IEICE TRANS. COMMUN., VOL.E107–B, NO.3 MARCH 2024

PAPER

Precoder Optimization Using Data Correlation for Wireless Data Aggregation

Ayano NAKAI-KASAI†a), *Member*, Naoyuki HAYASHI†, *Nonmember*, *and* Tadashi WADAYAMA†, *Senior Member*

SUMMARY In this paper, we consider precoder design for wireless data aggregation in sensor networks. The precoder optimization problem can be formulated as minimization of mean squared error under transmit power and block diagonal constraints. We include statistical correlation of data into the optimization problem, which is appeared in typical applications but is ignored in conventional designing methods. We propose precoder optimization algorithms based on projected gradient descent with projection onto the constraint sets. The proposed method can achieve better performance than the conventional methods that do not incorporate data correlation, especially when data are highly correlated. We also extend the proposed approach to the context of over-the-air computation.
key words: *wireless data aggregation, precoder optimization, data correlation, projected gradient descent, over-the-air computation*

1. Introduction

Internet-of-Things (IoT)-based automatic operations in 5th and 6th generation communication systems such as intelligent agriculture and industrial surveillance are supported by environmental data collection by distributed multiple sensors [1], [2]. The sensors measure environmental data such as temperature, humidity, and amount of chemicals, and transmit the data to a data aggregator (or central server) that performs data processing for the application via wireless communication. The process is called wireless data aggregation [3].

Data collected by sensors often have statistical correlation in typical applications of wireless data aggregation [4], [5]. One of the correlations is spatial correlation, where data collected by sensors located in close geographically are also close in the sensing values. There are also correlations in data types and temporal correlation. For example, higher temperatures lead to lower humidity and temperatures change slowly during the day. Such correlation is often used in signal processing to improve system performance [6].

Improvement of communication quality is essential for constructing reliable IoT applications using wireless data aggregation. The data aggregator is in general equipped with multiple receive antennas and an advanced computation system for data processing. The communication system between the aggregator and sensors can be modeled as a multiple-input multiple-output (MIMO) channel [7]. More advanced sensor devices transmitting vector values and with multiple transmit antennas such as drones have been recently considered [8], [9]. The system is then modeled as a multiuser MIMO channel. In the system, the received data at the aggregator deteriorate due to channel conditions and additive noise in general. To reconstruct the transmitted data as correctly as possible, precoder and decoder must be applied for the transmitted and received signals, respectively [10].

The precoder and decoder designing problem for wireless data aggregation can be formulated as a nonconvex optimization problem for deriving a precoding matrix with block diagonal structure and a decoding matrix under transmit power constraint. Huang et al. [11] proposed a non-iterative optimization method that first solves a convex optimization problem by ignoring the block diagonal constraint and data correlation, and then performs block diagonalization. Huh et al. [12] originally proposed a precoder design for over-the-air computation (AirComp) [13] but it can be applied to this problem. The precoder optimization problem in this method can be divided into multiple convex optimization problems by ignoring spatial correlation of data. The problem is iteratively solved by block coordinate descent method.

In these conventional methods, however, data correlation is ignored in the formulation of optimization problems, though it appears in typical applications of wireless data aggregation. In other words, the precoder and decoder derived from the conventional methods may not be suitable for practical situations of wireless data aggregation. To further improve communication quality, it is expected to formulate the optimization problem that can appropriately incorporate data correlation and construct an algorithm for solving the problem.

For these demands, this paper aims to develop a novel precoder and decoder design where data correlation is appropriately taken into account. The objective is to achieve better detection performance of transmitted data than the conventional methods ignoring data correlation, and to make the system more suitable for practical scenarios. We first formulate an optimization problem including statistical correlation of data. The problem is nonconvex minimization problem under the block diagonal constraint and the transmit power constraint. To solve the problem, we employ projected gradient descent method that uses projection onto the constraint sets. The proposed method is expected to be suitable for practical environments by adopting data correlation in the

Manuscript received June 1, 2023.
Manuscript revised August 31, 2023.
Manuscript publicized January 30, 2024.
†The authors are with the Graduate School of Engineering, Nagoya Institute of Technology, Nagoya-shi, 466-8555 Japan.
a) E-mail: nakai.ayano@nitech.ac.jp
DOI: 10.23919/transcom.2023EBT0007

design and by precisely satisfying the constraints. Performance of the proposed method is evaluated on synthetic and real environmental data [14].

This paper also extends the idea of the proposed method to the context of AirComp [13]. AirComp enables simultaneous transmission and processing of data to achieve immediate responsiveness in IoT applications. Sensor nodes simultaneously transmit signals over the same frequency band. The aggregator receives the sum of the transmitted signals with the analog-wave superposition property of wireless multiple-access channels and obtains function values that can be directly calculated from the received sum such as arithmetic mean, weighted sum, or Euclidean norm. The calculation of such functions is the typical objective in many applications of sensor networks including federated learning which is known as a recent distributed machine learning technique [15]. Also in AirComp, however, there occurs aggregation error, i.e., the error between the actual sum of transmitted data and the aggregated signals on the air, due to channel conditions and additive noise. Therefore, precoder and decoder are required to reduce the aggregation error. This paper applies the proposed approach to the context of AirComp and evaluates its performance.

2. Preliminalies

2.1 Notation

In the rest of the paper, we use the following notation. Superscripts $(\cdot)^{\mathrm{T}}$ and $(\cdot)^{\mathrm{H}}$ denote the transpose and the Hermitian transpose, respectively. The zero vector, zero matrix, and identity matrix are represented as $\mathbf{0}$, $\boldsymbol{O}$, and $\boldsymbol{I}$, respectively. Euclidean (ℓ_2) norm is represented by $\|\cdot\|$. The real Gaussian distribution $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ and complex circularly symmetric Gaussian distribution $\mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma})$ have mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$. The expectation and trace operators are $\mathbb{E}[\cdot]$ and $\mathrm{Tr}[\cdot]$, respectively. Hadamard product, i.e., element-wise multiplication of matrices, is represented as $\odot$.

