IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 14 October 2023, accepted 31 October 2023, date of publication 6 November 2023, date of current version 20 November 2023.

*Digital Object Identifier 10.1109/ACCESS.2023.3330158*

RESEARCH ARTICLE

# Fast Iterative Hybrid Precoding and Combining With Momentum Gradient Descent and Newton's Method for Millimeter Wave MIMO Systems

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

[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: Mohamed Alouzi (malou052@uottawa.ca)

**ABSTRACT** Millimeter-wave (mmWave) massive multiple-input multiple-output (MIMO) systems have attracted much attention from both researchers and industry professionals, as they are seen as a suitable solution to the growing demand for cellular services in fifth-generation (5G) and sixth-generation (6G) wireless communication systems. In mmWave MIMO systems, iterative hybrid precoding/combining algorithms, which use a combination of analog and digital precoders, have gained significant interest because they perform comparably to fully digital precoding/combining while operating at reduced complexity due to a lower number of radio frequency (RF) components. However, the problem with these algorithms is that their convergence requires a substantial number of iterations. This paper solves this problem and introduces a fast convergence iterative hybrid precoding/combining algorithm using momentum and Newton's method (FIHB-MN) for mmWave MIMO systems. Simulation results demonstrate the faster convergence of the algorithm's objective function compared to other iterative methods in the literature. Moreover, FIHB-MN provides performance similar to unconstrained digital precoding and combining with only a few iterations. The simulation results also confirm that the spectral efficiency as well as the bit error rate (BER) performances of the proposed FIHB-MN algorithm outperforms other hybrid beamforming methods in the literature, all while maintaining low computational complexity.

**INDEX TERMS** mmWave MIMO systems, hybrid design algorithms, momentum gradient descent, Newton's method.

## I. INTRODUCTION

For low-latency communications, millimeter wave (mmWave) technologies with hundreds of antennas have been a focus of research for 5G, and 6G wireless communication due to their ability to significantly increase spectral efficiency and system throughput [1], [2], [3], [4]. Although the path loss is significantly higher at these frequencies, large-scale antenna arrays can be deployed because the size of the antenna elements is much smaller than at lower frequencies, resulting in a high array gain with beamforming [5], [6].

Digital precoding/combining technology without constraints can achieve the maximum theoretical gain of massive MIMO systems that require each transmission/ reception antenna to be linked to a separate power-consuming radio frequency (RF) chain, which will result in a high hardware cost and power consumption. In order to avoid these problems, there is a need to significantly reduce the number of RF chains while ensuring minimal performance degradation. This has led to the emergence of hybrid analog/digital architectures as a critical research area in the development of

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

massive MIMO mmWave systems. By employing a significantly smaller number of RF chains compared to the quantity of antennas, these architectures circumvent the drawbacks of analog beamforming and offer a trade-off between complexity and performance [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20].

In the literature, there are two types of hybrid precoding and combining techniques: full array [7], [8] and subarray [21], [22], [23]. In the full array structure, every RF chain is linked to all the antenna elements, whereas in the subarray arrangement, each RF chain is connected to a portion of the antenna elements. This paper considers a full array configuration. The performance of hybrid precoding and combining in a full array configuration approaches that of the digital configuration, with a notable reduction in the required number of RF chains compared to the digital approach.

The use of simultaneous orthogonal matching pursuit (SOMP) technique was proposed in [7] for mmWave MIMO systems, presenting a full array hybrid design. This approach demonstrated performance comparable to the digital approach while significantly reducing hardware complexity. However, addressing the sparse optimization problem in [7] requires high computational complexity, particularly when considering the constraints of known array geometries. In [8], we introduced and examined a low-complexity hybrid design for mmWave MIMO systems using an iterative approach. Several SOMP-based hybrid designs have been proposed in [9], [10], and [11]. In [13], the authors proposed joint hybrid precoding strategies using the singular value decomposition and equivalent channel concept in mmWave massive MIMO systems. Hybrid design for the downlink multiuser mmWave MIMO channels was presented in [14]. The suggested scheme in [14] iteratively designed analog precoders and combiners using an alternating optimization technique and achieved rate balancing among users through a power allocation algorithm. The study in [15] investigated partially-connected hybrid beamforming for maximizing spectral efficiency. In [19], the authors investigate the application of modified user grouping and hybrid precoding methods to facilitate information decoding and energy harvesting in hardware-impaired mmWave massive MIMO non orthogonal multiple access systems. In [21], [22], and [23], we proposed several hybrid designs for subarray structure. In [22] and [23], an iterative hybrid designs for the non-overlapped SA and overlapped SA were proposed and studied. In [24], an approach to hybrid design is introduced that employs greedy method, without making any assumptions about the channel structure or array geometry. The authors in [25] present a low-complexity hybrid design called alternate minimization (HD-AM). While the spectral efficiency of HD-AM is high, it is applicable only when the number of data streams equals the number of RF chains. A hybrid precoding algorithm based on manifold optimization (MO-AltMin) is presented in [26]. While MO-AltMin offers high spectral efficiency, it comes with a high. The authors in [26] also introduced a low-complexity PE-AltMin algorithm. When the number of RF chains is equal to the number of data streams, the spectral efficiency of the PE-AltMin algorithm matches that of the MO-AltMin algorithm. However, it is outperformed by MO-AltMin with a larger number of RF chains. A heuristic hybrid beamforming algorithm was proposed in [27]. In [28] and [29], a gradient projection algorithm is used to design the hybrid beamforming, with the goal of minimizing the Euclidean distance between the digital and the hybrid precoding matrices. In [30], an iterative hierarchical hybrid precoding technique was proposed, where the digital and analog precoding are separately optimized, resulting in performance comparable to that of PE-AltMin. However, high computational complexity is the main drawback of all the solutions in [27], [28], [29], and [30]. Hybrid design algorithms in multi-user scenarios were proposed and studied in [31], [32], [33], [34], [35], and [36].

In the literature, there exists much research on iterative solutions to solve the optimization problem of the hybrid design. These works, such as the work in [8], yields a performance comparable to the digital precoder and combiner. However, the existing solutions require a large number of iterations to converge. The main objective of this paper is to propose a hybrid design solution for mmWave MIMO systems that requires a smaller number of iterations to converge to a performance that is comparable to the digital solution.

The novelty of this paper is as follows: To the best of our knowledge, there is no existing literature that has examined the application of gradient descent with momentum and Newton's method for designing hybrid precoding and combining in mmWave MIMO systems. The absence of prior research in this field has motivated us to study this topic.

The contributions of this paper can be summarized as follows:

- In this paper, a fast iterative hybrid design, named FIHB-MN, is proposed for mmWave wave MIMO systems. The key advantages of the proposed FIHB-MN approach include its similar performance to fully digital methods and its rapid convergence. We apply separate optimization of analog and digital precoding/combining. Based on the initial normalized analog precoding/combining, we apply the momentum term and Newton's method to the gradient descent in the design of the analog precoding/combining. The suggested method is guaranteed to converge to a local optimum point without requiring any assumptions about the channel. In this paper, a single-user scenario is assumed, and the extension to multi-user scenario will be considered in future work.
- Simulation results are carried out to evaluate the spectral efficiency as well as the BER performances of the proposed FIHB-MN hybrid design. The results verify that the proposed FIHB-MN algorithm outperforms the existing hybrid design algorithm in [7], [8], [24], and [25], while maintaining low complexity. Our results also demonstrate that gradient descent with momentum and

Newton's method can significantly improve the performance of the proposed hybrid design and reduce the objective function with only 10% of the number of iterations in [8].

The remainder of the paper is organized as follows. Section II describes the mmWave channel and system models, Section III introduces the proposed FIHB-MN scheme. The complexity of the proposed FIHB-MN scheme is studied and compared in Section IV, while Section V provides simulation results. Finally, our conclusions are drawn in Section VI.

*Notations:* Upper-case boldface letters represent matrices, while lower-case non-boldface letters denote vectors. The symbol '$b$' denotes a scalar. The notation $(.)^{\mathrm{H}}$ represents the conjugate transpose of a matrix, and $\mathbf{B}_i$ represents the $i^{th}$ column of matrix B. $\mathbf{B}_{i,j}$ denotes the entry located at the $i^{th}$ row and $j^{th}$ column of matrix A. The norm $\|.\|_F$ is the Frobenius norm of a matrix, and the function tr (.) represents the trace function. The operator diag(.) represents a diagonal matrix, $\oslash$ denotes element-wise division, and $\mathbf{I_M}$ is an identity matrix of size M $\times$ M.

