

Over-the-Air Computation via Reconfigurable Intelligent Surface with Phase-Dependent Amplitude Response

Deyou Zhang[†], Zhijie Zheng[†], Wanxi Zhang[†], Ming Xia[†], and Chuang Shi[§]

[†]*School of Electronics and Information Engineering, Beihang University, Beijing 100191, China*

[§]*School of Space and Earth Sciences, Beihang University, Beijing 100191, China*

email: {deyou, zhengzhijie, zhangwanxi, xiaming, shichuang}@buaa.edu.cn

Abstract—Over-the-air computation (AirComp) leverages the inherent superposition property of wireless multiple-access channels to enable direct signal aggregation from massive users. However, unfavorable channel conditions can severely degrade the computed mean square error (CMSE). To address this limitation, we introduce a reconfigurable intelligent surface (RIS) into the AirComp system. Unlike prior works that assume full signal reflection by each RIS element (RE) regardless of its phase shift, we adopt a practical model accounting for the coupling between the amplitude and phase of each RE. Based on this model, we formulate an optimization problem to minimize the CMSE by jointly optimizing transceiver design and the RIS reflection matrix. To tackle this highly nonconvex problem, we propose a dual-loop optimization framework, where the outer loop employs the genetic algorithm to obtain a near-optimal RIS reflection matrix and the inner loop uses an alternating optimization approach for transceiver design. Simulation results demonstrate that the proposed dual-loop algorithm outperforms baseline methods in reducing the CMSE.

Index Terms—Over-the-air computation, reconfigurable intelligent surface, phase-dependent amplitude response.

I. INTRODUCTION

The rapid development of artificial intelligence (AI), 5G technology, and edge computing has accelerated the growth of the Industrial Internet of Things (IIoT) [1], [2]. As a key component of Industry 4.0, IIoT drives the intelligent transformation of industries by connecting industrial equipment with information systems. As reported by [1], the rapid advancement of IIoT hinges on the deployment of interconnected edge devices, such as sensors, which continuously generate massive volumes of data. Aligning with this direction, we focus on a specific class of IIoT applications where an edge server aggregates data distributed across massive devices relying on wireless communications, therefore termed wireless data aggregation (WDA). Conventional “transmit-then-compute” approaches require an access point (AP) to first receive data from each device individually and then compute a specific function (e.g., arithmetic mean) of the received data. However,

this approach is inefficient, especially as the number of devices scales. Fortunately, over-the-air computation (AirComp) offers a promising solution for efficient WDA by seamlessly integrating communication and computation processes. Specifically, AirComp enables concurrent data transmissions from all devices over shared radio resources by exploiting the waveform superposition property of multiple-access channels, thus yielding a revolutionary paradigm of “compute when communicate” [3].

Recent works have extensively explored AirComp, particularly from the perspective of transceiver design [3]. Specifically, [4] and [5] focused on the single-input single-output (SISO) setup and derived optimal transceiver designs for AirComp. Subsequently, [6] investigated AirComp in the context of the single-input multiple-output (SIMO) setup, proposing a uniform-forcing transceiver design to mitigate non-uniform channel fading across devices. Additionally, [7] explored multiple-input multiple-output (MIMO)-enabled AirComp for multi-modal sensing, investigating associated transmit and receive beamforming designs. As reported in these studies [4]–[6], AirComp performance, often evaluated using the computation mean square error (CMSE), is significantly influenced by the quality of equivalent channels between edge devices and the AP.

Notably, in addition to optimizing transceiver design, reconfigurable intelligent surfaces (RISs) can be employed to directly strengthen wireless channels between edge devices and the AP [8]. Consequently, RIS-aided AirComp systems have garnered considerable attention. In particular, [9] first explored the RIS-aided AirComp system, investigating the joint design of AirComp transceivers and RIS phase shifts. Moreover, [10] proposed deploying two RISs to empower the AirComp system, with one RIS positioned near the AP and the other near edge devices. It is worth noting that the aforementioned works, along with as many others on RIS, have all assumed an ideal reflection model, i.e., unity reflection amplitude for each reflecting element (RE) regardless of its phase shift. However, this assumption is impractical due to hardware limitations. As demonstrated in [11], the amplitude response of an RE is generally non-uniform with respect to its phase shift, i.e., the amplitude and phase shift are coupled, and therefore the phase shift matrix of the RIS needs to be

This work was supported in part by the Fundamental Research Funds for the Central Universities, in part by the Beihang World TOP University Cooperation Program, and in part by the Beihang Ganwei Project under Grant WZ2024-2-31. (Corresponding author: Deyou Zhang.)

