Received 20 July 2024; revised 22 August 2024; accepted 3 September 2024. Date of publication 10 September 2024; date of current version 19 September 2024.

Digital Object Identifier 10.1109/OJCOMS.2024.3457507

Sparse Bayesian Learning Using Complex t-Prior for Beam-Domain Massive MIMO Channel Estimation

KENGO FURUTA (Graduate Student Member, IEEE), TAKUMI TAKAHASHI (Member, IEEE), AND HIDEKI OCHIAI (Fellow, IEEE)

Graduate School of Engineering, Osaka University, Suita 565-0871, Japan

CORRESPONDING AUTHOR: T. TAKAHASHI (e-mail: takahashi@comm.eng.osaka-u.ac.jp)

This work was supported in part by the JSPS KAKENHI under Grant JP23K13335, Grant JP23K22754, and Grant JP23K20935; in part by MIC/FORWARD under Grant JPMI240710001; and in part by JST SPRING under Grant JPMJSP2138.

ABSTRACT This paper proposes a novel beam-domain channel estimation (CE) algorithm via sparse Bayesian learning (SBL) using complex *t*-prior for massive multi-user multiple-input multiple-output (MIMO) systems. Due to the sidelobe leakage and insufficient observation resolution resulting from physical constraints, the equivalent channel after digital beamforming at the receiver has a structure with many small but non-zero elements, which cannot be modeled strictly as a sparse signal. To fully capture this *pseudo-sparse* structure characterized by the signal strength variations among elements, we design a novel SBL algorithm that incorporates a complex *t*-distribution using a hierarchical Bayesian model. By utilizing a high degree of adaptability of this heavy-tailed prior, it is possible to efficiently learn the signal strength, accounting for elements with non-zero but small values, which is verified by the regularization analysis based on an equivalent optimization problem. The efficacy of the proposed CE algorithm is confirmed by numerical simulations, which show that the proposed method not only significantly outperforms the state-of-the-art (SotA) sparse signal recovery (SSR)-based algorithms but also achieves the performance of a genie-aided scheme over a wide signal-to-noise ratio (SNR) range in both sub-6 GHz and millimeter-wave (mmWave) wireless communication scenarios.

INDEX TERMS Massive MIMO, channel estimation, sparse Bayesian learning, hierarchical Bayesian model, complex *t*-distribution, beam-domain signal processing.

I. INTRODUCTION

MASSIVE multiple-input multiple-output (MIMO) systems, where a base station (BS) is equipped with a massive number of antenna arrays, have been considered as one of the essential technologies in the fifth generation (5G) advanced and future sixth generation (6G) networks [1], [2], [3]. To take full advantage of the significant amount of spatial degrees of freedom (DoF) offered by massive MIMO antenna arrays, it is necessary to acquire highly accurate channel state information (CSI). However, reliable channel estimation (CE) using orthogonal pilot sequences inherently requires a large number of pilot/training symbols proportional to the number of antenna elements in the massive MIMO system, which increases the communication overhead required for CSI acquisition [4], [5], [6]. To tackle this issue, various beam-domain CE schemes that leverage the statistical structure of the massive MIMO channel have been proposed [7].

By using an appropriate basis as a digital beamformer, the beam-domain massive MIMO channel exhibits sparsity due to the enhanced spatial resolution provided by a large number of antenna elements mounted on a BS [8], [9]. For instance, in massive multi-user MIMO communications to accommodate a large number of wireless terminals [10], [11], it is well-known that the angular spread of the received signal from each user equipment (UE) as viewed from a BS is small as a result of local scatterers around the UE and high placement of the BS receiver, so that the wireless channel in the angular domain has a *sparse-like* structure even in the sub-6 GHz band [12], [13]. Therefore, by using a discrete Fourier transform (DFT) basis for full digital beamforming at the receiver [14], a similar structure can be found in

the equivalent beam-domain massive MIMO channel, which may be exploited to reduce pilot overhead [15].

Such a preferable structure in the beam-domain channel is particularly prominent in the millimeter-wave (mmWave) band, where there is no scattering and signals propagate through a small number of path clusters from the transmitter to the receiver [16], [17], [18]. Many low-overhead mmWave massive MIMO CE schemes that take advantage of this statistical property have been investigated. For instance, when the resolution of a small number of paths is sufficiently high and a low-rank representation of the wireless channel based on the steering vector is possible, the tensor-based CE via physical parameter estimation enables highly accurate estimation [19], [20]. In addition, a pilot overhead reduction method that exploits the sparsity of the MIMO channel in the virtual angular domain is proposed in [21], a CE method based on two-stage compressed sensing that exploits both the sparsity and low-rank structure of the angular domain channel is studied in [22], whereas in [23], the authors propose a beam-based statistical channel model (BSCM)-based method for estimating long-term statistics of the mmWave MIMO channel without knowledge of instantaneous CSI. Most of these existing studies have achieved a significant amount of reduction in pilot overhead while maintaining CE accuracy by solving problems induced by approximating the beam-domain CE as a sparse signal recovery (SSR) problem (or by modeling a beam-domain channel as a sparse signal).

However, in practice, it is not always possible to accurately replace the beam-domain channel by a strictly sparse signal model. Due to physical constraints on the size, weight, and especially power consumption of massive MIMO arrays, it may not be practical to set the number of radio frequency (RF) chains equal to that of antenna elements; hence, the actual observation resolution that can be achieved via a digital signal processor (DSP) unit may not be very high [24]. Consequently, in multi-user MIMO systems operating below 6 GHz, there are not many small beam-domain channel coefficients that can be treated as zeros with high accuracy due to angular spread and attenuated sidelobe leakage [13]. Even in mmWave MIMO systems, when the spatial sampling resolution is not sufficient, it is often inaccurate to apply a sparse signal model to the beam-domain channel since signal energy does not concentrate on the beam coefficient corresponding to the angle of arrival (AoA), but rather leaks into multiple beams in the surrounding region. The effect is more pronounced in multi-user MIMO scenarios where many UEs communicate with a BS simultaneously.

In light of the above, the beam-domain channel may not be precisely characterized by an explicit sparse model consisting of zero/non-zero elements, but by a model having differences in signal strength among elements with many small but non-zero elements, and such a statistical structure is henceforth referred to as *pseudo-sparsity* in this paper. The resulting distributions are usually difficult to define analytically, making it difficult to apply convex optimization via the fast iterative shrinkage thresholding algorithm (FISTA) [25], the gradient method [26], the primal-dual method [27], and/or Bayesian inference using message-passing algorithms [28], [29], [30]. Therefore, for the purpose of pseudo-sparse CE that cannot be attributed to the SSR problem, this paper designs a CE algorithm based on the framework of sparse Bayesian learning (SBL) [31], which can improve the estimation accuracy while estimating (learning) the prior distribution of unknown signals.

The SBL framework, which models the prior distribution of unknown signals and estimates them while learning the model parameters, was first proposed in [31] as a method for solving the SSR problem and has been used in various fields due to its broad applicability. A notable method that employs the SBL framework is the Bayesian compressive sensing algorithm proposed in [32], which is designed using the fact that maximum *a-posteriori* (MAP) estimation under Gaussian observation is equivalent to least absolute shrinkage and selection operator (LASSO) regression when a Laplace distribution is chosen as the prior distribution of the unknown signal. While the objective of [32] is the demonstration of the effectiveness of SSR based on the SBL framework in the context of image processing, another major contribution is the presentation of a systematic procedure for introducing a provisional prior distribution when designing SBL algorithms. Specifically, a hierarchical Bayesian modeling was proposed to represent the desired prior distribution by cascading multiple distributions based on the chain rule so that the relationship between the observation likelihood and the conjugate prior distribution can be maintained. As a result, in addition to SBL algorithm design based on classical Gaussian modeling [33], [34], there has been active research on improving performance by incorporating in the algorithm design a provisional prior distribution that can better reflect the statistical properties of the unknown variables, and in physical layer signal processing for wireless communications, the SBL algorithms have been applied mainly to CE [35], [36], channel tracking [37], [38], and direction of arrival (DoA) estimation [39], [40].

In light of the above literature, this paper proposes a novel SBL-based beam-domain massive MIMO CE algorithm that incorporates a complex *t*-distribution tailored for *pseudo-sparse* signal recovery via regularization analysis [41], [42] as a provisional prior distribution. To clarify the contribution of this paper, in the following paragraphs, we will describe the motivation for proposing this method and its novelty compared to the state-of-the-art (SotA) alternatives.

Turning to the literature dealing with CE that is closely related to this paper, relevance vector machine (RVM) [31], which is a classical Gaussian modeling-based SBL algorithm without explicitly setting a provisional prior distribution, was applied to mmWave massive MIMO CE in [33], [34]. In particular, it has been shown in [43] that accurate estimation performance can be achieved using a beam-domain massive MIMO CE algorithm designed based on RVM, with a problem setting almost identical to that of this

paper. On the other hand, there are many studies that aim to improve estimation accuracy by introducing a provisional prior distribution that reflects the statistical properties of the wireless channel, but the ones that are particularly closely related to this study are [35] and [36]. Although the problem setting and algorithm derivation methods are different in these studies, they have designed SBL algorithms that use the complex t-distribution [44] as a provisional prior distribution, similar to the CE algorithm proposed in this paper, and their performance has been shown to outperform the conventional alternatives.

At this point, it should be important to note that all the above SBL algorithms have been proposed as a means of solving the SSR problem, regardless of the provisional prior distribution. In designing the SBL algorithm for solving the SSR problem, it is natural to use the Laplace prior, but in fact, the t-distribution is also used as a prior for solving the SSR problem. Indeed, it is common practice to set the hyperparameters of the Gamma distribution to extremely small values (*e.g.*, 0, 10^{-4} in [31] and 10^{-6} in [36]) in the second layer of the hierarchical Bayesian model that constitutes the t-distribution, which is equivalent to setting the kurtosis of the t-distribution as large as possible. In other words, the complex t-distribution is currently only used as a provisional prior distribution for solving SSR problems.

Motivated by the current limited use of the provisional prior distributions, we hypothesize that it may be possible to design an SBL algorithm that achieves high estimation accuracy by taking advantage of a high degree of adaptability of the complex t-distribution, even for the CE problem that cannot be strictly formulated as an SSR problem. The complex t-distribution is a heavy-tailed distribution with DoF as a parameter, which is known to represent a wide range of probability distributions by appropriately adjusting the parameter [44]. Therefore, in this paper, we design a complex t-distribution with an appropriately adjusted DoF parameter so that it can reflect the pseudo-sparsity of the beam-domain massive MIMO channels that appears under realistic constraints. If we can construct a hierarchical Bayesian model using the resultant complex t-distribution as a provisional prior distribution and derive an SBL algorithm, it should be able to design a novel beam-domain massive MIMO CE scheme that can capture the pseudo-sparsity.

The contributions of the paper from an algorithmic perspective are summarized as follows:

- A novel CE method is designed based on an SBL algorithm that uses a complex t-distribution as the prior distribution, and this distribution is appropriately pre-adjusted via the regularization analysis so that it can capture the pseudo-sparsity of the beam-domain massive MIMO channel under realistic constraints. The two-layer hierarchical Bayesian model is constructed by assigning a complex Gaussian distribution to each channel coefficient and an inverse-Gamma distribution to its variance parameter, and then the MAP criterion-based estimation is performed by alternating between posterior estimation and approximate likelihood maximization of the distribution parameters. It should be noted that this hierarchical Bayesian modeling differs from the existing method (proposed in [31] and adopted in [35], [36]) that uses a Gamma distribution for the precision of the Gaussian distribution in the first layer. In addition, the proposed method derives an update formula for the distribution parameters based on maximizing the *marginal distribution* as in [45], which is also different from the cost function used in [31].
- To analyze the behavior of the proposed method, the regularization (*i.e.*, penalty term) in the optimization problem equivalent to the designed SBL algorithm is derived, and its variation with respect to (w.r.t.) the DoF parameter is investigated. Consequently, we clarify what statistical properties of unknown variables can be assumed by changing the DoF parameter.
- To confirm the efficacy of the proposed CE algorithm via SBL using the complex t-prior in massive MIMO systems, we compare the performance of the conventional and proposed methods in both sub-6 GHz and mmWave wireless communication scenarios. For comparison with the SotA SBL schemes that operate based on SSR, we also design an SBL algorithm using complex Laplace prior [46], [47] by recasting the compressive sensing algorithm in [32] on the complex domain.[1] Simulation results show that for all the system parameters evaluated, the proposed method outperforms the SotA alternatives and approaches the performance of the idealized scheme in which long-term channel statistics are perfectly known.

To the best of our knowledge, this is the first paper that focuses on the pseudo-sparse structure of beam-domain massive MIMO channels under realistic constraints and designs an SBL algorithm that uses a complex t-prior with appropriately adjusted hyperparameters, *i.e.*, the DoF parameter, for the purpose of utilizing a high degree of adaptability of the t-distribution rather than as a prior distribution for solving an SSR problem.

