

Hybrid Precoding in Millimeter Wave Systems: How Many Phase Shifters Are Needed?

Xianghao Yu*, Jun Zhang*, and Khaled B. Letaief^{*†}, *Fellow, IEEE*

^{*}Dept. of ECE, The Hong Kong University of Science and Technology, Hong Kong

[†]Hamad Bin Khalifa University, Doha, Qatar

Email: ^{*}{xyuam, eejzhang, eekhaled}@ust.hk, [†]kletaief@hbku.edu.qa

Abstract—Hybrid precoding has been recently proposed as a cost-effective transceiver solution for millimeter wave (mm-wave) systems. The analog component in such precoders, which is composed of a phase shifter network, is the key differentiating element in contrast to conventional fully digital precoders. While a large number of phase shifters with unquantized phases are commonly assumed in existing works, in practice the phase shifters should be discretized with a coarse quantization, and their number should be reduced to a minimum due to cost and power consideration. In this paper, we propose a new hybrid precoder implementation using a small number of phase shifters with quantized and fixed phases, i.e., a fixed phase shifter (FPS) implementation, which significantly reduces the cost and hardware complexity. In addition, a dynamic switch network is proposed to enhance the spectral efficiency. Based on the proposed FPS implementation, an effective alternating minimization (AltMin) algorithm is developed with closed-form solutions in each iteration. Simulation results show that the proposed algorithm with the FPS implementation outperforms existing ones. More importantly, it needs much fewer phase shifters than existing hybrid precoder proposals, e.g., ~ 10 fixed phase shifters are sufficient for practically relevant system settings.

I. INTRODUCTION

Uplifting the carrier frequency to millimeter wave (mm-wave) bands has been proposed to meet the capacity requirement of the upcoming 5G networks, and it thus has drawn extensive attention from both academia and industry [1]. Thanks to the small wavelength of mm-wave signals, large-scale antenna arrays can be leveraged at transceivers to support directional transmissions. As equipping each antenna element with a single radio frequency (RF) chain is costly, hybrid precoding has been put forward as a cost-effective solution, which utilizes a limited number of RF chains to incorporate a digital baseband precoder and an analog RF precoder [2].

In contrast to the conventional fully digital precoder, the additional component in the hybrid one is the analog precoder, which is usually implemented by a bunch of phase shifters in the RF domain. Several hybrid precoder structures and implementations have been proposed in existing works, e.g., the fully- and partially-connected structures [3], as well as the single phase shifter (SPS) [2] and double phase shifter (DPS) [4] implementations, to provide trade-off between spectral

efficiency, energy efficiency, and algorithmic complexity. The main differences among them are the connecting strategies from RF chains to antennas and the number of phase shifters in use to compose the beamforming gain for each of the connected paths. While existing hybrid precoder structures and implementations enjoy a small number of RF chains, the number of phase shifters scales linearly with the antenna size, which is a huge number and thus causes prohibitively high cost and power consumption. On the other hand, various hybrid precoding algorithms have been proposed assuming phase shifters with arbitrary precision, e.g., orthogonal matching pursuit (OMP) [2], manifold optimization [3], and successive interference canceling [5]. Although considering phase shifters with programmable high resolution eases the hybrid precoder design, it will weaken the practicality of the results since adaptively carrying out arbitrary phase shifts at mm-wave frequencies is highly impractical [6]. Therefore, it is of critical importance to develop effective design methodologies for hybrid precoders with a small number of quantized phase shifters.

There are a few works that attempted to consider quantized phase shifters [2], [7]–[11]. The main approach is either to determine all the phases at once [2], [7]–[10] or update one phase at a time [11] by ignoring the quantization effect at first. Then the phases are heuristically quantized into the finite feasible set according to a certain criterion. However, a simple quantization step is far from satisfactory, and the optimality and convergence of the proposed algorithms cannot be guaranteed [11]. On the other hand, the number of phase shifters in use was to some extent reduced in [10], which was determined for achieving a certain precision of the unquantized ones. Unfortunately, a large number of phase shifters are still needed for practical settings, i.e., 40 phase shifters for each RF chain, and the number will vary according to the precision requirement. More importantly, in existing works, the phases need to be adapted to the channel states, which brings high complexity for hardware implementation and also increases power consumption.

To overcome the above limitations, in this paper we propose a novel hybrid precoder implementation for general multiuser orthogonal frequency-division multiplexing (OFDM) mm-wave systems, where only a small number of phase shifters with fixed phases are available [6], namely the fixed phase

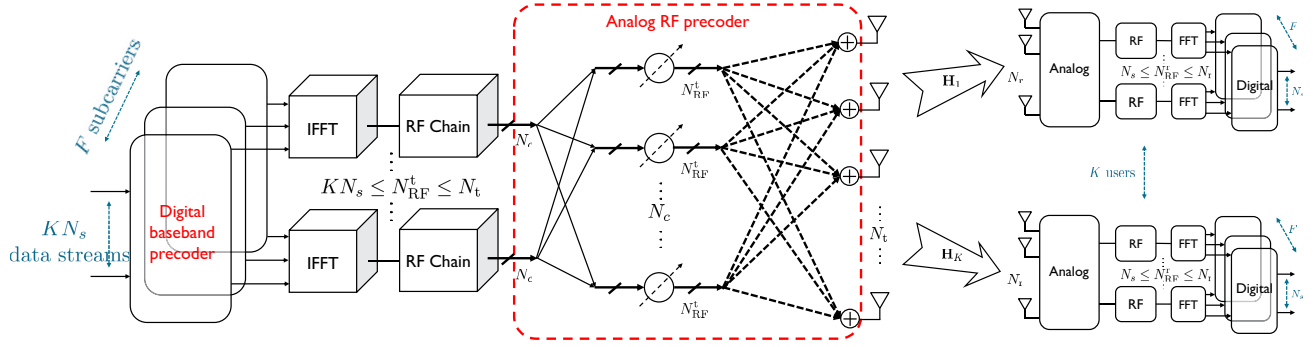


Fig. 1. A multiuser mm-wave MIMO-OFDM system with FPS hybrid precoder implementation. To simplify the figure, in the analog precoder, each solid line with a slash represents parallel signals transmission while each dotted line stands for N_{RF}^t switches.

shifter (FPS) implementation. To compensate the performance loss induced by the fixed phases, a switch network is proposed to provide dynamic mappings from phase shifters to antennas, which is easily implementable with adaptive switches [6], [9]. With the proposed FPS implementation, we develop an alternating minimization (AltMin) algorithm to design the hybrid precoder [3], where an upper bound of the objective function is derived as an effective surrogate. In particular, the large-scale binary constraints introduced by the switch network are delicately tackled with the help of the upper bound, which leads to closed-form solutions for both the dynamic switch network and the digital baseband precoder, and therefore enables a low-complexity hybrid precoding algorithm. Simulation results shall demonstrate that the proposed FPS-AltMin algorithm outperforms existing ones and approaches the performance of the fully digital precoder. What deserves a special mention is the sharp reduction of the number of phase shifters compared to existing hybrid precoder implementations, which indicates that the proposed FPS implementation is a promising candidate for hybrid precoding in 5G mm-wave communication systems.

