# Variational Bayes Estimation for Affine-Precoded Superimposed Pilots in Partially Connected Dual-Wideband Tera-Hertz MU-MIMO Systems

Abhisha Garg*, Suraj Srivastava†, Aditya K. Jagannatham*,
Department of Electrical Engineering, Indian Institute of Technology Kanpur, India*
Department of Electrical Engineering, Indian Institute of Technology Jodhpur, India†
(e-mail: abhisha20@iitk.ac.in*; surajsri@iitj.ac.in†; adityaj@iitk.ac.in*)

***Abstract*—This work conceives two affine precoding based system models, common precoding with joint channel estimation (CP-JCE) and user-specific precoding for decoupled channel estimation (USP-DCE). Considering a *dual-wideband* effected partially connected architecture, we rigorously model the terahertz (THz) multiple input multiple output (MIMO) channel for each subarray corresponding to each user by incorporating the absorption, reflection, and free-space losses. Next, to address the significant bandwidth overhead associated with conventional pilot-based channel estimation, we employ superimposed pilots. Building on this, we formulate a structured sparse channel model and develop a variational Bayesian inference algorithm that jointly estimates the channel coefficients and learns the underlying sparsity structure through hyperparameter inference, thereby enabling robust and high-precision superimposed pilot-based channel estimation under severe model uncertainty. Lastly, we compare our results for both systems and provide a trade-off analysis between them.**

***Index Terms*—TeraHertz, dual-wideband, superimposed pilot, variational Bayes, multi-user**

## I. Introduction

High frequency terahertz (THz) communication, spanning from $0.3-10$ THz band [1], is poised to revolutionize next-generation wireless communication. In addition to the traditional frequency selectivity challenge termed *frequency-wideband effect*, THz systems also experience differential delays across large antenna arrays termed *spatial-wideband effect*. The coexistence of both leads to the *dual-wideband effect* [2], giving rise to the *beam split effect* caused by frequency-dependent angle-of-arrival/ angle-of-departure (AoA/AoD) variations across subcarriers. Therefore, the precise knowledge of channel state information (CSI) is crucial in such systems to enable accurate channel estimation and optimization.

Conventional cellular systems allocate dedicated pilot resources, simplifying estimation but reducing data rate due to pilot overhead. In contrast, superimposed pilots (SIP) embed low-power pilots within data, allowing simultaneous data transmission and channel estimation without extra time slots [3]. Jesbin *et al.* [4], proposed a sparse SIP scheme for channel estimation in OTFS systems, where all bins in a frame carry data symbols while pilot symbols are sparsely superimposed over a subset of them. Jing *et al.* [5], proposed an iterative channel estimation technique based on Tikhonov regularization for multi-user multiple input multiple output (MU-MIMO) scenarios, aimed at improving estimation accuracy under interference-limited conditions. Moreover, the high dimensionality of antenna arrays and the limited number of multipath components in THz channels give rise to inherent sparsity, making sparse channel estimation well suited to a compressive sensing framework. Bayesian frameworks explicitly model the sparse structure by assigning a prior distribution and updating the information via a posterior distribution conditioned on the observations. For example, the Bernoulli-Gaussian distribution [6] is widely used to represent sparse signals. Moreover, sparsity-promoting priors include *Gaussian-gamma distribution* [7], which effectively captures the sparsity pattern of the underlying signal components. The next section presents the novel contributions of this work.

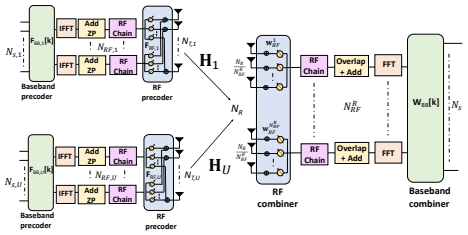



Figure 1. Schematic diagram of an MU-MIMO based OFDM system.

### A. Novel contributions of the paper

To the best of our knowledge, this is the first work that features a MU affine precoded SIP technique for a partially connected THz MIMO system impacted by the dual-wideband effect. Toward this, we propose two system models viz., the common precoding for joint channel estimation (CP-JCE) and user-specific precoding for decoupled channel estimation approaches (USP-DCE). We develop a sparse channel model for both scenarios and estimate the THz channel using *variational Bayesian inference*. Notably, the Bayesian framework employed does not rely solely on the prior distribution for inference; instead, it integrates information from the likelihood function during the inference process to refine the posterior distribution. Both the system models, i.e., CP-JCE and USP-DCE, lead to performance trade-offs, which are analyzed in detail.

## II. AP-SIP based THz MU-MIMO System Model

Consider a THz MU-MIMO orthogonal frequency division multiplexing (OFDM) uplink system where a base station (BS) is equipped with $N_{sub}$ subarrays and the $l$-th subarray consists of $N_R^l$ antennas which serves $U$ users simultaneously with each user possessing $N_{T,u}$ transmit antennas, as shown in the Fig.1. Furthermore, we have $N_{\rm RF}^R$ RF chains at the BS, while each user possess $N_{{\rm RF},u}$ RF chains and satisfies the relation $\sum_{u=1}^{U} N_{{\rm RF},u} \leq N_{\rm RF}^R = N_{sub} \ll N_R$. Therefore, the number of receive antennas at the BS is $N_R = \sum_{l=1}^{N_{\rm RF}^R} N_R^l$, while the total number of transmit antennas for all the users is $N_T = \sum_{u=1}^{U} N_{T,u}$ with total number of RF chains as $N_{\rm RF}^T = \sum_{u=1}^{U} N_{{\rm RF},u}$. Note that at the user equipment (UE), we have a fully-connected architecture, as shown in Fig. 1. At the receiver, a frequency-flat RF combiner $\mathbf{W}_{\rm RF} \in \mathbb{C}^{N_{\rm RF}^R \times N_R}$ is cascaded with a bank of frequency-selective baseband combiners $\mathbf{W}_{\rm BB}[k] \in \mathbb{C}^{N_s \times N_{\rm RF}^R}$ for each subcarrier $k$. Similarly, each user employs a frequency-flat RF precoder $\tilde{\mathbf{F}}_{{\rm RF},u} \in \mathbb{C}^{N_{T,u} \times N_{{\rm RF},u}}$ followed by frequency-selective baseband precoders $\tilde{\mathbf{F}}_{{\rm BB},u}[k] \in \mathbb{C}^{N_{{\rm RF},u} \times N_{s,u}}$. The RF

components satisfy constant-modulus constraints $|\mathbf{W}_{\mathrm{RF}}(a,b)| = \frac{1}{\sqrt{N_R}}$, $|\mathbf{F}_{\mathrm{RF}_u}(a,b)| = \frac{1}{\sqrt{N_{T,u}}}$.

