IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received April 21, 2021, accepted June 11, 2021, date of publication July 1, 2021, date of current version July 13, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3093880

Low Complexity Hybrid Precoding and Combining for Millimeter Wave Systems

MOHAMED ALOUZI[1], (Student Member, IEEE), FRANCOIS CHAN[1,2], (Senior Member, IEEE), AND CLAUDE D'AMOURS[1], (Member, IEEE)

[1]School of Electrical Engineering and Computer Science, University of Ottawa, Ottawa, ON K1N 6N5, Canada
[2]Department of Electrical and Computer Engineering, Royal Military College of Canada, Kingston, ON K7K 7B4, Canada

Corresponding author: Francois Chan (chan-f@rmc.ca)

ABSTRACT In millimeter-wave (mmWave) systems, hybrid precoding/combining architectures are an attractive low-complexity alternative to the use of fully digital precoding. This is because of the reduction in the number of Radio-Frequency (RF) and mixed-signal hardware components. Hybrid precoding/combining design involves a combination of analog and digital processing that enables both beamforming and spatial multiplexing gains in mmWave systems. This paper presents an algorithm for hybrid precoding and combining designs in mmWave systems. The proposed design does not depend on the antenna array geometry, unlike the orthogonal matching pursuit (OMP) hybrid precoding/combining approach. Moreover, this algorithm allows hybrid precoders/combiners to perform close to the known optimal performance of unconstrained digital precoders and combiners. Simulation results verify that the proposed hybrid precoding/combining scheme outperforms the previous methods in the literature in terms of bit error rate (BER) and achievable spectral efficiency with lower complexity.

INDEX TERMS mmWave communication, massive MIMO, hybrid precoding and combining design.

I. INTRODUCTION

By taking advantage of the vastly unused spectrum in the millimeter Wave (mmWave) band, a massive MIMO system with up to hundreds of antennas can transmit data at rates that reach several gigabits per second [1]–[7]. While the free-space path loss is considerably greater at these frequencies than at frequencies below 6 GHz, antenna arrays with hundreds of elements can be packed into small chips at mmWave frequencies, allowing the signal to be received with sufficient power [3]–[5]. Moreover, if beamforming is used to direct the signal towards the receiver, these large antenna arrays can yield a gain that alleviates the effects of the increased path loss [3]–[5], and [8]–[11].

Baseband or digital precoder/combiner architectures can achieve a high beamforming gain and allow multiple streams for multi-user communications. However, digital baseband precoders/combiners are impractical for mmWave systems due to the large number of antennas and the requirement of a power-hungry RF chain for each antenna [3], [4]. Hence, given its high power consumption and computational complexity it is impractical to implement a digital precoder/combiner architecture [3], [4].

Alternatively, beamforming can be implemented in the RF stage by using analog or RF precoders/combiners [3], [4]. The advantages of using the analog compared to the digital approach are that both the power consumption and the hardware complexity are reduced since the number of RF chains is considerably lower. At each antenna element, a network of analog phase shifters implemented in the RF domain controls the phase of the transmitted signal [3]–[11]. Analog beamforming is less complex than digital, but the single RF chain can only support a single-user's single-stream transmission. The beam is usually controlled by adjusting analog phase shifters, but the design of high-performance phase shifters in CMOS is quite challenging.

Hybrid beamforming architectures [6], [12]–[16] overcome the limitation of analog beamforming by dividing the process between the analog and digital domains and using a number of RF chains that is much lower than the antenna array size. Single user hybrid precoding was specially designed for mmWave communications in [17] to exploit the limited scattering nature of the mmWave channel with the presence of large antenna arrays. In addition, some

The associate editor coordinating the review of this manuscript and approving it for publication was Irfan Ahmed.

energy-efficiency devices such as low-resolution analog-to-digital converters (ADCs) and digital-to-analog converters (DACs) can also be adopted to reduce the power consumption [18]–[20]. In [17], the authors employ an algorithm based on the simultaneous orthogonal matching pursuit (SOMP) technique [21], [22] where the precoding and combining are done in both the baseband and RF domains. A hybrid precoder and combiner based on the SOMP algorithm can yield a performance close to that of an optimal digital beamformer while considerably decreasing the number of RF chains and hence, $N_{\mathrm{RF}} < N_{\mathrm{t}}$, where N_{RF} and N_{t} are the numbers of RF chains and transmit antennas, respectively. However, the main limitation of solving the sparse optimization problem is the high computational complexity and the need to assume known array geometries for both transmitter and receiver. In [23], the technique was refined, which reduced the computational complexity of the technique discussed in [17] while maintaining the same performance. However, the work in [23] does not address the channel array geometry limitations of [17]. The same approach was used in [24] to approximate the minimum mean squared error for the hybrid solution although the computational complexity is high. In [25], multiple antenna arrays with independent beamforming were employed to obtain diversity gains in mmWave systems. However, this work does not exploit the sparse nature of mmWave channels.

An iterative hybrid design that approaches the performance of the digital precoder was proposed in [26], but the computational complexity is high due to the large number of entries that need to be updated. The low complexity Greedy hybrid precoding/combining, which avoids any assumption on the array geometry or channel structure, is proposed in [27]. Although this method provides good experimental results, no theoretical proof is provided for its performance. The authors of [28] demonstrate that additional performance gains in hybrid precoding/combining can be obtained by optimizing the RF precoder/combiner without restricting them to predefined codebooks. However, this approach comes with higher computational complexity. Several hybrid precoding and combining algorithms were proposed in [29]. One of them, the Hybrid Design by Alternate Minimization (HD-AM) method, achieves high spectral efficiency with medium computational complexity, but it can only work in the special case when the number of data streams equals the number of RF chains. Another algorithm, the Fast Unitary Matching design, requires that the precoders and combiners are dictionary-based. Although its complexity is low, the spectral efficiency is lower than that of the other algorithms. The other algorithms do not make any assumption on the structure of the precoders/combiners, but the computational complexity is high [29]. Several algorithms alternately optimizing the digital precoder and the analog precoder were developed in [30]. The Manifold Optimization Based Hybrid Precoding (MO-AltMin) algorithm provides high spectral efficiency but requires high computational complexity. A low-complexity algorithm for the fully-connected structure (PE-AltMin algorithm) was also deployed with lower complexity. Its spectral efficiency is close to that of MO-AltMin when the number of RF chains equals the number of data streams, but there is a performance gap compared to MO-AltMin when the number of RF chains increases [30]. Sohrabi and Yu [31] provide a heuristic hybrid beamforming design that achieves good performance, but the technique's computational complexity is relatively high compared to other solutions. In [32] and [33], the Euclidean distance between the hybrid precoding matrix and full-digital precoding matrix is minimized by developing a gradient projection algorithm for hybrid beamforming design and solving the matrix factorization problem using gradient and Hessian information, respectively. Although it is a good trick to improve the performance, the main limitation of solving these problems is the high computational complexity. An iterative hierarchical hybrid precoding (HHP-Iterative) that divides the optimization problem into analog and digital parts has been proposed in [34]. It can achieve a spectral efficiency close to that of PE-AltMin, but the high computational complexity is the main limitation of this solution [34]. Furthermore, in multi-user scenarios, the digital beamformer is designed to eliminate inter-user interference and the analog precoder/combiner is designed by maximizing the user's signal power [35]–[38]. Hybrid beamforming with dynamic sub-connected structure and distributed architecture has been investigated in [39]–[46].

Many detection schemes have been applied to the SOMP-based hybrid precoder/combiner design. Sphere decoding (SD), maximum likelihood (ML), zero-forcing (ZF), space- time coding (STC), minimum mean square error (MMSE) detectors, and semi definite relaxation algorithms have been used with the SOMP approach [47]–[49]. In the SOMP-based design, ML detection is optimal in terms BER, but it is impractical due to its complexity which increases exponentially when we increase the number of data streams. This complexity is further increased when high-order modulation is used [47]–[49]. Sphere decoding, which was initially designed to decrease the high detection complexity of the ML detection while yielding a performance close to ML detection, provides a good BER performance for mmWave systems employing SOMP design, but the algorithm's computational complexity also increases exponentially with the number of data streams and modulation order [48], [49]. Hence, reduced complexity linear detection schemes such as ZF and MMSE detectors are preferred. However, compared to ML and SD based systems, their BER performance is noticeably degraded, especially at higher signal-to-noise ratios, and there is a large performance gap between these linear schemes and the SD or ML detectors [47]–[49].

In this paper, we propose a new hybrid precoding/combining design for mmWave systems with equal power allocation across data streams. Additionally, the proposed method provides near-optimal performance in terms of achievable spectral efficiency. The main advantage of

the proposed algorithms is that there is no need to make any assumption on the channel array geometry or channel structure while maintaining low computational complexity. Specifically, we optimize the analog and the digital precoders/combiners separately. Based on the history of the normalized analog precoders/combiners and by updating the digital precoders/combiners, the performance of the proposed hybrid design is close to that of the fully unconstrained digital one. In addition, it is guaranteed that this proposed approach converges to a local optimum point. Simulation results verify that the proposed hybrid precoding/combining scheme outperforms the other popular schemes and achieves a much better performance in terms of the achievable spectral efficiency and BER with low complexity.

The rest of the paper is organized as follows. Section II describes the system and channel models. The design of sparse hybrid beamforming is presented in Section III and the proposed algorithm, in Section IV. The complexity of the proposed algorithm is analyzed in Section V. Simulation results are provided in Section VI. Section VII concludes this paper.

The following notation is used throughout this paper: $\mathbb{C}$ denotes the set of complex numbers; $\mathbf{A}$ is a matrix; $\mathbf{a}$ is a vector; a is a scalar. The conjugate transpose of $\mathbf{A}$ is represented by $\mathbf{A}^{\mathrm{H}}$; $\mathbf{A}_i$ is the i^{th} column of $\mathbf{A}$; $\mathbf{A}_{i,j}$ is the entry in the i^{th} row and j^{th} column of $\mathbf{A}$; $\|\mathbf{A}\|_F$ is the Frobenius norm of $\mathbf{A}$; tr(A) is the trace of $\mathbf{A}$; diag(A) is a vector formed by the diagonal elements of $\mathbf{A}$; $\oslash$ stands for element-wise division; $\mathbf{I_N}$ is the $N \times N$ identity matrix.

II. SYSTEM AND CHANNEL MODELS

A. SYSTEM MODEL

The hybrid precoder at the base station (BS) and combiner at the mobile station (MS) are shown in Fig. 1 with N_{RF}^{BS} RF chains and N_{BS} antennas at the BS, and N_{RF}^{MS} RF chains and N_{MS} antennas at the MS [8]–[11], and [17]. Communication between the BS and a single MS is supported through N_S data streams with $N_S \leq N_{RF}^{BS} \ll N_{\mathrm{BS}}$ in the BS, and $N_S \leq N_{RF}^{MS} \ll N_{MS}$ in the MS. The downlink transmission path at the BS consists of an $N_{RF}^{BS} \times N_S$ baseband precoder $\mathbf{F_{BB}}$ followed by an $N_{BS} \times N_{RF}^{BS}$ RF precoder $\mathbf{F_{RF}}$. The resulting $N_{BS} \times N_S$ hybrid precoder $\mathbf{F}$ is given by $\mathbf{F_{RF}F_{BB}}$. The hybrid combiner $\mathbf{W} \in \mathbb{C}^{N_{MS} \times N_s}$ in the MS is equal to $\mathbf{W_{RF}W_{BB}}$, where $\mathbf{W_{RF}}$ is the RF combiner and $\mathbf{W_{BB}}$ is the baseband combiner.

