Optimal Analog Precoder Design for Hybrid Beamforming Is Possible

Jiaen Lv, Tianyu Wang, and Shaowei Wang

***Abstract*—Hybrid beamforming is a cost-effective solution for millimeter wave massive multiple-input-multiple-output communications, which highly reduces the number of required radio frequency chains by jointly using a digital precoder and an analog precoder. The digital precoder can be optimized by using the classical least squares method. However, due to the discrete resolution of analog phase shifters, the analog precoder design is a high-dimensional integer programming problem, for which the exsiting methods can only provide suboptimal solutions. In this paper, we first show that optimal analog precoder design is possible in typical antenna settings. Specifically, we show that this high-dimensional problem can be divided into multiple low-dimensional subproblems in parallel and there exists an analytical lower bound for each subproblem that can be efficiently utilized to obtain the optimal solution by using the branch and bound method. Numerical results show that the proposed hybrid precoder with an optimal analog precoder can achieve 98% performance of the fully digital solution in a typical antenna setting.**

***Index Terms*—Branch and bound, hybrid beamforming, low-resolution phase shifters, massive MIMO.**

I. Introduction

In millimeter-wave (mmWave) massive multiple-input-multiple-output (MIMO) communications, beamforming can be implemented by using fully digital precoders/combiners in the baseband, where each antenna requires a dedicated radio frequency (RF) chain to generate the corresponding RF signal [1]. However, as the number of antennas keeps increasing, the consequent hardware cost and power consumption increase dramatically, which quickly become unacceptable in practical systems [2], [3]. To address this issue, a joint analog-digital solution, referred to as hybrid beamforming, has been proposed [4]. Compared to the fully digital solution, hybrid beamforming introduces an additional analog precoder/combiner between the RF chains and antennas in the transmitter/receiver, which can highly reduce the number of required RF chains without greatly deteriorating the beamforming performance.

The analog precoder/combiner consists of multiple analog phase shifters, which modulate the phases of RF signals without changing the signal power. This unit modulus constraint makes the hybrid beamforming design intractable even for medium-sized cases [5]. In [6], the hybrid beamforming design is formulated as a matrix reconstruction problem, for which an orthogonal matching pursuit algorithm is proposed. However, the spectral efficiency is considerably decreased as compared to the fully digital solution. In [7], a cross-entropy algorithm is proposed to improve the spectral efficiency. In [8], a special case where the number of RF chains equals the number of data streams is considered and an iterative phase matching algorithm is proposed. However, the performance decreases when the number of RF chains is greater than the data streams, which is the common case in practical deployment.

Instead of regarding the hybrid beamforming design as a holistic optimization problem, the alternating minimization approach divides it into the digital precoder/combiner design and the analog precoder/combiner design and alternatively solves the analog and the digital precoders/combiners while fixing the other [9]–[12]. In [9], the digital precoder/combiner design is shown to be a least squares problem, for which an optimal solution can be obtained analytically. The analog precoder/combiner design can be formulated as a manifold optimization problem, for which a suboptimal solution is provided with high computational complexity. In [10], a Barzilai-Borwein gradient algorithm is proposed, which reduces the computational complexity of the manifold optimization without greatly decreasing the beamforming performance.

To reduce the power consumption and hardware cost of the analog precoder/combiner, low-resolution phase shifters are considered [13]. The assumption of limited resolution introduces additional integer constraints on the analog precoder/combiner, which further complicates the hybrid beamforming design. A straightforward approach is to quantify the continuous solutions given by algorithms considering infinite-resolution phase shifters. However, this approach may result in significant loss of spectral efficiency when the resolution is low [14]. In [11], the analog precoder/combiner design is formulated as an integer programming problem, for which a coordinate descent algorithm is proposed. In [12], the analog precoder/combiner design is formulated as a lattice decoding problem and a Babai's algorithm is proposed. Due to the intractable problem, all existing methods can only achieve suboptimal solutions to analog precoder/combiner design.

In this paper, we propose a parallel branch and bound algorithm to achieve the optimal analog precoder/combiner for hybrid beamforming with low-resolution phase shifters. The analog precoder/combiner design problem is an integer programming problem with high dimensionality. In fact, the search space size of the problem grows exponentially with the number of antennas, and the computational burden is

Manuscript received October 18, 2022; revised January 9, 2023; accepted February 7, 2023. This work was partially supported by the National Natural Science Foundation of China under Grants 61931023 and U1936202. The associate editor coordinating the review of this article and approving it for publication was K.-T. Feng. *(Corresponding author: Shaowei Wang.)*

The authors are with the School of Electronic Science and Engineering, Nanjing University, Nanjing 210023, China (email: dz21230022@smail.nju.edu.cn; tianyu.alex.wang@nju.edu.cn; wangsw@nju.edu.cn).

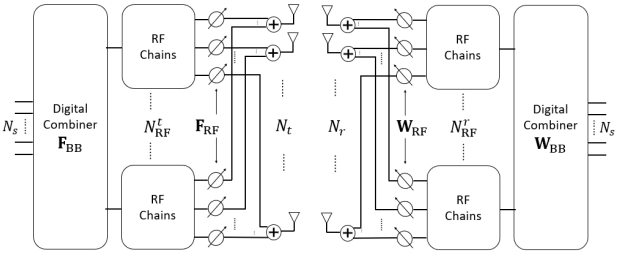

Fig. 1: mmWave massive MIMO system.

prohibitive as the number of antennas becomes large. We show that the optimal solution to analog precoder/combiner design problem can be constructed from the optimal solutions to multiple parallel subproblems. Thus, the overall search space size is reduced and increases only linearly with the number of antennas. Therefore, we can apply classical branch and bound method for each subproblem to show that the optimality is achievable with acceptable computational complexity. Specifically, the analog precoder/combiner design is divided into multiple tree search problems with reduced number of dimensions, each of which has an analytical lower bound that can be efficiently utilized by the branch and bound method to prune suboptimal branches. Numerical results show that the proposed algorithm outperforms the existing algorithms in terms of spectral efficiency and achieves 98% performance of the fully digital solution in a typical antenna setting.

