Reduced-Dimension Design of MIMO Over-the-Air Computing for Data Aggregation in Clustered IoT Networks

Dingzhu Wen, Guangxu Zhu\textsuperscript{c}, Student Member, IEEE, and Kaibin Huang\textsuperscript{c}, Senior Member, IEEE

Abstract—One basic operation of Internet-of-Things (IoT) networks is to acquire a function of distributed data collected from sensors over wireless channels, called wireless data aggregation (WDA). In the presence of dense sensors, low-latency WDA poses a design challenge for high-mobility or mission critical IoT applications. A promising solution is a latency multi-access scheme, called over-the-air computing (AirComp), that supports simultaneous transmission such that an access point (AP) can estimate and receive a summation-form function of the distributed sensing data by exploiting the waveform-superposition property of a multi-access channel. In this work, we propose a multiple-input-multiple-output (MIMO) AirComp framework for an IoT network with clustered multi-antenna sensors and an AP with large receive arrays. The framework supports low-complexity and low-latency AirComp of a vector-valued function. The contributions of this work are two-fold. Define the AirComp error as the error in the functional value received at AP due to channel noise. First, under the criterion of minimum error, the optimal receive beamformer at the AP, called decomposed aggregation beamformer (DAB), is shown to have a decomposed architecture: the inner component focuses on channel-dimension reduction and the outer component focuses on joint equalization of the resultant low-dimensional small-scale fading channels. In addition, an algorithm is designed to adjust the ranks of individual components of the DAB for a further performance improvement. Second, to provision DAB with the required channel state information (CSI), a low-latency channel feedback scheme is proposed by intelligently leveraging the AirComp principle to support simultaneous channel feedback by sensors. The proposed framework is shown by simulation to substantially reduce AirComp error compared with the existing design without considering channel structures.

Index Terms—massive MIMO, over-the-air computing, data aggregation, reduced-dimensional beamformer design, channel-rank selection, effective channel feedback.

I. INTRODUCTION

THE future Internet-of-Things (IoT) will collect distributed data from an enormous number of sensors to make decisions for the purpose of automating various operations of our society such as manufacturing, healthcare, and traffic control [1]. In a high-mobility environment (e.g., vehicle mounted readers), one challenge of designing IoT networks is fast wireless data aggregation (WDA), referring to estimating and receiving a function of distributed data acquired from sensors via wireless transmission. The low-latency requirement of fast WDA over dense sensors cannot be met by the traditional “transmit-then-compute” based on orthogonal access followed by functional computation. A more promising design approach is “transmit-and-compute” that integrates multiple access and computation. It can be realized using a scheme called over-the-air computation (AirComp), as illustrated in Fig. 1, which attracts increasing research interests recently [2]–[4]. By supporting simultaneous transmission, the principle of AirComp is to exploit the waveform-superposition property of a multi-access channel such that an access point (AP) directly receives a function of distributed sensing data. This results in low multi-access latency independent of the number of sensors, making AirComp particularly appealing in many applications such as fast sensing data aggregation using UAV mounted reader via wireless transmission [2] and distributed model averaging in federated edge learning [5]. To clarify, AirComp is not a real computing system but a low-latency multi-access scheme. The name was coined in the literature to emphasize the idea that the waveform superposition in a multi-access channel can be interpreted as the computing operation of averaging.

In this work, we design an AirComp framework for multi-antenna IoT networks to support vector-valued functional computation. The key feature is to exploit the clustered-channel structure for low-complexity and low-latency implementation and performance improvement.

A. Wireless Data Aggregation by AirComp

Consider a WDA system where an access point (AP) aims at obtaining the desired functional value. The particular class of functions exactly computable using AirComp is called nomographic functions [6], [7], as defined below.

Definition 1 (Nomographic Function [8]): The function is said to be nomographic, if there exit \( K \) preprocessing functions \( \{ f_k(\cdot) : \mathbb{R} \to \mathbb{R} \} \) along with a post-processing function \( q(\cdot) : \mathbb{R} \to \mathbb{R} \), such that it can be represented in the form:

\[
Z = q \left( \sum_k f_k(Z_k) \right),
\]

where \( Z_k \) is the input data of the \( k \)-th device.

In [9], [10], AirComp is generated for an arbitrary function via its decomposition to the sum of nomographic functions.
AirComp achieves the purpose via low-latency simultaneous transmission. The original idea of AirComp appeared in [7]. The design relies on structured codes (i.e., lattice codes) to cope with channel distortion introduced by the multi-access channel. It was subsequently discovered in [11] that simple analog transmission without coding but with channel pre-equalization can achieve the minimum distortion if the data sources are independent and identically distributed (i.i.d.) Gaussian. If this assumption does not hold, coding can be still beneficial e.g., as shown in [12] for the scenario where data sources follow the bivariate Gaussian distribution [12]. Nevertheless, the simplicity of the optimal design for the Gaussian case has inspired a series of follow-up research on making AirComp practical [13]–[15]. By measuring the AirComp distortion using mean squared error (MSE), the optimal power allocation and outage performance under a distortion constraint are studied in [13] and [14], respectively. The implementation of AirComp typically requires channel state information (CSI) at transmitters for channel pre-equalization. An attempt to relax the requirement was made in [15] where randomized transmission without CSI realizes AirComp at the cost of increased latency. Another practical issue for implementing AirComp is synchronizing the transmission of edge devices. One design addressing this issue is proposed in [16] that modulates the data into transmit power to relax the synchronization requirement. As a result, only coarse block-synchronization is required for realizing AirComp. An alternative scheme, called AirShare, is to broadcast a shared clock to all devices [17].

The prior work described above focuses on AirComp of scalar-valued functions. Most recent research in the area aims at MIMO AirComp using MIMO techniques to enable vector-valued functional computation [2]. In particular, receive beamforming targeting WDA, called aggregation beamforming, is proposed in [2] to compute vector-valued functions by spatial multiplexing and reduce AirComp distortion by spatial diversity. Along the same vein, the current work targets clustered IoT networks and addresses the issue of how to exploit channel structure for improving the performance of MIMO AirComp.

B. Reduced-Dimension Design for Massive MIMO Systems

In next-generation wireless systems, large-scale antenna arrays are expected to be deployed at APs (each with hundreds to thousands of elements) and mobile devices (each with tens of elements) [18]. In such massive MIMO systems, one research focus is to reduce complexity in transceiver designs and thereby also reduce overhead for CSI feedback. There exists a rich literature of such designs [19]. The phased-zero-forcing (PZF) precoding scheme proposed in [20] achieves complexity reduction by combining ZF precoding in the baseband domain and phase control in the radio frequency (RF) domain. On the other hand, a hierarchical architecture for implementing multiuser zero-forcing (ZF) receiver based on user clustering is shown in [21], [22] to yield complexity reduction. Another popular approach for reduced-dimension MIMO is called hybrid beamforming that decompose a MIMO transceiver into two cascaded components for analog and digital implementation [23]–[25]. For clustered MIMO channels, this implementation based architecture can dramatically reduce the number of required RF chains and the complexity of digital processing [26].

There exists one more key approach for reduced-dimension precoding design for massive MIMO downlink, which is closely related to the current work. The high spatial resolution of a large-scale arrays at an AP makes it possible to resolve the cluster structure embedded in multiuser MIMO channels. The main principle of the design approach is to decompose each MIMO channel into a slow-time-scale component, namely its (spatial) covariance matrix, and a fast-time-scale component, namely small-scale fading [27]–[31]. The covariance matrix is jointly determined by array and channel-topology parameters including the size and antenna-spacing of the transmit array, and angles of arrival (AoA) and angular spreads (AS) of user clusters. The channel decomposition leads to an efficient hierarchical beamformer structure cascading a slow-time-scale and a fast-time-scale components, which are computed based on the covariance and fading matrices, respectively [28]. The former is high-dimensional but requires infrequent or one-time computation. On the other hand, the latter is low-dimensional and hence supports efficient periodic computation and CSI feedback. The beamforming structure is proved in [28] and simultaneously in [27] to be capacity-achieving as the transmit-array size grows. The inspiring result has motivated a series of follow-up research that extends the mentioned beamforming design to millimeter-wave frequency bands [26], includes opportunistic user selection [30], and considers the minimum mean-square-error (MMSE) criterion [31].

![Fig. 1. (a) Traditional “transmit-then-compute” v.s. (b) over-the-air computation.](image-url)
The current work builds on the above prior work to design reduced-dimension aggregation beamforming for MIMO AirComp. In particular, we consider the same model of clustered massive MIMO channel and the same decomposed beamforming structure as in [27]–[29]. However, prior work targets rate-centric downlink systems and thus the objective for multiuser beamforming is sum-rate maximization. In contrast, we consider a computation-centric IoT system and the design criterion for aggregation beamforming is minimizing distortion in functional computation.

C. Contributions and Organization

In this paper, we consider WDA in a clustered massive MIMO IoT network, where an AP equipped with a large-scale array performs MIMO AirComp over distributed transmissions by sensors. The existing design of aggregation beamforming assuming structureless channels with rich scattering [2]. Its direct application in the current case would unnecessarily expose AirComp to strong noise from null space of low-dimensional cluster channels. This motivates the reduced-dimension aggregation beamforming for performance improvement by exploiting clustered-channel structure. In addition, to obtain the designed beamformers with required CSI, corresponding low-latency channel feedback schemes are proposed. The main contributions of the work are summarized below.

- **Decomposed Aggregation Beamforming (DAB) for Disjoint Clusters**: For the case of separable clusters with non-overlapping AoA ranges (see e.g., [27], [28]), we prove that the optimal aggregation beamformer has a decomposed architecture consisting inner and outer components. The inner components match the dominant eigen-subspaces of different clustered channels to receive low-dimensional signals from them. The outer components then aggregate the weighted signals to compute the desired vector function, where the weights are determined by minimum eigen-values of the said channel eigen-subspaces.