Simulation codes are provided to reproduce the results presented in this paper: <https://github.com/zheng977/Practical-RIS-Dual-Loop-Opt>.

properly designed to balance the reflected amplitude of each RE and phase alignment across all REs under this non-ideal model with phase-dependent amplitude response. Achieving such a balance is nontrivial and thus serves as the motivation for this work.

In this paper, we revisit the RIS-aided AirComp system by accounting for the coupling between the amplitude response and phase shift of REs. Our objective is to minimize the CMSE by jointly optimizing edge devices' transmit coefficients, the AP's denoising factor, and the RIS reflection matrix, which is a highly nonconvex optimization problem. To solve this, we establish a dual-loop optimization framework, where the outer loop employs a genetic algorithm (GA) to obtain a near-optimal RIS reflection matrix and the inner loop alternately optimizes edge devices' transmit coefficients and the AP's denoising factor. Simulation results demonstrate the effectiveness of our proposed algorithm in optimizing transceiver design and the RIS reflection matrix under the non-ideal reflection model with phase-dependent amplitude response.

Notations: We use regular, bold lowercase, and bold uppercase letters to denote scalars, vectors, and matrices, respectively. We use \mathbb{R} and \mathbb{C} to denote the sets of real and complex numbers, respectively. We use the notation x^* to denote the conjugate of x , $\text{Diag}(x)$ to denote a diagonal matrix with its diagonal entries specified by x , $\mathcal{CN}(\mu, \sigma)$ to denote the complex Gaussian distribution with mean μ and covariance σ , and \mathbb{E} to denote the expectation operator.

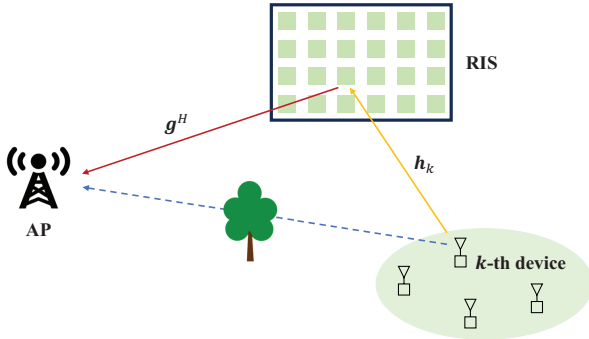


Fig. 1. The system model of RIS-aided AirComp.

II. SYSTEM MODEL AND PROBLEM FORMULATION

A. System Model

Consider a RIS-aided AirComp system consisting of a single-antenna AP, K single-antenna edge devices, and an N -element RIS, as shown in Fig. 1. Let $s_k \in \mathbb{C}$ denote the normalized information-bearing data generated by device k , satisfying $\mathbb{E}[s_k] = 0$, and $\mathbb{E}[|s_k|^2] = 1$, $\forall k \in \mathcal{K} \triangleq \{1, \dots, K\}$. Additionally, $\{s_1, \dots, s_K\}$ are assumed to be mutually uncorrelated, i.e., $\mathbb{E}[s_k s_j^*] = 0$, $k \neq j$, $k, j \in \mathcal{K}$. Following prior works [4]–[6], the target function that the AP aims to recover is the arithmetic mean value of all devices' data:

$$s = \frac{1}{K} \sum_{k=1}^K s_k. \quad (1)$$

Based on the principle of AirComp [3], all devices transmit their data concurrently to the AP. Consequently, the received signal at the AP can be expressed as

$$y = \mathbf{g}^H \mathbf{\Theta} \left(\sum_{k=1}^K \mathbf{h}_k b_k s_k \right) + z, \quad (2)$$

where $b_k \in \mathbb{C}$ denotes the transmit coefficient of device k , $\mathbf{h}_k \in \mathbb{C}^{N \times 1}$ denotes the channel from device k to RIS, $\mathbf{g}^H \in \mathbb{C}^{1 \times N}$ denotes the channel from RIS to AP, $\mathbf{\Theta}$ denotes the RIS reflection matrix, and $z \sim \mathcal{CN}(0, \sigma^2)$ denotes the additive Gaussian noise at the AP. Note that direct links between devices and the AP are assumed to be blocked in (2) for simplicity; the proposed algorithm can be readily extended to scenarios with direct links.

