

Bayesian Learning Aided Simultaneous Sparse Estimation of Dual-Wideband THz Channels in Multi-User Hybrid MIMO Systems

Abhisha Garg *Graduate Student Member, IEEE*, Akash Kumar, Suraj Srivastava, *Member, IEEE*, Nimish Yadav, Aditya K. Jagannatham, *Senior Member, IEEE* and Lajos Hanzo, *Life Fellow, IEEE*

Abstract—This work conceives the Bayesian Group-Sparse Regression (BGSR) for the estimation of a *spatial and frequency wideband*, i.e., a *dual wideband* channel in Multi-User (MU) THz hybrid MIMO scenarios. We develop a practical dual wideband THz channel model that incorporates absorption losses, reflection losses, diffused ray modeling and angles of arrival/departure (AoAs/AoDs) using a Gaussian Mixture Model (GMM). Furthermore, a low-resolution analog-to-digital converter (ADC) is employed at each RF chain, which is crucial for wideband THz massive MIMO systems to reduce power consumption and hardware complexity, given the high sampling rates and large number of antennas involved. The quantized MU THz MIMO model is linearized using the popular Bussgang decomposition followed by BGSR based channel learning framework that results in sparsity across different subcarriers, where each subcarrier has its unique dictionary matrix. Next, the Bayesian Cramér Rao Bound (BCRB) is devised for bounding the normalized mean square error (NMSE) performance. Extensive simulations were performed to assess the performance improvements achieved by the proposed BGSR method compared to other sparse estimation techniques. The metrics considered for quantifying the performance improvements include the NMSE and bit error rate (BER).

Index Terms—TeraHertz, beam squint, dual-wideband, Multi-User, Bussgang decomposition, Bayesian learning, Cramér-Rao bound

The work is supported by IEEE SPS scholarship grant for 2023, 2024 and 2025. The work of Aditya K. Jagannatham was supported in part by the Qualcomm Innovation Fellowship; in part by the Qualcomm 6G UR Gift; in part by the Arun Kumar Chair Professorship; and in part by the DST, Govt. of India. Lajos Hanzo would like to acknowledge the financial support of the following Engineering and Physical Sciences Research Council (EPSRC) projects is gratefully acknowledged: Platform for Driving Ultimate Connectivity (TITAN) under Grant EP/Y037243/1 and EP/X04047X/1; Robust and Reliable Quantum Computing (RoaRQ, EP/W032635/1); PerCom (EP/X012301/1); EP/X01228X/1; EP/Y037243/1. The work of S. Srivastava was supported in part by IIT Jodhpur's Research Grant No. I/RIG/SUS/20240043; in part by Anusandhan National Research Foundation's PM-ECRG/2024/478/ENS; and in part by Telecom Technology Development Fund (TTDF) under Grant TTDF/6G/368. S. Srivastava, A. K Jagannatham, and L. Hanzo jointly acknowledge the funding support provided to ICON-project by DST and UKRI-EPSRC under India-UK Joint opportunity in Telecommunications Research.

Abhisha Garg and Aditya K. Jagannatham are with the Department of Electrical Engineering, Indian Institute of Technology Kanpur, Kanpur-208016, India (e-mail: abhisha20@iitk.ac.in; adityaj@iitk.ac.in).

Akash Kumar is with Qualcomm India Pvt. Ltd., Hyderabad, Telangana, 500081, India (email: akkum@qti.qualcomm.com)

Suraj Srivastava is with the Department of Electrical Engineering, Indian Institute of Technology Jodhpur, Jodhpur, Rajasthan 342030, India (e-mail: surajsri@iitj.ac.in).

Nimish Yadav is with Samsung Semiconductor India Research, Bengaluru, Karnataka, 560048, India (email: nimish.y@samsung.com).

L. Hanzo is with the School of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, U.K. (email:lh@ecs.soton.ac.uk)

I. INTRODUCTION

TeraHertz (THz) communication stands out as a promising avenue to deliver ultra-high data rates and massive short-range connectivity in next-generation wireless communications. The THz band, spanning from (0.1 – 10) THz, offers spectral windows characterized by bandwidths of 10 GHz or higher [1]. In comparison to the mmWave band that spans 30 – 100 GHz [2], the major challenges in this band include having fewer spatial degrees-of-freedom (SDoF) [3], high blockages and the spatial wideband effect. The short wavelength of the THz signal leads to variation in the delay across different antenna elements within the array, resulting in deleterious spatial wideband effects [4]. Furthermore, a wideband THz signal exhibits frequency-selective characteristics stemming from the multipath delay spread, consequently inducing a hostile frequency-wideband effect. A general THz system experiencing simultaneous frequency- and spatial-wideband effects is termed a dual wideband system [5]. The so-called *beam squint* effect arises because of the variation in the effective angle of arrival (AoA)/ angle of departure (AoD) across the subcarriers, which in turn affects the array response vector [4]. This phenomenon leads to a significant difference with respect to its mmWave counterpart and gives rise to a fundamental challenge in this band. However, the presence of precise channel state information (CSI) is essential for successful beamforming in such systems. Furthermore, the conventional channel estimation methods like least-squares (LS) and linear minimum mean squared error (LMMSE) techniques necessitate a substantial pilot overhead because of the dense packing of antennas, leading to reduced spectral efficiency (SE). Therefore, the development of advanced CSI estimation techniques tackling the dual-wideband effect in THz hybrid MIMO systems is paramount, while also accounting for the beam squint effect. A popular technique of recovering such a sparse channel is sparse Bayesian learning (SBL), which exhibits a distinct advantage over other sparse recovery techniques such as LASSO and FOCUSS, as it automatically exploits sparsity, without user intervention [6]. Additionally, as a probabilistic method, SBL calculates the *a posteriori* distribution of the sparse weight vectors, including its mean and covariance, which enhances the overall accuracy of the estimation process. The following section will offer a critical appraisal of the pertinent previous studies.

