

# Energy-Efficient Hybrid Symbol-Level Precoding for Large-Scale mmWave Multiuser MIMO Systems

Alireza Haqiqatnejad<sup>1</sup>, Graduate Student Member, IEEE, Farbod Kayhan, and Björn Ottersten<sup>2</sup>, Fellow, IEEE

**Abstract**— We address the symbol-level precoding design problem for the downlink of a multiuser millimeter wave (mmWave) multiple-input multiple-output (MIMO) wireless system where the transmitter is equipped with a large-scale antenna array. The high cost and power consumption associated with the massive use of radio frequency (RF) chains prohibit fully-digital implementation of the precoder, and therefore, we consider a hybrid analog-digital architecture where a small-sized baseband precoder is followed by two successive networks of analog on-off switches and variable phase shifters according to a fully-connected structure. We jointly optimize the digital baseband precoder and the states of the switching network on a symbol-level basis, i.e., by exploiting both the channel state information (CSI) and the instantaneous data symbols, whereas the phase-shifting network is designed only based on the CSI due to practical considerations. Our approach to this joint optimization is to minimize the Euclidean distance between the optimal fully-digital and the hybrid symbol-level precoders. Remarkably, the use of a switching network allows for power-savings in the analog precoder by switching some of the phase shifters off according to the instantaneously optimized states of the switches. Our numerical results indicate that, on average, up to 50 percent of the phase shifters can be switched off. We provide an analysis of energy efficiency by adopting appropriate power dissipation models for the analog precoder, where it is shown that the energy efficiency of precoding can substantially be improved thanks to the phase shifter selection approach, compared to the fully-digital and the state-of-the-art hybrid symbol-level schemes.

**Index Terms**— Energy efficiency, hybrid analog-digital precoding, massive multiuser MIMO, mmWave communications, phase shifter selection, symbol-level precoding.

## I. INTRODUCTION

MILLIMETER wave (mmWave) communication has been widely accepted as a prime technology for

Manuscript received June 15, 2020; revised November 8, 2020 and January 6, 2021; accepted February 4, 2021. Date of publication February 12, 2021; date of current version May 18, 2021. This work was supported by the Luxembourg National Research Fund (FNR) under CORE Junior project: C16/IS/11332341 Enhanced Signal Space Optimization for satellite communication Systems (ESSTIMS). This article was presented in part at the IEEE Global Communications Conference (GLOBECOM), Taipei, Taiwan, 2020. The associate editor coordinating the review of this article and approving it for publication was S. Ma. (Corresponding author: Alireza Haqiqatnejad.)

Alireza Haqiqatnejad and Björn Ottersten are with the Interdisciplinary Centre for Security, Reliability and Trust (SnT), University of Luxembourg, 1855 Luxembourg City, Luxembourg (e-mail: alireza.haqiqatnejad@uni.lu; bjorn.ottersten@uni.lu).

Farbod Kayhan was with the Interdisciplinary Centre for Security, Reliability and Trust (SnT), University of Luxembourg, 1855 Luxembourg City, Luxembourg. He is now with the Institute for Communication Systems, University of Surrey, Guildford GU2 7XH, U.K. (e-mail: f.kayhan@surrey.ac.uk).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TCOMM.2021.3058967>.

Digital Object Identifier 10.1109/TCOMM.2021.3058967

the emerging outdoor/indoor wireless communication deployments, enabling multi-gigabit-per-second data rates thanks to the enormously available unregulated spectrum resources within 30-300 GHz band [1]–[3]. Communication in the mmWave band, however, suffers from an order-of-magnitude increase in the free-space path loss, higher shadow fading, and more severe penetration losses compared to the legacy lower-frequency systems [4]. On the other hand, the shorter wavelength of mmWave signals makes it possible to pack a larger number of antenna elements in the same physical dimension, allowing for large-scale spatial multiplexing and highly directional beamforming. Employing large antenna arrays, commonly known as massive multiple-input multiple-output (MIMO), can further provide considerable beamforming gain to compensate for severe propagation losses at mmWave frequencies [5], which is indispensable to achieve high-quality communication links in mmWave systems.

In traditional MIMO systems, the convention is to perform baseband precoding fully in the digital domain, which enables modification of both the amplitudes and phases of complex signals [6], [7]. This fully-digital signal processing, however, requires one dedicated radio frequency (RF) chain per antenna element, which is challenging to implement in practical systems with large antenna arrays due to the prohibitive cost and high power consumption of mixed-signal components, especially when operating at mmWave frequencies [8]. Given mmWave massive MIMO practical constraints, the design of cost-effective low-complexity precoding implementations has become an active line of research. Various precoding schemes, mostly aimed at either simplification of or reducing the number of RF chains, have been proposed for both single-user and multiuser MIMO systems, among which we refer to analog-only beamforming using RF phase shifters [9]–[11], antenna (sub-set) selection [12], [13], quantized fully-digital precoding via low-resolution (especially one-bit) digital-to-analog converters (DAC) [14], [15], and hybrid analog-digital beamforming [5], [8], [16]–[18].

Hybrid analog-digital precoding is a cost-effective alternative to enable both multi-stream transmission and large beamforming gains via splitting the signal processing operation between the digital and analog domains. In hybrid architectures, a small-sized digital precoder is followed by a high-dimensional analog precoder which is usually implemented using RF phase shifters and/or switches [16]. Such a setup allows for employing fewer RF chains, scaling with the number of multiplexed data streams rather than the number of antennas. Specifically, in multi-user mmWave systems,

the digital precoder is so designed to mitigate the inter-user interference, whereas the analog RF precoder is used to improve the antenna array gain [19]. Nevertheless, while designing the digital precoder is straightforward, the design and implementation of the analog precoder are usually nontrivial.

For large-scale multiuser mmWave systems, the design of block-level hybrid schemes where the precoding solution solely relies on the channel state information (CSI), has been extensively addressed. However, symbol-level approaches to hybrid precoding are not yet well studied. In the latter design approach, the data information is further exploited in optimizing the precoder such that the inter-user interference constructively contributes to the received signal of each user. This has been led to the advent of constructive interference (CI) which is based on the philosophy that a noise-free received signal can be decoded correctly not necessarily when it is close enough to the target symbol, rather, as long as it lies within the correct decision region even farther away from the intended symbol [20], [21].

Symbol-level hybrid precoding design under mmWave hardware limitations has been addressed in some recent work [22]–[24]. In [22], the authors adopt a disjoint sub-optimal approach to optimize the digital and analog precoders with a focus on the analog precoder design, where different techniques are studied and compared. Power-efficient transmitter architectures, including antenna selection and analog-only, are studied for symbol-level precoding in [25], where it has been shown that the analog-only design can outperform the other schemes especially when the transmit array size is much larger than the number of UEs. An even more cost-effective hybrid structure is considered in [23] where the baseband digitally precoded signal is subject to one-bit quantization due to the use of low-cost one-bit DACs for each RF chain. This excessive constraint, however, may limit the potential gain of symbol-level baseband signal processing. The joint optimization of digital and analog symbol-level precoders is addressed in [24], where the authors exploit the symbol-based design of the phase-shifting network to achieve the performance of the fully-digital precoder. In practice, the design needs to switch between the phases states of the variable phase shifters at the symbol rate. Keeping in mind target data rates of multi-Gbps in mmWave systems, such a high phase-switching speed requirement might be prohibitive in two aspects: first, it significantly increases the power consumption in the analog circuitry, and second yet, more importantly, it might be challenging from an implementation point of view considering the current RF semiconductor technologies [26]. Among the aforementioned symbol-level precoding techniques, those proposed in [23] and [24] are more related to the scope of this work, and therefore will be considered in the paper for comparison purposes.

Analog phase shifters and switches are two key components of the mmWave systems. A wide variety of hybrid precoding architectures are essentially based on employing either phase shifters or switches, or even a combination of both where the phase-shifting network is controlled by a preceding network of switches; see, e.g., [16] and [27] where several possible

architectures are described. Employing the combination of phase-shifting and switching networks in the analog RF precoder has a two-fold advantage. On the one hand, it can provide additional degrees-of-freedom (DoF) brought by the switching network when designing the analog precoder, and on the other hand, it allows for potential power-savings through switching some of the phase shifters off. From a power consumption perspective, one further need to take into account the excessive power consumed by the switching network. For this purpose, power consumption models such as those introduced in [27] and [14] can be used. However, roughly speaking, the excessive power consumption due to the operation of switches is relatively small compared to the power reduction in the phase-shifting network. One reason is that, in general, switches consume less power than phase shifters. Furthermore, recent advances in RF circuit design have enabled the implementation of low-power high-performance switches working at mmWave frequencies, making the switching operation even more energy-efficient; see, e.g., [28]–[30]. Therefore, the use of analog switches in combination with the phase-shifting network is an attractive architecture for hybrid mmWave systems. In this line of research, hybrid implementations with the so-called phase shifter selection, where a two-state on-off switch precedes each phase shifter, have been studied for conventional block-level precoding; see [31]–[33]. For example, in [31], it has been shown that the combination of phase shifters and switches offers noticeably higher energy efficiency compared to phase shifter-only architectures, while the spectral efficiency is almost preserved. More specifically, significant power consumption reductions are possible without sacrificing the spectral efficiency even when up to 50% of the phase shifters are turned off [32]. To the best of authors' knowledge, such an approach has not been investigated so far for hybrid symbol-level precoding.

In this paper, we consider a hybrid analog-digital architecture for symbol-level precoder where the analog precoder is implemented using a network of variable phase shifters preceded by an on-off switching network of the same dimension according to a fully-connected structure. As for the analog precoder, the phase states of the phase-shifting network are designed solely based on the instantaneous CSI, i.e., they stay unchanged within the duration of one channel coherence block. On the other hand, the on-off states of the switches as well as the baseband digital precoder are jointly optimized in our design on a symbol-level basis. Our approach to this optimization is to minimize the  $\ell_2$ -norm distance between the outputs of hybrid and optimal fully-digital symbol-level precoders. For the latter precoder, we adopt a power-constrained max-min signal-to-interference-plus-noise ratio (SINR) criterion subject to user-specific CI constraints, where the CI constraints are assumed to be distance-preserving, as characterized in [34]. Accordingly, the main contributions of this paper are as follows:

- i. We exploit the notion of CI along with the phase shifter selection approach in designing the hybrid precoder. The CI-based design can improve the symbol detection performance at the receiver side, while the phase shifter selection approach brings additional DoF to the design

problem and further enables the reduction of dissipated power in the phase-shifting network. The use of on-off switches, however, makes our design problem an NP-hard binary optimization. We deal with this difficulty by transforming the original problem into a biconvex form using an equivalent continuous-domain implication of the binary constraints. Efficient sub-optimal solutions can then be obtained via a standard block coordinate descent (BCD) algorithm.

- ii. We study the convergence of the proposed hybrid precoding algorithm, where it will be shown that convergence to a stationary point is guaranteed. We further analyze the required computational complexity in the large system limit. In our analysis, we consider both the Newton complexity, i.e., the number of iterations required till the BCD algorithm converges, and the per-iteration complexity. Moreover, we show via simulation results that the BCD algorithm usually converges within a few iterations for practical values of system parameters, i.e., array size, number of RF chains, and users.
- iii. We provide an analysis of energy efficiency, incorporating both performance and power consumption, to evaluate and compare different fully-digital/hybrid precoding architectures. For this purpose, we adopt appropriate power consumption models to take into account the power dissipated by the transmitter's RF circuitry. According to this analysis, the phase shifter selection mechanism offers significant improvements in the energy efficiency of precoding by switching off up to 50 percent of the phase shifters.
- iv. Our design approach is independent of the phase-shifting precision; however, to evaluate how this affects the ultimate precoding performance, in our simulations, we consider two implementations using infinite and finite resolution phase shifters. It will be shown that implementing the phase-shifter-selection-enabled analog precoder using low-resolution phase shifters can lead to gains of tens of Mbps/Joule per user in energy efficiency, compared to the case with infinite-resolution phase shifters.

*Organization:* The rest of this paper is organized as follows. In Section II, we describe the adopted system, signal, and channel model. We begin Section III by designing the phase-shifting network. Then, we study the symbol-level precoding problem for fully-digital architecture. This is followed by the derivation of the proposed hybrid precoding algorithm and analyses of its convergence and computational complexity. In Section IV, we provide energy efficiency analysis and explain the power consumption model. Simulation results are presented and discussed in Section V. Finally, we conclude the paper in Section VI.

*Notations:* We use bold-faced uppercase and lowercase letters to represent matrices and vectors, respectively. For matrices and vectors,  $\|\cdot\|$  respectively denotes the Frobenius norm and the  $\ell_2$ -norm. For vectors,  $\preceq$  and  $\succeq$  denote elementwise inequality. Operators  $\text{diag}(\cdot)$  and  $\text{blkdiag}(\cdot)$  represent diagonal and block-diagonal matrices, and  $\text{vec}(\cdot)$  and  $\text{vec}^{-1}(\cdot)$  denote the vectorization operation and its inverse, respectively. We use  $\mathbf{I}$  to represent the identity matrix, and

use  $\mathbf{0}$  and  $\mathbf{1}$  to represent, respectively, the all-zeros and the all-ones matrices (or vectors, depending on the context) of the appropriate dimensions. Operators  $\otimes$  and  $\circ$  stand for the Kronecker product and the Hadamard product, respectively. Furthermore,  $\mathbb{P}\{\cdot\}$  and  $\mathbb{E}\{\cdot\}$  denote the probability function and the statistical expectation.

## II. SYSTEM AND CHANNEL MODEL

In this section, we describe the system and channel model considered in the paper.

### A. System and Signal Model

We consider a single-cell single-carrier mmWave multiuser MIMO system where the base station (BS) is equipped with a large-scale antenna array of  $N_t$  elements and a (typically) much smaller number of transmit RF chains, denoted by  $N_1$ . The BS simultaneously communicates independent data streams to  $N_u$  single-antenna users, each supporting single-stream transmission. The maximum number of transmitted data streams (i.e., the maximum number of users scheduled within a transmission block) is limited by the number of available RF chains at the BS, which leads to the assumption  $N_u \leq N_1 < N_t$ . Due to the limited number of transmit RF chains, the fully-digital implementation of multiuser precoder is not possible, and therefore, a hybrid digital-analog architecture is employed where the digital baseband precoder is followed by the RF chains and an analog RF precoder, as shown in Fig. 1. It is worth noting that the baseband precoder is capable of modifying both the amplitudes and phases of the input symbols while the RF precoder adjusts only the phases of the upconverted RF signals.