We consider a wideband THz channel $\mathbf{H}_{d,u}^{l} \in \mathbb{C}^{N_R^l \times N_{T,u}}$, $0 \leq d \leq D$, for the $d$-th tap of user $u$ on subarray $l$. Therefore, the overall channel matrix of the partially connected architecture at the BS corresponding to $u$-th user is $\mathbf{H}_u[k] = \left[\left(\mathbf{H}_u^1[k]\right)^T \left(\mathbf{H}_u^2[k]\right)^T \cdots \left(\mathbf{H}_u^{N_{\mathrm{RF}}^R}[k]\right)^T\right]^T \in \mathbb{C}^{N_R \times N_T}$, where $\mathbf{H}_u^l[k] \in \mathbb{C}^{N_R^l \times N_{T,u}}$ denotes the channel between $u$-th user and $l$-th subarray. Moreover, at the receiver, the analog combiner possess the structure $\mathbf{W}_{\mathrm{RF}} = \mathrm{blkdiag}\big(\bar{\mathbf{w}}_{\mathrm{RF}}^{s_1}\ \bar{\mathbf{w}}_{\mathrm{RF}}^{s_2} \cdots \bar{\mathbf{w}}_{\mathrm{RF}}^{s_{N_{\mathrm{RF}}^R}}\big)$ where $\bar{\mathbf{w}}_{\mathrm{RF}}^{l} \in \mathbb{C}^{N_{\mathrm{BS}}^l \times 1}$. Thus, the frequency domain output $\tilde{\mathbf{Y}}_{\mathrm{MU},m}[k] \in \mathbb{C}^{N_{\mathrm{RF}}^R \times Q}$ corresponding to the $m$-th block at the $k$-th subcarrier is given as

$$\begin{aligned}\tilde{\mathbf{Y}}_{\mathrm{MU},m}[k] = \mathbf{W}_{\mathrm{RF},m}^H \mathbf{H}_{\mathrm{MU}}[k]\mathbf{F}_{\mathrm{RF},m}\tilde{\mathbf{S}}_{\mathrm{MU},m}[k] + \\ \mathbf{W}_{\mathrm{RF},m}^H \tilde{\mathbf{V}}_{\mathrm{MU},m}[k],\end{aligned} \tag{1}$$

where $\mathbf{H}_{\mathrm{MU}}[k] \in \mathbb{C}^{N_R \times N_T}$ represents the concatenated channel for all the $U$ users and is given as $\mathbf{H}_{\mathrm{MU}} = [\mathbf{H}_1[k]\,\mathbf{H}_2[k] \cdots \mathbf{H}_U[k]]$, $\mathbf{F}_{\mathrm{RF},m} \in \mathbb{C}^{N_T \times N_{\mathrm{RF}}^T}$ represents the stacked RF precoder corresponding to $U$ users for the $m$-th block and is formulated as $\mathbf{F}_{\mathrm{RF},m} = \mathrm{blkdiag}\big(\tilde{\mathbf{F}}_{\mathrm{RF},1,m} \cdots \tilde{\mathbf{F}}_{\mathrm{RF},U,m}\big)$. The quantity $\tilde{\mathbf{V}}_{\mathrm{MU},m}[k] \in \mathbb{C}^{N_{\mathrm{RF}}^R \times Q}$ represents the additive white Gaussian noise (AWGN) the elements of which are modeled as independent and identically distributed (i.i.d) complex Gaussian with $\mathcal{CN}(0,\sigma^2)$ whereas $\tilde{\mathbf{S}}_{\mathrm{MU},m}[k] \in \mathbb{C}^{N_{\mathrm{RF}}^T \times Q}$ represents the combined affine-precoded transmit superimposed training and data symbols corresponding to all the users and given as $\tilde{\mathbf{S}}_{\mathrm{MU},m}[k] = \big[\tilde{\mathbf{S}}_{1,m}[k]; \tilde{\mathbf{S}}_{2,m}[k];\ \cdots\ ; \tilde{\mathbf{S}}_{U,m}[k]\big]$. Moreover, $\tilde{\mathbf{S}}_{u,m}[k] \in \mathbb{C}^{N_{\mathrm{RF},u} \times Q}$ represents the superimposed training and data symbols corresponding to $u$-th user given by

$$\tilde{\mathbf{S}}_{u,m}[k] = \tilde{\mathbf{S}}_{d,u,m}[k]\mathbf{X}_{d,m} + \tilde{\mathbf{S}}_{p,u,m}[k]\mathbf{X}_{p,m}, \tag{2}$$

where $\tilde{\mathbf{S}}_{d,u,m}[k] \in \mathbb{C}^{N_{\mathrm{RF},u} \times Q_1}$ represents the data symbols corresponding to the $u$-th user while $\tilde{\mathbf{S}}_{p,u,m}[k] \in \mathbb{C}^{N_{\mathrm{RF},u} \times Q_2}$ represents the pilot symbols. Note that, $Q$ represents the superimposed block length, while $Q_1$ represents the block length corresponding to the data symbols and $Q_2$ represents the block length corresponding to pilot symbols, obeying $Q = Q_1 + Q_2$. The quantities $\mathbf{X}_{d,m} \in \mathbb{C}^{Q_1 \times Q}$, $\mathbf{X}_{p,m} \in \mathbb{C}^{Q_2 \times Q}$ represent the precoding matrices corresponding to data and training symbols, respectively, which can be readily obtained through unitary matrices as

$$\begin{aligned}\mathbf{X}_{d,m}\mathbf{X}_{d,m}^H &= \mathbf{I}_{Q_1 \times Q_1} \qquad \mathbf{X}_{p,m}\mathbf{X}_{p,m}^H = \mathbf{I}_{Q_2 \times Q_2} \\ \mathbf{X}_{d,m}\mathbf{X}_{p,m}^H &= \mathbf{0}_{Q_1 \times Q_2} \qquad \mathbf{X}_{p,m}\mathbf{X}_{d,m}^H = \mathbf{0}_{Q_2 \times Q_1}.\end{aligned} \tag{3}$$

However, it is interesting to note that these unitary matrices, $\mathbf{X}_{d,m}$ and $\mathbf{X}_{p,m}$, possess a special structure that can facilitate easier decoupling across users.

### *A. Common precoding with joint channel estimation (CP-JCE)*

For this scenario, we set $\mathbf{X}_{p,u,m} = \mathbf{X}_{p,m}$ and $\mathbf{X}_{d,u,m} = \mathbf{X}_{d,m}$, $\forall\, u$. In order to obtain the pilot model, we post-multiply the received output $\tilde{\mathbf{Y}}_{\mathrm{MU},m}[k]$ by the common training precoding matrix $\mathbf{X}_{p,m}^H$. Substituting Eq. (2) into Eq. (1), we get

$$\begin{aligned}\tilde{\mathbf{Y}}_{\mathrm{MU},m}[k]\mathbf{X}_{p,m}^H = \mathbf{W}_{\mathrm{RF},m}^H \mathbf{H}_{\mathrm{MU}}[k]\mathbf{F}_{\mathrm{RF},m}\tilde{\mathbf{S}}_{\mathrm{MU}\mathcal{P},m}[k] + \\ \mathbf{W}_{\mathrm{RF},m}^H \tilde{\mathbf{V}}_{\mathrm{MU},m}[k]\mathbf{X}_{p,m}^H,\end{aligned} \tag{4}$$