2.2 System Model

Assume a wireless data aggregation system with a single aggregator equipped with r receive antennas and L sensor nodes equipped with m transmit antennas, as illustrated in Fig. 1. Each sensor node s_i $(i = 1, 2, \ldots, L)$ obtains n-dimensional data, which are expressed as a vector $\boldsymbol{x}_i \in \mathbb{R}^n$. The data vector $\boldsymbol{x}_i$ is composed of, for example, n different types of data observed by a multi-modal sensor, such as temperature and humidity at the same time, or sequential data observed at a sensor, such as temperatures at n different times. Each node transmits a linearly precoded data vector

$$\boldsymbol{c}_i = \boldsymbol{A}_i \boldsymbol{x}_i \in \mathbb{C}^m \tag{1}$$

to the aggregator, where $\boldsymbol{A}_i \in \mathbb{C}^{m\times n}$ is a precoding matrix. The channel matrix in terms of node s_i and the aggregator is assumed to be $\boldsymbol{H}_i \in \mathbb{C}^{r\times m}$. The aggregator receives the sum of the transmitted data from all the nodes. The received signal is given by

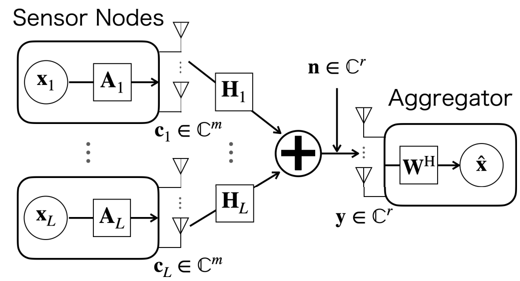


Fig. 1 System model for wireless data aggregation.

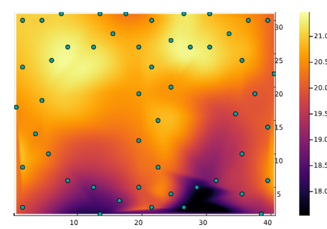


Fig. 2 Sensor positions and room temperature of real data [14].

$$\boldsymbol{y} = \left(\sum_{i=1}^{L} \boldsymbol{H}_i \boldsymbol{c}_i\right) + \boldsymbol{n} = \left(\sum_{i=1}^{L} \boldsymbol{H}_i \boldsymbol{A}_i \boldsymbol{x}_i\right) + \boldsymbol{n} \in \mathbb{C}^r, \tag{2}$$

where $\boldsymbol{n}$ is a complex additive Gaussian noise vector and is independent of $\boldsymbol{x}_i$. The equivalent model can be expressed as

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{A}\boldsymbol{x} + \boldsymbol{n}, \tag{3}$$

where $\boldsymbol{x} = [\boldsymbol{x}_1^{\mathrm{T}}, \ldots, \boldsymbol{x}_L^{\mathrm{T}}]^{\mathrm{T}}$, $\boldsymbol{H} = \left[\boldsymbol{H}_1, \ldots, \boldsymbol{H}_L\right] \in \mathbb{C}^{r\times mL}$, and

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_1 & & \boldsymbol{O} \\ & \ddots & \\ \boldsymbol{O} & & \boldsymbol{A}_L \end{bmatrix} \in \mathbb{C}^{mL\times nL}, \tag{4}$$

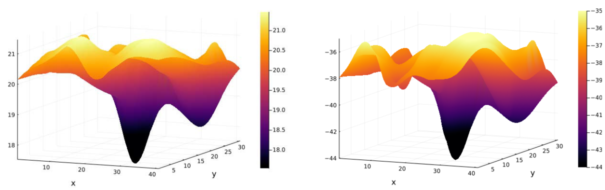
i.e., the precoding matrix $\boldsymbol{A}$ has a block diagonal structure.

The aggregator reconstructs the original data vector $\boldsymbol{x}$ by linearly multiplying a decoding matrix $\boldsymbol{W} \in \mathbb{C}^{r\times nL}$ with the received signal $\boldsymbol{y}$. The estimated data $\hat{\boldsymbol{x}}$ is given by $\hat{\boldsymbol{x}} = \boldsymbol{W}^{\mathrm{H}}\boldsymbol{y}$.

2.3 Assumptions on Correlation

Data collected by sensors tend to have statistical correlation in many applications. Figure 2 represents an example of real sensing data [14], where distributed sensors collect environmental data such as room temperature and humidity. From Fig. 2, we can see that sensors that are close in distance have close sensing values, that is, there is spatial correlation in sensing data. Moreover, it is also known that there are correlation between different data types (e.g., temperature and humidity as shown in Fig. 3) and temporal correlation [4].

With these facts, in this paper, we assume a correlation



(a) Temperature **(b)** Humidity

Fig. 3 Example of temperature and humidity data [14].

model for the data vector $\boldsymbol{x}$. The data vector $\boldsymbol{x}$ is assumed to follow Gaussian distribution $\mathcal{N}(\mathbf{0}, \boldsymbol{K})$. The positive definite covariance matrix $\boldsymbol{K} \in \mathbb{R}^{nL \times nL}$ has the following structure:

$$\boldsymbol{K} = \begin{bmatrix} \boldsymbol{K}_{11} & \boldsymbol{K}_{12} & \cdots & \boldsymbol{K}_{1L} \\ \boldsymbol{K}_{21} & \boldsymbol{K}_{22} & \cdots & \boldsymbol{K}_{2L} \\ \vdots & & \ddots & \vdots \\ \boldsymbol{K}_{L1} & \boldsymbol{K}_{L2} & \cdots & \boldsymbol{K}_{LL} \end{bmatrix}. \tag{5}$$

The diagonal block matrix $\boldsymbol{K}_{ii} \in \mathbb{R}^{n \times n}$ corresponds to correlation in sensing data of node s_i. When the data vector $\boldsymbol{x}_i$ is composed of n different types of data, the block includes correlation between the data types. When $\boldsymbol{x}_i$ is composed of data at n discrete times, the block includes temporal correlation. The non-diagonal block matrix $\boldsymbol{K}_{ij}$ $(i \neq j)$ includes spatial correlation between nodes s_i and s_j. Data encoding is not assumed in this paper to retain data correlation. It should be noted that the data cannot be whitened in each sensor due to spatial correlation among sensors.

In practical situations, the covariance matrix $\boldsymbol{K}$ can be estimated by using sample covariance matrix with respect to a part of the data. We will show the results of using the sample covariance matrix in Sect. 3.2.3.

We also assume that the noise vector $\boldsymbol{n}$ follows $\boldsymbol{n} \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{S})$. The positive definite covariance matrix $\boldsymbol{S} \in \mathbb{C}^{r \times r}$ represents the correlation of the noise vector.

2.4 Precoder Optimization Problems

In this paper, we explore the precoding matrix $\boldsymbol{A}$ and decoding matrix $\boldsymbol{W}$ that minimize mean squared error (MSE) in terms of the original data vector $\boldsymbol{x}$ and the estimated vector $\hat{\boldsymbol{x}}$. We name the context that aims to aggregate the original data vector $\boldsymbol{x}$ *full aggregation system*. We assume that the aggregator executes the optimization process of the precoder using all sensors' information and distributes the optimized precoder for each sensor.