1) *Digital Baseband Precoder:* We consider a (non-linear) symbol-level baseband precoder that calculates the digital outputs specifically for every set of input symbols. Accordingly, the discrete-time  $N_u \times 1$  complex modulated symbol vector  $\mathbf{s} = [s_1, s_2, \dots, s_{N_u}]^T$ , where  $\mathbb{E}\{\mathbf{s}\mathbf{s}^H\} = \mathbf{I}_{N_u}$ , is preprocessed in the digital domain using the symbol-level precoder, resulting in the output baseband signal  $\mathbf{u}_{\text{BB}} \in \mathbb{C}^{N_1 \times 1}$ . In contrast to linear precoding schemes, the nonlinear-precoded signal  $\mathbf{u}_{\text{BB}}$  is directly designed and thus may not be explicitly decomposable as a linear combination of the users' precoding vectors. This symbol-by-symbol processing will be described in more detail in Section VI. The baseband precoded signal  $\mathbf{u}_{\text{BB}}$  is then passed through the RF chains for upconversion to the carrier frequency.

2) *Analog RF Precoder:* We assume the analog precoder to be implemented following a fully-connected structure with two successive switching and phase-shifting networks of dimension  $N_1 \times N_t$  that map  $N_1$  digital outputs to  $N_t$  precoded analog signals feeding the transmit antennas; see [16], [32], [33] where similar hybrid architectures have been considered. To be more specific, each RF chain upconverts a digitally precoded signal and feeds it to a phased-array with  $N_t$  variable phase shifters, each of which preceded by a dedicated analog on-off switch determining whether the corresponding phase shifter is active or deactivated. The outputs of the phase-shifting network

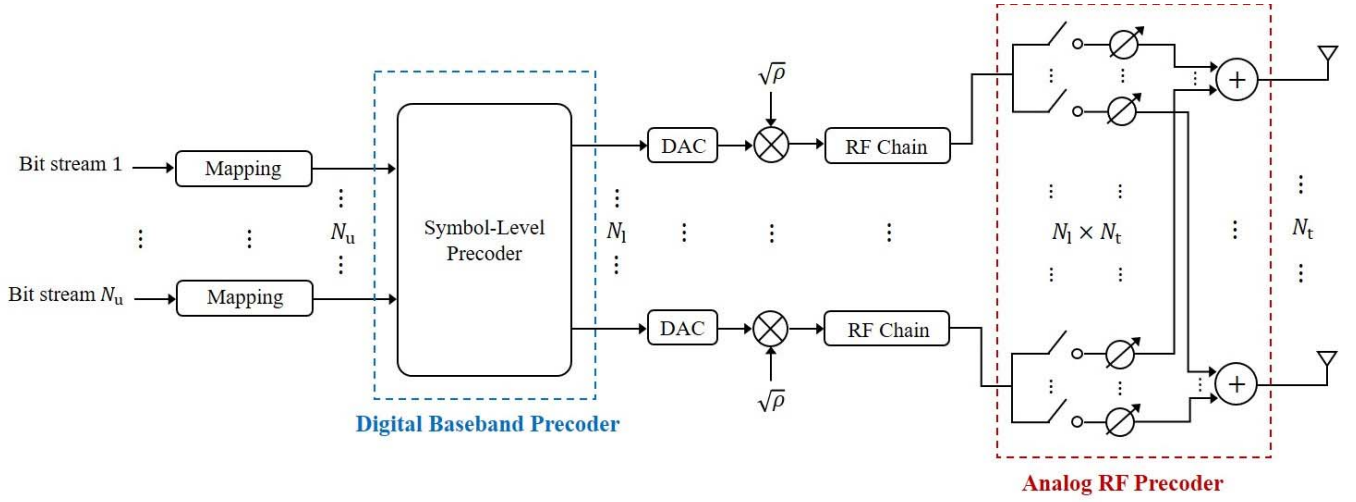


Fig. 1. The considered hybrid symbol-level precoding architecture with fully-connected switching and phase-shifting networks:  $N_1$  RF chains each connected to  $N_t$  switches and phase shifters drive an  $N_t$ -dimensional antenna array.

are then combined through  $N_t$  analog combiners before being fed to the antenna elements. Let  $\mathbf{F} \in \mathbb{C}^{N_t \times N_1}$  and  $\mathbf{T} \in \mathbb{B}$  represent the phase-shifting network and the on/off states of the switching network, respectively, where  $\mathbb{B} \triangleq \{\mathbf{X} \in \{0, 1\}^{N_t \times N_1} : \mathbf{X}\mathbf{1}_{N_1 \times 1} \succeq \mathbf{1}_{N_t \times 1}, \mathbf{X}^T \mathbf{1}_{N_t \times 1} \succeq \mathbf{1}_{N_1 \times 1}\}$ . Then, the entire RF precoder can be represented by  $\mathbf{F} \circ \mathbf{T}$ . Note that the set  $\mathbb{B}$  is defined such that the selection matrix  $\mathbf{T}$  has no all-zero row and column, where the former case corresponds to an antenna selection scheme and the latter case excludes an RF chain from the transmitter's analog circuitry, but neither is the focus of this paper. We further note that since  $\mathbf{F}$  is implemented using analog phase shifters, each element of  $\mathbf{F} \in \mathbb{C}^{N_t \times N_1}$  is normalized such that  $|f_{k,j}| = 1/\sqrt{N_t}$ , with  $|f_{k,j}|$  denoting the magnitude of the  $(k, j)$ -th element of  $\mathbf{F}$ .

Under the described system model, the vector collecting the baseband received signals for all  $N_u$  users is given by

$$\mathbf{y} = \sqrt{\rho} \mathbf{H} (\mathbf{F} \circ \mathbf{T}) \mathbf{u}_{\text{BB}} + \mathbf{z}, \quad (1)$$

where  $\mathbf{y} \in \mathbb{C}^{N_u \times 1}$  is the received signal vector,  $\rho$  is the instantaneous transmit power,  $\mathbf{H} \in \mathbb{C}^{N_u \times N_t}$  represents the mmWave multiuser channel, and  $\mathbf{z} \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Sigma})$  is a circularly symmetric complex Gaussian noise vector with  $\mathbf{\Sigma} \triangleq \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_{N_u}^2)$  where  $\sigma_j^2$  denotes the noise variance at the receiver of the  $j$ -th user. The instantaneous total transmit power is constrained by  $\rho$  through enforcing  $\|(\mathbf{F} \circ \mathbf{T}) \mathbf{u}_{\text{BB}}\|^2 = 1$ . It is further assumed that the BS has perfect knowledge of the instantaneous channel  $\mathbf{H}$ . In practical wireless systems, the CSI can be estimated at the receiver via, e.g., pilots or training sequences, and then fed back to the BS. Efficient mmWave channel estimation techniques that exploit the geometric nature of mmWave channels are presented in [35], [36]. At the receiver side, we assume that each user is capable of performing optimal single-user decoding of the received signal using the maximum likelihood (ML) detector.

### B. Multiuser mmWave Channel Model

The mmWave propagation environment is known to feature limited multipath components. To capture this sparse scattering nature, the narrowband clustered channel modeling based on the Saleh-Valenzuela model is commonly used [37]–[39]. Under this model, the channel vector corresponding to a single user is a summation over the contributions of  $N_c$  scattering clusters, with each cluster contributing  $N_p$  propagation paths between the BS and the user. Assuming the same number of clusters and scatterers to be seen by each user, the narrowband mmWave channel vector for the  $j$ -th user can be expressed as

$$\mathbf{h}_j^H = \sqrt{\frac{N_t}{N_c N_p}} \sum_{i=1}^{N_c} \sum_{l=1}^{N_p} \alpha_{j,i,l} \mathbf{a}^H(\phi_{j,i,l}, \theta_{j,i,l}), \quad (2)$$

where  $\mathbf{h}_j \in \mathbb{C}^{N_t \times 1}$  such that  $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_{N_u}]^H$ . For the  $l$ -th path in the  $i$ -th scattering cluster seen by the  $j$ -th user,  $\alpha_{j,i,l} \sim \mathcal{CN}(0, 1)$  denotes the circularly symmetric complex Gaussian gain of the path (i.e., the small-scale fading component),  $\phi_{j,i,l}$  and  $\theta_{j,i,l}$  are respectively the azimuth and elevation angles of departure (AoD), and  $\mathbf{a}(\phi_{j,i,l}, \theta_{j,i,l})$  represents the normalized transmit array response vector evaluated at specific azimuth and elevation angles  $\phi_{j,i,l}$  and  $\theta_{j,i,l}$ . The array response vector further depends on the array geometry. For uniform linear array (ULA), where the antenna elements are linearly and equally spaced, the array response vector is independent of the elevation angles  $\theta_{j,i,l}$  and follows the Vandermonde structure given by

$$\mathbf{a}(\phi_{j,i,l}) = \frac{1}{\sqrt{N_t}} \left[ 1, e^{j\frac{2\pi}{\lambda} d \sin(\phi_{j,i,l})}, \dots, e^{j(N_t-1)\frac{2\pi}{\lambda} d \sin(\phi_{j,i,l})} \right]^T, \quad (3)$$

where  $\lambda$  and  $d$  respectively denote the signal wavelength and the inter-element antenna spacing, and  $j = \sqrt{-1}$ . Note that the elevation angles  $\theta_{j,i,l}$  appear in the array response vector in case a uniform planar array (UPA) structure is employed [9].

The variance of the path gains  $\alpha_{j,i,l}$  and the normalization constant  $\sqrt{N_t/(N_c N_p)}$  are set such that  $\mathbb{E}\{\|\mathbf{H}\|_F^2\} = N_t N_u$ .

### III. HYBRID SYMBOL-LEVEL PRECODER DESIGN

We start by designing the analog phase-shifting network. The matrix  $\mathbf{F}$  representing the phase shifters' angles is usually considered to be solely dependent on the aggregate channel  $\mathbf{H}$ . Here, we adopt an analog design based on the singular value decomposition (SVD) of  $\mathbf{H}$ , which can be expressed as

$$\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H, \quad (4)$$

where  $\mathbf{\Sigma}$  is an  $N_u \times N_t$  rectangular diagonal matrix with the singular values on the diagonal in a descending order, and  $\mathbf{U}$  and  $\mathbf{V} \triangleq [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{N_t}]$  are respectively  $N_u \times N_u$  and  $N_t \times N_t$  unitary matrices with the columns representing the left and the right singular vectors. We align the angles of the phase-shifting network to those of the first  $N_1$  right singular vectors of  $\mathbf{H}$ , i.e.,  $\{\mathbf{v}_1, \dots, \mathbf{v}_{N_1}\}$ , with an element-wise normalization due to the constant modulus constraint of the phase shifters. Accordingly, denoting by  $f_{k,j}$  the  $k$ -th entry in the  $j$ -th column of  $\mathbf{F}$ , we set

$$f_{k,j} = \frac{1}{\sqrt{N_t}} e^{j\varphi_{k,j}}, \quad \forall k \in \{1, 2, \dots, N_t\}, \quad \forall j \in \{1, 2, \dots, N_1\}, \quad (5)$$

where  $\varphi_{k,j}$  is the phase of the  $k$ -th element in  $\mathbf{v}_j$ . Aligning the angles of the phase-shifting network according to the first  $N_1$  right singular vectors of  $\mathbf{H}$  enables the system to achieve larger array gains. Note that similar aligning schemes based on the SVD decomposition of the channel are used in, e.g., [23], [32]. Although infinite-resolution phase shifters are required for an accurate implementation of this approach, in practice, the use of finite-resolution phase shifters is preferred due to practical constraints of variable phase shifters, particularly in systems with large-scale antenna arrays as the number of phase shifters is proportional to the number of antenna elements. Therefore, in a more realistic implementation with discrete phase shifters, the phase states are quantized up to (typically) low bits of precision. We assume a quantization rule such that the phase of each entry of  $\mathbf{F}$  is mapped to the nearest phase value in the discrete set  $\{2m\pi/2^{b_{\text{PS}}}\} : m = 0, 1, \dots, 2^{b_{\text{PS}}} - 1\}$ . Accordingly, the quantized phase of  $f_{k,j}$ , denoted by  $\hat{\varphi}_{k,j}$ , can be obtained as

$$\hat{\varphi}_{k,j} = \frac{2\hat{m}\pi}{2^{b_{\text{PS}}}}, \quad \hat{m} = \underset{m \in \{0, 1, \dots, 2^{b_{\text{PS}}} - 1\}}{\operatorname{argmin}} \left| \varphi_{k,j} - \frac{2m\pi}{2^{b_{\text{PS}}}} \right|, \quad (6)$$

with  $b_{\text{PS}}$  denoting the number of phase shifter's resolution bits. Although our design process is independent of the precision of the entries of  $\mathbf{F}$ , we investigate in Section V the performance of the proposed hybrid precoding scheme for both finite-resolution and discrete phase shifters.

Accordingly, for given symbol vector  $\mathbf{s}$ , the channel matrix  $\mathbf{H}$  and the phase-shifting network matrix  $\mathbf{F}$ , our design objective is to jointly and instantaneously (i.e., on a symbol-level basis) optimize the digitally precoded signal  $\mathbf{u}_{\text{BB}}$  as well as the states of the switching network, represented by  $\mathbf{T}$ . To this end, we first study the symbol-level precoding solution

corresponding to a fully-digital architecture which will be of essential use in our subsequent elaboration of the hybrid precoding approach.

