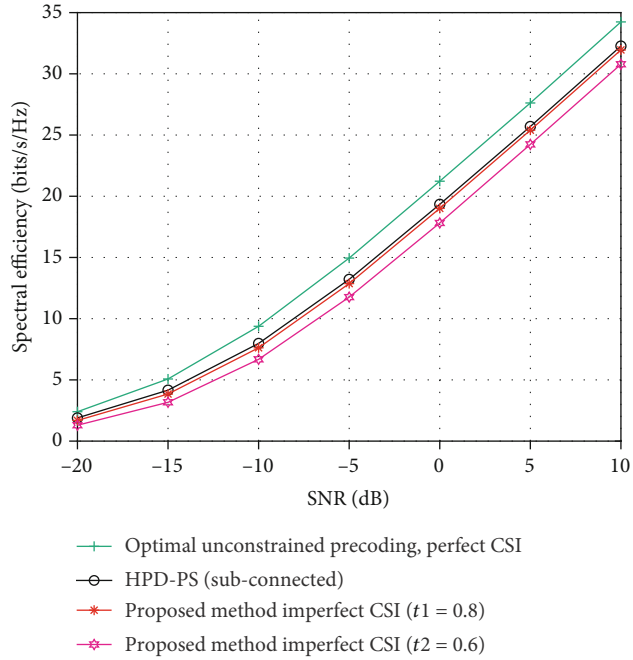FIGURE 4: Spectral efficiency comparison of different algorithms, when $NM \times N_r = 144 \times 16(N = 4)$.

FIGURE 5: Spectral efficiency by different algorithms under the perfect and the imperfect.
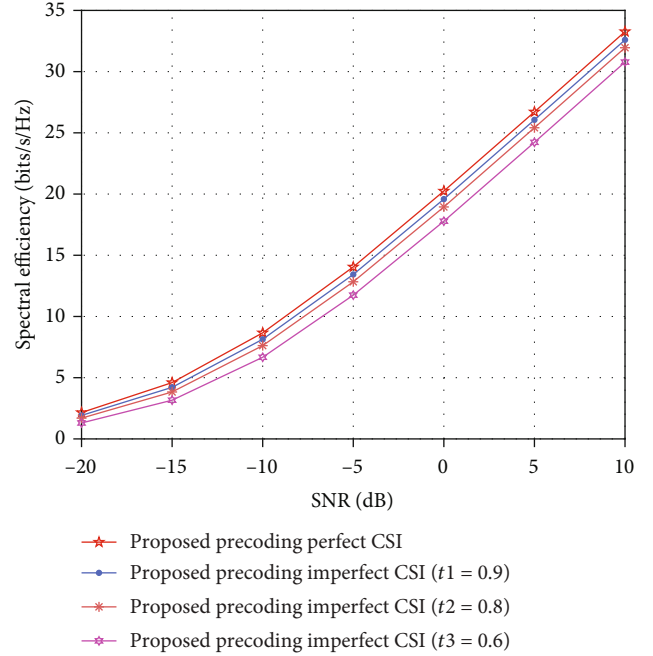

FIGURE 6: Impact of imperfect CSI on the proposed hybrid precoding algorithm.

the receiving end due to the limitation of the number of receiving antennas. No matter how accurate the channel estimation algorithm is, there will always be a fixed error bound, which cannot meet the specified arbitrary accuracy requirements [25, 26]. Therefore, it is of great research and practical value to analyze the effect of imperfect channel state information on downlink multiuser MIMO system performance and the method of obtaining and utilizing channel state information effectively.

The perfect result cannot be guaranteed by data acquisition and processing in communication network [27–30]. Although there are a lot of channel estimation algorithms, still with the ideal state, the perfect channel state information estimation has a certain gap. In the case of imperfect channel state information, the performance of the proposed hybrid precoding algorithm is compared. For the estimation of channel state information, the minimum mean square error method is used, and the expression is as [31] follows:

$$\widehat{\mathbf{H}} = t\mathbf{H} + \sqrt{1 - t^2}\mathbf{E}, \tag{17}$$

where $\widehat{\mathbf{H}}$ is the estimated channel matrix, $0 \le t \le 1$ presents the CSI accuracy, and $\mathbf{E}$ is the error matrix with entries following the distribution i.i.d. $\mathscr{CN}(0, 1)$.

