

Hybrid Beamforming for Massive MIMO Over-the-Air Computation

Xiongfei Zhai, Xihan Chen, Jie Xu, and Derrick Wing Kwan Ng

Abstract

Over-the-air computation (AirComp) has been recognized as a promising technique in Internet-of-Things (IoT) networks for fast data aggregation from a large number of wireless devices. However, as the number of devices becomes large, the computational accuracy of AirComp would seriously degrade due to the vanishing signal-to-noise ratio (SNR). To address this issue, we exploit the massive multiple-input multiple-output (MIMO) with hybrid beamforming, in order to enhance the computational accuracy of AirComp in a cost-effective manner. In particular, we consider the scenario with a large number of multi-antenna devices simultaneously sending data to an access point (AP) equipped with massive antennas for functional computation over the air. Under this setup, we jointly optimize the transmit digital beamforming at the wireless devices and the receive hybrid beamforming at the AP, with the objective of minimizing the computational mean-squared error (MSE) subject to the individual transmit power constraints at the wireless devices. To solve the non-convex hybrid beamforming design optimization problem, we propose an alternating-optimization-based approach, in which the transmit digital beamforming and the receive analog and digital beamforming are optimized in an alternating manner. In particular, we propose two computationally efficient algorithms to handle the challenging receive analog beamforming problem, by exploiting the techniques of successive convex approximation (SCA) and block coordinate descent (BCD), respectively. It is shown that for the special case with a fully-digital receiver at the AP, the achieved MSE of the massive MIMO AirComp system is inversely proportional to the number of receive antennas. Furthermore, numerical results

X. Zhai is with the School of Information Engineering, Guangdong University of Technology, Guangzhou 510006, China (e-mail: zhaixiongfei@gdut.edu.cn).

X. Chen is with the College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310000, China (e-mail: chenxihan@zju.edu.cn).

J. Xu is with the Future Network of Intelligence Institute (FNii) and the School of Science and Engineering, The Chinese University of Hong Kong, Shenzhen, Shenzhen 518172, China (e-mail: xujie@cuhk.edu.cn). J. Xu is the corresponding author.

D. W. K. Ng is with the School of Electrical Engineering and Telecommunications, University of New South Wales, Sydney, NSW 2052, Australia (email: w.k.ng@unsw.edu.au).

show that the proposed hybrid beamforming design substantially enhances the computation MSE performance as compared to other benchmark schemes, while the SCA-based algorithm performs closely to the performance upper bound achieved by the fully-digital beamforming.

Index Terms

Over-the-air computation (AirComp), Internet-of-Things (IoT) networks, massive multiple-input multiple-output (MIMO), hybrid beamforming, optimization.

I. INTRODUCTION

Future Internet-of-Things (IoT) networks need to support an enormous number of wireless devices for sensing the environment, aggregate massive sensing data for analysis, and accordingly take physical actions [1], [2]. Conventionally, such data aggregation is implemented via wireless devices individually sending their data to an access point (AP) or a fusion center over wireless multiple access channels, in which the AP may need to decode the individual messages from each device by treating messages from others as harmful interference. Nevertheless, for practical IoT applications, the AP may be interested in computing a certain function value (e.g., the sum value) of the aggregated data rather than the individual messages (e.g., in federated learning setups [3]). In this case, the above conventional multiple access scheme may not be energy- or spectral-efficient and may also lead to excessively long network latency, especially when the number of devices becomes significantly large. To overcome the drawback, the over-the-air computation (AirComp) technique has been proposed recently, which utilizes the co-channel interference among devices as a beneficial factor for functional computation [4], [5]. By exploiting the signal superposition property of multiple access channels, the AirComp technique is able to directly compute a class of nomographic functions (e.g., arithmetic mean, weighted sum, geometric mean, polynomial, and Euclidean norm) of distributed sensing data from the concurrent transmission of distributed wireless devices [6].

In general, AirComp can be implemented in both analog and digital modes. While the simple uncoded analog transmission was shown to achieve the minimum functional distortion when the data sources follow the independent and identically distributed (i.i.d.) Gaussian distribution [7], coding was shown to be necessary for improving the computation performance under the bivariate Gaussian [8] and correlated Gaussian [9] distributed data sources. Besides, for analog AirComp, the computation mean-squared error (MSE) is normally adopted as the performance metric. In the single-antenna setup, a proper power control

is essential for minimizing the computation MSE [10]–[12]. For instance, under the coherent multiple access channel, the optimal power control strategy for minimizing the computation MSE was proposed in [10] by using convex optimization, and optimal power allocation strategies were proposed in [11] for minimizing the distortion outage probability (defined as the probability that the computation MSE exceeds a given threshold). Furthermore, under fading channels, the optimal power allocation strategy for minimizing the average computation MSE was studied [12]. In particular, multi-antenna beamforming is an efficient technique to further enhance the computation MSE performance. For instance, the authors in [5] investigated the multiple-input multiple-output (MIMO) AirComp for computing multiple functions simultaneously, in which a closed-form equalization at the AP was proposed to minimize the computation MSE under the zero-forcing (ZF) transmit beamforming at wireless devices. Notice that the implementation of AirComp requires the synchronization among all devices; towards this end, the so-called AirShare design was developed in [13], where a shared clock was broadcast to all devices to facilitate synchronization.

On the other hand, digital AirComp was proposed to enhance computational accuracy via proper coding methods [14]–[19]. The idea of digital AirComp first appeared for functional computation in wireless sensor networks (see, e.g., [14]) and physical-layer network coding in a two-way relay channel (see, e.g., [15]). For AirComp, the achievable computation rate under different system setups was characterized in [18] and [19], which is defined as the number of functional values computed per unit time under a predefined computational accuracy. Furthermore, to enable multi-function computation with enhanced computation rate, the authors in [20] integrated the idea of non-orthogonal multiple access (NOMA) [21] in AirComp, in which multiple functions from different wireless devices are superposed in each resource block.

In this paper, we particularly focus our study on the analog AirComp, in which wireless devices send uncoded data to a single AP for functional computation. In practice, the implementation of AirComp over large-scale wireless networks faces several technical challenges. For instance, as the number of wireless devices increases, the computation performance in AirComp systems may seriously degrade, due to the vanishing of the received signal-to-noise ratio (SNR) at the AP. As a remedy, massive MIMO [22] is considered as a promising viable solution to improve the computational accuracy of AirComp systems by exploiting its rich spatial degrees of freedom and tremendous array gain. To the best of our knowledge, the exploitation of massive MIMO for AirComp has not been reported in the literature yet. Hence, it

motivates us to consider the combination of massive MIMO and AirComp to improve the computation accuracy.

Despite the potential benefits, the amalgamation of massive MIMO and AirComp would incur high fabrication cost and energy consumption due to the conditional large numbers of radio frequency (RF) chains as well as analog-to-digital converters (ADCs) associated with antennas. To address these issues, the hybrid beamforming structure has emerged as a promising solution, which allows the AP to employ a smaller number of RF chains than that of antenna elements [23]–[33]. Considering the tradeoff between performance versus complexity, different types of hybrid beamforming (e.g., fully-connected [24], [25], [27]–[29], [31], [32] and partially-connected [23], [26], [30], [33]) with different kinds of RF electronic modules (e.g., analog switches [23], [26] and analog phase shifters [24], [25], [27]–[33]) have been investigated. In particular, the fully-connected hybrid beamforming is appealing as it can achieve better performance with each RF chain connected to all antennas, while the partially-connected hybrid beamforming shows a lower complexity but compromised performance, since it allows every RF chain to connect to only part of antennas.

Although there have been a handful of prior works investigating hybrid beamforming for massive MIMO communication systems [23]–[33], these designs cannot be directly applied to massive MIMO AirComp systems, due to the following reasons. First, the design objectives are fundamentally different. In massive MIMO communication systems, the hybrid beamforming design aims to maximize the communication rate by eliminating the interference, while in massive MIMO AirComp systems, the hybrid beamforming design targets for minimizing the computation error by harnessing the “interference”. Second, inspired by the low-latency requirement of data-intensive IoT applications, the hybrid beamforming in massive MIMO AirComp systems should achieve good computational accuracy with a less computational complexity/latency. Motivated by the above observations, in this paper, we investigate the hybrid beamforming design for massive MIMO AirComp systems. As an initial attempt, we adopt the fully-connected hybrid beamforming with analog phase shifters to achieve full spatial degrees of freedom of massive MIMO for AirComp.

The main results of this paper are summarized as follows.

- We consider a massive MIMO AirComp system with a massive-antenna AP and a massive number of wireless devices, which aims to compute multiple arithmetic sum functions of the recorded signals

from all the devices. Our objective is to jointly optimize the transmit digital beamforming, the receive analog beamforming, and the receive digital beamforming to minimize the computation MSE. The formulated problem is generally intractable due to the highly coupled variables in the objective function and the constant modulus constraints for the receive analog beamforming.

- To address the non-convex hybrid beamforming optimization problem, we propose an alternating-optimization-based approach to alternately optimize the transmit beamforming and the receive analog and digital beamforming. In particular, we optimize the receive analog beamforming by applying two methods, namely the successive convex approximation (SCA) and block coordinate descent (BCD), respectively. While the SCA-based method leads to a better performance, the BCD-based method enjoys a lower computational complexity at the expense of a compromised performance.
- To gain more insights, we analyze the MSE performance for a special case with a fully-digital receiver at the AP. In this case, we show that the optimal (digital) receive beamforming follows the sum-minimization-MSE (sum-MMSE) structure. With the help of the sum-MMSE receive beamforming, we prove that the computation MSE is inversely proportional to the number of receive antennas when the AP adopts the techniques of massive MIMO.
- Furthermore, numerical results show that the proposed hybrid beamforming design substantially improves the computation MSE performance of multi-function/multi-modal massive MIMO AirComp systems as compared to other benchmark scheme inspired by the ZF-based fully-digital beamforming design in [5]. The SCA-based algorithm is shown to perform closely to the performance upper bound achieved by the fully-digital beamforming.

The remainder of this paper is organized as follows. Section II introduces the system model and formulates the computational MSE minimization problem. Section III presents the alternating-optimization-based approaches to address the formulated problem, where two algorithms are proposed to optimize the receive analog beamforming based on SCA and BCD, respectively. Section IV analyzes the MSE performance for the special case with fully-digital beamforming design. Section V presents the numerical results. Finally, Section VI draws the conclusion.

Notations: Throughout this paper, we adopt bold upper-case letters for matrices and bold lower-case letters for vectors. For a matrix \mathbf{A} , $\mathbf{A}(i, j)$ represents the entry on the i^{th} row and the j^{th} column, while \mathbf{A}^\dagger , \mathbf{A}^* , \mathbf{A}^T , and \mathbf{A}^H denote its Moore-Penrose pseudo inverse, conjugate, transpose, and Hermitian

transpose, respectively. Furthermore, \mathbf{I} denotes the identity matrix whose dimension will be clear from the context, and $\mathbb{C}^{m \times n}$ denotes the m -by- n dimensional complex space. The notations $\mathbb{E}(\cdot)$, $\text{Tr}(\cdot)$, $\det(\cdot)$, $\text{vec}(\cdot)$, $\Re(\cdot)$, and $\|\cdot\|$ represent the expectation, trace, determinant, vectorization, real part, and Frobenius norm of an input variable, respectively. $\nabla_{\mathbf{x}} f(\mathbf{x})$ denotes the gradient of $f(\mathbf{x})$ with respect to \mathbf{x} . \circ is the Hadamard product between two matrices. The circularly symmetric complex Gaussian (CSCG) distribution with mean Υ and covariance matrix Φ is denoted by $\mathcal{CN}(\Upsilon, \Phi)$. $\text{mod}(a, b)$ is the modulus operation of a with respect to b .

II. SYSTEM MODEL AND PROBLEM FORMULATION

A. System Model

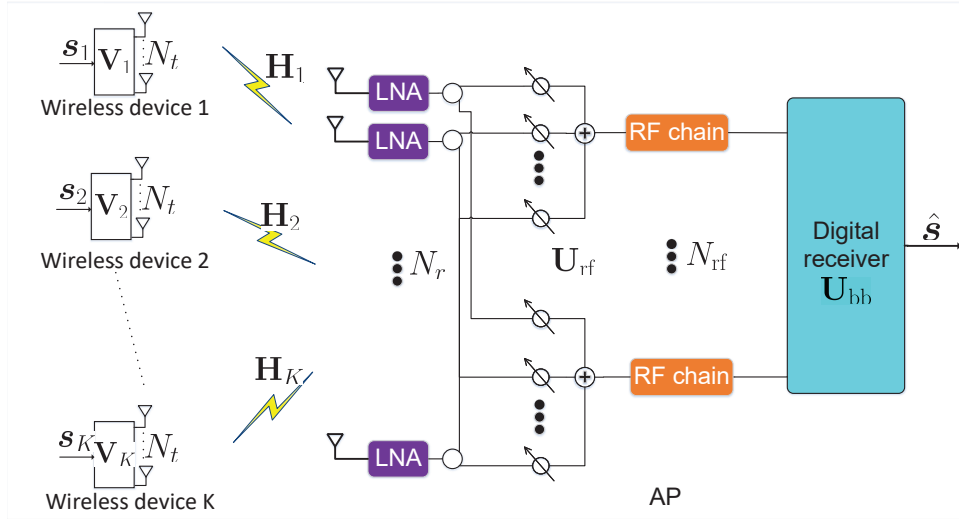


Fig. 1: A massive MIMO AirComp system with hybrid beamforming.