#### A. Optimal Fully-Digital Precoder

Let consider a fully-digital architecture for the symbol-level precoder where each antenna element is driven by a dedicated RF chain, i.e.,  $N_u \leq N_1 = N_t$ . The optimization problem of the symbol-level precoder corresponding to a power-constrained max-min fair design criterion subject to constructive interference (CI) constraint for each user is given in [40] as

$$\begin{aligned} & \max_{\mathbf{u}_{\text{FD}}, \mathbf{d} \geq \mathbf{0}} \min(\mathbf{d}) \\ & \text{s.t. } \sqrt{\rho} \bar{\mathbf{H}} \mathbf{u}_{\text{FD}} = \bar{\mathbf{\Sigma}} \bar{\mathbf{s}} + \mathbf{A}^{-1} \mathbf{W} \mathbf{d}, \\ & \quad \bar{\mathbf{u}}_{\text{FD}}^T \bar{\mathbf{u}}_{\text{FD}} \leq 1, \end{aligned} \quad (7)$$

where  $\mathbf{u}_{\text{FD}} \in \mathbb{C}^{N_t \times 1}$  denotes the fully-digital precoded signal and  $\min(\cdot)$  denotes element-wise minimum. Furthermore, the following real-valued definitions are used:  $\bar{\mathbf{u}}_{\text{FD}} \triangleq [\operatorname{Re}(\mathbf{u}_{\text{FD}})^T, \operatorname{Im}(\mathbf{u}_{\text{FD}})^T]^T$ ,  $\bar{\mathbf{H}} \triangleq [\bar{\mathbf{H}}_1^T, \bar{\mathbf{H}}_2^T, \dots, \bar{\mathbf{H}}_{N_u}^T]^T$  with

$$\bar{\mathbf{H}}_j \triangleq \begin{bmatrix} \operatorname{Re}(\mathbf{h}_j^T) & -\operatorname{Im}(\mathbf{h}_j^T) \\ \operatorname{Im}(\mathbf{h}_j^T) & \operatorname{Re}(\mathbf{h}_j^T) \end{bmatrix}, \quad j = 1, 2, \dots, N_u,$$

and  $\mathbf{A} \triangleq \operatorname{blkdiag}(\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_{N_u})$  with  $\mathbf{A}_j = [\mathbf{a}_{j,1}, \mathbf{a}_{j,2}]^T$  denoting an invertible  $2 \times 2$  matrix containing the normal vectors of the decision region boundaries associated with  $s_j$ ;  $\bar{\mathbf{\Sigma}} \triangleq \operatorname{diag}(\sigma_1, \sigma_2, \dots, \sigma_{N_u}) \otimes \mathbf{I}_2$ ;  $\bar{\mathbf{s}} \triangleq [\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_{N_u}]^T$  with  $\mathbf{s}_j \triangleq [\operatorname{Re}(s_j), \operatorname{Im}(s_j)]^T$ ;  $\mathbf{W} \triangleq \operatorname{diag}(w_1, w_2, \dots, w_{N_u}) \otimes \mathbf{I}_2$  is a  $2N_u \times 2N_u$  diagonal matrix with  $w_j = 1$  if the intended symbol for the  $j$ -th user is an outer constellation symbol<sup>1</sup> and  $w_j = 0$  otherwise; and the  $2N_u \times 1$  vector  $\mathbf{d} \triangleq [\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_{N_u}]^T$ , with  $\mathbf{d}_j = [d_{j,1}, d_{j,2}]^T$ , collects the orthogonal distances  $d_{i,1}$  and  $d_{i,2}$  of the noise-free received symbol  $\sqrt{\rho} \bar{\mathbf{H}}_j \mathbf{u}$  from the boundaries of its corresponding CI region, for all  $j = 1, 2, \dots, N_u$ .

The design formulation (7) uses the so-called distance-preserving CI regions, as introduced in [34], to exploit the CI at the receivers. By definition, any two points belonging to the distance-preserving CI regions of two different constellation points are distanced by at least the distance between the two constellation points. An illustrative example of distance-preserving CI regions is shown in Fig. 2 for a QPSK constellation. Accordingly, the CI constraint  $\sqrt{\rho} \bar{\mathbf{H}} \mathbf{u}_{\text{FD}} = \bar{\mathbf{\Sigma}} \bar{\mathbf{s}} + \mathbf{A}^{-1} \mathbf{W} \mathbf{d}$  in (7) attempts to instantaneously align the multiuser interference so that the noise-free signal received by each user locates within the distance-preserving CI region that corresponds to its intended symbol.

By introducing a slack variable  $\gamma$  and applying the change of variable  $\mathbf{d} \rightarrow \mathbf{t} + \gamma \mathbf{1}_{2N_u \times 1}$ , the optimization problem (7) can be equivalently written as

$$\begin{aligned} & \max_{\bar{\mathbf{u}}_{\text{FD}}, \mathbf{t} \geq \mathbf{0}, \gamma \geq 0} \gamma \\ & \text{s.t. } \sqrt{\rho} \bar{\mathbf{H}} \bar{\mathbf{u}}_{\text{FD}} = \bar{\mathbf{\Sigma}} \bar{\mathbf{s}} + \mathbf{A}^{-1} \mathbf{W} (\mathbf{t} + \gamma \mathbf{1}_{2N_u \times 1}), \\ & \quad \bar{\mathbf{u}}_{\text{FD}}^T \bar{\mathbf{u}}_{\text{FD}} \leq 1. \end{aligned} \quad (8)$$

<sup>1</sup>A symbol that resides on the convex hull of the constellation set is referred to as outer constellation symbol.

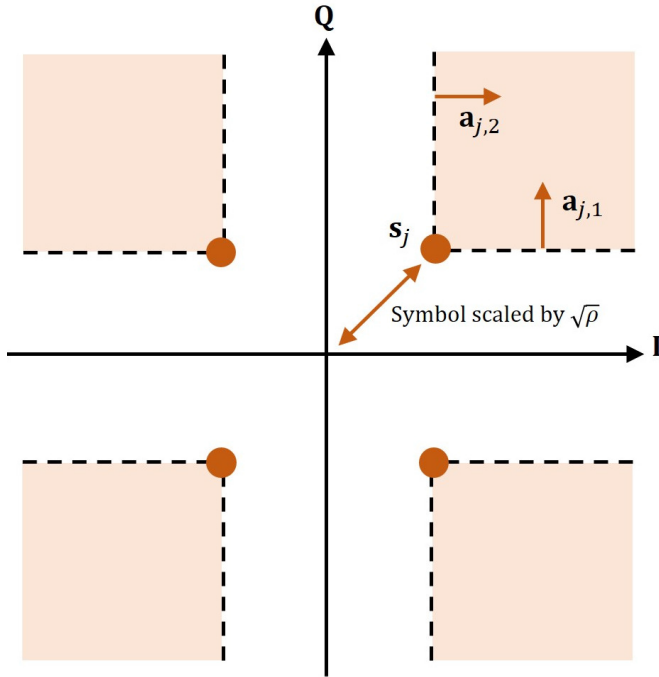


Fig. 2. Distance-preserving regions, depicted as orange areas, for a QPSK constellation.

Given  $\bar{\mathbf{u}}_{\text{FD}}$  and  $\mathbf{t}$ , maximization of (8) over  $\gamma$  is amenable to a provably positive solution, given in closed-form by

$$\gamma^* = \frac{1}{\|\mathbf{q}\|^2} \mathbf{q}^T (\sqrt{\rho} \bar{\mathbf{H}} \bar{\mathbf{u}}_{\text{FD}} - \bar{\Sigma} \bar{\mathbf{s}} - \mathbf{A}^{-1} \mathbf{W} \mathbf{t}), \quad (9)$$

where  $\mathbf{q} \triangleq \mathbf{A}^{-1} \mathbf{W} \mathbf{1}_{2N_u \times 1}$ . As a result, we can eliminate the decision variable  $\gamma$  from our design formulation by replacing  $\gamma$  in (8) with the closed-form expression for  $\gamma^*$  in (9). Denoting  $\mathbf{Q} \triangleq \|\mathbf{q}\|^2 \mathbf{I}_{N_u} - \mathbf{q} \mathbf{q}^T$ , after some straightforward algebraic steps, we can rewrite the optimization problem (8) in a convex form as

$$\begin{aligned} & \max_{\bar{\mathbf{u}}_{\text{FD}}, \mathbf{t} \geq \mathbf{0}} \mathbf{q}^T (\sqrt{\rho} \bar{\mathbf{H}} \bar{\mathbf{u}}_{\text{FD}} - \mathbf{A}^{-1} \mathbf{W} \mathbf{t}) \\ & \text{s.t. } \mathbf{Q} (\sqrt{\rho} \bar{\mathbf{H}} \bar{\mathbf{u}}_{\text{FD}} - \mathbf{A}^{-1} \mathbf{W} \mathbf{t} - \bar{\Sigma} \bar{\mathbf{s}}) = \mathbf{0}, \\ & \quad \bar{\mathbf{u}}_{\text{FD}}^T \bar{\mathbf{u}}_{\text{FD}} \leq 1, \end{aligned} \quad (10)$$

which can efficiently be solved via several off-the-shelf algorithms [41]. The optimal fully-digital precoded signal obtained from (10) is in fact a performance upper bound which can be achieved by the symbol-level precoder in case the number of BS's RF chains is equal to  $N_t$ . Therefore, we use this optimal yet impractical solution in developing our hybrid symbol-level precoding algorithm and also as a performance benchmark for comparisons in Section V.

### B. Hybrid Precoder With Phase Shifter Selection

We use the optimal fully-digital precoded signal in order to design the hybrid symbol-level precoder. More specifically, denoting by  $\mathbf{u}_{\text{FD}}^*$  the optimal solution to (10), we aim to find the digital-domain precoded signal  $\mathbf{u}_{\text{BB}}$  and the selection matrix  $\mathbf{T}$  such that the output of the hybrid precoder,

i.e.,  $(\mathbf{F} \circ \mathbf{T}) \mathbf{u}_{\text{BB}}$  is as close as possible to  $\mathbf{u}_{\text{FD}}^*$ . The corresponding optimization problem is therefore given by

$$\min_{\mathbf{u}_{\text{BB}}, \mathbf{T} \in \mathcal{B}} \|(\mathbf{F} \circ \mathbf{T}) \mathbf{u}_{\text{BB}} - \mathbf{u}_{\text{FD}}^*\|^2 \quad \text{s.t. } \|(\mathbf{F} \circ \mathbf{T}) \mathbf{u}_{\text{BB}}\|^2 = 1. \quad (11)$$

To proceed, by defining  $\mathbf{g} \triangleq \text{vec}(\mathbf{G})$  where  $\mathbf{G} \triangleq 2\mathbf{T} - \mathbf{1}_{N_t \times N_t}$ , we recast (11) in an equivalent form which is more convenient for our later use, i.e.,

$$\begin{aligned} & \min_{\mathbf{u}_{\text{BB}}, \mathbf{g}} \|(\bar{\mathbf{u}}_{\text{BB}}^T \otimes \mathbf{I}_{N_t}) \text{diag}(\text{vec}(\mathbf{F})) \mathbf{g} + \mathbf{F} \mathbf{u}_{\text{BB}} - 2 \mathbf{u}_{\text{FD}}^*\|^2 \\ & \text{s.t. } \frac{1}{4} \|(\bar{\mathbf{u}}_{\text{BB}}^T \otimes \mathbf{I}_{N_t}) \text{diag}(\text{vec}(\mathbf{F})) \mathbf{g} + \mathbf{F} \mathbf{u}_{\text{BB}}\|^2 = 1, \\ & \quad \mathbf{g} \in \left\{ \{-1, +1\}^{N_t N_t \times 1} \cap \bar{\mathcal{B}} \right\}, \end{aligned} \quad (12)$$

where

$$\begin{aligned} \bar{\mathcal{B}} \triangleq & \left\{ \mathbf{z} \in \mathbb{R}^{N_t N_t \times 1} : \right. \\ & \frac{1}{2} (\mathbf{1}_{N_t \times 1}^T \otimes \mathbf{I}_{N_t}) (\mathbf{z} + \mathbf{1}_{N_t N_t \times 1}) \succeq \mathbf{1}_{N_t \times 1}, \\ & \left. \frac{1}{2} (\mathbf{1}_{N_t \times 1}^T \otimes \mathbf{I}_{N_t}) (\mathbf{z} + \mathbf{1}_{N_t N_t \times 1}) \succeq \mathbf{1}_{N_t \times 1} \right\}. \end{aligned}$$

The new formulation in (12) is derived by using the well-known property  $\text{vec}(\mathbf{X} \mathbf{Y} \mathbf{Z}) = (\mathbf{Z}^T \otimes \mathbf{X}) \text{vec}(\mathbf{Y})$  for given matrices  $\mathbf{X}$ ,  $\mathbf{Y}$ ,  $\mathbf{Z}$ , along with the fact that  $\mathbf{T} = (\mathbf{G} + \mathbf{1}_{N_t \times N_t})/2$ . Problem (12) belongs to the class of minimization of quadratic forms over binary vectors (i.e., the binary constraints on the elements of  $\mathbf{g}$ ), which is known to be NP-hard in general [42]. To tackle this difficulty, we use an equivalent implication of the binary constraints given in a biconvex form.

*Lemma 1:* Let  $\mathbf{g}$  and  $\mathbf{e}$  be two real-valued vectors of equal length  $N_t N_t$ . Provided that  $-\mathbf{1} \preceq \mathbf{g} \preceq \mathbf{1}$  and  $\|\mathbf{e}\|^2 \leq N_t N_t$ , the condition  $\mathbf{g}^T \mathbf{e} = N_t N_t$  implies that  $\mathbf{g} = \mathbf{e}$  and further  $\mathbf{g} \in \{-1, +1\}^{N_t N_t}$ .

*Proof:* See [42].  $\square$

Using Lemma 1, we can rewrite problem (12) in an equivalent form with all the optimization variables being taken from continuous domains, i.e.,

$$\begin{aligned} & \min_{\mathbf{u}_{\text{BB}}, \mathbf{g}, \mathbf{e}} \|(\bar{\mathbf{u}}_{\text{BB}}^T \otimes \mathbf{I}_{N_t}) \text{diag}(\text{vec}(\mathbf{F})) \mathbf{g} + \mathbf{F} \mathbf{u}_{\text{BB}} - 2 \mathbf{u}_{\text{FD}}^*\|^2 \\ & \text{s.t. } \frac{1}{4} \|(\bar{\mathbf{u}}_{\text{BB}}^T \otimes \mathbf{I}_{N_t}) \text{diag}(\text{vec}(\mathbf{F})) \mathbf{g} + \mathbf{F} \mathbf{u}_{\text{BB}}\|^2 = 1, \\ & \quad -\mathbf{1} \preceq \mathbf{g} \preceq \mathbf{1}, \quad \mathbf{g} \in \bar{\mathcal{B}}, \\ & \quad \mathbf{g}^T \mathbf{e} = N_t N_t, \quad \mathbf{e}^T \mathbf{e} \leq N_t N_t, \end{aligned} \quad (13)$$

where  $\mathbf{g}^T \mathbf{e} = N_t N_t$  is often called the equilibrium constraint. Reformulation (13) is still a non-convex problem due to the biconvex equilibrium constraint. We use a well-known approach, namely, the exact penalty method, to efficiently solve (13). The interested readers are referred to [42] and [43] where studies on the accuracy and convergence characteristics of the exact penalty method are provided.

