IEICE TRANSACTIONS

on Communications

VOL. E104-B NO. 4
APRIL 2021

A PUBLICATION OF THE COMMUNICATIONS SOCIETY

The Institute of Electronics, Information and Communication Engineers
Kikai-Shinko-Kaikan Bldg., 5-8, Shibakoen 3chome, Minato-ku, TOKYO, 105-0011 JAPAN

PAPER

Optimization and Hole Interpolation of 2-D Sparse Arrays for Accurate Direction-of-Arrival Estimation*

Shogo NAKAMURA[†], **Sho IWAZAKI**[†], ***Student Members***, ***and*** **Koichi ICHIGE**[†a)], ***Member***

SUMMARY This paper presents a method to optimize 2-D sparse array configurations along with a technique to interpolate holes to accurately estimate the direction of arrival (DOA). Conventional 2-D sparse arrays are often defined using a closed-form representation and have the property that they can create hole-free difference co-arrays that can estimate DOAs of incident signals that outnumber the physical elements. However, this property restricts the array configuration to a limited structure and results in a significant mutual coupling effect between consecutive sensors. In this paper, we introduce an optimization-based method for designing 2-D sparse arrays that enhances flexibility of array configuration as well as DOA estimation accuracy. We also propose a method to interpolate holes in 2-D co-arrays by nuclear norm minimization (NNM) that permits holes and to extend array aperture to further enhance DOA estimation accuracy. The performance of the proposed optimum arrays is evaluated through numerical examples.

key words: *direction of arrival estimation, array signal processing, sparse array, mutual coupling, nuclear norm minimization*

1. Introduction

Antenna arrays are often employed as one of key devices in radar or communication applications [1], [2]. To accurately detect the incident directions of array input signals, direction of arrival (DOA) estimation is effective, and many DOA estimation algorithms have been proposed such as multiple signal classification (MUSIC) and estimation of signal parameters via rotational invariance techniques (ESPRIT) [3]–[6]. Most of those methods are based on the eigenvalue decomposition of a sample covariance matrix of an array input signal, which means that those algorithms are with the degree of freedom (DOF) of $O(N)$, where N denotes the number of antenna elements. Many methods have been developed to enhance the DOF like minimum redundancy array (MRA) [7], a method using 4th-order cumulants [8], or extended beamforming [9]–[11]. However, these approaches are very complicated.

2-D sparse arrays such as billboard arrays and 2-D nested arrays have already been studied that can accurately estimate 2-D DOAs (azimuth and elevation) [12], [13]. The arrays utilize Khatri-Rao product [14] to increase the DOF [15] and create virtual arrays, which are called difference co-arrays as they are composed of virtual sensors. The difference co-arrays possess the DOF to $O(N^2)$ that can resolve up to $O(N^2)$ uncorrelated signals. Thus, the co-arrays have no holes in the apertures, which means that they can form a complete uniform rectangular array (URA). This is called "hole-free" property.

One inherent problem of those arrays is that their configurations often trigger severe mutual coupling, resulting in considerable interference between sensor outputs [16], [17] and degrading DOA estimation performances. The hourglass array [18] can reduce the mutual coupling influence and creates a hole-free difference co-array. Large sensor separations help to reduce the mutual coupling effect, and the sensors are located more sparsely in the hourglass array than in the other 2-D sparse arrays such as open box arrays (OBAs) or billboard arrays [19]. However, the closed-form representation in the hourglass array also becomes a restriction and prevents further improvement of DOA estimation performance.

Recall that the array configuration can no longer be a closed-form representation or be further modified by any optimization, as demonstrated when one of the authors applied simulated annealing (SA) to optimize the configuration of a 1-D sparse array [8]. It is assumed that the hole-free property that restricts the array configuration and the DOA estimation performance can be further improved by introducing nuclear norm minimization (NNM) [20] so that we can interpolate hole elements. NNM can interpolate lacking matrix elements while preserving the low-rank property of the matrix and can be applied to estimate the received signal components by virtual elements. The optimum arrays are more robust to mutual coupling due of permitting holes in the aperture.

In this paper, we develop a method of constructing 2-D sparse arrays using optimization and hole interpolation for accurate DOA estimation. The array configuration is first optimized while fixing the array aperture [21]. We also introduce a way to interpolate holes in 2-D difference co-arrays by using NNM [22] to develop arrays with larger intervals and smaller mutual couplings. Then, we propose an extended optimum array whose aperture is extended by using the modified SA algorithm. Array elements can be located more sparsely by extending the array aperture. Also, the optimum extended array aperture is expected to increase DOFs and DOA estimation accuracy. The interpolation accuracy and DOA estimation accuracy of the proposed array configuration are evaluated through numerical examples.

Manuscript received March 10, 2020.
Manuscript revised July 31, 2020.
Manuscript publicized October 21, 2020.

[†]The authors are with Department of Electrical and Computer Engineering, Yokohama National University, Yokohama-shi, 240-8501 Japan.

*Preliminary versions of this paper are presented in [21], [22].

a) E-mail: koichi@ynu.ac.jp

DOI: 10.1587/transcom.2020EBP3035

2. Preliminaries

In this section, we review the data model and the conventional array geometry.

2.1 2-D Sparse Array Data Model

Suppose that D uncorrelated sources impinge on a 2-D sensor array in an additive white Gaussian noise (AWGN) environment, where the signals and noises are statistically independent. The array aperture and the sensor location are respectively given by $N_x \times N_y$ and $\boldsymbol{n}d$, where $\boldsymbol{n} = (n_x, n_y) \in \mathbb{Z}^2$ is an integer-valued vector, $d = \lambda/2$ is the minimum separation between sensors, and λ is the wavelength of incoming sources. Assume that the sensor location $\boldsymbol{n}$ forms a set $\mathbb{S}$. Then the sensor input $\boldsymbol{x}_{\mathbb{S}}$ on $\mathbb{S}$ can be modeled similarly to in the hourglass array [18] as

$$\boldsymbol{x}_{\mathbb{S}} = \sum_{i=1}^{D} A_i \boldsymbol{C} \boldsymbol{v}_{\mathbb{S}}(\bar{\theta}_i, \bar{\phi}_i) + \boldsymbol{u}_{\mathbb{S}}, \tag{1}$$

$$\bar{\theta}_i = \frac{d}{\lambda} \sin\theta_i \cos\phi_i, \tag{2}$$

$$\bar{\phi}_i = \frac{d}{\lambda} \sin\theta_i \sin\phi_i, \tag{3}$$

where the i-th source is with the complex amplitude $A_i \in \mathbb{C}$, the azimuth $\phi_i \in [0, 2\pi]$, and elevation $\theta_i \in [0, \pi/2]$. The element of the steering vector $\boldsymbol{v}_{\mathbb{S}}(\bar{\theta}_i, \bar{\phi}_i)$ corresponding to the sensor at $\boldsymbol{n} = (n_x, n_y)$ is given by $e^{j2\pi(\bar{\theta}_i n_x + \bar{\phi}_i n_y)}$. The mutual coupling matrix $\boldsymbol{C}$ [17] is characterized by its entries