## II. DOWNLINK SYSTEM MODEL

### A. mmWave CHANNEL MODEL

In this paper, we adopt a Saleh-Velenzuela (SV) model [3], [4], [5], [6], [7] to represent a mmWave channel and the narrowband clustered channel can be written as

$$\mathbf{H} = \sqrt{N_{BS}N_{MS}/N_{cl}N_{ray}} \times \sum\nolimits_{i,l} \alpha_{il}\Lambda_{\mathrm{r}}(\emptyset_{il}^{\mathrm{r}},\theta_{il}^{\mathrm{r}})\Lambda_t(\emptyset_{il}^{\mathrm{t}},\theta_{il}^{\mathrm{t}})\mathrm{a}_{\mathrm{r}}(\emptyset_{il}^{\mathrm{r}},\theta_{il}^{\mathrm{r}})\mathrm{a}_t(\emptyset_{il}^{\mathrm{t}},\theta_{il}^{\mathrm{t}})^* \quad (1)$$

where $N_{BS}$ and $N_{MS}$ are the number of antennas at the base station (BS) and the mobile station (MS), respectively. $N_{cl}$ and $N_{ray}$ are the number of clusters, and the number of paths in each cluster, respectively. $\alpha_{il} \in \mathbb{C}$ is the gain of the $l^{th}$ path and 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 $l^{\mathrm{th}}$ pat's azimuth (elevation) angles of departure and arrival (AODs/AOAs) in the $i^{\mathrm{th}}$ cluster, respectively, with uniform distribution. Also $\Lambda_{\mathrm{t}}\left(\emptyset_{il}^{\mathrm{t}},\theta_{il}^{\mathrm{t}}\right)$ and $\Lambda_{\mathrm{r}}(\emptyset_{il}^{\mathrm{r}},\theta_{il}^{\mathrm{r}})$ denote the transmit and receive antenna element gain at specific AOD and AOA. In this article, unity gain is assumed for all antenna elements. Finally, $\mathrm{a_t}(\emptyset_{il}^{\mathrm{t}},\theta_{il}^{\mathrm{t}})^{\mathrm{H}}$ and $\mathrm{a_r}(\emptyset_{il}^{\mathrm{r}},\theta_{il}^{\mathrm{r}})$ are the antenna array response vectors at the BS and MS, respectively. In this paper, we consider a uniform planar arrays (UPA) with $w_1$ and $w_2$ elements on width and height (Array size $= w_1 w_2$), and the array response vector can be expressed as [7], [8], [21], [22], and [23]

$$\mathrm{a}\,(\emptyset,\theta) = \frac{1}{\sqrt{N}}[1,\cdots,\mathrm{e}^{\mathrm{j}\left(\frac{2\pi}{\lambda}\right)d(m\sin(\emptyset)\sin(\theta)+n\cos(\theta))}, \cdots,\mathrm{e}^{\mathrm{j}\left(\frac{2\pi}{\lambda}\right)d((w_1-1)\sin(\emptyset)\sin(\theta)+(w_2-1)\cos(\theta))}]^{\mathrm{T}} \quad (2)$$

where $0 \le m < w_1$ and $0 \le n < w_2$. $\lambda$ is the wavelength of the signal, and $d$ is the antenna spacing.

### B. SYSTEM MODEL

Fig. 1 shows the block diagram of the mmWave MIMO system with hybrid precoding and combining. We assumed one MS with $N_{RF}^{MS}$ RF chains and $N_{MS}$ antennas and one BS with $N_{RF}^{BS}$ RF chains and $N_{BS}$ antennas [8]. $N_S$ data streams with $N_S \le N_{RF}^{MS} \ll N_{MS}$ in the MS and $N_S \le N_{RF}^{BS} \ll N_{\mathrm{BS}}$ in the BS are used to support communication between a single MS and the BS. The received signal at the MS can be written as [7] and [8]

$$\begin{aligned}\mathrm{y} &= (\mathbf{W_{RF}W_{BB}})^{\mathrm{H}}(\sqrt{\mathrm{P_r}}\mathbf{HP_{RF}P_{BB}}\mathrm{s}+\mathrm{n}) \\ &= \mathbf{W}^{\mathrm{H}}(\sqrt{\mathrm{P_r}}\mathbf{HP}\mathrm{s}+\mathrm{n})\end{aligned} \quad (3)$$

where $P_r$ is the average received power, $\mathbf{H} \in \mathbb{C}^{N_{MS}\mathrm{x}N_{BS}}$ denotes the channel matrix, $\mathrm{s} \in \mathbb{C}^{N_{\mathrm{s}}\mathrm{x}1}$ is the baseband transmitted signal with $\mathrm{E}\left[\mathrm{ss}^{\mathrm{H}}\right] = \frac{1}{N_S}\mathbf{I_{N_S}}$, and n is the $N_{MS}$x1 vector of i.i.d. $\mathcal{CN}(0,\sigma^2)$ additive complex Gaussian noise. $\mathbf{P_{BB}}$ is the $N_{RF}^{BS}\mathrm{x}N_S$ baseband precoder, $\mathbf{P_{RF}}$ is the $N_{BS}\mathrm{x}N_{RF}^{BS}$ RF precoder. $\mathbf{W_{BB}}$ is $N_{RF}^{MS}\mathrm{x}N_S$ baseband combiner and $\mathbf{W_{RF}}$ is the $N_{MS}\mathrm{x}N_{RF}^{MS}$ RF combiner. $\mathbf{P} = \mathbf{P_{RF}P_{BB}}$ provides the $N_{BS}\mathrm{x}N_S$ hybrid precoder, and $\mathbf{W} = \mathbf{W_{RF}W_{BB}}$ is the $N_{MS} \times N_s$ hybrid combiner.

Since $\mathbf{P_{RF}}$, and $\mathbf{W_{RF}}$ are implemented by the analog phase shifters, all elements of $\mathbf{P_{RF}}$, and $\mathbf{W_{RF}}$ should have the same amplitude but different phases, so that $\left|[\mathbf{P_{RF}}]_{i,j}\right|^2 = \frac{1}{N_{BS}}$ and $\left|[\mathbf{W_{RF}}]_{k,m}\right|^2 = \frac{1}{N_{MS}}$. Therefore, the norms of the elements of $\mathbf{P_{RF}}$ and $\mathbf{W_{RF}}$ are equal [7], [8]. In order to satisfy the total power constraint, we normalize the baseband precoder and combiner, so that $\|\mathbf{W_{RF}W_{BB}}\|_F^2 = \|\mathbf{P_{RF}P_{BB}}\|_F^2 = N_S$ [7], [8].

In this paper, the ZF and the MMSE are considered to detect the received signal and perfect channel estimation is assumed. The estimates of the transmitted symbols $\hat{s}$ using ZF and MMSE can be represented by (4) and (5) as follows

$$\hat{\mathrm{s}} = \mathbf{W_{ZF}}\mathrm{y} \quad (4)$$

$$\hat{\mathrm{s}} = \mathbf{W_{MMSE}}\mathrm{y} \quad (5)$$

where $\mathbf{W_{ZF}} = (\mathbf{H_{eff}})^{-1}$ is the ZF and $\mathbf{W_{MMSE}} = (\mathbf{H_{eff}^{H}H_{eff}} + \sigma^2\mathrm{I_{Nr}})^{-1}\mathbf{H_{eff}^{H}}$ is the MMSE. $\mathbf{H_{eff}} = \mathbf{W}^{\mathrm{H}}\mathbf{HP}$ is the effective channel [31]. Due to their significantly lower dimensions, which depend on the number of data streams $N_S$ [31], the effective channel $\mathbf{H_{eff}}$ has much lower complexity than the original mmWave channel matrix $\mathbf{H}$. Consequently, it becomes feasible to implement these detectors with reduced complexity. The main advantage of the ZF and MMSE schemes is their lower computational complexity compared to that of the ML and sphere decoding techniques.

The spectral efficiency can be written as [7] and [8]

$$\mathrm{R} = \log_2\left|\mathbf{I_{N_s}} + \frac{\mathrm{P}_r}{N_S}\mathbf{R_n}^{-1}\mathbf{W}^{\mathrm{H}}\mathbf{HPP}^{\mathrm{H}}\mathbf{H}^{\mathrm{H}}\mathbf{W}\right| \quad (6)$$

where $\mathbf{R_n} = \sigma^2\mathbf{W}^H\mathbf{W}$ denotes the noise covariance matrix.

## III. THE PROPOSED HYBRID BEAMFORMING

This section presents the description of the proposed algorithm, which combines gradient descent, momentum, and

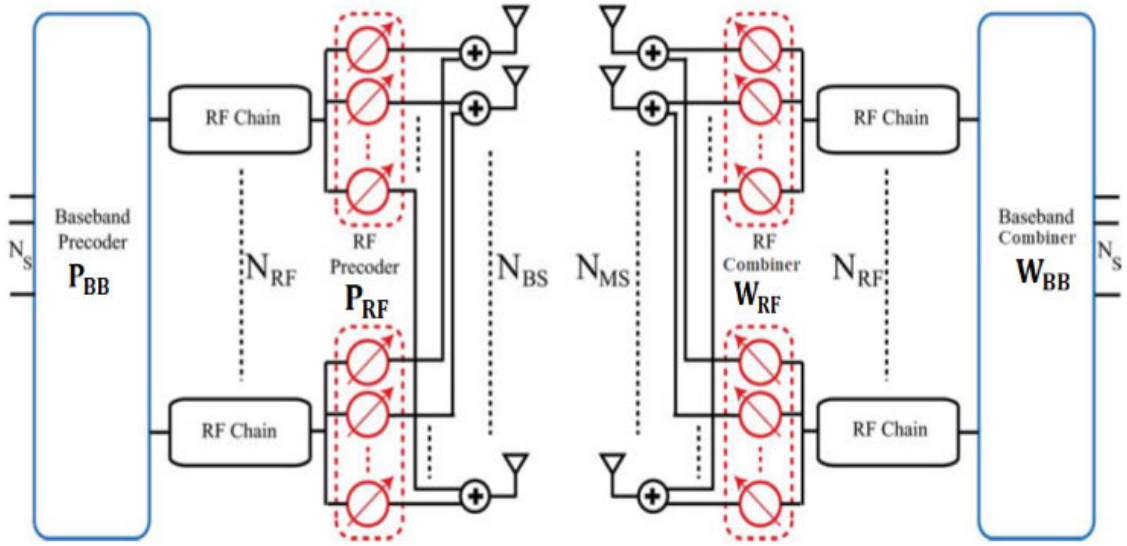



**FIGURE 1.** mmWave system with hybrid precoding and combining architectures [7].

Newton's method. The equations related to the precoder are derived, given that the derivation process for the combiner follows a comparable approach. Our objective is to maximize the spectral efficiency in (6) through the minimization of the Euclidean distance between the optimal unconstrained digital precoding and the hybrid precoding. This can be expressed as