The precoding matrix is required to have a block diagonal structure (4). This corresponds to the following constraint $C0$:

$$(C0) \quad \boldsymbol{A} \in \mathbb{B}_L^{m \times n}, \tag{6}$$

where $\mathbb{B}_L^{m \times n}$ is the set of complex block diagonal matrices with L diagonal blocks of $m \times n$ matrices. In addition, we consider a constraint on transmit power. We here employ the constraint that the total transmit power of all nodes, $\mathbb{E}[\|\boldsymbol{Ax}\|^2]$, is fixed to the constant P, i.e.,

Table 1 Precoder optimization problems in each section.

Sect.	Context	Constraints
3	Full aggregation	total transmit power ($C1$)
4	Full aggregation	transmit power /sensor ($C2$)
5	AirComp	total transmit power ($C1$)

$$(C1) \quad \mathbb{E}[\|\boldsymbol{Ax}\|^2] = P. \tag{7}$$

Another constraint can be considered and will be discussed later. Therefore, the optimization problem under the above two constraints $C0$ and $C1$ is summarized as

$$\begin{aligned} (\mathcal{P}1) \quad & \text{minimize}_{\boldsymbol{A},\boldsymbol{W}} \ \mathbb{E}\left[\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|^2\right] \\ & \text{s.t. } \mathbb{E}[\|\boldsymbol{Ax}\|^2] = P, \ \boldsymbol{A} \in \mathbb{B}_L^{m \times n}, \ \boldsymbol{W} \in \mathbb{C}^{r \times nL}. \end{aligned} \tag{8}$$

The problem $\mathcal{P}1$ is a nonconvex optimization problem so that it is difficult to solve analytically in general.

In the following, we use the decoding matrix that minimizes the MSE under fixed $\boldsymbol{A}$, i.e., minimum MSE (MMSE) decoding matrix. The MSE is represented by

$$\begin{aligned} \text{MSE} =& \mathbb{E}\left[\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|^2\right] = \text{Tr}[\boldsymbol{K} + \boldsymbol{W}^{\mathrm{H}}\boldsymbol{HAKA}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{W} \\ & + \boldsymbol{W}^{\mathrm{H}}\boldsymbol{SW} - \boldsymbol{W}^{\mathrm{H}}\boldsymbol{HAK} - \boldsymbol{KA}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{W}]. \end{aligned} \tag{9}$$

The MMSE decoding matrix $\boldsymbol{W}(\boldsymbol{A})$ can be obtained by the stationary condition $\partial \text{MSE}/\partial \boldsymbol{W} = 0$ and is given by

$$\boldsymbol{W}(\boldsymbol{A}) = \left(\boldsymbol{HAKA}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}} + \boldsymbol{S}\right)^{-1} \boldsymbol{HAK}. \tag{10}$$

By substituting it into the problem $\mathcal{P}1$ and using Woodbury matrix identity, we obtain the following precoder optimization problem under the constraints $C0$ and $C1$:

$$\begin{aligned} (\mathcal{P}1') \quad & \text{minimize}_{\boldsymbol{A}} \ \text{Tr}\left[\left(\boldsymbol{K}^{-1} + \boldsymbol{A}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{S}^{-1}\boldsymbol{HA}\right)^{-1}\right] \\ & \text{s.t. } \mathbb{E}[\|\boldsymbol{Ax}\|^2] = P, \ \boldsymbol{A} \in \mathbb{B}_L^{m \times n}. \end{aligned} \tag{11}$$

An algorithm for solving the problem $\mathcal{P}1'$ is derived in Sect. 3. This paper also includes other problem formulations for precoder optimization. We can replace the constraint $C1$ in the full aggregation system with another practical power constraint; transmit power constraints per sensor, which is discussed in Sect. 4. Moreover, we can deal with the precoder optimization problem for the context of AirComp as shown in Sect. 5. The constraint $C0$ is still required in these cases. Table 1 summarizes the organization of the paper.

3. Precoder Optimization Algorithm under Total Transmit Power Constraint

In this section, we derive an algorithm for solving the problem $\mathcal{P}1'$ and evaluate the performance of the algorithm.

3.1 Derivation

The precoder optimization problem $\mathcal{P}1'$ is still difficult to solve analytically because it includes the block diagonal constraint of the precoder. The problem cannot be separated into the problems in terms of each block A_i due to the existance of K.

Therefore, we propose a precoder optimization algorithm based on a projected gradient descent method that does not require any assumption of uncorrelation of data. The proposed algorithm allows precoder design to take the block diagonal constraint $C0$ and the total transmit power constraint $C1$ into consideration by using projection onto the constraint sets. The algorithm is derived from the precoder optimization problem $\mathcal{P}1'$ including the covariance matrix K of the data so that it can appropriately incorporate data correlation.

The proposed algorithm iterates gradient descent and projection onto the constraint sets. At tth iteration ($t = 1, 2, \ldots, J$), the following four steps are executed:

Step (A) Calculate the gradient matrix $G^{(t-1)}$ of the MSE in terms of $A^{(t-1)}$,

Step (B) Update $A^{(t-1)}$ by using gradient descent method and obtain $B^{(t)}$,

Step (C) Project $B^{(t)}$ onto the block diagonal constraint set $\mathbb{B}_L^{m\times n}$ and obtain $C^{(t)}$,

Step (D) Project $C^{(t)}$ onto the total transmit power constraint set and obtain the estimate $A^{(t)}$.

In Step (A), the gradient matrix $G^{(t-1)}$ can be calculated from the cost function of (11) as

$$G^{(t-1)} = \nabla_A \left(\mathrm{Tr}\left[F(A)^{-1}\right]\right)|_{A=A^{(t-1)}}, \tag{12}$$

where $F(A) = K^{-1} + A^{\mathrm{H}} H^{\mathrm{H}} S^{-1} H A$. It can be obtained by using automatic differentiation mechanisms in neural network frameworks or some matrix calculus [16].

In Step (B), the update equation is given by

$$B^{(t)} = A^{(t-1)} - \mu G^{(t-1)}, \tag{13}$$

where $\mu \in \mathbb{R}$ is the step-size parameter and $\mu > 0$.

In Step (C), the projection onto the block diagonal constraint set $\mathbb{B}_L^{m\times n}$ is performed by Hadamard product of $B^{(t)}$ and a mask matrix $M \in \mathbb{B}_L^{m\times n}$. The mask matrix has all 1 elements in the block diagonal matrices and 0 for the other elements. The projection is given by

$$C^{(t)} = B^{(t)} \odot M. \tag{14}$$

In Step (D), the total transmit power $\mathbb{E}[\|Ax\|^2]$ can be expanded as $\mathbb{E}[\|Ax\|^2] = \mathrm{Tr}[AKA^{\mathrm{H}}]$, so that the projection to satisfy $C1$ is given by

$$A^{(t)} = C^{(t)} \sqrt{\frac{P}{\mathrm{Tr}\left[C^{(t)} K (C^{(t)})^{\mathrm{H}}\right]}}. \tag{15}$$

The proposed method is summarized in Algorithm 1.

