Hybrid Precoding in mmWave MIMO Broadcast Channels With Dynamic Subarrays and Finite-Alphabet Inputs

Juening Jin*, Chengshan Xiao†, Wen Chen* and Yongpeng Wu*

*Department of Electronic Engineering, Shanghai Jiao Tong University
Shanghai, 200240, China, Email: jueningjin@gmail.com, wenchen@sjtu.edu.cn, yongpeng.wu@sjtu.edu.cn
†Department of Electrical and Computer Engineering, Lehigh University
Bethlehem, PA 18015, USA, Email: xiaoc@lehigh.edu

Abstract—Hybrid precoding provides a tradeoff between spectral efficiency and power consumption in millimeter wave (mmWave) multiple-input multiple-output (MIMO) systems. In this paper, we investigate the partially-connected hybrid precoding design for mmWave MIMO broadcast channels with finite alphabet inputs. To enhance the spectral efficiency, a new algorithm is proposed to dynamically optimize the mapping strategy from radio frequency (RF) chains to transmit antennas such that the weighted sum of channel gains is maximized. Then we adopt the inexact alternating minimization method to design hybrid precoding matrices with given optimal mapping strategy and finite-alphabet inputs. Simulation results demonstrate the good performance of our proposed algorithm.

I. INTRODUCTION

Millimeter wave (mmWave) multiple-input multiple-output (MIMO) communication is a promising technique for future generation wireless communication systems. To mitigate the prohibitive power consumption of radio frequency (RF) chains at mmWave frequencies, hybrid precoding is proposed, which divides the processing needed for precoding between analog and digital domains to reduce the number of RF chains [1].

Based on the connecting strategies from RF chains to antennas, hybrid precoding are typically realized by two structures, i.e., fully- and sub-connected structures. While the former enjoys the full precoding gain with each RF chain connected to all the antennas, the latter has drawn much attention recently due to its lower power consumption and lower hardware complexity. In particular, compared to the fully-connected structure, each RF chain in the sub-connected structure is connected to only a subset of the antennas. Thus the total number of phase shifters in the sub-connected structure is reduced by a factor of the RF chain number. Due to the limited number of phase shifters, there will be non-negligible degradation in the sub-connected structure. Therefore, it is of importance to develop effective design methodologies for hybrid precoding with limited phase shifters.

There exist a few studies on hybrid precoding in the sub-connected structure [2]–[5]. In [2], an iterative hybrid precoding algorithm based on successive interference cancellation was proposed for single user mmWave MIMO systems. The work in [3] formulated the hybrid precoding design as a matrix factorization problem, and then adopted the alternating minimization method to solve this problem. Recently, [4] developed a novel technique that dynamically constructs the sub-connected structure for MIMO-OFDM systems, and the proposed dynamic subarray structure outperforms the fixed subarray structure. Finally, reference [5] proposed a modified k-means algorithm for sub-connected hybrid precoding with dynamic double phase shifters implementation.

Most existing works on hybrid precoding assume Gaussian inputs, which cannot be realized in practice. It is well known that practical systems utilize finite-alphabet inputs, such as phase-shift keying (PSK) or quadrature amplitude modulation (QAM). Furthermore, precoding designs under Gaussian inputs are quite suboptimal for systems with finite-alphabet inputs [6]–[10]. Therefore, the precoding design with finite-alphabet inputs has drawn increasing research interest in recent years.

In this paper, we study the hybrid precoding design for mmWave MIMO broadcast channels with dynamic subarray and finite-alphabet inputs. The contributions of this paper are summarized as follows:

- We propose a simple algorithm to dynamically optimize the mapping strategy from radio frequency (RF) chains to transmit antennas such that the weighted sum of channel gains is maximized.
- With the given optimal mapping strategy, we propose a hybrid precoding algorithm to maximize the weighted sum rate under finite-alphabet inputs. The proposed algorithm has about 2.5dB performance gain over the successive interference cancelation (SIC)-based hybrid precoding [2].