$$\langle \boldsymbol{C} \rangle_{\boldsymbol{n}_1, \boldsymbol{n}_2} = \begin{cases} c(\|\boldsymbol{n}_1 - \boldsymbol{n}_2\|_2), & \|\boldsymbol{n}_1 - \boldsymbol{n}_2\|_2 \le B, \\ 0, & \text{otherwise}, \end{cases} \tag{4}$$

where $\boldsymbol{n}_1, \boldsymbol{n}_2 \in \mathbb{S}$ denote the sensor location, B is the maximum sensor separation where the mutual coupling effect exists, and $c(\cdot)$ is the mutual coupling coefficient given by $c(0) = 1$ and $|c(k)/c(\ell)| = \ell/k$ for $k, \ell > 0$ [18]. Here, the covariance matrix $\boldsymbol{R}_{\mathbb{S}}$ of the array $\mathbb{S}$ can be expressed as

$$\boldsymbol{R}_{\mathbb{S}} = E\left[\boldsymbol{x}_{\mathbb{S}} \boldsymbol{x}_{\mathbb{S}}^H\right]. \tag{5}$$

We also define the difference co-array $\mathbb{D} = \{\boldsymbol{n}_1 - \boldsymbol{n}_2 \mid \boldsymbol{n}_1, \boldsymbol{n}_2 \in \mathbb{S}\}$ for any 2-D sparse array $\mathbb{S}$. Then the input vector of the difference co-array $\boldsymbol{x}_{\mathbb{D}}$ can be obtained by vectorizing (5) while removing duplicated entries [18]. Note that the higher DOF of the difference co-array $\mathbb{D}$ enables us to identify $O(N^2)$ uncorrelated signals.

2.2 Conventional 2-D Sparse Arrays

The hourglass array [18] was developed on the basis of the OBA [19] to reduce the mutual coupling effect. The configurations of the OBA and the hourglass array are respectively illustrated in Fig. 1(a) and (b), where the red dots represent physical sensors with the minimum sensor separation $d = \lambda/2$. The sensor location of the hourglass array is defined in a closed-form expression and the difference co-array becomes a hole-free URA [18]. We also see from Fig. 1 that the hourglass array has fewer elements with the smallest separation than the OBA. That means, the hourglass array can reduce the mutual coupling effect more efficiently because of larger sensor separations.

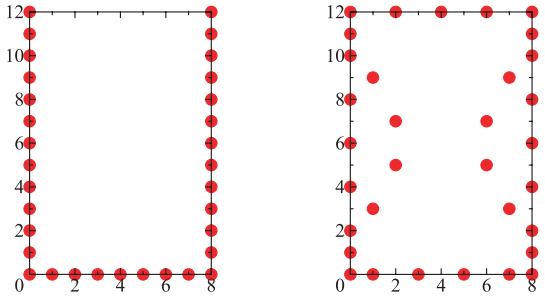


Fig. 1 Examples of conventional 2-D sparse arrays with 33 physical sensors. Apertures for both difference co-arrays are 17×25.

3. Configuration of the Proposed 2-D Sparse Arrays

In this section, we introduce SA-based optimum 2-D sparse arrays to further reduce the mutual coupling effect. First, we develop an optimization method to develop hole-free sparse arrays. However, the hole-free condition sometimes makes it difficult to distribute sensors more sparsely even though a sparse distribution is often necessary for accurate DOA estimation. Therefore, we also develop a method to allow holes in the difference co-array and then interpolate them. Then the interpolated hole-free virtual array can be applied to the DOA estimation algorithms that require continuous arrays such as ESPRIT.

We first develop a way to optimize the array configuration while given array apertures are fixed as given initial arrays in Sect. 3.1. Then, we introduce a method to interpolate holes for a 2-D co-array via NNM in Sect. 3.2. Finally, we propose a method to extend array apertures in Sect. 3.3.

3.1 Optimization of Array Configuration

In a previous study [8], we developed a 1-D sparse array that has a large aperture on the basis of the optimization by SA. In this work, we apply a similar approach to construct arrays that are less susceptible to the mutual coupling. Many optimization methods can be applied to develop array configuration optimization, such as genetic algorithm (GA), particle swarm optimization (PSO), and ant colony optimization (ACO). We evaluated those algorithms and found that the SA-based optimization has a very high probability of converging to a global solution and is quite simple to apply to array construction [8]. Because small sensor separations result in a large mutual coupling effect, we define the cost function to be minimized as

$$\mathcal{D}_1 = \sum_{i \neq j}^{N} \frac{1}{\|\boldsymbol{n}_i - \boldsymbol{n}_j\|_2}, \quad \text{such that} \quad \|\boldsymbol{n}_i - \boldsymbol{n}_j\|_2 \le B \tag{6}$$

where $\boldsymbol{n}_i, \boldsymbol{n}_j \in \mathbb{S}$ and N denotes the number of sensors, and B is shown in (4). Besides, $\|\cdot\|_2$ denotes ℓ_2 norm of vectors, and therefore $\|\boldsymbol{n}_i - \boldsymbol{n}_j\|_2$ means the distance between the i-th and j-th array elements. Equation (6) means the sum of a reciprocal of distances between two elements, and the function's behavior depends on the distance. In this optimization, the sensor locations are changed within a given aperture so as to minimize the cost function. The optimum array can be obtained by minimizing the cost function because a smaller value of $\mathcal{D}_1$ leads to sparser arrays.

The whole optimization algorithm is shown in Algorithm 1. The initial array $\mathbb{S}_0$ is the hourglass array, and the parameters related to temperature $(T, T_{min}, \Delta T)$ are set empirically. Note that the parameters $T, T_{min}, \Delta T$ are the temperature parameters used in SA; T denotes the temperature, and $\Delta T (< 1)$ is the temperature ratio in each optimization step. More specifically, the temperature goes down by $T \times \Delta T \to T$ and stops iteration when $T < T_{min}$. The number of holes for the final array $\mathbb{S}_{\text{final}}$, which is a solution, can be controlled by changing the parameter *pnum* (the permitted number of holes). For example, an array as a solution has fewer than 100 holes when *pnum* = 100.

