

RESEARCH

Open Access



Joint hybrid-precoding design for MU-MISO systems with a subconnected architecture

Eduard E. Bahingayi^{1,2} and Kyungchun Lee^{2*} 

*Correspondence:
kclee@seoultech.ac.kr

¹ Department of Electronics and Telecommunication Engineering, The University of Dodoma, Dodoma, Tanzania

² Department of Electrical and Information Engineering and the Research Center for Electrical and Information Technology, Seoul National University of Science and Technology, Nowon-gu, Seoul 01811, Republic of Korea

Abstract

In this study, we propose a joint hybrid-precoding algorithm for multiuser multiple-input single-output downlink systems. Specifically, we consider that the base station employs an energy-efficient hybrid-precoding subconnected (SC) architecture with fixed equal subarrays (FESA) (SC-FESA). Optimizing the analog precoding matrix in an SC-FESA architecture is challenging due to its unique constraint structure. In this study, to maximize system sum rate, we propose an efficient method to transform the system's sum-rate optimization problem into a continuous and differentiable objective function wherein only the nonzero elements of the analog precoding matrix are optimized. For the formulated problem, we develop an alternating optimization (AO) approach to jointly optimize the digital and analog precoders in succession by maximizing the system's sum rate. Specifically, in the proposed AO method, when the digital precoder is fixed, we employ the Riemannian conjugate gradient algorithm to generate the analog precoder. Furthermore, when the analog precoder is fixed, we use the minimum mean squared error method to obtain the digital precoder. Numerical simulation results show that the proposed AO algorithm improves the sum rate and energy efficiency of the SC-FESA architecture compared to existing algorithms.

Keywords: MU-MISO system, Subconnected architecture, Hybrid precoding, Millimeter wave

1 Introduction

A large multiple-input multiple-output (MIMO) system is one of the crucial technologies for fifth-generation (5G) communication systems because of its ability to provide ultrahigh data rates and massive device connectivity. Despite the considerable potential of large MIMO systems, the implementation of a conventional fully digital architecture, which requires a dedicated radio-frequency chain (RFC) per antenna, is typically prohibitive because of its high cost, complexity, and circuitry power consumption [1]. To address this issue, various hybrid analog–digital (HAD) architectures, which can achieve data rates closer to the fully digital architecture at a much lower cost and power consumption, have been proposed in the literature [2–8].

HAD architectures can be grouped into two categories: fully connected (FC) and subconnected (SC), depending on how the antennas and RFCs are connected [5, 9]. In an FC architecture, each RFC is connected to all antennas through phase shifters.

In contrast, in an SC architecture, each RFC is connected to a unique subset of antennas, forming an array of subarrays [5]. Moreover, FC schemes offer the maximum precoding gain, resulting in higher achievable rates than SC schemes. On the other hand, SC architectures provide lower power consumption and hardware complexity than FC architectures because of the reduced number of RF components. Consequently, as compared with FC architectures, the SC architectures are regarded as energy-efficient and low-complexity solutions for massive MIMO systems. Unfortunately, SC architectures have received less attention than FC architectures because high data rates are required for 5G systems [10].

However, studies have been devoted to developing algorithms to improve the performance of SC architectures [11, 12, 14–16]. The works of [11, 12] presented hybrid-precoding (HP) algorithms for single-user (SU)-MIMO systems. Specifically, [11] proposed a successive interference cancellation (SIC) approach for optimizing the columns of a precoding matrix via singular value decomposition, and [12] proposed a semidefinite relaxation-based method for jointly designing digital and analog precoders.

For SC architectures in the multiuser (MU)-MIMO scenario, [14] proposed the MU-SIC approach. In [14], the analog precoder and combiner were alternately updated via the SIC method, and a block diagonalization (BD) scheme was employed to generate the digital precoding and combining matrices. In [15], a coordinate ascent (CA)-based HP and combining design algorithm was proposed for a MU-MIMO environment. Recently, [16] proposed a nonlinear Tomlinson–Harashima (TH)-based HP and combining algorithm. In the TH-based algorithm, the columns of the analog precoding and combining matrices were successively optimized using the Schur complement, whereas the digital precoding and combining matrices were generated via QR decomposition. Although the algorithms in [15] and [16] have low complexity and achieve improved performance, they are restricted to a special case when the number of transmitted data symbols is equal to the number of RF chains at the BS. Additionally, the work in [14] achieves sub-optimal performance. Furthermore, previous works proposed dynamic SC schemes [17–20], where each subarray is connected to an arbitrary number of consecutive antennas. Although the dynamic SC schemes can achieve higher beamforming gains than the conventional fixed SC schemes, they require additional hardware, particularly switches, which leads to increased complexity and power consumption. Moreover, the switches may suffer from slow switching speeds and poor port isolation [19, 21], which makes the fixed SC schemes more practical than the dynamic SC schemes.

In this study, we aimed to further improve the performance of SC architectures. We propose an alternating optimization (AO) approach to jointly design the digital and analog precoders by directly solving the sum-rate maximization problem. Specifically, we consider a MU-multiple-input single-output (MISO) downlink system for an SC architecture with *fixed equal subarrays* (FESA) (SC-FESA). It is well known that an optimal digital precoder for MU-MISO systems can be obtained via traditional linear precoding schemes, such as zero-forcing (ZF), block diagonalization (BD), or minimum mean squared error (MMSE) [14, 22]. The primary goal of this study is to find an appropriate analog precoder that leads to improved sum rates. Moreover, the proposed algorithm is applicable to a scenario wherein the number of transmitted data symbols is less than or equal to that of RF chains at the BS, thus making it more practicable.

The SC-FESA architecture has a unique constraint structure on its analog precoding matrix, which is very challenging to tackle [12]. This paper proposes a novel technique to transform the system's sum-rate maximization problem into a tractable and differentiable objective function for optimizing the analog precoder. The formulated problem allows us to optimize the nonzero elements only in the analog precoding matrix. Consequently, we apply a manifold optimization (MO)-based algorithm to optimize the analog precoder as a single vector of all nonzero elements by iteratively searching on the complex circle manifold to find a local optimum with zero gradient of the objective function.

Our proposed method was motivated by the near-optimal performance achieved by the MO-based algorithm in [12, 13]. Reference [12] considers an SU-MIMO system with an FC architecture and solves the spectral efficiency-maximization problem by minimizing the Frobenius norm of the difference between the optimal fully digital precoder and the hybrid precoder. Moreover, Alluhaibi *et al.* in [13] also consider an SU-MIMO system with an SC architecture. In this study, we focus on an MU-MISO system with the SC-FESA architecture and directly solved the sum-rate maximization problem to generate digital and analog precoders; this has not been addressed in previous works [12, 13].

