Joint Antenna Selection and Beamforming for Parallel Over-the-Air Computing with Limited RF Chains

Yunting Chen*, Bowen Li†, Sai Zhao*, Junting Chen†, and Dong Tang*
* School of Electronics and Communication Engineering, Guangzhou University, Guangzhou 510006, China
† School of Science and Engineering (SSE) and Shenzhen Future Network of Intelligence Institute (FNii-Shenzhen)
The Chinese University of Hong Kong, Shenzhen, Guangdong 518172, China
Email: 1819500109@e.gzhu.edu.cn, bowenli6@link.cuhk.edu.cn, zhaosai@gzhu.edu.cn, juntingc@cuhk.edu.cn
and tangdong@gzhu.edu.cn

***Abstract*—This paper addresses the challenge of training multiple models in parallel via federated learning with over-the-air computation (AirComp) in multiuser multi-stream multiple-input multiple-output (MIMO) systems. With the rapid development of artificial intelligence (AI), many Internet-of-Things (IoT) and industrial applications require training and updating several specialized models. As state-of-the-art sensors usually equip with multiple antennas, it is possible to transmit multiple data streams to carry the parameters of multiple models and to be aggregated in parallel via AirComp at the access point (AP). However, existing AirComp techniques primarily focus on single-model training, and it remains a significant challenge for aggregating signals dedicated for training multiple models separately and simultaneously. In addition, due to hardware complexity in practice, a small number of radio frequency (RF) chains need to be shared among the array of antennas at the AP. As such, this paper proposes a joint antenna selection (AS) and beamforming design for AirComp that leverages antenna selection to minimize model-by-model signal aggregation error. Specifically, a mixed integer programming is formulated. Based on the trace inverse property and an integration of semidefinite relaxation (SDR), Schur decomposition, and Lagrangian relaxation (LR), a new optimization strategy is developed to transform and relax the problem into a series of easy-to-solve convex subproblems, leading to an efficient algorithm. Numerical results confirm that with AS, when doubling the antennas from the RF chains, more than 3dB gain can be achieved.**

***Index Terms*—Over-the-air computation (AirComp), multiple-input multiple-output (MIMO), antenna selection (AS), beamforming, semidefinite relaxation (SDR), Schur decomposition, Lagrangian relaxation (LR).**

I. Introduction

The strong demands on the development of AI drive the non-stopping collection and processing of data from various applications for model training. While the rapid growth of wireless communications and IoT have provided infrastructure that connects a large number of devices for sensing, data collection, and model training, transporting and fusing the vast amount of data by the massive devices imposes a significant burden on wireless communication networks. To address this challenge, federated learning with AirComp offers a promising solution to relieving the communication and computation burden in the communication network, where, instead of uploading the massive amount of raw data, the mobile devices may preprocess the data and transmit the updated model parameters in a synchronized way such that the parameters transmitted from difference devices are algebraically combined together when they arrive at the receiver of the fusion center; the fusion center then broadcasts the updated model parameters to the devices, thus forming a distributive model training process [1]–[4].

While there has been active research on AirComp in recent years, most of them focused on learning a single model, where one data stream is transmitted for each device. However, there are many application scenarios that require training or updating multiple models simultaneously. For example, in industrial IoT, many specialized models such as defect detection, environment sensing, and robot control are needed. On the other hand, many mobile devices and sensors are equipped with multiple antennas and have the capability to deliver multiple data streams. As a result, it is possible and will be more efficient to train multiple models *in parallel* using multiple data streams.

However, implementing multiuser multi-stream MIMO transmission can be very challenging. First, the channel condition can be poor for the worst data stream due to possibly insufficient scattering. Such an issue can be resolved by equipping more receiver antennas at the fusion center so that more propagation paths can be resolved. Second, the hardware complexity can be very high if each antenna is equipped with a dedicated RF chain. Thus, a compromise solution is to deploy more antennas at the AP while adopting AS, which connects multiple antennas to a limited number of RF chains, to boost the system performance without significantly increasing the costs.

Conventional AirComp focuses on the mean squared error (MSE) performance of signal aggregation at the receiver and centers on beamformer design, typically without considering AS. A dual-functional beamforming scheme is designed to simultaneously support MIMO integrated sensing and AirComp (ISAA) [5]. Combining wireless power transfer with aggregation beamforming optimizes AirComp computation accuracy under energy constraints [6]. AS plays a significant role in MIMO communication precoding, as it can reduce both hardware complexity and algorithm execution complexity. Reference [7] focuses on reconfigurable intelligent surface (RIS)-assisted multiuser scenarios, where low-complexity AS and phase-shift design are used to enhance channel gain and system throughput. Reference [8] proposes an AS scheme based on alternating optimization (AO) to reduce design complexity and address the secure beamforming design problem for large-scale MIMO eavesdropping channels. However, traditional design methods aim to recover individual data streams, making it challenging to directly extend to the AirComp scenario. Recent work in AirComp has discussed the joint issues related to AS and federated learning. Reference [9] investigates over-the-air federated learning techniques in large-scale MIMO systems with a limited number of RF chains. A multi-AS method combining the sparsification of matrix diagonals is proposed to optimize the receive scaling vector and AS, with an objective to reduce the MSE of AirComp [10]. While this method is effective in reducing the MSE, it overlooks the limitations of RF chains in massive MIMO systems, resulting in insufficient spatial diversity and reduced interference mitigation capabilities at the receiver, which directly impacts reception performance. Existing AS schemes in large-scale MIMO systems face high computational complexity and limited practicality. For example, in multiuser scenarios, they struggle to effectively balance performance optimization with hardware resource constraints and fail to achieve efficient performance optimization in complex environments.