where $\tilde{\mathbf{S}}_{\mathrm{MU}\mathcal{P},m}[k] \in \mathbb{C}^{N_{\mathrm{RF}}^T \times Q_2}$ represents the concatenated pilot matrix for all the $U$ users and is given as $\tilde{\mathbf{S}}_{\mathrm{MU}\mathcal{P},m}[k] = \big[\tilde{\mathbf{S}}_{p,1,m}[k];\ \tilde{\mathbf{S}}_{p,2,m}[k];\ \cdots\ ; \tilde{\mathbf{S}}_{p,U,m}[k]\big]$. Let $\mathbf{V}_{\mathrm{MU}\mathcal{P},m}[k] = \mathbf{W}_{\mathrm{RF},m}^H \tilde{\mathbf{V}}_{\mathrm{MU},m}[k]\mathbf{X}_{p,m}^H \in \mathbb{C}^{N_{\mathrm{RF}}^R \times Q_2}$ represent the equivalent noise matrix, $\mathbf{Y}_{\mathrm{MU}\mathcal{P},m}[k] = \tilde{\mathbf{Y}}_{\mathrm{MU},m}[k]\mathbf{X}_{p,m}^H \in \mathbb{C}^{N_{\mathrm{RF}}^R \times Q_2}$ represent the equivalent received signal and $\mathbf{S}_{\mathrm{MU}\mathcal{P},m}[k] = \mathbf{F}_{\mathrm{RF},m}\tilde{\mathbf{S}}_{\mathrm{MU}\mathcal{P},m}[k] \in \mathbb{C}^{UN_{T,u} \times Q_2}$ represent the equivalent pilot symbols. The noise covariance matrix $\mathbb{E}\{\mathbf{V}_{\mathrm{MU}\mathcal{P},m}[k]\mathbf{V}_{\mathrm{MU}\mathcal{P},m}^H[k]\} = \mathcal{CN}\big(\mathbf{0}_{N_{\mathrm{RF}}^R \times 1}, \mathbf{C}_m\big)$ with $\mathbf{C}_m \in \mathbb{C}^{N_{\mathrm{RF}}^R \times N_{\mathrm{RF}}^R}$.

### *B. User-specific precoding with decoupled channel estimation (USP-DCE)*

We enforce another condition in addition to (3) which ensures the orthogonality of pilot and data precoding matrices across users, i.e., $\mathbf{X}_{p,u_1,m}\mathbf{X}_{p,u_2,m}^H = \mathbf{0}$ and $\mathbf{X}_{d,u_1,m}\mathbf{X}_{d,u_2,m}^H = \mathbf{0}, \forall\, u_1 \neq u_2$. Substituting this condition into Eq. (1) and post-multiplying by the pilot precoding matrix $\mathbf{X}_{p,u,m}$ corresponding to the $u$-th user, we obtain

$$\begin{aligned}\tilde{\mathbf{Y}}_{\mathrm{MU},m}[k]\mathbf{P}_{p,m}^H = \mathbf{W}_{\mathrm{RF},m}^H \mathbf{H}_u[k]\mathbf{F}_{\mathrm{RF},u,m}\tilde{\mathbf{S}}_{p,u,m}[k] + \\ \mathbf{W}_{\mathrm{RF},m}^H \tilde{\mathbf{V}}_{\mathrm{MU},m}[k]\mathbf{P}_{p,m}^H,\end{aligned} \tag{5}$$

where $\mathbf{Y}_{p,u,m}[k] = \tilde{\mathbf{Y}}_{\mathrm{MU},m}[k]\mathbf{P}_{p,m}^H \in \mathbb{C}^{N_{\mathrm{RF}}^R \times Q_2}$ represents the received signal corresponding to the $u$-th user while $\mathbf{V}_{p,u,m} = \mathbf{W}_{\mathrm{RF},m}^H \tilde{\mathbf{V}}_{\mathrm{MU},m}[k]\mathbf{P}_{p,m}^H \in \mathbb{C}^{N_{\mathrm{RF}}^R \times Q_2}$ represents the effective noise. Additionally, the noise covariance matrix $\mathbb{E}\left\{\mathbf{V}_{p,u,m}[k]\mathbf{V}_{p,u,m}^H[k]\right\} = \mathcal{CN}\big(\mathbf{0}_{N_{\mathrm{RF}}^R \times 1}, \mathbf{C}_{u,m}\big)$ with $\mathbf{C}_{u,m} \in \mathbb{C}^{N_{\mathrm{RF}}^R \times N_{\mathrm{RF}}^R}$. Let $[\mathbf{A}_{p,m}, \cdots, \mathbf{P}_{p,m}]$ represents the precoding matrix corresponding to each user which are generated from a large discrete Fourier transform (DFT) matrix with orthogonal columns. These conditions can be practically satisfied by selecting the precoding matrices as sub-matrices of an unitary matrix, where each semiunitary submatrix corresponds to a specific user. Furthermore, it can be observed that the model in Eq. (5) decouples the received signals for individual users.

## III. MU-THz Channel Model and Sparse Model

The array response vector $\tilde{\mathbf{a}}_R^l(\theta, f_k)$ considering the spatial-wideband effect [2] for the $l$-th subarray with $N_R^l$ elements, can be expressed as

$$\mathbf{a}^l(\theta, f_k) = \frac{1}{\sqrt{N_R^l}}\left[1\ e^{-j\pi\delta_k \cos\theta}\ \cdots\ e^{-j(N_R^l-1)\pi\delta_k \cos\theta}\right]^T,$$

where $\delta_k = \frac{f_k}{f_c}$ represents the relative subcarrier frequency with $f_k \triangleq \left(k - \frac{K+1}{2}\right)\frac{B}{K}$. Furthermore, the quantity $\theta^l$ represents the incident angle, while $B$ denotes the bandwidth. Additionally, the THz channel includes a line-of-sight (LoS) component and few non-LoS (NLoS) components. Therefore, the channel model for subcarrier $k$ at the $l$-th subarray can be expressed as $\mathbf{H}_u^l[k] = \mathbf{H}_u^{l,\mathrm{LoS}} + \mathbf{H}_u^{l,\mathrm{NLoS}}$ given as

$$\mathbf{H}_u^{l,\mathrm{LoS}}[k] = \sqrt{N_{T,u}N_R^l}\,\alpha(f_k, d)\gamma_\tau \mathfrak{B}_{T,u}\mathfrak{B}_R^l \mathbf{a}_R^l(\theta, f_k)\,\mathbf{a}_u^H(\phi, f_k), \tag{6}$$

$$\begin{aligned}\mathbf{H}_u^{l,\mathrm{NLoS}}[k] = \sqrt{\frac{N_{T,u}N_R^l}{N_{\mathrm{NLoS}}N_{\mathrm{ray}}}} \sum_{\ell=1}^{N_{\mathrm{NLoS}}} \sum_{æ=1}^{N_{\mathrm{ray}}} \alpha_{\ell,æ}(f_k, d_{\ell,æ}) \\ \gamma_{\tau_{\ell,æ}} \mathfrak{B}_{T,u}\mathfrak{B}_R^l \mathbf{a}_R^l(\theta_{\ell,æ}, f_k)\,\mathbf{a}_u^H(\phi_{\ell,æ}, f_k),\end{aligned} \tag{7}$$

and $\gamma_{\tau_{\ell,æ}} = \sum_{\jmath=0}^{K-1} p(\jmath T_s - \tau_{\ell,æ}) e^{-j\frac{2\pi k \jmath}{K}}$, $\forall\, k, \jmath$. The quantities $\mathfrak{B}_{T,u}$ and $\mathfrak{B}_R$ represent the transmit and receive antenna gains, respectively. Furthermore, the variables $p(.), \tau_{(.)}, N_{\mathrm{NLoS}}, N_{\mathrm{ray}}$,