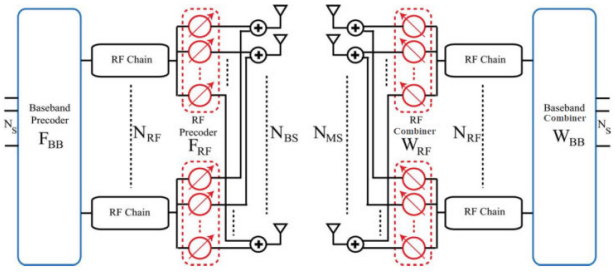


FIGURE 1. Block diagram of the transmitter-receiver single user mmWave system architecture that uses RF phase shifters and baseband beamformers at both ends [17].

The RF precoder and combiner are implemented using analog phase shifters, which are normalized to have the same amplitude, but different phases, i.e., $\left|[\mathbf{F_{RF}}]_{i,j}\right|^2 = \frac{1}{N_{BS}}$ and $\left|[\mathbf{W_{RF}}]_{k,m}\right|^2 = \frac{1}{N_{MS}}$. Hence, all elements of $\mathbf{F_{RF}}$ and $\mathbf{W_{RF}}$ have equal norm [6], [8]–[17]. Additionally, the baseband combiner and precoder are also normalized to satisfy the total power constraint, i.e., $\|\mathbf{F_{RF}F_{BB}}\|_F^2 = \|\mathbf{W_{RF}W_{BB}}\|_F^2 = N_S$ [8]–[11], and [17].

Considering a narrowband block-fading channel model, the signal vector y received at the MS is combined as follows [8]–[11], [17], and [50], [51]

$$\mathbf{y} = \mathbf{W}^{\mathrm{H}}(\sqrt{\mathrm{P_r}}\mathbf{HFs} + \mathbf{n}) \tag{1}$$

where $\mathbf{H}$ is the $N_{MS} \times N_{BS}$ mmWave channel matrix between BS and MS, $\mathbf{s} \in \mathbb{C}^{N_s \mathrm{x} 1}$ is the vector of transmitted symbols with $\mathrm{E}\left[\mathrm{ss}^{\mathrm{H}}\right] = \frac{1}{N_S}\mathbf{I_{N_s}}$, where $\mathbf{I_{N_s}}$ is the N_S by N_S identity matrix, P_r is the average received power, and $\mathbf{n}$ is a $N_{MS} \times 1$ Gaussian noise vector i.i.d, where each element has zero mean and variance σ^2. The uplink can be implemented in the same way as the downlink, with $\mathbf{H} \in \mathbb{C}^{N_{BS} \times N_{MS}}$ and by reversing the roles of the precoders and combiners. In this paper, perfect channel state information is assumed at the BS and MS as explained later, and the effective channels are used to detect the transmitted data streams at the MS given by [35], [52]

$$\mathbf{H_{eff}} = \mathbf{W}^{\mathrm{H}}\mathbf{HF} \tag{2}$$

The dimension of these effective channels is much smaller than that of the original mmWave channel matrix $\mathbf{H}$, depending on the number of data streams N_S [35], [52].

Since the dimension of $\mathbf{H_{eff}}$ is $N_S \times N_S$, low complexity ZF and MMSE detectors are implemented as follows

$$\mathbf{W_{ZF}} = (\mathbf{H_{eff}})^{-1} \tag{3}$$

$$\mathbf{W_{MMSE}} = (\mathbf{H_{eff}^{H}H_{eff}} + \sigma^2 \mathrm{I_{Nr}})^{-1}\mathbf{H_{eff}^{H}} \tag{4}$$

where $\sigma^2 = \mathrm{P}_r/\mathrm{SNR}$. The transmitted symbols $\hat{\mathbf{s}}$ can then be detected as follows

$$\hat{\mathbf{s}} = \mathbf{W_{ZF}y} \tag{5}$$

$$\hat{\mathbf{s}} = \mathbf{W_{MMSE}y} \tag{6}$$

Sphere Decoding and ML scheme use the diagonal matrix $\mathbf{H_{eff}}$ to find the solution $\hat{s}$ with the smallest Euclidean distance from the received signal, but their computational complexity is higher compared to the ZF and MMSE schemes.

Finally, for a given channel $\mathbf{H}$, the spectral efficiency that hybrid precoders/combiners can achieve is given by [8]–[11], and [17]

$$R = \log_2 \left| \mathbf{I_{N_s}} + \frac{P_r}{N_S}\mathbf{R_n}^{-1}\mathbf{W}^{\mathrm{H}}\mathbf{HFF}^{\mathrm{H}}\mathbf{H}^{\mathrm{H}}\mathbf{W} \right| \tag{7}$$

where $\mathbf{R_n} = \sigma^2 \mathbf{W}^H\mathbf{W}$ is the post-combiner noise covariance matrix.

B. CHANNEL MODE

The widely used Saleh-Velenzuela (SV) model is used to represent the limited spatial selectivity or scattering characteristic caused by the high path loss of a mmWave channel in outdoor scenarios [3]–[17]. Therefore, in this paper we adopt the narrowband clustered channel which is given by [17]–[30]

$$\mathbf{H} = \sqrt{N_{BS}N_{MS}/N_{cl}N_{ray}} \times \sum_{i,l} \alpha_{il}\, \Lambda_{\mathrm{r}}(\emptyset_{il}^{\mathrm{r}}, \theta_{il}^{\mathrm{r}})\Lambda_{t}(\emptyset_{il}^{\mathrm{t}}, \theta_{il}^{\mathrm{t}})\mathrm{a}_{\mathrm{r}}(\emptyset_{il}^{\mathrm{r}}, \theta_{il}^{\mathrm{r}})\mathrm{a}_{t}(\emptyset_{il}^{\mathrm{t}}, \theta_{il}^{\mathrm{t}})^{*} \tag{8}$$

where $N_{BS}, N_{MS}, N_{cl}, N_{ray}$ are the number of BS, MS antennas, the number of scattering clusters, and the number of propagation paths respectively. α_{il} is the complex gain of the l^{th} path and the i^{th} cluster and it is assumed to be Rayleigh distributed, i.e., $\alpha_{il} \sim$ i.i.d. $\mathrm{C}\aleph(0, \overline{\mathrm{P}_{\alpha,i}})$ with $\overline{\mathrm{P}_{\alpha,i}}$ the average power of the i^{th} cluster, and $\emptyset_{il}^{\mathrm{t}}\left(\theta_{il}^{\mathrm{t}}\right), \emptyset_{il}^{\mathrm{r}}\left(\theta_{il}^{\mathrm{r}}\right)$ are the l^{th} path's azimuth (elevation) angles of departure and arrival (AODs/AOAs) in the i^{th} cluster, respectively, with uniform distribution. Also, $\Lambda_{\mathrm{t}}\left(\emptyset_{il}^{\mathrm{t}}, \theta_{il}^{\mathrm{t}}\right)$ and $\Lambda_{\mathrm{r}}(\emptyset_{il}^{\mathrm{r}}, \theta_{il}^{\mathrm{r}})$ correspond to the transmit and receive antenna element gain at their departure and arrival angles. In this paper, it is assumed that each antenna element has unity gain. Finally, $\mathrm{a}_{\mathrm{t}}(\emptyset_{il}^{\mathrm{t}}, \theta_{il}^{\mathrm{t}})^{\mathrm{H}}$ and $\mathrm{a}_{\mathrm{r}}(\emptyset_{il}^{\mathrm{r}}, \theta_{il}^{\mathrm{r}})$ are the antenna array response vectors at their azimuth and elevation angles of departure and arrival at the BS and MS respectively. They are usually applied to uniform planar arrays (UPAs), but can be applied to different antennas arrays as well [8]–[11], and [17]. For a UPA [17]

$$\mathrm{a}_{\mathrm{t}}(\emptyset, \theta) = \frac{1}{\sqrt{N_{BS}}}[1, \cdots, e^{\mathrm{j}\left(\frac{2\pi}{\lambda}\right)d(m\sin(\emptyset)\sin(\theta)+n\cos(\theta))}, \cdots, e^{\mathrm{j}\left(\frac{2\pi}{\lambda}\right)d((w-1)\sin(\emptyset)\sin(\theta)+(h-1)\cos(\theta))}]^{\mathrm{T}} \tag{9}$$

where $0 \le m < w$ and $0 \le n < h$ are the width and height indices of an antenna element respectively and the antenna array size is wh. Also, λ is the wavelength of the mmWave signal, and d is the distance between the antenna elements, typically $d = \frac{\lambda}{2}$. The array response vector $\mathrm{a}_{\mathrm{r}}(\emptyset, \theta)$ for MS can also be obtained as above.

III. DESIGN OF SPARSE HYBRID BEAMFORMING

Sparse hybrid precoders/combiners or SOMP based-designs maximize the spectral efficiency R as defined in (7), taking into account the RF precoders/combiners constraint and baseband power constraint [17]. As mentioned in the previous section, the scattering in mmWave channels is limited and sparse hybrid precoders/combiners are built to approximate the unconstrained optimum digital precoder/combiner $\mathbf{F_{opt}}/\mathbf{W_{opt}}$ [17].

Designing the unconstrained optimum digital precoders/combiners, $\mathbf{F_{opt}}/\mathbf{W_{opt}}$, is usually done by using the channel singular value decomposition (SVD) defined by [8]–[11], [17]

$$\begin{aligned}\left[\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\mathbf{H}}\right] &= \mathrm{SVD}\,(\mathbf{H}) \\ &= [\mathbf{U_1}\mathbf{U_2}]\begin{bmatrix}\boldsymbol{\Sigma}_1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_2\end{bmatrix}\begin{bmatrix}\mathbf{V}_1^{\mathbf{H}} \\ \mathbf{V}_2^{\mathbf{H}}\end{bmatrix} \\ &= \mathbf{U_1}\boldsymbol{\Sigma}_1\mathbf{V}_1^{\mathbf{H}} + \mathbf{U_2}\boldsymbol{\Sigma}_2\mathbf{V}_2^{\mathbf{H}}\end{aligned} \tag{10}$$

where $\boldsymbol{\Sigma_1}=(\sigma_1, \ldots, \sigma_{N_s})$ is a diagonal matrix with the largest N_s singular values of $\mathbf{H}$ in decreasing order. Taking the left N_s columns of semi-unitary matrices $\mathbf{U} \in \mathbb{C}^{N_{MS} \times N_{MS}}$ and $\mathbf{V} \in \mathbb{C}^{N_{BS} \times N_{BS}}$ yields

$$\begin{aligned}\mathbf{F_{opt}} &= \mathbf{V_1} \in \mathbb{C}^{N_{BS} \times N_S} \\ \mathbf{W_{opt}} &= \mathbf{U_1^H} \in \mathbb{C}^{N_S \times N_{MS}}\end{aligned} \tag{11}$$