II. PRELIMINARIES

A. MATHEMATICAL NOTATIONS

The following notation is adopted throughout this paper, unless otherwise specified. Sets of real and complex numbers are denoted by $\mathbb{R}$ and $\mathbb{C}$, respectively. Vectors and matrices are denoted in lower-and upper-case bold-face fonts, respectively. The transpose and conjugate transpose operators are denoted by $\cdot^{\mathsf{T}}$ and $\cdot^{\mathsf{H}}$, respectively. The complex Gaussian distribution with a mean vector $\boldsymbol{a}$ and a covariance matrix $\boldsymbol{A}$ is denoted by $\mathcal{CN}(\boldsymbol{a}, \boldsymbol{A})$. The Gamma and inverse-Gamma distribution with a shape parameter a and a scale parameter b is denoted by $\Gamma(a, b)$ and $\Gamma^{-1}(a, b)$, respectively. The

[1]This method is one of the original contributions of this paper, but its detailed description is placed in Appendix-B for the sake of readability.

expected value of a random variable is denoted by $\mathbb{E}[\cdot]$. The $a \times a$ square identity matrix is denoted by $\boldsymbol{I}_a$. The diagonal matrix constructed by placing the elements of a vector $\boldsymbol{a}$ on its main diagonal is denoted by diag[$\boldsymbol{a}$]. The vector formed by stacking the column vectors of a matrix $\boldsymbol{A}$ vertically is denoted by vec[$\boldsymbol{A}$]. The determinant of a matrix $\boldsymbol{A}$ is denoted by $|\boldsymbol{A}|$. ℓ_0-, ℓ_1-, and ℓ_2-norms are denoted by $||\cdot||_0$, $||\cdot||_1$, and $||\cdot||_2$, respectively. The Kronecker product operation is denoted by $\otimes$. The Dirac delta function is denoted by $\delta(\cdot)$. The Gamma function is denoted by $\Gamma(\cdot)$, which is different from the Gamma distribution defined above. The imaginary unit is denoted by $\mathrm{j} \triangleq \sqrt{-1}$.

B. ANTENNA-DOMAIN SIGNAL MODEL

Consider an uplink multi-user MIMO system composed of a BS, having N receive (RX) antennas and M UE devices equipped with a single transmit (TX) antenna. The UE sends the known pilot (training) symbol matrix $\boldsymbol{X} \in \mathbb{C}^{K\times M}$ over a discrete time slot k ($1 \le k \le K$) via an antenna-domain MIMO channel expressed as $\tilde{\boldsymbol{H}} \in \mathbb{C}^{M\times N}$ and is observed as the RX symbol matrix, which is given as

$$\tilde{\boldsymbol{Y}} = \tilde{\boldsymbol{H}}\boldsymbol{X} + \tilde{\boldsymbol{W}}, \tag{1}$$

where $\tilde{\boldsymbol{W}} \in \mathbb{C}^{K\times N}$ is the additive white Gaussian noise (AWGN) matrix, each column vector of which follows $\mathcal{CN}(\boldsymbol{0}, N_0\boldsymbol{I}_N)$, and N_0 corresponds to the spectral density of AWGN, which is assumed to be known at the BS receiver. For later convenience, we define the precision as $\beta \triangleq \frac{1}{N_0}$.

C. BEAM-DOMAIN SIGNAL MODEL

At the receiver, full digital beamforming is performed. Using the DFT beamforming with the DFT matrix $\boldsymbol{D} \in \mathbb{C}^{N\times N}$, the RX symbol matrix in the beam domain can be expressed as [13], [48]

$$\boldsymbol{Y} = \boldsymbol{D}^{\mathsf{H}}\tilde{\boldsymbol{Y}} = \boldsymbol{D}^{\mathsf{H}}\tilde{\boldsymbol{H}}\boldsymbol{X} + \boldsymbol{D}^{\mathsf{H}}\tilde{\boldsymbol{W}} = \boldsymbol{H}\boldsymbol{X} + \boldsymbol{W}, \tag{2}$$

where $\boldsymbol{H} \triangleq \boldsymbol{D}^{\mathsf{H}}\tilde{\boldsymbol{H}}$ and $\boldsymbol{W} \triangleq \boldsymbol{D}^{\mathsf{H}}\tilde{\boldsymbol{W}}$ are the beam-domain channel matrix and noise, respectively. Using the identity $\mathrm{vec}[\boldsymbol{ABC}] = (\boldsymbol{C}^{\mathsf{T}} \otimes \boldsymbol{A})\mathrm{vec}[\boldsymbol{B}]$, the vector form of (2) can be expressed as

$$\begin{aligned}\boldsymbol{y} &\triangleq \left[y_1, \ldots, y_{KN}\right]^{\mathsf{T}} \\ &= \mathrm{vec}(\boldsymbol{Y}) = \left(\boldsymbol{X}^{\mathsf{T}} \otimes \boldsymbol{I}_N\right)\mathrm{vec}(\boldsymbol{H}) + \mathrm{vec}(\boldsymbol{W}), \\ &= \boldsymbol{S}\boldsymbol{h} + \boldsymbol{w},\end{aligned} \tag{3}$$

where $\boldsymbol{S}$, $\boldsymbol{h}$, and $\boldsymbol{w}$ are respectively defined as

$$\boldsymbol{h} \triangleq [h_1, \ldots, h_{MN}]^{\mathsf{T}} = \mathrm{vec}(\boldsymbol{H}) \in \mathbb{C}^{MN\times 1}, \tag{4a}$$

$$\boldsymbol{S} \triangleq \left(\boldsymbol{X}^{\mathsf{T}} \otimes \boldsymbol{I}_N\right) \in \mathbb{C}^{KN\times MN}, \tag{4b}$$

$$\boldsymbol{w} \triangleq [w_1, \ldots, w_{KN}]^{\mathsf{T}} = \mathrm{vec}(\boldsymbol{W}) \in \mathbb{C}^{KN\times 1}. \tag{4c}$$

Based on (3), the goal of this paper can now be stated concisely: Design a novel CE algorithm that enables the BS receiver to accurately estimate the beam-domain channel vector $\boldsymbol{h}$, given the effective observation matrix $\boldsymbol{S}$ and the beam-domain RX vector $\boldsymbol{y}$.

D. PSEUDO-SPARSITY OF BEAM-DOMAIN CHANNELS

Before moving on to the specific derivation of the CE algorithm, we describe the statistical structure of the beam-domain channels based on the two massive MIMO channel models corresponding to sub-6 GHz and mmWave wireless communications.

1) GEOMETRICAL ONE-RING MODEL

As a model to represent the spatial correlation among fading coefficients of wireless channels in a massive multi-user MIMO system operating in sub-6 GHz bands, we use the geometrical one-ring model [8], [49], without loss of generality in (3). The wireless channels between the BS and UE exhibit a small angular spread from the perspective of the BS as a result of local scatterers around the UE and the high placement of the BS antennas. Assume that isotropic scatterers exist uniformly on the two-dimensional (2D) plane around each UE and waves from the m-th UE arrive within an angular spread of $\Delta\psi_m \triangleq \psi_m^{\max} - \psi_m^{\min}$, where $\psi_m^{\max}$ and $\psi_m^{\min}$ are the maximum and minimum angles of arrival of the signal from the m-th UE. When the antenna element spacing is fixed to half the wavelength, the (i,j)-th element in the RX correlation matrix of the m-th UE $\boldsymbol{\Theta}_m \in \mathbb{C}^{N\times N}$ is expressed as

$$[\boldsymbol{\Theta}_m]_{i,j} \triangleq \frac{1}{\Delta\psi_m}\int_{\psi_m^{\min}}^{\psi_m^{\max}} e^{\mathrm{j}\pi(i-j)\cos(\psi)}d\psi, \tag{5}$$

which denotes the correlation coefficient between the i-th and j-th RX antenna elements. The m-th column vector of $\tilde{\boldsymbol{H}}$ can be expressed as

$$\tilde{\boldsymbol{h}}_m = \boldsymbol{\Theta}_m^{1/2}\boldsymbol{\epsilon}_m, \tag{6}$$

where $\boldsymbol{\epsilon}_m \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I}_N)$ represents small-scale fading for the m-th UE.

2) FINITE PATH MODEL

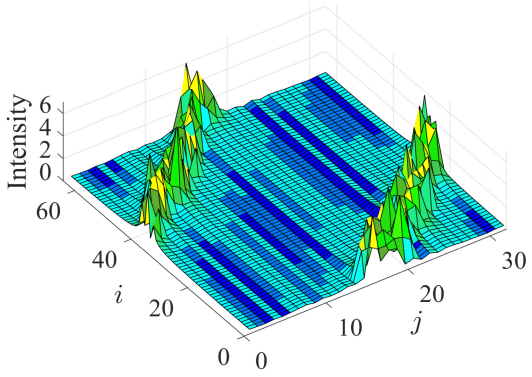
In mmWave wireless communications, where diffraction and scattering rarely occur, the number of paths arriving at the receiver is limited [16], [17], [18]. To represent such wireless channels, we use the finite path model [50], [51], where $\tilde{\boldsymbol{H}}$ is given by a geometrical model with L scatters with each scatter corresponding to a single propagation path between the BS and UE. Following the literature, we have

$$\tilde{\boldsymbol{h}}_m = \sum_{l=1}^{L} \zeta_{m,l}\boldsymbol{a}_{m,l}\left(\omega_{m,l}\right), \tag{7}$$

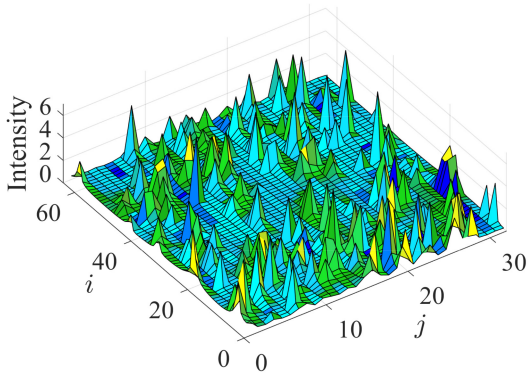
with the antenna steering vector

$$\boldsymbol{a}\left(\omega_{m,l}\right) \triangleq \left[1, \mathrm{e}^{\mathrm{j}\omega_{m,l}}, \ldots, \mathrm{e}^{\mathrm{j}(N-1)\omega_{m,l}}\right]^{\mathsf{T}}, \tag{8}$$

where $\zeta_{m,l} \sim \mathcal{CN}(0,1)$ represents the complex path gain along the l-th path of the m-th UE on the basis of slow TX power control [50], [52], and $\omega_{m,l} \triangleq \pi\sin\theta_{m,l}$ with $\theta_{m,l}$ denoting the azimuth angle of the l-th path of the m-th UE. The antenna element space is fixed to half the wavelength.



(a) Geometrical one-ring model



(b) Finite path model

FIGURE 1. Intensity of the elements in the beam-domain channel H, where the MIMO configuration is set to $(M, N) = (32, 64)$.

3) STATISTICAL STRUCTURE OF BEAM-DOMAIN CHANNELS

Fig. 1 visualizes the intensity of the elements of the MIMO channel generated according to the respective channel models, where the MIMO configuration is set to $(M, N) = (32, 64)$. The receiver is equipped with a uniform linear array (ULA) of sectors partitioned by a 120-degree opening, and the UE devices under slow TX power control are randomly and uniformly distributed inside the area.

Fig. 1(a) shows the beam-domain channel following the one-ring model with an RX angular spread of 15 degrees from each UE. In addition to the beam coefficients corresponding to the angular spread taking large values, the coefficients around them are also found to have non-zero values due to sidelobe leakage. Consequently, it can be seen that the number of elements that are as negligible as zero, shown in dark blue, is not very large.

Fig. 1(b) shows the beam-domain channel following the finite path model, where the number of effective clusters is set to $L = 6$ [50], [51]. Even in the absence of angular spreading, energy is also distributed around the beam coefficients corresponding to the AoA of each arrival path, confirming the pseudo-sparsity of the beam-domain channel in finite-dimensional observations. Obviously, it is infeasible to accurately replace these beam-domain MIMO channels by a strictly sparse signal model. Motivated by these results, in this paper, we consider a stochastic inference method to capture this statistical structure.

III. SBL USING COMPLEX STUDENT'S t-PRIOR

In this section, we construct a hierarchical Bayesian model that allows the complex t-distribution with DoF as a parameter to be incorporated as a provisional prior distribution and derive a novel SBL algorithm based on the resultant model. The complex t-distribution can represent a wide range of distributions depending on the DoF parameter and is used in the context of source separation of single-channel audio signals as a tractable model for distributions with heavy tails and high kurtosis [44]. This statistical feature is quite suitable for capturing pseudo-sparsity, in which small values close to zero occur frequently and the strength and weakness between realization values are pronounced. For later convenience, our proposed CE algorithm is hereafter referred to as complex t-distribution-based SBL (Ct-SBL).

A. HIERARCHICAL BAYESIAN MODEL

In this subsection, we construct the hierarchical Bayesian model of the complex t-distribution with DoF as a parameter.

From (1)–(3), the observation likelihood can be modeled by $\mathcal{CN}(\boldsymbol{Sh}, N_0\boldsymbol{I}_{KN})$ as follows:

$$\begin{aligned} p(\boldsymbol{y}|\boldsymbol{h}) &= \frac{1}{(\pi N_0)^{KN}} e^{-\frac{\|\boldsymbol{y}-\boldsymbol{Sh}\|_2^2}{N_0}} \\ &= \left(\frac{\beta}{\pi}\right)^{KN} e^{-\beta\|\boldsymbol{y}-\boldsymbol{Sh}\|_2^2}, \end{aligned} \tag{9}$$

where the complex t-distribution [44] is not conjugate to the Gaussian distribution in (9), making algorithm design infeasible if used directly as a provisional prior distribution for $\boldsymbol{h}$. To avoid this inconvenience, consider replacing the complex t-distribution with a hierarchical model represented by a cascade connection of multiple distributions, as in [31], [32].