Algorithm 1 Precoder Optimization Algorithm under Total Transmit Power Constraint

Input: $A^{(0)} \in \mathbb{B}_L^{m\times n}$ satisfying transmit power constraint
Output: $A^{(J)} \in \mathbb{B}_L^{m\times n}$

```
1: for t = 1, 2, ..., J do
2:     G^(t-1) = ∇_A (Tr[F(A)^(-1)])|_{A=A^(t-1)}
3:     B^(t) = A^(t-1) − μ G^(t-1)
4:     C^(t) = B^(t) ⊙ M
5:     A^(t) = C^(t) sqrt( P / Tr[C^(t) K (C^(t))^H] )
6: end for
```

The MMSE decoding matrix can be obtained by substituting $A^{(J)}$ into (10). The main factor of the computational costs is gradient calculation in Step (A). This is required in each step so that the complexity is comparable to the conventional iterative method based on [12] and is J times the order of the conventional non-iterative method based on eigenvalue decomposition [11].

3.2 Performance Evaluation

3.2.1 Settings

In this section, we evaluate estimation performance with the precoder obtained by the proposed algorithm via numerical evaluation. All the evaluations were implemented by Julia language. We used the automatic differentiation mechanism in Flux library for deriving the gradient matrix in Step (A). The number J of iterations of the proposed methods are set to yield sufficient convergence in each experiment.

The covariance matrix S of the noise vector was set to $S = \sigma^2 I$, where σ^2 is the noise variance. Signal-to-noise ratio (SNR) in the following results is defined as $\mathrm{SNR} = \mathbb{E}[\|Ax\|^2]/\mathbb{E}[\|n\|^2] = P/(r\sigma^2)$. Each element of the channel matrix H was generated from $\mathcal{CN}(0,1)$. We calculated the estimation of MSE in the following experiments by randomly generating H over 100 times.

We employed three baseline algorithms for comparing the performance with that of the proposed algorithm. The method shown as "Random" in legends is a random precoding method, where each element of the precoding matrix follows $\mathcal{CN}(0,1)$ and where the precoding matrix has a block diagonal structure and is normalized to satisfy the total transmit power constraint. The method of Huang et al. [11] is shown as "Huang" in legends. The method shown as "Huh" in legends is the precoding algorithm inspired by [12], whereas the original proposal of Huh et al. [12] is for the context of AirComp. For sufficient convergence of the method [12], the number of iterations of the method was set to 250 except for Fig. 6, where it was set to 300.

3.2.2 Evaluation on Synthetic Data

We first evaluate the performance on synthetic data. In this subsection, the following synthetic correlation model is assumed:

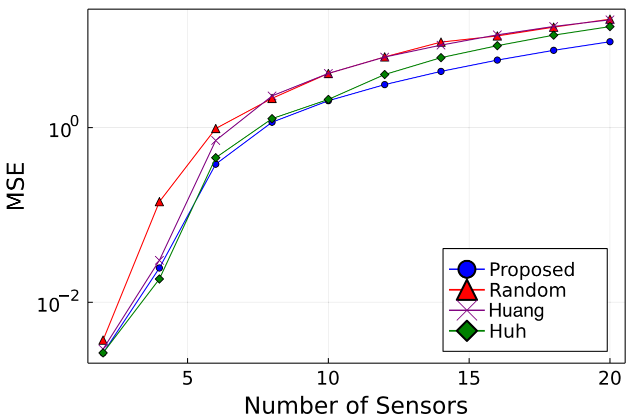


Fig. 4 MSE vs. the number of sensor nodes (Alg. 1), where $(n, m, r, a, P, \mathrm{SNR}) = (2, 2, 20, 0.75, 20, 20(\mathrm{dB}))$.

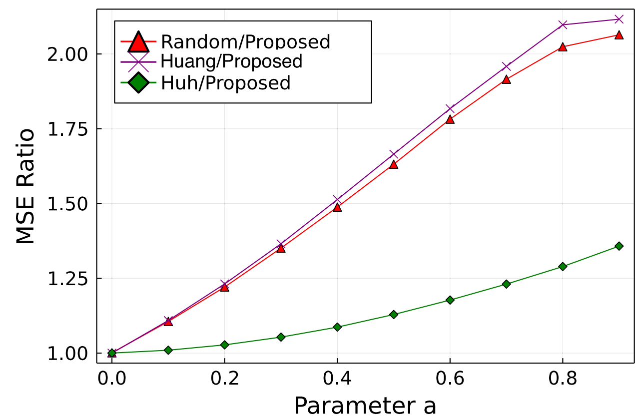


Fig. 5 MSE ratio vs. correlation parameter a (Alg. 1), where $(n, m, r, L, P, \mathrm{SNR}) = (2, 2, 20, 20, 20, 20(\mathrm{dB}))$.

$$\boldsymbol{K} = \begin{bmatrix} 1 & a & a^2 & \cdots & a^{nL-1} \\ a & 1 & a & \cdots & a^{nL-2} \\ a^2 & a & 1 & \cdots & a^{nL-3} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ a^{nL-1} & a^{nL-2} & a^{nL-3} & \cdots & 1 \end{bmatrix}, \quad (16)$$

where a is a correlation parameter and $0 \leq a \leq 1$. The larger a shows the stronger correlation of data. This model corresponds, for example, to the case where sensor nodes are arranged in a row with equal spacing in order of sensor ID, or the data vector $\boldsymbol{x}_i$ is composed of sequential data sampled at equally spaced times with correlation according to time interval.

Figure 4 shows MSE values (9) obtained by the precoder optimization methods in terms of the number L of sensor nodes. The system parameters were set to $(n, m, r, a, P, \mathrm{SNR}) = (2, 2, 20, 0.75, 20, 20(\mathrm{dB}))$. The total transmit power P was assumed to be constant regardless of the number of sensors. The number of iterations and step-size parameter of the proposed algorithm were $J = 5000$ and $\mu = 0.5$, respectively. In the cases of $L = 2, 4, \ldots, 10$, there is little difference between the performance of the proposed method and those of the conventional methods. However, it can be seen that the performance difference increases as the number of sensor nodes increases. In other words, the proposed method can achieve better performance for large-scale sensor networks.

