

MmWave MIMO Hybrid Precoding Design Using Phase Shifters and Switches

Qiang Liu*, Chenhao Qi*, Xianghao Yu[†] and Geoffrey Ye Li[‡]

*School of Information Science and Engineering, Southeast University, Nanjing, China

[†]Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, China

[‡]Department of Electrical and Electronic Engineering, Imperial College London, UK

Email: {220190764,qch}@seu.edu.cn, eexyu@ust.hk, geoffrey.li@imperial.ac.uk

Abstract—To reduce the number of phase shifters for analog precoding in millimeter wave massive multiple-input multiple-output communications, we investigate the hybrid use of expensive phase shifters and low-cost switches. Different from the existing fixed phase shifter (FPS) architecture where the phases are fixed and independent of the channel state information, we consider variable phase shifter (VPS) whose phases are variable and subject to the hardware constraint. Based on the VPS architecture, a hybrid precoding design (HPD) scheme named VPS-HPD is proposed to optimize the phases according to the channel state information. Specifically, we alternately optimize the analog precoder and the digital precoder, where the former is converted into several subproblems and each subproblem further includes the alternating optimization of the phase matrix and switch matrix. Simulation results show that the spectral efficiency of the VPS-HPD scheme is very close to that of the fully digital precoding, higher than that of the existing MO-AltMin scheme for the fully-connected architecture with much fewer phase shifters, and substantially higher than that of the existing FPS-AltMin scheme for the FPS architecture with the same number of phase shifters.

Index Terms—Alternating minimization, hybrid precoding, massive MIMO, millimeter wave (mmWave) communications

I. INTRODUCTION

Millimeter wave (mmWave) wireless communication has drawn extensive attention, as it can provide ultra-high-speed data rate [1], [2]. On the other side, uplifting the carrier frequency to such a high frequency band will inevitably bring several challenges such as significant path loss and unfavorable atmospheric absorption [3], making it difficult for the mmWave signal to propagate in a long distance. To deal with these drawbacks, massive multiple-input multiple-output (MIMO) is introduced for mmWave communications to form analog beams with high directivity and large beam gain.

The hardware cost of fully digital precoding becomes a prominent issue when increasing the working frequency from sub-6 GHz to mmWave band. To approach the achievable rate performance of the fully digital precoding meanwhile reducing the hardware cost, hybrid precoding and combining are proposed, where we can use much fewer radio frequency (RF) chains than the antennas [4]–[6]. Depending on whether each RF chain connecting to all the antennas or only part of the antennas, the hybrid precoding can be further detail

into fully-connected architecture [7] or partially-connected architecture [8], respectively, where the latter uses fewer phase shifters than the former but with some performance degradation. Nevertheless, either in the fully-connected or partially-connected architectures, the phase shifters play an important role. To reduce the hardware complexity and power consumption of phase shifters, the switches that only have on-off binary states are adopted for hybrid precoding and combining [9]. In [10], all the phase shifters are replaced by the low-cost switches, which can substantially reduce the hardware complexity. In [11], hybrid use of low-resolution phase shifters and switches is considered, where the switches are used for antenna selection to achieve further performance improvement. From an energy efficiency (EE) perspective [12], the closed-form expressions are provided to compare several promising hybrid beamforming architectures and the optimal numbers of antennas maximizing the EE are obtained, where significantly higher EE can be achieved by the combination of phase shifters and switches than the conventional phase shifter-only architectures. In [13], an architecture using switches to select antenna subsets and using constant phase shifters to control the phases of signals in the RF circuit is proposed, where three low-complexity algorithms are developed for per-RF chain antenna subset selection. In [14], a fixed phase shifter (FPS) architecture with the hybrid use of phase shifter with switch network is proposed, where a hybrid precoding algorithm based on alternating minimization (AltMin) is developed.

In this paper, different from the FPS architecture where the phases are fixed and independent of the channel state information (CSI), we consider variable phase shifter (VPS) whose phases are variable and subject to the hardware constraint. Based on the VPS architecture, a hybrid precoding design (HPD) scheme named VPS-HPD is proposed to optimize the phases according to the CSI. Specifically, we alternately optimize the analog precoder and the digital precoder, where the former is converted into several subproblems and each subproblem further includes the alternating optimization of the phase matrix and switch matrix. When determining the phase matrix, the hardware constraint on the phase shifters is temporarily relaxed and then the discrete optimization problem on the phases can be converted as a continuous one,