Based on the exact penalty method, the biconvex equilibrium constraint  $\mathbf{g}^T \mathbf{e} = N_t N_t$  can be handled by adding the difference  $N_t N_t - \mathbf{g}^T \mathbf{e}$  multiplied by  $\mu > 0$  as a penalty function to the objective function, where the difference  $N_t N_t - \mathbf{g}^T \mathbf{e}$  acts as a measure of deviation from the equilibrium constraint. Accordingly, denoting the objective function

of (13) by  $g(\mathbf{u}_{\text{BB}}, \mathbf{t})$ , we can write

$$\begin{aligned} & \min_{\mathbf{u}_{\text{BB}}, \mathbf{g}, \mathbf{e}} g(\mathbf{u}_{\text{BB}}, \mathbf{g}) + \mu (N_t N_1 - \mathbf{g}^T \mathbf{e}) \\ & \text{s.t. } \frac{1}{4} \left\| (\mathbf{u}_{\text{BB}}^T \otimes \mathbf{I}_{N_t}) \text{diag}(\text{vec}(\mathbf{F})) \mathbf{g} + \mathbf{F} \mathbf{u}_{\text{BB}} \right\|^2 = 1, \\ & \quad -\mathbf{1} \preceq \mathbf{g} \preceq \mathbf{1}, \quad \mathbf{g} \in \bar{\mathbb{B}}, \quad \mathbf{e}^T \mathbf{e} \leq N_t N_1, \end{aligned} \quad (14)$$

which is our final formulation for the proposed hybrid symbol-level precoding design. It is important to note that, in general, problems (14) and (13) are not equivalent. However, by monotonically increasing the penalty parameter  $\mu$  in each iteration up to a certain threshold, successive solutions of the penalized problem (13) eventually converge to the solution of the original biconvex problem. On the other hand, for a given  $\mathbf{u}_{\text{BB}}$ , it can be shown that if  $g(\mathbf{u}_{\text{BB}}, \mathbf{g})$  is an  $L$ -Lipschitz continuous convex function on  $-\mathbf{1} \preceq \mathbf{g} \preceq \mathbf{1}$ , problem (14) has the same local and global minima as those of (13) for  $\mu \geq 2L$ , where  $L$  denotes the Lipschitz constant of  $g(\mathbf{u}_{\text{BB}}, \mathbf{g})$  with respect to  $\mathbf{g}$ ; see [42, Th. 1]. In the following lemma, we show that function  $g(\mathbf{u}_{\text{BB}}, \mathbf{g})$  is Lipschitz continuous on the domain  $-\mathbf{1} \preceq \mathbf{g} \preceq \mathbf{1}$ .

*Lemma 2:* Let  $\mathbf{u}_{\text{BB}}$  be given, then  $g(\mathbf{u}_{\text{BB}}, \mathbf{g})$  is a Lipschitz continuous function on  $-\mathbf{1} \preceq \mathbf{g} \preceq \mathbf{1}$  with Lipschitz constant

$$\begin{aligned} L &= 2\sqrt{N_t N_1} \left\| (\mathbf{u}_{\text{BB}}^T \otimes \mathbf{I}_{N_t}) \text{diag}(\text{vec}(\mathbf{F})) \right\|_{\text{F}}^2 \\ & \quad + 2 \left\| ((\mathbf{u}_{\text{BB}}^T \otimes \mathbf{I}_{N_t}) \text{diag}(\text{vec}(\mathbf{F})))^H (\mathbf{F} \mathbf{u}_{\text{BB}} - 2 \mathbf{u}_{\text{FD}}^*) \right\|. \end{aligned} \quad (15)$$

*Proof:* See Appendix A.  $\square$

Finally, we exploit the fact that the objective function of the minimization problem (14), i.e.,  $g(\mathbf{u}_{\text{BB}}, \mathbf{g}) + \mu (N_t N_1 - \mathbf{g}^T \mathbf{e})$  is a biconvex quadratic function in  $\mathbf{g}$  and  $\mathbf{e}$ , i.e., fixing either  $\mathbf{g}$  or  $\mathbf{e}$  gives a convex function in the other variable. From Lemma 2, it then follows that one can use a standard block coordinate descent (BCD) algorithm to find at least a locally optimal solution to problem (13), where a coordinate block refers to either of the vectors  $\mathbf{u}_{\text{BB}}$ ,  $\mathbf{g}$  or  $\mathbf{e}$ . To be more specific, the objective function  $g(\mathbf{u}_{\text{BB}}, \mathbf{g}) + \mu (N_t N_1 - \mathbf{g}^T \mathbf{e})$  can be minimized over one of these vectors while the other two are fixed, and then, repeating the same procedure for the other two blocks. The penalty multiplier  $\mu$  should be increased monotonically every  $N$  cycles, where the Lipschitz constant  $L$  provided in Lemma 2 determines the limit for increasing  $\mu$  as a function of the other variables. Accordingly, the BCD algorithm solving (14) performs the following steps in the  $n$ -th cycle:

- i. **Updating  $\mathbf{g}$ :** Given  $\mathbf{u}_{\text{BB}}^{(n-1)}$  and  $\mathbf{e}^{(n-1)}$ , the value of  $\mathbf{g}$  in the  $n$ -th cycle is updated by solving

$$\mathbf{g}^{(n)} = \underset{-\mathbf{1} \preceq \mathbf{g} \preceq \mathbf{1}, \mathbf{g} \in \bar{\mathbb{B}}}{\text{argmin}} g(\mathbf{u}_{\text{BB}}^{(n-1)}, \mathbf{g}) + \mu (N_t N_1 - \mathbf{g}^T \mathbf{e}^{(n-1)}), \quad (16)$$

which is a linearly-constrained quadratic programming.

- ii. **Updating  $\mathbf{e}$ :** The value of  $\mathbf{e}^{(n)}$  can be obtained as the solution to the following problem:

$$\mathbf{e}^{(n)} = \underset{\|\mathbf{e}\|^2 \leq N_t N_1}{\text{argmin}} -\mathbf{e}^T \mathbf{g}^{(n)}, \quad (17)$$

---

**Algorithm 1** BCD Algorithm Solving (14)
 

---

- 1: **input:**  $\mathbf{F}$ ,  $\mathbf{u}_{\text{FD}}^*$
  - 2: **output:**  $\mathbf{u}_{\text{BB}}$ ,  $\mathbf{T}$
  - 3: **initialize:**  $\mathbf{g}^{(0)} = \mathbf{e}^{(0)} = \mathbf{1}_{N_t N_1 \times 1}$ ,  $\mathbf{u}_{\text{BB}}^{(0)} = \mathbf{0}_{N_1 \times 1}$ ,  $\mu^{(0)}, n=0$
  - 4: **set:**  $\vartheta > 1$
  - 5: **while** the terminating condition is met **do**
  - 6:    $n \leftarrow n + 1$
  - 7:   compute  $\mathbf{g}^{(n)}$  by solving (16)
  - 8:    $\mathbf{e}^{(n)} \leftarrow \sqrt{N_t N_1} \mathbf{g}^{(n)} / \|\mathbf{g}^{(n)}\|$
  - 9:   compute  $\mathbf{u}_{\text{BB}}^{(n)}$  using (20)
  - 10:   obtain  $L^{(n)}$  from (15)
  - 11:    $\mu^{(n)} \leftarrow \min\{2L^{(n)}, \vartheta \mu^{(n-1)}\}$
  - 12: **end while**
  - 13:  $\mathbf{T} \leftarrow (\text{vec}^{-1}(\mathbf{g}) + \mathbf{1}_{N_t \times N_1}) / 2$
- 

which is equivalent to a norm-constrained inner product maximization that admits a simple closed-form solution given by

$$\mathbf{e}^{(n)} = \sqrt{N_t N_1} \mathbf{g}^{(n)} / \|\mathbf{g}^{(n)}\|. \quad (18)$$

- iii. **Updating  $\mathbf{u}_{\text{BB}}$ :** Given  $\mathbf{g}^{(n)}$  and  $\mathbf{e}^{(n)}$ , the minimization problem (14) is equivalent to

$$\begin{aligned} & \min_{\mathbf{u}_{\text{BB}}} g(\mathbf{u}_{\text{BB}}, \mathbf{g}^{(n)}) \\ & \text{s.t. } \frac{1}{4} \left\| (\mathbf{u}_{\text{BB}}^T \otimes \mathbf{I}_{N_t}) \text{diag}(\text{vec}(\mathbf{F})) \mathbf{g}^{(n)} + \mathbf{F} \mathbf{u}_{\text{BB}} \right\|^2 = 1. \end{aligned} \quad (19)$$

Using the method of Lagrange multipliers, one can simply obtain the solution to (19) which is used as the  $n$ -th update of  $\mathbf{u}_{\text{BB}}$  and is given by

$$\mathbf{u}_{\text{BB}}^{(n)} = \frac{2((\mathbf{F} \circ \mathbf{G}^{(n)}) + \mathbf{F})^\dagger \mathbf{u}_{\text{FD}}^*}{\left\| ((\mathbf{F} \circ \mathbf{G}^{(n)}) + \mathbf{F}) ((\mathbf{F} \circ \mathbf{G}^{(n)}) + \mathbf{F})^\dagger \mathbf{u}_{\text{FD}}^* \right\|}, \quad (20)$$

where  $\text{vec}(\mathbf{G}^{(n)}) = \mathbf{g}^{(n)}$ , and  $(\cdot)^\dagger$  stands for the Moore-Penrose inverse.

- iv. **Updating  $\lambda$ :** Once every  $N$  cycles, the penalty parameter  $\mu$  is updated as

$$\mu^{(n)} = \min\{2L^{(n)}, \vartheta \mu^{(n-1)}\}, \quad (21)$$

where  $\vartheta > 1$  is a constant design parameter and  $L^{(n)}$  is the  $n$ -th update of the Lipschitz constant  $L$  which is computed by substituting  $\mathbf{u}_{\text{BB}}^{(n)}$  in (15).

The pseudocode of the described BCD algorithm is presented in Algorithm 1. In what follows, we analyze the convergence behavior of Algorithm 1.

### C. Convergence Analysis

The BCD algorithm is a successive optimization approach in which a certain approximate version of the objective function is optimized with respect to one block of variables at a time, while fixing the rest of block variables [44]. Let  $h(\mathbf{u}_{\text{BB}}, \mathbf{g}, \mathbf{e})$  denote the objective function of problem (14).

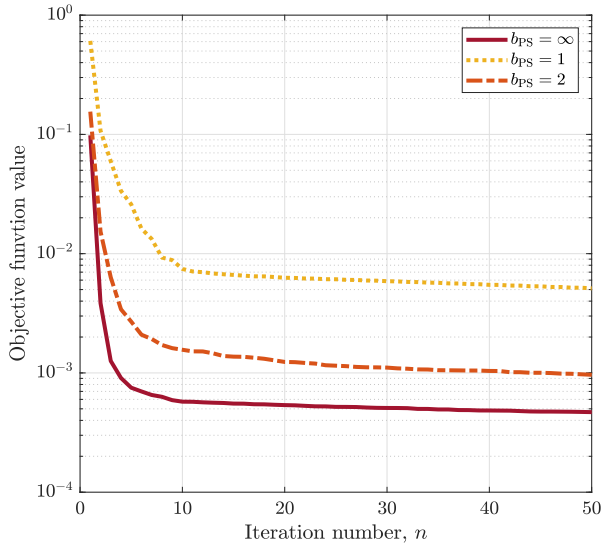


Fig. 3. Convergence behavior of Algorithm 1 as a function of iteration number for  $\mu = 10^{-4}$  and  $\vartheta = 1.1$ .

As mentioned earlier in this section, by fixing two variables among  $\mathbf{u}_{\text{BB}}$ ,  $\mathbf{g}$  and  $\mathbf{e}$ , function  $h(\mathbf{u}_{\text{BB}}, \mathbf{g}, \mathbf{e})$  becomes convex in the other variable. More precisely, the sub-problem (16) is a convex linearly-constrained quadratic programming (LCQP) which can be solved for the optimal solution. Furthermore, the sub-problems (17) and (19) are amenable to closed-form solutions, and therefore, can be solved for global optimality. This implies that, at the  $n$ -th iteration, we have

$$h(\mathbf{u}_{\text{BB}}^{(n-1)}, \mathbf{g}^{(n)}, \mathbf{e}^{(n-1)}) \leq h(\mathbf{u}_{\text{BB}}^{(n-1)}, \mathbf{g}^{(n-1)}, \mathbf{e}^{(n-1)}), \quad (22)$$

where  $\mathbf{g}^{(n)}$  denotes the  $n$ -th update of  $\mathbf{g}$ , and  $\mathbf{u}_{\text{BB}}^{(n-1)}$  and  $\mathbf{e}^{(n-1)}$  denote the updates of  $\mathbf{u}_{\text{BB}}$  and  $\mathbf{e}$  obtained from iteration  $n - 1$ . Similarly, we can write

$$h(\mathbf{u}_{\text{BB}}^{(n-1)}, \mathbf{g}^{(n)}, \mathbf{e}^{(n)}) \leq h(\mathbf{u}_{\text{BB}}^{(n-1)}, \mathbf{g}^{(n)}, \mathbf{e}^{(n-1)}), \quad (23)$$

and

$$h(\mathbf{u}_{\text{BB}}^{(n)}, \mathbf{g}^{(n)}, \mathbf{e}^{(n)}) \leq h(\mathbf{u}_{\text{BB}}^{(n-1)}, \mathbf{g}^{(n)}, \mathbf{e}^{(n)}). \quad (24)$$

As a result, the sequence of the objective function values after the update of each block is monotonically non-increasing, and therefore, convergence of Algorithm 1 to a stationary point (i.e., at least a local extremum) is guaranteed. We further note that the terminating condition for Algorithm 1 can be considered as

$$\left| h(\mathbf{u}_{\text{BB}}^{(n)}, \mathbf{g}^{(n)}, \mathbf{e}^{(n)}) - h(\mathbf{u}_{\text{BB}}^{(n-1)}, \mathbf{g}^{(n-1)}, \mathbf{e}^{(n-1)}) \right| \leq \epsilon_o, \quad (25)$$

where  $\epsilon_o$  denotes the threshold for the desired accuracy. In Fig. 3, we illustrate the convergence behavior of Algorithm 1 by plotting the value of the objective function  $h(\mathbf{u}_{\text{BB}}, \mathbf{g}, \mathbf{e})$  versus the number of outer iterations (cycles) for phase shifters with different precision bits  $b_{\text{PS}}$ , where it is shown that the proposed algorithm converges at a favorable rate. In particular, for a desired accuracy of  $\epsilon_o = 10^{-2}$ , it can be seen that, in all cases, Algorithm 1 converges in

no more than 10 iterations. It can further be seen that the algorithm shows a higher residual error for lower values of  $b_{\text{PS}}$ . This is due to the fact that discretizing the states of the phase shifters with lower number of precision bits induces a greater discontinuity in the feasible region of the optimization problem, and therefore, it may not be possible to reduce the Euclidean distance between the fully-digital and the hybrid precoders beyond a certain limit.