- **DAB for Overlapping Clusters**: For the more challenging case of inseparable clusters due to overlapping AoA ranges, we propose a DAB architecture consisting of a single inner and a single outer component. By solving an approximate AirComp-error minimization problem, we prove that the designs of inner and outer DABs can be separated and the separate optimization problems have identical forms. As a result, the inner DAB performs aggregation over reduced-dimensional covariance matrices of different clustered channels and the outer one over small-scale fading channels of different devices.

- **Clustered-Channel Rank Selection**: Signal-dimension reduction shortens the distances between the resultant channel precoders of different devices, thereby reducing the AirComp error. On the other hand, the operation also reduces received signal power and hence increases the error. Balancing these two effects of signal-dimension reduction gives rise to the channel-rank selection problem. Based on the DAB designs, practical algorithms are designed for choosing the ranks of reduced-dimension clustered channels (or received signals) under the criterion of minimum AirComp errors.

- **Low-Latency Channel Feedback**: To enable the preceding DAB design, schemes are presented for analog channel feedback for both the cases of disjoint and overlapping clusters. The schemes feature simultaneous reduced-dimension feedback by devices in a same cluster and sequential feedback for different clusters.

The paper is organized as follows. In Section II, the system model is introduced and the AirComp design problem is formulated. In Section III, the DAB designs are presented for both the cases of disjoint and overlapping channel clusters. The clustered-channel rank selection problem is solved in Section IV. The analog channel feedback schemes are proposed in Section V. Section VI presents the simulation results followed by concluding remarks in Section VII.

II. SYSTEM MODEL AND PROBLEM FORMULATION

A. System Model

Consider an IoT system with one AP and a large number of edge devices. Perfect local CSI is assumed to be available at all devices and channel reciprocity is considered. The system is designed to perform AirComp of distributed data transmitted by the devices. The system operations is illustrated in Fig. 2 and described as follows. The devices form \( G \) clusters each of which comprises \( K \) members. The \( k \)-th device (or channel) in the \( g \)-th cluster is identified by the index \((g, k)\). Each device, say device \((g, k)\), is provisioned with an array of \( N_t \) antennas for transmitting a \( L \)-dimensional pre-processed vector symbol by linear analog modulation, which is denoted as \( X_{g,k} \) representing \( f(Z_{g,k}) \) in Fig. 2, to the AP after precoding. The \( N_t \times L \) precoding matrix is represented by \( B_{g,k} \). For simplicity, we assume \( N_t = L \), namely exactly \( L \) antennas are used to transmit the \( L \)-dimensional vector symbol. Equipped with a large-scale array of \( N_r \) antennas \((N_r \gg N_t)\), the AP receives the simultaneous signals from all devices. The total received signal, denoted as \( Y \), is given as

\[
Y = \sum_{g=1}^{G} \sum_{k=1}^{K} Y_{g,k} = \sum_{g=1}^{G} \sum_{k=1}^{K} H_{g,k} B_{g,k} X_{g,k} + n, \tag{2}
\]

where \( Y_{g,k} \) is the received signal from device \((g, k)\), \( \{H_{g,k}\} \) represent the uplink MIMO channels and \( n \) is the channel-noise vector comprising \( CN(0, 1) \) elements. Then the received total signal is processed by aggregation beamforming, represented by the \( L \times N_r \) matrix \( A \), to yield the desired summation \( \sum_{g,k} X_{g,k} \), which gives the desired vector-valued function after post-processing (see Fig. 2). The current work focuses on designing \( A \) to minimize the distortion in functional computation. Due to the nature of analog transmission, the accuracy of AirComp is prone to the distortion by channel fading and noise. The goal of AirComp is to estimate and receive certain functions at the AP. Hence, the computation error becomes the natural performance metric. Following the existing literature (see, e.g., [13], [32]), the distortion is measured by the MSE \( E[||AX - \sum_{g,k} X_{g,k}||^2] \).
We adopt the model of clustered MIMO multi-access channels in [27], [28], characterized by clustered transmitters and rich local scattering. Consequently, for MIMO channels in the same cluster, there exists receive-antenna correlation but no transmit-antenna correlation. Specifically, the spatial correlation of the channels in the $g$-th cluster is represented by the covariance matrix $\Psi_g$ of rank denoted as $R_g$, namely $E[H_{g,k}H_{g,k}^H] = \Psi_g$ for given $g$ and any $k$. The rank $R_g$ satisfies the relation $L \leq R_g \leq N_r$. According to the one-ring model in [33], the matrix $\Psi_g$ is largely determined by angle-of-arrival (AoA) range $\Delta \theta_g = [\theta_g, \bar{\theta}_g]$ as well as array parameters (e.g., topology and antenna spacing), as illustrated in Fig. 3. Specifically, according to [33] and references therein, the channel correlation between the antennas $l$ and $p$ is given by

$$[\Psi_g]_{l,p} = \frac{1}{\bar{\theta}_g - \theta_g} \int_{\theta_g}^{\bar{\theta}_g} e^{jkT(\beta)(u_l - u_p)} d\beta,$$

(3)

where $k(\beta) = \frac{2\pi}{\lambda} [\cos(\beta), \sin(\beta)]^T$ is the wave vector for a planar wave impinging with AoA $\beta$, $\lambda$ is the carrier wavelength, and $u_l, u_p \in \mathbb{R}^2$ indicates the position of antennas $l, p$. As the AoA ranges of different clusters may be different, $R_g$ may vary in different clusters. Decompose the matrix $\Psi_g$ by singular-value decomposition as $U_g \Lambda_g U_g^H$. Then the channel matrix $H_{g,k}$ can be written as

$$H_{g,k} = U_g \Lambda_g^{\frac{1}{2}} W_{g,k},$$

(4)

where each element of $W_{g,k}$ is i.i.d. and follows $\mathcal{CN}(0,1)$. The array at the AP is assumed to be linear. Consider the case that the AoA ranges of different clusters are non-overlapping, which is referred to as the case of disjoint clusters, as illustrated in Fig. 3(a). Under this assumption, it is shown in [27], [28], as the array size $N_r$ grows, two channels belonging to different clusters approach being orthogonal as a result of $U_{m}^H U_{n} \to 0$. On the other hand, when the clusters’ AoA ranges overlap, as illustrated in Fig. 3(b), different clusters of channels cannot be orthogonalized by using a large-scale receive array, which is referred to as the case of overlapping clusters. Both cases are considered in the sequel.

### B. Problem of Decomposed Aggregation Beamforming

In this subsection, the AirComp problem for WDA is formulated as a joint DAB matrix, denoising factor (DF), and precoders design problem.
The aggregation beamforming matrix is designed under two constraints, namely the constraints of channel equalization and transmission power, described as follows. To output the desired summation \(\sum_{g,k} X_{g,k}\), the beamforming matrix need be jointly designed with the precoders \(\{B_{g,k}\}\) to overcome channel distortion. This leads to the constraints of channel equalization:

\[
\mathbf{A} \mathbf{H}_{g,k} \mathbf{B}_{g,k} = \eta \mathbf{I}, \quad \forall g,k.
\]  

(5)

where \(\eta\) is a positive scalar, called DF. Note that the solution of the precoder \(\mathbf{B}_{g,k}\) is the pseudo-inverse of \(\mathbf{A} \mathbf{H}_{g,k}\), as

\[
\mathbf{B}^*_g \mathbf{B}_g = \eta (\mathbf{A} \mathbf{H}_{g,k})^H (\mathbf{A} \mathbf{H}_{g,k} \mathbf{H}^H_{g,k} \mathbf{A}^H)^{-1}, \quad \forall g,k.
\]  

(6)

which is ZF pre-coding. It should be emphasized that the aggregation beamforming leverages “interference” in aggregation instead of suppressing it like the traditional ZF beamforming, thus requiring much fewer DoF than the latter. Consequently, the ZF constraints in (5), instead of resolving the inter-stream interference, are irrelevant for the beamforming design. Next, each device has finite transmission power, denoted as \(P_t\). The power of the pre-processed data symbol \(X_{g,k}\) is given as \(E[\|X_{g,k}\|^2]\). Without loss of generality, unit symbol power for all devices is assumed. Then the transmission-power constraints can be written as

\[
\text{tr}(\mathbf{B}_{g,k} \mathbf{B}_{g,k}^H) \leq P_t, \quad \forall g,k.
\]  

(7)

The objective of designing the beamforming matrix is to minimize the AirComp distortion. By substituting the precoders in (6) and under the constraints of (7), the problem can be formulated as the following,

\[
\begin{align*}
\text{(P1)} \quad \min_{\mathbf{A},\eta} & \quad E[\|\mathbf{A} \mathbf{Y} - \sum_{g=1}^{G} \sum_{k=1}^{K} \mathbf{X}_{g,k}\|^2], \\
\text{s.t.} & \quad \eta^2 \text{tr}((\mathbf{A} \mathbf{H}_{g,k} \mathbf{H}^H_{g,k} \mathbf{A}^H)^{-1}) \leq P_t, \quad \forall g,k.
\end{align*}
\]  

(8)

We remark that Problem (P1) is solved at the AP and it seems that global CSI is required. However, in the sequel, we will show that only a function of individual CSI is needed at AP based on our designs in Section III. And we exploit this fact to design in Section V a novel low-latency CSI-feedback scheme, the overhead of which does not scale with the number of devices.

### III. Decomposed Aggregation Beamforming

In this section, Problem (P1) is first reduced to an equivalent problem focusing on DAB design. Then, by deriving approximate solutions of the non-convex problem, the DAB matrices are designed for both the cases of disjoint and overlapping clusters.

#### A. An Equivalent DAB Design Problem

Problem (P1) is simplified to a DAB matrix design problem as follows.