$\alpha(f_k, d)$ denotes the pulse shaping filter, delay, number of NLoS components, number of diffused rays and complex-path gain respectively [8]. Additionally, the molecular absorption coefficient is calculated using HITRAN database [8].

The extended virtual channel representation corresponding to the $l$-th subarray can be expressed as $\mathbf{H}_u^l[k] = \mathbf{A}_R^l(\Theta_R, f_k)\,\mathbf{H}_{b,u}^l[k]\mathbf{A}_{T,u}^H(\Phi_{T,u}, f_k)$ where $\mathbf{A}_R^l(\Theta_R, f_k) \in \mathbb{C}^{N_R^l \times G_R^l}$, $\mathbf{A}_{T,u}(\Phi_{T,u}, f_k) \in \mathbb{C}^{G_{T,u} \times N_{T,u}}$ represents the receive and transmit array manifold given as

$$\mathbf{A}_R^l(\Theta_R, f_k) = \big[\mathbf{a}_R^l(\theta_1, f_k)\ \mathbf{a}_R^l(\theta_2, f_k)\ \cdots\ \mathbf{a}_R^l(\theta_{G_R}, f_k)\big], \quad (8)$$

$$\mathbf{A}_{T,u}(\Phi_{T,u}, f_k) = \big[\mathbf{a}_{T,u}(\phi_{1,u}, f_k)\ \cdots\ \mathbf{a}_{T,u}(\phi_{G_{T,u}}, f_k)\big], \quad (9)$$

and $G_{T,u}$ and $G_R^l$ represent the transmit and receive bin with $\Phi_{T,u}$ and $\Theta_R^l$ as directional cosines [8] and $\mathbf{H}_{b,u}^l$ represents the sparse channel corresponding to $l$-th subarray which needs to be estimated. Furthermore, the overall extended virtual channel corresponding to each user for all the subarrays can be given as $\mathbf{H}_u[k] = \mathbf{A}_R(\Theta_R, f_k)\,\mathbf{H}_{b,u}[k]\mathbf{A}_{T,u}^H(\Phi_{T,u}, f_k)$ where $\mathbf{A}_R(\Theta_R, f_k) = \mathrm{blkdiag}\big(\mathbf{A}_R^1(\Theta_R, f_k)\ \ \mathbf{A}_R^2(\Theta_R, f_k)\ \ \cdots\ \ \mathbf{A}_R^{N_{RF}^R}(\Theta_R, f_k)\big)$ and $\mathbf{H}_{b,u}[k] = \big[\mathbf{H}_{b,u}^1[k]; \mathbf{H}_{b,u}^2[k]; \cdots; \mathbf{H}_{b,u}^{N_{RF}^R}[k]\big]$. Let $\tilde{\mathbf{\Psi}}_u[k] \in \mathbb{C}^{N_R N_{T,u} \times G_R G_{T,u}}$ represent the *sparsifying dictionary* which is defined as $\mathbf{A}_{T,u}^*(\Phi_{T,u}, f_k) \otimes \mathbf{A}_R(\Theta_R, f_k)$. The vectorized channel frequency response (CFR) for all the $U$ users can be determined as $\mathbf{h}_{\mathrm{MU}}[k] = \mathrm{blkdiag}\underbrace{\big(\tilde{\mathbf{\Psi}}_1[k]\ \tilde{\mathbf{\Psi}}_2[k]\ \cdots\ \tilde{\mathbf{\Psi}}_U[k]\big)}_{\mathbf{\Psi}_{\mathrm{MU}}[k]}\underbrace{\big[\mathbf{h}_{b,1}^T[k]\mathbf{h}_{b,2}^T[k]\cdots\mathbf{h}_{b,U}^T[k]\big]^T}_{\mathbf{h}_{b,\mathrm{MU}}[k]}$, where $\mathbf{h}_{\mathrm{MU}}[k] = \mathrm{vec}(\mathbf{H}_{\mathrm{MU}}[k])$, $\mathbf{\Psi}_{\mathrm{MU}}[k] \in \mathbb{C}^{N_R N_T \times G_R G_T}$ represents the joint sparsifying dictionary and $\mathbf{h}_{b,\mathrm{MU}}[k] \in \mathbb{C}^{G_R G_T \times 1}$ represents the joint beamspace output for all the users where $\sum_{u=1}^{U} G_{T,u} = G_T$. After employing the vec(.) operator in (4), the MU equivalent sensing matrix, vectorized received signal, vectorized noise signal can be given as $\tilde{\mathbf{\Lambda}}_{\mathrm{MU}_{\mathcal{P}},m}[k] = \big(\mathbf{S}_{\mathrm{MU}_{\mathcal{P}},m}^T[k] \otimes \mathbf{W}_{\mathrm{RF},m}^H\big) \in \mathbb{C}^{Q_2 N_{\mathrm{RF}}^R \times N_T N_R}$, $\mathbf{y}_{\mathrm{MU}_{\mathcal{P}},m}[k] = \mathrm{vec}\,(\mathbf{Y}_{\mathrm{MU},m}[k])$, $\mathbf{v}_{\mathrm{MU}_{\mathcal{P}},m}[k] = \mathrm{vec}\,(\mathbf{V}_{\mathrm{MU}_{\mathcal{P}},m}[k])$ respectively. Therefore, the joint CP-JCE based channel model is given as

$$\mathbf{y}_{\mathrm{MU}_{\mathcal{P}},m}[k] = \underbrace{\tilde{\mathbf{\Lambda}}_{\mathrm{MU}_{\mathcal{P}},m}[k]\mathbf{\Psi}_{\mathrm{MU}}[k]}_{\mathbf{\Lambda}_{\mathrm{MU}_{\mathcal{P}},m}[k]}\mathbf{h}_{b,\mathrm{MU}}[k] + \mathbf{v}_{\mathrm{MU}_{\mathcal{P}},m}[k]. \quad (10)$$