$$\begin{aligned}\left(\mathbf{P}_{\mathrm{RF}}^{\mathrm{opt}}, \mathbf{P}_{\mathrm{BB}}^{\mathrm{opt}}\right) &= \underset{\mathbf{P}_{\mathrm{RF}}, \mathbf{P}_{\mathrm{BB}}}{\operatorname{argmin}} \left\|\mathbf{P}_{\mathbf{opt}} - \mathbf{P}_{\mathrm{RF}}\mathbf{P}_{\mathrm{BB}}\right\|_{\mathrm{F}}^2 \\ &\text{st.} \quad \mathbf{P}_{\mathrm{RF}} \varepsilon \mathcal{F} \\ &\quad \left\|\mathbf{P}_{\mathrm{RF}}\mathbf{P}_{\mathrm{BB}}\right\|_{\mathrm{F}}^2 = N_s \end{aligned} \tag{7}$$

where $\mathcal{F}$ includes all possible RF precoding matrices that satisfy the amplitude constraint. $\mathbf{P_{opt}} = \mathbf{V}_1 = \mathbf{V}(:, \mathbf{1}{:}N_s)$ is the optimal solution of the unconstrained hybrid precoding, and $\mathbf{V}$ can be obtained via the singular value decomposition (SVD) of $\mathbf{H}$ ($\mathbf{H} = \mathbf{U}\Sigma\mathbf{V}^{\mathrm{H}}$). The optimization problem in (7) can be solved in two steps as outlined below:

1. Solving for $\mathbf{P_{BB}}$ while keeping $\mathbf{P_{RF}}$ fixed, the problem can be rewritten as

$$\left(\mathbf{P}_{\mathbf{BB}}^{\mathbf{opt}}\right) = \arg\min_{\mathbf{P_{BB}}} \left\|\mathbf{P_{opt}} - \mathbf{P_{RF}}\mathbf{P_{BB}}\right\|_F^2 \tag{8}$$

It is possible to expand the objective function in (8) as

$$\begin{aligned}&\left\|\mathbf{P_{opt}} - \mathbf{P_{RF}}\mathbf{P_{BB}}\right\|_F^2 \\ &= N_S - 2\mathrm{tr}\left(\mathbf{P_{opt}}\mathbf{P_{RF}}\mathbf{P_{BB}}\right) + \mathrm{tr}(\mathbf{P}_{\mathbf{BB}}^{\mathbf{H}}\mathbf{P}_{\mathbf{RF}}^{\mathbf{H}}\mathbf{P_{RF}}\mathbf{P_{BB}})\end{aligned} \tag{9}$$

The least square solution of $\mathbf{P_{BB}}$ in (9) can be expressed as follows:

$$\mathbf{P_{BB}} = \left(\mathbf{P}_{\mathbf{RF}}^{\mathbf{H}}\mathbf{P_{RF}}\right)^{-1}\mathbf{P_{RF}}\mathbf{P_{opt}} \tag{10}$$

2. Solving for $\mathbf{P_{RF}}$ while keeping $\mathbf{P_{BB}}$ fixed, the problem can be rewritten as:

$$\left(\mathbf{P}_{\mathbf{RF}}^{\mathbf{opt}}\right) = \arg\min_{\mathbf{P_{RF}}} \left\|\mathbf{P_{opt}} - \mathbf{P_{RF}}\mathbf{P_{BB}}\right\|_F^2 \tag{11}$$

Similar to step 1, when the derivative of equation (9) is set to zero with respect to $\mathbf{P_{RF}}$ while holding $\mathbf{P_{BB}}$ constant, the result is:

$$\nabla_f\left(\mathbf{P_{RF}}\right) = -\mathbf{P_{opt}}\mathbf{P}_{\mathbf{BB}}^{\mathbf{H}} + \mathbf{P_{RF}}\mathbf{P_{BB}}\mathbf{P}_{\mathbf{BB}}^{\mathbf{H}} = 0 \tag{12}$$

Since $\mathbf{P_{BB}}\mathbf{P}_{\mathbf{BB}}^{\mathbf{H}}$ cannot be inverted when $N_S < N_{RF}^{BS}$, we use the gradient descent method to obtain an iterative solution for $\mathbf{P_{RF}}$ as follows

$$\begin{aligned}\mathbf{P}_{\mathbf{RF}}^{k+1} &= \mathbf{P}_{\mathbf{RF}}^{k} - \alpha\nabla_f\left(\mathbf{P}_{\mathbf{RF}}^{k}\right) \\ \mathbf{P}_{\mathbf{RF}}^{k+1} &= \mathbf{P}_{\mathbf{RF}}^{k} + \alpha(\mathbf{P_{opt}} - \mathbf{P}_{\mathbf{RF}}^{k}\mathbf{P_{BB}})\mathbf{P}_{\mathbf{BB}}^{\mathbf{H}} \\ \mathbf{P}_{\mathbf{RF}}^{k+1} &= \mathbf{P}_{\mathbf{RF}}^{k} + \alpha\mathbf{P_{res}}\mathbf{P}_{\mathbf{BB}}^{\mathbf{H}}\end{aligned} \tag{13}$$

where $\mathbf{P}_{\mathrm{res}} = \mathbf{P_{opt}} - \mathbf{P}_{\mathbf{RF}}^{k}\mathbf{P_{BB}}$ represents the residual precoding matrix. Note that equation (13) fulfills the characteristic of the gradient descent approach using a step size $\alpha$. This guarantees that $\mathbf{P_{RF}}$ will converge to a feasible local optimal solution.

The proposed FIHB-MN algorithm in this paper uses a combination of the momentum method and Newton's method to effectively eliminate the zig-zag effect of the negative gradient during convergence. Additionally, the algorithm automatically determines the learning rate $\alpha$, eliminating the need for manual selection. In the momentum method, the first couple of iterations will provide a crude moving average over the past gradients because we do not have enough values yet to average ove; the solution is to use what's called bias-corrected version. In Newton's method, the Hessian $\mathbf{P_{BB}}\mathbf{P}_{\mathbf{BB}}^{\mathbf{H}}$ of the objective function in (11) is $N_{BS}^{RF}$ x$N_{BS}^{RF}$ matrix. Because $N_{BS}^{RF}$ is small, then the inversion of the Hessian $(\mathbf{P_{BB}}\mathbf{P}_{\mathbf{BB}}^{\mathbf{H}})^{-1}$ is easy to compute. However, $\mathbf{P_{BB}}\mathbf{P}_{\mathbf{BB}}^{\mathbf{H}}$ cannot be inverted when $N_S < N_{RF}^{BS}$; therefore, we use the preconditioning method to avoid inverting the Hessian in its entirety but only inverting the diagonal entries. From (13), we improve the performance

of the proposed hybrid design by adding the momentum term $\mathbf{Z}_k$ and Newto's Method as follows:

$$\mathbf{Z}_k = (\beta \mathbf{Z}_{k-1} + (1-\beta)\nabla_f \left(\mathbf{P}_{\mathbf{RF}}^k\right))/(1-\beta^k) \quad (14)$$

$$\mathbf{P}_{\mathbf{RF}}^{k+1} = \mathbf{P}_{\mathbf{RF}}^k + \alpha \mathbf{Z}_k (diag(\mathbf{P}_{\mathbf{BB}}\mathbf{P}_{\mathbf{BB}}^{\mathbf{H}}))^{-1} \quad (15)$$

where $\alpha > 0$ is the step size or learning rate, that controls the distance of the negative gradient direction from the RF precoder $\mathbf{P}_{\mathbf{RF}}^k$, and $\beta$ is the momentum decay term. Large $\beta$ amounts to a long-range average, whereas small $\beta$ amounts to only a slight correction relative to a gradient method. We use the learning rate $\alpha$ with Newto's method because we do not use the full inversion of the Hessian, we only invert the diagonal entries. Notice that if $k$ is large, $\beta^k$ will be almost zero and thus, will not change the values of $\mathbf{Z}_k$ at all. By using (9) and (10) in the proposed hybrid design, we can improve its performance with fewer number of iterations. From our simulation results, we can confirm that the gradient descent with momentum can improve the proposed hybrid design and significantly reduce the objective function with a smaller number of iterations $K$.

This approach is applicable when $N_S = N_{RF}^{BS}$. However, it is important to note that when $N_S < N_{RF}^{BS}$, we must complete the $N_{BS} \text{x} N_{RF}^{BS}$ matrix $\mathbf{P}_{\mathbf{RF}}$ after initialization. In every iteration, we add a column to $\mathbf{P}_{\mathbf{RF}}$, achieving maximum reduction in the residual.