II. SYSTEM MODEL

A. Signal Model

Consider the downlink transmission for a multiuser mm-wave MIMO-OFDM system as shown in Fig. 1. A base station (BS) leverages an N_t -size antenna array to serve K users over F subcarriers using OFDM. Each user is equipped with N_r antennas and receives N_s data streams from the BS on each subcarrier. The numbers of available RF chains are N_{RF}^t and N_{RF}^r for the BS and each user, respectively, which are restricted as $K N_s \leq N_{\text{RF}}^t \leq N_t$ and $N_s \leq N_{\text{RF}}^r \leq N_r$.

The received signal of the k -th user on the f -th subcarrier is given by

$$\mathbf{y}_{k,f} = \mathbf{W}_{\text{BB},k,f}^H \mathbf{W}_{\text{RF},k}^H \left(\mathbf{H}_{k,f} \mathbf{F}_{\text{RF}} \sum_{k=1}^K \mathbf{F}_{\text{BB},k,f} \mathbf{s}_{k,f} + \mathbf{n}_{k,f} \right), \quad (1)$$

where $\mathbf{s}_{k,f}$ is the transmitted signal to the k -th user on the f -th subcarrier such that $\mathbb{E}[\mathbf{s}_{k,f} \mathbf{s}_{k,f}^H] = \frac{P}{K N_s F} \mathbf{I}_{N_s}$, and $\mathbf{n}_{k,f}$ is the circularly symmetric complex Gaussian noise with power

as σ_n^2 at the users. The digital baseband precoders and combiners are denoted as $\mathbf{F}_{\text{BB},k,f}$ and $\mathbf{W}_{\text{BB},k,f}$, respectively, with dimensions $N_{\text{RF}}^t \times N_s$ and $N_{\text{RF}}^r \times N_s$. Since the transmitted signals for all the users are mixed together by the digital precoders, and analog RF precoding is a post-IFFT operation, the RF analog precoder \mathbf{F}_{RF} with dimension $N_t \times N_{\text{RF}}^t$ is a common component for all the users and subcarriers. Correspondingly, the $N_r \times N_{\text{RF}}^r$ RF analog combiner $\mathbf{W}_{\text{RF},k}$ is subcarrier-independent for each user.

B. FPS Implementation

In earlier works on hybrid precoding [2], [3], [5], [7]–[9], a single phase shifter is adopted to adjust the phase of each of the paths from RF chains to antennas. Therefore, $N_t N_{\text{RF}}^t$ phase shifters are required, commonly assumed with arbitrary precision. Recently, it was shown in [4] that the performance of the hybrid precoder can be greatly improved by passing each signal through two unquantized phase shifters and then combining the outputs, which, however, induces high hardware complexity by employing $2 N_t N_{\text{RF}}^t$ adaptive phase shifters.

In this paper, we propose a hybrid precoder implementation using N_c phase shifters with fixed phases [6], where $N_c \ll N_t N_{\text{RF}}^t$, as shown in Fig. 1. Nevertheless, the limited number of fixed phase shifters, which cannot be adaptively adjusted according to the channel states, inevitably entail performance loss. To overcome this drawback brought by the simplified hardware implementation, we propose to cascade a dynamic switch network after the fixed phase shifters, which is adapted to the channel states.

In particular, N_c multichannel (N_{RF}^t -channel) fixed phase shifters [12] are deployed in the phase shifter network, each of which simultaneously processes the output signals from N_{RF}^t RF chains, i.e., in a parallel fashion. To clearly illustrate the proposed FPS implementation, we focus on one signal flow from an RF chain to an antenna, as shown in Fig. 2. The N_c fixed phase shifters generate N_c signals with different phases for the output signal of the given RF chain. Inspired by the idea of doubling phase shifters to achieve high spectral efficiency, as demonstrated in [4], we propose to adaptively combine a subset of the N_c signals to compose the analog precoding gain from the RF chain to the antenna, which is

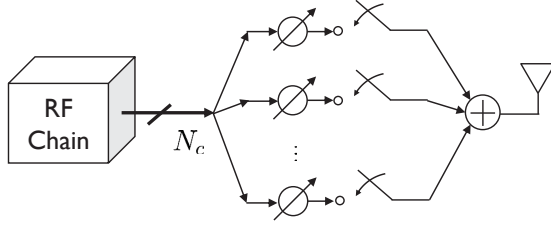


Fig. 2. The FPS implementation from an RF chain to each antenna.

implemented with N_c adaptive switches. As N_c switches are needed for each RF chain-antenna pair, in total $N_t N_{\text{RF}}^t N_c$ switches are required in the proposed FPS implementation. Note that adaptive switches with binary states are easier to implement in mm-wave bands than adaptive phase shifters with arbitrary precision [6], [9].

Accordingly, the analog RF precoding matrix \mathbf{F}_{RF} can be expressed as

$$\mathbf{F}_{\text{RF}} = \mathbf{S}\mathbf{C}, \quad (2)$$

where $\mathbf{S} \in \{0, 1\}^{N_t \times N_c N_{\text{RF}}^t}$ is the **switch matrix**, and the boolean constraints are induced by the switches with binary states. The matrix $\mathbf{C} \in \mathbb{C}^{N_c N_{\text{RF}}^t \times N_{\text{RF}}^t}$ stands for the **phase shift operation** carried out by the available fixed phase shifters, given by a block diagonal matrix as

$$\mathbf{C} = \text{diag} \left(\underbrace{\mathbf{c}, \mathbf{c}, \dots, \mathbf{c}}_{N_{\text{RF}}^t} \right), \quad (3)$$

where $\mathbf{c} = \frac{1}{\sqrt{N_c}} [e^{j\theta_1}, e^{j\theta_2}, \dots, e^{j\theta_{N_c}}]^T$ is the **normalized phase shifter vector** containing all N_c fixed phases $\{\theta_i\}_{i=1}^{N_c}$.