In this paper, we consider AS and beamforming design for federated learning with AirComp for the training of multiple models in parallel, where the users transmit multiple data streams and the receiver, the fusion center (FC), combines the signal using a subset of the antenna due to the RF-chain constraint. A mixed integer programming problem for joint AS and beamforming is formulated. For optimal reception, the minimum mean squared error (MMSE) method is employed to optimize the transmit beamformer. This approach introduces coupling between variables within a trace inverse function and, combined with the integer nature of the antenna selection variable, poses a significant challenge for solution derivation. To address these challenges, we employ the block coordinate descent (BCD) technique to decompose the non-convex integer programming problem into two subproblems for antenna selection and beamforming. Subsequently, we apply SDR, Schur decomposition, LR, and successive convex approximation (SCA) methods to transform and relax the complex non-convex problem into a series of simpler convex problems. An iterative algorithm is then developed to jointly optimize antenna selection and beamforming variables. Simulation results indicate that to achieve the same MSE of system, the proposed joint design requires significantly fewer RF chains compared to the case without AS, thereby substantially reducing hardware costs.

II. System Model and Problem Formulation

Consider a MIMO system with an AP equipped with M antennas and K sensors each equipped with N antennas. The AP has L RF chains, as shown in Fig. 1. Each sensor transmits $d \leq L$ data streams to the AP. For each data stream $i = 1, 2, \ldots, d$, the

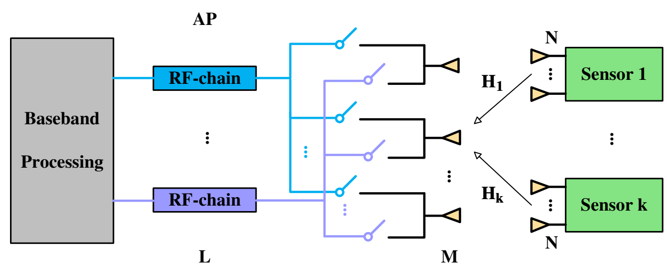

Fig. 1. MIMO AirComp system with joint AS and beamforming, with each RF chain selecting only one antenna.

time-frequency resources are shared among all the K users, and hence, the signals of the ith data stream from all the K users are superposed at the AP due to the nature of radio propagation. Specifically, consider a flat fading channel, and denote $\mathbf{H}_k \in \mathbb{C}^{M\times N}$ as the uplink channel from the kth sensor to the AP. Let $\mathbf{W}_k \in \mathbb{C}^{N\times d}$ be the precoder of the kth sensor, where d is the number of data streams. Denote $\mathbf{s}_k = [s_{k1}, s_{k2}, \ldots, s_{kd}]^T \in \mathbb{C}^{d\times 1}$ as the d-dimensional transmit signal of the kth user. It is assumed that the transmission power has been normalized as $E[\mathbf{s}_k\mathbf{s}_k^H] = \mathbf{I}$ and the data streams are orthogonal. As a result, the signal arrived at the antenna array of the AP is given by $\mathbf{y} = \sum_{k=1}^{K} \mathbf{H}_k\mathbf{W}_k\mathbf{s}_k + \mathbf{n}$, where $\mathbf{n} \in \mathbb{C}^{M\times 1}$ models the receiver noise. In the application of AirComp, the goal is to recover the aggregated signal $\mathbf{s} \triangleq \sum_k \mathbf{s}_k$.

Consider the limited RF chain case, where $L < M$. Thus, the AP needs to select L antennas from all M antennas and connect them to the L RF chains. Denote $\mathbf{\Delta} \in \mathbb{C}^{M\times L}$ as the antenna selection matrix, where $\triangle_{ml} \in \{0, 1\}$ denotes elements of $\mathbf{\Delta}$. Let $\mathbf{U} \in \mathbb{C}^{L\times d}$ be the decorrelation matrix. As a result, the decorrelated signal is given by $\hat{\mathbf{s}} = \mathbf{U}^H\mathbf{\Delta}^H \sum_{k=1}^{K} \mathbf{H}_k\mathbf{W}_k\mathbf{s}_k + \mathbf{U}^H\mathbf{\Delta}^H\mathbf{n}$. Our objective is to minimize the MSE defined as $\mathsf{MSE}(\mathbf{U}, \mathbf{\Delta}, \mathbf{W}) = \mathbb{E}\left[\mathrm{tr}\,(\hat{\mathbf{s}} - \mathbf{s})\,(\hat{\mathbf{s}} - \mathbf{s})^H\right]$, which can be derived as

$$\begin{aligned}\mathsf{MSE}(\mathbf{U}, \mathbf{\Delta}, \mathbf{W}) = \sum_{k=1}^{K} \mathrm{tr}\big[&(\mathbf{U}^H\mathbf{\Delta}^H\mathbf{H}_k\mathbf{W}_k - \mathbf{I})\\ &(\mathbf{U}^H\mathbf{\Delta}^H\mathbf{H}_k\mathbf{W}_k - \mathbf{I})^H\big] + \sigma^2\mathrm{tr}(\mathbf{\Delta}\mathbf{U}\mathbf{U}^H\mathbf{\Delta}^H) \qquad (1)\end{aligned}$$

where $\mathbf{W}$ is the collection of all beamforming matrices $\mathbf{W}_k$ at the sensors.