Algorithm 1 outlines the pseudo-code of the FIHB-MN solution. The algorithm takes $\mathbf{P}_{\mathbf{opt}} \in \mathrm{C}^{N_{BS} \text{x} N_S}$ and the maximum number of iterations $K$ as inputs. When $N_S < N_{RF}^{BS}$, the algorithm requires $K \geq N_{RF}^{BS} - N_S$ to calculate the $\mathbf{P}_{\mathbf{RF}}$, and $K \geq 1$ when $N_{RF}^{BS} = N_S$. When $1 \leq N_S \leq N_{RF}^{BS}$, the algorithm initializes $\mathbf{P}_{\mathbf{RF}}$ with the element-wise normalization of $\mathbf{P}_{\mathbf{opt}}$, i.e., $\mathbf{P}_{\mathbf{RF}} = \mathbf{P}_{\mathbf{opt}} \oslash \left(|\mathbf{P}_{\mathbf{opt}}| \sqrt{N_{BS}}\right)$, and the momentum term $\mathbf{Z} \in C^{N_{RF} \text{x} N_S}$ with zero matrix. Then in step 4, $\mathbf{P}_{\mathbf{BB}}$ is computed using (10). After that, the residual precoding matrix $\mathbf{P}_{\text{res}}$, the momentum term $\mathbf{Z}$, and the proposed $\mathbf{P}_{\mathbf{RF}}$ are updated in steps 5, 6 and 7 respectively. Step guarantees that $\mathbf{P}_{\mathbf{RF}}$ has constant-magnitude entries, suitable for analog phase shifters in RF implementation. When $N_S < N_{RF}^{BS}$, the element-wise normalization of the first column of $\mathbf{U}$zin $\mathbf{P}_{\text{res}} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathbf{H}}$ is added as a new column to the $\mathbf{F}_{\mathbf{RF}}$ in steps 10 and 11. Upon completing $K$ iterations, the algorithm obtain the proposed $\mathbf{F}_{\mathbf{RF}}$ and $\mathbf{P}_{\mathbf{BB}}$ matrice, minimizing $\left\|\mathbf{P}_{\mathbf{opt}} - \mathbf{P}_{\mathbf{RF}}\mathbf{P}_{\mathbf{BB}}\right\|_F$. Steps 1 and 1 satisf the transmit power constraint and output the proposed $\mathbf{P}^{\mathbf{P}} = \mathbf{P}_{\mathbf{RF}}\mathbf{P}_{\mathbf{BB}}$. The same approach can be used to calculate $\mathbf{W}_{\mathbf{RF}}^{\mathbf{P}}$.

The main difference between FIHB-MN algorithm and our hybrid design in [8] is the addition of the momentum term and the Newto's method to the gradient descent in the design of analog precoding/combining. It is worth noting that in step 8, the utilization of $\sqrt{N_{BS}}$ serves as an effective method to ensure that the diagonal elements of $\mathbf{P}_{\mathbf{RF}}^{\mathbf{H}}\mathbf{P}_{\mathbf{RF}}$ are normalized to one. This normalization is crucial due to the requirement that $\mathbf{P}_{\mathbf{opt}}^{\mathbf{H}}\mathbf{P}_{\mathbf{opt}} = \mathbf{I}_{\mathbf{N_s}}$. Therefore, its components should also exhibit a semi-unitary arrangement [25]. The RF precoder $\mathbf{P}_{\mathbf{RF}}$ can undergo another normalization step to achieve unit modulus and we have conducted verification and confirmed that this adjustment does not impact the results.

It is worth noting that the behavior of $\mathbf{P}_{\mathbf{BB}}$ depends on the relationship between $N_S$ and $N_{RF}^{BS}$. When $N_S$ is equal to $N_{RF}^{BS}$, $\mathbf{P}_{\mathbf{BB}}$ is a square matrix that approximates unitarity, characterized by $\mathbf{P}_{\mathbf{BB}}^{\mathbf{H}}\mathbf{P}_{\mathbf{BB}} \approx \mathbf{P}_{\mathbf{BB}}\mathbf{P}_{\mathbf{BB}}^{\mathbf{H}} \approx \mathbf{I}_{\mathbf{N_S}}$. Conversely, when $N_S$ is less than $N_{RF}^{BS}$, $\mathbf{P}_{\mathbf{BB}}$ becomes a non-square matrix that approximates semi-unitarity, indicated by $\mathbf{P}_{\mathbf{BB}}^{\mathbf{H}}\mathbf{P}_{\mathbf{BB}} \approx \mathbf{I}_{\mathbf{N_s}}$ [7], [25]. Consequently, each iteration of Algorithm 1 works towards minimizing the objective function $\left\|\mathbf{P}_{\mathbf{opt}} - \mathbf{P}_{\mathbf{RF}}\mathbf{P}_{\mathbf{BB}}\right\|_F^2$, leading to a monotonic decrease in the error term with every iteration. Given the existence of a lower bound in the objective function, it is assured that the proposed FIHB-MN algorithm will converge towards local optimum points. Simulation results section will confirm the monotonically decreasing nature of the objective function throughout the iterations and its eventual convergence to the lower bound. Moreover, the proposed design with momentum method makes the objective function converge faster compared to conventional gradient descent method.

## IV. COMPLEXITY ANALYSIS

In this section, we evaluate the complexity of the proposed FIHB-MN design using Algorithm 1. For the sake of simplicity in complexity analysis, let's denote $N$ as the maximum value of $\{N_{BS}, N_{MS}\}$.

Similarly, we define $N_{\text{RF}}$ as the maximum value of $\left\{N_{RF}^{BS}, N_{RF}^{MS}\right\}$. Additionally, $K$ and $M$ represent the maximum number of iterations for the proposed FIHB-MN algorithm and the hybrid design using alternating minimization (HD-AM) in [25]. Table 1 presents a comprehensive complexity analysis, which evaluates the total number of floating-point operations per second. Given the considerations regarding angular resolution and correlation functions within the sparse hybrid algorithm in [7], the proposed FIHB-MN algorithm and hybrid design algorithms in [8], [24], and [25] provide a notable reduction in complexity when compared with the sparse hybrid algorithm in [7], especially when $N_S \ll N$. When $K = M$, the complexity of the proposed FIHB-MN algorithm is twice that in [8] and [25]. However, as we will see in the simulation results section, the proposed FIHB-MN algorithm provides better performance than that in [8] and [25] with a smaller number of iterations. Furthermore, the HD-AM algorithm in [25] can only be applied in a scenario where $N_S = N_{RF}$, whereas the proposed FIHB-MN algorithm can be applied in all scenarios.The proposed FIHB-MN algorithm and the greedy algorithm in [24] have almost the same complexity. This is due to the fact that the proposed algorithm can outperform the algorithm in [24] with a relatively small value of $K$.

The complexity of the proposed FIHB-MN algorithm is also lower than that of the MO-AltMin algorithm in [26]. On the other hand, the MO-AltMin algorithm is characterized by its high complexity, and the update of the analog precoder

**Algorithm 1** Proposed FIHB-MN design

**Input:**

1: Input $\mathbf{P_{opt}} \in \mathbb{C}^{N_{BS} \times N_S}$ and the $K$.

**FIHB-MN Algorithm:**

2: Initialization: $\mathbf{P_{RF}} = \mathbf{P_{opt}} \oslash \left( |\mathbf{P_{opt}}| \sqrt{N_{BS}} \right)$ and $\mathbf{Z} = \mathbf{F_{RF}}.\mathbf{0}$

3: for $i = 1 : K$ do

4: Calculate: $\mathbf{P_{BB}} = \left( \mathbf{P_{RF}^H P_{RF}} \right)^{-1} \mathbf{P_{RF}^H P_{opt}}$

5: Update the residual: $\mathbf{P_{res}} = \mathbf{P_{opt}} - \mathbf{P_{RF}} \mathbf{P_{BB}}$

6: Update: $\mathbf{Z} = (\beta \mathbf{Z} + (\mathbf{1} - \beta) \mathbf{P_{res}} \mathbf{P_{BB}^H}) / (\mathbf{1} - \beta^i)$

7: Update: $\mathbf{P_{RF}} = \mathbf{P_{RF}} + \alpha \mathbf{Z} (\text{diag}(\mathbf{P_{BB}} \mathbf{P_{BB}^H}))^{-1}$

8: Element-Wise Normalization: $\mathbf{P_{RF}} = \mathbf{P_{RF}} \oslash \left( |\mathbf{P_{RF}}| \sqrt{N_{BS}} \right)$

9: If $i \leq N_{RF}^{BS} - N_S$

10: $\mathbf{P_{res}} = \mathbf{U \Sigma V^H}$

11: Normalized and add the first column of $\mathbf{U}$ as a new column to $\mathbf{P_{RF}}$:

$\mathbf{P_{RF}} = \left[ \mathbf{P_{RF}} \; (\mathbf{U})_1 \oslash \left( |(\mathbf{U})_1| \sqrt{N_{BS}} \right) \right]$

12: end if

13: end for

14: Compute $\mathbf{P_{BB}} = \left( \mathbf{P_{RF}^H P_{RF}} \right)^{-1} \mathbf{P_{RF}^H P_{opt}}$

15: Normalize $\mathbf{P_{BB}}$ as $\mathbf{P_{BB}} = \sqrt{N_S} \frac{\mathbf{P_{BB}}}{\|\mathbf{P_{RF}} \mathbf{P_{BB}}\|_F}$

16: return $\mathbf{P^P} = \mathbf{P_{RF}} \mathbf{P_{BB}}$

**Output:**

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

**TABLE 1.** Complexity comparison.

| Algorithm | Constraints | Complexity |
|---|---|---|
| Sparse Hybrid Design [7] | RF precoding/ Combining codebooks | $O(N^2 N_{RF} N_S)$ |
| HD-AM Design [25] | $N_S = N_{RF}$ | $O(N N_S^2 M)$ |
| Greedy Hybrid Design [24] | None | $O(N N_{RF}^2 N_S)$ |
| Hybrid Design [8] | None | $O(N N_{RF}^2 K)$ |
| Proposed FIHB-MN Design | None | $O(2 N N_{RF}^2 K)$ |

within MO-AltMin requires a line search algorithm [26]. Therefore, the algorithms in [7], [8], [24], and [25] are used as performance benchmarks in the simulation results section.