#### D. Analysis of Computational Complexity

Using the four-step BCD approach summarized in Algorithm 1, the overall computation cost of solving (14) in terms of the required number of arithmetic operations is composed of two main parts. The first part involves inner iterations to solve the sub-problem (16) over  $\mathbf{g}$  and updating  $\mathbf{u}_{\text{BB}}$  using (20), and the second part refers to the outer iterations (cycles) over coordinate blocks.

The computation cost of updating  $\mathbf{u}_{\text{BB}}$  via (20) is dominated by the arithmetic complexity of performing the matrix pseudo-inversion  $((\mathbf{F} \circ \mathbf{G}^{(n)}) + \mathbf{F})^\dagger$ , which is of order  $\mathcal{O}(N_t N_1^2)$  given the dimensions of  $\mathbf{F}$  and  $\mathbf{G}$ . Furthermore, to efficiently solve (16), one may use the off-the-shelf algorithms such as (accelerated) projected/proximal gradient methods [45], or quasi-Newton approaches, e.g., L-BFGS-B [46]. In particular, for a Lipschitz smooth (not necessarily strongly) convex objective function as in (14), all the aforementioned algorithms converge superlinearly at a rate of  $\mathcal{O}(1/\sqrt{\epsilon_i})$  to reach an  $\epsilon_i$ -optimal solution. For example, using the accelerated projected gradient descent algorithm, the per-iteration complexity associated with subproblem (16) is dominated by matrix multiplications of the limiting order  $N_1^2 N_t^2$ , as  $N_1, N_t \rightarrow \infty$ . Therefore, in the limiting case, the total number of operations needed to be performed in order to solve the inner subproblem (16) with an accuracy of  $\epsilon_i$  is of order  $\mathcal{O}(N_1^2 N_t^2)(1/\sqrt{\epsilon_i})$ . Putting this together with the complexity of computing  $\mathbf{u}_{\text{BB}}$ , every cycle of the BCD algorithm has a dominating complexity of  $\mathcal{O}(N_t N_1^2) + \mathcal{O}(N_1^2 N_t^2)(1/\sqrt{\epsilon_i})$ .

On the other hand, the reformulation (14), which is obtained based on the exact penalty method, is guaranteed to converge to a first-order Karush-Kuhn-Tucker (KKT) point with an accuracy of  $\epsilon_o$  in no more than  $\lceil (\ln(2L\sqrt{N_t}) - \ln(\mu^{(0)}\epsilon_o)) / \ln(\vartheta) \rceil$  iterations [42], where  $\mu^{(0)}$  is the initial value of the penalty parameter  $\mu$  and  $\lceil \cdot \rceil$  denotes the ceiling operation. To have a complete analysis of the complexity, we further need to evaluate the constant  $L$ . From (15), it is straightforward to show that

$$\begin{aligned} L &\leq 2\sqrt{N_1 N_t} \|\mathbf{u}_{\text{BB}}\|^2 \|\mathbf{F}\|_{\text{F}}^2 + 2 \|\mathbf{u}_{\text{BB}}\|^2 \|\mathbf{F}\|_{\text{F}}^2 \\ &\quad + 4 \|\mathbf{u}_{\text{BB}}\| \|\mathbf{F}\|_{\text{F}} \|\mathbf{u}_{\text{FD}}^*\|^2 \\ &= 2N_1 \sqrt{N_t N_1} \|\mathbf{u}_{\text{BB}}\|^2 + 2N_1 \|\mathbf{u}_{\text{BB}}\|^2 + 4\sqrt{N_1} \|\mathbf{u}_{\text{BB}}\|, \end{aligned} \quad (26)$$

where the equality can be justified considering the definition of matrix  $\mathbf{F}$  in (5), which yields  $\|\mathbf{F}\|_{\text{F}} = \sqrt{N_1}$ , along with the fact that  $\|\mathbf{u}_{\text{FD}}^*\| = 1$ ; see (10). It can further be verified that in the large system limit where  $N_t \rightarrow \infty$  with  $N_1 \ll N_t$ ,



we have  $\|\mathbf{u}_{\text{BB}}\| \rightarrow 1$ . Therefore, one can write

$$L \leq 2N_1\sqrt{N_t N_1} + \mathcal{O}\left(N_1 + \sqrt{N_1}\right), \quad (27)$$

As a result, in the limiting case with  $N_1 \rightarrow \infty$ , we have

$$L \leq N_1\sqrt{N_t N_1} \cdot \mathcal{O}(1). \quad (28)$$

Having the upper bound given by (28) on the dominating order of  $L$ , the worst-case computational complexity of Algorithm 1 solving the design problem (14) with accelerated inner gradient steps can be obtained as shown in Table I. In practice, however, the outer optimization usually converges in a few cycles, as we will see in Section V.

For comparison purposes, the complexities of hybrid symbol-level precoding techniques proposed in [23] and [24] is further reported in Table I. For the hybrid scheme in [23], the reported complexity order refers to the worst-case complexity of reaching an  $\epsilon_o$ -optimal solution to a linear problem via the interior-point method; see [47]. Note that, in Table I,  $\mathcal{C}_{\text{FD}}$  denotes the complexity of solving the fully-digital symbol-level precoding problem, which generally depends on the adopted optimization algorithm. This solution is used as a baseline in designing the hybrid precoder in [24] and this work. In section V, we will compare the complexities of the hybrid precoding schemes in Table I by incorporating the solutions to both the fully-digital and hybrid design problems and evaluating the corresponding execution times.

#### IV. ENERGY EFFICIENCY ANALYSIS

Hybrid precoding strategies predominantly focus on reducing hardware cost/complexity and power consumption by delegating part of the signal processing burden to the analog domain. In return, this may sacrifice the precoding performance, e.g., spectral efficiency, with respect to fully-digital systems. On the other hand, various hybrid implementations may differ from one another in their complexity and power consumption. In order to be able to compare different hybrid architectures and also to assess their efficiency versus the fully-digital alternative, one needs to incorporate both performance and complexity/power consumption aspects into one single figure of merit. A common choice is energy efficiency which can simply be expressed as the ratio between spectral efficiency and power consumption. Due to the assumption of finite-alphabet signaling, we measure the spectral efficiency in bits per symbol. Thereby, the energy efficiency of the precoding scheme, in bits per Joule, is defined as the ratio between goodput and power consumption, i.e.,

$$\eta \triangleq \frac{R(1 - P_e)}{P}, \quad (29)$$

where  $P_e \triangleq 1 - (1/N_u) \sum_{j=1}^{N_u} P_{e,j}$  is the average symbol error probability across all  $N_u$  users with  $P_{e,j}$  denoting the symbol error probability for the  $j$ -th user. The average per-user spectral efficiency  $R$  and the power consumption  $P$  are defined as follows.

#### A. Spectral Efficiency

Using an uncoded transmission scheme with finite-alphabet signaling, the communication rate towards the  $j$ -th user can be evaluated, in terms of bits per symbol per unit bandwidth, through calculating the average mutual information between the target symbol  $s_j$  and the received signal  $y_j$ , i.e.,

$$I(s_j; y_j) = \mathbb{E}_{s_j, y_j, \mathbf{H}} \left\{ \log_2 \frac{\mathbb{P}_{y_j|s_j, \mathbf{H}}(y_j|s_j, \mathbf{H})}{\mathbb{P}_{y_j|\mathbf{H}}(y_j|\mathbf{H})} \right\}. \quad (30)$$

Assuming transmission with Nyquist rate over a double-sided bandwidth of  $W$  Hz, the maximum allowable symbol rate is  $W$  symbols per second, which results in a bit rate of  $W \times I(s_j; y_j)$  for user  $j$ . Putting this together for all  $N_u$  users, the average per-user achievable rate of the downlink channel is given by

$$R = \frac{W}{N_u} \sum_{j=1}^{N_u} I(s_j; y_j). \quad (31)$$

It should be noted that deriving closed-form expressions for the conditional probability mass functions in (30) is a cumbersome task. As an alternative, one can obtain experimental probability distributions over sufficiently many independent realizations of the channel and the users' symbols to approximate the mutual information  $I(s_j; y_j)$  for each user  $j \in \{1, 2, \dots, N_u\}$ .

#### B. Power Consumption

The power dissipated by the BS's RF front-end components accounts for the power consumption at the BS. In the sequel, we first adopt power consumption models for typical components of an RF front-end and then specifically tailor the overall power consumption model according to each precoding architecture, namely, fully-digital and hybrid (with and without phase shifter selection). The transmit RF front-end of a multi-antenna system is commonly composed of one baseband processor, several RF chains, each preceded by a pair of DACs (i.e., one DAC for each I/Q channel), and power amplifiers (PA). The use of analog components such as dividers, combiners, switches, and/or phase shifters are limited to hybrid architectures.

As a rule of thumb, the power consumption of DAC scales linearly in sampling rate and exponentially in the number of bits per sample (i.e., resolution bits). We assume the DACs are of binary-weighted current-steering type [48], where its power consumption is approximately given in [49] as

$$P_{\text{DAC}} = \frac{3}{2} (2^{b_{\text{DAC}}} - 1) \times 10^{-5} + \frac{9}{2} b_{\text{DAC}} F_s \times 10^{-12}, \quad (32)$$

with  $b_{\text{DAC}}$  and  $F_s$  respectively denoting the number of precision bits and the sampling frequency.

A typical RF chain includes one mixer, one local oscillator, two low-pass filters and a baseband amplifier. We respectively denote by  $P_{\text{M}}$ ,  $P_{\text{LO}}$ ,  $P_{\text{LPF}}$  and  $P_{\text{BBA}}$ , the power dissipation of the RF chain components. Thereby, the power consumed by a single RF chain is equal to

$$P_{\text{RF}} = P_{\text{M}} + 2 P_{\text{LO}} + P_{\text{LPF}} + P_{\text{BBA}}. \quad (33)$$

In case all the RF streams are transmitted at the same frequency, it might be possible to share a single local oscillator

TABLE I  
COMPLEXITY COMPARISON OF HYBRID SYMBOL-LEVEL PRECODING SCHEMES

| Method     | Worst-case complexity   |
|------------|---|
| [23]       | $(4N_t + 2N_u + 1)^{3/2}(2N_t + 1)^2(1/\sqrt{\epsilon_o}) \cdot \mathcal{O}(1)$   |
| [24]       | $C_{FD} + \sqrt{N_t N_u}(N_t^3 N_u^2 + N_t^2 N_u^3) + N_t(N_t + 3)(1/\sqrt{\epsilon_o}) \cdot \mathcal{O}(1)$   |
| This paper | $C_{FD} + N_t N_t^2 (1 + N_t(1/\sqrt{\epsilon_i})) \left[ \left( \ln(2 N_t N_t \sqrt{N_t}) - \ln(\mu^{(0)} \epsilon_o) \right) / \ln(\vartheta) \right] \cdot \mathcal{O}(1)$ |

among all the chains and divide the power consumption  $P_{LO}$  accordingly [27]. Further, let  $P_{BB}$ ,  $P_{PA}$ ,  $P_{PS}$  and  $P_{SW}$  respectively denote the power consumption of the baseband processor, a single PA, a single phase shifter and a single analog switch. Note also that, in general, the power dissipation of the RF combining network is very low [50], and thus is ignored in our modeling.

The fully-digital BS architecture requires  $2N_t$  DACs, and  $N_t$  RF chains and PAs, and therefore its power consumption can be modeled as

$$P_{FD} = P_{BB} + N_t(2P_{DAC} + P_{RF} + P_{PA}). \quad (34)$$

On the other hand, the hybrid architecture with fully-connected phase-shifting network can be implemented using  $2N_t$  DACs,  $N_t$  RF chains,  $N_t$  PAs, and  $N_t \times N_t$  phase shifters. The resulting power dissipation is thus given by

$$P_H = P_{BB} + N_t(2P_{DAC} + P_{RF}) + N_t N_t P_{PS} + N_t P_{PA}. \quad (35)$$

To calculate the power consumption of the hybrid architecture with fully-connected networks of phase shifters and switches, i.e., with phase shifter selection, we assume the associated RF processes are turned off while a phase shifter is deactivated, and further, the phase shifter has negligible static power dissipation. Under this assumption, a deactivated phase shifter consumes no power. Denoting the average percentage of the active phase shifters at a symbol instant by  $\beta$ , the power consumed by the entire phase-shifting network is then  $\beta N_t N_t P_{PS}$ . As illustrated in Fig. 1, the phase shifter selection mechanism is implemented through a network of  $N_t \times N_t$  switches. Therefore, the power consumption of the hybrid precoder with phase shifter selection can be obtained as

$$P_{HPSS} = P_{BB} + N_t(2P_{DAC} + P_{RF}) + N_t N_t (\beta P_{PS} + P_{SW}) + N_t P_{PA}. \quad (36)$$

Recall from Section II that the selection matrix  $\mathbf{T}$  is constrained to have no all-zero row (column), i.e., at least one phase shifter corresponding to a specific antenna element (RF chain) must be active at a symbol instant. As a consequence, the number of active phase shifters during any symbol period is never less than  $\max\{N_t, N_t\} = N_t$ , from which it follows that  $1/N_t < \beta \leq 1$ . Our simulation results in Section V further indicate that  $\beta$  is usually smaller than 0.75 for the proposed hybrid symbol-level precoder in (14), regardless of the phase-shifting precision. This may lead to significant reductions in the power consumption of the analog phase-shifting network. It is also important to note that by employing

low-power yet efficient mmWave switches, the excessive power consumption due to the switching operation can be made negligible compared with the power reduction of the phase shifters.

Using the above power consumption models with appropriate parameter selection, we will compare the power consumed by different fully-digital/hybrid architectures in Section V.

## V. SIMULATION RESULTS

In this section, we present some simulation results to evaluate the performance of the proposed hybrid symbol-level precoding approach and to compare it with some other existing schemes. The simulation setup is as follows. We consider the hybrid analog-digital precoding architecture depicted in Fig. 1 for a downlink mmWave massive multiuser MIMO system, performing an uncoded transmission with QPSK signaling and a carrier frequency of 60 GHz over a bandwidth of 1 GHz. We assume unit noise variances at the receivers of all the users, i.e.,  $\sigma_j^2 = 1, \forall j = 1, 2, \dots, N_u$ . As described in Section II, we adopt a geometric model for the mmWave propagation environment with  $N_c = 1$  clusters and  $N_p = 12$  scatterers between the BS and each user. For all the propagation paths, the azimuth angles of departure  $\phi_{j,i,l}$  are drawn independently from a uniform distribution over  $[0, 2\pi)$ . To initialize Algorithm 1, we set  $N = 1$  and  $\vartheta = 1.1$  to avoid overshooting, and consider  $\mu^{(0)} = 10^{-4}$  to have a reasonable starting point.

