HOKKAIDO UNIVERSITY

Title	Radio Techniques Incorporating Sparse Modeling
Author(s)	Nishimura, Toshihiko; Ogawa, Yasutaka; Ohgane, Takeo et al.
Citation	IEICE transactions on fundamentals of electronics communications and computer sciences, E104A(3), 591-603 https://doi.org/10.1587/transfun.2020EAI0001
Issue Date	2021-03
Doc URL	https://hdl.handle.net/2115/81416
Rights	copyright©2021 IEICE
Type	journal article
File Information	E104.A_2020EAI0001.pdf

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

INVITED SURVEY PAPER

Radio Techniques Incorporating Sparse Modeling

Toshihiko NISHIMURA[†a)], ***Senior Member***, **Yasutaka OGAWA**[†], **Takeo OHGANE**[†], ***Fellows***, ***and*** **Junichiro HAGIWARA**[†], ***Member***

SUMMARY Sparse modeling is one of the most active research areas in engineering and science. The technique provides solutions from far fewer samples exploiting sparsity, that is, the majority of the data are zero. This paper reviews sparse modeling in radio techniques. The first half of this paper introduces direction-of-arrival (DOA) estimation from signals received by multiple antennas. The estimation is carried out using compressed sensing, an effective tool for the sparse modeling, which produces solutions to an underdetermined linear system with a sparse regularization term. The DOA estimation performance is compared among three compressed sensing algorithms. The second half reviews channel state information (CSI) acquisitions in multiple-input multiple-output (MIMO) systems. In time-varying environments, CSI estimated with pilot symbols may be outdated at the actual transmission time. We describe CSI prediction based on sparse DOA estimation, and show excellent precoding performance when using the CSI prediction. The other topic in the second half is sparse Bayesian learning (SBL)-based channel estimation. A base station (BS) has many antennas in a massive MIMO system. A major obstacle for using the massive MIMO system in frequency-division duplex mode is an overhead for downlink CSI acquisition because we need to send many pilot symbols from the BS and to get the feedback from user equipment. An SBL-based channel estimation method can mitigate this issue. In this paper, we describe the outline of the method, and show that the technique can reduce the downlink pilot symbols.

key words: *sparse modeling, compressed sensing, sparse Bayesian learning, DOA estimation, channel estimation*

1. Introduction

Radio techniques have developed rapidly for these decades. Especially, mobile communication has evolved almost every ten years. Many technologies supporting the fifth generation (5G) have been extensively studied [1]–[6], and commercial 5G services are being deployed in several countries. More recently, the evolution beyond 5G or toward 6G is being discussed [7], [8]. Various signal processing techniques enable advanced wireless communication. One of them is a multiple-input multiple-output (MIMO) system, in which both of the transmitter and the receiver have multiple antennas [9], [10]. The MIMO technique, space-domain signal processing, makes it possible to realize high-speed transmission using space division multiplexing [11]. Other than mobile communication, radio technologies have been widely studied in several fields such as radars and wireless sensor networks [12]–[16].

Manuscript received May 29, 2020.
Manuscript revised August 4, 2020.
Manuscript publicized September 1, 2020.

[†]The authors are with the Faculty of Information Science and Technology, Hokkaido University, Sapporo-shi, 060-0814 Japan.
a) E-mail: nishim@ist.hokudai.ac.jp

DOI: 10.1587/transfun.2020EAI0001

On the other hand, sparse modeling has been collecting considerable attention in several fields. The signal processing technique is a framework which provides solutions exploiting sparsity, that is, the majority of the data are equal to zero. Several important concepts are discussed in [17]. Compressed sensing or compressive sensing is an effective tool for the sparse modeling [18]–[30]. One of the biggest successes of the method is magnetic resonance imaging (MRI), which has been widely used for medical diagnosis due to its high resolution and noninvasive merits [31], [32]. Another breathtaking result is black hole imaging achieved by the Event Horizon Telescope Collaboration [33]–[39].

Compressed sensing plays a crucial role also in radio techniques. We need channel state information (CSI) for wireless communication. CSI is essential for detection at a receiver, and is also required for MIMO precoding at a transmitter. In radio engineering, information is transmitted by radio waves. The waves are inevitably affected by propagation environments, that is, they are reflected, scattered, and diffracted from the surrounding objects, and are received as multipath signals. These multipath signals determine channels. According to channel measurement campaigns [40], [41], multipath components do not arrive uniformly from all angles of arrival but as clusters. This propagation property holds also in the delay domain, and we can assume that wireless channels are sparse in spatial and temporal domains. Thus, many researches have been done for channel estimation with compressed sensing such as [42]–[50].

Another research area is high-resolution direction-of-arrival (DOA) estimation of radio waves, which is of great importance in such as radar, position estimation, radio astronomy, and CSI estimation in MIMO systems. The research dates back to the 1960s [51], and has been continued until today [52]–[55]. In sparse radio environments, DOA estimation based on compressed sensing has significant benefits such as smaller number of snapshots required and has ability to deal with coherent signals, therefore many references have been published for example [56]–[61]. Furthermore, the DOA estimation technique that uses compressed sensing and a generalized MUSIC criterion has been proposed [62]. The method can reduce the number of antenna elements.

Based on these research backgrounds, we review sparse modeling in radio techniques. The first half of this paper introduces DOA estimation using compressed sensing. After taking a quick look at sparse modeling and formulating the

Copyright © 2021 The Institute of Electronics, Information and Communication Engineers

problem as array signal processing, we will explain sparse DOA estimation algorithms. The second half reviews MIMO CSI acquisition methods. In time-varying environments, CSI estimated with pilot symbols may be outdated at the actual transmission time. One of the solutions to this issue is channel prediction based on sparse DOA estimation [63], which can provide a long prediction range. We describe the technique in detail. The other topic to be introduced in the second half is sparse Bayesian learning (SBL)-based channel estimation. A base station (BS) has many antennas in a massive MIMO system [64], [65]. A major obstacle for using the massive MIMO system in frequency-division duplex (FDD) mode is an overhead for downlink CSI acquisition, that is, we need to send many pilot symbols from the BS and to get the feedback from user equipment (UE) [66]–[68]. Dai et al. [69] have proposed an SBL-based channel estimation method to mitigate this issue, which will be discussed later. Introduction of the above two channel acquisition methods will be beneficial to readers who are interested in sparse modeling in radio techniques.

The rest of the paper is organized as follows. In Sect. 2, we briefly review sparse modeling and compressed sensing. In Sect. 3, we explain the basics of array signal processing and DOA estimation based on compressed sensing. In Sect. 4, we introduce channel prediction using the DOA estimation in time-varying environments and downlink SBL-based channel estimation in an FDD massive MIMO system. Conclusions follow in Sect. 5.

2. Brief Review of Sparse Modeling and Compressed Sensing

This section briefly describes sparse modeling and compressed sensing algorithms used for DOA estimation in Sects. 3 and 4. For interested readers, refer to review papers or books such as [27], [28], [70], [71], [73] for further details on this research area.

We consider the following simultaneous linear equations

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}, \tag{1}$$

where $\boldsymbol{y}$ and $\boldsymbol{x}$ are M-dimensional and N-dimensional complex-valued vectors, respectively, and $\boldsymbol{A}$ is an $M \times N$ complex-valued matrix. We want to solve the unknown vector $\boldsymbol{x}$ for given $\boldsymbol{y}$ and $\boldsymbol{A}$. We assume $M < N$, that is, the number of equations is less than that of unknowns. In this case, (1) is called underdetermined systems, and its solution is not unique. Infinitely many vectors $\boldsymbol{x}$ satisfy (1).

Here, we assume that $\boldsymbol{x}$ is sparse, that is, most of the elements are zero. Due to the sparsity, it is natural to obtain the solution $\boldsymbol{x}$ that has the least non-zero elements. We define a vector 0-norm (ℓ_0-norm) as the number of non-zero elements in $\boldsymbol{x}$, and denote it as $\|\boldsymbol{x}\|_0$. Then, the solution is given by

$$\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_0 \quad \text{subject to} \quad \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}. \tag{2}$$

Unfortunately, the 0-norm minimization is unrealistic because of the expected combinatorial explosion as the dimension of $\boldsymbol{x}$ increases.

We introduce a vector p-norm (ℓ_p-norm) defined by

$$\|\boldsymbol{x}\|_p = \left(\sum_{i=1}^{N} |x_i|^p \right)^{\frac{1}{p}} \quad (p > 0). \tag{3}$$

We replace $\|\boldsymbol{x}\|_0$ in (2) with $\|\boldsymbol{x}\|_p$. Furthermore, using a Lagrange multiplier α, we have the following unconstrained minimization

$$\arg\min_{\boldsymbol{x}} \left(\frac{1}{2} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \alpha \|\boldsymbol{x}\|_p^p \right), \tag{4}$$

The above equation can also treat a case where $\boldsymbol{y}$ includes noise as stated in later sections. When $\boldsymbol{p} \geq 1$, the objective function of (4) is a convex function of the variable $\boldsymbol{x}$, which allows us to efficiently find an approximate solution by numerical computations. However, it is difficult to obtain a sparse solution with 2-norm (ℓ_2-norm) minimization. By using 1-norm (ℓ_1-norm) minimization, the sparse solution can be obtained with realistic computational complexities, and many methods have been presented so far [27]. As will be mentioned later, some algorithm uses the value of $0 < p < 1$.

One of the most typical algorithms for the 1-norm minimization is the fast iterative shrinkage thresholding algorithm (FISTA) [74] which is an accelerated version of the iterative shrinkage thresholding algorithm (ISTA), a kind of proximity gradient method. The solution of ISTA is given by the following iteration

$$\boldsymbol{x}[k+1] = S_{\frac{\alpha}{L}} \left(\boldsymbol{x}[k] + L^{-1}\boldsymbol{A}^H \left(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}[k] \right) \right), \tag{5}$$

where ^{H} denotes Hermitian transpose and $S_a(b)$ is a shrinkage-thresholding function, also known as a soft threshold function, defined by

$$S_a(b) = \begin{cases} \left(\frac{|b|-a}{|b|} \right) b & (a < |b|) \\ 0 & (\text{otherwise}) \end{cases}. \tag{6}$$

Note that the function is applied to each element of the vector. Furthermore, L is a parameter satisfying

$$L \geq \|\boldsymbol{A}^H\boldsymbol{A}\|. \tag{7}$$

Here, $\|\boldsymbol{A}^H\boldsymbol{A}\|$ denotes the maximum eigenvalue of $\boldsymbol{A}^H\boldsymbol{A}$, which is the Lipschitz constant of the gradient of $(1/2)\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2$. Note that when $\boldsymbol{x}, \boldsymbol{y}$, and $\boldsymbol{A}$ are real-valued, the Hermitian transpose in the above expressions is replaced with transpose [27].