## V. SIMULATION RESULTS

In this section, we present the simulation results of the spectral efficiency as well as the BER to evaluate the performance of the proposed FIHB-MN algorithm. In our simulations, we utilize a 16QAM modulation scheme without error-control coding for the ZF detection algorithm. Furthermore, we assume a constant mmWave channel throughout the transmission of one block of data for all cases examined. The channel model, described by (1), assumes the average power is equal to 1 for all clusters. To implement the sparse hybrid design, we use the AoD/AoA beamforming codebooks, which represent the exact array response of the mmWave channel, at the BS and MS.

**TABLE 2.** Simulation parameters.

| Description | Parameter |
|---|---|
| Number of BS | 1 |
| Number of MS | 1 |
| Distance between BS and MS | 100 meters |
| Antenna Spacing | $d = \lambda/2$ |
| Path loss exponent | $n = 3.4$ |
| Carrier frequency | 28 GHz |
| Signal-to-Noise Ratio | SNR $= P_r/\sigma^2$. |
| AoAs/AoDs of the rays within a cluster | Laplacian distributed. |
| AoAs/AoDs of the cluster means | Uniformly distributed |
| Number of Clusters | $N_{cl} = 8$ |
| Number of rays per cluster | $N_{ray} = 10$ |
| Angular spread | $7.5^o$ |
| Transmitter's sector angle in the azimuth domain | $60^o$ |
| Transmitter's sector angle in the elevation domain | $20^o$ |
| Channel Estimation | Perfect and imperfect |

For the implementation of the sparse hybrid algorithm in [7], we employ the AoD/AoA beamforming codebooks at both the BS and MS. For fairness, all precoding/combining solutions are subjected to the same total power constraint. Further details about the simulation setup are summarized in Table 2.

### A. SPECTRAL EFFICIENCY PERFORMANCE

This subsection evaluates the sum rate achieved by the proposed FIHB-MN algorithm and the hybrid design algorithms in [7], [8], [24], and [25].

The best values of $\beta$ and $\alpha$ are determined simultaneously using a 3-dimensional graph as shown in Fig. 2. The graph illustrates the spectral efficiency performance of the proposed FIHB-MN design in a $64 \times 16$ UPAs mmWave system with $N_S = 3, N_{RF}^{BS} = N_{RF}^{MS} = 4$, and $K = 15$, across different values of the parameters $\alpha$ and $\beta$. It is observed that when $\beta$ is set close to 1, the proposed FIHB-MN design exhibits poorer performance due to the momentum term $Z_k$ collecting very small values of the past gradients. The figure highlights that the maximum spectral efficiency is achieved when $\beta$=0.455 and $\alpha$=1.366, which closely aligns with the values chosen for the rest of the experiments.

Figs. 3 and 4 depict the spectral efficiency performance versus the number of RF chains for various hybrid precoding/combining algorithms, including the proposed FIHB-MN design with different values of $K$, the optimal digital algorithm, and the hybrid design algorithms in [7], [24], and [25] for a $64 \times 16$ mmWave system. In Fig. 3, $N_S = N_{RF}^{BS} = N_{RF}^{MS}$, $SNR = 0\ dB$ and $K = M$ are considered. The proposed FIHB-MN algorithm outperforms that in [25] for $M > 1$ and the greedy algorithm in [24] for $N_s = N_{RF}^{BS} = N_{RF}^{MS} = 2, 3,$ and 4.

Fig. 3 shows that the proposed FIHB-MN algorithm provides an improvement over that in [7] for all numbers of

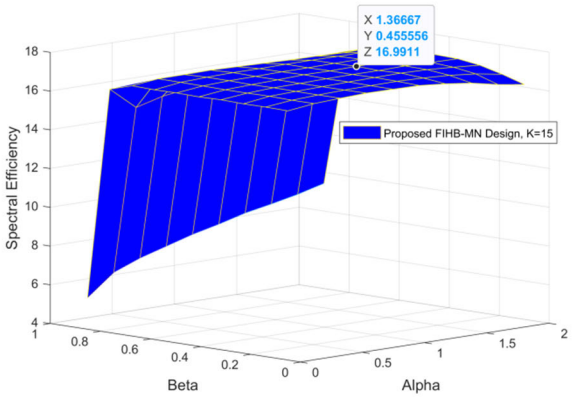



**FIGURE 2.** The spectral efficiency performance versus $\alpha$ and $\beta$ of the proposed FIHB-MN design in a 64 × 16 UPAs mmWave system with $N_S = 3, N_{RF}^{BS} = N_{RF}^{MS} = 4$, and $K = 15$.

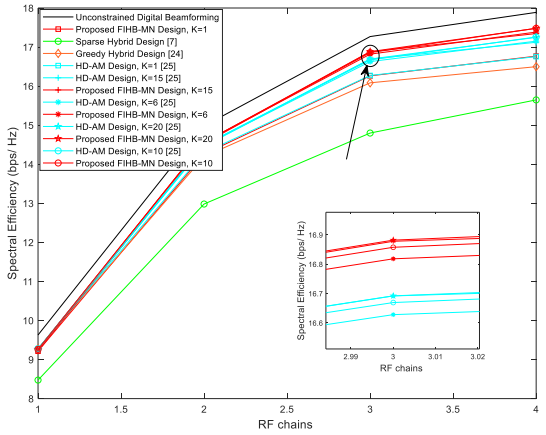



**FIGURE 3.** The spectral efficiency versus the number of RF Chains.

RF chains and outperforms the HD-AM algorithm in [25], even with increasing $K$ and $M$. This indicates that utilizing the proposed FIHB-MN algorithm is more favorable than employing the HD-AM algorithm in [25] when dealing with a scenario where $N_S = N_{RF}^{BS} = N_{RF}^{MS}$.

Fig. 4 studies the spectral efficiency performance for scenarios where $N_{RF}^{BS} = N_{RF}^{MS} \geq N_S$. $N_S \in \{1, 2, 4\}$, and a $SNR = 0dB$ are assumed. It is evident that the proposed FIHB-MN algorithm and the greedy algorithm in [24] provide superior performance compared to the sparse algorithm in [7] and the proposed FIHB-MN algorith closely approximates the optimal solution with only $K = 6$ iterations. It requires a small number of RF chains ($N_{RF}^{BS} = N_{RF}^{MS} < 2N_S$) to provide optimal performance when compared to the sparse algorithm in [7] and the greedy algorithm in [24]. The proposed FIHB-MN algorith demonstrates significant performance improvement over the greedy algorithm in [24] when $N_S \in \{2, 4\}$. Also, it is clearly seen that the proposed algorithm with $K = 6$ outperforms the greedy algorithm in [24] as well as the sparse algorithm in [7] and its performance improves gradually when $K > 6$.

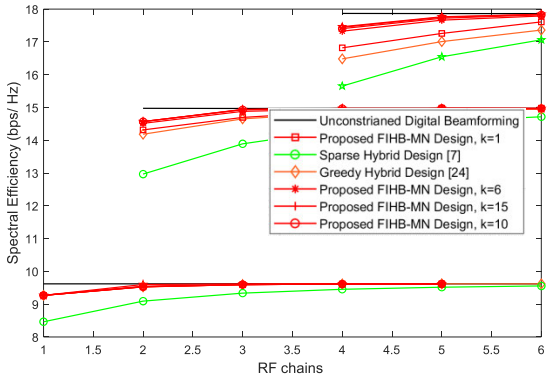



**FIGURE 4.** The spectral efficiency versus the number of RF Chains.

In summary, the proposed FIHB-MN algorith outperforms the HD-AM in [25] when $N_S = N_{RF}^{BS} = N_{RF}^{MS}$ and $K = M > 1$ and outperforms the hybrid algorithms in [7] and [24] when $N_{RF}^{BS} = N_{RF}^{MS} \geq N_S$ with a small value of $K$.

Figs. 5 and 6 demonstrate the spectral efficiency of the proposed FIHB-MN algorithm with $K = 15$, the iterative hybrid design algorithm in [8] with different values of $K$, and the unconstrained digital algorithm for a 64 × 16 mmW system with $N_S > N_{RF}^{BS} = N_{RF}^{MS}$, and $N_S = N_{RF}^{BS} = N_{RF}^{MS}$, respectively. From Figs. 5 and 6, it is evident that the proposed FIHB-MN algorithm with $K = 15$ achieves a spectral efficiency very close to that of the unconstrained digital algorithm. Conversely, the iterative hybrid design algorithm in [8] needs a significantly larger number of iterations ($K = 150$) to attain performance similar to the proposed FIHB-MN algorithm with $K = 15$, particularly for $N_S \in \{2, 4\}$ in Fig. 5 and for all data stream ranges in Fig. 6. Therefore, the iterative hybrid design algorithm in [8] requires high computational complexity to match the performance of the proposed FIHB-MN algorithm in this paper.

