Joint Secure Transceiver Design for an Untrusted MIMO Relay Assisted Over-the-Air Computation Networks With Perfect and Imperfect CSI

Hualiang Luo, Quanzhong Li, Qi Zhang, *Member, IEEE*, and Jiayin Qin

***Abstract*— In this paper, we investigate the physical layer security of an untrusted relay assisted over-the-air computation (AirComp) network, where each node is equipped with multiple antennas and the relay is operated in an amplify-and-forward mode. The relay receives the data from each sensor and sends them to the access point (AP) in the first and second time slot, respectively. The AP applies artificial noise (AN) to protect the aggregation of sensors' data from being wiretapped by the untrusted relay in the first time slot. In particular, we are interested in minimizing the computation distortion measured by the mean-squared error (MSE) via jointly optimizing beamforming matrices at all nodes, subject to the MSE constraint at the relay and individual power constraints at the AP, the relay and each sensor. In the case of the perfect channel state information (CSI), we convert the nonconvex MSE minimization problem into a difference-of-convex (DC) form and propose a constrained concave-convex procedure that can obtain a local minimum to solve the DC problem. We also generalize the framework to an imperfect CSI case where the additional interference term due to incomplete interference cancellation is considered, and the nonconvex robust MSE minimization problem is solved by a proposed inexact block coordinate descent algorithm. Numerical results are presented to show the effectiveness of our proposed schemes.**

***Index Terms*— Over-the-air computation, physical layer security, untrusted relay, transceiver design, optimization.**

I. Introduction

With the increasing number of Internet-of-Things (IoT) devices [1], [2], the communication overhead between mobile devices and the access point (AP) that collects the data becomes dominant. Unlike traditional communication base stations, the AP is more interested in aggregated observations than individual observations [3]. Over-the-air computation (AirComp) is a promising solution for wireless data aggregation, and the main idea of AirComp is to utilize the signal-superposition property of a multiple access channel for "over-the-air aggregation". Benefiting from the ultra-low-latency and high-mobility [4], [5], [6] of the AirComp in massive wireless data aggregation, much attention has been paid to AirComp [3], [7], [8], [9], [10], [11]. The authors in [9] propose a scheme combining the AirComp and orthogonal frequency division multiplexing, where the desired function is divided into sub-functions and allocated to multiple subcarriers to prevent a vanishing computation rate from the increase in the number of sensors. In [10], the tradeoff between AirComp and delay has been investigated. A cooperative wide-band spectrum sensing scheme using AirComp has been studied in [11], which utilizes the superposition property of wireless channel to implement the computation of Fourier transform. However, both AP in the AirComp network and the base station (BS) in the traditional wireless communication network may suffer from severe fading when sensors or wireless users are far from them. In order to address this issue, relays have been widely applied in the traditional wireless communication network. In [12], an amplify-and-forward (AF) relay is used to assist the BS in downlink communication, and the AF transformation matrix has been jointly optimized with the beamforming matrix of BS. A millimeter wave full-duplex multiple-input-multiple-output (MIMO) relay system is considered in [13] to increase coverage and guarantee high transmission rates for mm-wave communications. Additionally, RIS can also enhance the channel conditions between the information source and sink [14]. The authors in [15] devise and analyze a multiple-relay-aided massive NOMA network to aid the wireless communication and improve spectral efficiency. Motivated by the great success of relaying in wireless communications [15], [16], [17], relays have been introduced in the evolving AirComp networks [18], [19], [20], [21], [22], [23].

A. Related Works

1) Studies on Relay Assisted AirComp: Recently, relay assisted AirComp has become a popular investment spot since relay can improve transmission quality and increase the sensor coverage area. The authors in [18] consider an AF assisted AirComp network and propose a hierarchical communication framework to minimize the computation MSE. The authors

Manuscript received 28 August 2022; revised 8 February 2023 and 20 March 2023; accepted 9 April 2023. Date of publication 13 April 2023; date of current version 24 April 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 62272493, Grant 61802447, and Grant 61672549; and in part by the Guangdong Basic and Applied Basic Research Foundation under Grant 2023A1515011201. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Ragnar Thobaben. *(Corresponding author: Quanzhong Li.)*

Hualiang Luo and Qi Zhang are with the School of Electronics and Information Technology, Sun Yat-sen University, Guangzhou 510006, China (e-mail: luohliang@mail2.sysu.edu.cn; zhqi26@mail.sysu.edu.cn).

Quanzhong Li is with the School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou 510006, China (e-mail: liquanzh@mail.sysu.edu.cn).

Jiayin Qin, deceased, was with the School of Electronics and Information Technology, Sun Yat-sen University, Guangzhou 510006, China.

Digital Object Identifier 10.1109/TIFS.2023.3266928

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

in [19] consider an AirComp network that each sensor can transmit its signal with or without relay assistance. A relay selection scheme in relay-aided AirComp is studied in [20], which chooses the best AF relay to send the received source nodes' signal to the AP. A hierarchical AirComp network using multiple relays is investigated in [21], and the computation rate via a multi-hop system has been derived. The relays in [18], [19], [20], and [21] are assumed to be equipped with a single antenna, while some studies have considered the MIMO relays, which have multiple anttennas and can bring the benefits of higher spectral and energy efficiency, broader coverage, and lower mutual interference. A MIMO relay assisted AirComp network is studied in [22], where AP can receive the sensors' data directly from themselves or the data amplified and forwarded by the relay. In [23], a more complicated model has been considered, where sensors send their data to the AP with the help of one selected from multiple MIMO relays, while considering the direct links from the sensors to the AP.

2) Studies on Security Using Untrusted MIMO Relays: Even though a relay can bring many benefits to system performance, it may wiretap the aggregation of the sensors' data under the relay process if the relay is untrusted, or to be more specific, is a potential eavesdropper [25]. This scenario of using untrusted MIMO relays has been widely considered in previous works [26], [27], [28], [29], [30], [31], [32], [33]. In [26], [27], [28], and [29], a MIMO relay operating in a one-way amplify-and-forward mode is considered to be untrusted. In [30], [31], and [32], secure beamforming design problems with an untrusted two-way MIMO relay have been studied. The authors in [33] develop a joint relay selection and power allocation scheme to maximize the instantaneous secrecy rate of a wireless communication network, where there are multiple passive eavesdroppers and untrusted relays.

3) Studies on Security Design for AirComp: With an untrusted relay which may cause information leakage, it is rational to enhance the security of AirComp networks. In [34], physical layer security (PLS) optimization problem for the AirComp network has been studied, and artificial noise (AN) is applied to degrade the eavesdropper's links while receiving sensors' pre-processed signals. To against passive eavesdropping, the authors in [35] propose a scheme that uses a friendly jammer whose signal can be reconstructed and fully canceled by the legitimate receiver but deteriorate the eavesdropper's signal-noise-ratio (SNR), and thus inhibit the illegitimate receiver's ability to estimate the value of the objective function. Besides, different from the physical layer security rate defined in wireless communication, a δ-semantically secure based on the total variation norm on signed measures and a V-MSE-secure are defined in [35]. The V-MSE-secure guarantees that the estimation MSE at the eavesdropper is at least V under a uniformly distributed objective, regardless of which estimator the eavesdropper uses.

B. Motivations and Contributions

First, due to the broadcast nature of wireless communications, the data from sensors may be wiretapped by potential eavesdroppers. Thus, developing a scheme that protects the vulnerable data aggregation from eavesdroppers is reasonable. While quite a few advanced physical layer security strategies (including power allocation [24], [25], [26], [27], [28], [29], [30], [31], [32], [33], AN [24], [29], [32], cooperative jamming [24], [25], [27], zero-forcing beamforming [24], relay selection [24], [33], etc.) have been proposed to protect the transmission security from potential eavesdroppers in traditional wireless communication systems, they cannot be applied to the AirComp network directly. Because, in many AirComp applications, the AP only requires a function of data available at the distributed sensors rather than the complete information about the data themselves, which makes sensors' secrecy rate commonly used in traditional communications become an unsuitable mensuration in AirComp scenarios.

Second, inspired by AN can protect the signal from being wiretapped by jamming and distorting the observation of the eavesdropper in wireless communication networks [25], [34], [36], we may utilize AN to prevent the untrusted relay from wiretapping the value of the aggregation function and improve the transmission security and covertness in AirComp networks.

Third, perfect channel state information (CSI) may not always be available in practice, and a robust AirComp in the presence of imperfect CSI is a more general and realistic assumption. Considering the case of imperfect CSI, [43] uses a reconfigurable intelligent surface to assist the AirComp in IoT networks, and [44] investigates a joint beamforming design of energy supply and data aggregation for wirelessly powered AirComp systems. However, none of them consider a relay assisted AirComp network under a practical condition that perfect CSI may not be obtained.

Motivated by the above observations, in this paper, we study an AN-aided scheme for enhancing the security of an AirComp network that contains an AP, an untrusted relay and several sensors. Each node in the network is equipped with multiple antennas, and there is no direct link between the AP and each sensor. In the first time slot, each sensor sends its own data to the relay, and in the second slot, the relay amplifies and forwards the received signal to the AP.

In summary, the main contributions of this paper are:

- We propose an AN-aided scheme to protect the aggregation of sensors' data being wiretapped by the potential eavesdropper. In the first time slot, when sensors send their data to the untrusted relay, the AP sends an AN to interfere with the eavesdropper from obtaining the result of the aggregated observation function at AP. In the second time slot, the relay amplifies and forwards the received signal in the previous time slot to the AP. Note that the AN can be canceled from the received signal at the AP, thereby the AP can estimate the value of the aggregated observation function with high quality but not be interfered with AN.
- We formulate an MSE minimization problem such that the MSE at the untrusted relay should be greater than a threshold, and the power of each node should be less than a preset budget. The formulated optimization problem is highly nonconvex due to coupling optimization variables. Thus, we reformulate the optimization problem into a difference-of-convex (DC) form and propose

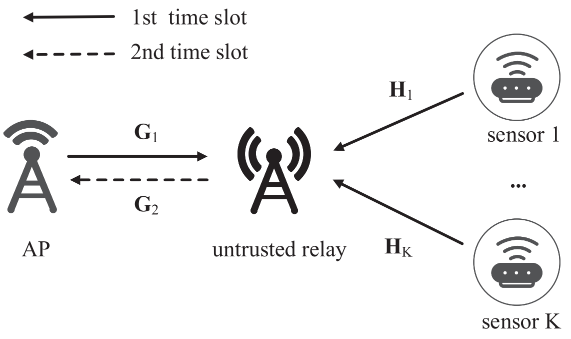

Fig. 1. The untrusted relay-assisted AirComp network.

a constrained concave-convex procedure (CCCP) based algorithm to solve it, and the proposed algorithm can obtain a local minimum.

- We also consider the case of imperfect CSI. In this scenario, the interference caused by AN can not be eliminated perfectly. Therefore, we formulate a robust AP MSE minimization problem that considers the remaining interference term. To solve the nonconvex robust optimization problem, we proposed an inexact block coordinate descent (IBCD) algorithm that can converge to a Karush-Kuhn-Tucker (KKT) point.
- Numerical results demonstrate that the proposed algorithms are effective and superior in terms of MSE performance compared to other existing schemes.

C. Organizations and Notations

The rest of this paper is organized as follows. In Section II, we describe the system model. In Section III, we formulate the optimization problem with perfect CSI and propose a CCCP algorithm to solve the problem. In Section IV, we extend the optimization problem to the case of imperfect CSI and propose an IBCD algorithm to solve it. We present the numerical results in Section V and conclude our paper in Section VI.

The $\|\mathbf{X}\|$, $\mathrm{tr}(\mathbf{X})$, $\mathbf{X}^*$, $\mathbf{X}^\dagger$, $\mathbf{X}^\ddagger$, $\mathbf{X}^T$ and $\mathrm{vec}(\mathbf{X})$ denote Frobenius norm, trace, conjugate, conjugate transpose, pseudo inverse, transpose and vectorization, respectively. $\mathbf{X} \succeq (\succ)\mathbf{0}$ denotes that $\mathbf{X}$ is positive semidefinite (positive definite). $\mathbf{x}(i)$ represents the i-th entry of vector $\mathbf{x}$. $\Re\{x\}$ means the real part of x. $\otimes$ denotes the Kronecker product. The $\mathbb{C}^{m\times n}(\mathbb{R}^{m\times n})$ denotes a complex(real) matrix with m rows and n columns. $\lambda_{\max}(\mathbf{X})$ denotes the maximum eigenvalue of matrix $\mathbf{X}$.

II. System Model

Consider an untrusted relay-assisted AirComp network as illustrated in Fig. 1, which consists of K sensors each with N_s antennas, an untrusted MIMO relay with N_r antennas and an AP with N_a antennas.[1] Denote the channel from the k-th sensor to relay as $\mathbf{H}_k \in \mathbb{C}^{N_r\times N_s}$, the channel from AP to relay as $\mathbf{G}_1 \in \mathbb{C}^{N_r\times N_a}$ and the channel from relay to AP as $\mathbf{G}_2 \in \mathbb{C}^{N_a\times N_r}$. We assume that there is no direct link between AP and any sensor, which occurs when the direct link is blocked due to long-distance path loss or obstacle [18], [20]. Each sensor transmits its data, which consists of K heterogeneous time-varying data, to AP with the help of the relay., each sensor's pre-processing data $\mathbf{s}_k \in \mathbb{C}^{N_s}, k \in \mathcal{K} = \{1\dots K\}$, is satisfied that $\mathbb{E}[\mathbf{s}_k\mathbf{s}_k^\dagger] = \mathbf{I}$ and $\mathbb{E}[\mathbf{s}_k\mathbf{s}_j^\dagger] = \mathbf{0}$, where $k \neq j$. The relay is untrusted and regarded as a potential eavesdropper which tends to eavesdrop on the data aggregation from all sensors. The relay is considered to be trusted at the service level while untrusted at the data level[2], which has been commonly adopted in the literature on secure cooperative communications with untrusted relay [25], [26], [30], [33]. The relay is operated in a half-duplex mode and the AirComp from the sensors to the AP is completed in two time slots. We assume that the system is perfectly synchronised[3] [20], [23], and the AP have the global CSI.[4] The proposed framework could be implemented in many realistic IoT networks, e.g., in heterogeneous networks (or integrated aerial-terrestrial networks), the small-cell BS (or a low-altitude unmanned aerial vehicle (UAV)) acts as an untrusted relay and the macro-cell BS (or high-altitude platform station or airship) acts as the AP to collect data [18].

We consider the sum operation[5] as the target-function [20], [43], [44]

$$\mathbf{s} = \sum_{k=1}^{K} \mathbf{s}_k. \tag{1}$$

In the first time slot, the k-th sensor's data is multiplied by a transmit beamforming matrix $\mathbf{W}_k \in \mathbb{C}^{N_s\times N_s}$ and then sent to the relay. To enhance the security of the aggregation of all sensor's data, the AP sends an AN signal $\mathbf{s}_a \in \mathbb{C}^{N_a}$ to interfere with the potential eavesdropper, i.e., the untrusted relay. The AN signal $\mathbf{s}_a$ is satisfied $\mathbb{E}[\mathbf{s}_a\mathbf{s}_a^\dagger] = \mathbf{I}$ and also multiplied by a beanforming matrix $\mathbf{V} \in \mathbb{C}^{N_a\times N_a}$. Thus, in the first time slot,

[1] Each node in the framework can be equipped with a single antenna or multiple antennas arbitrarily, and the case of multiple-input and single-output (MISO) or single-input and single-output (SISO) can be regarded as a special case of the framework. Usually, the sensor has fewer antennas because the number of sensors is generally large, and its limitation on the physical size, weight, and deployment cost need to be considered. The AP and relay are equipped with more antennas to obtain higher spectral efficiency and better performance [15].

[2] Service level trust implies the relay follows the AF protocol as expected, which involves that the relay has to send back true CSI and forward the amplified version of the received signal without modification. Being untrusted at the data level means that the relay may decipher the confidential aggregation of the sensors' data from its received signal, that is to say, being a potential passive eavesdropper.

[3] An alternative scheme called AirShare [37] could be used for synchronizing sensors by broadcasting a reference-clock signal and its effectiveness demonstrated using a prototype, while the synchronization phase offset (SPO) can be compensated by the SPO estimation and equalization method designed in [11].

[4] We assume that the CSI from each sensor to the relay and the CSI between the relay and the AP can be obtained by implementing a suitable channel estimation method [25], [36], [38], [39], [40]. Many works have been studied to reduce the overhead for CSI estimation [10], [41], [42], and it will be a promising future work to extend those techniques to relay assisted AirComp network. Additionally, since the AP has the global CSI, the AP can act as a central processor to perform the optimization process, and the optimized results can be assigned to the relay and sensors through the feedback channel [38], [40].

[5] The aggregation function can be many nomographic functions such as arithmetic mean, weighted sum, euclidean norm, polynomial, etc.

the received signal at the relay is

$$\mathbf{y}_r = \sum_{k=1}^{K} \mathbf{H}_k \mathbf{W}_k \mathbf{s}_k + \mathbf{G}_1 \mathbf{V} \mathbf{s}_a + \mathbf{n}_r, \tag{2}$$

where $\mathbf{n}_r \sim \mathcal{CN}(0, \sigma^2 \mathbf{I})$ denotes the noise at the relay.

In the second time slot, the relay amplifies and forwards the received signal to the AP. Specifically, the received signal $\mathbf{y}_r$ is amplified by a beamforming matrix $\mathbf{F} \in \mathbb{C}^{N_r \times N_r}$ and sent to the AP. Then, the received signal at the AP is

$$\mathbf{y}_d = \mathbf{G}_2 \mathbf{F} \left(\sum_{k=1}^{K} \mathbf{H}_k \mathbf{W}_k \mathbf{s}_k + \mathbf{G}_1 \mathbf{V} \mathbf{s}_a + \mathbf{n}_r \right) + \mathbf{n}_a, \tag{3}$$

where $\mathbf{n}_a \sim \mathcal{CN}(0, \sigma^2 \mathbf{I})$ denotes the noise at the AP.