In FISTA, to accelerate the convergence, we slightly change the iteration as follows

---

$\boldsymbol{x}[0] := \boldsymbol{A}^H\boldsymbol{y}$.
$\boldsymbol{w}[1] := \boldsymbol{x}[0]$.
$\beta := 1$.
$k := 1$.

loop
$\boldsymbol{x}[k] := S_{\alpha/L}(\boldsymbol{w}[k] + L^{-1}\boldsymbol{A}^H(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{w}[k]))$.
$\beta[k+1] := \frac{1}{2} + \sqrt{\frac{1}{4} + \beta[k]^2}$.
$\boldsymbol{w}[k+1] := \boldsymbol{x}[k] + \frac{\beta[k]-1}{\beta[k+1]}(\boldsymbol{x}[k] - \boldsymbol{x}[k-1])$.
$k := k + 1$.
end loop

It is seen that $\boldsymbol{A}\boldsymbol{w}[k]$ requires MN complex multiplications. Since $\boldsymbol{A}^H$ and $(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{w}[k])$ are an $N \times M$ matrix and an M-dimensional vector, respectively, the matrix-vector multiplication $\boldsymbol{A}^H(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{w}[k])$ needs NM complex multiplications. Thus, the computational complexity at each step is $O(MN)$.

Now, we consider the half-quadratic regularization (HQR) method [76]. In HQR, $p < 1$ is usually used in (4). The algorithm is formulated deleting the constant $1/2$ in (4) in [76]. In order to avoid problems due to non-differentiability of the ℓ_p-norm around the origin when $p \leq 1$, the following smooth approximation is applied to (3).

$$\|\boldsymbol{x}\|_p \approx \left(\sum_{i=1}^{N}(|x_i|^2 + \epsilon)^{p/2}\right)^{1/p}. \tag{8}$$

Here, $\epsilon \geq 0$ is a small constant. HQR obtains the solution by the iteration

$$\boldsymbol{x}[k+1] = \boldsymbol{H}(\boldsymbol{x}[k])^{-1}\boldsymbol{A}^H\boldsymbol{y}, \tag{9}$$

where

$$\boldsymbol{H}(\boldsymbol{x}) = \boldsymbol{A}^H\boldsymbol{A} + \alpha\boldsymbol{\Lambda}(\boldsymbol{x}) \tag{10}$$

$$\boldsymbol{\Lambda}(\boldsymbol{x}) = \mathrm{diag}\left\{\frac{\frac{p}{2}}{(|x_i|^2 + \epsilon)^{1-\frac{p}{2}}}\right\}. \tag{11}$$

In (11), diag$\{\cdot\}$ denotes a diagonal matrix whose ith diagonal element is the expression inside the brackets.

It is seen that we can calculate $\boldsymbol{A}^H\boldsymbol{y}$ and $\boldsymbol{A}^H\boldsymbol{A}$ in advance before the iteration. Since $\boldsymbol{H}(\boldsymbol{x})$ is an $N \times N$ matrix, we need complex multiplications in the order of N^3 to carry out the inversion $\boldsymbol{H}(\boldsymbol{x}[k])^{-1}$ in (9). Furthermore, $\boldsymbol{A}^H\boldsymbol{y}$ is an N-dimensional vector, and the matrix-vector multiplication $\boldsymbol{H}(\boldsymbol{x}[k])^{-1}\boldsymbol{A}^H\boldsymbol{y}$ requires N^2 complex multiplications. Consequently, we need $O(N^3)$ complex multiplications at each step. The computational load is heavy because N usually has a large value.

Finally, we introduce the Orthogonal Matching Pursuit (OMP), which is one of the greedy algorithms [22], [23]. OMP is not based on (4) but obtains the support set of $\boldsymbol{x}$ denoted by $\boldsymbol{\Omega}$. We express the K-dimensional vector whose elements consist of $\boldsymbol{x}$ in $\boldsymbol{\Omega}$ as $\boldsymbol{x}_{\boldsymbol{\Omega}}$, where K denotes the number of non-zero elements in $\boldsymbol{x}$. We also represent the matrix deleting the columns except $\boldsymbol{\Omega}$ from $\boldsymbol{A}$ as $\boldsymbol{A}_{\boldsymbol{\Omega}}$, which is an $M \times K$ matrix. Using $\boldsymbol{x}_{\boldsymbol{\Omega}}$ and $\boldsymbol{A}_{\boldsymbol{\Omega}}$, (1) is rewritten as

$$\boldsymbol{y} = \boldsymbol{A}_{\boldsymbol{\Omega}}\boldsymbol{x}_{\boldsymbol{\Omega}}. \tag{12}$$

From the sparsity of $\boldsymbol{x}$, $K < M$ holds, and we can solve (12) for $\boldsymbol{x}_{\boldsymbol{\Omega}}$. The solution is given as

$$\boldsymbol{x}_{\boldsymbol{\Omega}} = \left(\boldsymbol{A}_{\boldsymbol{\Omega}}^H\boldsymbol{A}_{\boldsymbol{\Omega}}\right)^{-1}\boldsymbol{A}_{\boldsymbol{\Omega}}^H\boldsymbol{y}. \tag{13}$$

We iteratively obtain the support set $\boldsymbol{\Omega}$. Let $\boldsymbol{\Omega}[k]$ and $\boldsymbol{r}[k]$ denote the kth support set and residual $\boldsymbol{y} - \boldsymbol{A}_{\boldsymbol{\Omega}[k]}\boldsymbol{x}_{\boldsymbol{\Omega}[k]}$, respectively. We start with $\boldsymbol{\Omega}[0] = \phi$ and $\boldsymbol{x}_{\boldsymbol{\Omega}[0]} = 0$, i.e., $\boldsymbol{r}[0] = \boldsymbol{y}$. Moreover, we express the jth column of $\boldsymbol{A}$ as $\boldsymbol{a}_j$. At each iteration, we find $\boldsymbol{a}_j$ most correlated with the residual, and include the column number into the support. Based on this concept, we update the support set in the following manner [27], [71].

$\boldsymbol{\Omega}[0] := \phi$.
$\boldsymbol{x}_{\boldsymbol{\Omega}[0]} := \boldsymbol{0}$. i.e., $\boldsymbol{r}[0] = \boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}_{\boldsymbol{\Omega}[0]} = \boldsymbol{y}$.
$k := 0$.
loop
$J[k+1] := \arg\max_j \frac{|\boldsymbol{a}_j^H\boldsymbol{r}[k]|}{\|\boldsymbol{a}_j\|_2}$.
$\boldsymbol{\Omega}[k+1] := \boldsymbol{\Omega}[k] \cup \{J[k+1]\}$.
$\boldsymbol{x}_{\boldsymbol{\Omega}[k+1]} := \left(\boldsymbol{A}_{\boldsymbol{\Omega}[k+1]}^H\boldsymbol{A}_{\boldsymbol{\Omega}[k+1]}\right)^{-1}\boldsymbol{A}_{\boldsymbol{\Omega}[k+1]}^H\boldsymbol{y}$.
$\boldsymbol{r}[k+1] := \boldsymbol{y} - \boldsymbol{A}_{\boldsymbol{\Omega}[k+1]}\boldsymbol{x}_{\boldsymbol{\Omega}[k+1]}$.
$k := k + 1$.
end loop

Since $\boldsymbol{a}_j$ and $\boldsymbol{r}[k]$ are M-dimensional vectors, we need M complex multiplications to calculate the scalar product $\boldsymbol{a}_j^H\boldsymbol{r}[k]$. In the kth loop of the iteration, we need $M(N-k)$ complex multiplications to obtain $\boldsymbol{a}_j$ that maximizes $|\boldsymbol{a}_j^H r[k]|/\|\boldsymbol{a}_j\|_2$. Note that $(k-1)$ vectors have been chosen up to the $(k-1)$th loop, and we only need to search for the vector $\boldsymbol{a}_j$ among the remaining $(N-k)$ ones. Furthermore, $\boldsymbol{x}_{\boldsymbol{\Omega}[k+1]}$ and $\boldsymbol{r}[k+1]$ require $k(kM + M + k) + O(k^3)$ and kM complex multiplications for their updates, respectively. In total, we need $O(\max(MN, k^2M, k^3))$ complex multiplications in the kth loop.

We have considered computational complexities at each step of the algorithms. It would depend on an algorithm how many steps we need for convergence. The authors advise interested readers to try simulations.

3. DOA Estimation of Electromagnetic Waves

Estimating DOAs of radio waves from signals received by multiple antennas is a typical example of array signal processing. The handling of signals linearly observed by an array antenna can be described by simultaneous linear equations. We usually need to treat an underdetermined linear system with less observation data than the number of unknowns (original signals). However, when the original signal is sparse, the exact solution can be obtained by a concept called compressed sensing. In this section, we introduce DOA estimation with compressed sensing and compare the performance of three algorithms.

3.1 Basics of DOA Estimation with Array Antenna

Figure 1 is a diagram of a DOA estimation system using an

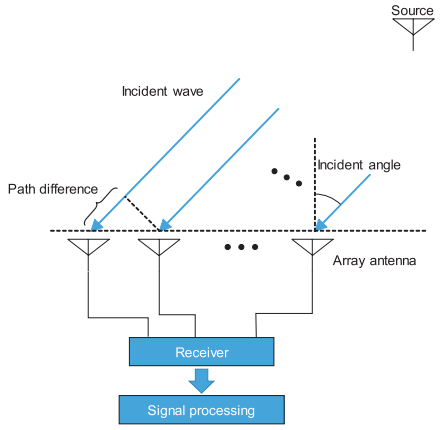


Fig. 1 DOA estimation with array antenna.

array antenna for a single source case. It is assumed that the source is so far away that the radio wave is incident as a plane wave on the array antenna, and the radio wave experiences phase rotation between antenna elements according to the path difference shown in Fig. 1. Therefore, the DOA of the radio wave can be estimated by calculating the path difference between the elements from the phase difference. As can be seen from Fig. 1, to calculate the DOA of a single wave, the array antenna needs only two elements. If the positions of the two antenna elements are known exactly, the DOA can be readily calculated. However, the received signal contains noise in addition to errors such as IQ imbalance due to the fabrication process of the device. Although the latter can be offset or mitigated by calibration, it is impossible to completely remove the effects of noise, and the estimated DOA inevitably has an error. Furthermore, in general, we have plural sources and multipaths. DOA estimation has been a challenging issue for radio researchers, and has a long research history [72]. In the next subsection, we introduce compressed sensing-based DOA estimation.