We consider a massive MIMO AirComp system as shown in Fig. 1, where the AP simultaneously serves K wireless devices. Suppose each device is equipped with N_t transmit antennas and the AP is equipped with N_r receive antenna elements, each of which is connected to a low-noise amplifier (LNA). With massive MIMO, it is assumed that $N_r \gg N_t$. Besides, fully-digital beamforming is adopted at the wireless devices while hybrid beamforming with only N_{rf} RF chains is implemented at the AP, with $N_{\text{rf}} \leq N_r$, to reduce the hardware cost and implementation complexity [27], [28]. For the purpose of initial investigation, we consider the fully-connected hybrid beamforming with phase shifters to achieve full spatial degrees of freedom of massive MIMO. For the convenience of expression, we denote $\mathcal{N}_r \triangleq \{1, 2, \dots, N_r\}$ and $\mathcal{N}_{\text{rf}} \triangleq \{1, 2, \dots, N_{\text{rf}}\}$ as the antenna and RF chain sets, respectively.

In this massive MIMO AirComp system, every device records L heterogeneous time-varying parameters (e.g., humidity, temperature, noise) of the environment with $L \leq \min(N_{\text{rf}}, N_t)$. In particular, since we focus on studying multi-function/multi-modal massive MIMO AirComp systems, we have $L > 1$. The devices simultaneously transmit the recorded data to the AP for computation. At a particular time slot, we denote s_{kl} as the recorded data of the l th parameter at device k , $k \in \mathcal{K} \triangleq \{1, 2, \dots, K\}$, $l \in \mathcal{L} \triangleq \{1, 2, \dots, L\}$, and $\mathbf{s}_k = [s_{k1}, \dots, s_{kL}]^T \in \mathbb{C}^{L \times 1}$ as the composite record vector at that device. Without loss of generality, the collected data vector is assumed to be normalized and independent from each other, i.e., $\mathbb{E}(\mathbf{s}_k \mathbf{s}_k^H) = \mathbf{I}$, $\mathbb{E}(\mathbf{s}_k \mathbf{s}_j^H) = \mathbf{0}$, $\forall k, j \in \mathcal{K}, k \neq j$, where the normalization factor for each data type is uniform for all devices and can be inverted at the AP for recovering the original data.

In order to support ultrafast data computation, the AP exploits the superposition property of the multiple access channel to directly compute the target nomographic function with reduced communication overheads. In this paper, we are interested in the sum operation, while the design is also extendable for other nomographic functions [6]. Towards this end, each device transmits a vector \mathbf{s}_k and the AP is interested in estimating $\mathbf{s} = \sum_{k=1}^K \mathbf{s}_k$, which is referred to as the target-function vector [5].

Referring to the proposed system model, the transmitted signal by device k is denoted by

$$\mathbf{x}_k = \mathbf{V}_k \mathbf{s}_k, \quad (1)$$

where $\mathbf{V}_k \in \mathbb{C}^{N_t \times L}$ denotes the transmit beamforming matrix. Let P denote the maximum transmit power at each device. Accordingly, we have $\mathbb{E}(\|\mathbf{x}_k\|^2) = \text{Tr}(\mathbf{V}_k \mathbf{V}_k^H) \leq P, \forall k \in \mathcal{K}$.

It is assumed that the channel state information (CSI) is perfectly known at both the AP and the devices¹. Then the received signal vector at the AP is given by

$$\mathbf{y} = \sum_{k=1}^K \mathbf{H}_k \mathbf{V}_k \mathbf{s}_k + \mathbf{n}, \quad (2)$$

where $\mathbf{H}_k \in \mathbb{C}^{N_r \times N_t}$ denotes the channel matrix from device k to the AP and $\mathbf{n} \in \mathbb{C}^{N_r \times 1}$ is the AWGN vector with $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I})$.

Next, the AP adopts the hybrid beamforming for AirComp. Here, the hybrid beamforming needs to be properly designed for not only harnessing part of the inter-device interference to facilitate the computation, but also eliminating the inter-function interference. Let $\mathbf{U}_{\text{rf}} \in \mathbb{C}^{N_r \times N_{\text{rf}}}$ denote the receive

¹Practical channel estimation methods, such as random vector quantization codebook training, limited feedback, and over-the-air signaling procedure, have been proposed in [34]–[37]. Assuming time division duplexing protocol, both the devices and the AP can obtain the channel by applying the above channel estimation methods and exploiting the uplink-downlink channel reciprocity.

analog beamforming, whose entries have constant modulus, i.e., $|\mathbf{U}_{\text{rf}}(i, j)| = 1, \forall i \in \mathcal{N}_r, j \in \mathcal{N}_{\text{rf}}$, and $\mathbf{U}_{\text{bb}} \in \mathbb{C}^{N_{\text{rf}} \times L}$ denote the low-dimension receive digital beamforming. Therefore, the processed signal after the adopted hybrid beamforming can be expressed as

$$\hat{\mathbf{s}} = \mathbf{U}_{\text{bb}}^H \mathbf{U}_{\text{rf}}^H \mathbf{y}. \quad (3)$$

Consequently, the computational accuracy is measured by the MSE between $\hat{\mathbf{s}}$ and \mathbf{s} , which is given by [5]

$$\begin{aligned} & \text{MSE}(\{\mathbf{V}_k\}, \mathbf{U}_{\text{rf}}, \mathbf{U}_{\text{bb}}) \\ &= \mathbb{E}[\|\mathbf{s} - \hat{\mathbf{s}}\|^2] \\ &= \sum_{k=1}^K \text{Tr} [(\mathbf{U}_{\text{bb}}^H \mathbf{U}_{\text{rf}}^H \mathbf{H}_k \mathbf{V}_k - \mathbf{I})(\mathbf{U}_{\text{bb}}^H \mathbf{U}_{\text{rf}}^H \mathbf{H}_k \mathbf{V}_k - \mathbf{I})^H] + \sigma^2 \text{Tr}(\mathbf{U}_{\text{bb}}^H \mathbf{U}_{\text{rf}}^H \mathbf{U}_{\text{rf}} \mathbf{U}_{\text{bb}}). \end{aligned} \quad (4)$$

B. Problem Formulation

In this work, we are interested in minimizing the MSE defined in (4) by jointly optimizing the transmit beamforming $\{\mathbf{V}_k\}$ at the devices and the receive hybrid beamforming \mathbf{U}_{rf} and \mathbf{U}_{bb} at the AP, subject to the constant modulus constraints on \mathbf{U}_{rf} and the maximum power budget constraints on $\{\mathbf{V}_k\}$. In particular, the MSE minimization problem can be formulated as

$$\begin{aligned} \mathcal{P}1: & \underset{\mathbf{V}_k, \mathbf{U}_{\text{rf}}, \mathbf{U}_{\text{bb}}}{\text{minimize}} \quad \text{MSE}(\{\mathbf{V}_k\}, \mathbf{U}_{\text{rf}}, \mathbf{U}_{\text{bb}}) \\ & \text{subject to} \quad \text{Tr}(\mathbf{V}_k \mathbf{V}_k^H) \leq P, \forall k \in \mathcal{K} \\ & \quad |\mathbf{U}_{\text{rf}}(i, j)| = 1, \forall i \in \mathcal{N}_r, j \in \mathcal{N}_{\text{rf}}. \end{aligned} \quad (5)$$

Problem $\mathcal{P}1$ is difficult to solve, as the optimization variables $\{\mathbf{V}_k\}$, \mathbf{U}_{rf} , and \mathbf{U}_{bb} are highly coupled in the objective function while the unit modulus constraints of \mathbf{U}_{rf} are highly non-convex. Furthermore, problem $\mathcal{P}1$ aims to minimize the computation MSE for recovering $\mathbf{s} = \sum_{k=1}^K \mathbf{s}_k$ by exploiting the interference from various wireless devices. Note that this is significantly different from the conventional hybrid beamforming design problems in massive MIMO systems which mainly maximize the communication rate (for decoding \mathbf{s}_k 's individually) by eliminating the inter-device interference. Hence, the conventional designs are not directly applicable to the considered problem $\mathcal{P}1$. Besides, due to the requirement of ultrafast computation for data aggregation, attaining an efficient solution to problem $\mathcal{P}1$ with low computational complexity is also desirable. To the best of our knowledge, however, there lacks computationally efficient and systematic algorithms to solve such non-convex problems optimally. As a compromise approach, in

the next section, by exploiting the structure of problem, we propose an alternating-optimization-based method to iteratively optimize $\{\mathbf{V}_k\}$, \mathbf{U}_{rf} , and \mathbf{U}_{bb} .

III. HYBRID BEAMFORMING DESIGNS

In this section, a novel hybrid beamforming approach is proposed to handle problem $\mathcal{P}1$, by updating $\{\mathbf{V}_k\}$, \mathbf{U}_{rf} , and \mathbf{U}_{bb} in an alternating manner. In the following, we first optimize the transmit beamforming $\{\mathbf{V}_k\}$ by using the Lagrange duality method, then we update the receive analog beamforming \mathbf{U}_{rf} by using the techniques of SCA or BCD, and finally optimize \mathbf{U}_{bb} by exploiting the first order optimality condition.

A. Optimization of Transmit Beamforming $\{\mathbf{V}_k\}$

First, we focus on the optimization of $\{\mathbf{V}_k\}$ under given \mathbf{U}_{rf} and \mathbf{U}_{bb} . In this case, problem $\mathcal{P}1$ can be equivalently decomposed into the following K subproblems each for one device $k \in \mathcal{K}$, by ignoring the constant term $\sigma^2 \text{Tr}(\mathbf{U}_{\text{rf}} \mathbf{U}_{\text{bb}} \mathbf{U}_{\text{bb}}^H \mathbf{U}_{\text{rf}}^H)$:

$$\begin{aligned} \mathcal{P}2: \underset{\mathbf{V}_k}{\text{minimize}} \quad & \text{Tr}[(\mathbf{U}_{\text{bb}}^H \mathbf{U}_{\text{rf}}^H \mathbf{H}_k \mathbf{V}_k - \mathbf{I})(\mathbf{U}_{\text{bb}}^H \mathbf{U}_{\text{rf}}^H \mathbf{H}_k \mathbf{V}_k - \mathbf{I})^H] \\ \text{subject to} \quad & \text{Tr}(\mathbf{V}_k \mathbf{V}_k^H) \leq P, \forall k \in \mathcal{K}. \end{aligned} \quad (6)$$

Problem $\mathcal{P}2$ is a convex quadratic optimization problem that satisfies the Slater's constraint condition, and therefore, this problem can be optimally solved by using standard convex optimization techniques [38]. To gain more insights, we apply the Lagrange duality method to find a semi-closed-form optimal solution, which is summarized in the following lemma.

Lemma 1: The optimal transmit beamforming solution to problem $\mathcal{P}2$ for device k is given by:

$$\mathbf{V}_k^{\text{opt}} = (\mathbf{H}_k^H \mathbf{U}_{\text{rf}} \mathbf{U}_{\text{bb}} \mathbf{U}_{\text{bb}}^H \mathbf{U}_{\text{rf}}^H \mathbf{H}_k + \mu_k^{\text{opt}} \mathbf{I})^{-1} \mathbf{H}_k^H \mathbf{U}_{\text{rf}} \mathbf{U}_{\text{bb}}, \quad (7)$$

where $\mu_k^{\text{opt}} \geq 0$, $k \in \mathcal{K}$, denotes the optimal Lagrange multiplier associated with the power constraint for device k in problem $\mathcal{P}2$. Here, if $\mathbf{H}_k^H \mathbf{U}_{\text{rf}} \mathbf{U}_{\text{bb}} \mathbf{U}_{\text{bb}}^H \mathbf{U}_{\text{rf}}^H \mathbf{H}_k$ is invertible and

$$\text{Tr}((\mathbf{H}_k^H \mathbf{U}_{\text{rf}} \mathbf{U}_{\text{bb}} \mathbf{U}_{\text{bb}}^H \mathbf{U}_{\text{rf}}^H \mathbf{H}_k)^{-2} \mathbf{H}_k^H \mathbf{U}_{\text{rf}} \mathbf{U}_{\text{bb}} \mathbf{U}_{\text{bb}}^H \mathbf{U}_{\text{rf}}^H \mathbf{H}_k) < P \quad (8)$$

holds, we have $\mu_k^{\text{opt}} = 0$; otherwise, μ_k^{opt} is chosen such that the equality in (9) holds.

$$\text{Tr}((\mathbf{H}_k^H \mathbf{U}_{\text{rf}} \mathbf{U}_{\text{bb}} \mathbf{U}_{\text{bb}}^H \mathbf{U}_{\text{rf}}^H \mathbf{H}_k + \mu_k^{\text{opt}} \mathbf{I})^{-2} \mathbf{H}_k^H \mathbf{U}_{\text{rf}} \mathbf{U}_{\text{bb}} \mathbf{U}_{\text{bb}}^H \mathbf{U}_{\text{rf}}^H \mathbf{H}_k) = P. \quad (9)$$

Proof: See Appendix A. ■

From Lemma 1, we can see that $\{\mathbf{V}_k\}$ is optimized by considering the following two cases. If the transmit power budget P is sufficiently large, then we choose $\mu_k^{\text{opt}} = 0$ such that the objective function value of problem $\mathcal{P}2$ is forced to be zero; otherwise, if the transmit power budget P is limited, then we choose μ_k^{opt} such that the transmit power is fully used to minimize the computation MSE.