First, the prior distribution of $\boldsymbol{h}$ given $\boldsymbol{\lambda}$ follows $\mathcal{CN}(\boldsymbol{0}, \boldsymbol{\Lambda})$, *i.e.*,

$$\begin{aligned} p(\boldsymbol{h}|\boldsymbol{\lambda}) &= \frac{1}{\pi^{MN}|\boldsymbol{\Lambda}|} e^{-\boldsymbol{h}^{\mathsf{H}}\boldsymbol{\Lambda}^{-1}\boldsymbol{h}} \\ &= \frac{1}{\pi^{MN}} \left(\prod_{i=1}^{MN} \frac{1}{\lambda_i}\right) e^{-\sum_{i=1}^{MN} \frac{|h_i|^2}{\lambda_i}}, \end{aligned} \tag{10}$$

where $\boldsymbol{\lambda} \triangleq [\lambda_1, \lambda_2, \ldots, \lambda_{MN}]^{\mathsf{T}}$ is the variance vector and $\boldsymbol{\Lambda} \triangleq \mathrm{diag}[\boldsymbol{\lambda}]$ is the covariance matrix.

Next, using the fact that the inverse-Gamma distribution [53] is the conjugate prior to the Gaussian distribution, we assume that each element of $\boldsymbol{\lambda}$, given the DoF parameter ν, has the inverse-Gamma distribution, *i.e.*,

$$\begin{aligned} p(\boldsymbol{\lambda}|\nu) &= \prod_{i=1}^{MN} p(\lambda_i|\nu) \\ &= \prod_{i=1}^{MN} \left(\frac{\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)}\right) \lambda_i^{-\left(\frac{\nu}{2}+1\right)} e^{-\frac{\nu}{2}\sum_{i=1}^{MN}\frac{1}{\lambda_i}}, \end{aligned} \tag{11}$$

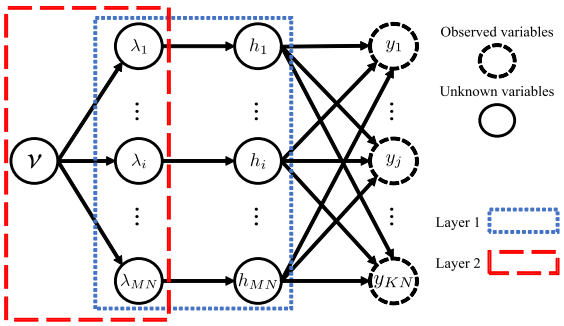


FIGURE 2. Two-layer hierarchical Bayesian model based on the complex *t*-distribution.

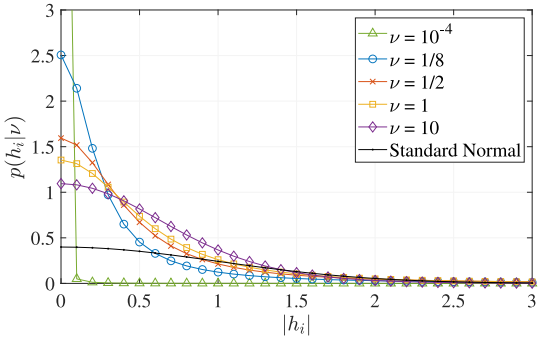


FIGURE 3. PDFs of the complex *t*-distributions for various DoF parameters ν.

where ν is a DoF parameter. This hierarchical structure differs from that in [31], which uses a Gamma distribution for the precision of the Gaussian distribution in the first layer.

Finally, from (10) and (11), the constructed hierarchical Bayesian model is expressed as

$$p(\boldsymbol{y}, \boldsymbol{h}, \boldsymbol{\lambda}|\nu) \propto p(\boldsymbol{y}|\boldsymbol{h})p(\boldsymbol{h}, \boldsymbol{\lambda}|\nu) \propto p(\boldsymbol{y}|\boldsymbol{h})p(\boldsymbol{h}|\boldsymbol{\lambda})p(\boldsymbol{\lambda}|\nu). \quad (12)$$

Fig. 2 shows the relationship among variables in (12), where each node represents a variable and the directed edges represent dependencies between variables. It is found that the probability distribution of the unknown variables $h_i, \forall i \in \mathcal{I} \triangleq \{1, 2, \ldots, MN\}$, is represented by a graphical model consisting of two layers each corresponding to the variables $\lambda_i, \forall i \in \mathcal{I}$, and ν, namely, a two-layer hierarchical Bayesian model.

To confirm whether the constructed hierarchical Bayesian model is able to represent the desired distribution, the joint probability density function (PDF) conditioned on the DoF parameter ν is marginalized by the intermediate variable $\boldsymbol{\lambda}$, yielding

$$p(\boldsymbol{h}|\nu) = \int p(\boldsymbol{h}, \boldsymbol{\lambda}|\nu)d\boldsymbol{\lambda} \quad (13a)$$

$$= \frac{1}{(\pi\nu)^{MN}} \prod_{i=1}^{MN} \left(\frac{2}{\nu}|h_i|^2 + 1\right)^{-\left(\frac{\nu}{2}+1\right)}, \quad (13b)$$

where for the sake of readability the detailed derivation of (13b) is given in Appendix-A. The PDF given by (13b) is in fact a complex *t*-distribution[44]; thus, from (10)–(12), the desired prior distribution is constructed via the intermediate variable $\boldsymbol{\lambda}$ while maintaining the relationship between the conjugate prior distributions.

Fig. 3 visualizes the PDFs of the complex *t*-distributions drawn with several different DoF parameters ν, showing that the probability mass in the tails is controlled by ν. For $\nu \to \infty$, it becomes a Gaussian distribution with *thin* tails, and for $\nu = 1$, it becomes a Cauchy distribution with *fat* tails, from which the kurtosis increases as ν approaches 0. When ν is set to extremely small values ($\nu = 10^{-4}$ as in [31]), the kurtosis of the complex *t*-distribution becomes extremely large, confirming that the distribution assumes the sparsity of the unknown variables. Because of this flexibility in the shape of the distribution, as we will show in the regularization analysis in Section IV, one can find a provisional prior distribution that captures the pseudo-sparsity by appropriately adjusting the DoF parameter ν.

B. DERIVATION OF SBL ALGORITHM

In this subsection, we derive the SBL algorithm from the hierarchical Bayesian model presented in Section III-A using the evidence procedure (type-II maximum likelihood approach) in [32]. As is widely known, Bayesian inference is performed based on the posterior distribution, *i.e.*,

$$p(\boldsymbol{h}|\boldsymbol{y}) = \int p(\boldsymbol{h}, \boldsymbol{\lambda}|\boldsymbol{y})d\boldsymbol{\lambda}$$

$$= \int p(\boldsymbol{h}|\boldsymbol{y}, \boldsymbol{\lambda})p(\boldsymbol{\lambda}|\boldsymbol{y})d\boldsymbol{\lambda}. \quad (14)$$

Since it is difficult in general to solve (14) analytically, the PDF of the intermediate variable, *i.e.*, the PDF of the variance vector $\boldsymbol{\lambda}$, is approximated by the Dirac delta function as follows [31]:

$$p(\boldsymbol{h}|\boldsymbol{y}) \approx \int p(\boldsymbol{h}|\boldsymbol{y}, \boldsymbol{\lambda})\delta\left(\boldsymbol{\lambda} - \hat{\boldsymbol{\lambda}}\right)d\boldsymbol{\lambda} = p\left(\boldsymbol{h}|\boldsymbol{y}, \hat{\boldsymbol{\lambda}}\right), \quad (15)$$

where $\hat{\boldsymbol{\lambda}}$ denotes the MAP estimate of $\boldsymbol{\lambda}$, given $\boldsymbol{y}$. This approximation is based on the assumption that the approximated PDF using $\hat{\boldsymbol{\lambda}}$ is nearly identical to that obtained by sampling from the full posterior distribution[54]. Since the posterior distribution of $\boldsymbol{\lambda}$ has only one peak and the MAP estimation is equivalent to a point estimate of that peak, the posterior distribution calculated by fixing the value of $\boldsymbol{\lambda}$ to $\hat{\boldsymbol{\lambda}}$ can be regarded as a highly accurate approximation of the posterior distribution obtained by marginalizing the joint posterior distribution by $\boldsymbol{\lambda}$.

Fig. 4 shows a block diagram of our proposed scheme, where the estimation is performed by iterative processing of posterior estimation based on (15) and variance vector update based on (12). The iterative process starts from the posterior estimation step based on the RX vector $\boldsymbol{y}$. Let $\check{\boldsymbol{\lambda}}$ be the variance vector obtained in the previous iteration, and $\check{\boldsymbol{\Lambda}} \triangleq \mathrm{diag}[\check{\boldsymbol{\lambda}}]$ be its diagonal matrix representation. At the first iteration ($t = 1$), the variance vector is appropriately

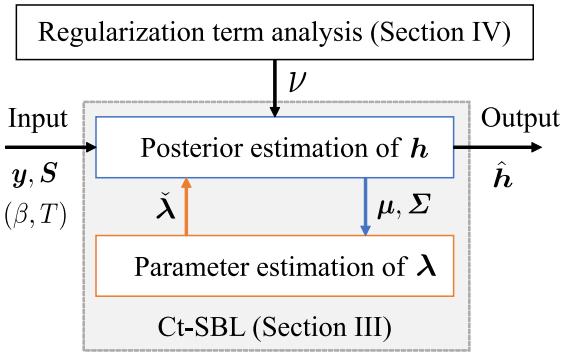


FIGURE 4. Block diagram of proposed channel estimation scheme.

initialized. From (9) and (10), the approximated posterior distribution in (15) can be expressed by a multivariate complex Gaussian distribution $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ as

$$p(\boldsymbol{h}|\boldsymbol{y}) \approx p\left(\boldsymbol{h}|\boldsymbol{y}, \check{\boldsymbol{\lambda}}\right) \propto p(\boldsymbol{y}|\boldsymbol{h}) p\left(\boldsymbol{h}|\check{\boldsymbol{\lambda}}\right) \tag{16}$$

$$= \frac{1}{\pi^{MN}|\boldsymbol{\Sigma}|} e^{-\boldsymbol{\mu}^{\mathsf{H}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}} \tag{17}$$

with

$$\boldsymbol{\mu} = \beta \boldsymbol{\Sigma} \boldsymbol{S}^{\mathsf{H}} \boldsymbol{y}, \tag{18a}$$

$$\boldsymbol{\Sigma} = \left[\beta \boldsymbol{S}^{\mathsf{H}} \boldsymbol{S} + \check{\boldsymbol{\Lambda}}^{-1}\right]^{-1}. \tag{18b}$$

Next, let us describe the estimation process for updating $\boldsymbol{\lambda}$. From (10)–(12), the joint PDF after being marginalized for $\boldsymbol{h}$, *i.e.*, *marginal distribution* [45], is expressed as

$$p(\boldsymbol{y}, \boldsymbol{\lambda}|\nu) = \int p(\boldsymbol{y}|\boldsymbol{h}) p(\boldsymbol{h}|\boldsymbol{\lambda}) p(\boldsymbol{\lambda}|\nu) d\boldsymbol{h}$$

$$= A \cdot \left(\prod_{i=1}^{MN} \lambda_i^{-\left(\frac{\nu}{2}+1\right)}\right) e^{-\frac{\nu}{2}\sum_{i=1}^{MN}\frac{1}{\lambda_i}} \cdot |\boldsymbol{\Lambda}|^{-1} |\boldsymbol{\Sigma}| \exp[-\xi], \tag{19}$$

with

$$\xi \triangleq \beta ||\boldsymbol{y} - \boldsymbol{S}\boldsymbol{\mu}||_2^2 + \boldsymbol{\mu}^{\mathsf{H}} \boldsymbol{\Lambda}^{-1} \boldsymbol{\mu}, \tag{20a}$$

$$A \triangleq \frac{\beta^{KN}}{\pi^{N^2}} \left(\frac{\frac{\nu}{2}^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)}\right)^{MN} \tag{20b}$$

where we use the complex Gaussian integral [55].

In [31], the log distribution of the parameters is incorporated into the cost function, whereas in our method, the logarithm of the marginal distribution in (19) is used as the cost function as in [45]. By taking the logarithm of (19), we define the cost function $\mathcal{L}$ for $\boldsymbol{\lambda}$ as follows:

$$\ln p(\boldsymbol{y}, \boldsymbol{\lambda}|\nu)$$

$$= \ln A - \left(\frac{\nu}{2}+1\right) \sum_{i=1}^{MN} \ln \lambda_i - \frac{\nu}{2} \sum_{i=1}^{MN} \frac{1}{\lambda_i} - \ln |\boldsymbol{\Lambda}| + \ln |\boldsymbol{\Sigma}| - \xi$$

$$= -\left(\frac{\nu}{2}+2\right) \sum_{i=1}^{MN} \ln \lambda_i - \frac{\nu}{2} \sum_{i=1}^{MN} \frac{1}{\lambda_i} + \ln |\boldsymbol{\Sigma}| + C \tag{21}$$

$$= \underbrace{-\sum_{i=1}^{MN} \left\{\left(\frac{\nu}{2}+2\right) \ln \lambda_i + \left(\frac{\nu}{2} + |\mu_i|^2\right) \frac{1}{\lambda_i}\right\} + \ln |\boldsymbol{\Sigma}|}_{\triangleq \mathcal{L}(\boldsymbol{\lambda}):\ \text{Cost function}} + C$$

$$= \mathcal{L}(\boldsymbol{\lambda}) + C \tag{22}$$

where we use in (21) the following relationship:

$$\ln |\boldsymbol{\Lambda}| = \sum_{i=1}^{MN} \ln \lambda_i, \tag{23}$$

and the terms that are not dependent on $\boldsymbol{\lambda}$ are summarized as the constant C in (22).