C. Problem Formulation

It has been shown in [2]–[4], [7], [11] that minimizing the Euclidean distance between the fully digital precoder and the hybrid precoder is an effective and tractable alternative objective for maximizing the spectral efficiency of mm-wave systems. In this paper, we resort to this approach and the **hybrid precoder design** is correspondingly formulated as

$$\begin{aligned} \mathcal{P}_1 : \quad & \underset{\mathbf{S}, \mathbf{F}_{\text{BB}}}{\text{minimize}} \quad \|\mathbf{F}_{\text{opt}} - \mathbf{S}\mathbf{C}\mathbf{F}_{\text{BB}}\|_F^2 \\ & \text{subject to} \quad \begin{cases} \mathbf{S} \in \mathcal{B} \\ \|\mathbf{S}\mathbf{C}\mathbf{F}_{\text{BB}}\|_F^2 \leq K N_s F, \end{cases} \end{aligned} \quad (4)$$

where $\mathbf{F}_{\text{opt}} = [\mathbf{F}_{\text{opt},1,1}, \dots, \mathbf{F}_{\text{opt},k,f}, \dots, \mathbf{F}_{\text{opt},K,F}]$ is the **combined fully digital precoder** with dimension $N_t \times K N_s F$, and $\mathbf{F}_{\text{BB}} = [\mathbf{F}_{\text{BB},1,1}, \dots, \mathbf{F}_{\text{BB},k,f}, \dots, \mathbf{F}_{\text{BB},K,F}]$ is the concatenated digital baseband precoder with dimension $N_{\text{RF}}^t \times K N_s F$. The set of binary matrices is denoted as \mathcal{B} , and the second constraint is the transmit power constraint. Note that the combiners at the user side can be designed in the same way without the power constraint [3], [13] and thus are omitted due to space limitation.

Remark 1: Since the switch matrix \mathbf{S} is with finite possibilities, the cardinality of the constraint set for \mathbf{F}_{RF} is finite,

which means the OMP algorithm [2] is applicable to this problem \mathcal{P}_1 . However, the dimension of the dictionary in the OMP algorithm is oversized, i.e., $\left[\sum_{i=1}^{N_c} \binom{N_c}{i} \right]^{N_t}$, which prevents its practical implementation.

Remark 2: Alternating minimization can be directly applied to \mathcal{P}_1 where the binary constraints can be tackled with **semidefinite relaxation** [3]. However, an $N_t N_{\text{RF}}^t N_c + 1$ -dimension semidefinite programming (SDP) problem should be solved in each iteration, which causes prohibitive computational complexity. Moreover, the optimality of the relaxation in each iteration cannot be ensured and hence the overall convergence of the AltMin algorithm cannot be guaranteed.

As illustrated above, the main difficulty to solve \mathcal{P}_1 is the binary constraints of \mathbf{S} , and it is the main obstacle for designing an efficient algorithm with performance guarantee. In this paper, by deriving an effective surrogate and adopting alternating minimization, we shall come up with a **low-complexity hybrid precoding algorithm** that well tackles the **binary constraints**.

III. HYBRID PRECODER DESIGN IN SINGLE-CARRIER SYSTEMS WITH THE FPS IMPLEMENTATION

In this section, we first present the hybrid precoder design in single-carrier systems¹, i.e., when $F = 1$. In particular, an upper bound of the objective function is firstly derived, based on which an alternating minimization algorithm is then developed.

A. Objective Upper Bound

In [3], [4], [13], imposing a **semi-orthogonal structure** for \mathbf{F}_{BB} is shown to achieve near-optimal performance. Inspired by these results, we take a similar approach. In single-carrier systems, the digital precoder matrix \mathbf{F}_{BB} is with dimension $N_{\text{RF}}^t \times K N_s$. Recall that the number of RF chains is limited as $K N_s \leq N_{\text{RF}}^t < N_t$, which forces \mathbf{F}_{BB} as to be tall matrix, and thus the semi-orthogonal constraint is specified as

$$\mathbf{F}_{\text{BB}}^H \mathbf{F}_{\text{BB}} = \alpha^2 \mathbf{F}_{\text{DD}}^H \mathbf{F}_{\text{DD}} = \alpha^2 \mathbf{I}_{K N_s}, \quad (5)$$

where $\mathbf{F}_{\text{BB}} = \alpha \mathbf{F}_{\text{DD}}$ and \mathbf{F}_{DD} is a **semi-unitary matrix**. Then, the objective function in \mathcal{P}_1 can be rewritten as

$$\|\mathbf{F}_{\text{opt}}\|_F^2 - 2\alpha \Re \text{tr}(\mathbf{F}_{\text{DD}}^H \mathbf{F}_{\text{opt}}^H \mathbf{S}\mathbf{C}) + \alpha^2 \|\mathbf{S}\mathbf{C}\mathbf{F}_{\text{DD}}\|_F^2. \quad (6)$$

Note that, according to (3), the **phase shifter matrix \mathbf{C} is also a semi-unitary matrix**, i.e., $\mathbf{C}^H \mathbf{C} = \mathbf{I}_{N_{\text{RF}}^t}$. Therefore, we can derive an upper bound for the last term in (6), given by

$$\begin{aligned} \|\mathbf{S}\mathbf{C}\mathbf{F}_{\text{DD}}\|_F^2 &= \text{tr}(\mathbf{F}_{\text{DD}}^H \mathbf{C}^H \mathbf{S}^H \mathbf{S} \mathbf{C} \mathbf{F}_{\text{DD}}) \\ &\stackrel{(a)}{=} \text{tr} \left(\begin{bmatrix} \mathbf{I}_{K N_s} & \mathbf{0} \end{bmatrix} \mathbf{K}^H \mathbf{S}^H \mathbf{S} \mathbf{K} \right) \\ &< \text{tr}(\mathbf{K}^H \mathbf{S}^H \mathbf{S} \mathbf{K}) = \|\mathbf{S}\|_F^2, \end{aligned} \quad (7)$$

where (a) follows the singular value decomposition (SVD) of $\mathbf{C}\mathbf{F}_{\text{DD}}\mathbf{F}_{\text{DD}}^H \mathbf{C}^H = \mathbf{K} \text{diag}(\mathbf{I}_{K N_s}, \mathbf{0}) \mathbf{K}^H$ by utilizing the

