

RESEARCH

Open Access



GAMP-SBL-based channel estimation for millimeter-wave MIMO systems

Jianfeng Shao^{*}, Xianpeng Wang, Xiang Lan, Zhiguang Han and Ting Su

^{*}Correspondence:
shaojianfeng@hainanu.
edu.cn
College of Information
and Communication
Engineering, University
of Hainan, Haikou, China

Abstract

Based on the finite scattering characters of the millimeter-wave multiple-input multiple-output (MIMO) channel, the mmWave channel estimation problem can be considered as a sparse signal recovery problem. However, most traditional channel estimation methods depend on grid search, which may lead to considerable precision loss. To improve the channel estimation accuracy, we propose a high-precision two-stage millimeter-wave MIMO system channel estimation algorithm. Since the traditional expectation-maximization-based sparse Bayesian learning algorithm can be applied to handle this problem, it spends lots of time to calculate the E-step which needs to compute the inversion of a high-dimensional matrix. To avoid the high computation of matrix inversion, we combine damp generalized approximate message passing with the E-step in SBL. We then improve a refined algorithm to handle the dictionary matrix mismatching problem in sparse representation. Numerical simulations show that the estimation time of the proposed algorithm is greatly reduced compared with the traditional SBL algorithm and better estimation performance is obtained at the same time.

Keywords: Channel estimation, Millimeter-wave multiple-input multiple-output (MIMO), Sparse Bayesian learning (SBL), DGAMP

1 Introduction

In the 5G cellular communication, millimeter wave (mmWave) has been drawing enormous attention from academia, colleges and governments due to the wide available spectrum [1, 2]. However, challenges also exist in the application of mmWave [3, 4]. With frequency increasing, various attenuation losses of the mmWave channel, for example, the path loss and showing effects, become more hostile. One solution to the power loss problem in the channel is to use large MIMO antennas. For both the base station and the mobile station, large antennas are equipped to get transceiver powerful communication. However, another challenge is that the conventional training overhead for the channel state information (CSI) acquisition grows proportionally with the BS and MS antenna size. Fortunately, the massive MIMO system has a high correlation due to the finite number of scatters in the propagation path, so that the effective dimension is much less than the actual dimension [5]. Hence, with the uniform linear array (ULA) assumption, the mmWave channel can be approximately transformed into the sparse representation under the discrete Fourier transform (DFT) basis when large-scale antennas are equipped [6–8]. Compressive sensing (CS)

theory was developed by Candes et al. [9] and Donoho [10] which can be very useful when signals are sparse or compressible. Several algorithms have been presented in [5, 11–13] to exploit such hidden sparsity. The above methods contribute to achieving a significant reduction in the training overhead. However, there are still some limits in the algorithms. For example, in the expectation–maximization (EM) sparse Bayesian learning algorithm, the computation complexity increases exponentially with the signal dimension [14], which may meet great challenges in practical implementations. Another generally existed problem is that although the sparse algorithms can achieve super estimation performance for sparse signals, there also exists a grid mismatch issue. Some super-resolution (off-grid) compressed sensing methods and two-stage channel estimation algorithms have been applied to improve channel estimation accuracy [14–20] aiming to overcome the grid mismatch caused by conventional compressed sensing techniques. However, the super-resolution dictionary learning algorithm has two main shortcomings: the convergence of the results cannot be guaranteed; a large amount of prior training is required for the sparse dictionary. These two shortcomings limit the practical design and implementation.

In this paper, we adopt a two-step algorithm. The first step is coarse channel estimate by the SBL algorithm, which simplifies operations using DGAMP. The second step is an accurate angle estimate by the interpolation of three 2-DFT spectral lines. The following summarizes the contributions of this paper.

- We formulate a sparse recovery problem and develop a DGAMP-SBL algorithm for the coarse estimation channel of the hybrid MIMO system. Based on the algorithm, we accelerate the computation speed of the original EM sparse Bayesian learning algorithm for mmWave channel estimation. The computation complexity shows that we significantly reduce operation time comparing to the original SBL algorithm with large-scale transmission and reception antennas.
- To address the sparse reconstruction based channel estimation schemes suffering from the grid-off problem, we improve a fast refined algorithm without iteration. In the discrete Fourier transform domain, the influence of Gaussian noise decreases with the increasing antenna number, which helps obtain a more accurate angle estimation.

The rest of the paper is organized as follows. In Sect. 2, we outline our algorithm and briefly explain the operation steps. The system and channel model are discussed in Sect. 3. We formulate the sparse recovery-based channel estimation with the proposed DGAMP-SBL algorithm and get more accurate channel estimation using a refined algorithm in Sect. 4. The performance of the proposed algorithm is compared with that of the existing algorithms, and the superiority of the proposed algorithm is concluded in V. At the end of the paper, a conclusion is given in Sect. 6.

The notations related to this paper are shown in Table 1.

2 Method

In this section, an off-grid channel model is implied for sparse representation of massive mmWave MIMO systems and was propose a computationally efficient and accurate mmWave channel estimation algorithm. Generalized approximate message passing

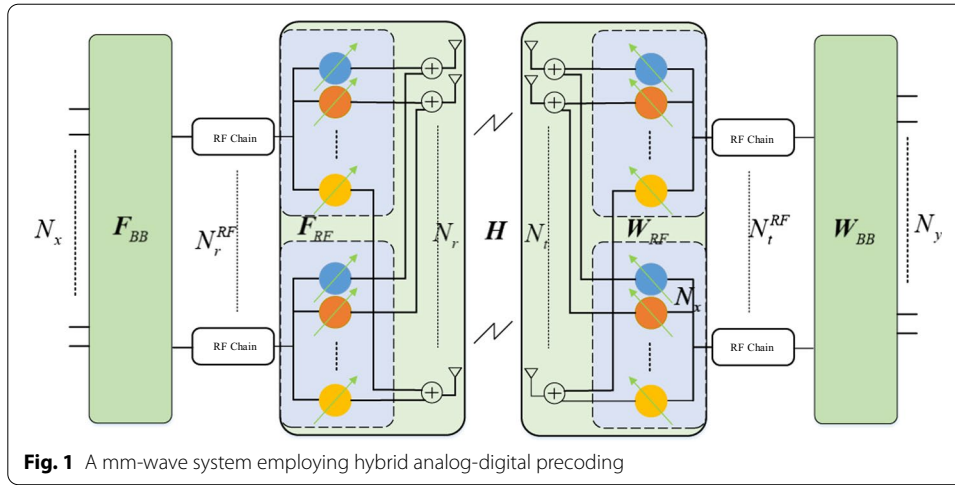
Table 1 Related notations

Notations	Definitions
capital bold letters	Matrices
lowercase bold letters	Vectors
$[\cdot]^*$	Conjugate operator
$[\cdot]^H$	Conjugate transpose operator
$[\cdot]^T$	Transpose operator
$[\cdot]^{-1}$	Inverse operator
$\ \mathbf{X}\ _F$	Frobenius norm of \mathbf{X}
$\ \mathbf{x}\ _2$	l_2 norm of \mathbf{x}
\otimes	Kronecker product
\odot	Khatri–Rao product
$\text{diag}(\mathbf{x})$	Diagonal matrix with elements of \mathbf{x} on its diagonal
$\text{vec}(\mathbf{X})$	Vectorization operation of \mathbf{X}
$\text{vecd}(\mathbf{X})$	Vectorization operation of diagonal elements of \mathbf{X}
$\text{Re}\{\cdot\}$	Real part operator
$\text{Im}\{\cdot\}$	Imaginary part operator
$E(\cdot)$	Expectation operator
$\det(\cdot)$	Determinant operator

(GAMP) is an extension of the approximate message passing algorithm [21] to more comprehensive scenarios involving arbitrary priors and observation noise [22]. We use a GAMP based low complexity SBL algorithm to avoid the high-dimension matrix inversion. This algorithm can be embedded in the expectation–maximization (EM) framework to approximate the true posterior distribution of sparse. Then we improve a fast refined algorithm that utilizes the interpolation of three 2-DFT spectral lines to improve the accuracy of the algorithm. Significantly, the improved algorithm can get accurate angles of departures/arrivals (AoAs/AoDs) estimation without iterations. Consequently, we can obtain both computational efficiency and estimation precision. Finally, the lower bound of normalized mean-squared error of the proposed algorithm is derived analytically: the lower bound of estimation is derived by the least square (LS) algorithm with the assumption that the angles of AoAs/AoDs are known. And then, we estimate the channel by using the LS algorithm. The result indicates our proposed algorithm is close to the lower bound at high signal-to-noise ratios.

3 System model

In this paper, we consider a typical hybrid mmWave MIMO systems consisting of a BS and a MS, where the BS and MS are equipped with N_t and N_r antennas, respectively. Let N_t^{RF} and N_r^{RF} be the number of transmitter radio frequency (RF) chains and receiver RF chains, respectively. We consider a hybrid analog–digital communication system that has inherent sparse characteristics as shown in Fig. 1. For mmWave massive MIMO with hybrid precoding, the amount of RF chains is far less than the amount of antennas, i.e., $N_t^{\text{RF}} < N_t$, $N_r^{\text{RF}} < N_r$ [23–25]. We assume that $\mathbf{F} = \mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}$ is the $N_t \times N_s^t$ hybrid



precoding matrix in BS, where $\mathbf{F}_{RF} \in \mathbb{C}^{N_t \times N_t^{RF}}$ and $\mathbf{F}_{BB} \in \mathbb{C}^{N_t^{RF} \times N_s^t}$ denote the analog and digital precoders, respectively. $\mathbf{W} = \mathbf{W}_{RF}\mathbf{W}_{BB} \in \mathbb{C}^{N_r \times N_s^r}$ is the hybrid combiner. $\mathbf{W}_{RF} \in \mathbb{C}^{N_r \times N_r^{RF}}$ and $\mathbf{W}_{BB} \in \mathbb{C}^{N_r^{RF} \times N_s^r}$ denote the analog and digital combiners, respectively, and the received signal $\mathbf{r} \in \mathbb{C}^{N_s^r \times 1}$ can be rewritten as

$$\begin{aligned} \mathbf{r} &= \mathbf{W}_{BB}^H \mathbf{W}_{RF}^H \mathbf{H} \mathbf{F}_{RF} \mathbf{F}_{BB} \mathbf{s} \\ &\quad + \mathbf{W}_{BB}^H \mathbf{W}_{RF}^H \mathbf{n} \end{aligned} \tag{1}$$

$$= \mathbf{W}^H \mathbf{H} \mathbf{F} \mathbf{s} + \mathbf{W}^H \mathbf{n},$$

where $\mathbf{r} \in \mathbb{C}^{N_s^r \times 1}$ is the received vector, $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$ is the channel matrix, $\mathbf{s} \in \mathbb{C}^{N_s^t \times 1}$ and $\mathbf{n} \in \mathbb{C}^{N_s^r \times 1}$ is the Gaussian noise matrix.