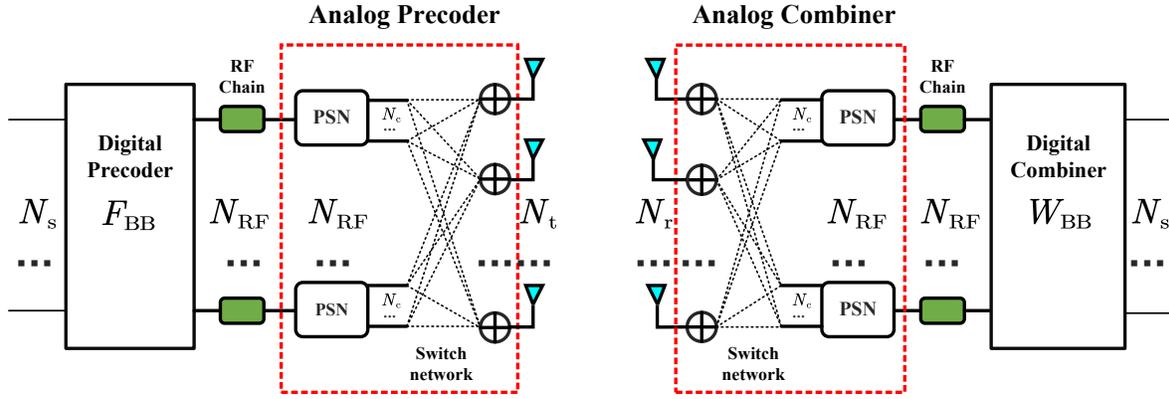


Fig. 1. A point-to-point mmWave massive MIMO system with the FPS/VPS architecture.

so that the Riemannian manifold optimization followed by the phase quantization can be used to solve it.

Notations: Symbols for vectors (lower case) and matrices (upper case) are in boldface. For a vector \mathbf{a} , $[\mathbf{a}]_m$ denotes its m th entry. For a matrix \mathbf{A} , $[\mathbf{A}]_{m,:}$, $[\mathbf{A}]_{:,n}$, $[\mathbf{A}]_{m,n}$, \mathbf{A}^T , \mathbf{A}^{-1} , \mathbf{A}^H and $\|\mathbf{A}\|_F^2$ denote the m th row, the n th column, the entry on the m th row and n th column, the transpose, the inverse, the conjugate transpose (Hermitian), and Frobenius norm, respectively. \mathbf{I}_L denotes an $L \times L$ identity matrix. The symbols $\angle(\cdot)$, $\mathbb{E}(\cdot)$, $\mathcal{O}(\cdot)$ and $\mathcal{CN}(m, \mathbf{R})$ denote the angle of a complex-valued number, the expectation, the order of complexity, and the complex Gaussian distribution with the mean of m and the covariance matrix being \mathbf{R} , respectively. The symbols \mathbb{C} and \mathbb{Z} denote the set of complex-valued numbers and the set of integers, respectively.

II. PROBLEM FORMULATION

Consider a point-to-point mmWave massive MIMO system with N_t and N_r antennas at the transmitter and receiver, respectively. The transmitter and receiver employ the hybrid precoding and combining with N_{RF} RF chains, where $N_{\text{RF}} \ll N_t$ and $N_{\text{RF}} \ll N_r$. We assume that N_s independent data streams are transmitted in parallel, where $N_s \leq N_{\text{RF}}$. The hybrid precoder includes a digital precoder at baseband (BB) and an analog precoder in the RF domain, denoted as $\mathbf{F}_{\text{BB}} \in \mathbb{C}^{N_{\text{RF}} \times N_s}$ and $\mathbf{F}_{\text{RF}} \in \mathbb{C}^{N_t \times N_{\text{RF}}}$, respectively. We normalize the power gain of the hybrid precoder by setting $\|\mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}\|_F^2 = N_s$. Similarly, the hybrid combiner includes a digital combiner and an analog combiner, denoted as $\mathbf{W}_{\text{BB}} \in \mathbb{C}^{N_{\text{RF}} \times N_s}$ and $\mathbf{W}_{\text{RF}} \in \mathbb{C}^{N_r \times N_{\text{RF}}}$, respectively. Then the received signal after hybrid combining can be expressed as

$$\mathbf{y} = \sqrt{P}\mathbf{W}_{\text{BB}}^H\mathbf{W}_{\text{RF}}^H\mathbf{H}\mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}\mathbf{s} + \mathbf{W}_{\text{BB}}^H\mathbf{W}_{\text{RF}}^H\boldsymbol{\eta}, \quad (1)$$

where $\mathbf{s} \in \mathbb{C}^{N_s}$ is the transmit signal subject to the constraint of maximum transmit power P , i.e., $\mathbb{E}(\mathbf{s}\mathbf{s}^H) = \frac{P}{N_s}\mathbf{I}_{N_s}$; $\boldsymbol{\eta} \sim \mathcal{CN}(0, \sigma^2\mathbf{I}_{N_r})$ is a noise term which obeys the complex Gaussian distribution with zero mean and variance of σ^2 .