Notations: Boldface lowercase letters, boldface uppercase letters, and calligraphic letters are used to denote vectors, matrices and sets, respectively. The real and complex number fields are denoted by $\mathbb{R}$ and $\mathbb{C}$, respectively. The superscripts $(\cdot)^T$, $(\cdot)^*$ and $(\cdot)^H$ stand for transpose, conjugate, and conjugate transpose operations, respectively. $\text{tr}(\cdot)$ is the trace of a matrix; $\| \cdot \|$ denotes the Euclidean norm of a vector; $\| \cdot \|_F$ represents the Frobenius norm of a matrix; $E_x(\cdot)$
represents the statistical expectation with respect to \( \mathbf{x} \). \( I \) and \( 0 \) denote an identity matrix and a zero matrix, respectively, with appropriate dimensions; \( \odot \) denotes the Hadamard matrix products; \( I(\cdot) \) represents the mutual information; \( \Re \) and \( \Im \) are the real and image parts of a complex value; \( \log(\cdot) \) is used for the base two logarithm.

II. Problem Formulation

A. System Model

We consider a multiuser downlink channel with one base station (BS) and \( K \) noncooperative mobile stations (MSs). The BS is equipped with \( n_T \) antennas and \( n_{RF} \leq n_T \) radio frequency (RF) chains. The \( k \)th MS is equipped with \( n_R \) antennas. The total number of receive antennas is defined to be \( n_R = \sum_k n_{Rk} \). We will use the notation \( \{n_{R1}, \ldots, n_{RK}\} \times n_T \times n_{RF} \) to represent such a channel. For example, a \((2, 2, 2)\times 64 \times 6 \) channel has a 64-antenna BS with 6 RF chains and three 2-antenna MSs.

The BS sends independent data vectors \( \{\mathbf{x}_k\}_{k=1}^K \) to \( K \) MSs, where \( \mathbf{x}_k \in \mathbb{C}^{d_k \times 1} \) is the data vector intended for the \( k \)th MS. Without loss of generality, we assume that \( \{\mathbf{x}_k\}_{k=1}^K \) are zero-mean vectors with the same covariance matrix \( \mathbf{I} \). In the hybrid precoding architecture, each data vector \( \mathbf{x}_k \) is first precoded by a digital precoding matrix \( \mathbf{B}_k \in \mathbb{C}^{n_{RF} \times d_k} \). After passing through RF chains, the baseband signal \( \sum_k \mathbf{B}_k \mathbf{x}_j \) is furthered precoded in the analog-domain by an analog precoding matrix \( \mathbf{F} \in \mathbb{C}^{n_{RF} \times n_{RF}} \). Then the received signal \( \mathbf{y}_j \in \mathbb{C}^{n_R \times 1} \) at the \( k \)th MS in a narrowband system can be represented as

\[
\mathbf{y}_k = \mathbf{H}_k \mathbf{B}_k \mathbf{x}_k + \sum_{j=1}^K \mathbf{H}_k \mathbf{B}_{j} \mathbf{x}_j + \mathbf{n}_k, \quad k = 1, \ldots, K
\]

where \( \mathbf{H}_k \in \mathbb{C}^{n_R \times n_{RF}} \) is the channel matrix from the BS to the \( k \)th MS, and \( \mathbf{n}_k \in \mathbb{C}^{n_R \times 1} \) is the independent and identically distributed (i.i.d.) complex Gaussian noise with zero-mean and covariance \( \sigma_n^2 \mathbf{I} \).

In this paper, the analog precoding matrix \( \mathbf{F} \) is implemented by a dynamic phase shifter subarray, where each RF chain is connected to a subset of the antennas. We denote \( \mathcal{S}_j \) by the collection of BS antennas connected to \( j \)th RF chain. Then we need to partition \( n_T \) BS antennas into \( n_{RF} \) subsets \( \{\mathcal{S}_j\}_{j=1}^{n_{RF}} \) satisfying

\[
\bigcup_{j=1}^{n_{RF}} \mathcal{S}_j = \{1, 2, \ldots, n_T\}
\]

\[
\mathcal{S}_j \cap \mathcal{S}_\ell = \emptyset, \quad \forall j \neq \ell
\]

From (2), we conclude that if \( i \in \mathcal{S}_j \) (the \( j \)th BS antenna is connected to the \( j \)th RF chain), the \((i, j)\)th entry of \( \mathbf{F} \) has unit modulus, otherwise it is zero. Therefore, the constraints on \( \mathbf{F} \) can be expressed by

\[
|F_{ij}| = 1_{\mathcal{S}_j}(i), \quad \forall (i, j)
\]

where \( 1_{\mathcal{S}_j}(i) \) is the indicator function:

\[
1_{\mathcal{S}_j}(i) = \begin{cases} 
1 & \text{if } i \in \mathcal{S}_j \\
0 & \text{otherwise.}
\end{cases}
\]