¹In this paper, single-carrier systems refer to single-carrier transmissions assuming flat-fading channels. The choice of such systems is for the ease of presentation, and the algorithm will be later extended to the more realistic multicarrier case with frequency-selective fading channels.

semi-unitary property of $\mathbf{C}\mathbf{F}_{\text{DD}}$. Thus, we obtain an upper bound for the original objective function, expressed as

$$\|\mathbf{F}_{\text{opt}}\|_F^2 - 2\alpha \Re \text{tr}(\mathbf{F}_{\text{DD}} \mathbf{F}_{\text{opt}}^H \mathbf{S} \mathbf{C}) + \alpha^2 \|\mathbf{S}\|_F^2. \quad (8)$$

B. Alternating Minimization

By adopting the upper bound (8) as the surrogate objective function and dropping the constant term $\|\mathbf{F}_{\text{opt}}\|_F^2$, the hybrid precoder design problem is reformulated as

$$\mathcal{P}_2 : \begin{aligned} & \underset{\alpha, \mathbf{S}, \mathbf{F}_{\text{DD}}}{\text{minimize}} && \alpha^2 \|\mathbf{S}\|_F^2 - 2\alpha \Re \text{tr}(\mathbf{F}_{\text{DD}} \mathbf{F}_{\text{opt}}^H \mathbf{S} \mathbf{C}) \\ & \text{subject to} && \begin{cases} \mathbf{S} \in \mathcal{B} \\ \mathbf{F}_{\text{DD}}^H \mathbf{F}_{\text{DD}} = \mathbf{I}_{KN_s}. \end{cases} \end{aligned} \quad (9)$$

Alternating minimization, as an effective tool for optimization problems involving different subsets of variables, has been widely applied and shown empirically successful in hybrid precoder design [3], [4], [13]. **In this section, we apply it this effective rule of thumb to the hybrid precoder design with the FPS implementation.**

In each step of the AltMin algorithm, one subset of the optimization variables is optimized while keeping the other parts fixed. **When the switch matrix \mathbf{S} and α are fixed,** the optimization problem can be written as

$$\begin{aligned} & \underset{\mathbf{F}_{\text{DD}}}{\text{maximize}} && \alpha \Re \text{tr}(\mathbf{F}_{\text{DD}} \mathbf{F}_{\text{opt}}^H \mathbf{S} \mathbf{C}) \\ & \text{subject to} && \mathbf{F}_{\text{DD}}^H \mathbf{F}_{\text{DD}} = \mathbf{I}_{KN_s}. \end{aligned} \quad (10)$$

According to the definition of the **dual norm** [14], we have

$$\begin{aligned} \alpha \Re \text{tr}(\mathbf{F}_{\text{DD}} \mathbf{F}_{\text{opt}}^H \mathbf{S} \mathbf{C}) & \leq |\text{tr}(\alpha \mathbf{F}_{\text{DD}} \mathbf{F}_{\text{opt}}^H \mathbf{S} \mathbf{C})| \\ & \stackrel{(b)}{\leq} \|\mathbf{F}_{\text{DD}}\|_\infty \|\alpha \mathbf{F}_{\text{opt}}^H \mathbf{S} \mathbf{C}\|_1 \\ & = \|\alpha \mathbf{F}_{\text{opt}}^H \mathbf{S} \mathbf{C}\|_1 = \sum_{i=1}^{KN_s} \sigma_i, \end{aligned} \quad (11)$$

where $\|\cdot\|_\infty$ and $\|\cdot\|_1$ stand for the infinite and one Schatten norms [14], and (b) follows the Hölder's inequality. The equality is established only when

$$\mathbf{F}_{\text{DD}} = \mathbf{V}_1 \mathbf{U}^H, \quad (12)$$

where $\alpha \mathbf{F}_{\text{opt}}^H \mathbf{S} \mathbf{C} = \mathbf{U} \Sigma \mathbf{V}_1^H$ follows the SVD and Σ is a diagonal matrix with non-zero singular values $\sigma_1, \dots, \sigma_{KN_s}$.

While we can divide the optimization of the two variables α and \mathbf{S} into two separate subproblems, we propose to update them in parallel to save the number of subproblems involved in the AltMin algorithm and therefore reduce the computational complexity. By adding a **constant term** $\|\Re(\mathbf{F}_{\text{opt}} \mathbf{F}_{\text{DD}}^H \mathbf{C}^H)\|_F^2$ to the objective function of \mathcal{P}_2 , the subproblem of updating α and \mathbf{S} can be recast as

$$\begin{aligned} & \underset{\alpha, \mathbf{S}}{\text{minimize}} && \|\Re(\mathbf{F}_{\text{opt}} \mathbf{F}_{\text{DD}}^H \mathbf{C}^H) - \alpha \mathbf{S}\|_F^2 \\ & \text{subject to} && \mathbf{S} \in \mathcal{B}. \end{aligned} \quad (13)$$