The AP and the relay apply the aggregation beamforming to the received signals as the computing output, so as to reduce the computation distortion as much as possible. Then the computing output of AP and relay can be expressed as

$$\hat{\mathbf{s}}_d = \mathbf{U}_d^{\dagger} \mathbf{y}_d, \tag{4a}$$

$$\hat{\mathbf{s}}_r = \mathbf{U}_r^{\dagger} \mathbf{y}_r, \tag{4b}$$

where $\mathbf{U}_d$ and $\mathbf{U}_r$ are the aggregation beamforming matrix of the relay and the AP, respectively.

The computation distortion between $\hat{\mathbf{s}}$ and $\mathbf{s}$ can be measured by the MSE [23], i.e. $\mathbf{MSE}(\hat{\mathbf{s}}, \mathbf{s}) = \mathbb{E}(\|\hat{\mathbf{s}} - \mathbf{s}\|^2)$. Thus, the MSE at the relay is given by

$$\mathbf{E}_r = \sum_{k=1}^{K} \|\mathbf{U}_r^{\dagger} \mathbf{H}_k \mathbf{W}_k - \mathbf{I}\|^2 + \|\mathbf{U}_r^{\dagger} \mathbf{G}_1 \mathbf{V}\|^2 + \sigma^2 \|\mathbf{U}_r^{\dagger}\|^2. \tag{5}$$

III. Optimization With Perfect CSI

In this section, we consider the case that all the CSI in the AirComp network is known perfectly. Since the perfect CSI is available and AN is known at the AP, it can eliminate the interference caused by AN before aggregating sensors' data. After that, the computing output of AP can be expressed as

$$\tilde{\mathbf{s}}_d = \mathbf{U}_d^{\dagger} \left(\mathbf{G}_2 \mathbf{F} \left(\sum_{k=1}^{K} \mathbf{H}_k \mathbf{W}_k \mathbf{s}_k + \mathbf{n}_r \right) + \mathbf{n}_a \right), \tag{6}$$

and the MSE at the AP is given by

$$\mathbf{E}_d = \sum_{k=1}^{K} \|\mathbf{U}_d^{\dagger} \mathbf{G}_2 \mathbf{F} \mathbf{H}_k \mathbf{W}_k - \mathbf{I}\|^2 + \sigma^2 \|\mathbf{U}_d^{\dagger} \mathbf{G}_2 \mathbf{F}\|^2 + \sigma^2 \|\mathbf{U}_d^{\dagger}\|^2. \tag{7}$$

To enhance the AirComp performance, we are interested in minimizing the MSE at AP under the transmit power constraints at each node. Meanwhile, to protect the sensors' data aggregation from being wiretapped by the untrusted relay, the minimum MSE at the relay needs to be greater than a given threshold. Thus, the MSE minimization problem is formulated as

$$\min_{\mathbf{W}_k, \mathbf{V}, \mathbf{F}, \mathbf{U}_d} \quad \mathbf{E}_d \tag{8a}$$

$$\text{s.t.} \quad \min_{\mathbf{U}_r} \mathbf{E}_r \geq \xi, \tag{8b}$$

$$\|\mathbf{V}\|^2 \leq P_a, \tag{8c}$$

$$\sum_{k=1}^{K} \|\mathbf{F} \mathbf{H}_k \mathbf{W}_k\|^2 + \|\mathbf{F} \mathbf{G}_1 \mathbf{V}\|^2 + \sigma^2 \|\mathbf{F}\|^2 \leq P_r, \tag{8d}$$

$$\|\mathbf{W}_k\|^2 \leq P_k, \forall k \in \mathcal{K}, \tag{8e}$$

where ξ is the minimum secure MSE threshold of the untrusted relay, P_a, P_r and P_k are the maximum transmit power at the AP, the relay and the k-th sensor.

When $\mathbf{W}_k$,$\mathbf{V}$ and $\mathbf{F}$ are fixed, the optimal solution of $\mathbf{U}_r$ and $\mathbf{U}_d$ are given by [45]

$$\mathbf{U}_r^{\star} = (\mathbf{A}_r + \mathbf{B}_r + \sigma^2 \mathbf{I})^{-1} \mathbf{C}_r, \tag{9a}$$

$$\mathbf{U}_d^{\star} = (\mathbf{A}_d + \sigma^2 \mathbf{B}_d + \sigma^2 \mathbf{I})^{-1} \mathbf{C}_d, \tag{9b}$$

where

$$\mathbf{A}_r = \sum_{k=1}^{K} \mathbf{H}_k \mathbf{W}_k \mathbf{W}_k^{\dagger} \mathbf{H}_k^{\dagger}, \tag{10a}$$

$$\mathbf{B}_r = \mathbf{G}_1 \mathbf{V} \mathbf{V}^{\dagger} \mathbf{G}_1^{\dagger}, \tag{10b}$$

$$\mathbf{C}_r = \sum_{k=1}^{K} \mathbf{H}_k \mathbf{W}_k, \tag{10c}$$

$$\mathbf{A}_d = \sum_{k=1}^{K} \mathbf{G}_2 \mathbf{F} \mathbf{H}_k \mathbf{W}_k \mathbf{W}_k^{\dagger} \mathbf{H}_k^{\dagger} \mathbf{F}^{\dagger} \mathbf{G}_2^{\dagger}, \tag{10d}$$

$$\mathbf{B}_d = \mathbf{G}_2 \mathbf{F} \mathbf{F}^{\dagger} \mathbf{G}_2^{\dagger}, \tag{10e}$$

$$\mathbf{C}_d = \sum_{k=1}^{K} \mathbf{G}_2 \mathbf{F} \mathbf{H}_k \mathbf{W}_k. \tag{10f}$$

Substituting (9) into the problem (8), it can be reformulated as

$$\min_{\substack{\mathbf{W}_k, \mathbf{V}, \mathbf{F}, \\ \mathbf{A}_r, \mathbf{B}_r, \mathbf{C}_r, \\ \mathbf{A}_d, \mathbf{B}_d, \mathbf{C}_d}} \quad \mathrm{tr}\left(-\mathbf{C}_d^{\dagger} (\mathbf{A}_d + \sigma^2 \mathbf{B}_d + \sigma^2 \mathbf{I})^{-1} \mathbf{C}_d + K\mathbf{I} \right) \tag{11a}$$

$$\text{s.t.} \quad \mathrm{tr}\left(\mathbf{C}_r^{\dagger} (\mathbf{A}_r + \mathbf{B}_r + \sigma^2 \mathbf{I})^{-1} \mathbf{C}_r - K\mathbf{I} \right) + \xi \leq 0, \tag{11b}$$

$$(10a) - (10f), \tag{11c}$$

$$(8c) - (8e). \tag{11d}$$

Due to the existence of coupling optimization variables and nonlinear equality constraints, the optimal solution of (11) is hard to obtain. Therefore, we will need the following lemmas to convert the problem (11) into a DC form, and aim to obtain a local optimum.

Lemma 1: The following two sets of expressions are equivalent to each other,

$$\acute{\mathbf{X}} = \acute{\mathbf{A}} \acute{\mathbf{A}}^{\dagger}, \acute{\mathbf{Y}} = \acute{\mathbf{A}} \acute{\mathbf{B}} \acute{\mathbf{C}} \tag{12}$$

$$\begin{bmatrix} \acute{\mathbf{X}} & \acute{\mathbf{Y}} & \acute{\mathbf{A}} \\ \acute{\mathbf{Y}}^{\dagger} & \acute{\mathbf{S}} & \acute{\mathbf{C}}^{\dagger} \acute{\mathbf{B}}^{\dagger} \\ \acute{\mathbf{A}}^{\dagger} & \acute{\mathbf{B}} \acute{\mathbf{C}} & \mathbf{I} \end{bmatrix} \succeq 0, \ \mathrm{tr}\left(\acute{\mathbf{X}} - \acute{\mathbf{A}} \acute{\mathbf{A}}^{\dagger} \right) \leq 0. \tag{13}$$

Lemma 2: The following two sets of expressions are equivalent to each other,

$$\acute{\mathbf{B}} \succ 0,\ \ \acute{\mathbf{X}} = \acute{\mathbf{A}}^{\dagger}\acute{\mathbf{B}}^{-1}\acute{\mathbf{A}} \tag{14}$$

$$\begin{bmatrix} \acute{\mathbf{X}} & \acute{\mathbf{A}}^{\dagger} \\ \acute{\mathbf{A}} & \acute{\mathbf{B}} \end{bmatrix} \succeq 0,\ \ \mathrm{tr}\left(\acute{\mathbf{X}} - \acute{\mathbf{A}}^{\dagger}\acute{\mathbf{B}}^{-1}\acute{\mathbf{A}}\right) \leq 0. \tag{15}$$

Lemma 3: If $\acute{\mathbf{Y}} \succ 0$, $\mathrm{tr}(\acute{\mathbf{X}}^{\dagger}\acute{\mathbf{Y}}^{-1}\acute{\mathbf{X}})$ is convex with respect to (w.r.t) $(\acute{\mathbf{X}}, \acute{\mathbf{Y}})$

Proof: See Appendix A. ■

To apply the above lemmas to the equality constraints in (10), we define the auxiliary variables as follow,

$$\mathbf{X} = [\mathbf{H}_1\mathbf{W}_1, \ldots, \mathbf{H}_K\mathbf{W}_K], \tag{16a}$$

$$\mathbf{Y} = \mathbf{G}_2\mathbf{F}\mathbf{X}, \tag{16b}$$

$$\mathbf{Q} = \mathbf{F}\mathbf{X}, \tag{16c}$$

$$\mathbf{R} = \mathbf{F}\mathbf{G}_1\mathbf{V}, \tag{16d}$$

$$\mathbf{T} = \mathbf{F}\mathbf{F}^{\dagger}. \tag{16e}$$

Applying Lemma 1, Lemma 2, and Lemma 3 and denoting $\Theta = \{\mathbf{A}_r, \mathbf{B}_r, \mathbf{C}_r, \mathbf{A}_d, \mathbf{B}_d, \mathbf{C}_d, \mathbf{X}, \mathbf{Y}, \mathbf{Q}, \mathbf{R}, \mathbf{T}\}$, problem (11) can be further recast as

$$\min_{\substack{\mathbf{W}_k, \mathbf{V}, \\ \mathbf{F}, \Theta}} \ \mathrm{tr}\left(-\mathbf{C}_d^{\dagger}(\mathbf{A}_d + \sigma^2\mathbf{B}_d + \sigma^2\mathbf{I})^{-1}\mathbf{C}_d + K\mathbf{I}\right) \tag{17a}$$

$$\text{s.t.} \ \ \mathrm{tr}\left(\mathbf{C}_r^{\dagger}(\mathbf{A}_r + \mathbf{B}_r + \sigma^2\mathbf{I})^{-1}\mathbf{C}_r - K\mathbf{I}\right) + \xi \leq 0, \tag{17b}$$

$$\begin{bmatrix} \mathbf{A}_r & \mathbf{X} \\ \mathbf{X}^{\dagger} & \mathbf{I} \end{bmatrix} \succeq 0, \tag{17c}$$

$$\begin{bmatrix} \mathbf{B}_r & \mathbf{G}_1\mathbf{V} \\ \mathbf{V}^{\dagger}\mathbf{G}_1^{\dagger} & \mathbf{I} \end{bmatrix} \succeq 0, \tag{17d}$$

$$\mathbf{C}_r = \sum_{k=1}^{K} \mathbf{H}_k\mathbf{W}_k, \tag{17e}$$

$$\begin{bmatrix} \mathbf{A}_d & \mathbf{Y} \\ \mathbf{Y}^{\dagger} & \mathbf{I} \end{bmatrix} \succeq 0, \tag{17f}$$

$$\begin{bmatrix} \mathbf{B}_d & \mathbf{Y} & \mathbf{G}_2\mathbf{F} \\ \mathbf{Y}^{\dagger} & \mathbf{S} & \mathbf{X}^{\dagger} \\ \mathbf{F}^{\dagger}\mathbf{G}_2^{\dagger} & \mathbf{X} & \mathbf{I} \end{bmatrix} \succeq 0, \tag{17g}$$

$$\begin{bmatrix} \mathbf{B}_d & \mathbf{G}_2\mathbf{F} \\ \mathbf{F}^{\dagger}\mathbf{G}_2^{\dagger} & \mathbf{I} \end{bmatrix} \succeq 0, \tag{17h}$$

$$\begin{bmatrix} \mathbf{B}_d & \mathbf{C}_d & \mathbf{G}_2\mathbf{F} \\ \mathbf{C}_d^{\dagger} & \mathbf{S} & \mathbf{C}_r^{\dagger} \\ \mathbf{F}^{\dagger}\mathbf{G}_2^{\dagger} & \mathbf{C}_r & \mathbf{I} \end{bmatrix} \succeq 0, \tag{17i}$$

$$\mathrm{tr}(\mathbf{V}\mathbf{V}^{\dagger}) \leq P_a, \tag{17j}$$

$$\mathrm{tr}(\mathbf{Q}\mathbf{Q}^{\dagger} + \mathbf{R}\mathbf{R}^{\dagger} + \sigma^2\mathbf{F}\mathbf{F}^{\dagger}) \leq P_r, \tag{17k}$$

$$\begin{bmatrix} \mathbf{T} & \mathbf{Q} & \mathbf{F} \\ \mathbf{Q}^{\dagger} & \mathbf{S} & \mathbf{X}^{\dagger} \\ \mathbf{F}^{\dagger} & \mathbf{X} & \mathbf{I} \end{bmatrix} \succeq 0, \tag{17l}$$

$$\begin{bmatrix} \mathbf{T} & \mathbf{R} & \mathbf{F} \\ \mathbf{R}^{\dagger} & \mathbf{S} & \mathbf{V}^{\dagger}\mathbf{G}_1^{\dagger} \\ \mathbf{F}^{\dagger} & \mathbf{G}_1\mathbf{V} & \mathbf{I} \end{bmatrix} \succeq 0, \tag{17m}$$

$$\|\mathbf{W}_k\|^2 \leq P_k, \forall k \in \mathcal{K}, \tag{17n}$$

$$\mathbf{X} = [\mathbf{H}_1\mathbf{W}_1, \ldots, \mathbf{H}_K\mathbf{W}_K], \tag{17o}$$

$$\mathrm{tr}\left(\mathbf{A}_r - \mathbf{X}\mathbf{X}^{\dagger}\right) \leq 0, \tag{17p}$$

$$\mathrm{tr}\left(\mathbf{B}_r - \mathbf{G}_1\mathbf{V}\mathbf{V}^{\dagger}\mathbf{G}_1^{\dagger}\right) \leq 0, \tag{17q}$$

$$\mathrm{tr}\left(\mathbf{A}_d - \mathbf{Y}\mathbf{Y}^{\dagger}\right) \leq 0, \tag{17r}$$

$$\mathrm{tr}\left(\mathbf{B}_d - \mathbf{G}_2\mathbf{F}\mathbf{F}^{\dagger}\mathbf{G}_2^{\dagger}\right) \leq 0, \tag{17s}$$

$$\mathrm{tr}\left(\mathbf{T} - \mathbf{F}\mathbf{F}^{\dagger}\right) \leq 0. \tag{17t}$$

Note that the constraints (17p)-(17t) have DC form and they are nonconvex. Employing the exact penalty method [12], [46], problem (17) can be reformulated as

$$\begin{aligned} \min_{\substack{\mathbf{W}_k, \mathbf{V}, \\ \mathbf{F}, \Theta}} \ & \mathrm{tr}\left(-\mathbf{C}_d^{\dagger}(\mathbf{A}_d + \sigma^2\mathbf{B}_d + \sigma^2\mathbf{I})^{-1}\mathbf{C}_d + K\mathbf{I}\right) \\ & + \tau\mathrm{tr}\left(\mathbf{A}_r + \mathbf{B}_r + \mathbf{T} - \mathbf{X}\mathbf{X}^{\dagger} - \mathbf{G}_1\mathbf{V}\mathbf{V}^{\dagger}\mathbf{G}_1^{\dagger} - \mathbf{F}\mathbf{F}^{\dagger}\right) \\ & + \tau\mathrm{tr}\left(\mathbf{A}_d + \mathbf{B}_d - \mathbf{Y}\mathbf{Y}^{\dagger} - \mathbf{G}_2\mathbf{F}\mathbf{F}^{\dagger}\mathbf{G}_2^{\dagger}\right) \end{aligned} \tag{18a}$$

$$\text{s.t.} \ \ (17\text{b}) - (17\text{o}), \tag{18b}$$

where τ is a positive penalty factor that is large enough. In problem (18), the objective function is DC form and the constraints are convex.

Proposition 1: There exist $0 < \tau_l < \infty$ such that when $\tau > \tau_l$, problem (18) is equivalent to problem (17).

Proof: See Appendix B. ■

Since problem (18) is a DC program, we employ CCCP to solve it.[6] By applying the first-order Taylor approximation to the concave term in (18a), in $(t+1)$-th iteration, we need to solve the problem (19), as shown at the bottom of the next page, where $\rho > 0$ is a constant coefficient of the regular term, $\mathbf{A}_d^t$, $\mathbf{B}_d^t$, $\mathbf{C}_d^t$, $\mathbf{X}^t$, $\mathbf{Y}^t$, $\mathbf{V}^t$ and $\mathbf{F}^t$ denote the optimal solution in the t-th iteration, and $\Theta_{all} = \{\mathbf{W}_1, \ldots, \mathbf{W}_K, \mathbf{V}, \mathbf{F}, \Theta\}$.