The transmitted signal at the BS is restricted by a total power constraint \( P \):

\[
E \left| \sum_k \mathbf{F} \mathbf{B}_k \mathbf{x}_k \right|^2 = \sum_{k=1}^K \text{tr} (\mathbf{F}^H \mathbf{F} \mathbf{B}_k \mathbf{B}_k^H) \leq P.
\]

To eliminate the effect of \( \mathbf{F} \) on (5) and make our problem more tractable, we consider the following change of variables:

\[
\mathbf{F} = (\mathbf{F}^H \mathbf{F})^{-\frac{1}{2}} \Rightarrow \mathbf{B}_k = (\mathbf{F}^H \mathbf{F})^{-\frac{1}{2}} \mathbf{B}_k.
\]

Then the power constraint in (5) becomes

\[
\sum_{k=1}^K \text{tr} (\mathbf{B}_k^H \mathbf{B}_k) \leq P
\]

and the constraints on \( \mathbf{F} \) can be expressed as

\[
|F_{ij}| = \frac{1}{\sqrt{\left|\mathcal{S}_j\right|}} 1_{\mathcal{S}_j}(i), \quad \forall (i, j).
\]

Furthermore, plugging \( \mathbf{F} \) and \( \mathbf{B}_k \) into the system model in (1), we have

\[
\mathbf{y}_k = \mathbf{H}_k \mathbf{B}_k \mathbf{x}_k + \sum_{j=1 \atop j \neq k}^K \mathbf{H}_k \mathbf{B}_j \mathbf{x}_j + \mathbf{n}_k, \quad k = 1, \ldots, K
\]

where \( \mathbf{H}_k = \mathbf{H}_k \mathbf{F} \). From (7) and (9), we observe that \( \mathbf{H}_k \) and \( \mathbf{B}_k \) can be regarded as the effective channel matrix and precoding matrix for typical MIMO Gaussian broadcast channels, respectively. Therefore, the role of \( \mathbf{F} \) is to increase the gain of the effective channels \( \{\mathbf{H}_k\}_{j=1}^{K} \), and this is the motivation of our problem formulation in the subsection D.

B. Channel Model

Due to the high free space omnidirectional path loss and signal attenuation in the mmWave frequency bands, the scattering is limited such that there might be only a small number of paths over which the signals from the BS can reach the \( k \)th MS. This allows to use a multipath channel model, where the channel matrix is given by

\[
\mathbf{H}_k = \sqrt{\frac{n_{RF} \sqrt{P}}{L_k}} \sum_{\ell=1}^{L_k} \alpha_{k, \ell} \mathbf{a}(\theta_{k, \ell}^{MS}) \mathbf{a}(\theta_{k, \ell}^{BS})^H.
\]

Here, \( L_k \) denotes the number of physical propagation paths between the BS and the \( k \)th MS. Each path \( \ell \) is described by three parameters: complex gain \( \alpha_{k, \ell} \), angle of arrival \( \theta_{k, \ell}^{MS} \), and angle of departure \( \theta_{k, \ell}^{BS} \). The angles \( \{\theta_{k, \ell}^{MS}\}_{k, \ell} \) and \( \{\theta_{k, \ell}^{BS}\}_{k, \ell} \) are i.i.d. uniformly distributed over \([0, 2\pi)\), and the complex gains \( \{\alpha_{k, \ell}\}_{k, \ell} \) are i.i.d. complex Gaussian distributed with zero-mean and unit-variance. The array steering vectors of the antenna arrays deployed at the BS and MS are denoted by \( \mathbf{a}(\theta_{k, \ell}^{BS}) \) and \( \mathbf{a}(\theta_{k, \ell}^{MS}) \), respectively. In this paper, the BS and
MSs adopt uniform linear arrays, whose array steering vector \( \mathbf{a}(\theta) \) is given by
\[
\mathbf{a}(\theta) = \sqrt{\frac{\lambda}{N}} \left[ 1, e^{-j \frac{2\pi}{d_S} \sin \theta}, \ldots, e^{-j \frac{2\pi}{d_S} (N-1) \sin \theta} \right]^T.
\]
where \( N \) is the number of antenna element, \( \lambda \) is the wavelength of the carrier frequency and \( d_S = \frac{1}{2} \lambda \) is the antenna spacing.