Assuming that the elements of $\boldsymbol{\lambda}$ are independent of each other, we can estimate λ_1 through λ_{MN} separately so that the cost function in (22) is maximized. Thus, we consider the following partial derivative w.r.t. λ_i:

$$\frac{\partial \mathcal{L}(\boldsymbol{\lambda})}{\partial \lambda_i}$$

$$= -\sum_{i=1}^{MN} \left\{\left(\frac{\nu}{2}+2\right) \frac{\partial (\ln \lambda_i)}{\partial \lambda_i} + \left(\frac{\nu}{2} + |\mu_i|^2\right) \frac{\partial}{\partial \lambda_i}\left(\frac{1}{\lambda_i}\right)\right\} + \frac{\partial}{\partial \lambda_i} \ln |\boldsymbol{\Sigma}|$$

$$= -\left(\frac{\nu}{2}+2\right) \frac{1}{\lambda_i} + \left(\frac{\nu}{2} + |\mu_i|^2\right) \frac{1}{\lambda_i^2} + \frac{\Sigma_{i,i}}{\lambda_i^2}, \tag{24}$$

where μ_i and $\Sigma_{i,i}$ denote the i-th element of $\boldsymbol{\mu}$ and the i-th diagonal element of $\boldsymbol{\Sigma}$. By solving for $\frac{\partial \mathcal{L}(\boldsymbol{\lambda})}{\partial \lambda_i} = 0$ and taking into account the fact that $\lambda_i > 0$, the update formula can be obtained by

$$\check{\lambda}_i = \frac{\frac{\nu}{2} + \Sigma_{i,i} + |\mu_i|^2}{\frac{\nu}{2}+2}, \quad \forall i \in \mathcal{I}. \tag{25}$$

For the sake of completeness, the pseudo-code of the proposed Ct-SBL algorithm described above is given in Algorithm 1[2]

IV. REGULARIZATION TERM ANALYSIS

In this section, we analyze how the designed Ct-SBL algorithm learns the distribution parameters and searches for the solution according to the provisional prior distribution. More specifically, as in [41], [42], we focus on the optimization problem that the proposed algorithm attempts to solve via the iterative process and investigate its regularization term (*i.e.*, penalty term) in order to gain insight into the statistical properties of the solution obtained.

[2]The proposed method can be extended to time-varying channel estimation by appropriately modifying the algorithm. For instance, an extension method is proposed in [37] to modify the Bayesian model using a first-order auto-regressive (AR) model to capture the dynamic characteristics of the MIMO channel. A simpler method is proposed in [38], which incorporates a Markov process describing time variation into a hierarchical Bayesian model. These methods can be directly applied to our algorithm as well.

Algorithm 1 Proposed Ct-SBL Algorithm

Input: $\boldsymbol{y}$, $\boldsymbol{S}$, β, ν, T (Num. of iterations)
Output: $\hat{\boldsymbol{h}}$
/* ———— Initialization ———— */
$\breve{\boldsymbol{\Lambda}} = \boldsymbol{I}_{MN}$
/* ———— Iterative process ———— */
for $t = 1$ to T **do**
  $\boldsymbol{\Sigma} = \left[\beta \boldsymbol{S}^{\mathsf{H}}\boldsymbol{S} + \breve{\boldsymbol{\Lambda}}^{-1}\right]^{-1}$
  $\boldsymbol{\mu} = \beta \boldsymbol{\Sigma}\boldsymbol{S}^{\mathsf{H}}\boldsymbol{y}$
  for $i = 1$ to MN **do**
    $\breve{\lambda}_i = \frac{\frac{\nu}{2}+\Sigma_{i,i}+|\mu_i|^2}{\frac{\nu}{2}+2}$
  end for
  $\breve{\boldsymbol{\lambda}} = \left[\breve{\lambda}_1, \breve{\lambda}_2, \ldots, \breve{\lambda}_{MN}\right]^{\mathsf{T}}$
  $\breve{\boldsymbol{\Lambda}} = \mathrm{diag}\left[\breve{\boldsymbol{\lambda}}\right]$
end for
$\hat{\boldsymbol{h}} = \boldsymbol{\mu}$
/* ———— Termination ———— */

First, the minimization problem solved by the approximate posterior estimation using the knowledge of estimated parameters in (25) can be expressed from (16) as

$$\boldsymbol{\mu} = \arg\min_{\boldsymbol{h}} \mathcal{F}_{\mathrm{h}}(\boldsymbol{h}), \tag{26a}$$

with

$$F_{\mathrm{h}}(\boldsymbol{h}) \triangleq \beta||\boldsymbol{y} - \boldsymbol{S}\boldsymbol{h}||_2^2 + \boldsymbol{h}^{\mathsf{H}}\breve{\boldsymbol{\Lambda}}^{-1}\boldsymbol{h}, \tag{26b}$$

where $\boldsymbol{\mu}$ is given by (18a).

Similarly, the minimization problem solved by the type-II maximum likelihood approach using the knowledge of the MAP estimates in (18) can be expressed from (21) as

$$\breve{\boldsymbol{\lambda}} = \arg\min_{\boldsymbol{\lambda}} \mathcal{F}_{\lambda}(\boldsymbol{\lambda}), \tag{27a}$$

with

$$\mathcal{F}_{\lambda}(\boldsymbol{\lambda}) \triangleq \beta||\boldsymbol{y} - \boldsymbol{S}\boldsymbol{\mu}||_2^2 + \boldsymbol{\mu}^{\mathsf{H}}\boldsymbol{\Lambda}^{-1}\boldsymbol{\mu} + \sum_{i=1}^{NM}\left\{\left(\frac{\nu}{2}+2\right)\ln\lambda_i + \frac{\nu}{2\lambda_i}\right\} - \ln|\boldsymbol{\Sigma}|. \tag{27b}$$

From the above, the proposed algorithm, in effect, performs estimation by solving the minimization problems of (26) and (27) in an iterative manner.

Next, further formulating the two minimization problems as a nested optimization problem, we obtain

$$\hat{\boldsymbol{h}} = \arg\min_{\boldsymbol{h}} \mathcal{G}_{\mathrm{h}}(\boldsymbol{h}) \tag{28a}$$

with

$$\mathcal{G}_{\mathrm{h}}(\boldsymbol{h}) \triangleq \beta||\boldsymbol{y} - \boldsymbol{S}\boldsymbol{h}||_2^2 + \Phi(\boldsymbol{h}), \tag{28b}$$

where $\Phi(\boldsymbol{h})$ is the penalty term of the optimization problem for $\boldsymbol{h}$. From (27b), this term is expressed as

$$\Phi(\boldsymbol{h}) = \min_{\boldsymbol{\lambda}}(\mathcal{G}_{\lambda}(\boldsymbol{\lambda}, \boldsymbol{h})), \tag{29a}$$

with

$$\mathcal{G}_{\lambda}(\boldsymbol{\lambda}, \boldsymbol{h}) \triangleq \left(\sum_{i=1}^{MN}\left\{\left(\frac{\nu}{2}+2\right)\ln\lambda_i + \left(\frac{\nu}{2}+|h_i|^2\right)\frac{1}{\lambda_i}\right\} - \ln|\boldsymbol{\Sigma}|\right). \tag{29b}$$

Accordingly, the statistical properties of the solution to be estimated can be evaluated by analyzing the behavior of $\Phi(\boldsymbol{h})$ w.r.t. $\boldsymbol{h}$.

It should be noted here that if one attempts to solve the minimization problem in (29) directly, the derivatives are expressed in a form that includes the diagonal elements of $\boldsymbol{\Sigma}$, as shown in (24). In order to evaluate the penalty term $\Phi(\boldsymbol{h})$ as a function of $\boldsymbol{h}$, the minimum value must also be calculated by considering $\boldsymbol{\lambda}$ contained in this $\boldsymbol{\Sigma}$, which significantly complicates the analysis.

To deal with this, we assume the column vectors of the observation matrix $\boldsymbol{S}$ to be *orthonormal* to each other, *i.e.*, [41], [42],

$$-\ln|\boldsymbol{\Sigma}| = \sum_{i=1}^{NM}\left(\beta + \frac{1}{\lambda_i}\right). \tag{30}$$

Substituting (30) into (29b) yields

$$\mathcal{G}_{\lambda}(\boldsymbol{\lambda}, \boldsymbol{h}) = \sum_{i=1}^{MN}\Bigg\{\underbrace{\left(\frac{\nu}{2}+2\right)\ln\lambda_i + \left(\frac{\nu}{2}+|h_i|^2\right)\frac{1}{\lambda_i} + \ln\left(\beta + \frac{1}{\lambda_i}\right)}_{\triangleq \mathcal{G}_{\lambda_i}(\lambda_i, h_i)}\Bigg\}, \tag{31}$$

and its partial derivative w.r.t. λ_i can be expressed as

$$\frac{\partial \mathcal{G}_{\lambda}(\boldsymbol{\lambda}, \boldsymbol{h})}{\partial \lambda_i} = \left(\frac{\nu}{2}+2\right)\frac{1}{\lambda_i} - \left(\frac{\nu}{2}+|h_i|^2\right)\frac{1}{\lambda_i^2} + \frac{\beta}{\beta\lambda_i+1}. \tag{32}$$

Solving for $\frac{\partial \mathcal{G}_{\lambda}(\boldsymbol{\lambda}, \boldsymbol{h})}{\partial \lambda_i} = 0$ and taking into account the fact that $\lambda_i > 0$, the solution is given by

$$\breve{\lambda}_i = \frac{-\alpha + \sqrt{\alpha^2 + \beta(\nu+2)\left(\nu+2|h_i|^2\right)}}{(\nu+2)\beta}, \quad \forall i \in \mathcal{I}, \tag{33}$$

with

$$\alpha \triangleq \left\{\left(\frac{\nu}{2}+1\right) - \left(\frac{\nu}{2}+|h_i|^2\right)\beta\right\}. \tag{34}$$

From (33), the penalty term of (29a) can be expressed as

$$\Phi(\boldsymbol{h}) = \mathcal{G}_{\lambda}\left(\breve{\boldsymbol{\lambda}}, \boldsymbol{h}\right), \tag{35}$$

where from (31) we may define the element-wise penalty as

$$\phi(h_i) \triangleq \mathcal{G}_{\lambda_i}\left(\breve{\lambda}_i, h_i\right). \tag{36}$$

Fig. 5 shows the behavior of $\phi(h_i)$ as a function of $|h_i|$ for different DoF parameters $\nu = 10$, 1, 1/2, 1/8, and 10^{-4}, with SNR = 36 dB. Note that in order to clarify the behavior of the penalty term, we are focusing here on the high SNR region with the aim of reducing the effect of noise. As reference properties, we also show results for the case where the ℓ_0-, ℓ_1-, and ℓ_2-norms are used as the

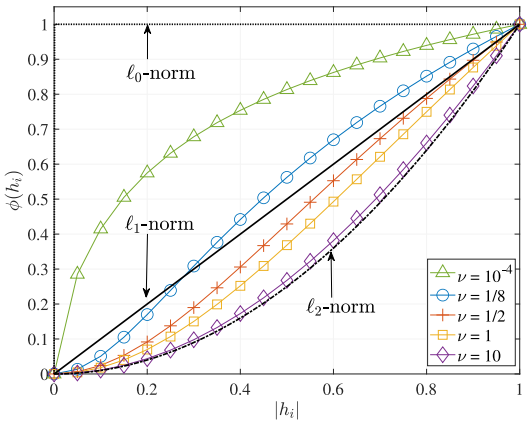


FIGURE 5. The penalty term $\phi(h_i)$ as a function of $|h_i|$, where the SNR was set to 36 dB.

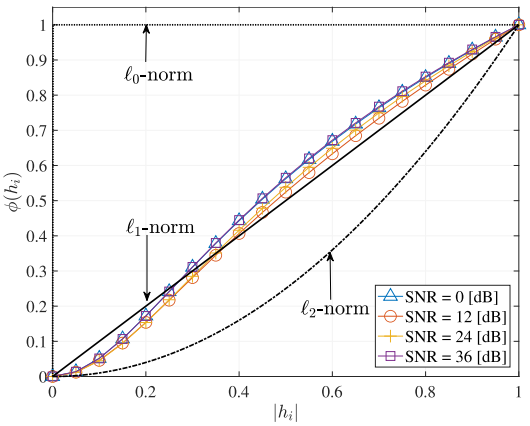


FIGURE 6. The penalty term $\phi(h_i)$ with different values of SNRs, where $\nu = 1/8$.

penalty terms. The shape of the curve changes significantly as the DoF parameter ν changes. Specifically, for large values of ν, the curve is convex like the ℓ_2-norm, and as the value of ν decreases, the curve approaches the ℓ_1-norm, indicating that the penalty term works to promote solution sparsification. Furthermore, with the conventional extremely small parameter settings (*i.e.*, $\nu = 10^{-4}$ [31]), the shape of the curve becomes concave and approaches the ℓ_0-norm, confirming that the resultant algorithm is functioning as an SSR method. It is noteworthy here that the penalty for values in the vicinity of zero is relaxed in the proposed method compared to the ℓ_1-norm even when ν is quite small (*i.e.*, $\nu = 1/8$). This property is precisely what facilitates the estimation of the pseudo-sparse signal described in Section II-D, which suggests that the intended algorithm design can be achieved by using the complex t-distribution with ν appropriately adjusted, *e.g.*, $\nu = 1/8$, as the provisional prior distribution.