A. Review of existing works

The THz band is associated with significant challenges primarily due to the inevitable propagation and molecular absorption (MA) losses. Pioneering research conducted by Jornet and Akyildiz [1] has led to the development of a cutting-edge THz channel model that effectively incorporates both propagation and MA losses. The pioneering work of Lin and Li [7] proposed a distance- and frequency-dependent hybrid beamforming scheme for an indoor scenario considering multiple users (MU). They modeled the AoA/AoD using a Gaussian Mixture Model (GMM).

The problem of finding sparse solutions from a single measurement vector (SMV) has received extensive attention in prior research [8]. Despite its NP-hard nature, numerous near-optimal algorithms designed for single-measurement scenarios have been devised, and these algorithms have proven to be valuable in diverse applications [9]. The algorithms conceived for SMV can broadly be classified into greedy algorithms [10], mixed-norm optimization [11], iterative re-weighted [12] and Bayesian learning algorithms. As a result of this groundwork, numerous multiple measurement vector (MMV) algorithms have emerged as straightforward extensions of these SMV algorithms. These MMV algorithms leverage the foundational principles established in the SMV context to address more complex scenarios, owing to their efficiency and adaptability. Wipf and Rao [6] proposed the SBL algorithm for SMV and later on extended to MMV in form of the MMV-SBL (MSBL) algorithm. Within the realm of MMV algorithms, the family of Bayesian learning frameworks has gained significant popularity among researchers due to their demonstrated can be ability of achieving optimal recovery performance. Cotter *et al.* [13], in their path breaking work expanded the application of MMV techniques to encompass a range of sparse recovery algorithms, including MMV based matching pursuit (MBMP), MMV orthogonal matching pursuit (MOMP), MMV order-recursive matching pursuit (MORMP), and MFOCUSS with unknown sparsity structure. The key advantage of using MSBL over the other MMV algorithms is that the global minima of MSBL is always the sparsest solution and exhibits very few local minima compared to other MMV based algorithms [14].

The versatility and relevance of MSBL techniques are not limited to quasi-static channels; they also lend themselves to employment in the domains of online learning and data recovery processes. Joseph *et al.* [15], introduce a pair of online MSBL algorithms designed for recovering temporally correlated sparse data, using a sequential EM approach. They model the correlation using an auto-regressive (AR) process and achieve online estimation with a small delay. Srivastava *et al.* [16], proposed a MMV sparse Kalman filtering based approach for data-aided channel estimator tailored for doubly selective mmWave MIMO-OFDM systems. They model the temporal variation of the path gain using an AR-1 process.

The authors of [17] developed an SBL based TD THz channel estimation framework considering optimal pilot design. Sha and Wang [18] in their pioneering work proposed a two-stage channel estimator for frequency wideband THz channels considering RF impairments. Dovelos *et al.* [19], modeled

the OFDM-based dual-wideband THz channel of a uniform planar array (UPA) and estimated the channel using both OMP and generalized MOMP (GSOMP) based approaches. They consider a single-user (SU) configuration in a single antenna scenario. The recent literature only has a limited number of works addressing the dual wideband effect for THz channel estimation. In this work, Chou *et al.* [20], proposed a framework for dual wideband channel modeling in the sub-THz regime, utilizing MMV least squares (MMV-LS-CS) techniques for acquiring a time-varying MIMO-OFDM channel. Li and Madhukumar [21] proposed a channel-estimation framework for hybrid (near- and far)-field THz UM-MIMO that integrates dictionary learning with Bayesian sparse recovery. A sparsifying dictionary is learned to capture the joint hybrid-field channel structure, and Bayesian inference is then applied to recover the sparse coefficients, yielding improved estimation performance under constrained pilot resources relative to fixed-basis approaches.

Wang *et al.* [4], in their groundbreaking treatise suggest a new massive MIMO channel model that considers both the spatial-wideband and frequency-wideband effects. However, they left the problem of dual-wideband reception under low-resolution ADCs as an open issue. Zhang *et al.* [22], compare the performance of finite-resolution ADCs and DACs in narrow-band MU THz hybrid MIMO systems. They derived a closed-form expression for the achievable lower bound. Nikbakht and Lozano [23] proposed an unsupervised learning-based technique for transmit beamforming in THz systems considering low-resolution ADCs. However, none of the existing literature investigating low-resolution ADCs considers the dual-wideband channel, which is predominant in the THz band. The authors in their seminal work [24] proposed a data-driven de-quantizer to combat hardware imperfections in OFDM systems. They utilized a data-driven approach to mitigate signal distortion caused by quantization and additive white Gaussian noise (AWGN), whereas compressive sensing techniques do not account for system properties. Kim and Choi in their pioneering work [25] proposed a channel estimation scheme considering low-resolution ADCs in the face of spatial-wideband effect only. However, they consider a mmWave massive MIMO system, where each user is equipped with a single antenna. Furthermore, they employ a time-domain system that does not account for frequency selectivity, a common consideration for wideband THz systems.