Using the widely-used Saleh-Valenzuela model, the channel matrix \mathbf{H} in (1) is defined as

$$\mathbf{H} = \sqrt{\frac{N_t N_r}{L}} \sum_{l=1}^L \alpha_l \mathbf{a}(N_r, \phi_l) \mathbf{a}^H(N_t, \theta_l), \quad (2)$$

where L , α_l , $\theta_l \in (0, 2\pi)$ and $\phi_l \in (0, 2\pi)$ are the number of channel paths, complex channel coefficient, physical angle-of-departure (AoD) and physical angle-of-arrival (AoA) of the l th path for $l = 1, 2, \dots, L$, respectively. The channel steering vector, as a function of antenna number N and AoD or AoA φ , is defined as

$$\mathbf{a}(N, \varphi) \triangleq \frac{1}{\sqrt{N}} [1, e^{j\pi \sin \varphi}, e^{j2\pi \sin \varphi}, \dots, e^{j(N-1)\pi \sin \varphi}]^T \quad (3)$$

for uniform linear arrays (ULAs).

For the fully-connected architecture where each RF chain connects to all the antennas via phase shifters, we need totally $N_t N_{\text{RF}}$ phase shifters. To reduce the hardware complexity, the partially-connected architecture decreases the number of phase shifters to N_t , where each RF chain only connects to a subset of the antennas. Since fewer phase shifters are used, there is some performance degradation comparing to the fully-connected one.

Note that the switches with on-off binary states are much cheaper and faster than the phase shifters [9]. Therefore, the phase shifters in the fully-connected architecture can be replaced by low-cost switches with some sacrifice of achievable rate performance [10]. To balance the hardware complexity and achievable rate performance, an architecture named FPS with hybrid use of phase shifters and switches is proposed [14]. The block diagram of the FPS architecture is shown in Fig. 1, where the analog precoder or combiner is formed by N_{RF} phase shifter networks (PSNs), a switch network and N_t signal adders. Each PSN includes N_c phase shifters and therefore the FPS architecture includes $N_c N_{\text{RF}}$ phase shifters in total. In Table I, we list different numbers of phase shifters required for the fully-connected, partially-connected and FPS architectures. Since we normally set $N_c \ll N_t$, we may use much fewer phase shifters in the FPS

TABLE I
OVERHEAD COMPARISON FOR DIFFERENT MMWAVE MIMO
ARCHITECTURES.

Different architectures	Number of phase shifters	Number of switches
Fully-connected	$N_t N_{\text{RF}}$	0
Partially-connected	N_t	0
FPS / VPS	$N_c N_{\text{RF}}$	$N_c N_{\text{RF}} N_t$

architecture than that in the fully-connected architecture. The switch network of the FPS architecture includes $N_c N_{\text{RF}} N_t$ switches, where each switch determines the on-off state of the link between a PSN output and a signal adder. Each signal adder adds at most $N_c N_{\text{RF}}$ different signals together.

For the FPS architecture, the phase of the i th phase shifter in each PSN is $2\pi(i-1)/N_c$ for $i = 1, 2, \dots, N_c$. Note that the phases are determined by N_c . Moreover, the phases are fixed and independent of mmWave CSI, which does not fully exploit the phase-space freedom of the phase shifters.

Different with the FPS architecture, in this work we will consider a VPS architecture, where the phases are optimized according to the mmWave CSI and subject to the hardware constraints. Note that the numbers of phase shifters and switches are the same in both the VPS and FPS architectures, as shown in Table I. However, the VPS uses variable phase shifters while the FPS uses fixed ones. Due to the difference of these two architectures, their HPD schemes are completely different.

According to Fig. 1, \mathbf{F}_{RF} can be expressed as

$$\mathbf{F}_{\text{RF}} = \mathbf{S}_t \mathbf{P}_t, \quad (4)$$

where $\mathbf{P}_t \in \mathbb{C}^{N_c N_{\text{RF}} \times N_{\text{RF}}}$ is a phase matrix determined by N_{RF} PSNs and $\mathbf{S}_t \in \mathbb{Z}^{N_t \times N_c N_{\text{RF}}}$ is a switch matrix determined by the switch network. Each entry of \mathbf{S}_t is selected from a binary set

$$\mathcal{A} \triangleq \{0, 1\} \quad (5)$$

where 1 and 0 indicate the on and off states of a switch, respectively. In fact, \mathbf{P}_t is a generalized block diagonal matrix, where each column of \mathbf{P}_t has only N_c nonzero entries representing the values of N_c phase shifters. The entry on the i th column and $((i-1)N_c + l)$ th row of \mathbf{P}_t can be denoted as

$$[\mathbf{P}_t]_{(i-1)N_c+l,i} = \frac{1}{\sqrt{N_c}} e^{j\theta_{li}}, \quad (6)$$

$$i = 1, 2, \dots, N_{\text{RF}}, \quad l = 1, 2, \dots, N_c,$$

where θ_{li} is the phase of the l th phase shifter in the i th PSN. Suppose the resolution of the phase shifters is b bits. Then all available phases form a set

$$\mathcal{B} \triangleq \left\{ \frac{2\pi i}{2^b} \mid i = 1, 2, \dots, 2^b \right\} \quad (7)$$

and we have $\theta_{li} \in \mathcal{B}$.