Figure 5 shows MSE ratio obtained by the precoder optimization methods when the correlation parameter a varies as $a = 0, 0.1, 0.2, \ldots, 0.9$. The MSE ratio means the ratio of MSE values (9) of methods to that of the proposed method. For example, the MSE ratio 2 means that the MSE value of the method is twice the MSE value of the proposed method. The higher the value of the MSE ratio, the worse the performance of the method with respect to the proposed method. The system parameters were set to $(n, m, r, L, P, \mathrm{SNR}) = (2, 2, 20, 20, 20, 20(\mathrm{dB}))$. The number of iterations and the step-size parameter of the proposed algorithm were $J = 5000$ and $\mu = 0.5$, respectively. From the figure, as the correlation parameter a increases, the MSE ratio of the three conventional methods also increases. The three baseline algorithms do not incorporate data correlation but the proposed algorithm appropriately incorporates it so that the proposed method achieves better performance, especially in the case where the correlation parameter a is large, i.e., the data are highly correlated.

Table 2 Example of real data [14].

date	time	id	temperature	humidity	light
2004-02-28	03:16:3	1	19.3024	38.4629	45.08
2004-02-28	06:16:0	1	19.1652	38.8039	45.08

3.2.3 Evaluation on Real Data

In this subsection, we evaluate the performance on real data sensors placed indoors collected. We have used Intel Berkeley Research Lab Sensor Data [14], where 54 sensors placed in the laboratory collect environmental data from February 28, 2004 to April 5. A part of the data is shown in Table 2. We have removed missing values and picked up the daily average of temperature and humidity data (data from $L = 47$ sensors remained and $n = 2$). The sample covariance matrix $\hat{\boldsymbol{K}}$ was calculated by using daily average data from February 28 to March 3. We used the data on March 4 and 5 as the original data vector $\boldsymbol{x}$. The mean of the data was preprocessed to become zero.

We have calculated the average of squared errors $\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|^2$ obtained by the precoder optimization methods. Figure 6 shows the performance in each SNR. The system parameters were set to $(m, r, P) = (2, 141, 30)$. The number of iterations and the step-size parameter were $J = 10000$ and $\mu = 15$, respectively. The result of "Huh" is removed from the figure because it shows considerably worse performance than the other methods. From Fig. 6, the proposed method shows comparable or better MSE performance than the method of Huang et al. It is confirmed that the proposed method performs better also in realistic situations with correlated data.

It should be noted that the proposed method may not necessarily perform well with the sample covariance matrix

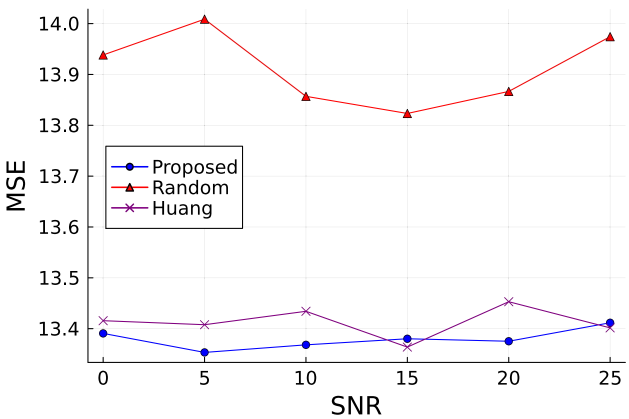


Fig. 6 MSE vs. SNR on real data [14] (Alg. 1), where $(m, r, P) = (2, 141, 30)$.

if the statistical properties of the data vary rapidly, or if sufficient data for estimating the true covariance matrix are not available for constructing the sample covariance matrix.

4. Precoder Optimization Algorithm under Transmit Power Constraints per Sensor

This section deals with the precoder optimization problem under transmit power constraints per sensor.

4.1 Derivation

In many applications, sensor network is composed of small sensors with small batteries so that it requires severe power constraint per sensor. The transmit power of sensor s_i $(i = 1, \ldots, L)$ can be expressed as $\mathbb{E}[\|\boldsymbol{A}_i \boldsymbol{x}_i\|^2]$, because the transmit signal is (1). We employ the constraint that the transmit power of sensor s_i is fixed to the constant P_i, i.e.,

$$(C2) \quad \mathbb{E}[\|\boldsymbol{A}_i \boldsymbol{x}_i\|^2] = P_i \ (i = 1, \ldots, L). \tag{17}$$

The precoder optimization problem in this case is under the block diagonal constraint $C0$ and the transmit power constraints per sensor $C2$, and is summarized as

$$(\mathcal{P}2) \quad \text{minimize}_{\boldsymbol{A}} \ \mathrm{Tr}\left[\left(\boldsymbol{K}^{-1} + \boldsymbol{A}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{S}^{-1}\boldsymbol{H}\boldsymbol{A}\right)^{-1}\right]$$
$$\text{s.t.} \ \ \boldsymbol{A} \in \mathbb{B}_L^{m\times n}, \mathbb{E}[\|\boldsymbol{A}_i \boldsymbol{x}_i\|^2] = P_i \ \ (i = 1, \ldots, L). \tag{18}$$

We propose a precoder optimization algorithm under transmit power constraints per sensor based on the projected gradient descent method, which is derived from the same approach as in Sect. 3. The only different point from the previous optimization problem $\mathcal{P}1'$ is the transmit power constraint. The transmit power per sensor can be expanded as $\mathbb{E}[\|\boldsymbol{A}_i \boldsymbol{x}_i\|^2] = \mathrm{Tr}[\boldsymbol{A}_i \boldsymbol{K}_i \boldsymbol{A}_i^{\mathrm{H}}]$. We replace Step(D) of the previous algorithm, the projection onto the transmit power constraint set, with the projection to satisfy $C2$:

Algorithm 2 Precoder Optimization Algorithm under Transmit Power Constraints per Sensor

Input: $\boldsymbol{A}^{(0)} \in \mathbb{B}_L^{m\times n}$ satisfying transmit power constraint
Output: $\boldsymbol{A}^{(J)} \in \mathbb{B}_L^{m\times n}$
1: **for** $t = 1, 2, \ldots, J$ **do**
2: $\quad \boldsymbol{G}^{(t-1)} = \nabla_{\boldsymbol{A}}(\mathrm{Tr}[\boldsymbol{F}(\boldsymbol{A})^{-1}])|_{\boldsymbol{A}=\boldsymbol{A}^{(t-1)}}$
3: $\quad \boldsymbol{B}^{(t)} = \boldsymbol{A}^{(t-1)} - \mu \boldsymbol{G}^{(t-1)}$
4: $\quad \boldsymbol{C}^{(t)} = \boldsymbol{B}^{(t)} \odot \boldsymbol{M}$
5: $\quad$ **for** $i = 1, 2, \ldots, L$ **do**
6: $\quad\quad \boldsymbol{A}_i^{(t)} = \boldsymbol{C}_i^{(t)} \sqrt{\dfrac{P_i}{\mathrm{Tr}\left[\boldsymbol{C}_i^{(t)} \boldsymbol{K}_{ii} \left(\boldsymbol{C}_i^{(t)}\right)^{\mathrm{H}}\right]}}$
7: $\quad$ **end for**
8: **end for**

