

Beamforming Design Based on Two-Stage Stochastic Optimization for RIS-Assisted Over-the-Air Computation Systems

Xiongfei Zhai, *Member, IEEE*, Guojun Han, *Senior Member, IEEE*, Yunlong Cai, *Senior Member, IEEE*, and Lajos Hanzo, *Fellow, IEEE*

Abstract—Over-the-air computation (AirComp) has been recognized as a promising technique of enabling the fusion center (FC) to aggregate the data gleaned from massive distributed wireless devices (WDs). Nevertheless, the computational performance of AirComp is significantly affected by the potentially poor channel conditions between the WDs and FC due to physical obstacles. For mitigating this limitation, we employ reconfigurable intelligent surfaces (RISs) for enhancing the reception quality and thus improve the computational performance of AirComp. Moreover, the previous studies of RIS-assisted AirComp tend to rely on the real-time channel state information (CSI), leading to excessive overhead since the number of RIS elements is large. To mitigate the above issue, a mixed-timescale penalty-dual-decomposition (MTPDD) algorithm is proposed, in which the transmit power of each WD, the receive beamforming vector at the FC, and the passive beamforming matrix of the RIS are jointly optimized. We aim to minimize the average computation mean-squared error (MSE) over time with reduced signaling overhead. Specifically, at each time slot, we optimize the short-term transmit power and receive beamforming vector based on the real-time low-dimensional CSI vectors. By contrast, in each frame, we update the long-term passive RIS beamforming matrix based on the channel statistics. Besides, we analyzed both the convergence and the computational complexity of the proposed algorithms. Simulation results verify the benefits of our proposed MTPDD beamforming algorithm. It is also shown that the performance of the MTPDD algorithm approaches that achieved by the scheme using real-time perfect CSI with reduced signal overhead.

Index Terms—Internet-of-Things (IoT) networks, over-the-air computation (AirComp), reconfigurable intelligent surface (RIS), mixed-timescale beamforming.

The work of X. Zhai was supported by the Science and Technology Program of Guangzhou under Grant 202102020869. The work of G. Han was supported in part by the Joint Funds of the National Natural Science Foundation of China and Guangdong under Grant U2001203, and the National Natural Science Foundation of China under Grant 61871136. The work of Y. Cai was supported in part by the National Natural Science Foundation of China under Grants 61971376 and 61831004, and in part by the Zhejiang Provincial Natural Science Foundation for Distinguished Young Scholars under Grant LR19F010002. L. Hanzo would like to acknowledge the financial support of the Engineering and Physical Sciences Research Council projects EP/P034284/1 and EP/P003990/1 (COALESCE) as well as of the European Research Council's Advanced Fellow Grant QuantCom (Grant No. 789028). (Corresponding author: Yunlong Cai)

X. Zhai and G. Han are with the School of Information Engineering, Guangdong University of Technology, Guangzhou 510006, China (e-mail: zhaixiongfei@gdut.edu.cn; gjhan@gdut.edu.cn).

Y. Cai is with the College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310027, China (e-mail: ylc@zju.edu.cn).

L. Hanzo is with the Department of Electronics and Computer Science, University of Southampton, Southampton, UK (Email: lh@ecs.soton.ac.uk).

Copyright (c) 2021 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

I. INTRODUCTION

The swift data aggregation from massive wireless devices (WDs) is required in the Internet-of-things (IoT) networks [1]–[3]. In conventional wireless communication systems, the fusion center (FC) has to extract each and every WD's signal from the received signal for data aggregation. Hence the signals gleaned from different WDs constitute harmful interference. However, the computation of a specific function of the aggregated data is required at the FC in practical IoT applications, e.g., the sum or the mean, rather than the recovery of each individual message. Hence, as the number of WDs becomes extremely large, the traditional wireless techniques are not energy-efficient in this context and may result in excessive latency. To deal with the above impediments, over-the-air computation (AirComp) has emerged as a promising solution [4]–[8], which finds a way for exploiting the co-channel interference among WDs during its computations. By exploiting the signal superposition imposed by wireless channels, the direct computation of the arithmetic mean, weighted sum, and polynomial of the parameter/data gathered from the WDs is workable at the FC. Due to the above advantages, AirComp has beneficial applications in federated learning, distributed sensing, and consensus control [9]. Based on federated learning, the edge server can obtain the average model from edge devices via wireless channels for updating the global model by exploiting AirComp. For the applications of distributed sensing, AirComp can be harnessed by unmanned aerial vehicles (UAVs) carrying the FC, which can perform ultra-fast data aggregation of a large number of sensors mounted on buildings. The computation results support the ubiquitous city-wide surveillance in smart-city applications. Furthermore, for the application of consensus control (e.g., for UAV swarm formation control and vehicular platooning), an agreement needs to be reached based on the interaction among all the vehicles. Hence, the information state of each vehicle can be updated with the average of those of others based on AirComp.

In the literature, there are two lines of research, which focus on digital coded AirComp (see, e.g., [10]–[18]) and analog uncoded AirComp (see, e.g., [4]–[8], [19]–[24]). Simple uncoded transmission was proven to achieve a sufficiently low computation distortion for independent and identically Gaussian distributed data sources [19]. As a further advance, the authors of [10] and [11] have shown that coding is necessary for enhancing the computation performance for bivariate and correlated Gaussian distributed data sources. The concept

of digital AirComp was first investigated in [12] for the computation of certain functions in wireless sensor networks and for physical-layer network coding aided two-way relaying channels in [13]. As a further development, linear coding and lattice coding were studied in [14] and [15], respectively. Based on information theory, the achievable computation rate of AirComp was investigated in [16] and [17]. To improve the computation rate of multi-function AirComp, non-orthogonal multiple access (NOMA) was studied in [18], where various functions arriving from different WDs were superimposed in each resource block.

On the other hand, the computation mean-squared error (MSE) is typically adopted as the accuracy metric for analog AirComp. The optimal power control strategies for AirComp under fading channels and coherent multiple access channels were investigated in [4], [20] and [21] by using convex optimization, respectively. The authors of [22] developed optimal power allocation for minimizing the probability that the computation MSE exceeds a given threshold. Recently, multiple-input multiple-output (MIMO) techniques have also been investigated in the context of multi-function computation in AirComp. In particular, the authors of [7] investigated MIMO-aided AirComp employed for simultaneously computing multiple functions and proposed a closed-form equalization solution at the access point (AP) based on zero-forcing (ZF) transmit beamforming at the WDs. Furthermore, a powerful massive MIMO assisted AirComp solution was designed in [6], [8], [23] for mitigating the computation accuracy erosion upon increasing the number of WDs. It is important to note however that accurate synchronization of all WDs is pivotal in AirComp. Hence a novel synchronization technique referred to as AirShare was proposed in [24], where the synchronization is implemented by broadcasting a shared clock to all WDs.

But naturally, the computational performance of AirComp may be significantly affected by the propagation conditions between the WDs and FC due to physical obstacles. As a remedy, it has been shown that the spatial degrees of freedom and the resultant array gain can be improved by the reconfigurable intelligent surfaces (RIS) at a low energy and hardware cost [25]–[29]. In particular, an RIS is able to independently change both the amplitude and phase shift of the incident signal. In this way, a favorable propagation environment can be created to enhance the system performance. Furthermore, no radio frequency (RF) chains are required by the passive RISs, which significantly reduces the fabrication cost and energy consumption as compared to that of the active antennas at the relays. Furthermore, it was shown in [30] that no modification of the existing infrastructure and operating standards is needed for the implementation of RIS in cellular systems.

Given its advantages, RISs have been studied in various contexts, such as RIS-assisted orthogonal frequency division multiplexing (OFDM) [31], [32], physical layer security [33]–[35], and wireless power transfer [36]–[40], etc. More recently, RISs have also received substantial attention in AirComp [41]–[43]. Specifically, an alternating difference-of-convex (DC) programming algorithm based on matrix lifting was proposed in [41] for the joint beamforming design problem of RIS-assisted AirComp systems. Moreover, a new scheme com-

binning RIS-assisted AirComp with energy beamforming was studied in [43], where the problems of AirComp and energy beamforming are addressed separately based on semidefinite relaxation (SDR) and DC programming. It was shown that the computation performance is significantly improved with the aid of RISs. However, the RIS-related channel state information (CSI) was routinely assumed to be known in the prior contributions, which is not realistic for the following two reasons. 1) The RIS usually consists of a very large number of reflecting elements, hence it is quite challenging to obtain accurate CSI at the RIS, and the CSI estimation error may lead to poor computation performance; 2) the optimization of the passive beamforming matrix by utilizing the instantaneous CSI may result in heavy signaling overhead.

To address the aforementioned challenges, the mixed-timescale beamforming philosophy has emerged [44]–[47]. In this scheme, the short-term variables are optimized by exploiting the instantaneous CSI. In the contrast, the long-term variables are updated based on the channel statistics. In particular, an online algorithm based on stochastic optimization was proposed in [45] for addressing the mixed-timescale hybrid beamforming problem of a massive MIMO aided cloud radio access network (C-RAN). However, we cannot directly apply most of the designs to RIS-assisted AirComp systems, since the design objectives are fundamentally different. In [44]–[46], the conventional RIS-assisted wireless communication systems focus on detecting the data of each user, in which the data from others is treated as the harmful interference. Hence, the mixed-timescale beamforming in [44]–[46] was designed for suppressing the multi-user interference. However, the RIS-assisted AirComp systems generally aim to recovery the sum of the data from all WDs by exploiting the signal superposition. There is no multi-WD interference and all WDs contribute to the computation accuracy. Therefore, the mixed-timescale beamforming for RIS-assisted AirComp systems should concurrently and beneficially exploit the signals from all WDs to assist functional computation. Based on stochastic successive convex approximation (SSCA), a mixed-timescale hybrid beamforming algorithm for massive MIMO AirComp systems was developed in [47]. It has been shown that the mixed-timescale scheme in [47] can significantly reduce the signaling overhead as compared to [8]. In contrast to [8] and [47], in this contribution we investigate a more refined RIS-assisted AirComp system, which offers a more energy-efficient solution. Specifically, we first equivalently transform the stochastic beamforming problem into a deterministic one by fully exploiting the channel statistics. Then a novel mixed-timescale beamforming algorithm is developed based on solving the resultant problem. It is verified that the proposed algorithm outperforms the design in [47] in terms of the computation MSE and signaling overhead.

Explicitly, our main contributions are summarized as follows.

- We investigate an RIS-assisted AirComp system relying on a multi-antenna FC, an RIS, and a number of WDs, which aims for computing the sum of the data gleaned from the WDs. We conceive a mixed-timescale scheme, where the transmit power at the WDs, the passive beamforming matrix at the RIS, and the receive beamforming

vector at the FC are jointly optimized for minimizing the average computation MSE with the reduced CSI signaling overhead. This challenging problem is generally intractable due to facts that the objective function is stochastic and the constraints are non-convex.

- We first transform this challenging stochastic optimization problem into an equivalent but more tractable deterministic problem based on the channel statistics. Then, we propose a mixed-timescale penalty-dual-decomposition (MTPDD) algorithm for optimizing the variables at different time scales. Specifically, we optimize the short-term variables (i.e., the transmit power at the WDs and receive beamforming vector at the FC) at each time slot by exploiting the real-time low-dimensional CSI vector. Then we update the long-term RIS passive beamforming matrix based on the channel statistics in a frame-based manner. The convergence and computational complexity of the proposed MTPDD algorithm are also analyzed.
- Our numerical results show that the proposed mixed-timescale beamforming design substantially outperforms other benchmarks inspired by the block coordinate descent (BCD) method and the SSCA method. Finally, the proposed algorithm is shown to match the performance upper bound associated with the perfect real-time RIS-related CSI.

The remainder of this paper is organized as follows. Section II describes the system model. Section III formulates the average computation MSE minimization problem under the transmit power and constant modulus constraints. Section IV presents the MTPDD approach to address the investigated problem. Section V presents the simulation results. Finally, Section VI draws the conclusion.

The acronyms in this paper are summarized in Table I.

TABLE I: The acronyms appeared in the paper.