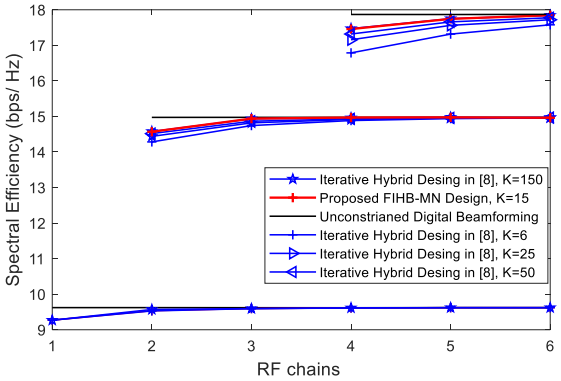



**FIGURE 5.** Average spectral efficiency versus the number of RF chains when $SNR = 0$ dB.

In summary, the performance results indicate that when $N_S > 1$ and $N_{RF} \geq N_S$, the proposed FIHB-MN algorithm can achieve higher performance with a small value of $K$ compared to the iterative hybrid design algorithm in [8]. In scenario where $N_S = 1$, $K$=1 is the appropriate value for

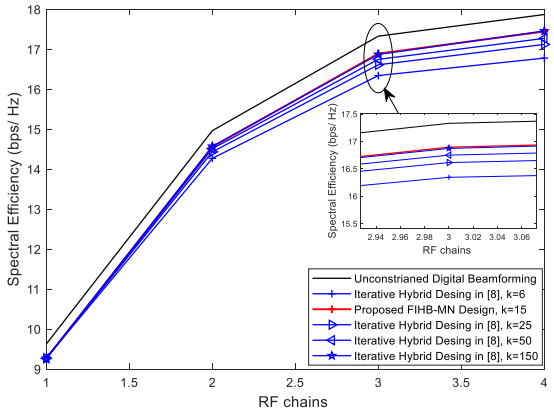



**FIGURE 6.** Spectral efficiency versus the number of RF Chains.

the proposed FIHB-MN algorith when $N_{RF} = N_S$. When $N_{RF} > N_S$, the appropriate value of K for the proposed FIHB-MN algorithm is $K = N_{RF} - N_S$.

Fig. 7 studies the average value of the objective function $\left\|\mathbf{P_{opt}} - \mathbf{P_{RF}}\mathbf{P_{BB}}\right\|_F^2$ against the number of iterations $K$ for the proposed FIHB-MN algorith and the iterative hybrid design algorithm in [8]. We assume mmWave system with $N_S = 3, N_{RF}^{BS} = N_{RF}^{MS} = 4$ and $N_S = N_{RF}^{BS} = N_{RF}^{MS} = 4$.

The results indicate that the proposed FIHB-MN algorithm achieves a lower objective function value than that in [8] across different numbers of iterations. As $K$ exceeds 1000, the objective functions for all designs converge to a lower bound. Notice that the objective function of the proposed design only employing the momentum method (without Newton's method) falls between the iterative hybrid design algorithm in [8] and the proposed FIHB-MN algorithm with both the momentum method and Newton's method. After the number of iterations $K$ exceeds 1000, the objective functions for all designs converge to a lower bound. The objective function of the proposed FIHB-MN algorithm with only the momentum method (without using the Newto's method) is in the middle between the hybrid design in [8] and the proposed design, with the momentum method and Newto's method. Although the momentum method decreases the objective function compared the hybrid design in [8], without using the Newton's method, we need to find the appropriate learning rate $\alpha$ through trial and error which is not a trivial task. Note that when $N_{RF} > N_S$, the value of $\beta$ increases to 0.4 instead of 0.3 because the optimization problem is ill-conditioned, which means that there are some directions where progress is much slower than in others. Thus, we use a larger value of $\beta$ to obtain more stable directions of descent. In scenarios where $N_{RF} > N_S$, the average value of the objective function is notably lower than the scenarios where $N_{RF} = N_S$. This is due to the enhanced capability of the proposed FIHB-MN algorithm and the iterative hybrid design algorithm in [8] to more precisely approximate the optimal design.

The proposed design with only the momentum method can improve the performance compared to that in [8] as shown in Fig. 7, but it is better to use the momentum method and Newton's method in our proposed design to improve the performance much more with fewer number of iterations $K$. Fig. 7 confirms the convergence property of the proposed FIHB-MN algorithm.

Figs. 8 (a) and (b) show a closer look at the average value of the objective function $\left\|\mathbf{P_{opt}} - \mathbf{P_{RF}}\mathbf{P_{BB}}\right\|_F^2$ versus the number of iterations $K$ for a $64 \times 16$ mmW system, which was displayed in Fig 7.

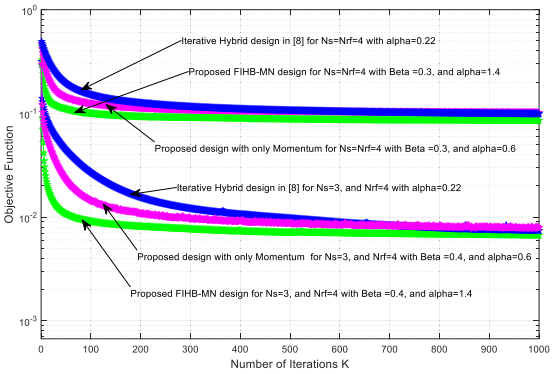


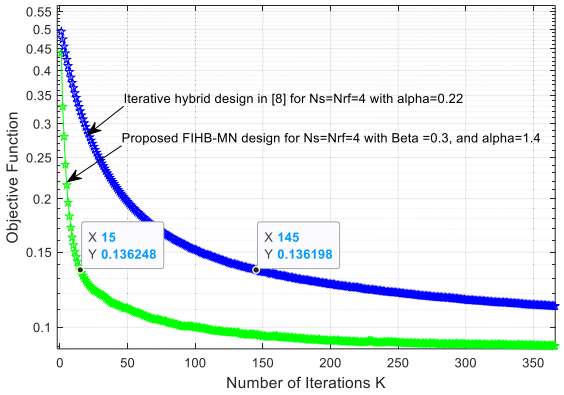

**FIGURE 7.** Average value of the objective function versus the number of iteration $K$.

As we can see from both Figs. 8 (a) and (b), the iterative hybrid algorithm in [8] needs a very large value of $K$, approximately $K$=150, to achieve a performance similar to that of the proposed FIHB-MN algorithm with $K = 15$. Therefore, the iterative hybrid algorithm in [8] needs high computational complexity in order to match the performance of the proposed FIHB-MN algorithm. This plot confirms the results presented in Figs. 5, and 6.

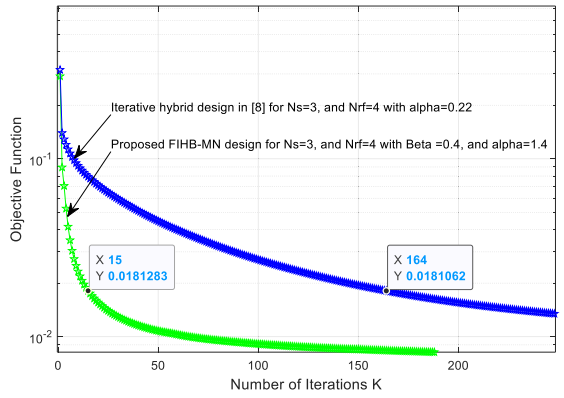
Table 3 presents the average simulation running time of the proposed FIHB-MN algorithm and the iterative hybrid design algorithm in [8] with $K$=15, and $K$=150, respectively. The simulations consider a $64 \times 16$ mmW system with $N_S = N_{RF}^{BS} = N_{RF}^{MS} = 3$. It is clear that the complexity of both the proposed FIHB-MN algorithm and the hybrid design in [8] increases with the number of iterations $K$. However, as illustrated in Figs. 5, and 6, the proposed hybrid FIHB-MN algorithm achieves superior performance while requiring only $K/10$ iterations compared to the algorithm in [8]. On the other hand, the average simulation running time of the proposed FIHB-MN algorithm when $K$=15 is lower than that of the algorithm in [8] when $K$=150. This indicates that, the proposed FIHB-MN algorithm is much faster than that in [8].

**TABLE 3.** Average running time of the proposed FIHB-MN algorithm and the algorithm in [8].

| Method | Number of Iteration $K$ | Average running time in seconds |
|---|---|---|
| Hybrid design algorithm in [8] | 150 | 0.0026 |
| Proposed FIHB-MN algorithm | 15 | $9.3405 \times 10^{-4}$ |




(a)




(b)

**FIGURE 8.** Average value of the objective function versus the number of iteration $K$ for the proposed algorithm and the iterative hybrid algorithm in [8]. (a) $N_{RF}^{BS} = N_{RF}^{MS} = N_S = 4$. (b) $N_{RF}^{BS} = N_{RF}^{MS} = 4$ and $N_S = 3$.

The spectral efficiency of different hybrid design algorithms in a 64 × 16 mmWave system with $N_S = 4$, and $N_{RF}^{BS} = N_{RF}^{MS} = 4$ is presented in Fig. 9. A value of $K = 15$ is assumed for the proposed FIHB-MN, the iterative hybrid design algorithm in [8], and the HD-AM algorithm in [25]. Comparing the results, the proposed FIHB-MN demonstrates optimal performance similar to the optimal digital solution. It outperforms the HD-AM algorithm in [25], and iterative algorithm in [8] with a slight gain and performs better than the greedy algorithm in [24] and the sparse algorithm in [7], especially at higher SNR values, where it exhibits higher gains. The proposed FIHB-MN design shows promising results and outperforms the other techniques in various SNR scenarios, making it a compelling solution for mmWave systems with 64 × 16 UPAs.