3.2 Application of Compressed Sensing

As shown in Fig. 2, we assume that K narrowband signals are incident on a uniform linear array (ULA) of M omnidirectional elements. Also, as seen from Fig. 2, we divide the angular domain into N small regions, each of which is called a "bin". We define $x_n(t)$ as a signal arriving from the direction of the nth bin at time t. If no signal arrives from it, $x_n(t) = 0$ holds. We express an N-dimensional column vector consisting of $x_n(t)$ as $\boldsymbol{x}(t)$, which is called an original signal vector. For example, when signals are incident from the directions of the third and nth bins, the original signal vector is $\boldsymbol{x}(t) = [0,\ 0,\ x_3(t),\ 0, \ldots 0,\ x_n(t),\ 0, \ldots 0]^{\mathrm{T}}$, where $[\ ]^{\mathrm{T}}$denotes transpose. We see that the DOAs of arriving signals can be estimated from the support of $\boldsymbol{x}(t)$.

Next, we introduce an $M \times N$ matrix $\boldsymbol{A}$ whose the (m, n)th element is given by

$$a_{mn} = e^{j\frac{2\pi}{\lambda} d_m \sin\theta_n}, \tag{14}$$

where λ, d_m, and θ_n are the wavelength corresponding to the frequency of the arriving wave, the distance of the mth antenna element from a reference point, and the angle of the center of the nth bin, respectively. We refer to the center of each bin as a grid point. Each column vector of $\boldsymbol{A}$ is a steering vector for the angle θ_n. Note that $\boldsymbol{A}$ is called a steering matrix, an observation matrix, or a sensing matrix.

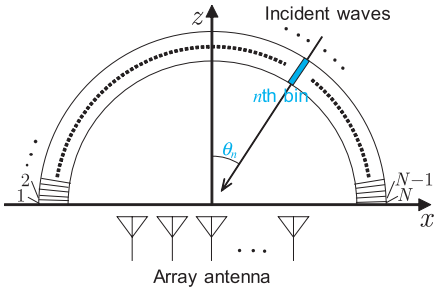


Fig. 2 DOA estimation model for compressed sensing.

We denote an M-dimensional column vector whose mth element is a received signal at the mth antenna at time t as $\boldsymbol{y}(t)$. The vector is given by

$$\boldsymbol{y}(t) = \boldsymbol{A}\boldsymbol{x}(t) + \boldsymbol{n}(t), \tag{15}$$

where $\boldsymbol{n}(t)$ is an M-dimensional noise vector. Note that throughout this paper, we assume additive Gaussian noise that is independent spatially and temporally.

In (15), $\boldsymbol{y}(t)$, $\boldsymbol{x}(t)$, and $\boldsymbol{n}(t)$ are continuous-time waves. We assume that the DOAs of arriving signals do not change during a limited time period. Sampling (15), we can estimate $\boldsymbol{x}(t)$. Moreover, if the complex envelopes of $\boldsymbol{x}(t)$ do not change temporally during the period, we can improve SNR averaging plural snapshot measured data. We assume that the number of arriving signals K is much less than that of bins N, so $\boldsymbol{x}(t)$ is a sparse vector. Thus, the Eq. (15) becomes an underdetermined system with a sparse solution. The relationship among the number of arriving signals K and the dimensions $\boldsymbol{x}(t)$ and $\boldsymbol{y}(t)$ is given by

$$K < M < N. \tag{16}$$

From the above, it is seen that we can obtain $\boldsymbol{x}(t)$ using compressed sensing stated in Sect. 2, and we can estimate the DOAs of arrival signals.

3.3 Example of DOA Estimation Using Compressed Sensing Algorithms

In this subsection, we show an example of DOA estimation using FISTA, HQR, and OMP explained in Sect. 2. The width of each bin shown in Fig. 2 is 1°, and the grid points are located at integer multiples of 1° such as $\ldots, -1°, 0°, 1°, 2° \ldots$. Two uncorrelated narrow-band waves with the same power impinge on a five-element ULA with half-wavelength spacing. The DOAs are $-10°$ and 30° that coincide with grid points. We assumed that we knew in advance the number of arrival signals 2. We applied each algorithm with a certain convergence condition, and found

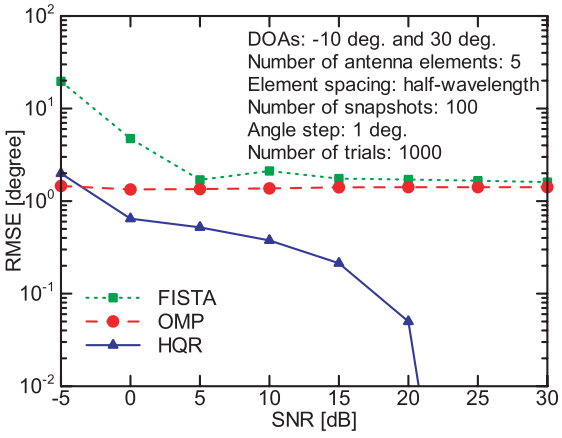


Fig. 3 Comparison among three algorithms.

two largest elements in the estimated vector $\boldsymbol{x}(t)$. We considered them as the incident waves, and estimated the DOAs from the supports of them.

Figure 3 shows the root mean square error (RMSE) of the DOA estimation versus SNR. Here, RMSE is defined by

$$\mathrm{RMSE} = \sqrt{\frac{1}{2I}\sum_{k=1}^{2}\sum_{i=1}^{I}\left|\hat{\theta}_{k,i} - \theta_k\right|^2}, \tag{17}$$

where I is the number of trials, which was 1000 in the simulation. $\hat{\theta}_{k,i}$ and θ_k are the estimated incident angle and the true one of the kth incident wave ($k = 1, 2$), respectively. Here, SNR denotes the signal-to-noise power ratio of each incident wave. From the figure, it can be seen that HQR shows the best estimation result in almost the whole region of the SNR. The reason for this may be due to the value of $0 < p < 1$. The effects of parameters including p to the DOA estimation accuracy are discussed in [77]. RMSE in the region of SNR larger than 20 dB was 0. The estimated DOAs coincided with grid points of $-10°$ and $30°$, and we had no estimation error in the 1000 trials. As stated previously, however, HQR needs an inversion of $N \times N$ matrix every step, and the computational load is heavy. On the other hand, the performance of FISTA was significantly poor at low SNR. OMP had an error of about 1.5°, regardless of the SNR. It is interesting that although the computational complexity of OMP is very low, the DOA estimation performance is much better than that of FISTA at the low SNR region. The authors advise readers to carefully consider algorithms before they apply the compressed sensing to their issue.

4. MIMO CSI Acquisition Using Sparse Modeling

In this section, we introduce two examples of MIMO CSI acquisition techniques based on DOA estimation. The first is multi-user CSI prediction using compressed sensing in time-varying environments. The second is SBL-based downlink CSI estimation in an FDD massive MIMO system.

4.1 Multi-User MIMO CSI Prediction

Here, we review the basic concept of MIMO CSI prediction using two-step compressed sensing [63]. In multi-user MIMO downlink transmission, precoding is done at a BS to prevent inter-user interference and to enhance transmission rates. Downlink CSI is required at the BS for the precoding, and is obtained using pilot symbols. In time-varying environments, however, the MIMO channels change during the time interval between channel estimation and actual signal transmission. This causes interference and deteriorates communication quality. If we accurately predict channels from past CSI, we can mitigate the problem. Many prediction techniques have been presented so far [78]. Among them, the sum-of-sinusoids (SOS) method predicts channels by resolving an arrival signal to a UE into individual multipath components and summing the predicted ones. If the Doppler frequency and complex amplitude of each multipath component are estimated accurately, we can predict reliable channels for a long prediction range. The SOS method stated below is thus promising. Note that we assume a narrowband system in which we can ignore the delay between multipath components. The considerations below, however, also hold for each subcarrier of OFDM in a broadband system.

A time-varying channel $h(t)$ at a reference point of a UE array from a certain BS antenna is given by

$$h(t) = \sum_{k=1}^{K} A_k e^{j(2\pi f_k t + \phi_k)}, \tag{18}$$

where K denotes the number of multipath components. Also, A_k, ϕ_k, and f_k denote the amplitude, phase, and Doppler frequency for the kth multipath component, respectively. From (18), if we obtain the complex amplitudes $A_k e^{j\phi_k}$ and the Doppler frequencies f_k at time $t = 0$ for the multipath components, we can predict future channels for the UE by summing the predicted multipath components.

We resolve the signal into multipath components by estimating the DOAs. Because the DOAs change relatively slowly, we can assume that they are constant for the channel prediction period. Moreover, we assume that UEs have a uniform circular array (UCA) with M omnidirectional antennas, and that the UEs move at a constant speed.

Here, we state the formulation for one UE. For the other UEs, the formulation is the same. We divide the angular domain surrounding the UE into N bins. We represent the observation vector $\boldsymbol{y}(t)$, original signal vector $\boldsymbol{x}(t)$, and thermal noise vector $\boldsymbol{n}(t)$ as

$$\boldsymbol{y}(t) = [y_1(t),\ y_2,(t)\ \ldots,\ y_M(t)]^{\mathrm{T}} \tag{19}$$
$$\boldsymbol{x}(t) = [x_1(t),\ x_2(t), \ldots,\ x_N(t)]^{\mathrm{T}} \tag{20}$$
$$\boldsymbol{n}(t) = [n_1(t),\ n_2(t), \ldots,\ n_M(t)]^{\mathrm{T}}. \tag{21}$$

Note that each element in $\boldsymbol{x}(t)$ is a complex amplitude of an incident multipath component. Then, obtaining the original signal vector $\boldsymbol{x}(t)$ corresponds to DOA estimation as stated in

Sect. 3. The observation vector $\boldsymbol{y}(t)$ consists of the received signal at each UE antenna, and is given by

$$\boldsymbol{y}(t) = \boldsymbol{A}\boldsymbol{x}(t) + \boldsymbol{n}(t), \tag{22}$$

where $\boldsymbol{A}$ is the $M \times N$ observation matrix.

We use a channel prediction method for a higher frequency band such as 20 GHz, where we can consider that the number of multipath components K is much fewer than N [79]. Thus, the original signal vector $\boldsymbol{x}(t)$ is sparse, and we can apply the compressed sensing technique to obtain $\boldsymbol{x}(t)$. The HQR method is used for the DOA estimation. This is the first-step compressed sensing.