The following steps after setting these parameters are iterated until the temperature becomes sufficiently cool ($T > T_{min}$). In the annealing iteration, the first step is to calculate the value of the initial cost function $\mathcal{D}_{1,\text{ini}}$ on the basis of (6) and the number of holes (*num*) for array $\mathbb{S}$. Then, a sensor randomly selected from $\mathbb{S}_{\text{tmp}}$ is moved in the aperture until the condition *num* > *pnum* is met, where four sensors that determine the aperture are fixed at corners. Note that if the condition is not met, the picked sensor is returned to the original location and the operation is repeated. An annealing test is conducted after sensor movement. Here two conditions are assumed for adopting new arrays: the value of the cost function for $\mathbb{S}_{\text{tmp}}$ is smaller than that of $\mathbb{S}$ ($\mathcal{D}_{1,\text{tmp}} < \mathcal{D}_{1,\text{ini}}$) or the value calculated by $\exp((\mathcal{D}_{1,\text{ini}} - \mathcal{D}_{1,\text{tmp}})/T) > x$ is larger than $x \in [0, 1]$, where x is a random value. The latter is characteristic of SA, a new array is often adopted when T is large, and the probability of adoption gradually decreases as T decreases. By controlling the possibility of adopting a new array configuration in this manner, we can often reach a global-optimum solution without converging to local-optimums.

The annealing test helps to avoid convergence to the local minimum. The final step is to cool the temperature as the equation: $T \Leftarrow T \times \Delta T$. Through these steps, an optimum 2-D sparse array $\mathbb{S}_{\text{final}}$ that is more robust to mutual coupling can be obtained, where the value of the cost function of $\mathbb{S}_{\text{final}}$ is equal to $\mathcal{D}_{1,\text{ini}}$.

Algorithm 1 Optimization based on SA

```
set initial array S_ini as hourglass array, initial temperature T, end temperature T_min, changing rate of temperature ΔT, permitted # of holes (pnum)
S ⇐ S_ini
calculate value of cost function D_1,ini for S_0 by (6)
while T > T_min do
    calculate # of holes (num) for S
    while num > pnum do
        S_tmp ⇐ S
        move a sensor randomly rather than 4 sensors at corners in S_tmp
    end while
    calculate value of cost function D_1,tmp for S_tmp by (6)
    set a random value x ∈ [0, 1]
    if D_1,tmp < D_1,ini then
        S ⇐ S_tmp
        D_1,ini ⇐ D_1,tmp
    else if exp((D_1,ini − D_1,tmp)/T) > x then
        S ⇐ S_tmp
        D_1,ini ⇐ D_1,tmp
    end if
    T ⇐ T × ΔT
end while
S_final ⇐ S
```

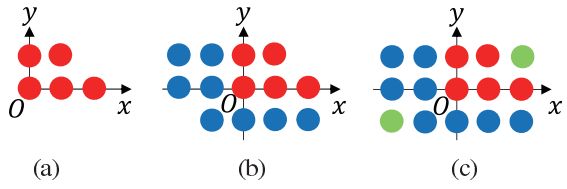


Fig. 2 Example 2-D sparse array in case of $N_x = 3$, $N_y = 2$, with the (a) physical elements $\mathbb{S}$ (red), (b) difference co-array $\mathbb{D}$ (red + blue), and (c) interpolated co-array $\mathbb{V}$ (red + blue + green).

3.2 2-D Co-Array Interpolation via NNM

The holes in optimum arrays in Sect. 3.1 must be interpolated to accurately estimate DOA. An interpolation method for 1-D sparse arrays has been proposed [20] that utilizes the property that a correlation matrix is a low rank and the best relaxation problem for low rank optimization: NNM. We extend this method to 2-D cases for accurate interpolation.

In the 1-D sparse array interpolation method [20], the correlation matrix involving hole elements for NNM is designed by assuming the matrix is Hermitian Toeplitz when the array is linear, that is, how to interpolate holes can be interpreted as how to design the matrix. We will present several methods on how to interpolate holes, in other words, how to design a matrix for NNM in the 2-D case.

Figure 2 shows an example of 2-D sparse array whose aperture is $N_x = 3, N_y = 2$: (a) the physical array $\mathbb{S}$, (b) the difference co-array $\mathbb{D}$, and (c) the interpolated difference co-array $\mathbb{V}$. Red and blue circles represent physical and virtual sensors, respectively, while green circles denote holes to be interpolated. The origin O is the bottom-left corner in Fig. 2(a), and the sensor output corresponding to the sensor at $\boldsymbol{n}_1 - \boldsymbol{n}_2$ ($\boldsymbol{n}_1, \boldsymbol{n}_2 \in \mathbb{S}$) is expressed as $\langle \tilde{\boldsymbol{x}}_{\mathbb{D}} \rangle_{\boldsymbol{n}_1, \boldsymbol{n}_2}$, which are known virtual array outputs.

The form of NNM in a 2-D case is modeled as

$$\tilde{\boldsymbol{R}}_{\mathbb{V}}^{\star} = \arg\min_{\tilde{\boldsymbol{R}}_{\mathbb{V}}} \|\tilde{\boldsymbol{R}}_{\mathbb{V}}\|_* \tag{7}$$

where

$$\tilde{\boldsymbol{R}}_{\mathbb{V}} = \tilde{\boldsymbol{R}}_{\mathbb{V}}^{H}, \tag{8}$$

$$\langle \tilde{\boldsymbol{R}}_{\mathbb{V}} \rangle_{\boldsymbol{n}_1, \boldsymbol{n}_2} = \langle \tilde{\boldsymbol{x}}_{\mathbb{D}} \rangle_{\boldsymbol{n}_1, \boldsymbol{n}_2}. \tag{9}$$

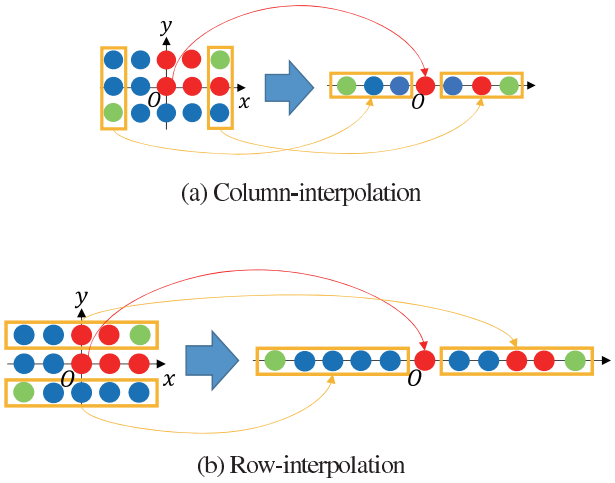


Fig. 3 Basic concept to apply 1-D NNM to 2-D case. The two sensor outputs have complex conjugate values when they are odd-symmetrically located.

The optimal solution $\tilde{\boldsymbol{R}}_{\mathbb{V}}^{\star}$ in (7) denotes the interpolated autocorrelation matrix. The method to construct the initial matrix $\tilde{\boldsymbol{R}}_{\mathbb{V}}$ with holes is divided into several ways by utilizing the co-array property shown in (10), and we will describe them.