The proposed CCCP algorithm for solving the problem (17) is summarized below.

Algorithm 1 Proposed CCCP Algorithm

```
initiate $\mathbf{X}^t$, $\mathbf{V}^t$, $\mathbf{Y}^t$, $\mathbf{F}^t$, $\mathbf{A}_d^t$, $\mathbf{B}_d^t$ and $\mathbf{C}_d^t$ by a feasible point, initiate $\tau$ and $\rho$;
$t \leftarrow 0$
repeat
    Obtain $\{\mathbf{W}_k^t\}_{k=1}^K$, $\mathbf{V}^t$,$\mathbf{F}^t$, $\mathbf{A}_r^t$, $\mathbf{B}_r^t$, $\mathbf{C}_r^t$, $\mathbf{A}_d^t$, $\mathbf{B}_d^t$, $\mathbf{C}_d^t$, $\mathbf{X}^t$, $\mathbf{Y}^t$, $\mathbf{Q}^t$, $\mathbf{R}^t$, $\mathbf{T}^t$ by solving (19);
    $\tau \leftarrow 2\tau$;
    $t \leftarrow t + 1$;
until Convergence;
```

Proposition 2: Algorithm 1 converges to a local optimal solution.

Proof: See Appendix C. ■

[6] While a general penalty CCCP algorithm is also proposed to solve DC problem in [12], our algorithm has the following differences: 1) the system model is quite different, which leads to distinct optimization problems; 2) we do not introduce any auxiliary variable when linearizing the nonconvex terms, which does not increase the dimension of the optimization problem; 3) we use Schur complement to convert equality constraints into linear matrix inequalities (LMIs) and DC constraints, whose form and kind are more complex than that in [12], and we use epigraph to convert the MSE constraint into a more traceable form.

Complexity: Problem (19) is a semi-definite programming problem, and the interior-point method can be used to solve it efficiently with the worst-case complexity of $\mathcal{O}(n^{3.5}\log(\frac{1}{\epsilon}))$, where n is the number of optimization variables and ϵ is the preset solution accuracy [47]. Thus, the computational complexity of Algorithm 1 is $\mathcal{O}(T_{iter_1}(KN_s^2+5N_d^2+5N_r^2+(2K+1)N_rN_s+(K+1)N_aN_s+N_rN_a)^{3.5}\log(\frac{1}{\epsilon}))$, where T_{iter_1} is the iteration number of Algorithm 1.

Remark 1: The outer polyblock approximation algorithm in [48] is employed to obtain a global optimal solution for a DC program. However, it is hard to be applied to problem (18) due to the non-convex quadratic terms and the inversion of the matrix in the objective function of problem (18). Obtaining a globally optimal solution to problem (18) is still an open problem.

IV. Optimization With Imperfect CSI

In this section, we generalize the AirComp network to a more practical case that all CSI contains channel estimation error.

Consider the presence of uncertainties in the channels and model them as [49] and [50]

$$\mathbf{H}_k = \bar{\mathbf{H}}_k + \Delta\mathbf{H}_k,\ \forall k \in \mathcal{K}, \tag{20a}$$

$$\mathbf{G}_i = \bar{\mathbf{G}}_i + \Delta\mathbf{G}_i,\ \forall i \in \{1,2\}, \tag{20b}$$

where $\bar{\mathbf{H}}_k$ and $\bar{\mathbf{G}}_i$ are the estimated CSI, $\Delta\mathbf{H}_k$ and $\Delta\mathbf{G}_i$ are the corresponding channel uncertainties whose elements are independent and identically distributed (i.i.d.) zero mean complex Gaussian random variables. In general, the channel uncertainties can be modeled as [49] and [50]

$$\Delta\mathbf{H}_k = \Sigma_{H,k}^{\frac{1}{2}}\breve{\mathbf{H}}_k\Psi_{H,k}^{\frac{1}{2}},\ \forall k \in \mathcal{K}, \tag{21a}$$

$$\Delta\mathbf{G}_i = \Sigma_{G,i}^{\frac{1}{2}}\breve{\mathbf{G}}_i\Psi_{G,i}^{\frac{1}{2}},\ \forall i \in \{1,2\}, \tag{21b}$$

where $\breve{\mathbf{H}}_k$ and $\breve{\mathbf{G}}_i$ are i.i.d. Gaussian random variables with zero mean and unit variance, $\Sigma_{H,k}$ and $\Sigma_{G,i}$ are the row covariance matrices of $\Delta\mathbf{H}_k$ and $\Delta\mathbf{G}_i$, and $\Psi_{H,k}^T$ and $\Psi_{G,i}^T$ are the column covariance matrices of $\Delta\mathbf{H}_k$ and $\Delta\mathbf{G}_i$.

When the CSI is imperfect, AP can no longer eliminate the interference caused by AN perfectly, and only a part of the interference can be eliminated. Therefore, the computing output of AP can be further expressed as

$$\breve{\mathbf{s}}_d = \mathbf{U}_d^\dagger\left(\mathbf{G}_2\mathbf{F}\left(\sum_{k=1}^K\mathbf{H}_k\mathbf{W}_k\mathbf{s}_k + \Delta\mathbf{G}_1\mathbf{V}\mathbf{s}_a + \mathbf{n}_r\right) + \mathbf{n}_a\right), \tag{22}$$

and the MSE at the AP is given by

$$\begin{aligned}
&\breve{\mathbf{E}}_d\\
&= \mathbb{E}\left\{\|\mathbf{U}_d^\dagger\mathbf{G}_2\mathbf{F}\sum_{k=1}^K\mathbf{H}_k\mathbf{W}_k\mathbf{s}_k + \mathbf{U}_d^\dagger\mathbf{G}_2\mathbf{F}\Delta\mathbf{G}_1\mathbf{V}\mathbf{s}_a\right.\\
&\quad\left.+\mathbf{U}_d^\dagger\mathbf{G}_2\mathbf{F}\mathbf{n}_r + \mathbf{U}_d^\dagger\mathbf{n}_a - \mathbf{s}\|^2\right\} && \text{(23a)}\\
&= \mathbb{E}\left\{\sum_{k=1}^K\left(\operatorname{tr}\left((\mathbf{U}_d^\dagger\mathbf{G}_2\mathbf{F}\mathbf{H}_k\mathbf{W}_k)(\mathbf{U}_d^\dagger\mathbf{G}_2\mathbf{F}\mathbf{H}_k\mathbf{W}_k)^\dagger\right)\right.\right.\\
&\quad\left.-\operatorname{tr}\left(\mathbf{U}_d^\dagger\mathbf{G}_2\mathbf{F}\mathbf{H}_k\mathbf{W}_k + (\mathbf{U}_d^\dagger\mathbf{G}_2\mathbf{F}\mathbf{H}_k\mathbf{W}_k)^\dagger\right)\right) + \operatorname{tr}(K\mathbf{I})\\
&\quad+\operatorname{tr}\left((\mathbf{U}_d^\dagger\mathbf{G}_2\mathbf{F}\Delta\mathbf{G}_1\mathbf{V})(\mathbf{U}_d^\dagger\mathbf{G}_2\mathbf{F}\Delta\mathbf{G}_1\mathbf{V})^\dagger\right)\\
&\quad\left.+\sigma^2\operatorname{tr}\left((\mathbf{U}_d^\dagger\mathbf{G}_2\mathbf{F})(\mathbf{U}_d^\dagger\mathbf{G}_2\mathbf{F})^\dagger\right) + \sigma^2\operatorname{tr}\left(\mathbf{U}_d^\dagger\mathbf{U}_d\right)\right\} && \text{(23b)}\\
&= \operatorname{tr}\left(\mathbf{U}_d^\dagger\mathbb{E}_{\Delta\mathbf{G}_2}\left\{\mathbf{G}_2\mathbf{F}\mathbb{E}_{\Delta\mathbf{H}}\left\{\sum_{k=1}^K(\mathbf{H}_k\mathbf{W}_k)(\mathbf{H}_k\mathbf{W}_k)^\dagger\right\}\right.\right.\\
&\quad\left.\left.\times\mathbf{F}^\dagger\mathbf{G}_2^\dagger\right\}\mathbf{U}_d\right)\\
&\quad-\sum_{k=1}^K\operatorname{tr}\left(\mathbf{U}_d^\dagger\bar{\mathbf{G}}_2\mathbf{F}\bar{\mathbf{H}}_k\mathbf{W}_k + (\mathbf{U}_d^\dagger\bar{\mathbf{G}}_2\mathbf{F}\bar{\mathbf{H}}_k\mathbf{W}_k)^\dagger\right) + \operatorname{tr}(K\mathbf{I})\\
&\quad+\operatorname{tr}\left(\mathbf{U}_d^\dagger\mathbb{E}_{\Delta\mathbf{G}_2}\left\{\mathbf{G}_2\mathbf{F}\mathbb{E}_{\Delta\mathbf{G}_1}\left\{\Delta\mathbf{G}_1\mathbf{V}\mathbf{V}^\dagger\Delta\mathbf{G}_1^\dagger\right\}\mathbf{F}^\dagger\mathbf{G}_2^\dagger\right\}\mathbf{U}_d^\dagger\right)\\
&\quad+\sigma^2\operatorname{tr}\left(\mathbf{U}_d^\dagger\mathbb{E}_{\Delta\mathbf{G}_2}\{\mathbf{G}_2\mathbf{F}\mathbf{F}^\dagger\mathbf{G}_2^\dagger\}\mathbf{U}_d\right) + \sigma^2\operatorname{tr}\left(\mathbf{U}_d^\dagger\mathbf{U}_d\right). && \text{(23c)}
\end{aligned}$$

For ease of exposition, we denote the expectation term of $\Delta\mathbf{H}$, $\Delta\mathbf{G}_1$ and $\Delta\mathbf{G}_2$ in (23c) as $\lambda_{\mathbf{H}}(\{\mathbf{W}_k\}_{k=1}^K)$, $\lambda_{\Delta\mathbf{G}_1}(\mathbf{V})$ and $\lambda_{\mathbf{G}_2}(\mathbf{F})$, and the details are presented in Appendix D for brevity.

Further, the MSE $\breve{\mathbf{E}}_d$ at the AP can be recast as

$$\begin{aligned}
\breve{\mathbf{E}}_d &= \operatorname{tr}\left(\mathbf{U}_d^\dagger\left\{\bar{\mathbf{G}}_2\mathbf{F}(\lambda_{\mathbf{H}}(\{\mathbf{W}_k\}_{k=1}^K) + \lambda_{\Delta\mathbf{G}_1}(\mathbf{V}))\mathbf{F}^\dagger\bar{\mathbf{G}}_2^\dagger\right.\right.\\
&\quad+\operatorname{tr}\left(\mathbf{F}(\lambda_{\mathbf{H}}(\{\mathbf{W}_k\}_{k=1}^K) + \lambda_{\Delta\mathbf{G}_1}(\mathbf{V}))\mathbf{F}^\dagger\Psi_{G,2}\right)\Sigma_{G,2}\\
&\quad\left.\left.+\sigma^2\lambda_{\mathbf{G}_2}(\mathbf{F}) + \sigma^2\mathbf{I}\right\}\mathbf{U}_d\right)\\
&\quad-\sum_{k=1}^K\operatorname{tr}\left(\mathbf{U}_d^\dagger\bar{\mathbf{G}}_2\mathbf{F}\bar{\mathbf{H}}_k\mathbf{W}_k + (\mathbf{U}_d^\dagger\bar{\mathbf{G}}_2\mathbf{F}\bar{\mathbf{H}}_k\mathbf{W}_k)^\dagger\right)\\
&\quad+\operatorname{tr}(K\mathbf{I}).
\end{aligned}\tag{24}$$

Similarly, denote $\lambda_{\mathbf{G}_1}(\mathbf{V})$ as an expectation term of $\Delta\mathbf{G}_1$, and the definition is presented in Appendix D. Based on the

$$\begin{aligned}
\min_{\substack{\mathbf{W}_k,\mathbf{V},\\ \mathbf{F},\Theta}}\quad & -\left(\operatorname{tr}(\mathbf{C}_d^{t\dagger}(\mathbf{A}_d^t+\sigma^2\mathbf{B}_d^t+\sigma^2\mathbf{I})^{-1}\mathbf{C}_d^t) + 2\Re\left\{\operatorname{tr}(\mathbf{C}_d^{t\dagger}(\mathbf{A}_d^t+\sigma^2\mathbf{B}_d^t+\sigma^2\mathbf{I})^{-1}(\mathbf{C}_d-\mathbf{C}_d^t))\right\}\right.\\
& \left.-\Re\left\{\operatorname{tr}(\mathbf{C}_d^{t\dagger}(\mathbf{A}_d^t+\sigma^2\mathbf{B}_d^t+\sigma^2\mathbf{I})^{-1}(\mathbf{A}_d+\sigma^2\mathbf{B}_d-\mathbf{A}_d^t-\sigma^2\mathbf{B}_d^t)(\mathbf{A}_d^t+\sigma^2\mathbf{B}_d^t+\sigma^2\mathbf{I})^{-1}\mathbf{C}_d^t)\right\}\right) + \operatorname{tr}(K\mathbf{I})\\
& +\tau\left(\operatorname{tr}\left(\mathbf{A}_r+\mathbf{B}_r+\mathbf{T}\right)+\operatorname{tr}\left(\mathbf{A}_d+\mathbf{B}_d\right)\right) - 2\tau\Re\left\{\operatorname{tr}(\mathbf{X}\mathbf{X}^{t\dagger}+\mathbf{G}_1\mathbf{V}\mathbf{V}^{t\dagger}\mathbf{G}_1^\dagger+\mathbf{F}\mathbf{F}^{t\dagger}) + \operatorname{tr}(\mathbf{Y}\mathbf{Y}^{t\dagger}+\mathbf{G}_2\mathbf{F}\mathbf{F}^{t\dagger}\mathbf{G}_2^\dagger)\right\}\\
& +\tau\left(\operatorname{tr}(\mathbf{X}^t\mathbf{X}^{t\dagger}+\mathbf{G}_1\mathbf{V}^t\mathbf{V}^{t\dagger}\mathbf{G}_1^\dagger+\mathbf{F}^t\mathbf{F}^{t\dagger}) + \operatorname{tr}(\mathbf{G}_2\mathbf{F}^t\mathbf{F}^{t\dagger}\mathbf{G}_2^\dagger+\mathbf{Y}^t\mathbf{Y}^{t\dagger})\right) + \rho\|\Theta_{all}-\Theta_{all}^t\|^2 && \text{(19a)}\\
\text{s.t.}\quad & (17\text{b}) - (17\text{o}) && \text{(19b)}
\end{aligned}$$

above, the MSE $\breve{\mathbf{E}}_r$ at the relay can be formulated as

$$\begin{aligned}\breve{\mathbf{E}}_r =& \mathrm{tr}\left(\mathbf{U}_r^\dagger(\lambda_{\mathbf{H}}(\{\mathbf{W}_k\}_{k=1}^K) + \lambda_{\mathbf{G_1}}(\mathbf{V}) + \sigma^2\mathbf{I})\mathbf{U}_r\right) \\ &- \mathrm{tr}\left(\mathbf{U}_r^\dagger\sum_{k=1}^K(\mathbf{H}_k\mathbf{W}_k) + \sum_{k=1}^K(\mathbf{W}_k^\dagger\mathbf{H}_k^\dagger)\mathbf{U}_r\right) + \mathrm{tr}(K\mathbf{I}).\end{aligned} \tag{25}$$

Similar to $\breve{\mathbf{E}}_d$ and $\breve{\mathbf{E}}_r$, the left-hand-side of constraint (8d) can be expressed as

$$\begin{aligned}&\mathbb{E}\left\{\sum_{k=1}^K\|\mathbf{F}\mathbf{H}_k\mathbf{W}_k\|^2 + \|\mathbf{F}\mathbf{G}_1\mathbf{V}\|^2 + \sigma^2\|\mathbf{F}\|^2\right\} \\ &= \mathrm{tr}\left(\mathbf{F}(\lambda_{\mathbf{H}}(\{\mathbf{W}_k\}_{k=1}^K) + \lambda_{\mathbf{G}_1}(\mathbf{V}) + \sigma^2\mathbf{I})\mathbf{F}^\dagger\right).\end{aligned} \tag{26}$$

So far, we can generalize the MSE minimization problem (8) into a robust optimization problem, which can be formulated as

$$\min_{\mathbf{W}_k,\mathbf{V},\mathbf{F},\mathbf{U}_d} \quad \breve{\mathbf{E}}_d \tag{27a}$$

$$\text{s.t.} \quad \min_{\mathbf{U}_r}\breve{\mathbf{E}}_r \geq \xi, \tag{27b}$$

$$\|\mathbf{V}\|^2 \leq P_a, \tag{27c}$$

$$\mathrm{tr}(\mathbf{F}(\lambda_{\mathbf{H}}(\{\mathbf{W}_k\}_{k=1}^K) + \lambda_{\mathbf{G}_1}(\mathbf{V}) + \sigma^2\mathbf{I})\mathbf{F}^\dagger) \leq P_r, \tag{27d}$$

$$\|\mathbf{W}_k\|^2 \leq P_k, \forall k \in \mathcal{K}. \tag{27e}$$

Since the optimization problem (27) is highly nonconvex due to coupling optimization variables, we will propose an IBCD algorithm to solve it effectively.