Suppose we have already obtained a channel estimate of \mathbf{H} , denoted as $\widehat{\mathbf{H}}$, using existing channel estimation

methods. The singular value decomposition (SVD) of $\widehat{\mathbf{H}}$ can be expressed as

$$\widehat{\mathbf{H}} = \mathbf{V} \mathbf{\Lambda} \mathbf{U}^H \quad (8)$$

where \mathbf{U} is a unitary matrix with the dimension of N_t and \mathbf{V} is a unitary matrix with the dimension of N_r . The fully digital precoder \mathbf{F}_{opt} is determined by selecting the first N_s columns from \mathbf{U} . The fully digital combiner \mathbf{W}_{opt} is determined by selecting the first N_s columns from \mathbf{V} .

The design of \mathbf{F}_{RF} and \mathbf{F}_{BB} aims to approximate \mathbf{F}_{opt} by minimizing their Euclidean distance, which can be written as

$$\min_{\mathbf{S}_t, \mathbf{P}_t, \mathbf{F}_{\text{BB}}} \|\mathbf{F}_{\text{opt}} - \mathbf{S}_t \mathbf{P}_t \mathbf{F}_{\text{BB}}\|_F^2 \quad (9a)$$

$$\text{s.t.} \quad \angle([\mathbf{P}_t]_{i,l}) \in \mathcal{B}, \quad \forall i, l, \quad (9b)$$

$$[\mathbf{S}_t]_{m,n} \in \mathcal{A}, \quad \forall m, n, \quad (9c)$$

$$\|\mathbf{S}_t \mathbf{P}_t \mathbf{F}_{\text{BB}}\|_F^2 = N_s. \quad (9d)$$

Note that the constraint in (9d) can be temporarily neglected. Then (9) can be rewritten as

$$\min_{\mathbf{S}_t, \mathbf{P}_t, \mathbf{F}_{\text{BB}}} \|\mathbf{F}_{\text{opt}} - \mathbf{S}_t \mathbf{P}_t \mathbf{F}_{\text{BB}}\|_F^2 \quad (10a)$$

$$\text{s.t.} \quad \angle([\mathbf{P}_t]_{i,l}) \in \mathcal{B}, \quad \forall i, l, \quad (10b)$$

$$[\mathbf{S}_t]_{m,n} \in \mathcal{A}, \quad \forall m, n. \quad (10c)$$

Once the solutions of (10) is obtained, we can adjust \mathbf{F}_{BB} to satisfy (9d) [15]. Note that (10) is a non-convex NP-hard optimization problem. In the following, we will elaborate on how to solve it.

III. HYBRID PRECODER DESIGN

The optimization problem of (10) involves three matrices \mathbf{S}_t , \mathbf{P}_t and \mathbf{F}_{BB} , which are coupled and difficult to handle. Therefore, we consider using alternating optimization to iteratively solve it. Specifically, we alternately optimize the analog precoder and the digital precoder, where the analog precoder optimization is converted into several subproblems and each subproblem further includes the alternating optimization of the phase matrix and switch matrix. In the following, we first determine \mathbf{S}_t and \mathbf{P}_t for given \mathbf{F}_{BB} in the first subsection, then determine \mathbf{F}_{BB} for given \mathbf{S}_t and \mathbf{P}_t in the second subsection, and finally analyze the convergence and computational complexity in the last subsection.

A. Phase Matrix and Switch Matrix Optimization for Analog Precoder Design

Given \mathbf{F}_{BB} , an estimate of \mathbf{F}_{RF} , denoted as $\widehat{\mathbf{F}}_{\text{RF}}$, can be expressed as

$$\widehat{\mathbf{F}}_{\text{RF}} = \mathbf{F}_{\text{opt}} \mathbf{F}_{\text{BB}}^H (\mathbf{F}_{\text{BB}} \mathbf{F}_{\text{BB}}^H)^{-1}. \quad (11)$$

Then we can determine \mathbf{S}_t and \mathbf{P}_t by the following optimization problem as

$$\min_{\mathbf{S}_t, \mathbf{P}_t} \|\widehat{\mathbf{F}}_{\text{RF}} - \mathbf{S}_t \mathbf{P}_t\|_F^2$$

$$\text{s.t.} \quad \angle([\mathbf{P}_t]_{i,l}) \in \mathcal{B}, \quad \forall i, l,$$

$$[\mathbf{S}_t]_{m,n} \in \mathcal{A}, \quad \forall m, n. \quad (12)$$

We denote the nonzero entries of the i th column of \mathbf{P}_t as \mathbf{p}_i , i.e.,

$$[\mathbf{p}_i]_k = [\mathbf{P}_t]_{(i-1)N_c+k,i}, \quad k = 1, 2, \dots, N_c, \quad i = 1, 2, \dots, N_{\text{RF}}. \quad (13)$$