Use $\mathbf{x} = \mathbf{F} \mathbf{s} \in \mathbb{C}^{N_t \times 1}$ to denote the transmit signal, and the transmit vector elements correspond to the transmit antennas position. Firstly the channel estimation is considered in M time slots, and the channel matrix remains unchanged during the M time slots. N_x ($N_x < N_t$) pilot sequences $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{N_x}$ are sent by transmitters, and we also use N_y dimensional received pilot \mathbf{y}_p by receiving antennas for each transmits pilot sequence \mathbf{x}_p ($1 \leq p \leq N_x$)

$$\mathbf{y}_p = \mathbf{W}_r^H \mathbf{H} \mathbf{x}_p + \mathbf{W}_r^H \mathbf{n}_p, \tag{2}$$

where $\mathbf{W}_r = [\mathbf{W}_1, \mathbf{W}_2, \dots, \mathbf{W}_M] \in \mathbb{C}^{N_r \times N_y}$, $\mathbf{n}_p \in \mathbb{C}^{N_r \times 1}$ is the noise. Denote $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_{N_x}]$, $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{N_x}]$, $\mathbf{N} = [\mathbf{n}_1, \mathbf{n}_2, \dots, \mathbf{n}_{N_x}]$, we have

$$\mathbf{Y} = \mathbf{W}_r^H \mathbf{H} \mathbf{X} + \widehat{\mathbf{N}}, \tag{3}$$

where $\widehat{\mathbf{N}} = \mathbf{W}_r^H \mathbf{N}$ and \mathbf{N} represents the noise matrix with each entry following the distribution $\mathcal{CN}(0, \sigma_n^2)$. In this paper, we suppose that the matrices \mathbf{W}_r and \mathbf{X} are known as prior knowledge at the receiver without loss of generality. As mentioned in [16], the

entries in \mathbf{W}_r and \mathbf{X} are optimally chosen from the set $w_{i,j} = \sqrt{1/N_t} e^{j\beta_{i,j}}$, $x_{i,j} = \sqrt{\rho/N_t} e^{j\beta_{i,j}}$ where $\beta_{i,j}$ is the random phase uniformly distributed in $[0, 2\pi)$. This assumption is also employed in [26, 27]. Further details of how to design \mathbf{W}_r and \mathbf{X} can be found from [28].

It is not available to observe \mathbf{H} directly. Instead, a noisy edition $\mathbf{W}_r^H \mathbf{H} \mathbf{X}$ can be observed from the receiver. This is referred to as channel subspace sampling and a detailed analysis can be found in [29, 30]. Fortunately, as mentioned above, the channel estimation problem can be solved as a sparse signal recovery. The typical model of mmWave channel is modeled as [31]

$$\mathbf{H} = \sum_{l=1}^{\tilde{L}} \alpha_l \mathbf{a}_r(\theta_l) \mathbf{a}_t^H(\phi_l), \tag{4}$$

where \tilde{L} denotes the number of communication paths. There are usually a few reflected path clusters in mmWave multipath propagation channel [26–28]. Therefore, we usually have $\tilde{L} \ll \min\{N_r, N_t\}$. α_l is the complex-valued gain corresponding to the l -th communication path. $\theta_l \in [-\pi, \pi]$ and $\phi_l \in [-\pi, \pi]$ are the corresponding azimuth AoAs and AoDs [32–34], respectively, and $\mathbf{a}_r \in \mathbb{C}^{N_r \times 1}$ ($\mathbf{a}_t \in \mathbb{C}^{N_t \times 1}$) is the array response vector corresponding to the receiver (transmitter).

In this paper, we employ ULA both at the transmitter and the receiver. Then, the steering vectors at the BS and MS can be written as

$$\mathbf{a}_r(\theta_l) = \left[1, e^{-j2\pi \frac{d}{\lambda} \sin(\theta_l)}, \dots, e^{-j2\pi \frac{(N_r-1)d}{\lambda} \sin(\theta_l)} \right]^T, \tag{5}$$

$$\mathbf{a}_t(\phi_l) = \left[1, e^{-j2\pi \frac{d}{\lambda} \sin(\phi_l)}, \dots, e^{-j2\pi \frac{(N_t-1)d}{\lambda} \sin(\phi_l)} \right]^T, \tag{6}$$

where λ is the signal wavelength and d denotes the spacing between every two adjacent antenna elements with $d = \lambda/2$. For convenience, each scatterer is assumed to possess a transmission path, and the channel gains are modeled as independent identically distributed random variables with the distribution $\mathcal{CN}(0, \sigma_\alpha^2)$. By defining the steering matrices $\Phi_r = [\mathbf{a}_r(\theta_1), \dots, \mathbf{a}_r(\theta_{\tilde{L}})] \in \mathbb{C}^{N_r \times \tilde{L}}$, $\Phi_t = [\mathbf{a}_t(\phi_1), \dots, \mathbf{a}_t(\phi_{\tilde{L}})] \in \mathbb{C}^{N_t \times \tilde{L}}$ in a matrix form, the mmWave channel matrix \mathbf{H} in (4) can be rewritten in a matrix form as follows

$$\mathbf{H} = \Phi_r \mathbf{H}_v \Phi_t^H, \tag{7}$$

where \mathbf{H}_v is a diagonal matrix with diagonal elements $\alpha = [\alpha_1, \dots, \alpha_{\tilde{L}}]^T$. To transform the signal estimation problem into the sparse recovery problem, we firstly express the channel in a sparse matrix form

$$\mathbf{H} = \mathbf{A}_r \mathbf{C} \mathbf{A}_t^H, \tag{8}$$

where both $\mathbf{A}_r \triangleq \frac{1}{\sqrt{N_r}} [\mathbf{a}_r(\psi_1) \cdots \mathbf{a}_r(\psi_{N_1})] \in \mathbb{C}^{N_r \times N_1}$ and $\mathbf{A}_t \triangleq \frac{1}{\sqrt{N_t}} [\mathbf{a}_t(\omega_1) \cdots \mathbf{a}_t(\omega_{N_2})] \in \mathbb{C}^{N_t \times N_2}$ are overcomplete beamforming matrices

($N_1 \geq N_r, N_2 \geq N_t$) which are assumed as channel-invariant unitary DFT matrices, and $\mathbf{A}_r \mathbf{A}_r^H = \mathbf{I}_{N_1}, \mathbf{A}_t \mathbf{A}_t^H = \mathbf{I}_{N_2}$, respectively. $\mathbf{C} \in \mathbb{C}^{N_1 \times N_2}$ is a sparse representation matrix of mmWave channel with the \tilde{L} nonzero elements correspond to each scatterer under the condition that each AoA/AoD is defined in quantization grids. Note that the practical direction is usually invalid to locate on the predefined grids, so it is actually that the quantizing error exists in the sparse channel matrix. Hence, a refined channel estimation scheme is given in the later sections to solve the off-grid problem. What is more, it is worth noting that \mathbf{C} is no longer diagonal.

The signal model in (3) can be further denoted as follows

$$\begin{aligned} \mathbf{y} &= \sqrt{P} \text{vec}(\mathbf{W}_r^H \mathbf{H} \mathbf{X}) + \text{vec}(\widehat{\mathbf{N}}) \\ &= \sqrt{P} (\mathbf{X}^T \otimes \mathbf{W}_r^H) \text{vec}(\mathbf{H}) + \widehat{\mathbf{n}} \\ &= \sqrt{P} (\mathbf{X}^T \mathbf{A}_t^* \otimes \mathbf{W}_r^H \mathbf{A}_r) \mathbf{c} + \widehat{\mathbf{n}}, \end{aligned} \tag{9}$$

where $\mathbf{c} = \text{vec}(\mathbf{C}), \mathbf{y} = \text{vec}(\mathbf{Y}) \in \mathbb{C}^{N_x N_y \times 1}$ and $\widehat{\mathbf{n}} = \text{vec}(\widehat{\mathbf{N}})$. The estimation of \mathbf{c} can be handled as a sparse signal recovery problem with 2D dictionary matrix ($\mathbf{A}_t^* \otimes \mathbf{A}_r$), and the sensing matrix also can be denoted as $\Phi = \mathbf{F}_t^T \mathbf{A}_t^* \otimes \mathbf{W}_r^H \mathbf{A}_r$. We consider the number of grids $N_1 = N_2$ for simplicity, and the channel estimation problem can be summarized as compressed sensing problem as follows

$$\arg \min_{\mathbf{c}} \|\mathbf{c}\|_0 \quad \text{subject to} \left\| \mathbf{y} - \sqrt{P} \Phi \mathbf{c} \right\|_2 < \zeta, \tag{10}$$

where ζ is a tolerance constant which is determined by noise power. Since recovery signal from l_0 -norm is a NP hard question, we replace the l_1 -norm with l_0 -norm as follows

$$\arg \min_{\mathbf{c}} \|\mathbf{c}\|_1 \quad \text{subject to} \left\| \mathbf{y} - \sqrt{P} \Phi \mathbf{c} \right\|_2 < \zeta. \tag{11}$$

4 Proposed GAMP-SBL-based high precision channel estimation

This section describes the proposed SBL scheme for sparse beamspace channel vector estimation. We begin by introducing the proposed hierarchical prior model. Then, we use the DGAMP-SBL to estimate the mmWave channel. After that, we propose a refined algorithm to get the accurate estimation of AoAs and AoDs. And then we propose a refined algorithm to estimate the AoAs and AoDs accurately. Finally, we analyze the computation complexity of the proposed algorithm and make a comparison to the complexity of SBL.