We consider the fully-digital Wiener filter (WF) precoding [51], and the optimal fully-digital symbol-level precoding (SLP) as our performance benchmarks, and further, provide comparisons with the block-level hybrid precoding technique PZF and its quantized variant QPZF in [18], the block-level hybrid precoding with phase shifter selection in [32], and the hybrid symbol-level precoders in [23] and [24]. Note that the application of PZF and QPZF techniques is limited only to fully-loaded systems, i.e., when  $N_u = N_t$ , and therefore, their performances have been evaluated only in the relevant scenarios. We further note that the method in [23] performs a symbol-based optimization of the digital baseband precoder subject to one-bit DACs, and adopts a CSI-only design for the phase-shifting network. On the other hand, the hybrid scheme in [24] jointly optimizes both the digital baseband precoder and the phase-shifting network on a symbol-level basis. Accordingly, we refer to the methods in [23], [24], and the proposed scheme based on the adopted hybrid architecture and the precoder design approach. To summarize, throughout this section, the hybrid precoding techniques of interest are referred to as:

- Hybrid PZF: hybrid block-level precoding based on ZF solution in [18]
- Hybrid QPZF: quantized hybrid block-level precoding based on ZF solution in [18]
- Hybrid PSS BLP: hybrid block-level precoding with phase shifter selection in [32]
- Hybrid BB SLP: hybrid symbol-level precoding with baseband precoder optimization in [23]
- Hybrid BB+PS SLP: hybrid symbol-level precoding with joint baseband precoder and phase-shifting network optimization in [24]
- Hybrid BB+SW SLP: the proposed hybrid symbol-level precoding with joint baseband precoder and switching network optimization (Algorithm 1)
- Hybrid BB+SW SLP-NOPSS: the proposed hybrid symbol-level precoding with baseband precoder optimization and no phase shifter selection

In our simulations, the power consumption is calculated according to the model introduced in Section IV, in which we consider reference values of  $P_{RF} = 40$  mW,  $P_{PA} = 20$  mW,  $P_{PS} = 30$  mW, and  $P_{BB} = P_{DAC}$ , as in [27]. As for the power consumption of switches, it is well known that nFET switches have zero static power dissipation. On the other hand, silicon-germanium (SiGe) based switches are shown to be capable of achieving high performance while consuming powers of less than 1 mW [29]. Therefore, based on the available technology for the implementation of RF switches, a fairly conservative choice would be  $P_{SW} = 1$  mW. Moreover, the power consumption of DACs is calculated via (32) assuming a sampling frequency of  $F_s = 1$  GHz which should be sufficient for mmWave systems. We further assume  $b_{DAC} = 12$  for those architectures employing high-resolution DACs.

In Fig. 4, we compare the power consumption of various hybrid precoding implementations with that of the fully-digital architecture as a function of the number of BS's RF chains  $N_1$ , while fixing the number of transmit antennas and users to be  $N_t = 64$  and  $N_u = 4$ , respectively. As might be expected, the power consumption values associated with the hybrid implementations increase with  $N_1$ , which is a consequence of requiring more RF elements, phase shifters, and/or switches. This implies that increasing the number of RF chains beyond a certain limit makes the hybrid implementation a more power-consuming approach than the fully-digital architecture. Nevertheless, for  $N_1 \leq 10$ , all the hybrid implementations consume less power than the fully-digital architecture. Remarkably, the proposed hybrid precoder in this paper, i.e., the hybrid BB+SW SLP, offers smaller power consumption amounts, with either infinite-precision or discrete phase shifters, among the other hybrid symbol-level precoding schemes in Fig. 4. This is brought by the adopted phase shifter selection mechanism in implementing the hybrid precoder. In particular, the proposed hybrid precoder has the smallest power consumption with  $b_{PS} = 1$  due to the large percentage of deactivated phase shifters, as we will see later in this section. Note also that the differences in power consumption of different hybrid precoding schemes in Fig. 4 increase with  $N_1$ .

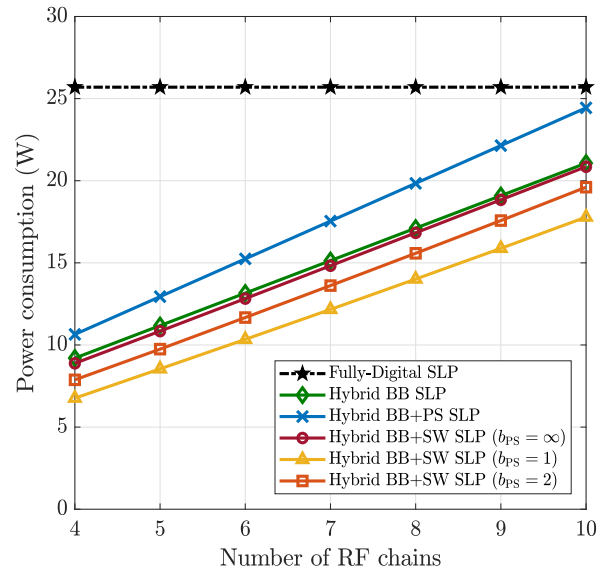


Fig. 4. Power consumption of different hybrid precoding schemes as a function of  $N_1$  with  $N_t = 64$  and  $N_u = 4$  at SNR = -5 dB.

In a scenario with  $N_t = 64$  and  $N_u = 4$ , the percentages of deactivated phase shifters, i.e.,  $(1 - \beta) \times 100$ , for the proposed hybrid precoding approach are shown versus the number of RF chains in Fig. 5 for different values of  $b_{PS}$ . It follows from the results that in the case of using low-resolution phase shifters, a higher percentage of the phase shifters can be switched off, and hence more power-savings are possible. In particular, in the case where  $N_1 = 4$ , up to 55% of the phase shifters with  $b_{PS} = 1$  can be turned off, which can be roughly translated to a power reduction of  $\beta N_t N_1 P_{PS} \approx 4200$  mW in the phase-shifting network. It can further be observed from Fig. 5 that the percentage of deactivated phase shifters decreases with increasing  $N_1$ . One can justify these observations by considering that increasing the number of phase-shifting precision bits and the number of RF chains, respectively, reduces the discontinuity in the feasible region of the optimization problem (14) and increases the design degrees of freedom. In both cases, this enables the algorithm to achieve lower values for the objective function by activating a larger ratio of the phase shifters. The former case can also be verified from Fig. 3 where the residual error in the objective function is shown to be smaller for phase shifters with higher resolution bits.

We plot the average users' symbol error rate (SER) achieved by the precoding techniques of interest with either fully-digital or hybrid architecture versus the transmit SNR for a system with  $(N_t, N_1, N_u) = (64, 8, 8)$  in Fig. 6. The proposed hybrid symbol-level precoder is evaluated for various implementations with infinite-precision and discrete phase shifters, where in the latter case we assume  $b_{PS} = 1$  and  $b_{PS} = 2$  bits of precision. It can be seen that, for the case with  $b_{PS} = 2$ , both the hybrid BB+PS SLP and the hybrid BB+SW SLP schemes are capable of performing well close to the fully-digital SLP, though requiring far less RF chains to process the transmitted signal. The corresponding losses at SER =  $10^{-2}$  are respectively around 0.1 dB and 0.5 dB. Using phase shifters with  $b_{PS} = 1$ , the hybrid BB+SW SLP scheme

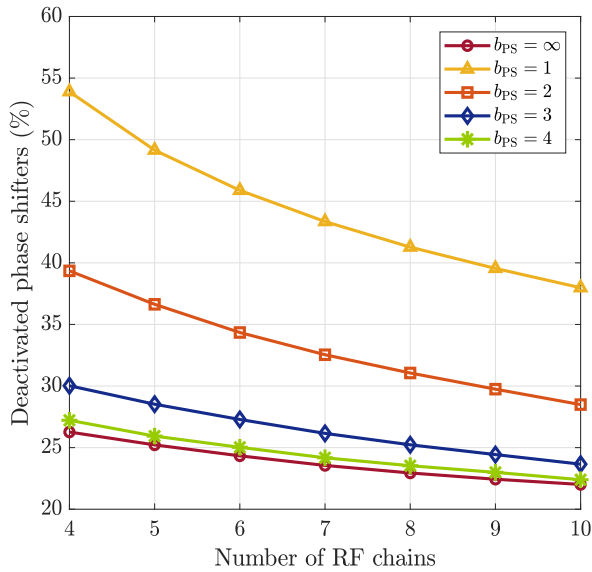


Fig. 5. Average percentage of deactivated phase shifters as a function of  $N_1$  with  $N_t = 64$  and  $N_u = 4$  at  $\text{SNR} = -5$  dB.

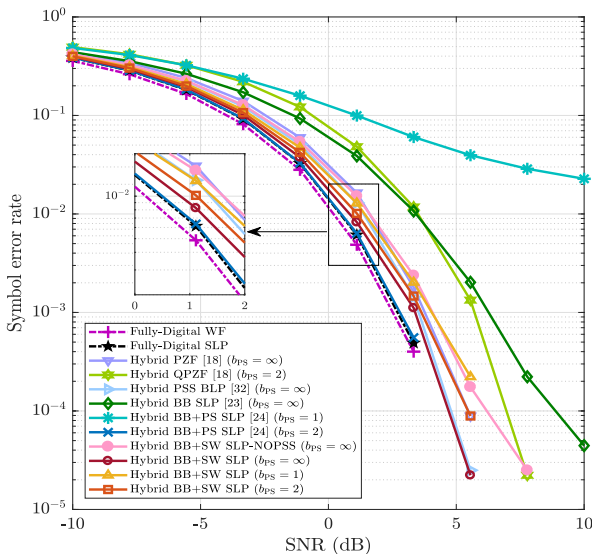


Fig. 6. Average per-user symbol error rate versus transmit SNR for a system with  $(N_t, N_1, N_u) = (64, 8, 8)$ .

still offers a reasonable performance with a loss smaller than 1 dB at  $\text{SER} = 10^{-2}$  compared with the fully-digital SLP, as opposed to the hybrid BB+PS SLP scheme which shows a significantly deteriorated performance. It can further be seen from Fig. 6 that both hybrid BB+PS SLP and hybrid BB+SW SLP approaches outperform the PZF technique, which is a result of designing the precoded signal specifically for each instantaneous combination of the users' target symbols. Overall, from Fig. 6, it follows that the hybrid BB+PS SLP scheme offers the best SER performance compared to the other hybrid SLP schemes of interest. Nevertheless, as demonstrated in Fig. 4, this superior performance comes with increased power consumption.

In Fig. 7, the average per-user spectral efficiencies of the precoding schemes of interest are shown for the system parameter sets  $(N_t, N_1, N_u) = (64, 8, 8)$ . As can be seen,

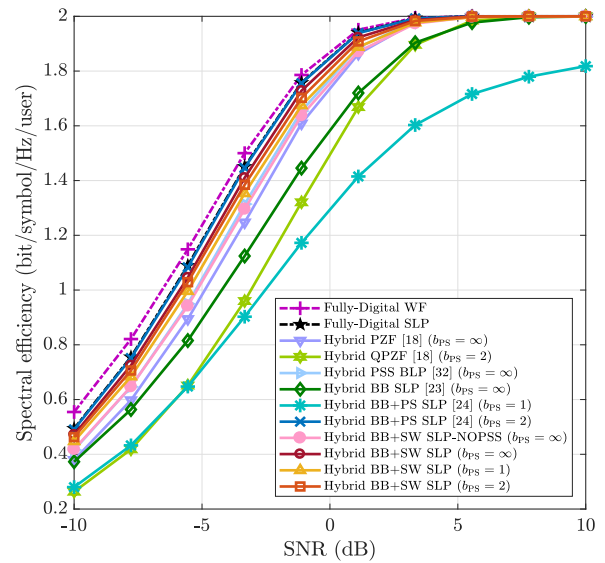


Fig. 7. Per-user spectral efficiency versus transmit SNR for a system with  $(N_t, N_1, N_u) = (64, 8, 8)$ .

the spectral efficiency plot follows the same relative trend as that of the SER plot. The hybrid BB+PS and BB+SW SLP schemes are more spectrally-efficient than the PZF and QPZF techniques, which is a result of the CI-based positioning of the received signals. Remarkably, the achievable spectral efficiencies by the hybrid BB+PS SLP scheme with  $b_{\text{PS}} = 2$  and by the hybrid BB+SW SLP scheme with either  $b_{\text{PS}} = 1$ ,  $b_{\text{PS}} = 2$  or  $b_{\text{PS}} = \infty$  are close to those of the fully-digital WF and SLP. The maximum loss with respect to the fully-digital SLP corresponds to the Hybrid BB+SW SLP scheme with  $b_{\text{PS}} = 1$ , which is around 0.06 bps/symbol/Hz at  $\text{SNR} = 0$  dB. On the other hand, hybrid BB+PS SLP is shown in Fig. 7 to be the most spectrally-efficient approach among the hybrid symbol-level precoders of interest.