In [63], the following procedures are used to estimate more accurate DOAs, complex amplitudes, and Doppler frequencies, and to predict future MIMO channels:

1. We apply Khatri-Rao processing [80] before the DOA estimation. The processing improves the DOA estimation accuracy because it works as if we had M^2 antennas at the UE. Khatri-Rao processing requires wide-sense quasi-stationarity, which means that the second-order statistics of the multipath components are time-varying, but that they remain static over the channel prediction period. This assumption holds because the channel prediction can be done for a short period of time.

2. We have P sets of observation signals obtained at multiple times ($t = t_1, t_2, \ldots, t_P$). Khatri-Rao processing works properly when the arrival signals are uncorrelated, but all the multipath components are coherent in narrow band systems. Using the P sets of signals, we do preprocessing to decrease the correlations between the multipath components. Moreover, we need phases of each multipath component at different times to obtain Doppler frequency.

3. Using the estimated DOAs, we obtain the complex amplitudes and Doppler frequencies of the multipath components with the second-step compressed sensing. The two-step technique enables us to deal with the channel changes due to the motion of the UE and/or surrounding scatterers.

4. The above process is done for the channels between a single BS antenna and all UE antennas. For the remaining BS antennas, we calculate the complex amplitudes of the multipath components using the least squares method with less computational complexity.

Up to this point, we have estimated the Doppler frequencies and complex amplitudes of the multipath components to the UE antennas from all BS antennas. Using these values, we can predict the MIMO channels at the actual transmission time.

Here, we briefly show simulation results of the proposed two-step prediction method. As shown in Fig. 4, we assumed three UEs in the multipath environment. We used the Jakes channel model, and there were nine scatterers around each

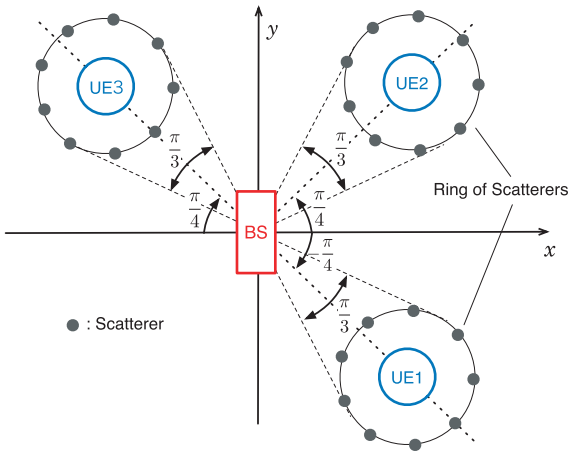


Fig. 4 Positions of BS and UEs in multipath environment. Reprinted from [63] with permission (©2019 IEEE).

UE. The BS had a 24-element ULA with half-wavelength spacing. The array was aligned along the y-axis. The UEs had an eight-element UCA whose radius was a wavelength. In Fig. 4, UE2 and UE3 are symmetrically located about the y-axis, the array axis. In the Jakes channel model, multipath components from the scatterers have equal amplitude but different phases. Thus, the channels for UE2 and UE3 are not symmetric about the y-axis, and the ULA can treat UE2 and UE3. The UEs moved at a constant speed of 1.5 m/s. The scatterers moved at a constant speed that was distributed independently and uniformly in the range from 0 to 1.5 m/s. Also, the moving directions of the UEs and scatterers were uniformly random from 0 to 2π. The channels were observed every 5 ms. We used a single snapshot at each observation for DOA estimation using Khatri-Rao processing. Refer to [63] for the other simulation parameters.

Figure 5 plots the average packet error rates (PERs) for UE1 versus the normalized transmit (TX) power. The prediction interval was 2.5 ms, that is, we predicted the MIMO channels at future time of 2.5 ms from the last observation time t_P. Note that we added the performance of the proposed channel prediction method for $P = 5$ on the assumption that the DOAs of the multipath components were perfectly known "Perfect DOA Estimation". In this case, the complex amplitudes and Doppler frequencies were obtained with the compressed sensing technique. The performance deteriorated when transmitting without channel prediction. The improvement in the linear extrapolation was small, and the performance of the autoregressive (AR)-model-based channel prediction method was worse than that of the case without prediction. On the other hand, when using the proposed channel prediction method, the PER performance greatly improved. The deterioration for the proposed channel prediction method ($P = 5$) was less than 3 dB from "Ideal Case", and 1 dB from "Perfect DOA Estimation" at the PER of 10^{-2}. Also, the larger P, the better the PER was obtained. From these results, we can say that the channel prediction method is effective in sparse channel environments.

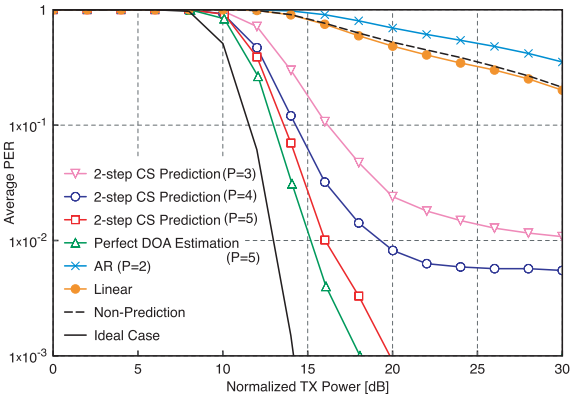


Fig. 5 Average packet error rate for UE1 versus normalized TX power. Reprinted from [63] with permission (©2019 IEEE).

4.2 SBL-Based FDD Massive MIMO Downlink CSI Estimation

As stated in the precedent subsection, DOA information is effective for CSI estimation. Also, as we have described so far, we discretize the angular domain into bins. The center of each bin is called a grid point, which was defined in Sect. 3.2. We have performance degradation in practical situations where true DOAs are not on the grid point, and a dense grid is preferable for accurate DOA estimation. On the other hand, a dense grid leads to a high mutual coherence [27] that violates the condition for the sparse signal recovery. Against the backdrop of the problem, off-grid methods have been studied in [81]–[84] where the estimated DOAs are no longer constrained in the sampling grid set.

In this subsection, we introduce the basic concept of SBL-based off-grid estimation proposed in [69]. This paper addresses downlink CSI estimation in FDD massive MIMO systems, where the base stations are equipped with many antennas. The conventional training overhead for the downlink channel acquisition grows proportionally with the number of BS antennas, which can be quite large. The main points of [69] are as follows.

We assume that the BS has an M-element ULA with half-wavelength spacing, and that a UE has a single antenna. The BS sends a sequence of T training pilot symbols. We denote them by a $T \times M$ matrix $\boldsymbol{X}$. Representing the M-dimensional downlink channel vector as $\boldsymbol{h}$, the T-dimensional received signal vector $\boldsymbol{v}$ at the UE is given by

$$\boldsymbol{v} = \boldsymbol{X}\boldsymbol{h} + \boldsymbol{n}, \tag{23}$$

where $\boldsymbol{n}$ is the T-dimensional additive Gaussian noise vector. We assume that the elements are independent, and have mean 0 and variance σ^2. Since M is large and $M > T$, the conventional channel estimation techniques do not work.

Here, we assume that K multipath components depart from the BS array, and denote their azimuth directions-of-departure (DODs) as $\{\theta_k', \ k = 1, 2, \ldots, K\}$. Since we estimate downlink channels and transmit training symbols from the BS, we treat DODs instead of DOAs. However, the concept is the same as the case where we consider DOAs. We represent a fixed sampling grid that uniformly covers the angular domain $[-\pi/2, \pi/2]$ as $\boldsymbol{\theta} = \{\theta_n, \ n = 1, 2, \ldots, N\}$. That is, we discretize the angular domain into N grid points. Note that each angle θ_n is shown in Fig. 2. If all the K DODs θ_k's are on the grid points, we have

$$\boldsymbol{h} = \boldsymbol{A}\boldsymbol{w}, \tag{24}$$

where $\boldsymbol{A}$ is an $M \times N$ matrix defined as $[\boldsymbol{a}(\theta_1), \boldsymbol{a}(\theta_2), \ldots, \boldsymbol{a}(\theta_N)]$ and $\boldsymbol{a}(\theta_n)$ is given by[†]

$$\boldsymbol{a}(\theta_n) = \left[1, e^{j\pi \sin\theta_n}, \ldots, e^{j\pi(N-1)\sin\theta_n}\right]^T. \tag{25}$$

Also, $\boldsymbol{w}$ is an N-dimensional sparse vector whose non-zero elements are the complex amplitudes of multipath components corresponding to the true directions of $\{\theta_k', \ k = 1, 2, \ldots, K\}$.

Unfortunately, the above assumption that all the K DODs θ_k's are on the grid points does not hold in general. Then, we adopt an off-grid model. If θ_{n_k} is the nearest grid point to θ_k', we write θ_k' as

$$\theta_k' = \theta_{n_k} + \beta_{n_k}, \tag{26}$$

where β_{n_k} is the off-grid gap. From (25) and (26), we have $\boldsymbol{a}(\theta_k') = \boldsymbol{a}(\theta_{n_k} + \beta_{n_k})$. Then, we can rewrite (24) as

$$\boldsymbol{h} = \boldsymbol{A}(\boldsymbol{\beta})\boldsymbol{w}, \tag{27}$$

where

$$\boldsymbol{\beta} = [\beta_1, \beta_2, \ldots, \beta_N]^T, \tag{28}$$

$$\boldsymbol{A}(\boldsymbol{\beta}) = [\boldsymbol{a}(\theta_1 + \beta_1), \boldsymbol{a}(\theta_2 + \beta_2), \ldots, \boldsymbol{a}(\theta_N + \beta_N)] \tag{29}$$

and

$$\beta_{n_k} = \begin{cases} \theta_k' - \theta_{n_k}, & k = 1, 2, \ldots, K \\ 0, & \text{otherwise} \end{cases} \tag{30}$$

Substituting (27) into (23), we have

$$\boldsymbol{v} = \boldsymbol{X}\boldsymbol{A}(\boldsymbol{\beta})\boldsymbol{w} + \boldsymbol{n} = \boldsymbol{\Phi}(\boldsymbol{\beta})\boldsymbol{w} + \boldsymbol{n}, \tag{31}$$

where $\boldsymbol{\Phi}(\boldsymbol{\beta})$ is given as

$$\boldsymbol{\Phi}(\boldsymbol{\beta}) = \boldsymbol{X}\boldsymbol{A}(\boldsymbol{\beta}). \tag{32}$$

Although $\boldsymbol{w}$ is a sparse vector, we cannot apply compressed sensing algorithms to (31). The reason for this is because the observation matrix $\boldsymbol{\Phi}(\boldsymbol{\beta})$ includes the unknowns $\boldsymbol{\beta}$. To jointly obtain $\boldsymbol{w}$ and $\boldsymbol{\beta}$, we apply the SBL algorithm [85], [86] to this problem as below.