A. Optimization of $\{\mathbf{U}_r$ and $\mathbf{U}_d\}$

When $\mathbf{W}_k$,$\mathbf{V}$ and $\mathbf{F}$ are fixed, the optimization problem of $\mathbf{U}_r$ and $\mathbf{U}_d$ is an unconstrained optimization problem, and the optimal solutions of $\mathbf{U}_r$ and $\mathbf{U}_d$ are given by [45]

$$\begin{aligned}\mathbf{U}_d^\star =& \Big(\bar{\mathbf{G}}_2\mathbf{F}(\lambda_{\mathbf{H}}(\{\mathbf{W}_k\}_{k=1}^K) + \lambda_{\Delta\mathbf{G}_1}(\mathbf{V}))\mathbf{F}^\dagger\bar{\mathbf{G}}_2^\dagger \\ &+ \mathrm{tr}\left(\mathbf{F}(\lambda_{\mathbf{H}}(\{\mathbf{W}_k\}_{k=1}^K) + \lambda_{\Delta\mathbf{G}_1}(\mathbf{V}))\mathbf{F}^\dagger\Psi_{G,2}\right)\Sigma_{G,2} \\ &+ \sigma^2\lambda_{\mathbf{G}_2}(\mathbf{F}) + \sigma^2\mathbf{I}\Big)^{-1} \\ &\times\left(\bar{\mathbf{G}}_2\mathbf{F}\sum_{k=1}^K\bar{\mathbf{H}}_k\mathbf{W}_k\right),\end{aligned} \tag{28a}$$

$$\mathbf{U}_r^\star = (\lambda_{\mathbf{H}}(\{\mathbf{W}_k\}_{k=1}^K) + \lambda_{\mathbf{G_1}}(\mathbf{V}) + \sigma^2\mathbf{I})^{-1}\times\left(\sum_{k=1}^K\bar{\mathbf{H}}_k\mathbf{W}_k\right). \tag{28b}$$

B. Optimization of $\{\mathbf{F}\}$

When $\mathbf{W}_k$,$\mathbf{V}$, $\mathbf{U}_r$ and $\mathbf{U}_d$ are fixed, there is only the constraint (27d) that is relevant to the optimization variable $\mathbf{F}$, and the optimization problem of $\mathbf{F}$ can be written as

$$\begin{aligned}\min_{\mathbf{F}} \quad & \mathrm{tr}\Big(\mathbf{F}(\lambda_{\mathbf{H}}(\{\mathbf{W}_k\}_{k=1}^K) + \lambda_{\Delta\mathbf{G}_1}(\mathbf{V}) + \sigma^2\mathbf{I})\mathbf{F}^\dagger \\ & \times\left(\bar{\mathbf{G}}_2^\dagger\mathbf{U}_d\mathbf{U}_d^\dagger\bar{\mathbf{G}}_2 + \mathrm{tr}\left\{\mathbf{U}_d^\dagger\Sigma_{G,2}\mathbf{U}_d\right\}\Psi_{G,2}\right)\Big) \\ & - \sum_{k=1}^K\mathrm{tr}\left(\mathbf{U}_d^\dagger\bar{\mathbf{G}}_2\mathbf{F}\bar{\mathbf{H}}_k\mathbf{W}_k + (\mathbf{U}_d^\dagger\bar{\mathbf{G}}_2\mathbf{F}\bar{\mathbf{H}}_k\mathbf{W}_k)^\dagger\right) \\ & + \mathrm{tr}(\sigma^2\mathbf{U}_d^\dagger\mathbf{U}_d) + \mathrm{tr}(K\mathbf{I})\end{aligned} \tag{29a}$$

$$\text{s.t.} \quad \mathrm{tr}\left(\mathbf{F}(\lambda_{\mathbf{H}}(\{\mathbf{W}_k\}_{k=1}^K) + \lambda_{\mathbf{G}_1}(\mathbf{V}) + \sigma^2\mathbf{I})\mathbf{F}^\dagger\right) \leq P_r. \tag{29b}$$

As in [51], using the property of the Kronecker product $\mathrm{Tr}(\mathbf{ABCD}) = \mathrm{vec}\left(\mathbf{A}^\dagger\right)^\dagger\left(\mathbf{D}^T\otimes\mathbf{B}\right)\mathrm{vec}(\mathbf{C})$ and the diagonal matrix vectorization [52], problem (29) can be further recast as

$$\min_{\mathbf{f}} \ \mathbf{f}^\dagger\breve{\mathbf{P}}\mathbf{f} + 2\Re\{\breve{\mathbf{q}}^\dagger\mathbf{f}\} + \mathrm{tr}(\sigma^2\mathbf{U}_d^\dagger\mathbf{U}_d) + \mathrm{tr}(K\mathbf{I}) \tag{30a}$$

$$\text{s.t.} \ \mathbf{f}^\dagger\breve{\mathbf{R}}\mathbf{f} \leq P_r, \tag{30b}$$

where $\mathbf{f} = \mathrm{vec}(\mathbf{F})$, $\breve{\mathbf{P}}$, $\breve{\mathbf{q}}$ and $\breve{\mathbf{R}}$ are constant in problem (30), and their definitions are given in Appendix D for brevity.

Since problem (30) is a convex quadratically constrained quadratic problem (QCQP), the Lagrange multiplier method can be used to solve it efficiently, and the expression of the optimal $\mathbf{f}$ can be obtained as

$$\mathbf{f}^\star = -(\breve{\mathbf{P}} + \mu\breve{\mathbf{R}})^\ddagger\breve{\mathbf{q}} \tag{31a}$$

$$= -\mathrm{vec}\left((\breve{\mathbf{P}} + \mu\mathbf{I})^\ddagger\left(-\sum_{k=1}^K\bar{\mathbf{H}}_k\mathbf{W}_k\mathbf{U}_d^\dagger\bar{\mathbf{G}}_2\right)^\dagger\breve{\mathbf{R}}^{-1}\right), \tag{31b}$$

where μ is the optimal Lagrange multiplier. Thus, the optimal $\mathbf{F}^\star$ is given by

$$\mathbf{F}^\star = (\breve{\mathbf{P}} + \mu\mathbf{I})^\ddagger\left(-\sum_{k=1}^K\bar{\mathbf{H}}_k\mathbf{W}_k\mathbf{U}_d^\dagger\bar{\mathbf{G}}_2\right)^\dagger\breve{\mathbf{R}}^{-1}. \tag{32}$$

Note that there are two possible cases for the optimal μ. Once $\breve{\mathbf{q}}((\breve{\mathbf{P}})^\ddagger)^\dagger\breve{\mathbf{R}}(\breve{\mathbf{P}})^\ddagger\breve{\mathbf{q}} \leq P_r$ is satisfied, we have $\mu = 0$, otherwise, μ can be obtained by bisection method as in [53].

C. Optimization of $\{\mathbf{W}_k$ and $\mathbf{V}\}$

When $\mathbf{F}$, $\mathbf{U}_r$ and $\mathbf{U}_d$ are given, the problem of optimizing $\mathbf{W}_k$ and $\mathbf{V}$ can be written as

$$\begin{aligned}\min_{\mathbf{W}_k,\mathbf{V}} \quad & \sum_{k=1}^K\mathrm{tr}(\mathbf{W}_k\mathbf{W}_k^\dagger\Xi_k) + \mathrm{tr}(\mathbf{V}\mathbf{V}^\dagger\Xi_v) \\ & - \sum_{k=1}^K\mathrm{tr}\left(\mathbf{U}_d^\dagger\bar{\mathbf{G}}_2\mathbf{F}\bar{\mathbf{H}}_k\mathbf{W}_k + (\mathbf{U}_d^\dagger\bar{\mathbf{G}}_2\mathbf{F}\bar{\mathbf{H}}_k\mathbf{W}_k)^\dagger\right) \\ & + \mathrm{tr}\left(\mathbf{U}_d^\dagger(\sigma^2\lambda_{\mathbf{G}_2}(\mathbf{F}) + \sigma^2\mathbf{I})\mathbf{U}_d\right) + \mathrm{tr}(K\mathbf{I})\end{aligned} \tag{33a}$$

$$\begin{aligned}\text{s.t.} \quad & -\Bigg(\sum_{k=1}^K\mathrm{tr}(\mathbf{W}_k\mathbf{W}_k^\dagger\Gamma_k) + \mathrm{tr}(\mathbf{V}\mathbf{V}^\dagger\Gamma_v) \\ & - \sum_{k=1}^K\mathrm{tr}\left(\mathbf{U}_r^\dagger\bar{\mathbf{H}}_k\mathbf{W}_k + (\mathbf{U}_r^\dagger\bar{\mathbf{H}}_k\mathbf{W}_k)^\dagger\right) \\ & + \mathrm{tr}(\mathbf{U}_r^\dagger\sigma^2\mathbf{U}_r) + \mathrm{tr}\,(K\mathbf{I})\Bigg) + \xi \leq 0,\end{aligned} \tag{33b}$$

$$\sum_{k=1}^K\mathrm{tr}(\mathbf{W}_k\mathbf{W}_k^\dagger\Upsilon_k) + \mathrm{tr}(\mathbf{V}\mathbf{V}^\dagger\Upsilon_v) + \mathrm{tr}(\sigma^2\mathbf{F}\mathbf{F}^\dagger) \leq P_r, \tag{33c}$$

$$\|\mathbf{V}\|^2 \leq P_a, \tag{33d}$$

$$\|\mathbf{W}_k\|^2 \leq P_k, \forall k \in \mathcal{K}, \tag{33e}$$

where Ξ_k, Ξ_v, Γ_k, Γ_v, Υ_k and Υ_v are constant in problem (33), and their definitions are given in Appendix D for brevity.

Note that the constraint (33b) is nonconvex, thus we apply first-order Taylor approximation to tackle the nonconvex terms. In the $(t+1)$-th iteration, we solve

$$\min_{\mathbf{W}_k, \mathbf{V}} \quad (33a) \tag{34a}$$

$$\begin{aligned} \text{s.t.} \quad & -\Bigg(\sum_{k=1}^{K}\Big(\mathrm{tr}(\mathbf{W}_k^{t\dagger}\Gamma_k\mathbf{W}_k^t) + \mathrm{tr}(\mathbf{W}_k^{t\dagger}\Gamma_k(\mathbf{W}_k - \mathbf{W}_k^t)) \\ & +\mathrm{tr}\left((\mathbf{W}_k - \mathbf{W}_k^t)^\dagger\Gamma_k\mathbf{W}_k^t\right)\Big) + \mathrm{tr}\left(\mathbf{V}^{t\dagger}\Gamma_v\mathbf{V}^t\right) \\ & +\mathrm{tr}\left(\mathbf{V}^{t\dagger}\Gamma_v(\mathbf{V} - \mathbf{V}^t)\right) + \mathrm{tr}\left((\mathbf{V} - \mathbf{V}^t)^\dagger\Gamma_v\mathbf{V}^t\right) \\ & -\sum_{k=1}^{K}\mathrm{tr}\left(\mathbf{U}_r^\dagger\bar{\mathbf{H}}_k\mathbf{W}_k + (\mathbf{U}_r^\dagger\bar{\mathbf{H}}_k\mathbf{W}_k)^\dagger\right) \\ & +\mathrm{tr}(\mathbf{U}_r^\dagger\sigma^2\mathbf{U}_r) + \mathrm{tr}(K\mathbf{I})\Bigg) + \xi \leq 0, \end{aligned} \tag{34b}$$

$$(33c) - (33e), \tag{34c}$$

where $\mathbf{W}_k^t$ and $\mathbf{V}^t$ are the optimal solution in the t-th iteration. Incidentally, since problem (34) is convex, the consensus-ADMM [53] could be used to covert it into several subproblems, and in each iteration of consensus-ADMM, each subproblem can be solved via a closed-form solution which is quite similar to equation (31).

The proposed IBCD algorithm for solving the problem (27) is summarized in Algorithm 2.

Algorithm 2 Proposed IBCD Algorithm

```
1: $t \leftarrow 0$;
2: initiate $\mathbf{U}_d^t$, $\mathbf{U}_r^t$, $\mathbf{W}^t$, $\mathbf{V}^t$ and $\mathbf{F}^t$ by a feasible point;
3: repeat
4:     Updata $\mathbf{U}_d^{t+1}$ and $\mathbf{U}_r^{t+1}$ by (28);
5:     Updata $\mathbf{F}^{t+1}$ by (32);
6:     Updata $\mathbf{W}_k^{t+1}$ and $\mathbf{V}^{t+1}$ by solving (34);
7:     $t \leftarrow t+1$;
8: until Convergence;
```

Proposition 3: Algorithm 2 converges to a KKT point.

Proof: See Appendix E. ■

Complexity: The computation burden of the proposed Algorithm 2 mainly comes from solving the convex problem (34), which can be cast as a second-order cone program (SOCP) and takes the complexity of $\mathcal{O}\Big(k_{soc}^{0.5}\Big(m_{soc}^3 + m_{soc}^2 k_{soc} n_{soc} + k_{soc} n_{soc}^2\Big)\log(\frac{1}{\epsilon})\Big)$ with a given accuracy ϵ [54], where k_{soc}, m_{soc} and n_{soc} are the number of SOC constraints, the dimension of the optimization problem and the dimension of each SOC. Thus, the computational complexity of Algorithm 2 is $\mathcal{O}(T_{iter_2}(K+2)^{0.5}[(N_a^2 + KN_s^2)^3 + (N_a^2 + KN_s^2)(K+2)N_{as} + (K+2)N_{as}^2]\log(\frac{1}{\epsilon}))$, where $N_{as} = \max(N_a, N_s)$ and T_{iter_2} is the iteration number of Algorithm 2.

Remark 2: Although the stochastic CSI model appears in many studies [14], [49], [50], bounded CSI model is also commonly adopted. In this scenario, the extension of S-lemma [55], [56] can be applied to convert the power constraints and the relaxed constraint corresponding to the objective function into LMIs. While the conservative approximation [57] can be used to handle the secure MSE constraint.

V. Simulation Results

In this section, we provide the simulation results to evaluate the performances of our proposed algorithms with perfect and imperfect CSI. The wireless channels $\mathbf{H}_k$, $\mathbf{G}_1$ and $\mathbf{G}_2$ are modeled as i.i.d. circularly symmetric complex Gaussian random variables with zero mean and variances $\sigma_H^2 = d_H^{-4}$ and $\sigma_G^2 = d_G^{-4}$, where $d_H = d_G = 1$ [22], [23] denote the corresponding distance from AP to relay and the corresponding distance from relay to sensors. The AP and the relay are equipped with $N_a = N_r = 10$ antennas [43], and the sensors are equipped with $N_s = 4$ antennas [7]. For fair performance comparison with different numbers of sensors K, the performance of the AirComp network is given by the normalized MSE at AP, defined by MSE$/K$ [8], [34], [44]. All simulations are performed in MATLAB R2019b on a Windows desktop with Intel i7 and 8 GB of RAM, and the CVX [58] is used to solve convex optimization problems. The simulation results are averaged over 100 randomly generated channel realizations.

A. The MSE With Perfect CSI

In the perfect CSI scenario, we compare the proposed Algorithm 1 with a scheme without beamforming (denoted as "w/oBF") [3], a scheme of Maximal Ratio Transmission (denoted as "MRT") [59], a scheme that only optimizes sensors' beamforming (denoted as "OSBF"), a scheme that only optimizes the relay's beamforming (denoted as "ORBF") and a scheme that without secure MSE constraint (denoted as "w/oSec"). The "w/oBF", "MRT" "OSBF" and "ORBF" schemes are assumed to use the entire power budget to send the AN signal and successfully cancel the interference caused by AN at AP. The transmit powers at AP, the relay and the sensors are $P_a = P_r = 20$ dB, $P_k = 10$ dB [3], [44], [49], the number of the sensors is $K = 30$ [6], and the minimum secure MSE threshold of the relay is $\xi = 3K$, if not specified.

In Fig. 2, we show the convergence performance of the proposed CCCP Algorithm. From Fig. 2, we see that the proposed algorithm converges as the iteration progresses.

Fig. 3 depicts the normalized MSE at AP versus the number of sensors K. It is observed that the normalized MSE performance of all schemes decreases as K increases except the "ORBF" scheme, and the proposed algorithm outperforms all other schemes. Because the proposed algorithm also optimizes the beamforming of the relay, it achieves better performance than the "OSBF" scheme. When the sensors' beamforming has not been optimized, for any sensor, the signals sent by other sensors can be regarded as interference. As the number of sensors increases, the interference will also increase; this may be the reason why the MSE of the "ORBF" scheme increases as K increases. Also illustrated in the figure, the performances of the "w/oBF" and "MRT" schemes are inferior to other schemes since they do not use the channel state

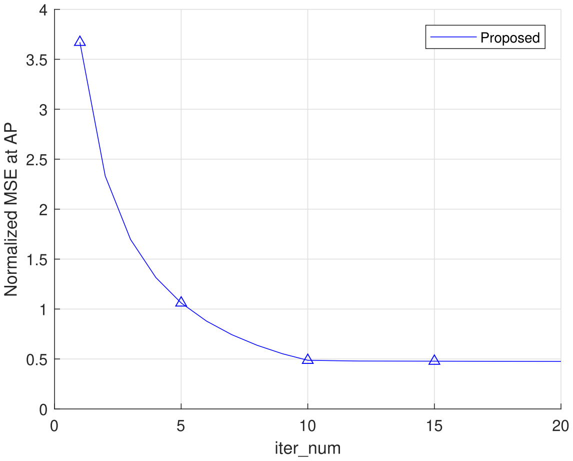

Fig. 2. The convergence performance of the proposed CCCP Algorithm.

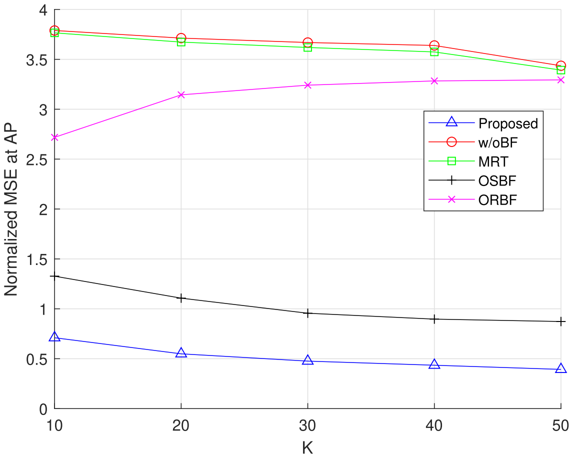

Fig. 3. The normalized MSE at AP under different number of sensors.