According to [11], the RIS reflection matrix $\mathbf{\Theta}$ can be written as follows:

$$\mathbf{\Theta} = \text{Diag}(\alpha_1 e^{j\omega_1}, \alpha_2 e^{j\omega_2}, \dots, \alpha_N e^{j\omega_N}), \quad (3)$$

where α_n and ω_n denote the amplitude response and phase shift of the n -th RE, respectively, $\forall n \in \mathcal{N} \triangleq \{1, \dots, N\}$. Unlike the ideal model that assumes full signal reflection by each RE, we consider a practical phase-dependent amplitude response for each RE, as given in [11]

$$\alpha_n(\omega_n) = (1 - \alpha_{\min}) \left(\frac{\sin(\omega_n - \phi) + 1}{2} \right)^\gamma + \alpha_{\min}, \quad (4)$$

where $\alpha_{\min} \geq 0$, $\phi \geq 0$, and $\gamma \geq 0$ are circuit implementation parameters. As evident from (4), the amplitude response of an RE is uniquely determined by its phase shift. In other words, optimizing the RIS reflection matrix $\mathbf{\Theta}$ is equivalent to optimizing the phase shifts of all REs, $\boldsymbol{\omega} \triangleq [\omega_1, \dots, \omega_N]$. Moreover, $\alpha_{\min} = 1$ leads to an ideal amplitude response, i.e., $\alpha_n = 1$, $\forall n \in \mathcal{N}$.

Let $\eta \in \mathbb{R}$ denote the denoising factor at the AP. Then, the estimated target variable \hat{s} can be expressed as

$$\hat{s} = \eta y = \sum_{k=1}^K \eta h_{e,k} b_k s_k + \eta z, \quad (5)$$

where $h_{e,k} = \mathbf{g}^H \mathbf{\Theta} \mathbf{h}_k$ is termed the equivalent channel between device k and AP, $\forall k \in \mathcal{K}$.

B. Problem Formulation

Our objective is to minimize the difference between s and \hat{s} , evaluated by the CMSE as defined below:

$$\begin{aligned} \text{CMSE} &\triangleq \mathbb{E}[|\hat{s} - s|^2] \\ &= \mathbb{E} \left[\left| \sum_{k=1}^K \left(\eta h_{e,k} b_k - \frac{1}{K} \right) s_k + \eta z \right|^2 \right] \\ &= \sum_{k=1}^K \left| \eta h_{e,k} b_k - \frac{1}{K} \right|^2 + \eta^2 \sigma^2. \end{aligned} \quad (6)$$

To this end, we construct the following optimization problem:

$$\min_{\eta, \mathbf{b}, \boldsymbol{\omega}} \text{CMSE} \quad (7a)$$

$$\text{s.t. } |b_k|^2 \leq P_0, \forall k \in \mathcal{K}, \quad (7b)$$

$$\omega_n \in [-\pi, \pi], \forall n \in \mathcal{N}, \quad (7c)$$

where (7b) accounts for the power constraint for each device, and $\mathbf{b} \triangleq [b_1, \dots, b_K]$.

III. JOINT DESIGN OF TRANSCEIVER AND RIS REFLECTION MATRIX

Since the objective function is highly nonconvex over $\boldsymbol{\omega}$ and variables η , \mathbf{b} , and $\boldsymbol{\omega}$ are mutually coupled, solving (7) is challenging. To tackle this problem, we propose a dual-loop optimization framework to derive a suboptimal solution. The outer loop employs a GA to optimize $\boldsymbol{\omega}$, where the fitness function of each chromosome (i.e., a realization of $\boldsymbol{\omega}$) is evaluated in the inner loop by alternately optimizing η and \mathbf{b} until convergence.