From (13), if \mathbf{p}_i for $i = 1, 2, \dots, N_{\text{RF}}$ is given, we can determine \mathbf{P}_t by replacing the i th diagonal entry of $\mathbf{I}_{N_{\text{RF}}}$ with \mathbf{p}_i . We denote the i th submatrix of \mathbf{S}_t , which is made up of the columns from $(i-1)N_c+1$ to iN_c of \mathbf{S}_t as \mathbf{Q}_i , i.e.,

$$[\mathbf{Q}_i]_{k,l} = [\mathbf{S}_t]_{k,(i-1)N_c+l}, \quad k = 1, 2, \dots, N_t, \quad l = 1, 2, \dots, N_c. \quad (14)$$

Based on (14), if \mathbf{Q}_i for $i = 1, 2, \dots, N_{\text{RF}}$ is given, we can determine \mathbf{S}_t by

$$\mathbf{S}_t = [\mathbf{Q}_1, \mathbf{Q}_2, \dots, \mathbf{Q}_{N_{\text{RF}}}] . \quad (15)$$

Then (12) can be converted into N_{RF} independent subproblems, where the i th subproblem for $i = 1, 2, \dots, N_{\text{RF}}$ is written as

$$\min_{\mathbf{Q}_i, \mathbf{p}_i} \left\| [\widehat{\mathbf{F}}_{\text{RF}}]_{:,i} - \mathbf{Q}_i \mathbf{p}_i \right\|_F^2 \quad (16a)$$

$$\text{s.t. } [\mathbf{Q}_i]_{m,n} \in \mathcal{A}, \quad \forall m, n, \quad (16b)$$

$$\angle([\mathbf{p}_i]_k) \in \mathcal{B}, \quad k = 1, 2, \dots, N_c. \quad (16c)$$

From (16), \mathbf{Q}_i and \mathbf{p}_i are coupled. Again we resort to alternating minimization to solve it.

1) *Determination of \mathbf{p}_i given \mathbf{Q}_i* : Given \mathbf{Q}_i , the determination of \mathbf{p}_i based on (16) can be rewritten as

$$\min_{\mathbf{p}_i} \left\| [\widehat{\mathbf{F}}_{\text{RF}}]_{:,i} - \mathbf{Q}_i \mathbf{p}_i \right\|_F^2 \quad (17)$$

$$\text{s.t. } \angle([\mathbf{p}_i]_k) \in \mathcal{B}, \quad k = 1, 2, \dots, N_c.$$

Each entries in \mathbf{p}_i is selected from a candidate set. For the exhaustive search, it needs to find a best combination from 2^{bN_c} ones. For example, if $b = 3$ and $N_c = 8$, it needs to exhaustively search from $2^{24} = 16777216$ combinations, implying that it is computationally inefficient. To reduce the computational complexity, we may temporarily relax the hardware constraint in (17) by converting the discrete entries of \mathbf{p}_i into continuous ones, which can be expressed as

$$\min_{\mathbf{p}_i} \left\| [\widehat{\mathbf{F}}_{\text{RF}}]_{:,i} - \mathbf{Q}_i \mathbf{p}_i \right\|_F^2 \quad (18)$$

$$\text{s.t. } |[\mathbf{p}_i]_k| = \frac{1}{\sqrt{N_c}}, \quad k = 1, 2, \dots, N_c.$$

Note that (18) is a typical Riemannian manifold optimization problem and can be solved by the existing toolbox. Suppose we obtain a solution $\tilde{\mathbf{p}}_i$ from (18). Then the phases of $\tilde{\mathbf{p}}_i$ is denoted as $\tilde{\theta}_{ki} \triangleq \angle([\tilde{\mathbf{p}}_i]_k)$ for $k = 1, 2, \dots, N_c$. We make the phase quantization of θ_{ki} by solving

$$\tilde{\theta}_{ki} = \arg \min_{\theta \in \mathcal{B}} \left| \tilde{\theta}_{ki} - \theta \right| \quad (19)$$

so that we can obtain a feasible solution for (17).

Algorithm 1 VPS-HPD Scheme

```

1: Input:  $\widehat{\mathbf{H}}$ 
2: Obtain  $\mathbf{F}_{\text{opt}}$  via (8).
3: Initialize  $\mathbf{F}_{\text{BB}}$  as a random full-column-rank matrix.
4: repeat
5:   Obtain  $\widehat{\mathbf{F}}_{\text{RF}}$  via (11).
6:   for  $i = 1 : N_{\text{RF}}$  do
7:     Initialize  $\mathbf{Q}_i$  as a random binary matrix.
8:     repeat
9:       Obtain  $\mathbf{p}_i$  via (18) and (19).
10:      Obtain  $\mathbf{Q}_i$  via (21).
11:     until stop condition (1) is satisfied
12:   end for
13:   Obtain  $\mathbf{P}_t$  and  $\mathbf{S}_t$  via (13) and (15), respectively.
14:   Obtain  $\mathbf{F}_{\text{BB}}$  by (22).
15: until stop condition (2) is satisfied
16: Normalize  $\mathbf{F}_{\text{BB}}$  by (23).
17: Output:  $\mathbf{F}_{\text{BB}}, \mathbf{S}_t, \mathbf{P}_t$ .