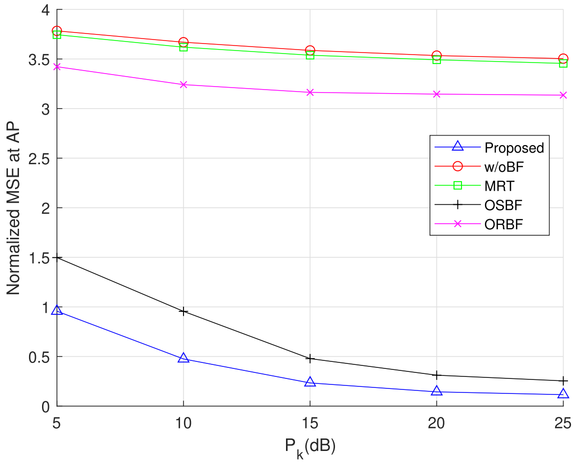

Fig. 4. The normalized MSE at AP under different transmit power of sensors.

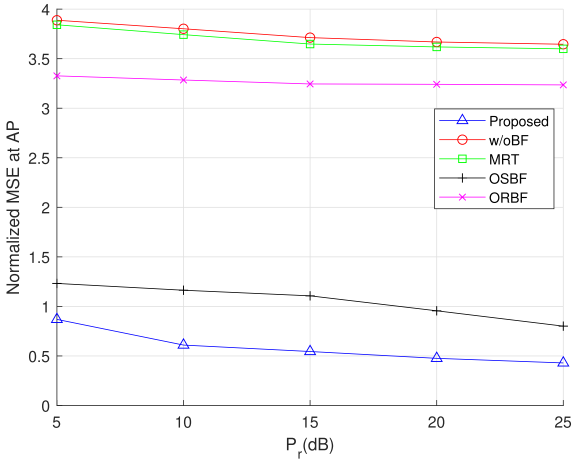

Fig. 5. The normalized MSE at AP under different transmit power of the relay.

information efficiently, since the “w/oBF” scheme does not use any CSI and the “MRT” scheme may not properly utilize the CSI of each node, both “w/oBF” and “MRT” schemes yield poor performance solutions which are close to the upper bound of the MSE at the AP (corresponding to the aggregation beamforming matrix of the AP is set to zero matrix). Moreover, the simulation result also implies that the schemes designed for information communication may not be directly applied to the AirComp network.

Fig. 4 shows the normalized MSE versus the transmit power of sensors P_k. The figure shows that as P_k increases, the normalized MSE obtained by each scheme decreases. That is because a higher transmit power of each sensor P_k leads to a higher transmit SNR, which is beneficial to the aggregation at AP. Among all schemes, the proposed algorithm achieves the best performance. Besides, although the normalized MSE gap between the proposed algorithm and “OSBF” decreases as the transmit power of each sensor grows, the gap is not disappearing. This phenomenon indicates that a larger transmit power of sensors P_k brings a lower MSE at the AP, but the impact introduced by not optimizing relay beamforming cannot be ignored (even when the transmit power of a single sensor P_k is already greater than the transmit power of the relay).

Fig. 5 plots the normalized MSE versus the transmit power of the relay P_r. Racing the transmit power of the relay leads to reduced normalized MSE, according to predicted results, and the proposed algorithm obtains the smallest normalized MSE. In addition, the decline rate of the normalized MSE is lower when the transmit power of the relay P_r is relatively large. At this time, the performance of the AirComp network mainly depends on the transmit power of sensors P_k which is much smaller than the transmit power of the relay P_r.

Fig. 6 shows the normalized MSE versus the secure MSE threshold. With the increase of the secure MSE threshold at the relay, the MSE at the AP obtained by each scheme increases very slowly, which implies that with the help of AN, the MSE at the AP is not significantly affected by the increase of the secure MSE threshold.

Fig. 7 compares the normalized MSE at AP and the secure MSE constraint violation rate of the two schemes with and without secure MSE constraint under the number of the relay antennas. The alternating optimization (AO) method proposed in [22] and [23] is used to solve the optimization problem without secure MSE constraint. The violation rate of a scheme is defined as the proportion of the channels whose solution obtained by the scheme violates the secure MSE constraint to the total number of channel realizations. In Fig. 7, the secure

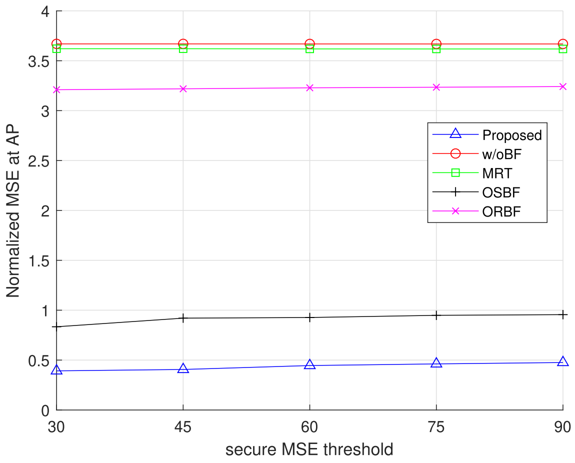

Fig. 6. The normalized MSE at AP under different secure MSE threshold.

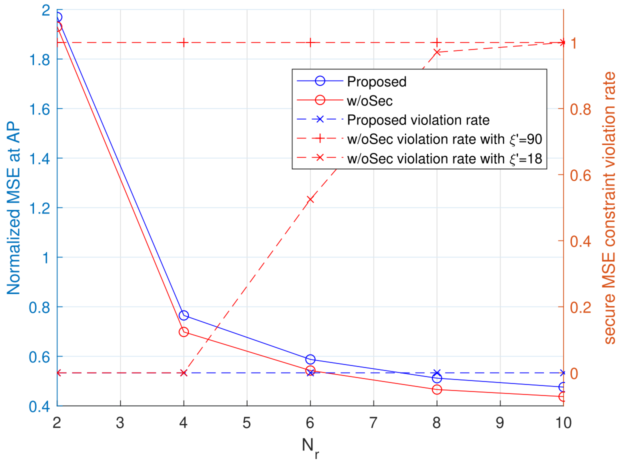

Fig. 7. The normalized MSE at AP and the secure MSE constraint violation rate under different number of relay antennas.

MSE threshold of the proposed algorithm is $\xi = 90$, and we say that for the "w/oSec" scheme, an obtained result violates the secure MSE constraint if the MSE at the relay is greater than ξ'. It can be seen from the figure that the normalized MSE degrades as the number of relay antennas N_r increases. Both schemes have similar MSE performances, which implies that with the help of the AN, the secure MSE constraint does not significantly impact the AP's performance. However, when the secure MSE threshold is set to the same as 90, the violation rate of the "w/oSec" scheme is always one. Even though the secure MSE threshold is quite lower than the proposed scheme, the violation rate of the "w/oSec" scheme increases with the N_r, eventually reaching one. This shows that compared to the proposed algorithm, the existing AirComp techniques that do not consider security may not protect the aggregation of sensors' data from being wiretapped by the untrusted relay.

B. The MSE With Imperfect CSI

In the imperfect CSI scenario, we apply the proposed IBCD Algorithm under different channel uncertainties and compare the proposed algorithm with "OSBF", "ORBF" and sample average approximation (denoted as "SAA") [60] schemes. The "OSBF", "ORBF" and "SAA" schemes are assumed to cancel a part of the interference caused by AN at AP. The row and column covariance matrices of the channel uncertainty are defined as [49]

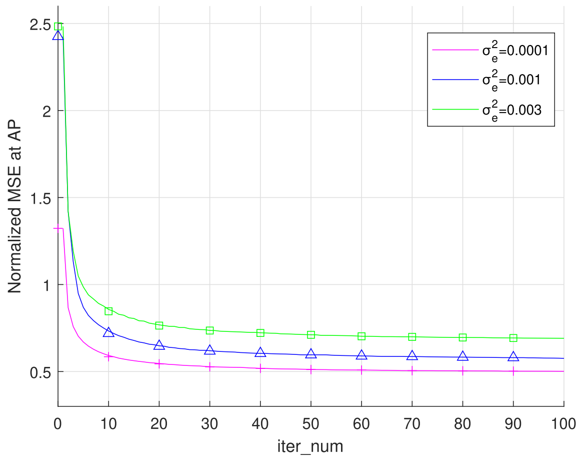

Fig. 8. The convergence performance of the proposed IBCD Algorithm.

$$\Sigma = \begin{bmatrix} 1 & \alpha & \cdots & \alpha^n \\ \alpha & 1 & & \vdots \\ \vdots & & \ddots & \vdots \\ \alpha^n & \cdots & \cdots & 1 \end{bmatrix}, \tag{35a}$$

$$\Psi = \sigma_e^2 \begin{bmatrix} 1 & \beta & \cdots & \beta^m \\ \beta & 1 & & \vdots \\ \vdots & & \ddots & \vdots \\ \beta^m & \cdots & \cdots & 1 \end{bmatrix}, \tag{35b}$$

where $\alpha = 0.6$ and $\beta = 0.5$ are the correlation coefficients, σ_e^2 denotes the estimation error variance, and the dimensions of Σ and Ψ are related to the corresponding channel, to be more specified, the $\mathbf{H}_k$, $\mathbf{G}_1$ and $\mathbf{G}_2$ in (21). Each scheme can apply the successive interference cancellation to eliminate a part of the interference caused by AN. The transmit powers of the AP, the relay and the sensors are $P_a = P_r = 20$ dB, $P_k = 10$ dB [3], [44], [49], the number of the sensors is $K = 30$ [6], the minimum secure MSE threshold of relay is $\xi = 3K$, the estimation error variance is $\sigma_e^2 = 0.001$, and the number of channel realizations of "SAA" scheme is 100, if not specified.

In Fig. 8, we show the convergence performance of the proposed IBCD Algorithm with different channel uncertainties, i.e. $\sigma_e^2 = 0.0001$, 0.001 and 0.003 [49]. It is observed from Fig. 8 that our proposed IBCD Algorithm can converge with the iteration increasing. Under different channel uncertainties, the number of iterations for the proposed IBCD Algorithm to achieve convergence is almost the same. As expected, a greater channel uncertainty leads to a greater normalized MSE.

Fig. 9 presents the normalized MSE at AP versus the number of sensors K with different channel uncertainties. We use the performance in the case of perfect CSI as a benchmark and show it in the figure with the dotted line. It can be observed that all the normalized MSE reduces as the number of sensors increases, which means that more sensing data aggregated at the AP can reduce the normalized computation

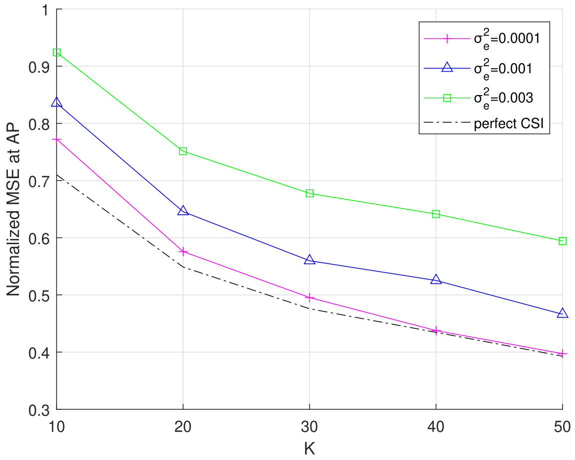

Fig. 9. The normalized MSE at AP versus number of sensors under different channel uncertainties.

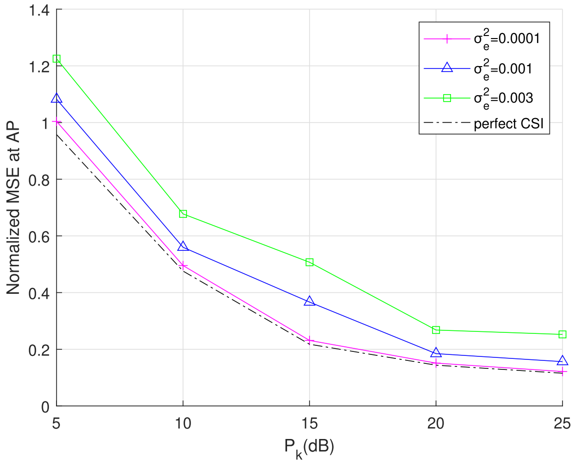

Fig. 10. The normalized MSE at AP versus transmit power of each sensor under different channel uncertainties.

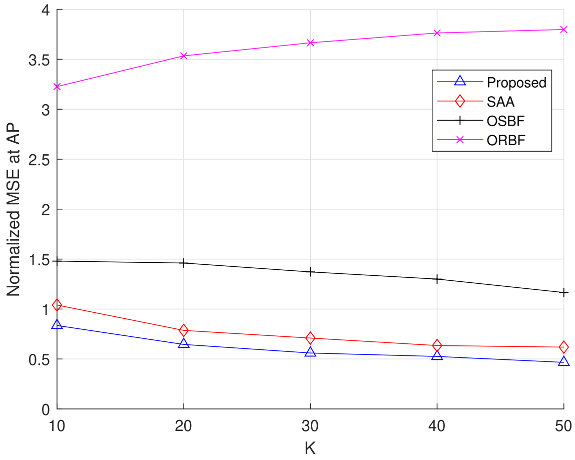

Fig. 11. The normalized MSE at AP versus number of sensors.

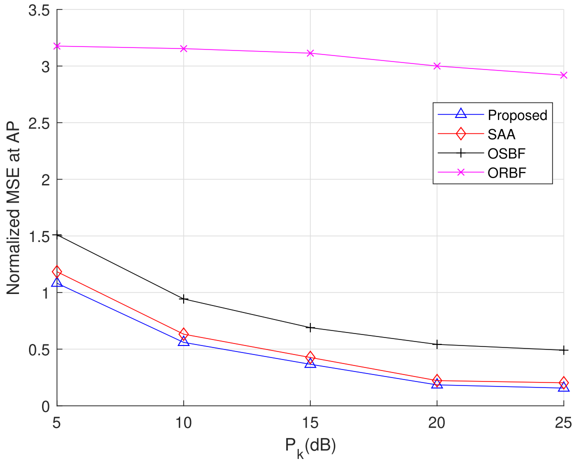

Fig. 12. The normalized MSE at AP versus transmit power of each sensor.

distortion. Meanwhile, a smaller channel uncertainty leads to a smaller normalized MSE. Besides, it can be observed that the normalized MSE obtained under the smallest channel uncertainty is close to the normalized MSE under perfect CSI, especially when the K is relatively large, which indicates that the proposed IBCD Algorithm can also provide a relatively efficient solution. The gap between perfect CSI and imperfect CSI becomes smaller with the increase of K, which may be because the gap between $\mathbf{E}_d$ and $\breve{\mathbf{E}}_d$ (the interference caused by the residual AN) is averaged by the number of sensors K and becomes smaller with the increase of K.

Fig. 10 presents the normalized MSE at AP versus transmit power of each sensor P_k with different channel uncertainties. It can be seen from Fig. 10 that with different channel uncertainty, the normalized MSE decreases as the transmit power P_k increases. When the channel uncertainty is small, the performance of the proposed IBCD algorithm is generally close to the perfect CSI case, which also reflects the effectiveness of the proposed IBCD algorithm.

Fig. 11 shows the normalized MSE at AP versus the number of sensors. Similar to the case of perfect CSI, the normalized MSE performance of each scheme decreases as K increases except the "ORBF" scheme, and the proposed algorithm outperforms all other schemes. The performance of the "SAA" scheme is closer to the proposed algorithm than other schemes, and if the number of channel realizations increases, the gap between the two schemes will be smaller. Nevertheless, this will increase the computation complexity of the "SAA" scheme and consume more system storage resources.

Fig. 12 plots the normalized MSE at AP versus transmit power of each sensor. The normalized MSE performance of all schemes decreases as the transmit power of each sensor P_k increases. The performance of the proposed algorithm is better than all the other schemes.

Fig. 13 depicts the normalized MSE versus the secure MSE threshold. It can be observed that the growth rate of the MSE at the AP with the secure MSE threshold is still very slow, which signifies that the AN still works in the case of imperfect CSI.

Fig. 14 compares the normalized MSE at AP and the secure MSE constraint violation rate of the two schemes with and without secure MSE constraint versus the number of the relay antennas. The definition of violation rate is the same as in the case of perfect CSI. The secure MSE threshold of the proposed algorithm is $\xi = 90$, and the threshold of the "w/oSec" scheme is $\xi' = 90$ or 22.5. It can be seen that the MSE at the AP of "w/oSec" scheme is slightly lower than that of the proposed

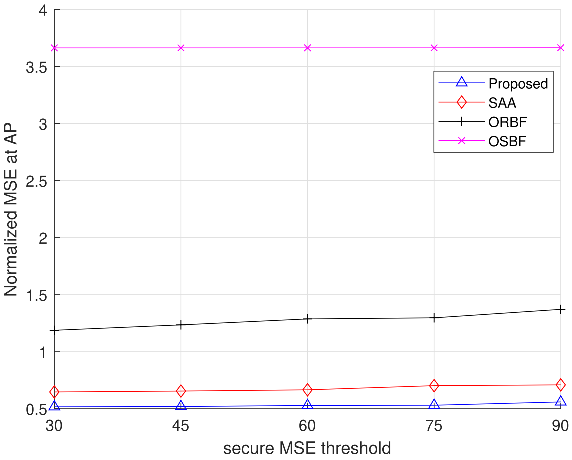

Fig. 13. The normalized MSE at AP versus the secure MSE threshold.