Moreover, in contrast to [5], which primarily focuses on angular-domain processing, this work addresses multiple key challenges inherent in THz communication. Specifically, we incorporate off-grid-based dictionary estimation to enhance the angular resolution, account for low-resolution ADCs which are the dominant sources of power consumption and extension of the framework to construct the MU dual-wideband channel formulation. Additionally, we integrate the GMM based approach for AoA/AoD generation to better capture spatial variations. These enhancements collectively provide a more comprehensive and practical framework for robust channel estimation and system design in THz-band communications. To address this research gap, we propose a novel CSI estimation framework that considers a dual wideband channel

Table I
COMPARISON OF THE NOTABLE CONTRIBUTIONS OF THIS STUDY WITH THE ALREADY EXISTING WORKS

Features	[7]	[5]	[16]	[17]	[19]	[20]	[21]	[25]	[26]	[18]	This Paper
sub-THz/ THz Band	✓	✓		✓	✓	✓				✓	✓
Reflection losses & molecular losses	✓	✓		✓	✓	✓	✓				✓
SC-FDE system		✓								✓	✓
Dual-Wideband Effect		✓			✓	✓		✓			✓
Multi-User MIMO	✓					✓		✓	✓		✓
Low Resolution ADCs						✓	✓	✓	✓		✓
BCRB		✓	✓	✓	✓	✓					✓
Taylor-based off-grid estimation											✓
MU AoA/AoD with GMM											✓
RRC-PSF based dual-wideband channel											✓

for MU scenarios using an SC-FDE system. It is important to note that, for low-resolution ADCs, single-carrier systems are generally preferred, since OFDM-based systems cannot preserve orthogonality [27]. Table-I boldly contrasts the salient contributions of the proposed work to the literature at a glance.

B. Contributions

- 1) We commence by formulating a practical THz channel model that incorporates both frequency- and spatial-wideband effects, while also accounting for their distance-dependent characteristics. Furthermore, we consider the effects of diffused rays associated with each multipath component, a practical challenge that has been overlooked in previous contributions [1], [7], [20]. Additionally, we compare the root raised cosine pulse shaping filter (RRC-PSF) and the rectangular pulse shaping filter (Rect-PSF) in dual-wideband channel formulations. A key challenge is the generation of distinct and realistic spatial signatures for AoA/AoD pairs using the GMM corresponding to each user, which has been overlooked in [7], [28] and addressed in this work.
- 2) The approaches in [17] and [29] assume a common dictionary across all subcarriers, which is not feasible due to the beam squint effect altering the traditional array manifold. To address this, we leverage the sparsity in the angular domain to formulate a compressive sensing (CS) problem and solve it using the Bayesian Group Sparse Regression (BGSR) approach. Additionally, we construct both the on-grid and Taylor-based off-grid (TBoD) dictionary matrices and demonstrate that the channel exhibits a shared support across different subcarriers, unlike the case in [5].
- 3) We quantify the accuracy of the proposed BGSR based technique by comparing the estimation error variance to the Bayesian Cramér-Rao Bound (BCRB). Therefore, another significant contribution of this work involves deriving the BCRB for MU scenarios. It provides a valuable benchmark for assessing the performance of the CSI frameworks.
- 4) Another key contribution of this work involves the combined study of the grave challenges occurring in the THz band, including the hybrid architecture, low-resolution ADCs, and dual-wideband channels, to enhance the end-to-end system performance. To the best of our knowledge, no existing work considers an SC-FDE-based system utilizing low-resolution ADCs in

dual-wideband channels. This is because single-carrier systems are inherently more robust to low-resolution ADCs than OFDM systems. Therefore, integrating these elements within a unified SC-FDE framework further exacerbates their interdependencies, which we aim to address in this work.

- 5) Our simulation results conclusively illustrate the improved performance of the proposed BGSR algorithm, also verifying that the normalized mean squared error (NMSE) of the proposed algorithm approaches the BCRB, demonstrating its effectiveness. Furthermore, the MA coefficient is evaluated using the High Resolution Transmission (HITRAN) database [30], which exhibits reliability throughout the entire THz band.

C. Notation:

Matrices are represented by uppercase letters \mathbf{B} , while lowercase letters \mathbf{b} represent vectors. Various operators are indicated by superscripts such as $(\cdot)^T$, $(\cdot)^H$, $(\cdot)^{-1}$ which represent the transpose, Hermitian and inverse, respectively. Let $\{\mathbf{H}_u(0), \mathbf{H}_u(1), \dots, \mathbf{H}_u(N-1)\}$ represent a sequence of matrices and $\{\mathbf{x}_u(0), \mathbf{x}_u(1), \dots, \mathbf{x}_u(N-1)\}$ represent a sequence of vectors. The circular convolution $\{\mathbf{r}_u(n)\}_{n=0}^{N-1}$ can be defined as

$$\mathbf{r}_u(n) = \sum_{l=0}^{N-1} \mathbf{H}_u(l) \mathbf{x}_u[(n-l)]_N + \check{\mathbf{v}}_u(n),$$

where $[\cdot]_N$ represents modulo-N operation. Let $\{\mathbf{x}_u(0), \mathbf{x}_u(1), \dots, \mathbf{x}_u(N-1)\}$ represent the input sequence, where $\mathbf{x}(n, k)$ denotes the k th element of $\mathbf{x}(n)$. Let $\{\mathbf{r}_u(0), \mathbf{r}_u(1), \dots, \mathbf{r}_u(N-1)\}$ constitute the output sequence, where $\mathbf{r}(n, k)$ denotes the k th element of $\mathbf{r}(n)$. Then, the N -point FFT of the vector sequence is defined as $\mathbf{r}(p, q) = \sum_{n=0}^{N-1} \mathbf{x}(n, q) e^{-j \frac{2\pi np}{N}}$. The uniform distribution is denoted by $\mathcal{U}(\cdot)$ while the Gaussian distribution is denoted with $\mathcal{CN}(\mu, \Sigma)$ where μ represents mean while Σ represents variance.