Furthermore, to develop a compatible model for the $k$-th subcarrier, we concatenate the outputs for all the $M$ blocks

$$\underbrace{\begin{bmatrix}\mathbf{y}_{\mathrm{MU}_{\mathcal{P}},1}[k]\\ \vdots \\ \mathbf{y}_{\mathrm{MU}_{\mathcal{P}},M}[k]\end{bmatrix}}_{\mathbf{y}_{\mathrm{MU}_{\mathcal{P}}}[k]} = \underbrace{\begin{bmatrix}\mathbf{\Lambda}_{\mathrm{MU}_{\mathcal{P}},1}[k]\\ \vdots \\ \mathbf{\Lambda}_{\mathrm{MU}_{\mathcal{P}},M}[k]\end{bmatrix}}_{\mathbf{\Lambda}_{\mathrm{MU}_{\mathcal{P}}}[k]}\mathbf{h}_{b,\mathrm{MU}} + \underbrace{\begin{bmatrix}\mathbf{v}_{\mathrm{MU}_{\mathcal{P}},1}[k]\\ \vdots \\ \mathbf{v}_{\mathrm{MU}_{\mathcal{P}},M}[k]\end{bmatrix}}_{\mathbf{v}_{\mathrm{MU}_{\mathcal{P}}}[k]},$$

where $\mathbf{C}_w = \mathbb{E}\{\mathbf{v}_{\mathrm{MU}_{\mathcal{P}}}[k]\mathbf{v}_{\mathrm{MU}_{\mathcal{P}}}^H[k]\}$ is the noise covariance. Following a similar procedure, one can obtain the sparse channel estimation model for the USP-DCE approach, which is given as

$$\mathbf{y}_{p,u,m}[k] = \mathbf{\Lambda}_{p,u,m}[k]\mathbf{h}_u[k] + \mathbf{v}_{p,u,m}[k], \quad (11)$$

where $\mathbf{y}_{p,u,m}[k] = \mathrm{vec}(\mathbf{Y}_{u,m}[k])$, $\mathbf{h}_u[k] = \mathrm{vec}(\mathbf{H}_u[k])$ and $\mathbf{v}_{p,u,m}[k] = \mathrm{vec}(\mathbf{V}_{u,m}[k])$. Furthermore, $\mathbf{\Lambda}_{p,u,m}[k] = (\tilde{\mathbf{S}}_{p,u,m}^T[k]\mathbf{F}_{\mathrm{RF},u,m}^T \otimes \mathbf{W}_{\mathrm{RF},m}^H) \in \mathbb{C}^{Q_2 N_{\mathrm{RF}}^R \times N_{T,u} N_R}$ represents the equivalent sensing matrix corresponding to $u$-th user. Similarly, one can obtain the stacked model for the USP-DCE approach for all the $M$ blocks which is omitted here due to space constraints. Interestingly, both (10) and (11) represent sparse channel estimation models for the joint channel $\mathbf{h}_{b,\mathrm{MU}}$ of all the users and $\mathbf{h}_{b,u}$ of the $u$-th user. Consequently, we provide a detailed sparse estimation approach for the CP-JCE model of (10). The next section introduces a variational Bayesian inference algorithm for sparse CSI estimation.

## IV. Variational Inference based sparse estimation

Bayesian probabilistic models are learning models that assume the unknown variables obeying certain posterior distribution, which is composed of prior distributions reflecting our belief on the variables before any observation, and the likelihood function describing the observations. In this regard, let the parameterized Gaussian prior $g(\mathbf{h}_{b,\mathrm{MU}}[k]; \mathbf{\Gamma}_{k,\mathrm{MU}})$ corresponding to the $k$-th subcarrier be defined as

$$g(\mathbf{h}_{b,\mathrm{MU}}[k]; \mathbf{\Gamma}_{k,\mathrm{MU}}) = \prod_{\imath=1}^{UG_RG_{T,u}} (\pi\gamma_{k,\imath})^{-1}\exp\Big(-\frac{|\mathbf{h}_{b,\mathrm{MU}}[k](\imath)|^2}{\gamma_{k,\imath}}\Big), \quad (12)$$

where $\gamma_{k,\imath}$ represents the $\imath$-th hyperparameter corresponding to $k$-th subcarrier and $\mathbf{\Gamma}_{k,\mathrm{MU}} = \mathrm{diag}(\gamma_{k,1}, \gamma_{k,2}, \cdots, \gamma_{k,UG_RG_{T,u}}) \in \mathbb{R}^{UG_RG_{T,u} \times UG_RG_{T,u}}$ denotes the hyperparameter matrix. To promote sparsity, we adopted the *Gaussian-Gamma* prior, which is widely used in the machine learning literature [7], where the Gamma prior is conjugate to the Gaussian prior.

**Sparsity-enhancing Prior:** We constrained the hyperparameters $\bar{\gamma}_k = (\gamma_{k,1}, \cdots, \gamma_{k,UG_RG_{T,u}})^T$ by treating them as random variables and imposing a *Gamma prior distribution* as

$$\begin{aligned} g(\bar{\gamma}_k; \kappa, \chi) &= \mathrm{Gamma}(\gamma_{k,\imath}|\kappa_\imath, \chi_\imath), \\ &= \frac{\chi^\kappa}{\Upsilon(\kappa)}\gamma_{k,\imath}^{(\kappa-1)}\exp(-\chi\gamma_{k,\imath}),\ \gamma_{k,\imath} > 0, \end{aligned} \quad (13)$$

where $\kappa > 0$ and $\chi > 0$ while $\Upsilon(\kappa) = \int_0^\infty u^{\kappa-1}e^{-u}du$. Therefore, the joint distribution over all the hyperparameters can be given as $g(\bar{\gamma}_k) = \prod_{\imath=1}^{UG_RG_{T,u}} g(\gamma_{k,\imath}|\kappa_\imath, \chi_\imath)$.

**Modelling unknown noise precision:** We assume the Gamma prior distribution $g(\zeta; \varphi, \varsigma) = \mathrm{Gamma}(\zeta|\varphi, \varsigma)$ over the noise precision $\frac{1}{\sigma^2}$ to effectively adapt the actual noise condition in the observed data. Note that, the high dimensional integral $g(\mathbf{y}_{\mathrm{MU}_{\mathcal{P}}}[k]) = \int g(\mathbf{y}_{\mathrm{MU}_{\mathcal{P}}}[k], \mathcal{J})d\mathcal{J}$ is mathematically intractable, where $\mathcal{J} = \{\mathbf{h}_{b,\mathrm{MU}}[k], \bar{\gamma}_k, \zeta\}$ represents the set of latent variables. To address this, we employ variational Bayes inference, which seeks an approximate distribution that is close to the posterior distribution in Kullback-Leibler (KL) divergence and formulated as Eq. (14), where $f(\mathcal{J})$ represents a tractable distribution. However, such sophisticated probabilistic model does not admit exact Bayesian inference, as the posterior distribution of the unknown variables cannot be obtained in close-form. Consequently, we approximate the Bayesian inference by applying the *mean-field* approximation as outlined in Eq. (16), with their individual posterior updates detailed below.