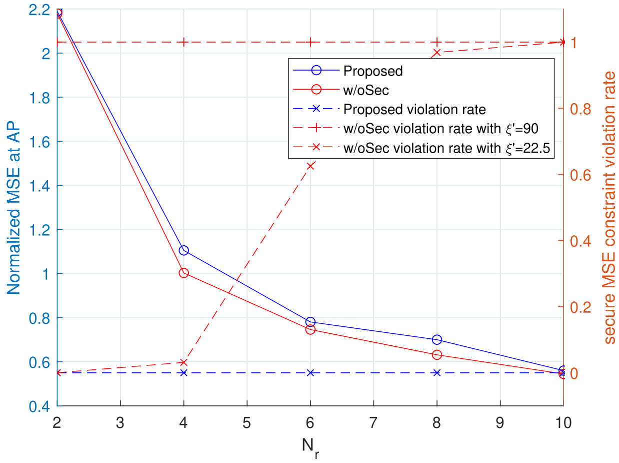

Fig. 14. The normalized MSE at AP and the secure MSE constraint violation rate versus the number of the relay antennas.

scheme. However, the “w/oSec” scheme still cannot guarantee that the MSE at the relay is below a given threshold in the case of imperfect CSI, even when the threshold is small.

VI. Conclusion

In this paper, we have studied the physical layer security of an untrusted relay assisted AirComp network. To protect the sensors’ data aggregation from being wiretapped by the untrusted relay, AN is applied to ensure that the MSE at the relay is greater than a given threshold. We propose a CCCP based iterative algorithm for the case of perfect CSI and propose an IBCD Algorithm for the case of imperfect CSI. Numerical results show that the proposed scheme is superior to other existing schemes, and AN can help improve the security of the sensors’ data aggregation, indirectly improving the AirComp network’s performance.

Appendix A
Proof of Lemma 1, Lemma 2 and Lemma 3

First, we prove lemma 1. Applying the Schur complement to (12), we have

$$\begin{bmatrix} \acute{\mathbf{X}} & \acute{\mathbf{Y}} & \acute{\mathbf{A}} \\ \acute{\mathbf{Y}}^\dagger & \acute{\mathbf{S}} & \acute{\mathbf{C}}^\dagger\acute{\mathbf{B}}^\dagger \\ \acute{\mathbf{A}}^\dagger & \acute{\mathbf{B}}\acute{\mathbf{C}} & \mathbf{I} \end{bmatrix} \succeq 0 \tag{36a}$$

$$\Leftrightarrow \begin{bmatrix} \acute{\mathbf{X}} & \acute{\mathbf{Y}} \\ \acute{\mathbf{Y}}^\dagger & \acute{\mathbf{S}} \end{bmatrix} - \begin{bmatrix} \acute{\mathbf{A}} \\ \acute{\mathbf{C}}^\dagger\acute{\mathbf{B}}^\dagger \end{bmatrix} \begin{bmatrix} \acute{\mathbf{A}}^\dagger & \acute{\mathbf{B}}\acute{\mathbf{C}} \end{bmatrix} \succeq 0 \tag{36b}$$

$$\Leftrightarrow \begin{bmatrix} \acute{\mathbf{X}} - \acute{\mathbf{A}}\acute{\mathbf{A}}^\dagger & \acute{\mathbf{Y}} - \acute{\mathbf{A}}\acute{\mathbf{B}}\acute{\mathbf{C}} \\ \acute{\mathbf{Y}}^\dagger - \acute{\mathbf{C}}^\dagger\acute{\mathbf{B}}^\dagger\acute{\mathbf{A}}^\dagger & \acute{\mathbf{S}} - \acute{\mathbf{C}}^\dagger\acute{\mathbf{B}}^\dagger\acute{\mathbf{B}}\acute{\mathbf{C}} \end{bmatrix} \succeq 0. \tag{36c}$$

Combining $\mathrm{tr}\left(\acute{\mathbf{X}} - \acute{\mathbf{A}}\acute{\mathbf{A}}^\dagger\right) \leq 0$ with (36c), we have $\acute{\mathbf{X}} = \acute{\mathbf{A}}\acute{\mathbf{A}}^\dagger$, which indicates that $\acute{\mathbf{Y}} = \acute{\mathbf{A}}\acute{\mathbf{B}}\acute{\mathbf{C}}$.

Then, we prove lemma 2. Applying the Schur complement again,

$$\begin{bmatrix} \acute{\mathbf{X}} & \acute{\mathbf{A}}^\dagger \\ \acute{\mathbf{A}} & \acute{\mathbf{B}} \end{bmatrix} \succeq 0 \tag{37a}$$

$$\Leftrightarrow \acute{\mathbf{X}} - \acute{\mathbf{A}}^\dagger\acute{\mathbf{B}}^{-1}\acute{\mathbf{A}} \succeq 0. \tag{37b}$$

Combining $\mathrm{tr}\left(\acute{\mathbf{X}} - \acute{\mathbf{A}}^\dagger\acute{\mathbf{B}}^{-1}\acute{\mathbf{A}}\right) \leq 0$ with (37b), we have

$$\mathrm{tr}\left(\acute{\mathbf{X}} - \acute{\mathbf{A}}^\dagger\acute{\mathbf{B}}^{-1}\acute{\mathbf{A}}\right) = 0. \tag{38}$$

With (37b) and (38), we have $\acute{\mathbf{X}} = \acute{\mathbf{A}}^\dagger\acute{\mathbf{B}}^{-1}\acute{\mathbf{A}}$.

Finally, we prove lemma 3. The epigraph of the function $\mathrm{tr}(\acute{\mathbf{X}}^\dagger\acute{\mathbf{Y}}^{-1}\acute{\mathbf{X}})$ can be expressed as

$$\begin{aligned} &\mathbf{epi}\ \mathrm{tr}(\acute{\mathbf{X}}^\dagger\acute{\mathbf{Y}}^{-1}\acute{\mathbf{X}}) && (39a)\\ &= \left\{(\acute{\mathbf{X}}, \acute{\mathbf{Y}}, t) | \acute{\mathbf{Y}} \succ 0, \mathrm{tr}(\acute{\mathbf{X}}^\dagger\acute{\mathbf{Y}}^{-1}\acute{\mathbf{X}}) \leq t\right\} && (39b)\\ &= \left\{(\acute{\mathbf{X}}, \acute{\mathbf{Y}}, t) | \acute{\mathbf{Y}} \succ 0, \mathbf{vec}^\dagger(\acute{\mathbf{X}}^\dagger)(\acute{\mathbf{Y}}^{-T} \otimes \mathbf{I})\mathbf{vec}(\acute{\mathbf{X}}^\dagger) \leq t\right\} && (39c)\\ &= \left\{(\acute{\mathbf{X}}, \acute{\mathbf{Y}}, t) | \acute{\mathbf{Y}} \succ 0, \begin{bmatrix} \acute{\mathbf{Y}}^T \otimes \mathbf{I} & \mathbf{vec}(\acute{\mathbf{X}}^\dagger) \\ \mathbf{vec}^\dagger(\acute{\mathbf{X}}^\dagger) & t \end{bmatrix} \succ 0\right\}, && (39d) \end{aligned}$$

where (39c) uses the property of the Kronecker product $\mathrm{tr}(\mathbf{ABCD}) = \mathrm{vec}\left(\mathbf{A}^\dagger\right)^\dagger \left(\mathbf{D}^T \otimes \mathbf{B}\right) \mathrm{vec}(\mathbf{C})$ [51], (39d) uses $(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}$ and the Schur complement [51]. Note that, in (39d), the condition is an LMI in $(\acute{\mathbf{X}}, \acute{\mathbf{Y}}, t)$, thus the epigraph is convex, which means the function $\mathrm{tr}(\acute{\mathbf{X}}^\dagger\acute{\mathbf{Y}}^{-1}\acute{\mathbf{X}})$ is convex [46].

Appendix B
Proof of Proposition 1

For notational simplicity, denote the optimal value of problem (17) and (18) as η_1 and η_2, respectively. Denote $\Theta_p = (\mathbf{A}_r, \mathbf{B}_r, \mathbf{A}_d, \mathbf{Z}, \mathbf{B}_d, \mathbf{T})$, $\Theta_n = \{\mathbf{X}, \mathbf{V}, \mathbf{Y}, \mathbf{F}\}$, and

$$\mathbf{f}_p(\Theta_p) = \mathrm{tr}\left(\mathbf{A}_r + \mathbf{B}_r + \mathbf{A}_d + \mathbf{Z} + \mathbf{B}_d + \mathbf{T}\right), \tag{40a}$$

$$\begin{aligned} \mathbf{f}_n(\Theta_n) = \mathrm{tr}\Big(&\mathbf{X}\mathbf{X}^\dagger + \mathbf{G}_1\mathbf{V}\mathbf{V}^\dagger\mathbf{G}_1^\dagger + \mathbf{Y}\mathbf{Y}^\dagger \\ &+ (1+\sigma^2)\mathbf{G}_2\mathbf{F}\mathbf{F}^\dagger\mathbf{G}_2^\dagger + \mathbf{F}\mathbf{F}^\dagger\Big). \end{aligned} \tag{40b}$$

Since the feasible set of problem (17) is a subset of the feasible set of problem (18), we have $\eta_2 \leq \eta_1$, that is, η_1 is an upper bound on the problem (18).

Then we show that $\eta_2 \geq \eta_1$. To prove that, we only need to prove that when $\tau \geq \tau_l$, the optimal solution $\left(\Theta_p^\tau, \Theta_n^\tau\right)$ of the problem (18) is a feasible solution to the problem (17). That is to say, $\mathbf{f}_p(\Theta_p^\tau) - \mathbf{f}_n(\Theta_n^\tau) \leq 0$ is a necessary condition. We prove this by contradiction. Assume the τ_l satisfying the conditions in Proposition 1 does not exist, which means that $\mathbf{f}_p(\Theta_p^\tau) - \mathbf{f}_n(\Theta_n^\tau) > 0$ for any positive τ. Thus, when $\tau \to +\infty$, we have $\tau(\mathbf{f}_p(\Theta_p^\tau) - \mathbf{f}_n(\Theta_n^\tau))) \to +\infty$, which indicates that the optimal value of problem (18) is unbounded. This contradicts the fact that problem (17) is bounded and $\eta_2 \leq \eta_1$.

Appendix C
Proof of Proposition 2

For ease of exposition, in the t-th iteration, denote the objective value and solution of problem (19) as η^t and $\Omega^t = \left\{\{\mathbf{W}_k^t\}_{k=1}^K, \mathbf{V}^t, \mathbf{F}^t, \Theta^t\right\}$, and denote the objective function of problem (18) and problem (19) as $\mathbf{f}(\Omega)$ and $\tilde{\mathbf{f}}(\Omega;\Omega^t)$, respectively. Assume that the feasible set of each optimization problem is not empty.

First, we show that the objective value of problem (19) obtained by Algorithm 1 converges as the iteration of the CCCP proceeds. With the property of the first order Taylor expansion, it is easy to verify that Ω^t is a feasible solution of the optimal solution in $(t+1)$-th iteration, and the corresponding objective value is η^t, which means that the objective value obtained in $(t+1)$-th iteration is not greater than η^t, i.e., $\eta^{t+1} \leq \eta^t$. In short, the objective value is non-increase as the iteration of the CCCP proceeds. Meanwhile, since the feasible set of the problem (19) is compact, the objective value is bounded. Thus, Algorithm 1 produces a convergent non-descending objective value sequence.

Then, we show that Algorithm 1 converges to a stationary point. Since the objective function of problem (19) is strictly convex, the solution of it is unique [46], which means the entries of the two sequences $\{\eta^t\}$ and $\{\Omega^t\}$ are one-to-one correspondence. Therefore, Ω^t is convergence. Denote the convergence point as $\Omega^\star = \lim_{t\to\infty}\Omega^t$.

Since $\tilde{\mathbf{f}}(\Omega;\Omega^t)$ is constructed by the first order Taylor expansion, for any t-th iteration, it is easy to verify that

$$\mathbf{f}(\Omega) \leq \tilde{\mathbf{f}}(\Omega;\Omega^t), \tag{41a}$$

$$\mathbf{f}(\Omega^t) = \tilde{\mathbf{f}}(\Omega^t;\Omega^t), \tag{41b}$$

$$\nabla_\Omega \mathbf{f}(\Omega) = \nabla_\Omega \tilde{\mathbf{f}}(\Omega;\Omega^t), \tag{41c}$$

where the gradient in (41c) is w.r.t Ω. The equations in (41) fulfill the properties (P1), (P2), and (P3) in [61], and a constraint qualification condition named linear independence constraint qualification (LICQ) [61] is satisfied for the problem (19) naturally. Therefore, according to the result (R2) in [61], the limit point $\Omega^\star$ is a stationary point of the problem (18).

Finally, we show that $\Omega^\star$ is a local minimum. In general, a stationary point could be a saddle point, a local maximum, or a local minimum of a nonlinear program [62]. According to [63], Algorithm 1 will converge to either a local optimum or a saddle point of the problem (18). Because the objective of the problem (19) is twice-continuously differentiable and strictly convex, Algorithm 1 will not converge to a saddle point [46] but a local optimal solution, i.e., $\Omega^\star$ is a local minimum.

Appendix D
The Definitions of Some Expectation Terms and Coefficients

A. The Definitions of the Expectation Terms in $\breve{\mathbf{E}}_d$

$$\begin{aligned}&\lambda_{\mathbf{H}}(\{\mathbf{W}_k\}_{k=1}^K)\\ &\quad\triangleq \mathbb{E}_{\Delta\mathbf{H}}\left\{\sum_{k=1}^K(\mathbf{H}_k\mathbf{W}_k)(\mathbf{H}_k\mathbf{W}_k)^\dagger\right\}\\ &\quad= \sum_{k=1}^K\left(\bar{\mathbf{H}}_k\mathbf{W}_k\mathbf{W}_k^\dagger\bar{\mathbf{H}}_k^\dagger + \mathrm{tr}(\mathbf{W}_k\mathbf{W}_k^\dagger\Psi_{H,k})\Sigma_{H,k}\right).\end{aligned} \tag{42}$$

$$\begin{aligned}\lambda_{\Delta\mathbf{G}_1}(\mathbf{V}) &\triangleq \mathbb{E}_{\Delta\mathbf{G}_1}\left\{\Delta\mathbf{G}_1\mathbf{V}\mathbf{V}^\dagger\Delta\mathbf{G}_1^\dagger\right\}\\ &= \mathrm{tr}\left(\mathbf{V}\mathbf{V}^\dagger\Psi_{G,1}\right)\Sigma_{G,1}.\end{aligned} \tag{43}$$

$$\begin{aligned}\lambda_{\mathbf{G}_1}(\mathbf{V}) &\triangleq \mathbb{E}_{\Delta\mathbf{G}_1}\left\{\mathbf{G}_1\mathbf{V}\mathbf{V}^\dagger\mathbf{G}_1^\dagger\right\}\\ &= \bar{\mathbf{G}}_1\mathbf{V}\mathbf{V}^\dagger\bar{\mathbf{G}}_1^\dagger + \mathrm{tr}\left(\mathbf{V}\mathbf{V}^\dagger\Psi_{G,1}\right)\Sigma_{G,1}.\end{aligned} \tag{44}$$

$$\begin{aligned}\lambda_{\mathbf{G}_2}(\mathbf{F}) &\triangleq \mathbb{E}_{\Delta\mathbf{G}_2}\left\{\mathbf{G}_2\mathbf{F}\mathbf{F}^\dagger\mathbf{G}_2^\dagger\right\}\\ &= \bar{\mathbf{G}}_2\mathbf{F}\mathbf{F}^\dagger\bar{\mathbf{G}}_2^\dagger + \mathrm{tr}\left(\mathbf{F}\mathbf{F}^\dagger\Psi_{G,2}\right)\Sigma_{G,2}.\end{aligned} \tag{45}$$

B. The Definitions of Coefficients in Problem (30)

$$\begin{aligned}\breve{\mathbf{P}} &= (\lambda_{\mathbf{H}}(\{\mathbf{W}_k\}_{k=1}^K) + \lambda_{\Delta\mathbf{G}_1}(\mathbf{V}) + \sigma^2\mathbf{I})^T\\ &\quad\otimes\left(\bar{\mathbf{G}}_2^\dagger\mathbf{U}_d\mathbf{U}_d^\dagger\bar{\mathbf{G}}_2 + \mathrm{tr}\left(\mathbf{U}_d^\dagger\Sigma_{G,2}\mathbf{U}_d\right)\Psi_{G,2}\right).\end{aligned} \tag{46}$$

$$\breve{\mathbf{q}} = \mathbf{vec}\left(\left(-\sum_{k=1}^K\bar{\mathbf{H}}_k\mathbf{W}_k\mathbf{U}_d^\dagger\bar{\mathbf{G}}_2\right)^\dagger\right). \tag{47}$$

$$\breve{\mathbf{R}} = (\lambda_{\mathbf{H}}(\{\mathbf{W}_k\}_{k=1}^K) + \lambda_{\mathbf{G}_1}(\mathbf{V}) + \sigma^2\mathbf{I})^T \otimes \mathbf{I}. \tag{48}$$