Acronym	Full name
AirComp	Over-the-air computation
FC	Fusion center
WD	Wireless device
CSI	Channel state information
MSE	Mean-squared error
IoT	Internet-of-things
UAV	Unmanned aerial vehicle
NOMA	Non-orthogonal multiple access
MIMO	Multiple-input multiple-output
ZF	Zero-forcing
RF	Radio frequency
OFDM	Orthogonal frequency division multiplexing
DC	Difference-of-convex
SDR	Semidefinite relaxation
C-RAN	Cloud radio access network
SSCA	Stochastic successive convex approximation
BCD	Block coordinate descent
CSCG	Circularly symmetric complex Gaussian
LTE	Long term evolution
AWGN	Additive white Gaussian noise
LoS	Line-of-sight
NLoS	Non-line-of-sight
PDD	Penalty duality decomposition
KKT	Karush-Kuhn-Tucker
ULA	Uniform linear array
URA	Uniform rectangular array

Notations: Throughout this paper, bold upper-case letters and bold lower-case letters are adopted for matrices and

vectors, respectively. $\mathbf{A}(i, j)$ denotes the entry on the i^{th} row and the j^{th} column for a matrix \mathbf{A} , while \mathbf{A}^T , \mathbf{A}^H , and \mathbf{A}^* denote its transpose, Hermitian transpose, and conjugate, respectively. Furthermore, we denote \mathbf{I} as the identity matrix whose dimension will be clarified from the context, and denote $\mathbb{C}^{m \times n}$ ($\mathbb{R}^{m \times n}$) as the m -by- n dimensional complex (real) space. The notations $\mathbb{E}(\cdot)$, $\text{vec}(\cdot)$, $\|\cdot\|$, and $\|\cdot\|_{\infty}$ represent the expectation, vectorization, Frobenius norm, and infinite norm of an input variable, respectively [48]. We denote \circ and \otimes as the Hadamard product and Kronecker product between two matrices, respectively. $\mathcal{CN}(\Upsilon, \Phi)$ denotes the circularly symmetric complex Gaussian (CSCG) distribution with mean Υ and covariance matrix Φ .

II. SYSTEM MODEL

We investigate the RIS-assisted AirComp system as shown in Fig. 1, where one FC and one RIS simultaneously serve K single-antenna WDs. We assume that the FC and the RIS are equipped with N and M antenna elements, respectively¹. A smart controller is employed for adjusting the passive reflection coefficients at the RIS, which communicates with the FC via a backhaul link for enhancing the computation accuracy and exchanging information (this link is capacity limited and cannot support the transmission of massive data from all WDs). Besides, we assume that the synchronization of different WDs relies on accurate timing advance control mechanism², which is commonly considered in long term evolution (LTE) systems [9]. For convenience, the variables in this paper are summarized in Table II.

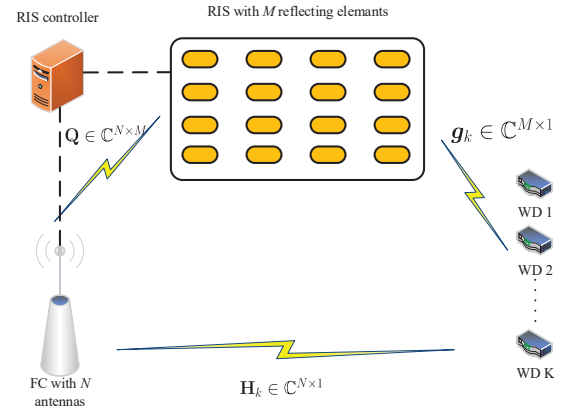


Fig. 1: RIS-assisted AirComp system.

In this AirComp system, the heterogeneous time-varying parameter of the environment (e.g., humidity, temperature,

¹As illustrated in Section I, the general AirComp system investigated in the literature (e.g., [4], [6]–[8], [21]–[23]) can be exploited in various areas. However, there exists the case that the channel condition of the direct link between the FC and the WDs is pretty poor due to long distance and obstacles. Hence, in this paper, we consider the deployment of a RIS to deal with the poor channel condition. It is shown that the RIS can improve the computation performance efficiently [41]–[43].

²Accordingly, each WD estimates the propagation delay and then transmit ahead of time, thus the FC receives the transmitted signal in the allocated time slot - regardless of the WDs' location. Since the cyclic prefix (about 5 microseconds) in LTE systems is much longer than the typical synchronization offset/error (about 0.1 microseconds) [49], the FC can correct the phase shift of the received signal by sub-channel equalization. Therefore, the time synchronization proposed for our system is feasible.

TABLE II: The variables appeared in the paper.

Variables	Meaning	Domain/Dimension
N	The number of antennas at the FC	Positive integer
M	The number of reflecting elements at the RIS	Positive integer
K	The number of WDs	Positive integer
\mathcal{K}	The set of WDs	Define as $\mathcal{K} \triangleq \{1, 2, \dots, K\}$
s_k	The collected parameter at the WD k	\mathbb{C}
s	The sum of the collected parameters	\mathbb{C}
x_k	The signal transmitted by the WD k	\mathbb{C}
v_k	The transmit coefficient at WD k	\mathbb{C}
P	The maximum transmit power of each WD	Positive number
Θ	The RIS diagonal passive beamforming matrix at the RIS	$\mathbb{C}^{M \times M}$
θ_m	The phase shift of the m th reflecting element	Positive number in the set of $[0, 2\pi)$
\mathbf{g}_k	The channel vector from WD k to the RIS	$\mathbb{C}^{M \times 1}$
\mathbf{Q}	The channel matrix from the RIS to the FC	$\mathbb{C}^{N \times M}$
\mathbf{h}_k	The channel vector from WD k to the FC	$\mathbb{C}^{N \times 1}$
\mathbf{y}	The received signal vector at the FC	$\mathbb{C}^{N \times 1}$
\mathbf{n}	The AWGN vector	$\mathbb{C}^{N \times 1}$
σ^2	The noise power	Positive number
\mathbf{u}	The receive beamforming vector at the FC	$\mathbb{C}^{N \times 1}$
\hat{s}	The processed signal after the receive beamforming	\mathbb{C}
β_{WI}	The Rician factor for the WD-RIS link	\mathbb{R}
$\tilde{\mathbf{z}}_{g,k}$	The LoS component of \mathbf{g}_k	$\mathbb{C}^{M \times 1}$
$\mathbf{z}_{g,k}$	The NLoS component of \mathbf{g}_k	$\mathbb{C}^{M \times 1}$
$\Phi_{g,k}$	The spatial correlation matrix between WD k and the RIS	$\mathbb{R}^{M \times M}$
β_{IF}	The Rician factor for the RIS-FC link	\mathbb{R}
\mathbf{F}	The LoS component of \mathbf{Q}	$\mathbb{C}^{N \times M}$
\mathbf{F}	The NLoS component of \mathbf{Q}	$\mathbb{C}^{N \times M}$
Φ_t	The RIS transmit correlation matrix	$\mathbb{R}^{N \times N}$
Φ_r	The RIS receive correlation matrix	$\mathbb{R}^{M \times M}$
β_{WF}	The Rician factor for the WD-FC link	\mathbb{R}
$\tilde{\mathbf{z}}_{h,k}$	The LoS component of \mathbf{h}_k	$\mathbb{C}^{N \times 1}$
$\mathbf{z}_{h,k}$	The NLoS component of \mathbf{h}_k	$\mathbb{C}^{N \times 1}$
$\Phi_{h,k}$	The correlation matrix between WD k and the FC	$\mathbb{R}^{N \times N}$
R	The number of frames in a long time block	Positive integer
T	The number of time slots in a frame	Positive integer
\mathcal{F}	The feasible set of Θ	Defined as $\mathcal{F} \triangleq \{\Theta = \text{diag}(e^{j\theta_1}, \dots, e^{j\theta_M}), \theta_m \in [0, 2\pi), m \in \mathcal{M}\}$
ϕ	The diagonal vector of Θ^*	$\mathbb{C}^{M \times 1}$
ϕ_m	The m th element in ϕ	\mathbb{C}
\mathbf{f}	The auxiliary variable in the PDD framework	$\mathbb{C}^{M \times 1}$
f_m	The m th element in \mathbf{f}	\mathbb{C}
ρ	The penalty parameter	Positive number
λ	The dual variable vector	$\mathbb{C}^{M \times 1}$
ϵ	The threshold to control the termination of the short-term algorithm	Positive number
ϵ_{in}	The threshold to control the termination of the inner-loop of the PDD-based method	Positive number
ϵ_{out}	The threshold to control the termination of the outer-loop of the PDD-based method	Positive number
L	The path loss	Positive number
C_0	The path loss at the referent distance	Positive number
D_0	The referent distance	Positive number
α_{WI}	The path loss exponent for the WD-RIS link	Positive number
α_{WF}	The path loss exponent for the WD-FC link	Positive number
α_{IF}	The path loss exponent for the RIS-FC link	Positive number
d_x	The distant between the reference antenna at the FC and the origin along the x -axis	Positive number
d_y	The distant between the reference element at the RIS and the origin along the y -axis	Positive number
d_r	The radius of the circle on which the WDs locate	Positive number
κ_r	The correlation coefficient for Θ_r	Positive number in the set of $[0, 1]$
κ_t	The correlation coefficient for Θ_t	Positive number in the set of $[0, 1]$
$\kappa_{g,k}$	The correlation coefficient for $\Theta_{g,k}$	Positive number in the set of $[0, 1]$

noise) is first recorded at every WD, and then the recorded data from all the WDs is simultaneously transmitted to the FC for computation. At a given time slot, $s_k \in \mathbb{C}$ denotes the recorded parameter at the WD $k \in \mathcal{K}$. Without loss of generality, we assume that the collected parameters are independent of each other and normalized, i.e., $\mathbb{E}(s_k s_k^*) = 1$, $\mathbb{E}(s_k s_j^*) = 0$, $\forall k, j \in \mathcal{K}, k \neq j$. The uniform normalization factor of each parameter for all WDs can be inverted at the FC for data recovery.

To support ultra-fast data aggregation, by exploiting the superposition encountered in multiple access channels, the FC directly computes the target nomographic function at a reduced communication overhead. Again, we assume that the FC is interested in the sum operation, i.e. estimating $s = \sum_{k=1}^K s_k$ [7], but this design is also readily extendable to other nomographic functions [50]. Nevertheless, the poor channel conditions between the WDs and FC due to physical obstacles significantly affect the computational performance of AirComp. To address the above issue, we employ an RIS for enhancing the overall performance.

Based on the considered system model, we formulate the signal transmitted by the WD k as

$$x_k = v_k s_k, \quad (1)$$

where we denote $v_k \in \mathbb{C}$ as the transmit coefficient at WD k . By denoting P as the maximum transmit power of each WD, we have $\mathbb{E}(|x_k|^2) = |v_k|^2 \leq P$, $\forall k \in \mathcal{K}$ accordingly.

We denote $\Theta \triangleq \text{diag}(e^{j\theta_1}, e^{j\theta_2}, \dots, e^{j\theta_M}) \in \mathbb{C}^{M \times M}$ as the RIS diagonal passive beamforming matrix, where $\theta_m \in [0, 2\pi)$, $m \in \mathcal{M} \triangleq \{1, 2, \dots, M\}$, represents the phase shift coefficient of the reflecting element. Let $\mathbf{g}_k \in \mathbb{C}^{M \times 1}$, $\mathbf{Q} \in \mathbb{C}^{N \times M}$, and $\mathbf{h}_k \in \mathbb{C}^{N \times 1}$ denote the channel vectors/matrices from WD k to the RIS, from the RIS to the FC, and from WD k to the FC, respectively. Therefore, the received signal vector at the FC is given by

$$\mathbf{y} = \sum_{k \in \mathcal{K}} (\mathbf{Q} \Theta \mathbf{g}_k + \mathbf{h}_k) v_k s_k + \mathbf{n}, \quad (2)$$

where we denote \mathbf{n} as the additive white Gaussian noise (AWGN) vector with $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I})$.