C. System Performance Metric

Throughout the paper, we assume that the BS has the knowledge of all MSs’ channels \( \{H_k\}_{k=1}^K \), and the \( k \)th MS only knows its own channel matrix \( H_k \). In addition, each entry of the data vectors \( \{x_k\}_{k=1}^K \) is drawn from an equiprobable constellation set \( \mathcal{A} = \{a_1, a_2, \ldots, a_M\} \) with cardinality \( M \), i.e., \( x_k \in \mathcal{A}^{d_k \times 1}, k = 1, \ldots, K \). Under these assumptions, the mutual information between \( x_k \) and \( y_k \), denoted by \( I(x_k; y_k) \), can be used to characterize the data rate from the BS to the \( k \)th MS. However, since \( \{x_k\}_{k=1}^K \) are drawn from a discrete constellation set, the computational complexity for evaluating \( I(x_k; y_k) \) grows exponentially with respect to the total number of data streams \( d \). To address this issue, we apply the finite-alphabet signal Gaussian interference (FASGI) approximation proposed in [9]. The key idea behind the FASGI approximation is to model the sum interference as a Gaussian distributed signal with the same covariance matrix. Consider the equivalent system model in (9)
\[
y_k = H_k B_k x_k + \sum_{j \neq k} H_k B_j x_j + n_k, \quad k = 1, \ldots, K.
\]

According to the central limit theorem, when \( K \) is large, the sum interference plus noise \( \sum_{j \neq k} H_k B_j x_j + n_k \) tends toward the Gaussian distribution with the covariance matrix
\[
C_k = \sigma_k^2 I + \sum_{j \neq k} H_k B_j H_k^H.
\]

Based on the FASGI approximation, the system model in (12) is reduced to
\[
C_k^{-\frac{1}{2}} y_k = C_k^{-\frac{1}{2}} H_k B_k x_k + z_k
\]
where \( z_k \sim \mathcal{CN}(0, I) \) is the normalized sum interference plus noise. The corresponding constellation-constrained mutual information is given by
\[
I(x_k; C_k^{-\frac{1}{2}} y_k) = \log N_k - \frac{1}{N_k} \sum_{m=1}^{N_k} E_{z_k} \left\{ \log \sum_{n=1}^{N_k} \exp \left( -\| C_k^{-\frac{1}{2}} H_k B_k e^{(m)}_{mn} + z_k \|^2 - \| z_k \|^2 \right) \right\}
\]
where \( N_k = M^{d_k} \) and \( e^{(k)}_{mn} \in \mathbb{C}^{d_k \times 1} \) is the difference of two possible transmit signals from \( \mathcal{A}^{d_k \times 1} \).

The constellation-constrained mutual information as well as its gradient is difficult to compute directly because they have no closed form expressions. In order to estimate \( I(x_k; C_k^{-\frac{1}{2}} y_k) \) and its gradient, we need to use monte carlo methods, whose complexity is prohibitively high especially when \( N_k \) is large. This issue can be mitigated by using an approximation of \( I(x_k; C_k^{-\frac{1}{2}} y_k) \) derived in [8]
\[
\hat{I}_k(F, B_k) = \log N_k - \frac{1}{N_k} \sum_{m=1}^{N_k} \log \sum_{n=1}^{N_k} \exp \left( \zeta^{(k)}_{mn} \right)
\]
where \( \zeta^{(k)}_{mn} \) is given by
\[
\zeta^{(k)}_{mn} = -\frac{1}{2} (e^{(k)}_{mn})^H F^H H_k^H C_k^{-1} H_k F B_k e^{(k)}_{mn}.
\]

This approximation is accurate for arbitrary channel and precoding matrices, and its computational complexity is several orders of magnitude lower than that of the original mutual information in (15).

D. Hybrid Precoding Design

Using the approximated mutual information in (16), the hybrid precoding problem, which maximizes the weighted sum rate under the power constraint, is formulated as
\[
\max_{F, \{B_k\}, \{S_j\}} \sum_{k=1}^{K} w_k \hat{I}_k(F, B_k)
\]
subject to
\[
\sum_{k=1}^{K} \text{tr}(B_k^H B_k) \leq P
\]
\[
\{F\}_{ij} = [S_j]^{-\frac{1}{2}} 1_{S_j(i), \forall (i,j)}
\]
\[
\{S_j\}_{j=1}^{n_RF} \text{satisfies (2)}
\]
where \( w_k \geq 0 \) with \( \sum_{k=1}^{K} w_k = 1 \). To make the structure of problem (18) more clear, we rewrite it as follows
\[
\max_{\{S_j\}} \quad R(\{S_j\})
\]
subject to
\[
\{S_j\}_{j=1}^{n_RF} \text{satisfies (2)}
\]
where \( R(\{S_j\}) \) is the optimal value to problem (18) with given partition of subsets, i.e.,
\[
R(\{S_j\}) = \max_{F, \{B_k\}} \sum_{k=1}^{K} w_k \hat{I}_k(F, \{B_k\})
\]
subject to
\[
\sum_{k=1}^{K} \text{tr}(B_k^H B_k) \leq P
\]
\[
\{F\}_{ij} = [S_j]^{-\frac{1}{2}} 1_{S_j(i), \forall (i,j)}
\]