C. The Definitions of Coefficients in Problem (33)

$$\begin{aligned}&\Xi_k\\ &\;= \bar{\mathbf{H}}_k^\dagger\mathbf{F}^\dagger\bar{\mathbf{G}}_2^\dagger\mathbf{U}_d\mathbf{U}_d^\dagger\bar{\mathbf{G}}_2\mathbf{F}\bar{\mathbf{H}}_k\\ &\quad+\ \mathrm{tr}\left(\mathbf{U}_d^\dagger\Sigma_{G,2}\mathbf{U}_d\right)\bar{\mathbf{H}}_k^\dagger\mathbf{F}^\dagger\Psi_{G,2}\mathbf{F}\bar{\mathbf{H}}_k\\ &\quad+\mathrm{tr}\left(\mathbf{U}_d^\dagger\left(\bar{\mathbf{G}}_2\mathbf{F}\Sigma_{H,k}\mathbf{F}^\dagger\bar{\mathbf{G}}_2^\dagger+\mathrm{tr}\{\mathbf{F}\Sigma_{H,k}\mathbf{F}^\dagger\Psi_{G,2}\}\Sigma_{G,2}\right)\mathbf{U}_d\right)\\ &\quad\times\ \Psi_{H,k}.\end{aligned} \tag{49}$$

$$\begin{aligned}&\Xi_v\\ &\;= \mathrm{tr}\left(\mathbf{U}_d^\dagger\left(\bar{\mathbf{G}}_2\mathbf{F}\Sigma_{G,1}\mathbf{F}^\dagger\bar{\mathbf{G}}_2^\dagger + \mathrm{tr}\{\mathbf{F}\Sigma_{G,1}\mathbf{F}^\dagger\Psi_{G,2}\}\Sigma_{G,2}\right)\mathbf{U}_d\right)\\ &\quad\times\ \Psi_{G,1}.\end{aligned} \tag{50}$$

$$\Gamma_k = \bar{\mathbf{H}}_k^\dagger\mathbf{U}_r\mathbf{U}_r^\dagger\bar{\mathbf{H}}_k + \mathrm{tr}\left(\mathbf{U}_r^\dagger\Sigma_{H,k}\mathbf{U}_r\right)\Psi_{H,k}. \tag{51}$$

$$\Gamma_v = \bar{\mathbf{G}}_1^\dagger\mathbf{U}_r\mathbf{U}_r^\dagger\bar{\mathbf{G}}_1 + \mathrm{tr}\left(\mathbf{U}_r^\dagger\Sigma_{G,1}\mathbf{U}_r\right)\Psi_{G,1}. \tag{52}$$

$$\Upsilon_k = \bar{\mathbf{H}}_k^\dagger\mathbf{F}^\dagger\mathbf{F}\bar{\mathbf{H}}_k + \mathrm{tr}\left(\mathbf{F}\Sigma_{H,k}\mathbf{F}^\dagger\right)\Psi_{H,k}. \tag{53}$$

$$\Upsilon_v = \bar{\mathbf{G}}_1^\dagger\mathbf{F}^\dagger\mathbf{F}\bar{\mathbf{G}}_1 + \mathrm{tr}\left(\mathbf{F}\Sigma_{G,1}\mathbf{F}^\dagger\right)\Psi_{G,1}. \tag{54}$$

Appendix E
Proof of Proposition 3

For simplicity, in the $t-$th iteration, denote the left hand side of the constraint (33b) as $g(\{\mathbf{W}_k\}_{k=1}^K, \mathbf{V})$, the left hand side of the constraint (34b) as $g(\{\mathbf{W}_k\}_{k=1}^K, \mathbf{V}; \{\mathbf{W}_k^t\}_{k=1}^K, \mathbf{V}^t)$, the solution set of problem (34) as $\mathbb{S}(\mathbf{U}_r^t, \mathbf{F}^t, \{\mathbf{W}_k^t\}_{k=1}^K, \mathbf{V}^t)$, the objective function of problem (27) as $f(\mathbf{U}_d, \mathbf{F}, \{\mathbf{W}_k\}_{k=1}^K, \mathbf{V})$. The limit point of the IBCD algorithm is denoted as $(\mathbf{U}_d^\star, \mathbf{F}^\star, \{\mathbf{W}_k^\star\}_{k=1}^K, \mathbf{V}^\star) = \lim_{t\to\infty}(\mathbf{U}_d^t, \mathbf{F}^t, \{\mathbf{W}_k^t\}_{k=1}^K, \mathbf{V}^t)$. Assume that the feasible set of each optimization problem is not empty.

First, we show that $(\{\mathbf{W}_k\}_{k=1}^K, \mathbf{V})$ obtained in the $(t+1)$-th iteration is feasible to the problem (27). Assume $(\{\mathbf{W}_k^t\}_{k=1}^K, \mathbf{V}^t)$ is feasible. With the property of the first order Taylor expansion, we can easily verify that $g(\{\mathbf{W}_k\}_{k=1}^K, \mathbf{V}) \le g(\{\mathbf{W}_k\}_{k=1}^K, \mathbf{V}; \{\mathbf{W}_k^t\}_{k=1}^K, \mathbf{V}^t)$. When the feasible set of (34) is not empty, any solution of (34) satisfies that $g(\{\mathbf{W}_k\}_{k=1}^K, \mathbf{V}; \{\mathbf{W}_k^t\}_{k=1}^K, \mathbf{V}^t) \le 0$, then we have $g(\{\mathbf{W}_k\}_{k=1}^K, \mathbf{V}) \le 0$, which means $(\{\mathbf{W}_k\}_{k=1}^K, \mathbf{V})$ satisfies the constraint (33b) and is a feasible point of problem (27).

Next, we show that the objective value of problem (27) w.r.t $(\{\mathbf{W}_k\}_{k=1}^K, \mathbf{V})$ is monotonically convergent as IBCD proceeds. As follows,

$$
\begin{aligned}
& f(\mathbf{U}_d^{t+1}, \mathbf{F}^{t+1}, \{\mathbf{W}_k^{t+1}\}_{k=1}^K, \mathbf{V}^{t+1}) && \text{(55a)}\\
&\quad \le f(\mathbf{U}_d^{t+1}, \mathbf{F}^{t+1}, \{\mathbf{W}_k^t\}_{k=1}^K, \mathbf{V}^t) && \text{(55b)}\\
&\quad \le f(\mathbf{U}_d^{t+1}, \mathbf{F}^t, \{\mathbf{W}_k^t\}_{k=1}^K, \mathbf{V}^t) && \text{(55c)}\\
&\quad \le f(\mathbf{U}_d^t, \mathbf{F}^t, \{\mathbf{W}_k^t\}_{k=1}^K, \mathbf{V}^t), && \text{(55d)}
\end{aligned}
$$

where the first inequality is due to the property of the first order Taylor expansion mentioned above, and the second and third inequality are due to Step 5 and Step 4 in Algorithm 2, respectively. Since $\mathbf{f}(\mathbf{U}_d, \mathbf{F}, \{\mathbf{W}_k\}_{k=1}^K, \mathbf{V})$ is bounded and continuity w.r.t $(\{\mathbf{W}_k\}_{k=1}^K, \mathbf{V})$, the inequalities in (55) lead to the monotonic convergence of the objective value w.r.t $(\{\mathbf{W}_k\}_{k=1}^K, \mathbf{V})$.

Then, we show the limit point $(\{\mathbf{W}_k^\star\}_{k=1}^K, \mathbf{V}^\star) \in \mathbb{S}(\mathbf{U}_r^\star, \mathbf{F}^\star, \{\mathbf{W}_k^\star\}_{k=1}^K, \mathbf{V}^\star)$.

As the definition of $(\mathbf{U}_d^\star, \mathbf{F}^\star, \{\mathbf{W}_k^\star\}_{k=1}^K, \mathbf{V}^\star)$, there exist a convergent subsequence $(\mathbf{U}_d^{t_j}, \mathbf{F}^{t_j}, \{\mathbf{W}_k^{t_j}\}_{k=1}^K, \mathbf{V}^{t_j})$ such that $\lim_{j\to\infty}(\{\mathbf{W}_k^{t_j}\}_{k=1}^K, \mathbf{V}^{t_j}) = (\{\mathbf{W}_k^\star\}_{k=1}^K, \mathbf{V}^\star)$. Since the feasible set of problem (27) is compact, by restricting to a subsequence, it is reasonable to assume that $(\{\mathbf{W}_k^{t_j}\}_{k=1}^K, \mathbf{V}^{t_j})$ converges to a limit point $(\{\mathbf{W}_k^{\star\star}\}_{k=1}^K, \mathbf{V}^{\star\star})$.

Define the constraint set of problem (34) as $\mathbb{C}_\le(\{\mathbf{W}_k^t\}_{k=1}^K, \mathbf{V}^t)$, and define another set as follow

$$
\begin{aligned}
&\mathbb{C}_<(\{\mathbf{W}_k^t\}_{k=1}^K, \mathbf{V}^t)\\
&\quad = \Big\{(\{\mathbf{W}_k\}_{k=1}^K, \mathbf{V}) | g(\{\mathbf{W}_k\}_{k=1}^K, \mathbf{V}; \{\mathbf{W}_k^t\}_{k=1}^K, \mathbf{V}^t) < 0,\\
&\quad (33c) - (33e)\Big\}. \qquad (56)
\end{aligned}
$$

Obviously, $\mathbb{C}_<(\{\mathbf{W}_k^t\}_{k=1}^K, \mathbf{V}^t) \subset \mathbb{C}_\le(\{\mathbf{W}_k^t\}_{k=1}^K, \mathbf{V}^t)$. Since $g(\{\mathbf{W}_k\}_{k=1}^K, \mathbf{V}; \{\mathbf{W}_k^t\}_{k=1}^K, \mathbf{V}^t)$ is continuous and $\lim_{j\to\infty}$ $(\{\mathbf{W}_k^{t_j}\}_{k=1}^K, \mathbf{V}^{t_j}) = (\{\mathbf{W}_k^\star\}_{k=1}^K, \mathbf{V}^\star)$, for any fixed $(\{\mathbf{W}_k\}_{k=1}^K, \mathbf{V}) \in \mathbb{C}_<(\{\mathbf{W}_k^\star\}_{k=1}^K, \mathbf{V}^\star)$, there exist a large enough integer I_t such that

$$
g(\{\mathbf{W}_k\}_{k=1}^K, \mathbf{V}; \{\mathbf{W}_k^{t_j}\}_{k=1}^K, \mathbf{V}^{t_j}) < 0, \forall j \ge I_t, \qquad (57)
$$

This implies that

$$
\begin{aligned}
&\mathbb{C}_<(\{\mathbf{W}_k^\star\}_{k=1}^K, \mathbf{V}^\star)\\
&\quad \subseteq \mathbb{C}_<(\{\mathbf{W}_k^{t_j}\}_{k=1}^K, \mathbf{V}^{t_j})\\
&\quad \subset \mathbb{C}_\le(\{\mathbf{W}_k^{t_j}\}_{k=1}^K, \mathbf{V}^{t_j}). \qquad (58)
\end{aligned}
$$

Since $(\{\mathbf{W}_k^{t_j+1}\}_{k=1}^K, \mathbf{V}^{t_j+1})$ is the optimal solution in the t_j-th iteration, we have

$$
\begin{aligned}
&f(\mathbf{U}_d^{t_j+1}, \mathbf{F}^{t_j+1}, \{\mathbf{W}_k\}_{k=1}^K, \mathbf{V})\\
&\quad \ge f(\mathbf{U}_d^{t_j+1}, \mathbf{F}^{t_j+1}, \{\mathbf{W}_k^{t_j+1}\}_{k=1}^K, \mathbf{V}^{t_j+1}),\\
&\qquad \forall(\{\mathbf{W}_k\}_{k=1}^K, \mathbf{V}) \in \mathbb{C}_<(\{\mathbf{W}_k^\star\}_{k=1}^K, \mathbf{V}^\star)\\
&\quad \subset \mathbb{C}_\le(\{\mathbf{W}_k^{t_j}\}_{k=1}^K, \mathbf{V}^{t_j}). \qquad (59)
\end{aligned}
$$

And because $f(\cdot)$ is continuity w.r.t $(\{\mathbf{W}_k\}_{k=1}^K, \mathbf{V})$, let $j \to \infty$, we have

$$
\begin{aligned}
&f(\mathbf{U}_d^\star, \mathbf{F}^\star, \{\mathbf{W}_k\}_{k=1}^K, \mathbf{V}) \ge f(\mathbf{U}_d^\star, \mathbf{F}^\star, \{\mathbf{W}_k^{\star\star}\}_{k=1}^K, \mathbf{V}^{\star\star}),\\
&\quad \forall(\{\mathbf{W}_k\}_{k=1}^K, \mathbf{V}) \in \mathbb{C}_<(\{\mathbf{W}_k^\star\}_{k=1}^K, \mathbf{V}^\star). \qquad (60)
\end{aligned}
$$

With the continuity of $g(\{\mathbf{W}_k\}_{k=1}^K, \mathbf{V}; \{\mathbf{W}_k^t\}_{k=1}^K, \mathbf{V}^t)$, we have

$$
\begin{aligned}
&f(\mathbf{U}_d^\star, \mathbf{F}^\star, \{\mathbf{W}_k\}_{k=1}^K, \mathbf{V}) \ge f(\mathbf{U}_d^\star, \mathbf{F}^\star, \{\mathbf{W}_k^{\star\star}\}_{k=1}^K, \mathbf{V}^{\star\star})\\
&\quad \forall(\{\mathbf{W}_k\}_{k=1}^K, \mathbf{V}) \in \mathbb{C}_\le(\{\mathbf{W}_k^\star\}_{k=1}^K, \mathbf{V}^\star). \qquad (61)
\end{aligned}
$$

Combining with (55), we have

$$
f(\mathbf{U}_d^\star, \mathbf{F}^\star, \{\mathbf{W}_k^\star\}_{k=1}^K, \mathbf{V}^\star) = f(\mathbf{U}_d^\star, \mathbf{F}^\star, \{\mathbf{W}_k^{\star\star}\}_{k=1}^K, \mathbf{V}^{\star\star}). \quad (62)
$$

With the fact that $(\{\mathbf{W}_k^{t_j}\}_{k=1}^K, \mathbf{V}^{t_j})$ is feasible to problem (27) and $g(\{\mathbf{W}_k^{t_j}\}_{k=1}^K, \mathbf{V}^{t_j}) = g(\{\mathbf{W}_k^{t_j}\}_{k=1}^K, \mathbf{V}^{t_j}; \{\mathbf{W}_k^{t_j}\}_{k=1}^K, \mathbf{V}^{t_j})$, we know $(\{\mathbf{W}_k^{t_j}\}_{k=1}^K, \mathbf{V}^{t_j}) \in \mathbb{C}_\le(\{\mathbf{W}_k^{t_j}\}_{k=1}^K, \mathbf{V}^{t_j})$. It follows that $(\{\mathbf{W}_k^\star\}_{k=1}^K, \mathbf{V}^\star) \in \mathbb{C}_\le(\{\mathbf{W}_k^\star\}_{k=1}^K, \mathbf{V}^\star)$. Combining this with (61) and (62), we have $(\{\mathbf{W}_k^\star\}_{k=1}^K, \mathbf{V}^\star) \in \mathbb{S}(\mathbf{U}_r^\star, \mathbf{F}^\star, \{\mathbf{W}_k^\star\}_{k=1}^K, \mathbf{V}^\star)$.

Finally, we show that any limit point $(\mathbf{U}_d^\star, \mathbf{F}^\star, \{\mathbf{W}_k^\star\}_{k=1}^K, \mathbf{V}^\star)$ is a KKT point of the problem (27). Since the Slater's condition holds for problem (34) and $(\{\mathbf{W}_k^\star\}_{k=1}^K, \mathbf{V}^\star) \in \mathbb{S}(\mathbf{U}_r^\star, \mathbf{F}^\star, \{\mathbf{W}_k^\star\}_{k=1}^K, \mathbf{V}^\star)$, the KKT condition of problem (34) is satisfied, i.e.,

$$
\begin{aligned}
&\sum_{k=1}^K \Big((\Xi_k - \lambda_1 \Gamma_k + \lambda_2 \Upsilon_k + \lambda_{k+3}\mathbf{I})\,\mathbf{W}_k^\star - \bar{\mathbf{H}}_k^\dagger \mathbf{F}^\dagger \bar{\mathbf{G}}_2^\dagger \mathbf{U}_d \Big)\\
&\quad + (\Xi_v - \lambda_1 \Gamma_v + \lambda_2 \Upsilon_v + \lambda_3 \mathbf{I})\,\mathbf{V}^\star = 0, && \text{(63a)}\\
&\lambda_1 g(\{\mathbf{W}_k^\star\}_{k=1}^K, \mathbf{V}^\star) = 0, && \text{(63b)}\\
&\lambda_2 \Bigg(\sum_{k=1}^K \mathrm{tr}\left(\mathbf{W}_k^\star \mathbf{W}_k^{\star\dagger} \Upsilon_k\right) + \mathrm{tr}\left(\mathbf{V}^\star \mathbf{V}^{\star\dagger} \Upsilon_v\right) + \mathrm{tr}\left(\sigma^2 \mathbf{F}^\star \mathbf{F}^{\star\dagger}\right)\\
&\quad - P_r) = 0, && \text{(63c)}\\
&\lambda_3 \left(\|\mathbf{V}^\star\|^2 - P_a \right) = 0, && \text{(63d)}\\
&\lambda_{k+3} \left(\|\mathbf{W}_k^\star\|^2 - P_k \right) = 0, \forall k \in \mathcal{K}. && \text{(63e)}
\end{aligned}
$$

By the continuity, we have

$$
\mathbf{U}_d^\star = \Big(\bar{\mathbf{G}}_2 \mathbf{F}^\star (\lambda_{\mathbf{H}}(\{\mathbf{W}_k\}_{k=1}^K) + \lambda_{\Delta \mathbf{G}_1}(\mathbf{V})) \mathbf{F}^{\star\dagger} \bar{\mathbf{G}}_2^\dagger
$$

$$+\mathrm{tr}\left(\mathbf{F}^{\star}(\lambda_{\mathbf{H}}(\{\mathbf{W}_k\}_{k=1}^{K})+\lambda_{\Delta\mathbf{G}_1}(\mathbf{V}))\mathbf{F}^{\star\dagger}\Psi_{G,2}\right)\Sigma_{G,2}$$

$$+\sigma^2\lambda_{\mathbf{G}_2}(\mathbf{F})+\sigma^2\mathbf{I}\Big)^{-1}\left(\bar{\mathbf{G}}_2\mathbf{F}^{\star}\sum_{k=1}^{K}\bar{\mathbf{H}}_k\mathbf{W}_k^{\star}\right), \quad (64a)$$

$$\mathbf{U}_r^{\star}=(\lambda_{\mathbf{H}}(\{\mathbf{W}_k\}_{k=1}^{K})+\lambda_{\mathbf{G_1}}(\mathbf{V})+\sigma^2\mathbf{I})^{-1}\left(\sum_{k=1}^{K}\bar{\mathbf{H}}_k\mathbf{W}_k^{\star}\right), \quad (64b)$$

$$\mathbf{F}^{\star}=(\breve{\mathbf{P}}+\mu\mathbf{I})^{\ddagger}\left(-\sum_{k=1}^{K}\bar{\mathbf{H}}_k\mathbf{W}_k^{\star}\mathbf{U}_d^{\star\dagger}\bar{\mathbf{G}}_2\right)^{\dagger}\breve{\mathbf{R}}^{-1}. \quad (65)$$

Together with (63)-(65), $(\mathbf{U}_d^{\star},\mathbf{F}^{\star},\{\mathbf{W}_k^{\star}\}_{k=1}^{K},\mathbf{V}^{\star})$ is a KKT point of problem (27).