B. Optimization of Receive Analog Beamforming \mathbf{U}_{rf}

In this subsection, we optimize \mathbf{U}_{rf} under given $\{\mathbf{V}_k\}$ and \mathbf{U}_{bb} , for which the problem is given by

$$\begin{aligned} \mathcal{P}3: \quad & \underset{\mathbf{U}_{\text{rf}}}{\text{minimize}} \text{MSE}(\{\mathbf{V}_k\}, \mathbf{U}_{\text{rf}}, \mathbf{U}_{\text{bb}}) \\ & \text{subject to } |\mathbf{U}_{\text{rf}}(i, j)| = 1, \forall i \in \mathcal{N}_r, j \in \mathcal{N}_{\text{rf}}. \end{aligned} \quad (10)$$

Problem $\mathcal{P}3$ is still challenging to solve mainly due to the constant modulus constraints which are intrinsically non-convex. To address this issue, we propose two algorithms by using SCA and BCD, respectively.

1) *SCA*: To gain more insights, motivated by [32], we transform problem $\mathcal{P}3$ into a more tractable form by exploiting the SCA method. To start with, we first rewrite the constant modulus constraints in its exponential form. Let $\mathbf{u}_{\text{rf}} = \text{vec}(\mathbf{U}_{\text{rf}}) \in \mathbb{C}^{N_r N_{\text{rf}} \times 1}$ and $\boldsymbol{\theta} \triangleq [\theta_1, \theta_2, \dots, \theta_{N_r N_{\text{rf}}}]^T$ denote the vectorization of \mathbf{U}_{rf} and the corresponding phase vector of \mathbf{u}_{rf} , respectively. Then problem $\mathcal{P}3$ is transformed to the following equivalent problem:

$$\begin{aligned} \mathcal{P}4: \quad & \underset{\boldsymbol{\theta}}{\text{minimize}} f(\boldsymbol{\theta}) \\ & \text{subject to } -\pi \leq \boldsymbol{\theta}(i) \leq \pi, \forall i \in \mathcal{Y}, \end{aligned} \quad (11)$$

where

$$f(\boldsymbol{\theta}) = \text{MSE}(\{\mathbf{V}_k\}, \mathbf{U}_{\text{rf}}(\boldsymbol{\theta}), \mathbf{U}_{\text{bb}}), \quad (12)$$

$$\mathcal{Y} \triangleq \{1, 2, \dots, N_r N_{\text{rf}}\}. \quad (13)$$

Note that $\mathbf{U}_{\text{rf}}(\boldsymbol{\theta})$ in (12) means that \mathbf{U}_{rf} is a function of $\boldsymbol{\theta}$, which can be written element-wisely as:

$$\mathbf{U}_{\text{rf}}(i, j) = e^{\sqrt{-1}\boldsymbol{\theta}((j-1)N_r + i)}, \forall i \in \mathcal{N}_r, j \in \mathcal{N}_{\text{rf}}, \quad (14)$$

where $\sqrt{-1}$ denotes the imaginary unit.

With the above derivation, we transform the intractable constant modulus constraints into linear constraints equivalently. From problem $\mathcal{P}4$, we can see that the objective function is non-convex. To address this issue, according to the technique of SCA, we need to find a surrogate function of $f(\boldsymbol{\theta})$ first. Let r

and $\boldsymbol{\theta}_r$ denote the iteration number and the current point in the r th iteration. Under the given local point $\boldsymbol{\theta}_r$, we can obtain one of the corresponding surrogate functions denoted by $\hat{f}(\boldsymbol{\theta}, \boldsymbol{\theta}_r)$ via exploiting the first-order Taylor approximation, which is given by

$$\hat{f}(\boldsymbol{\theta}, \boldsymbol{\theta}_r) = f(\boldsymbol{\theta}_r) + \boldsymbol{\gamma}_{\boldsymbol{\theta}_r}^H (\boldsymbol{\theta} - \boldsymbol{\theta}_r) + \tau \|\boldsymbol{\theta} - \boldsymbol{\theta}_r\|^2, \quad (15)$$

$$\boldsymbol{\gamma}_{\boldsymbol{\theta}_r} = \nabla_{\boldsymbol{\theta}} f(\boldsymbol{\theta})|_{\boldsymbol{\theta}=\boldsymbol{\theta}_r} = -\text{vec}\{2\Re[\sqrt{-1}\mathbf{U}_{\text{rf},r}^* \circ \mathbf{F}_r]\}, \quad (16)$$

$$\mathbf{F}_r = \left(\sum_{k=1}^K \mathbf{H}_k \mathbf{V}_k \mathbf{V}_k^H \mathbf{H}_k^H + \sigma^2 \mathbf{I} \right) \mathbf{U}_{\text{rf},r} \mathbf{U}_{\text{bb}} \mathbf{U}_{\text{bb}}^H - \sum_{k=1}^K \mathbf{H}_k \mathbf{V}_k \mathbf{U}_{\text{bb}}^H, \quad (17)$$

$$\mathbf{U}_{\text{rf},r}(i, j) = e^{\sqrt{-1}\boldsymbol{\theta}_r((j-1)N_r+i)}, \forall i \in \mathcal{N}_r, j \in \mathcal{N}_{\text{rf}}. \quad (18)$$

Note that the third term in (15) is a proximal regularization term with $\tau > 0$ being a small positive number to guarantee the strong convexity and to control the convergence rate [32]. $\boldsymbol{\gamma}_{\boldsymbol{\theta}_r}$ is the gradient of $f(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ at point $\boldsymbol{\theta}_r$, which is calculated by the chain rule. With the above derivation, we can see that $\hat{f}(\boldsymbol{\theta}, \boldsymbol{\theta}_r)$ is the upper bound of $f(\boldsymbol{\theta}_r)$. Besides, $\hat{f}(\boldsymbol{\theta}, \boldsymbol{\theta}_r)$ and $f(\boldsymbol{\theta}_r)$ have the same values and gradient at point $\boldsymbol{\theta}_r$. Thus, $\hat{f}(\boldsymbol{\theta}, \boldsymbol{\theta}_r)$ is a valid surrogate function at point $\boldsymbol{\theta}_r$ [38] and the issue of non-convexity of $f(\boldsymbol{\theta})$ is addressed.

According to the procedure of SCA, we can update $\boldsymbol{\theta}$ and \mathbf{U}_{rf} by solving the following approximated problem of $\mathcal{P}4$:

$$\begin{aligned} \mathcal{P}5: \quad & \underset{\boldsymbol{\theta}}{\text{minimize}} \quad \hat{f}(\boldsymbol{\theta}, \boldsymbol{\theta}_r) \\ & \text{subject to} \quad -\pi \leq \boldsymbol{\theta}(i) \leq \pi, \forall i \in \mathcal{Y}. \end{aligned} \quad (19)$$

Since problem $\mathcal{P}5$ is convex with respect to $\boldsymbol{\theta}$, we can optimize $\boldsymbol{\theta}$ by checking the first-order optimality condition, for which the optimal solution is given by

$$\boldsymbol{\theta}_{r+1}(i) = \text{mod} \left(\boldsymbol{\theta}_r(i) - \frac{\boldsymbol{\gamma}_{\boldsymbol{\theta}_r}(i)}{2\tau}, 2\pi \right), \forall i \in \mathcal{Y}. \quad (20)$$

Finally, the updated variable $\mathbf{U}_{\text{rf},r+1}$ can be obtained by

$$\mathbf{U}_{\text{rf},r+1}(i, j) = e^{\sqrt{-1}\boldsymbol{\theta}_{r+1}((j-1)N_r+i)}, \forall i \in \mathcal{N}_r, j \in \mathcal{N}_{\text{rf}}. \quad (21)$$

The SCA-based algorithm to address problem $\mathcal{P}3$ is summarized in Algorithm 1. According to the analysis in [32] and [39], the proposed SCA-based algorithm can guarantee the convergence of a local optimum theoretically when τ is chosen properly. However, since the surrogate function is chosen based on the Taylor expansion which does not fully exploit the special structure of problem $\mathcal{P}3$. Hence, it

Algorithm 1 The SCA-based Algorithm for Solving Problem $\mathcal{P}3$

Set $r = 0$, $\tau > 0$, and $\epsilon > 0$

Repeat

Step 1: Calculate γ_{θ_r} according to (16), (17), and (18);

Step 2: Udata θ_{r+1} according to (20);

Step 3: Update $\mathbf{U}_{\text{rf},r+1}$ according to (21);

Step 4: $r = r + 1$;

until the decrease of the objective function in problem $\mathcal{P}3$ is less than ϵ .

may lead to high computational complexity and slow convergence rate. In the following, we propose an alternative effective approach with lower complexity in handling (10).

2) *Low-complexity Design via BCD:* Considering the tradeoff between the performance and complexity, we develop an alternative low-complexity algorithm to address problem $\mathcal{P}3$ by exploiting BCD. Based on further manipulation, problem $\mathcal{P}3$ can be equivalently converted as:

$$\begin{aligned} \mathcal{P}6: \quad & \underset{\mathbf{U}_{\text{rf}}}{\text{minimize}} \quad \text{Tr}(\mathbf{U}_{\text{rf}}^H \mathbf{A} \mathbf{U}_{\text{rf}} \mathbf{C}) - 2\Re\{\text{Tr}(\mathbf{U}_{\text{rf}}^H \mathbf{B})\} \\ & \text{subject to} \quad |\mathbf{U}_{\text{rf}}(i, j)| = 1, \forall i \in \mathcal{N}_r, j \in \mathcal{N}_{\text{rf}}, \end{aligned} \quad (22)$$

where $\mathbf{A} \triangleq \sum_{k=1}^K \mathbf{H}_k \mathbf{V}_k \mathbf{V}_k^H \mathbf{H}_k^H + \sigma^2 \mathbf{I}$, $\mathbf{B} \triangleq \sum_{k=1}^K \mathbf{H}_k \mathbf{V}_k \mathbf{U}_{\text{bb}}^H$, and $\mathbf{C} \triangleq \mathbf{U}_{\text{bb}} \mathbf{U}_{\text{bb}}^H$. Since the unit modulus constraints are separable, inspired by [40], we can update \mathbf{U}_{rf} by applying the BCD type algorithm, i.e., in each step we only update one entry of \mathbf{U}_{rf} by fixing others. Without loss of generality, by defining

$$\phi(\mathbf{U}_{\text{rf}}) = \text{Tr}(\mathbf{U}_{\text{rf}}^H \mathbf{A} \mathbf{U}_{\text{rf}} \mathbf{C}) - 2\Re\{\text{Tr}(\mathbf{U}_{\text{rf}}^H \mathbf{B})\}, \quad (23)$$

we investigate the problem of minimizing $\phi(\mathbf{U}_{\text{rf}})$ with respect to $\mathbf{U}_{\text{rf}}(i, j)$ for a particular $i \in \mathcal{N}_r$ and $j \in \mathcal{N}_{\text{rf}}$ subject to the unit modulus constraint $|\mathbf{U}_{\text{rf}}(i, j)| = 1$, i.e.,

$$\mathcal{P}7: \quad \underset{|\mathbf{U}_{\text{rf}}(i, j)|=1}{\text{minimize}} \quad \phi(\mathbf{U}_{\text{rf}}). \quad (24)$$

It can be observed that the objective function $\phi(\mathbf{U}_{\text{rf}})$ can be re-expressed as a quadratic function with respect to $\mathbf{U}_{\text{rf}}(i, j)$, i.e., $\tilde{\phi}(\mathbf{U}_{\text{rf}}(i, j)) \triangleq a|\mathbf{U}_{\text{rf}}(i, j)|^2 - 2\Re\{b^* \cdot \mathbf{U}_{\text{rf}}(i, j)\}$ for some real number a and some complex number b that will be explained later. Due to the unit modulus constraint, the first term of $\tilde{\phi}(\mathbf{U}_{\text{rf}}(i, j))$ is a constant. Then problem $\mathcal{P}7$ can be simplified as

$$\mathcal{P}8: \quad \underset{|\mathbf{U}_{\text{rf}}(i, j)|=1}{\text{maximize}} \quad \Re\{b^* \cdot \mathbf{U}_{\text{rf}}(i, j)\}. \quad (25)$$

It is clear that the optimal solution of $\mathbf{U}_{\text{rf}}(i, j)$ to problem $\mathcal{P}8$ is equal to $\frac{b}{|b|}$. Hence, we only need to obtain b for the update of $\mathbf{U}_{\text{rf}}(i, j)$.