Fig. 10 illustrates the impact of the channel estimation errors on the spectral efficiency of four different hybrid designs in a 64 × 16 mmWave system with $K = 15$, $N_S = 4$, and $N_{RF}^{BS} = N_{RF}^{MS} = 4$. From the figure, it is evident that both the proposed FIHB-MN algorithm and the iterative hybrid design algorithm in [8] exhibit higher robustness compared to the sparse algorithm in [7]. Furthermore, both algorithms closely approach the performance of the optimal design across the entire range of variance. Regarding the spectral efficiency performance, the proposed FIHB-MN algorithm outperforms the iterative algorithm in [8] for small variance values. However, as the variance increases, their performance starts to overlap, indicating that both approaches are highly effective in handling higher variance values.

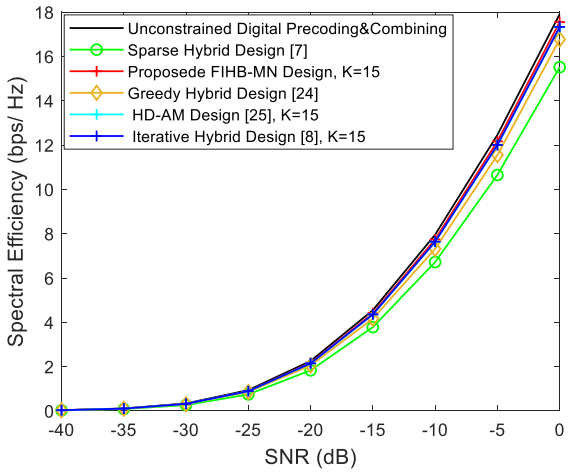



**FIGURE 9.** Spectral Efficiency versus the *SNR* when $N_{RF}^{BS} = N_{RF}^{MS} = 4$.

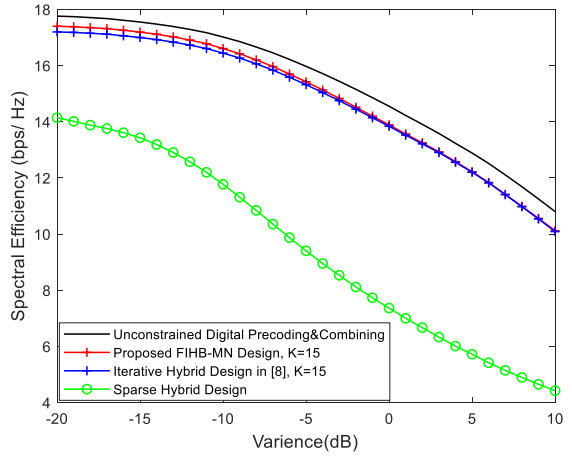



**FIGURE 10.** Spectral Efficiency performance versus the variance of the error in the CSI for different hybrid designs.

### B. BER PERFORMANCE

This subsection evaluates the BER performance of the proposed FIHB-MN algorithm and compares it with other hybrid designs.

In Fig. 11, the BER of proposed FIHB-MN algorithm with ZF detector is studied for different values of $K$, and compared to the sparse algorithm in [7], the optimal digital design, and the greedy algorithm in [24]. The simulation is performed for a 16 QAM 64 × 16 mmWave systems with $N_{RF}^{BS} = N_{RF}^{MS} = 4$ and $N_S = 3$. Fig. 11 shows that the proposed FIHB-MN algorithm with ZF provides a performance that closely matches that of the digital solution. It also outperforms the

sparse algorithm in [7] with ZF detector across the entire SNR range. Comparing it to the greedy algorithm in [24] with ZF detector, the proposed FIHB-MN algorithm performs similarly when $K$ is the smallest ($K = N_{RF}^{BS} - N_S = 1$), but outperforms the greedy algorithm as $K$ increases. Fig. 11 also illustrates that moderate values of $K$, such as $K = 6$ or 10, are adequate to outperform other hybrid algorithms and attain BER levels closely approaching the optimal one.

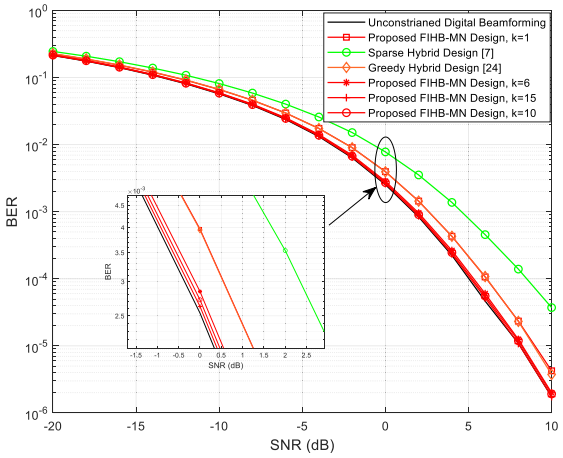



**FIGURE 11.** BER versus the *SNR* when $N_{RF}^{BS} = N_{RF}^{MS} = 4$.

Furthermore, the performance gain of the proposed FIHB-MN algorithm is significant when K increases from 1 to 6, but becomes marginal after $K = 6$.

Fig. 12 demonstrates the BER performance of different hybrid design algorithms with ZF detector in a 16 QAM $64 \times 16$ mmW system with $N_{RF}^{BS} = N_{RF}^{MS} = 3$ and $N_S = 3$ data streams. To ensure fairness, we assume that $K$ equals M. The results show that the proposed FIHB-MN algorithm ($K = 1$), the HD-AM algorithm ($M = 1$) in [25], and the greedy algorithm in [24] have overlapping performance and superior performance compared to sparse algorithm in [7]. With increasing values of $K$ and $M$, the proposed FIHB-MN algorithm exhibits superior performance compared to the HD-AM algorithm in [25]. Note that in this paper, the proposed FIHB-MN algorithm outperforms the HD-AM algorithm in [25] when $K = M > 1$, achieving significant improvements of more than 6 dB in comparison to the sparse algorithm in [7], particularly for $K > 1$.

Fig. 13 illustrates the BER performance of different hybrid design algorithm with ZF detector in a 16 QAM $64 \times 16$ mmW system. The algorithms include the proposed FIHB-MN with $K = 15$, the iterative hybrid design in [8] with different values of $K$, and the unconstrained digital precoding/combining. From the results obtained in Fig. 13, it is clear that the iterative hybrid design in [8] requires a high value of $K$ ($K = 150$) to achieve a performance similar to the proposed FIHB-MN in this paper with a significantly lower number of iterations ($K = 15$). This indicates that the computational complexity of the iterative hybrid algorithm in

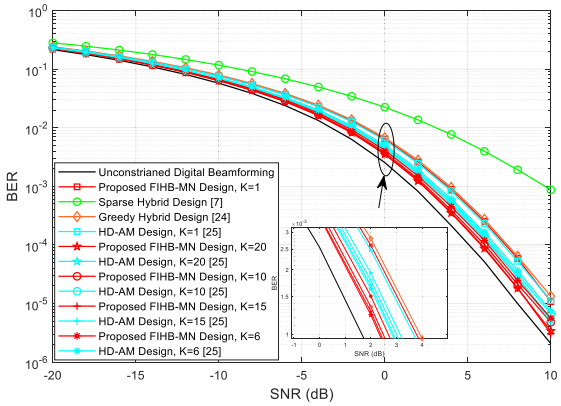



**FIGURE 12.** BER versus the *SNR* when $N_{RF}^{BS} = N_{RF}^{MS} = 3$.

[8] is significantly higher than that of the proposed FIHB-MN algorithm presented in this paper.

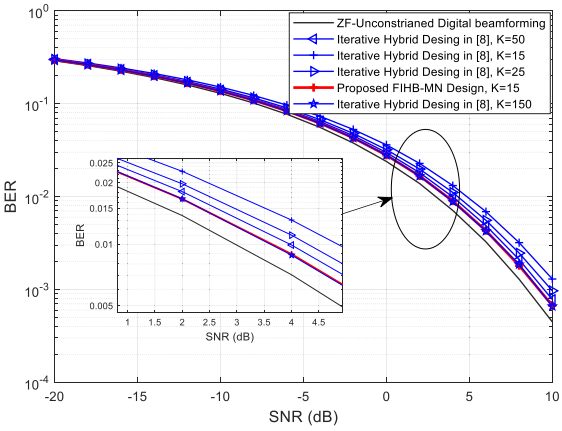



**FIGURE 13.** BER versus the *SNR* when $N_S = 4$.

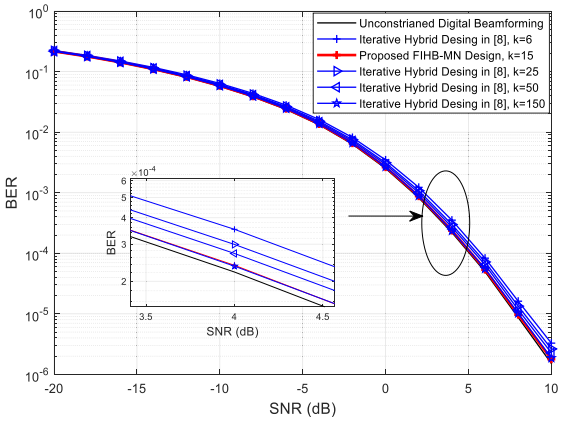



**FIGURE 14.** BER versus the *SNR* when $N_S = 3$.

Fig. 14 compares the BER performance of different hybrid design algorithms with ZF detector in a 16 QAM $64 \times 16$ mmW system when $N_{RF}^{BS} = N_{RF}^{MS} = 4$ and $N_S = 3$. Results show that the proposed FIHB-MN algorithm achieves

a BER close to that of the unconstrained digital algorithm. It is also clear that the iterative hybrid design algorithm in [8] requires a high computational complexity ($K = 150$) to achieve a performance similar to that of the proposed FIHB-MN algorithm with $K$=15.