3.2.1 Column/Row Interpolation

The basic idea of this interpolation is to apply 1-D NNM [20] to a 2-D case directly. In a 1-D case, the matrix $\tilde{\boldsymbol{R}}_{\mathbb{V}}$ in (7) is Hermitian Toeplitz and can be designed by using the following property. For 2-D difference co-arrays, an odd-symmetrically located sensor pair for the origin O has complex conjugate signals as shown in (10), e.g., the signal on $(1, 1)$ is a complex conjugate of the signal on $(-1, -1)$. That means the equality $x_{1,1} = x^*_{-1,-1}$ holds, where x_{n_x,n_y} denotes sensor output on (n_x, n_y):

$$\begin{aligned} x_{n_x,n_y} &= \exp\left\{an_x + bn_y\right\} \\ &= \overline{\exp\left\{a(-n_x) + b(-n_y))\right\}} \\ &= x^*_{-n_x,-n_y}, \end{aligned} \tag{10}$$

where a, b are appropriate complex values. Figure 3(a) and (b) show the basic concept to apply 1-D NNM to a 2-D case. Designing a Hermitian Toeplitz matrix $\tilde{\boldsymbol{R}}_{\mathbb{V}}$ by using 2-D array outputs corresponds to abstracting a linear array as a correlation matrix of a linear array is Hermitian Toeplitz. How to abstract sensors is not determined uniquely. Figure 3(a) shows column interpolation, which utilizes a part located as a column where holes exist. Similarly, we also propose row interpolation that uses sensor output located as in Fig. 3(b). The sensors in yellow rectangles are odd-symmetrically located, hence, they have complex conjugate signals. Note that the origin output has real value so that it is lined up at a principal diagonal in the matrix $\tilde{R}_{\mathbb{V}}$ for any interpolation method.

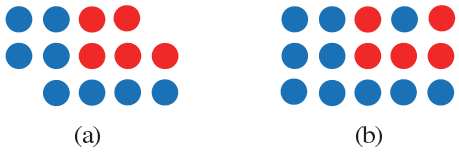


Fig. 4 Example 2-D sparse array in case of $N_x = 3$, $N_y = 2$, red and blue circles represent physical and virtual sensors, respectively. (a) Array with holes and (b) reference array (OBA) that has same aperture as (a) and hole-free property.

3.2.2 2-D Interpolation

The correlation matrix for NNM in Sect. 3.2.1 must be a Hermitian Toepliz and is for linear arrays. That is, it is a method to abstract linear arrays from 2-D arrays. Hence, NNM has to be performed at each part where there are holes, and this incurs calculation cost. In this subsection, we propose a method to construct a correlation matrix of 2-D spare arrays with holes in which we prepare the matrix for NNM only once. A hole-free reference array $\mathbb{S}_{\text{ref}}$ is prepared in advance, and the index information is used to realize this method. For example, we consider the array with holes shown in Fig. 4(a) and the reference array with hole-free property shown in Fig. 4(b). These two arrays must have the same aperture (but not necessarily the same number of sensors), which is 3×2 for a physical array in this example. The correlation matrices $R_{\mathbb{S}}$ and $R_{\mathbb{S}_{\text{OBA}}}$, which respectively correspond to Figs. 4(a) and 4(b), are given by

$$R_{\mathbb{S}_{\text{OBA}}} = \begin{bmatrix} x_{0,0} & x_{0,-1} & x_{-1,0} & x_{-2,0} & x_{-2,-1} \\ x_{0,1} & x_{0,0} & x_{-1,1} & x_{-2,1} & x_{-2,0} \\ x_{1,0} & x_{1,-1} & x_{0,0} & x_{-1,0} & x_{-1,-1} \\ x_{2,0} & x_{2,-1} & x_{-1,0} & x_{0,0} & x_{0,-1} \\ x_{2,1} & x_{2,0} & x_{1,1} & x_{0,1} & x_{0,0} \end{bmatrix}, \tag{11}$$

$$R_{\mathbb{S}} = \begin{bmatrix} x_{0,0} & x_{0,-1} & x_{-1,0} & x_{-2,0} & x_{-2,0} \\ x_{0,1} & x_{0,0} & x_{-1,1} & x_{-1,0} & x_{-2,1} \\ x_{1,0} & x_{1,-1} & x_{0,0} & x_{0,-1} & x_{-1,0} \\ x_{1,1} & x_{1,0} & x_{0,1} & x_{0,0} & x_{-1,1} \\ x_{2,0} & x_{2,-1} & x_{1,0} & x_{1,-1} & x_{0,0} \end{bmatrix}, \tag{12}$$

Comparing (11) and (12), the signal information of difference co-array $x_{-2,-1}, x_{2,1}$ in (11) is lacking. Thus, the signals in (12) are sorted the same as in (11) while the holes $x_{-2,1}, x_{2,1}$ are set to 0. NNM can be applied to this correlation matrix with holes $\tilde{\boldsymbol{R}}_{\mathbb{V}}$ directly, and NNM is operated only once. There must be a reference array $\mathbb{S}_{\text{ref}}$ with any aperture such as OBA, and a correlation matrix for NNM can be constructed with a size of $|\mathbb{S}_{\text{ref}}| \times |\mathbb{S}_{\text{ref}}|$. It includes all information of difference co-array's signals rather than information of holes, so more accurate interpolation can be expected. In this paper, we use OBA as a reference array. The interpolation accuracy does not depend on which arrays are used as a reference array as we only use the index information of the reference array.

3.3 Extension of Array Aperture

In this subsection, we propose an extended optimum array that has a larger aperture than the initial array. For the SA algorithm referred to in Sect. 3.1, the aperture determined by corner sensors is fixed as the initial condition. On the other hand, the modified SA algorithm can make a larger array aperture and element intervals, which will lead to larger DOFs and more accurate DOA estimation. Here we fixed the number of physical elements, i.e., the target of this work is to enhance DOA estimation performance by optimizing the array configuration under the condition of the same number of physical elements. The flow of the modified SA algorithm is shown in Algorithm 2 and its details are as follows.

The initial parameters used in Algorithm 2 are the same as in Algorithm 1; the initial array location, temperature parameters, the number of holes of difference co-array, and the permitted number of holes are defined as $\mathbb{S}_0, T, T_{min}, \Delta T, num$, and $pnum$, respectively. The cost function is defined as

$$\mathcal{D}_2 = \frac{1}{2}\mathcal{D}_1(1+r), \tag{13}$$

where $\mathcal{D}_1$ is calculated by (6), and r means the DOF ratio modeled as

$$r = \frac{\mathrm{DOF}}{(2N_x-1)(2N_y-1)}. \tag{14}$$

In (14), the numerator DOF is the number of sensors of the difference co-array and N_x, N_y is the aperture. The cost $\mathcal{D}_2$ in (13) matches $\mathcal{D}_1$ in (6) when r is equal to 1, in the case the difference co-array is a complete URA whose aperture is $(2N_x-1)(2N_y-1)$. The reason we define the cost function as this form is that a large number of holes makes the performance worse and the renewed cost function leads to sparser arrays and fewer holes (large DOF). It is suitable as the modified SA because there must be holes when extending the aperture.