**Computing the posterior of $\mathbf{h}_{b,\mathrm{MU}}[k]$:** Taking the logarithm on both sides of Eq. (17) and solving further yields Eq. (18). Therefore, the resulting function is quadratic and can be approximated by a standard Gaussian, expressed as $f(\mathbf{h}_{b,\mathrm{MU}}[k]) \approx \mathcal{CN}(\boldsymbol{\mu}_k, \mathbf{\Sigma}_k)$ where the *a posterior* mean and covariance are

$$\begin{aligned} \mathbf{\Sigma}_k &= \big(\mathbb{E}(\mathbf{\Gamma}_{k,\mathrm{MU}}) + \mathbb{E}(\zeta)\mathbf{\Lambda}_{\mathrm{MU}_{\mathcal{P}}}^H[k]\mathbf{C}_w^{-1}\mathbf{\Lambda}_{\mathrm{MU}_{\mathcal{P}}}^H[k]\big)^{-1}, \\ \boldsymbol{\mu}_k &= \mathbb{E}(\zeta)\mathbf{\Sigma}_k\mathbf{\Lambda}_{\mathrm{MU}_{\mathcal{P}}}^H[k]\mathbf{C}_w^{-1}\mathbf{y}_{\mathrm{MU}_{\mathcal{P}}}[k]. \end{aligned} \quad (22)$$

$$\ln g(\mathbf{y}_{\mathrm{MU}_{\mathcal{P}}}[k]) = \int \underbrace{f(\mathcal{J}) \ln \frac{g(\mathbf{y}_{\mathrm{MU}_{\mathcal{P}}}[k], \mathcal{J}) d\mathcal{J}}{f(\mathcal{J})}}_{\mathcal{L}(f)} - \int \underbrace{f(\mathcal{J}) \ln \frac{g(\mathcal{J}|\mathbf{y}_{\mathrm{MU}_{\mathcal{P}}}[k]) d\mathcal{J}}{f(\mathcal{J})}}_{\mathrm{KL}(f(\mathcal{J}) \| g(\mathcal{J}|\mathbf{y}_{\mathrm{MU}_{\mathcal{P}}}[k]))}, \tag{14}$$

$$g(\mathbf{h}_{b,\mathrm{MU}}[k], \bar{\gamma}_k, \zeta | \mathbf{y}_{\mathrm{MU}_{\mathcal{P}}}[k]) = \frac{g(\mathbf{y}_{\mathrm{MU}_{\mathcal{P}}}[k] | \mathbf{h}_{b,\mathrm{MU}}[k], \zeta) g(\mathbf{h}_{b,\mathrm{MU}}[k] | \gamma_k) g(\bar{\gamma}_k) g(\zeta)}{g(\mathbf{y}_{\mathrm{MU}_{\mathcal{P}}}[k])}, \tag{15}$$

$$g(\mathbf{h}_{b,\mathrm{MU}}[k], \bar{\gamma}_k, \zeta | \mathbf{y}_{\mathrm{MU}_{\mathcal{P}}}[k]; \kappa, \chi, \varphi, \varsigma) \approx \prod_{k=1}^{K} f(\mathbf{h}_{b,\mathrm{MU}}[k], \bar{\gamma}_k, \zeta) = \prod_{k=1}^{K} f(\mathbf{h}_{b,\mathrm{MU}}[k]) f(\bar{\gamma}_k) f(\zeta). \tag{16}$$

$$\ln f(\mathbf{h}_{b,\mathrm{MU}}[k]) = \mathbb{E}_{\bar{\gamma}_k, \zeta} \big\{ \ln g(\mathbf{y}_{\mathrm{MU}_{\mathcal{P}}}[k] | \mathbf{h}_{b,\mathrm{MU}}[k], \zeta) + \ln g(\mathbf{h}_{b,\mathrm{MU}}[k] | \bar{\gamma}_k) + \ln g(\bar{\gamma}_k) + \ln g(\zeta) \big\}, \tag{17}$$

$$\ln f(\mathbf{h}_{b,\mathrm{MU}}[k]) = \mathbb{E}_{\bar{\gamma}_k, \zeta} \Big\{ -\zeta (\mathbf{y}_{\mathrm{MU}_{\mathcal{P}}}[k] - \mathbf{\Lambda}_{\mathrm{MU}_{\mathcal{P}}}[k] \mathbf{h}_{b,\mathrm{MU}}[k])^H \mathbf{C}_w^{-1} (\mathbf{y}_{\mathrm{MU}_{\mathcal{P}}}[k] - \mathbf{\Lambda}_{\mathrm{MU}_{\mathcal{P}}}[k] \mathbf{h}_{b,\mathrm{MU}}[k]) - \frac{1}{2} \mathrm{Tr} \big[ (\mathbf{\Gamma}_{k,\mathrm{MU}}) \mathbf{h}_{b,\mathrm{MU}}[k] \mathbf{h}_{b,\mathrm{MU}}^T[k] \big] \Big\}. \tag{18}$$

$$\begin{aligned} \ln f(\bar{\gamma}_k) &= \mathbb{E}_{\mathbf{h}_{b,\mathrm{MU}}[k]} \big\{ \ln g(\mathbf{h}_{b,\mathrm{MU}}[k] | \bar{\gamma}_k) + \ln g(\bar{\gamma}_k) \big\} + \mathrm{const} \\ &= \big(\kappa - \frac{1}{2}\big) \sum_{\imath=1}^{UG_R G_{T,u}} \ln(\gamma_{k,\imath}) - \sum_{\imath=1}^{UG_R G_{T,u}} (\chi + \mathbb{E}(|\mathbf{h}_{b,\mathrm{MU}}[k](\imath)|^2) \gamma_{k,\imath}) + \mathrm{const}. \end{aligned} \tag{19}$$

$$\ln f(\zeta) = \mathbb{E}_{\mathbf{h}_{b,\mathrm{MU}}[k]} \big\{ \ln g(\mathbf{y}_{\mathrm{MU}_{\mathcal{P}}}[k] | \mathbf{h}_{b,\mathrm{MU}}[k], \zeta) + \ln g(\zeta) \big\} + \mathrm{const}, \tag{20}$$

$$= \mathbb{E}_{\mathbf{h}_{b,\mathrm{MU}}[k]} \big\{ (\varphi - 1 + UG_R G_{T,u}) \ln(\zeta) + \zeta (\mathbf{y}_{\mathrm{MU}_{\mathcal{P}}}[k] - \mathbf{\Lambda}_{\mathrm{MU}_{\mathcal{P}}}[k] \mathbf{h}_{b,\mathrm{MU}}[k])^H \mathbf{C}_w^{-1} (\mathbf{y}_{\mathrm{MU}_{\mathcal{P}}}[k] - \mathbf{\Lambda}_{\mathrm{MU}_{\mathcal{P}}}[k] \mathbf{h}_{b,\mathrm{MU}}[k]) + \varsigma \big\} + \mathrm{const}. \tag{21}$$

**Computing the posterior of $\bar{\gamma}_k$:** By isolating the terms dependent on $\bar{\gamma}_k$ and taking the logarithm on both sides, we derive Eq. (19). Comparing with the Gamma distribution parameters, we obtain

$$\tilde{\kappa} = \kappa; \ \tilde{\chi} = \chi + \mathbb{E}\{|\mathbf{h}_{b,\mathrm{MU}}[k](\imath)|^2\}, \tag{23}$$

where $\mathbb{E}\big\{|\mathbf{h}_{b,\mathrm{MU}}[k](\imath)|^2\big\} = |\boldsymbol{\mu}_k(\imath)|^2 + \mathbf{\Sigma}_k(\imath, \imath)$.