By denoting $\mathbf{u} \in \mathbb{C}^{N \times 1}$ as the receive beamforming vector at the FC, the processed signal after the receive beamforming is given by

$$\hat{s} = \mathbf{u}^H \mathbf{y}. \quad (3)$$

In this paper, we assume the tightly spaced antennas and the limited angular spread of the scattering environment. Hence, a general spatially correlated Rician fading channel model is considered for the WD-RIS, RIS-FC, and WD-FC links [51], which consists of both line-of-sight (LoS) and non-LoS (NLoS) components. In this channel model, the channel statistics of the WD-RIS link are assumed to be dependent on the WD-location and its second-order statistics are assumed to be identical for each WD. Such channel model is widely used in literature [41], [43], [44]. In particular, the channel vector from WD k to the RIS is given by

$$\mathbf{g}_k = \sqrt{\frac{\beta_{WI}}{1 + \beta_{WI}}} \bar{\mathbf{z}}_{g,k} + \sqrt{\frac{1}{1 + \beta_{WI}}} \Phi_{g,k}^{1/2} \mathbf{z}_{g,k}, \quad (4)$$

where we denote $\beta_{WI} \in \mathbb{R}$ and $\bar{\mathbf{z}}_{g,k} \in \mathbb{C}^{M \times 1}$ as the Rician factor and the deterministic (LoS) component, respectively. $\mathbf{z}_{g,k} \in \mathbb{C}^{M \times 1}$ is the NLoS component, which describes the Rayleigh fading coefficients. Moreover, its elements follow independently and identically CSCG distribution with zero mean and unit variance. The spatial correlation matrix between WD k and the RIS is denoted as $\Phi_{g,k} \in \mathbb{R}^{M \times M}$. Similarly, we respectively formulate the channel matrices/vectors for the RIS-FC link and the WD-FC link as

$$\mathbf{Q} = \sqrt{\frac{\beta_{IF}}{1 + \beta_{IF}}} \bar{\mathbf{F}} + \sqrt{\frac{1}{1 + \beta_{IF}}} \Phi_t^{1/2} \mathbf{F} \Phi_r^{1/2}, \quad (5)$$

$$\mathbf{h}_k = \sqrt{\frac{\beta_{WF}}{1 + \beta_{WF}}} \bar{\mathbf{z}}_{h,k} + \sqrt{\frac{1}{1 + \beta_{WF}}} \Phi_{h,k}^{1/2} \mathbf{z}_{h,k}, k \in \mathcal{K}, \quad (6)$$

where $\bar{\mathbf{F}} \in \mathbb{C}^{N \times M}$ and $\bar{\mathbf{z}}_{h,k} \in \mathbb{C}^{N \times 1}$ represent the LoS components, and $\mathbf{F} \in \mathbb{C}^{N \times M}$ and $\mathbf{z}_{h,k} \in \mathbb{C}^{N \times 1}$ are the NLoS components. $\Phi_{h,k} \in \mathbb{R}^{N \times N}$ and $\Phi_r \in \mathbb{R}^{M \times M}$ ($\Phi_t \in \mathbb{R}^{N \times N}$) denote the correlation matrix between WD k and the FC and the RIS receive (transmit) correlation matrix, respectively. It is assumed that the correlation matrices (i.e., $\Phi_{g,k}$, Φ_r , Φ_t , and $\Phi_{h,k}$) and the LoS components (i.e., $\bar{\mathbf{z}}_{g,k}$, $\bar{\mathbf{F}}$ and $\bar{\mathbf{z}}_{h,k}$) vary quite slowly, which are viewed as the channel statistics. On the other hand, the real-time CSI (the NLoS components) varies rapidly. For the ease of presentation, we involve the terms of Rician factors into $\{\bar{\mathbf{z}}_{g,k}, \mathbf{z}_{g,k}, \bar{\mathbf{F}}, \mathbf{F}, \bar{\mathbf{z}}_{h,k}, \mathbf{z}_{h,k}\}$. Hence the channel matrices/vectors can be equivalently reduced to $\mathbf{g}_k = \bar{\mathbf{z}}_{g,k} + \Phi_{g,k}^{1/2} \mathbf{z}_{g,k}$, $\mathbf{Q} = \bar{\mathbf{F}} + \Phi_t^{1/2} \mathbf{F} \Phi_r^{1/2}$, and $\mathbf{h}_k = \bar{\mathbf{z}}_{h,k} + \Phi_{h,k}^{1/2} \mathbf{z}_{h,k}$.

III. PROBLEM FORMULATION

In this section, we first introduce the frame structure used and then mathematically formulate the problem investigated.

A. Frame Structure

In practice, the estimation and exchange of real-time CSI lead to huge overhead. Thus it is unrealistic to jointly optimize the required variables for each channel realization. Therefore, to circumvent this problem, we conceive a mixed-timescale algorithm. As shown in Fig. 2, we consider a sufficiently long time block, where we assume that the channel statistics of all links (e.g., the LoS components) are constant. This block consists of R frames, each of which is further divided into T time slots. Note that the obtained CSI varies at different time slots due to the small scale fading. Besides, we assume that the CSI remains time-invariant within each time slot. In our algorithm, to achieve the tradeoff between the performance and signaling overhead, we only make use of the channel statistics to optimize the passive RIS beamforming matrix at the end of each frame. By contrast, we exploit the low-dimensional effective real-time CSI for determining the transmit power and receive beamforming vector within each time slot. This scheme would have a low signaling overhead of CSI acquisition, as compared with the assumption of perfect real-time full CSI of all links [41]–[43].

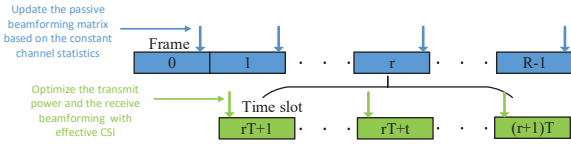


Fig. 2: The transmission scheme for the mixed-timescale design algorithm.

B. Problem Formulation

We measure the computation accuracy by the MSE between \hat{s} and s with given $\{\mathbf{h}_k\}$, $\{\mathbf{g}_k\}$, and \mathbf{Q} , which can be expressed as

$$\begin{aligned}
& g(\{v_k\}, \Theta, \mathbf{u}; \{\mathbf{h}_k\}, \{\mathbf{g}_k\}, \mathbf{Q}) \\
& \triangleq \mathbb{E}(|s - \hat{s}|^2) \\
& = \mathbb{E}(|\sum_{k \in \mathcal{K}} s_k - \mathbf{u}^H \sum_{k \in \mathcal{K}} (\mathbf{Q}\Theta\mathbf{g}_k + \mathbf{h}_k)v_k s_k - \mathbf{u}^H \mathbf{n}|^2) \\
& = \mathbb{E}(|\sum_{k \in \mathcal{K}} (1 - \mathbf{u}^H (\mathbf{Q}\Theta\mathbf{g}_k + \mathbf{h}_k)v_k) s_k - \mathbf{u}^H \mathbf{n}|^2) \\
& = \sum_{k \in \mathcal{K}} |1 - \mathbf{u}^H (\mathbf{Q}\Theta\mathbf{g}_k + \mathbf{h}_k)v_k|^2 \mathbb{E}(s_k s_k^*) \\
& \quad + \mathbf{u}^H \mathbb{E}(\mathbf{n}\mathbf{n}^H) \mathbf{u} \\
& = \sum_{k \in \mathcal{K}} |\mathbf{u}^H (\mathbf{Q}\Theta\mathbf{g}_k + \mathbf{h}_k)v_k - 1|^2 + \sigma^2 |\mathbf{u}|^2, \quad (7)
\end{aligned}$$

where the second last equality holds due to fact that the collected parameters are independent of the AWGN noise among the WDs. The last equality holds due to $\mathbb{E}(s_k s_k^*) = 1$ and $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I})$. Then the average computation MSE is given by

$$\bar{g}(\{v_k\}, \Theta, \mathbf{u}) = \mathbb{E}[g(\{v_k\}, \Theta, \mathbf{u}; \{\mathbf{h}_k\}, \{\mathbf{g}_k\}, \mathbf{Q})], \quad (8)$$

where the expectation is taken over $\{\mathbf{h}_k\}$, $\{\mathbf{g}_k\}$, and \mathbf{Q} .

In order to achieve accurate computation, we aim for minimizing the average computation MSE over the channels. Thus, the stochastic problem formulated may be written as

$$\begin{aligned}
\mathcal{P}1: & \text{minimize}_{v_k, \mathbf{u}, \Theta} \bar{g}(\{v_k\}, \Theta, \mathbf{u}) \\
& \text{subject to } |v_k|^2 \leq P, \forall k \in \mathcal{K}, \quad (9) \\
& \Theta \in \mathcal{F},
\end{aligned}$$

where we denote the feasible set of Θ as $\mathcal{F} \triangleq \{\Theta = \text{diag}(e^{j\theta_1}, \dots, e^{j\theta_M}), \theta_m \in [0, 2\pi), m \in \mathcal{M}\}$. Furthermore, the first constraint denotes the transmit power constraint of WD k .

Problem $\mathcal{P}1$ is quite challenging to solve, since the long-term variable (i.e., Θ) and short-term variables (i.e., $\{v_k\}$ and \mathbf{u}) are highly coupled in the objective function. Moreover, the average computation MSE is neither a convex function nor a concave function. Even worse, it also involves expectation over $\{\mathbf{h}_k\}$, $\{\mathbf{g}_k\}$, and \mathbf{Q} . To the best of our knowledge, there is a paucity of efficient algorithms for handling this non-convex stochastic problem.

IV. PROPOSED MTPDD ALGORITHM

In this section, a MTPDD algorithm is proposed for dealing with $\mathcal{P}1$. In particular, we optimize the receive beamforming

vector \mathbf{u} at the FC and the transmit coefficients $\{v_k\}$ at the WDs in each time slot, while the passive beamforming matrix Θ at the RIS is updated in a frame-based manner. More specifically, we optimize $\{v_k\}$ and \mathbf{u} with the aid of the Lagrange duality method and the first order optimality condition. Then Θ is updated by exploiting the penalty duality decomposition (PDD) based algorithm.

A. Short-term Optimization Over Time Slot

In the following, in order to minimize the computation MSE at each time slot, the transmit coefficients $\{v_k\}$ at the WDs and the receive beamforming \mathbf{u} at the FC are optimized. With given Θ , $\{\mathbf{h}_k\}$, $\{\mathbf{g}_k\}$, and \mathbf{Q} , problem $\mathcal{P}1$ is reformulated as a deterministic problem as follows:

$$\begin{aligned}
\mathcal{P}2: & \text{minimize}_{v_k, \mathbf{u}} g(\{v_k\}, \Theta, \mathbf{u}; \{\mathbf{h}_k\}, \{\mathbf{g}_k\}, \mathbf{Q}) \\
& \text{subject to } |v_k|^2 \leq P, \forall k \in \mathcal{K}. \quad (10)
\end{aligned}$$

Based on the objective function of (10), we can address problem $\mathcal{P}2$ in an alternating manner. Firstly, given \mathbf{u} , we can equivalently decompose problem $\mathcal{P}2$ into the following K independent subproblems:

$$\begin{aligned}
\mathcal{P}3: & \text{minimize}_{v_k} |\mathbf{u}^H (\mathbf{Q}\Theta\mathbf{g}_k + \mathbf{h}_k)v_k - 1|^2 \\
& \text{subject to } |v_k|^2 \leq P. \quad (11)
\end{aligned}$$

It is obvious that problem $\mathcal{P}3$ is a convex quadratic optimization problem. To obtain more insights, we propose the following lemma to obtain a closed-form optimal solution by employing the Lagrange duality method [52].

Lemma 1: By denoting $\mu_k^* \geq 0$ as the optimal Lagrange multiplier associated with the transmit power constraint in problem $\mathcal{P}3$, the optimal solution for problem $\mathcal{P}3$ is given by:

$$v_k^*(\mu_k^*) = \frac{\tilde{\mathbf{h}}_k^H \mathbf{u}}{\mathbf{u}^H \tilde{\mathbf{h}}_k \tilde{\mathbf{h}}_k^H \mathbf{u} + \mu_k^*}, \quad (12)$$

where $\tilde{\mathbf{h}}_k = \mathbf{Q}\Theta\mathbf{g}_k + \mathbf{h}_k$. If $\mathbf{u}^H \tilde{\mathbf{h}}_k \tilde{\mathbf{h}}_k^H \mathbf{u} \neq 0$ and the following inequality holds

$$|v_k^*(0)|^2 < P, \quad (13)$$

then the optimal Lagrange multiplier must be zeros; otherwise, we choose μ_k^* to satisfy the following equation

$$|v_k^*(\mu_k^*)|^2 = P. \quad (14)$$

We can prove Lemma 1 by exploiting the Karush-Kuhn-Tucker (KKT) conditions [52]. To save space, we omit the details of the proof. According to Lemma 1, it can be observed that we can solve problem $\mathcal{P}3$ by considering two different cases: If (13) is met, we set $\mu_k^* = 0$ and update v_k according to (12); otherwise, we then determine μ_k^* based on (14) using the bisection method and update v_k according to (12).