II. System Model

Consider an mmWave massive MIMO system shown in Fig. 1, where a transmitter with N_t antennas and N_{RF}^t RF chains communicates N_s data streams to a receiver with N_r antennas and N_{RF}^r RF chains. The RF chains at the transmitter and receiver are subject to constraints $N_s \leqslant N_{\mathrm{RF}}^t \leqslant N_t$ and $N_s \leqslant N_{\mathrm{RF}}^r \leqslant N_r$, respectively [6]. $\mathbf{s}$ denotes an $N_s \times 1$ symbol vector representing N_s normalized data streams, i.e., $\mathbb{E}[\mathbf{s}\mathbf{s}^H] = \mathbf{I}_{N_s}$. The transmitted signal is then given by

$$\mathbf{x} = \mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\mathbf{s}, \tag{1}$$

where $\mathbf{F}_{\mathrm{BB}}$ is the $N_{\mathrm{RF}}^t \times N_s$ digital precoder in the baseband and $\mathbf{F}_{\mathrm{RF}}$ is the $N_t \times N_{\mathrm{RF}}^t$ analog precoder in the RF domain. The normalized transmit power constraint is given by $\|\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\|_F^2 = N_s$, where $\|\cdot\|_F$ denotes the Frobenius norm of a matrix. The analog precoder is implemented by using $N_t \times N_{\mathrm{RF}}^t$ low-resolution phase shifters connecting each RF chain and each antenna, each of which has N_{PS} discrete phases, i.e., $\mathbf{F}_{\mathrm{RF}} \in \mathcal{P}^{N_t \times N_{\mathrm{RF}}^t}$ with $\mathcal{P} = \{e^{2j\pi n/N_{\mathrm{PS}}} | n = 1, 2, ..., N_{\mathrm{PS}}\}$.

The received signal after combining is then given by

$$\mathbf{y} = \sqrt{\rho}\mathbf{W}_{\mathrm{BB}}^H\mathbf{W}_{\mathrm{RF}}^H\mathbf{H}\mathbf{x} + \mathbf{W}_{\mathrm{BB}}^H\mathbf{W}_{\mathrm{RF}}^H\mathbf{n}, \tag{2}$$

where ρ is the average received power, $\mathbf{W}_{\mathrm{BB}}$ is the $N_{\mathrm{RF}}^r \times N_s$ digital combiner in the baseband and $\mathbf{W}_{\mathrm{RF}}$ is the $N_r \times N_{\mathrm{RF}}^r$ analog combiner in the RF domain, $\mathbf{H}$ is the $N_r \times N_t$ channel matrix, and $\mathbf{n}$ is the $N_r \times 1$ Gaussian noise vector. The analog combiner is also implemented by using $N_r \times N_{\mathrm{RF}}^r$ low-resolution phase shifters connecting each RF chain and each antenna, i.e., $\mathbf{W}_{\mathrm{RF}} \in \mathcal{P}^{N_r \times N_{\mathrm{RF}}^r}$.

Thus, the achievable spectral efficiency is given by [6]

$$\begin{aligned} R = \log\det\Big(\mathbf{I}_{N_s} + \frac{\rho}{\sigma_n^2 N_s}\left(\mathbf{W}_{\mathrm{RF}}\mathbf{W}_{\mathrm{BB}}\right)^{\dagger}\mathbf{H}\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}} \\ \times \mathbf{F}_{\mathrm{BB}}^H\mathbf{F}_{\mathrm{RF}}^H\mathbf{H}^H\mathbf{W}_{\mathrm{RF}}\mathbf{W}_{\mathrm{BB}}\Big). \end{aligned} \tag{3}$$

Hybrid beamforming aims at maximizing the achievable spectral efficiency R by designing the above precoders and combiners, which is formulated as

$$\begin{aligned} \max_{\mathbf{F}_{\mathrm{RF}},\mathbf{F}_{\mathrm{BB}},\mathbf{W}_{\mathrm{RF}},\mathbf{W}_{\mathrm{BB}}} \quad & R, \\ \text{s.t.} \quad & \|\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\|_F^2 = N_s, \\ & \mathbf{F}_{\mathrm{RF}} \in \mathcal{P}^{N_t \times N_{\mathrm{RF}}^t}, \\ & \mathbf{W}_{\mathrm{RF}} \in \mathcal{P}^{N_r \times N_{\mathrm{RF}}^r}. \end{aligned} \tag{4}$$

Problem (4) is a non-convex mixed integer programming problem, which is generally intractable for practical antenna scales. In the literature, this problem is usually divided into two subproblems, i.e., the precoder design and the combiner design [6]. The objective of each subproblem is modified to minimize the Euclidean distance between the hybrid precoder/combiner and the fully digital precoder/combiner.