The joint antenna selection and beamforming problem $P1$ is formulated as follows:

$$\begin{aligned}\underset{\mathbf{U},\{\mathbf{W}_k\},\mathbf{\Delta}}{\text{minimize}}\quad & \mathrm{tr}\big[(\mathbf{U}^H\mathbf{\Delta}^H\mathbf{H}_k\mathbf{W}_k - \mathbf{I})(\mathbf{U}^H\mathbf{\Delta}^H\mathbf{H}_k\mathbf{W}_k\\ & - \mathbf{I})^H\big] + \sigma^2\mathrm{tr}(\mathbf{\Delta}\mathbf{U}\mathbf{U}^H\mathbf{\Delta}^H) && (2a)\\ \text{subject to}\quad & \mathrm{tr}\left(\mathbf{W}_k\mathbf{W}_k^H\right) \leq P, \quad \forall k && (2b)\\ & \sum_m \Delta_{ml} = 1, \quad \forall l && (2c)\\ & \sum_l \Delta_{ml} \leq 1, \quad \forall m && (2d)\\ & \Delta_{ml} \in \{0, 1\}, \quad \forall m, \forall l. && (2e)\end{aligned}$$

Constraint (2b) represents the transmit power constraint to ensure the transmit power is less than P. Constraint (2c) indicates that each RF chain must select one antenna, while constraint (2d) indicates that each antenna can be assigned to at most one RF chain. Constraint (2e) represents the binary entries of the antenna selection matrix.

III. Antenna Selection and Beamforming Design

For the AirComp system with antenna selection at the AP, the transmit beamformer can be written as a function of the receive beamformer $\mathbf{U}$ and antenna selection matrix $\boldsymbol{\Delta}$, by applying the MMSE method. Since the MSE function over the transmit beamformer $\mathbf{W}_k$, as defined in (1), is convex. The optimal transmit beamformer, denoted as $(\mathbf{W}_k)_{\text{opt}}$, is the solution to the following equation

$$\frac{\partial}{\partial \mathbf{W}_k}\mathsf{MSE}(\mathbf{U}, \boldsymbol{\Delta}, \mathbf{W}) = 0.$$

Consequently, we obtain the optimal transmit beamforming value $(\mathbf{W}_k)_{\text{opt}}$ of the sensors as follows

$$(\mathbf{W}_k)_{\text{opt}} = \left(\mathbf{H}_k^H \boldsymbol{\Delta}\mathbf{U}\mathbf{U}^H\boldsymbol{\Delta}^H\mathbf{H}_k\right)^{-1}\mathbf{H}_k^H\boldsymbol{\Delta}\mathbf{U}, \quad \forall k. \tag{3}$$

Substituting (3) into $P1$, we have

$$\begin{aligned}
P2: \quad & \underset{\mathbf{U},\boldsymbol{\Delta}}{\text{minimize}} \quad \sigma^2\,\text{tr}(\boldsymbol{\Delta}\mathbf{U}\mathbf{U}^H\boldsymbol{\Delta}^H) && \text{(4a)}\\
& \text{subject to} \quad \text{tr}\left[\left(\mathbf{H}_k^H\boldsymbol{\Delta}\mathbf{U}\mathbf{U}^H\boldsymbol{\Delta}^H\mathbf{H}_k\right)^{-1}\right] \le P,\ \forall k && \text{(4b)}\\
& \sum_m \Delta_{ml} = 1,\ \forall l && \text{(4c)}\\
& \sum_l \Delta_{ml} \le 1,\ \forall m && \text{(4d)}\\
& \Delta_{ml} \in \{0, 1\},\ \forall m, \forall l. && \text{(4e)}
\end{aligned}$$

The problem $P2$ is non-trivial to solve due to the coupling of variables in both the non-convex objective function (4a) and constraint (4b), as well as the integer nature of the antenna selection variable $\boldsymbol{\Delta}$. To tackle this non-convex integer programming problem, we first utilize the BCD technique, decomposing the problem $P2$ into two sub-problems. Specifically, we first address the beamforming component, applying the SDR method to manage the non-convexity. Next, we tackle the antenna selection sub-problem. To resolve issues with matrix trace inverse, we employ Shure decomposition. Additionally, we apply LR to relax the equality constraints and SCA to iteratively transform the non-convex problem into a series of convex problems for more efficient optimization.

A. Receive Beamforming Matrix Optimization

Given the antenna selection matrix $\boldsymbol{\Delta}$, the optimization problem $P2$ for the beamforming matrix $\mathbf{U}$ becomes

$$P3: \quad \min_{\mathbf{U}} \quad \sigma^2\,\text{tr}[\boldsymbol{\Delta}\mathbf{U}\mathbf{U}^H\boldsymbol{\Delta}^H] \quad \text{s.t. (4b).}$$

The problem $P3$ is non-convex because the function $\text{tr}[(\mathbf{H}_k^H\boldsymbol{\Delta}\mathbf{U}\mathbf{U}\boldsymbol{\Delta}\mathbf{H}_k)^{-1}]$ is non-convex over variable $\mathbf{U}$.