$$\boldsymbol{A}_i^{(t)} = \boldsymbol{C}_i^{(t)} \sqrt{\frac{P_i}{\mathrm{Tr}\left[\boldsymbol{C}_i^{(t)} \boldsymbol{K}_{ii} \left(\boldsymbol{C}_i^{(t)}\right)^{\mathrm{H}}\right]}}, \tag{19}$$

where $\boldsymbol{A}_i^{(t)} \in \mathbb{C}^{m\times n}$ and $\boldsymbol{C}_i^{(t)} \in \mathbb{C}^{m\times n}$ are the diagonal block matrices of $\boldsymbol{A}^{(t)}$ and $\boldsymbol{C}^{(t)}$, respectively. The gradient descent in Step (A) and (B) and the projection onto the block diagonal constraint set in Step (C) are the same as Algorithm 1. The proposed precoder optimization algorithm under transmit power constraints per sensor is summarized in Algorithm 2.

4.2 Performance Evaluation

We evaluate estimation performance with Algorithm 2 on synthetic data. Most of the experimental settings are the same as in Sect. 3.2.1 and 3.2.2. The synthetic correlation model (16) was assumed. Transmit power of all sensors was set to the constant $P_i = P_0$. The conventional method of Huang et al. [11] cannot be applied to problem with the transmit power constraints per sensor so that it is removed from the following results.

Figure 7 shows MSE values (9) in terms of the number L of sensor nodes. The system parameters were set to $(n, m, r, a, P_0, \mathrm{SNR}) = (2, 2, 40, 0.85, 5, 30(\mathrm{dB}))$. The number of iterations and the step-size parameter of Algorithm 2 were $J = 50000$ and $\mu = 0.5$, respectively. The proposed method achieves the lowest MSE when $L = 2$ to 18. When $L = 20$, however, it shows the same MSE performance as the conventional method of Huh et al.

Figure 8 shows MSE ratio when the correlation parameter a varies as $a = 0, 0.1, 0.2, \ldots, 0.9$. The system parameters were set to $(n, m, r, L, P_0, \mathrm{SNR}) = (2, 2, 20, 20, 1, 20(\mathrm{dB}))$. The number of iterations and the step-size parameter of Algorithm 2 were $J = 5000$ and $\mu = 0.5$, respectively. From Fig. 8, the MSE ratio in terms of the method of Huh et al. is almost one for any value of a although it is never less than one. This means that there is only a slight performance difference between the proposed method and the method of Huh et al. when $L = 20$. By comparing Fig. 8 with Fig. 5, it can be seen that the proposed approach does not necessarily yield a large performance difference over the conventional method under the power constraints for each sensor.

The performance behavior of the proposed method in

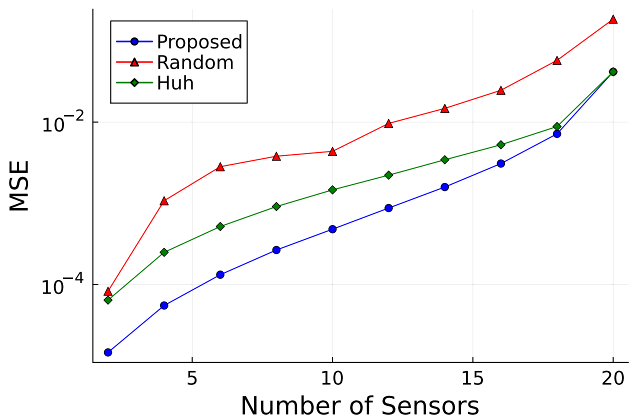


Fig. 7 MSE vs. the number of sensor nodes (Alg. 2), where $(n, m, r, a, P_0, \mathrm{SNR}) = (2, 2, 40, 0.85, 5, 30(\mathrm{dB}))$.

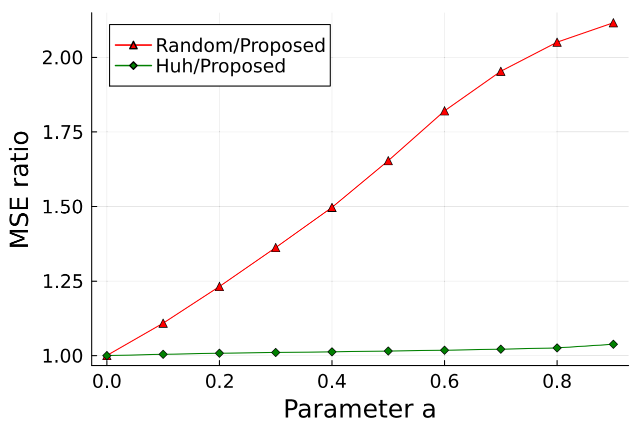


Fig. 8 MSE ratio vs. correlation parameter a (Alg. 2), where $(n, m, r, L, P_0, \mathrm{SNR}) = (2, 2, 20, 20, 1, 20(\mathrm{dB}))$.

Figs. 7 and 8 is different from the results in Sect. 3.2.2. This can be interpretable by the difference between Algorithms 1 and 2. The most important contribution of the proposed algorithms over the conventional methods is the incorporation of correlation, especially spatial correlation included in $\boldsymbol{K}_{ij}(i \neq j)$, into the algorithm design. In Algorithm 1, information on the matrices $\boldsymbol{K}_{ij}(i \neq j)$ is included in two steps: the gradient descent step (Step (B)) and the projection step (Step (D)) onto the transmit power constraint set. This is considered to be the reason for the large performance difference from the conventional methods. On the other hand, in Algorithm 2, it is included only in the gradient descent step. Therefore, the advantage of using the information of the matrices $\boldsymbol{K}_{ij}(i \neq j)$ is more prominent in Algorithm 1.

5. Precoder Optimization Algorithm for Over-the-air Computation

The same idea as the proposed algorithms can be applied to precoder design in the context of AirComp [13].

5.1 Derivation

Typical functions required in applications of sensor networks can be directly calculated by the sum of transmitted data from sensors [17]. To calculate such a function at the aggregator, it requires obtaining the sum $\boldsymbol{s} = \sum_{i=1}^{L} \boldsymbol{x}_i = \boldsymbol{Q}\boldsymbol{x} \in \mathbb{R}^n$ as correctly as possible, where $\boldsymbol{Q} = [\boldsymbol{I}, \ldots, \boldsymbol{I}] \in \mathbb{R}^{n \times nL}$. Therefore, the objective for AirComp is to design precoder and decoder that minimizes MSE in terms of the sum $\boldsymbol{s}$ instead of the data vector $\boldsymbol{x}$.