We denote the precoder and combiner of the fully digital solution as $\mathbf{F}_{\mathrm{opt}}$ and $\mathbf{W}_{\mathrm{opt}}$, respectively. Thus, for any channel $\mathbf{H}$, we have $\mathbf{F}_{\mathrm{opt}} = \mathbf{V}(:, 1 : N_s)$ and $\mathbf{W}_{\mathrm{opt}} = \mathbf{U}(:, 1 : N_s)$, where $\mathbf{V}$ and $\mathbf{U}$ are the unitary matrixes derived from the singular value decomposition of $\mathbf{H}$, i.e., $\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$. Therefore, the precoder and combiner design can be formulated as

$$\begin{aligned} \min_{\mathbf{F}_{\mathrm{RF}},\mathbf{F}_{\mathrm{BB}}} \quad & \|\mathbf{F}_{\mathrm{opt}} - \mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\|_F^2, \\ \text{s.t.} \quad & \|\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\|_F^2 = N_s, \\ & \mathbf{F}_{\mathrm{RF}} \in \mathcal{P}^{N_t \times N_{\mathrm{RF}}^t}, \end{aligned} \tag{5}$$

and

$$\begin{aligned} \min_{\mathbf{W}_{\mathrm{RF}},\mathbf{W}_{\mathrm{BB}}} \quad & \|\mathbf{W}_{\mathrm{opt}} - \mathbf{W}_{\mathrm{RF}}\mathbf{W}_{\mathrm{BB}}\|_F^2, \\ \text{s.t.} \quad & \mathbf{W}_{\mathrm{RF}} \in \mathcal{P}^{N_r \times N_{\mathrm{RF}}^r}, \end{aligned} \tag{6}$$

Due to the similarity between problems (5) and (6), we only discuss the precoder design given by (5), and the proposed algorithm can also be applied to the combiner design given by (6).

III. Overall Hybrid Beamforming Design

We adopt a widely used alternating minimization framework to solve problem (5), in which $\mathbf{F}_{\mathrm{RF}}$ and $\mathbf{F}_{\mathrm{BB}}$ are optimized alternately when the other is fixed. Thus, the overall precoder design can be decoupled into the digital precoder design and the analog precoder design. Specifically, the digital precoder design can be formulated as

$$\begin{aligned} \min_{\mathbf{F}_{\mathrm{BB}}} \quad & \|\mathbf{F}_{\mathrm{opt}} - \mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\|_F^2, \\ \text{s.t.} \quad & \|\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\|_F^2 = N_s. \end{aligned} \tag{7}$$

and the analog precoder design can be formulated as

$$\begin{aligned} \min_{\mathbf{F}_{\mathrm{RF}}} \quad & \|\mathbf{F}_{\mathrm{opt}} - \mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\|_F^2, \\ \text{s.t.} \quad & \mathbf{F}_{\mathrm{RF}} \in \mathcal{P}^{N_t \times N_{\mathrm{RF}}^t}, \\ & \|\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\|_F^2 = N_s. \end{aligned} \tag{8}$$

As proved in [9] as long as we can make the Euclidean distance between the optimal digital precoder and the hybrid precoders (i.e., $\left\|\mathbf{F}_{\text{opt}}-\mathbf{F}_{\text{RF}}\widehat{\mathbf{F}}_{\text{BB}}\right\|_F$) sufficiently small when ignoring the power constraint, the normalization step will also achieve a near-optimal digital precoder. Thus, we temporarily remove the power constraint. Problem (7) can be reformulated as

$$\min_{\widehat{\mathbf{F}}_{\text{BB}}} \quad \left\|\mathbf{F}_{\text{opt}}-\mathbf{F}_{\text{RF}}\widehat{\mathbf{F}}_{\text{BB}}\right\|_F^2, \tag{9}$$

and problem (8) can be reformulated as

$$\begin{aligned}\min_{\mathbf{F}_{\text{RF}}} \quad & \left\|\mathbf{F}_{\text{opt}}-\mathbf{F}_{\text{RF}}\widehat{\mathbf{F}}_{\text{BB}}\right\|_F^2, \\ \text{s.t.} \quad & \mathbf{F}_{\text{RF}}\in\mathcal{P}^{N_t\times N_{\text{RF}}^t}.\end{aligned} \tag{10}$$

After the alternating minimization converges, the obtained $\widehat{\mathbf{F}}_{\text{BB}}$ should be further normalized by a factor of $\sqrt{N_s}/\left\|\mathbf{F}_{\text{RF}}\widehat{\mathbf{F}}_{\text{BB}}\right\|_F$ to satisfy the power constraint, and $\mathbf{F}_{\text{BB}}$ can be given by

$$\mathbf{F}_{\text{BB}}=\frac{\sqrt{N_s}}{\left\|\mathbf{F}_{\text{RF}}\widehat{\mathbf{F}}_{\text{BB}}\right\|_F}\widehat{\mathbf{F}}_{\text{BB}}. \tag{11}$$

Problem (9) is a least squares problem, to which the optimal solution can be given by

$$\widehat{\mathbf{F}}_{\text{BB}}=\mathbf{F}_{\text{RF}}^{\dagger}\mathbf{F}_{\text{opt}}, \tag{12}$$

where $\mathbf{F}_{\text{RF}}^{\dagger}$ is the Moore-Penrose pseudo inverse of $\mathbf{F}_{\text{RF}}$.