Proposition 1. *The optimal solution to (13) is given by*

$$\alpha^* = \arg \min_{\{\tilde{x}_i, \bar{x}_i\}_{i=1}^n} \{f(\tilde{x}_i), f(\bar{x}_i)\}, \quad (14)$$

$$\mathbf{S}^* = \begin{cases} \mathbf{1} \left\{ \Re(\mathbf{F}_{\text{opt}} \mathbf{F}_{\text{DD}}^H \mathbf{C}^H) > \frac{\alpha}{2} \mathbf{1}_{N_t \times N_c N_{\text{RF}}^t} \right\} & \alpha > 0 \\ \mathbf{1} \left\{ \Re(\mathbf{F}_{\text{opt}} \mathbf{F}_{\text{DD}}^H \mathbf{C}^H) < \frac{\alpha}{2} \mathbf{1}_{N_t \times N_c N_{\text{RF}}^t} \right\} & \alpha < 0, \end{cases} \quad (15)$$

where $\mathbf{x} = \text{vec}\{\Re(\mathbf{F}_{\text{opt}} \mathbf{F}_{\text{DD}}^H \mathbf{C}^H)\}$, $\mathbf{1}(\cdot)$ is the indicator function, and $\mathbf{1}_{m \times n}$ denotes an $m \times n$ matrix with all entries equal to one. The objective function in (13) can be rewritten as $f(\alpha)$ as (26) in the proof. In addition, **\tilde{x}_i is the i -th smallest entry in \mathbf{x} , and²**

$$\bar{x}_i \triangleq \begin{cases} \frac{\sum_{j=1}^i \tilde{x}_j}{i} & \alpha < 0 \text{ and } \frac{\sum_{j=1}^i \tilde{x}_j}{i} \in [2\tilde{x}_i, 2\tilde{x}_{i+1}] \\ \frac{\sum_{j=i+1}^n \tilde{x}_j}{n-i} & \alpha > 0 \text{ and } \frac{\sum_{j=i+1}^n \tilde{x}_j}{n-i} \in [2\tilde{x}_i, 2\tilde{x}_{i+1}] \\ +\infty & \text{otherwise.} \end{cases} \quad (16)$$

Proof: See Appendix A. ■

Basically, the optimal α^* is obtained via a closed-form solution by comparing the optimal solutions of α in all the intervals $\{\mathcal{R}_i\}_{i=1}^n$, where $\mathcal{R}_i \triangleq [2\tilde{x}_i, 2\tilde{x}_{i+1}]$. Nevertheless, since the number of intervals that need to compare is **$n = N_t N_c N_{\text{RF}}^t$** , it will incur high computational complexity when N_t is large in mm-wave systems. **In the following lemma, we show that there is no need to compute the optimal α in all the intervals \mathcal{R}_i , which will further reduce the complexity of the proposed algorithm.**

Lemma 1. *The optimal α^* is obtained at one of the points $\{\tilde{x}_i\}_{i=1}^n$.*

Proof: See Appendix B. ■

Lemma 1 indicates that any endpoint of the intervals cannot be the optimal solution for α . Therefore, we only need to pick the \tilde{x}_i 's that have finite values of $f(\tilde{x}_i)$, i.e., the ones that satisfy the first two conditions in (16), denoted as a set \mathcal{X} , and the optimal solution for α is given by

$$\alpha^* = \arg \min_{\tilde{x}_i \in \mathcal{X}} f(\tilde{x}_i). \quad (17)$$

By Lemma 1, the number of intervals we need to compare to obtain the optimal α^* is shrunk from n to $|\mathcal{X}|$, which is empirically shown to be less than 5 via simulations in Section V and hence further reduces the computational complexity of the proposed AltMin algorithm.

Remark 3: It is shown that, with the help of the upper bound derived in (7), **the large-scale binary switch matrix \mathbf{S} can be efficiently optimized by a closed-form solution**, which verifies the benefits and superiority of the surrogate objective function adopted in \mathcal{P}_2 .

With the closed-form solutions derived in (12), (15), and (17) at hands, the AltMin algorithm for the FPS implementation is summarized as **FPS-AltMin Algorithm**. The FPS-AltMin algorithm is essentially a **block coordinate descent (BCD)** algorithm with two blocks that have globally optimal solutions in Steps 3 and 4, and the algorithm is guaranteed to converge to a stationary point of \mathcal{P}_2 [15]. **The algorithm may be sensitive to the initial point $\mathbf{F}_{\text{DD}}^{(0)}$.** Note that the fully

² $f(\alpha)$ is a coercive function, i.e., $f(+\infty) \rightarrow +\infty$.

FPS-AltMin Algorithm: A Low-Complexity Hybrid Precoding Algorithm for the FPS Implementation

Input: \mathbf{F}_{opt}

- 1: Construct an initial point for $\mathbf{F}_{\text{DD}}^{(0)}$ and set $k = 0$;
 - 2: **repeat**
 - 3: Fix $\mathbf{F}_{\text{DD}}^{(k)}$, optimize $\alpha^{(k)}$ and $\mathbf{S}^{(k)}$ according to (17) and (15), respectively;
 - 4: Fix $\mathbf{S}^{(k)}$ and $\alpha^{(k)}$, update $\mathbf{F}_{\text{DD}}^{(k)}$ with (12);
 - 5: $k \leftarrow k + 1$;
 - 6: **until** convergence.
 - 7: Compute the additional BD precoder at the baseband to cancel the inter-user interference [4].
 - 8: For the digital precoder at the transmit end, normalize $\mathbf{F}_{\text{BB}} = \frac{\sqrt{KN_s F}}{\|\mathbf{S}\mathbf{C}\mathbf{F}_{\text{DD}}\|_F} \mathbf{F}_{\text{DD}}$.
-

digital precoding matrix \mathbf{F}_{opt} can be decomposed as follows according to its SVD $\mathbf{F}_{\text{opt}} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$, i.e.,

$$\mathbf{F}_{\text{opt}} = [\mathbf{U}\mathbf{\Sigma} \quad \mathbf{F}] \begin{bmatrix} \mathbf{V}^H \\ \mathbf{0} \end{bmatrix}, \quad (18)$$

where $\mathbf{U}\mathbf{\Sigma}$ is an $N_t \times KN_s$ full rank matrix, \mathbf{V}^H is a KN_s dimension square matrix, and \mathbf{F} is an arbitrary $N_t \times (N_{\text{RF}}^t - KN_s)$ matrix. In (18), the fully digital precoding matrix \mathbf{F}_{opt} is decomposed into two matrices that satisfy the dimensions of \mathbf{F}_{RF} and \mathbf{F}_{DD} , respectively. In this way, we propose to construct the initial point $\mathbf{F}_{\text{DD}}^{(0)}$ as

$$\mathbf{F}_{\text{DD}}^{(0)} = [\mathbf{V} \quad \mathbf{0}_{KN_s \times (N_{\text{RF}}^t - KN_s)}]^H. \quad (19)$$

To cancel the inter-user interference, similar to [4], we cascade an additional block diagonal precoder at the baseband in the Step 7 based on the effective channel including the hybrid precoder and physical channel. In the final step, we normalize the digital precoder to maximize the signal to noise ratio (SNR) while satisfying the transmit power constraint.