A. Inner-Loop Optimization

In the inner-loop of the proposed algorithm, to determine the fitness value of one chromosome, we need to solve the following problem by jointly optimizing η and \mathbf{b} :

$$\min_{\eta, \mathbf{b}} \sum_{k=1}^K \left| \eta h_{e,k} b_k - \frac{1}{K} \right|^2 + \eta^2 \sigma^2 \quad (8a)$$

$$\text{s.t. } |b_k|^2 \leq P_0, \forall k \in \mathcal{K}. \quad (8b)$$

To resolve the coupling between η and \mathbf{b} , we decouple (8) into two sub-problems and solve them alternately until convergence, as detailed below.

1) Optimization of \mathbf{b} . Given η , the optimization problem over \mathbf{b} is formulated as

$$\min_{\mathbf{b}} \sum_{k=1}^K \left| \eta h_{e,k} b_k - \frac{1}{K} \right|^2 \quad (9a)$$

$$\text{s.t. } |b_k|^2 \leq P_0, \forall k \in \mathcal{K}, \quad (9b)$$

where $\eta^2 \sigma^2$ is ignored since it is unrelated with \mathbf{b} . Note that problem (9) can be decoupled into K independent sub-problems, each corresponding to one device's transmit coefficient. The problem associated with b_k is formulated as

$$\min_{b_k} \left| \eta h_{e,k} b_k - \frac{1}{K} \right|^2 \quad (10a)$$

$$\text{s.t. } |b_k|^2 \leq P_0. \quad (10b)$$

Problem (10) is observed to be a quadratically constraint quadratic program (QCQP), and it satisfies the Slater's conditions. Therefore, it can be solved optimally by using the Karush-Kuhn-Tucker (KKT) conditions, as detailed below.

The associated Lagrangian function of (10) can be written as follows:

$$\mathcal{L} = \left| \eta h_{e,k} b_k - \frac{1}{K} \right|^2 + \lambda_k (|b_k|^2 - P_0), \quad (11)$$

where $\lambda_k \geq 0$ is the Lagrangian multiplier. The corresponding KKT conditions of (11) are given by

$$\frac{\partial \mathcal{L}}{\partial b_k} = 0, \quad (12a)$$

$$\lambda_k (|b_k|^2 - P_0) = 0, \quad (12b)$$

$$|b_k|^2 \leq P_0. \quad (12c)$$

By solving (12a), we obtain

$$b_k^* = \frac{1}{K} \frac{\eta h_{e,k}^*}{\eta^2 |h_{e,k}|^2 + \lambda_k^*}. \quad (13)$$

Considering (12b) and (12c), the optimal Lagrangian multiplier λ_k^* is computed as

$$\lambda_k^* = \max \left(\frac{1}{K} \frac{\eta |h_{e,k}|}{\sqrt{P_0}} - \eta^2 |h_{e,k}|^2, 0 \right). \quad (14)$$

2) Optimization of η . Given \mathbf{b} , the optimization problem corresponding to η is given by

$$\min_{\eta} \text{CMSE} = \sum_{k=1}^K \left| \eta h_{e,k} b_k - \frac{1}{K} \right|^2 + \eta^2 \sigma^2. \quad (15)$$

Problem (15) is observed to be an unconstrained convex quadratic programming (QP) problem, and its optimal solution can be obtained by setting the derivative of (15) with respect to η to zero, i.e.,

$$\frac{\partial \text{CMSE}}{\partial \eta} = 0. \quad (16)$$

By solving (16), we obtain:

$$\eta^* = \frac{1}{K} \frac{\sum_{k=1}^K |h_{e,k} b_k|}{\sigma^2 + \sum_{k=1}^K |h_{e,k} b_k|^2}. \quad (17)$$

B. Outer-Loop Optimization

In the outer loop of the proposed algorithm, we employ the GA as an efficient approach to optimize $\boldsymbol{\omega}$, which consists of the following steps.

Step 0: A random population of $I \gg 1$ chromosomes is generated, denoted by $\{\boldsymbol{\omega}_1, \dots, \boldsymbol{\omega}_I\}$, where each chromosome represents a potential solution for $\boldsymbol{\omega}$.