Now we propose a handy method to update the complex number b . First, the following equality holds [41]:

$$\nabla_{\mathbf{U}_{\text{rf}}^*(i, j)} \tilde{\phi}(\mathbf{U}_{\text{rf}}(i, j)) \Big|_{\mathbf{U}_{\text{rf}}(i, j) = \tilde{\mathbf{U}}_{\text{rf}}(i, j)} = \frac{1}{2} \left(a \tilde{\mathbf{U}}_{\text{rf}}(i, j) - b \right). \quad (26)$$

Besides, we have [41]

$$\nabla_{\mathbf{U}_{\text{rf}}^*} \tilde{\phi}(\mathbf{U}_{\text{rf}}) \Big|_{\mathbf{U}_{\text{rf}} = \tilde{\mathbf{U}}_{\text{rf}}} = \frac{1}{2} \left(\mathbf{A} \tilde{\mathbf{U}}_{\text{rf}} \mathbf{C} - \mathbf{B} \right). \quad (27)$$

Combining (26) and (27), we have $[\mathbf{A} \tilde{\mathbf{U}}_{\text{rf}} \mathbf{C} - \mathbf{B}]_{ij} = a \tilde{\mathbf{U}}_{\text{rf}}(i, j) - b$. By expanding $[\mathbf{A} \tilde{\mathbf{U}}_{\text{rf}} \mathbf{C}]_{ij}$ and checking the coefficient of $\tilde{\mathbf{U}}_{\text{rf}}(i, j)$, we have

$$a \tilde{\mathbf{U}}_{\text{rf}}(i, j) = \mathbf{A}(i, i) \tilde{\mathbf{U}}_{\text{rf}}(i, j) \mathbf{C}(j, j). \quad (28)$$

Hence, b can be updated according to the following equation:

$$b = \mathbf{A}(i, i) \tilde{\mathbf{U}}_{\text{rf}}(i, j) \mathbf{C}(j, j) - [\mathbf{A} \tilde{\mathbf{U}}_{\text{rf}} \mathbf{C}]_{ij} + \mathbf{B}(i, j). \quad (29)$$

Considering the above analysis, we can update the entries of \mathbf{U}_{rf} iteratively. The corresponding algorithm for solving problem $\mathcal{P}6$ is summarized in Algorithm 2, where we need to accordingly update \mathbf{Q} in Step 3 once $\mathbf{U}_{\text{rf}}(i, j)$ is updated (which is done in Step 4). As we can see, Step 3 is the most costly step requiring complexity $\mathcal{O}(N_r N_{\text{rf}})$. Hence, it can be shown that the algorithm has complexity of $\mathcal{O}(N_r^2 N_{\text{rf}}^2)$.

From the above derivation, it is shown that we can obtain the optimal solution of each subproblem for one element of the receive analog beamforming while fixing the others. Considering the concept of the BCD algorithm [42], the proposed algorithm in Algorithm 2 can converge to a stationary point of problem $\mathcal{P}6$. Furthermore, as compared with the SCA-based algorithm, this element-wise update in the BCD method can reduce the computational complexity by exploiting the special structures of the constant modulus constraints at the expense of certain performance degradation, which will be discussed in Section III-D.

C. Optimization of Receive Digital Beamforming \mathbf{U}_{bb}

Then, we optimize \mathbf{U}_{bb} under fixed \mathbf{U}_{rf} and $\{\mathbf{V}_k\}$, for which we need to solve an unconstrained convex optimization problem given as:

$$\mathcal{P}9: \underset{\mathbf{U}_{\text{bb}}}{\text{minimize}} \text{MSE}(\{\mathbf{V}_k\}, \mathbf{U}_{\text{rf}}, \mathbf{U}_{\text{bb}}). \quad (30)$$

Algorithm 2 BCD-type Algorithm for Solving Problem $\mathcal{P}6$

Set $r = 0$, $\mathbf{Q}_r = \mathbf{A}\mathbf{U}_{\text{rf},r}\mathbf{C}$, and $\epsilon > 0$

Repeat

For $i \in \mathcal{N}_r$ and $j \in \mathcal{N}_{\text{rf}}$

Step 1: Calculate $b = \mathbf{A}(i, i)\mathbf{U}_{\text{rf},r}(i, j)\mathbf{C}(j, j) - \mathbf{Q}_r(i, j) + \mathbf{B}(i, j)$;

Step 2: Calculate $x = \frac{b}{|b|}$;

Step 3: Update $\mathbf{Q}_{r+1} = \mathbf{Q}_r + (x - \mathbf{U}_{\text{rf},r}(i, j))\mathbf{A}(:, i)\mathbf{C}(j, :)$;

Step 4: Update $\mathbf{U}_{\text{rf},r+1}(i, j) = x$;

end

$r = r + 1$;

until the decrease of the objective function in problem $\mathcal{P}6$ is less than ϵ .

The receive digital beamforming to problem $\mathcal{P}9$ can be updated by applying the first-order optimality condition, which is given by

$$\mathbf{U}_{\text{bb}}^{\text{opt}} = \left(\mathbf{U}_{\text{rf}}^H \left(\sum_{k=1}^K \mathbf{H}_k \mathbf{V}_k \mathbf{V}_k^H \mathbf{H}_k^H + \sigma^2 \mathbf{I} \right) \mathbf{U}_{\text{rf}} \right)^{-1} \mathbf{U}_{\text{rf}}^H \left(\sum_{k=1}^K \mathbf{H}_k \mathbf{V}_k \right). \quad (31)$$

From (31), it can be observed that the expression of \mathbf{U}_{bb} has a sum-MMSE structure, which is different from the conventional MMSE receiver for multiuser massive MIMO communication systems in the form of $\left(\mathbf{U}_{\text{rf}}^H \left(\sum_{k=1}^K \mathbf{H}_k \mathbf{V}_k \mathbf{V}_k^H \mathbf{H}_k^H + \sigma^2 \mathbf{I} \right) \mathbf{U}_{\text{rf}} \right)^{-1} \mathbf{U}_{\text{rf}}^H \mathbf{H}_k \mathbf{V}_k$ for estimating the individual message \mathbf{s}_k from device k [43]. More specifically, for the term outside the matrix inversion, we have $\mathbf{U}_{\text{rf}}^H \left(\sum_{k=1}^K \mathbf{H}_k \mathbf{V}_k \right)$ in (31) for estimating \mathbf{s} in AirComp but $\mathbf{U}_{\text{rf}}^H \mathbf{H}_k \mathbf{V}_k$ in conventional MMSE receiver for individually detecting \mathbf{s}_k 's in communications. This is due to the fact that the signals from all the devices are exploited concurrently and beneficially to assist functional computation in massive MIMO AirComp systems, which is in shape contrast to the conventional multi-user massive MIMO communication systems by treating signals from different devices as harmful inter-device interference.

D. Overall Algorithms

According to the aforementioned results, the proposed hybrid beamforming designs for massive MIMO AirComp systems, named Lagrange-SCA and Lagrange-BCD, are summarized in Algorithm 3.

Now, we investigate the complexity of Algorithm 3 for designing the hybrid beamforming in the massive MIMO AirComp system, where only the dominant computational complexity with respect to

Algorithm 3 Pseudo-code of Proposed Hybrid Beamforming with SCA/BCD

Initialize \mathbf{V}_k , \mathbf{U}_{rf} , and \mathbf{U}_{bb} , such that they meet all the constraints;

Repeat

Step 1: Optimize $\mathbf{V}_k, \forall k \in \mathcal{K}$, using the Lagrange duality method;

Step 2: Optimize \mathbf{U}_{rf} using SCA or BCD;

Step 3: Optimize \mathbf{U}_{bb} according to (31);

until a stopping criterion is satisfied.

N_r is considered. In Step 1 of Algorithm 3, the bisection method for solving μ_k requires a complexity independent of N_r . Then, the complexity in calculating the matrix inverse in (7) of Step 1 is $\mathcal{O}(K N_t N_{\text{rf}} N_r)$. Similarly, the dominant computational complexity of the SCA-based algorithm is caused by the calculation of the gradient of the objective function, which can be expressed by $\mathcal{O}(N_r^2 K (L + N_{\text{rf}}))$. Also, the complexity of the BCD-based algorithm is $\mathcal{O}(N_r^2 N_{\text{rf}}^2)$ [40]. Since the number of wireless devices can be large, the complexity of the Lagrange-SCA algorithm is generally higher than that of the Lagrange-BCD algorithm. Finally, the complexity of optimization of \mathbf{U}_{bb} is $\mathcal{O}(N_r^2 (N_t + N_{\text{rf}}))$.

Furthermore, the convergence of the proposed algorithms in Algorithm 3 is obtained in the following theorem.

Theorem 1: Any limiting point of the sequence generated by the Lagrange-SCA algorithm or the Lagrange-BCD algorithm in Algorithm 3 is a stationary point of problem $\mathcal{P}1$.

Proof: See Appendix B. ■

Considering Theorem 1, the objective function values of $\mathcal{P}1$ generated by the Lagrange-SCA and the Lagrange-BCD algorithms both decrease monotonically with respect to the number of iterations, for which the convergence speed will be validated in Section V.

IV. ANALYSIS OF MASSIVE MIMO AIRCOMP SYSTEMS

In this section, we consider the special case with a fully-digital receiver (i.e., $N_{\text{rf}} = N_r$ and $\mathbf{U}_{\text{rf}} = \mathbf{I}$) to gain more design insights. This generally serves as an performance upper bound for other cases with $N_{\text{rf}} < N_r$. For ease of analysis, we first focus on the case with a fixed transmit beamforming given by

$$\mathbf{V}_k^H \mathbf{V}_k = \frac{P}{L} \mathbf{I}. \quad (32)$$

Moreover, since we consider massive MIMO AirComp systems, we impose the following assumption.

Assumption 1: When N_r is sufficiently large, the channel matrices between different wireless devices and the AP is asymptotically orthogonal, i.e.,

$$\mathbf{H}_k^H \mathbf{H}_{k'} \approx \mathbf{0}, \forall k, k' \in \mathcal{K}, k \neq k', \quad (33)$$

$$\mathbf{H}_k^H \mathbf{H}_k \approx \beta N_r \mathbf{I}, \forall k \in \mathcal{K}, \quad (34)$$

where β denotes the path loss from device to the AP.

This assumption is reasonable due to the properties of the massive MIMO technologies [22]. In the following, we analyze the computation MSE performance and the corresponding receive digital beamforming design for the case with N_r being sufficiently large. First, we derive the optimal fully-digital sum-MMSE receiver by setting $\mathbf{U}_{\text{rf}} = \mathbf{I}$ in (31), which is given by

$$\mathbf{U}_{\text{summse}} = \left(\sum_{k=1}^K \mathbf{H}_k \mathbf{V}_k \mathbf{V}_k^H \mathbf{H}_k^H + \sigma^2 \mathbf{I} \right)^{-1} \left(\sum_{k=1}^K \mathbf{H}_k \mathbf{V}_k \right). \quad (35)$$

With the help of (35) and the aforementioned assumptions, then we demonstrate how massive MIMO technologies affect the performance of the AirComp systems and how the receiver in (35) can be simplified via the following theorem and lemma.

Theorem 2: When N_r is sufficiently large, the computation MSE of massive MIMO AirComp systems, given by

$$\text{MSE}(\{\mathbf{V}_k\}, \mathbf{U}_{\text{bb}}) = \sum_{k=1}^K \text{Tr}[(\mathbf{U}_{\text{bb}}^H \mathbf{H}_k \mathbf{V}_k - \mathbf{I})(\mathbf{U}_{\text{bb}}^H \mathbf{H}_k \mathbf{V}_k - \mathbf{I})^H] + \sigma^2 \text{Tr}(\mathbf{U}_{\text{bb}}^H \mathbf{U}_{\text{bb}}), \quad (36)$$

is inversely proportional to N_r and can be written as follows:

$$\text{MSE}(\{\mathbf{V}_k\}, \mathbf{U}_{\text{bb}}) \approx \frac{KL^2\sigma^2}{\beta N_r P}. \quad (37)$$

In particular, as $N_r \rightarrow \infty$, we have $\text{MSE}(\{\mathbf{V}_k\}, \mathbf{U}_{\text{bb}}) \rightarrow 0$.