Problem (19) is a combinatorial optimization problem for which finding the optimal solution requires an exhaustive search over all possible partition of subsets. Let \( \{S_j^{(l)}\}_{j=1}^{n_RF} \) denotes the \( l \)th given partition of subsets satisfying (2), and \( K_S \) denotes the total number of ways to partition \( n_T \) antennas into \( n_{RF} \) nonempty subsets:
\[
K_S = \frac{1}{n_{RF}!} \sum_{k=0}^{n_{RF}} (-1)^{n_{RF}-k} \binom{n_T}{k} k^{n_T}.
\]
Then we can rewrite problem (19) as

\[
\max_{\ell \in \{1, \ldots, K_\ell\}} R(S^{(\ell)}_1, \ldots, S^{(\ell)}_{n_{\text{RF}}}).
\]

(22)

Although (22) provides a theoretically possible way for solving the hybrid precoding problem (18), its computational complexity is prohibitive even for a small number of antennas and RF chains. For example, when \(n_T = 16\) and \(n_{\text{RF}} = 4\), \(K_\ell\) is equal to \(1.718 \times 10^8\), which means that we need to solve problem (20) over ten million times to obtain the optimal analog and digital precoding matrices.

We propose a new formulation to reduce the computational burden of problem (22). Recall that the role of \(F\) is to increase the gain of the effective channel matrices. Therefore, instead of maximizing the weighted sum rate, we design \(F\) and \(\{S^{(\ell)}_j\}_{j=1}^{n_{\text{RF}}}\) such that the weighted sum of the effective channel gains \(\sum_k w_k \text{tr}(H^*_k H_k)\) is maximized. The dynamic subarray partitioning problem is then formulated as

\[
\begin{align*}
\max_{F, \{S_j\}} & \quad \sum_{k=1}^K w_k \text{tr}(F^H W F) \\
\text{subject to} & \quad |F_{ij}| = |S_j|^{-\frac{1}{2}} 1_{S_j}(i), \forall (i, j) \quad \{S_j\}_{j=1}^{n_{\text{RF}}} \text{satisfies (2)}
\end{align*}
\]

where \(W = \sum_k w_k H^*_k H_k\).

The complete procedure of our proposed hybrid precoding design is summarized as follows. First, solve problem (23) to obtain its optimal solution \(\hat{F}_{\text{init}}\), and \(\{\hat{S}^*_j\}_{j=1}^{n_{\text{RF}}}\). Second, insert \(\{\hat{S}^*_j\}_{j=1}^{n_{\text{RF}}}\) into problem (20) and then solve problem (20) to obtain the optimal \(\hat{F}^*\) and \(\{\hat{B}^*_k\}_{k=1}^K\). Note that \(\hat{F}_{\text{init}}\) serves as a good initial point for solving (20). Third, recover the corresponding analog and digital precoding matrices \(\hat{F}^*\) and \(\{\hat{B}^*_k\}_{k=1}^K\) from \(\hat{F}^*\) and \(\{\hat{B}^*_k\}_{k=1}^K\).

### III. Dynamic Subarray Partitioning Design

In this section, we propose a low complexity algorithm to solve problem (23). Plugging \(\hat{F} = F(H^H F)^{-\frac{1}{2}}\) into problem (23), we obtain

\[
\max_{F \in \mathcal{F}} \quad \text{tr}\left((H^H F)^{-\frac{1}{2}} F^H W F (H^H F)^{-\frac{1}{2}}\right)
\]

(24)

where the feasible set \(\mathcal{F}\) is given by

\[
\mathcal{F} = \{F; |F_{ij}| = 1_{S_j}(i), \forall (i, j); \{S_j\}_{j=1}^{n_{\text{RF}}} \text{satisfies (2)}\}.
\]

(25)

The main difficulty of problem (24) lies in the fact that the constraints in \(\mathcal{F}\) depend on \(\{S_j\}_{j=1}^{n_{\text{RF}}}\). The following proposition address this issue by providing a set of new constraints to characterize \(\mathcal{F}\).