Step 1: Evaluate the fitness $\text{CMSE}(\eta_i^*, \mathbf{b}_i^*, \boldsymbol{\omega}_i)$ of each chromosome $\boldsymbol{\omega}_i$, where $(\eta_i^*, \mathbf{b}_i^*)$ denotes the locally optimal solution obtained in the inner loop for a given $\boldsymbol{\omega}_i$, $\forall i \in \mathcal{I}$, where $\mathcal{I} \triangleq \{1, \dots, I\}$.

Step 2: Create an offspring population by executing parent chromosome selection, crossover, and mutation. Additionally, a projection function $\Pi(\cdot)$ is applied to every newly generated

chromosome to ensure satisfaction of constraint (7c), defined as

$$\Pi(\omega_n) = \begin{cases} -\pi + [(\omega_n + \pi) \bmod 2\pi], & \text{if } \omega_n < -\pi, \\ -\pi + [\omega_n \bmod 2\pi], & \text{if } \omega_n \geq \pi, \\ \omega_n, & \text{otherwise.} \end{cases} \quad (18)$$

Furthermore, to prevent degradation of solution quality across generations, we adopt elitism selection, preserving the chromosome with the best fitness in each generation. This ensures monotonic improvement of the best fitness (i.e., objective function) and accelerates convergence toward optimal solutions.

Step 3: Check if the termination condition is satisfied. If yes, stop and return the best solution in current population; otherwise, proceed to **Step 1**.

C. Overall Algorithm

The overall algorithm for solving (7) is summarized in **Algorithm 1**, where T denotes the maximum number of iterations in the outer loop. Note that the convergence of this algorithm is guaranteed since the objective function is non-increasing with each iteration and has a lower bound of zero.

Algorithm 1 Proposed Dual-Loop Algorithm

```

1: for  $t = 1$  to  $T$  do                                ▷ Outer Loop
2:   Create  $I$  chromosomes  $\omega_1, \dots, \omega_I$  following Step 2.
3:   for  $i = 1$  to  $I$  do
4:     Set  $b_k = \sqrt{P_0}$ ,  $\forall k \in \mathcal{K}$ , and  $\omega = \omega_i$ .
5:     repeat                                              ▷ Inner Loop
6:       Update  $\eta$  via (17).
7:       Update  $b_k$  via (13),  $\forall k \in \mathcal{K}$ .
8:     until Convergence
9:   end for
10: end for
11: Obtain  $j = \arg \min_{1 \leq i \leq N} \text{CMSE}(\eta_i^*, b_i^*, \omega_i)$ .
12: Return  $\eta_j^*$ ,  $b_j^*$ , and  $\omega_j$ .
```

IV. SIMULATION RESULTS

In this section, we present numerical results to demonstrate the effectiveness of our proposed dual-loop optimization framework for solving (7). We consider a three-dimensional (3D) coordinate setup, where the AP and the RIS are positioned at $(-50, 0, 10)$ and $(0, 0, 10)$ meters, respectively, and the $K = 20$ devices are uniformly distributed within the region spanning $([0, 20], [-10, 10], 0)$ meters. The communication channels \mathbf{h}_k and \mathbf{g} experience both path loss and small-scale fading. Specifically, the path loss is modeled as $\text{PL}(d) = \kappa_0(d/d_0)^{-\varrho}$, where $\kappa_0 = 30$ dB accounts for the path loss at a reference distance of $d_0 = 1$ meter, d is the link distance, and ϱ denotes the path loss exponent. Throughout the simulations, we set the path loss exponents for \mathbf{h}_k and \mathbf{g} to 2.8 and 2.2, respectively. For the small-scale fading, we employ the standard Rician channel model, assigning Rician factors of 0 and 2 to \mathbf{h}_k and \mathbf{g} , respectively. For the GA parameters,

the population size I is set to 80, crossover rate and mutation rate are set to 0.8 and 0.2, respectively. Regarding the phase-dependent amplitude response model in (4), we set $\phi = 0$, and $\gamma = 3$. The other parameters will be specified when used.