4.1 Hierarchical prior model

We have the following results using the assumption of circular symmetric complex Gaussian noises

$$p(\mathbf{y}|\mathbf{c}, \lambda) = \mathcal{CN}(\mathbf{y}|\Phi \mathbf{c}, \lambda^{-1} \mathbf{I}), \tag{12}$$

where $\lambda = \sigma^{-2}$ stands for the noise precision. In order to leverage the sparse structure of the underlying mmWave channel, we enforce sparsity constraints on the channel vector which is commonly used in the sparse Bayesian model. The sparse channel vector \mathbf{c} is assumed to follow a two hierarchical prior Gaussian distribution. In the first layer, \mathbf{c} is assumed that the entries of \mathbf{c} are independent and identically distributed, i.e.

$$p(\mathbf{c}|\gamma) \sim \prod_{i=1}^{N_1N_2} \mathcal{CN}(c_i; 0, \gamma_i^{-1}), \tag{13}$$

where γ_i denotes the inverse variance, and $\gamma = [\gamma_1, \gamma_2, \dots, \gamma_{N_1N_2}]$. Meanwhile, a Gamma prior is employed for the inverse variance vector

$$\begin{aligned} p(\gamma) &= \prod_{i=1}^{N_1N_2} \text{Gamma}(\gamma_i | a + 1, b) \\ &= \prod_{i=1}^{N_1N_2} \Gamma^{-1}(a + 1) b^a \gamma_i^a e^{-b\gamma_i}, \end{aligned} \tag{14}$$

where a, b are parameters associated with the above distribution, and $\Gamma(a) = \int_0^\infty t^{a-1} e^{-t}$ is the Gamma function. Besides, \mathbf{n} is Gaussian noise with zero mean and covariance matrix $(1/\lambda)\mathbf{I}$. We set a Gamma hyperprior over λ : $p(\lambda) = \text{Gamma}(\lambda | c + 1, d) = \Gamma(c + 1)^{-1} d^{c+1} \lambda^c e^{-d\lambda}$.

To obtain a broad hyperprior, we set $a, b \rightarrow 0$ [35, 36]. This two-stage hierarchical prior gives

$$p(c_i) = \int_0^\infty p(c_i | \gamma_i^{-1}) p(\gamma_i) d\gamma_i, \tag{15}$$

which is helpful to obtain sparse solutions due to the sharp peak and heavy tails with small a and b . Actually, according to paper [37], the maximum posterior estimation of \mathbf{c} is consistent with l_0 -norm solution in formula (10) by FOCUSS with $p \rightarrow 0$. To update the parameter $\theta = \{\gamma, \lambda\}$, we can also use maximum posteriori estimation to achieve the most probable values, i.e.,

$$(\gamma^*, \lambda^*) = \arg \max_{\gamma, \lambda} p(\gamma, \lambda | \mathbf{y}), \tag{16}$$

or, equivalently,

$$(\gamma^*, \lambda^*) = \arg \max_{\gamma, \lambda} \ln p(\gamma, \lambda, \mathbf{y}). \tag{17}$$

Then, the EM-SBL algorithm is employed to learn the sparse vector \mathbf{c} and iteratively update the hyper-parameters $\theta = \{\gamma, \lambda\}$. Note that the key step is to update the hyper-parameters $p(\gamma, \lambda | \mathbf{y})$ by maximizing the posterior probability when the EM-SBL is in the updating phase. And in the stage of E-step, the likelihood function can be written as follows

$$p(\mathbf{y}|\mathbf{c}; \sigma^2) = \frac{1}{(2\pi\sigma^2)^{N_x N_y}} \exp\left(-\frac{1}{2\sigma^2} \|\mathbf{y} - \Phi \mathbf{c}\|^2\right). \tag{18}$$

As noted before, the conditional density $p(\mathbf{c}|\mathbf{r})$ shows \mathbf{c} is Gaussian distribution. When we treat \mathbf{c} as a hidden variable, its posterior distribution we need to compute conditioned on the observed vector \mathbf{y} and the updated hyper-parameters θ is a complex Gaussian [35] function

$$\begin{aligned} p(\mathbf{c}|\mathbf{y}, \gamma^{(t)}, \lambda^{(t)}) &= \frac{p(\mathbf{c}|\gamma^{(t)})p(\mathbf{y}|\mathbf{c}, \lambda^{(t)})}{p(\mathbf{y}|\gamma^{(t)}, \lambda^{(t)})} \\ &= (2\pi)^{-\frac{(N_1 N_2 + 1)}{2}} |\Sigma_c|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\mathbf{c} - \mu_c)^T \Sigma^{-1}(\mathbf{c} - \mu_c)\right\}, \end{aligned} \tag{19}$$

where the posterior covariance Σ_c and mean μ_c are given, respectively,

$$\begin{aligned} \Sigma_c &= (\lambda \Phi^T \Phi + \mathbf{D})^{-1} \\ \mu_c &= \lambda \Sigma_c \Phi^T \mathbf{y}, \end{aligned} \tag{20}$$

where $\mathbf{D} = \text{diag}(\gamma_0, \gamma_1, \dots, \gamma_{N_1 N_2})$, Σ_c and μ_c are the posterior mean and variance with relevant for $p(\mathbf{c} | \mathbf{y}, \gamma^{(t)}, \lambda^{(t)})$, respectively. We assume that τ_c is the vector whose elements are composed of the diagonal of the covariance matrix Σ_c . As mentioned above, in the EM algorithm iterative process, the hyper-parameters are updated by iteratively maximizing the R -function, i.e.,

$$\begin{aligned} \theta^{(t+1)} &\triangleq \arg \max_{\theta} R(\theta|\theta^{(t)}) \\ &\triangleq \arg \max_{\theta} E_{\mathbf{c}|\mathbf{y}, \theta^{(t)}}[\log p(\theta|\mathbf{c}, \mathbf{y})]. \end{aligned} \tag{21}$$

Using Bayesian rule, (21) can be rewritten by ignoring part unrelated to θ as follows

$$\begin{aligned} E_{\mathbf{c}|\mathbf{y}, \gamma^t, \lambda}[\log p(\gamma^t, \lambda|\mathbf{c}, \mathbf{y})] \\ = E_{\mathbf{c}|\mathbf{y}, \gamma^t, \lambda}[-\log p(\mathbf{y}|\mathbf{c}; \lambda) - \log p(\mathbf{c}|\gamma) - \log p(\gamma)]. \end{aligned} \tag{22}$$

Firstly, the algorithm carries out the M-step for the hyper-parameters $\{\gamma_n\}$. We take the partial derivative of the R-function with respect to γ_n with eliminating independent terms. Since the first term in (22) does not depend on γ , it can be ignored as it will not be relevant for the M-Step. The objective function in (22) becomes

$$\begin{aligned} E_{\mathbf{c}|\mathbf{y}, \theta^{(t)}}[\log p(\theta|\mathbf{c}, \mathbf{y})] \\ = E_{\mathbf{c}|\mathbf{y}, \theta^{(t)}}[-\log p(\mathbf{c}|\gamma) - \log p(\gamma)] \\ = \sum_{n=1}^{N_1 N_2} \left(\frac{\gamma_n (\hat{c}_n^2 + \tau_{c_n})}{2} - \frac{1}{2} \log \gamma_n + \log p(\gamma_n) \right) \\ = \sum_{n=1}^{N_1 N_2} \left(\left(\frac{\gamma_n (\hat{c}_n^2 + \tau_{c_n})}{2} - \frac{1}{2} \log \gamma_n + a \log \gamma_n - b \gamma_n \right) \right). \end{aligned} \tag{23}$$

We take the partial derivative of the R-function with respect to γ_n with eliminating independent terms, and the iteration of γ_n can be denoted by

$$\gamma_n^{i+1} = \arg \min_{\gamma_n} \left(\frac{\gamma_n(\hat{c}_n^2 + \tau_{c_n})}{2} + \log p(\gamma_n) - \frac{1}{2} \log \gamma_n \right). \tag{24}$$

According to the hyperprior $p(\gamma_n)$ which possesses a non-informative when the parameter a and b tend to zero, we can simplify the update formalization as

$$\gamma_n^{i+1} = \frac{1}{\hat{c}_n^2 + \tau_{c_n}}. \tag{25}$$

Similarly, we then compute the estimation of the scalar hyper-parameter λ and it can be updated as

$$\begin{aligned} \lambda^{i+1} &= \arg \max_{\lambda} E_{c|\mathbf{y}, \gamma; \lambda} [p(\mathbf{y}, \mathbf{c}, \gamma; \lambda)] \\ &= \frac{\|\mathbf{y} - \Phi \mathbf{c}\|^2 + (\lambda)^i \sum_{n=1}^{N_1 N_2} (1 - \frac{\tau_{c_n}}{\gamma_n})}{N}, \end{aligned} \tag{26}$$

where $N = N_x N_y$ represents the dimension of \mathbf{y} . According to the references and practical implementation tips, we can consistently set the constant as follows: $a = b = c = d = 0.0001$, since the result of constant initialization has little effect on estimation performance.

4.2 Update by DGAMP

As mentioned earlier, it is obvious that the calculation of the posterior mean and posterior variance which involve the inversion of high-dimensional matrices is extensive. Hence, the high computational complexity characteristic of the EM-SBL algorithm causes that it is impractical to be adopted by the massive MIMO channel estimation. To simplify the calculation, we replace posterior calculation with the GAMP algorithm which is a very-low-complexity Bayesian iterative technique. It is noted that the hyper-parameters $\{\gamma, \lambda\}$ are considered as known constants during the iterative process of the GAMP algorithm.

GAMP is a fast heuristic algorithm and can be utilized for simplifying matrix inversion within the SBL framework [38, 39]. GAMP algorithm obtains the maximum posterior estimation of \mathbf{c} by Taylor approximation. Specifically, the process of iteratively computing the marginal posterior $p(c_n | \mathbf{y}, \gamma, \lambda)$ is performed by message passing on the GAMP factor graph. By utilizing the condition that all posteriors are Gauss, the process can be simplified by replacing posterior probability with expectation and variance of the sparse variables $\{c_n\}$ and mixture variables $\{z_m\}$ whose elements are denoted by $\mathbf{z} = \Phi \mathbf{c}$. To detour the convergence of GAMP whose measurement matrix satisfies independent Gaussian distribution, paper [40] proposes a DGAMP algorithm to improve the robustness of the measurement matrix through importing damping factors $\rho_s, \rho_c \in (0, 1]$, but it will also slow down the convergence speed. Nevertheless, the process is computationally

efficient since that it only contains scalar operations. Then, we summarize the key steps of DGAMP in Algorithm 1. The readers can refer to [22] to learn more about details of the derivation of DGAMP.