Secondly, with the transmit coefficients $\{v_k\}$ given, we optimize the receive beamforming \mathbf{u} at the FC for minimizing the computation MSE. Hence, we can equivalently rewrite the corresponding subproblem as

$$\mathcal{P}4: \text{minimize}_{\mathbf{u}} g(\{v_k\}, \Theta, \mathbf{u}; \{\mathbf{H}_k\}, \{\mathbf{G}_k\}, \mathbf{Q}). \quad (15)$$

Algorithm 1 The Short-term Alternating-optimization Algorithm for Address Problem $\mathcal{P}2$

Input $\epsilon > 0$, Θ , $\{\mathbf{h}_k\}$, $\{\mathbf{g}_k\}$, and \mathbf{Q}

Repeat

Step 1: Optimize v_k by exploiting Lemma 1.

Step 2: Optimize \mathbf{u} according to (16).

Until the decrease of the objective function in problem $\mathcal{P}2$ is less than ϵ .

Since problem $\mathcal{P}4$ is an unconstrained convex problem, by exploiting the first-order optimality condition [52], the closed-form solution for \mathbf{u} is given by

$$\mathbf{u}^* = \left(\sum_{k \in \mathcal{K}} |v_k|^2 \tilde{\mathbf{h}}_k \tilde{\mathbf{h}}_k^H + \sigma^2 \mathbf{I} \right)^{-1} \left(\sum_{k \in \mathcal{K}} \tilde{\mathbf{h}}_k v_k \right). \quad (16)$$

From (16), one can see that there exists a sum-MMSE structure in the expression of \mathbf{u} . The above sum-MMSE receiver is quite different from the conventional MMSE receiver in multi-user communication systems. In particular, the conventional MMSE receiver is in the form of $\left(\sum_{k \in \mathcal{K}} |v_k|^2 \tilde{\mathbf{h}}_k \tilde{\mathbf{h}}_k^H + \sigma^2 \mathbf{I} \right)^{-1} \tilde{\mathbf{h}}_k v_k$ aiming for estimating the dedicated message s_k of WD k . For computing the sum of s_k , $k \in \mathcal{K}$ in AirComp, the term outside the matrix inversion in (16) is $\sum_{k \in \mathcal{K}} \tilde{\mathbf{h}}_k v_k$. However, in the conventional MMSE receiver, this term becomes $\tilde{\mathbf{h}}_k v_k$ aiming for individually estimating s_k 's of each WD. This sum-MMSE structure is better to concurrently and beneficially exploit the signals of all the WDs for assisting function computation in RIS-aided AirComp systems. Therefore, with the aid of the sum-MMSE structure, there are only the terms of signal misalignment and noise in the expression of the computation MSE in (7). Nevertheless, the signals transmitted by different WDs are viewed as harmful inter-WD interference in the conventional multi-user communication systems. In other words, the receive beamforming relying on the sum-MMSE structure is capable of exploiting the co-channel interference among WDs as a beneficial factor in order to better deal with the issue of signal misalignment.

We summarize the short-term alternating-optimization procedure in Algorithm 1. Observe from (12) and (16) that v_k and \mathbf{u} are both updated based on $\tilde{\mathbf{h}}_k$, which is viewed as the effective low-dimensional real-time CSI between the WD and the FC. As discussed in the Section II, the estimation of $\tilde{\mathbf{h}}_k$ relying on a fixed passive beamforming matrix Θ is much easier to obtain in practice than that of \mathbf{Q} and $\{\mathbf{g}_k\}$. In other words, the update of $\{v_k\}$ and \mathbf{u} based on the effective real-time CSI is much more efficient than the methods used in [41]–[43], where the high-dimensional RIS-related CSI matrices \mathbf{Q} and $\{\mathbf{g}_k\}$ have to be acquired at each time slot.

B. Long-term Optimization Over Frames

Given the transmit coefficients $\{v_k\}$ at the WDs and the receive beamforming vector \mathbf{u} , we can reformulate problem $\mathcal{P}1$ with respect to Θ as follows:

$$\mathcal{P}5: \underset{\Theta \in \mathcal{F}}{\text{minimize}} \bar{g}(\{v_k\}, \Theta, \mathbf{u}), \quad (17)$$

which aims for minimizing the average computation MSE over channels by optimizing the passive beamforming matrix Θ at

the RIS. As it can be observed in problem $\mathcal{P}5$, the objective function (shown in (8)) is stochastic and the feasible set is non-convex. To solve problem $\mathcal{P}5$, by exploiting the following theorem, we equivalently reformulate $\mathcal{P}5$ into a more tractable form by exploiting the channel statistics in all RIS-related links.

Theorem 1: The objective function of problem $\mathcal{P}5$ can be equivalently rewritten as $\phi^H \mathbf{A} \phi + \phi^H \mathbf{b} + \mathbf{b}^H \phi + c$, where

$$\phi \triangleq [\phi_1, \dots, \phi_M] = \text{diag}(\Theta^*), \quad (18)$$

$$\begin{aligned} \mathbf{A} \triangleq & \sum_{k \in \mathcal{K}} |v_k|^2 \left(\text{diag}(\mathbf{u}^H \bar{\mathbf{F}}) \bar{\mathbf{z}}_{g,k} \bar{\mathbf{z}}_{g,k}^H \text{diag}(\bar{\mathbf{F}}^H \mathbf{u}) \right. \\ & + \frac{\text{diag}(\mathbf{u}^H \bar{\mathbf{F}}) \Phi_{g,k} \text{diag}(\bar{\mathbf{F}}^H \mathbf{u})}{1 + \beta_{WI}} \\ & \left. + \frac{\sum_{i \in \mathcal{M}} \lambda_i}{1 + \beta_{IF}} \left(\Phi_r \circ (\bar{\mathbf{z}}_{g,k} \bar{\mathbf{z}}_{g,k}^H) + \frac{\Phi_r \circ \Phi_{g,k}}{1 + \beta_{WI}} \right) \right), \end{aligned} \quad (19)$$

$$\begin{aligned} \mathbf{b} \triangleq & \sum_{k \in \mathcal{K}} \left(|v_k|^2 \text{diag}(\mathbf{u}^H \bar{\mathbf{F}}) \bar{\mathbf{z}}_{g,k} \bar{\mathbf{z}}_{h,k}^H \mathbf{u} \right. \\ & \left. - \text{diag}(\mathbf{u}^H \bar{\mathbf{F}}) \bar{\mathbf{z}}_{g,k} v_k \right), \end{aligned} \quad (20)$$

$$\begin{aligned} c \triangleq & \sum_{k \in \mathcal{K}} \left(|v_k|^2 \mathbf{u}^H \left(\bar{\mathbf{z}}_{h,k} \bar{\mathbf{z}}_{h,k}^H + \frac{\Phi_{h,k}}{1 + \beta_{WF}} \right) \mathbf{u} \right. \\ & \left. - 2\Re(\mathbf{u}^H \bar{\mathbf{z}}_{h,k} v_k) + 1 \right) + \sigma^2 |\mathbf{u}|^2, \end{aligned} \quad (21)$$

λ_i denotes the i th eigenvalue of $\Phi_r^{1/2H} \mathbf{u} \mathbf{u}^H \Phi_r^{1/2}$.

Proof: See Appendix A. ■

Inspired by [44], we first expand the objective function of $\mathcal{P}5$ by considering the independence of \mathbf{F} , $\mathbf{z}_{g,k}$, and $\mathbf{z}_{h,k}$ in Appendix A. Then we remove the expectation operator by joint exploiting the matrix identities in [48] (e.g., $\mathbf{q}^H \mathbf{D} = \text{diag}^T(\mathbf{D}) \text{diag}(\mathbf{q}^H)$, $\mathbf{D} \mathbf{q} = \text{diag}(\mathbf{q}) \text{diag}^*(\mathbf{D})$, and $\text{diag}(\mathbf{q}^H) \mathbf{p} \mathbf{p}^H \text{diag}(\mathbf{q}) = (\mathbf{q} \mathbf{q}^H) \circ (\mathbf{p} \mathbf{p}^H)$ for any vectors \mathbf{q} , \mathbf{p} and diagonal matrix \mathbf{D}) and eigendecomposition. In this way, we can obtain the result of Theorem 1 and transform the stochastic objective function of problem $\mathcal{P}5$ to a deterministic one, which is more tractable.

According to Theorem 1, problem $\mathcal{P}5$ can be equivalently reformulated as the following deterministic problem:

$$\begin{aligned} \mathcal{P}6: \underset{\phi}{\text{minimize}} \quad & \phi^H \mathbf{A} \phi + \phi^H \mathbf{b} + \mathbf{b}^H \phi \\ \text{subject to} \quad & |\phi_m| = 1, \forall m \in \mathcal{M}. \end{aligned} \quad (22)$$

As we can see, problem $\mathcal{P}6$ is more tractable than problem $\mathcal{P}5$ but it is still non-convex, which is NP-hard in general. The authors of [25] and [26] proposed two methods to address a similar problem based on the techniques of SDR and BCD. However, there are several drawbacks in these methods, namely the high complexity of the SDR method and the one-by-one iterative optimization of the elements of the BCD method. To solve $\mathcal{P}6$ with reduced computational complexity, we propose a PDD-based method by updating the elements of ϕ in parallel. Specifically, the proposed algorithm consists of two loops. We iteratively optimize the primal variables in the inner loop, while updating the penalty parameter and dual variables in the outer loop.

To facilitate the PDD optimization framework [53], we introduce an auxiliary variable $\mathbf{f} = [f_1, \dots, f_M]^T \in \mathbb{C}^{M \times 1}$

and set $\phi = \mathbf{f}$. Then problem $\mathcal{P}6$ is equivalently reformulated as

$$\begin{aligned} \mathcal{P}7: \text{ minimize}_{\phi, \mathbf{f}} \quad & \phi^H \mathbf{A} \phi + \phi^H \mathbf{b} + \mathbf{b}^H \phi \\ \text{subject to} \quad & \phi = \mathbf{f}, \\ & |f_m| = 1, \forall m \in \mathcal{M}. \end{aligned} \quad (23)$$

In the PDD framework, we optimize ϕ by solving the augmented Lagrangian problem of $\mathcal{P}7$ [52], which is given by:

$$\begin{aligned} \mathcal{P}8: \text{ minimize}_{\phi, \mathbf{f}} \quad & \phi^H \mathbf{A} \phi + \phi^H \mathbf{b} + \mathbf{b}^H \phi \\ & + \frac{1}{2\rho} \|\phi - \mathbf{f} + \rho \boldsymbol{\lambda}\|^2 \\ \text{subject to} \quad & |f_m| = 1, \forall m \in \mathcal{M}, \\ & \|\phi\|^2 \leq M, \end{aligned} \quad (24)$$

where we denote ρ and $\boldsymbol{\lambda} = [\lambda_1, \dots, \lambda_M]^T$ as the penalty parameter and the dual variable vector associated with the constraint $\phi = \mathbf{f}$, respectively. Without loss of optimality, we add the constraint $\|\phi\|^2 \leq M$ due to $\phi_m \leq 1$. Furthermore, this new constraint is necessary to guarantee the efficient search space of the proposed algorithm, which can accelerate the convergence. In the following, we detail the optimization in the inner and outer loops.