```

2) *Determination of \mathbf{Q}_i given \mathbf{p}_i* : For a given \mathbf{p}_i , (16) can be rewritten as

$$\min_{\mathbf{Q}_i} \left\| [\widehat{\mathbf{F}}_{\text{RF}}]_{:,i} - \mathbf{Q}_i \mathbf{p}_i \right\|_F^2 \quad (20)$$

$$\text{s.t. } [\mathbf{Q}_i]_{m,n} \in \mathcal{A}, \quad \forall m, n.$$

In fact, (20) can be converted into N_t independent subproblems, where the m th subproblem for $m = 1, 2, \dots, N_t$ can be expressed as

$$\min_{[\mathbf{Q}_i]_{m,:}} \left| [\widehat{\mathbf{F}}_{\text{RF}}]_{m,i} - [\mathbf{Q}_i]_{m,:} \mathbf{p}_i \right| \quad (21)$$

$$\text{s.t. } [\mathbf{Q}_i]_{m,n} \in \mathcal{A}, \quad n = 1, 2, \dots, N_c.$$

Each subproblem can be solved by the exhaustive search to find a best combination from 2^{N_c} ones. If $N_c = 8$, we needs to search from $2^8 = 256$ combinations, which is computationally tractable.

Given a random and binary initialization of \mathbf{Q}_i , we can iteratively run the procedures described by the equations from (17) to (21) until *stop condition (1)* is satisfied, where *stop condition (1)* can be set as equaling a predefined number of iterations. Then optimized \mathbf{Q}_i and \mathbf{p}_i can be obtained. Consequently, we can obtain \mathbf{S}_t and \mathbf{P}_t via (14) and (13), respectively.

B. Digital Precoder Design

Given \mathbf{S}_t and \mathbf{P}_t , which are equivalent as given \mathbf{F}_{RF} according to (4), we can compute \mathbf{F}_{BB} in (10) by the least square method as

$$\mathbf{F}_{\text{BB}} = (\mathbf{P}_t^H \mathbf{S}_t^H \mathbf{S}_t \mathbf{P}_t)^{-1} \mathbf{P}_t^H \mathbf{S}_t^H \mathbf{F}_{\text{opt}}. \quad (22)$$

Given an random initialization of \mathbf{F}_{BB} which is full column rank, we can determine \mathbf{S}_t and \mathbf{P}_t , based on which we can further obtain an optimized \mathbf{F}_{BB} . In this way, we can iteratively optimize \mathbf{S}_t , \mathbf{P}_t and \mathbf{F}_{BB} until *stop condition (2)*

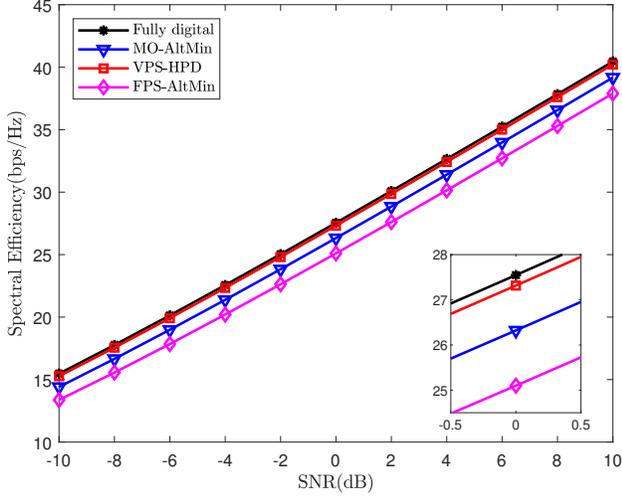


Fig. 2. SE comparisons of the VPS architecture with the other ones.

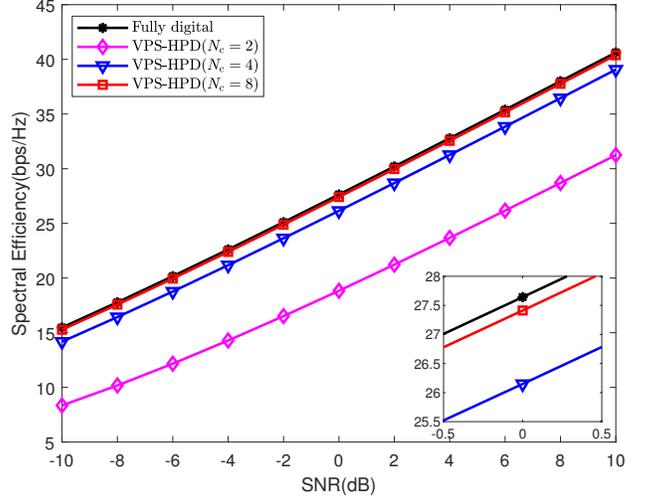


Fig. 3. SE comparisons of the VPS-HPD scheme with different N_c .

is satisfied where *stop condition* (2) can be set as equaling a predefined number of iterations.