Algorithm 1 DGAMP-SBL Algorithm

- 1: Initialization
 - 2: $\mathbf{S} \leftarrow |\Phi^2|$
 - 3: Initialize $\tilde{\tau}_c^0, \gamma^0, \lambda^0 > 0$
 - 4: $\tilde{\mathbf{s}}, \tilde{\mathbf{c}} \leftarrow \mathbf{0}$
 - 5: for $i = 1, 2, \dots, I_{\max}$
 - 6: $\tau_c^1 \leftarrow \tilde{\tau}_c^{i-1}, \hat{\mathbf{c}}^1 \leftarrow \tilde{\mathbf{c}}^{i-1}, \mathbf{s}^1 \leftarrow \tilde{\mathbf{s}}^{i-1}$
 - 7: E-Step
 - 8: for $k = 1, 2, \dots, K_{\max}$
 - 9: $1/\tau_p^k \leftarrow \mathbf{S}\tau_x^k$
 - 10: $\mathbf{p}^k \leftarrow \mathbf{s}^{k-1} + \tau_p^k \Phi \hat{\mathbf{c}}^k$
 - 11: $\tau_s^k \leftarrow \tau_p^k g'_s(\mathbf{p}^k, \tau_p^k)$
 - 12: $\mathbf{s}^k \leftarrow (1 - \theta_s) \mathbf{s}^{k-1} + \theta_s g_s(\mathbf{p}^k, \tau_p^k)$
 - 13: $1/\tau_r^k \leftarrow \mathbf{S}^T \tau_s^k$
 - 14: $\mathbf{r}^k \leftarrow \hat{\mathbf{c}}^k - \tau_r^k \Phi^T \mathbf{s}^k$
 - 15: $\tau_c^{k+1} \leftarrow \tau_r^k g'_c(\mathbf{r}^k, \tau_r^k)$
 - 16: $\hat{\mathbf{c}}^{k+1} \leftarrow (1 - \theta_x) \hat{\mathbf{c}}^k + \theta_x g_c(\mathbf{r}^k, \tau_r^k)$
 - 17: $\|\hat{\mathbf{c}}^{k+1} - \hat{\mathbf{c}}^k\|^2 / \|\hat{\mathbf{c}}^{k+1}\|^2 < \varepsilon_{gamp}$
 - 18: $\tilde{\mathbf{s}}^i \leftarrow \mathbf{s}^k, \tilde{\mathbf{c}}^i \leftarrow \hat{\mathbf{c}}^{k+1}, \tilde{\tau}_c^i \leftarrow \tau_c^{k+1}$
 - 19: M-Step
 - 20: $\gamma^{i+1} \leftarrow \left| \tilde{\mathbf{c}}^i \right|^2 + \tilde{\tau}_c^i$
 - 21: $(\sigma^2)^{i+1} \leftarrow \frac{\|\mathbf{y} - \Phi \tilde{\mathbf{c}}^i\|^2 + \lambda^i \sum_{n=1}^N \left(1 - \frac{\tilde{\tau}_c^i}{\gamma_n^i}\right)}{M}$
 - 22: if $\left\| \tilde{\mathbf{c}}^i - \mathbf{c}^{i-1} \right\|^2 / \left\| \tilde{\mathbf{c}}^i \right\|^2 < \varepsilon_{em}, \text{break}$
 - 23: end
-

There are two versions which are the max-sum version and sum-product versions damp-GAMP algorithm. The input and output functions $g_s(\mathbf{p}, \tau_p)$ and output functions $g_x(\mathbf{r}, \tau_r)$ in Algorithm 1 are distinguished according to whether the max-sum or the sum-product version of GAMP. Coincidentally, both the sum-product and max-sum version simplify the same equation. We only introduce the functions of the input and output of the sum-product, and readers can refer [41] to get furthermore detail. The intermediate variables \mathbf{r} and \mathbf{p} are explained as approximations of Gaussian noise corrupted of \mathbf{c} and $\mathbf{z} = \Phi \mathbf{c}$ with the noise levels of τ_r and τ_p , respectively. The difference between the sum-product and max-sum version is the estimation strategy. The sum-product version uses the vector minimum mean-squared error (MMSE) estimation which is reduced to a sequence of scalar MMSE estimation. The input and output functions are shown as follows

$$[g_s(\mathbf{p}, \tau_p)]_m = \frac{\int z_m p(y_m|z_m) \mathcal{N}\left(z_m; \frac{p_m}{\tau_{pm}}, \frac{1}{\tau_{pm}}\right) dz_m}{\int p(y_m|z_m) \mathcal{N}\left(z_m; \frac{p_m}{\tau_{pm}}, \frac{1}{\tau_{pm}}\right) dz_m}, \tag{27}$$

$$[g_c(\mathbf{r}, \tau_r)]_n = \frac{\int c_n p(c_n) \mathcal{N}(c_n; r_n, \tau_{rn}) dc_n}{\int p(c_n) \mathcal{N}(c_n; r_n, \tau_{rn}) dc_n}. \tag{28}$$

According to the (12–15) which are the prior information and the posterior information we imposed on \mathbf{c} and $p(\mathbf{y}|\mathbf{c})$, respectively, the input function and its derivative function can be rewritten as follows

$$g_c(\mathbf{r}, \tau_r) = \frac{\gamma}{\gamma + \tau_r} \mathbf{r}, \tag{29}$$

$$g'_c(\mathbf{r}, \tau_r) = \frac{\gamma}{\gamma + \tau_r}. \tag{30}$$

Similarly, the output function and its derivative function can be rewritten as follows

$$g_s(\mathbf{p}, \tau_p) = \frac{(\mathbf{p}/\tau_p - \mathbf{y})}{(\lambda + 1/\tau_p)}, \tag{31}$$

$$g'_s(\mathbf{p}, \tau_p) = \frac{\lambda^{-1}}{\lambda^{-1} + \tau_p}. \tag{32}$$

Upon convergence, DGAMP-SBL yields a sparse estimate of \mathbf{c} according to the prior information given by the posterior mean. This also gets a coarse estimate of the channel matrix, which can be represented by a paired frequency of AoAs and AoDs using multiplying the steering vectors. The initial channel estimation can be obtained as

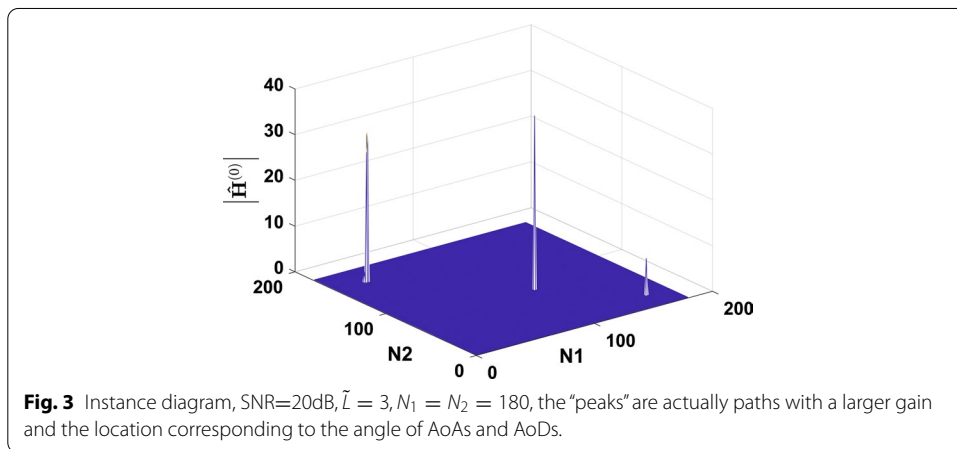
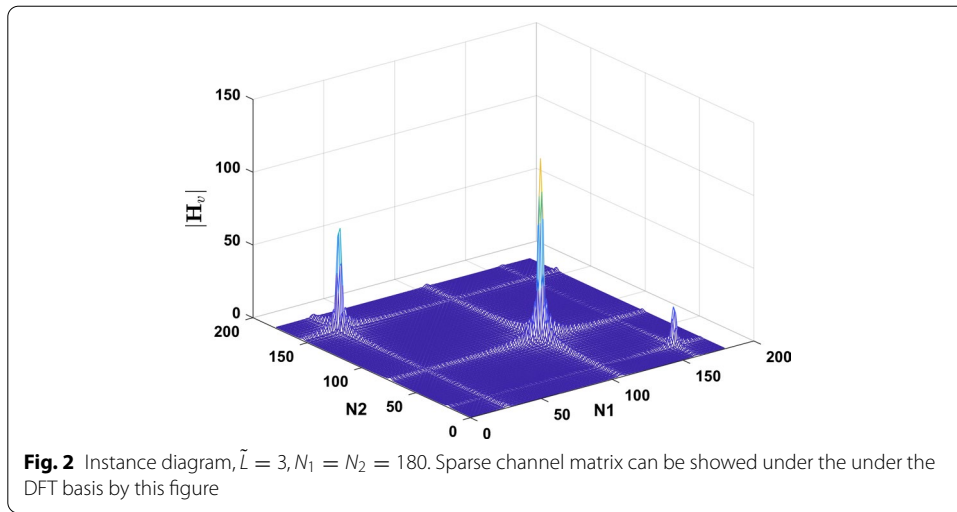
$$\hat{\mathbf{H}}^{(0)} = \mathbf{A}_r \hat{\mathbf{C}} \mathbf{A}_t^H, \tag{33}$$

where $\hat{\mathbf{C}}$ denote order by columns as \mathbf{H} .

It is worth noting that GAMP is a low complexity algorithm that transforms the vector estimation into the scalar estimation; therefore, (27), (28) and the operations in the Algorithm 1, all vector squares, divisions and multiplications are taken element-wise. Figure 2 represents sparse channel matrix based on discrete Fourier basis. Figure 3 represents the SBL-GAMP algorithm estimation without considering the sparse off-grid problem when the number of grids is chosen as $N_1 = N_2 = 180$ for comparison.

4.3 Refined estimation

In the following, we propose an exact and fast 2D frequency estimation method base on the interpolation of three 2-DFT spectral lines [42]. The SBL-DGAMP algorithm,



represented by the frequency estimate $\hat{f} = \{\hat{f}_1, \hat{f}_2, \dots, \hat{f}_{\tilde{L}}\}$ where the $\hat{f}_i = \{f_{i1}, f_{i2}\}$ ($i = 1, 2, \dots, \tilde{L}$), is only a coarse estimate due to grid mismatch.