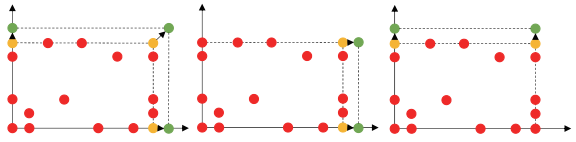
The aperture is extended as seen in Fig. 5 if the condition $num \le pnum$ is satisfied. We introduce three ways to extend the aperture: modes 1, 2, and 3. In mode 1, the sensor located at the origin $(0,0)$ is fixed throughout the steps, and sensors located at $(0,N_y), (N_x,0), (N_x,N_y)$ are moved to $(0,N_y+1), (N_x+1,0), (N_x+1,N_y+1)$ respectively. For modes 2 and 3, the apertures are extended as in Fig. 5(b) and (c). In this case, they are enlarged one at a time in only the x or y direction. Here, the sensors defining the aperture are picked up in only this step, otherwise they are fixed.

In the next step, a sensor is chosen rather than all four sensors at corners, is moved to a space in the aperture, and then calculates the cost function (13). If either $\mathcal{D}_{2,\mathrm{tmp}} < \mathcal{D}_{2,\mathrm{ini}}$ or $\exp((\mathcal{D}_{2,\mathrm{ini}} - \mathcal{D}_{2,\mathrm{tmp}})/T) > x$ is satisfied, a new array is adopted. The temperature T is reduced and if T does not meet the termination $T < T_{\min}$, these steps are iterated. Otherwise it stops. The final solution $\mathbb{S}_{\mathrm{final}}$ is an extended optimum array that has a larger aperture than and the same number of sensors as the initial array $\mathbb{S}_{\mathrm{ini}}$.

Algorithm 2 Modified SA algorithm for extending aperture

set initial array $\mathbb{S}_{\mathrm{ini}}$ as hourglass array, initial temperature T, end temperature T_{min}, changing rate of temperature ΔT, permitted # of holes ($pnum$)

$\mathbb{S} \Leftarrow \mathbb{S}_{\mathrm{ini}}$
calculate value of cost function $\mathcal{D}_{2,\mathrm{ini}}$ by (13)
while $T > T_{\min}$ **do**
  calculate # of holes (num) for $\mathbb{S}$
  if $num \le pnum$ **then**
    extend the aperture: the extended temporary array is $\mathbb{S}_{\mathrm{tmp}}$
  end if
  move a sensor randomly rather than 4 sensors at corners in $\mathbb{S}_{tmp}$
  calculate value of cost function $\mathcal{D}_{2,\mathrm{tmp}}$ for $\mathbb{S}_{\mathrm{tmp}}$
  set a random value $x \in [0,1]$
  if $\mathcal{D}_{2,\mathrm{tmp}} < \mathcal{D}_{2,\mathrm{ini}}$ **then**
    $\mathbb{S} \Leftarrow \mathbb{S}_{tmp}$
    $\mathcal{D}_{2,\mathrm{ini}} \Leftarrow \mathcal{D}_{2,\mathrm{tmp}}$
  else if $\exp((\mathcal{D}_{2,\mathrm{ini}} - \mathcal{D}_{2,\mathrm{tmp}})/T) > x$ **then**
    $\mathbb{S} \Leftarrow \mathbb{S}_{\mathrm{tmp}}$
    $\mathcal{D}_{2,\mathrm{ini}} \Leftarrow \mathcal{D}_{2,\mathrm{tmp}}$
  end if
  $T \Leftarrow T \times \Delta T$
end while
$\mathbb{S}_{\mathrm{final}} \Leftarrow \mathbb{S}$



(a) Mode 1 (b) Mode 2 (c) Mode 3

Fig. 5 Example way to extend array aperture. Sensors located at corners without the sensor at the origin $(0,0)$ are moved as shown in this figure.

4. Numerical Results

In this section, we evaluate the DOA estimation accuracy in the presence of mutual coupling for hourglass arrays [18], the optimum arrays developed in Sect. 3.1, and the extended optimum arrays developed in Sect. 3.3.

After showing specifications of simulation in Sect. 4.1, we first evaluate the performance of extended optimum arrays in Sect. 4.2. We also evaluate the values of cost function $\mathcal{D}_2$ in (13) as a function of $pnum$ shown in Sect. 3.3. By changing the value of $pnum$, the arrays with different apertures or DOFs that have the smallest value of $\mathcal{D}_2$ can be obtained. Accordingly, in Sect. 4.3, we evaluate the performances of different interpolation approaches: column interpolation, row interpolation, and 2-D interpolation for the arrays. Finally, the arrays are evaluated through root mean square error (RMSE) as a function of signal-to-noise ratio SNR and the number of snapshots in Sect. 4.4.

4.1 Simulation Specifications

Tables 1 and 2 show the simulation parameters used in the following subsections. In Sect. 4.2, arrays are constructed

Table 1 Simulation specifications in Sect. 4.2.

minimum element interval, d	0.5λ
initial array	hourglass array
initial array size, (N_x, N_y)	(26, 13)
# of physical sensors, N	50
initial temperature, T	10^5
end temperature, T_{min}	10^{-5}
changing rate of temperature, ΔT	0.999
permitted # of holes, *pnum*	0 to 5,000
# of trials, M	1,000

Table 2 Simulation specifications in Sects. 4.3 and 4.4.

Section	4.3	4.4
# of incident signals, D	16 uncorrelated sources	
DOAs	as in Fig. 6	
SNR (in dB)	0	−25 to 5 / 0
# of snapshots, K	200	200 / 20 to 1,000
coupling of elements, $c(1)$	0.3	
coupling effected interval, B	5	
DOA estimation method	2D unitary ESPRIT	
hole interpolation method	column/row/2-D	2-D
# of trials M	1,000	

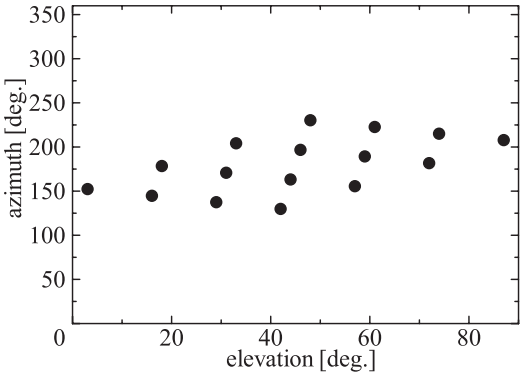


Fig. 6 Distribution of incident signals.

on different conditions: parameter *pnum* is changed per 100 from 0 to 5,000 at each mode. The initial array is an hourglass array whose aperture and the number of sensors are 26×13 and 50, respectively. Parameters related to temperature in SA are shown in Table 1. Then, a comparison of interpolation methods, SNR-RMSE, and Snapshots-RMSE characteristics are simulated under the conditions in Table 2. We assume 16 uncorrelated, narrowband and equal-power sources whose DOAs are shown in Fig. 6. SNR is defined as the ratio of the source signal power σ_s^2 to the noise power σ_n^2 in decibels, i.e., $\mathrm{SNR\,[dB]} = 10\log(\sigma_s^2/\sigma_n^2)$. The mutual coupling model is given by (4), where $c(1) = 0.3$, $B = 5$ and $c(\ell) = \exp[j\pi(\ell - 1)/4]/\ell$. The parameter $c(1)$ means the mutual coupling magnitude between two nearest sensors, where the distance between the nearest sensors is $\lambda/2$.