Proof: See Appendix C. ■

Lemma 2: When N_r is sufficiently large, the optimal sum-MMSE receiver in (35) can be approximated as

$$\tilde{\mathbf{U}}_{\text{summse}} = \sum_{k=1}^K \mathbf{H}_k \mathbf{V}_k (\sigma^2 \mathbf{I} + \beta N_r \mathbf{V}_k^H \mathbf{V}_k)^{-1}. \quad (38)$$

Under the transmit beamforming $\{\mathbf{V}_k\}$ with $\mathbf{V}_k^H \mathbf{V}_k = \frac{P}{L}$ in (32), $\tilde{\mathbf{U}}_{\text{summse}}$ in (38) is rewritten as

$$\tilde{\mathbf{U}}_{\text{summse}} = \sum_{k=1}^K \frac{L}{\beta N_r P} \mathbf{H}_k \mathbf{V}_k. \quad (39)$$

Proof: See Appendix D. ■

From Theorem 2, we know that $\text{MSE}(\{\mathbf{V}_k\}, \mathbf{U}_{\text{bb}})$ increases with K , which indicates that the computation MSE performance is seriously degraded as the number of devices increases. To combat the vanishing computation MSE performance due to large K , we have two options: increasing the transmit power at the wireless devices or increasing the number of receive antennas at the AP. As such, the exploitation of massive MIMO techniques in AirComp systems is a practical solution since the former option is not energy-efficient. It is also observed from Theorem 2 that the computation MSE is not sensitive to L due to our presumption that $L \leq \min(N_{\text{rf}}, N_t) \ll N_r$. Besides, exploiting Lemma 2, the sum-MMSE receiver can be significantly simplified under the considered special case with large N_r .

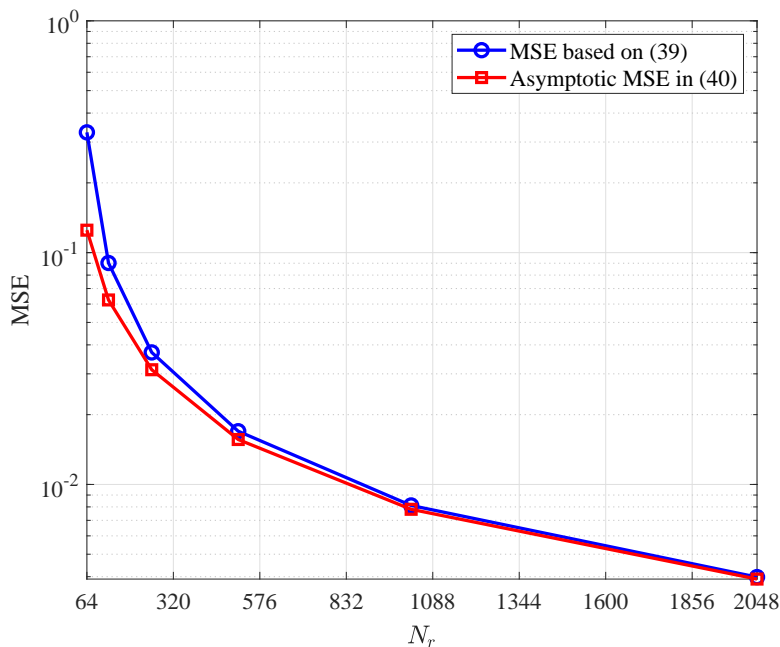


Fig. 2: The computation MSE performance of (36) and (37) versus N_r when $N_t = 2$, $K = 20$, $L = 2$, and $\text{SNR} = 10$ dB.

For illustration, in Fig. 2, we verify the accuracy of the derived asymptotic MSE in (37) by comparing with the practical one obtained based on (36), where we set $\text{SNR} = \frac{\beta P}{\sigma^2} = 10$ dB, $\sigma = 1$, $N_t = 2$, $K = 20$, and $L = 2$. It can be seen in Fig. 2 that the gap between the asymptotic MSE and the practical one decreases as N_r increases. Specifically, the gap becomes negligible when $N_r \geq 512$.

V. SIMULATION RESULTS

In this section, we evaluate the computation MSE performance of the proposed hybrid beamforming design approaches for massive MIMO AirComp systems, as compared with following benchmark schemes.

- FD-ZF: For this scheme, inspired by [5], a fully-digital orthogonal receive beamforming is adopted at the AP. Then, we alternately optimize the transmit beamforming and the receive beamforming. Specifically, the transmit beamforming is updated by the ZF method, i.e., forcing the first term of (4) to be zero. Second, by tightening the power constraints, an approximated problem of the fully-digital receive beamforming is formulated and solved by exploiting differential geometry [44];
- FD: For this scheme, we alternately optimize the transmit beamforming and the fully-digital receive beamforming by exploiting the Lagrange duality method and the sum-MMSE receiver in (35), respectively.

Among the simulation experiments, each channel is assumed to be normalized i.i.d. Rayleigh fading. The initial phases of the receive analog beamforming, \mathbf{U}_{rf} , follow an uniform distribution over $[-\pi, \pi]$. Also, we set $N_t = L$, $\tau = 0.2$, and $\epsilon = 10^{-3}$. Besides, all simulation results are averaged over 500 channel realizations.

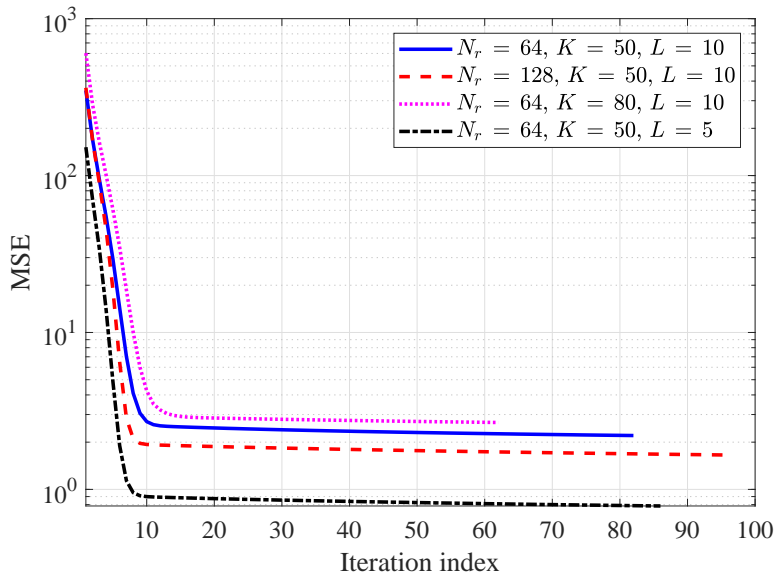


Fig. 3: The convergence performance of the Lagrange-SCA algorithm with $N_{\text{rf}} = 16$ and $\text{SNR} = 10$ dB.

Fig. 3 and Fig. 4 illustrate the convergence behavior for the Lagrange-SCA and the Lagrange-BCD algorithms in Algorithm 3, where $N_{\text{rf}} = 16$ and $\text{SNR} = 10$ dB, respectively. From these figures, we conclude that the proposed Lagrange-SCA algorithm and Lagrange-BCD algorithm both converge rapidly in a few iterations. Besides, it can be observed that the Lagrange-BCD algorithm always converges faster, while the Lagrange-SCA algorithm shows better performance. This unveils the tradeoff between the system

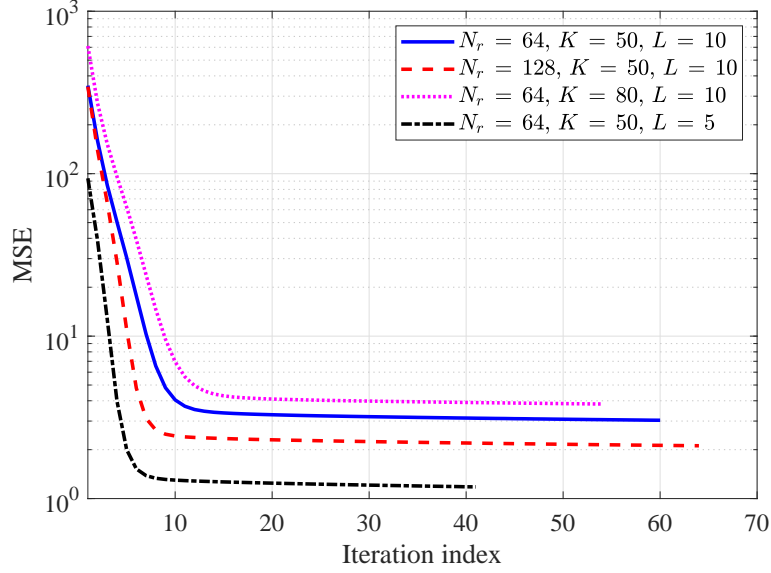


Fig. 4: The convergence performance of the Lagrange-BCD algorithm with $N_{\text{rf}} = 16$ and $\text{SNR} = 10$ dB.

performance and the computational complexity.

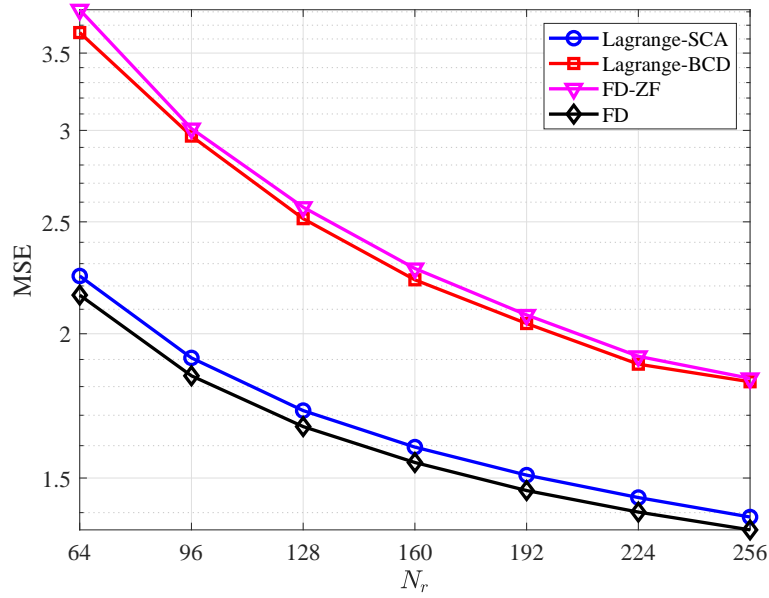


Fig. 5: The computation MSE performance versus N_r when $N_t = 10$, $N_{\text{rf}} = 10$, $K = 50$, $L = 10$, and $\text{SNR} = 10$ dB.

Fig. 5 shows the computation MSE performance of multi-function/multi-modal ($L > 1$) massive MIMO AirComp systems versus N_r under different setups. From Fig. 5, we can see that the computation MSE value of both proposed algorithms decreases considerably when the number of receive antennas increases, showing the effectiveness of applying massive MIMO. In particular, the proposed Lagrange-SCA and

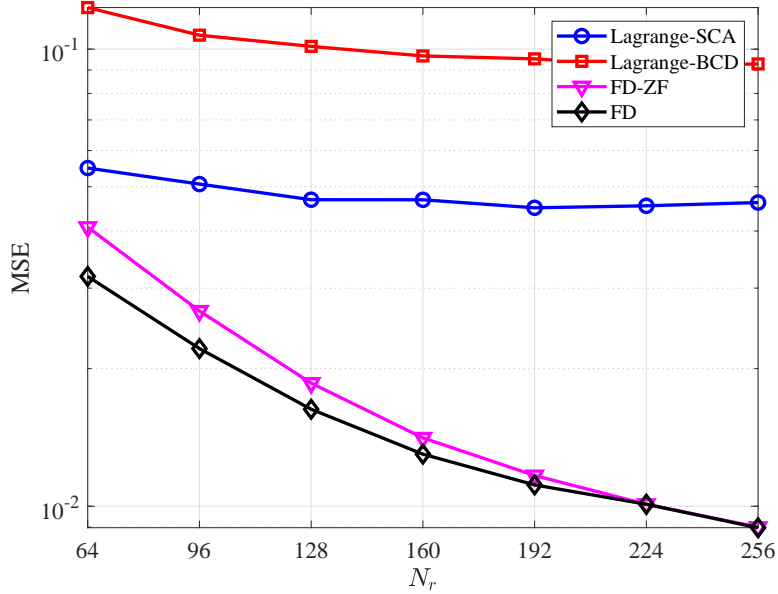


Fig. 6: The computation MSE performance versus N_r when $N_t = 1$, $N_{\text{rf}} = 1$, $K = 20$, $L = 1$, and $\text{SNR} = 10$ dB.