References

[1] 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, Sep. 2021.

[2] G. Yu, X. Chen, and D. W. K. Ng, "Low-cost design of massive access for cellular Internet of Things," *IEEE Trans. Commun.*, vol. 67, no. 11, pp. 8008–8020, Nov. 2019.

[3] L. Chen, N. Zhao, Y. Chen, F. R. Yu, and G. Wei, "Over-the-air computation for IoT networks: Computing multiple functions with antenna arrays," *IEEE Internet Things J.*, vol. 5, no. 6, Dec. 2018, Art. no. 52965306.

[4] B. Nazer and M. Gastpar, "Computation over multiple-access channels," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, Oct. 2007, Art. no. 34983516.

[5] M. Gastpar, "Uncoded transmission is exactly optimal for a simple Gaussian sensor network," *IEEE Trans. Inf. Theory*, vol. 54, no. 11, Nov. 2008, Art. no. 52475251.

[6] O. Abari, H. Rahul, and D. Katabi, "Over-the-air function computation in sensor networks," 2016, *arXiv:1612.02307*.

[7] X. Chen, A. Liu, and M.-J. Zhao, "High-mobility multi-modal sensing for IoT network via MIMO aircomp: A mixed-timescale optimization approach," *IEEE Commun. Lett.*, vol. 24, no. 10, pp. 2295–2299, Oct. 2020.

[8] X. Li, G. Zhu, Y. Gong, and K. Huang, "Wirelessly powered data aggregation for IoT via over-the-air function computation: Beamforming and power control," *IEEE Trans. Wireless Commun.*, vol. 18, no. 7, pp. 3437–3452, Jul. 2019.

[9] F. Wu, L. Chen, N. Zhao, Y. Chen, F. R. Yu, and G. Wei, "Computation over wide-band multi-access channels: Achievable rates through subfunction allocation," *IEEE Trans. Wireless Commun.*, vol. 18, no. 7, pp. 3713–3725, Jul. 2019.

[10] L. Chen, N. Zhao, Y. Chen, F. R. Yu, and G. Wei, "Toward optimal rate-delay tradeoff for computation over multiple access channel," *IEEE Trans. Commun.*, vol. 69, no. 7, pp. 4335–4346, Jul. 2021.

[11] L. Chen, N. Zhao, Y. Chen, F. R. Yu, and G. Wei, "Over-the-air computation for cooperative wideband spectrum sensing and performance analysis," *IEEE Trans. Veh. Technol.*, vol. 67, no. 11, pp. 10603–10614, Nov. 2018.

[12] Y. Cai, Q. Shi, B. Champagne, and G. Ye Li, "Joint transceiver design for secure downlink communications over an amplify-and-forward MIMO relay," *IEEE Trans. Commun.*, vol. 65, no. 9, pp. 3691–3704, Sep. 2017.

[13] Y. Cai, Y. Xu, Q. Shi, B. Champagne, and L. Hanzo, "Robust joint hybrid transceiver design for millimeter wave full-duplex MIMO relay systems," *IEEE Trans. Wireless Commun.*, vol. 18, no. 2, pp. 1199–1215, Feb. 2019.

[14] X. Zhai, G. Han, Y. Cai, and L. Hanzo, "Beamforming design based on two-stage stochastic optimization for RIS-assisted over-the-air computation systems," *IEEE Internet Things J.*, vol. 9, no. 7, pp. 5474–5488, Apr. 2022.

[15] X. Chen, R. Jia, and D. W. K. Ng, "The application of relay to massive non-orthogonal multiple access," *IEEE Trans. Commun.*, vol. 66, no. 11, pp. 5168–5180, Nov. 2018.

[16] J. Laneman, D. Tse, and G. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, Dec. 2004, Art. no. 30623080.

[17] A. Goldsmith, *Wireless Communications* Cambridge, U.K.: Cambridge Univ. Press, 2005.

[18] F. Wang, J. Xu, V. K. N. Lau, and S. Cui, "Amplify-and-forward relaying for hierarchical over-the-air computation," *IEEE Trans. Wireless Commun.*, vol. 21, no. 12, pp. 10529–10543, Dec. 2022.

[19] S. Tang, H. Yin, C. Zhang, and S. Obana, "Reliable over-the-air computation by amplify-and-forward based relay," *IEEE Access*, vol. 9, pp. 53333–53342, 2021.

[20] R. Jiang, S. Zhou, and K. Huang, "Achieving cooperative diversity in over-the-air computation via relay selection," in *Proc. IEEE 92nd Veh. Technol. Conf. (VTC-Fall)*, Nov. 2020, pp. 1–6.

[21] F. Wu, L. Chen, N. Zhao, Y. Chen, F. R. Yu, and G. Wei, "Computation over multi-access channels: Multi-hop implementation and resource allocation," *IEEE Trans. Commun.*, vol. 69, no. 2, pp. 1038–1052, Feb. 2021.

[22] Y. Li, M. Jiang, G. Zhang, and M. Cui, "Joint optimization for multi-antenna AF-relay aided over-the-air computation," *IEEE Trans. Veh. Technol.*, vol. 71, no. 6, pp. 6744–6749, Jun. 2022.

[23] M. Jiang, Y. Li, G. Zhang, and M. Cui, "Joint beamforming optimization in multi-relay assisted MIMO over-the-air computation for multi-modal sensing data aggregation," *IEEE Commun. Lett.*, vol. 25, no. 12, pp. 3937–3941, Dec. 2021.

[24] X. Chen, D. W. K. Ng, W. H. Gerstacker, and H. Chen, "A survey on multiple-antenna techniques for physical layer security, *IEEE Commun. Surveys Tuts.*, vol. 19, no. 2, pp. 1027–1053, 2nd Quart., 2017.

[25] L. Lv, F. Zhou, J. Chen, and N. Al-Dhahir, "Secure cooperative communications with an untrusted relay: A NOMA-inspired jamming and relaying approach," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 12, pp. 3191–3205, Dec. 2019.

[26] C. Jeong, I.-M. Kim, and D. Kim, "Joint secure beamforming design at the source and the relay for an amplify-and-forward MIMO untrusted relay system," *IEEE Trans. Signal Process.*, vol. 60, no. 1, pp. 310–325, Jan. 2012.

[27] J. Xiong, L. Cheng, D. Ma, and J. Wei, "Destination-aided cooperative jamming for dual-hop amplify-and-forward MIMO untrusted relay systems," *IEEE Trans. Veh. Technol.*, vol. 65, no. 9, pp. 7274–7284, Sep. 2016.

[28] W. Jiang, Y. Gong, Q. Xiao, and Y. Liao, "Secrecy rate maximization for untrusted relay networks with nonorthogonal cooperative transmission protocols," *IEEE Trans. Veh. Technol.*, vol. 67, no. 7, pp. 6325–6339, Jul. 2018.

[29] Q. Li, L. Yang, Q. Zhang, and J. Qin, "Robust an-aided secure precoding for an AF MIMO untrusted relay system," *IEEE Trans. Veh. Technol.*, vol. 66, no. 11, pp. 10572–10576, Nov. 2017.

[30] J. Mo, M. Tao, Y. Liu, and R. Wang, "Secure beamforming for MIMO two-way communications with an untrusted relay," *IEEE Trans. Signal Process.*, vol. 62, no. 9, pp. 2185–2199, May 2014.

[31] S. Zhao, J. Liu, Y. Shen, X. Jiang, and N. Shiratori, "Secure beamforming for full-duplex MIMO two-way untrusted relay systems," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 3775–3790, 2020.

[32] Q. Li and L. Yang, "Artificial noise aided secure precoding for MIMO untrusted two-way relay systems with perfect and imperfect channel state information," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 10, pp. 2628–2638, Oct. 2018.

[33] A. Kuhestani, A. Mohammadi, and M. Mohammadi, "Joint relay selection and power allocation in large-scale MIMO systems with untrusted relays and passive eavesdroppers," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 2, pp. 341–355, Feb. 2018.

[34] C. Hu, Q. Li, Q. Zhang, and J. Qin, "Secure transceiver design and power control for over-the-air computation networks," *IEEE Commun. Lett.*, vol. 26, no. 7, pp. 1509–1513, Jul. 2022.

[35] M. Frey, I. Bjelaković, and S. Stańczak, "Towards secure over-the-air computation," in *Proc. IEEE ISIT.*, Jul. 2021, pp. 700–705.

[36] L. Lv, H. Jiang, Z. Ding, L. Yang, and J. Chen, "Secrecy-enhancing design for cooperative downlink and uplink NOMA with an untrusted relay," *IEEE Trans. Commun.*, vol. 68, no. 3, pp. 1698–1715, Mar. 2020.

[37] O. Abari, H. Rahul, D. Katabi, and M. Pant, "AirShare: Distributed coherent transmission made seamless, in *Proc. IEEE Conf. Comput. Commun.*, Apr. 2015, Art. no. 17421750.

[38] X. Cao, G. Zhu, J. Xu, and K. Huang, "Cooperative interference management for over-the-air computation networks," *IEEE Trans. Wireless Commun.*, vol. 20, no. 4, pp. 2634–2651, Apr. 2021.

[39] X. Cao, G. Zhu, J. Xu, and K. Huang, "Optimized power control for over-the-air computation in fading channels," *IEEE Trans. Wireless Commun.*, vol. 19, no. 11, pp. 7498–7513, Nov. 2020.

[40] Z. Wang, Y. Zhou, Y. Shi, and W. Zhuang, “Interference management for over-the-air federated learning in multi-cell wireless networks,” *IEEE J. Sel. Areas Commun.*, vol. 40, no. 8, pp. 2361–2377, Aug. 2022.

[41] J. Dong, Y. Shi, and Z. Ding, “Blind over-the-air computation and data fusion via provable Wirtinger flow,” *IEEE Trans. Signal Process.*, vol. 68, Jan. 2020, Art. no. 11361151.

[42] G. Zhu and K. Huang, “MIMO over-the-air computation for high-mobility multi-modal sensing,” *IEEE Internet Things J.*, vol. 6, no. 4, Aug. 2018, Art. no. 60896103.

[43] Q. An, Y. Zhou, and Y. Shi, “Robust design for reconfigurable intelligent surface assisted over-the-air computation,” in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Mar. 2021, pp. 1–6.

[44] Q. Qi, X. Chen, L. Lei, C. Zhong, and Z. Zhang, “Robust convergence of energy and computation for B5G cellular Internet of Things,” in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2019, pp. 1–6.

[45] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1993.

[46] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[47] Y. Nesterov and A. Nemirovskii, *Interior-Point Polynomial Algorithms in Convex Programming*. Philadelphia, PA, USA: SIAM, 1994.

[48] Y. Sun, D. W. K. Ng, Z. Ding, and R. Schober, “Optimal joint power and subcarrier allocation for full-duplex multicarrier non-orthogonal multiple access systems,” *IEEE Trans. Commun.*, vol. 65, no. 3, pp. 1077–1091, Mar. 2017.

[49] C. Xing, S. Ma, and Y.-C. Wu, “Robust joint design of linear relay precoder and destination equalizer for dual-hop amplify-and-forward MIMO relay systems,” *IEEE Trans. Signal Process.*, vol. 58, no. 4, pp. 2273–2283, Apr. 2010.

[50] E. G. Larsson and P. Stoica, *Space-Time Block Coding for Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2003.

[51] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.

[52] M. Jiang et al., “MIMO beamforming design in nonorthogonal multiple access downlink interference channels, *IEEE Trans. Veh. Technol.*, vol. 67, no. 8, Aug. 2018, Art. no. 69516959.

[53] K. Huang and N. D. Sidiropoulos, “Consensus-ADMM for general quadratically constrained quadratic programming,” *IEEE Trans. Signal Process.*, vol. 64, no. 20, pp. 5297–5310, Oct. 2016.

[54] A. Ben-Tal and A. Nemirovski, *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications*. Philadelphia, PA, USA: SIAM, 2001.

[55] J. Yang, B. Champagne, Y. Zou, and L. Hanzo, “Joint optimization of transceiver matrices for MIMO-aided multiuser AF relay networks: Improving the QoS in the presence of CSI errors,” *IEEE Trans. Veh. Technol.*, vol. 65, no. 3, pp. 1434–1451, Mar. 2016.

[56] Y. C. Eldar, A. Ben-Tal, and A. Nemirovski, “Robust mean-squared error estimation in the presence of model uncertainties, *IEEE Trans. Signal Process.*, vol. 53, no. 1, Jan. 2005, Art. no. 168181.

[57] Y. Huo, X. Fan, L. Ma, X. Cheng, Z. Tian, and D. Chen, “Secure communications in tiered 5G wireless networks with cooperative jamming, *IEEE Trans. Wireless Commun.*, vol. 18, no. 6, Jun. 2019, Art. no. 32653280.

[58] M. Grant and S. Boyd. *CVX: MATLAB Software for Disciplined Convex Programming*. Accessed: Oct. 16, 2021. [Online]. Available: http://cvxr.com/cvx

[59] S. Yan, X. Wang, Z. Li, B. Li, and Z. Fei, “Cooperative jamming for physical layer security in hybrid satellite terrestrial relay networks, *China Commun.*, vol. 16, no. 12, Dec. 2019, Art. no. 154164.

[60] A. Shapiro, D. Dentcheva, and A. Ruszczynski, *Lectures on Stochastic Programming: Modeling and Theory* (MPS-SIAM Series on Optimization). Philadelphia, PA, USA: SIAM, Sep. 2009.

[61] A. Aubry, A. De Maio, A. Zappone, M. Razaviyayn, and Z.-Q. Luo, “A new sequential optimization procedure and its applications to resource allocation for wireless systems,” *IEEE Trans. Signal Process.*, vol. 66, no. 24, pp. 6518–6533, Dec. 2018.

[62] H. A. Taha, *Operations Research: An Introduction*, 8th ed. Englewood Cliffs, NJ, USA: Prentice-Hall, 2007.

[63] A. J. Smola, S. V. N. Vishwanathan, and T. Hofmann, “Kernel methods for missing variables,” in *Proc. 10th Int. Workshop Artif. Intell. Stat.* Mar. 2005, Art. no. 325332.

Hualiang Luo received the B.Eng. and M.S. degrees from the Wuhan University of Technology, Wuhan, China, in 2017 and 2020, respectively. He is currently pursuing the Ph.D. degree with the School of Electronics and Information Technology, Sun Yat-sen University, Guangzhou, China. His research interests include the optimization design of wireless communications systems.

Quanzhong Li received the B.S. and Ph.D. degrees from Sun Yat-sen University (SYSU), Guangzhou, China, in 2009 and 2014, respectively. He is currently an Associate Professor with the School of Computer Science and Engineering, SYSU. His research interests include wireless communications and signal processing, with main focus on optimization techniques for resource allocation and physical layer security.

Qi Zhang (Member, IEEE) received the B.Eng. (Hons.) and M.S. degrees from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 1999 and 2002, respectively, and the Ph.D. degree in electrical and computer engineering from the National University of Singapore (NUS), Singapore, in 2007.

From 2007 to 2008, he was a Research Fellow with the Communications Laboratory, Department of Electrical and Computer Engineering, NUS. From 2008 to 2011, he was with the Center for Integrated Electronics, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China. He is currently an Associate Professor with the School of Electronics and Information Technology, Sun Yat-sen University, Guangzhou, China. His research interests include UAV communications, non-orthogonal multiple access, wireless communications powered by energy harvesting, cooperative communications, and ultra-wideband communications.

Jiayin Qin received the M.S. degree in radio physics from Huazhong Normal University, Wuhan, China, in 1992, and the Ph.D. degree in electronics from Sun Yat-sen University (SYSU), Guangzhou, China, in 1997.

From 2002 to 2004, he was the Head of the Department of Electronics and Communication Engineering, SYSU. From 2003 to 2008, he was the Vice Dean of the School of Information Science and Technology, SYSU. His research interests include wireless communications and submillimeter wave technology.

Dr. Qin was a recipient of the IEEE Communications Society Heinrich Hertz Award for Best Communications Letter in 2014, the Second Young Teacher Award of Higher Education Institutions, Ministry of Education (MOE), China, in 2001, the Seventh Science and Technology Award for Chinese Youth in 2001, and the New Century Excellent Talent, MOE, China, in 1999.