Note that mutual coupling exists in the measurement but the 2D unitary ESPRIT method does not deal with the mutual coupling effect. Therefore, the baseline performance of DOA estimation can be evaluated. The RMSE of the DOA estimation for azimuth ϕ and elevation θ are respectively evaluated by

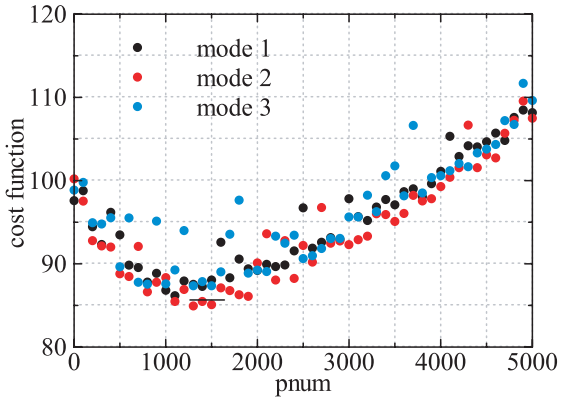


Fig. 7 Cost function in (13) as a function of *pnum*.

$$\mathrm{RMSE}(\phi)[\mathrm{deg.}] = \sqrt{\frac{1}{DM}\sum_{i=1}^{D}\sum_{m=1}^{M}(\hat{\phi}_i - \phi_i)^2}, \tag{15}$$

$$\mathrm{RMSE}(\theta)[\mathrm{deg.}] = \sqrt{\frac{1}{DM}\sum_{i=1}^{D}\sum_{m=1}^{M}(\hat{\theta}_i - \theta_i)^2}, \tag{16}$$

where M and D stand for the numbers of Monte-Carlo trials and received signals, respectively.

4.2 Examination of Extended Optimum Arrays

First, we start by determining which extended optimum arrays have better performance. They are evaluated by the values of the cost function in (13) and RMSE as a function of *pnum*. The results are shown in Fig. 7. The graphs for each mode have minimum values around *pnum* = 1,000 to 1,500 although they are not completely smooth. At each *pnum*, the cost function is minimized to optimize the array configuration. However, the configuration as well as the value of the cost function depends on the permitted number of holes in the optimization. Considering (13) including distance and DOF term, the array tends to be more balanced when the cost function becomes smaller. That is, the array with the smallest cost function value should be better in the presence of mutual coupling. Hence, we pick out arrays with such values at each mode and compare them with conventional methods. Figure 8 shows the tested array configurations: hourglass array, hole-free optimum array, optimum array with holes, mode-1 optimum array, mode-2 optimum array, and mode-3 optimum array.

4.3 Comparison of Interpolation Methods

In this subsection, we evaluate which methods in Sects. 3.2.1 and 3.2.2 can interpolate holes more accurately. The methods to be compared include column, row, and 2-D interpolation. RMSEs are computed under the following conditions: SNR= 0dB, the number of snapshots = 200,

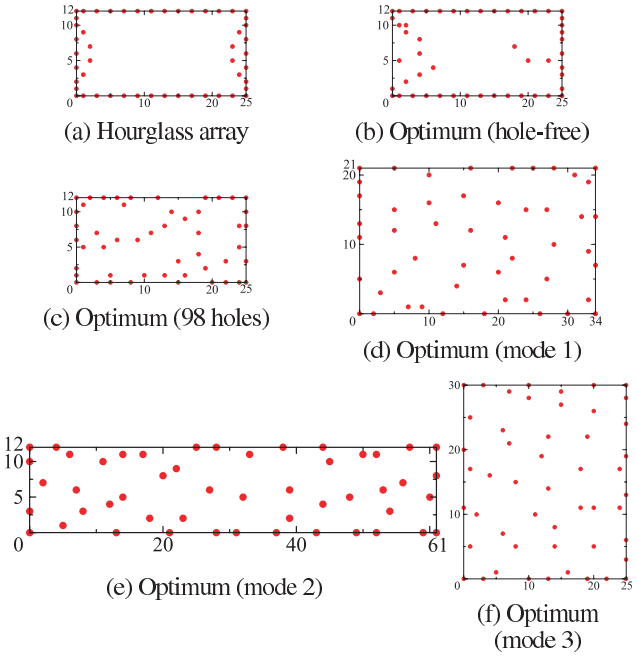


Fig. 8 Arrays' configurations for simulation.

Table 3 Comparison of array aperture and the value of the cost function.

kind of mode	aperture	value of cost function
Hourglass array	26×13	120.11
Optimum (hole-free)	26×13	114.27
Optimum (98 holes)	26×13	104.50
Mode 1	35×22	86.164
Mode 2	62×13	84.950
Mode 3	26×31	87.356

$c(1) = 0.3$, and $B = 5$. All the arrays have the smallest possible cost function value at each mode, and their aperture and the values of cost functions are shown in Table 3. Note that the values of *pnum* in modes 1, 2, and 3 are $1,100, 1,300$, and $1,300$, respectively. We see from Table 3 that the optimum arrays with larger apertures can make values of the cost function smaller, and mode 2 has the smallest value.

Figure 9 shows the results for azimuth and elevation angles. The RMSEs are the smallest for any arrays interpolated by 2-D interpolation. Therefore, we conclude that the 2-D interpolation method is more optimal than other approaches. This fact makes more accurate interpolation possible. Among the three modes, mode 2 has the worst RMSEs in both azimuth and elevation angles when the same interpolation method is used although its cost function has the minimum value. Also, it achieves the aperture of 62×13, but its corresponding unbalanced aperture may not be appropriate for 2-D DOA estimation in this case. In contrast, the balanced arrays in modes 1 and 3 have better RMSEs, and in azimuth angles, mode 3 outperforms mode 1, whereas in elevation angles, mode 1 is better. Basically, the longer the x axis, the better the elevation estimation performance, and azimuth estimation accuracy depends on the y axis length. Mode 2 has a small array aperture along the y direction,