Next, we focus on the effect of β. Fig. 6 shows the behavior of $\phi(h_i)$ with $\nu = 1/8$ as a function of $|h_i|$ when the SNR is set as 0, 12, 24, and 36 dB. As expected, it is found that the shape of the curve is less sensitive to SNR, which suggests that the DoF parameter ν should be adjusted according to the statistics of the unknown variables.

V. NUMERICAL RESULTS

Computer simulations were conducted to demonstrate the performance of the proposed CE algorithm via SBL using the t-prior in massive MIMO systems. Simulated results are offered in terms of the normalized mean square error (NMSE) of the estimated channel vector, defined by

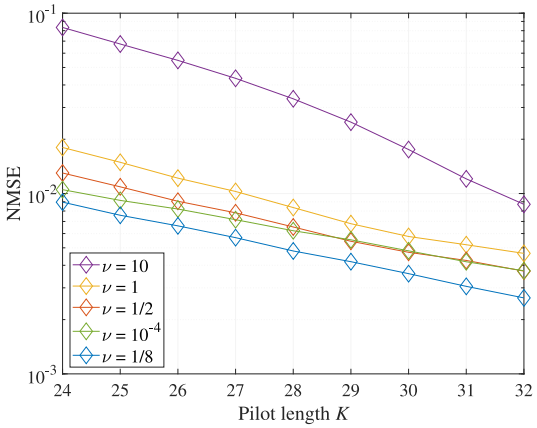
$$\mathrm{NMSE}(\boldsymbol{h}) \triangleq \mathbb{E}\left[\frac{\left\|\boldsymbol{h}-\hat{\boldsymbol{h}}\right\|_2^2}{\|\boldsymbol{h}\|_2^2}\right]. \tag{37}$$

The MIMO configuration was set to $(M, N) = (32, 64)$. Without loss of generality, we used simple Gaussian pilots, each element of which follows $\mathcal{CN}(0, 1)$ [56], [57].

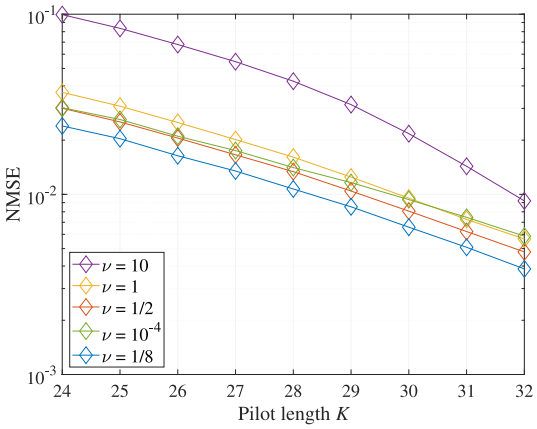
The two channel models described in Section II-D, *i.e.*, the geometrical one-ring model [8], [49], and the finite path model [50], [51], were used, and the parameters of each model are the same as those employed in Fig. 1. As described in Section II-D, the receiver is equipped with a ULA of sectors partitioned by a 120-degree opening, and the single-antenna UE devices under slow TX power control are randomly and uniformly distributed inside the area. In the geometrical one-ring model, the RX angular spread was set to 15 degrees from each UE. Let θ_m^{UE} be the azimuth position of the m-th UE, where its spatial frequency is defined as $\Omega_m = \pi \sin(\theta_m^{\mathrm{UE}})$. Assuming time-frequency resource allocation, we enforced a minimum separation in spatial frequency between any two users, *i.e.*, $|\Omega_i - \Omega_j| \leq \frac{2.783}{N}$ for $\forall i, j \in \{1, 2, \ldots, M\}$, to avoid the excessive interference among UEs [58]. In the finite path model, the number of effective clusters was set to $L = 6$, as in Section II-D.

A. ASSESSMENT OF CHANGES IN THE DOF PARAMETER ν

We first investigate how the estimation accuracy of the proposed Ct-SBL algorithm changes for different DoF parameters $\nu = 10$, 1, 1/2, 1/8, and 10^{-4} used in Fig. 5. Fig. 7 shows the NMSE performance as a function of the pilot length K for the two channel models described above. From the results, it is found that the order of NMSE performances for both channel models is very consistent with what we would expect from the behavior of the penalty term in Fig. 5. For $\nu = 10$, where the shape of the curve is similar to the ℓ_2-norm, the estimation accuracy is poor because the value of the solution cannot be suppressed to a small value. As the value of ν decreases, the estimation accuracy improves, and as expected, it can be confirmed that the best NMSE performance is achieved at $\nu = 1/8$, which most reflects the pseudo-sparsity. It is also worth noting here that the performance deteriorates when the proposed method is run as an SSR algorithm with $\nu = 10^{-4}$. This is because, as shown in Fig. 5, as the penalty term approaches the ℓ_0-norm, it becomes difficult to take into account the

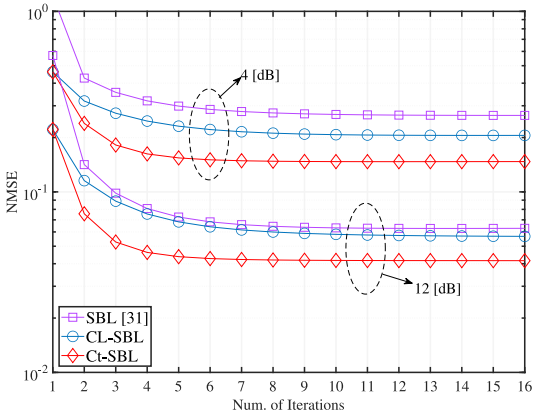


(a) Geometrical one-ring model

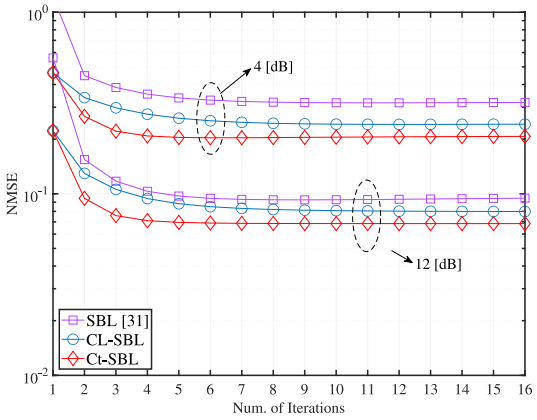


(b) Finite path model

FIGURE 7. NMSE performances w.r.t. the pilot length K for various DoF parameters ν, where $(M, N) = (32, 64)$ and SNR = 36 dB.



(a) Geometric one-ring model



(b) Finite path model

FIGURE 8. NMSE performances of the SBL-based CE algorithms w.r.t. the number of iterations, where $(M, N, K) = (32, 64, 32)$.

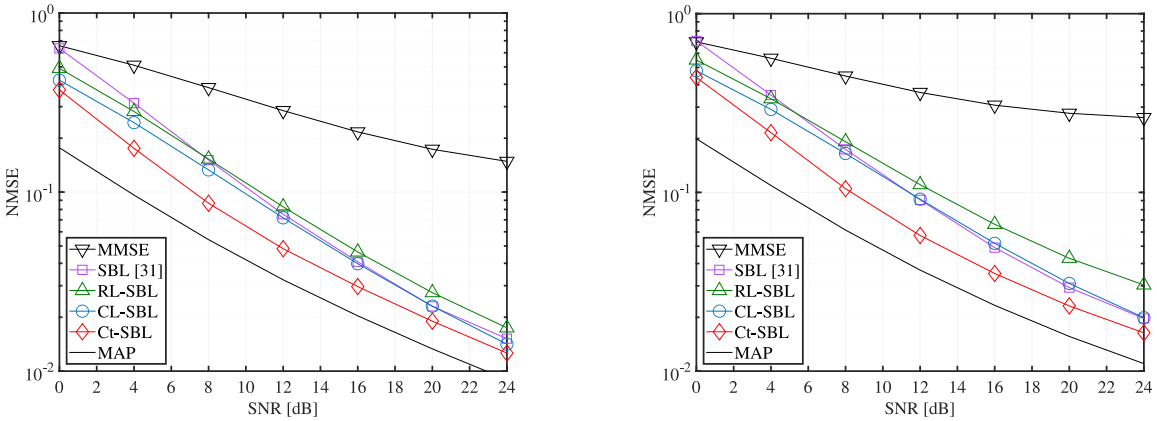
existence of many small but non-zero elements and the difference in strength among them. The results clarify the performance degradation that occurs when the beam-domain CE is approximated by the SSR problem and imply the importance of designing the provisional prior distribution appropriately according to the statistical properties of the unknown variables.

Based on the above results, in the subsequent simulation results, the NMSE performance of the proposed Ct-SBL algorithm is evaluated with the value of $\nu = 1/8$.

B. CONVERGENCE ANALYSIS

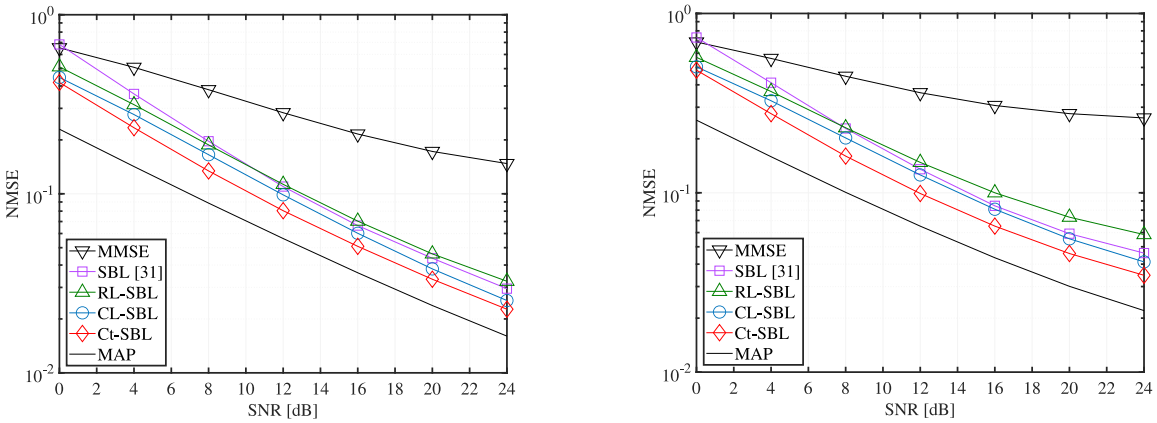
Next, we briefly analyze the convergence behavior of the SBL algorithms. In the following simulations, the NMSE performances of the following schemes are compared:

- *MMSE:* Baseline linear minimum mean square error (LMMSE) filtering, providing a reference performance to verify the gain achieved by considering the statistical structure of the beam-domain channel.
- *SBL [31]:* Baseline CE scheme based on the classical and well-known SBL using t-prior [31], providing a performance comparison with the existing SBL-based method that has been cited in numerous related studies as the SSR algorithm.
- *RL-SBL:* SotA CE scheme via SBL using real Laplace prior [32], providing a reference performance of the SSR approach based on the real-domain LASSO. After transforming the beam-domain signal model given by (3) into a real-valued double-size equivalent model [29], a hierarchical Bayesian model is constructed using the real Laplace distribution as a provisional prior distribution for both $\Re\{\boldsymbol{h}\}$ and $\Im\{\boldsymbol{h}\}$.
- *CL-SBL:* SotA CE scheme via SBL using complex Laplace prior [46], [47] derived by recasting "RL-SBL" on the complex domain, providing a reference performance of the SSR approach based on the complex-domain LASSO. The detailed derivation is given in Appendix-B. This performance allows accurate verification of the gain achieved by considering the pseudo-sparsity of the beam-domain channel.
- *Ct-SBL:* Proposed CE scheme (Algorithm 1).
- *MAP:* Idealized genie-aided scheme in which the perfect variance vector is known at the receiver, providing an absolute lower bound that can only be achieved if



(a) $K = 28$ (b) $K = 24$

FIGURE 9. NMSE performances of massive MIMO systems in correlated channels following the geometrical one-ring model, where $(M, N) = (32, 64)$.



(a) $K = 28$ (b) $K = 24$

FIGURE 10. NMSE performances of massive MIMO systems in correlated channels following the finite path model, where $(M, N) = (32, 64)$.

the model parameters (*e.g.*, variance and precision) are estimated without error using the SBL algorithm.