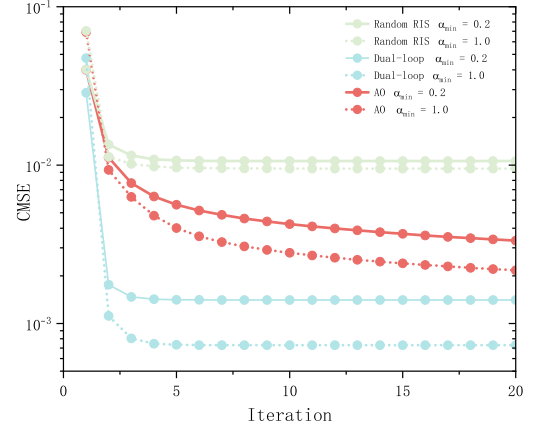


Fig. 2. Convergence behavior of the proposed algorithm, where $N = 200$, and $\text{SNR} \triangleq P_0/\sigma^2 = 110$ dB.

Fig. 2 presents the convergence behavior of our proposed dual-loop optimization algorithm. For comparison, two base-lines are also provided. **1) Random RIS:** the phase shifts of all REs are randomly selected without optimization. **2) AO:** Alternately optimizes η , \mathbf{b} , and ω in one loop, where η and \mathbf{b} are updated via (17) and (13), respectively, and ω is optimized via the successive convex approximation approach. As evident from Fig. 2, our proposed dual-loop algorithm is superior to the AO approach in reducing the CMSE. It also converges faster than the AO approach. Additionally, compared to full signal reflection by each RE (i.e., $\alpha_{\min} = 1.0$), the phase-dependent amplitude response leads to notable performance loss.

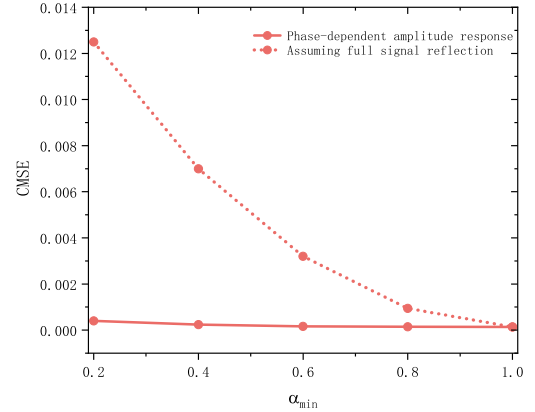


Fig. 3. Impact of the phase-dependent amplitude response on the CMSE, where $\text{SNR} = 110$ dB, and $N = 200$.

Fig. 3 further depicts the impact of the phase-dependent amplitude response model on the achieved CMSE. From this figure, we clearly observe that model mismatch severely degrades the performance.

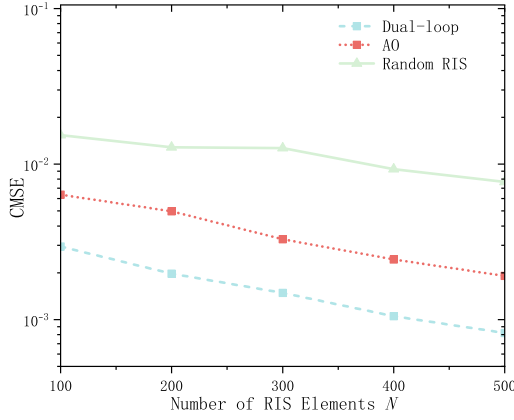


Fig. 4. CMSE versus number of RIS elements N , where $\alpha_{\min} = 0.2$, and SNR = 110 dB.

Fig. 4 shows the impact of varying the number of RIS elements on the CMSEs of the aforementioned three schemes. As shown in this figure, the MSEs achieved by all three schemes decrease as N increases, and our proposed dual-loop algorithm significantly outperforms the AO approach.