IV. HYBRID PRECODER DESIGN IN MULTICARRIER SYSTEMS WITH THE FPS IMPLEMENTATION

Multicarrier techniques such as OFDM are often utilized to overcome the multipath fading caused by the large available bandwidth in mm-wave systems. Compared with the narrowband hybrid precoder design in Section III, the main difference in OFDM systems is that the analog precoder is shared by all the subcarriers. In particular, the digital precoding matrix $\mathbf{F}_{\text{BB}} \in \mathbb{C}^{N_{\text{RF}}^t \times KN_s F}$ in \mathcal{P}_1 is no longer a tall matrix since $KN_s F \geq N_{\text{RF}}^t$ for practical OFDM system settings.

In this section, we modify the FPS-AltMin algorithm for OFDM systems. Similar to (5), we enforce a semi-orthogonal constraint on the digital precoding matrix, i.e.,

$$\mathbf{F}_{\text{BB}} \mathbf{F}_{\text{BB}}^H = \alpha^2 \mathbf{F}_{\text{DD}} \mathbf{F}_{\text{DD}}^H = \alpha^2 \mathbf{I}_{N_{\text{RF}}^t}. \quad (20)$$

In this way, the upper bound of the objective function derived in (7) still holds since

$$\begin{aligned} \|\mathbf{S}\mathbf{C}\mathbf{F}_{\text{DD}}\|_F^2 &= \text{tr}(\mathbf{C}^H \mathbf{S}^H \mathbf{S} \mathbf{C}) \\ &\stackrel{(c)}{=} \text{tr} \left(\begin{bmatrix} \mathbf{I}_{N_{\text{RF}}^t} \\ \mathbf{0} \end{bmatrix} \mathbf{K}^H \mathbf{S}^H \mathbf{S} \mathbf{K} \right) \\ &< \text{tr}(\mathbf{K}^H \mathbf{S}^H \mathbf{S} \mathbf{K}) = \|\mathbf{S}\|_F^2, \end{aligned} \quad (21)$$

where (c) comes from the SVD of $\mathbf{C}\mathbf{C}^H$ since \mathbf{C} is a semi-unitary matrix. In the AltMin algorithm, the update of α and \mathbf{S} is the same as that in Section III-B. Since the dimension of \mathbf{F}_{DD} is different in OFDM systems, the optimization of \mathbf{F}_{DD} is modified as

$$\mathbf{F}_{\text{DD}} = \mathbf{V}\mathbf{U}_1^H, \quad (22)$$

where $\mathbf{F}_{\text{opt}}^H \mathbf{S} \mathbf{C} = \mathbf{U}_1 \mathbf{\Sigma} \mathbf{V}^H$ and $\mathbf{\Sigma}$ is a diagonal matrix with non-zero singular values $\sigma_1, \dots, \sigma_{N_{\text{RF}}^t}$, which is the SVD of $\mathbf{F}_{\text{opt}}^H \mathbf{S} \mathbf{C}$. Correspondingly, the construction of the initial $\mathbf{F}_{\text{DD}}^{(0)}$ is given by

$$\mathbf{F}_{\text{DD}}^{(0)} = \mathbf{V}_{[1:N_{\text{RF}}^t]}^H, \quad (23)$$

where $\mathbf{F}_{\text{opt}} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$ is the SVD of \mathbf{F}_{opt} and the subscript $[1:n]$ denotes the first to the n -th columns of a matrix.

By substituting (23) and (22) into the Steps 1 and 4, we obtain the FPS-AltMin algorithm for OFDM mm-wave systems.

V. SIMULATION RESULTS

In this section, we will evaluate the performance of the proposed FPS-AltMin algorithm via simulations. The BS and each user are equipped with 144 and 16 antennas, respectively, while all the transceivers are equipped with uniform planar arrays (UPAs). Four users and 128 subcarriers are assumed when considering multiuser OFDM systems. To reduce the cost and power consumption, the minimum number of RF chains is adopted according to the assumptions in Section II-A, i.e., $N_{\text{RF}}^t = KN_s$ and $N_{\text{RF}}^r = N_s$. The phases of the available fixed phase shifters are uniformly separated within $[0, 2\pi]$ by N_c equal length intervals. Furthermore, the Saleh-Valenzuela model is adopted in simulations to characterize mm-wave channels [2], [3]. The nominal SNR is defined as $\frac{P}{KN_s F \sigma_n^2}$, and all the simulation results are averaged over 1000 channel realizations.

A. Single-User Single-Carrier (SU-SC) Systems

As a great number of previous efforts have been spent on point-to-point systems, it is intriguing to test the performance of the proposed algorithm by comparing with existing works as benchmarks. The OMP algorithm proposed in [2] has been widely used as a low-complexity algorithm with the analog precoder selected from a predefined set. The MO-AltMin algorithm was then proposed in [3] to improve the performance of the OMP algorithm, yet with high computational complexity of performing the manifold optimization. Both of these algorithms are applied with the SPS implementation. Fig. 3 shows that the proposed FPS-AltMin algorithm achieves the highest spectral efficiency with the simulation time comparable to the OMP algorithm. The performance

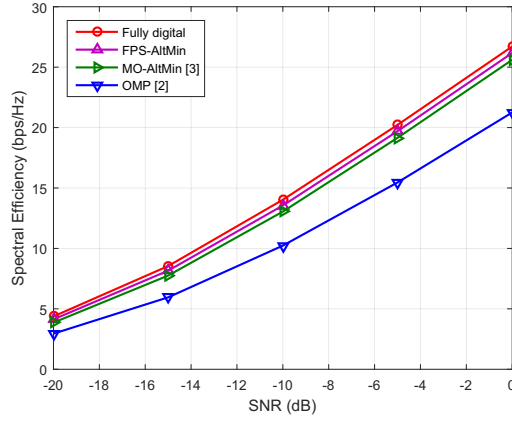


Fig. 3. Spectral efficiency achieved by different hybrid precoding algorithms in SU-SC systems when $N_{\text{RF}}^t = N_{\text{RF}}^r = N_s = 4$ and $N_c = 30$.