1.1 Main contributions

In this study, our main contributions are as follows:

1. Considering the SC-FESA architecture in MU-MISO downlink systems, we propose a novel method to transform the system's original sum-rate maximization problem into a more tractable problem for optimizing the nonzero elements in the analog precoding matrix. The formulated analog precoding optimization problem belongs to a class of MO problems defined by a unit modulus constraint optimized using the MO-based algorithms.
2. To solve the formulated analog precoding optimization problem, we develop an AO-based algorithm that jointly designs the digital and analog precoders to maximize the system's sum rate. Specifically, the analog precoder is generated via the Riemannian conjugate gradient (RCG) algorithm when the digital precoder is fixed. Furthermore, by fixing the analog precoder, the digital precoder is generated via the MMSE algorithm.

1.2 Notations

A boldface capital letter, \mathbf{X} , is used to denote a matrix, and a boldface lowercase letter, \mathbf{x} , denotes a vector. The n th row and m th column entry of \mathbf{X} is denoted by $x_{n,m}$. We use \mathbf{X}^H , \mathbf{X}^T , and \mathbf{X}^{-1} to denote the Hermitian transpose, transpose, and inverse of \mathbf{X} , respectively. $\text{diag}(a_1, \dots, a_N)$ is a diagonal matrix containing a_1, \dots, a_N as its diagonal elements, $\text{blkdiag}(\mathbf{x}_1, \dots, \mathbf{x}_N)$ is a block-diagonal matrix formed by vectors $(\mathbf{x}_1, \dots, \mathbf{x}_N)$, and \mathbf{I}_N is an $N \times N$ identity matrix. $|x|$ denotes the magnitude of a complex number x , and $\text{Re}\{x\}$ is its real part. The Frobenius norm of \mathbf{X} and the Euclidean norm of \mathbf{x} are denoted by $\|\mathbf{X}\|_F$ and $\|\mathbf{x}\|$, respectively. Finally, \circ denotes the Hadamard (elementwise) product and a calligraphic letter, \mathcal{X} , denotes a set.

2 System model and problem formulation

We consider a conventional SC-FESA architecture for a MU-MISO downlink system, which is illustrated in Fig. 1. The BS is equipped with $N_t = NM$ transmit antennas and $M (< N_t)$ RFCs transmitting K data streams to K single-antenna users, whereas $K \leq M$. Each RFC is connected to a fixed subarray consisting of N antennas.

Let $\mathbf{s} \in \mathbb{C}^{K \times 1}$ denote the transmitted data symbol, which satisfies $\mathbb{E}\{\mathbf{s}\mathbf{s}^H\} = \mathbf{I}_K$. The received vector for all K users, $\mathbf{y} = [y_1, \dots, y_K]^T$, where y_k denotes the signal received by the k th user, can be written as

$$\mathbf{y} = \mathbf{H}\mathbf{A}\mathbf{D}\mathbf{s} + \mathbf{z}, \tag{1}$$

where $\mathbf{H} = [\mathbf{h}_1, \dots, \mathbf{h}_K]^T \in \mathbb{C}^{K \times N_t}$ is a channel matrix consisting of $\mathbf{h}_k \in \mathbb{C}^{N_t \times 1}$, $k = 1, \dots, K$, which represents the channel column vector between the BS and k th user. $\mathbf{A} = \text{blkdiag}(\mathbf{a}_1, \dots, \mathbf{a}_M) \in \mathbb{C}^{N_t \times M}$ denotes an analog precoding matrix.

Here, $\mathbf{a}_m = [a_{1,m}, \dots, a_{N,m}]^T \in \mathbb{C}^{N \times 1}$, $m = 1, 2, \dots, M$ is an analog weighting vector for the m th subarray, whose elements have a unit magnitude, i.e., $|a_{n,m}| = 1$. Moreover, $\mathbf{D} = [\mathbf{d}_1, \dots, \mathbf{d}_K] \in \mathbb{C}^{M \times K}$ indicates the digital precoding matrix, where $\mathbf{d}_k \in \mathbb{C}^{M \times 1}$ is the digital precoding vector for the k th user and $\mathbf{z} \in \mathbb{C}^{K \times 1}$ is an independent and identically distributed additive white Gaussian noise vector with $z_i \sim \mathcal{CN}(0, \sigma^2)$. For the signal model in (1), the sum rate for all K users is given by [14]

$$R = \sum_{k=1}^K \log_2 \left(1 + \frac{|\mathbf{h}_k^H \mathbf{A} \mathbf{d}_k|^2}{\sum_{i \neq k} |\mathbf{h}_k^H \mathbf{A} \mathbf{d}_i|^2 + \sigma^2} \right), \tag{2a}$$

$$\text{s.t. } \mathbf{A} \in \mathcal{A}, \tag{2b}$$

$$\|\mathbf{A}\mathbf{D}\|_F^2 \leq P_t, \tag{2c}$$

where P_t is the transmit power at the BS and \mathcal{A} is a set of feasible RF precoders, which can be defined as

$$\mathcal{A} = \{ \mathbf{A} \in \mathbb{C}^{N_t \times M}, |a_{n,m}| = \begin{cases} 1 & \text{if } n \in \mathbb{N}, m \in \mathbb{M}, \\ 0 & \text{otherwise,} \end{cases}$$

where

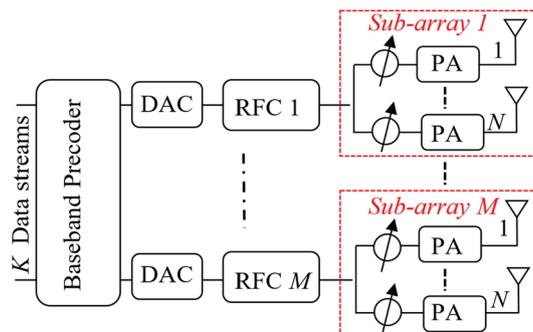


Fig. 1 Illustration of an SC-FESA architecture for a MU-MISO downlink system

$$\mathbb{M} = \{1, 2, \dots, M\} \text{ and}$$

$$\mathbb{N} = \{(m-1)N + 1, \dots, mN | m \in \mathbb{M}\}.$$

2.1 Channel model

In this study, we adopted a widely used geometric-channel model for millimeter-wave communication systems [23, 25]. We assume that the BS is equipped with a uniform linear array (ULA) [23]. Hence, the physical-channel model between the BS and the k th user is given by:

$$\mathbf{h}_k = \sqrt{\frac{N_t}{L_k}} \sum_{l=1}^{L_k} \alpha_l^k \mathbf{a}_t(\phi_l^k), \quad (3)$$

where $\sqrt{\frac{N_t}{L_k}}$ is the normalization factor, L_k denotes the number of propagation paths for the k th user, and $\alpha_l^k \sim \mathcal{CN}(0, \rho_l^k)$ denotes the complex gain associated with the l th path as seen by the k th user. The variance ρ_l^k includes the path loss of the k th user and $\frac{1}{L_k} \sum_{l=1}^{L_k} \rho_l^k = 1$. Thus, without loss of generality, we consider $\rho_l^k = \rho = 1$ [12, 24, 25]. In addition, $\phi_l^k \in [0, 2\pi)$ denotes the angle of departure (AoD) of the l th path from the BS to the k th user, and $\mathbf{a}_t(\cdot)$ denotes the array response vector at the BS. We assume that the BS deploys a uniform linear array, for which the corresponding normalized array response vector is given by [23, Eq. (34)], where λ is the wavelength and d is the antenna spacing.

2.2 Problem formulation

Our goal is to maximize the sum rate by jointly optimizing the digital and analog precoding matrices using an AO approach under constraints on the total transmit power at the BS and unit-modulus analog phase shifters. The optimization problem is expressed as follows:

$$\mathcal{P}1 : \max_{\mathbf{A}, \mathbf{D}} \sum_{k=1}^K \log_2 \left(1 + \frac{|\mathbf{h}_k^H \mathbf{A} \mathbf{d}_k|^2}{\sum_{i \neq k} |\mathbf{h}_k^H \mathbf{A} \mathbf{d}_i|^2 + \sigma^2} \right), \quad (4a)$$

$$\text{s.t. } \mathbf{A} \in \mathcal{A}, \quad (4b)$$

$$\|\mathbf{A} \mathbf{D}\|_F^2 \leq P_t. \quad (4c)$$

Unfortunately, optimizing $\mathcal{P}1$ is a difficult task because of the non-convexity in \mathbf{A} , which makes it unlikely to obtain a global solution. In the next section, we discuss the proposed AO algorithm, which provides an efficient solution to $\mathcal{P}1$.

3 Proposed AO method

In this section, we describe the proposed AO method for solving problem $\mathcal{P}1$. We decompose $\mathcal{P}1$ into two subproblems: one for optimizing the digital precoder \mathbf{D} and the other for the analog precoder \mathbf{A} . In particular, the joint optimization employs an

alternating updating rule; i.e., for a fixed analog precoder \mathbf{A} , the digital precoder \mathbf{D} is optimized, and then vice versa, until convergence is reached.

3.1 Digital precoding design

Given that the analog precoding matrix \mathbf{A} is fixed, we adopted the MMSE algorithm for the digital precoding design. Using the MMSE method, the k th user's digital precoder \mathbf{d}_k , which satisfies the total power constraint in (4c), is given in the following two steps [22]:

$$\hat{\mathbf{d}}_k = \frac{\left(\hat{\mathbf{H}}^H \hat{\mathbf{H}} + \frac{K\sigma^2}{P_t} \mathbf{I}_M\right)^{-1} \hat{\mathbf{h}}_k}{\left\| \left(\hat{\mathbf{H}}^H \hat{\mathbf{H}} + \frac{K\sigma^2}{P_t} \mathbf{I}_M\right)^{-1} \hat{\mathbf{h}}_k \right\|}, \tag{5a}$$

$$\mathbf{d}_k = \sqrt{P_t} \hat{\mathbf{d}}_k / \left\| \mathbf{A} \hat{\mathbf{D}} \right\|_F, \forall k = 1, 2, \dots, K. \tag{5b}$$

In (5), $\hat{\mathbf{H}} = \mathbf{H}\mathbf{A} = [\hat{\mathbf{h}}_1, \dots, \hat{\mathbf{h}}_K]^T \in \mathbb{C}^{K \times M}$ is the effective channel matrix, where $\hat{\mathbf{h}}_k \in \mathbb{C}^{M \times 1}$, and $\hat{\mathbf{D}} = [\hat{\mathbf{d}}_1, \dots, \hat{\mathbf{d}}_K] \in \mathbb{C}^{M \times K}$.

3.2 Analog precoding design

We now focus on optimizing the analog precoder \mathbf{A} for a fixed digital precoder \mathbf{D} . However, because of the unique structure of the constraint on \mathbf{A} , we first transform $\mathcal{P}1$ into an efficient form that allows for the optimization of nonzero elements in \mathbf{A} only. Hence, let $\mathbf{x} = [a_{1,1}, \dots, a_{n,m}, \dots, a_{NM,M}]^H \in \mathbb{C}^{N_t \times 1}$, $\forall n \in \mathbb{N}, m \in \mathbb{M}$ denote a vector of all nonzero elements in \mathbf{A} ; $\mathbf{X} = \text{diag}(\mathbf{x}^H)$ be an $N_t \times N_t$ diagonal matrix, and $\mathbf{G} = \text{blkdiag}(\mathbf{g}_1, \dots, \mathbf{g}_M)$ represent a block-diagonal matrix of size $N_t \times M$, where \mathbf{g}_m is an N -dimensional vector of all ones. Subsequently, the objective function for optimizing \mathbf{x} is given by

$$\mathcal{P}2: \max_{\mathbf{x}} \sum_{k=1}^K \log_2 \left(1 + \frac{|\mathbf{h}_k^H \mathbf{X} \mathbf{G} \mathbf{d}_k|^2}{\sum_{i \neq k} |\mathbf{h}_k^H \mathbf{X} \mathbf{G} \mathbf{d}_i|^2 + \sigma^2} \right), \tag{6a}$$

$$\text{s.t. } |x_\ell| = 1, \ell = 1, 2, \dots, N_t. \tag{6b}$$

Note that, in $\mathcal{P}2$, the analog precoding matrix is

$$\mathbf{A} = \mathbf{X} \mathbf{G}. \tag{7}$$

Moreover, by defining $\tilde{\mathbf{h}}_k = \text{diag}(\mathbf{h}_k^H) \mathbf{G} \mathbf{d}_k \in \mathbb{C}^{N_t \times 1}$, the optimization problem $\mathcal{P}2$ can be rewritten as

$$\mathcal{P}3: \max_{\mathbf{x}} f(\mathbf{x})$$

$$\text{s.t. } |x_\ell| = 1, \forall \ell = 1, 2, \dots, N_t,$$

where

$$f(\mathbf{x}) = \sum_{k=1}^K \log_2 \left(1 + \frac{|\mathbf{x}^H \tilde{\mathbf{h}}_k|^2}{\sum_{i \neq k} |\mathbf{x}^H \tilde{\mathbf{h}}_i|^2 + \sigma^2} \right).$$