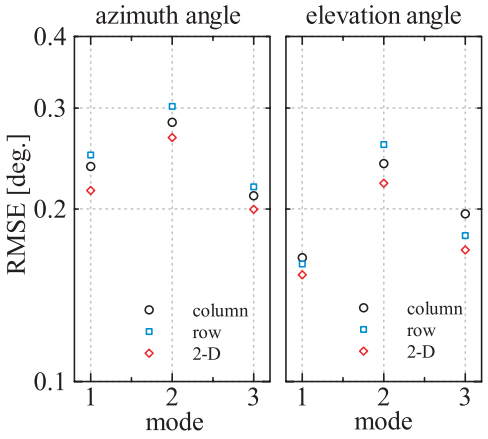


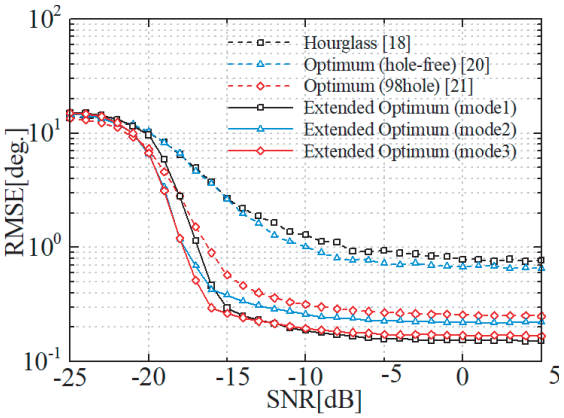
Fig. 9 Comparison of interpolation methods.

and therefore the estimation error of azimuth angles often becomes large. Note that 2-D Unitary ESPRIT estimates DOA by (2) and (3), which include both azimuth and elevation angles. The azimuth estimation error in 2-D Unitary ESPRIT will affect the elevation estimation accuracy, hence the RMSE by mode 2 becomes worse. The aperture of mode 1 is 35×22, and that of mode 3 is 26×31. Thus, the differences in the mode aperture accounts for the differences in their performance. In real situations, it would be better to use an array whose aperture is close to a square to reduce the difference between azimuth and elevation.

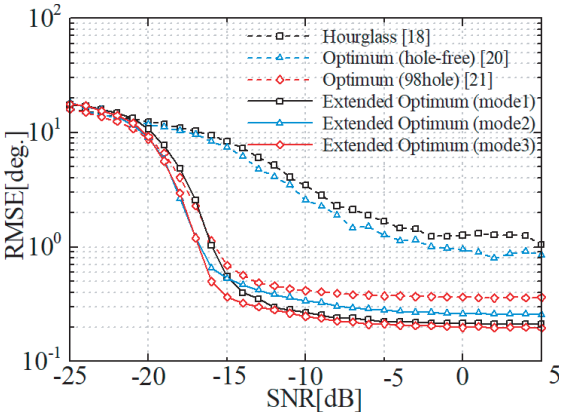
4.4 SNR-RMSE and Snapshots-RMSE Characteristics

In this subsection, proposed optimum arrays are evaluated through RMSE as a function of SNR and the number of snapshots. For comparison, the hourglass array, hole-free optimum array [21], optimum array with 98 holes [22], and extended optimum array modes 1-3 are used. Here the optimum array's aperture is fixed as introduced in this paper and in our previous paper [22]. In the paper [22], the optimum array with 98 holes performed the best. Note that the hourglass array, hole-free optimum array, and optimum array with 98 holes have the same aperture: 26×13. The extended optimum arrays are characterized in Table 3.

First, the simulation results of SNR-RMSE characteristics are shown in Fig. 10(a) and (b). In Fig. 10, Hourglass [18], Optimum [21], [22], Optimum (98 holes) [19], and Extended Optimum (modes 1-3) represent hourglass array, hole-free optimum array, optimum array with 98 holes, and extended optimum arrays (modes 1-3), respectively. The optimum array with hole-free property has almost the same estimation accuracy as the hourglass array due to the restriction that the array must be hole-free. However, in both azimuth and elevation angles, the estimation performance for the fixed optimum array is greatly improved in the low or high SNR domains and the improvement is especially outstanding at SNR values over -15 dB. As a result, this array configuration can suppress the mutual coupling more efficiently by permitting holes in the aperture. As for the ex-

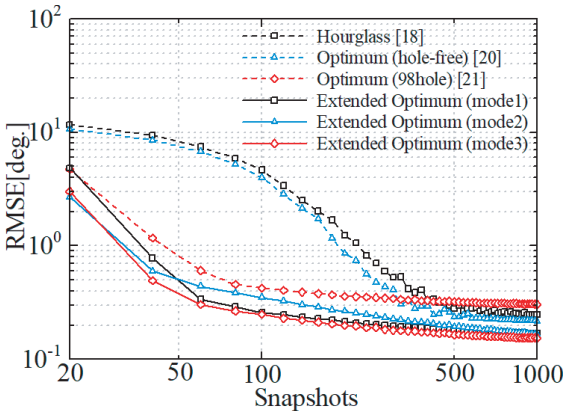


(a) RMSE of azimuth angles

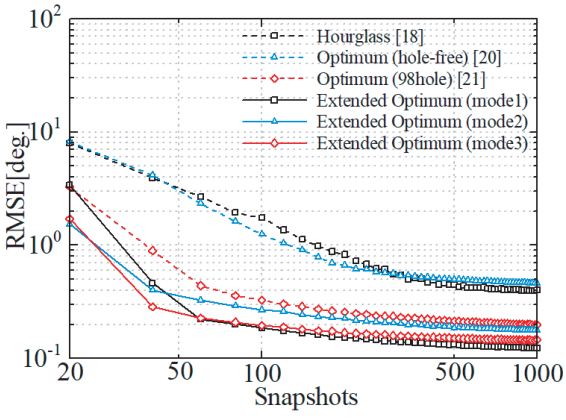


(b) RMSE of elevation angles

Fig. 10 SNR-RMSE characteristics in case of extended optimum arrays.



(a) RMSE of azimuth angles



(b) RMSE of elevation angles

Fig. 11 Snapshots-RMSE characteristics.

tended optimum arrays, especially modes 1 and 3, the RMSEs in both angles under high SNR (more than −15dB) are better than for Optimum (98 holes). The threshold values for modes 2 and 3 are almost the same while that of mode 1 is slightly larger. Therefore, mode 3 has the best performance and the most balanced aperture among these three modes.

The obtained configuration also depends on the initial array. In this example, the initial aperture is 26×13 and mode 3, which is extended in the y direction only to make it more balanced (the aperture of the mode 3 is 26×30). It is important to choose the mode that leads to a balanced array.

Snapshots-RMSE characteristics are shown in Fig. 11(a) and (b). The results resemble SNR-RMSE characteristics in terms of mode 3, which has the lowest RMSE in a small number of snapshots (around 40) and good accuracy in a large number of snapshots. The results emphasize the superiority of optimum arrays, especially extended optimum arrays.