which are the fully digital SVD-based precoder and combiner, respectively. Therefore, the sparse hybrid precoder is found by approximating the unconstrained optimal precoder $\mathbf{F_{opt}}$ as follows [17]

$$\begin{aligned}(\mathbf{F_{BB}}) = \arg\min_{\mathbf{F_{BB}}} &\left\|\mathbf{F_{opt}} - \mathbf{A_{BS}}\mathbf{F_{BB}}\right\|_F \\ \text{s.t.} &\left\|\mathrm{diag}(\mathbf{F_{BB}}\mathbf{F_{BB}^H})\right\|_{\mathbf{0}} = N_{BS}^{RF} \\ &\|\mathbf{A_{BS}}\mathbf{F_{BB}}\|_F^2 = N_S\end{aligned} \tag{12}$$

where $\mathbf{A_{BS}}$ is the set of feasible quantized RF precoding codebooks with constant-magnitude entries which can be implemented in the RF circuitry using analog phase shifters. The design of RF precoding/combining codebooks $\mathbf{A_{BS}}$ have the same form of the array response of the mmWave channel and thus can be quantized in the azimuth and elevation angles with angular resolution L, where $L = 2^{N_\emptyset + N_\theta}$, and $N_\emptyset$ and N_θ are the numbers of azimuth and elevation quantization bits, respectively [17]. The analog combining codebook $\mathbf{A_{MS}}$ can be defined in the same way. However, to build the quantized RF precoding codebooks, some assumptions on the mmWave channel structure and array geometry have to be made. The sparsity constraint $\left\|\mathrm{diag}\left(\mathbf{F_{BB}}\mathbf{F_{BB}^H}\right)\right\|_{\mathbf{0}} = N_{BS}^{RF}$ states that $\mathbf{F_{BB}}$ cannot have more than N_{BS}^{RF} non-zero rows. Equation (12) can be solved by the well-known concept of orthogonal matching pursuit (OMP) [17]. It works by finding the vector in the RF precoding codebooks matrix $\mathbf{A_{BS}}$ along which the optimal precoder $\mathbf{F_{opt}}$ has the maximum projection. It then appends the selected column vector to the RF precoder $\mathbf{F_{RF}}$ matrix. After that, $\mathbf{F_{BB}}$ is calculated using the least squares solution and the contribution of the selected vector from $\mathbf{A_{BS}}$ is removed. The process continues until all N_{BS}^{RF} beamforming vectors have been selected from RF precoding codebooks matrix $\mathbf{A_{BS}}$. At the end of the N_{BS}^{RF} iterations, the RF precoding matrix $\mathbf{F_{RF}}$, and the baseband precoder $\mathbf{F_{BB}}$, which minimize $\left\|\mathbf{F_{opt}} - \mathbf{F_{RF}}\mathbf{F_{BB}}\right\|_F$, are found. Finally, $\mathbf{F_{BB}}$ is normalized so that the transmit power constraint is satisfied, and the sparse hybrid precoder $\mathbf{F} = \mathbf{F_{RF}}\mathbf{F_{BB}}$ is returned. The sparse hybrid combiner is designed by maximizing the data rate using $\mathbf{W_{opt}}$ or minimizing the mean squared error between transmitted and received signals ($\mathbf{W_{MMSE}}$) [17].

The main drawback of the sparse hybrid design is the use of the set of feasible quantized RF precoding codebooks and the costly correlation operations. Moreover, the accuracy of the algorithm depends on the angular resolution L. Therefore, in this paper we propose an algorithm to solve (12) without making any assumption on the array geometry and without any other constraint, such as the number of streams must be equal to the number of RF chains, or the baseband precoder/combiner is scaled unitary, while achieving a good spectral efficiency with low computational complexity compared to the other popular designs in the literature.

IV. PROPOSED HYBRID BEAMFORMING

In this section, we propose a low complexity-precoding algorithm with equal power allocation per stream. In addition, we do not assume any constraint on the optimization problem, which is related to (12). The derivation of the combiner is similar. We note that the optimal unconstrained semi-unitary precoder for $\mathbf{H}$ is simply given by $\mathbf{F_{opt}} = \mathbf{V_1}$. Also, we need the hybrid precoder $\mathbf{F_{RF}F_{BB}}$ to be sufficiently "close" [17] to the optimal precoder $\mathbf{V_1}$ by using its digital precoder to construct linear combinations of the RF precoder vectors. By formulating an optimization problem, we first need to find the baseband precoder $\mathbf{F_{BB}}$ that minimizes the Euclidean distance by using the initialization of the proposed RF precoder, which is calculated by taking the first N_s columns from $\mathbf{F_{opt}}$ and then normalizing them such that each entry has constant magnitude, i.e., $\mathbf{F_{RF}}=\mathbf{F_{opt}}\oslash\left(\left|\mathbf{F_{opt}}\right|\sqrt{N_{BS}}\right)$. One should note that the element-wise normalization of $\mathbf{F_{RF}}$ satisfies the normalization constraint $\left|[\mathbf{F_{RF}}]_{i,j}\right|^2 = \frac{1}{N_{BS}}$ [17]. We then find the RF precoder $\mathbf{F_{RF}}$ such that the hybrid precoder $\mathbf{F_{RF}F_{BB}}$ is sufficiently "close" to the optimal unconstrained digital precoder $\mathbf{V_1}$. Specifically, we would like to solve the following alternating optimization problem first, which is related to (12):

$$(\mathbf{F_{BB}}) = \arg\min_{\mathbf{F_{BB}}} \left\|\mathbf{F_{opt}}-\mathbf{F_{RF}F_{BB}}\right\|_F^2 \tag{13}$$

The objective function can be expanded as follows

$$\begin{aligned}&\left\|\mathbf{F_{opt}}-\mathbf{F_{RF}F_{BB}}\right\|_F^2\\ &= \mathrm{tr}\left(\mathbf{F_{opt}^H F_{opt}}\right) - 2\mathrm{tr}\left(\mathbf{F_{opt}^H F_{RF} F_{BB}}\right) + \mathrm{tr}(\mathbf{F_{BB}^H F_{RF}^H F_{RF} F_{BB}})\\ &= N_S - 2\mathrm{tr}\left(\mathbf{F_{opt}^H F_{RF} F_{BB}}\right) + \mathrm{tr}(\mathbf{F_{BB}^H F_{RF}^H F_{RF} F_{BB}})\end{aligned} \tag{14}$$

To minimize over $\mathbf{F_{BB}}$, we set the derivative of (14) with respect to $\mathbf{F_{BB}}$ equal to zero, which yields the following minimized proposed baseband precoder $\mathbf{F_{BB}}$ (least squares solution)

$$\mathbf{F_{BB}} = \left(\mathbf{F_{RF}^H F_{RF}}\right)^{-1}\mathbf{F_{RF}F_{opt}} \tag{15}$$

Then, we keep $\mathbf{F_{BB}}$ fixed and solve the same optimization problem, but now minimizing over $\mathbf{F_{RF}}$

$$(\mathbf{F_{RF}}) = \arg\min_{\mathbf{F_{RF}}} \left\|\mathbf{F_{opt}}-\mathbf{F_{RF}F_{BB}}\right\|_F^2 \tag{16}$$

Similarly to (14), expanding the objective function yields

$$\begin{aligned}&\left\|\mathbf{F_{opt}}-\mathbf{F_{RF}F_{BB}}\right\|_F^2\\ &= N_S - 2\mathrm{tr}\left(\mathbf{F_{opt}^H F_{RF} F_{BB}}\right) + \mathrm{tr}(\mathbf{F_{BB}^H F_{RF}^H F_{RF} F_{BB}})\end{aligned} \tag{17}$$

We again set the derivative of (17) with respect to $\mathbf{F_{RF}}$ equal to zero, which yields the following

$$\mathbf{F_{RF}F_{BB}F_{BB}^H} = \mathbf{F_{opt}F_{BB}^H} \tag{18}$$

The problem in (18) is that $\mathbf{F_{BB}F_{BB}^H}$ cannot be inverted when $N_S < N_{RF}^{BS}$. Thus, to solve it we add $\mathbf{F_{RF}}$ to both sides of (18) such that

$$\begin{aligned}\mathbf{F_{RF}} &= \mathbf{F_{opt}F_{BB}^H} - \mathbf{F_{RF}F_{BB}F_{BB}^H} + \mathbf{F_{RF}}\\ \mathbf{F_{RF}} &= (\mathbf{F_{opt}} - \mathbf{F_{RF}F_{BB}})\mathbf{F_{BB}^H} + \mathbf{F_{RF}}\\ \mathbf{F_{RF}} &= \mathbf{F_{res}F_{BB}^H} + \mathbf{F_{RF}}\end{aligned} \tag{19}$$

where the residual precoding matrix $\mathbf{F}_{\mathrm{res}} = \mathbf{F_{opt}} - \mathbf{F_{RF}F_{BB}}$. Equation (19) yields an iterative solution for $\mathbf{F_{RF}}$ such that in the current iteration, the updated $\mathbf{F_{RF}}$ is equal to $\mathbf{F_{RF}}$ from the previous iteration plus $\mathbf{F_{res}F_{BB}^H}$. Moreover, (19) satisfies the property of the gradient descent method with a step size equal to one and without vectorizing the RF precoder $\mathbf{F_{RF}}$ [32]. Thus, $\mathbf{F_{RF}}$ is guaranteed to converge to a feasible local optimal solution. This solution can also be applied when $N_S = N_{RF}^{BS}$. Note that when $N_S < N_{RF}^{BS}$, we need to complete the $N_{BS} \times N_{RF}^{BS}\mathbf{F_{RF}}$ after the initialization. In each iteration, we add a column to $\mathbf{F_{RF}}$ which leads to the highest reduction of the residual. This column can be selected from the basis of the range of the residual matrix, which is the element-wise normalization of the first singular vector of the residual.

The pseudo-code for the proposed hybrid precoder $\mathbf{F^P}$ solution is given in Algorithm 1. In a mmWave system using the hybrid precoding design, the BS or MS can support up to N_{RF} data streams, i.e., $N_S \leq \min(N_{RF}^{BS}, N_{RF}^{MS})$ [8]–[11], and [17]. The inputs of the algorithm are $\mathbf{F_{opt}} \in \mathbb{C}^{N_{BS}\times N_S}$ and the maximum number of iterations K, where $K \geq N_{RF}^{BS}$ - N_S when $N_S < N_{RF}^{BS}$, in order to calculate the $N_{BS} \times N_{RF}^{BS}\mathbf{F_{RF}}$ matrix, and $K \geq 1$ when $N_{RF}^{BS} = N_S$. In the general case of $N_S \geq 1$ where $N_S \leq N_{RF}^{BS}$, the algorithm starts by initializing $\mathbf{F_{RF}}$ with the element-wise normalization of the first N_s columns of $\mathbf{F_{opt}}$ i.e., $\mathbf{F_{RF}}= \mathbf{F_{opt}}\oslash\left(\left|\mathbf{F_{opt}}\right|\sqrt{N_{BS}}\right)$. Then, $\mathbf{F_{BB}}$ is computed using least squares in step 2. After that, the algorithm proceeds to calculate the residual precoding matrix $\mathbf{F}_{\mathrm{res}}$ and the proposed RF precoder $\mathbf{F_{RF}}$ in steps 3 and 4 respectively. Step 5 ensures that the proposed RF precoder $\mathbf{F_{RF}}$ is satisfied exactly with constant-magnitude entries which can be applied at RF using analog phase shifters. In steps 7 and 8, when $N_S < N_{RF}^{BS}$, we need to complete the $N_{BS} \times N_{RF}^{BS}\mathbf{F_{RF}}$, which can be done by adding the element-wise normalization of the first singular vector of the residual matrix $\mathbf{F}_{\mathrm{res}}$ to $\mathbf{F_{RF}}$. After K iterations the process is completed and the algorithm will find the $N_{BS} \times N_{RF}^{BS}$ proposed RF precoding matrix $\mathbf{F_{RF}}$ and the optimal $N_{RF}^{BS} \times N_S$ baseband precoder $\mathbf{F_{BB}}$ such that $\left\|\mathbf{F_{opt}}-\mathbf{F_{RF}F_{BB}}\right\|_F$ is

Algorithm 1 Proposed Hybrid Precoding.