1) *Optimization in the Inner Loop:* In the inner loop, we optimize ϕ and \mathbf{f} in an alternating manner. Given \mathbf{f} , the subproblem with respect to ϕ is expressed as

$$\begin{aligned} \mathcal{P}9: \text{ minimize}_{\phi} \quad & \phi^H \mathbf{A} \phi + \phi^H \mathbf{b} + \mathbf{b}^H \phi \\ & + \frac{1}{2\rho} \|\phi - \mathbf{f} + \rho \boldsymbol{\lambda}\|^2 \\ \text{subject to} \quad & \|\phi\|^2 \leq M. \end{aligned} \quad (25)$$

It is readily seen that $\mathcal{P}9$ is convex. Thus we can directly solve this problem based on the first-order optimality condition [52] if the obtained solution satisfies the constraint. Otherwise, a factor is needed to scale the solution. Hence, we have the following the optimal solution of problem $\mathcal{P}9$:

$$\begin{aligned} \phi^* &= \begin{cases} \Phi_a^{-1} \phi_b, & \text{if } \|\Phi_a^{-1} \phi_b\|^2 \leq M, \\ \frac{\Phi_a^{-1} \phi_b}{\phi_c}, & \text{otherwise.} \end{cases} \\ \Phi_a &= \mathbf{A} + \frac{1}{2\rho} \mathbf{I}, \\ \phi_b &= \frac{1}{2\rho} \mathbf{f} - \frac{\boldsymbol{\lambda}}{2} - \mathbf{b}, \\ \phi_c &= \|\Phi_a^{-1} \phi_b\|^2 / M. \end{aligned} \quad (26)$$

Next, we optimize \mathbf{f} by fixing ϕ . The subproblem is given by (ignoring the constant terms)

$$\begin{aligned} \mathcal{P}10: \text{ minimize}_{\mathbf{f}} \quad & \|\phi - \mathbf{f} + \rho \boldsymbol{\lambda}\|^2 \\ \text{subject to} \quad & |f_m| = 1, \forall m \in \mathcal{M}. \end{aligned} \quad (27)$$

Since $\{f_m\}$ are separated in both the constraints and the objective function of problem $\mathcal{P}10$, the optimal solution can be easily obtained [52], which is given by

$$f_m^* = \frac{\phi_m + \rho \lambda_m}{|\phi_m + \rho \lambda_m|}, \forall m \in \mathcal{M}. \quad (28)$$

Algorithm 2 The PDD-Based Method for Solving Problem $\mathcal{P}6$

Initialize $\epsilon_{\text{in}} > 0$, $\epsilon_{\text{out}} > 0$, ϕ , \mathbf{f} , ρ , and $\boldsymbol{\lambda}$ such that they meet all the constraints;

Repeat

Inner loop:

Optimize ϕ and \mathbf{f} in an alternating manner according to (26) and (28), until the decrease of the objective function in problem $\mathcal{P}9$ is less than ϵ_{in} .

Outer loop:

Update $\boldsymbol{\lambda}$ according to (29) and let $\rho = \eta \rho$.

Until the constraint violation $\|\phi - \mathbf{f}\|_{\infty}$ is below ϵ_{out} .

Algorithm 3 The MTPDD Algorithm for Solving Problem $\mathcal{P}1$

Initialize v_k , \mathbf{u} , and ϕ , such that they meet all the constraints;

Short-term optimization:

In each time slot, update $\{v_k\}$ and \mathbf{u} according to Algorithm 1.

Long-term optimization:

At the end of each frame, update ϕ by solving problem $\mathcal{P}6$ according to Algorithm 2.

2) *Updation in the Outer Loop:* In the outer loop, the dual variable $\boldsymbol{\lambda}$ is updated by

$$\boldsymbol{\lambda} = \boldsymbol{\lambda} + \frac{1}{\rho} (\phi - \mathbf{f}), \quad (29)$$

which is viewed as a dual ascend step [53]. Furthermore, we decrease the penalty parameter ρ in the outer loop as $\rho = \eta \rho$, where $0 < \eta < 1$ is a constant. We summarize the proposed PDD-based method in Algorithm 2.

C. Complete Algorithm

With the aid of the above derivation, the proposed overall MTPDD procedure is summarized in Algorithm 3. Then, let us analyze the complexity of Algorithm 1. According to Lemma 1, the complexity of updating $\{v_k\}$ is $\mathcal{O}(K(M^2N + MN + N))$. Similarly, the complexity of optimizing \mathbf{u} is given by $\mathcal{O}(K(M^2N + MN + N) + N^3)$. The complexity of the PDD-based method is dominated by solving problem $\mathcal{P}9$, where the complexities of updating ϕ and \mathbf{f} are $\mathcal{O}(M^3)$ and $\mathcal{O}(M)$, respectively. Therefore, the overall complexity of Algorithm 2 is $\mathcal{O}(I_i I_o (M^3 + M))$, where I_i and I_o denote the maximum iteration numbers for the inner and outer loop, respectively. As illustrated in [25], the complexity of the SDR method in the worst case is $\mathcal{O}((M+1)^{6.5})$. It can be clearly observed that our proposed PDD-based method significantly reduces the complexity. Therefore, we can obtain the overall complexity. Furthermore, we summarize the convergence of the proposed MTPDD algorithm in the following theorem.

Theorem 2: Any limit point of the sequence generated by the MTPDD procedure of Algorithm 3 is guaranteed to converge to a stationary solution of problem $\mathcal{P}1$.

Proof: See Appendix B. ■

V. SIMULATION RESULTS

In this section, the performance of the MTPDD beamforming design proposed for RIS-assisted AirComp systems is

evaluated and compared with that of the following benchmark schemes.

- **MTSDR**: In this MTSDR scheme, we consider a mixed-timescale beamformer based on the SDR technique of [25]. Specifically, we optimize the passive beamforming matrix at the RIS by exploiting the SDR method in the long term, while updating the receive beamforming vector at the FC and the transmit coefficients at the WDs in the short term according to Algorithm 1.
- **MTBCD**: In this MTBCD scheme, we consider a mixed-timescale beamformer based on the BCD technique. Specifically, we optimize the passive beamforming matrix at the RIS by exploiting the BCD method [26] in the long term, while updating the receive beamforming vector at the FC and the transmit coefficients at the WDs in the short term according to Algorithm 1.
- **SSCA**: In this SSCA scheme, we consider a mixed-timescale beamformer by exploiting the method of stochastic successive convex approximation [45], [47].
- **PCSI**: In this PCSI scheme, we consider a single-timescale beamformer by assuming that the FC, the RIS and the WDs can perfectly estimate the complete CSI in each time slot. The transmit coefficients are updated according to Lemma 1. Then we design the receive beamformer by (16). Finally, the passive beamforming matrix is optimized by minimizing the instantaneous computation MSE via the SDR method.

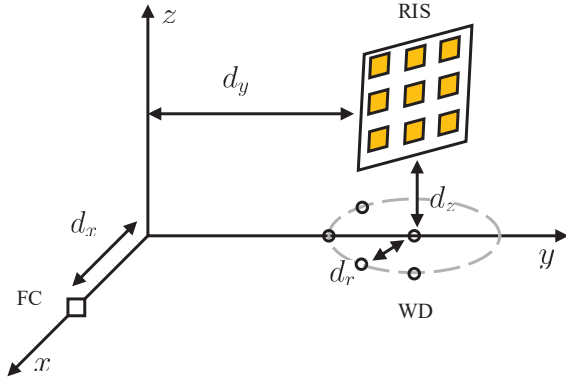


Fig. 3: The location setup in the simulations.

In the following simulations, we define the distance-dependent path loss as $L = C_0 \left(\frac{d}{D_0}\right)^{-\alpha}$, where we denote C_0 , d , and α as the path loss at the reference distance $D_0 = 1$ meter (m), the individual distance, and the path loss exponent (i.e., α_{WI} , α_{WF} , and α_{IF} denote the path loss exponents for the WD-RIS, WD-FC and RIS-FC links, respectively). Since the RIS is deployed for the WDs that suffer from severe channel condition in the WD-FC link, the path loss exponent of the WD-FC link is assumed to be larger than those of other links by setting $\alpha_{WF} = 3.4$, $\alpha_{WI} = 2.2$, and $\alpha_{IF} = 3$. In the simulation setup, we consider the three-dimensional system of Fig. 3, where we deploy the FC and the RIS on the x -axis and y - z plane, respectively. We assume that the FC is equipped with a uniform linear array (ULA). A uniform rectangular array (URA) with $M = M_y M_z$ reflecting elements

are deployed at the RIS, where we denote M_y and M_z as the number of elements along the y -axis and z -axis, respectively. Without loss of generality, we set $M_y = 4$. We deploy the reference antenna/element at the FC/RIS at $(d_x = 2 \text{ m}, 0, 0)$ and $(0, d_y = 50 \text{ m}, d_z = 3 \text{ m})$. Besides, the WDs uniformly locate on a circle with radius $d_r = 3 \text{ m}$ whose center is at $(0, d_y = 50 \text{ m}, 0)$. Furthermore, we consider the exponential correlation model in [54]–[56] for $\Phi_{h,k}$, which is given by

$$\Phi_{h,k}(i, j) = \begin{cases} \kappa_{h,k}^{j-i}, & \text{if } i \leq j, \\ \Phi_{h,k}(j, i), & \text{otherwise,} \end{cases} \quad (30)$$

where $0 \leq \kappa_{h,k} \leq 1, \forall k \in \mathcal{K}$ is the correlation coefficient. Similarly, we can also define Φ_t according to (30) with the correlation coefficient κ_t , while we model Φ_r and $\Phi_{h,k}$ as $\Phi_r = \Phi_r^h \otimes \Phi_r^v$ and $\Phi_{g,k} = \Phi_{g,k}^h \otimes \Phi_{g,k}^v$. Φ_r^h ($\Phi_{g,k}^h$) and Φ_r^v ($\Phi_{g,k}^v$) denotes the spatial correlation matrices of the horizontal and vertical domains, respectively, which are also defined as according to (30) with κ_r and $\kappa_{g,k}$ being the correlation coefficients. We define the signal-to-noise-ratio (SNR) as $\text{SNR} = \frac{P}{\sigma^2}$. Furthermore, the simulation parameter settings in Table II are considered unless otherwise specified.

TABLE III: The settings of the simulation parameters.

Parameter	Value	Parameter	Value
M	32	d_y	50 m
M_y	4	d_z	3 m
M_z	M/M_y	d_r	3 m
N	32	β_{WI}	3 dB
K	64	β_{IF}	3 dB
R	100	β_{WF}	-3 dB
T	10	κ_r	0.5
σ^2	-80 dBm	κ_t	0.5
C_0	-30 dB	$\kappa_{g,k}$	$\frac{k-1}{K}$
D_0	1 m	$\kappa_{h,k}$	$\frac{0.2(k-1)}{K}$
α_{WI}	2.2	ϵ	10^{-4}
α_{WF}	3.4	ϵ_{in}	10^{-4}
α_{IF}	3	ϵ_{out}	10^{-4}
d_x	2 m		

A. Convergence

Fig. 4 illustrates the convergence behavior of Algorithm 1 and Algorithm 2 with SNR = 5 dB, 15 dB, and 25 dB. As we can see, the computation MSE in Algorithm 1 at the analyzed SNRs monotonically converges to a stationary point within a few number of iterations. Besides, Algorithm 1 converges faster in the low SNR region. Furthermore, the constraint violation of Algorithm 2 maintains decreasing and tends to be zero, which means that the unit modulus constraint in $\mathcal{P}6$ is satisfied eventually. Figs. 5-7 illustrate the convergence behaviors of the MTPDD algorithm over frames with SNR = 5 dB, 15 dB, and 25 dB, respectively. One can see that the computation MSE at different SNRs all converges to a stationary point of $\mathcal{P}1$ over a number of frames.

B. Impact of the Number of WDs

We show the computation MSE of the RIS-assisted Air-Comp systems versus the number of WDs K for the different algorithms at SNR = 15 dB in Fig. 8. We can observe

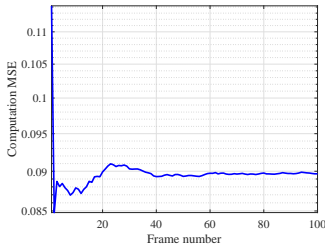


Fig. 5: Convergence behavior of the MTPDD algorithm over frames with SNR = 5 dB.

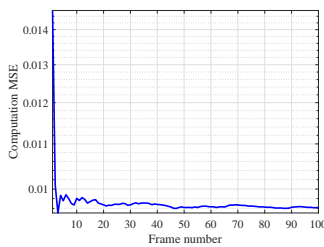


Fig. 6: Convergence behavior of the MTPDD algorithm over frames with SNR = 15 dB.

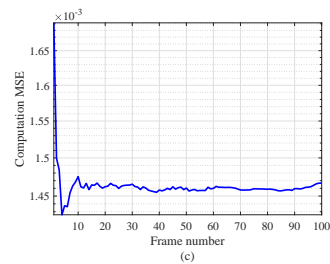


Fig. 7: Convergence behavior of the MTPDD algorithm over frames with SNR = 15 dB.