II. MU THz HYBRID MIMO SYSTEM AND CHANNEL ESTIMATION MODELS

Consider an MU THz hybrid MIMO system as shown in Fig. 1(a), where a BS is fitted with N_R receive antennas and serves U users simultaneously, with each user possessing $N_{T,u}$ transmit antennas. Therefore, the total of all the transmit antennas for all the users equals $\sum_{u=1}^U N_{T,u} = U N_{T,u} = N_T$.

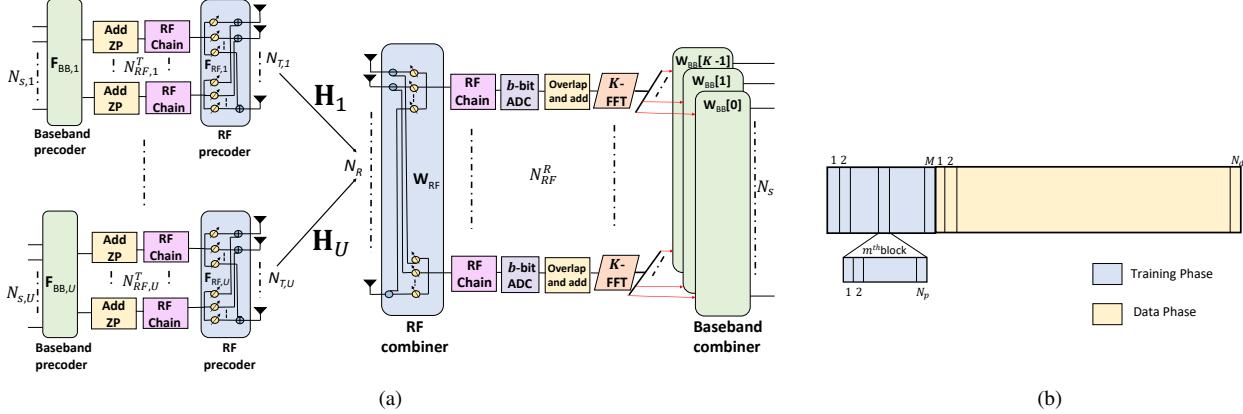


Figure 1. (a) Schematic diagram of SC-FDE based MU THz hybrid MIMO systems with low-resolution ADCs. (b) Frame structure utilized for wideband MU THz hybrid MIMO system using SC-FDE corresponding to each user.

The BS has \$N_{RF}^R\$ RF chains, while each user possess \$N_{RF,u}^T\$ RF chains. Furthermore, we consider a practical scenario in which the RF chain count follows the constraint \$N_{RF}^R \ll N_R\$ and \$N_{RF,u}^T < N_{T,u}\$ in order to support the hybrid MIMO architecture. The aggregate number of data streams, denoted as \$N_s\$, is subject to the constraint \$N_s \leq N_{RF}^R\$, where \$N_s\$ is defined as \$N_s = \sum_{u=1}^U N_{s,u} = U N_{s,u}\$, and \$N_{s,u}\$ represents the number of data streams corresponding to each user. Moreover, at the receiver, the hybrid processing module consists of a set of frequency-selective baseband receiver combiners (RC) \$\mathbf{W}_{BB}[k] \in \mathbb{C}^{N_{RF}^R \times N_s}\$ cascaded with a frequency-flat RF RC \$\mathbf{W}_{RF} \in \mathbb{C}^{N_R \times N_{RF}^R}\$. At the transmitter, the hybrid processing module is comprised of a frequency-flat baseband transmit precoder (TPC) \$\mathbf{F}_{BB,u} \in \mathbb{C}^{N_{RF,u}^T \times N_{s,u}}\$ for each user coupled with a frequency-flat RF TPC \$\mathbf{F}_{RF,u} \in \mathbb{C}^{N_{T,u} \times N_{RF,u}^T}\$. Furthermore, the RF TPC and RC are composed of digitally controlled networks of phase-shifters, which consist of only constant-amplitude phase elements following the constraints \$|\mathbf{W}_{RF}(\kappa, \ell)| = \frac{1}{\sqrt{N_R}}\$, \$|\mathbf{F}_{RF,u}(\kappa, \ell)| = \frac{1}{\sqrt{N_{T,u}}}\$.

The frame structure of each user is shown in Fig. 1(b). Let \$\mathbf{H}_{l,u} \in \mathbb{C}^{N_R \times N_{T,u}}\$ represent the complex wideband THz MIMO channel corresponding to the \$l\$-th delay tap of \$u\$-th user. Furthermore, let \$\mathbf{a}_{m,u}^{(p)} \in \mathbb{C}^{N_{RF,u}^T \times 1}\$ represent the \$p\$-th complex pilot vector of the \$m\$-th block corresponding to the \$u\$-th user. Prior to transmission through individual RF chains, the \$N_p\$ pilot vectors undergo zero-padding (ZP), where \$L-1\$ zeros are appended to each block. This process results in ZP blocks of length \$K\$, which is further given by \$\{\mathbf{a}_{m,u}^{(q)}\}_{q=0}^{K-1} = \{\mathbf{a}_{m,u}^{(0)}, \mathbf{a}_{m,u}^{(1)}, \dots, \mathbf{a}_{m,u}^{(N_p-1)}, \underbrace{\mathbf{0}, \dots, \mathbf{0}}_{L-1}\}\$.