To avoid this issue, we introduce a new variable to slove this point which is $\hat{\mathbf{U}} \triangleq \mathbf{U}\mathbf{U}^H$, then the objective function becomes $\sigma^2\,\text{tr}(\boldsymbol{\Delta}\hat{\mathbf{U}}\boldsymbol{\Delta}^H)$, and the allocated power in (4b) is transformed in the same way. Accordingly, the problem $P3$ is transformed to

$$\begin{aligned}
P4: \quad & \underset{\hat{\mathbf{U}}}{\text{minimize}} \quad \sigma^2\text{tr}(\boldsymbol{\Delta}\hat{\mathbf{U}}\boldsymbol{\Delta}^H) && \text{(5a)}\\
& \text{subject to} \quad \text{tr}\left[\left(\mathbf{H}_k^H\boldsymbol{\Delta}\hat{\mathbf{U}}\boldsymbol{\Delta}^H\mathbf{H}_k\right)^{-1}\right] \le P,\ \forall k && \text{(5b)}\\
& \text{rank}(\hat{\mathbf{U}}) = d && \text{(5c)}\\
& \hat{\mathbf{U}} \succeq 0 && \text{(5d)}
\end{aligned}$$

where the objective function (5a) and the constraint (5b) over $\hat{\mathbf{U}}$ are convex, constraint (5c) must satisfy because the rank of $\mathbf{U}$ is equal to the rank of $\hat{\mathbf{U}}$, that is, $\min d, L = d$, and constraint (5d) is because $\hat{\mathbf{U}}$ is hermitian. Note the rank constraint (5c) is non-convex.

Then we use SDR technique [11] to relax the non-convex constraint (5c), then problem $P4$ becomes

$$P5: \quad \min_{\hat{\mathbf{U}}} \quad \sigma^2\,\text{tr}[\boldsymbol{\Delta}\hat{\mathbf{U}}\boldsymbol{\Delta}^H] \quad \text{s.t. (5b), (5d).}$$

The problem $P5$ is convex since the trace of a matrix and the trace inverse of a matrix is convex.

Denote the solution to the problem $P5$ as $\hat{\mathbf{U}}_{\text{opt}}$. Then, we construct the beamforming vector $\mathbf{U}$ by singular value decomposition (SVD). Specifically, we select the largest d eigenvalues and their corresponding eigenvectors to form a beamforming matrix $\mathbf{U}$,[1] and the scaling method is applied to ensure that the power constraint (5b) is satisfied.

B. Antenna Selection Matrix Optimization

Given the beamforming matrix $\mathbf{U}$, the optimization problem $P2$ for the antenna selection matrix $\boldsymbol{\Delta}$ is

$$P6: \quad \min_{\boldsymbol{\Delta}} \quad \sigma^2\,\text{tr}[\boldsymbol{\Delta}\mathbf{U}\mathbf{U}^H\boldsymbol{\Delta}^H] \quad \text{s.t. (4b) – (4e).}$$

In addition to the non-convex constraint trace inverse function in constraint (4b), problem $P6$ has an additional integer-programming issue.

Similar to Section III-A deals with $\mathbf{U}$, we introduce a new variable $\mathbf{A} \triangleq \boldsymbol{\Delta}\mathbf{U}\mathbf{U}^H\boldsymbol{\Delta}^H$ to handle non-convexity in the objective function (4a) and constraint (4b), where $\mathbf{A} \succeq 0$. Then the objective function (4a) becomes convex over $\mathbf{A}$, as $\sigma^2\,\text{tr}[\mathbf{A}]$, and the constraint (4b) becomes convex over $\mathbf{A}$, as

$$\text{tr}\left[\left(\mathbf{H}_k^H\mathbf{A}\mathbf{H}_k\right)^{-1}\right] \le P,\ \forall k. \tag{6}$$

Note that the matrix $\mathbf{A}$ must be non-full rank because of the antenna selection constraint (4d). Thus, there is no inverse for the matrix $\mathbf{H}_k^H\mathbf{A}\mathbf{H}_k$. Therefore, we must avoid the existence of the matrix inverse. Here, we use Shure decomposition to transform the (6) with matrix inverse for a non-full-rank matrix, as

$$\begin{bmatrix} t\mathbf{I} & \mathbf{I} \\ \mathbf{I} & (\mathbf{H}_k^H\mathbf{A}\mathbf{H}_k) \end{bmatrix} \succeq 0,\ \forall k,\ \text{tr}[t\mathbf{I}] \le P \tag{7}$$

[1]Here, we select the largest eigenvalues, rather than relying on Gaussian randomization, to ensure stable iterations on $\boldsymbol{\Delta}$.

where t is a new variable.

According to Theorem 4.3 in [12], $[t\mathbf{I}, \mathbf{I}; \mathbf{I}, (\mathbf{H}_k^H \mathbf{A} \mathbf{H}_k)] \succeq 0$ leads to $\mathbf{H}_k^H \mathbf{A} \mathbf{H}_k - t\mathbf{I} \succeq 0$. Then $\mathrm{tr}\left[\left(\mathbf{H}_k^H \mathbf{A} \mathbf{H}_k\right)^{-1}\right] \leq P$ must hold if $\mathrm{tr}[t\mathbf{I}] \leq P$. Therefore constraint (7) is equivalent to constraint (6), but without matrix inverse.

The introduction of $\mathbf{A}$ and t transpose the non-convex objective (4a) and constraint (4b) to be convex, but one equality constraint is introduced, that is $\mathbf{A} = \boldsymbol{\Delta}\mathbf{U}\mathbf{U}^H\boldsymbol{\Delta}^H$. Together with equality constraint (4c) and integer constraint (4e), we have three problems to solve.