Finally, to satisfy the constraint of (9d), we normalize the digital precoder \mathbf{F}_{BB} as the new one by

$$\mathbf{F}_{\text{BB}} \leftarrow \frac{\sqrt{N_s}}{\|\mathbf{S}_t \mathbf{P}_t \mathbf{F}_{\text{BB}}\|_F^2} \mathbf{F}_{\text{BB}}. \quad (23)$$

The detailed steps of the proposed VPS-HPD scheme is summarized in **Algorithm 1**. Note that the hybrid combining design including \mathbf{W}_{RF} and \mathbf{W}_{BB} are similar.

C. Convergence and Computational Complexity Analysis

1) *Convergence*: For the VPS-HPD scheme, when using alternating optimization to solve N_{RF} subproblems in (16), we adopt the Riemannian manifold optimization and exhaustive search to determine \mathbf{p}_i and \mathbf{Q}_i , respectively. Note that the Riemannian manifold optimization can guarantee the monotonic decreasing of the objective function of (18) in each iteration. The exhaustive search can also guarantee the result of \mathbf{Q}_i is optimal in each iteration. Therefore, the convergence of the VPS-HPD scheme can be verified.

2) *Computational Complexity*: Suppose the predefined numbers of iterations for *stop condition* (1) and *stop condition* (2) are $N_{\text{max}}^{(1)}$ and $N_{\text{max}}^{(2)}$. The alternating optimization of the analog precoder and the digital precoder is performed for $N_{\text{max}}^{(2)}$ iterations. During each iteration, the analog precoder design is converted into N_{RF} subproblems, where each subproblem further includes the alternating optimization of the phase matrix and switch matrix for $N_{\text{max}}^{(1)}$ iterations. When optimizing the switch matrix, it is further converted into N_t subproblems and each subproblem needs 2^{N_c} iterations for the exhaustive search. Then the computational complexity for the VPS-HPD is

$$\mathcal{O}(N_{\text{max}}^{(1)} N_{\text{RF}} N_{\text{max}}^{(2)} (N_t 2^{N_c} + \xi)) \quad (24)$$

where ξ represents the number of complex-valued multiplication to obtain a solution using the Riemannian manifold optimization.

IV. SIMULATION RESULTS

The considered mmWave massive MIMO system includes a transmitter equipped with $N_t = 128$ antennas and a receiver equipped with $N_r = 16$ antennas. For both the transmitter and receiver, we use $N_{\text{RF}} = 4$ RF chains to support $N_s = 4$ independent data streams. The resolution of phase shifters is $b = 3$ bits. The mmWave MIMO channel matrix is generated based on the Saleh-Valenzuela model, where the number of channel path is set to be $L = 4$ with $\alpha_1 \sim \mathcal{CN}(0, 1)$ and $\alpha_l \sim \mathcal{CN}(0, 0.1)$ for $l = 2, 3, 4$.

A. Spectral Efficiency Comparison

We evaluate the spectral efficiency (SE) of different HPD schemes for different architectures, including the VPS-HPD scheme for the VPS architecture, the FPS-AltMin scheme for the FPS architecture [14], the MO-AltMin scheme for the fully-connected architecture [15], and the fully digital precoding. Note that the fully digital precoding needs the same number of RF chains as that of the antennas, and therefore is hardware-expensive and only used as the performance upper bound. For fair comparisons between VPS-HPD and FPS-AltMin, we set $N_c = 8$.

As shown in Fig. 2, the SE of VPS-HPD is very close to the performance bound. The SE of VPS-HPD is better than that of MO-AltMin with much fewer phase shifters, owing to the use of low-cost switch networks in VPS-HPD. Moreover, VPS-HPD substantially outperforms FPS-AltMin with the same number of phase shifters, since we can sufficiently exploit the flexibility of phase shifters by optimizing the phase matrix according to the CSI. The number of phase shifters for both VPS-HPD and FPS-AltMin is 64, while that for MO-AltMin

TABLE II
POWER CONSUMPTION COMPARISON FOR DIFFERENT MMWAVE MIMO ARCHITECTURES.

Different architectures	Number / Power of phase shifters	Number / Power of switches	Total power
Fully-connected	576 / 17.28 W	0 / 0 W	17.28 W
FPS / VPS	64 / 1.92 W	4608 / 4.61 W	6.53 W

is 576, implying that we can save up to 88.9% phase shifters by introducing the switch networks.

We also evaluate the SE of VPS-HPD for different N_c . It is seen from Fig. 3 that the increasing N_c leads to better performance. When $N_c = 8$, the performance gap between VPS-HPD and the upper bound is 0.23 bps/Hz. The performance gap between $N_c = 8$ and $N_c = 4$ is 1.26 bps/Hz, while the latter can save 32 phase shifters compared to the former.