**Proposition 1:** The feasible set \(\mathcal{F}\) of problem (24) can be characterized by

\[
|F_{ij}| \in \{0, 1\}, \forall (i, j)
\]

\[
|F_{\bullet \bullet}|_0 = 1, \forall i
\]

(26)

where \(F_{\bullet \bullet}\) denotes the \(i\)th row of \(F\), and \(|\cdot|_0\) measures the number of nonzero elements in a vector.

**Proof:** We start with the necessary condition, i.e., if \(F \in \mathcal{F}\), then \(F\) satisfies (26). Since \(|F_{ij}| = 1_{S_j}(i)\), we must have \(|F_{ij}| \in \{0, 1\}\). In addition, based on equations (2), the \(i\)th antenna belongs to exactly one subset, thus \(|F_{\bullet \bullet}|_0 = 1\).

Next, we prove the sufficient condition, i.e., if \(F\) satisfies (26), then \(F \in \mathcal{F}\). Let \(S_j\) denotes the collection of nonzero entries in the \(j\)th column of \(F\). Then \(F\) can be expressed as

\[
|F_{ij}| = \begin{cases} 
1 & \text{if } i \in S_j \\
0 & \text{otherwise.}
\end{cases}
\]

(27)

Since \(|F_{\bullet \bullet}|_0 = 1\) for all \(i \in \{1, 2, \ldots, n_T\}\), we must have

\[
\bigcup_{j=1}^{n_{\text{RF}}} S_j = \{1, 2, \ldots, n_T\}
\]

(28)

\[
S_j \cap S_l = \emptyset, \forall j \neq l.
\]

(29)

Therefore, \(\{S_j\}_{j=1}^{n_{\text{RF}}} \) satisfies (2) and this completes the proof.

According to proposition 1, we rewrite problem (24) as

\[
\begin{align*}
\max_{F} & \quad \text{tr}\left((H^H F)^{-\frac{1}{2}} F^H W F (H^H F)^{-\frac{1}{2}}\right) \\
\text{subject to} & \quad |F_{ij}| \in \{0, 1\}, \forall (i, j) \\
& \quad |F_{\bullet \bullet}|_0 = 1, \forall i.
\end{align*}
\]

(30)

Problem (30) is intractable due to the discrete constraints \(|F_{ij}| \in \{0, 1\}\) and \(|F_{\bullet \bullet}|_0 = 1\). Therefore, we first drop the constraints of problem (30) and consider the unconstrained problem

\[
\max_{F} \quad \text{tr}\left((H^H F)^{-\frac{1}{2}} F^H W F (H^H F)^{-\frac{1}{2}}\right)
\]

(31)

Problem (31) is a generalized eigenvalue problem, and its optimal solution is given below.

**Proposition 2:** Let \(U_W \in C^{n_T \times n_{\text{RF}}}\) be the eigenvectors of \(W\) corresponding to the largest \(n_{\text{RF}}\) eigenvalues. For any unitary matrix \(R \in C^{n_{\text{RF}} \times n_{\text{RF}}}\), \(F = U_W R\) is a globally optimal solution of problem (31).

**Proof:** The proof is omitted due to space limits.

In general, there does not exist \(R\) such that the unconstrained solution \(U_W R\) is feasible to problem (30). However, we can use \(U_W R\) to find a nearby feasible solution. Specifically, consider the following minimization problem

\[
\begin{align*}
\min_{F, R \in \mathcal{U}} & \quad \|F - U_W R\|_F^2 \\
\text{subject to} & \quad |F_{ij}| \in \{0, 1\}, \forall (i, j) \\
& \quad |F_{\bullet \bullet}|_0 = 1, \forall i
\end{align*}
\]

(32)

where \(\mathcal{U}\) denotes the set of unitary matrices. Since the optimization variables \(F\) and \(R\) are separate, we adopt the alternating minimization approach to solve problem (32).