**Computing the posterior of $f(\zeta)$:** Similarly, by rearranging the terms involving $\zeta$ and applying the logarithm to both sides, we establish Eq. (20), which consequently leads to Eq. (21). Therefore, by aligning the resulting expression with the standard Gamma distribution parameters, we obtain

$$\begin{aligned} \tilde{\varphi} &= (\varphi - 1 + UG_R G_{T,u}), \\ \tilde{\varsigma} &= \varsigma + \mathbb{E}_{\mathbf{h}_{b,\mathrm{MU}}[k]} \big\{ \big(\mathbf{y}_{\mathrm{MU}_{\mathcal{P}}}[k] - \mathbf{\Lambda}_{\mathrm{MU}_{\mathcal{P}}}[k] \mathbf{h}_{b,\mathrm{MU}}[k]\big)^H \\ &\quad \mathbf{C}_w^{-1} \big(\mathbf{y}_{\mathrm{MU}_{\mathcal{P}}}[k] - \mathbf{\Lambda}_{\mathrm{MU}_{\mathcal{P}}}[k] \mathbf{h}_{b,\mathrm{MU}}[k]\big) \big\}. \end{aligned} \tag{24}$$

Note that the resulting variational inference algorithm not only provides closed-form updates for each variable, but also accounts for the uncertainties associated with the estimated parameters.

**Hyperparameter setting:** We assume no prior knowledge where the hyperparameters $\{\kappa, \chi\}_{t=1}^{UG_R G_{T,u}}$ and $\{\varphi, \varsigma\}$ are set as $10^{-6}$ to represent non-informative priors. A similar procedure can be applied to the USP-DCE-based approach for sparse estimation; however, we have omitted it here due to space constraints. Algorithm-1 provides a summary of the sparse estimation procedure. The final estimate is given by

$$\widehat{\mathbf{H}}_{\mathrm{VBI}} = \mathbf{A}_R(\Theta_R, f_k) \mathrm{vec}^{-1}(\widehat{\mathbf{h}}_{\mathrm{VBI}}[k]) \mathbf{A}_{T,\mathrm{MU}}^H(\Phi_T, f_k),$$

where the concatenated transmit vector $\mathbf{A}_{T,\mathrm{MU}}^H(\Phi_T, f_k) = \mathrm{blkdiag}(\mathbf{A}_{T,1}(\Phi_T, f_k), \cdots, (\mathbf{A}_{T,U}(\Phi_{T,U}, f_k)) \in \mathbb{C}^{N_T G_T \times N_T G_T}$.

**Algorithm 1:** VBI based sparse channel estimation

**Input:** $\mathbf{y}_{\mathrm{MU}_{\mathcal{P}}}[k], \mathbf{\Lambda}_{\mathrm{MU}_{\mathcal{P}}}[k], \mathbf{C}_w^{-1}, \epsilon, N_{\max}$

**Initialization:** $\gamma_{k,\imath}^{(0)} = 1 \ \forall \ 0 \le \imath \le UG_R G_{T,u}; \widehat{\mathbf{\Gamma}}_{k,\mathrm{MU}}^{(0)} = \mathbf{I}_{UG_R G_{T,u}}; \widehat{\mathbf{\Gamma}}_{k,\mathrm{MU}}^{(-1)} = \mathbf{0}; j = 0$

**while** $\| \widehat{\mathbf{\Gamma}}_{k,\mathrm{MU}}^{(j)} - \widehat{\mathbf{\Gamma}}_{k,\mathrm{MU}}^{(j-1)} \|_{\mathcal{F}}^2 \ge \epsilon$ *and* $j < N_{\max}$ **do**

  $j = j + 1$

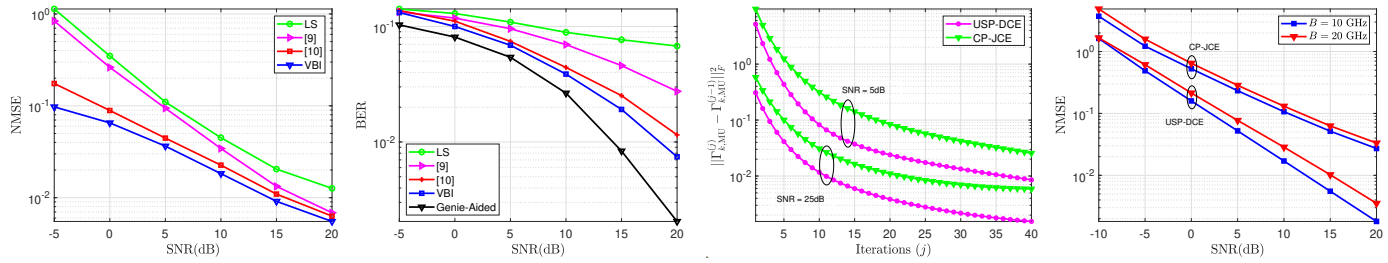
  Update the *a posteriori* mean and covariance using Eq. (22)

  Update the hyperparameters and noise covariance using Eq. (23) and (24) respectively.

**Output:** $\widehat{\mathrm{h}}_{\mathrm{VBI}}[k] = \boldsymbol{\mu}_k$

## V. Simulation Results

We consider a practical scenario of $U = 3$ users each employing $N_{T,u} = 4$ transmit antennas possessing $N_{\mathrm{RF},u} = 2$ RF chains. The BS is equipped with $N_R = 64$ antennas with $N_{\mathrm{RF}}^R = 8$ RF chains, where each subarray possess $N_R^l = \frac{N_R}{N_{\mathrm{RF}}^R} = 8$ antennas. The carrier frequency is $f_c = 0.671\,\mathrm{THz}$ and transmission distance is $d = 12\,\mathrm{m}$ with molecular absorption $k_{\mathrm{abs}}(f_k) = 0.01358\,\mathrm{m}^{-1}$. We consider a single LoS path and 3 NLoS path with $N_{\mathrm{ray}} = 1$ diffused ray. Furthermore, we consider $K = 64$ subcarriers with $B = 10$ GHz and $M = 15$ pilot blocks. The transmit dictionary includes $G_{T,u} = 8\,\mathrm{atoms}$ while each subarray includes $G_R^l = 16\,\mathrm{atoms}$. The transmit and receive antenna gains are $\sum_{u=1}^{U} \mathfrak{B}_{T,u} = 16\,\mathrm{dBi}$ and $\mathfrak{B}_R^l = 16\,\mathrm{dBi}$. The frequency-independent phase shifters are designed as described in [8]. Furthermore, we use Gaussian mixture model (GMM) to generate the AoAs/AoDs due to its accurate representation in the THz band [8]. The NMSE metric is defined as $\mathrm{NMSE} = \frac{\sum_{k=0}^{K-1} \|\widehat{\mathbf{H}}[k] - \mathbf{H}[k]\|_F^2}{\sum_{k=0}^{K-1} \|\mathbf{H}[k]\|_F^2}$. Fig. 2(a) illustrates the NMSE performance as a function of SNR. The poor performance of the least squares (LS) method is attributed to its inability to incorporate sparsity priors into the estimation process, which poses a significant limitation compared to CS-based approaches. Moreover, the