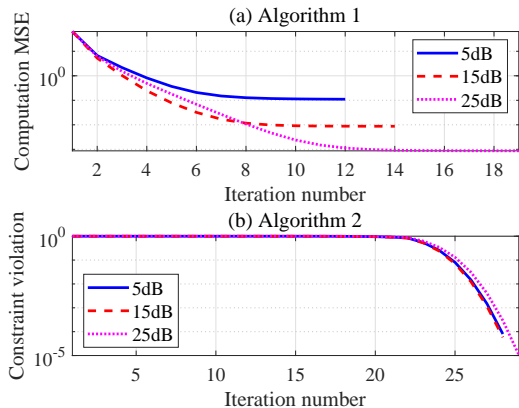


Fig. 4: The convergence of Algorithm 1 and Algorithm 2 with SNR = 15 dB.

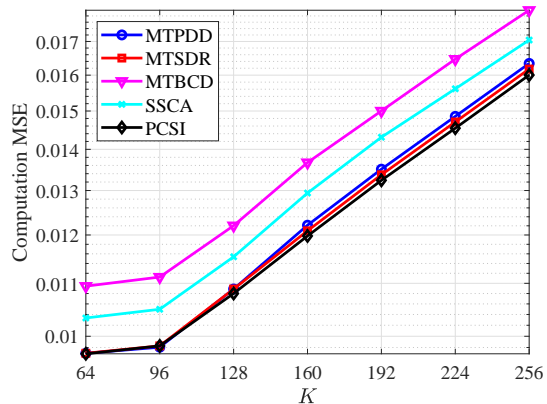


Fig. 8: The computation MSE performance versus K with SNR = 15 dB.

that the computation MSE of all the considered schemes increases as K becomes larger. This matches our discussion that enabling the connection of more WDs leads to worse computation MSE, where the design of a common passive beamforming matrix at the RIS to assist the data aggregation of more WDs is quite challenging. Our proposed MTPDD algorithm outperforms the MTBCD and SSCA algorithms with the increasing performance gaps due to the fact that our proposed algorithm can better exploit the channel statistics. Besides, the MTPDD algorithm approaches the performance of the MTSDR method with much reduced computational complexity, since the MTSDR method requires to deal with the Rank-1 issue. Furthermore, the computation MSE of the MTPDD algorithm is also close to that of the PCSI method, which is viewed as the performance lower bound. This verifies the efficiency of our proposed mixed-timescale beamforming design.

C. Impact of the Number of Receive Antennas at the FC

The computation MSE performance versus the number of receive antennas at the FC N is shown in Fig. 9 at SNR = 15 dB. When the number of receive antennas increases, we can observe that the computation MSE of the proposed algorithm decreases considerably. In particular, the computation MSE of the PCSI algorithm decreases almost linearly with N . It coincides with the discussion in [8], stating that with sufficiently large number of receive antennas N the computation MSE is inversely proportional to N . Besides, the MTPDD algorithm

still outperforms the MTBCD and SSCA algorithms with the increase of N . When $N \geq 96$, the performance gaps between our proposed algorithm and the above two benchmark schemes become much smaller due to the rich spatial degree of freedom brought by the large-scale receive antenna array of the FC. Furthermore, when N varies from 32 to 128, the MTPDD algorithm approaches the computation MSE of the MTSDR and PCSI methods with much less complexity and overhead. This further verifies the efficiency of our proposed algorithm in RIS-assisted large-scale MIMO AirComp systems.

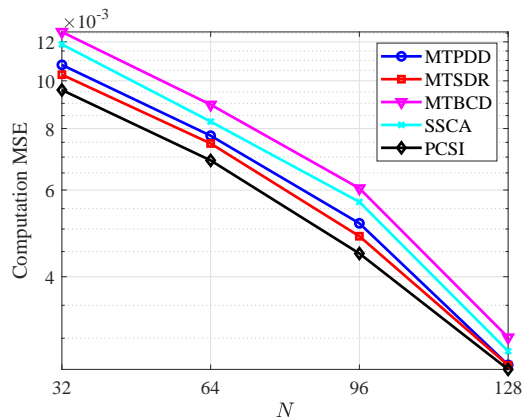


Fig. 9: The computation MSE performance versus N with SNR = 15 dB.

D. Impact of SNR

We show the computation MSE performance versus SNR in Fig. 10. It can be observed that the computation MSE achieved by all the analyzed schemes decreases monotonically as SNR increases. Within the considered region of SNR, the proposed algorithm outperforms the MTBCD and SSCA algorithms with an increasing performance gap. Specifically, the performance gap between the MTPDD and MTBCD algorithms is roughly 3.5 dB, while that between the MTPDD and SSCA algorithm is 2 dB. As SNR varies from 0 dB to 20 dB, our proposed algorithm approaches the MTSDR algorithm with a small performance gap.

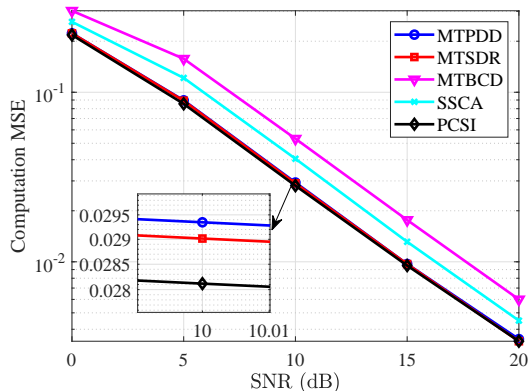


Fig. 10: The computation MSE performance versus SNR.

E. Comparison of Signaling Overhead and Complexity

We compare the number of CSI feedback bits required by the proposed MTPDD algorithm, the SSCA algorithm and the single-timescale beamforming scheme versus the number of reflecting elements at the RIS M in Fig. 11. Assuming that the quantization of each element of the CSI matrix requires $B = 6$ bits, the total number of signaling bits of the proposed MTPDD is $MN + KM + BRTKN$, while that of the SSCA scheme is $BR(TKN + MN + KM)$ since it requires to obtain a full CSI sample in each frame. Similarly, the number of signaling bits of the single-timescale beamforming scheme is $BRT(KN + MN + KM)$. One can see that these two analyzed mixed-timescale beamforming schemes (MTPDD and SSCA) both can significantly reduce the overhead (about 87% when $M = 256$) as compared to the single-timescale scheme. Furthermore, our proposed MTPDD scheme requires fewer signaling bits than that of the SSCA scheme for the RIS-assisted AirComp systems (about 45% when $M = 256$). Moreover, the ratios of CPU time of other schemes to the PCSI scheme with SNR = 15 dB are presented in Table IV. We can see that the ratios are about 70% to 92%. This unveils the tradeoff between the system performance and computation complexity. Furthermore, based on Figs. 8-11 and Table IV, we can conclude that our proposed algorithm matches the computation MSE performance of the PCSI algorithm with much lower complexity and signaling overhead.

VI. CONCLUSION

In this work, the beamforming design of RIS-assisted AirComp systems was investigated. To minimize the average com-

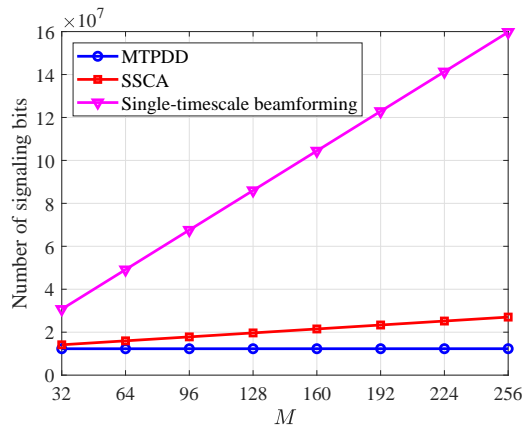


Fig. 11: The number of signaling bits versus M .

TABLE IV: The ratios of the CPU time of other schemes to the PCSI scheme with SNR = 15 dB.

Scheme	The ratio of the CPU time
MTPDD	72%
MTSDR	92%
MTBCD	65%
SSCA	70%

putation MSE over time with reduced CSI signaling overhead, we proposed a MTPDD algorithm for jointly optimizing the WDs' transmit power, the receive beamforming vector at the FC and the passive RIS beamforming matrix. In particular, the short-term transmit power and receive beamforming vector are optimized at each time slot by exploiting the Lagrange duality method and the first-order optimality condition. By contrast, we update the long-term passive RIS beamforming matrix in a frame-based manner by exploiting the PDD framework. Our simulation results showed that the proposed MTPDD beamforming design outperforms the MTBCD and SSCA benchmark schemes. Besides, the performance of the MTPDD algorithm approaches the computation MSE lower bound achieved by the single-timescale beamforming scheme relied on perfect real-time CSI with much lower signaling overhead and complexity. Our proposed mixed-timescale beamforming design can serve as an excellent candidate for RIS-assisted AirComp systems, achieving the tradeoff between the performance and signaling overhead.

In the following, some important issues that are not addressed yet in this work are discussed to motivate future research.

- For simplicity, we only considered the single-antenna WDs in this work. In general, the WDs may collect several types of parameters simultaneously and be equipped with multiple antennas. Then the RIS-assisted AirComp systems need to perform multi-function computation. Therefore, the mixed-timescale beamforming design for RIS-assisted MIMO AirComp systems is an important problem to investigate in the future.
- In this work, we assumed that there is no channel estimation error when acquiring the real-time CSI or the channel statistics. In practice, since the number of reflecting elements at the RIS is large, it is challenging to perform channel estimation and the large errors may exist.

Hence, the robust mixed-timescale beamforming design against CSI errors is worthy of further investigation.

- In practice, due to the hardware limitation, the phase shifts at the reflecting elements of the RIS are discrete. Thus, the mixed-timescale beamforming with discrete phase shifts for RIS-assisted AirComp systems is an interesting topic."

APPENDIX A PROOF OF THEOREM 1

From problem $\mathcal{P}5$, we can rewrite its objective function as

$$\begin{aligned}
& \mathbb{E} \left[\sum_{k \in \mathcal{K}} |\mathbf{u}^H (\mathbf{Q}\Theta \mathbf{g}_k + \mathbf{h}_k) v_k - 1|^2 + \sigma^2 |\mathbf{u}|^2 \right] \\
&= \sum_{k \in \mathcal{K}} \mathbb{E} (|\mathbf{u}^H (\mathbf{Q}\Theta \mathbf{g}_k + \mathbf{h}_k) v_k - 1|^2) + \mathbb{E} (\sigma^2 |\mathbf{u}|^2) \\
&= \sum_{k \in \mathcal{K}} \mathbb{E} \left(|\mathbf{u}^H (\mathbf{Q}\Theta \mathbf{g}_k + \mathbf{h}_k) v_k|^2 \right. \\
&\quad \left. - \mathbf{u}^H (\mathbf{Q}\Theta \mathbf{g}_k + \mathbf{h}_k) v_k - v_k^* (\mathbf{Q}\Theta \mathbf{g}_k + \mathbf{h}_k)^H \mathbf{u} + 1 \right) \\
&\quad + \mathbb{E} (\sigma^2 |\mathbf{u}|^2) \\
&= \sum_{k \in \mathcal{K}} \mathbb{E} (|\mathbf{u}^H ((\bar{\mathbf{F}} + \Phi_t^{1/2} \mathbf{F} \Phi_r^{1/2}) \Theta (\bar{\mathbf{z}}_{g,k} + \Phi_{g,k}^{1/2} \mathbf{z}_{g,k})) \\
&\quad + \bar{\mathbf{z}}_{h,k} + \Phi_{h,k}^{1/2} \mathbf{z}_{h,k}) v_k|^2 \\
&\quad - \mathbf{u}^H ((\bar{\mathbf{F}} + \Phi_t^{1/2} \mathbf{F} \Phi_r^{1/2}) \Theta (\bar{\mathbf{z}}_{g,k} + \Phi_{g,k}^{1/2} \mathbf{z}_{g,k})) \\
&\quad + \bar{\mathbf{z}}_{h,k} + \Phi_{h,k}^{1/2} \mathbf{z}_{h,k}) v_k \\
&\quad - v_k^* ((\bar{\mathbf{F}} + \Phi_t^{1/2} \mathbf{F} \Phi_r^{1/2}) \Theta (\bar{\mathbf{z}}_{g,k} + \Phi_{g,k}^{1/2} \mathbf{z}_{g,k})) \\
&\quad + \bar{\mathbf{z}}_{h,k} + \Phi_{h,k}^{1/2} \mathbf{z}_{h,k})^H \mathbf{u} + 1) + \mathbb{E} (\sigma^2 |\mathbf{u}|^2) \\
&= \sum_{k \in \mathcal{K}} (|x_{1,k}|^2 + \mathbb{E}(|x_{2,k}|^2) + \mathbb{E}(|x_{3,k}|^2) + \mathbb{E}(|x_{4,k}|^2) \\
&\quad + \mathbb{E}(|x_{5,k}|^2) - x_{1,k} - x_{1,k}^* + 1) + \sigma^2 |\mathbf{u}|^2, \tag{31}
\end{aligned}$$