It is seen that $\boldsymbol{v}$ is a T-dimensional complex Gaussian random vector with a mean vector $\boldsymbol{\Phi}(\boldsymbol{\beta})\boldsymbol{w}$ and a covariance matrix $\sigma^2\boldsymbol{I}$. Here, $\boldsymbol{I}$ denotes a $T \times T$ identity matrix. Then, we have

[†]In [69], the sign of the phase is opposite to (25). This is due to the difference of definition of the positive direction of angle θ_n. Our definition is shown in Fig. 2.

$$p(\boldsymbol{v}|\boldsymbol{w},\alpha,\boldsymbol{\beta}) = \mathcal{CN}\left(\boldsymbol{v}|\boldsymbol{\Phi}(\boldsymbol{\beta})\boldsymbol{w},\alpha^{-1}\boldsymbol{I}\right), \tag{33}$$

where $\alpha = \sigma^{-2}$ is the noise precision. Since α is unknown, we regard α as a random variable. Given the observed signals $\boldsymbol{v}$, we estimate the random variables $\boldsymbol{w}, \alpha, \boldsymbol{\beta}$ that maximize the posterior $p(\boldsymbol{w}, \alpha, \boldsymbol{\beta}|\boldsymbol{v})$. Note that this idea will be revised later. Since $\boldsymbol{w}, \alpha, \boldsymbol{\beta}$ are independent, from Bayes' theorem, we have

$$p(\boldsymbol{w},\alpha,\boldsymbol{\beta}|\boldsymbol{v}) \propto p(\boldsymbol{v}|\boldsymbol{w},\alpha,\boldsymbol{\beta})p(\alpha)p(\boldsymbol{\beta})p(\boldsymbol{w}). \tag{34}$$

Here, $p(\boldsymbol{v}|\boldsymbol{w},\alpha,\boldsymbol{\beta})$ is the likelihood given by (33). Furthermore, $p(\alpha)$, $p(\boldsymbol{\beta})$, and $p(\boldsymbol{w})$ are priors for α, $\boldsymbol{\beta}$ and $\boldsymbol{w}$, respectively. $p(\alpha)$ is expressed by a Gamma hyperprior $\Gamma(\alpha; 1+c, d)$ [69]†. We fix the hyperparameters c and d to small non-negative values [85]. As for the off-grid gap, we do not have prior information, and we do not know in advance the nearest grid point for each multipath component. Then, $p(\boldsymbol{\beta})$ is given by an N-dimensional uniform distribution $[-\pi/2, \pi/2]$ as

$$p(\boldsymbol{\beta}) = U\left([-\pi/2,\pi/2]^N\right). \tag{35}$$

Next, we consider the prior $p(\boldsymbol{w})$. We assume that $\boldsymbol{w}$ given the variance is an N-dimensional complex Gaussian vector with mean $\boldsymbol{0}$. The elements are independent each other, and we express the precision of the ith component, the inverse of variance, as γ_i. Moreover, we define the N-dimensional vector $\boldsymbol{\gamma}$ by

$$\boldsymbol{\gamma} = [\gamma_1, \gamma_2, \ldots, \gamma_N]^T. \tag{36}$$

Then, we have

$$p(\boldsymbol{w}|\boldsymbol{\gamma}) = \mathcal{CN}\left(\boldsymbol{w}|\boldsymbol{0}, \mathrm{diag}(\boldsymbol{\gamma}^{-1})\right), \tag{37}$$

where $\mathrm{diag}(\boldsymbol{\gamma}^{-1})$ is an $N \times N$ diagonal matrix, and the ith diagonal component is γ_i^{-1}. As in the case of α, we express $p(\boldsymbol{\gamma})$ by independent Gamma hyperpriors, and we have

$$p(\boldsymbol{\gamma}) = \prod_{i=1}^{N} \Gamma(\gamma_i; 1+a, b). \tag{38}$$

From (37) and (38), the prior $p(\boldsymbol{w})$ is given as

$$\begin{aligned} p(\boldsymbol{w}) &= \int_0^\infty p(\boldsymbol{w}|\boldsymbol{\gamma})p(\boldsymbol{\gamma})d\boldsymbol{\gamma} \\ &= \int_0^\infty \mathcal{CN}\left(\boldsymbol{w}|\boldsymbol{0}, \mathrm{diag}(\boldsymbol{\gamma}^{-1})\right) p(\boldsymbol{\gamma})d\boldsymbol{\gamma} \\ &= \prod_{i=1}^{N} (b + |w_i|^2/2)^{-(a+3/2)}. \end{aligned} \tag{39}$$

Here, we omitted detailed derivations for brevity††. From (39), we see the following:

i) The prior of w_i follows the Student's t-distribution.

ii) It has a sharp peak at zero with a small value of b, that is, $p(\boldsymbol{w})$ is a sparse prior.

iii) It has heavy tails, that is, the tails do not decay so fast compared with an exponential distribution. This is preferable because we do not know in advance which value non-zero w_i has.

It is seen from the above, we have obtained all the priors and likelihood for (34). When we estimate the distribution (34), the prior $p(\boldsymbol{w})$ is implemented based on the hierarchical specification via $p(\boldsymbol{w}|\boldsymbol{\gamma})$ and $p(\boldsymbol{\gamma})$. Then, we extend the random variables to be estimated to $\boldsymbol{w}, \boldsymbol{\gamma}, \alpha, \boldsymbol{\beta}$. The posterior $p(\boldsymbol{w}, \boldsymbol{\gamma}, \alpha, \boldsymbol{\beta}|\boldsymbol{v})$ is given by

$$\begin{aligned} p(\boldsymbol{w},\boldsymbol{\gamma},\alpha,\boldsymbol{\beta}|\boldsymbol{v}) &\propto p(\boldsymbol{v}|\boldsymbol{w},\boldsymbol{\gamma},\alpha,\boldsymbol{\beta})p(\boldsymbol{w},\boldsymbol{\gamma},\alpha,\boldsymbol{\beta}) \\ &= p(\boldsymbol{v}|\boldsymbol{w},\alpha,\boldsymbol{\beta})p(\boldsymbol{w}|\boldsymbol{\gamma})p(\boldsymbol{\gamma})p(\alpha)p(\boldsymbol{\beta}) \end{aligned} \tag{40}$$

Note that since $\boldsymbol{v}$ is conditionally independent of $\boldsymbol{\gamma}$ when $\boldsymbol{w}$ is given, we have $p(\boldsymbol{v}|\boldsymbol{w},\boldsymbol{\gamma},\alpha,\boldsymbol{\beta}) = p(\boldsymbol{v}|\boldsymbol{w},\alpha,\boldsymbol{\beta})$. We can estimate $\boldsymbol{w}, \boldsymbol{\gamma}, \alpha, \boldsymbol{\beta}$ using an algorithm such as a Markov chain Monte Carlo or a variational Bayesian analysis [86]. However, in [69], the authors first estimate $\alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}$ with the in-exact block majorization-minimization algorithm. Using the estimated $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, the channel $\boldsymbol{h}$ is obtained as follows:

From (37), we see that w_n given the variance ${\gamma_n}^{-1}$ is a Gaussian random variable with mean 0. So, if γ_n has a large value, w_n is around 0 and we regard w_n as 0. If γ_n is not large, w_n has a non-zero value. Thus, from $\boldsymbol{\gamma}$ we can obtain the support of $\boldsymbol{w}$ denoted by $\boldsymbol{\Omega}$. We express the K-dimensional vector whose components consist of $\boldsymbol{w}$ in $\boldsymbol{\Omega}$ as $\boldsymbol{w}_{\boldsymbol{\Omega}}$. Next, we represent the matrix deleting the columns except $\boldsymbol{\Omega}$ from $\Phi(\boldsymbol{\beta})$ as $\Phi_\Omega(\boldsymbol{\beta})$. Then, from (31), we have

$$\boldsymbol{v} = \Phi_\Omega(\boldsymbol{\beta})\boldsymbol{w}_\Omega + \boldsymbol{n}. \tag{41}$$

Similarly, we represent the matrix deleting the columns except Ω from $\boldsymbol{A}(\boldsymbol{\beta})$ as $A_\Omega(\boldsymbol{\beta})$. We see from (27) that

$$\boldsymbol{h} = \boldsymbol{A}_\Omega(\boldsymbol{\beta})\boldsymbol{w}_\Omega, \tag{42}$$

holds.

If the estimation of $\alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}$ is accurate, $\Phi_\Omega(\boldsymbol{\beta})$ is a $T \times K$ matrix. Since the channel is sparse, K is so small that we have $T > K$. Thus, from (41), we can obtain

$$\boldsymbol{w}_\Omega = \Phi_\Omega(\boldsymbol{\beta})^+\boldsymbol{v}, \tag{43}$$

where $\Phi_\Omega(\boldsymbol{\beta})^+$ denotes the Moore-Penrose generalized inverse matrix of $\Phi_\Omega(\boldsymbol{\beta})$. Substituting (43) into (42), the downlink channel is estimated as

$$\boldsymbol{h}^e = \boldsymbol{A}_\Omega(\boldsymbol{\beta})\Phi_\Omega(\boldsymbol{\beta})^+\boldsymbol{v}. \tag{44}$$

Here, we show an example of the downlink channel estimation performance. We define the normalized mean square error (NMSE) as

†The Gamma hyperprior is usually given by $\Gamma(\alpha; c, d)$ as in [85]. However, for $c > 0$, we have $\Gamma(0; 1+c, d) = 0$, which means the probability density that $\alpha = 0$ or $\sigma^2 = \infty$ holds is 0. This matches physical phenomena, and we use the expression $p(\alpha) = \Gamma(\alpha; 1+c, d)$ in this paper. This is the same also for (38).

††(39) is slightly different from (22) in [69]. It seems that the authors of [69] regarded $2\gamma_i$ in the Gamma distribution as γ_i, which does not cause any contradiction.