Lagrange-BCD algorithms, where hybrid receivers are adopted, outperform the FD-ZF algorithm with a fully-digital receiver. In fact, our proposed algorithms can exploit the structure of the beamforming problem, offering a better solution for massive MIMO AirComp systems than the heuristic algorithm based on the ZF method. It is also observed that the performance gaps between the FD-ZF algorithm and our proposed algorithms become smaller as N_r increases. This is intuitive, as the FD-ZF can exploit the increased spatial degrees of freedom for enhancing the performance due to its fully-digital beamforming structure. Moreover, the computation MSE performance of the Lagrange-BCD algorithm and the FD-ZF algorithm coincides when $N_r = 256$. It suggests that the diversity gain of the massive MIMO techniques can compensate for the performance degradation due to the drawbacks of the FD-ZF algorithm, which confirms the correctness of our analysis in Section IV. Combining Figs. 3, 4 and 5, it is clear that the Lagrange-SCA algorithm outperforms the Lagrange-BCD algorithm, but at the cost of a slower convergence and a high complexity. Besides, the computation MSE performance of the proposed Lagrange-SCA algorithm is close to the performance upper bound achieved by the FD method, which verifies the effectiveness of the proposed hybrid beamforming design.

To unveil more insights, Fig. 6 shows MSE performance for the special case with single-function/single-modal with a single RF chain equipped at the AP (i.e., $L = 1$ and $N_{\text{rf}} = 1$). It can be seen that the considered hybrid beamforming designs, including the Lagrange-SCA algorithm and the Lagrange-BCD

algorithm, perform poorly in this case. It is mainly due to that the hybrid beamforming design reduces to the simple analog beamforming design when $N_{\text{rf}} = 1$, which can not exploit the spatial degrees of freedom and the array gain brought by massive MIMO.

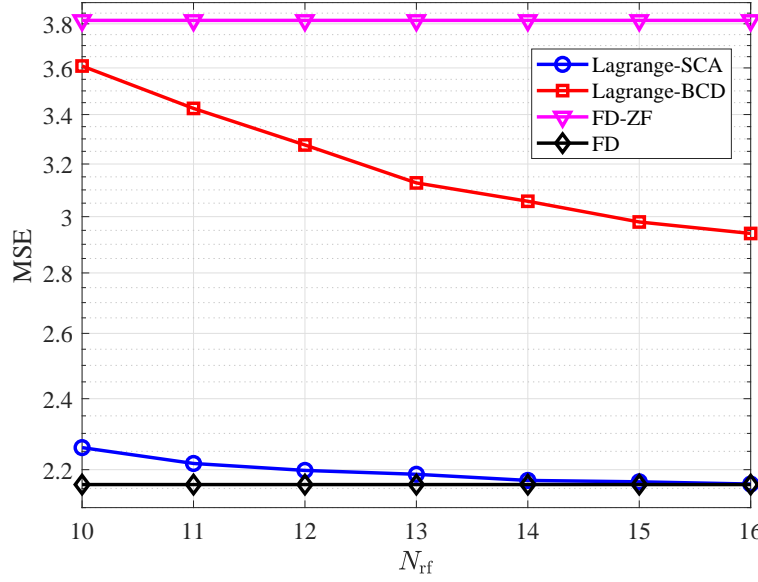


Fig. 7: The computation MSE performance versus N_{rf} when $N_r = 64$, $N_t = 10$, $K = 50$, $L = 10$, and $\text{SNR} = 10$ dB.

Fig. 7 shows the computation MSE performance versus N_{rf} when $N_r = 64$, $N_t = 10$, $K = 50$, $L = 10$, and $\text{SNR} = 10$ dB. As a property of hybrid beamforming schemes, it can be observed that the MSE values achieved by the Lagrange-SCA and Lagrange-BCD algorithms reduce with increasing N_{rf} . By adopting the minimum number of RF chains ($N_{\text{rf}} \geq L$ and $L = 10$), the Lagrange-SCA algorithm shows only slight performance degradation compared with the FD method while both algorithms achieve the same MSE when $N_{\text{rf}} \geq 15$. Besides, the Lagrange-SCA and the Lagrange-BCD algorithms both outperform the FD-ZF algorithm even when $N_{\text{rf}} = 10$. This illustrates that our proposed algorithm can exploit the spatial degrees of freedom efficiently with much less RF chains. In contrast, the FD-ZF algorithm shows a constant computation MSE value when N_{rf} increases, as the performance of a fully-digital beamforming design is independent of the number of RF chains.

The computation MSE performance versus K is shown in Fig. 8 under $N_r = 64$, $N_t = 10$, $N_{\text{rf}} = 16$, $L = 10$, and $\text{SNR} = 10$ dB. One can see that the MSE performance of all the considered schemes increases with K . This coincides with our discussion that supporting the connection of more wireless devices is at the cost of degrading computational accuracy which makes it more challenging to design a

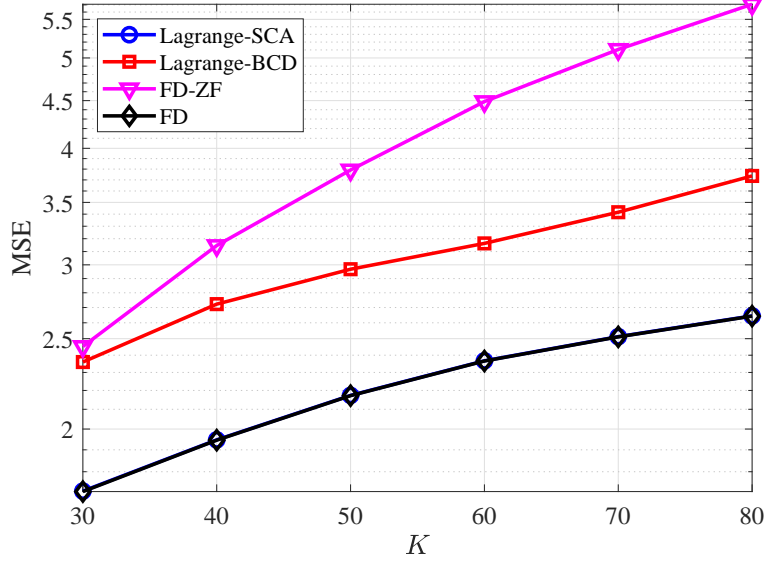


Fig. 8: The computation MSE performance versus K when $N_r = 64$, $N_t = 10$, $N_{rf} = 16$, $L = 10$, and $\text{SNR} = 10$ dB.

common receive fully-digital/hybrid beamforming to equalize all the wireless devices' channels. Besides, the performance of the Lagrange-SCA algorithm still approaches that of the benchmark FD method. Also, our proposed algorithms outperform the FD-ZF algorithm. Furthermore, the performance gap between our proposed algorithms and the FD-ZF algorithm increases as K increases from 30 to 80. It is mainly due to the fact that the FD-ZF algorithm adopts an orthogonal receive beamforming which becomes a highly suboptimal solution jeopardizing the system performance when K is large.

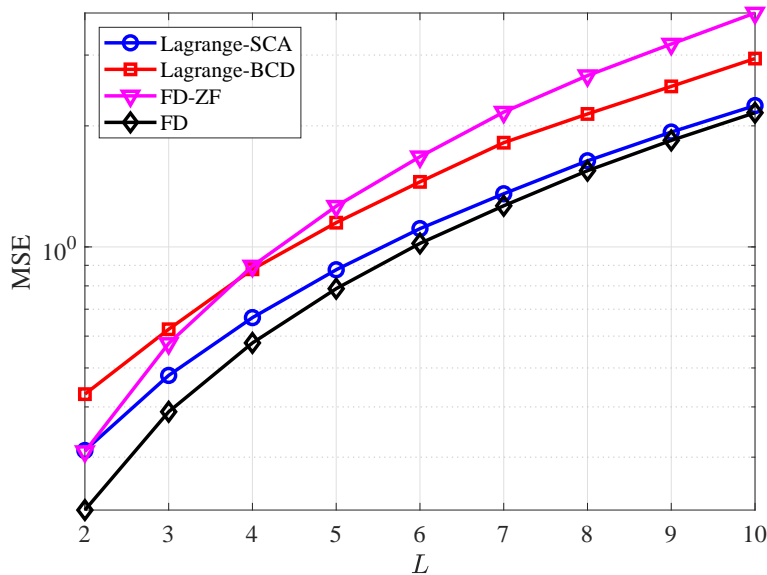


Fig. 9: The computation MSE performance versus L when $N_r = 64$, $N_{rf} = 10$, $K = 50$, and $\text{SNR} = 10$ dB.

Fig. 9 compares the MSE performance of different algorithms versus L when $N_r = 64$, $N_{\text{rf}} = 10$, $K = 50$, and $\text{SNR} = 10$ dB. From this figure, the computation MSEs of all algorithms increase with the number of functions operated at the AP, implying that the multi-function operation at the AP leads to increased computation error. Besides, we can see that the FD-ZF algorithm shows less computation MSE than the Lagrange-BCD algorithm when $L \leq 3$, since the inter-function interference is insignificant and can be handled by the FD-ZF algorithm in such a case. In particular, the fully-digital beamforming structure in the FD-ZF algorithm shows its advantages of exploiting the spatial degrees and the array gain compared with the hybrid one in the Lagrange-BCD algorithm. However, these advantages are marginal as the inter-function interference becomes more severe in the regime of large L . This further verifies the effectiveness of our proposed algorithms in multi-function/multi-modal massive MIMO AirComp systems.

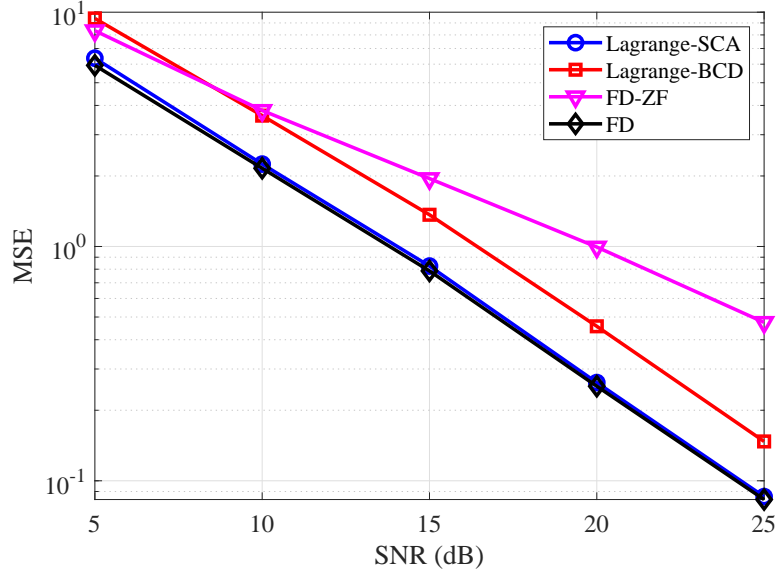


Fig. 10: The computation MSE performance versus SNR when $N_r = 64$, $N_t = 10$, $N_{\text{rf}} = 10$, $K = 50$, and $L = 10$.

Fig. 10 shows the MSE performance versus SNR under $N_r = 64$, $N_t = 10$, $N_{\text{rf}} = 10$, $K = 50$, and $L = 10$. We can see that the MSE values by all our considered schemes decreases monotonically as SNR increases. Due to the advantages of the fully-digital beamforming structure, the FD-ZF algorithm achieves a smaller computation MSE than that of the Lagrange-BCD algorithm when SNR is less than 10 dB, while the opposite holds when SNR is larger than 10 dB. However, our proposed Lagrange-SCA algorithm outperforms the FD-ZF algorithm over the whole SNR regime due to the proposed resource optimization.

VI. CONCLUSION

In this paper, we exploited massive MIMO with hybrid beamforming for AirComp systems. We jointly optimized the transmit digital beamforming at devices, the receive analog and digital beamforming at the AP to minimize the computation MSE. To solve the non-convex hybrid beamforming problem for MSE minimization, we proposed alternating-optimization-based designs, in which we first optimized the transmit digital beamforming using the Lagrange duality method, then proposed two approaches to update the receive analog beamforming via SCA and BCD, respectively, and finally optimized the receive digital beamforming by applying the first-order optimality condition. To gain more insights, we analyzed the system performance for the special case with a fully-digital receive beamforming at AP and proved that the computation MSE is inversely proportional to the number of receive antennas for this special case. Our numerical results showed that the proposed algorithms achieve an outstanding performance that is close to the performance upper bound achieved under a fully-digital receiver but with a much smaller number of RF chains required.