Problem (10) is an integer programming problem and the objective function can be rewritten as

$$\begin{aligned}&\left\|\mathbf{F}_{\text{opt}}-\mathbf{F}_{\text{RF}}\widehat{\mathbf{F}}_{\text{BB}}\right\|_F^2 \\ =&\sum_{i=1}^{N_t}\left\|\mathbf{f}_{\text{opt},i}-\mathbf{f}_{\text{RF},i}\widehat{\mathbf{F}}_{\text{BB}}\right\|_2^2 \\ =&\sum_{i=1}^{N_t}\left\|\left(\mathbf{f}_{\text{opt},i}\right)^T-\left(\widehat{\mathbf{F}}_{\text{BB}}\right)^T\left(\mathbf{f}_{\text{RF},i}\right)^T\right\|_2^2,\end{aligned} \tag{13}$$

where $\|\cdot\|_2^2$ denotes the 2-norm of a vector, $\mathbf{f}_{\text{opt},i}$ denotes the i-th row of $\mathbf{F}_{\text{opt}}$, and $\mathbf{f}_{\text{RF},i}$ denotes the i-th row of $\mathbf{F}_{\text{RF}}$. Thus, the original objective function is the sum of N_t independent functions corresponding to each row of $\mathbf{F}_{\text{RF}}$. In addition, the constraint can also be divided into N_t independent constraints corresponding to each row of $\mathbf{F}_{\text{RF}}$, given by

$$\mathbf{f}_{\text{RF},i}\in\mathcal{P}^{N_{\text{RF}}^t}, i=1,2,...,N_t. \tag{14}$$

Therefore, the analog precoder design given by (10) can be divided into N_t independent subproblems corresponding to each row of $\mathbf{F}_{\text{RF}}$. The subproblem of the i-th row is given by

$$\begin{aligned}\min_{\mathbf{f}_{\text{RF},i}} \quad & \left\|\left(\mathbf{f}_{\text{opt},i}\right)^T-\left(\widehat{\mathbf{F}}_{\text{BB}}\right)^T\left(\mathbf{f}_{\text{RF},i}\right)^T\right\|_2^2, \\ \text{s.t.} \quad & \mathbf{f}_{\text{RF},i}\in\mathcal{P}^{N_{\text{RF}}^t}.\end{aligned} \tag{15}$$

These subproblems have the same mathematical formulations. To simplify mathematical symbols, we denote $\mathbf{a}=(\mathbf{f}_{\text{opt},i})^T$, $\mathbf{B}=(\widehat{\mathbf{F}}_{\text{BB}})^T$ and $\mathbf{q}=(\mathbf{f}_{\text{RF},i})^T$. Thus, these subproblems can be uniformly represented by

$$\begin{aligned}\min_{\mathbf{q}} \quad & \|\mathbf{a}-\mathbf{B}\mathbf{q}\|_2^2, \\ \text{s.t.} \quad & \mathbf{q}\in\mathcal{P}^{N_{\text{RF}}^t}.\end{aligned} \tag{16}$$

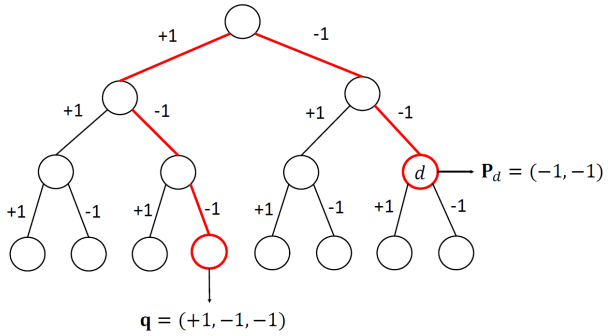

Fig. 2: Illustration of a search tree with $N_{\text{PS}}=2, N_{\text{RF}}^t=3$.

IV. Optimal Analog Precoder Design

As illustrated in Fig. 2, the solution space of problem (16) constitutes an N_{PS}-ary tree T with N_{RF}^t+1 levels, where each branch in level n represents the phase of the n-th phase shifter, and each leaf node represents a feasible $\mathbf{q}$. We denote the set of all nodes in level n as $\mathcal{D}_n$. For each node $d\in\mathcal{D}_n$, the corresponding phase sequence of the first n given phase shifters is denoted as $\mathbf{p}_d\in\mathcal{P}^n$.

Due to a large number of RF chains, it is impractical to calculate the optimal solution of problem (16) by exhaustive search. Thus, we leverage the classical branch and bound method to solve problem (16) [15]. Specifically, we perform a breadth-first search with N_{RF}^t iterations, during which it maintains three global variables, i.e., the available set $\mathcal{G}$ representing the set of nodes that need to be explored, the currently best solution $\mathbf{q}^*\in\mathcal{P}^{N_{\text{RF}}^t}$ and the smallest known upper bound UB^*. $\mathcal{G}$ is initialized as the set consisting of the root node d_0. $\mathbf{q}^*$ is initialized as $\mathbf{q}_0$ by randomly choosing the phases of all phase shifters. UB^* is initialized as $\text{UB}_0=\|\mathbf{a}-\mathbf{B}\mathbf{q}_0\|_2^2$.