where

$$x_{1,k} = \mathbf{u}^H (\bar{\mathbf{F}} \Theta \bar{\mathbf{z}}_{g,k} + \bar{\mathbf{z}}_{h,k}) v_k, \tag{32}$$

$$x_{2,k} = \mathbf{u}^H \bar{\mathbf{F}} \Theta \Phi_{g,k}^{1/2} \mathbf{z}_{g,k} v_k, \tag{33}$$

$$x_{3,k} = \mathbf{u}^H \Phi_t^{1/2} \mathbf{F} \Phi_r^{1/2} \Theta \bar{\mathbf{z}}_{g,k} v_k, \tag{34}$$

$$x_{4,k} = \mathbf{u}^H \Phi_t^{1/2} \mathbf{F} \Phi_r^{1/2} \Theta \Phi_{g,k}^{1/2} \mathbf{z}_{g,k} v_k, \tag{35}$$

$$x_{5,k} = \mathbf{u}^H \Phi_{h,k}^{1/2} \mathbf{z}_{h,k} v_k. \tag{36}$$

Note that the last equality in (31) holds since \mathbf{F} , $\mathbf{z}_{g,k}$, and $\mathbf{z}_{h,k}$ are independent of each other. For $|x_{1,k}|^2$, $x_{1,k}$, and $x_{1,k}^*$ in (31), according to [48], we have

$$\begin{aligned}
|x_{1,k}|^2 &= |v_k|^2 (\phi^H \text{diag}(\mathbf{u}^H \bar{\mathbf{F}}) \bar{\mathbf{z}}_{g,k} \bar{\mathbf{z}}_{g,k}^H \text{diag}(\bar{\mathbf{F}}^H \mathbf{u}) \phi \\
&\quad + \phi^H \text{diag}(\mathbf{u}^H \bar{\mathbf{F}}) \bar{\mathbf{z}}_{g,k} \bar{\mathbf{z}}_{h,k}^H \mathbf{u} \\
&\quad + \mathbf{u}^H \bar{\mathbf{z}}_{h,k} \bar{\mathbf{z}}_{g,k}^H \text{diag}(\bar{\mathbf{F}}^H \mathbf{u}) \phi + \mathbf{u}^H \bar{\mathbf{z}}_{h,k} \bar{\mathbf{z}}_{h,k}^H \mathbf{u}), \tag{37}
\end{aligned}$$

$$x_{1,k} = \phi^H \text{diag}(\mathbf{u}^H \bar{\mathbf{F}}) \bar{\mathbf{z}}_{g,k} v_k + \mathbf{u}^H \bar{\mathbf{z}}_{h,k} v_k, \tag{38}$$

$$x_{1,k}^* = v_k^* \bar{\mathbf{z}}_{g,k}^H \text{diag}(\bar{\mathbf{F}}^H \mathbf{u}) \phi + v_k^* \bar{\mathbf{z}}_{h,k}^H \mathbf{u}. \tag{39}$$

Next, by expanding the term $\mathbb{E}(|x_{2,k}|^2)$, we can obtain

$$\begin{aligned}
\mathbb{E}(|x_{2,k}|^2) &= |v_k|^2 \mathbb{E}(\phi^H \text{diag}(\mathbf{u}^H \bar{\mathbf{F}}) \Phi_{g,k}^{1/2} \mathbf{z}_{g,k} \\
&\quad \times \mathbf{z}_{g,k}^H \Phi_{g,k}^{1/2H} \text{diag}(\bar{\mathbf{F}}^H \mathbf{u}) \phi) \\
&= \frac{|v_k|^2}{1 + \beta_{WI}} \phi^H \text{diag}(\mathbf{u}^H \bar{\mathbf{F}}) \Phi_{g,k} \text{diag}(\bar{\mathbf{F}}^H \mathbf{u}) \phi. \tag{40}
\end{aligned}$$

For the term $\mathbb{E}(|x_{3,k}|^2)$, we have

$$\begin{aligned}
\mathbb{E}(|x_{3,k}|^2) &= |v_k|^2 \mathbb{E}(\phi^H \text{diag}(\mathbf{u}^H \Phi_t^{1/2} \mathbf{F} \Phi_r^{1/2}) \bar{\mathbf{z}}_{g,k} \\
&\quad \times \bar{\mathbf{z}}_{g,k}^H \text{diag}(\Phi_r^{1/2H} \mathbf{F}^H \Phi_t^{1/2H} \mathbf{u}) \phi) \\
&= |v_k|^2 \mathbb{E}(\phi^H (\Phi_r^{1/2H} \mathbf{F}^H \Phi_t^{1/2H} \mathbf{u} \mathbf{u}^H \Phi_t^{1/2} \mathbf{F} \Phi_r^{1/2} \\
&\quad \circ (\bar{\mathbf{z}}_{g,k} \bar{\mathbf{z}}_{g,k}^H)) \phi) \\
&= |v_k|^2 \phi^H (\Phi_r^{1/2H} \mathbb{E}(\mathbf{F}^H \Phi_t^{1/2H} \mathbf{u} \mathbf{u}^H \Phi_t^{1/2} \mathbf{F}) \Phi_r^{1/2} \\
&\quad \circ (\bar{\mathbf{z}}_{g,k} \bar{\mathbf{z}}_{g,k}^H)) \phi. \tag{41}
\end{aligned}$$

Since $\Phi_t^{1/2}$ is a Hermitian matrices, by denoting $\Phi_t^{1/2H} \mathbf{u} \mathbf{u}^H \Phi_t^{1/2} \triangleq \mathbf{U} \Sigma \mathbf{U}^H$ according to the eigendecomposition (\mathbf{U} is a unitary matrix) [48], we can show that

$$\begin{aligned}
& \mathbb{E}(\mathbf{F}^H \Phi_t^{1/2H} \mathbf{u} \mathbf{u}^H \Phi_t^{1/2} \mathbf{F}) \\
&= \mathbb{E}(\mathbf{F}^H \mathbf{U} \Sigma \mathbf{U}^H \mathbf{F}) \\
&= \mathbb{E}(\mathbf{F}^H \Sigma \mathbf{F}) \\
&= \left(\sum_{i \in \mathcal{M}} \lambda_i \right) \frac{1}{1 + \beta_{IF}} \mathbf{I}, \tag{42}
\end{aligned}$$

where the second equality is due to the fact that the distribution of a matrix does not change after multiplying a unitary matrix [48]. Hence, we can rewrite (41) as

$$\begin{aligned}
\mathbb{E}(|x_{3,k}|^2) &= \left(\sum_{i \in \mathcal{M}} \lambda_i \right) \frac{|v_k|^2}{1 + \beta_{IF}} \phi^H (\Phi_r \circ (\bar{\mathbf{z}}_{g,k} \bar{\mathbf{z}}_{g,k}^H)) \phi. \tag{43}
\end{aligned}$$

Similarly, we have

$$\begin{aligned}
\mathbb{E}(|x_{4,k}|^2) &= \left(\sum_{i \in \mathcal{M}} \lambda_i \right) \frac{|v_k|^2 \phi^H (\Phi_r \circ \Phi_{g,k}) \phi}{(1 + \beta_{WI})(1 + \beta_{IF})} \tag{44}
\end{aligned}$$

$$\mathbb{E}(|x_{5,k}|^2) = \frac{|v_k|^2}{1 + \beta_{WF}} \mathbf{u}^H \Phi_{h,k} \mathbf{u}. \tag{45}$$

Combining (37)-(40) and (43)-(45), the results can be obtained.

APPENDIX B PROOF OF THEOREM 2

We first outline the main procedures of the proof for Theorem 2. Based on the fixed long-term variable Θ and given CSI, the short-term alternating-optimization algorithm applied to problem $\mathcal{P}2$ is proven to converge to a stationary point. Besides, the sequence generated by the PDD-based algorithm in Algorithm 2 also converges to a stationary point of the problem $\mathcal{P}5$ by fixing $\{v_k\}$ and \mathbf{u} . With the above results, we finally demonstrate the convergence of the MTPDD algorithm (Algorithm 3) applied to problem $\mathcal{P}1$.

At the beginning of the proof, we prove that there exists at least one limit point. In this paper, one can see that the feasible sets of v_k , \mathbf{u} , and Θ are all compact. Then problem $\mathcal{P}1$ over their Cartesian product set is bounded. Hence, the sequence generated by the MTPDD algorithm is compact and bounded, which must have one limit point at least.

In the following, we prove the convergence of Algorithm 1 in the short-term optimization. As we can see that all the objective functions of the subproblems are equivalent to that of problem $\mathcal{P}2$. Thus, the objective function of problem $\mathcal{P}2$ is lower bounded at each iteration when we apply Algorithm 1. Next, we show that at each iteration the objective function value of problem $\mathcal{P}2$ is non-increasing. We denote q as the iteration index. According to the derivation in Section III-A, for given Θ , $\{\mathbf{h}_k\}$, $\{\mathbf{g}_k\}$, and \mathbf{Q} , we have

$$\begin{aligned} & g(\{v_{k,q+1}\}, \Theta, \mathbf{u}_q; \{\mathbf{h}_k\}, \{\mathbf{g}_k\}, \mathbf{Q}) \\ & \leq g(\{v_{k,q}\}, \Theta, \mathbf{u}_q; \{\mathbf{h}_k\}, \{\mathbf{g}_k\}, \mathbf{Q}), \end{aligned} \quad (46)$$

$$\begin{aligned} & g(\{v_{k,q}\}, \Theta, \mathbf{u}_{q+1}; \{\mathbf{h}_k\}, \{\mathbf{g}_k\}, \mathbf{Q}) \\ & \leq g(\{v_{k,q}\}, \Theta, \mathbf{u}_q; \{\mathbf{h}_k\}, \{\mathbf{g}_k\}, \mathbf{Q}). \end{aligned} \quad (47)$$

We can be observed that the objective function value of problem $\mathcal{P}2$ is non-increasing and also lower bounded by zeros at each iteration. Therefore, Algorithm 1 converges to a set of stationary solution of problem $\mathcal{P}2$.

Moreover, we obtain that the PDD-based algorithm in Algorithm 2 is guaranteed to converge to a set of stationary points of problem $\mathcal{P}6$. The detailed proof can be found in [53], which is omitted to save space. Considering the fact that problem $\mathcal{P}6$ is equivalent to problem $\mathcal{P}5$ (i.e., the global optimal solutions for the two problems are identical), the convergence of the PDD-based algorithm applied in problem $\mathcal{P}5$ is also guaranteed.

Finally, we prove the convergence of the MTPDD algorithm. Let ϕ^* denote the stationary point generated by Algorithm 2. From the first-order optimality condition, we have

$$\nabla_{\phi}^T \bar{g}(\{v_k^*\}, \phi^*, \mathbf{u}^*)(\phi - \phi^*) \geq 0, |\phi_m| = 1, \forall m \in \mathcal{M}, \quad (48)$$

where $\bar{g}(\{v_k^*\}, \phi^*, \mathbf{u}^*)$ is equivalent to $\bar{g}(\{v_k^*\}, \Theta^*, \mathbf{u}^*)$ by replacing Θ^* with ϕ^* , where $\{v_k^*\}$ and \mathbf{u}^* are generated by Algorithm 1. Combining (46), (47), and (48), the proposed MTPDD algorithm converges to a set of stationary points of problem $\mathcal{P}1$.