First, by substituting (6), the objective function of Problem (P1) can be rewritten as

\[
\begin{align*}
E[&\|\mathbf{A} (\sum_{g=1}^{G} \sum_{k=1}^{K} \mathbf{H}_{g,k} \mathbf{B}_{g,k} \mathbf{X}_{g,k} + \mathbf{n}) - \sum_{g=1}^{G} \sum_{k=1}^{K} \mathbf{X}_{g,k}\|^2], \\
= &\frac{1}{\eta^2} E[\|\mathbf{A} \mathbf{n}\|^2], \\
= &\frac{1}{\eta^2} N_0 \text{tr}(\mathbf{A} \mathbf{A}^H),
\end{align*}
\]  

(9)

where \(N_0\) is the noise power. One can observe from (9) that the computation error due to channel noise \(n\), given by the objective, decreases as \(\eta\) grows, giving its name DF.

Next, we derive the optimal \(\eta\) in terms of \(\mathbf{A}\) based on the constraints in Problem (P1), which can be equivalently derived as

\[
\eta \leq \sqrt{\frac{P_t}{\text{tr}((\mathbf{A} \mathbf{H}_{g,k} \mathbf{H}^H_{g,k} \mathbf{A}^H)^{-1})}}, \quad \forall g,k.
\]  

(10)

Since the objective function in (9) decreases with increasing \(\eta\), the optimal DF \(\eta^*\), constrained by (10), is given as

\[
\eta^* = \max_{g,k} \eta = \min_{g,k} \sqrt{\frac{P_t}{\text{tr}((\mathbf{A} \mathbf{H}_{g,k} \mathbf{H}^H_{g,k} \mathbf{A}^H)^{-1})}}.
\]  

(11)

By substituting \(\eta^*\) in (11) into (6), we can get the optimal precoders \(\mathbf{B}^*_g\). The results are summarized as follows.

**Lemma 1 (Optimal DF and precoding):** Given an aggregation beamforming \(\mathbf{A}\), the optimal conditional DF and precoders are

- **Optimal DF**:
  \[
  \eta^* = \min_{g,k} \sqrt{\frac{P_t}{\text{tr}((\mathbf{A} \mathbf{H}_{g,k} \mathbf{H}^H_{g,k} \mathbf{A}^H)^{-1})}},
  \]  

(12)

- **Optimal precoders**:
  \[
  \mathbf{B}^*_g = \eta^* (\mathbf{A} \mathbf{H}_{g,k})^H (\mathbf{A} \mathbf{H}_{g,k} \mathbf{H}^H_{g,k} \mathbf{A}^H)^{-1}, \forall g,k.
  \]

**Remark 1 (Weakest Link Dominant Performance):** As mentioned, the AirComp error is inverse proportional to \(\eta^2\) and thus it is desirable to enhance the DF \(\eta^*\). One can observe from (12), that \(\eta^*\) is limited by the weakest link. To be specific, a weak link is characterized by small channel gains, the largest misalignment between the channel matrix and DAB \(\mathbf{A}\), or both. Note that the alignment between a channel, say \(\mathbf{H}_{g,k}\), and \(\mathbf{A}\) can be measured by a subspace distance [2]. It follows that the weakest link corresponds to \(\max_{g,k} \text{tr}((\mathbf{A} \mathbf{H}_{g,k} \mathbf{H}^H_{g,k} \mathbf{A}^H)^{-1})\) in (12).

Last, by substituting the optimal design in (12) into (9), the unconstrained DAB design problem, equivalent to (P1), is derived as

\[
\begin{align*}
\text{(P2)} \quad \min_{\mathbf{A} \in \mathbb{C}^{K \times N_r}} & \quad \max_{g,k} \text{tr}(\mathbf{A} \mathbf{A}^H) \text{tr}((\mathbf{A} \mathbf{H}_{g,k} \mathbf{H}^H_{g,k} \mathbf{A}^H)^{-1}),
\end{align*}
\]  

(13)

The problem is non-convex. The classic solution approach is semi-definite relaxation (SDR), which is, however, too complex in the current context of massive MIMO due to its iterative algorithms and the dimensionality, \(N_r \to \infty\).
A more efficient approach as we pursue is to exploit high-dimensional but low rank characteristics of clustered channels to design efficient DAB matrices in closed form. The details are presented in the following sub-sections.

B. DAB Design for Disjoint Clusters

Consider the case of disjoint clusters where the AoA ranges of any two clusters are disjoint. With large-scale receive arrays \( (N_r \to \infty) \), it is well known that the column spaces of the covariance matrices of any two different cluster channels are orthogonal: \( U_g^H U_{g'} = 0, \forall g \neq g' \). Exploiting this property, we first prove that the optimal DAB has a summation form, where each term depends on only one cluster. Furthermore, each summation term is decomposed into a product form cascoding an inner and an outer per-cluster beamforming, where the former reduces signal-space dimension and the latter performs AirComp in the reduced-dimensional signal-space.

Given the said orthogonality between cluster channels, the received signals from different clusters of devices can be decoupled without inter-cluster interference. This fact allows us to derive the structure of the optimal DAB as shown below.

**Proposition 1 (Decomposed DAB structure): In the case of disjoint clusters, the optimal DAB matrix solves Problem (P2) has the following decomposed structure,**

\[
A^* = \sum_{g=1}^{G} C_g^H U_g^H,
\]

where the size of \( C_g \) is \( R_g \times L \).

**Proof:** See Appendix A.

Several observations can be made from the optimal DAB structure in (14). Let \( \{ U_g \} \) and \( \{ C_g \} \) be referred to as the inner and the outer per-cluster DABs, respectively. Each inner term \( U_g \) is matched to one cluster and extracts the signal from the dominant \( R_g \)-dimensional eigenspace of the high-dimensional cluster channels \( \{ H_{g,k}, k \in [1, K] \} \). This yields a reduced-dimensional signal-space, where performing AirComp using \( \{ C_g \} \) has two advantages. The signal-to-noise ratio (SNR) therein are high and the subspace distances between the effective channels \( \{ U_g H_{g,k}, k \in [1, K] \} \) are small, leading to AirComp-error reduction. Furthermore, AirComp in a reduced-dimensional space results in dramatic complexity reduction.

Next, building on the optimal DAB structure on (14), we focus on designing the outer per-cluster DABs \( \{ C_g \} \). By substituting (14), Problem (P2) can be derived as

\[
\min_{\{ C_g \}} \max_{g,k} \sum_{m=1}^{G} \text{tr} \left( C_m^H C_m \right) \frac{1}{\lambda_{1}} \left( C_g^H F_{g,k} F_{g,k}^H C_g \right),
\]

where \( F_{g,k} = \Lambda_{1}^{\frac{1}{2}} W_{g,k} \) is the effective channel after dimension reduction, the function \( \lambda_{1} \) acquires the \( i \)-th eigenvalue of a matrix, and the eigenvalues are arranged in a decreasing order, i.e., \( \lambda_{1} \geq \lambda_{2} \geq \ldots \geq \lambda_{L} \).

The problem in (15) reduces high-dimensional design in Problem (P2) to the design of reduced-dimensional DAB matrices \( \{ C_g \} \). (15) is non-convex. As its solution is intractable, we derive an approximate solution in closed form to obtain an efficient design of \( \{ C_g \} \). The approximation consists of two steps. The first is to replace the objective function in problem (15) by an upper bound based on the following inequalities, \( \lambda_{1} \left( C_g^H F_{g,k} F_{g,k}^H C_g \right) \geq \lambda_{\min} \left( C_g^H F_{g,k} F_{g,k}^H C_g \right) \). The bounds are tight when the eigenvalues of the matrix \( C_g^H F_{g,k} F_{g,k}^H C_g \) are similar. It follows that the problem in (15) can be approximated as the following outer-DAB-design problem,

\[
(P3) \min_{\{ C_g \}} \max_{g,k} \sum_{m=1}^{G} \text{tr} \left( C_m^H C_m \right) \lambda_{\min}^{-1} \left( C_g^H F_{g,k} F_{g,k}^H C_g \right) L.
\]

However, Problem (P3) is still non-convex. To overcome the difficulty, the second approximation step adopts a general approach in beamforming literatures (see e.g., [34]–[36]), that constrains the beamforming matrices \( \{ C_g \} \) to be unitary. This is reasonable as it is the subspace spanned by \( C_g \) that has a dominant effect on the AirComp performance. With the constraint \( (C_g^H C_g = I) \), Problem (P3) can be further approximated as

\[
\min_{\{ C_g \}} \max_{g,k} \lambda_{\min}^{-1} \left( C_g^H F_{g,k} F_{g,k}^H C_g \right) L, \quad \text{s.t.} \quad C_g^H C_g = I, \quad \forall g.
\]

In (17), it can be shown that the design of \( C_g \) depends on solely the \( g \)-th cluster channel matrices \( \{ F_{g,k}, 1 \leq k \leq K \} \) and is independent of other clusters \( (g' \neq g) \). Therefore, the inner beamforming design can be decoupled, as shown in the following lemma.

**Lemma 2 (Outer per-cluster DAB): The joint design of \( \{ C_g \} \) in (17) can be decoupled to solve \( C_g \) in the following problem for all \( g \),**

\[
\min_{C_g} \max_{g,k} \lambda_{\min}^{-1} \left( C_g^H F_{g,k} F_{g,k}^H C_g \right), \quad \text{s.t.} \quad C_g^H C_g = I.
\]

The problem in (18) has the same form as the Problem (P5) in [2]. Following the approach in [2], \( C_g' \) that solves the problem in (18), can be obtained as the weighted subspace centroid of the column spaces of \( \{ F_{g,k} \} \), which is the \( L \)-dimensional principal eigenspace of the following matrix,

\[
S_{g}^{(a)} = \sum_{k=1}^{K} \lambda_{\min} \left( F_{g,k}^H F_{g,k} \right) U_{F_{g,k}} U_{F_{g,k}}^H,
\]

where \( U_{F_{g,k}} \) is the column space of \( F_{g,k} \). In other words, the solution is \( C_g' = \left[ U_{S_{g}^{(a)}} \right]_{1:L} \), where \( \left[ U_{S_{g}^{(a)}} \right]_{1:L} \) denotes the \( L \)-dimensional principal eigenspace of \( S_{g}^{(a)} \).