Input: The optimum unconstrained solution $\mathbf{F_{opt}} \in \mathbb{C}^{N_{BS} \times N_S}$ and the maximum number of iterations K.

Output: Analog $\mathbf{F_{RF}} \in \mathbb{C}^{N_{BS} \times N_{RF}^{BS}}$ with the element-wise normalization and baseband $\mathbf{F_{BB}} \in \mathbb{C}^{N_{RF}^{BS} \times N_S}$ such that $\left\|\mathbf{F_{opt}} - \mathbf{F^P}\right\|_F$ is reduced and $\left\|\mathbf{F^P}\right\|_F^2 = N_S$, where $\mathbf{F^P} = \mathbf{F_{RF}}\mathbf{F_{BB}}$.

Initialization: analog precoder $\mathbf{F_{RF}} = \mathbf{F_{opt}} \oslash \left(\left|\mathbf{F_{opt}}\right| \sqrt{N_{BS}}\right)$.

1: for $i = 1{:}K$ do
2: Update: $\mathbf{F_{BB}} = \left(\mathbf{F_{RF}^H F_{RF}}\right)^{-1} \mathbf{F_{RF}^H F_{opt}}$
3: Update the residual: $\mathbf{F_{res}} = \mathbf{F_{opt}} - \mathbf{F_{RF}F_{BB}}$
4: Update: $\mathbf{F_{RF}} = \mathbf{F_{res}F_{BB}^H} + \mathbf{F_{RF}}$
5: Element-Wise Normalization: $\mathbf{F_{RF}} = \mathbf{F_{RF}} \oslash \left(\left|\mathbf{F_{RF}}\right| \sqrt{N_{BS}}\right)$
6: If $i \leq N_{RF}^{BS} - N_S$
7: $\mathbf{F_{res}} = \mathbf{U}\Sigma\mathbf{V^H}$
8: Append the element-wise normalization of the first vector of $\mathbf{U}$ as a new column to $\mathbf{F_{RF}}$: $\mathbf{F_{RF}} = \left[\mathbf{F_{RF}} \left|\mathbf{U}_1 \oslash \left(\left|\mathbf{U}_1\right| \sqrt{N_{BS}}\right)\right.\right]$
9: end if
10: end for
11: $\mathbf{F_{BB}} = \left(\mathbf{F_{RF}^H F_{RF}}\right)^{-1} \mathbf{F_{RF}^H F_{opt}}$
12: $\mathbf{F_{BB}} = \sqrt{N_S} \frac{\mathbf{F_{BB}}}{\left\|\mathbf{F_{RF}^P F_{BB}}\right\|_F}$
13: return $\mathbf{F^P} = \mathbf{F_{RF}F_{BB}}$

minimized. In steps 12 and 13, we ensure that the transmit power constraint is satisfied and return the proposed hybrid precoder $\mathbf{F^P} = \mathbf{F_{RF}F_{BB}}$. The proposed hybrid combiner $\mathbf{W_{RF}^P}$ can be calculated in the same way.

Remark 1 - Convergence of the Proposed Hybrid Precoder to Local Minimum Points: Note that when $N_S = N_{RF}^{BS}$ or $N_S < N_{RF}^{BS}$, $\mathbf{F_{BB}}$ is a square matrix that is approximately unitary $\mathbf{F_{BB}^H F_{BB}} \approx \mathbf{F_{BB}F_{BB}^H} \approx \mathbf{I_{N_S}}$ or a non-square matrix that is approximately semi-unitary $\mathbf{F_{BB}^H F_{BB}} \approx \mathbf{I_{N_s}}$, respectively [17], [29]. Thus, each iteration in Algorithm 1 minimizes the objective function $\left\|\mathbf{F_{opt}} - \mathbf{F_{RF}F_{BB}}\right\|_F$ and the error term decreases monotonically with each iteration. Since the objective function has a lower bound, the proposed method must converge to local optimum points. In Section VI, simulations results will confirm that the objective function decreases monotonically with each iteration and it tends asymptotically to its lower bound.

In Table 1, we compare the proposed hybrid design with other popular approaches in the literature. The hybrid design by alternating minimization (HD-AM) method [29] assumes two additional constraints. The first one is that the number of streams must be equal to the number of RF chains, and the second one is that the baseband precoder is scaled unitary and designed by orthonormal Procrustes solution, which is given by $\mathbf{F_{BB}} = \mathbf{VU^H}$ with the use of the singular value decomposition of $\mathbf{F_{opt}^H F_{RF}} = \mathbf{U}\Sigma\mathbf{V^H}$ where the initial precoder is given by $\mathbf{F_{RF}} = \mathbf{F_{opt}} \oslash \left(\left|\mathbf{F_{opt}}\right|\right)$. Moreover, the solution of analog precoder is based on solving an optimization problem which is given by $\mathbf{F_{RF}} = \mathbf{F_{opt}F_{BB}^H} \oslash \left(\left|\mathbf{F_{opt}F_{BB}^H}\right|\right)$. This process is repeated until the maximum number of iterations M is reached. After the last iteration, $\mathbf{F_{BB}}$ is updated via the least squares. The same method can be applied for the hybrid combiners. In the sparse hybrid precoder [17], the vectors of the analog precoders are found from the RF precoding codebooks matrix $\mathbf{A_{BS}}$ along which the optimal precoder $\mathbf{F_{opt}}$ has the maximum projection, and the baseband precoder is calculated using the least squares solution. The same method can be applied for the hybrid combiners or by minimizing the mean square error between transmitted and received signals [17]. For the greedy hybrid precoding [27], $\mathbf{F_{BB}}$ is computed using the least squares solution after the initialization of the analog precoders, i.e., $\mathbf{F_{RF}} = \mathbf{F_{opt}} \oslash \left(\left|\mathbf{F_{opt}}\right|\right)$. If the number of streams is smaller than the number of RF chains, the columns of $\mathbf{F_{RF}}$ are selected from the element-wise normalization of the first singular vectors of the residual matrices. The same method can be applied for the hybrid combiners. The proposed algorithm is based on an alternating optimization procedure, which leads to direct solution without assuming any additional constraint. Moreover, the proposed solution can be applied in the general case (the number of RF chains can be greater than or equal to the number of data streams). The design of baseband precoder is not made exactly unitary unlike the HD-AM approach. Specifically, after the initialization of the analog precoder i.e., $\mathbf{F_{RF}} = \mathbf{F_{opt}} \oslash \left(\left|\mathbf{F_{opt}}\right| \sqrt{N_{BS}}\right)$, the baseband precoder of the proposed design is solved using the least squares solution, but it is approximately unitary or approximately semi-unitary [17], [29]. After that, the analog precoder is designed by solving an optimization problem separately, which is given by $\mathbf{F_{RF}} = \mathbf{F_{res}F_{BB}^H} + \mathbf{F_{RF}}$. Then, the analog precoder $\mathbf{F_{RF}}$ is normalized element-wise to satisfy the hardware constraint, i.e., $\mathbf{F_{RF}} = \mathbf{F_{RF}} \oslash \left(\left|\mathbf{F_{RF}}\right| \sqrt{N_{BS}}\right)$ where each entry has constant magnitude, which can be implemented in the RF circuitry using analog phase shifters. Notice that the $\sqrt{N_{BS}}$ in the normalization step is a good way to make the diagonal of $\mathbf{F_{RF}^H F_{RF}}$ normalized to one, which is needed because $\mathbf{F_{opt}^H F_{opt}} = \mathbf{I_{N_s}}$. Hence, its factors should also have a semi-unitary structure [29]. The analog precoder $\mathbf{F_{RF}}$ can be normalized again to the unit modulus between step 10 and 11 in Algorithm 1 and we have verified that the results are not affected by that. This process is repeated up to K iterations. After the last iteration, $\mathbf{F_{BB}}$ is updated via the least squares. The same method can be applied for the hybrid combiners.

V. COMPLEXITY ANALYSIS OF THE PROPOSED ALGORITHM

In this section, we analyze the complexity in implementing the proposed hybrid precoding/combining design using Algorithm 1 as compared to previous methods from the literature. To simplify the complexity analysis, let us denote $N = L = \max\{N_{BS}, N_{MS}\}$, $N_{\mathrm{RF}} = \max\left\{N_{RF}^{BS}, N_{RF}^{MS}\right\}$, K is

TABLE 1. Comparison between the proposed method and other approaches.

Method	Sparse Hybrid Design [17]	Greedy Hybrid Design [27]	Hybrid Design by Alternating Minimization (HD-AM) [29]	Proposed Hybrid Design
Year	2014	2015	2016	2021
Constraints	RF precoding /combining codebooks	None	1. $N_S = N_{RF}$ 2.Baseband precoders /combiners are exactly unitary	None
Initial analog precoder	None	$\mathrm{F_{RF}} = \mathrm{F_{opt}} \oslash \left(\left\lvert \mathrm{F_{opt}} \right\rvert\right)$	$\mathrm{F_{RF}} = \mathrm{F_{opt}} \oslash \left(\left\lvert \mathrm{F_{opt}} \right\rvert\right)$	$\mathrm{F_{RF}} = \mathrm{F_{opt}} \oslash \left(\left\lvert \mathrm{F_{opt}} \right\rvert \sqrt{N_{\mathrm{BS}}}\right)$
Baseband precoder design	Least squares solution $\mathrm{F_{BB}} = (\mathrm{F_{RF}^H F_{RF}})^{-1} \mathrm{F_{RF}^H F_{opt}}$	Least squares solution $\mathrm{F_{BB}} = (\mathrm{F_{RF}^H F_{RF}})^{-1} \mathrm{F_{RF}^H F_{opt}}$	$\mathrm{F_{opt}}^{\mathrm{H}} \mathrm{F_{RF}} = \mathrm{U}\,\Sigma\,\mathrm{V}^{\mathrm{H}}$ $\mathrm{F_{BB}} = \mathrm{V}\,\mathrm{U}^{\mathrm{H}}$	Least squares solution $\mathrm{F_{BB}} = (\mathrm{F_{RF}^H F_{RF}})^{-1} \mathrm{F_{RF}^H F_{opt}}$
Analog precoder design	The vectors of $\mathrm{F_{RF}}$ are found from the RF precoding codebooks matrix $\mathrm{A_{BS}}$ along which the optimal precoder $\mathrm{F_{opt}}$ has the maximum projection	When $N_S < N_{RF}$, the columns of $\mathrm{F_{RF}}$ are selected from the element-wise normalization of the first singular vectors of the residual matrices $\mathrm{F_{res}}$	$\mathrm{F_{RF}} = \mathrm{F_{opt}} \mathrm{F_{BB}}^{\mathrm{H}} \oslash \left(\left\lvert \mathrm{F_{opt}} \mathrm{F_{BB}}^{\mathrm{H}} \right\rvert\right)$	$\mathrm{F_{RF}} = \mathrm{F_{res}} \mathrm{F_{BB}^H} + \mathrm{F_{RF}}$ $\mathrm{F_{RF}} = \mathrm{F_{RF}} \oslash \left(\left\lvert \mathrm{F_{RF}} \right\rvert \sqrt{N_{BS}}\right)$ When $N_S < N_{RF}$, the columns of $\mathrm{F_{RF}}$ are selected from the element-wise normalization of the first singular vectors of the residual matrices $\mathrm{F_{res}}$
Max # of iterations	N_{RF}	$N_{RF} - N_S$	M	K
Baseband/Analog combiner design	Find $\mathrm{W_{RF}}$ and $\mathrm{W_{BB}}$ that minimize the mean-squared-error between transmitted and received signals or use same design technique as Baseband/Analog precoder.	Same as Baseband/Analog precoder design	Same as Baseband/Analog precoder design	Same as Baseband/Analog precoder design