To achieve consistent dimensions, \$N_p-1\$ zero-matrices of size \$N_R \times N_{T,u}\$ are appended to the channel taps \$\mathbf{H}_{l,u}\$, and the resultant taps are given by \$\{\mathbf{H}_{q,u}\}_{q=0}^{K-1} = \{\mathbf{H}_{0,u}, \mathbf{H}_{1,u}, \dots, \mathbf{H}_{L-1,u}, \underbrace{\mathbf{0}, \dots, \mathbf{0}}_{N_p-1}\}\$. Therefore, the received signal vector \$\mathbf{r}_m(q) \in \mathbb{C}^{N_{BS} \times 1}\$ corresponding to all the \$U\$ users is given by

$$\mathbf{r}_m(q) = \sum_{u=1}^U \mathbf{H}_{q,u} \otimes_K (\mathbf{F}_{RF,m,u} \mathbf{a}_{m,u}^{(q)}) + \tilde{\mathbf{v}}_m(q), \quad (1)$$

where \$\tilde{\mathbf{v}}_m(q)\$ represents the AWGN, which obeys

\$\mathcal{CN}(\mathbf{0}_{N_{BS} \times 1}, \sigma^2 \mathbf{I}_{N_{BS}})\$, while \$\otimes_K\$ denotes the circular convolution of length \$K\$. The quantity \$\mathbf{F}_{RF,m,u}\$ represents the RF TPC for the \$u\$-th user corresponding to the \$m\$-th block. Furthermore, the ZP operation converts the linear convolution to circular convolution and after applying the fast Fourier transform (FFT) at the receiver, the system is decoupled into different frequency bins. At this juncture it is also worth noting that while the overall system model is decomposed into different frequency bins at the receiver, it can be still viewed as an SC system at the transmitter. This is because, the transmitter does not employ an inverse FFT (IFFT) block and the signals \$\mathbf{a}_{m,u}^{(q)}\$ are transmitted sequentially with the aid of ZP. Therefore, the received signal \$\mathbf{y}_{m,MU}(q) \in \mathbb{C}^{N_{RF}^R \times 1}\$ for all the \$U\$ users after performing RF combining corresponding to the \$m\$-th block and passing through low-resolution ADCs \$\mathcal{Q}(\cdot)\$ can be formulated as

$$\begin{aligned} \mathbf{y}_{m,MU}(q) &= \mathcal{Q}\left(\mathbf{W}_{RF,m}^H \sum_{u=1}^U \mathbf{H}_{q,u} \otimes (\mathbf{F}_{RF,m,u} \mathbf{a}_{m,u}^{(q)}) + \mathbf{W}_{RF,m}^H \tilde{\mathbf{v}}_m(q)\right), \\ &\approx \mathbf{D} \mathbf{W}_{RF,m}^H \sum_{u=1}^U \mathbf{H}_{q,u} \otimes_K (\mathbf{F}_{RF,m,u} \mathbf{a}_{m,u}^{(q)}) + \mathbf{D} \mathbf{W}_{RF,m}^H \tilde{\mathbf{v}}_m(q) + \check{\mathbf{v}}. \end{aligned} \quad (2)$$

Assuming each RF chain to be quantized independently, the quantization matrix \$\mathbf{D} \in \mathbb{C}^{N_{RF}^R \times N_{RF}^R}\$ may be considered to be diagonal. Moreover, the entries of \$\mathbf{D}\$ are represented as \$\mathbf{D} = \varepsilon \mathbf{I}\$, where \$\varepsilon = 1 - v\$ and \$v\$ represents the quantization noise-to-signal power [31]. Let \$\mathbb{E}\{\mathbf{a}_{m,u}^{(q)}(\mathbf{a}_{m,u}^{(q)})^H\} = \sigma_b^2 \mathbf{I}_{N_{s,u}}\$ where \$\sigma_b^2\$ represents the pilot power. Additionally, the quantization vector \$\check{\mathbf{v}}\$ follows the distribution \$\mathcal{CN}(\mathbf{0}_{N_{RF}^R \times 1}, \mathbf{C}_m)\$, where \$\mathbf{C}_m = \varepsilon(1 - \varepsilon) \text{diag}(\mathbf{W}_{RF,m}^H \mathbf{Q}_m \mathbf{W}_{RF,m} + \sigma_n^2 \mathbf{W}_{RF,m}^H \mathbf{W}_{RF,m})\$. The derivation of this along with that for matrix \$\mathbf{Q}_m\$ is detailed in Appendix-A. Moreover, the diagonal error covariance matrix results from quantizing each RF chain independently, which is followed from the property of Bussgang decomposition for Gaussian inputs [32]. Furthermore, the values of \$v\$ are given in Table II for \$b < 5\$ and can be approximated as \$v = \frac{\pi\sqrt{3}}{2} 2^{-2b}\$ for \$b > 5\$ [31]. Let \$\mathbf{v}_m(q) = \mathbf{D} \mathbf{W}_{RF,m}^H \tilde{\mathbf{v}}_m(q) + \check{\mathbf{v}} \in \mathbb{C}^{N_{RF}^R \times 1}\$ represent the equivalent noise. Therefore, the equivalent noise covariance matrix of \$\mathbf{R}_{vv} = \mathbb{E}\{\mathbf{v}_m(q)\mathbf{v}_m^H(q)\} \in \mathbb{C}^{N_{RF}^R \times N_{RF}^R}\$ can be expressed as \$\mathbf{R}_{vv} = \varepsilon^2 \sigma_n^2 \mathbf{W}_{RF,m}^H \mathbf{W}_{RF,m} + \mathbf{C}_m\$. Therefore, the received signal \$\mathbf{y}_{m,MU}[k] \in \mathbb{C}^{N_{RF}^R \times 1}\$ corresponding