By combining the results in Lemma 1 and 2, the DAB matrix design for the AirComp in disjoint-cluster case is given as

\[
\text{(Optimal DAB)} \quad A^* = \sum_{g=1}^{G} C_g^H U_g^H,
\]

where \( C_g = \left[ U_{S_{g}^{(a)}} \right]_{1:L} \), and \( S_{g}^{(a)} \) is defined in (19), respectively.
C. DAB Design for Overlapping Clusters

In this subsection, DAB is designed for the case of overlapping clusters. Unlike the preceding case of disjoint clusters, it is impossible to decouple the received signals from different clusters due to their overlapping [28]. Then, the optimal DAB form in (20) derived for the former no longer for the current case. Nevertheless, inspired by the result, we propose that the DAB design should have the decomposed form: \( \mathbf{A} = \mathbf{A}(o)\mathbf{A}(i) \), where the inner DAB \( \mathbf{A}(i) \) is a \( R_s \times N_r \) matrix with

\[
R_s = \min \left( R_1, \ldots, R_g, \ldots, R_Q \right),
\]

and the outer DAB \( \mathbf{A}(o) \) is a \( L \times R_s \) matrix. In other words, \( \mathbf{A}(i) \) is responsible for the dimension reduction of the signal space. Nevertheless, the operation of AirComp is distributed over outer and inner beamformers instead of relying only on the former as in the preceding case. For tractability, following the same reason as for designing \( \mathbf{C}_g \) in Problem (P3), we constrain both \( \mathbf{A}(o) \) and \( \mathbf{A}(i) \) to be unitary. We show in the sequel that the designs of outer and inner DAB can be reduced to optimization problems having the identical form.

1) Inner Beamforming Design: Under the criterion of minimizing AirComp error, the inner DAB should be matched to the \( R_s \)-dimensional dominant eigenspace of each cluster channel, which is obtained as follows.

Denote the \( R_s \)-dimensional dominant eigenspace, the \( R_s \)-dimensional dominant eigenvalue matrix, and the corresponding small-scale fading matrix of device \((g,k)\) as \( \hat{\mathbf{U}}_g = [\hat{\mathbf{U}}_g]_{1:R_s}, \hat{\mathbf{A}}_g = [\hat{\mathbf{A}}_g]_{1:R_s,1:R_s}, \) and \( \hat{\mathbf{W}}_{g,k} = [\hat{\mathbf{W}}_{g,k}]_{1:R_s,1:R_s} \), respectively. Then, the dominant \( R_s \)-dimensional subspace of the channel model in (4) is

\[
\hat{\mathbf{H}}_{g,k} = \hat{\mathbf{U}}_g \hat{\mathbf{A}}_g \hat{\mathbf{W}}_{g,k}.
\]

To solve Problem (P2) in this case, we first derive a useful inequality as follows.

**Lemma 3:** With \( \mathbf{A} = \mathbf{A}(o)\mathbf{A}(i) \), the following inequality holds.

\[
\text{tr} \left( \left( \mathbf{A} \hat{\mathbf{H}}_{g,k} \hat{\mathbf{H}}_{g,k}^H \mathbf{A}^H \right)^{-1} \right) \leq \lambda_{\min}^{-1} \left( \mathbf{A}(i) \hat{\mathbf{U}}_g \hat{\mathbf{U}}_g^H \mathbf{A}(i)^H \right) \quad \times \sum_{i=1}^{L} \lambda_i^{-1} \left( \hat{\mathbf{A}}_g^H \hat{\mathbf{W}}_{g,k} \hat{\mathbf{W}}_{g,k}^H \hat{\mathbf{A}}_g \right),
\]

(23)

where the bound is tight when the eigenvalues of the matrix \( \left( \mathbf{A}(i) \hat{\mathbf{U}}_g \hat{\mathbf{U}}_g^H \mathbf{A}(i)^H \right) \) are similar.

**Proof:** See Appendix B.

By substituting (23) into Problem (P2), the objective function is replaced by its upper bound, and Problem (P2) can be approximated as the following inner-DAB-design problem,

\[
\begin{aligned}
\text{(P4)} \quad \min_{\mathbf{A}(i)} & \quad \max_g \alpha_g' \lambda_{\min}^{-1} \left( \mathbf{A}(i) \hat{\mathbf{U}}_g \hat{\mathbf{U}}_g^H \mathbf{A}(i)^H \right), \\
\text{s.t.} & \quad \mathbf{A}(i) \mathbf{A}(i)^H = \mathbf{I},
\end{aligned}
\]

(24)

where \( \alpha_g' = \max_k \sum_{i=1}^{L} \lambda_i^{-1} \left( \hat{\mathbf{A}}_g^H \hat{\mathbf{W}}_{g,k} \hat{\mathbf{W}}_{g,k}^H \hat{\mathbf{A}}_g \right) \). The problem has the same form as Problem (P5) in [2]. Following the approach in [2], \( \mathbf{A}(i)^H \) is solved as the \( R_s \)-dimensional principal eigenspace of the following matrix,

\[
\mathbf{S}^{(b)} = \sum_{g=1}^{G} \alpha_g' \hat{\mathbf{U}}_g \hat{\mathbf{U}}_g^H.
\]

That’s to say, the solution is \( \mathbf{A}^*_i = [\mathbf{U}^{(b)}_{S^{(b)}}]_{1:R_s} \), where \([\mathbf{U}^{(b)}_{S^{(b)}}]_{1:R_s} \) is the \( R_s \)-dimensional principal eigenspace of \( \mathbf{S}^{(b)} \).

2) Outer Beamforming Design: By substituting \( \mathbf{A}^*_i = [\mathbf{U}^{(b)}_{S^{(b)}}]_{1:R_s} \) into Problem (P2), it can be derived as the following outer-DAB-design problem,

\[
\begin{aligned}
\text{(P5)} \quad \min_{\mathbf{A}(o)} & \quad \max_{g,k} \text{tr} \left( \left( \mathbf{A}(o) \mathbf{F}_{g,k} \mathbf{F}_{g,k}^H \mathbf{A}(o)^H \right)^{-1} \right), \\
\text{s.t.} & \quad \mathbf{A}(o) \mathbf{A}(o)^H = \mathbf{I},
\end{aligned}
\]

(26)

where \( \mathbf{F}_{g,k} = \mathbf{A}^*_i \hat{\mathbf{U}}_g \hat{\mathbf{A}}_g^* \hat{\mathbf{W}}_{g,k} \). Using the following inequality to replace the objective function of (P5) with its upper bound,

\[
\text{tr} \left( \left( \mathbf{A}(o) \mathbf{F}_{g,k} \mathbf{F}_{g,k}^H \mathbf{A}(o)^H \right)^{-1} \right) \leq L \lambda_{\min}^{-1} \left( \mathbf{A}(o) \mathbf{F}_{g,k} \mathbf{F}_{g,k}^H \mathbf{A}(o)^H \right),
\]

(27)

whose bound is tight when the eigenvalues of the matrix \( \mathbf{A}(o) \mathbf{F}_{g,k} \mathbf{F}_{g,k}^H \mathbf{A}(o)^H \) are similar, Problem (P5) can be further approximated to

\[
\begin{aligned}
\min_{\mathbf{A}(o)} & \quad \max_{g,k} L \lambda_{\min}^{-1} \left( \mathbf{A}(o) \mathbf{F}_{g,k} \mathbf{F}_{g,k}^H \mathbf{A}(o)^H \right), \\
\text{s.t.} & \quad \mathbf{A}(o) \mathbf{A}(o)^H = \mathbf{I}.
\end{aligned}
\]

(28)

Then \( \mathbf{A}^*_o \) can be solved by the same approach with the problem in (18) as the \( L \)-dimensional principal eigenspace of the following matrix,

\[
\mathbf{S}^{(c)} = \sum_{g=1}^{G} \sum_{k=1}^{K} \lambda_{\min} \left( \mathbf{F}_{g,k} \mathbf{F}_{g,k}^H \right) \mathbf{U}_{F_{g,k}} \mathbf{U}_{F_{g,k}}^H.
\]

(29)

That’s to say, the solution is \( \mathbf{A}^*_o = [\mathbf{U}^{(c)}_{S^{(c)}}]_{1:R_s} \), where \([\mathbf{U}^{(c)}_{S^{(c)}}]_{1:R_s} \) is the \( L \)-dimensional principal eigenspace of \( \mathbf{S}^{(c)} \).

3) Overall DAB Design: In summary, the overall DAB design in overlapping-cluster case is comprised of inner beamformer \( \mathbf{A}^*_i \) and outer beamformer \( \mathbf{A}^*_o \), which are given as

\[
\begin{aligned}
\mathbf{A}^*_i &= [\mathbf{U}^{(b)}_{S^{(b)}}]_{1:R_s}, \\
\mathbf{A}^*_o &= [\mathbf{U}^{(c)}_{S^{(c)}}]_{1:R_s}.
\end{aligned}
\]

(30)

where \([\mathbf{U}^{(b)}_{S^{(b)}}]_{1:R_s} \) and \([\mathbf{U}^{(c)}_{S^{(c)}}]_{1:R_s} \) are the \( R_s \)-dimensional and \( L \)-dimensional principal eigenspaces of \( \mathbf{S}^{(b)} \) and \( \mathbf{S}^{(c)} \), and \( \mathbf{S}^{(b)} \) and \( \mathbf{S}^{(c)} \) are defined in (25) and (29), respectively.