For $(k, j) \in [(k_1, j_1), (k_2, j_2) \dots, (k_{\tilde{L}}, j_{\tilde{L}})]$ related to the corresponding index set of the coarse estimate frequency. And then, we present a interpolation on Fractional Fourier Coefficient. The Fractional Fourier transform can be formulated as follows

$$D(k, j) = \sum_{m=0}^{N_r} \sum_{n=0}^{N_t} [\tilde{H}]_{n,m} e^{-j2\pi(im+kn)/N_{\text{DFT}}}, \tag{34}$$

$$D(k \pm \delta, j) = \sum_{m=0}^{N_r} \sum_{n=0}^{N_t} [\tilde{H}]_{n,m} e^{-j2\pi(im+(k\pm\delta)n)/N_{\text{DFT}}}, \tag{35}$$

$$D(k, \delta \pm j) = \sum_{m=0}^{N_r} \sum_{n=0}^{N_t} [\tilde{H}]_{n,m} e^{-j2\pi(i(m\pm\delta)+kn)/N_{DFT}}, \tag{36}$$

where δ is a real number, and we formulate the $D(k, j), D(k \pm \delta, j), D(k, j \pm \delta)$ as $D, D_{K+\delta}, D_{K-\delta}, D_{J+\delta}, D_{J-\delta}$, respectively. Then, we propose the interpolation algorithm as follows

$$\hat{\lambda}_x = \frac{N_r}{\pi} \tan^{-1} \left(\operatorname{Re} \left\{ \frac{(D_{K+\delta} E_r^+ - D_{K-\delta} E_r^-) \cdot \sin(\pi i / N_r)}{(D_{K+\delta} E_r^+ + D_{K-\delta} E_r^-) \cdot \cos(\pi i / N_r) - 2D \cos(\pi i)} \right\} \right), \tag{37}$$

$$\hat{\lambda}_y = \frac{N_t}{\pi} \tan^{-1} \left(\operatorname{Re} \left\{ \frac{(D_{J+\delta} E_t^+ - D_{J-\delta} E_t^-) \cdot \sin(\pi i / N_t)}{(D_{J+\delta} E_t^+ + D_{J-\delta} E_t^-) \cdot \cos(\pi i / N_t) - 2D \cos(\pi i)} \right\} \right), \tag{38}$$

where $\hat{\lambda}_x, \hat{\lambda}_y$ are frequency deviation normalized by grid distance Δf which is defined by $\Delta f = 1/(N_{DFT}/2)$.

And then we could get the frequency estimations that

$$\hat{f}_x = \begin{cases} -(\hat{\lambda}_x + k)/(N_{DFT}/2) & k < N_{DFT}/2 \\ 2 - (\hat{\lambda}_x + k)/(N_{DFT}/2) & k \geq N_{DFT}/2 \end{cases}, \tag{39}$$

and

$$\hat{f}_y = \begin{cases} (\hat{\lambda}_y + j)/(N_{DFT}/2) & j < N_{DFT}/2 \\ -(2 - (\hat{\lambda}_y + j)/(N_{DFT}/2)) & j \geq N_{DFT}/2 \end{cases}. \tag{40}$$

According to [43], in order to get a more accurate frequency estimate, we adopt a parabolic interpolation algorithm. For each frequency dimension $d \in \{x, y\}$, we calculate three periodogram sample D_{d1}, D_{d2} and D_{d3} at frequency $\theta_{d1} = \hat{f}_d - \Delta_d, \theta_{d2} = \hat{f}_d$ and $\theta_{d3} = \hat{f}_d + \Delta_d$. Middle frequency \hat{f}_d given by (39) and (40), and the sides of frequency θ_{d1} and θ_{d3} are excused by Δ_d which is chosen to satisfy $\Delta_d \in (0, \frac{1}{2N_{DFT}})$, as well as high estimation accuracy. The last step of frequency estimation along the d th dimension is achieved by calculating the vertex of a parabola fitted through points $(\theta_{d1}, D_{d1}), (\theta_{d2}, D_{d2})$ and (θ_{d3}, D_{d3}) , i.e.

$$\theta_d^{fin} = \frac{1}{2} \frac{\theta_{d3}^2(\Delta_{12}) + \theta_{d2}^2(\Delta_{31}) + \theta_{d1}^2(\Delta_{23})}{\theta_{d3}(\Delta_{12}) + \theta_{d2}(\Delta_{31}) + \theta_{d1}(\Delta_{23})}, \tag{41}$$

where $d = x, y, \Delta_{12} = (D_{d1} - D_{d2}), \Delta_{31} = D_{d3} - D_{d1}, \Delta_{23} = D_{d2} - D_{d3}$. Eventually, according to (41) the final estimation frequency, the accurate angles AoAs and AoDs can be obtained, i.e., $\{\hat{\theta}_l\}_{l=1}^{\tilde{L}}$ and $\{\hat{\phi}_l\}_{l=1}^{\tilde{L}}$.

Algorithm 2 PROPOSED CHANNEL ESTIMATION SCHEME

First stage:Initialization:

- 1: Based on Algorithm 1, we use the DGAMP-SBL to obtain $\hat{\mathbf{c}}$;
- 2: Use (31) to obtain the initial channel estimate $\hat{\mathbf{H}}^{(0)}$;

Second stage:Refine:

- 1: According to the coarse estimation $\hat{\mathbf{H}}^{(0)}$ to get the frequency initial $\hat{f} = \{\hat{f}_1, \hat{f}_2, \dots, \hat{f}_{\bar{L}}\}$;
 - 2: Compute the 2DFT, update the frequency \hat{f}_x and \hat{f}_y using (37) and (38);
 - 3: Compute, recoder the final frequency θ_x^{fin} and θ_y^{fin} by (39), obtain the AoAs and AoDs $\{\hat{\theta}_l\}_{l=1}^{\bar{L}}$ and $\{\hat{\phi}_l\}_{l=1}^{\bar{L}}$;
 - 4: Running (42) to obtain \mathbf{h}_v , and obtain $\hat{\mathbf{H}} = \Phi_r \text{diag}(\mathbf{h}_v) \Phi_t^H$;
 - 5: End
-

4.4 Reconstruct mmWave MIMO channel

In this subsection, according to the obtained AoAs $\{\hat{\theta}_l\}_{l=1}^{\bar{L}}$ and AoDs $\{\hat{\phi}_l\}_{l=1}^{\bar{L}}$, the mmWave MIMO channel will be reconstructed as follow. Firstly, we recover the steering vectors $\mathbf{a}_r(\hat{\theta}_l)$ and $\mathbf{a}_t(\hat{\phi}_l)$ in (5) and (6) using the estimated AoAs and AoDs. Secondly, we rewrite the expression (6) as follows

$$\mathbf{H} = (\Phi_t^* \odot \Phi_r) \text{vecd}(\mathbf{H}_v). \tag{42}$$

Thus, the receive signal also can be rewritten using vector form as follows

$$\begin{aligned} \mathbf{y} &= \sqrt{P} \mathbf{W}_r^H ((\Phi_t^* \odot \Phi_r) \text{vecd}(\mathbf{H}_v)) \mathbf{X} + \hat{\mathbf{n}} \\ &= \sqrt{P} (\mathbf{X}^T \Phi_t^* \odot \mathbf{W}_r^H \Phi_r) \mathbf{h}_v + \hat{\mathbf{n}} \\ &= \sqrt{P} \mathbf{Q}^o \mathbf{h}_v + \hat{\mathbf{n}} \end{aligned} \tag{43}$$

where $\mathbf{h}_v = \text{vecd}(\mathbf{H}_v)$, $\mathbf{Q}^o = \mathbf{X}^T \Phi_t^* \odot \mathbf{W}_r^H \Phi_r$. The estimator estimates \mathbf{h}_v in the LS sense. From (35), the LS estimate of \mathbf{h}_v , denoted as $\hat{\mathbf{h}}_v$, is given as follows

$$\hat{\mathbf{h}}_v = \frac{1}{\sqrt{P}} \left((\mathbf{Q}^o)^H \mathbf{Q}^o \right)^{-1} (\mathbf{Q}^o)^H \mathbf{y}. \tag{44}$$

Finally, according to the obtained exact angle estimation, and the gains of path $\hat{\mathbf{h}}_v$ above, we can recover the high-dimensional mmWave MIMO channel as $\hat{\mathbf{H}} = \Phi_r \text{diag}(\mathbf{h}_v) \Phi_t^H$.

4.5 Analysis of computational complexity

Apparently, the complexity of the DGAMP-SBL algorithm is dominated by the E-step, and the matrix multiplications are a big part of it which matrix multiplications by \mathbf{S} , \mathbf{S}^T , Φ , and Φ^T at each iteration. The complexity of each iteration is $O(4 \cdot N_1 N_2 (N_x N_y))$, since we should convert the complex signal to a real signal. It is worth noting that separate operations can reduce the single time, so we also can neglect the coefficient 4. Nonetheless, it is mentioned above that the multiplications operation in Algorithm 1 is taken element-wise. The complexity is much smaller than $O(N_1 N_2 (N_x N_y)^2)$. The complexity of SBL iteration when the dimension of $N_x N_y$ is large. And the refined part does not need