5. Conclusion

In this paper, we introduced a method to optimize array configurations and a method to design a correlation matrix for hole interpolation via NNM. The optimum arrays are made on the basis of SA, and a modified SA algorithm enlarges the aperture to obtain extended optimum arrays. The interpolation methods are based on NNM and optimization orientations were categorized as all-sensors, column, row, and 2-D interpolation. Among the four orientations, 2-D interpolation has the best performance in interpolation methods even the computational cost becomes large. Besides, the extended optimum arrays (with holes and larger aperture) have lower RMSEs than that the conventional method. We emphasize that the larger aperture works effectively even the existence of holes, which can be interpolated by NNM.

For future work, the optimum cost function for array optimization will be studied with considerations for the balance of aperture. Also the application of 2-D sparse arrays to beamforming should be studied, as we have already discussed in [23].

Acknowledgments

The authors thank Prof. Chun-Lin Liu (National Taiwan University) and Mr. Steven Wandale (Yokohama National University) for their helpful comments.

References

[1] M.I. Skolnik, Introduction to Radar Systems, 3rd ed., McGraw-Hill,

2002.
[2] H.L. Van Trees, Optimum Array Processing: Part IV of Detection, Estimation, and Modulation Theory, Wiley, 2002.
[3] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," IEEE Trans. Acoust., Speech, Signal Process., vol.37, no.7, pp.984–995, July 1989.
[4] R.O. Schmidt, "Multiple emitter location and signal parameter estimation," IEEE Trans. Antennas Propag., vol.34, no.3, pp.276–280, March 1986.
[5] T.E. Tuncer and B. Friedlander, Classical and Modern Direction of Arrival Estimation, Academic Press, 2009.
[6] S. Haykin, Array Signal Processing, Prentice-Hall, 1984.
[7] A. Moffet, "Minimum-redundancy linear arrays," IEEE Trans. Antennas Propag., vol.16, no.2, pp.172–175, March 1968.
[8] Y. Iizuka and K. Ichige, "Optimum linear array geometry for 2q-th order cumulant based array processing," Proc. Int. Workshop. Compressed Sensing and its application to Radar, Multimodal Sensing, and Imaging, pp.198–201, May 2016.
[9] S. Iwazaki and K. Ichige, "Underdetermined direction of arrival estimation by sum and difference composite co-array," Proc. 25th IEEE International Conference on Electronics, Circuits and Systems (ICECS), pp.669–672, Dec. 2018.
[10] S. Iwazaki and K. Ichige, "Extended beamforming by sum and difference composite co-array for real-valued signals," IEICE Trans. Fundamentals, vol.E102-A, no.7, pp.918–925, July 2019.
[11] S. Iwazaki, S. Nakamura, and K. Ichige, "DOA-based weighted spatial filter design for sum and difference composite co-array," IEICE Trans. Commun., vol.E103-B, no.10, pp.1147–1154, Oct. 2020.
[12] P. Pal and P.P. Vaidyanathan, "Nested arrays in two dimensions, Part II: Application in two dimensional array processing," IEEE Trans. Signal Process., vol.60, no.9, pp.4706–4718, Sept. 2012.
[13] P. Pal and P.P. Vaidyanathan, "Array design," Bull Seismol. Soc. Amer., vol.58, no.3, pp.977–991, June 1968.
[14] W. Ma, T. Hsieh, and C. Chi, "DOA estimation of quasi-stationary signals with less sensors than sources and unknown spatial noise covariance: A Khatri-Rao subspace approach," IEEE Trans. Signal Process., vol.58, no.4, pp.2168–2180, April 2010.
[15] P. Chevalier, L. Albera, A. Ferreol, and P. Comon, "On the virtual array concept for higher order array processing," IEEE Trans. Signal Process., vol.53, no.4, pp.1254–1271, April 2005.
[16] C.A. Balanis, Antenna Theory: Analysis and Design, 4th ed., Wiley, 2016.
[17] B. Friedlander and A. Weiss, "Direction finding in the presence of mutual coupling," IEEE Trans. Antennas Propag., vol.39, no.3, pp.273–284, March 1991.
[18] C.L. Liu and P.P. Vaidyanathan, "Hourglass arrays and other novel 2D sparse arrays with reduced mutual coupling," IEEE Trans. Signal Process., vol.65, no.13, pp.3369–3383, July 2017.
[19] C.L. Liu and P.P. Vaidyanathan, "Two-dimensional sparse arrays with hole-free coarray and reduced mutual coupling," Proc. IEEE Asilomar Conf. Signal, Syst., Comput., pp.1508–1512, 2016.
[20] C.L. Liu and P.P. Vaidyanathan, "Coprime coarray interpolation for DOA estimation via nuclear norm minimization," Proc. IEEE Int. Sympo. Circuit and Systems, pp.2369–2642, May 2016.
[21] S. Nakamura, S. Iwazaki, and K. Ichige, "An optimum 2D sparse array configuration with reduced mutual coupling," Proc. Int. Sympo. Antennas and Propagation, pp.617–618, Oct. 2018.
[22] S. Nakamura, S. Iwazaki, and K. Ichige, "Optimum 2D sparse array and its interpolation via nuclear norm minimization," Proc. Int. Sympo. Circuits and Systems, pp.1–5, May 2019.
[23] S. Nakamura, S. Iwazaki, and K. Ichige, "Extended beamforming by optimum 2-D sparse arrays," IEICE Communication Express, vol.9, no.6, pp.200–206, June 2020.

Shogo Nakamura received a B.E. and M.E. degrees in Electrical and Computer Engineering from Yokohama National University, Japan, in 2018 and 2020. He joined Denso Corporation in 2020. His research interests include array antenna, DOA estimation, and digital signal processing.

Sho Iwazaki received B.E., M.E. and Dr. Eng. degrees in Electrical and Computer Engineering from Yokohama National University, Japan, in 2009, 2011 and 2020. He joined Sony Corporation in 2011 and has been on loan to Sony Interactive Entertainment Inc. since 2015. His research interests include array signal processing and its applications. He received the Best Student Paper Award in Int. Conf. on Electrical, Electronic and Systems Engineering (ICAEESE) in 2016.

Koichi Ichige received B.E., M.E., and Dr. Eng. degrees in Electronics and Computer Engineering from the University of Tsukuba, Japan, in 1994, 1996, and 1999. He became a research associate at the Department of Electrical and Computer Engineering, Yokohama National University, Japan in 1999, where he is currently a professor. He was a visiting researcher at the Swiss Federal Institute of Technology Lausanne (EPFL), Switzerland, in 2001–2002. His research interests include digital signal processing, approximation theory, and their applications to image processing and mobile communication. He served as an associate editor of IEEE Transactions on Industrial Electronics in 2004–2008, an associate editor of Journal of Circuits, Systems and Computers (JCSC) in 2012–2014, and an associate editor of IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences (IEICE-EA) in 2015–2018. He currently serves as a editor of IEICE-EA. He received the Meritorious Award on Radio from the Association of Radio Industries and Businesses (ARIB) in 2006 and the Best Letter Award from IEICE Communication Society in 2007. He is a member of IEEE and IEICE.