APPENDIX A

PROOF OF LEMMA 1

Let $\mu_k \geq 0$ denote the Lagrange multiplier associated with the power constraint for device k in problem $\mathcal{P}2$. The Lagrangian of $\mathcal{P}2$ is denoted as:

$$\mathcal{L}(\mathbf{V}_k) \triangleq \text{Tr}[(\mathbf{U}_{\text{bb}}^H \mathbf{U}_{\text{rf}}^H \mathbf{H}_k \mathbf{V}_k - \mathbf{I})(\mathbf{U}_{\text{bb}}^H \mathbf{U}_{\text{rf}}^H \mathbf{H}_k \mathbf{V}_k - \mathbf{I})^H] + \mu_k [\text{Tr}(\mathbf{V}_k \mathbf{V}_k^H) - P]. \quad (40)$$

According to the Karush-Kuhn-Tucker (KKT) conditions, the following equalities hold:

$$\nabla_{\mathbf{V}_k} \mathcal{L}(\mathbf{V}_k) = \mathbf{0}, \quad (41)$$

$$\text{Tr}(\mathbf{V}_k \mathbf{V}_k^H) - P \leq 0, \quad (42)$$

$$\mu_k \geq 0, \quad (43)$$

$$\mu_k (\text{Tr}(\mathbf{V}_k \mathbf{V}_k^H) - P) = 0. \quad (44)$$

(41) is the first-order optimality condition of $\mathcal{L}(\mathbf{V}_k)$ with respect to \mathbf{V}_k , which yields

$$\mathbf{V}_k = (\mathbf{H}_k^H \mathbf{U}_{\text{rf}} \mathbf{U}_{\text{bb}} \mathbf{U}_{\text{bb}}^H \mathbf{U}_{\text{rf}}^H \mathbf{H}_k + \mu_k \mathbf{I})^{-1} \mathbf{H}_k^H \mathbf{U}_{\text{rf}} \mathbf{U}_{\text{bb}}. \quad (45)$$

Therefore, once we obtain the optimal μ_k , denoted by μ_k^{opt} , \mathbf{V}_k can be updated by substituting μ_k^{opt} into (45). Combining (43) and (44), we optimize μ_k by considering two cases with $\mu_k = 0$ and $\mu_k > 0$.

Let $\mathbf{V}_k(\mu_k)$ denote the right-hand side of (45). For the first case, if $\mathbf{H}_k^H \mathbf{U}_{\text{rf}} \mathbf{U}_{\text{bb}} \mathbf{U}_{\text{bb}}^H \mathbf{U}_{\text{rf}}^H \mathbf{H}_k$ is invertible and

$$\text{Tr}[\mathbf{V}_k(0) \mathbf{V}_k^H(0)] < P, \quad (46)$$

then the optimal transmit beamforming is given by $\mathbf{V}_k(0)$. For the second case, the equality of constraint in problem $\mathcal{P}2$ holds according to (44).

$$\text{Tr}[\mathbf{V}_k(\mu_k) \mathbf{V}_k^H(\mu_k)] = P. \quad (47)$$

Substituting (45) into (47), we have

$$\text{Tr}[(\mathbf{H}_k^H \mathbf{U}_{\text{rf}} \mathbf{U}_{\text{bb}} \mathbf{U}_{\text{bb}}^H \mathbf{U}_{\text{rf}}^H \mathbf{H}_k + \mu_k \mathbf{I})^{-2} \mathbf{H}_k^H \mathbf{U}_{\text{rf}} \mathbf{U}_{\text{bb}} \mathbf{U}_{\text{bb}}^H \mathbf{U}_{\text{rf}}^H \mathbf{H}_k] = P. \quad (48)$$

Note that μ_k must be positive in this case and the left-hand side is a decreasing function with respect to μ_k for $\mu_k > 0$. Then we can obtain the optimal Lagrange multiplier in (48) via a bisection search. Finally, the transmit beamforming can be updated by substituting μ_k^{opt} into (45).

APPENDIX B

PROOF OF THEOREM 1

We first prove the existence of at least one limiting point before stating that any limit point of the sequence generated by the proposed algorithms is a stationary solution. In this paper, the feasible set of each variable ($\{\mathbf{V}_k\}$, \mathbf{U}_{rf} , and \mathbf{U}_{bb}) is compact, respectively. Then, problem $\mathcal{P}1$ over their Cartesian product set is bounded. Therefore, the sequence generated by Algorithm 3 is compact and bounded. Since any compact and bounded sequence must have at least one limiting point, we can claim the existence of a limiting point of our proposed algorithms.

Then, let r denote the iteration number. Clearly, given $\mathbf{U}_{\text{rf},r}$ and $\mathbf{U}_{\text{bb},r}$, the optimal solution for $\{\mathbf{V}_{k,r}\}$ can be obtained by using the Lagrange duality method and the KKT conditions, which leads to

$$\text{MSE}(\{\mathbf{V}_{k,r+1}\}, \mathbf{U}_{\text{rf},r}, \mathbf{U}_{\text{bb},r}) \leq \text{MSE}(\{\mathbf{V}_{k,r}\}, \mathbf{U}_{\text{rf},r}, \mathbf{U}_{\text{bb},r}). \quad (49)$$

Considering the discussion in Section III-B, Algorithm 1 and Algorithm 2 both establish the local convergence to the stationary solutions of problem $\mathcal{P}3$ when $\{\mathbf{V}_{k,r}\}$ and $\mathbf{U}_{\text{bb},r}$ are fixed. Then the following inequality holds:

$$\text{MSE}(\{\mathbf{V}_{k,r}\}, \mathbf{U}_{\text{rf},r+1}, \mathbf{U}_{\text{bb},r}) \leq \text{MSE}(\{\mathbf{V}_{k,r}\}, \mathbf{U}_{\text{rf},r}, \mathbf{U}_{\text{bb},r}). \quad (50)$$

Similarly, when we update \mathbf{U}_{bb} by using (31) with given $\{\mathbf{V}_{k,r}\}$ and $\mathbf{U}_{\text{rf},r}$, we have

$$\text{MSE}(\{\mathbf{V}_{k,r}\}, \mathbf{U}_{\text{rf},r}, \mathbf{U}_{\text{bb},r+1}) \leq \text{MSE}(\{\mathbf{V}_{k,r}\}, \mathbf{U}_{\text{rf},r}, \mathbf{U}_{\text{bb},r}). \quad (51)$$

Note that in each iteration of Algorithm 3, the objective function value is non-increasing and also lower bounded by zero. Hence, the convergence of Algorithm 3 follows.

APPENDIX C

PROOF OF THEOREM 2

Substituting the optimal sum-MMSE receiver in (35) into (36), we have

$$\begin{aligned} & \text{MSE}(\{\mathbf{V}_k\}, \mathbf{U}_{\text{bb}}) \\ &= \sum_{k=1}^K \text{Tr}[(\mathbf{U}_{\text{bb}}^H \mathbf{H}_k \mathbf{V}_k \mathbf{V}_k^H \mathbf{H}_k^H \mathbf{U}_{\text{bb}} - \mathbf{U}_{\text{bb}}^H \mathbf{H}_k \mathbf{V}_k - \mathbf{V}_k^H \mathbf{H}_k^H \mathbf{U}_{\text{bb}} + \mathbf{I}) + \sigma^2 \mathbf{U}_{\text{bb}}^H \mathbf{U}_{\text{bb}}] \\ &= KL + \text{Tr}[\mathbf{U}_{\text{bb}}^H \left(\sum_{k=1}^K \mathbf{H}_k \mathbf{V}_k \mathbf{V}_k^H \mathbf{H}_k^H + \sigma^2 \mathbf{I} \right) \mathbf{U}_{\text{bb}} - \sum_{k=1}^K (\mathbf{U}_{\text{bb}}^H \mathbf{H}_k \mathbf{V}_k + \mathbf{V}_k^H \mathbf{H}_k^H \mathbf{U}_{\text{bb}})] \\ &= KL - \text{Tr} \left[\left(\sum_{k=1}^K \mathbf{V}_k^H \mathbf{H}_k^H \right) \left(\sum_{k=1}^K \mathbf{H}_k \mathbf{V}_k \mathbf{V}_k^H \mathbf{H}_k^H + \sigma^2 \mathbf{I} \right)^{-1} \left(\sum_{k=1}^K \mathbf{H}_k \mathbf{V}_k \right) \right]. \end{aligned} \quad (52)$$

As we can see that the term $\left(\sum_{k=1}^K \mathbf{H}_k \mathbf{V}_k \mathbf{V}_k^H \mathbf{H}_k^H + \sigma^2 \mathbf{I} \right)^{-1}$ is intractable. To simplify the below derivations, let us define the following matrix sequence:

$$\mathbf{A}_n = \sum_{k=1}^n \mathbf{H}_k \mathbf{V}_k \mathbf{V}_k^H \mathbf{H}_k^H + \sigma^2 \mathbf{I}, \forall n \in \mathcal{K}. \quad (53)$$

Then we have $\mathbf{A}_n = \mathbf{A}_{n-1} + \mathbf{H}_n \mathbf{V}_n \mathbf{V}_n^H \mathbf{H}_n^H$ and $\mathbf{A}_K = \sum_{k=1}^K \mathbf{H}_k \mathbf{V}_k \mathbf{V}_k^H \mathbf{H}_k^H + \sigma^2 \mathbf{I}$. According to the Kailath Variant identity [41], the inverse of \mathbf{A}_n is given by

$$\mathbf{A}_n^{-1} = \mathbf{A}_{n-1}^{-1} - \mathbf{A}_{n-1}^{-1} \mathbf{H}_n \mathbf{V}_n (\mathbf{I} + \mathbf{V}_n^H \mathbf{H}_n^H \mathbf{A}_{n-1}^{-1} \mathbf{H}_n \mathbf{V}_n)^{-1} \mathbf{V}_n^H \mathbf{H}_n^H \mathbf{A}_{n-1}^{-1}. \quad (54)$$

Letting n be 1 in (53) and using (54), we have

$$\mathbf{A}_1^{-1} = \frac{1}{\sigma^2} \left(\mathbf{I} - \mathbf{H}_1 \mathbf{V}_1 (\sigma^2 \mathbf{I} + \mathbf{V}_1^H \mathbf{H}_1^H \mathbf{H}_1 \mathbf{V}_1)^{-1} \mathbf{V}_1^H \mathbf{H}_1^H \right). \quad (55)$$

Then we check the inverse of \mathbf{A}_2

$$\begin{aligned} \mathbf{A}_2^{-1} &= \mathbf{A}_1^{-1} - \mathbf{A}_1^{-1} \mathbf{H}_2 \mathbf{V}_2 (\mathbf{I} + \mathbf{V}_2^H \mathbf{H}_2^H \mathbf{A}_1^{-1} \mathbf{H}_2 \mathbf{V}_2)^{-1} \mathbf{V}_2^H \mathbf{H}_2^H \mathbf{A}_1^{-1} \\ &\approx \mathbf{A}_1^{-1} - \frac{1}{\sigma^2} \mathbf{H}_2 \mathbf{V}_2 \left(\mathbf{I} + \frac{1}{\sigma^2} \mathbf{V}_2^H \mathbf{H}_2^H \mathbf{H}_2 \mathbf{V}_2 \right)^{-1} \mathbf{V}_2^H \mathbf{H}_2^H \\ &= \frac{1}{\sigma^2} \left(\mathbf{I} - \sum_{n=1}^2 \mathbf{H}_n \mathbf{V}_n (\sigma^2 \mathbf{I} + \mathbf{V}_n^H \mathbf{H}_n^H \mathbf{H}_n \mathbf{V}_n)^{-1} \mathbf{V}_n^H \mathbf{H}_n^H \right), \end{aligned} \quad (56)$$

where the approximation is due to Assumption 1 stated in (33) and (34). Hence, from (56), we can obtain the inverse of \mathbf{A}_K as follows:

$$\mathbf{A}_K^{-1} \approx \frac{1}{\sigma^2} \left(\mathbf{I} - \sum_{n=1}^K \mathbf{H}_n \mathbf{V}_n (\sigma^2 \mathbf{I} + \mathbf{V}_n^H \mathbf{H}_n^H \mathbf{H}_n \mathbf{V}_n)^{-1} \mathbf{V}_n^H \mathbf{H}_n^H \right). \quad (57)$$

Substituting (57) into (52) yields

$$\begin{aligned} & \text{MSE}(\{\mathbf{V}_k\}, \mathbf{U}_{\text{bb}}) \\ &= KL - \frac{1}{\sigma^2} \text{Tr} \left[\left(\sum_{k=1}^K \mathbf{V}_k^H \mathbf{H}_k^H \right) \mathbf{A}_K^{-1} \left(\sum_{k=1}^K \mathbf{H}_k \mathbf{V}_k \right) \right] \\ &= KL - \frac{1}{\sigma^2} \text{Tr} \left[\sum_{k=1}^K (\mathbf{V}_k^H \mathbf{H}_k^H \mathbf{H}_k \mathbf{V}_k - \mathbf{V}_k^H \mathbf{H}_k^H \mathbf{H}_k \mathbf{V}_k (\sigma^2 \mathbf{I} + \mathbf{V}_k^H \mathbf{H}_k^H \mathbf{H}_k \mathbf{V}_k)^{-1} \mathbf{V}_k^H \mathbf{H}_k^H \mathbf{H}_k \mathbf{V}_k) \right] \\ &\approx KL - \frac{1}{\sigma^2} \sum_{k=1}^K \text{Tr} (\beta N_r \mathbf{V}_k^H \mathbf{V}_k - \beta^2 N_r^2 \mathbf{V}_k^H \mathbf{V}_k (\sigma^2 \mathbf{I} + \beta N_r \mathbf{V}_k^H \mathbf{V}_k)^{-1} \mathbf{V}_k^H \mathbf{V}_k) \\ &= KL - \sum_{k=1}^K \text{Tr} (\beta N_r \mathbf{V}_k^H \mathbf{V}_k (\sigma^2 \mathbf{I} + \beta N_r \mathbf{V}_k^H \mathbf{V}_k)^{-1}) \\ &= KL - \sum_{k=1}^K \text{Tr} \left(\frac{\beta N_r P}{L} \left(\sigma^2 + \frac{\beta N_r P}{L} \right)^{-1} \mathbf{I} \right) \\ &= KL - \frac{KL \beta N_r P}{L \sigma^2 + \beta N_r P} \\ &= \frac{KL^2 \sigma^2}{L \sigma^2 + \beta N_r P}, \end{aligned} \quad (58)$$

where the approximation and the fourth equality are due to (34) and (32), respectively. When N_r is sufficiently large, $\beta N_r P$ becomes the dominated term of denominator such that (58) is reduced to

$$\text{MSE}(\{\mathbf{V}_k\}, \mathbf{U}_{\text{bb}}) \approx \frac{KL^2 \sigma^2}{\beta N_r P}. \quad (59)$$

Obviously, when $N_r \rightarrow \infty$, (59) goes to zeros. The result follows immediately.