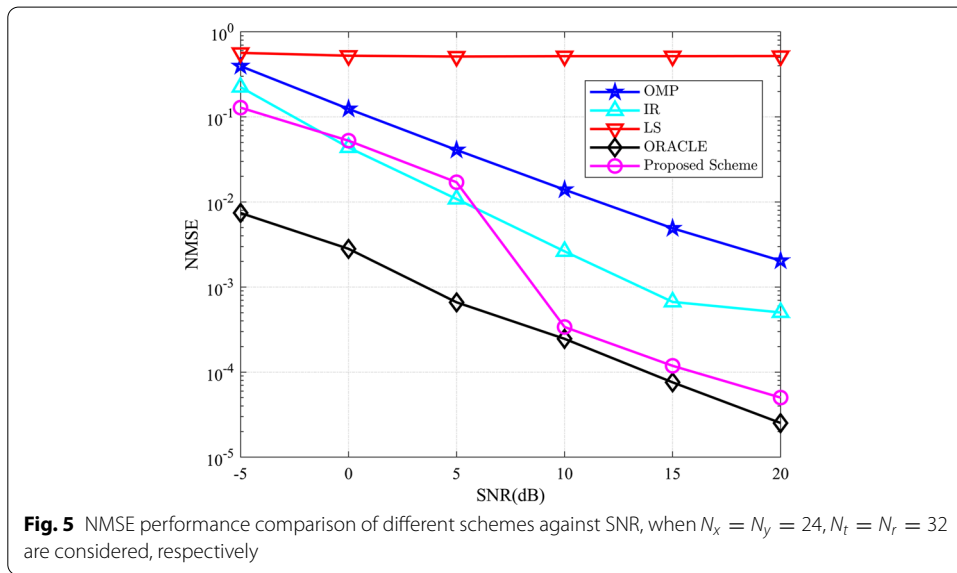
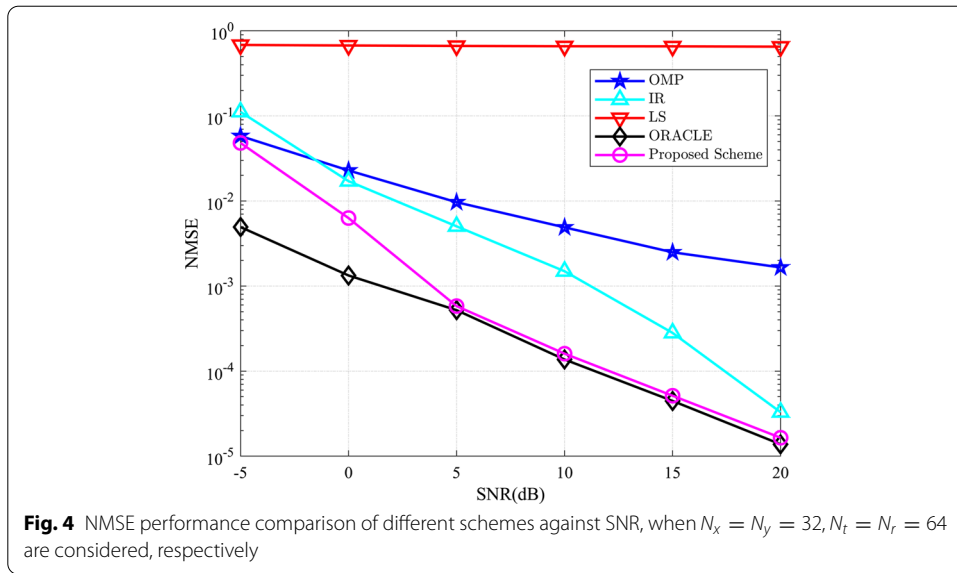
iteration so that it can ignore the complexity. For the OMP scheme, the main computational complexity is $O(N_1 N_2 (N_x N_y))$ which lies in the correlation operation. According to [16], a preprocessing step is proposed to reduce the computational cost of each iteration greatly. For the iterative reweight (IR) scheme, the main computational cost that is contributed by gradient operation is $O(L^2 \cdot (N_x N_y (N_r + N_t)))$. Actually, according to the author's description in [16], the first L maximum correlation angles extracted only according to preprocessing are likely to miss the real angle, especially when L is large. That means the first $2L$ maximum correlation angles are often taken so that the computational complexity can be written as $O(4L^2 \cdot (N_x N_y (N_r + N_t)))$. Finally, for the least square algorithm, the total complexity of LS is $O((N_x N_y)^2 (N_r N_t) + (N_x N_y)^3)$. Due to the large antenna dimension of the channel, it is difficult to use LS for mm-Wave channel estimation.

According to the above analysis, we can find that the computational complexity of DGAMP-SBL is proportional to $N_1 N_2 (N_x N_y)$, which is similar to the OMP algorithm. Obviously, for large $N_x N_y$, the computational costs of DGAMP-SBL are much smaller than SBL. When L is small, the computational complexity of IR is the smallest, but as the value of L becomes larger, the complexity of IR becomes close to that of DGAMP-SBL and OMP.

5 Results and discussion

In this section, we will prove the performance of the proposed algorithm superiority-based channel estimation scheme through MATLAB simulation with the following parameters. The transmitter and the receiver are equipped with $N_r = N_t \in \{32, 64\}$, $N_{\text{RF}} = N_t^{\text{RF}} = N_r^{\text{RF}} = 2$, $N_x = N_y \in \{24, 32\}$ and $N_1 = N_2 = 120$. Each item of the transmitted pilots \mathbf{X} is defined as $x_{i,j} = \sqrt{\rho/N_t} e^{j\beta_{i,j}}$, where ρ is the transmitted power, $\beta_{i,j}$ is the random phase uniformly distributed in $[0, 2\pi)$. The signal-to-noise (SNR) is defined by $\text{SNR} = \frac{\rho \sigma_a^2}{\sigma^2}$. We consider the ULA geometry, and we take the following algorithm as comparison algorithms, i.e., the OMP-based channel estimation [44], the IR-based super-resolution channel estimation scheme [16], the Oracle estimator in [44] and the LS estimator. The normalized mean-squared error (NMSE) which is denoted as $E \left[\left\| \mathbf{H} - \hat{\mathbf{H}} \right\|_F^2 / \left\| \mathbf{H} \right\|_F^2 \right]$ is used for performance comparison. And the achievable spectral efficiency efficiency (ASE) can be defined as $\log_2 \left| \mathbf{I}_{N_{\text{RF}}} + \frac{P}{N_{\text{RF}}} \mathbf{R}_n^{-1} \tilde{\mathbf{W}}^H \mathbf{H} \tilde{\mathbf{F}} \tilde{\mathbf{F}}^H \mathbf{H}^H \tilde{\mathbf{W}} \right|$, for the channel estimation $\hat{\mathbf{H}}$. And $\tilde{\mathbf{F}}$ and $\tilde{\mathbf{W}}$ are defined as the optimal precoder and combiner, respectively, which are designed via the singular value decomposition of $\hat{\mathbf{H}}$, denoted as $\hat{\mathbf{H}} = \hat{\mathbf{U}} \hat{\mathbf{\Sigma}} \hat{\mathbf{V}}^H$. Here the diagonal entries of the singular value matrix $\hat{\mathbf{\Sigma}}$ are arranged in decreasing order. $\tilde{\mathbf{F}}$ and $\tilde{\mathbf{W}}$ consist of the first N_{RF} columns of $\hat{\mathbf{V}}$ and $\hat{\mathbf{U}}$, respectively. The \mathbf{R}_n can be defined as $\mathbf{R}_n \doteq \sigma^2 \tilde{\mathbf{W}}^H \tilde{\mathbf{W}}$. The results in this simulation are all obtained through 200 Monte Carlo experiments [45, 46].

Figures 4 and 5 compare NMSE performance against SNR with $N_r = N_t = 32$, $N_x = 24$ and $N_r = N_t = 64$, $N_x = 32$, respectively. In both scenarios, it is apparent that the curves of the proposed algorithm are below that of comparing algorithms under most circumstances. Comparison results indicate that the proposed scheme has remarkable performance improvement. Note that our proposed scheme almost coincides with the curve of Oracle in the condition of high SNR. Under the same



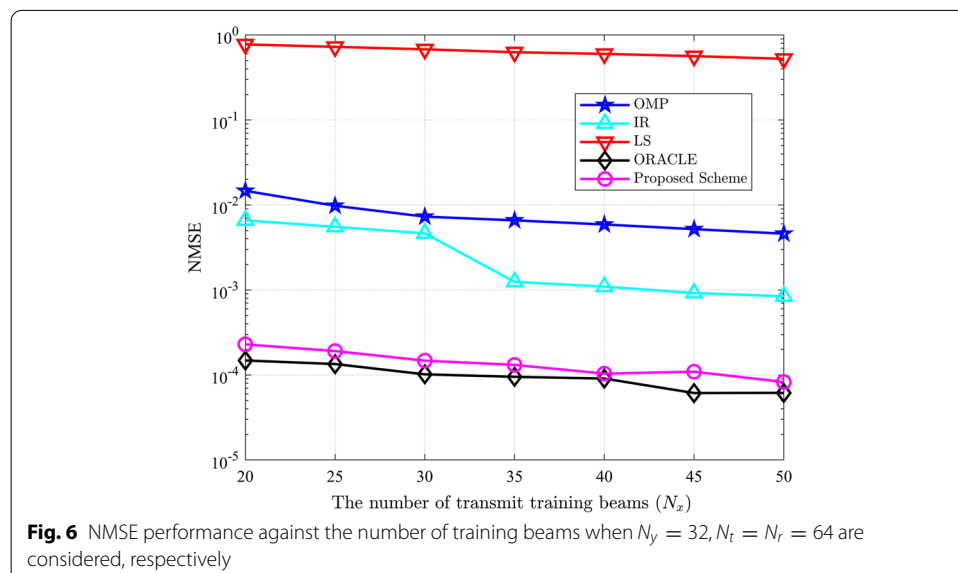
SNR, our algorithm achieves a huge advantage since our algorithm considers the off-grid influence which can estimate a more accurate angle. By comparing Fig. 4, our proposed algorithm achieves a great advantage for all of SNR. And with the increase in SNR, the advantage of our proposed algorithm performance over OMP becomes greater. Moreover, our proposed algorithm is superior to the IR algorithm, and it almost coincides with the Oracle curve. The NMSE performance degrades in low SNR levels because the suffer noise interference exists. And under the low SNR, our proposed algorithm coarse estimation faces a challenge which is brought by noise and off-grid, so the performance does not achieve the expectation. Although the number of antennas and pilots decline, our proposed algorithm still have superior performance. Taking the SNR of 10dB, 15dB, 20dB as an example, our proposed algorithm