Now, the primary obstacle to solving $\mathcal{P}3$ is the non-convex unit modulus constraint $|x_\ell| = 1$, which lacks a standard method. However, we observe that $\mathcal{P}3$ is continuous and differentiable, and its constraint set of \mathbf{x} forms a complex circle manifold. Hence, this study considers an MO-based approach for obtaining the local optimal solution of $\mathcal{P}3$. Specifically, this study employs a type of MO approach called the RCG algorithm [26] to address $\mathcal{P}3$. The RCG algorithm is used as it has shown good performance in addressing hybrid-precoding design problems [11, 13]. The RCG algorithm follows three main steps for each iteration.

The first step is to determine the *Riemannian gradient*, defined as an orthogonal projection onto the tangent space of the gradient of $f(\mathbf{x})$. The Riemannian gradient is obtained by

$$\text{grad}f = \nabla f - \text{Re}\{\nabla f(\mathbf{x}) \circ (\mathbf{x})^H\} \circ \mathbf{x}, \quad (8)$$

where ∇f denotes the Euclidean gradient given by

$$\nabla f = \sum_{k=1}^K \frac{2}{\ln 2} \left(\frac{\sum_i \tilde{\mathbf{h}}_i \tilde{\mathbf{h}}_i^H \mathbf{x}}{\sum_i |\mathbf{x}^H \tilde{\mathbf{h}}_i|^2 + \sigma^2} - \frac{\sum_{i \neq k} \tilde{\mathbf{h}}_i \tilde{\mathbf{h}}_i^H \mathbf{x}}{\sum_{i \neq k} |\mathbf{x}^H \tilde{\mathbf{h}}_i|^2 + \sigma^2} \right).$$

The second step is to find the *conjugate gradient direction* in Euclidean space via the following update rule:

$$\boldsymbol{\eta} = -\text{grad}f + \varpi \mathcal{T}(\tilde{\boldsymbol{\eta}}), \quad (9)$$

where ϖ is chosen as the Polak–Ribiere parameter [27, p. 2], $\tilde{\boldsymbol{\eta}}$ represents the previous search direction, and $\mathcal{T}(\cdot)$ is given by

$$\mathcal{T}(\boldsymbol{\eta}) = \tilde{\boldsymbol{\eta}} - \text{Re}\{\boldsymbol{\eta} \circ (\mathbf{x})^H\} \circ \mathbf{x}. \quad (10)$$

The third step is to perform a *retraction*, which is the process of mapping points from a tangent space back to the complex circle manifold. The following update rule is applied:

$$x_\ell \leftarrow \frac{(\mathbf{x} + \tau \boldsymbol{\eta})_\ell}{|(\mathbf{x} + \tau \boldsymbol{\eta})_\ell|}, \quad (11)$$

where τ is the Armijo step size obtained according to [28, Definition 4.2.2]. The overall procedure for the RCG algorithm to solve $\mathcal{P}3$ is summarized in Algorithm 1. We note that input vector \mathbf{x} is obtained from the initialization and update procedures of Algorithm 2.

Algorithm 1 RCG-based method for solving $\mathcal{P}3$

Input: H, G, D, x , and σ^2 .
1: Compute $\text{grad}f$ according to (8).
2: Set $\eta = -\text{grad}f$.
3: **repeat**
4: Choose the Armijo step size τ [28].
5: Update x using (11).
6: Compute $\text{grad}f$ according to (8).
7: Choose Polak–Ribiere parameter ϖ [27].
8: Update η using (9) and (10).
9: **until** convergence.
10: **return** x .

Having discussed the designs of the digital and analog precoders, where the digital precoder is obtained using the MMSE algorithm and the analog precoder is generated via the RCG-based algorithm, their joint optimization via the AO approach is summarized in Algorithm 2.

Algorithm 2 Proposed AO method for the joint optimization of A and D

Input: P_t, σ^2, H , and G .
Output: A, D .
1: Set $j = 0$.
2: Initialize $x^{(j)}$ with random phases.
3: Generate the initial analog precoder $A^{(j)}$ using (7).
4: Generate the initial digital precoder $D^{(j)}$ using (5b).
5: Calculate the initial sum rate $R^{(j)}$ using (2a).
6: **repeat**
7: Find $x^{(j+1)}$ using Algorithm 1.
8: Update the analog precoder $A^{(j+1)}$ using (7).
9: Update the digital precoder $D^{(j+1)}$ using (5b).
10: Update the sum rate $R^{(j+1)}$ using (2a).
11: $j \leftarrow j + 1$.
12: **until** convergence.

3.3 Complexity analysis

In this subsection, we analyze the complexity of the proposed AO algorithm. We compare the complexity of the proposed AO algorithm to the existing algorithms by considering the dominant computational loads of each algorithm. In the proposed algorithm, the digital precoding matrix is generated via the MMSE method, which requires a complexity of $\mathcal{O}(K^3)$. Then, Algorithm 1 is used to generate the analog precoder in step 6 of Algorithm 2, which is dominated by the $\mathcal{O}(N_t^2 K^2)$ complexity required to compute the Euclidean gradient. Therefore, the dominant complexity of the proposed AO algorithm is $\mathcal{O}(I_O I_I N_t^2 K^2)$, where I_O and I_I denote the numbers of iterations required for Algorithm 1 and Algorithm 2, respectively, to converge. In comparison, the dominant complexities of the SIC-HP [14], CA-HP [15], TH-HP [16], and phase extraction–based HP (PE-HP) [25] are $\mathcal{O}(\max\{K^4, KM^3\})$, $\mathcal{O}(IKN_t^3)$, $\mathcal{O}(IK^2 N_t)$, and $\mathcal{O}(K^3)$, respectively, where I is the number of iterations required for convergence. In the FD architecture, the MMSE algorithm is applied to the channel matrix H . Hence, its complexity becomes $\mathcal{O}(N_t^3)$.