Firstly, we equivalently transform the equality constraint $\mathbf{A} = \boldsymbol{\Delta}\mathbf{U}\mathbf{U}^H\boldsymbol{\Delta}^H$ to

$$\begin{bmatrix} \mathbf{A} & \boldsymbol{\Delta}\mathbf{U} \\ [\boldsymbol{\Delta}\mathbf{U}]^H & \mathbf{I} \end{bmatrix} \succeq 0,\ \mathrm{tr}(\mathbf{A} - \boldsymbol{\Delta}\mathbf{U}\mathbf{U}^H\boldsymbol{\Delta}^H) \leq 0$$

according to Lemma 2 in [13]. Then, using the LR technique [14], the integer constraint (4e) is equivalent to

$$0 \leq \Delta_{ml} \leq 1,\ \Delta_{ml}(1 - \Delta_{ml}) \leq 0.$$

Finally, convert the equality constraint (4b) to two inequality constraints as

$$\sum_m \Delta_{ml} \leq 1,\ \sum_m \Delta_{ml} \geq 1, \quad \forall l.$$

Combine the transportation above, the Problem $P6$ is equivalently transformed to

$$\begin{aligned}
P7: \quad & \underset{\boldsymbol{\Delta},\mathbf{A},t}{\text{minimize}} \quad \sigma^2\,\mathrm{tr}[A] && \text{(8a)}\\
& \text{subject to} \quad \begin{bmatrix} t\mathbf{I} & \mathbf{I} \\ \mathbf{I} & (\mathbf{H}_k^H \mathbf{A} \mathbf{H}_k) \end{bmatrix} \succeq 0,\ \forall k && \text{(8b)}\\
& \mathrm{tr}[t\mathbf{I}] \leq P && \text{(8c)}\\
& \mathrm{tr}[\mathbf{A} - \boldsymbol{\Delta}\mathbf{U}\mathbf{U}^H\boldsymbol{\Delta}^H] \leq 0 && \text{(8d)}\\
& \begin{bmatrix} \mathbf{A} & \boldsymbol{\Delta}\mathbf{U} \\ [\boldsymbol{\Delta}\mathbf{U}]^H & \mathbf{I} \end{bmatrix} \succeq 0 && \text{(8e)}\\
& \sum_l \Delta_{ml} \leq 1, \quad \forall m && \text{(8f)}\\
& \sum_l \Delta_{ml} \geq 1, \quad \forall m && \text{(8g)}\\
& 0 \leq \Delta_{ml} \leq 1, \quad \forall m, \forall l && \text{(8h)}\\
& \Delta_{ml}(1 - \Delta_{ml}) \leq 0, \quad \forall m, \forall l && \text{(8i)}\\
& \mathbf{A} \succeq 0. && \text{(8j)}
\end{aligned}$$

The Lagrange multiplier method is performed to ensure that the problem can gradually converge to feasible points. Specifically, introduce three Lagrangian variables $\tau \geq 0$, $\beta_m \geq 0$, and $\rho_{ml} \geq 0$, corresponding to constraints (8d), (8g), and (8i). Then, the problem $P7$ can be converted to

$$\begin{aligned}
P8: \quad \underset{\boldsymbol{\Delta},\mathbf{A},t}{\text{minimize}} \quad & \sigma^2\,\mathrm{tr}[A] + \tau\,\mathrm{tr}[\mathbf{A} - \boldsymbol{\Delta}\mathbf{U}\mathbf{U}^H\boldsymbol{\Delta}^H] \\
& + \sum_m \beta_m (1 - \sum_l \Delta_{ml}) \\
& + \sum_m \sum_l \rho_{ml}\Delta_{ml}(1 - \Delta_{ml}) \\
\text{subject to} \quad & \text{(8b), (8c), (8e), (8f), (8h), (8j).}
\end{aligned}$$

Note that as $\tau \to \infty$, $\beta_m \to \infty$, and $\rho_{ml} \to \infty$, the solution to problem $P8$ converges to the solution to problem $P7$. This is because: 1) $\mathrm{tr}[\mathbf{A} - \boldsymbol{\Delta}\mathbf{U}\mathbf{U}^H\boldsymbol{\Delta}^H]$ is constrained to be greater or equal to 0 because of the constraint (8e), which lets $\mathbf{A} - \boldsymbol{\Delta}\mathbf{U}\mathbf{U}^H\boldsymbol{\Delta}^H \succeq 0$, *i.e.*, $\mathrm{tr}[\mathbf{A} - \boldsymbol{\Delta}\mathbf{U}\mathbf{U}^H\boldsymbol{\Delta}^H] \geq 0$; 2) $(1 - \sum_l \Delta_{ml})$ is constrained to be greater or equal to 0 because of the constraint (8f); and 3) $\Delta_{ml}(1 - \Delta_{ml})$ is constrained to be greater or equal to 0 because of the constraint (8h).