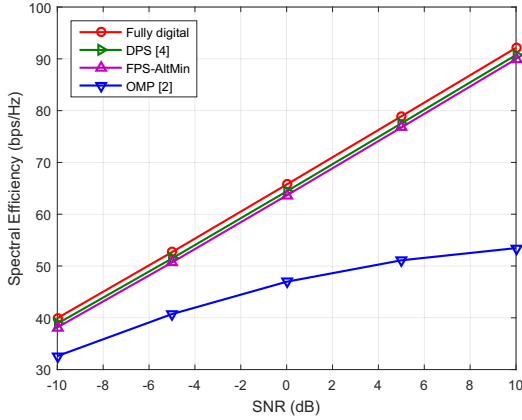


Fig. 4. Spectral efficiency achieved by different hybrid precoding algorithms in MU-MC systems when $N_{\text{RF}}^t = 8$, $N_{\text{RF}}^r = N_s = 2$, and $N_c = 30$.

gain is mainly attributed to the proposed FPS implementation, where the unit modulus constraints in the SPS implementation are relaxed. Furthermore, the proposed algorithm leads to an effective design of the dynamic switch network, and provides a better approximation of the fully digital precoder than existing algorithms.

B. Multiuser Multicarrier (MU-MC) Systems

In [4], the DPS implementation was proposed for MU-MC systems to approach the performance of the fully digital precoder by sacrificing the hardware complexity of employing a large number of phase shifters, i.e., $2N_t N_{\text{RF}}^t$ phase shifters. As shown in Fig. 4, the proposed FPS-AltMin algorithm only entails little performance loss compared to the DPS implementation when only 30 fixed phase shifters are adopted. On the other hand, it enjoys significant improvement in terms of spectral efficiency compared to the OMP algorithm. This result demonstrates the effectiveness of both the newly proposed implementation and algorithm. In addition, it indicates that the number of phase shifters can be sharply reduced even if the analog precoder is shared by all the subcarriers and users in MU-MC systems.

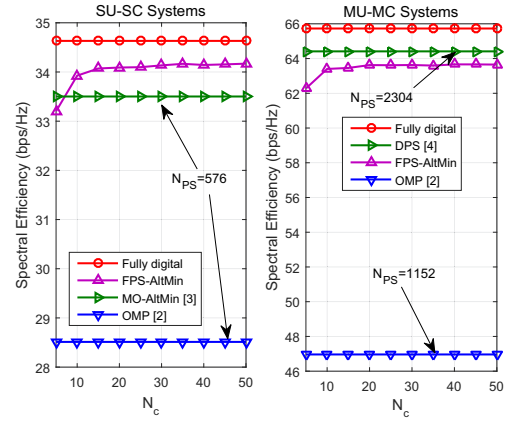


Fig. 5. Spectral efficiency achieved by different hybrid precoding algorithms in mm-wave MIMO systems given SNR = 0 dB.

C. How Many Phase Shifters Are Needed?

Fig. 5 plots the spectral efficiency achieved with different numbers of fixed phase shifters, i.e., N_c . The simulation parameters are the same as those in Figs. 3 and 4 for SU-SC and MU-MC systems, respectively. Fig. 5 shows that in SU-SC systems 15 phase shifters are enough for achieving a satisfactory performance as the spectral efficiency almost saturates when we further increase the number of fixed phase shifters. By contrast, 576 phase shifters are needed in the SPS implementation. Moreover, the OMP algorithm achieves a lower spectral efficiency and the MO-AltMin algorithm suffers from the high computational complexity. A similar phenomenon is found in MU-MC systems, i.e., around 10 fixed phase shifters are sufficient, which has not been revealed in existing works. Although the performance of the DPS implementation slightly outperforms the proposed FPS-AltMin algorithm, it employs 200 times more phase shifters. This illustrates that the proposed FPS implementation is much more cost-effective than existing hybrid precoder implementations, and with satisfactory performance.

VI. CONCLUSIONS

In this paper, we proposed a cost-effective hybrid precoder implementation with a small number of fixed phase shifters. To enhance the performance, a dynamic switch network was adopted, for which a low-complexity AltMin algorithm was developed. The proposed implementation is able to approach the performance of the fully digital precoder, remarkably, with small numbers of RF chains and phase shifters. Thus, this proposal stands out as a promising candidate for hybrid precoders for 5G mm-wave systems.

APPENDIX A PROOF OF PROPOSITION 1

Note that each entry in the switch matrix \mathbf{S} is either 0 or 1, and we discover that they can be optimally determined individually once α is given. In particular, to minimize the objective function, $s_{m,n}$ should take value 1 if the corresponding (m,n) -th entry in the matrix $\Re(\mathbf{F}_{\text{opt}} \mathbf{F}_{\text{DD}}^H \mathbf{C}^H)$ is closer to

$$\begin{aligned}
f(\alpha) &= \|\tilde{\mathbf{x}} - \alpha \mathbf{s}\|_2^2 \\
&= \begin{cases} \sum_{j=1}^i (\tilde{x}_j - \alpha)^2 + \sum_{j=i+1}^n \tilde{x}_j^2 & \alpha < 0 \text{ and } \frac{\alpha}{2} \in \mathcal{R}_i \\ \sum_{j=1}^i \tilde{x}_j^2 + \sum_{j=i+1}^n (\tilde{x}_j - \alpha)^2 & \alpha > 0 \text{ and } \frac{\alpha}{2} \in \mathcal{R}_i \end{cases} = \begin{cases} i\alpha^2 - 2\sum_{j=1}^i \tilde{x}_j\alpha + \sum_{j=1}^n \tilde{x}_j^2 & \alpha < 0 \text{ and } \alpha \in [2\tilde{x}_i, 2\tilde{x}_{i+1}] \\ (n-i)\alpha^2 - 2\sum_{j=i+1}^n \tilde{x}_j\alpha + \sum_{j=1}^n \tilde{x}_j^2 & \alpha > 0 \text{ and } \alpha \in [2\tilde{x}_i, 2\tilde{x}_{i+1}] \end{cases}
\end{aligned} \tag{26}$$

α than 0 in the Euclidean space, and take value 0 otherwise, as given in (15).