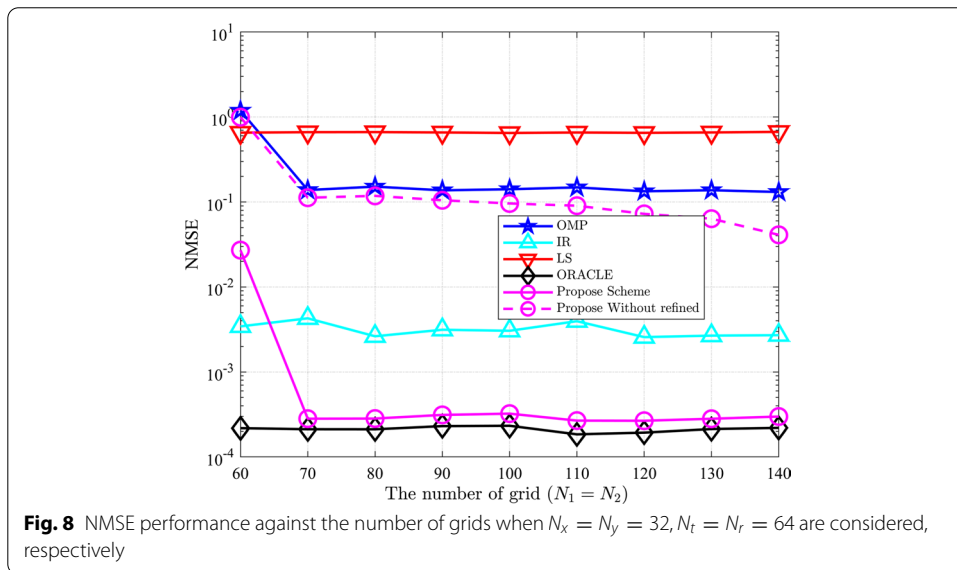
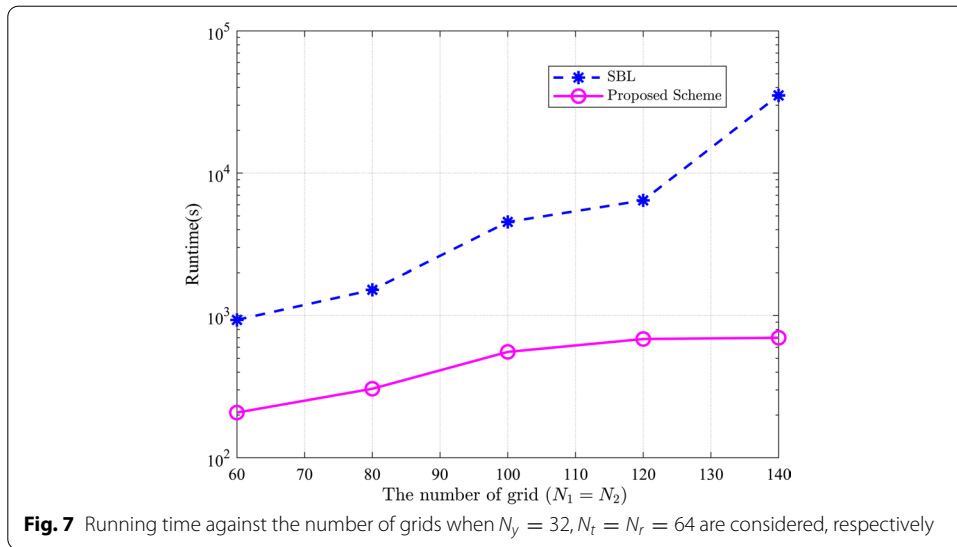
is also very close to the Oracle curve. It can be seen that the channel estimation accuracy based on our proposed algorithm is better than that of comparing algorithm, which verifies the advantages of our proposed schemes. To summarize, the proposed scheme is superior to another scheme in channel estimation accuracy regardless of the antennas and pilots are more or less.

Figure 6 compares the NMSE performance against the number of training beams when SNR=10dB, $N_t = N_r = 64$. As the number of training beams with N_x increases, the NMSE curves of all algorithms decrease monotonically. It is obvious that the comparing algorithm curves tends to decrease slowly. Even so, our proposed algorithm still achieves a huge advantage under different pilots. When the SNR is 10dB, our proposed algorithm achieves high estimation accuracy, and it is very close to the Oracle curve. And we have tested that our algorithm may get an accurate result when N_x is very small. Still, a probability of failure exists comparing with high N_x . Again, among the comparing algorithm, the proposed algorithm still performs better outperforms than others.

Figure 7 compares the difference in runtime between the original SBL and the proposed scheme. In order to facilitate comparison, we fix the number of iterations 200 times. Note that although we apply 200 iterations, the algorithm convergence only needs 100 times and even less in most cases. It is obvious that the proposed algorithm needs much less time than the original SBL algorithm as the dimension increases, which suggests that the proposed algorithm is more practical for large-size sparse matrix recovery.

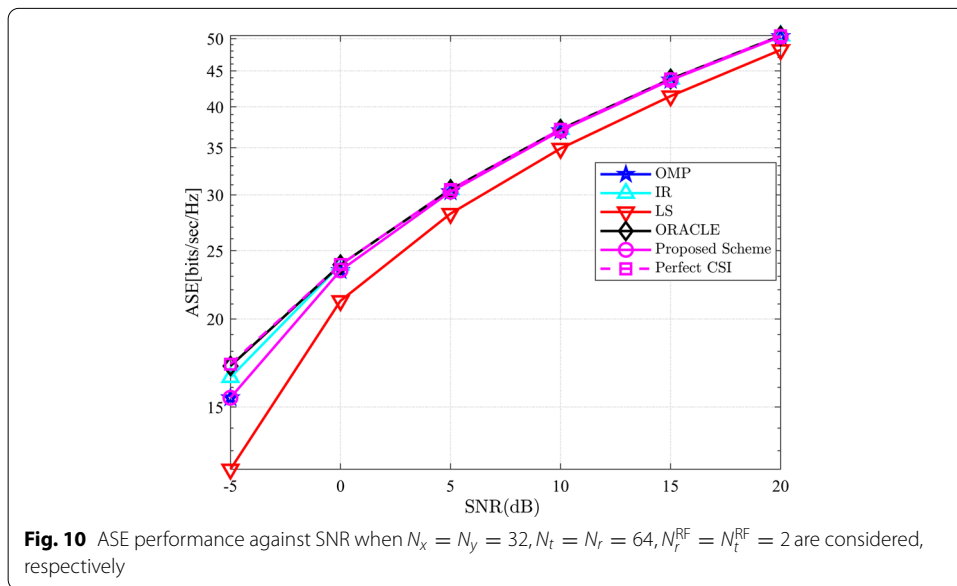
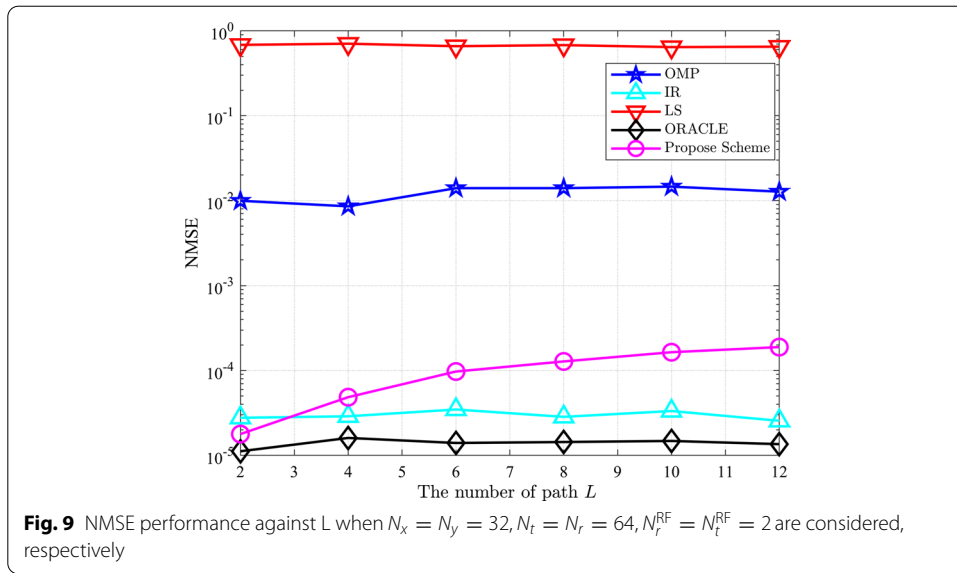
Figure 8 shows the NMSE curves against the number of grids $N_1 = N_2$ when the SNR is 10 dB. We refine the mesh and compare the performance of different algorithms at different mesh spacing. In order to compare the advantages of the second refining algorithm, we add the unrefined DGAMP-SBL to the comparison. It is obvious that the performance of the refined results is greatly improved under any grid spacing. The proposed algorithm outperforms the IR when $N_1 = N_2 \geq 70$. After the number of grids is greater than 70, the NMSE of the proposed algorithm is slightly down. The NMSE curves of the sparse matrix algorithms monotonically decrease with the number of the grids





increasing. It is interesting that the number of grids only needs to be set to 80 to balance the complexity of the computation and the accuracy of estimation. For comparison, the NMSE curves of the Oracle, LS and IR algorithms remain unchanged for different values of grids since all of those work are off-grid.

Figure 9 compares the NMSE curves against the number of paths. When L varies from 2 to 12, the curve of the IR scheme is basically stable since it can still find the most relevant columns. Nevertheless, the number of iterations increases greatly, and as mentioned in Sect. 4.5, with the increase in L , the calculation cost of the IR scheme is proportional to $4L^2$. When L is large, the computation costs required for a single calculation are much bigger than our algorithm to maintain the estimation accuracy. For our proposed algorithm, the NMSE curve rises slightly. This is because when the number of L is large, due to the small adjacent angular distance, it may



lead to the ambiguity of adjacent peaks, which affects the performance of the second step estimation. It is worth noting that our algorithm still has good estimation ability and is much higher than the comparison algorithm OMP.

Figure 10 compares the ASE against SNR when different channel estimation schemes are used with $N_r = N_t = 64, N_x = 32$. We introduce a perfect CSI as a reference upper limit curve. It is noted that although the performance of our proposed algorithm is slightly worse than the IR scheme at -5dB, the ASE curve of the proposed scheme almost overlaps that of CSI for high SNR values.

6 Conclusion

In this paper, we have proposed a high-precision channel estimation scheme for mmWave massive MIMO with hybrid precoding. Specifically, a two-stage strategy is used to estimate the channel. In the first phase of the algorithm, the DGAMP-SBL which simplifies operations is used to estimate the mmWave MIMO channel coarsely. In the second phase of the algorithm, the previously obtained channel which considers off-grid error is refined by 2DFT-interpolation algorithm. Experimental results have indicated that the proposed high-precision channel estimation strategy can perform better than state-of-art estimation precision. For future work, we focus on the sparse channel estimation 2D high-dimension matrix reduction, since the sparse recovery problem still be limited by the 2D sparse dictionary matrix which has a huge matrix dimension.

Abbreviations

EM-SBL: Expectation–maximization sparse Bayesian learning; DGAMP: Damp generalized approximate message passing; mmWave: Millimeter wave; MIMO: Multiple-input multiple-output; SBL: Sparse Bayesian learning; AoDs/AoAs: Angles of departures/arrivals; CSI: Channel state information; ULA: Uniform linear array; DFT: Discrete Fourier transform; CS: Compressive sensing; EM: Expectation–maximization; GAMP: Generalized approximate message passing; LS: Least square; MMSE: Minimum mean-squared error; IR: Iterative reweight; NMSE: Normalized mean-squared error; ASE: Average spectral efficiency.

Acknowledgements

This work was supported by the Hainan University.