the maximum number of iterations of the proposed hybrid design, and M is the maximum number of iterations of the hybrid design by alternating minimization (HD-AM) [29]. In Table 2, we present the complexity analysis by evaluating the total number of floating-point operations (flops) for each hybrid precoding/combining approach.

The primary factor in the sparse hybrid precoder/combiner design algorithm's complexity depends on the selection of RF precoders/combiners [17]. The complexity increases when the process involves an exhaustive search for the codebooks at both the BSs and the MSs. The sparse hybrid precoders/combiners need higher angular resolution L to improve the performance. Therefore, the complexity of the sparse hybrid design is $\mathrm{O}(N^2 N_{RF} N_S)$ operations for a resolution $\mathrm{O}(N)$. Because of the angular resolution and the correlation function for sparse hybrid design, the proposed design, greedy hybrid design [27], and the HD-AM provide a great complexity reduction especially when $N_S \ll N$.

The construction of the proposed hybrid design only requires complex matrix multiplication and element wise division operations. Moreover, the computational complexity of the proposed hybrid design is the same as the HD-AM when the maximum number of iterations is the same, i.e., $K = M$, but the drawback of the HD-AM is that it can only be used when the number of RF chains equals the number of data streams, $N_S = N_{RF}$. The greedy hybrid design and the proposed hybrid design have almost similar computational complexity and our hybrid design only needs a small number of iterations K to outperform the greedy hybrid design.

Therefore, the complexity of the proposed hybrid precoders/combiners is lower than that of the sparse hybrid precoders/combiners, similar to the HD-AM with the same performance when the maximum number of iterations is the same, i.e., $K = M$, and close to the greedy hybrid design with better performance.

TABLE 2. Complexity of the proposed algorithm compared to previous methods from the literature.

Method	Constraints	Complexity
Sparse Hybrid Design [17]	RF precoding/Combining codebooks	$O(N^2 N_{RF} N_S)$
Hybrid Design by Alternating Minimization (HD-AM) [29]	$N_S = N_{RF}$	$O(NN_S^2 M)$
Greedy Hybrid Design [27]	None	$O(NN_{RF}^2 N_S)$
Proposed Hybrid Design	None	$O(NN_{RF}^2 K)$

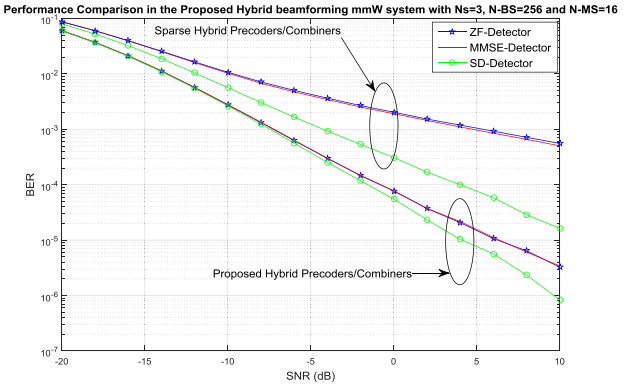


FIGURE 2. BER performance for uncoded 16 QAM single-cell proposed hybrid precoders/combiners with $K = 6$ compared to the sparse hybrid precoders/combiners [17], [48], [49] in mmWave systems for $N_{BS} = 256$ and $N_{MS} = 16$ with 3 and 3 RF chains respectively, and $N_S = 3$.

VI. SIMULATION RESULTS

This section presents the numerical results to show the performance advantages of the proposed hybrid precoders /combiners implemented as described in Algorithm 1. Specifically, we present the BER performance of the ZF, MMSE, ML, and sphere decoding. Moreover, we provide the sum rate for the proposed design compared to other popular methods from the literature. We show numerical simulations of the proposed method's performance when maximizing the spectral efficiency as defined in (7). For all detection algorithms, the QPSK and 16 QAM modulation schemes are used without error-control coding. In addition, for all cases, the mmWave channel remains constant during the transmission of one block of data.

In these simulations, we use the system architecture presented in Fig. 1. We consider the case where there is only one BS and one MS at a distance of 100 meters. The spacing between antenna elements is equal to $\lambda/2$. The system is assumed to operate at a 28 GHz carrier frequency in an outdoor scenario, and with a path loss exponent $n = 3.4$. The channel model is described in (8), with $\overline{P_{\alpha,i}} = 1$ for all clusters. The azimuth and elevation angles AoAs/AoDs of the rays within a cluster are assumed to be randomly Laplacian distributed. The AoAs/AoDs azimuths and elevations of the cluster means are assumed to be uniformly distributed. We use the AoD/AoA beamforming codebooks (exact array response of mmWave channel) at the BSs and MSs respectively for the sparse hybrid design. The SNR in all the plots is defined as SNR $= P_r/\sigma^2$. We assume perfect CSI at the BS and MS. For fairness, the same total power constraint is enforced on all precoding/combining solutions.

In Fig. 2 and Fig. 3, the channel model has five paths, which is equivalent to 5 clusters with an angular spread of zero. The AoAs/AoDs azimuths of the cluster means are assumed to be uniformly distributed in $[0, 2\pi]$, while their AoAs/AoDs elevations are uniformly distributed in $[\frac{-\pi}{2}, \frac{\pi}{2}]$. In Fig. 2, we evaluate the BER performance of the ZF, MMSE, and sphere decoding algorithms for a 16 QAM 256×16 UPAs mmWave systems with $N_{RF}^{BS} = N_{RF}^{MS} = 3$ RF chains and $N_S = 3$ data streams.

Fig. 2 shows that the BER performance is significantly improved by using the proposed hybrid precoding/combining design with $K = 6$ compared to the BER performance of the sparse hybrid precoding/combining designs [17], and [48], [49]. Furthermore, the performance gap between the sphere decoding algorithm and the ZF and MMSE detectors with the proposed hybrid precoding/combining over the whole range of SNR is considerably reduced compared to the performance gap of the same schemes with the sparse hybrid precoding/combining designs [17], and [48], [49]. The sphere decoding algorithm, ZF and MMSE detectors perform similarly over the whole range of SNR for the proposed hybrid beamforming design, but the ZF and MMSE detectors perform significantly worse than the SD when the sparse hybrid beamformer is used, especially for a BER below 10^{-2}. A bit error probability of 10^{-5} can be obtained with SNR $= 4dB$ using sphere decoding with the proposed hybrid precoding/combining design, whereas the ZF and MMSE detectors with the proposed hybrid precoding/combining design need about 6 dB to obtain the same BER, which represents a small difference of 2 dB. However, with the sparse hybrid precoding/combining design, the difference between sphere decoding and ZF or MMSE exceeds 15 dB to obtain a BER of 10^{-4}. Hence, the proposed beamformer allows the use of the ZF or MMSE detectors, which have a significantly lower computational complexity than SD [48], [49], without incurring the performance penalty observed with the sparse hybrid precoder/combiner.

Fig. 3 shows the performance with the same system parameters used in Fig. 2, but with QPSK modulation and ML detection instead of sphere decoding. Similarly to Fig. 2, the proposed hybrid precoding/combining design with $K = 6$ significantly improves the BER performance compared to the sparse hybrid precoding/combining design [17], and [48], [49]. It can be seen from Fig. 3 that by using the proposed hybrid precoding/combining design with $K = 6$ the ZF and MMSE detectors can achieve almost the same performance as the ML detector for low SNRs, with a small degradation for high SNRs. In contrast, the performance gap between the ML detector and the ZF and MMSE detectors with the sparse hybrid precoding/combining design is

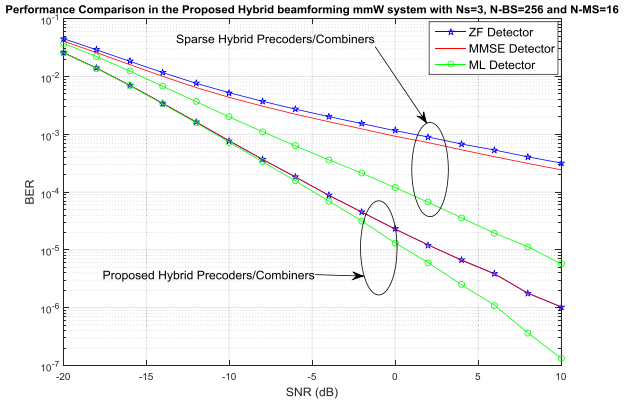


FIGURE 3. BER performance for uncoded QPSK single-cell proposed hybrid precoders/combiners with $K = 6$ compared to the sparse hybrid precoders/combiners [17], [48], [49] in mmWave systems for $N_{BS} = 256$ and $N_{MS} = 16$ with 3 and 3 RF chains respectively, and $N_S = 3$.

significant and increases with SNR. Hence, the proposed hybrid precoding/combining design with $K = 6$ can reduce considerably the performance gap between the ML detector and the ZF and MMSE detectors compared to the sparse hybrid precoding/combining design. Furthermore, there is a small difference between the ZF and MMSE detectors with the sparse hybrid precoding/combining designs, whereas their performance is identical with the proposed hybrid precoding/combining designs. Notice that the complexity of sphere decoding increases with the number of data streams and modulation order. Although the speed of sphere decoding (SD) increases in the high SNR region, it still requires a considerably longer time compared to the low complexity ZF/MMSE detectors for any SNR [48], [49].