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] X. Chen, D. W. K. Ng, W. Yu, E. G. Larsson, N. Al-Dhahir, and R. Schober, "Massive access for 5G and beyond," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 3, pp. 615-637, Mar. 2021.
- [4] X. Cao, G. Zhu, J. Xu, and K. Huang, "Optimal power control for over-the-air computation in fading channels," *IEEE Trans. Wireless Commun.*, vol. 19, no. 11, pp. 7498-7513, Nov. 2020.
- [5] X. Zang, W. Liu, Y. Li, and B. Vucetic, "Over-the-Air computation systems: Optimal design with sum-power constraint," *IEEE Wireless Commun. Lett.*, vol. 9, no. 9, pp. 1524-1528, Sep 2020.
- [6] D. Wen, G. Zhu, and K. Huang, "Reduced-dimension design of MIMO over-the-air computing for data aggregation in clustered IoT networks," *IEEE Trans. Wireless Commun.*, vol. 18, no. 11, pp. 5255-5268, Nov. 2019.
- [7] 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.
- [8] X. Zhai, X. Chen, J. Xu, and D. W. K. Ng, "Hybrid beamforming for massive MIMO over-the-air computation," *IEEE Trans. Commun.*, vol. 69, no. 4, pp. 2737-2751, Apr. 2021.
- [9] G. Zhu, J. Xu, and K. Huang, "Over-the-air computing for 6G-Turning Air into a computer," [Online]. <https://arxiv.org/pdf/2009.02181.pdf>, 2020.
- [10] 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.
- [11] 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.
- [12] B. Nazer and M. Gastpar, "Computation over multiple-access channels," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3498-3516, Oct. 2007.
- [13] 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.
- [14] 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.
- [15] 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.
- [16] 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.
- [17] 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.
- [18] 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.
- [19] 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.
- [20] W. Liu, X. Zang, Y. Li, and B. Vucetic, "Over-the-Air computation systems: Optimization, analysis and scaling laws," *IEEE Trans. Wireless Commun.*, vol. 19, no. 8, pp. 5488-5502, Aug. 2020.
- [21] 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.
- [22] 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.
- [23] Y.-S. Jeon, M. M. Amiri, J. Li, and H. V. Poor, "Gradient estimation for federated learning over massive MIMO communication systems," [Online]. <https://arxiv.org/pdf/2003.08059.pdf>, 2020.
- [24] O. Abari, H. Rahul, D. Katabi, and M. Pant, "Airshare: Distributed coherent transmission made seamless," in *Proc. IEEE INFOCOM*, pp. 1742-1750, Apr. 2015.
- [25] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming," *IEEE Trans. Wireless Commun.*, vol. 18, no. 11, pp. 5394-5409, Nov. 2019.
- [26] Q. Wu and R. Zhang, "Beamforming optimization for wireless network aided by intelligent reflecting surface with discrete phase shifts," *IEEE Trans. Commun.*, vol. 68, no. 3, pp. 1838-1851, Mar. 2020.
- [27] C. Huang, A. Zappone, G. C. Alexandropoulos, M. Debbah, and C. Yuen, "Reconfigurable intelligent surfaces for energy efficiency in wireless communication," *IEEE Trans. Wireless Commun.*, vol. 18, no. 8, pp. 4157-4170, Aug. 2019.
- [28] M. Di Renzo *et al.*, "Smart radio environments empowered by AI reconfigurable meta-surfaces: An idea whose time has come," *EURASIP J. Wireless Commun. Netw.*, May 2019. [Online]. <https://doi.org/10.1186/s13638-019-1438-9>.
- [29] E. Basar, M. Di Renzo, J. De Rosny, M. Debbah, M. Alouini, and R. Zhang, "Wireless communications through reconfigurable intelligent surfaces," *IEEE Access*, vol. 7, pp. 116753-116773, Aug. 2019.
- [30] Q. Wu and R. Zhang, "Towards smart and reconfigurable environment: Intelligent reflecting surface aided wireless network," *IEEE Commun. Mag.*, vol. 58, no. 1, pp. 106-112, Nov. 2019.
- [31] Y. Yang, B. Zheng, S. Zhang, and R. Zhang, "Intelligent reflecting surface meets OFDM: Protocol design and rate maximization," *IEEE Trans. Commun.*, vol. 68, no. 7, pp. 4522-4535, Jul. 2020.
- [32] B. Zheng and R. Zhang, "Intelligent reflecting surface-enhanced OFDM: Channel estimation and reflection optimization," *IEEE Wireless Commun. Lett.*, vol. 9, no. 4, pp. 518-522, Apr. 2020.

- [33] M. Cui, G. Zhang, and R. Zhang, "Secure wireless communication via intelligent reflecting surface," *IEEE Wireless Commun. Lett.*, vol. 8, no. 5, pp. 1410-1414, Oct. 2019.
- [34] D. Xu, X. Yu, Y. Sun, D. W. K. Ng, and R. Schober, "Resource allocation for secure IRS-assisted multiuser MISO systems," in *Proc. IEEE GLOBECOM Wkshps.*, pp. 1-6, Dec. 2019.
- [35] X. Guan, Q. Wu, and R. Zhang, "Intelligent reflecting surface assisted secrecy communication: Is artificial noise helpful or not?" *IEEE Wireless Commun. Lett.*, vol. 9, no. 6, pp. 778-782, Jun. 2020.
- [36] Q. Wu and R. Zhang, "Weighted sum power maximization for intelligent reflecting surface aided SWIPT," *IEEE Wireless Commun. Lett.*, vol. 9, no. 5, pp. 587-590, May 2020.
- [37] C. Pan, H. Ren, K. Wang, M. Elkashlan, A. Nallanathan, J. Wang, and L. Hanzo, "Intelligent reflecting surface enhanced MIMO broadcasting for simultaneous wireless information and power transfer," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1719-1734, Aug. 2020.
- [38] Q. Wu and R. Zhang, "Joint active and passive beamforming optimization for intelligent reflecting surface assisted SWIPT under QoS constraints," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1735-1748, Aug. 2020.
- [39] X. Li, C. Zhang, C. He, G. Chen, and J. A. Chambers, "Sum rate maximization in IRS-assisted wireless power communication networks," *IEEE Internet Things J.*, Early Access, Apr. 2021.
- [40] S. Jung, J.-W. Lee, and C. Lee, "RSS-based channel estimation for IRS-aided wireless energy transfer system," *IEEE Internet Things J.*, Early Access, Apr. 2021.
- [41] T. Jiang and Y. Shi, "Over-the-air computation via intelligent reflecting surfaces," in *Proc. IEEE GLOBECOM*, pp. 1-6, Feb. 2020.
- [42] D. Yu, S.-H. Park, O. Simeone, S. S. Shitz, "Optimizing over-the-air computation in IRS-aided C-RAN systems," in *Proc. IEEE SPAWC*, pp. 1-5, Aug. 2020.
- [43] Z. Wang, Y. Shi, Y. Zhou, H. Zhou, and N. Zhang, "Wireless-powered over-the-air computation in intelligent reflecting surface aided IoT networks," *IEEE Internet Things J.*, vol. 8, no. 3, pp. 1585-1598, Feb. 2021.
- [44] M.-M. Zhao, Q. Wu, M.-J. Zhao, and R. Zhang, "Intelligent reflecting surface enhanced wireless networks: Two-timescale beamforming optimization," *IEEE Trans. Wireless Commun.*, vol. 20, no. 1, pp. 2-17, Jan. 2021.
- [45] 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. 2484-2498, May 2019.
- [46] X. Chen, A. Liu, Y. Cai, V. K. N. Lau, and M. Zhao, "Randomized two-timescale hybrid precoding for downlink multicell massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 67, no. 16, pp. 4152-4167, Aug. 2019.
- [47] X. Zhai, X. Chen, and Y. Cai, "Power minimization for massive MIMO over-the-air computation with two-timescale hybrid beamforming," *IEEE Wireless Commun. Lett.*, vol. 10, no. 4, pp. 873-877, Apr. 2021.
- [48] K. B. Petersen and M. S. Pedersen, "The matrix cookbook," Nov. 2012 [Online]. Available: <https://www.math.uwaterloo.ca/hwolkowi/matrixcookbook.pdf>.
- [49] G. Arunabha, J. Zhang, J. G. Andrews, and R. Muhamed, "Fundamentals of LTE," *The Prentice Hall communications engineering and emerging technologies series*, 2010.
- [50] M. Goldenbaum, H. Boche and S. Stańczak, "Nomographic functions: Efficient computation in clustered Gaussian sensor networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 4, pp. 2093-2105, Apr. 2015.
- [51] M. R. McKay and I. B. Collings, "General capacity bounds for spatially correlated Rician MIMO channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 9, pp. 3121-3145, Sep. 2005.
- [52] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge U.K.: Cambridge Univ. Press, 2004.
- [53] Q. Shi and M. Hong, "Penalty dual decomposition method for nonsmooth nonconvex optimization-Part I: Algorithms and convergence analysis," *IEEE Trans. Signal Process.*, vol. 68, pp. 4108-4122, Jun. 2020.
- [54] S. L. Loyka, "Channel capacity of MIMO architecture using the exponential correlation matrix," *IEEE Commun. Lett.*, vol. 5, no. 9, pp. 369-371, Sep. 2001.
- [55] X. Xia, D. Zhang, K. Xu, W. Ma, and Y. Xu, "Hardware impairments aware transceiver for full-duplex massive MIMO relaying," *IEEE Trans. Signal Process.*, vol. 63, no. 24, pp. 6565-6580, Dec. 2015.
- [56] Y. Wei, M.-M. Zhao, M. Hong, M.-J. Zhao, and M. Lei, "Learned conjugate gradient descent network for massive MIMO detection," *IEEE Trans. Signal Process.*, vol. 68, pp. 6336-6349, Nov. 2020.



Xiongfei Zhai (M'19) received the B.S. and Ph.D. degrees from Zhejiang University, Hangzhou, China, in 2009 and 2018, respectively. He was a visiting student with the Department of Electrical and Computer Engineering, Iowa State University, Ames, IA, USA. He is currently an assistant professor with the School of Information Engineering, Guangdong University of Technology, Guangzhou, China. His research interests include over-the-air computation, massive MIMO communications, millimeter-wave communications, and application of optimization

algorithm in communications.



Guojun Han (M'12-SM'14) received his Ph.D. from Sun Yatsen University, Guangzhou, China, and the M.E. degree from South China University of Technology, Guangzhou, China. From March 2011 to August 2013, he was a Research Fellow at the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. From October 2013 to April 2014, he was a Research Associate at the Department of Electrical and Electronic Engineering, Hong Kong University of Science and Technology. He is now a Full Professor and Executive Dean at the School of Information Engineering, Guangdong University of Technology, Guangzhou, China. He has been a Senior Member of IEEE since 2014. His research interests are in the areas of wireless communications, signal processing, coding and information theory. He has more than 14 years experience on research and development of advanced channel coding and signal processing algorithms and techniques for various data storage and communication systems.



Yunlong Cai (S'07-M'10-SM'16) received the M.Sc. degree in electronic engineering from the University of Surrey, Guildford, U.K., in 2006, and the Ph.D. degree in electronic engineering from the University of York, York, U.K., in 2010. From 2010 to 2011, he was a Post-Doctoral Fellow with the Electronics and Communications Laboratory, Conservatoire National des Arts et Metiers, Paris, France. Since February 2011, he has been with the College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou, China,

where he is currently a Full Professor. From August 2016 to January 2017, he was a Visiting Scholar at the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, USA. His research interests include transceiver design for multiple-antenna systems, cooperative and relay communications, UAV communications, and machine learning for communications. He is currently an Associate Editor of IEEE SIGNAL PROCESSING LETTERS.



Lajos Hanzo (<http://www-mobile.ecs.soton.ac.uk>, https://en.wikipedia.org/wiki/Lajos_Hanzo)

(FIEEE'04) received his Master degree and Doctorate in 1976 and 1983, respectively from the Technical University (TU) of Budapest. He was also awarded the Doctor of Sciences (DSc) degree by the University of Southampton (2004) and Honorary Doctorates by the TU of Budapest (2009) and by the University of Edinburgh (2015). He is a Foreign Member of the Hungarian Academy of Sciences and a former Editor-in-Chief of the IEEE

Press. He has served several terms as Governor of both IEEE ComSoc and of VTS. He has published 1900+ contributions at IEEE Xplore, 19 Wiley-IEEE Press books and has helped the fast-track career of 123 PhD students. Over 40 of them are Professors at various stages of their careers in academia and many of them are leading scientists in the wireless industry. He is also a Fellow of the Royal Academy of Engineering (FREng), of the IET and of EURASIP.