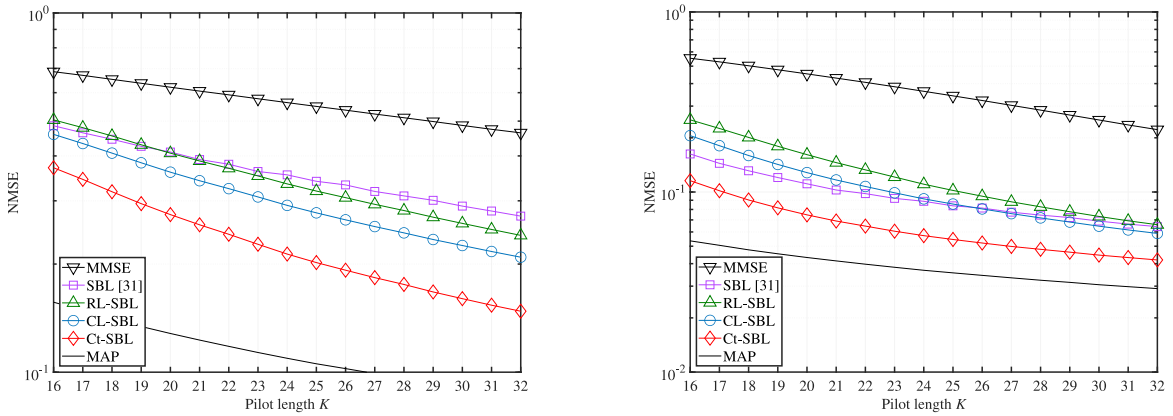
In "SBL [31]," the DoF parameter ν was set to 0 as in [31]. In both "RL-SBL" and "CL-SBL," the initial value of γ was set to 0.6 from preliminary simulations.

Since the stopping criterion for the SBL algorithm operating in the complex domain is not yet clear [47], the appropriate number of iterations T must be determined from preliminary simulations. Fig. 8 shows the NMSE performances of "SBL [31]," "CL-SBL," and "Ct-SBL" as a function of the number of iterations with SNR $= 4$ dB and 12 dB. Based on the results, the number of iterations for the three algorithms was henceforth fixed as $T = 8$ to allow for sufficient iterative convergence, even though other criteria could also be employed.

C. NMSE PERFORMANCE

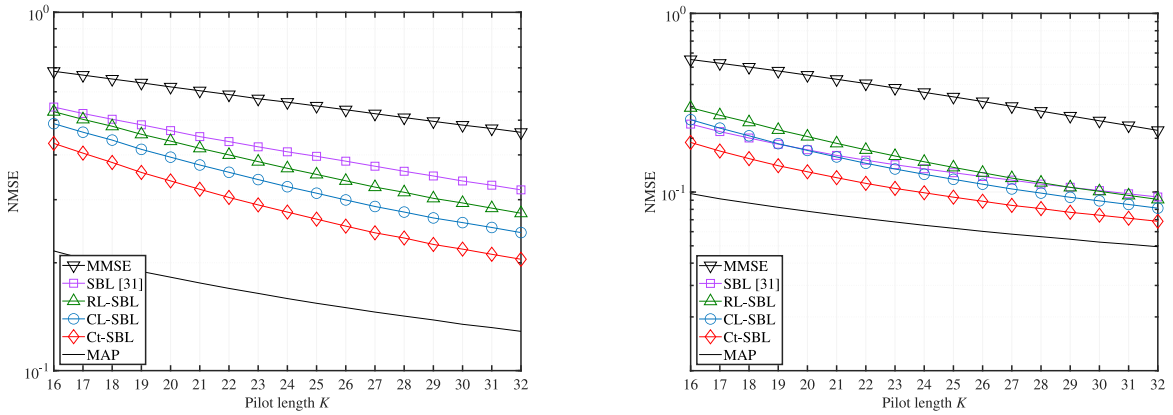
Figs. 9 and 10 show the NMSE performance of massive MIMO systems as a function of SNR, for the two channel models described above. The pilot lengths were set to $K = 28$ and 24 (corresponding to compression ratios $\rho \triangleq \frac{K}{M} = 0.875$ and 0.75, respectively).

As expected, the SBL-based CE algorithms, which can improve estimation accuracy while learning the prior distribution of the beam-domain channel coefficients, can significantly improve estimation accuracy compared to classical "MMSE," which cannot exploit the statistical structure of the beam-domain channel. It is worth noting here that the performance varies greatly among them depending on the provisional prior distribution. First, "CL-SBL," which takes into account the correlation between real and imaginary parts of $\boldsymbol{h}$, can improve performance over conventional "RL-SBL" using the conventional real Laplace prior [32]. This improvement is due in part to the fact that "RL-SBL," which requires separate estimation of the variance of the real and imaginary parts, has more variables to be estimated, making the distribution learning unstable. Indeed, the performance of "RL-SBL" deteriorates in the low SNR region, where distribution learning is more challenging. In addition, "SBL [31]," which operates as an SSR algorithm similar to these methods, achieves better performance than "RL-SBL" but is slightly inferior to "CL-SBL." However, due to errors in approximating the CE problem to SSR, the performance gap between these



(a) SNR $= 4$ dB (b) SNR $= 12$ dB

FIGURE 11. NMSE performances w.r.t. K in correlated channels following the geometrical one-ring model, where $(M, N) = (32, 64)$.



(a) SNR $= 4$ dB (b) SNR $= 12$ dB

FIGURE 12. NMSE performances w.r.t. K in correlated channels following the finite path model, where $(M, N) = (32, 64)$.

SSR methods and "MAP" remains large, leaving room for performance improvement.

In contrast, the proposed "Ct-SBL" achieves improved performance by mitigating the approximation error by using the complex t-prior that accounts for pseudo-sparsity. The improvement from "CL-SBL" is more pronounced in the geometric one-ring model of Fig. 9, where many small nonzero elements due to the angular spread occur, but an improvement of about 2 dB can also be seen in the finite path model of Fig. 10, especially in the high SNR region. The performance degradation from the genie-aided performance of "MAP" is suppressed to less than 4 dB at NMSE $= 10^{-1}$ in any simulation configuration.

D. ASSESSMENT OF CHANGES IN THE PILOT LENGTH K

Finally, we focus on the NMSE performance to changes in the number of pilot symbols K in order to demonstrate the efficacy in terms of overhead reduction. Figs. 11 and 12 show the NMSE performance as a function of K. The SNR is fixed at either 4 dB or 12 dB, and the other system parameters are the same as those in Figs. 9 and 10.

Clearly, the SBL-based CE algorithms can significantly reduce the number of pilot symbols required to achieve the same level of estimation accuracy compared to "MMSE." In addition, it is found that at SNR $= 4$ dB, the performance gap between the SBL schemes and the ideal genie-aided reference "MAP" becomes very large since the noise hinders the learning process of the prior distribution, while at SNR $= 12$ dB, the performance gap decreases significantly as K increases. The most attractive feature is that the proposed "Ct-SBL" outperforms the SotA alternatives over all K ranges, regardless of SNR and channel model, which suggests that the proposed method is quite robust against changes in system parameters. This contrasts with "SBL [31]," which cannot find a sparse solution due to increasing noise and has a significant drop in performance in the low SNR region. The results above show that the proposed method, which uses the adaptability of the complex t-distribution, is effective for inference problems that cannot be strictly reduced to the SSR problem.

VI. CONCLUSION

In this paper, we proposed a novel beam-domain CE algorithm for practical massive MIMO systems. In the proposed SBL algorithm, a complex t-distribution with an appropriately adjusted DoF parameter is pre-designed as a provisional prior distribution so that the pseudo-sparsity of the beam-domain massive MIMO channels that appears under realistic constraints can be reflected. As a framework for adjusting the DoF parameter, we presented the regularization analysis that investigates the penalty term of the equivalent optimization problem solved by the proposed SBL algorithm. A hierarchical Bayesian model is constructed to represent the designed prior distribution by cascading the Gaussian distribution and inverse-Gamma distribution, and a parameter update formula is derived based on the type-II maximum likelihood method. The efficacy of the proposed algorithm is confirmed by computer simulations for both sub-6 GHz and mmWave wireless communication scenarios, which show that the proposed method not only significantly outperforms the SotA alternatives using Laplace prior but also outperforms the existing SSR algorithm based on SBL using t-prior. These results clarify the performance degradation that occurs when the beam-domain CE is unreasonably approximated as an SSR problem and show the importance of designing provisional prior distributions that take into account the statistical properties of unknown variables in SBL design.

APPENDIX

A. DERIVATION OF EQUATION (13b)

From (12), (13a) can be rewritten as

$$p(\boldsymbol{h}|\nu) = \int p(\boldsymbol{h}|\boldsymbol{\lambda})p(\boldsymbol{\lambda}|\nu)d\boldsymbol{\lambda}. \tag{38}$$

Substituting (10) and (11) into (38) yields

$$\begin{aligned} p(\boldsymbol{h}|\nu) &= \int p(\boldsymbol{h}|\boldsymbol{\lambda})p(\boldsymbol{\lambda}|\nu)d\boldsymbol{\lambda} \\ &= \left(\frac{\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}}}{\pi\Gamma\left(\frac{\nu}{2}\right)}\right)^{MN} \int \left(\prod_{i=1}^{MN}\frac{1}{\lambda_i}\right)\left(\prod_{i=1}^{MN}\lambda_i^{-\left(\frac{\nu}{2}+1\right)}\right) \\ &\quad \cdot e^{-\sum_{i=1}^{MN}\frac{|h_i|^2}{\lambda_i}} e^{-\frac{\nu}{2}\sum_{i=1}^{MN}\frac{1}{\lambda_i}} d\boldsymbol{\lambda} \\ &= \left(\frac{\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}}}{\pi\Gamma\left(\frac{\nu}{2}\right)}\right)^{MN} \prod_{i=1}^{MN}\int \lambda_i^{-1-\left(\frac{\nu}{2}+1\right)} e^{-\frac{|h_i|^2+\frac{\nu}{2}}{\lambda_i}} d\lambda_i \\ &= \left(\frac{\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}}}{\pi\Gamma\left(\frac{\nu}{2}\right)}\right)^{MN} \prod_{i=1}^{MN}\frac{\Gamma\left(\frac{\nu}{2}+1\right)}{\left(|h_i|^2+\frac{\nu}{2}\right)^{\frac{\nu}{2}+1}} \\ &\quad \int \lambda_i^{-1-\left(\frac{\nu}{2}+1\right)} \frac{\left(|h_i|^2+\frac{\nu}{2}\right)^{\frac{\nu}{2}+1}}{\Gamma\left(\frac{\nu}{2}+1\right)} e^{-\frac{|h_i|^2+\frac{\nu}{2}}{\lambda_i}} d\lambda_i, \end{aligned} \tag{39a}$$

where we note that the integrand in (39a) is the PDF of the inverse-Gamma distribution with shape parameter $\frac{\nu}{2}+1$ and scale parameter $|h_i|^2+\frac{\nu}{2}$. Using the fact that the integral of a PDF equals 1, we obtain the following result:

$$p(\boldsymbol{h}|\nu) = \left(\frac{\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}+1\right)}{\pi\Gamma\left(\frac{\nu}{2}\right)}\right)^{MN} \cdot \prod_{i=1}^{MN}\left(|h_i|^2+\frac{\nu}{2}\right)^{-\left(\frac{\nu}{2}+1\right)}. \tag{39b}$$

Using the property of the Gamma function:

$$\Gamma\left(\frac{\nu}{2}+1\right) = \frac{\nu}{2}\Gamma\left(\frac{\nu}{2}\right), \tag{40}$$

Eq. (39) can be rewritten as

$$\begin{aligned} p(\boldsymbol{h}|\nu) &= \left(\frac{\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}+1}\Gamma\left(\frac{\nu}{2}\right)}{\pi\Gamma\left(\frac{\nu}{2}\right)}\right)^{MN} \prod_{i=1}^{MN}\left(|h_i|^2+\frac{\nu}{2}\right)^{-\left(\frac{\nu}{2}+1\right)} \\ &= \frac{1}{(\pi\nu)^{MN}} \prod_{i=1}^{MN}\left(\frac{2}{\nu}|h_i|^2+1\right)^{-\left(\frac{\nu}{2}+1\right)}. \end{aligned} \tag{41}$$

B. DERIVATION OF CL-SBL ALGORITHM

In order to accurately validate the efficacy of the proposed method considering pseudo-sparsity against the SotA SBL algorithm based on SSR, we design a novel CE algorithm via SBL using complex Laplace prior [46]. The method is designed by recasting the RL-SBL algorithm [32] on the complex domain with the use of a hierarchical Bayesian model for complex-valued SSR proposed in [47].

1) HIERARCHICAL BAYESIAN MODEL

First, the conditional distribution of $\boldsymbol{h}$, given the variance vector $\boldsymbol{\lambda}$, is set as in (10), which is the first layer of a hierarchical Bayesian model.

Next, as the second layer of the hierarchical Bayesian model, we assign the Gamma distribution to each variance, *i.e.*, $\lambda_i, \forall i \in \mathcal{I}$, separately as follows: [47]

$$\begin{aligned} p(\boldsymbol{\lambda}|\gamma) &= \prod_{i=1}^{MN}\frac{\lambda_i^{\frac{1}{2}}\left(\frac{\gamma}{4}\right)^{\frac{3}{2}}}{\Gamma\left(\frac{3}{2}\right)} e^{-\frac{\gamma}{4}\lambda_i} \\ &= \left(\frac{\left(\frac{\gamma}{4}\right)^{\frac{3}{2}}}{\frac{1}{2}\Gamma\left(\frac{1}{2}\right)}\right)^{MN}\left(\prod_{i=1}^{MN}\lambda_i^{\frac{1}{2}}\right) e^{-\frac{\gamma}{4}\sum_{i=1}^{MN}\lambda_i} \\ &= \left(\frac{\gamma^{\frac{3}{2}}}{4\sqrt{\pi}}\right)^{MN}\left(\prod_{i=1}^{MN}\lambda_i^{\frac{1}{2}}\right) e^{-\frac{\gamma}{4}\sum_{i=1}^{MN}\lambda_i}, \end{aligned} \tag{42}$$

where $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ and γ is a scaled rate parameter.