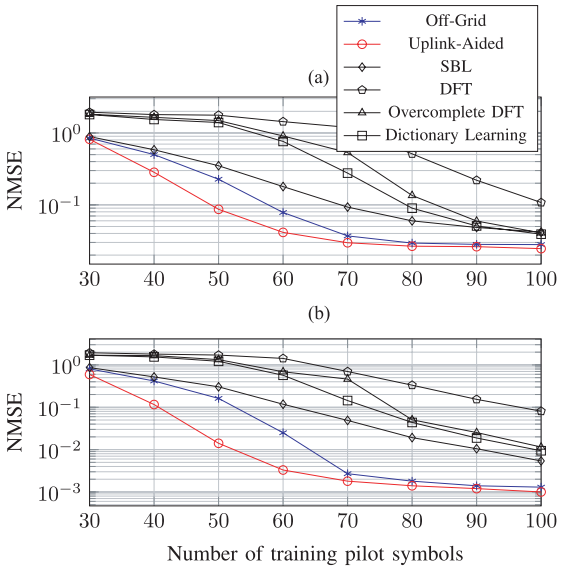


Fig. 6 NMSE of dowlink channel estimation versus the number of training pilot symbols. (a) SNR = 0 dB; (b) SNR = 10 dB. Reprinted from [69] with permission (©2018 IEEE).

$$\frac{1}{M_c}\sum_{m=1}^{M_c}\frac{\|\boldsymbol{h}_m^e-\boldsymbol{h}_m\|_2^2}{\|\boldsymbol{h}_m\|_2^2}, \tag{45}$$

where $\boldsymbol{h}_m^e$ is the estimated value of $\boldsymbol{h}_m$ at the mth Monte Carlo trial, and M_c is the number of trials. Figure 6 shows NMSE versus number of training pilot symbols T for the ULA with $M = 150$ antennas and for $M_c = 200$.

The graph legends are as follows:

Off-Grid Proposed method stated above

Uplink-Aided Improved version of the proposed method using both of uplink and downlink signals

The others below ignore the off-grid gaps.

SBL $\boldsymbol{h}$ is estimated using the standard SBL [85]

DFT $\boldsymbol{h}$ is estimated using the ℓ_1-norm minimization [18], [19] with a DFT basis

Overcomplete DFT $\boldsymbol{h}$ is estimated using the ℓ_1-norm minimization [18], [19] with the matrix $\boldsymbol{A}$ in (24)

Dictionary Learning $\boldsymbol{h}$ is estimated using the method proposed in [87] with the matrix $\boldsymbol{A}$ defined in (24)

The number of grid points N is 200 for all except for the DFT method. Refer to [69] for the other simulation parameters.

From Fig. 6, we see that as the number of training pilot symbols increases, the channel estimation performance tends to be better. The off-grid technique outperforms the methods that ignore the off-grid gaps. Although the performance of the off-grid technique is worse than that of the uplink-aided one, the difference is very small when the number of the training pilot symbols is 70 or more. Since the BS has 150 antennas, conventional channel estimation techniques such as least squares need 150 or more pilot symbols. From this, we can see that the SBL-based off-grid estimation technique can reduce the training pilot symbols. However, the training overhead is still large, and there is room for improvement.

5. Conclusions

We reviewed sparse modeling in wireless techniques. Beginning with the basics of array signal processing, we introduced DOA estimation algorithms based on compressed sensing. After that, we considered two MIMO CSI acquisitions using the DOA estimation. One is channel prediction in time-varying environments. It was shown that the degradation of MIMO transmission performance is small. The other is downlink SBL-based channel estimation in an FDD massive MIMO system. The estimation technique can decrease the number of training pilot symbols. Note that SBL can be applied also to other radio research fields such as wireless sensor networks [88], [89].

On the other hand, terahertz (THz)-band communication is envisioned as one of the key wireless technologies in 6G systems. Since the wavelength is extremely short, we can integrate a very large number of antennas in very small footprints to increase the communication distance, which enables ultra-massive MIMO communications [90], [91]. In ultra-massive MIMO systems, CSI acquisition or beam selection may be a challenging issue. Also, the Doppler frequency becomes high in THz bands. Consequently, channels rapidly change. The sparse modeling is envisaged to play an important role in these situations.

At the end of this paper, we state a problem of a sparse modeling technique in wireless engineering. According to propagation measurements, there are a few clusters in environments, but each cluster may scatter a number of subpaths [92]. This phenomenon may violate the condition for the sparse signal recovery. How we avoid the problem is a critical question in this research field.

References

[1] F. Boccardi, R.W. Heath, Jr., A. Lozano, T.L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," IEEE Commun. Mag., vol.52, no.2, pp.74–80, Feb.. 2014. DOI: 10.1109/MCOM.2014.6736746

[2] J.G. Andrews, S. Buzzi, W. Choi, S.V. Hanly, A. Lozano, A.C. K. Soong, and J.C. Zhang, "What will 5G be?," IEEE J. Sel. Areas Commun., vol.32, no.6, pp.1065–1082, June 2014. DOI: 10.1109/JSAC.2014.2328098

[3] Y. Okumura, S. Suyama, J. Mashino, and K. Muraoka, "Recent activities of 5G experimental trials on massive MIMO technologies and 5G system trials toward new services creation," IEICE Trans. Commun., vol.E102-B, no.8, pp.1352–1362, Aug. 2019. DOI: 10.1587/transcom.2018TTI0002

[4] T. Okuyama, S. Suyama, J. Mashino, K. Muraoka, K. Izui, K. Yamazaki, and Y. Okumura, "Performance evaluation of low complexity digital beamforming algorithms by link-level simulations and outdoor experimental trials for 5G low-SHF-band massive MIMO," IEICE Trans. Commun., vol.E102-B, no.8, pp.1382–1389, Aug. 2019. DOI: 10.1587/transcom.2018TTP0022

[5] T. Kobayashi, M. Tsutsui, T. Dateki, H. Seki, M. Minowa, C. Akiyama, T. Okuyama, J. Mashino, S. Suyama, and Y. Okumura, "Experimental study of large-scale coordinated multi-user MIMO for 5G ultra high-density distributed antenna systems," IEICE Trans. Commun., vol.E102-B, no.8, pp.1390–1400, Aug. 2019. DOI:

10.1587/transcom.2018TTP0012

[6] M. Mikami and H. Yoshino, "Field trial on 5G low latency radio communication system towards application to truck platooning," IEICE Trans. Commun., vol.E102-B, no.8, pp.1447–1457, Aug. 2019. DOI: 10.1587/transcom.2018TTP0021

[7] K.B. Letaief, W. Chen, Y. Shi, J. Zhang, and Y.-J.A. Zhang, "The roadmap to 6G: AI empowered wireless networks," IEEE Commun. Mag., vol.7, no.8, pp.84–90, Aug. 2019. DOI: 10.1109/MCOM.2019.1900271

[8] M. Giordani, M. Polese, M. Mezzavilla, S. Rangan, and M. Zorzi, "Toward 6G networks: Use cases and technologies," IEEE Commun. Mag., vol.58, no.3, pp.55–61, March 2020. DOI: 10.1109/MCOM.001.1900411

[9] E. Telatar, "Capacity of multi-antenna Gaussian channels," European Trans. Telecommn., vol.10, no.6, pp.585–595, Nov./Dec. 1999. DOI: 10.1002/ett.4460100604

[10] E. Biglieri, R. Caldebank, A. Constantinides, A. Goldsmith, A. Paulraj, and H.V. Poor, MIMO Wireless Communications, Cambridge University Press, New York, 2007.

[11] T. Ohgane, T. Nishimura, and Y. Ogawa, "Applications of space division multiplexing and those performance in a MIMO channel," IEICE Trans. Commun., vol.E88-B, no.5, pp.1843–1851, May 2005. DOI: 10.1093/ietcom/e88-b.5.1843

[12] Y.-C. Lin, T.-S. Lee, Y.-H. Pan, and K.-H. Lin, "Low-complexity high-resolution parameter estimation for automotive MIMO radars," IEEE Access, vol.8, pp.16127–16138, Jan. 2020. DOI: 10.1109/ACCESS.2019.2926413

[13] Z. Qiu, Q. Zhang, M. Sun, and J. Zhu, "Parameter estimation for multiple chirp signals based on single channel Nyquist folding receiver," IEICE Trans. Fundamentals, vol.E103-A, no.3, pp.623–628, March 2020. DOI: 10.1587/transfun.2019EAL2097

[14] K. Jimi, I. Matsunami, and R. Nakamura, "Improvement of ranging accuracy during interference avoidance for stepped FM radar using Khatri-Rao product extended-phase processing," IEICE Trans. Commun., vol.E102-B, no.1, pp.156–164, Jan. 2019. DOI: 10.1587/transcom.2018EBP3042

[15] W. Osamy, A.A. El-Sawy, and A. Salim, "CSOCA: Chicken swarm optimization based clustering algorithm for wireless sensor networks," IEEE Access, vol.8, pp.60676–60688, March 2020. DOI: 10.1109/ACCESS.2020.2983483

[16] H. Asahina, K. Toyoda, P.T. Mathiopoulos, I. Sasase, and H. Yamamoto, "RPL-based tree construction scheme for target-specific code dissemination in wireless sensors networks," IEICE Trans. Commun., vol.E103-B, no.3, pp.190–199, March 2020. DOI: 10.1587/transcom.2019EBP3066

[17] D.L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ^1 minimization," Proc. Natil. Acad. Sci. USA., vol.100, no.5, pp.2197–2202, March 2003. DOI: 10.1073/pnas.0437847100

[18] D.L. Donoho, "Compressed sensing," IEEE Trans. Inf. Theory, vol.52, no.4, pp.1289–1306, Apr. 2006. DOI: 10.1109/TIT.2006.871582

[19] E.J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," IEEE Trans. Inf. Theory, vol.52, no.2, pp.489–509, Feb. 2006. DOI: 10.1109/TIT.2005.862083

[20] E.J. Candès and T. Tao, "Near optimal signal recovery from random projections: Universal encoding strategies?," IEEE Trans. Inf. Theory, vol.52, no.12, pp.5406–5425, Dec. 2006. DOI: 10.1109/TIT.2006.885507

[21] D.L. Donoho, "For most large underdetermined systems of linear equations the minimal ℓ_1-norm solution is also the sparsest solution," Comm. Pure Appl. Math., vol.56, no.6, pp.797–829, June 2006. DOI: 10.1002/cpa.20132

[22] J.A. Tropp, "Greed is good: Algorithmic results for sparse approximation," IEEE Trans. Inf. Theory, vol.50, no.10, pp.2231–2242, Oct. 2004. DOI: 10.1109/TIT.2004.834793

[23] J.A. Tropp, A.C. Gilbert, and M.J. Strauss, "Algorithms for simultaneous sparse approximation: Part I: Greedy pursuit," Signal Process., vol.86, no.3, pp.572–588, March 2006. DOI: 10.1016/j.sigpro.2005.05.030

[24] Special Issue on Compressive Sampling, IEEE Signal Process. Mag., vol.25, no.2, March 2008.

[25] Special Issue on Applications of Sparse Representation & Compressive Sensing, Proc. IEEE, vol.98, no.6, June 2010.