Table II
 v FOR DIFFERENT ADC BITS b [31]

b	1	2	3	4	5
v	0.3634	0.1175	0.03454	0.009497	0.002499

to the k -th subcarrier for all the users, obtained via K -point FFT $\{\mathbf{y}_{m,\text{MU}}[k]\}_{k=0}^{K-1} = \text{FFT}(\{\mathbf{y}_{m,\text{MU}}[q]\}_{q=0}^{K-1})$ is given by

$$\mathbf{y}_{m,\text{MU}}[k] \approx \mathbf{DW}_{\text{RF},m}^H \mathbf{H}_{\text{MU}}[k] \mathbf{F}_{\text{RF},m,u} \mathbf{a}_{m,u}[k] + \mathbf{v}_m[k], \quad (3)$$

where $\mathbf{H}_{\text{MU}}[k] = [\mathbf{H}_1[k] \ \mathbf{H}_2[k] \ \cdots \ \mathbf{H}_U[k]] \in \mathbb{C}^{N_R \times N_T}$ represents the concatenated channel for all the users while $\mathbf{H}_u[k] \in \mathbb{C}^{N_R \times N_{T,u}}$ represents the K -point FFT of $\{\mathbf{H}_{l,u}\}_{l=0}^{L-1}$. The quantity $\mathbf{a}_{m,u}[k] \in \mathbb{C}^{N_{RF,u} \times 1}$ is obtained by taking the K -point FFT of the input signal $\{\mathbf{a}_{m,u}^{(q)}\}$ whereas $\mathbf{v}_m[k] \in \mathbb{C}^{N_{RF} \times 1}$ results from applying the same transformation to the effective noise signal $\mathbf{v}_m(q)$. Let $\mathbf{s}_{m,\text{MU}}[k] = \mathbf{F}_{\text{RF},m,u} \mathbf{a}_{m,u}[k] \in \mathbb{C}^{N_{T,u} \times 1}$ for notational simplicity. Upon employing the $\text{vec}(\cdot)$ operator property in Eq. (3), we obtain

$$\mathbf{y}_{m,\text{MU}}[k] = \underbrace{(\mathbf{s}_{m,\text{MU}}^T[k] \otimes \mathbf{DW}_{\text{RF},m}^H)}_{\mathbf{A}_{m,\text{MU}}[k]} \underbrace{\text{vec}(\mathbf{H}_{\text{MU}}[k])}_{\mathbf{h}_{\text{MU}}[k]} + \mathbf{v}_m[k].$$

Note that, although the Bussgang approximation is sub-optimal under low-resolution ADCs, it remains a practical choice in THz systems [3]. Due to the extremely high bandwidth and large antenna arrays, exact nonlinear quantization models become computationally expensive and scale poorly with system dimensions. However, Bussgang linearization provides a tractable *trade-off* sacrificing some modeling accuracy at higher resolutions, while enabling large-scale multi-subcarrier, multi-antenna channel estimation in THz systems, where beam-squint and high dimensionality already render exact quantization approaches intractable. The next section will discuss the dual-wideband effected channel model.

III. MU THZ CHANNEL MODELING WITH DUAL-WIDEBAND EFFECTS

This section delves into the characteristics of the dual-wideband THz MIMO channel for a generalized pulse shaping filter (PSF) and further describes the channel model for a THz hybrid MIMO system. Table-III provides the description of the channel parameters considered. The generalized array steering vector $\mathbf{a}(\phi, f_c)$ can be formulated as

$$\mathbf{a}(\phi, f_c) = \frac{1}{\sqrt{N}} \left[1, e^{-j \frac{2\pi}{\lambda} d_c \cos \phi}, \dots, e^{-j(N-1) \frac{2\pi}{\lambda} d_c \cos \phi} \right]^T, \quad (4)$$

where $\lambda = \frac{c}{f}$ while $d_c = \frac{c}{2f_c}$. It is worth noting that the frequently employed approximation of $d_c \approx \frac{\lambda}{2}$ used for antenna arrays holds true only for narrowband channels, where the B is much smaller than the center frequency f_c , i.e., $B \ll f_c$. Naturally, it breaks down for wideband signals occupying a significant frequency spectrum [33]. Let f_k represent the frequency of the k th subcarrier, which is defined

$$f_k \triangleq f_c + \left(k - \frac{K+1}{2} \right) \frac{B}{K}. \quad (5)$$

Furthermore, we define the spatial phase offset $\Delta\vartheta_{n,k}$ for the n -th antenna at the k -th subcarrier as

$$\Delta\vartheta_{n,k} = \frac{2\pi}{\lambda} (n-1) d_c \cos \phi = \pi (n-1) \varrho_k \cos \phi, \quad (6)$$

where $\varrho_k = \frac{f_k}{f_c}$. Consequently, the effective spatial AoA $\widehat{\phi}(f_k)$ at the k -th subcarrier is defined as

$$\widehat{\phi}(f_k) = \arccos(\varrho_k \cos \phi). \quad (7)$$

Thus, the modified array steering vector associated with frequency-dependent angles from Equations (4) and (7), can be further expressed as

$$\tilde{\mathbf{a}}(\phi, f_k) = \frac{1}{\sqrt{N}} \left[1, e^{-j\pi\varrho_k \cos(\phi)}, \dots, e^{-j(N-1)\pi\varrho_k \cos(\phi)} \right]^T. \quad (8)$$

Moreover, the THz channel can be characterized as a combination of line-of-sight (LoS) and non-line-of-sight (NLoS) components. The LoS component captures the direct propagation path between the BS and UE, while the NLoS component encompasses the multiple reflections caused by various scatters within the environment. Therefore, the dual-wideband THz MIMO channel can be mathematically expressed as the merger of the LoS and NLoS components, which can be described for the subcarrier k as

$$\mathbf{H}_u[k] = \mathbf{H}_u^{\text{LoS}}[k] + \mathbf{H}_u^{\text{NLoS}}[k]. \quad (9)$$