However, the problem $P8$ is still non-convex because of the non-convexity of $\mathrm{tr}[-\boldsymbol{\Delta}\mathbf{U}\mathbf{U}^H\boldsymbol{\Delta}^H]$ and $-\Delta_{ml}^2$ in the objective function. Then, we solve this problem by iteratively approximating them with more straightforward and convex problems using SCA [15]. Specifically, given the local antenna selection variable $\tilde{\boldsymbol{\Delta}}$, using Taylor expansion, we have

$$\mathrm{tr}[\boldsymbol{\Delta}\mathbf{U}\mathbf{U}^H\boldsymbol{\Delta}^H] \approx 2\mathrm{tr}[\mathbf{U}\mathbf{U}^H\tilde{\boldsymbol{\Delta}}^H\boldsymbol{\Delta}] - \mathrm{tr}[\tilde{\boldsymbol{\Delta}}\mathbf{U}\mathbf{U}^H\tilde{\boldsymbol{\Delta}}^H]$$

$$\Delta_{ml}^2 \approx 2\tilde{\Delta}_{ml}\Delta_{ml} - \tilde{\Delta}_{ml}^2$$

Then, the objective function in problem $P8$ is approximated at the point $\tilde{\boldsymbol{\Delta}}$ to

$$\begin{aligned}
& \sigma^2\,\mathrm{tr}[A] + \tau\left(\mathrm{tr}[\mathbf{A}] - \mathrm{tr}[\mathbf{U}\mathbf{U}^H\tilde{\boldsymbol{\Delta}}^H\boldsymbol{\Delta}] + \mathrm{tr}[\tilde{\boldsymbol{\Delta}}\mathbf{U}\mathbf{U}^H\tilde{\boldsymbol{\Delta}}^H]\right) \\
& + \sum_m \beta_m (1 - \sum_l \Delta_{ml}) \\
& + \sum_m \sum_l \rho_{ml}\left(\left(1 - 2\tilde{\Delta}_{ml}\right)\Delta_{ml} + \tilde{\Delta}_{ml}^2\right). \qquad (10)
\end{aligned}$$

Accordingly, the non-convex problem $P8$ becomes a convex problem locally as

$$P9: \quad \min_{\boldsymbol{\Delta},\mathbf{A},t} \quad (10) \quad \text{s.t. (8b), (8c), (8e), (8f), (8h), (8j).}$$

C. Antenna Selection and Beamforming Algorithm

Based on the problem decomposition, transformation, and relaxation in Section III-A and III-B, we proposed a joint antenna selection and beamforming optimization algorithm, as shown in Algorithm 1.

To avoid being trapped in a local optimum due to the integer nature of the variable $\boldsymbol{\Delta}$, we group the Lagrangian parameters

Algorithm 1 Antenna selection and beamforming optimization

INPUT: $\mathbf{H}_k$, $k \in \{1, 2, \cdots, K\}$;
1: Initialize: $\boldsymbol{\Delta}^{(0)}$, $\beta_m^{(0)}$, and $\rho_{ml}^{(0)}$, and $i \leftarrow 1$;
2: Obtain $\hat{\mathbf{U}}^{(i)}$ by solving problem $P5$ and construct $\mathbf{U}^{(i)}$ by SVD decomposition.
3: Initialize $\tau^{(0)}$, set $\boldsymbol{\Delta}^{(i,0)} \leftarrow \boldsymbol{\Delta}^{(i-1)}$ and set $j \leftarrow 1$;
4: Obtain $\boldsymbol{\Delta}^{(i,j)}$ and $\mathbf{A}^{(i,j)}$ by solving problem $P9$, where $\tilde{\boldsymbol{\Delta}} \leftarrow \boldsymbol{\Delta}^{(i-1,j)}$;
5: Update $\tau^{(j)}$ according to (13);
6: Go to step 4 and set $j \leftarrow j + 1$ until $|\mathbf{A}^{(i,j)} - \boldsymbol{\Delta}^{(i,j)}\mathbf{U}^{(i)}\mathbf{U}^{(i)H}\boldsymbol{\Delta}^{(i)H}| \to 0$
7: Update $\beta_m^{(i)}$ and $\rho_{ml}^{(i)}$ according to (12) and (11); Set $\boldsymbol{\Delta}^{(i)} \leftarrow \boldsymbol{\Delta}^{(i,j)}$;
8: Go to step 2 and $i \leftarrow i + 1$ until $|\boldsymbol{\Delta}^{(i)} - \boldsymbol{\Delta}^{(i-1)}| \to 0$
OUTPUT: $\boldsymbol{\Delta}^{(i)}$ and $\mathbf{U}^{(i)}$;

into two parts: $\{\tau\}$ and $\{\beta_m, \rho_{ml}\}$, and design a two-layer optimization framework. In the inner layer, we iteratively update $\{\tau\}$ to ensure that the equality constraint $\mathbf{A} = \mathbf{\Delta}\mathbf{U}\mathbf{U}^H\mathbf{\Delta}^H$ is satisfied at each outer iteration, as shown in steps 3 to 5 in Algorithm 1. In the outer layer, we iteratively update $\{\beta_m, \rho_{ml}\}$ to ensure that $\sum_m \Delta_{ml} = 1, \forall l$and $\Delta_{ml} \in \{0, 1\}, \forall m, \forall l$, ultimately, meet these requirements, as shown in steps 2 and 7 to 8 in Algorithm 1.

In detail, the Lagrangian parameters are updated based on their gradients, that is

$$\rho_{ml}^{(i+1)} = \rho_{ml}^{(i)} + \lambda_1 \Delta_{ml}^{(i)}(1 - \Delta_{ml}^{(i)}) \tag{11}$$

$$\beta_m^{(i+1)} = \beta_m^{(i+1)} + \lambda_2(1 - \sum_l \Delta_{ml}^{(i)}) \tag{12}$$

$$\tau^{(j+1)} = \tau^{(j+1)} + \lambda_3 \operatorname{tr}[\mathbf{A}^{(i,j)} - \mathbf{\Delta}^{(i,j)}\mathbf{U}^{(i)}\mathbf{U}^{(i)H}\mathbf{\Delta}^{(i,j)H}] \tag{13}$$

where $\{\lambda_1, \lambda_2, \lambda_3\}$ is the iteration step size.