Figure 5 shows the performance comparison of the proposed algorithm with HPD-PS and optimal unconstrained full-digital precoding under imperfect channel state information and perfect channel state information. Obviously, the performance of this algorithm is not as good as that of HPD-PS when $t = 0.6$, but it is close to that of HPD-PS when $t = 0.8$.

Other than that, Figure 6 shows the detailed performance simulation results of the proposed algorithm under imperfect

base station and the user, and some uncertain factors, such as uncertain channel estimation and limited feedback, perfect channel state information is difficult to obtain. Thus, the preprocessing of the transmit signal of the base station cannot completely eliminate the interference between the users as under the accurate channel state information, and the users cannot be used as the interference elimination strategy at

channel state information. As can be seen from the graph, the spectral efficiency of the algorithm is relatively stable from the perfect channel state information to $t = 0.8$, and even if $t = 0.6$, the performance of the algorithm does not decrease greatly.

5. Conclusions

In this paper, we mainly focus on the hybrid precoding of millimeter-wave large-scale MIMO systems based on sub-connection structures. Firstly, the hybrid precoding optimization problem is decoupled into analog precoding and digital precoding. In the optimization process, the digital precoding matrix is first fixed to solve the optimal analog precoder. Thanks to the block diagonal form of the analog precoding matrix; the optimization problem can be decomposed into a series of optimization subproblems and solved in turn. Finally, the optimal digital precoding matrix can be obtained by MMSE. The simulation results show that the spectrum efficiency of the proposed algorithm is better than that of SIC and HPD-PS. Considering the fact that the channel state information cannot be obtained completely, the performance of the algorithm under imperfect channel state information is simulated and analyzed. It can be seen that the performance degradation of the algorithm is acceptable even under imperfect CSI and is similar to the performance of full-digital precoding under perfect channel state information.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

References

[1] S. K. Yong and C.-C. Chong, "An overview of multigigabit wireless through millimeter wave technology: potentials and technical challenges," *EURASIP Journal on Wireless Communications and Networking*, vol. 2007, no. 1, 10 pages, 2006.

[2] W. H. Chin, Z. Fan, and R. Haines, "Emerging technologies and research challenges for 5G wireless networks," *IEEE Wireless Communications*, vol. 21, no. 2, pp. 106–112, 2014.

[3] F. Al-Ogaili and R. M. Shubair, "Millimeter-wave mobile communications for 5G: challenges and opportunities," in *2016 IEEE International Symposium on Antennas and Propagation (APSURSI)*, pp. 1003-1004, Fajardo, PR, USA, 2016.

[4] S. A. Busari, K. M. S. Huq, S. Mumtaz, L. Dai, and J. Rodriguez, "Millimeter-wave massive MIMO communication for future wireless systems: a survey," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 2, pp. 836–869, 2018.

[5] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 13, no. 3, pp. 1499–1513, 2014.

[6] R. Méndez-Rial, C. Rusu, N. González-Prelcic, and R. W. Heath, "Dictionary-free hybrid precoders and combiners for mmWave MIMO systems," in *2015 IEEE 16th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pp. 151–155, Stockholm, Sweden, 2015.

[7] F. Zhang and M. Wu, "Hybrid analog-digital precoding for millimeter wave MIMO Systems," in *2017 IEEE 17th International Conference on Communication Technology (ICCT)*, Chengdu, China, 2017.

[8] X. Liu, X. Li, S. Cao et al., "Hybrid precoding for massive mmWave MIMO systems," *IEEE Access*, vol. 7, pp. 33577–33586, 2019.

[9] X. Yu, J. Shen, J. Zhang, and K. B. Letaief, "Alternating minimization algorithms for hybrid precoding in millimeter wave MIMO systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 485–500, 2016.

[10] K. Xu, F. Zheng, P. Cao, H. Xu, and X. Zhu, "A divide-and-conquer precoding scheme for sub-connected massive MIMO systems," in *2019 IEEE 90th Vehicular Technology Conference (VTC2019-Fall)*, pp. 1–5, Honolulu, HI, USA, 2019.

[11] X. Gao, L. Dai, S. Han, C. L. I, and R. W. Heath, "Energy-efficient hybrid analog and digital precoding for MmWave MIMO systems with large antenna arrays," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 4, pp. 998–1009, 2016.

[12] Y. Chen, Y. Xia, Y. Xing, and L. Yang, "Low complexity hybrid precoding for mmWave massive MIMO systems," in *2017 26th Wireless and Optical Communication Conference (WOCC)*, pp. 1–5, Newark, NJ, USA, 2017.