**Remark 2 (DAB Design for Overlapping Clusters):** One can observe from (30) that the DAB design performs two-tier AirComp. To be specific, the inner DAB \( \mathbf{A}^*_i \) performs AirComp over channel covariance matrices. Subsequently, in the reduced-dimension signal space created by the inner DAB, the outer DAB \( \mathbf{A}^*_o \) performs AirComp over small-scale fading channels.
IV. CLUSTERED-CHANNEL RANK SELECTION

In the preceding section, DAB is designed with fixed ranks for its components. Adjusting the ranks according to clustered channel covariance provides another dimension for reducing AirComp error. Relevant algorithms are presented in this section.

A. Channel-Rank Selection for Disjoint Clusters

Modifying the DAB design in (20) to allow variable ranks for inner components:

\[ \mathbf{A}^* = \sum_{g=1}^{G} \mathbf{C}_g^* \mathbf{U}_g^H, \]  

(31)

where \( \mathbf{U}_g^H \) selects the \( r_g \)-dimensional dominant eigenspace of the channel covariance matrix \( \mathbf{P}_g \), and \( \mathbf{C}_g^* \) is computed in the same way as \( \mathbf{C}_g \) with \( \mathbf{U}_g \) replaced by \( \mathbf{U}_g^* \). There exists a tradeoff in setting the ranks \( \{ r_g \} \) of \( \{ \mathbf{U}_g \} \). On one hand, as can be proved, increasing the ranks \( \{ r_g \} \) receives more signal energy from the channels and helps reduce AirComp error. On the other hand, increasing an inner-DAB rank, says \( r_g \), increases the dimensionality of the reduced-dimension subspace, where small-scale-fading channels of cluster \( g \) are jointly equalized for the purpose of AirComp, thereby increasing its error. The above tradeoff is leveraged in the sequel to formulate an optimization problem for channel-rank selection and to derive an algorithmic solution.

The problem of channel-rank selection, namely optimizing the ranks \( \{ r_g \} \) of inner DAB can be formulated by substituting the optimal design in (20) into (18):

\[
\begin{align*}
(\text{P6}) \quad \min_{\{ r_g \}} & \max_{g,k} \lambda_{\min}^{-1} \left( \mathbf{C}_g^* \mathbf{H}_g \mathbf{F}_{g,k} \mathbf{F}_{g,k}^H \mathbf{C}_g^* \right), \\
\text{s.t.} & \quad L \leq r_g \leq R_g, \quad \forall g.
\end{align*}
\]

(32)

where \( \mathbf{F}_{g,k} = \frac{1}{\sqrt{\mathbf{\lambda}_g}} \mathbf{W}_{g,k} \) is the \( r_g \times L \) effective channel after dimension reduction using inner DAB. Decompose \( \mathbf{F}_{g,k} \) using SVD as \( \mathbf{F}_{g,k} = \mathbf{U}_{F_{g,k}} \mathbf{C}_{F_{g,k}} \). Then, the objective function of Problem (P6) can be bounded as

\[
\begin{align*}
& \lambda_{\min}^{-1} \left( \mathbf{C}_g^* \mathbf{H}_g \mathbf{F}_{g,k} \mathbf{F}_{g,k}^H \mathbf{C}_g^* \right) \\
& = \lambda_{\min}^{-1} \left( \mathbf{C}_g^* \mathbf{U}_{F_{g,k}} \mathbf{\Sigma}_{F_{g,k}}^2 \mathbf{U}_{F_{g,k}}^H \mathbf{C}_g^* \right), \\
& \leq \lambda_{\min}^{-1} \left( \mathbf{\Sigma}_{F_{g,k}}^2 \right) \lambda_{\min}^{-1} \left( \mathbf{C}_g^* \mathbf{U}_{F_{g,k}} \mathbf{U}_{F_{g,k}}^H \mathbf{C}_g^* \right), \\
& = \lambda_{\min}^{-1} \left( \mathbf{H}_{g,k}^H \mathbf{F}_{g,k} \right) \left( 1 - d_{F_{g,k}}^2 \right)^{-1},
\end{align*}
\]

(33)

where \( d_{F_{g,k}}^2 = \mathbf{C}_g^* \mathbf{U}_{F_{g,k}} \) is the projection 2-norm subspace distance between the subspaces spanned by \( \mathbf{C}_g^* \) and \( \mathbf{U}_{F_{g,k}} \) [37]. Using the inequality in (33), whose bound is tight when the eigenvalues of the matrix \( \mathbf{C}_g^* \mathbf{H}_g \mathbf{F}_{g,k} \mathbf{F}_{g,k}^H \mathbf{C}_g^* \) are similar, to replace the objective function with the upper bound, Problem (P6) can be further approximated for tractability as

\[
\begin{align*}
& \min_{\{ r_g \}} \max_{g,k} \lambda_{\min}^{-1} \left( \mathbf{F}_{g,k}^H \mathbf{F}_{g,k} \right) \left( 1 - d_{F_{g,k}}^2 \right)^{-1}, \\
& \text{s.t.} \quad L \leq r_g \leq R_g, \quad \forall g.
\end{align*}
\]

(34)

The objective function in (34) represents a component of AirComp error measured using MSE. A useful result is obtained as follows.

**Lemma 4:** Consider the \( r_g \times N_t \) effective channel \( \mathbf{F}_{g,k} \) of device \( k \) in cluster \( g \) after dimension reduction. The eigenvalue \( \lambda_{\min} \left( \mathbf{F}_{g,k}^H \mathbf{F}_{g,k} \right) \) is a monotone increasing function of \( r_g \).

**Proof:** See Appendix C.

**Remark 3** (Tradeoff in Channel-Rank Selection): The said tradeoff is reflected in the objective function in (34). To be specific, as \( r_g \) grows, \( \lambda_{\min} \left( \mathbf{F}_{g,k}^H \mathbf{F}_{g,k} \right) \) increases according to Lemma 4, reducing AirComp error. On the other hand, the dimensionality \( (r_g) \) of the subspace of \( \mathbf{H}_{g,k} \) after dimension reduction grows. Note that in this subspace, the outer DAB \( \hat{C}_g \) equals the cluster of channels \( \{ \mathbf{F}_{g,k} \} \) for the purpose of AirComp. As the dimensionality grows, the subspaces distance, \( d_{F_{g,k}}^2 \) \( (\mathbf{C}_g^*, \mathbf{U}_{F_{g,k}}) \), increases [37], thereby elevating the AirComp error.

For simplicity, denote \( \lambda_{\min}^{-1} \left( \mathbf{F}_{g,k}^H \mathbf{F}_{g,k} \right) \left( 1 - d_{F_{g,k}}^2 \right)^{-1} \) as \( \text{MSE}_{g,k} \) and define \( \text{MSE} = \max_{g,k} \{ \text{MSE}_{g,k} \} \). Hence, the problem in (34) can be simplified as

\[
\begin{align*}
& \min_{r_g} \text{MSE}, \\
& \text{s.t.} \quad L \leq r_g \leq R_g, \quad \forall g.
\end{align*}
\]

(35)

In the sequel, the problem in (35) is solved to yield two schemes: homogeneous and heterogeneous channel-rank selection.

1) Homogeneous Rank-Selection Scheme: To simplify design, apply the constraints of homogeneous rank selection: \( \{ r_g = r, \quad \forall g \} \). Then, the problem in (35) can be re-written as

\[
\begin{align*}
(\text{P7}) \quad \min_{r_g} & \text{MSE}, \\
\text{s.t.} & \quad r_g = r, \quad \forall g, \\
& \quad L \leq r \leq \min_{g} \{ R_g \}.
\end{align*}
\]

(36)

Since the rank of the covariance matrix \( r \) is an integer variable and its range, \( N_t \leq r \leq \min_{g} \{ R_g \} \), is usually small, the optimal value of \( r \) can be found by one-dimensional search with the complexity of \( O(R_g) \).

2) Heterogeneous Rank-Selection Scheme: In this case, inner DAB components \( \{ \mathbf{U}_g \} \) are allowed to have different ranks. The corresponding problem in (35) is an integer problem, whose solution is NP-hard. To address this issue, we propose a sub-optimal design based on the following procedure.

First, the channel cluster that is the bottleneck of AirComp is identified and the rank of the corresponding outer DAB component is optimized. Next, the preceding step is repeated till the algorithm converges. The details of the algorithm are in Algorithm 1.

B. Channel-Rank Selection for Overlapping Clusters

Due to overlapping clusters, it is no longer feasible to match the ranks of individual DAB components according to those of individual clusters. However, it is possible to optimize the rank of inner DAB \( r \) in the design in (30) over the range \( L \leq r \leq \min_{g} R_g \), namely performing homogeneous rank selection similarly as in Problem (P7). The resultant problem of channel-rank selection for outer DAB can be formulated by...
Algorithm 1 Heterogeneous Rank Selection Algorithm

1: Initialize \( r_g = N, \forall g \).
2: \textbf{Loop}
3: \quad Solve \( A^* \) with (20) and calculate \( \text{MSE}_{g,k}, \forall g,k \).
4: \quad Find \((G_0, K_0) = \arg \max_{g,k} \text{MSE}_{g,k} \) by search,
5: \quad Initialize \( \text{MSE}_c = \infty \).
6: \quad For \( r_{G_0} = N \), \( \forall g \):
7: \quad \quad Solve \( A^* \) with (20) and calculate \( \text{MSE}_{G_0,k}, \forall k \),
8: \quad \quad \quad MSE_{e} = \text{MSE}_{c},
9: \quad \quad MSE_{e} = \max_k \text{MSE}_{G_0,k}, \forall k,
10: \quad \quad If \( MSE_{e} < \text{MSE}_{e} \),
11: \quad \quad \quad \hat{R} = r_{G_0},
12: \quad \quad EndIf
13: \quad EndFor
14: \quad Update \( r_{G_0} = \hat{R} \).
15: \quad Until convergence.

substituting the design in (30) into (28):

\[
(P8) \quad \min_r \quad \lambda_{min}^{-1} \left( \begin{bmatrix} A_{(o)}^* \end{bmatrix} F_{g,k} F_{g,k}^H A_{(o)}^* \right),
\]

s.t. \( r_g = r, \forall g \), \( L \leq r \leq \min_g \{R_g\} \). \hfill (37)

which can also be solved by one-dimensional search, since the ranks’ range, \( L \leq r \leq \min_g \{R_g\} \), is usually small.