IV. Simulation Results

A. Parameter Setting

In this section, we evaluate the performance of the proposed AirComp system with joint beamforming and AS through simulations. Unless specified otherwise, the simulation parameters are set as follows. The number of sensors is $K = 2$, the transmission power is $P = 10$, the number of transmit antennas is $N = 3$, the data stream length is $d = 2$, the total number of receive antennas is $M = 10$, and the number of selected antennas is $L = 5$. In the simulations, we maintain the following relationship: $M \geq L \geq N \geq d$. We assume that each MIMO channel experiences independent and identically distributed (i.i.d.) Rayleigh fading and is modeled as i.i.d. complex Gaussian random variables with nonzero mean $\nu = 0$, variance $\sigma_1^2 = 10$ and $\sigma_2^2 = 1$ respectively.

In the simulations, we also set up the following comparison algorithm. Different from our proposed scheme with antenna selection (W\ Antenna selection) which selects the optimal L receive antennas out of M to maximize system accuracy, the comparison algorithm without antenna selection (W\o Antenna selection) simply fixes the selection of the first L receive antennas.

B. Simulation Result

Fig. 2 shows the relationship between the MSE of the AirComp system and the transmission power of the sensors. From Fig. 2, it can be observed that the system MSE of both schemes decreases as the transmit power increases. This is because as the transmit power P increases, the effective power of the signal also grows and hence the signal-to-noise ratio (SNR) for each sensor signal also enhances, thereby reducing the system MSE. Meanwhile, increasing the total number of receive antennas M enhances the spatial diversity gain in AirComp, allowing more effective MMSE-based optimal beamforming, which further reduces the MSE. As shown in Fig. 2, the W\ Antenna selection scheme we adopted outperforms the comparison algorithm W\o Antenna selection.

Fig. 3 illustrates the trend of system MSE as the number of selected antennas L varies under a fixed total number of receive antennas $M = 10$. It can be observed that the system MSE decreases as the number of selected antennas increases. The reason is that selecting the optimal L antennas in an MIMO system maximizes signal quality and reduces MSE. Moreover, choosing an optimal subset of antennas reduces the reliance on RF chains of system, significantly lowering hardware costs. As shown in Fig. 3, the W\ Antenna selection scheme we adopted is better than the comparison algorithm W\o Antenna selection. It is worth noting that for the cases of $L = 6$ and $L = 8$, the system MSE at a transmission power of $P = 2$ is slightly higher than the baseline MSE at $P = 4$. This is because, for $P > 5$, the transmission power has a dominant effect on the MSE of system. As shown in Fig. 3, the difference in MSE between the

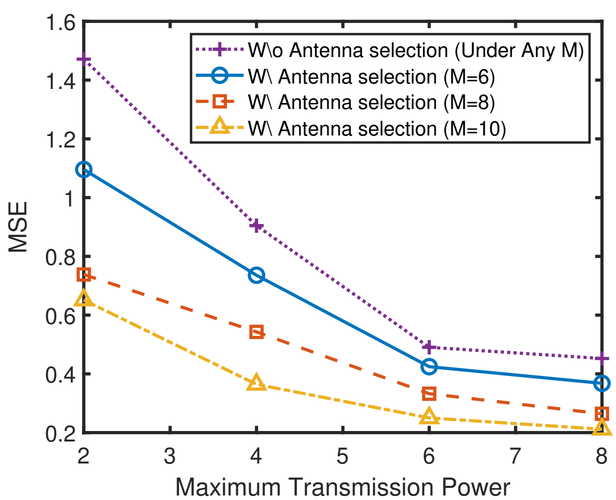

Fig. 2. The effects of sensor transmission power on the computation error of AirComp, with $L = 5$.

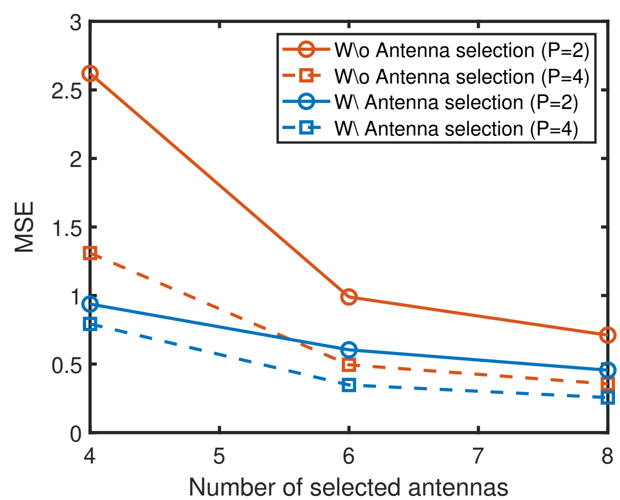

Fig. 3. The effects of sensor selected antenna numbers on the computation error of AirComp, with $M = 10$.