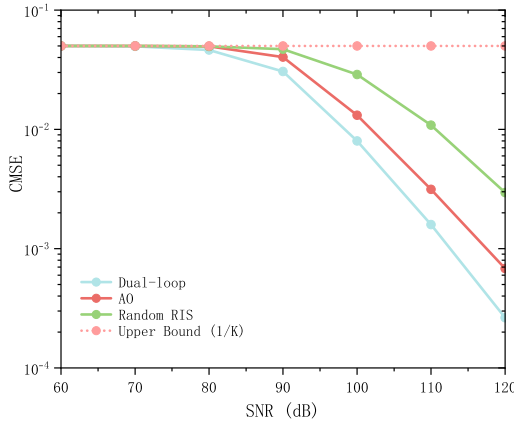


Fig. 5. CMSE versus SNR, where $\alpha_{\min} = 0.2$, and $N = 200$.

Fig. 5 depicts the impact of varying transmit SNR, defined as P_0/σ^2 , on the CMSE of the considered RIS-aided AirComp system. As evident from this figure, the CMSE achieved by using our proposed algorithm reduces notably as SNR increases. Furthermore, it is observed that the CMSE has an upper limit, $1/K$, as proven in Appendix A.

V. CONCLUSION

In this paper, we revisited the RIS-aided AirComp system by accounting for the non-ideal reflection model of the RIS. Under this new model, we proposed a dual-loop optimization framework to jointly optimize the AirComp transceiver and the RIS reflection matrix. Simulation results showed that the coupling between phase shift and amplitude of REs notably degraded the CMSE, and therefore properly designing the AirComp transceiver and the RIS reflection matrix was crucial.

APPENDIX A

Below we derive the upper bound on the MSE achieved by the RIS-assisted AirComp system. By substituting (17) into (6), we have

$$\begin{aligned} \overline{\text{CMSE}}(\mathbf{b}, \boldsymbol{\omega}) &= \frac{1}{K} - \frac{1}{K^2} \frac{(\sum_{k=1}^K |h_{e,k} b_k|)^2}{\sigma^2 + \sum_{k=1}^K |h_{e,k} b_k|^2} \\ &\leq \frac{1}{K}. \end{aligned}$$

REFERENCES

- [1] L. Xu *et al.*, "Internet of things in industries: A survey," *Transactions on Industrial Informatics*, vol. 10, no. 4, pp. 2233-2243, Nov. 2014.
- [2] E. Sisinni *et al.*, "Industrial internet of things: Challenges, opportunities, and directions," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 11, pp. 4724-4734, Nov. 2018.
- [3] Z. Wang *et al.*, "Over-the-air computation for 6G: Foundations, technologies, and applications," *IEEE Internet of Things Journal*, vol. 11, no. 14, pp. 24634-24658, July, 2024.
- [4] X. Cao *et al.*, "Optimized power control for over-the-air computation in fading channels," *IEEE Transactions on Wireless Communications*, vol. 19, no. 11, pp. 7498-7513, Nov. 2020.
- [5] W. Liu *et al.*, "Over-the-air computation systems: Optimization, analysis and scaling laws," *IEEE Transactions on Wireless Communications*, vol. 19, no. 8, pp. 5488-5502, Aug. 2020.
- [6] L. Chen *et al.*, "A uniform-forcing transceiver design for over-the-air function computation," *IEEE Wireless Communications Letters*, vol. 7, no. 6, pp. 942-945, Dec. 2018.
- [7] G. Zhu *et al.*, "MIMO over-the-air computation for high mobility multimodal sensing," *IEEE Internet of Things Journal*, vol. 6, no. 4, pp. 6089-6103, Aug. 2019.
- [8] Q. Wu *et al.*, "Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming," *IEEE Transactions on Wireless Communications*, vol. 18, no. 11, pp. 5394-5409, Nov. 2019.
- [9] T. Jiang *et al.*, "Over-the-air computation via intelligent reflecting surfaces," *Proceedings of the IEEE Global Communications Conference (GLOBECOM)*, pp. 1-6, Waikoloa, HI, USA, 2019.
- [10] X. Zhai *et al.*, "Joint beamforming aided over-the-air computation systems relying on both BS-side and user-side reconfigurable intelligent surfaces," *IEEE Transactions on Wireless Communications*, vol. 21, no. 12, pp. 10766-10779, Dec. 2022.
- [11] S. Abeywickrama *et al.*, "Intelligent reflecting surface: Practical phase shift model and beamforming optimization," *IEEE Transactions on Communications*, vol. 68, no. 9, pp. 5849-5863, Sept. 2020.