In Figs. 4 and 5, the channel model has $N_{cl} = 8$ clusters, and the number of rays $N_{ray} = 10$ per cluster with an angular spread of 7.5° which is the same at the transmitter and receiver. Also, the transmitter's sector angle is 60° wide in the azimuth domain and 20° wide in elevation, but the receivers have relatively smaller antenna arrays of omni-directional elements. In Fig. 4, we evaluate the BER performance of the ZF detector using the proposed hybrid precoders/combiners with different numbers of iterations K, the sparse hybrid precoders/combiners [17], the unconstrained digital precoders/combiners, and the greedy hybrid precoders/combiners [27] for a 16 QAM 64×16 UPAs mmW systems with $N_{RF}^{BS} = N_{RF}^{MS} = 4$ RF chains and $N_S = 3$ data streams. Fig. 4 shows that the ZF detector with the proposed hybrid precoders/combiners achieves a BER, which is very close to that of the unconstrained digital precoders/combiners and performs better than the ZF detector with sparse hybrid precoders/combiners over the whole range of SNR. The ZF detector with the proposed hybrid precoders/combiners performs similarly to the ZF detector with the greedy hybrid precoders/combiners with the smallest number of iteration $K = N_{RF}^{BS} - N_S = 1$, but outperforms the greedy hybrid precoders/combiners when K is increased. Moreover, as we can see from Fig. 4, a number of iterations $K = 6$ or 10 is sufficient to outperform the other hybrid designs and to achieve a performance very close to that of the unconstrained digital precoders/combiners. Furthermore, the performance improvement of the proposed hybrid precoders/combiners is noticeably fast by increasing K from 1 to 6, but it becomes slow after $K = 6$. Notice that the HD-AM [29] cannot be applied when the number of RF chains is greater than the number of data streams.

Fig. 5 shows the BER performance of the ZF detector with the proposed hybrid precoders/combiners with different numbers of iterations K, ZF detector with the sparse hybrid precoders/combiners, ZF detector with the unconstrained digital precoders/combiners, ZF detector with the greedy hybrid

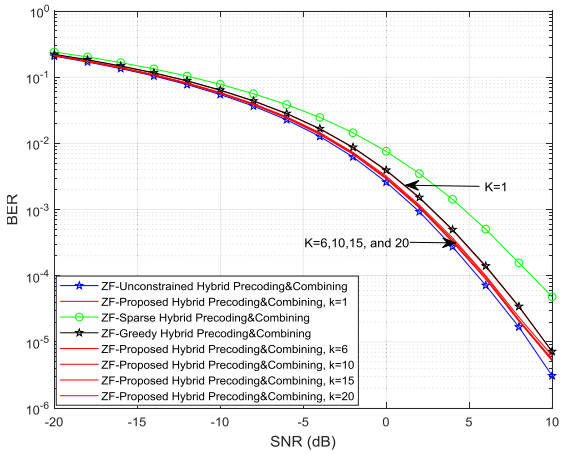


FIGURE 4. BER performance achieved by the proposed design with different K compared to the sparse hybrid precoders/combiners [17], the optimal unconstrained digital precoders/combiners, and the greedy hybrid precoders/combiners [27] for ZF detector in uncoded 16 QAM single-cell in UPAs mmWave systems for $N_{BS} = 64$ and $N_{MS} = 16$ with 4 and 4 RF chains respectively, and $N_S = 3$.

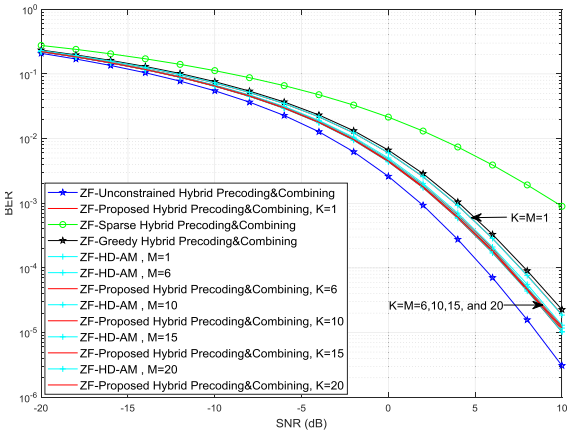


FIGURE 5. BER performance achieved by the proposed design with different K compared to the sparse hybrid precoders/combiners [17], the optimal unconstrained digital precoders/combiners, the greedy hybrid precoders/combiners [27], and the HD-AM [29] with different M for ZF detector in uncoded 16 QAM single-cell in UPAs mmWave systems for $N_{BS} = 64$ and $N_{MS} = 16$ with 3 and 3 RF chains respectively, and $N_S = 3$.

precoders/combiners, and ZF detector with the HD-AM [29] with different numbers of iterations M for a 16 QAM 64 × 16 UPAs mmW system with $N_{RF}^{BS} = N_{RF}^{MS} = 3$ RF chains and $N_S = 3$ data streams. For the sake of fairness, we use the same number of iterations for the proposed hybrid precoders/combiners and the HD-AM design, i.e., $K = M$. As we can see from Fig. 5, the ZF detector with the proposed hybrid design and the ZF detector with HD-AM are overlapped and outperform the ZF detector with sparse hybrid design and the ZF detector with greedy hybrid design by increasing the number of iterations K and M. The improvement over the sparse hybrid design is significant and exceeds 6 dB even with $K = M = 1$. In addition, the number of iterations K or M should be equal to 6 since the performance is improved compared to $K = M = 1$, but after $K = M = 6$ the performance of proposed hybrid precoders/combiners and HD-AM improves slowly.

Figs. 6, 7, and 8 show the simulated sum rate of the proposed hybrid precoders/combiners with different numbers of iterations, sparse hybrid precoders/combiners, optimal unconstrained digital precoders/combiners, greedy hybrid design, and HD-AM with different number of iterations M. The channel model has $N_{cl} = 8$ clusters and the number of rays $N_{ray} = 10$ per cluster with an angular spread of 7.5° which is the same at the transmitter and receiver. Also, the transmitter's sector angle is 60° wide in the azimuth domain and 20° wide in elevation, but the receivers have relatively smaller antenna arrays of omni-directional elements.

Fig. 6 shows the spectral efficiency achieved by the proposed hybrid precoders/combiners, the sparse hybrid precoders /combiners, the optimal unconstrained digital precoders/combiners, greedy hybrid design in a 64 × 16 UPAs mmWave system for different SNR values with $N_S \in \{1, 2\}$, and $N_{RF}^{BS} = N_{RF}^{MS} = 4$. The maximum number of iterations K for the proposed hybrid design is equal to N_{RF}^{BS} -N_S, which is 3 and 2 for $N_S = 1$ and 2 respectively. The proposed hybrid precoders/combiners and the greedy hybrid design outperform the sparse hybrid precoders /combiners for $N_S = 2$ and are almost similar for $N_S = 1$. The proposed hybrid design overlaps the greedy hybrid design for any number of data streams with the smallest number of iterations K and both achieve the optimal performance of the unconstrained digital precoder/combiner.

When the number of RF chains equals the number of data streams, Fig. 7 shows the spectral efficiency achieved by the proposed hybrid precoders/combiners with different numbers of iterations K, the sparse hybrid precoders /combiners, the optimal unconstrained digital precoders/combiners, greedy hybrid design, and the HD-AM with different numbers of iterations M for different numbers of RF chains and data streams varying from 1 to 4 in a 64 × 16 UPAs mmWave systems where $N_s = N_{RF}^{BS} = N_{RF}^{MS}$. The SNR is fixed to 0 dB for any number of RF chains. To make a fair comparison between the proposed hybrid design and the HD-AM, the number of iterations for both designs is the same, i.e., $K = M$. We can see that the proposed hybrid design and the HD-AM are overlapped and yield an improvement over the sparse hybrid design and the greedy hybrid design. Moreover, by increasing the number of iterations K and M, the overall performance of the proposed hybrid precoders/combiners and of the HD-AM is improved; the maximum number of iterations K and M for both designs should be between 1 and 6 because the performance improves slowly after $K = M = 6$. Notice that the HD-AM can only be applied when $N_S = N_{RF}^{BS} = N_{RF}^{MS}$, but the proposed design, and the greedy hybrid design can be applied in the general case (the number of RF chains is greater than or equal to the number of data streams).

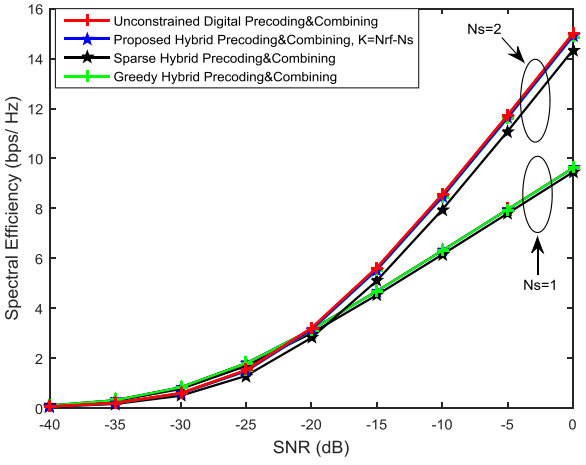


FIGURE 6. Average spectral efficiency achieved by the proposed design with $K = N_{RF}^{BS} - N_S$ compared to the sparse hybrid precoders/combiners [17], the optimal unconstrained digital precoders/combiners, and the greedy hybrid precoders/combiners [27] for a 64 × 16 UPAs mmWave systems for different SNR values with $N_S \in \{1, 2\}$, and $N_{RF}^{BS} = N_{RF}^{MS} = 4$.

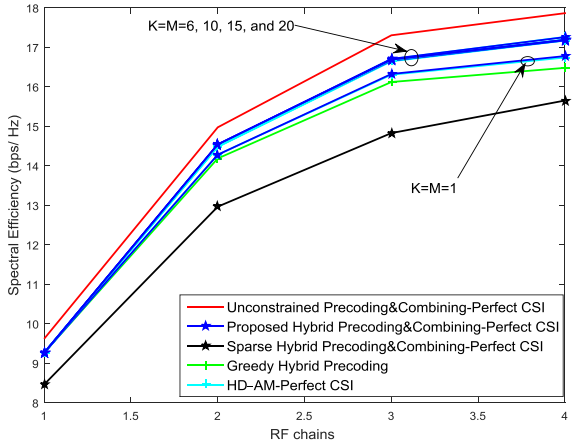


FIGURE 7. Average spectral efficiency achieved by the proposed design with different K compared to the sparse hybrid precoders/combiners [17], the optimal unconstrained digital precoders/combiners, greedy hybrid precoders/combiners [27], and the HD-AM [29] with different M in 64 × 16 UPAs mmWave systems for SNR = 0 dB with $N_S = N_{RF}^{BS} = N_{RF}^{MS}$.

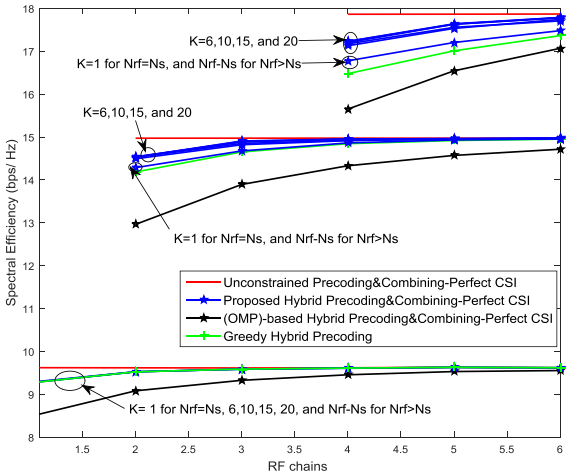


FIGURE 8. Average spectral efficiency achieved by the proposed design with different K compared to the sparse hybrid precoders/combiners [17], the optimal unconstrained digital precoders/combiners, and the greedy hybrid precoders/combiners [27] in 64 × 16 UPAs mmWave systems for SNR = 0 dB with $N_S \in \{1, 2, 4\}$ and different RF chains.