In the n-th iteration, the proposed method searches in the n-th level. The available set $\mathcal{G}$ is updated by replacing each node with its children. For each node $d\in\mathcal{G}$, it represents a state with n given phase shifters and N_{RF}^t-n unknown phase shifters. We rewrite $\mathbf{q}$, $\mathbf{B}$ and $\mathbf{a}$ as

$$\mathbf{q}=\begin{bmatrix}\mathbf{q}_{\text{L}}\\ \mathbf{q}_{\text{R}}\end{bmatrix}, \tag{17}$$

$$\mathbf{B}=\begin{bmatrix}\mathbf{B}_{\text{L}} & \mathbf{B}_{\text{R}}\end{bmatrix}, \tag{18}$$

$$\mathbf{a}=\mathbf{a}_{\text{L}}+\mathbf{a}_{\text{R}}, \tag{19}$$

where subscript L represents the dimensions corresponding to the first n phase shifters, and subscript R represents the dimensions corresponding to the last N_{RF}^t-n phase shifters. Thus, the objective function of problem (16) can be divided into

$$f_{\text{L}}(\mathbf{q}_{\text{L}})=\|\mathbf{a}_{\text{L}}-\mathbf{B}_{\text{L}}\mathbf{q}_{\text{L}}\|_2^2, \tag{20}$$

Algorithm 1 Parallel Branch and Bound for Analog Precoder Design

for $i = 1 : N_t$ **do**
  Initialize $\mathcal{G} = \{d_0\}$, $\mathbf{q}^* = \mathbf{q}_0$ and $\mathrm{UB}^* = \mathrm{UB}_0$;
  for $n = 1 : N_{\mathrm{RF}}^t$ **do**
    Replace each node in $\mathcal{G}$ with its child nodes;
    for each node $d \in \mathcal{G}$ **do**
      Calculate $f_{\mathrm{L}}(\mathbf{p}_d)$, $f_{\mathrm{R,LB}}(d)$ and $f_{\mathrm{R,UB}}(d)$ as in (20), (27) and (28), respectively;
      Update UB^* and $\mathbf{q}^*$ as (29) and (30), respectively;
    end for
    for each node $d \in \mathcal{G}$ **do**
      Update $\mathcal{G}$ as in (31);
    end for
  end for
  $\mathbf{f}_{\mathrm{RF},i} = (\mathbf{q}^*)^T$.
end for
return $\mathbf{F}_{\mathrm{RF}}$.

and

$$f_{\mathrm{R}}(\mathbf{q}_{\mathrm{R}}) = \|\mathbf{a}_{\mathrm{R}} - \mathbf{B}_{\mathrm{R}}\mathbf{q}_{\mathrm{R}}\|_2^2 . \tag{21}$$

The constraint of problem (16) can be divided into

$$\mathbf{q}_{\mathrm{L}} = \mathbf{p}_d, \tag{22}$$

and

$$\mathbf{q}_{\mathrm{R}} \in \mathcal{P}^{N_{\mathrm{RF}}^t - n}. \tag{23}$$

Since $f_{\mathrm{L}}(\mathbf{q}_{\mathrm{L}})$ and $\mathbf{q}_{\mathrm{L}}$ are fixed, each node $d \in \mathcal{G}$ corresponds to a local problem given by

$$\begin{aligned} \min_{\mathbf{q}_{\mathrm{R}}} \quad & f_{\mathrm{R}}(\mathbf{q}_{\mathrm{R}}), \\ \text{s.t.} \quad & \mathbf{q}_{\mathrm{R}} \in \mathcal{P}^{N_{\mathrm{RF}}^t - n}. \end{aligned} \tag{24}$$

Problem (24) can be transformed into a least squares problem by relaxing the discrete constraint, to which the optimal solution can be given by

$$\widehat{\mathbf{q}}_{\mathrm{R}} = (\mathbf{B}_{\mathrm{R}})^{\dagger}\mathbf{a}_{\mathrm{R}}. \tag{25}$$

We can round $\widehat{\mathbf{q}}_{\mathrm{R}}$ to give a feasible solution to problem (24), given by

$$\widetilde{\mathbf{q}}_{\mathrm{R}} = \underset{\mathbf{q}_{\mathrm{R}} \in \mathcal{P}^{N_{\mathrm{RF}}^t - n}}{\arg\min} \|\mathbf{q}_{\mathrm{R}} - \widehat{\mathbf{q}}_{\mathrm{R}}\|_2 . \tag{26}$$

Thus, problem (24) has a lower bound

$$f_{\mathrm{R,LB}}(d) = f_{\mathrm{R}}(\widehat{\mathbf{q}}_{\mathrm{R}}), \tag{27}$$

and an upper bound

$$f_{\mathrm{R,UB}}(d) = f_{\mathrm{R}}(\widetilde{\mathbf{q}}_{\mathrm{R}}). \tag{28}$$

Each upper bound of problem (24) corresponds to an upper bound of problem (16). Thus, for each node $d \in \mathcal{G}$, the smallest known upper bound UB^* is updated as

$$\mathrm{UB}^* = \min\{\mathrm{UB}^*, f_{\mathrm{L}}(\mathbf{p}_d) + f_{\mathrm{R,UB}}(d)\}, \tag{29}$$

and $\mathbf{q}^*$ is updated as

$$\mathbf{q}^* = \begin{cases} \mathbf{q}^*, & \text{if } \mathrm{UB}^* \le f_{\mathrm{L}}(\mathbf{p}_d) + f_{\mathrm{R,UB}}(d), \\ \begin{bmatrix} \mathbf{p}_d \\ \widetilde{\mathbf{q}}_{\mathrm{R}} \end{bmatrix}, & \text{if } \mathrm{UB}^* > f_{\mathrm{L}}(\mathbf{p}_d) + f_{\mathrm{R,UB}}(d). \end{cases} \tag{30}$$

For each node d, we can use the updated bound UB^* to decide whether the corresponding subtree needs further exploration. Specifically, if the lower bound of node d exceeds the global upper bound, i.e., $f_{\mathrm{L}}(\mathbf{p}_d) + f_{\mathrm{R,LB}}(d) > \mathrm{UB}^*$, all leaves below node d are suboptimal, and the corresponding branches should be pruned. Thus, $\mathcal{G}$ is updated as

$$\mathcal{G} = \mathcal{G} \backslash \{d\}, \text{ if } f_{\mathrm{L}}(\mathbf{p}_d) + f_{\mathrm{R,LB}}(d) > \mathrm{UB}^*. \tag{31}$$

After N_{RF}^t iterations, the optimal solution is given by the final $\mathbf{q}^*$, and the minimum of the objective function value of problem (16) is given by the final upper bound UB^*. The overall algorithm for the analog precoder design is summarized in Algorithm 1, in which N_t subproblems given by problem (16) are solved in parallel.