V. ANALOG CHANNEL FEEDBACK

In this section, the principle of AirComp is applied to design efficient schemes for CSI feedback to enable DAB designed in the preceding sections. Specifically, given channel reciprocity and reliable feedback channel, the schemes feature low-latency simultaneous analog feedback such that the desired DAB \( A^* \) can be computed as

\[
A^* = q \left( \sum_{g=1}^{G} q_g \left( Y_g \right) \right) = q \left( \sum_{g=1}^{G} q_g \left( \sum_{k=1}^{K} H_{g,k} Z_{g,k} \right) \right), \hfill (38)
\]

where \( Y_g \) is the aggregated feedback signals from all devices in cluster \( g \), \( q_g(\cdot) \) is the cluster-based post-processing function, and \( q(\cdot) \) is the overall post-processing function. The principle was first applied in [2] to design feedback for AirComp targeting rich-scattering channels. Based on the same principle, we design feedback schemes for reduced-dimensional AirComp for clustered MIMO channels.

In practical systems such as 3GPP LTE, CSI feedback is part of control signaling and protected against channel fading and noise by high transmission power and coding, creating reliable feedback channels. Such channels are also assumed in this work, where reliable analog feedback is ensured by high power and linear analog coding. As a result, noise as well as analog feedback detection [38] are omitted in the exposition for brevity. In the sequel, we focus on the design of feedback signals, pre-processing, and post-processing.

A. Analog Feedback for Disjoint Clusters

Based on the design in (20), the objective for feedback is to obtain at the AP the desired DAB \( A^* = \sum_{g=1}^{G} C_g^H U_g^H \)

The matrix \( U_g \) is the eigenspace of the channel covariance matrix \( \Psi_g \), which can be estimated reliably at both the AP and devices from past transmission [27], [28]. It follows that the feedback purpose is for the AP to acquire \( \{C_g\} \), which depend on small-scale fading.

Based on the principle in (38), we propose the following “one-shot” analog feedback scheme as Algorithm 2, where the notation follows that in Section III-B. It is straightforward to verify that the DAB obtained using the above feedback scheme is the desired one.

Remark 4 (One-Shot Feedback): The key feature of the above analog feedback scheme is simultaneous analog transmission (or “one-shot” feedback). The minimum feedback duration is a single symbol duration. Therefore, the feedback overhead is low and its latency is independent of the number of devices.

Algorithm 2 Analog Feedback for Disjoint Clusters

1: Design feedback signal for device \((g,k)\) as
\[
Y_{g,k} = \lambda_{min} \left( F_{g,k}^H F_{g,k} \right) \Psi_{F_{g,k}} \sum_{m=1}^{M} \Psi_{F_{g,k}}^{-1} U_{g,k}^H, \forall g,k.
\]
2: The received aggregated feedback signal at AP is given by
\[
Y_g = \sum_{g=1}^{G} U_g \sum_{k=1}^{K} F_{g,k} Z_{g,k}.
\]
3: Extract the signal from cluster \( g \) and perform cluster-based post-processing \( q_g(\cdot) \) as
\[
Y_g = U_g^H Y_g = \sum_{k=1}^{K} F_{g,k} Z_{g,k}.
\]
4: Perform overall post-processing \( q(\cdot) \) to get \( A^* \), as
\[
A^* = \sum_{g=1}^{G} \left[ U Y_g \right]_{1:L} U_g^H,
\]

where \( [U Y_g]_{1:L} \) is the \( L \)-dimensional principal eigenspace of \( Y_g \).
5: Broadcast \( A^* \) to each user and use the maximum-AirComp algorithm in [7] to get \( \eta^* \) at AP.
6: Broadcast \( A^* \) and \( \eta^* \) and design the precoder \( \{B_{g,k}\} \) for each user according to (12).

B. Analog Feedback for Overlapping Clusters

Following the design in Section III-C, the desired DAB in the current case is \( A^* = A_{(o)}^* A_{(i)}^* \). In the preceding case of disjoint clusters, one-shot feedback is feasible due to the fact that signals from different clusters are separable at the AP. This does not hold in the current case while feedback signals from different clusters still need be separated. Consequently, the feedback scheme requires \( G \) slots where feedback in each slot targets one specific cluster of channels.

Consider feedback of the inner DAB \( A_{(i)}^* \). One can observe from (25) that the inner DAB \( A_{(i)}^* \) depends on 1) the channel covariance matrices, which are known to both the AP and devices, and 2) aggregation weights (scalars) \( \{\alpha_g^i\} \) with one
for each cluster. A particular weight, say $\alpha_g'$, requires computation of the maximum over $K$ scalars transmitted by $K$ devices in cluster $g$. This can be realized using the existing AirComp algorithm in [7], referred to as maximum-AirComp algorithm. As the scalars depend on small-scale fading, their feedback need be periodic and repeated for every channel coherence time. Next, consider feedback of the outer DAB $A^*_g$. The design depends on small-scale fading according to (29). For the reason mentioned earlier, the outer-DAB feedback requires $G$ slots.

The proposed scheme combining feedback of outer and inner DAB are shown in Algorithm 3, where the notation follows that in Section III-C.

Algorithm 3 Analog Feedback for Overlapping Clusters

1: For cluster $g = 1, 2, \cdots, G$,
2: Get $\alpha_g'$ with maximum-AirComp algorithm in [7] at AP,
3: Design the feedback signal for user $(g, k)$ as
4: \[ Z^*_{g,k} = \lambda_{\min} \left( F^H_{g,k} F_{g,k} \right) V_{F_{g,k}} \Sigma_{F_{g,k}}^{-1} U^H_{F_{g,k}}, \]
5: Receive the aggregative signal from cluster $g$,
6: Compute $S^{(b)} = \sum_{g=1}^{G} \Omega_g^t \hat{U}_g \hat{U}_g^H$, at the AP,
7: Compute the inner DAB as the $R_g$-dimensional and principal eigenspace of $S^{(b)}$, as
8: $A^*_g (i) = [U_{S^{(b)}}]^H R_g$
9: Perform cluster-based post-processing $q_g$ as $\Omega_g = A^*_g (i) \hat{U}_g \hat{U}_g^H Y_g$
10: Compute $S^{(c)} = \sum_{g=1}^{G} \Omega_g$
11: Broadcast $A^*$ to each user and use the maximum-AirComp algorithm in [7] to get $\eta^*$ at AP
12: Broadcast $A^*$ and $\eta^*$ and design the precoder $\{B_{g,k}\}$ for each user according to (12).

Remark 5 (Cluster-Based Feedback): In the case of overlapping clusters, even though the feedback is not one-shot and divided into separate feedback for different clusters, the feedback for devices within the same cluster are simultaneous. Consequently, the total feedback overhead depends on the number of clusters and does not scale with the number of devices.

Remark 6 (Implementation Complexity of DAB Designs): The DAB designs in Section III are implemented by the effective channel feedback schemes in Algorithms 2 and 3, respectively, which require different operations at devices and at AP, as described below.

- (Device Complexity) At each device $(g, k)$, operations such as SVD of the reduced-dimension channel matrix $F_{g,k}$, matrix multiplication, and inverse of diagonal matrix, are involved. The computational complexity is dominant by the SVD of $F_{g,k}$, which is $O(R_g^2 N_t)$, where $R_g$ is the rank of cluster $g$’s covariance matrix and a small number in general.
- (AP Complexity) At the AP side, operations such as eigenvalue decomposition of the high-dimensional channel covariance matrix, weighted subspace centroid calculation, and the matrix multiplications, are involved. The computational complexity depends on the eigenvalue decomposition and the multiplications, which is $O(N_r^3 + N_r R_g N_r)$.

VI. SIMULATION RESULTS

Consider an IoT network with an AP and $G$ clusters of devices. There are $K$ devices in each cluster. The AP performs AirComp over the data transmitted by the devices. The performance metric MSE is normalized by the unit symbol power. The simulation parameters are summarized in Table I, where the receive antenna spacing $D$ is normalized by the wavelength.

A. Channel-Rank-Selection Gain

1) Channel-Rank Selection for Disjoint Clusters: The homogeneous channel-rank selection of two disjoint clusters is presented in Fig. 4(a). In this case, the number of receive arrays is $N_r = 48$. The number of transmit antennas is $N_t = 1$. The AOA ranges are $\Delta \theta_1 = [-49^\circ, -1^\circ]$ and $\Delta \theta_2 = [1^\circ, 49^\circ]$, respectively. According to [28], the number of ranks of both clusters can be calculated as $R_1 = R_2 = 18$. In Fig. 4(a), the optimal homogeneous rank selection scheme can significantly improve the performance compared with no rank selection. Besides, the MSE increases with the number of devices in each cluster. The reason is that the subspace distances between the effective channels after dimension reduction, on which the AirComp performance depends, increases with the number of devices.