B. Power Consumption Comparison

In Table II, we compare the power consumption for different architectures, including the fully-connected, FPS and VPS architectures. Since the same numbers of antennas and RF chains are used for different architectures, we focus on the total power consumption coming from the phase shifters and switches. The number of phase shifters used by the transmitter and receiver for FPS and VPS is the same $2N_c N_{\text{RF}} = 64$, while that for the fully-connected architecture is $(N_t + N_r)N_{\text{RF}} = 576$. The number of switches used by the transmitter and receiver for FPS and VPS is the same $(N_t + N_r)N_{\text{RF}}N_c = 4608$. Since the power consumption of each phase shifter or each switch is 30 mW or 1 mW [12], respectively, the total power consumption of the fully-connected, FPS and VPS architectures is 17.28 W, 6.53 W and 6.53 W, respectively.

From the above discussion, the total power consumption of the fully-connected architecture is almost three times of that of the FPS or VPS architectures, while the SE performance of the fully-connected architecture is still worse than that of VPS and only slightly better than that of FPS according to Fig. 2. Therefore, the effectiveness of replacing high-resolution phase shifters by the low-cost switches is verified.

V. CONCLUSIONS

In this paper, we have proposed the VPS-HPD scheme to alternately optimize the analog precoder and the digital precoder, where the former has been converted into several subproblems and each subproblem further includes the alternating optimization of the phase matrix and switch matrix. Simulation results have shown that the spectral efficiency of the VPS-HPD scheme is very close to that of the fully digital precoding, higher than that of the existing MO-AltMin scheme for the fully-connected architecture with much fewer phase shifters, and substantially higher than that of the existing FPS-AltMin scheme for the FPS architecture with the same number of phase shifters. The future work will be continued with the focus on hardware efficient architecture for mmWave massive MIMO.

ACKNOWLEDGMENT

This work is supported in part by the National Natural Science Foundation of China (NSFC) under Grant 62071116.

REFERENCES

- [1] R. W. Heath, N. Gonzalez-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Top. Signal Process.*, vol. 10, no. 3, pp. 436–453, Apr. 2016.
- [2] X. Yang, M. Matthaiou, J. Yang, C. Wen, F. Gao, and S. Jin, "Hardware constrained millimeter-wave systems for 5G: Challenges, opportunities, and solutions," *IEEE Commun. Mag.*, vol. 57, no. 1, pp. 44–50, Jan. 2019.
- [3] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, May 2013.
- [4] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.
- [5] K. Chen, C. Qi, and G. Y. Li, "Two-step codeword design for millimeter wave massive MIMO systems with quantized phase shifters," *IEEE Trans. Signal Process.*, vol. 68, no. 1, pp. 170–180, Jan. 2020.
- [6] W. Ma, C. Qi, Z. Zhang, and J. Cheng, "Sparse channel estimation and hybrid precoding using deep learning for millimeter wave massive MIMO," *IEEE Trans. Commun.*, vol. 68, no. 5, pp. 2838–2849, May 2020.
- [7] C. Chen, C. Tsai, Y. Liu, W. Hung, and A. Wu, "Compressive sensing (CS) assisted low-complexity beamspace hybrid precoding for millimeter-wave MIMO systems," *IEEE Trans. Signal Process.*, vol. 65, no. 6, pp. 1412–1424, Mar. 2017.
- [8] X. Yu, J. Zhang, and K. B. Letaief, "Partially-connected hybrid precoding in mmWave systems with dynamic phase shifter networks," in *Proc. IEEE Int. Workshop Signal Process. Adv. Wireless Commun.*, Sapporo, Japan, Jul. 2017, pp. 129–133.
- [9] R. Méndez-Rial, C. Rusu, N. González-Prelcic, A. Alkhateeb, and R. W. Heath, "Hybrid MIMO architectures for millimeter wave communications: Phase shifters or switches?" *IEEE Access*, vol. 4, pp. 247–267, Jan. 2016.
- [10] S. Zhang, C. Guo, T. Wang, and W. Zhang, "On-off analog beamforming for massive MIMO," *IEEE Trans. Veh. Technol.*, vol. 67, no. 5, pp. 4113–4123, May 2018.
- [11] H. Li, Z. Wang, Q. Liu, and M. Li, "Transmit antenna selection and analog beamforming with low-resolution phase shifters in mmWave MISO systems," *IEEE Commun. Lett.*, vol. 22, no. 9, pp. 1878–1881, Sep. 2018.
- [12] S. Payami, N. Mysore Balasubramanya, C. Masourous, and M. Sellathurai, "Phase shifters versus switches: An energy efficiency perspective on hybrid beamforming," *IEEE Wireless Commun. Lett.*, vol. 8, no. 1, pp. 13–16, Feb. 2019.
- [13] E. E. Bahingayi and K. Lee, "Hybrid combining based on constant phase shifters and active/inactive switches," *IEEE Trans. Veh. Technol.*, vol. 69, no. 4, pp. 4058 – 4068, Apr. 2020.
- [14] X. Yu, J. Zhang, and K. B. Letaief, "A hardware-efficient analog network structure for hybrid precoding in millimeter wave systems," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 2, pp. 282–297, May 2018.
- [15] X. Yu, J. C. Shen, J. Zhang, and K. B. Letaief, "Alternating minimization algorithms for hybrid precoding in millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 485–500, Apr. 2016.