In each iteration, the optimal solution to problem (9) can be produced, based on which problem (10) can be solved too. The existing methods for problem (10) only provide suboptimal solutions, while the proposed algorithm can solve problem (10) optimally. Thus, we refer to the solution as optimal analog precoder. As both the digital and analog precoders can achieve the optimal solutions, the alternative minimization process can converge to a feasible solution [9]. Although the optimality of alternating minimization is still an open problem, it is still intuitive and works well in a surprising number of cases [16]. And we will show that the proposed algorithm outperforms other conventional algorithms in practical antenna settings.

TABLE I: Computational Complexity of the Algorithms

Algorithm	Computational Complexity
PBB	$\mathcal{O}\left(N_{iter}^o \left(N_t N_s N_{\mathrm{RF}} (N_{\mathrm{PS}})^{N_{\mathrm{RF}}}\right)\right)$
MO	$\mathcal{O}\left(N_{iter}^o \left(N_{\mathrm{t}} N_{\mathrm{RF}}^2\right) + N_{iter}^i \left(N_{\mathrm{t}}^2 N_{\mathrm{RF}} N_{\mathrm{S}}\right)\right)$
CDM	$\mathcal{O}\left(N_{iter}^o \left(N_{\mathrm{t}}^3 N_{\mathrm{RF}}^2 N_{\mathrm{S}}\right)\right)$
OMP	$\mathcal{O}\left(N_{\mathrm{t}} N_{\mathrm{RF}}^3\right)$
IPM	$\mathcal{O}\left(N_{iter}^o \left(N_{\mathrm{t}} N_{\mathrm{RF}}^3\right) + N_{iter}^i \left(N_{\mathrm{t}}^2 N_{\mathrm{r}} N_{\mathrm{RF}}\right)\right)$

In Table I, we show the computational complexity of the proposed parallel branch and bound (PBB) algorithm, the orthogonal matching pursuit (OMP) algorithm [6], the iterative phase matching (IPM) algorithm [8], the manifold optimization (MO) algorithm [9] and the coordinate descent method (CDM) algorithm [11]. N_{iter}^o denotes the number of outer iterations and N_{iter}^i denotes the accumulation of the number of inner iterations at each outer iteration. We assume the number of RF chains is the same for both the transmitter and the receiver, i.e., $N_{\mathrm{RF}}^t = N_{\mathrm{RF}}^r = N_{\mathrm{RF}}$. For the IPM algorithm, Table I lists the computational complexity of hybrid precoder and combiner designs as the precoder and combiner are designed in pairs. For other algorithms, Table I only lists the computational complexity of hybrid precoder design and that of hybrid combiner design has similar expressions.

This article has been accepted for publication in IEEE Transactions on Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVT.2023.3245091

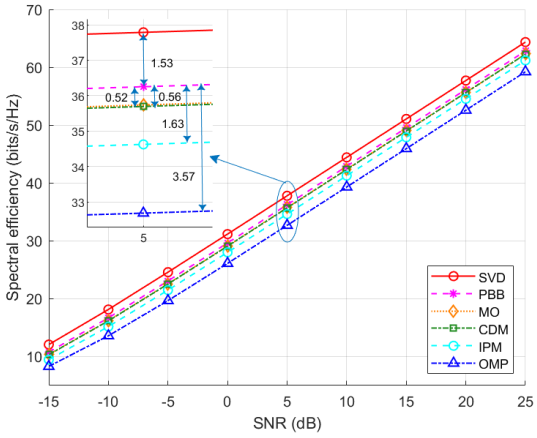

Fig. 3: Spectral efficiency against SNR with $N_t = 144$, $N_r = 36$ and $N_{\mathrm{RF}} = 6$.

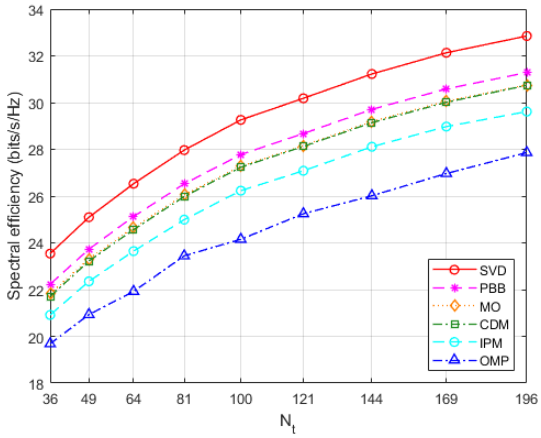

Fig. 5: Spectral efficiency against N_t with $N_r = 36$, $N_{\mathrm{RF}} = 6$ and SNR $= 0$ dB.