The received signal in the case of AirComp is also given by (2). We suppose the assumption on data correlation as in Sect. 2.3. In addition, we assume the use of an MMSE decoder for AirComp [18], that is, the estimate $\hat{\boldsymbol{s}}$ of the sum is given by

$$\hat{\boldsymbol{s}} = \boldsymbol{Q}\boldsymbol{K}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}} \left(\boldsymbol{H}\boldsymbol{A}\boldsymbol{K}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}} + \boldsymbol{S}\right)^{-1} \boldsymbol{y}. \tag{20}$$

The MSE in terms of the sum can be calculated as

$$\mathbb{E}[\|\hat{\boldsymbol{s}} - \boldsymbol{s}\|^2] = \mathrm{Tr}\left[\boldsymbol{Q}\left(\boldsymbol{K}^{-1} + \boldsymbol{A}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{S}^{-1}\boldsymbol{H}\boldsymbol{A}\right)^{-1}\boldsymbol{Q}^{\mathrm{T}}\right] \tag{21}$$

by using Woodbury matrix identity.

The block diagonal constraint $C0$ of the precoder is required in this case because the received signal model is the same as in Sect. 3. In addition, we consider the total transmit power constraint $C1$. Note that the transmit power constraint per sensor is also applicable. The precoder optimization problem for AirComp under the constraints $C0$ and $C1$ is then given by [18]

$$(\mathcal{P}3) \quad \min_{\boldsymbol{A}} \mathrm{Tr}\left[\boldsymbol{Q}\left(\boldsymbol{K}^{-1} + \boldsymbol{A}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{S}^{-1}\boldsymbol{H}\boldsymbol{A}\right)^{-1}\boldsymbol{Q}^{\mathrm{T}}\right]$$
$$\text{s.t. } \mathrm{E}[\|\boldsymbol{A}\boldsymbol{x}\|^2] = P,\ \boldsymbol{A} \in \mathbb{B}_L^{m \times n}. \tag{22}$$

The constraints are the same as for the previous problem $\mathcal{P}1'$ and the difference from $\mathcal{P}1'$ is only the cost function. Therefore, the algorithm for solving $\mathcal{P}3$ can be constructed in the same way as Algorithm 1 except for the gradient calculation step, Step (A). The gradient calculation step is formulated from the cost function of (22) as

$$\boldsymbol{G}^{(t-1)} = \nabla_{\boldsymbol{A}}(\mathrm{Tr}\,[\boldsymbol{T}(\boldsymbol{A})])|_{\boldsymbol{A}=\boldsymbol{A}^{(t-1)}}, \tag{23}$$

where $\boldsymbol{T}(\boldsymbol{A}) = \boldsymbol{Q}\left(\boldsymbol{K}^{-1} + \boldsymbol{A}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{S}^{-1}\boldsymbol{H}\boldsymbol{A}\right)^{-1}\boldsymbol{Q}^{\mathrm{T}}$. The algorithm is summarized in Algorithm 3.

5.2 Performance Evaluation

We evaluate detection performance of Algorithm 3 on synthetic data. Figure 9 shows MSE values of AirComp in terms of the number L of sensor nodes. The synthetic correlation model (16) was used and the system parameters were set to $(n, m, r, a, P, \mathrm{SNR}) = (2, 2, 20, 0.9, 20, 30)$. The number of iterations and the step-size parameter were $J = 50000$ and

Algorithm 3 Precoder Optimization Algorithm for AirComp

Input: $\boldsymbol{A}^{(0)} \in \mathbb{B}_L^{m\times n}$ satisfying transmit power constraint
Output: $\boldsymbol{A}^{(J)} \in \mathbb{B}_L^{m\times n}$
for $t = 1, 2, \ldots, J$ **do**
$\boldsymbol{G}^{(t-1)} = \nabla_{\boldsymbol{A}}(\mathrm{Tr}\,[\boldsymbol{T}(\boldsymbol{A})])|_{\boldsymbol{A}=\boldsymbol{A}^{(t-1)}}$
$\boldsymbol{B}^{(t)} = \boldsymbol{A}^{(t-1)} - \mu \boldsymbol{G}^{(t-1)}$
$\boldsymbol{C}^{(t)} = \boldsymbol{B}^{(t)} \odot \boldsymbol{M}$
$\boldsymbol{A}^{(t)} = \boldsymbol{C}^{(t)} \sqrt{\frac{P}{\mathrm{Tr}[\boldsymbol{C}^{(t)}\boldsymbol{K}(\boldsymbol{C}^{(t)})^{\mathrm{H}}]}}$
end for

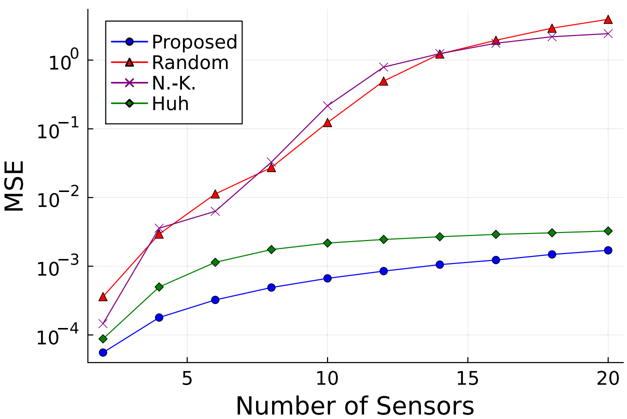


Fig. 9 MSE vs. the number of sensor nodes (Alg. 3), where $(n, m, r, a, P, \mathrm{SNR}) = (2, 2, 20, 0.9, 20, 30)$.

$\mu = 0.5$, respectively. We used three baseline algorithms for AirComp, a random precoding method ("Random"), the method of Nakai-Kasai and Wadayama [18] ("N.-K."), and the method of Huh et al. [12] ("Huh"). The method of Nakai-Kasai and Wadayama [18] is a non-iterative optimization method for the context of AirComp, which is similar to the method of Huang et al. [11]. This shows good performance when the system is overloaded. The proposed method achieves the lowest MSE among the methods. The proposed precoder optimization using correlation effectively performs also in the context of AirComp.