The performance of the heterogeneous channel-rank selection of three disjoint clusters is shown in Fig. 4(b). In this case, the number of receive arrays is $N_r = 48$. The number of transmit antennas is $N_t = 1$. The AOA ranges are $\Delta \theta_1 = [-51^\circ, -15^\circ]$, $\Delta \theta_2 = [-14^\circ, 14^\circ]$, and $\Delta \theta_3 = [15^\circ, 41^\circ]$, respectively. The corresponding ranks are $R_1 = 12$, $R_2 = 12$, and $R_3 = 10$, respectively. In Fig. 4(b), the sub-optimal channel-rank-selection scheme proposed in Algorithm 1 can improve the performance. Again, the MSE increases with the number of devices in each cluster.

2) Rank Selection for Overlapping Clusters: The performance of the channel-rank selection for overlapping clusters of two overlapping clusters is presented in Fig. 5. In this case, the number of receive arrays is $N_r = 48$. The number of transmit antennas is $N_t = 4$. The AOA ranges are $\Delta \theta_1 = [-45^\circ, 15^\circ]$ and $\Delta \theta_2 = [-15^\circ, 45^\circ]$, respectively. The ranks of both
clusters are $R_1 = R_2 = 15$. Fig. 5 shows that the channel-rank selection in this case can improve the performance.

The simulation results above verify that channel-rank selection is needed to achieve the best performance of AirComp in massive MIMO systems. Compared with non-channel-rank selection, channel-rank selection can provide a better reduced dimension of the channel matrices and hence decreases the AirComp error. Besides, increasing the transmit power and selecting the devices with most correlated small-scale fading can also improve the performance.

B. Reduced-Dimension Gain

To show the reduced-dimension gain of the proposed designs, the one-tier aggregation beamforming (OTAB) AirComp scheme in [2] is used for comparison. In [2], the beamforming matrix is designed as the weighted subspace centroid of all devices’ channel matrices, as $\mathbf{A} = \sum_g \sum_k \lambda_{\min}(\mathbf{H}_{g,k}^H \mathbf{H}_{g,k}) \mathbf{U}_{H_{g,k}} \mathbf{U}_{H_{g,k}}^H$, where $\mathbf{H}_{g,k}$ is the channel matrix of device $(g, k)$, $\lambda_{\min}(\cdot)$ represents the minimum eigenvalue, and $\mathbf{U}_{H_{g,k}}$ is the column space of $\mathbf{H}_{g,k}$, respectively.

In Fig. 6, we show the reduced-dimension gains of the proposed designs. Two clusters of devices are considered. To investigate the impact of channels’ correlation, the number of receive arrays is set to $N_r = 30$. The number of transmit antennas is set to $N_t = 5$. In the figure, “DDAB” and “ODAB” represent the beamforming design for disjoint clusters and overlapping clusters, respectively.

In Fig. 6 (a), the transmit power is $P_t = 24$ dBm. The AoA ranges of the two clusters are $\Delta \theta_1 = -\delta + [-35^\circ, 25^\circ]$, $\Delta \theta_2 = \delta + [-30^\circ, 30^\circ]$, respectively, where $\delta$ is the AoA ranges change. As $\delta$ increases, the AoA ranges of the two clusters change from highly overlapping to nearly disjoint. In the figure, when the AoA ranges are highly overlapping, the DAB design for disjoint clusters has best performance. Otherwise, the performance of the DAB design for disjoint clusters is the best. Besides, in highly overlapping case, the disjoint DAB design can still has good performance since the two clusters can nearly be regarded as one cluster.

In Fig. 6 (b), The AoA ranges are $\Delta \theta_1 = [-50^\circ, 10^\circ]$, $\Delta \theta_2 = [-15^\circ, 45^\circ]$, respectively. It shows the the performance
of DAB design for disjoint clusters is the best, because AoA ranges of the two clusters are not highly overlapping.

The simulation results above show that our proposed DAB designs can achieve better performance than the existing approaches, since the proposed DAB designs reduce channel dimensionality to suppress the noise from the channel null-spaces.

C. Multi-Antenna Gain

To show the multi-antenna gain of the proposed designs, the channel-inversion (CI) AirComp scheme developed in [6] targeting single-antenna transceiver is compared with the proposed MIMO AirComp solution in Fig. 7. In the experiment, we consider two clusters of devices. The AoA ranges are \( \Delta \theta_1 = [-50^\circ, 10^\circ] \), \( \Delta \theta_2 = [-15^\circ, 45^\circ] \), respectively. We set the number of transmit antennas \( N_t = 1 \) for both schemes, while the number of receive antennas is set to \( N_r = 30 \) for the proposed schemes and \( N_r = 1 \) for the scheme in [6]. It is observed that the equipping with multiple antennas can significantly improve the MSE performance of AirComp, showing the multi-antenna gain and the effectiveness of the proposed aggregation beamforming.

D. Low Multi-Access Latency of AirComp

In Fig. 8, we show that AirComp can achieve low transmission latency compared to the traditional “transmit-then-compute” scheme. One cluster of devices is considered. In both schemes, linear analog modulation is used for transmission. For fair comparison, the SNRs of all devices in the “transmit-then-compute” scheme are assumed to be equal to that of AirComp. Besides, the latency is measured by the number of transmit symbol required. It is observed that, the transmission latency of the “transmit-then-compute” scheme linearly scales with the number of devices, while for AirComp, the transmission latency keeps low. This verifies the low-latency property of the proposed AirComp solution.

VII. CONCLUSION

In this paper, we have presented the framework of reduced-dimension MIMO AirComp for clustered IoT networks.
The design exploits the structure of clustered MIMO channel to reduce AirComp errors and channel-feedback overhead. The key feature of the framework is the design of decomposed aggregation beamforming, which comprises outer components performing channel dimension reduction and joint equalization of channel covariances and the inner components jointly equalize small-scale fading channels components.

The current work opens several directions for further investigation. One direction is algorithmic design for MIMO AirComp. In particular, sensor clustering algorithms can be designed to improve the performance of MIMO AirComp. Another interesting direction is to apply MIMO AirComp to specific IoT or distributed-learning applications such as high-mobility UAV networks, federated learning or cloud coordinated vehicular platooning.

APPENDIX

A. Proof of Lemma 1

We first construct an orthonormal basis of the $N_r$-dimensional space using the column vectors of $\{U_{g_i} \in [1, G]\}$. Then, each column vector of $A$ is presented as a linear combination of the basis. Finally, the decomposition form of $A$ is proved by combining the column vectors.

For notation simplicity, let $R_{G+1} = (N_r - \sum_{i=1}^{G} R_g)$. Define a $N_r \times R_{G+1}$ dimensional unitary matrix, $U_{G+1}$, which satisfies $U_{G+1}^H U_{G+1} = 0$ for all $g \in [1, G]$. Then, the column vectors of $\{U_{g_i} \in [1, G+1]\}$ forms an orthonormal basis of the $N_r$-dimensional space. Denote the $i$-th column vector of $U_J$ and $A^H$ as $u_{g,i}$ and $a_i$, respectively. According to the projection theory, we have

$$a_i = \sum_{g=1}^{G+1} \sum_{j=1}^{R_g} c_{i,j}^{(g)} u_{g,j}, \forall i \in [1, L], \quad (39)$$

where $c_{i,j}^{(g)}$ the coefficient and $c_{g,i} = [c_{i,1}^{(g)}, c_{i,2}^{(g)}, \ldots, c_{i,R_g}^{(g)}]^T$, respectively. Thereby, $A^H$ can be presented as

$$A^H = [a_1, a_2, \ldots, a_L] = \sum_{g=1}^{G+1} U_g [c_{g,1}, \ldots, c_{g,L}], \quad (40)$$

where $C_g = [c_{g,1}, \ldots, c_{g,L}]$.

Besides, with the channel, $H_{g,k}$, defined in (4), we have $C_{H, G+1}^H U_{G+1} H_{g,k} = 0$, for all $g \in [1, G]$. That’s to say, the component, $C_{H, G+1}^H$, of $A$ has no contribution solve Problem (P2). Then, let $C_{H, G+1} = 0$ for simplicity. Hence, we have

$$A = \sum_{g=1}^{G} C_g U_g^H. \quad (41)$$

This completes the proof.

B. Proof of Lemma 3

Denoting $J_{g,k} = A_{H,g,k} A_{H,g,k}^H$ and substituting $A = A_{(o)} A_{(i)}$ and (22), we have

$$J_{g,k} = A_{(o)} A_{(i)} \hat{U}_{g,k} \hat{W}_{g,k} A_{(o)} A_{(i)}^H \hat{W}_{g,k}^H \hat{U}_{g,k}^H A_{(i)} A_{(i)}^H A_{(o)}^H, \quad (42)$$

whose eigenvalues can be approximated to

$$\lambda_i (J_{g,k}) \geq \min \left( A_{(o)} A_{(i)} \hat{U}_{g,k} \hat{W}_{g,k} A_{(i)}^H, A_{(i)} A_{(i)}^H \right) \lambda_i \left( \hat{A}_{g,k} \hat{W}_{g,k} \hat{A}_{g,k}^H \right),$$

$$\geq \lambda_{\min} \left( A_{(i)} \hat{U}_{g,k} \hat{W}_{g,k} A_{(i)}^H \right) \lambda_{\min} \left( A_{(o)} A_{(o)}^H \right) \times \lambda_i \left( \hat{A}_{g,k} \hat{W}_{g,k} \hat{A}_{g,k}^H \right),$$

$$= \lambda_{\min} \left( A_{(i)} \hat{U}_{g,k} \hat{W}_{g,k} A_{(i)}^H \right) \lambda_i \left( \hat{A}_{g,k} \hat{W}_{g,k} \hat{A}_{g,k}^H \right),$$

where the last equality above is because $A_{(o)} A_{(o)}^H = I$. Besides, $tr \left( (A_{H,g,k} H_{g,k} A_{H,g,k})^{-1} \right) = \sum_{i=1}^{L} \lambda_i^{-1} (J_{g,k})$, by substituting the above inequality, we have

$$tr \left( (A_{H,g,k} H_{g,k} A_{H,g,k})^{-1} \right) \leq \min \left( A_{(i)} \hat{U}_{g,k} \hat{W}_{g,k} A_{(i)}^H \right) \times \sum_{i=1}^{L} \lambda_i^{-1} \left( \hat{A}_{g,k} \hat{W}_{g,k} \hat{A}_{g,k}^H \right).$$

This completes the proof.