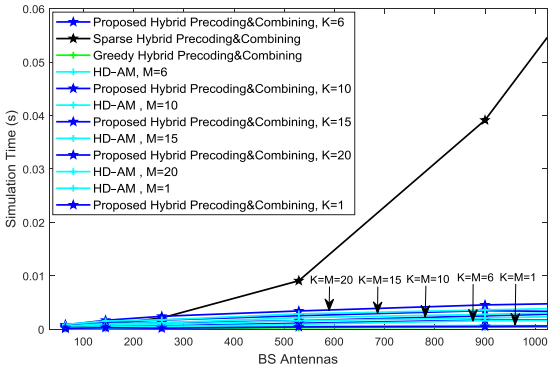


FIGURE 9. Average running time of the proposed hybrid design with different K compared to the HD-AM with different M [29], greedy hybrid design [27], and sparse hybrid method [17] with different number of BS antennas for UPAs mmWave systems - $N_{RF}^{BS}=N_{RF}^{MS}N_S=3$.

Fig. 8 evaluates the performance when the number of RF chains is greater than the number of data streams, where $N_S \in \{1, 2, 4\}$ and the SNR is fixed to 0 dB over the whole range of RF chains. We note again that the HD-AM cannot be applied when the number of RF chains is greater than the number of data streams. The other parameters are the same. The proposed hybrid design and the greedy hybrid design perform better than the sparse hybrid method in all cases. The proposed hybrid design with $K=6$ can accurately approximate the optimal unconstrained one when the number of RF chains slightly exceeds the number of data streams. Moreover, it only needs a small number of RF chains (less than twice the number of streams) to achieve the optimal performance compared to the sparse hybrid precoders / combiners, and the greedy hybrid design. The gain of the proposed hybrid solution over the greedy hybrid design is noticeable for $N_S \in \{2, 4\}$. In practice, the number of RF chains will be limited because of the high power consumption and cost per RF chain [5]–[17].

Notice that the number of iterations $K=6$ of the proposed hybrid design is sufficient to outperform both the greedy hybrid design and the sparse hybrid design. Moreover, the total performance of the proposed hybrid design improves slowly when the number of iterations K increases beyond 6.

Therefore, the performance of the proposed method is the same as that of the HD-AM when $N_S = N_{RF}^{BS} = N_{RF}^{MS}$ and better than that of the greedy or sparse hybrid design when the number of RF chains is greater than or equal to the number of data streams with a reasonable number of iterations K.

Figs. 9 and 10 show the average simulation running time of the proposed hybrid design with different numbers of iterations K compared to the other methods vs. the number of BS antennas. In Fig. 9 and Fig. 10, we consider UPAs mmWave systems with $N_S = N_{RF}^{BS} = N_{RF}^{MS} = 3$ and $N_S = 3, N_{RF}^{BS} = N_{RF}^{MS} = 4$ respectively. As we can see from both figures, the complexity of the sparse hybrid method with an angular resolution $L = N_{BS}$ increases with the number of BS antennas. In addition, the proposed hybrid with any number of iterations K, HD-AM with any number of iteration M, and greedy hybrid have very low complexity compared to the sparse hybrid method especially when $N_S \ll N_{BS}$. In Fig. 9, the average simulation running time of the proposed hybrid design and HD-AM is the same for any number of iteration, where $K = M$ for fair comparison. In addition, the proposed hybrid design, HD-AM, and the greedy hybrid design have almost the same simulation time for small numbers of iteration $K = M = 1$ and 6 but the greedy hybrid design becomes faster with an increase in the number of iterations. In Fig. 10, the average simulation running times of the proposed hybrid design with $K=6$ or 10 and of the greedy hybrid design are almost the same, but the simulation time of the proposed hybrid design increases slightly with the increase of the number of iterations K.

To summarize the performance results, when $N_S > 1$, the appropriate number of iterations for the proposed hybrid design is $N_{RF} - N_S \leq K \leq 6$ when the number of RF chains is greater than or equal to the number of data streams. In practice, the number of RF chains and data streams will be limited and hence, the difference between them should be less than 6. Moreover, the gain of the proposed hybrid design improves slowly by increasing the number of iteration above 6. In the case when $N_S = 1$, the appropriate number of iterations in the proposed hybrid design is $K = 1$ when the number of RF chains equals the number of data streams, i.e., $N_{RF} = N_S$ and $K = N_{RF} - N_S$ when the number of RF chains is greater than the number of data streams.

Figs. 11 shows the average value of the objective function $\left\|\mathbf{F_{opt}} - \mathbf{F_{RF}}\mathbf{F_{BB}}\right\|_F$ vs. the number of iterations K for one million simulation runs of the proposed hybrid beamformer. In Fig. 11, we use UPAs mmWave systems with $N_S = 3, N_{RF}^{BS} = N_{RF}^{MS} = 4$ and $N_S = N_{RF}^{BS} = N_{RF}^{MS} = 4$. As we can see from both figures, the objective function

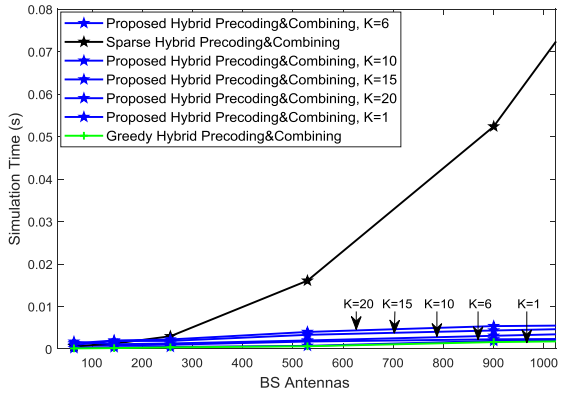


FIGURE 10. Average running time of the proposed hybrid design with different K compared to the greedy hybrid design [27], and sparse hybrid method [17] with different number of BS antennas for UPAs mmWave systems - N_{RF}^{BS}=N_{RF}^{MS}4, and N_S=3.

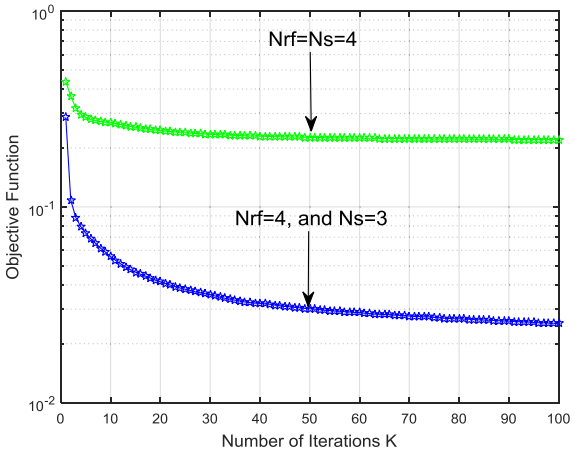


FIGURE 11. Average value of the objective function $\left\|\mathbf{F_{opt}}-\mathbf{F_{RF}F_{BB}}\right\|_F^2$ obtained by simulating the proposed hybrid design with different numbers of iteration K for a 64 × 16 UPAs mmWave system with N_{RF}^{BS}=N_{RF}^{MS} $N_S = 4$ and N_{RF}^{BS}=N_{RF}^{MS} $= 4$ and $N_S = 3$.

$\left\|\mathbf{F_{opt}}-\mathbf{F_{RF}F_{BB}}\right\|_F^2$ decreases monotonically with the number of iterations K. In addition, the objective function of the proposed hybrid design decreases rapidly when the number of iterations K increases from 1 to 6 approximately and after that, decreases slowly. Thus, the performance of the proposed hybrid design improves slowly when the number of iterations K increases beyond 6. When $N_{RF} > N_S$, the objective function is smaller compared to the case when $N_{RF} = N_S$ and that is because the proposed hybrid design can more accurately approximate the optimal unconstrained beamformer. This plot confirms the convergence property in *Remark 1*.

VII. CONCLUSION

In this paper, we have proposed a low complexity hybrid precoder/combiner design for single-user communication in mmWave systems. Assuming the channel state information is known, we consider an optimization problem that finds the hybrid precoder/combiner pair that best approximates the optimal unconstrained digital ones. In addition, there is no need to make any assumption on the antenna array geometry or for any other constraint, such as unitary baseband precoder/combiner or equal numbers of streams and RF chains. The computational complexity of our proposed solution is comparable to that of the HD-AM technique and close to that of the greedy hybrid design when the number of iterations K is reasonable, and it is much lower than the complexity of the sparse hybrid design. Our proposed solution can also be applied in all cases whereas the HD-AM technique can only be applied when the number of RF chains is equal to the number of data streams. The spectral efficiency simulation results show that our proposed low complexity solution with a reasonable number of iterations K is similar to the HD-AM technique when $N_S = N_{RF}^{BS} = N_{RF}^{MS}$ and $K = M$, and outperforms the greedy or sparse hybrid design in all cases. Furthermore, our simulation results have shown that the proposed low complexity hybrid precoding/combining design improves the BER performance of simple ZF/MMSE detectors and considerably reduces the performance gap between these detectors and the high computational complexity ML detector or Sphere Decoding. Moreover, for a given detection scheme, the BER of the proposed hybrid beamformer is lower than that of the sparse hybrid and greedy beamformers, similar to the HD-AM and close to the unconstrained hybrid beamformer. Finally, because of the large available bandwidth in the mmWave band, it will be interesting to extend the proposed hybrid design to multicarrier techniques such as orthogonal frequency division multiplexing (OFDM).

REFERENCES

[1] 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.

[2] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 74–80, Feb. 2014.

[3] T. E. Bogale and L. B. Le, "Massive MIMO and mmWave for 5G wireless HetNet: Potential benefits and challenges," *IEEE Veh. Technol. Mag.*, vol. 11, no. 1, pp. 64–75, Mar. 2016.

[4] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work," *IEEE Access*, vol. 1, pp. 335–349, 2013.

[5] R. W. Heath, N. Gonzalez-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 436–453, Apr. 2016.

[6] W. Roh, J.-Y. Seol, J. Park, B. Lee, J. Lee, Y. Kim, J. Cho, K. Cheun, and F. Aryanfar, "Millimeter-wave beamforming as an enabling technology for 5G cellular communications: Theoretical feasibility and prototype results," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 106–113, Feb. 2014.

[7] A. L. Swindlehurst, E. Ayanoglu, P. Heydari, and F. Capolino, "Millimeter-wave massive MIMO: The next wireless revolution?" *IEEE Commun. Mag.*, vol. 52, no. 9, pp. 56–62, Sep. 2014.

[8] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 742–758, Oct. 2014.