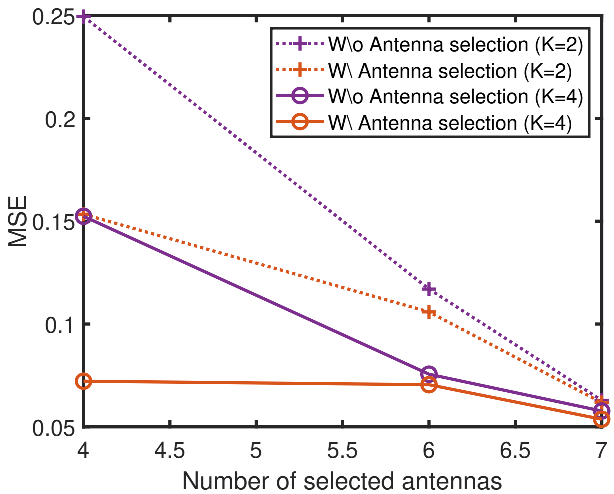

Fig. 4. The effects of sensor numbers on the computation error of AirComp, with $P = 4$ and $M = 10$.

two cases is less than 0.2, but our proposed algorithm reduces the dependency of system on the RF-chain, significantly lowering the signal transmission cost. This is a key focus of our research.

Fig. 4 depicts the trend of system MSE as the number of selected antennas L increases, for sensor counts of $K = 2$ and $K = 4$, with a fixed transmission power of $P = 4$ and a total of $M = 10$ receive antennas. It can be observed that the MSE of system is significantly reduced when the number of sensor is 4 compared to when it is 2. This reduction occurs because, with the increase in the number of sensors, the system can obtain more effective signals. These signals processed using the AirComp technique, reduce the impact of noise on the system, thereby enhancing the signal processing ability of the system and improving its accuracy, which results in a lower MSE. Last but not least, the W\ Antenna selection scheme we adopted outperforms the comparison algorithm W\o Antenna selection.

V. Conclusion

In this paper, we reduce the dependency of parallel AirComp systems on the RF-chain by proposing a joint AS and beamforming scheme. This study presents four key technical contributions. First, we replace the commonly used zero forcing (ZF) algorithm with the MMSE algorithm to solve the optimal transmission beamforming problem, as the MMSE algorithm takes both channel gain and noise into account during the optimization process, it enables the AirComp system to achieve better computation accuracy under low SNR conditions. Second, we adopt the BCD method instead of the traditional AO algorithm to solve the joint optimization problem of the receive beamformer and AS matrix. This approach effectively prevents the problem from prematurely converging to a local optimum. Next, when applying SDR and separating the optimal receive beamforming values, we choose eigenvalue decomposition over the traditional Gaussian randomization method, which helps stabilize the antenna matrix iteration. Finally, we employ a joint scheme of Schur decomposition, LR, and SCA, which further aids in obtaining an optimal solution for the antenna matrix. Numerical results demonstrate that the MSE performance of the proposed scheme outperforms the baseline algorithm, effectively improving the accuracy of the AirComp network while reducing the dependency of system on the RF-chain. This significantly lowers communication hardware costs, achieving the core objective of this study.

References

[1] 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, 2022.

[2] D. Wen, G. Zhu, and K. Huang, "Reduced-dimension design of MIMO over-the-air computing for data aggregation in clustered IoT networks," *IEEE Trans. Wireless Commun.*, vol. 18, no. 11, pp. 5255–5268, 2019.

[3] M. Kim, A. L. Swindlehurst, and D. Park, "Beamforming vector design and device selection in over-the-air federated learning," *IEEE Trans. Wireless Commun.*, vol. 22, no. 11, pp. 7464–7477, 2023.

[4] J. Du, B. Jiang, C. Jiang, Y. Shi, and Z. Han, "Gradient and channel aware dynamic scheduling for over-the-air computation in federated edge learning systems," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 4, pp. 1035–1050, 2023.

[5] X. Li, F. Liu, Z. Zhou, G. Zhu, S. Wang, K. Huang, and Y. Gong, "Integrated sensing and over-the-air computation: Dual-functional MIMO beamforming design," in *Proc. 6GNet'22, Paris, France*, 2022, pp. 1–8.

[6] 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, 2019.

[7] Z. Abdullah, G. Chen, S. Lambotharan, and J. A. Chambers, "Low-complexity antenna selection and discrete phase-shifts design in IRS-assisted multiuser massive MIMO networks," *IEEE Trans. Veh. Technol.*, vol. 71, no. 4, pp. 3980–3994, 2022.

[8] M. Tian, W. Sun, P. Zhang, L. Huang, and Q. Li, "Joint beamforming design and receive antenna selection for large-scale mimo wiretap channels," *IEEE Trans. Veh. Technol.*, vol. 69, no. 3, pp. 2716–2730, 2020.

[9] S. Asaad, H. Tabassum, C. Ouyang, and P. Wang, "Joint antenna selection and beamforming for massive MIMO-enabled over-the-air federated learning," *IEEE Trans. Wireless Commun.*, vol. 23, no. 8, pp. 8603–8618, 2024.

[10] S. Tang and S. Obana, "Antenna selection for multiantenna over-the-air computation," *IEEE Wireless Commun. Lett.*, pp. 1–1, 2024.

[11] Z.-q. Luo, W.-k. Ma, A. M.-c. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 20–34, 2010.

[12] J. Gallier, "The schur complement and symmetric positive semidefinite (and definite) matrices," *Penn Engineering*, pp. 1–12, 2010.

[13] H. Luo, Q. Li, Q. Zhang, and J. Qin, "Joint secure transceiver design for an untrusted MIMO relay assisted over-the-air computation networks with perfect and imperfect CSI," *IEEE Trans. Inf. Forensics Security*, vol. 18, pp. 2508–2523, 2023.

[14] L. A. Wolsey, *Integer Programming*. John Wiley & Sons, 2020.

[15] M. Razaviyayn, "Successive convex approximation: Analysis and applications," Ph.D. dissertation, University of Minnesota, 2014.