Further, the proposed AO algorithm requires I_I and I_O iterations for the inner and outer loops, respectively. These iterations increase the overall complexity of the proposed algorithm than those of the TH-HP and CA-HP algorithms, which require only I iterations to generate their analog precoding matrices. However, as observed later in Sect. 4, I_I is slightly higher than I , whereas I_O requires only few tens of iterations. Hence, the complexity of the proposed algorithm is still within the range of existing algorithms with a slightly increased complexity. Moreover, the increased complexity of the proposed AO algorithm can be justified, owing to its benefits in terms of the sum rate and EE, which we verify through numerical simulation results in Sect. 4.

3.4 Energy efficiency

The energy efficiency (EE) is defined as the ratio of the achievable sum rate to the total power consumption at the BS, i.e., $EE = \frac{R}{P}$ [5]. The total power consumption of the FD, FC, and SC architectures is expressed as follows [5]:

$$P_{FD} = N_t(P_{PA} + P_{RFC} + P_{DAC}) + P_{BB}, \quad (12a)$$

$$P_{FC} = N_t(P_{PA} + MP_{PS}) + M(P_{RFC} + P_{DAC}) + P_{BB}, \quad (12b)$$

$$P_{SC} = N_t(P_{PA} + P_{PS}) + M(P_{RFC} + P_{DAC}) + P_{BB}, \quad (12c)$$

where P_{PA} , P_{PS} , P_{RFC} , P_{DAC} , and P_{BB} denote the power consumption of the power amplifier, phase shifter, RFC, digital-to-analog converter, and baseband processing, respectively.

4 Simulation results

In this section, we present the simulation results to evaluate the effectiveness of the proposed AO algorithm by comparing it with existing algorithms. For a fair comparison, the MMSE algorithm is adopted for the digital precoding designs of all benchmark schemes. The compared algorithms and their corresponding abbreviations are summarized as follows:

- *FD*: Fully digital architecture with the MMSE digital precoding algorithm [22].
- *PE-HP (FC)*: Fully connected architecture with the channel phase extraction-based analog precoding and MMSE digital precoding [25].
- *PE-HP (SC)*: Subconnected architecture with the channel phase extraction-based analog precoding and MMSE digital precoding [25].
- *SIC-HP (SC)*: Subconnected architecture with successive interference cancellation-based analog precoding and MMSE digital precoding [14].
- *CA-HP (SC)*: Subconnected architecture with the coordinate ascent-based analog precoding and MMSE digital precoding [15].
- *TH-HP (SC)*: Subconnected architecture with the Tomlinson–Harashima-based analog precoding and MMSE digital precoding [16].
- *Proposed AO (SC)*: The proposed method for the subconnected architecture with the RCG-based analog precoding and MMSE digital precoding.

In the simulation, we assume a millimeter-wave channel with $L = 10$, $d = \frac{\lambda}{2}$ [23]. Furthermore, we set $I = I_O = I_I = 10$, unless otherwise stated. For the EE comparison, we set $P_{PA} = 20$ mW, $P_{PS} = 30$ mW, $P_{RFC} = 40$ mW, and $P_{DAC} = P_{BB} = 200$ mW [5]. Finally, all simulation results are averaged over 10^3 channel realizations, and the signal-to-noise ratio (SNR) in the plots is defined as $\frac{P_t}{\sigma^2}$, where the noise variance at each user σ^2 is set to 1.

Figure 2 compares the sum rates of the different algorithms for various SNRs. We assume a system with $N_t = 60$ and $K = M = 12$. In Fig. 2, we observe a large performance gap between the FC and SC architectures owing to the reduced precoding gain of the SC architecture. However, considering only the schemes based on SC architecture, Fig. 2 demonstrates that the proposed AO scheme achieves a higher performance than conventional algorithms over the entire SNR range. It is also observed that if the analog precoder is not properly optimized—for instance, in the PE-HP scheme, which extracts only the phases of the channel coefficients to generate the analog precoding matrix—we obtain poor performance compared to the other schemes. This suggests that both the digital and analog precoding matrices have a significant impact on improving the achievable sum rate of the SC architecture.

Using the results reported in Fig. 3, we study the effect of the number of transmit antennas, namely N_t , located at the BS on the achievable sum rate of a system. We assume that the BS employs a number of RFCs equal to the number of users, that is, $K = M = 12$ and SNR = 0 dB. Figure 3 shows that the sum rates of all algorithms can be improved by increasing N_t . Figure 3 also indicates that the proposed AO scheme maintains a performance gap over TH-HP and CA-HP schemes, whereas the proposed AO demonstrates more significant performance gains over PE-HP and SIC-HP schemes as N_t increases, which further indicates the effectiveness of our proposed method for the SC architecture.

Figure 4 compares the achievable sum rate of different algorithms as a function of the number of users. We assume that the BS employs RFCs equal to the number of

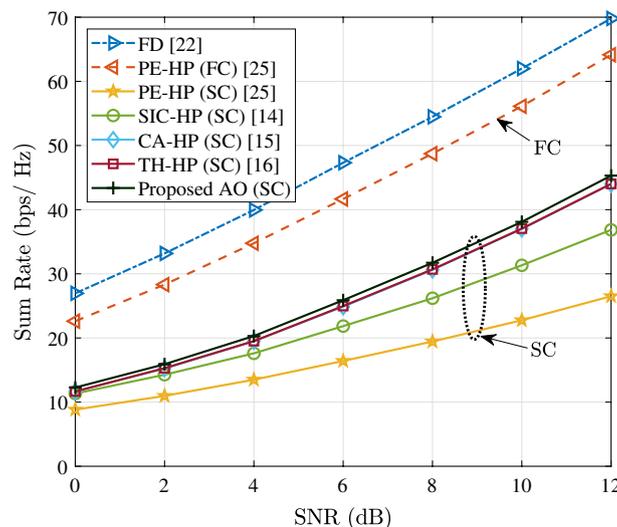


Fig. 2 Sum rate versus SNR. We set $N_t = 60$ and $K = M = 12$

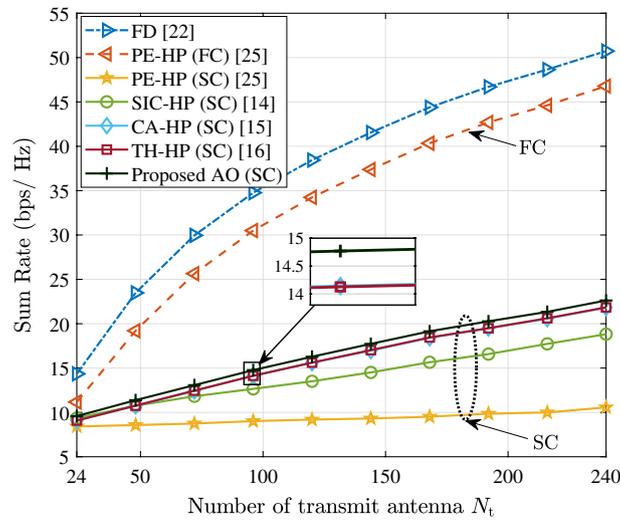


Fig. 3 Sum rate versus number of transmit antennas (N_t). We set $K = M = 12$ and $\text{SNR} = 0$ dB