[13] Y. Wang and W. Zou, "Hybrid precoding design for millimetre wave systems with the partially-connected structure," *IET Communications*, vol. 14, no. 4, pp. 561–567, 2020.

[14] W. Dong, T. Zhang, Z. Hu, Y. Liu, and X. Han, "Energy-efficient hybrid precoding for mmWave massive MIMO systems," in *2018 IEEE/CIC International Conference on Communications in China (ICCC Workshops)*, pp. 6–10, Beijing, China, 2018.

[15] Y. Xie, B. Li, Z. Yan, J. Fan, and M. Yang, "A general hybrid precoding method for mmWave massive MIMO systems," *Radioengineering*, vol. 27, no. 2, pp. 439–446, 2019.

[16] P. Xia, S.-K. Yong, J. Oh, and C. Ngo, "A practical SDMA protocol for 60 GHz millimeter wave communications," in *2008 42nd Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, USA, 2008.

[17] S. Han, C.-l. I, Z. Xu, and C. Rowell, "Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G," *IEEE Communications Magazine*, vol. 53, no. 1, pp. 186–194, 2015.

[18] S. Hur, T. Kim, D. J. Love, J. V. Krogmeier, T. A. Thomas, and A. Ghosh, "Millimeter wave beamforming for wireless backhaul and access in small cell networks," *IEEE Transactions on Communications*, vol. 61, no. 10, pp. 4391–4403, 2013.

[19] H. Xu, V. Kukshya, and T. S. Rappaport, "Spatial and temporal characteristics of 60-GHz indoor channels," *IEEE Journal on Selected Areas in Communications*, vol. 20, no. 3, pp. 620–630, 2002.

[20] V. Raghavan and A. M. Sayeed, "Sublinear capacity scaling laws for sparse MIMO channels," *IEEE Transactions on Information Theory*, vol. 57, no. 1, pp. 345–364, 2011.

[21] L. M. Correia and P. F. M. Smulders, "Characterisation of propagation in 60 GHz radio channels," *Electronics & Communication Engineering Journal*, vol. 9, no. 2, pp. 73–80, 1997.

[22] D. Fan, F. Gao, Y. Liu, Y. Deng, and A. Nallanathan, "Angle domain channel estimation in hybrid mmWave massive MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 17, no. 12, pp. 8165–8179, 2018.

[23] J. Zhao, F. Gao, W. Jia, S. Zhang, S. Jin, and H. Lin, "Angle domain hybrid precoding and channel tracking for mmWave massive MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 16, no. 10, pp. 6868–6880, 2017.

[24] R. Bansal, "Antenna theory; analysis and design," *Proceedings of the IEEE*, vol. 72, no. 7, pp. 989-990, 1984.

[25] J. Li, S. Cheng, Z. Cai, J. Yu, C. Wang, and Y. Li, "Approximate holistic aggregation in wireless sensor networks," *ACM Transactions on Sensor Networks*, vol. 13, no. 2, pp. 1–24, 2017.

[26] S. Cheng, Z. Cai, J. Li, and H. Gao, "Extracting kernel dataset from big sensory data in wireless sensor networks," *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 4, pp. 813–827, 2017.

[27] Z. Cai, S. Ji, J. He, A. G. Bourgeois, and L. Wei, "Distributed and asynchronous data collection in cognitive radio networks with fairness consideration," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 8, pp. 2020–2029, 2014.

[28] Z. Cai, Y. Duan, and A. G. Bourgeois, "Delay efficient opportunistic routing in asynchronous multi-channel cognitive radio networks," *Journal of Combinatorial Optimization*, vol. 29, no. 4, pp. 815–835, 2015.

[29] Z. Cai, S. Ji, J. He, and A. G. Bourgeois, "Optimal distributed data collection for asynchronous cognitive radio networks," in *2012 IEEE 32nd International Conference on Distributed Computing Systems*, pp. 245–254, Macau, China, 2012.

[30] Q. Chen, Z. Cai, L. Cheng, and H. Gao, "Low-latency data aggregation scheduling for cognitive radio networks with non-predetermined structure," *IEEE Transactions on Mobile Computing*, 2020.

[31] F. Rusek, D. Persson, Buon Kiong Lau, E. G. Larsson, T. L. Marzetta, and F. Tufvesson, "Scaling up MIMO: opportunities and challenges with very large arrays," *Signal Processing Magazine*, vol. 30, no. 1, pp. 40–60, 2013.