Authors' contributions

JS wrote the manuscript, simulated and modified the experiments; XW and XL helped to conceive the idea and revise the manuscript, the theory and article writing; ZH and TS contributed to the development of the ideas. All authors read and approved the final manuscript.

Funding

This work was supported by the Natural Science Foundation of Hainan Province(620RC555), National Natural Science Foundation of China (Nos. 61861015 and 61961013), Key Research and Development Program of Hainan Province (No. ZDYF2019011), National Key Research and Development Program of China (No. 2019CXTD400), Young Elite Scientists Sponsorship Program by CAST (No. 2018QNR001) and the Scientific Research Setup Fund of Hainan University (No. KYQD(ZR) 1731).

Availability of data and materials

The datasets simulated and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 20 June 2021 Accepted: 6 September 2021

Published online: 24 September 2021

References

1. I.A. Hemadeh, K. Satyanarayana, M. El-Hajjar, L. Hanzo, Millimeter-wave communications: physical channel models, design considerations, antenna constructions, and link-budget. *IEEE Commun. Surv. Tutor.* **20**(2), 870–913 (2017)
2. C. Xue, S. He, Y. Huang, Y. Wu, L. Yang, An efficient beam-training scheme for the optimally designed subarray structure in mmWave LoS MIMO systems. *EURASIP J. Wirel. Commun. Netw.* **2017**(1), 1–12 (2017)
3. R.W. Heath, N. Gonzalez-Prelcic, S. Rangan, W. Roh, A.M. Sayeed, An overview of signal processing techniques for millimeter wave MIMO systems. *IEEE J. Sel. Top. Signal Process.* **10**(3), 436–453 (2016)
4. I.-S. Kim, J. Choi, Channel estimation via gradient pursuit for mmWave massive MIMO systems with one-bit ADCs. *EURASIP J. Wirel. Commun. Netw.* **2019**(1), 1–16 (2019)
5. J. Dai, L. Zhou, C. Chang, W. Xu, Robust Bayesian learning approach for massive MIMO channel estimation. *Signal Process.* **168**, 107345 (2020)

6. X. Rao, V.K. Lau, Distributed compressive CSIT estimation and feedback for FDD multi-user massive MIMO systems. *IEEE Trans. Signal Process.* **62**(12), 3261–3271 (2014)
7. Z. Chen, C. Yang, Pilot decontamination in wideband massive MIMO systems by exploiting channel sparsity. *IEEE Trans. Wirel. Commun.* **15**(7), 5087–5100 (2016)
8. J.-C. Shen, J. Zhang, E. Alsusa, K.B. Letaief, Compressed CSI acquisition in FDD massive MIMO: how much training is needed? *IEEE Trans. Wirel. Commun.* **15**(6), 4145–4156 (2016)
9. E.J. Candès, J. Romberg, T. Tao, Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inf. Theory* **52**(2), 489–509 (2006)
10. Y. Tsaig, D.L. Donoho, Compressed sensing (2004)
11. C. Huang, L. Liu, C. Yuen, S. Sun, Iterative channel estimation using LSE and sparse message passing for mmWave MIMO systems. *IEEE Trans. Signal Process.* **67**(1), 245–259 (2018)
12. Z. Zhang, B.D. Rao, Sparse signal recovery with temporally correlated source vectors using sparse Bayesian learning. *IEEE J. Sel. Top. Signal Process.* **5**(5), 912–926 (2011)
13. R. Tibshirani, Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B (Methodol.)* **58**(1), 267–288 (1996)
14. J. Dai, A. Liu, V.K. Lau, FDD massive MIMO channel estimation with arbitrary 2D-array geometry. *IEEE Trans. Signal Process.* **66**(10), 2584–2599 (2018)
15. J. Fang, F. Wang, Y. Shen, H. Li, R.S. Blum, Super-resolution compressed sensing for line spectral estimation: an iterative reweighted approach. *IEEE Trans. Signal Process.* **64**(18), 4649–4662 (2016)
16. C. Hu, L. Dai, T. Mir, Z. Gao, J. Fang, Super-resolution channel estimation for mmWave massive MIMO with hybrid precoding. *IEEE Trans. Veh. Technol.* **67**(9), 8954–8958 (2018)
17. X. Cheng, C. Tang, Z. Zhang, Accurate channel estimation for millimeter-wave MIMO systems. *IEEE Trans. Veh. Technol.* **68**(5), 5159–5163 (2019)
18. L. Wan, X. Kong, F. Xia, Joint range-Doppler-angle estimation for intelligent tracking of moving aerial targets. *IEEE Internet Things J.* **5**(3), 1625–1636 (2017)
19. X. Wang, L.T. Yang, D. Meng, M. Dong, K. Ota, H. Wang, Multi-UAV cooperative localization for marine targets based on weighted subspace fitting in SAGIN environment. *IEEE Internet Things J.* (2021)
20. L. Wan, Y. Sun, L. Sun, Z. Ning, J.J. Rodrigues, Deep learning based autonomous vehicle super resolution DOA estimation for safety driving. *IEEE Trans. Intell. Transp. Syst.* (2020)
21. D.L. Donoho, A. Maleki, A. Montanari, Message-passing algorithms for compressed sensing. *Proc. Natl. Acad. Sci.* **106**(45), 18914–18919 (2009)
22. S. Rangan, Generalized approximate message passing for estimation with random linear mixing. In: 2011 IEEE International Symposium on Information Theory Proceedings (IEEE, 2011), pp. 2168–2172
23. S. Mumtaz, J. Rodriguez, L. Dai, *MmWave Massive MIMO: A Paradigm for 5G* (Academic Press, London, 2016)
24. O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, R.W. Heath, Spatially sparse precoding in millimeter wave MIMO systems. *IEEE Trans. Wirel. Commun.* **13**(3), 1499–1513 (2014)
25. X. Gao, L. Dai, S. Han, I. Chih-Lin, R.W. Heath, Energy-efficient hybrid analog and digital precoding for mmWave MIMO systems with large antenna arrays. *IEEE J. Sel. Areas Commun.* **34**(4), 998–1009 (2016)
26. R. Méndez-Rial, C. Rusu, A. Alkhateeb, N. González-Prelcic, R.W. Heath, Channel estimation and hybrid combining for mmWave: Phase shifters or switches? In: 2015 Information Theory and Applications Workshop (ITA) (IEEE, 2015), pp. 90–97
27. R. Méndez-Rial, C. Rusu, N. González-Prelcic, A. Alkhateeb, R.W. Heath, Hybrid MIMO architectures for millimeter wave communications: phase shifters or switches? *IEEE Access* **4**, 247–267 (2016)
28. X. Cheng, M. Wang, S. Li, Compressive sensing-based beamforming for millimeter-wave OFDM systems. *IEEE Trans. Commun.* **65**(1), 371–386 (2016)
29. S. Hur, T. Kim, D.J. Love, J.V. Krogmeier, T.A. Thomas, A. Ghosh, Millimeter wave beamforming for wireless backhaul and access in small cell networks. *IEEE Trans. Commun.* **61**(10), 4391–4403 (2013)
30. T. Kim, D.J. Love, Virtual AoA and AoD estimation for sparse millimeter wave MIMO channels. In: 2015 IEEE 16th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC) (IEEE, 2015), pp. 146–150
31. A. Alkhateeb, O. El Ayach, G. Leus, R.W. Heath, Channel estimation and hybrid precoding for millimeter wave cellular systems. *IEEE J. Sel. Top. Signal Process.* **8**(5), 831–846 (2014)
32. J. Shi, F. Wen, T. Liu, Nested MIMO radar: coarrays, tensor modeling and angle estimation. *IEEE Trans. Aerosp. Electron. Syst.* (2020)
33. X. Wang, M. Huang, L. Wan, Joint 2D-DOD and 2D-DOA estimation for coprime EMVS-MIMO radar. *Circuits Syst. Signal Process.* **1–17**(2021)
34. L. Wan, K. Liu, Y.-C. Liang, T. Zhu, DOA and polarization estimation for non-circular signals in 3-D millimeter wave polarized massive MIMO systems. *IEEE Trans. Wirel. Commun.* **20**(5), 3152–3167 (2021)
35. M.E. Tipping, Sparse Bayesian learning and the relevance vector machine. *J. Mach. Learn. Res.* **1**(Jun), 211–244 (2001)
36. S. Ji, Y. Xue, L. Carin, Bayesian compressive sensing. *IEEE Trans. Signal Process.* **56**(6), 2346–2356 (2008)
37. B.D. Rao, K. Kreutz-Delgado, An affine scaling methodology for best basis selection. *IEEE Trans. Signal Process.* **47**(1), 187–200 (1999)
38. M. Al-Shoukairi, B. Rao, Sparse Bayesian learning using approximate message passing. In: 2014 48th Asilomar Conference on Signals, Systems and Computers (IEEE, 2014), pp. 1957–1961
39. F. Li, J. Fang, H. Duan, Z. Chen, H. Li, Computationally efficient sparse Bayesian learning via generalized approximate message passing. [arXiv:1501.04762](https://arxiv.org/abs/1501.04762) (2015)
40. S. Rangan, P. Schniter, A.K. Fletcher, S. Sarkar, On the convergence of approximate message passing with arbitrary matrices. *IEEE Trans. Inf. Theory* **65**(9), 5339–5351 (2019)
41. M. Al-Shoukairi, P. Schniter, B.D. Rao, A GAMP-based low complexity sparse Bayesian learning algorithm. *IEEE Trans. Signal Process.* **66**(2), 294–308 (2017)
42. L. Fan, G. Qi, Frequency estimator of sinusoid based on interpolation of three DFT spectral lines. *Signal Process.* **144**, 52–60 (2018)

43. V. Popović-Bugarin, S. Djukanović, A low complexity model order and frequency estimation of multiple 2-D complex sinusoids. *Digital Signal Process.* **104**, 102794 (2020)
44. J. Lee, G.-T. Gil, Y.H. Lee, Channel estimation via orthogonal matching pursuit for hybrid MIMO systems in millimeter wave communications. *IEEE Trans. Commun.* **64**(6), 2370–2386 (2016)
45. J. Cong, X. Wang, M. Huang, L. Wan, Robust DOA estimation method for MIMO radar via deep neural networks. *IEEE Sens. J.* (2020)
46. J. Cong, X. Wang, X. Lan, M. Huang, L. Wan, Fast target localization method for FMCW MIMO radar via VDSR neural network. *Remote Sens.* **13**(10), 1956 (2021)

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)