Given \(R\), problem (32) is reduced to

\[
\begin{align*}
\min_{F} & \quad \|F - U_W R\|_F^2 \\
\text{subject to} & \quad |F_{ij}| \in \{0, 1\}, \forall (i, j) \\
& \quad |F_{\bullet \bullet}|_0 = 1, \forall i
\end{align*}
\]

(33)
Let \( \ell^*(i) = \arg\max_{\ell \leq \ell_{\text{max}}} |U_i R_{i,\ell}| \), then the optimal solution of problem (33) can be expressed as
\[
F_{ij} = \begin{cases} 
|U_i R_{i,\ell}| & \text{if } j = \ell^*(i) \\
0 & \text{otherwise.} 
\end{cases}
\tag{34}
\]

Given \( F \), problem (32) is reduced to an orthogonal procrustes problem
\[
\min_{R \in \mathbb{C}^{L \times L}} ||F - U_i R||_F^2. 
\tag{35}
\]
Denote the singular value decomposition of \( F^H U W \) by
\[
F^H U W = \tilde{U} \Sigma \tilde{V}^H
\]
then the optimal solution of problem (35) can be expressed as [11]
\[
R = \tilde{V} \tilde{U}^H. 
\tag{37}
\]
Combining (34) and (37), we propose a simple algorithm for problem (32), which is summarized in Algorithm 1.

**Algorithm 1 Dynamic subarray partitioning**

1. Given \( U_i W \). Set initial unitary matrix \( R \).
2. repeat
   - Update \( F \) by (34).
   - Update \( R \) by (37).
until a stopping criterion triggers.
3. Return \( F_{\text{init}} = F(\mathcal{F}^H F)^{-1} \) and the corresponding \( \{S_j^*\} \).

Note that when \( F_{\text{init}}^* \) is determined, the corresponding \( \{S_j^*\}_{j=1}^{n_{RF}} \) is given by
\[
S_j^* = \left\{ i \left| \left| F_{\text{init},ij}^* \right| \neq 0, \forall i \right. \right\}, \quad j = 1, \ldots, n_{RF}. 
\tag{38}
\]

IV. HYBRID PRECODING WITH FINITE-ALPHABET INPUTS

In this section, we propose a hybrid precoding algorithm to solve problem (20) with \( \{S_j^*\}_{j=1}^{n_{RF}} \) computed by Algorithm 1. Note that the constraint \( \left| F_{ij} \right| = \left| S_j^* \right|^{-1/2} |S_j^*| (i) \) implies that only the phase of nonzero \( F_{ij} \) can be changed. Therefore, instead of using \( F \) as the optimization variable, it is more convenient to optimize the phase of nonzero entries in \( F \). Define a phase matrix \( \Phi \) as
\[
\Phi_{ij} = \begin{cases} 
\angle F_{ij} & \text{if } |F_{ij}| \neq 0 \\
0 & \text{otherwise.} 
\end{cases}
\tag{39}
\]
where \( \angle F_{ij} \) represents the phase of a nonzero \( F_{ij} \). Then \( F \) can be expressed as
\[
F_{ij} = |S_j^*|^{-1/2} \exp(p_{ij} \Phi_{ij}) 1_{S_j^*}(i), \quad \forall (i,j).
\tag{40}
\]

Using \( \Phi \) as the optimization variable and rewriting \( F \) as \( F(\Phi) \), we can rewrite problem (20) as the following problem
\[
\begin{align*}
\text{maximize} & \quad \sum_{k=1}^K w_k \hat{I}_k(\Phi, \{B_k\}) \\
\text{subject to} & \quad \sum_{k=1}^K \text{tr}(B_k^H B_k) \leq P. 
\end{align*}
\tag{41}
\]

We propose an inexact alternating minimization algorithm to solve problem (41). Based on the results in [12], the gradients of \( \hat{I}_k(\Phi, \{B_k\}) \) with respect to \( \Phi \) and \( \{B_k\}_{k=1}^K \) are
\[