APPENDIX D

PROOF OF LEMMA 2

According to (57), (35) can be rewritten as

$$\begin{aligned}
\tilde{\mathbf{U}}_{\text{smmse}} &= \frac{1}{\sigma^2} \left(\mathbf{I} - \sum_{k=1}^K \mathbf{H}_k \mathbf{V}_k (\sigma^2 \mathbf{I} + \mathbf{V}_k^H \mathbf{H}_k^H \mathbf{H}_k \mathbf{V}_k)^{-1} \mathbf{V}_k^H \mathbf{H}_k^H \right) \left(\sum_{k=1}^K \mathbf{H}_k \mathbf{V}_k \right) \\
&\approx \sum_{k=1}^K \frac{1}{\sigma^2} (\mathbf{H}_k \mathbf{V}_k - \mathbf{H}_k \mathbf{V}_k (\sigma^2 \mathbf{I} + \mathbf{V}_k^H \mathbf{H}_k^H \mathbf{H}_k \mathbf{V}_k)^{-1} \mathbf{V}_k^H \mathbf{H}_k^H \mathbf{H}_k \mathbf{V}_k), \\
&= \sum_{k=1}^K \frac{1}{\sigma^2} \mathbf{H}_k \mathbf{V}_k (\mathbf{I} - (\sigma^2 \mathbf{I} + \mathbf{V}_k^H \mathbf{H}_k^H \mathbf{H}_k \mathbf{V}_k)^{-1} \mathbf{V}_k^H \mathbf{H}_k^H \mathbf{H}_k \mathbf{V}_k), \\
&= \sum_{k=1}^K \mathbf{H}_k \mathbf{V}_k (\sigma^2 \mathbf{I} + \mathbf{V}_k^H \mathbf{H}_k^H \mathbf{H}_k \mathbf{V}_k)^{-1} \\
&\approx \sum_{k=1}^K \mathbf{H}_k \mathbf{V}_k (\sigma^2 \mathbf{I} + \beta N_r \mathbf{V}_k^H \mathbf{V}_k)^{-1},
\end{aligned} \tag{60}$$

where the approximations are due to (33) and (34), respectively. For the special case where $\mathbf{V}_k^H \mathbf{V}_k = \frac{P}{L} \mathbf{I}$ and N_r is sufficiently large, (60) can be simplified as:

$$\begin{aligned}
\tilde{\mathbf{U}}_{\text{smmse}} &= \sum_{k=1}^K \frac{L}{L\sigma^2 + \beta N_r P} \mathbf{H}_k \mathbf{V}_k \\
&= \sum_{k=1}^K \frac{L}{\beta N_r P} \mathbf{H}_k \mathbf{V}_k.
\end{aligned} \tag{61}$$

The result follows immediately.

REFERENCES

- [1] M. Agiwal, A. Roy, and N. Saxena, "Next generation 5G wireless networks: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 3, pp. 1617-1655, 3rd quarter, 2016.
- [2] J. Lin, W. Yu, N. Zhang, X. Yang, H. Zhang and W. Zhao, "A survey on internet of things: Architecture, enabling technologies, security and privacy, and applications," *IEEE Internet Things J.*, vol. 4, no. 5, pp. 1125-1142, Oct. 2017.
- [3] K. Yang, T. Jiang, Y. Shi, and Z. Ding, "Federated learning via over-the-air computation," *IEEE Trans Wireless Commun.*, vol. 19, no. 3, pp. 2022-2035, Mar. 2020.
- [4] O. Abari, H. Rahul, and D. Katabi, "Over-the-air function computation in sensor networks," [online]. <http://arxiv.org/abs/1612.02307>, 2016.
- [5] G. Zhu and K. Huang, "MIMO over-the-air computation for high-mobility multi-modal sensing," *IEEE Internet Things J.*, vol. 6, no. 4, pp. 6089-6103, Aug. 2019.
- [6] M. Goldenbaum, H. Boche and S. Stańczak, "Nomographic functions: Efficient computaion in clustered Gaussian sensor networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 4, pp. 2093-2105, Apr. 2015.

- [7] M. Gastpar, "Uncoded transmission is exactly optimal for a simple Gaussian 'sensor' network," *IEEE Trans. Inf. Theory*, vol. 54, no. 11, pp. 5247-5251, Nov. 2008.
- [8] A. B. Wagner, S. Tavildar, and P. Viswanath, "Rate region of the quadratic Gaussian two-encoder source-coding problem," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 1938-1961, May 2008.
- [9] R. Soundararajan and S. Vishwanath, "Communicating linear functions of correlated Gaussian sources over a MAC," *IEEE Trans. Inf. Theory*, vol. 58, no. 3, pp. 1853-1860, Mar. 2012.
- [10] J.-J. Xiao, S. Cui, Z.-Q. Luo, and A. J. Goldsmith, "Linear coherent decentralized estimation," *IEEE Trans. Signal Process.*, vol. 56, no. 2, pp. 757-770, Feb. 2008.
- [11] C.-H. Wang, A. S. Leong, and S. Dey, "Distortion outage minimization and diversity order analysis for coherent multiaccess," *IEEE Trans. Signal Process.*, vol. 59, no. 12, pp. 6144-6159, Dec. 2011.
- [12] X. Cao, G. Zhu, J. Xu, and K. Huang, "Optimal power control for over-the-air computation in fading channels," [online]. <https://arxiv.org/pdf/1906.06858.pdf>, 2019.
- [13] O. Abari, H. Rahul, D. Katabi, and M. Pant, "Airshare: Distributed coherent transmission made seamless," in *Proc. IEEE INFOCOM*, pp. 1742-1750, Apr. 2015.
- [14] B. Nazer and M. Gastpar, "Computation over multiple-access channels," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3498-3516, Oct. 2007.
- [15] S. Zhang, S. C. Liew, and P. P. Lam, "Hot topic: Physical-layer network coding," in *Proc. Int. Conf. Mobile Comput. Netw.*, pp. 358-365, Sep. 2006.
- [16] R. Appuswamy and M. Franceschetti, "Computing linear functions by linear coding over networks," *IEEE Trans. Inf. Theory*, vol. 60, no. 1, pp. 422-431, Jan. 2014.
- [17] U. Erez, S. Litsyn, and R. Zamir, "Lattices which are good for (almost) everything," *IEEE Trans. Inf. Theory*, vol. 51, no. 10, pp. 3401-3416, Oct. 2005.
- [18] B. Nazer and M. Gastpar, "Compute-and-forward: Harnessing interference through structured codes," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6463-6486, Oct. 2011.
- [19] S. W. Jeon and C. J. Bang, "Opportunistic function computation for wireless sensor networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 6, pp. 4045-4059, Jun. 2016.
- [20] F. Wu, L. Chen, N. Zhao, Y. Chen, F. R. Yu, and G. Wei, "NOMA-enhanced computation over multi-access channels," *IEEE Trans. Wireless Commun.*, vol. 19, no. 4, pp. 2252-2267, Apr. 2020.
- [21] L. Dai, B. Wang, Y. Yuan, S. Han, I. Chin-Lin, and Z. Wang, "Non-orthogonal multiple access for 5G: solutions, challenges, opportunities, and future research trends," *IEEE Commun. Mag.*, vol. 53, no. 9, pp. 74-81, Sep. 2015.
- [22] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590-3600, Nov. 2010.
- [23] P. Sudarshan, N. B. Mehta, A. F. Molisch, and J. Zhang, "Channel statistics-based RF pre-processing with antenna selection," *IEEE Trans. Wireless Commun.*, vol. 5, no. 12, pp. 3501-3511, Dec. 2006.
- [24] S. Hur, T. Kim, D. Love, J. Krogmeier, T. Thomas, and A. Ghosh, "Millimeter wave beamforming for wireless backhaul and access in small cell networks," *IEEE Trans. Commun.*, vol. 61, no. 10, pp. 4391-4403, Nov. 2013.
- [25] A. Alkhateeb, O. Ayach, G. Leus, and R. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 831-846, Oct. 2014.

- [26] R. Méndez-Rial, C. Rusu, A. Alkhateeb, N. González-Prelcic, and R. Heath, "Hybrid MIMO architectures for millimeter wave communications: Phase shifters or switches?" *IEEE Access*, vol. 4, pp. 247-267, Jan. 2016.
- [27] 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.
- [28] X. Zhai, Y. Cai, Q. Shi, M. Zhao, G. Y. Li, and B. Champagne, "Joint transceiver design with antenna selection for large-scale MU-MIMO mmWave systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 8, pp. 2085-2096, Sep. 2017.
- [29] L. Yang and W. Zhang, "Hybrid precoding design achieving fully digital performance for millimeter wave communications," *Journal of Communication and Information Networks*, vol. 3, no. 4, pp. 74-84, Dec. 2018.
- [30] K. B. Dsouza, K. N. R. S. V. Prasad, and V. K. Bhargava, "Hybrid precoding with partially connected structure for millimeter wave massive MIMO OFDM: A parallel framework and feasibility analysis," *IEEE Trans. Wireless Commun.*, vol. 17, no. 12, pp. 8108-8122, Dec. 2018.
- [31] C. Han, J. Wang, J. Wang, and L. Bai, "Hybrid beamforming design for uplink mmWave systems with a predefined low-resolution codebook," *Journal of Communication and Information Networks*, vol. 4, no. 4, pp. 1-8, Sep. 2019.
- [32] A. Liu, X. Chen, W. Yu, V. K. N. Lau, and M. Zhao, "Two-timescale hybrid compression and forward for massive MIMO aided C-RAN," *IEEE Trans. Signal Process.*, vol. 67, no. 9, pp. 2484-2496, May 2019.
- [33] X. Song, T. Kühne, and G. Caire, "Fully-/partially-connected hybrid beamforming architectures for mmWave MU-MIMO," *IEEE Trans. Wireless Commun.*, vol. 19, no. 3, pp. 1754-1769, Mar. 2020.
- [34] X. Rao and V. K. N. Lau, "Distributed compressive CSIT estimation and feedback for FDD multi-user massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3261-3271, Jun. 2014.
- [35] M. Soltanalian, M. M. Naghsh, N. Shariati, P. Stoica and B. Hassibi, "Training signal design for correlated massive MIMO channel estimation," *IEEE Trans. Wireless Commun.*, vol. 16, no. 2, pp. 1135-1143, Feb. 2017.
- [36] A. Alkhateeb, G. Leus, and R. W. Heath, "Limited feedback hybrid precoding for multi-user millimeter wave systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 11, Nov. 2015.
- [37] M. Goldenbaum and S. Stańczak, "On the channel estimation effort for analog computation over wireless multiple-access channels," *IEEE Wireless Commun. Lett.*, vol. 3, pp. 261-264, Jun. 2014.
- [38] S. Boyd and L. Vandenberghe, *Convex Optimization*, 1st ed. Cambridge, UK: Cambridge University Press, 2004.
- [39] M. Razaviyayn, M. Hong, and Z.-Q. Luo, "A unified convergence analysis of block successive minimization methods for nonsmooth optimization", *SIAM J. Optim.*, vol. 23, no. 2, pp. 1126-1153, 2013.
- [40] 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.
- [41] K. B. Pertersen and M. S. Pedersen, "The matrix cookbook," Nov. 2012 [Online]. Available: <https://www.math.uwaterloo.ca/~hwolkowi/matrixcookbook.pdf>
- [42] D. Bertsekas, *Nonlinear Programming*, 3rd ed. Belmont, MA: Athena Scientific, 2016.
- [43] N. Krishnan, R. D. Yates, N. B. Mandayam, "Uplink linear receivers for multi-cell multiuser MIMO with pilot contamination: Large system analysis," *IEEE Trans. Wireless Commun.*, vol. 13, no. 8, pp. 4360-4373, Aug. 2014.
- [44] V. A. Toponogov, *Differential Geometry of Curves and Surfaces*. New York, NY: Birkhäuser Basel, 2006.