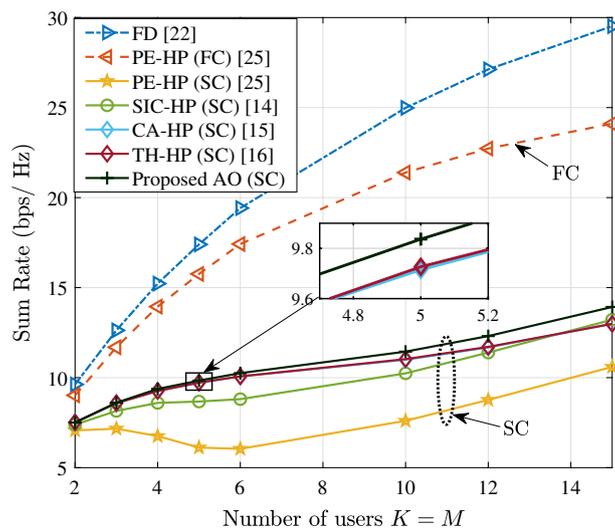


Fig. 4 Sum rate versus number of users K . We set $N_t = 60, M = K$, and $\text{SNR} = 0$ dB

users, i.e., $K = M$. Furthermore, $N_t = 60$ and $\text{SNR} = 0$ dB are considered. Figure. 4 shows that the proposed AO scheme performs better than the conventional algorithms based on the SC architecture. Additionally, the performance gap between the proposed scheme and existing algorithms based on SC architecture slightly widens as K increases. This indicates that the proposed AO scheme effectively eliminates inter-user interference, resulting in an improved performance of the system.

Figure 5 illustrates the sum rate of the proposed algorithm versus the number of users with a fixed number of RF chains $M = 12$. Note that we only consider the algorithms that can be generalized for $K \leq M$. Moreover, we assume $N_t = 60$ and $\text{SNR} = 0$ dB. The simulation results in Fig. 5 verify that the proposed algorithm can achieve substantial performance gains over the SIC-HP scheme in the case of $K \leq M$.

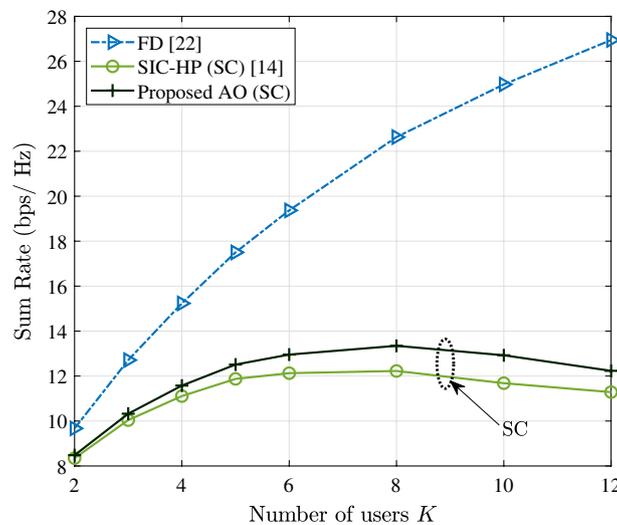


Fig. 5 Sum rate versus number of users K . We set $N_t = 60, M = 12$, and $\text{SNR} = 0$ dB

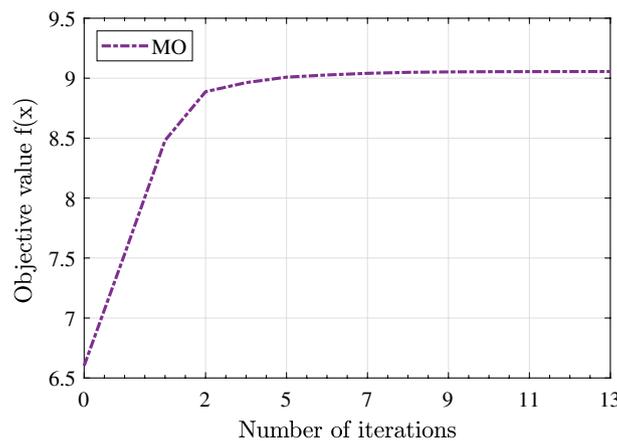


Fig. 6 Convergence behavior of the proposed algorithm for solving $\mathcal{P}3$

Next, we show the convergence behavior of the proposed AO, TH-HP, and CA-HP algorithms for $N_t = 60, K = M = 12$, and $\text{SNR} = 0$ dB. In Fig. 6, we illustrate the convergence behavior of the proposed MO-based algorithm for optimizing the analog precoder in $\mathcal{P}3$, which terminates in approximately 13 iterations because there is a negligible increase in the objective value $f(x)$ of $\mathcal{P}3$. In Fig. 7, we plot the achievable sum rates versus the number of iterations for the proposed AO, TH-HP, and CA-HP algorithms. This shows that the convergence speed of the proposed AO scheme is slightly slower than those of the TH-HP and CA-HP schemes, which can cause increased complexity. However, we note that the proposed algorithm presents substantially higher sum rates than the existing schemes.