[26] M. Nagahara, T. Matusda, and K. Hayashi, "Compressive sampling for remote control systems," Trans. IEICE Fundamentals, vol.E95-A, no.4, pp.713–722, April 2012. DOI: 10.1587/transfun.E95.A.713

[27] K. Hayashi, M. Nagahara, and T. Tanaka, "A user's guide to compressed sensing for communications systems," IEICE Trans. Commun., vol.E96-B, no.3, pp.685–712, March 2013. DOI: 10.1587/transcom.E96.B.685

[28] J. Otsuki, M. Ohzeki, H. Shinaoka, and K. Yoshimi, "Sparse modeling in quantum many-body problems," J. Phys. Soc. Jpn. 89, vol.89, no.1, 012001, 2020. DOI: 10.7566/JPSJ.89.012001

[29] X. Zhang, W. Xu, Y. Tian, J. Lin, and W. Xu, "Improved analysis for SOMP algorithm in terms of restricted isometry property," IEICE Trans. Fundamentals, vol.E103-A, no.2, pp.533–537, Feb. 2020. DOI: 10.1587/transfun.2019EAL2055

[30] W. Zhang and F. Yu, "Parameterized L1-minimization algorithm for off-the-gird spectral compressive sensing," IEICE Trans. Fundamentals, vol.E100-A, no.9, pp.2026–2030, Sept. 2017. DOI: 10.1587/transfun.E100.A.2026

[31] M. Lustig, D.L. Donoho, and J.M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," Magn. Resonance Med., vol.58, no.6, pp.1182–1195, Dec. 2007. DOI: 10.1002/mrm.21391

[32] M. Lustig, D.L. Donoho, J.M. Santos, and J.M. Pauly, "Compressed sensing MRI," IEEE Signal Process. Mag., vol.25, no.2, pp.72–82, March 2008. DOI: 10.1109/MSP.2007.914728

[33] M. Honma, K. Akiyama, M. Uemura, and S. Ikeda, "Super-resolution imaging with radio interferometry using sparse modeling," Publ. Astron. Soc. Japan, vol.66, no.5, pp.95(1–14), Oct. 2014. DOI: 10.1093/pasj/psu070

[34] K. Akiyama, A. Alberdi, W. Alef, et al., "First M87 event horizon telescope results. I. — The shadow of the supermassive black hole —," Astrophys. J. Lett., vol.875, no.1, pp.1–17, April 2019. DOI: 10.3847/2041-8213/ab0ec7

[35] K. Akiyama, A. Alberdi, W. Alef, et al., "First M87 event horizon telescope results. II. — Array and instrumentation —," Astrophys. J. Lett., vol.875, no.2, pp.1–28, April 2019. DOI: 10.3847/2041-8213/ab0c96

[36] K. Akiyama, A. Alberdi, W. Alef, et al., "First M87 event horizon telescope results. III. — Data processing and calibration —," Astrophys. J. Lett., vol.875, no.3, pp.1–32, April 2019. DOI: 10.3847/2041-8213/ab0c57

[37] K. Akiyama, A. Alberdi, W. Alef, et al., "First M87 event horizon telescope results. IV. — Imaging the central supermassive black hole —," Astrophys. J. Lett., vol.875, no.4, pp.1–52, April 2019. DOI: 10.3847/2041-8213/ab0e85

[38] K. Akiyama, A. Alberdi, W. Alef, et al., "First M87 event horizon telescope results. V. — Physical origin of the asymmetric ring —," Astrophys. J. Lett., vol.875, no.5, pp.1–31, April 2019. DOI: 10.3847/2041-8213/ab0f43

[39] K. Akiyama, A. Alberdi, W. Alef, et al., "First M87 event horizon telescope results. VI. — The shadow and mass of the central black hole —," Astrophys. J. Lett., vol.875, no.6, pp.1–44, April 2019. DOI: 10.3847/2041-8213/ab1141

[40] N. Czink, X. Yin, H. Ozcelik, M. Herdin, E. Bonek, and B.H. Fleury, "Cluster characteristics in a MIMO indoor propagation environment," IEEE Trans. Wireless Commun., vol.6, no.4, pp.1465–1475, April 2007. DOI: 10.1109/TWC.2007.348343

[41] L. Vuokko, V.-M. Kolmonen, J. Salo, and P. Vainikainen, "Measurement of large-scale cluster power characteristics for geometric chan-

nel models," IEEE Trans. Antennas Propag., vol.55, no.11, pp.3361–3365, Nov. 2007. DOI: 10.1109/TWC.2007.348343
[42] W.U. Bajwa, J. Haupt, A.M. Sayeed, and R. Nowak, "Compressed channel sensing: A new approach to estimating sparse multipath channels," Proc. IEEE, vol.98, no.6, pp.1058–1076, June 2010. DOI: 10.1109/JPROC.2010.2042415
[43] Y. Liu, W. Mei, and H. Du, "Model-based compressive channel estimation over rapidly time-varying channels in OFDM systems," IEICE Trans. Commun., vol.E97-B, no.8, pp.1709–1716, Aug. 2014. DOI: 10.1587/transcom.E97.B.1709
[44] Y. Liu, W. Mei, and H. Du, "Compressive channel estimation using distribution agnostic Bayesian method," IEICE Trans. Commun., vol.E98-B, no.8, pp.1672–1679, Aug. 2015. DOI: 10.1587/transcom.E98.B.1672
[45] Z. Gao, L. Dai, Z. Wang, and S. Chen, "Spatially common sparsity based adaptive channel estimation and feedback for FDD massive MIMO," IEEE Trans. Signal Process., vol.63, no.23, pp.6169–6183, Dec. 2015. DOI: 10.1109/TSP.2015.2463260
[46] L. Xiao, Z. Liang, and K. Liu, "A novel compressed sensing-based channel estimation method for OFDM system," IEICE Trans. Fundamentals, vol.E100-A, no.1, pp.322–326, Jan. 2017. DOI: 10.1587/transfun.E100.A.322
[47] X. Liu, X. Chen, L. Xue, and Z. Xie, "Channel estimation of OQAM/OFDM based on compressed sensing," IEICE Trans. Commun., vol.E100-B, no.6, pp.955–961, June 2017. DOI: 10.1587/transcom.2016EBP3280
[48] Q. Qin, L. Gui, B. Gong, and S. Luo, "Sparse channel estimation for massive MIMO-OFDM systems over time-varying channels," IEEE Access, vol.6, pp.33740–33751, June 2018. DOI: 10.1109/ACCESS.2018.2843783
[49] K.-H. Kim, "A low-complexity path delay searching method in sparse channel estimation for OFDM systems," IEICE Trans. Commun., vol.E101-B, no.11, pp.2297–2303, Nov. 2018. DOI: 10.1587/transcom.2018EBP3026
[50] F. Kulsoom, A. Vizziello, H.N. Chaudhry, and P. Savazzi, "Joint sparse channel recovery with quantized feedback for multi-user massive MIMO systems," IEEE Access, vol.8, pp.11046–11060, Jan. 2020. DOI: 10.1109/ACCESS.2020.2965280
[51] J. Capon, "High-resolution frequency-wave number spectrum analysis," Proc. IEEE, vol.57, no.8, pp.1408–1418, Aug. 1969. DOI: 10.1109/PROC.1969.7278
[52] H. Krim and M. Viberg, "Two decades of array signal processing research: The parametric approach," IEEE Signal Process. Mag., vol.13, no.4, pp.67–94, July 1996. DOI: 10.1109/79.526899
[53] W. Jhang, S.W. Chen, and A.C. Chang, "Efficient hybrid DOA estimation for massive uniform linear array," IEICE Trans. Fundamentals, vol.E102-A, no.5, pp.721–724, May 2019. DOI: 10.1587/transfun.E102.A.721
[54] W. Jhang, S.W. Chen, and A.C. Chang, "Computationally efficient DOA estimation for massive uniform linear array," IEICE Trans. Fundamentals, vol.E103-A, no.1, pp.361–365, Jan. 2020. DOI: 10.1587/transfun.2019EAL2107
[55] Q. Tian, T. Qiu, R. Cai, and J. Ma, "A novel DOA estimation for distributed sources in an impulsive noise environment," IEEE Access, vol.8, pp.61405–61420, March 2020. DOI: 10.1109/ACCESS.2020.2983046
[56] D. Malioutov, M. Çetin, and A.S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," IEEE Trans. Signal Process., vol.53, no.8, pp.3010–3022, Aug. 2005. DOI: 10.1109/TSP.2005.850882
[57] P. Stoica, P. Babu, and J. Li, "SPICE: A sparse covariance-based estimation method for array processing," IEEE Trans. Signal Process., vol.59, no.2, pp.629–638, Feb. 2011. DOI: 10.1109/TSP.2010.2090525
[58] Z.M. Liu, Z.T. Huang, and Y.Y. Zhou, "Direction-of-arrival estimation of wideband signals via covariance matrix sparse representation," IEEE Trans. Signal Process., vol.59, no.9, pp.4256–4270, Sept. 2011. DOI: 10.1109/TSP.2011.2159214
[59] K. Lee, Y. Bresler, and M. Junge, "Subspace methods for joint sparse recovery," IEEE Trans. Inf. Theory, vol.58, no.6, pp.3613–3641, June 2012. DOI: 10.1109/TIT.2012.2189196
[60] M. Rossi, A.M. Haimovich, and Y.C. Eldar, "Spatial compressive sensing for MIMO radar," IEEE Trans. Signal Process., vol.62, no.2, pp.419–430, Jan. 2014. DOI: 10.1109/TSP.2013.2289875
[61] Q. Shen, W. Liu, W. Cui, and S. Wu, "Underdetermined DOA estimation under the compressive sensing framework: A review," IEEE Access, vol.4, pp.8865–8878, Nov. 2016. DOI: 10.1109/ACCESS.2016.2628869
[62] J.M. Kim, O.K. Lee, and J.C. Ye, "Compressive MUSIC: Revisiting the link between compressive sensing and array signal processing," IEEE Trans. Inf. Theory, vol.58, no.1, pp.278–301, Jan. 2012. DOI: 10.1109/TIT.2011.2171529
[63] S. Uehashi, Y. Ogawa, T. Nishimura, and T. Ohgane, "Prediction of time-varying multi-user MIMO channels based on DOA estimation using compressed sensing," IEEE Trans. Veh. Technol., vol.68, no.1, pp.565–577, Jan. 2019. DOI: 10.1109/TVT.2018.2882214
[64] F. Rusek, D. Persson, B.K. Lau, E.G. Larsson, T.L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," IEEE Signal Process. Mag., vol.30, no.1, pp.40–60, Jan. 2013. DOI: 10.1109/MSP.2011.2178495
[65] E.G. Larsson, O. Edfors, F. Tufvesson, and T.L. Marzetta, "Massive MIMO for next generation wireless systems," IEEE Commun. Mag., vol.52, no.2, pp.186–195, Feb. 2014. DOI: 10.1109/MCOM.2014.6736761
[66] A. Adhikary, E.A. Safadi, M.K. Samimi, R. Wang, G. Caire, T.S. Rappaport, and A.F. Molisch, "Joint spatial division and multiplexing for mm-wave channels," IEEE J. Sel. Areas Commun., vol.32, no.6, pp.1239–1255, June 2014. DOI: 10.1109/JSAC.2014.2328173
[67] E. Björnson, E.G. Larsson, and T.L. Marzetta, "Massive MIMO: Ten myths and one critical question," IEEE Commun. Mag., vol.54, no.2, pp.114–123, Feb. 2016. DOI: 10.1109/MCOM.2016.7402270
[68] F. Liu, X. He, C. Li, and Y. Xu, "CsiNet-plus model with truncation and noise on CSI feedback," IEICE Trans. Fundamentals, vol.E103-A, no.1, pp.376–381, Jan. 2020. DOI: 10.1587/transfun.2019EAL2123
[69] J. Dai, A. Liu, and V.K. N. Lau, "FDD massive MIMO channel estimation with arbitrary 2D-array geometry," IEEE Trans. Signal Process., vol.66, no.10, pp.2584–2599, May 2018. DOI: 10.1109/TSP.2018.2807390
[70] M. Elad, Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing, Springer, 2010.
[71] I. Rish and G. Grabarnik, Sparse Modeling: Theory, Algorithms, and Applications, CRC Press, 2014.
[72] L.C. Godara, "Application of antenna arrays to mobile communications, Part II: Beam-forming and direction-of-arrival considerations," Proc. IEEE, vol.85, no.8, pp.1195–1245, Aug. 1997. DOI: 10.1109/5.622504
[73] E.J. Candes, M.B. Wakin, "An introduction to compressive sampling," IEEE Signal Process. Mag., vol.25, no.2, pp.21–30, March 2008. DOI: 10.1109/MSP.2007.914731
[74] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," SIAM Journal on Imaging Sciences., vol.2, no.1, pp.183–202, March 2009. DOI: 10.1137/080716542
[75] Y.C. Pati, R. Rezaiifar, P.S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with application to wavelet decomposition," Proc. 27th Asilomar Conf. Sign. Syst. Comput., vol.1, pp.40–44, Nov. 1993. DOI: 10.1109/ACSSC.1993.342465
[76] M. Çetin, D.M. Malioutov, and A.S. Willsky, "A variational technique for source localization based on a sparse signal reconstruction perspective," Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., vol.3, pp.2965–2968, May 2002. DOI: 10.1109/