C. Proof of Lemma 4

By substituting $F_{g,k} = \hat{A}_{g,k} \hat{W}_{g,k}$, we have

$$\lambda_{\min} (F_{g,k}) = \lambda_{\min} \left( W_{g,k} A_{H,g,k} \right).$$

Besides, each element of $W_{g,k}$ is i.i.d., and follows $CN(0,1)$. For notation simplicity, let $y_g = r$. Let $T_r = W_{g,k} A_{g,r} W_{r,g,k}$ and $T_{r+1} = W_{g,k,r+1} A_{g,r+1} W_{r,g,k,r+1}$, where the size of $W_{g,k,r}$ and $W_{r,g,k,r+1}$ are $r \times N_i$ and $(r+1) \times N_i$, and $A_{g,r}$ and $A_{g,r+1}$ are the $r$ and $r+1$ dominant eigenvalue matrix of $A_g$, respectively. Therefore, we have

$$T_{r+1} = T_r + \Delta, \quad (43)$$

where $\Delta$ is a non-negative matrix. According to Weyl’s inequality in matrix theory [39],

$$\lambda_i (T_r) \leq \lambda_i (T_{r+1}). \quad (44)$$

Therefore, $\lambda_{\min} (T_r) \leq \lambda_{\min} (T_{r+1})$. That’s to say, $\lambda_{\min} (F_{g,k})$ increases with $r$. This completes the proof.

REFERENCES

[1] M. Agiwal, A. Roy, and N. Saxena, “Next generation 5G wireless networks: A comprehensive survey,” IEEE Commun. Surveys Tut., vol. 18, no. 3, pp. 1617–1655, 3rd Quart., 2016.

[2] G. Zhu and K. Huang, “MIMO over-the-air computation for high-mobility multimodal sensing,” IEEE Internet Things J., vol. 6, no. 4, pp. 6089–6103, Aug. 2019.

[3] X. Li, G. Zhu, Y. Hong, 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.

[4] J. Dong, Y. Shi, and Z. Ding, “Blind over-the-air computation and data fusion via provable wittington flow,” Nov. 2018, arXiv:1811.04644. [Online]. Available: https://arxiv.org/abs/1811.04644

[5] G. Zhu, Y. Wang, and K. Huang, “Broadband analog aggregation for low-latency federated edge learning (extended version),” Dec. 2018, arXiv:1812.11494. [Online]. Available: https://arxiv.org/abs/1812.11494

[6] O. Abari, H. Rahul, and D. Katabi, “Over-the-air function computation in sensor networks,” Dec. 2016, arXiv:1612.02307. [Online]. Available: https://arxiv.org/abs/1612.02307

[7] B. Nazer and M. Gastpar, “Computation over multiple-access channels,” IEEE Trans. Inf. Theory, vol. 55, no. 10, pp. 3498–3516, Oct. 2007.

[8] R. C. Buck, “Nomographic functions are nowhere dense,” Proc. Amer. Math. Soc., vol. 85, no. 2, pp. 195–199, 1982.
[9] M. Goldenbaum, H. Boche, and S. Stanczak, “Nonomographic gossiping for f-consensus,” in Proc. 10th Int. Symp. Modeling Optim. Mobile, Ad Hoc Wireless Netw. (WiOpt), May 2012, pp. 130–137.

[10] R. C. Buck, “Approximate complexity and functional representation,” J. Math. Anal. Appl., vol. 70, pp. 280–298, Jul. 1979.

[11] M. Gastpar, “Uncoded transmission is exactly optimal for a simple Gaussian ‘sensor’ network,” IEEE Trans. Inf. Theory, vol. 54, no. 11, pp. 5247–5251, Nov. 2008.

[12] A. B. Wagner, S. Tavildar, and P. Viswanath, “Rate region of the quadratic Gaussian two-encoder source-coding problem,” IEEE Trans. Inf. Theory, vol. 54, no. 5, pp. 1938–1961, May 2008.

[13] J. J. Xiao, S. Cui, Z. Q. Luo, and A. J. Goldsmith, “Linear coherent decentralized estimation,” IEEE Trans. Signal Process., vol. 56, no. 2, pp. 777–770, Feb. 2008.

[14] C.-H. Wang, A. S. Leong, and S. Dey, “Distortion outage minimization and diversity order analysis for coherent multiaccess,” IEEE Trans. Signal Process., vol. 59, no. 12, pp. 6144–6159, Dec. 2011.

[15] M. Goldenbaum and S. Stanczak, “On the channel estimation effort for analog computation over wireless multiple-access channels,” IEEE Wireless Commun. Lett., vol. 3, no. 3, pp. 261–264, Jun. 2014.

[16] M. Goldenbaum and S. Stanczak, “Robust analog function computation via wireless multiple-access channels,” IEEE Trans. Commun., vol. 61, no. 9, pp. 3863–3877, Sep. 2013.

[17] O. Abari, H. Rahul, D. Katabi, and M. Pant, “AirShare: Distributed decentralized estimation,” in Proc. IEEE Conf. Comput. Commun. (INFOCOM), Apr./May 2015, pp. 1742–1750.

[18] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, “Massive MIMO for next generation wireless systems,” Apr. 2013, arXiv:1304.6690 [Online]. Available: https://arxiv.org/abs/1304.6690

[19] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, “An overview of massive MIMO: Benefits and challenges,” IEEE J. Sel. Topics Signal Process., vol. 8, no. 5, pp. 742–758, Oct. 2014.

[20] L. Liang, W. Xu, and X. Dong, “Low-complexity hybrid precoding in massive multiuser MIMO systems,” IEEE Wireless Commun. Lett., vol. 3, no. 6, pp. 653–656, Dec. 2014.

[21] R. Zakhour and S. Hanly, “Min-max fair coordinated beamforming via wireless multiple-access channels,” IEEE Trans. Commun., vol. 66, no. 9, pp. 3863–3877, Sep. 2018.

[22] A. Adhikary et al., “Joint spatial division and multiplexing—the large-scale array regime,” J. Sel. Areas Commun., vol. 35, no. 9, pp. 2558–2570, Sep. 2017.

[23] J. Yu, J.-C. Shen, J. Zhang, and K. B. Letaief, “Alternating minimization algorithms for hybrid precoding in millimeter wave MIMO systems,” IEEE J. Sel. Topics Signal Process., vol. 10, no. 3, pp. 485–500, Apr. 2016.

[24] O. El Ayach, S.-R. Lee, S.-R. Lee, S. Maeng, J. Jung, and I. Lee, “New interference mitigation algorithms for multiuser detection,” IEEE Trans. Commun., vol. 70, pp. 280–298, Jul. 1979.

[25] A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire, “Joint spatial division and multiplexing: the large-scale array regime,” IEEE Trans. Inf. Theory, vol. 59, no. 10, pp. 6441–6463, Oct. 2013.

[26] A. Liu and V. Lau, “Hierarchical interference mitigation for massive MIMO cellular networks,” IEEE Trans. Signal Process., vol. 62, no. 18, pp. 4786–4797, Sep. 2014.

[27] J. Nam, A. Adhikary, J.-Y. Ahn, and G. Caire, “Joint spatial division and multiplexing: Opportunistic beamforming, user grouping and simplified downlink scheduling,” IEEE J. Sel. Topics Signal Process., vol. 8, no. 5, pp. 876–890, Oct. 2014.

[28] Y. Jeon, C. Song, S.-R. Lee, S. Maeng, J. Jung, and I. Lee, “New beamforming designs for joint spatial division and multiplexing in large-scale MISO multi-user systems,” IEEE Trans. Wireless Commun., vol. 16, no. 5, pp. 3029–3041, May 2017.

[29] A. Klein, G. K. K Taleh, and P. W. Baier, “Zero forcing and minimum mean-square-error equalization for multiuser detection in code-division multiple-access channels,” IEEE Trans. Veh. Technol., vol. 45, no. 2, pp. 276–287, May 1996.

[30] D.-S. Shiu, G. J. Foschini, M. J. Gans, and J. M. Kahn, “Fading correlation and its effect on the capacity of multielement antenna systems,” IEEE Trans. Commun., vol. 48, no. 3, pp. 502–513, Mar. 2000.

[31] D. J. Love and R. W. Heath, Jr., “Limited feedback unitary precoding for spatial multiplexing systems,” IEEE Trans. Inf. Theory, vol. 51, no. 8, pp. 2967–2976, Aug. 2005.

[32] J. Choi, B. Mondal, and R. W. Heath, “Interpolation based unitary precoding for spatial multiplexing MIMO-OFDM with limited feedback,” IEEE Trans. Signal Process., vol. 54, no. 12, pp. 4730–4740, Dec. 2006.

[33] S. W. Peters and R. W. Heath, “Cooperative algorithms for MIMO interference channels,” IEEE Trans. Veh. Technol., vol. 60, no. 1, pp. 206–218, Jan. 2011.

[34] A. Edelman, T. A. Arias, and S. T. Smith, “The geometry of algorithms with orthogonality constraints,” SIAM J. Matrix Anal. Appl., vol. 20, no. 2, pp. 303–353, 1998.

[35] T. L. Marzetta and B. M. Hochwald, “Fast transfer of channel state information in wireless systems,” IEEE Trans. Signal Process., vol. 54, no. 4, pp. 1268–1278, Apr. 2006.

[36] R. Bhatia, Matrix Analysis, 2nd ed. New York, NY, USA: Springer, 2013.