6. Conclusions

This paper has considered precoder design in wireless data aggregation using statistical data correlation that appeared in typical applications of sensor networks. We introduced covariance matrix of data into the precoder optimization problem and proposed novel algorithms on the basis of projected gradient descent method. The proposed methods achieve better performance than the conventional precoder optimization algorithms that ignore data correlation. We have also proposed the precoder optimization algorithm for the context of AirComp and it shows better performance. From these results, the proposed method enables superior precoder design by appropriately incorporating data correlation. Future work includes the combination of the proposed algorithm with federated learning and its performance evaluation.

Acknowledgments

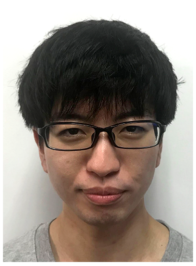
This work was supported by JSPS KAKENHI Grant-in-Aid for Young Scientists Grant Number JP23K13334 (to A. Nakai-Kasai) and for Scientific Research(A) Grant Number JP22H00514 (to T. Wadayama).

References

[1] L. Krishnamachari, D. Estrin, and S. Wicker, "The impact of data aggregation in wireless sensor networks," Proc. IEEE ICDCS Workshops, pp.575–578, 2002.
[2] L. Dan, C. Xin, H. Chongwei, and J. Liangliang, "Intelligent agriculture greenhouse environment monitoring system based on IoT technology," Proc. ICITBS, pp.487–490, Dec. 2015.
[3] T. Salam, W.U. Rahman, and X. Tao, "Data aggregation in massive machine type communication: Challenges and solutions," IEEE Access, vol.7, pp.41921–41946, April 2019.
[4] M.C. Vuran, Ö.B. Akan, and I.F. Akyildiz, "Spatio-temporal correlation: Theory and applications for wireless sensor networks," Computer Networks, vol.45, no.3, pp.245–259, 2004.
[5] A. Singh, F. Ramos, H.D. Whyte, and W.J. Kaiser, "Modeling and decision making in spatio-temporal processes for environmental surveillance," Proc. ICRA, pp.5490–5497, May 2010.
[6] D. Gu and H. Hu, "Spatial Gaussian process regression with mobile sensor networks," IEEE Trans. Neural Netw. Learn. Syst., vol.23, no.8, pp.1279–1290, Aug. 2012.
[7] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO networks: Spectral, energy, and hardware efficiency," Found. Trends Signal Process., vol.11, no.3–4, pp.154–655, 2017.
[8] G. Zhu and K. Huang, "MIMO over-the-air computation for high-mobility multi-modal sensing," IEEE Internet Things J., vol.6, no.4, pp.6089–6103, Aug. 2019.
[9] D. Wen, G. Zhu, and K. Huang, "Reduced-dimension design of MIMO AirComp for data aggregation in clustered IoT networks," Proc. IEEE GLOBECOM, Dec. 2019.
[10] V. Stankovic and M. Haardt, "Generalized design of multi-user MIMO precoding matrices," IEEE Trans. Wireless Commun., vol.7, no.3, pp.953–961, March 2008.
[11] J. Huang, E. Au, and V. Lau, "Linear precoder and equalizer design for uplink multiuser MIMO systems with imperfect channel state information," Proc. IEEE WCNC, pp.1295–1300, March 2007.
[12] M. Huh, D. Yu, and S.-H. Park, "Signal processing optimization for federated learning over multiuser MIMO uplink channel," Proc. ICOIN, pp.495–498, Jan. 2021.
[13] B. Nazer and M. Gastpar, "Computation over multiple-access channels," IEEE Trans. Inf. Theory, vol.53, no.10, pp.3498–3516, Oct. 2007.
[14] Intel Berkeley Research Lab Sensor Data, http://db.csail.mit.edu/labdata/labdata.html
[15] P. Kairouz, H.B. McMahan, B. Avent, A. Bellet, M. Bennis, A.N. Bhagoji, K. Bonawitz, Z. Charles, G. Cormode, R. Cummings, H. Eichner, S. El Rouayheb, D. Evans, J. Gardner, Z. Garrett, A. Gascón, B. Ghazi, P.B. Gibbons, M.O. Gruteser, Z. Harchaoui, C. He, L. He, Z. Huo, B. Hutchinson, J. Hsu, M. Jaggi, T. Javidi, G. Joshi, M. Khodak, J. Konecný, A. Korolova, F. Koushanfar, S. Koyejo, T. Lepoint, Y. Liu, P. Mittal, M. Mohri, R. Nock, A. Özgür, R. Pagh, H. Qi, D. Ramage, R. Raskar, M. Raykova, D. Song, W. Song, S.U. Stich, Z. Sun, A.T. Suresh, F.S. Tramèr P. Vepakomma, J. Wang, L. Xiong, Z. Xu, Q. Yang, F.X. Yu, H. Yu, and S. Zhao, "Advances and open problems in federated learning," Found. Trends Mach. Learn., vol.14, no.1–2, pp.1–210, June 2021.
[16] K.B. Petersen and M.S. Pedersen, "The matrix cookbook," Technical University of Denmark, Nov. 2012.

[17] M. Goldenbaum and S. Stanczak, “Robust analog function computation via wireless multiple-access channels,” IEEE Trans. Commun., vol.61, no.9, pp.3863–3877, Aug. 2013.
[18] A. Nakai-Kasai and T. Wadayama, “Precoder design for correlated data aggregation via over-the-air computation in sensor networks,” Proc. IEEE GLOBECOM Workshop, Dec. 2022.

Ayano Nakai-Kasai received the bachelor’s degree in engineering, the master’s degree in informatics, and Ph.D. degree in informatics from Kyoto University, Kyoto, Japan, in 2016, 2018, and 2021, respectively. She is currently an Assistant Professor at Graduate School of Engineering, Nagoya Institute of Technology. Her research interests include signal processing, wireless communication, and machine learning. She received the Young Researchers’ Award from the Institute of Electronics, Information and Communication Engineers in 2018 and APSIPA ASC 2019 Best Special Session Paper Nomination Award. She is a member of IEEE and IEICE.



Naoyuki Hayashi received the B.E. and the M.E. degrees from Nagoya Institute of Technology in 2020 and 2022, respectively. His research interest includes signal processing for wireless communications.

Tadashi Wadayama was born in Kyoto, Japan, on May 9, 1968. He received the B.E., the M.E., and the D.E. degrees from Kyoto Institute of Technology in 1991, 1993 and 1997, respectively. On 1995, he started to work with Faculty of Computer Science and System Engineering, Okayama Prefectural University as a research associate. From April 1999 to March 2000, he stayed in Institute of Experimental Mathematics, Essen University (Germany) as a visiting researcher. On 2004, he moved to Nagoya Institute of Technology as an associate professor. Since 2010, he has been a full professor of Nagoya Institute of Technology. His research interests are in coding theory, information theory, and signal processing for wireless communications. He is a member of IEEE and a senior member of IEICE.