[9] J. G. Andrews, T. Bai, M. N. Kulkarni, A. Alkhateeb, A. K. Gupta, and R. W. Heath, "Modeling and analyzing millimeter wave cellular systems," *IEEE Trans. Commun.*, vol. 65, no. 1, pp. 403–430, Jan. 2017.

[10] 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.

[11] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[12] J. Nsenga, A. Bourdoux, and F. Horlin, "Mixed analog/digital beamforming for 60 GHz MIMO frequency selective channels," in *Proc. IEEE Int. Conf. Commun.*, May 2010, pp. 1–6.

[13] F. Gholam, J. Via, and I. Santamaria, "Beamforming design for simplified analog antenna combining architectures," *IEEE Trans. Veh. Technol.*, vol. 60, no. 5, pp. 2373–2378, Jun. 2011.

[14] Z. Pi, "Optimal transmitter beamforming with per-antenna power constraints," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2012, pp. 3779–3784.

[15] X. Zhang, A. F. Molisch, and S.-Y. Kung, "Variable-phase-shift-based RF-baseband codesign for MIMO antenna selection," *IEEE Trans. Signal Process.*, vol. 53, no. 11, pp. 4091–4103, Nov. 2005.

[16] V. Venkateswaran and A.-J. van der Veen, "Analog beamforming in MIMO communications with phase shift networks and online channel estimation," *IEEE Trans. Signal Process.*, vol. 58, no. 8, pp. 4131–4143, Aug. 2010.

[17] 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.

[18] L. Fan, S. Jin, C.-K. Wen, and H. Zhang, "Uplink achievable rate for massive MIMO systems with low-resolution ADC," *IEEE Commun. Lett.*, vol. 19, no. 12, pp. 2186–2189, Dec. 2015.

[19] Q. Hou, R. Wang, E. Liu, and D. Yan, "Hybrid precoding design for MIMO system with one-bit ADC receivers," *IEEE Access*, vol. 6, pp. 48478–48488, 2018.

[20] J. Mo, A. Alkhateeb, S. Abu-Surra, and R. W. Heath, "Hybrid architectures with few-bit ADC receivers: Achievable rates and energy-rate tradeoffs," *IEEE Trans. Wireless Commun.*, vol. 16, no. 4, pp. 2274–2287, Apr. 2017.

[21] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit," *Signal Process.*, vol. 86, no. 3, pp. 572–588, Mar. 2006.

[22] K. Engan, S. O. Aase, and J. Hakon Husoy, "Method of optimal directions for frame design," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing. (ICASSP)*, Mar. 1999, pp. 2443–2446.

[23] C. Rusu, R. Mendez-Rial, N. Gonzalez-Prelciy, and R. W. Heath, "Low complexity hybrid sparse precoding and combining in millimeter wave MIMO systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2015, pp. 1340–1345.

[24] M. Kim and Y. H. Lee, "MSE-based hybrid RF/baseband processing for millimeter-wave communication systems in MIMO interference channels," *IEEE Trans. Veh. Technol.*, vol. 64, no. 6, pp. 2714–2720, Jun. 2015.

[25] J. Singh and S. Ramakrishna, "On the feasibility of codebook-based beamforming in millimeter wave systems with multiple antenna arrays," *IEEE Trans. Wireless Commun.*, vol. 14, no. 5, pp. 2670–2683, May 2015.

[26] C.-E. Chen, "An iterative hybrid transceiver design algorithm for millimeter wave MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 4, no. 3, pp. 285–288, Jun. 2015.

[27] R. Mendez-Rial, C. Rusu, N. Gonzalez-Prelcic, and R. W. Heath, "Dictionary-free hybrid precoders and combiners for mmWave MIMO systems," in *Proc. IEEE 16th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jun. 2015, pp. 151–155.

[28] W. Ni, X. Dong, and W.-S. Lu, "Near-optimal hybrid processing for massive MIMO systems via matrix decomposition," *IEEE Trans. Signal Process.*, vol. 65, no. 15, pp. 3922–3933, Aug. 2017.

[29] C. Rusu, R. Mendez-Rial, N. Gonzalez-Prelcic, and R. W. Heath, "Low complexity hybrid precoding strategies for millimeter wave communication systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 12, pp. 8380–8393, Dec. 2016.

[30] X. Yu, J.-C. Shen, J. Zhang, and K. B. Letaief, "Alternating minimization algorithms for hybrid precoding in millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 485–500, Apr. 2016.

[31] F. Sohrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale antenna arrays," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 501–513, Apr. 2016.

[32] J.-C. Chen, "Gradient projection-based alternating minimization algorithm for designing hybrid beamforming in millimeter-wave MIMO systems," *IEEE Commun. Lett.*, vol. 23, no. 1, pp. 112–115, Jan. 2019.

[33] J. Jin, Y. R. Zheng, W. Chen, and C. Xiao, "Hybrid precoding for millimeter wave MIMO systems: A matrix factorization approach," *IEEE Trans. Wireless Commun.*, vol. 17, no. 5, pp. 3327–3339, May 2018.

[34] R. Zhang, W. Zou, Y. Wang, and M. Cui, "Hybrid precoder and combiner design for single-user mmWave MIMO systems," *IEEE Access*, vol. 7, pp. 63818–63828, 2019.

[35] 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.

[36] W. Ni and X. Dong, "Hybrid block diagonalization for massive multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 64, no. 1, pp. 201–211, Jan. 2016.

[37] D. H. N. Nguyen, L. B. Le, T. Le-Ngoc, and R. W. Heath, "Hybrid MMSE precoding and combining designs for mmWave multiuser systems," *IEEE Access*, vol. 5, pp. 19167–19181, 2017.

[38] R. A. Stirling-Gallacher and M. S. Rahman, "Linear MU-MIMO precoding algorithms for a millimeter wave communication system using hybrid beam-forming," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2014, pp. 5449–5454.

[39] S. Han, C.-L. I, Z. Xu, and C. Rowell, "Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G," *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 186–194, Jan. 2015.

[40] J. Du, W. Xu, H. Shen, X. Dong, and C. Zhao, "Hybrid precoding architecture for massive multiuser MIMO with dissipation: Sub-connected or fully connected structures?" *IEEE Trans. Wireless Commun.*, vol. 17, no. 8, pp. 5465–5479, Aug. 2018.

[41] X. Gao, L. Dai, S. Han, C.-L. I, and R. W. Heath, "Energy-efficient hybrid analog and digital precoding for mmWave MIMO systems with large antenna arrays," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 998–1009, Apr. 2016.

[42] R. Magueta, D. Castanheira, A. Silva, R. Dinis, and A. Gameiro, "Hybrid multi-user equalizer for massive MIMO millimeter-wave dynamic subconnected architecture," *IEEE Access*, vol. 7, pp. 79017–79029, 2019.

[43] S. Park, A. Alkhateeb, and R. W. Heath, "Dynamic subarrays for hybrid precoding in wideband mmWave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 2907–2920, May 2017.

[44] D. Castanheira, P. Lopes, A. Silva, and A. Gameiro, "Hybrid beamforming designs for massive MIMO millimeter-wave heterogeneous systems," *IEEE Access*, vol. 5, pp. 21806–21817, 2017.

[45] A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing—The large-scale array regime," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6441–6463, Oct. 2013.

[46] A. F. Molisch, V. V. Ratnam, S. Han, Z. Li, S. L. H. Nguyen, L. Li, and K. Haneda, "Hybrid beamforming for massive MIMO: A survey," *IEEE Commun. Mag.*, vol. 55, no. 9, pp. 134–141, Sep. 2017.

[47] M. Alouzi and F. Chan, "Millimeter wave massive MIMO with Alamouti code and imperfect channel state information," in *Proc. IEEE 5G World Forum (5GWF)*, Jul. 2018, pp. 507–511.

[48] M. Alouzi, F. Chan, and C. D'Amours, "Sphere decoding for millimeter wave massive MIMO," in *Proc. IEEE 90th Veh. Technol. Conf. (VTC-Fall)*, Sep. 2019, pp. 1–6.

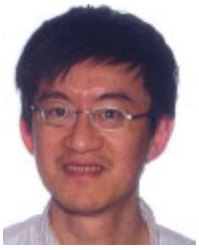
[49] M. Alouzi, F. Chan, and C. D'Amours, "Semidefinite relaxation for millimeter wave massive MIMO detection," in *Proc. 11th IEEE Annu. Inf. Technol., Electron. Mobile Commun. Conf. (IEMCON)*, Nov. 2020, pp. 0734–0739.

[50] E. Torkildson, U. Madhow, and M. Rodwell, "Indoor millimeter wave MIMO: Feasibility and performance," *IEEE Trans. Wireless Commun.*, vol. 10, no. 12, pp. 4150–4160, Dec. 2011.

[51] J. Brady, N. Behdad, and A. M. Sayeed, "Beamspace MIMO for millimeter-wave communications: System architecture, modeling, analysis, and measurements," *IEEE Trans. Antennas Propag.*, vol. 61, no. 7, pp. 3814–3827, Jul. 2013.

[52] T. Xie, L. Dai, X. Gao, M. Z. Shakir, and J. Li, "Geometric mean decomposition based hybrid precoding for millimeter-wave massive MIMO," *China Commun.*, vol. 15, no. 5, pp. 229–238, May 2018.

MOHAMED ALOUZI (Student Member, IEEE) received the B.S. degree in electrical and computer engineering from Zawiya University, Al Zawiya, Libya, in 2009, and the M.A.Sc. degree in electrical engineering from the Royal Military College of Canada, Kingston, Canada, in 2017. He is currently pursuing the Ph.D. degree with the University of Ottawa, Ottawa, Canada. His research interests include millimeter wave, 5G systems, and hybrid beamforming.



FRANCOIS CHAN (Senior Member, IEEE) received the B.Eng. degree in electrical engineering from McGill University, Montréal, Canada, and the M.Sc.A. and Ph.D. degrees in electrical engineering from the Ecole Polytechnique de Montréal. He was a Visiting Researcher with the University of California, Irvine, in 2002 and 2005. He is currently an Associate Professor with the Department of Electrical and Computer Engineering, Royal Military College of Canada, Kingston, ON, Canada. He is also an Adjunct and a Visiting Professor with the University of Ottawa. His research interests include digital communications, wireless communications, and digital signal processing.

CLAUDE D'AMOURS (Member, IEEE) received the B.A.Sc., M.A.Sc., and Ph.D. degrees in electrical engineering from the University of Ottawa, in 1990, 1992, and 1995, respectively. In 1992, he worked as a Systems Engineer with Calian Communications Ltd. In 1995, he joined as a Systems Engineer with the Communications Research Centre, Ottawa, ON, Canada. Later in 1995, he joined as an Assistant Professor with the Department of Electrical and Computer Engineering, Royal Military College of Canada, Kingston, ON. He joined as an Assistant Professor with the School of Information Technology and Engineering (SITE), which has since been renamed as the School of Electrical Engineering and Computer Science (EECS), University of Ottawa, in 1999. From 2007 to 2011, he served as the Vice Dean of undergraduate studies for the Faculty of Engineering, University of Ottawa, where he has been serving as the Director for the School of EECS, since 2013. His research interests include physical layer technologies for wireless communications systems, notably in multiple access techniques, and interference cancellation.