ICASSP.2002.5745271
[77] T. Nishimura, T. Endo, T. Ohgane, and Y. Ogawa, "Parameter settings on DOA estimation of multi-band signals using a compressed sensing technique," Proc. APSIPA ASC 2016, Dec. 2016.
[78] A. Duel-Hallen, "Fading channel prediction for mobile radio adaptive transmission systems," Proc. IEEE, vol.95, no.12, pp.2299–2313, Dec. 2007. DOI: 10.1109/JPROC.2007.904443
[79] M.R. Akdeniz, Y. Liu, M.K. Samimi, S. Sun, S. Rangan, T.S. Rappaport, and E. Erkip, "Millimeter wave channel modeling and cellular capacity evaluation," IEEE J. Sel. Areas Commun., vol.32, no.6, pp.1164–1179, June 2014. DOI: 10.1109/JSAC.2014.2328154
[80] W.K. Ma, T.H. Hsieh, and C.Y. 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. DOI: 10.1109/TSP.2009.2034935
[81] H. Zhu, G. Leus, and G. Giannakis, "Sparsity-cognizant total least-squares for perturbed compressive sampling," IEEE Trans. Signal Process., vol.59, no.5, pp.2002–2016, May 2011. DOI: 10.1109/TSP.2011.2109956
[82] Z. Yang, L. Xie, and C. Zhang, "Off-grid direction of arrival estimation using sparse Bayesian inference," IEEE Trans. Signal Process., vol.61, no.1, pp.38–43, Jan. 2013. DOI: 10.1109/TSP.2012.2222378
[83] L. Liu, P. Wei, and H.S. Liao, "Off-grid DOA estimation based on analysis of the convexity of maximum likelihood function," IEICE Trans. Fundamentals, vol.E98-A, no.12, pp.2705–2708, Dec. 2015. DOI: 10.1587/transfun.E98.A.2705
[84] Z. Yang and L. Xie, "Enhancing sparsity and resolution via reweighted atomic norm minimization," IEEE Trans. Signal Process., vol.64, no.4, pp.995–1006, Feb. 2016. DOI: 10.1109/TSP.2015.2493987
[85] M.E. Tipping, "Sparse Bayesian learning and the relevance vector machine," J. Mach. Learn. Res., vol.1, no.3, pp.211–244, June 2001.
[86] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," IEEE Trans. Signal Process., vol.56, no.6, pp.2346–2356, June. 2008. DOI: 10.1109/TSP.2007.914345
[87] Y. Ding and B.D. Rao, "Dictionary learning-based sparse channel representation and estimation for FDD massive MIMO systems," IEEE Trans. Wireless Commun., vol.17, no.8, pp.5437– 5451, Aug. 2018. DOI: 10.1109/TWC.2018.2843786
[88] B. Xue, L. Zhang, and Y. Yu, "Multi-target localization based on sparse Bayesian learning in wireless sensor networks," IEICE Trans. Commun., vol.E99-B, no.5, pp.1093–1110, May 2016. DOI: 10.1587/transcom.2015EBP3250
[89] Y. Guo, P. Qian, and N. Li, "Several bits are enough: Off-grid target localization in WSNs using variational Bayesian EM algorithm," IEICE Trans. Fundamentals, vol.E102-A, no.7, pp.926–929, July 2019. DOI: 10.1587/transfun.E102.A.926
[90] C. Han, J.M. Jornet, and I.F. Akyildiz, "Ultra-massive MIMO channel modeling for graphene-enabled terahertz-band communications," Proc. IEEE Veh. Technol. Conf., June 2018. DOI: 10.1109/VTCSpring.2018.8417893
[91] A. Faisal, H. Sarieddeen, H. Dahrouj, T.Y. Al-Naffouri, and M.-S. Alouini, "Ultra-massive MIMO systems at terahertz bands: Prospects and challenges," arXiv: 1902.11090, 2019.
[92] M.K. Samimi and T.S. Rappaport, "3-D millimeter-wave statistical channel model for 5G wireless system design," IEEE Trans. Microw. Theory Tech., vol.64, no.7, pp.2207–2225, July 2016. DOI: 10.1109/TMTT.2016.2574851



Toshihiko Nishimura received the B.S. and M.S. degrees in physics and Ph.D. degree in electronics engineering from Hokkaido University, Sapporo, Japan, in 1992, 1994, and 1997, respectively. Since 1998, he has been with Hokkaido University, where he is currently an Associate Professor. His current research interests are in MIMO systems using smart antenna techniques. He received the Young Researchers' Award of IEICE in 2000, the Best Paper Award from IEICE in 2007, and TELECOM System Technology Award from the Telecommunications Advancement Foundation of Japan in 2008, the best magazine paper award from IEICE Communications Society in 2011, and the Best Tutorial Paper Award from the IEICE Communications Society in 2018. He is a member of the IEEE.



Yasutaka Ogawa received the B.E., M.E., and Ph.D. degrees from Hokkaido University, Sapporo, Japan, in 1973, 1975, and 1978, respectively. Since 1979, he has been with Hokkaido University, where he is currently a Professor Emeritus. During 1992–1993, he was with ElectroScience Laboratory, the Ohio State University, as a Visiting Scholar, on leave from Hokkaido University. His professional expertise encompasses super-resolution estimation techniques, applications of adaptive antennas for mobile communication, multiple-input multiple-output (MIMO) techniques, and measurement techniques. He proposed a basic and important technique for time-domain super-resolution estimation for electromagnetic wave measurement such as antenna gain measurement, scattering/diffraction measurement, and radar imaging. Also, his expertise and commitment to advancing the development of adaptive antennas contributed to the realization of space division multiple accesses (SDMA) in the Personal Handy-phone System (PHS). He received the Yasujiro Niwa Outstanding Paper Award in 1978, the Young Researchers' Award of IEICE in 1982, the Best Paper Award from IEICE in 2007, TELECOM system technology award from the Telecommunications Advancement Foundation of Japan in 2008, the Best Magazine Paper Award from IEICE Communications Society in 2011, the Achievement Award from IEICE in 2014, and the Best Tutorial Paper Award from IEICE Communications Society in 2018. He also received the Hokkaido University Commendation for excellent teaching in 2012. He is a Life Fellow of the IEEE.



Takeo Ohgane received the B.E., M.E., and Ph.D. degrees in electronics engineering from Hokkaido University, Sapporo, Japan, in 1984, 1986, and 1994, respectively. From 1986 to 1992, he was with Communications Research Laboratory, Ministry of Posts and Telecommunications. From 1992 to 1995, he was on assignment at ATR Optical and Radio Communications Research Laboratory. Since 1995, he has been with Hokkaido University, where he is currently a Professor. During 2005–2006, he was at Centre for Communications Research, University of Bristol, U.K., as a Visiting Fellow. His research interests are in MIMO signal processing for wireless communications. He received the IEEE AP-S Tokyo Chapter Young Engineer Award in 1993, the Young Researchers' Award of IEICE in 1990, the Best Paper Award from IEICE in 2007, TELECOM System Technology Award from the Telecommunications Advancement Foundation of Japan in 2008, the Best Magazine Paper Award from IEICE Communications Society in 2011, and the Best Tutorial Paper Award from IEICE Communications Society in 2018. He is a member of the IEEE.



Junichiro Hagiwara received the B.E., M.E., and Ph.D. degrees from Hokkaido University, Sapporo, Japan, in 1990, 1992, and 2016, respectively. He joined the Nippon Telegraph and Telephone Corporation in April 1992 and transferred to NTT Mobile Communications Network, Inc. (currently NTT DOCOMO, INC.) in July 1992. Later, he became involved in the research and development of mobile communication systems. His current research interests are in the application of stochastic theory to the communication domain. He is currently a visiting professor at Hokkaido University.