Up until this point in the simulation results, we have seen that among the hybrid symbol-level precoders of interest, one approach outperforms the other in terms of either power consumption, symbol error rate, or spectral efficiency. To have an all-inclusive comparison, we use the energy efficiency measure, as defined in Section IV, that incorporates all the aforementioned figures of merit in evaluating the overall precoding performance. The results are shown in Fig. 8, where the energy efficiencies of different fully-digital and hybrid multiuser precoders are plotted as a function of the transmit SNR for a system with  $(N_t, N_1, N_u) = (64, 8, 8)$ . As can be seen, almost all of the hybrid symbol-level precoders are more energy-efficient than the fully-digital SLP, while the proposed hybrid BB+SW SLP approach with phase shifter selection outperforms the other schemes with either infinite or finite resolution phase shifters. The most energy-efficient scheme is shown to be hybrid BB+SW SLP with  $b_{\text{PS}} = 1$ , using which energy efficiency gains of up to 75 Mbps/Joule per user can be achieved against the fully-digital SLP. In contrast to the Hybrid BB+PS SLP scheme, employing phase shifters with lower precision bits improves the energy-efficiency of Hybrid BB+SW SLP. This is because more phase shifters can be switched off using low-precision phase shifters, which leads

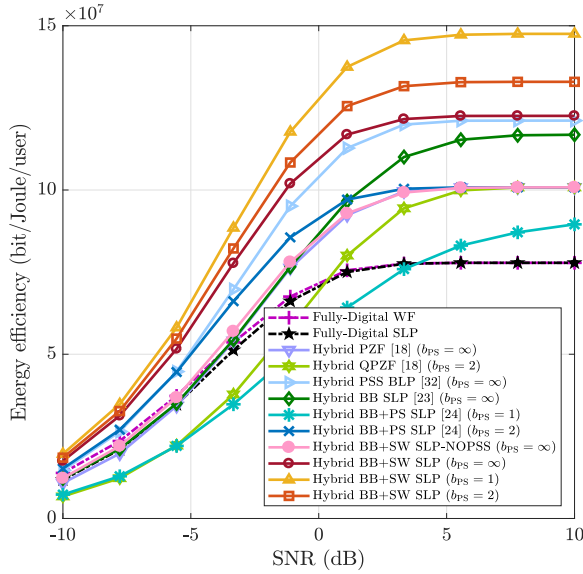


Fig. 8. Energy efficiency versus transmit SNR with  $(N_t, N_1, N_u) = (64, 8, 8)$ .

to larger reductions in power consumption. It is important to note that in our power consumption model, we consider the same reference value for phase shifters with any number of precision bits. This is rather a simplistic approach as, in practice, higher-resolution phase shifters consume more power. In such a case, the results for power consumption and energy efficiency of the proposed hybrid precoder with low-resolution phase shifters would show an even higher gain compared to the other schemes of interest. It can further be seen from Fig. 8 that the proposed hybrid algorithm with  $b_{PS} = \infty$  outperforms the hybrid PSS BLP scheme, where both techniques employ a phase shifter selection mechanism via a switching network but on a symbol-level and block-level basis, respectively. In particular, the hybrid BB+SW technique shows higher energy efficiency gains against the hybrid PSS BLP scheme at low SNRs. We are further interested in the behavior of energy efficiency as a function of the number of available RF chains  $N_1$ , which is plotted in Fig. 9 for fixed numbers of transmit antennas  $N_t = 64$  and users  $N_u = 4$  at an SNR of  $-5$  dB. A common trend across all the hybrid symbol-level precoders is that their energy efficiency becomes lower as  $N_1$  increases. This is in accordance with the power consumption results in Fig. 4, indicating that for a fixed number of antennas, a hybrid precoding implementation becomes less energy-efficient than its fully-digital counterpart whenever the number of RF chains exceeds an upper limit. This upper limit is shown in Fig. 9 to be larger for the proposed hybrid BB+SW SLP approach. On the other hand, comparing the proposed hybrid symbol-level precoder with the case where all the phase shifters are active, i.e., with no phase shifter selection, we can conclude that applying the phase shifter selection mechanism can substantially improve the energy efficiency of hybrid symbol-level precoding. The results in Fig. 9 shows that gains of up to 37 Mbps/Joule per user can be achieved using the hybrid BB+SW SLP method compared to its counterpart scheme without phase shifter selection.

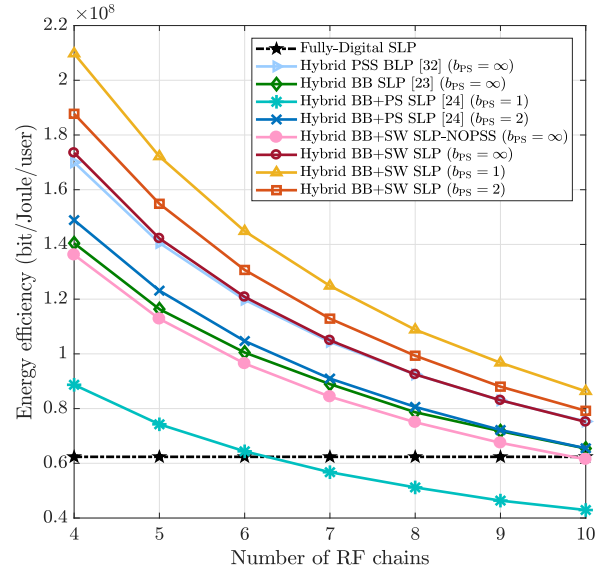


Fig. 9. Energy efficiency at SNR =  $-5$  dB as a function of  $N_1$  with  $N_t = 64$  and  $N_u = 4$ .

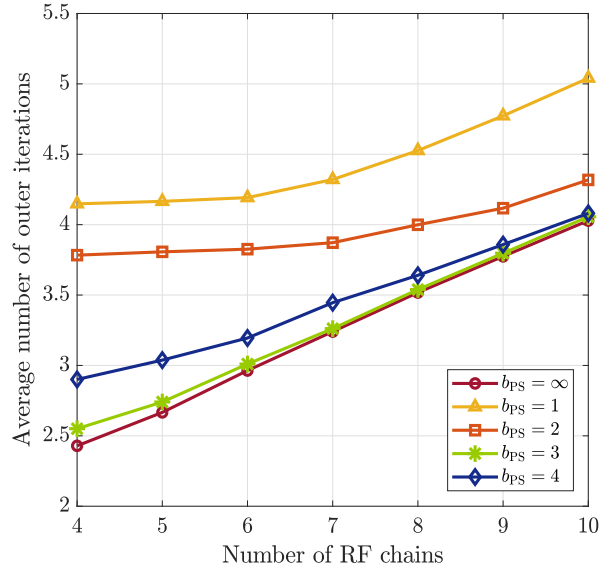


Fig. 10. Average number of outer iterations till convergence of the proposed hybrid precoding algorithm versus  $N_1$  with  $N_t = 64$ .

Following the analytic complexity analysis provided in Section III, we numerically evaluate the computational complexity of the proposed hybrid SLP algorithm in both scenarios with infinite-resolution and discrete phase shifters. The complexity results, in terms of the required number of outer iterations (i.e., block cycles) for convergence, is shown in Fig. 10 as a function of the number of RF chains  $N_1$ . It is, however, important to note that the complexity of solving the inner sub-problem (16) is not of our interest since, as mentioned earlier, this problem is a typical linearly-constrained quadratic program which can efficiently be solved using many existing algorithms. As might be expected, the number of outer iterations until convergence of the proposed hybrid SLP algorithm increases with  $N_1$  in all the cases, which is due to the corresponding growth in the problem size. On the other hand, the computation cost increases with reducing the precision of

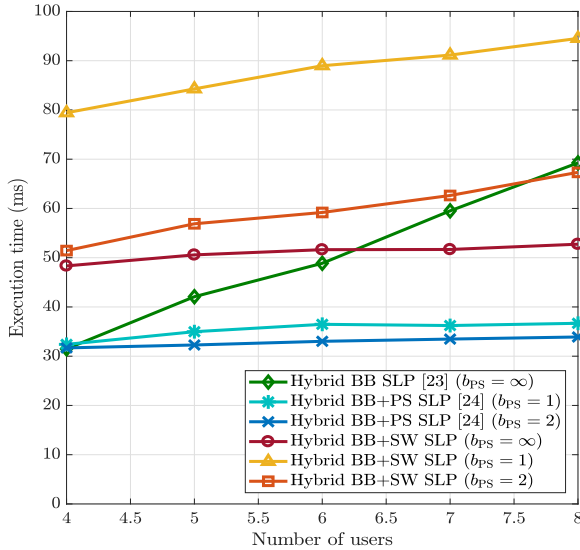


Fig. 11. Average execution time of different hybrid symbol-level precoding schemes versus  $N_u$  with  $N_t = 64$  and  $N_1 = 8$ .

the phase shifters. Such an observation, however, is not surprising since having discrete possible phase states causes a discontinuity in the feasible region of the optimization problem, and consequently, more cycles are needed for convergence to a stationary point. We further compare the complexities of different hybrid symbol-level precoding techniques in Fig. 11, where the execution time of each hybrid algorithm is plotted as a function of the number of users while the number of transmit antennas and RF chains are fixed as  $N_t = 64$  and  $N_1 = 8$ . It should be noted that for the hybrid algorithm in [24] and the proposed scheme, the results in Fig. 11 incorporate the solutions to both the fully-digital and hybrid symbol-level precoding problems. Note further that the execution times shown in Fig. 11 have been obtained using the MATLAB timing functions, and the fully-digital symbol-level precoding problems and the optimization problem corresponding to the hybrid scheme in [23] have been solved using the *fmincon* tool. It can be seen that the execution times of the hybrid BB+PS SLP algorithm and the proposed hybrid BB+SW SLP technique (with high-resolution phase shifters) slightly grows with increasing the number of users, whereas the complexity of the hybrid BB SLP method shows a larger growth rate with the number of users. The proposed hybrid BB+SW SLP technique requires higher execution times than the other two methods, particularly for a small number of users. Moreover, the proposed hybrid precoding scheme shows higher execution times with low-precision phase shifters, which is in accordance with the results shown in Fig. 10. Similarly, this observation is a consequence of having increased discontinuity in the solution space of the precoding problem due to discrete phase shifters.

## VI. CONCLUDING REMARKS

In this paper, we proposed a hybrid analog-digital precoding scheme for large-scale multiuser mmWave downlink systems. The multiuser precoding operation is split between the digital and analog domains, where processing in the analog domain is carried out through fully-connected networks of switches and phase shifters. The use of on-off switches

enables us to perform phase shifter selection in the analog precoder. We adopted a CSI-only design approach for the phase-shifting network, whereas the digital baseband precoder and the switching network are optimized in a symbol-level manner, i.e., by exploiting the instantaneous data symbols to enable constructive interference at the receiver side. We formulated our design problem to minimize the Euclidean distance between the hybrid symbol-level precoder and its optimal fully-digital counterpart, where a power-constrained max-min SINR design criterion subject to constructive interference constraints was adopted. Our design approach led us to an intractable binary optimization problem. We tackled this difficulty by transforming the original problem to an equivalent continuous-domain biconvex form, which can efficiently be solved for a sub-optimal solution via the standard block coordinate descent (BCD) algorithm. We evaluated the computational complexity of the proposed scheme, where it was shown by numerical results that the adopted BCD algorithm needs only a few (normally less than ten) cycles to converge. To assess and compare different fully-digital/hybrid precoding schemes from both performance and power consumption points of view, we provided an analysis of energy efficiency by considering appropriate models for the power dissipation of the RF elements. Our simulation results indicated that applying the phase shifter selection approach, up to half of the phase shifters can be switched off, allowing for reductions of multi-Watts in the power consumption of the analog circuitry. This power consumption reduction can significantly improve the energy efficiency of precoding, as compared to the fully-digital and the state-of-the-art hybrid symbol-level techniques. Moreover, we evaluated the proposed hybrid precoding scheme with both infinite and finite precision phase shifters. It was shown that using phase shifters with lower precision bits, on the one hand, degrades the spectral efficiency, but on the other hand, allows for more power-savings due to a larger number of deactivated phase shifters, and therefore, is more energy-efficient.

## APPENDIX A PROOF OF LEMMA 2

Given  $\mathbf{u}_{BB}$ ,  $g(\mathbf{u}_{BB}, \mathbf{g})$  is an  $\ell_2$ -norm function of  $\mathbf{g}$  and therefore is continuously differentiable everywhere. Let  $\{\mathbf{g}_1, \mathbf{g}_2\} \in \mathbf{R}^{N_t N_1 \times 1}$  be any two distinct inputs to the function  $g(\mathbf{u}_{BB}, \mathbf{g})$  such that  $-\mathbf{1} \preceq \{\mathbf{g}_1, \mathbf{g}_2\} \preceq \mathbf{1}$ . Let us further denote  $\Theta \triangleq (\mathbf{u}_{BB}^T \otimes \mathbf{I}_{N_t}) \text{diag}(\text{vec}(\mathbf{F}))$ . Then, we can write

$$\begin{aligned}
 & \left| g(\mathbf{u}_{BB}, \mathbf{g}_1) - g(\mathbf{u}_{BB}, \mathbf{g}_2) \right| \\
 &= \left| \left\| \Theta \mathbf{g}_1 + \mathbf{F} \mathbf{u}_{BB} - 2 \mathbf{u}_{FD}^* \right\|^2 - \left\| \Theta \mathbf{g}_2 + \mathbf{F} \mathbf{u}_{BB} - 2 \mathbf{u}_{FD}^* \right\|^2 \right| \\
 &= \left| \left\| \Theta \mathbf{g}_1 \right\|^2 - \left\| \Theta \mathbf{g}_2 \right\|^2 + (2 \mathbf{u}_{BB}^H \mathbf{F}^H \Theta - 4 \mathbf{u}_{FD}^{*H} \Theta)(\mathbf{g}_1 - \mathbf{g}_2) \right|. \tag{37}
 \end{aligned}$$

According to the matrix/vector operator norm inequality, the following chain of inequalities holds true:

$$\begin{aligned}
 & \left| g(\mathbf{u}_{BB}, \mathbf{g}_1) - g(\mathbf{u}_{BB}, \mathbf{g}_2) \right| \\
 & \stackrel{(a)}{\leq} \left| \left\| \Theta \mathbf{g}_1 \right\|^2 - \left\| \Theta \mathbf{g}_2 \right\|^2 \right| + \left| (2 \mathbf{u}_{BB}^H \mathbf{F}^H \Theta - 4 \mathbf{u}_{FD}^{*H} \Theta)(\mathbf{g}_1 - \mathbf{g}_2) \right|
 \end{aligned}$$

$$\begin{aligned}
& \stackrel{(b)}{\leq} \|\Theta \mathbf{g}_1 - \Theta \mathbf{g}_2\|^2 + \left\| 2\Theta^H \mathbf{F} \mathbf{u}_{\text{BB}} - 4\Theta^H \mathbf{u}_{\text{FD}}^* \right\| \|\mathbf{g}_1 - \mathbf{g}_2\| \\
& \leq \|\Theta\|_{\text{F}}^2 \|\mathbf{g}_1 - \mathbf{g}_2\|^2 + \left\| 2\Theta^H \mathbf{F} \mathbf{u}_{\text{BB}} - 4\Theta^H \mathbf{u}_{\text{FD}}^* \right\| \|\mathbf{g}_1 - \mathbf{g}_2\|,
\end{aligned} \tag{38}$$

where inequality (a) and inequality (b) are both followed from the (reverse) triangle inequality. Since  $-\mathbf{1} \preceq \{\mathbf{g}_1, \mathbf{g}_2\} \preceq \mathbf{1}$ , we always have  $\|\mathbf{g}_1 - \mathbf{g}_2\|_2 \leq 2\sqrt{N_t N_1}$ , where equality is achieved in case either  $\mathbf{g}_1$  or  $\mathbf{g}_2$  is equal to  $\mathbf{1}$  while the other equals  $-\mathbf{1}$ . Using the latter inequality and the one in (38), we obtain the following upper bound:

$$\begin{aligned}
& \left| g(\mathbf{u}_{\text{BB}}, \mathbf{g}_1) - g(\mathbf{u}_{\text{BB}}, \mathbf{g}_2) \right| \\
& \leq \left( 2\sqrt{N_t N_1} \|\Theta\|_{\text{F}}^2 + \left\| 2\Theta^H \mathbf{F} \mathbf{u}_{\text{BB}} - 4\Theta^H \mathbf{u}_{\text{FD}}^* \right\| \right) \|\mathbf{g}_1 - \mathbf{g}_2\|.
\end{aligned} \tag{39}$$

It then immediately follows that

$$\begin{aligned}
& \left| g(\mathbf{u}_{\text{BB}}, \mathbf{g}_1) - g(\mathbf{u}_{\text{BB}}, \mathbf{g}_2) \right| / \|\mathbf{g}_1 - \mathbf{g}_2\| \\
& \leq 2\sqrt{N_t N_1} \|\Theta\|_{\text{F}}^2 + 2 \left\| \Theta^H (\mathbf{F} \mathbf{u}_{\text{BB}} - 2 \mathbf{u}_{\text{FD}}^*) \right\| \triangleq L,
\end{aligned} \tag{40}$$

where  $L$  is a positive real constant in either case with  $\Theta \neq \mathbf{0}$  or  $\mathbf{F} \mathbf{u}_{\text{BB}} - 2 \mathbf{u}_{\text{FD}}^* \neq \mathbf{0}$ , implying the Lipschitz continuity property for the function  $g(\mathbf{u}_{\text{BB}}, \mathbf{g})$  on  $-\mathbf{1} \preceq \mathbf{g} \preceq \mathbf{1}$ .