The remaining problem is to choose an optimal α that minimizes the objective function. Since $\mathbf{S} \in \mathcal{B}$ is an element wise constraint, to simplify the notations, it is equivalent to consider the vectorization version of (13), given by

$$\begin{aligned}
&\underset{\alpha, \mathbf{s}}{\text{minimize}} \quad \|\mathbf{x} - \alpha \mathbf{s}\|_2^2 \\
&\text{subject to} \quad \mathbf{s} \in \{0, 1\}^n,
\end{aligned} \tag{24}$$

where $n = N_t N_c N_{\text{RF}}^t$, $\mathbf{x} \triangleq \text{vec}\{\Re(\mathbf{F}_{\text{opt}} \mathbf{F}_{\text{DD}}^H \mathbf{C}^H)\}$, and $\mathbf{s} = [s_1, s_2, \dots, s_n] \triangleq \text{vec}\{\alpha \mathbf{S}\}$.

First, we sort the entries of \mathbf{x} in the ascending order as $\tilde{\mathbf{x}} = [\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n]$, where $\tilde{x}_1 \leq \tilde{x}_2 \leq \dots \leq \tilde{x}_n$. Then all the entries split the real line into $n+1$ intervals $\{\mathcal{I}_i\}_{i=0}^n$, where $\mathcal{I}_i \triangleq [\tilde{x}_i, \tilde{x}_{i+1}]$. Furthermore, we can obtain some insights from (15) to optimize α . Specifically, if $\frac{\alpha}{2}$ falls into a certain interval \mathcal{I}_i , the corresponding optimal \mathbf{s} can be determined as

$$\{s_k\}_{k=1}^{i-1} = \begin{cases} 0 & \alpha > 0 \\ 1 & \alpha < 0, \end{cases} \quad \{s_k\}_{k=i}^n = \begin{cases} 1 & \alpha > 0 \\ 0 & \alpha < 0. \end{cases} \tag{25}$$

Therefore, the objective function in (13) can be rewritten as (26) at the top of this page. Note that within each interval $\mathcal{R}_i = [2\tilde{x}_i, 2\tilde{x}_{i+1}]$, the objective function is a quadratic function in terms of α , and hence it is easy to give the optimal solution for α in Proposition 1.

APPENDIX B PROOF OF LEMMA 1

We prove Lemma 1 by contradictory. Since in each interval \mathcal{R}_i the objective function is a quadratic function of α . The optimal α^* can only be obtained at the two endpoints of \mathcal{R}_i or at the axis of symmetry if the objective is not monotonic in \mathcal{R}_i . When $\alpha < 0$, the axis of symmetry of the quadratic function is given by

$$\bar{x}_i = \frac{\sum_{j=1}^i \tilde{x}_j}{i}, \tag{27}$$

which is the mean value of the first i entries in $\tilde{\mathbf{x}}$.

A hypothesis is firstly made that a certain endpoint \tilde{x}_i is the optimal solution to α . It means that the axis of symmetry of the objective function in \mathcal{R}_{i-1} is on the right hand side of \tilde{x}_i , and the axis of symmetry of the objective function in \mathcal{R}_i is on the left hand side of \tilde{x}_i , i.e.,

$$\bar{x}_i < \tilde{x}_i < \bar{x}_{i-1}. \tag{28}$$

Note that the entries in $\tilde{\mathbf{x}}$ are ordered in the ascending order. Hence, \bar{x}_i , as the mean value of the first i entries in $\tilde{\mathbf{x}}$, is an increasing function with respect to i , i.e., $\bar{x}_i \geq \bar{x}_{i-1}$, which is contradictory with (28) and completes the proof for $\alpha < 0$. The scenario of $\alpha > 0$ can be similarly proved.

REFERENCES

- [1] 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.
- [2] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, Jr., "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.
- [3] 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.
- [4] X. Yu, J. Zhang, and K. B. Letaief, "Alternating minimization for hybrid precoding in multiuser OFDM mmWave systems," in *Proc. Asilomar Conf. Signals Syst. Comput.*, Pacific Grove, CA, USA, Nov. 2016, pp. 281–285.
- [5] X. Gao, L. Dai, S. Han, C.-L. I, and R. W. Heath, Jr., "Energy-efficient hybrid analog and digital precoding for mmwave MIMO systems with large antenna arrays," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 998–1009, Apr. 2016.
- [6] A. M. Abbosh, "Broadband fixed phase shifters," *IEEE Microw. Compon. Lett.*, vol. 21, no. 1, pp. 22–24, Jan. 2011.
- [7] J. Mirza, B. Ali, S. S. Naqvi, and S. Saleem, "Hybrid precoding via successive refinement for millimeter wave mimo communication systems," *IEEE Commun. Lett.*, vol. 21, no. 5, pp. 991–994, May 2017.
- [8] L. Liang, W. Xu, and X. Dong, "Low-complexity hybrid precoding in massive multiuser MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 3, no. 6, pp. 653–656, Dec. 2014.
- [9] R. Méndez-Rial, C. Rusu, N. González-Prelcic, A. Alkhateeb, and R. W. Heath, Jr., "Hybrid MIMO architectures for millimeter wave communications: Phase shifters or switches?" *IEEE Access*, vol. 4, pp. 247–267, Jan. 2016.
- [10] T. E. Bogale, L. B. Le, A. Haghighat, and L. Vandendorpe, "On the number of RF chains and phase shifters, and scheduling design with hybrid analog-digital beamforming," *IEEE Trans. Wireless Commun.*, vol. 15, no. 5, pp. 3311–3326, May 2016.
- [11] J.-C. Chen, "Hybrid beamforming with discrete phase shifters for millimeter wave massive MIMO systems," *IEEE Trans. Veh. Technol.*, to appear.
- [12] Z. Feng, S. Fu, T. Ming, and D. Liu, "Multichannel continuously tunable microwave phase shifter with capability of frequency doubling," *IEEE Photon. J.*, vol. 6, no. 1, pp. 1–8, Feb. 2014.
- [13] F. Sohrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale antenna arrays," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 501–513, Apr. 2016.
- [14] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2012.
- [15] L. Grippo and M. Sciandrone, "On the convergence of the block nonlinear gauss–seidel method under convex constraints," *Operat. Res. Lett.*, vol. 26, no. 3, pp. 127–136, Mar. 2000.