Figure 8 plots the EE variations for different numbers of transmit antennas. We set $\text{SNR} = 0$ dB and fix the number of users and RFCs to $K = M = 12$, whereas the number of transmit antennas N_t varies from 24 to 240. In Fig. 8, we observe that the EE performance of the HP schemes based on the SC architecture is higher than that of

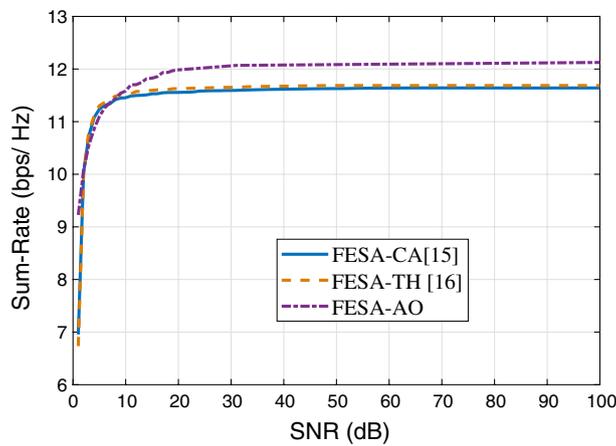


Fig. 7 Sum rate versus the number of iterations. We set $N_t = 60, K = M = 12$, and SNR = 0 dB

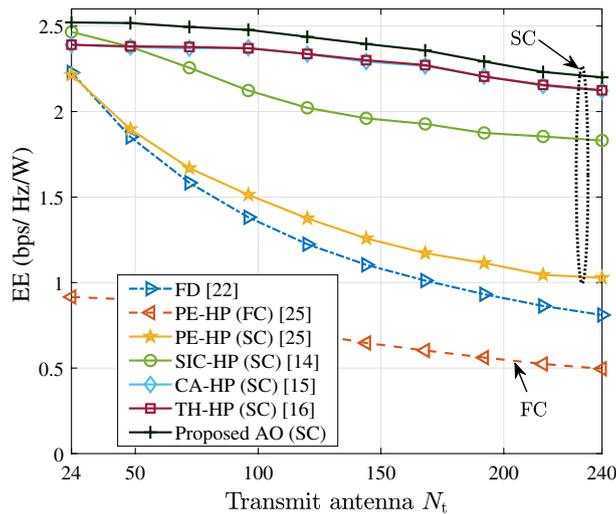


Fig. 8 EE versus the number of transmit antennas (N_t). We set $K = M = 12$ and SNR = 0 dB

the PE-HP (FC) and FD schemes. Moreover, Fig. 8 shows that the EE gap between the SC and FC schemes increases with N_t . The proposed AO scheme achieves a higher EE than that of the conventional algorithms over the entire range of N_t . In Fig. 8, we also observe that as N_t increases, the EE of all the compared SC schemes decreases. However, the EE gains of the proposed AO over the SIC-HP and PE-HP schemes increase drastically with N_t , while maintaining nearly constant performance gains over the CA-HP and TH-HP schemes.

Finally, in Fig. 9, we show the EE versus number of users for $N_t = 60$ and SNR = 0 dB. For $K = M$, the power consumption of a system increases with K due to the increased number of RFCs M . Consequently, in Fig. 9, we observe that the EE tends to decrease for all schemes as K increases. We also observe that the EE of the FD scheme increases with K since its sum rate increases with K , whereas its power

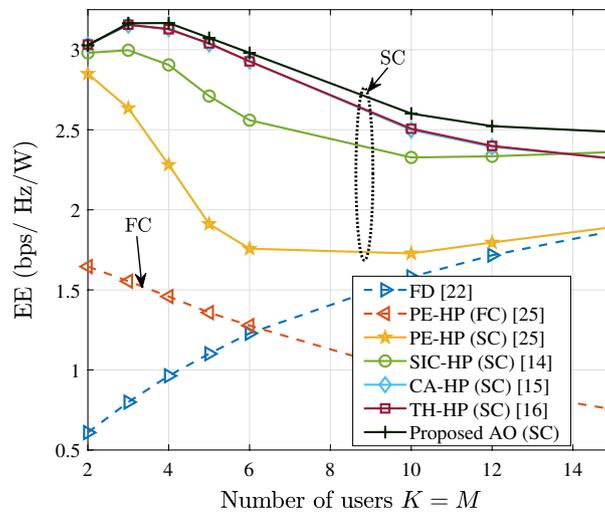


Fig. 9 EE versus number of users K . We set $N_t = 60, M = K$, and $\text{SNR} = 0$ dB

consumption remains constant. More importantly, the proposed AO scheme always achieves better EE performance than the existing schemes for different K .

5 Conclusion

In this study, we proposed a joint HP design algorithm for MU-MISO downlink systems, where the BS is equipped with a SC-FESA architecture. We proposed a novel technique to transform the system’s sum-rate optimization problem into a tractable objective function, which allows the efficient adoption of the MO-based algorithm to optimize the nonzero elements of the analog precoding matrix. The proposed algorithm enhanced the achievable sum rate by jointly optimizing the digital and analog precoding matrices using an AO approach. Specifically, we employed the MMSE algorithm to generate the digital precoder, whereas the analog precoder was generated using an MO-based technique. Through numerical simulation, we showed that the proposed AO algorithm attained higher sum rates than the existing algorithms for SC architecture. In addition, the proposed AO algorithm achieved higher EE than conventional algorithms for both SC and FC architectures. Notably, in this study, we considered only the narrowband channels and MU-MISO system; however, the hybrid precoding for wideband channels and MU-MIMO wireless communication systems is one of the areas that require investigation. It would be interesting to extend our proposed algorithm to wideband channels and MU-MIMO systems in future studies. Moreover, the convergence analysis of the proposed AO algorithm will require further investigation.

Abbreviations

MIMO	Multiple-input multiple-output
5G	Fifth-generation
RFC	Radio-frequency chain
HAD	Hybrid analog–digital
FC	Fully connected
SC	Subconnected
HP	Hybrid precoding
SU-MIMO	Single-user MIMO
SIC	Successive interference cancellation
MU-MIMO	Multi-user MIMO

BD	Block diagonalization
CA	Coordinate ascent
TH	Tomlinson–Harashima
AO	Alternating optimization
MISO	Multiple-input single-output
FESA	Fixed equal subarrays
SC-FESA	Subconnected architecture with FESA
MU-MISO	Multiuser MISO
ZF	Zero-forcing
MMSE	Minimum mean squared error
BS	Base station
MO	Manifold optimization
RCG	Riemannian conjugate gradient
EE	Energy efficiency
ULA	Uniform linear array
PE	Phase extraction
PA	Power amplifier
DAC	Digital-to-analog converter
PS	Phase shifter
BB	Baseband
SNR	Signal-to-noise ratio

Author contributions

EEB and KL both contributed to formulation of the research problem, development of algorithms, numerical simulations, and writing of the paper. Both authors read and approved the manuscript.

Funding

This work was supported in part by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education under Grant NRF-2019R1A6A1A03032119 and in part by the NRF Grant funded by the Korean Government (MSIT) under Grant NRF-2022R1A2C1006566.

Availability of data and materials

Not applicable.

Declarations

Ethics approval and consent to participate

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable.