## REFERENCES

- [1] T. S. Rappaport *et al.*, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, 2013.
- [2] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 101–107, Jun. 2011.
- [3] S. K. Yong and C.-C. Chong, "An overview of multigigabit wireless through millimeter wave technology: Potentials and technical challenges," *EURASIP J. Wireless Commun. Netw.*, vol. 2007, no. 1, pp. 1–10, Dec. 2006.
- [4] C.-X. Wang *et al.*, "Cellular architecture and key technologies for 5G wireless communication networks," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 122–130, Feb. 2014.
- [5] A. Alkhateeb, G. Leus, and R. W. Heath, "Limited feedback hybrid precoding for multi-user millimeter wave systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 11, pp. 6481–6494, Nov. 2015.
- [6] C. B. Peel, B. M. Hochwald, and A. L. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication—Part I: Channel inversion and regularization," *IEEE Trans. Commun.*, vol. 53, no. 1, pp. 195–202, Jan. 2005.
- [7] A. Wiesel, Y. C. Eldar, and S. Shamai, "Zero-forcing precoding and generalized inverses," *IEEE Trans. Signal Process.*, vol. 56, no. 9, pp. 4409–4418, Sep. 2008.
- [8] 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.
- [9] S. Hur, T. Kim, D. J. Love, J. V. Krogmeier, T. A. Thomas, and A. Ghosh, "Millimeter wave beamforming for wireless backhaul and access in small cell networks," *IEEE Trans. Commun.*, vol. 61, no. 10, pp. 4391–4403, Oct. 2013.
- [10] A. M. Sayeed and V. Raghavan, "Maximizing MIMO capacity in sparse multipath with reconfigurable antenna arrays," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 1, pp. 156–166, Jun. 2007.
- [11] O. E. Ayach, R. W. Heath, S. Abu-Surra, S. Rajagopal, and Z. Pi, "The capacity optimality of beam steering in large millimeter wave MIMO systems," in *Proc. IEEE 13th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jun. 2012, pp. 100–104.
- [12] X. Gao, O. Edfors, J. Liu, and F. Tufvesson, "Antenna selection in measured massive MIMO channels using convex optimization," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Dec. 2013, pp. 129–134.
- [13] B. Moo Lee, J. Choi, J. Bang, and B.-C. Kang, "An energy efficient antenna selection for large scale green MIMO systems," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May 2013, pp. 950–953.
- [14] L. N. Ribeiro, S. Schwarz, M. Rupp, and A. L. F. de Almeida, "Energy efficiency of mmWave massive MIMO precoding with low-resolution DACs," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 2, pp. 298–312, May 2018.
- [15] S. Jacobsson, G. Durisi, M. Coldrey, T. Goldstein, and C. Studer, "Quantized precoding for massive MU-MIMO," *IEEE Trans. Commun.*, vol. 65, no. 11, pp. 4670–4684, Nov. 2017.
- [16] S. S. Ioushua and Y. C. Eldar, "A family of hybrid analog–digital beamforming methods for massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 67, no. 12, pp. 3243–3257, Jun. 2019.
- [17] A. F. Molisch *et al.*, "Hybrid beamforming for massive MIMO: A survey," *IEEE Commun. Mag.*, vol. 55, no. 9, pp. 134–141, Sep. 2017.
- [18] 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.
- [19] 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.
- [20] C. Masouros and G. Zheng, "Exploiting known interference as green signal power for downlink beamforming optimization," *IEEE Trans. Signal Process.*, vol. 63, no. 14, pp. 3628–3640, Jul. 2015.
- [21] M. Alodeh, S. Chatzinotas, and B. Ottersten, "Constructive multiuser interference in symbol level precoding for the MISO downlink channel," *IEEE Trans. Signal Process.*, vol. 63, no. 9, pp. 2239–2252, May 2015.
- [22] G. Hegde, C. Masouros, and M. Pesavento, "Analog beamformer design for interference exploitation based hybrid beamforming," in *Proc. IEEE 10th Sensor Array Multichannel Signal Process. Workshop (SAM)*, Jul. 2018, pp. 109–113.
- [23] A. Li, C. Masouros, and F. Liu, "Hybrid analog-digital precoding for interference exploitation (Invited Paper)," in *Proc. 26th Eur. Signal Process. Conf. (EUSIPCO)*, Sep. 2018, pp. 812–816.
- [24] R. Liu, H. Li, and M. Li, "Symbol-level hybrid precoding in mmWave multiuser MISO systems," *IEEE Commun. Lett.*, vol. 23, no. 9, pp. 1636–1639, Sep. 2019.
- [25] S. Domouchtsidis, C. G. Tsinos, S. Chatzinotas, and B. Ottersten, "Symbol-level precoding for low complexity transmitter architectures in large-scale antenna array systems," *IEEE Trans. Wireless Commun.*, vol. 18, no. 2, pp. 852–863, Feb. 2019.
- [26] L. A. Belov, S. M. Smolskiy, and V. N. Kochemasov, *Handbook of RF, Microwave, and Millimeter-wave Components*. Norwood, MA, USA: Artech House, 2012.
- [27] R. Mendez-Rial, C. Rusu, N. Gonzalez-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, 2016.
- [28] D. Liu, U. Pfeiffer, J. Grzyb, and B. Gaucher, *Advanced Millimeter-wave Technologies: Antennas, Packaging and Circuits*. Hoboken, NJ, USA: Wiley, 2009.
- [29] R. L. Schmid, P. Song, C. T. Coen, A. C. Ulusoy, and J. D. Cressler, "On the analysis and design of low-loss single-pole double-throw W-band switches utilizing saturated SiGe HBTs," *IEEE Trans. Microw. Theory Techn.*, vol. 62, no. 11, pp. 2755–2767, Nov. 2014.
- [30] I. Kalfass, S. Diebold, H. Massler, S. Koch, M. Seelmann-Eggebert, and A. Leuther, "Multiple-throw millimeter-wave FET switches for frequencies from 60 up to 120 GHz," in *Proc. Eur. Microw. Integr. Circuit Conf.*, Oct. 2008, pp. 1453–1456.
- [31] S. Payami, N. Mysore Balasubramanya, C. Masouros, 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.
- [32] S. Payami, M. Ghoraiishi, and M. Dianati, "Hybrid beamforming for large antenna arrays with phase shifter selection," *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7258–7271, Nov. 2016.
- [33] Y. Jiang, Y. Feng, and M. K. Varanasi, "Hybrid beamforming for massive MIMO: A unified solution for both phase shifter and switch networks," in *Proc. 10th Int. Conf. Wireless Commun. Signal Process. (WCSP)*, Oct. 2018, pp. 1–5.
- [34] A. Haqiqatnejad, F. Kayhan, and B. Ottersten, "Constructive interference for generic constellations," *IEEE Signal Process. Lett.*, vol. 25, no. 4, pp. 586–590, Apr. 2018.
- [35] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 831–846, Oct. 2014.
- [36] W. U. Bajwa, J. Haupt, A. M. Sayeed, and R. Nowak, "Compressed channel sensing: A new approach to estimating sparse multipath channels," *Proc. IEEE*, vol. 98, no. 6, pp. 1058–1076, Jun. 2010.

- [37] T. S. Rappaport, R. W. Heath, Jr., R. C. Daniels, and J. N. Murdock, *Millimeter Wave Wireless Communications*. London, U.K.: Pearson, 2015.
- [38] H. Xu, V. Kukshya, and T. S. Rappaport, "Spatial and temporal characteristics of 60-GHz indoor channels," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 3, pp. 620–630, Apr. 2002.
- [39] P. F. M. Smulders and L. M. Correia, "Characterisation of propagation in 60 GHz radio channels," *Electron. Commun. Eng. J.*, vol. 9, no. 2, pp. 73–80, Apr. 1997.
- [40] A. Haqiqatnejad, F. Kayhan, and B. Ottersten, "Symbol-level precoding design based on distance preserving constructive interference regions," *IEEE Trans. Signal Process.*, vol. 66, no. 22, pp. 5817–5832, Nov. 2018.
- [41] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
- [42] G. Yuan and B. Ghanem, "Binary optimization via mathematical programming with equilibrium constraints," 2016, *arXiv:1608.04425*. [Online]. Available: <http://arxiv.org/abs/1608.04425>
- [43] X. M. Hu and D. Ralph, "Convergence of a penalty method for mathematical programming with complementarity constraints," *J. Optim. Theory Appl.*, vol. 123, no. 2, pp. 365–390, Nov. 2004.
- [44] D. P. Bertsekas, *Nonlinear Programming*. Belmont, MA, USA: Athena Scientific, 1999.
- [45] R. A. Polyak, "Projected gradient method for non-negative least square," *Contemp. Math.*, vol. 636, pp. 167–179, Mar. 2015.
- [46] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu, "A limited memory algorithm for bound constrained optimization," *SIAM J. Sci. Comput.*, vol. 16, no. 5, pp. 1190–1208, Sep. 1995.
- [47] A. Ben-Tal and A. Nemirovski, *Lectures on Modern Convex Optimization*, vol. 2. Philadelphia, PA, USA: SIAM, 2001.
- [48] M. Gustavsson, J. J. Wikner, and N. Tan, *CMOS Data Converters for Communications*, vol. 543. Cham, Switzerland: Springer, 2000.
- [49] S. Cui, A. J. Goldsmith, and A. Bahai, "Energy-constrained modulation optimization," *IEEE Trans. Wireless Commun.*, vol. 4, no. 5, pp. 2349–2360, Sep. 2005.
- [50] A. Natarajan *et al.*, "A fully-integrated 16-element phased-array receiver in SiGe BiCMOS for 60-GHz communications," *IEEE J. Solid-State Circuits*, vol. 46, no. 5, pp. 1059–1075, May 2011.
- [51] M. Joham, W. Utschick, and J. A. Nossek, "Linear transmit processing in MIMO communications systems," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 2700–2712, Aug. 2005.



**Alireza Haqiqatnejad** (Graduate Student Member, IEEE) received the B.Sc. degree in electrical engineering from the Isfahan University of Technology, Iran, in 2012, the M.Sc. degree in telecommunications engineering from the University of Isfahan, Iran, in 2015, and the Ph.D. degree in computer science from the Interdisciplinary Centre for Security, Reliability, and Trust (SnT), University of Luxembourg, in 2021. He is currently a Research Associate (Post-Doctoral Researcher) with the SnT, University of Luxembourg. His research

interests include signal processing and optimization for wireless and satellite communication systems, interference mitigation in multiuser MIMO channels, and advanced precoding techniques for future generation wireless communications.



**Farbod Kayhan** received the B.Sc. degree in mathematics from the Shahid Bahonar University of Kerman in 1999, the M.Sc. degree in modeling and simulation of complex realities from the International Centre for Theoretical Physics (ICTP) in 2003, and the Ph.D. degree in electrical engineering and telecommunication from the Politecnico di Torino in 2007. He has spent one-year post-doctoral position with the Department of Systems Science, Kyoto University, from 2007 to 2008. From 2009 to 2015, he was a Research Assistant with the Politecnico di Torino. From 2015 to 2020, he was a Research Associate with the SIGCOM Group, University of Luxembourg. Since September 2020, he has been a Lecturer of Satellite Communication with the 5G and 6G Innovation Centre, University of Surrey. His main research interests include modern coding theory and combinatorial optimization with particular emphasis on iterative channel and source coding and constellation space optimization, signal processing, and satellite communication systems.



**Björn Ottersten** (Fellow, IEEE) received the M.S. degree in electrical engineering and applied physics from Linköping University, Linköping, Sweden, in 1986, and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, USA, in 1990. He has held research positions with the Department of Electrical Engineering, Linköping University, the Information Systems Laboratory, Stanford University, Katholieke Universiteit Leuven, Leuven, Belgium, and the University of Luxembourg, Luxembourg. In 1991, he was appointed as a

Professor of Signal Processing with the Royal Institute of Technology (KTH), Stockholm, Sweden. From 1996 to 1997, he was the Director of Research with ArrayComm, Inc., a start-up in San Jose, CA, USA, based on his patented technology. He has been the Head of the Department for Signals, Sensors, and Systems, KTH, and the Dean of the School of Electrical Engineering, KTH. He is currently the Director for the Interdisciplinary Centre for Security, Reliability and Trust, University of Luxembourg. He is also a fellow of EURASIP. He has been a Board Member of the IEEE Signal Processing Society and the Swedish Research Council. He currently serves for the boards of EURASIP and the Swedish Foundation for Strategic Research. He was a recipient of the IEEE Signal Processing Society Technical Achievement Award, the EURASIP Group Technical Achievement Award, and the European Research Council advanced research grant twice. He has coauthored journal papers that received the IEEE Signal Processing Society Best Paper Award in 1993, 2001, 2006, 2013, and 2019, and eight IEEE conference papers best paper awards. He has served as the Editor-in-Chief for EURASIP Signal Processing, and acted on the editorial boards of IEEE TRANSACTIONS ON SIGNAL PROCESSING, *IEEE Signal Processing Magazine*, IEEE OPEN JOURNAL FOR SIGNAL PROCESSING, *EURASIP Journal of Advances in Signal Processing*, and *Foundations and Trends in Signal Processing*.