\begin{align*}
\nabla_{\Phi} \hat{I}_k & = 2 \Re(\nabla_{\Phi} \hat{I}_k \hat{I}_k^*) \\
\nabla_{B_k} \hat{I}_k & = \left\{ \begin{array}{ll}
H_k^* G_k E_k & \text{if } \ell = k \\
-H_k^* G_k E_k G_k^H H_k B_k & \text{if } \ell \neq k.
\end{array} \right.
\end{align*}
\tag{42}
\]
where \( H_k = H_k \hat{F}, \quad G_k = C_k^{-1/2} H_k B_k \),
\[
E_k = \frac{1}{2} \sum_{m,n} e_m(k) e_n(k)^H
\tag{43}
\]
\[
\nabla_{F} \hat{I}_k = H_k^* G_k E_k \left[ I - G_k^H H_k F \sum_{j \neq k} (B_j B_j^H) \right].
\tag{44}
\]
The inexact alternating minimization algorithm updates \( \Phi \) and \( B = [B_1^T, \ldots, B_K^T]^T \) using gradient information. Given \( B \), we update \( \Phi \) by the following rule
\[
\Phi := \Phi + \rho_w \sum_{k=1}^K w_k \nabla_{\Phi} \hat{I}_k
\tag{45}
\]
where \( \rho \) is the stepsize for \( \Phi \). Then we fix \( \Phi \) and update \( B \) according to the projected gradient method, i.e.,
\[
B := \text{Proj}_{B} \left[ B + \rho_B \sum_{k=1}^K w_k \nabla_{B} \hat{I}_k \right]
\tag{46}
\]
where \( \nabla_{B} \hat{I}_k = [\nabla_{B_1} \hat{I}_k^T, \ldots, \nabla_{B_K} \hat{I}_k^T]^T \), \( \rho_B \) is the stepsize for \( B \), and \( \text{Proj}_{B} [\cdot] \) is defined as
\[
\text{Proj}_{B} [X] = \begin{cases} 
X & \text{if } ||X||_F \leq P \\
\frac{\sqrt{P}}{||X||_F} X & \text{otherwise.}
\end{cases}
\tag{47}
\]
The details of our proposed hybrid precoding algorithm is summarized in Algorithm 2.

**Algorithm 2 Hybrid precoding algorithm**

1. Given \( \{S_j^*\}_{j=1}^{n_{RF}} \). Set initial \( \Phi \) and \( B \) based on \( F_{\text{init}}^* \).
2. repeat
   - Compute \( \rho_B \) via backtracking line search [13].
   - Update \( B \) according to (46).
   - Compute \( \rho_B \) via backtracking line search [13].
   - Update \( B \) according to (47).
until a stopping criterion triggers.
3. Return \( F(\Phi) \) and \( \{B_k\}_{k=1}^K \).

V. SIMULATION RESULTS

In this section, we provide numerical examples to evaluate the performance of our proposed hybrid precoding algorithm. We first consider a \( 2 \times 64 \times 2 \) point-to-point MIMO channel. The number of physical propagation paths \( L \) is set to 12. The input signal is drawn from QPSK modulation, and the signal-to-noise ratio (SNR) is defined as \( \text{SNR} = \frac{E_b}{N_0} \), where \( \sigma^2 \) is the
noise power at the MS. Finally, the average mutual information is plotted versus SNR over 500 channel realizations.

We set the optimal unconstrained precoding [7] as a benchmark, and then make comparison between our proposed hybrid precoding with the SIC-based hybrid precoding proposed in [2]. The results in Fig. 1 shows that our proposed hybrid precoding algorithm has about 2.5dB performance gain over the SIC-based hybrid precoding because 1) our proposed algorithm is designed based on finite-alphabet inputs; 2) our proposed algorithm utilizes the dynamic subarray while the SIC-based precoding considers fixed subarray. Then we consider a \( \{2, 2, 2, 2\} \times 64 \times 8 \) MIMO broadcast channel. The number of physical propagation paths \( L \) is set as 12, and the input signal is drawn from QPSK modulation. Fig.2 depicts the comparison result with the fixed subarray hybrid precoding, which utilize Algorithm 2 to solve problem (20) with the following given \( \{S_j\} \):

\[
S_j = \{ (j - 1)q + 1, (j - 1)q + 2, \ldots, (j - 1)q + g \}, \quad \forall j
\]

where \( q = \frac{2g}{M^2} \). The result in Fig. 2 shows that our proposed (dynamic subarray) hybrid precoding has about 1.5dB performance gain over the fixed subarray hybrid precoding in the high SNR regime.

VI. CONCLUSION

This paper considers the hybrid precoding design for mmWave MIMO broadcast channels with dynamic subarray and finite-alphabet inputs. We first proposes a simple algorithm to dynamically optimize the mapping strategy from RF chains to transmit antennas such that the weighted sum of channel gains is maximized. Then we design an inexact alternating minimization based hybrid precoding algorithm to maximize the weighted sum rate under finite-alphabet inputs. The good performance of the proposed algorithm is demonstrated by simulation results.

ACKNOWLEDGMENT

The work of C. Xiao was supported in part by US National Science Foundation under grants ECCS-1231848, ECCS-1408316 and ECCS-1539316. The work of W. Chen was supported by Shanghai STCSM 16JC1402900 and 17510740700, National Science and Technology Major Project 2017ZX03001002-005 and 2018ZX03001009-002, NSF China 61671294, and Guangxi NSF 2015GXNSFDA139037.

REFERENCES