Finally, using the fact that the Gamma distribution is conjugate prior to the Gamma distribution w.r.t. the rate parameter, the third layer of the hierarchical Bayesian model is constructed as

$$p(\gamma|\boldsymbol{c}) = \frac{\gamma^{c_1-1}c_2^{c_1}}{\Gamma(c_1)} e^{-c_2\gamma}, \tag{43}$$

where $\boldsymbol{c} \triangleq [c_1, c_2]^{\mathsf{T}}$ is a predetermined constant parameter.

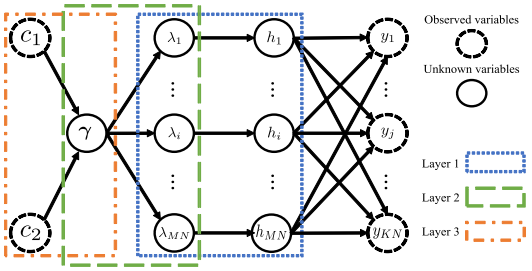


FIGURE 13. Three-layer hierarchical Bayesian model based on the complex Laplace distribution.

From (10), (42), and (43), the constructed hierarchical Bayesian model is expressed as

$$p(\boldsymbol{y}, \boldsymbol{h}, \boldsymbol{\lambda}, \gamma|\boldsymbol{c}) \propto p(\boldsymbol{y}|\boldsymbol{h})p(\boldsymbol{h}|\boldsymbol{\lambda})p(\boldsymbol{\lambda}|\gamma)p(\gamma|\boldsymbol{c}). \tag{44}$$

Fig. 13 shows the relationship among variables in (44). It is found that the probability distribution of $\boldsymbol{h} = [h_1, \ldots, h_{MN}]^{\mathsf{T}}$ is represented by a graphical model consisting of three layers corresponding to the three variables $\boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_{MN}]^{\mathsf{T}}$, γ, and $\boldsymbol{c}$, namely, a three-layer hierarchical Bayesian model.

To confirm the provisional prior distribution, the joint PDF conditioned on γ is marginalized by $\boldsymbol{\lambda}$, yielding

$$p(\boldsymbol{h}|\gamma) = \int p(\boldsymbol{h}, \boldsymbol{\lambda}|\gamma)d\boldsymbol{\lambda} = \int p(\boldsymbol{h}|\boldsymbol{\lambda})p(\boldsymbol{\lambda}|\gamma)d\boldsymbol{\lambda}. \tag{45}$$

Substituting (10) and (42) into (45) yields

$$\begin{aligned} p(\boldsymbol{h}|\gamma) &= \acute{A} \int \left(\prod_{i=1}^{MN} \frac{1}{\lambda_i}\right)\left(\prod_{i=1}^{MN} \lambda_i^{\frac{1}{2}}\right) \\ &\quad \cdot \left(e^{-\sum_{i=1}^{MN} \frac{|h_i|^2}{\lambda_i}}\right)\left(e^{-\frac{\gamma}{4}\sum_{i=1}^{MN}\lambda_i}\right)d\boldsymbol{\lambda} \\ &= \acute{A} \prod_{i=1}^{MN} \underbrace{\int \frac{1}{\sqrt{\lambda_i}} e^{-\left(\frac{|h_i|^2}{\lambda_i} + \frac{\gamma}{4}\lambda_i\right)} d\lambda_i}_{\triangleq D(h_i|\gamma)}, \end{aligned} \tag{46}$$

with $\acute{A} \triangleq (\frac{1}{4}(\frac{\gamma}{\pi})^{\frac{3}{2}})^{MN}$. Noticing that $0 < \lambda_i < \infty, \forall i \in \mathcal{I}$, we have

$$\begin{aligned} D(h_i|\gamma) &= \int_0^\infty \frac{1}{\sqrt{\lambda_i}} e^{-\left(\frac{|h_i|^2}{\lambda_i} + \frac{\gamma}{4}\lambda_i\right)} d\lambda_i \\ &= e^{-\sqrt{\gamma}|h_i|} \\ &\quad \cdot \int_0^\infty \frac{1}{\sqrt{\lambda_i}} e^{-\left(\frac{\sqrt{\gamma}}{2}\sqrt{\lambda_i} - \frac{|h_i|}{\sqrt{\lambda_i}}\right)^2} d\lambda_i \qquad (47a) \\ &= 2e^{-\sqrt{\gamma}|h_i|} \\ &\quad \cdot \int_0^\infty e^{-\left(\frac{\sqrt{\gamma}}{2}t - \frac{|h_i|}{t}\right)^2} dt \qquad (47b) \\ &= B \int_{-\infty}^{\infty} \left(e^{-u^2} + \frac{u}{\sqrt{u^2 + 2\sqrt{\gamma}|h_i|}}\right) du, \qquad (47c) \end{aligned}$$

with

$$B \triangleq \frac{2e^{-\sqrt{\gamma}|h_i|}}{\sqrt{\gamma}}, \tag{48}$$

where the following variable transformations were performed: $\sqrt{\lambda_i} = t$ in (47b) and $\frac{\sqrt{\gamma}}{2}t - \frac{|h_i|}{t} = u$ in (47c). Noticing that the second term in (47c) is an odd function, we obtain

$$D(h_i|\gamma) = B \underbrace{\int_{-\infty}^{\infty} e^{-u^2} du}_{=\sqrt{\pi}} = 2\left(\frac{\pi}{\gamma}\right)^{\frac{1}{2}} e^{-\sqrt{\gamma}|h_i|}. \tag{49}$$

Substituting (49) into (46) yields

$$\begin{aligned} p(\boldsymbol{h}|\gamma) &= \left\{\frac{1}{4}\left(\frac{\gamma}{\pi}\right)^{\frac{3}{2}}\right\}^{MN} \prod_{i=1}^{MN} \left\{2\left(\frac{\pi}{\gamma}\right)^{\frac{1}{2}} e^{-\sqrt{\gamma}|h_i|}\right\} \\ &= \left(\frac{\gamma}{2\pi}\right)^{MN} e^{-\sqrt{\gamma}\|\boldsymbol{h}\|_1}. \end{aligned} \tag{50}$$

The PDF given by (50) is a complex generalized Gaussian distribution presented in [46], and we employ this as a complex Laplace distribution as in [47].

2) DERIVATION OF SBL ALGORITHM

From the hierarchical Bayesian model presented in the previous subsection, the CL-SBL algorithm is derived using the evidence procedure as in Section III-B. The posterior distribution of $\boldsymbol{h}$ can be expressed as

$$\begin{aligned} p(\boldsymbol{h}|\boldsymbol{y}) &= \int\int p(\boldsymbol{h}, \boldsymbol{\lambda}, \gamma|\boldsymbol{y})d\boldsymbol{\lambda}d\gamma \\ &= \int\int p(\boldsymbol{h}|\boldsymbol{y}, \boldsymbol{\lambda}, \gamma)p(\boldsymbol{\lambda}|\gamma, \boldsymbol{y})p(\gamma|\boldsymbol{y})d\boldsymbol{\lambda}d\gamma. \end{aligned} \tag{51}$$

Similar to the derivation of (15), (51) is approximated by [31]

$$\begin{aligned} p(\boldsymbol{h}|\boldsymbol{y}) &\approx \int\int p(\boldsymbol{h}|\boldsymbol{y}, \boldsymbol{\lambda}, \gamma)\delta\left(\boldsymbol{\lambda} - \hat{\boldsymbol{\lambda}}\right)\delta\left(\gamma - \hat{\gamma}\right)d\boldsymbol{\lambda}d\gamma \\ &= p\left(\boldsymbol{h}|\boldsymbol{y}, \hat{\boldsymbol{\lambda}}, \hat{\gamma}\right), \end{aligned} \tag{52}$$

where $\hat{\boldsymbol{\lambda}}$ and $\hat{\gamma}$ denote the MAP estimates of $\boldsymbol{\lambda}$ and γ, respectively, given $\boldsymbol{y}$. Similar to (16), from (9), (10), and (52), the approximated posterior estimate and its covariance matrix are given by (18).

Next, from (10), (42), and (43), the joint PDF after marginalization of $\boldsymbol{h}$ is expressed as

$$\begin{aligned} p(\boldsymbol{y}, \boldsymbol{\lambda}, \gamma|\boldsymbol{c}) &= \int p(\boldsymbol{y}|\boldsymbol{h})p(\boldsymbol{h}|\boldsymbol{\lambda})p(\boldsymbol{\lambda}|\gamma)p(\gamma|\boldsymbol{c})d\boldsymbol{h} \\ &= C' \cdot \frac{\gamma^{c_1 - 1}}{\Gamma(c_1)^{-1}} \left(\prod_{i=1}^{MN} \lambda_i^{\frac{1}{2}}\right) e^{-\left(\frac{\gamma}{4}\sum_{i=1}^{MN}\lambda_i + c_2\gamma\right)} \\ &\quad \cdot |\boldsymbol{\Lambda}|^{-1}|\boldsymbol{\Sigma}|e^{-\xi}, \end{aligned} \tag{53}$$

with

$$C' \triangleq \frac{c_2^{c_1}\beta^{KN}}{\pi^{N^2}} \left(\frac{\gamma^{\frac{3}{2}}}{4\sqrt{\pi}}\right)^{MN}. \tag{54}$$

Following [32], the update formula for $\boldsymbol{\lambda}$ and γ can be obtained by

$$\check{\lambda}_i = -\frac{1}{\gamma} + \frac{1}{\gamma}\sqrt{1 + 4\gamma\left(\Sigma_{i,i} + |\mu_i|^2\right)}, \quad \forall i \in \mathcal{I}, \tag{55}$$

Algorithm 2 CE Algorithm via CL-SBL

Input: $\boldsymbol{y}$, $\boldsymbol{S}$, β, γ_{ini} (Initial value of γ), T (Num. of Iterations)

Output: $\hat{\boldsymbol{h}}$

/* ———— Initialization ———— */

$\check{\boldsymbol{\Lambda}} = \boldsymbol{I}_{MN}$, $\check{\gamma} = \gamma_{ini}$

/* ———— Iterative process ———— */

for $t = 1$ to T **do**

  $\boldsymbol{\Sigma} = \left[\beta \boldsymbol{S}^{\mathsf{H}}\boldsymbol{S} + \check{\boldsymbol{\Lambda}}^{-1}\right]^{-1}$

  $\boldsymbol{\mu} = \beta \boldsymbol{\Sigma}\boldsymbol{S}^{\mathsf{H}}\boldsymbol{y}$

  for $i = 1$ to MN **do**

    $\check{\lambda}_i = -\frac{1}{\gamma} + \frac{1}{\gamma}\sqrt{1 + 4\gamma\left(\Sigma_{i,i} + |\mu_i|^2\right)}$

  end for

  $\check{\gamma} = \frac{\frac{3}{2}MN + c_1 - 1}{\frac{1}{4}\sum_{i=1}^{MN}\lambda_i + c_2}$

  $\check{\boldsymbol{\Lambda}} = \text{diag}\left[\check{\boldsymbol{\lambda}}\right]$

end for

$\hat{\boldsymbol{h}} = \boldsymbol{\mu}$

/* ———— Termination ———— */

and

$$\check{\gamma} = \frac{\frac{3}{2}MN + c_1 - 1}{\frac{1}{4}\sum_{i=1}^{MN}\lambda_i + c_2}, \tag{56}$$

respectively. The parameter $\boldsymbol{c}$ was set to $\boldsymbol{0}$ in our simulations since it is empirically known that a small value of $\boldsymbol{c}$ is sufficient and that its variation will hardly affect the performance [31], [32].

For the sake of completeness, the pseudo-code of the CE algorithm described above, named complex Laplace distribution-based SBL (CL-SBL), is given in Algorithm 2.

REFERENCES

[1] P. Yang, Y. Xiao, M. Xiao, and S. Li, "6G wireless communications: Vision and potential techniques," *IEEE Netw.*, vol. 33, no. 4, pp. 70–75, Jul./Aug. 2019.

[2] I. F. Akyildiz, A. Kak, and S. Nie, "6G and beyond: The future of wireless communications systems," *IEEE Access*, vol. 8, pp. 133995–134030, 2020.

[3] W. Jiang, B. Han, M. A. Habibi, and H. D. Schotten, "The road towards 6G: A comprehensive survey," *IEEE Open J. Commun. Soc.*, vol. 2, pp. 334–366, 2021.

[4] B. Hassibi and B. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.

[5] T. Takahashi, H. Iimori, K. Ando, K. Ishibashi, S. Ibi, and G. T. F. de Abreu, "Bayesian receiver design via bilinear inference for cell-free massive MIMO with low-resolution ADCs," *IEEE Trans. Wireless Commun.*, vol. 22, no. 7, pp. 4756–4772, Jul. 2023.