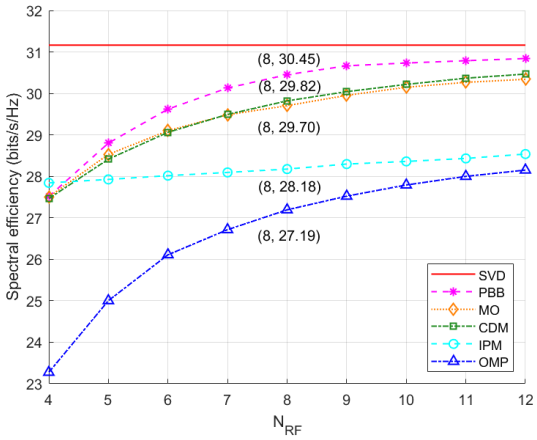

Fig. 4: Spectral efficiency against N_{RF} with $N_t = 144$, $N_r = 36$ and SNR $= 0$ dB.

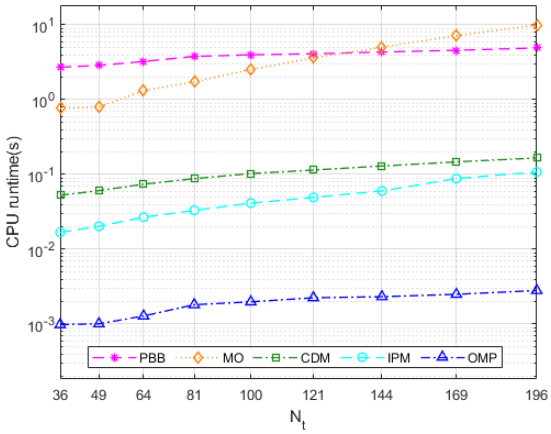

Fig. 6: CPU runtime against N_t with $N_r = 36$ and $N_{\mathrm{RF}} = 6$.

As shown in Table I, the OMP algorithm achieves the lowest computational complexity. The computational complexity of the PBB algorithm is proportional to N_t. The computational complexity of the MO and the IPM algorithms is proportional to N_t^2. The computational complexity of the CDM algorithm is proportional to N_t^3. As N_t is much larger than N_{RF} in practical implementations, we can expect that the complexity of the PBB method will be much lower than other algorithms, except for the OMP algorithm, when N_t is extremely large, e.g., in ultra-massive MIMO systems. In addition, the elapsed time of the PBB algorithm can be further reduced as the PBB can be executed in parallel.

V. Numerical Results

We consider a hybrid beamforming system with $N_s = 4$ data streams. The number of RF chains is the same for both the transmitter and the receiver, i.e., $N_{\mathrm{RF}}^t = N_{\mathrm{RF}}^r = N_{\mathrm{RF}}$ [6]. We assume $N_{\mathrm{PS}} = 4$ and the set of feasible phases for the analog precoder is given by $\mathcal{P} = \left\{1, e^{j\pi/2}, e^{j\pi}, e^{j3\pi/2}\right\}$. Channel $\mathbf{H}$ is given by the Saleh−Valenzuela mmWave channel model with a sum of 5 scattering clusters, each of which contributes 10 propagation paths [17].

The proposed PBB algorithm is compared with the OMP algorithm [6], the IPM algorithm [8], the MO algorithm [9] and the CDM algorithm [11]. All the algorithms assume the knowledge of the optimal beamforming matrix. Each algorithm is applied to both the transmitter and the receiver for hybrid precoding and combining, respectively. Specifically, the continuous solutions given by the OMP and the MO algorithms are quantified according to the available phases given by $\mathcal{P}$. The fully digital solution achieved by singular value decomposition (SVD) is taken as the upper bound of the achievable spectrum efficiency. The numerical results are averaged over 1000 channel realizations.

Fig. 3 shows the spectral efficiency as a function of signal-to-noise ratio (SNR) with $N_t = 144$, $N_r = 36$ and $N_{\mathrm{RF}} = 6$. Due to the optimal analog precoder design, the proposed hybrid precoder outperforms all other hybrid precoders in terms of spectral efficiency by 0.52-3.57 bits/s/Hz. Whereas, due to the integer constraints on analog phase shifters and the suboptimality of alternative minimization, the proposed PBB algorithm can only achieve 96% performance of the fully

digital solution in terms of spectral efficiency.

Fig. 4 shows the spectral efficiency as a function of the number of RF chains N_{RF} with $N_t = 144$, $N_r = 36$ and $\mathrm{SNR} = 0$ dB. As the number of RF chains increases, the hybrid precoder will approach the fully digital precoder and the spectral efficiency increases for all considered algorithms. The proposed PBB algorithm achieves the highest spectral efficiency in all settings except for $N_{\mathrm{RF}} = N_s$, for which the IPM algorithm can utilize the scaled unitary nature of the digital precoder when $N_{\mathrm{RF}} = N_s$ and outperforms the proposed algorithm. As N_{RF} increases, the scaled unitary nature is not valid and the performance of the IPM algorithm is inferior to other algorithms. Thus, the cross points between 4 and 5 imply thresholds of N_{RF} , above which the IPM algorithm is inferior to other algorithms in terms of spectral efficiency. However, as N_{RF} must be an integer, there are no practical meanings of these thresholds. The proposed PBB algorithm achieves 98% performance of the fully digital solution in a typical antenna setting with $N_t = 144$, $N_r = 36$ and $N_{\mathrm{RF}} = 8$.