## VI. CONCLUSION

In this paper, we proposed a new low-computational complexity hybrid precoders and combiners called FIHB-MN, for a single-user mmWave MIMO systems. Under the assumption of perfect channel estimation, we formulate an optimization problem to find the hybrid precoding/combining pair that closely approximates the optimal unconstrained digital pair. The proposed design incorporates the momentum method to mitigate the zig-zag effect of the negative gradient and Newton's method to automatically determine the parameter $\alpha$. The proposed hybrid design approach does not rely on any assumptions and works in all scenarios, even when $N_{RF}^{BS} = N_{RF}^{MS} \neq N_S$. Our simulation results show that the proposed solution exhibits significantly lower computational complexity compared to the sparse algorithm in [7], the iterative algorithm in [8], and the HD-AM algorithm in [25] due to its fast convergence. It achieves superior performance with only one-tenth the number of iterations when compared to the hybrid design in [8] and the HD-AM technique.

Moreover, when selecting a reasonable value of $K$, the computational complexity of the proposed FIHB-MN algorithm closely aligns with that of the greedy algorithm in [24], while still delivering improved performance.

Results also demonstrated that the proposed FIHB-MN algorithm, with a reasonable value of $K$, outperforms the iterative hybrid design algorithm in [8], the HD-AM algorithm in [25] when $N_S = N_{RF}^{BS} = N_{RF}^{MS}$ and $K = M$, and the sparse algorithm in [7], and the greedy algorithm in [24] in all cases. Additionally, the proposed FIHB-MN algorithm achieves better BER performance when compared with that in [7], [8], and [24], and close to the optimal solution. For future work, it would be interesting to explore extending our proposed FIHB-MN algorithm to the multiuser scenario.

## REFERENCES

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

[2] 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.

[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] M. Xiao, S. Mumtaz, Y. Huang, L. Dai, Y. Li, M. Matthaiou, G. K. Karagiannidis, E. Björnson, K. Yang, C.-L. I, and A. Ghosh, "Millimeter wave communications for future mobile networks," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1909–1935, Sep. 2017.

[5] 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.

[6] 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.

[7] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath Jr., "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.

[8] M. Alouzi, F. Chan, and C. D'Amours, "Low complexity hybrid precoding and combining for millimeter wave systems," *IEEE Access*, vol. 9, pp. 95911–95924, 2021.

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

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

[11] 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)*, Vancouver, BC, Canada, Nov. 2020, pp. 0734–0739.

[12] W. Park and J. Choi, "Hybrid precoding and combining strategy for MMSE-based rate balancing in mmWave multiuser MIMO systems," *IEEE Access*, vol. 10, pp. 88043–88057, 2022.

[13] S. Wang, Z. Li, M. He, T. Jiang, R. Ruby, H. Ji, and V. C. M. Leung, "A joint hybrid precoding/combining scheme based on equivalent channel for massive MIMO systems," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 10, pp. 2882–2893, Oct. 2022.

[14] W.-H. Lim, S. Jang, W. Park, and J. Choi, "ZF-based downlink hybrid precoding and combining for rate balancing in mmWave multiuser MIMO systems," *IEEE Access*, vol. 9, pp. 162731–162742, 2021.

[15] X. Zhao, T. Lin, Y. Zhu, and J. Zhang, "Partially-connected hybrid beamforming for spectral efficiency maximization via a weighted MMSE equivalence," *IEEE Trans. Wireless Commun.*, vol. 20, no. 12, pp. 8218–8232, Dec. 2021.

[16] T. Mir, U. Abbasi, R. Ali, S. M. Hussain, and U. Mir, "Joint hybrid precoder and combiner for wideband millimeter-wave massive MIMO systems," *IEEE Access*, vol. 8, pp. 196375–196385, 2020.

[17] J. Deng, O. Tirkkonen, and C. Studer, "MmWave multiuser MIMO precoding with fixed subarrays and quantized phase shifters," *IEEE Trans. Veh. Technol.*, vol. 68, no. 11, pp. 11132–11145, Nov. 2019.

[18] A. N. Uwaechia, N. M. Mahyuddin, M. F. Ain, N. M. A. Latiff, and N. F. Za'bah, "On the spectral-efficiency of low-complexity and resolution hybrid precoding and combining transceivers for mmWave MIMO systems," *IEEE Access*, vol. 7, pp. 109259–109277, 2019.

[19] W. M. Audu and O. O. Oyerinde, "Modified user grouping and hybrid precoding for information decoding and energy harvesting in hardware impaired point-to-point MM-wave massive MIMO NOMA," *IEEE Access*, vol. 11, pp. 33741–33756, 2023.

[20] A. Alkhateeb, J. Mo, N. Gonzalez-Prelcic, and R. W. Heath Jr., "MIMO precoding and combining solutions for millimeter-wave systems," *IEEE Commun. Mag.*, vol. 52, no. 12, pp. 122–131, Dec. 2014.

[21] M. Alouzi, F. Al-Kamali, C. D'Amours, and F. Chan, "Direct conversion of hybrid precoding and combining from full array architecture to subarray architecture for mmWave MIMO systems," *IEEE Access*, vol. 11, pp. 35457–35468, 2023.

[22] F. Al-Kamali and C. D'Amours, "Low-complexity hybrid precoding for subarray architecture mmWave MIMO systems," *IEEE Access*, vol. 10, pp. 74921–74930, 2022.

[23] F. Al-Kamali, C. D'Amours, and F. Chan, "Hybrid precoding for mmWave MIMO systems with overlapped subarray architecture," *IEEE Access*, vol. 10, pp. 130699–130707, 2022.

[24] R. Méndez-Rial, C. Rusu, N. González-Prelcic, and R. W. Heath Jr., "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.

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

[26] 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.

[27] 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.
[28] 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.
[29] 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.
[30] 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.
[31] A. Alkhateeb, G. Leus, and R. W. Heath Jr., "Limited feedback hybrid precoding for multi-user millimeter wave systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 11, pp. 6481–6494, Nov. 2015.
[32] 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.
[33] D. H. N. Nguyen, L. B. Le, T. Le-Ngoc, and R. W. Heath Jr., "Hybrid MMSE precoding and combining designs for mmWave multiuser systems," *IEEE Access*, vol. 5, pp. 19167–19181, 2017.
[34] 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.
[35] F. Dong, W. Wang, and Z. Wei, "Low-complexity hybrid precoding for multi-user mmWave systems with low-resolution phase shifters," *IEEE Trans. Veh. Technol.*, vol. 68, no. 10, pp. 9774–9784, Oct. 2019.
[36] Q. Shi and M. Hong, "Spectral efficiency optimization for millimeter wave multiuser MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 3, pp. 455–468, Jun. 2018.

**MOHAMED ALOUZI** (Student Member, IEEE) received the B.S. degree in electrical and computer engineering from Zawiya University, Al Zawiya, Libya, in 2009, the M.A.Sc. degree in electrical engineering from the Royal Military College of Canada, Kingston, Canada, in 2017, and the Ph.D. degree in electrical engineering from the University of Ottawa, Ottawa, Canada, in 2023. His current research interests include millimeter wave, 5G and beyond systems, and hybrid beamforming.

**FRANCOIS CHAN** (Senior Member, IEEE) received the B.Eng. degree in electrical engineering from McGill University, Montreal, Canada, and the M.Sc.A. and Ph.D. degrees in electrical engineering from Ecole Polytechnique de Montréal, Canada. 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 Professor and a Visiting Professor with the University of Ottawa. He was a Visiting Researcher with the University of California at Irvine, in 2002 and 2005. His current 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 was a Systems Engineer with Calian Communications Ltd. In 1995, he joined the Communications Research Centre, Ottawa, ON, Canada, as a Systems Engineer. In 1995, he joined the Department of Electrical and Computer Engineering, Royal Military College of Canada, Kingston, ON, Canada, as an Assistant Professor. He joined 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, as an Assistant Professor, in 1999. From 2007 to 2011, he served as the Vice Dean of Undergraduate Studies with the Faculty of Engineering, University of Ottawa, where he has been the Director of the School of EECS, since 2013. His current research interests include physical layer technologies for wireless communications systems, notably in multiple access techniques and interference cancellation.

**FAISAL AL-KAMALI** received the B.Sc. degree in electronics and communication engineering from the University of Baghdad, Baghdad, Iraq, in 2001, and the M.Sc. and Ph.D. degrees in communication engineering from the University of Minufia, Egypt, in 2008 and 2011, respectively. He joined as an Assistant Professor with the Electrical Engineering Department, University of Ibb, Ibb, Yemen, in 2011, where he was an Associate Professor, in January 2017. Since November 2021, he has been a Visiting Scholar with the School of Electrical Engineering and Computer Science, Faculty of Engineering, University of Ottawa, Canada. He was the Head of the Electrical Engineering Department, Faculty of Engineering, University of Ibb, from 2013 to 2014, and the Vice Dean of the Academic Affairs with the Faculty of Engineering, University of Ibb, from 2018 to 2020. His current research interests include physical layer technologies for wireless communication systems, such as interference cancellation, channel equalization and channel estimation, image transmission over wireless communication systems, hybrid precoding and combining, mmWave systems, and massive MIMO systems. He was a recipient of the Prize of Supervising the Best Graduation Project in Ibb University, in 2016.