[6] K. Ito, T. Takahashi, S. Ibi, and S. Sampei, "Bilinear gaussian belief propagation for massive MIMO detection with non-orthogonal pilots," *IEEE Trans. Commun.*, vol. 72, no. 2, pp. 1045–1061, Feb. 2024.

[7] Y. Zhu, H. Guo, and V. K. N. Lau, "Bayesian channel estimation in multi-user massive MIMO with extremely large antenna array," *IEEE Trans. Signal Process.*, vol. 69, pp. 5463–5478, Sep. 2021.

[8] A. Adhikary, J. Nam, J. Ahn, and G. Caire, "Joint spatial division and multiplexing—The large-scale array regime," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6441–6463, Oct. 2013.

[9] X. Rao and V. K. N. Lau, "Distributed compressive CSIT estimation and feedback for FDD multi-user massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3261–3271, Jun. 2014.

[10] X. Lin et al., "5G new radio: Unveiling the essentials of the next generation wireless access technology," *IEEE Commun. Stand. Mag.*, vol. 3, no. 3, pp. 30–37, Sep. 2019.

[11] X. Chen, D. W. K. Ng, W. Yu, E. G. Larsson, N. Al-Dhahir, and R. Schober, "Massive access for 5G and beyond," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 3, pp. 615–637, Mar. 2021.

[12] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 264–273, Feb. 2013.

[13] T. Takahashi, A. Tölli, S. Ibi, and S. Sampei, "Low-complexity large MIMO detection via layered belief propagation in beam domain," *IEEE Trans. Wireless Commun.*, vol. 21, no. 1, pp. 234–249, Jan. 2022.

[14] A. Sayeed, "Deconstructing multiantenna fading channels," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2563–2579, Oct. 2002.

[15] T. Takahashi, A. Tolli, S. Ibi, and S. Sampei, "Low-complexity beam-domain channel estimation with long-term statistics for large MIMO detection," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, 2019, pp. 1–6.

[16] M. R. Akdeniz et al., "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1164–1179, Jun. 2014.

[17] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter-wave cellular wireless networks: Potentials and challenges," *Proc. IEEE*, vol. 102, no. 3, pp. 366–385, Mar. 2014.

[18] T. S. Rappaport, G. R. MacCartney, M. K. Samimi, and S. Sun, "Wideband millimeter-wave propagation measurements and channel models for future wireless communication system design," *IEEE Trans. Commun.*, vol. 63, no. 9, pp. 3029–3056, Sep. 2015.

[19] Z. Zhou, J. Fang, L. Yang, H. Li, Z. Chen, and R. S. Blum, "Low-rank tensor decomposition-aided channel estimation for Millimeter wave MIMO-OFDM systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 7, pp. 1524–1538, Jul. 2017.

[20] R. Zhang et al., "Integrated sensing and communication with massive MIMO: A unified tensor approach for channel and target parameter estimation," *IEEE Trans. Wireless Commun.*, vol. 23, no. 8, pp. 8571–8587, Aug. 2024.

[21] P. Zhao, K. Ma, Z. Wang, and S. Chen, "Virtual angular-domain channel estimation for FDD based massive MIMO systems with partial orthogonal pilot design," *IEEE Trans. Veh. Technol.*, vol. 69, no. 5, pp. 5164–5178, May 2020.

[22] X. Li, J. Fang, H. Li, and P. Wang, "Millimeter wave channel estimation via exploiting joint sparse and low-rank structures," *IEEE Trans. Wireless Commun.*, vol. 17, no. 2, pp. 1123–1133, Feb. 2018.

[23] A.-A. Lu, Y. Chen, and X. Gao, "2D beam domain statistical CSI estimation for massive MIMO uplink," *IEEE Trans. Wireless Commun.*, vol. 23, no. 1, pp. 749–761, Jan. 2024.

[24] F. Sohrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale antenna arrays," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 501–513, Apr. 2016.

[25] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, 2009. [Online]. Available: https://doi.org/10.1137/080716542

[26] D. P. Bertsekas, "Incremental gradient, subgradient, and proximal methods for convex optimization: A survey," in *Optimization for Machine Learning*. Cambridge, MA, USA: MIT Press, Sep. 2011. [Online]. Available: https://doi.org/10.7551/mitpress/8996.003.0006

[27] L. Condat, "A primal—Dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms," *J. Optim. Theory Appl.*, vol. 158, no. 2, pp. 460–479, Aug. 2013. [Online]. Available: https://doi.org/10.1007/s10957-012-0245-9

[28] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proc. Nat. Acad. Sci.*, vol. 106, no. 45, pp. 18914–18919, Nov. 2009.

[29] T. Takahashi, S. Ibi, and S. Sampei, "Design of adaptively scaled belief in multi-dimensional signal detection for higher-order modulation," *IEEE Trans. Commun.*, vol. 67, no. 3, pp. 1986–2001, Mar. 2019.

[30] K. Takeuchi, "Rigorous dynamics of expectation-propagation-based signal recovery from unitarily invariant measurements," *IEEE Trans. Inf. Theory*, vol. 66, no. 1, pp. 368–386, Jan. 2020.

[31] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *J. Mach. Learn. Res.*, vol. 1, pp. 211–244, Sep. 2001.

[32] S. D. Babacan, R. Molina, and A. K. Katsaggelos, "Bayesian compressive sensing using laplace priors," *IEEE Trans. Image Process.*, vol. 19, pp. 53–63, 2010.

[33] X. Cheng, Y. Yang, B. Xia, N. Wei, and S. Li, "Sparse channel estimation for millimeter wave massive MIMO systems with lens antenna array," *IEEE Trans. Veh. Technol.*, vol. 68, no. 11, pp. 11348–11352, Nov. 2019.

[34] S. Srivastava, A. Mishra, A. Rajoriya, A. K. Jagannatham, and G. Ascheid, "Quasi-static and time-selective channel estimation for block-sparse Millimeter wave hybrid MIMO systems: Sparse Bayesian learning (SBL) based approaches," *IEEE Trans. Signal Process.*, vol. 67, no. 5, pp. 1251–1266, Mar. 2019.

[35] R. Zhang, B. Shim, and H. Zhao, "Downlink compressive channel estimation with phase noise in massive MIMO systems," *IEEE Trans. Commun.*, vol. 68, no. 9, pp. 5534–5548, Sep. 2020.

[36] L. Xu, L. Cheng, N. Wong, Y.-C. Wu, and H. V. Poor, "Overcoming beam squint in mmWave MIMO channel estimation: A Bayesian multi-band sparsity approach," *IEEE Trans. Signal Process.*, vol. 72, pp. 1219–1234, Feb. 2024.

[37] J. Ma, S. Zhang, H. Li, F. Gao, and S. Jin, "Sparse Bayesian learning for the time-varying massive MIMO channels: Acquisition and tracking," *IEEE Trans. Commun.*, vol. 67, no. 3, pp. 1925–1938, Mar. 2019.

[38] G. Liu, A. Liu, R. Zhang, and M. Zhao, "Angular-domain selective channel tracking and doppler compensation for high-mobility mmWave massive MIMO," *IEEE Trans. Wireless Commun.*, vol. 20, no. 5, pp. 2902–2916, May 2021.

[39] Y. Jin, D. He, S. Wei, and W. Yu, "Off-grid DOA estimation method based on sparse Bayesian learning with clustered structural-aware prior information," *IEEE Trans. Veh. Technol.*, vol. 73, no. 4, pp. 5469–5483, Apr. 2024.

[40] N. Li, X.-K. Zhang, B. Zong, F. Lv, J. Xu, and Z. Wang, "An off-grid direction-of-arrival estimator based on sparse Bayesian learning with three-stage hierarchical laplace priors," *Signal Process.*, vol. 218, May 2024, Art. no. 109371. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0165168423004450

[41] D. Wipf and S. Nagarajan, "A new view of automatic relevance determination," in *Proc. 20th Int. Conf. Neural Inf. Process. Syst.*, 2007, pp. 1625–1632.

[42] D. P. Wipf, B. D. Rao, and S. Nagarajan, "Latent variable Bayesian models for promoting sparsity," *IEEE Trans. Inf. Theory*, vol. 57, no. 9, pp. 6236–6255, Sep. 2011.

[43] J. Ji, H. Chang, C.-X. Wang, J. Huang, and C. Huang, "A beam domain channel estimation algorithm based on sparse Bayesian learning," in *Proc. IEEE 23rd Int. Conf. Commun. Technol. (ICCT)*, 2023, pp. 384–389.

[44] K. Yoshii, K. Itoyama, and M. Goto, "Student's T nonnegative matrix factorization and positive semidefinite tensor factorization for single-channel audio source separation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, 2016, pp. 51–55.

[45] J. O. Berger, *Statistical Decision Theory and Bayesian Analysis* (Springer Series in Statistics), 2nd ed. New York, NY, USA: Springer, 1985.

[46] M. Novey, T. Adali, and A. Roy, "A complex generalized Gaussian distribution—Characterization, generation, and estimation," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1427–1433, Mar. 2010.

[47] S. Afrifa, T. Zhang, Z. Xin, P. Appiahene, and M. S. Yaw, "Space alternating variational estimation based sparse Bayesian learning for complex-value sparse signal recovery using adaptive laplace priors," *IET Signal Process.*, vol. 17, no. 3, pp. 1–17, 2023.

[48] T. Yoshida, D. Shirase, T. Takahashi, S. Ibi, and S. Sampei, "Low-complexity large MIMO detection based on beam-domain local LMMSE filters," in *Proc. IEEE Int. Conf. Commun.*, 2022, pp. 1312–1317.

[49] W. C. Jakes and D. C. Cox, *Microwave Mobile Communications*. Hoboken, NJ, USA: Wiley, 1994.

[50] L. Xu, C. Qian, F. Gao, W. Zhang, and S. Ma, "Angular domain channel estimation for mmWave massive MIMO with one-bit ADCs/DACs," *IEEE Trans. Wireless Commun.*, vol. 20, no. 2, pp. 969–982, Feb. 2021.

[51] K. Ito, T. Takahashi, K. Igarashi, S. Ibi, and S. Sampei, "AoA estimation-aided Bayesian receiver design via bilinear inference for mmWave massive MIMO," in *Proc. IEEE Int. Conf. Commun.*, 2023, pp. 6474–6479.

[52] D. Fan et al., "Angle domain channel estimation in hybrid Millimeter wave massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 12, pp. 8165–8179, Dec. 2018.

[53] S. H. Abid and S. Al-Hassany, "On the inverted gamma distribution," *Int. J. Syst. Sci. Appl. Math.*, vol. 1, no. 3, pp. 16–22, 2016.

[54] J. Shen, F. Gini, M. Greco, and T. Zhou, "Off-grid DOA estimation using improved root sparse Bayesian learning for non-uniform linear arrays," *EURASIP J. Adv. Signal Process.*, vol. 34, pp. 1–22, Mar. 2023.

[55] J. K. Jain, *Gaussian Integral*. Cambridge, U.K.: Cambridge Univ., 2007.

[56] Y. Gu and Y. D. Zhang, "Pilot design for gaussian mixture channel estimation in massive MIMO," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, 2018, pp. 3266–3270.

[57] D. Dogan and G. Leus, "Correlated sparse Bayesian learning for recovery of block sparse signals with unknown borders," *IEEE Open J. Signal Process.*, vol. 5, pp. 421–435, Jan. 2024.

[58] C. A. Balanis, *Antenna Theory: Analysis and Design*. Hoboken, NJ, USA: Wiley, 2005.

KENGO FURUTA (Graduate Student Member, IEEE) received the B.E. and M.E. degrees in communication engineering from Osaka University, Osaka, Japan, in 2022 and 2024, respectively, where he is currently pursuing the Ph.D. degree with the Graduate School of Engineering. His research interests include sparse Bayesian learning, compressed sensing, sparse signal reconstruction, and wireless communication.

TAKUMI TAKAHASHI (Member, IEEE) received the B.E., M.E., and Ph.D. degrees in communication engineering from Osaka University, Osaka, Japan, in 2016, 2017, and 2019, respectively, where he joined as an Assistant Professor with the Graduate School of Engineering in 2019. From 2018 to 2019, he was a Visiting Researcher with the Center for Wireless Communications, University of Oulu, Finland. His current research interests include Bayesian inference, belief propagation, signal processing, and wireless communications.

HIDEKI OCHIAI (Fellow, IEEE) received the B.E. degree in communication engineering from Osaka University, Osaka, Japan, in 1996 and the M.E. and Ph.D. degrees in information and communication engineering from The University of Tokyo, Tokyo, Japan, in 1998 and 2001, respectively. From 1994 to 1995, he was with the Department of Electrical Engineering, University of California at Los Angeles, under the scholarship of the Ministry of Education, Science and Culture. From 2001 to 2003, he was a Research Associate with The University of Electro-Communications, Tokyo, Japan. From 2003 to 2024, he was with Yokohama National University, Yokohama, Japan, where he was a Professor since 2015. In April 2024, he joined Osaka University as a Professor. From 2003 to 2004, he was a Visiting Scientist with Harvard University, Cambridge, MA, USA. From 2019 to 2020, he was a Visiting Professor with the University of Waterloo, ON, Canada, and a Visiting Fellow with Princeton University, NJ, USA. His research interests are in the areas of wireless communications and networks. He served as an Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS from 2007 to 2011 and IEEE WIRELESS COMMUNICATIONS LETTERS from 2011 to 2016.