Fig. 5 shows the spectral efficiency as a function of the number of transmit antennas N_t with $N_r = 36$, $N_{\mathrm{RF}} = 6$ and $\mathrm{SNR} = 0$ dB. As the number of transmit antenna increases, the spatial degrees-of-freedom increases which improves the spectral efficiency. Due to the optimal analog precoder design, the proposed hybrid precoder outperforms all other hybrid precoders in terms of spectral efficiency by 0.41-3.62 bits/s/Hz and achieves 94%-96% performance of the fully digital solution.

Fig. 6 shows the CPU runtime as a function of the number of transmit antennas N_t with $N_r = 36$ and $N_{\mathrm{RF}} = 6$. The OMP and the IPM algorithms which do not leverage alternating minimization achieve short elapsed time. The MO, the CDM and the proposed PBB algorithms leverage alternating minimization to improve the spectral efficiency but lead to longer elapsed time. As the computational complexity of the MO algorithm is proportional to N_t^2 while the complexity of our proposed algorithm is proportional to N_t, our proposed algorithm achieves shorter CPU runtime than the MO algorithm when N_t is large. Thus, the cross point between 121 and 144 implies a threshold of N_t, above which the proposed algorithm outperforms the MO algorithm in terms of CPU runtime. However, as we assume the uniform square planar array with $\sqrt{N_t} \times \sqrt{N_t}$ antenna elements, the threshold has no practical meaning.

VI. Conclusion

In this paper, we have considered the analog precoder design in hybrid beamforming with low-resolution phase shifters, for which we propose an optimal analog precoding scheme based on the branch and bound method. Specifically, we show that the analog precoder design can be divided into multiple low-dimensional subproblems in parallel and there exists an analytical lower bound for each subproblem that can be efficiently utilized to obtain the optimal solution. Numerical results show that the proposed algorithm outperforms the existing algorithms in terms of spectral efficiency and achieves 98% performance of the fully digital solution in a typical antenna setting.

Acknowledgement

The authors would like to thank the editors and the anonymous reviewers, whose invaluable comments helped improve the presentation of this paper substantially.

References

[1] W. Yu *et al.*, "Multi-label learning based antenna selection in massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 70, no. 7, pp. 7255–7260, Jul. 2021.

[2] Z. Kuai and S. Wang, "Thompson sampling-based antenna selection with partial CSI for TDD massive MIMO systems," *IEEE Trans. Commun.*, vol. 68, no. 12, pp. 7533–7546, Dec. 2020.

[3] Z. Tao and S. Wang, "Improved downlink rates for FDD massive MIMO systems through bayesian neural networks-based channel prediction," *IEEE Trans. Wireless Commun.*, vol. 21, no. 3, pp. 2122–2134, Mar. 2022.

[4] A. F. Molisch *et al.*, "Hybrid beamforming for massive MIMO: A survey," *IEEE Commun. Mag.*, vol. 55, no. 9, pp. 134–141, Sep. 2017.

[5] A. Khaled *et al.*, "Hybrid analog-digital beamforming design for SE and EE maximization in massive MIMO networks," *IEEE Trans. Veh. Technol.*, vol. 69, no. 1, pp. 377–389, Jan. 2020.

[6] O. E. Ayach, *et al.*, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.

[7] Y. Zhang *et al.*, "Machine learning-based hybrid precoding with low-resolution analog phase shifters," *IEEE Commun. Lett.*, vol. 25, no. 1, pp. 186–190, Sep. 2021.

[8] Z. Wang *et al.*, "Hybrid precoder and combiner design with low-resolution phase shifters in mmWave MIMO systems," *IEEE J. Sel. Top. Signal Process.*, vol. 12, no. 2, pp. 256–269, May 2018.

[9] X. Yu *et al.*, "Alternating minimization algorithms for hybrid precoding in millimeter wave MIMO systems," *IEEE J. Sel. Top. Signal Process.*, vol. 10, no. 3, pp. 485–500, Apr. 2016.

[10] M. Mulla *et al.*, "Barzilai-Borwein gradient algorithm based alternating minimization for single user millimeter wave systems," *IEEE Wireless Commun. Lett.*, vol. 9, no. 4, pp. 508–512, Dec. 2020.

[11] J.-C. Chen, "Hybrid beamforming with discrete phase shifters for millimeter-wave massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 66, no. 8, pp. 7604–7608, Aug. 2017.

[12] S. Lyu *et al.*, "Lattice-based mmWave hybrid beamforming," *IEEE Trans. Commun.*, vol. 69, no. 7, pp. 4907–4920, Jul. 2021.

[13] X. Cui and Q. Li, "Hybrid beamforming with finite-resolution phase shifters for multiuser millimeter-wave downlink," *IEEE Wireless Commun. Lett.*, vol. 9, no. 2, pp. 219–222, Feb. 2020.

[14] F. Dong *et al.*, "Low-complexity hybrid precoding for multi-user mmWave systems with low-resolution phase shifters," *IEEE Trans. Veh. Technol.*, vol. 68, no. 10, pp. 9774–9784, Aug. 2019.

[15] Morrison *et al.*, "Branch-and-bound algorithms: A survey of recent advances in searching, branching, and pruning," *Discr. Optim.*, vol. 19, pp. 79–102, Feb. 2016.

[16] J. C. Bezdek and R. J. Hathaway, "Some notes on alternating optimization," in *Proc. AFSS*, 2002, pp. 288–300.

[17] H. Xu *et al.*, "Spatial and temporal characteristics of 60-GHz indoor channels," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 3, pp. 620–630, Aug. 2002.