Received: 2 October 2022 Accepted: 18 January 2023

Published online: 25 January 2023

References

1. C.H. Doan, S. Emami, D.A. Sobel, A.M. Niknejad, R.W. Brodersen, Design considerations for 60 GHz CMOS radios. *IEEE Commun. Mag.* **42**(12), 132–140 (2004)
2. R.W. Heath, N. González-Prelcic, S. Rangan, W. Roh, A.M. Sayeed, An overview of signal processing techniques for millimeter wave MIMO systems. *IEEE J. Sel. Top. Signal Process.* **10**(3), 436–453 (2016)
3. I. Ahmed, H. Khammari, A. Shahid, A. Musa, K.S. Kim, E. De Poorter, I. Moerman, A survey on hybrid beamforming techniques in 5G: architecture and system model perspectives. *IEEE Commun. Surv. Tuts.* **20**(4), 3060–3097 (2018)
4. E.E. Bahingayi, K. Lee, Hybrid combining based on constant phase shifters and active/inactive switches. *IEEE Trans. Veh. Technol.* **69**(4), 4058–4068 (2020)
5. R. Méndez-Rial, C. Rusu, N. González-Prelcic, A. Alkhateeb, R.W. Heath, Hybrid MIMO architectures for millimeter wave communications: phase shifters or switches? *IEEE Access* **4**, 247–267 (2016)
6. W.B. Abbas, F. Gomez-Cuba, M. Zorzi, Millimeter wave receiver efficiency: a comprehensive comparison of beamforming schemes with low resolution ADCs. *IEEE Trans. Wirel. Commun.* **16**(12), 8131–8146 (2017)
7. O. Alluhaibi, Q.Z. Ahmed, E. Kampert, M.D. Higgins, J. Wang, Revisiting the energy-efficient hybrid D-A precoding and combining design for mmWave systems. *IEEE Trans. Green Commun. Net.* **4**(2), 340–354 (2020)
8. F. Dong, W. Wang, Z. Wei, Two-stage BFGS-based hybrid precoding for mmWave multiuser MIMO systems. *IET Commun.* **13**(9), 1271–1277 (2019)
9. O.E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, R.W. Heath, Spatially sparse precoding in millimeter wave MIMO systems. *IEEE Trans. Wirel. Commun.* **13**(3), 1499–1513 (2014)
10. S. Yang, L. Hanzo, Fifty years of MIMO detection: the road to large-scale MIMOs. *IEEE Commun. Surv. Tuts.* **17**(4), 1941–1988 (2015)

11. X. Gao, L. Dai, S. Han, I. Chih-Lin, R.W. Heath, Energy-efficient hybrid analog and digital precoding for mmWave MIMO systems with large antenna arrays. *IEEE J. Sel. Areas Commun.* **34**(4), 998–1009 (2016)
12. X. Yu, J.-C. Shen, J. Zhang, K.B. Letaief, Alternating minimization algorithms for hybrid precoding in millimeter wave MIMO systems. *IEEE J. Sel. Top. Signal Process.* **10**(3), 485–500 (2016)
13. O. Alluhaibi, Q. Z. Ahmed, J. Wang, H. Zhu, Hybrid digital-to-analog precoding design for mmWave systems. In *2017 IEEE International Conference on Communications (ICC)*, pp. 1–6 (2017)
14. Z. Zhang, X. Wu, D. Liu, Joint precoding and combining design for hybrid beamforming systems with subconnected structure. *IEEE Syst. J.* **14**(1), 184–195 (2020)
15. K. Ardah, G. Fodor, Y.C.B. Silva, W.C. Freitas, F.R.P. Cavalcanti, A Unifying design of hybrid beamforming architectures employing phase shifters or switches. *IEEE Trans. Veh. Technol.* **67**(11), 11243–11247 (2018)
16. X. Bai, F. Liu, R. Du, Y. Xu, Z. Sun, Hybrid TH precoding and combining with sub-connected structure for mmwave systems. *IEEE Commun. Lett.* **24**(8), 1821–1824 (2020)
17. G.M. Gadiel, N.T. Nguyen, K. Lee, Dynamic unequally sub-connected hybrid beamforming architecture for massive MIMO systems. *IEEE Trans. Veh. Technol.* **70**(4), 3469–3478 (2021)
18. N.T. Nguyen, K. Lee, Unequally sub-connected architecture for hybrid beamforming in massive MIMO systems. *IEEE Trans. Wirel. Commun.* **19**(2), 1127–1140 (2020)
19. A. Alkhateeb, Y. Nam, J. Zhang, R.W. Heath, Massive MIMO combining with switches. *IEEE Wirel. Commun. Lett.* **5**(3), 232–235 (2016)
20. R. Du, H. Liu, Z. Guan, F. Liu, HP-SPC algorithm with dynamic partially connected structure for mmWave MIMO systems. *Trans. Emerg. Telecom. Technol.* **33**(7), e4474 (2022)
21. H. Tataria, M. Matthaoui, P. J. Smith, G. C. Alexandropoulos, V. F. Fusco, Impact of RF processing and switching errors in lens-based massive MIMO systems. In *2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pp. 1–5 (2018)
22. E. Björnson, M. Bengtsson, B. Ottersten, Optimal multiuser transmit beamforming: a difficult problem with a simple solution structure [Lecture notes]. *IEEE Signal Process. Mag.* **31**(4), 142–148 (2014)
23. F. Sohrabi, W. Yu, Hybrid digital and analog beamforming design for large-scale antenna arrays. *IEEE J. Sel. Top. Signal Process.* **10**(3), 501–513 (2016)
24. J. Li, L. Xiao, X. Xu, S. Zhou, Robust and low complexity hybrid beamforming for uplink multiuser mmWave MIMO systems. *IEEE Commun. Lett.* **20**(6), 1140–1143 (2016)
25. L. Liang, W. Xu, X. Dong, Low-complexity hybrid precoding in massive multiuser MIMO systems. *IEEE Commun. Lett.* **3**(6), 653–656 (2014)
26. N. Boumal, B. Mishra, P.-A. Absil, R. Sepulchre, Manopt, a matlab toolbox for optimization on manifolds. *J. Mach. Learn. Res.* **15**(1), 1455–1459 (2014)
27. W.W. Hager, H. Zhang, A survey of nonlinear conjugate gradient methods. *Pacific J. Opt.* **2**(1), 35–58 (2006)
28. P.-A. Absil, R. Mahony, R. Sepulchre, *Optimization Algorithms on Matrix Manifolds* (Princeton University Press, 2009)

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)