The LoS and NLoS components for the dual-wideband channel are given by

$$\mathbf{H}_u^{\text{LoS}}[k] = \sqrt{N_{T,u} N_R} \alpha(f_k, d) \beta_\tau \mathfrak{B}_{T,u} \mathfrak{B}_R \tilde{\mathbf{a}}_R(\phi, f_k) \tilde{\mathbf{a}}_{T,u}^H(\theta, f_k), \quad (10)$$

$$\mathbf{H}_u^{\text{NLoS}}[k] = \sqrt{\frac{N_{T,u} N_R}{N_{\text{NLoS}} N_{\text{ray}}}} \sum_{z=1}^{N_{\text{NLoS}}} \sum_{j=1}^{N_{\text{ray}}} \alpha_{z,j}(f_k, d_{z,j}) \beta_{\tau_{z,j}} \mathfrak{B}_{T,u} \mathfrak{B}_R \tilde{\mathbf{a}}_R(\phi_{z,j}, f_k) \tilde{\mathbf{a}}_{T,u}^H(\theta_{z,j}, f_k), \quad (11)$$

$$\beta_{\tau_{z,j}} = \sum_{l=0}^{K-1} p(lT_s - \tau_{z,j}) e^{-j \frac{2\pi k l}{K}}, \forall k, l, \quad (12)$$

where the parameters of $\mathbf{H}_u^{(\cdot)}[k]$ are defined in Table-III. Moreover, THz wireless channels are susceptible to substantial attenuation due to various factors. Among these are absorption losses arising from molecular interactions and noise introduced by water vapor molecules [1]. These considerations are integrated into the modeling of the complex-path gains $\alpha(f_k, d)$ and $\alpha_{z,j}(f_k, d_{z,j})$, which depend on both the subcarrier frequency f_k and transmission distance $d_{(\cdot)}$. Moreover, the intricate path gain $\alpha_{(\cdot)}(f_k, d_{(\cdot)})$, can be described by its magnitude and phase components as $\alpha_{(\cdot)}(f_k, d_{(\cdot)}) = |\alpha_{(\cdot)}(f_k, d_{(\cdot)})| e^{j\varpi}$. Thus, the modeling of the complex path gain for the LoS component as discussed in [1] is given by

$$|\alpha(f_k, d)|^2 = L_{\text{free}}(f_k, d) L_{\text{abs}}(f_k, d), \quad (13)$$

where $L_{\text{free}}(f_k, d)$ and $L_{\text{abs}}(f_k, d)$ [20] are respectively given as

$$L_{\text{free}}(f_k, d) = \left(\frac{c}{4\pi f_k d} \right)^2, \quad L_{\text{abs}} = e^{-k_{\text{abs}}(f_k)d}, \quad (14)$$

while $k_{\text{abs}}(f_k) = \sum_{p,q} k_{\text{abs}}^{p,q}(f_k)$. The quantity $k_{\text{abs}}^{p,q}(f_k)$ represents the attenuation of radiation at the subcarrier frequency f_k due to the absorption by the p -th isotopologue of the q -th gas and can be efficiently calculated from the HITRAN database [1]. In a similar vein, the complex path gain associated with the

Table III
NOTATION AND DESCRIPTION OF CHANNEL PARAMETERS CONSIDERED

Parameter	Description	Parameter	Description
λ	wavelength	ϖ	phase of complex path gain
f_k	subcarrier frequency	K_{abs}	molecular-absorption loss
$\Delta\vartheta$	spatial phase offset	Υ	first-order reflection coefficient
d_c	inter-antenna spacing	ϱ_k	relative subcarrier frequency
N_{ray}	# of diffused rays	μ_0/ϵ_0	free-space permeability/ permittivity
$\phi(\cdot)$	AoA	\varkappa	absorption coefficient of reflecting medium
$\theta(\cdot)$	AoD	L_{free}	free-space loss
L_{abs}	absorption loss	Z_0	intrinsic impedance of free space
$\alpha(\cdot)$	complex path gain	ν_i/ν_r	angle of incidence/ refraction of medium
T_s	sampling time	Z	characteristic impedance
$\tau(\cdot)$	delay	η	index of refraction
$p(\cdot)$	pulse-shaping filter	σ_r	standard roughness deviation of reflecting medium

Note that, the notation (\cdot) represents a generalized index placeholder corresponding to specific parameter, depending on the context.

z -th NLoS component corresponding to the j -th diffuse ray can be mathematically modelled as

$$|\alpha_{z,j}(f_k, d_{z,j})|^2 = \Upsilon_{z,j}^2(f_k) L_{\text{free}}(f_k, d_{z,j}) L_{\text{abs}}(f_k, d_{z,j}). \quad (15)$$

Here, $\Upsilon_{z,j}^2$ signifies the first-order reflection coefficient pertaining to the j -th diffuse ray within the z -th NLoS cluster. The quantity $\Upsilon_{z,j}^2$ is the multiplication of the Fresnel reflection coefficient and Rayleigh roughness factor [34], as

$$\Upsilon_{z,j}^2 = \frac{Z(f_k) \cos(\nu_{i_{z,j}}) - Z_0(\cos \nu_{r_{z,j}})}{Z(f_k) \cos(\nu_{i_{z,j}}) + Z_0(\cos \nu_{r_{z,j}})} e^{-\frac{1}{2} \left(\frac{4\pi f_k \sigma_r \cos(\nu_{i_{z,j}})}{c} \right)^2}$$

where $\nu_{r_{z,j}} = \arcsin \left(\sin(\nu_{i_{z,j}}) \frac{Z(f_k)}{Z_0} \right)$. The characteristic impedance $Z(f_k)$ of the reflecting medium given by $Z(f_k) = \sqrt{\frac{\mu_0}{\epsilon_0 (\eta^2 - (\frac{\nu_r c}{4\pi f_k})^2) - j \frac{2\eta \nu_r c}{4\pi f_k}}}$ where the specific parameter values

are detailed in Table-VII. It is worth noting that the highly directional nature of propagation in the THz band results in a sparsely populated multipath channel in the AoA/AoD domain. In such scenarios, the recently developed sparse signal processing paradigm can yield excellent results for channel estimation, which leads to improved bandwidth efficiency. The next section introduces the sparse THz recovery model.