(a) (b) (c) (d)

Figure 2. Performance comparison between the proposed approach and existing state-of-the-art techniques $(a)$ NMSE vs SNR $(b)$ BER vs SNR $(c)$ $\| \mathbf{\Gamma}_{k,\mathrm{MU}}^{(j)} - \mathbf{\Gamma}_{k,\mathrm{MU}}^{(j-1)} \|_{\mathcal{F}}^2$ vs iterations for the proposed CP-JCE and USP-DCE system model $(d)$ NMSE vs SNR for the proposed CP-JCE and USP-DCE system model by varying bandwidth $B \in \{10, 20\}$ GHz.

algorithm in [9] is prone to both structural and convergence errors. In particular, setting a low stopping threshold introduces structural errors, whereas a high stopping threshold results in convergence errors. Additionally, the algorithm in [10] frequently converges to suboptimal local minima, leading to convergence errors. In contrast, variational Bayes inference assumes parameterized prior distributions and optimizes the variational parameters to approximate the posterior distribution, thereby enabling the joint estimation of all unknown variables and making it well-suited for diverse propagation environments. Fig. 2(b) depicts the accuracy of data detection achieved by employing the MMSE receiver with the proposed Bayesian inference, along with other conventional sparse estimation based techniques and the Genie-aided detector as benchmarks. As illustrated, the resulting BER decreases consistently with increasing SNR, which closely aligns with the NMSE results presented previously, underscoring the high quality of the CSI estimated by the Bayesian inference. Additionally, the BER of the proposed channel learning approach closely matches the hypothetical Genie-aided detector (derive from true channel knowledge) at higher SNRs, highlighting its efficacy and robustness in realistic scenarios. Fig. 2(c) depicts the convergence of the variational Bayes inference as a function of iterations for the CP-JCE and USP-DCE based models. To achieve a target NMSE of $0.1$, the USP-DCE model requires $9$ iterations, while the CP-JCE model requires $19$ iterations at $\mathrm{SNR} = 5$ dB. Note that, the complexity increase in CP-JCE remains moderate. This is because CP-JCE effectively exploits the common angular support shared among users, reducing redundancy and preventing complexity from scaling linearly with user count.

### A. Trade-off between the CP-JCE and USP-DCE approaches

*1) Complexity:* The CP-JCE approach utilizes a large channel matrix that includes the channels of all users, resulting in a higher computational complexity. In contrast, the USP-DCE method employs smaller channel matrices specific to individual users, leading to reduced complexity. Specifically, the complexity of CP-JCE is of the order $\mathcal{O}(U^3 G_{T,u}^3 G_R^3)$, while the complexity of USP-DCE is of the order of $\mathcal{O}(U G_{T,u}^3 G_R^3)$. This demonstrates that USP-DCE has a complexity advantage over CP-JCE, particularly as the number of users increases.

*2) NMSE performance:* From Fig. 2(d), it can be observed that the NMSE performance of the CP-JCE approach is poorer compared to the USP-DCE approach. This is attributed to the fact that the CP-JCE approach handles $U$ users simultaneously, whereas the USP-DCE method is designed for a single user. Consequently, the increased number of users in CP-JCE leads to higher estimation errors due to the added complexity and interference in the estimation process.

*3) Coherence-Time:* In the CP-JCE approach, only two precoding matrices are required i.e., one for pilot decoupling and other for data decoupling for all the users. In contrast, the USP-DCE approach requires $2 \times U$ precoding matrices i.e., $U$ matrices for pilot decoupling and $U$ matrices for data decoupling. As a result, the CP-JCE approach benefits from a *shorter required coherence time*, making it more suitable for rapidly changing channels, whereas the longer coherence time of the USP-DCE approach can be a disadvantage in dynamic environments.

## VI. Conclusion

This study proposes affine precoded SIP models based on CP-JCE and USP-DCE for partially connected MU THz systems and analyzes the trade-offs between them in terms of complexity and estimation accuracy. A variational Bayes inference framework is developed for efficient estimation by considering the dual-wideband channel. Simulation results demonstrate the performance gains achieved by the channel estimation approach.

## References


[1] J. M. Jornet and I. F. Akyildiz, "Channel modeling and capacity analysis for electromagnetic wireless nanonetworks in the TeraHertz band," *IEEE Transactions on Wireless Communications*, 2011.

[2] B. Wang, F. Gao, S. Jin, H. Lin, and G. Y. Li, "Spatial-and frequency-wideband effects in millimeter-wave massive MIMO systems," *IEEE Transactions on Signal Processing*, pp. 3393–3406, 2018.

[3] R. Shafin and L. Liu, "Superimposed pilot for multi-cell multi-user massive FD-MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 19, no. 5, pp. 3591–3606, 2020.

[4] F. Jesbin, S. R. Mattu, and A. Chockalingam, "Sparse superimposed pilot based channel estimation in OTFS systems," in *2023 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE, 2023.

[5] X. Jing, M. Li, H. Liu, S. Li, and G. Pan, "Superimposed pilot optimization design and channel estimation for multiuser massive MIMO systems," *IEEE Transactions on Vehicular Technology*, 2018.

[6] S. Wu, Z. Ni, X. Meng, and L. Kuang, "Block expectation propagation for downlink channel estimation in massive MIMO systems," *IEEE communications letters*, vol. 20, no. 11, pp. 2225–2228, 2016.

[7] V. Cevher, "Learning with compressible priors," *Advances in Neural Information Processing Systems*, vol. 22, 2009.

[8] A. Garg, A. Kumar, S. Srivastava, N. Yadav, A. K. Jagannatham, and L. Hanzo, "Bayesian Learning Aided Simultaneous Sparse Estimation of Dual-Wideband THz Channels in Multi-User Hybrid MIMO Systems," *IEEE Transactions on Vehicular Technology*, 2025.

[9] O. A. Aghda, M. J. Omidi, and H. Saeedi-Sourck, "Superimposed channel estimation in OTFS modulation using compressive sensing," in *31st International Conference on Electrical Engineering*. IEEE, 2023.

[10] F. Talaei and X. Dong, "Hybrid mmWave MIMO-OFDM channel estimation based on the multi-band sparse structure of channel," *IEEE Transactions on Communications*, vol. 67, no. 2, pp. 1018–1030, 2018.