IV. SPARSE CHANNEL ESTIMATION FOR MU THZ HYBRID MIMO SYSTEMS

Let $G_{T,u}$ and G_R represent the number of transmit and receive angular bins, satisfying the relationship $(G_{T,u}, G_R) \geq \max(N_{T,u}, N_R)$. Furthermore, let $\Theta_{T,u}$ and Φ_R represent the transmit and receive angular grids, which are formed by the directional cosines in the range $[-1, 1]$ and are given by

$$\begin{aligned} \Phi_R &= \{ \phi_r : \cos(\phi_r) = \frac{2}{G_R} (r-1) - 1, 1 \leq r \leq G_R \}, \\ \Theta_{T,u} &= \{ \theta_{t,u} : \cos(\theta_{t,u}) = \frac{2}{G_{T,u}} (t-1) - 1, 1 \leq t \leq G_{T,u} \}. \end{aligned} \quad (16)$$

The extended virtual channel model [35] for the u -th user is given by

$$\mathbf{H}_u[k] = \mathbf{A}_R[k] \mathbf{H}_{b,u}[k] \mathbf{A}_{T,u}^H[k], \quad (17)$$

where $\mathbf{A}_R(\Phi_R, f_k) \in \mathbb{C}^{N_R \times G_R}$ represents the receive array manifold which is given by

$$\mathbf{A}_R(\Phi_R, f_k) = [\tilde{\mathbf{a}}_R(\phi_1, f_k), \tilde{\mathbf{a}}_R(\phi_2, f_k), \dots, \tilde{\mathbf{a}}_R(\phi_{G_R}, f_k)]. \quad (18)$$

The quantity $\mathbf{A}_{T,u}(\Theta_{T,u}, f_k) \in \mathbb{C}^{N_{T,u} \times G_{T,u}}$ denotes the transmit array manifold vector for the u -th user, which is given

$$\mathbf{A}_{T,u}(\Theta_{T,u}, f_k) = [\tilde{\mathbf{a}}_{T,u}(\theta_{1,u}, f_k), \tilde{\mathbf{a}}_{T,u}(\theta_{2,u}, f_k), \dots, \tilde{\mathbf{a}}_{T,u}(\theta_{G_{T,u}}, f_k)]. \quad (19)$$

Furthermore, $\mathbf{H}_{b,u}[k] \in \mathbb{C}^{G_R \times G_{T,u}}$ represents the beamspace channel frequency response (CFR) matrix corresponding to $\mathbf{H}_u[k]$. Leveraging the $\text{vec}(\cdot)$ operator and the \otimes relationship, the THz MIMO channel corresponding to the u -th user can be recast as

$$\text{vec}(\mathbf{H}_u[k]) = \underbrace{(\mathbf{A}_{T,u}^*(\Theta_{T,u}, f_k) \otimes \mathbf{A}_R(\Phi_R, f_k))}_{\tilde{\Psi}_u[k]} \text{vec}(\mathbf{H}_{b,u}[k]),$$

where $\tilde{\Psi}_u[k] \in \mathbb{C}^{N_R N_{T,u} \times G_R G_{T,u}}$ represents the *sparsifying-dictionary* of the u -th user. Furthermore, with $\text{vec}(\mathbf{H}_{b,u}[k])$ denoting $\mathbf{h}_{b,u}[k]$, the concatenated vectorized CFR across all the U users for the k -th subcarrier is given by $\mathbf{h}_{\text{MU}}[k] = \text{blkdiag}(\underbrace{\tilde{\Psi}_1[k] \tilde{\Psi}_2[k] \dots \tilde{\Psi}_U[k]}_{\mathbf{\Psi}_{\text{MU}}[k]} \underbrace{[\mathbf{h}_{b,1}^T[k] \mathbf{h}_{b,2}^T[k] \dots \mathbf{h}_{b,U}^T[k]]^T}_{\mathbf{h}_{b,\text{MU}}[k]})$,

where $\mathbf{\Psi}_{\text{MU}}[k] \in \mathbb{C}^{U N_R N_{T,u} \times G_R \sum_{u=1}^U G_{T,u}}$ represents the concatenated sparsifying dictionary and $\mathbf{h}_{b,\text{MU}}[k] \in \mathbb{C}^{G_R \sum_{u=1}^U G_{T,u} \times 1}$ is the concatenated beamspace output for all the U users. Consider a simplistic scenario associated with $\sum_{u=1}^U G_{T,u} = U G_{T,u}$. The beamspace received output vector for all the U users corresponding to the m -th block is given by

$$\mathbf{y}_{m,\text{MU}}[k] = \underbrace{\Lambda_{m,\text{MU}}[k] \mathbf{\Psi}_{\text{MU}}[k]}_{\Omega_{m,\text{MU}}[k]} \mathbf{h}_{b,\text{MU}}[k] + \mathbf{v}_m[k]. \quad (20)$$

Furthermore, to develop a compatible pilot model [29] for the k -th subcarrier, we concatenate the outputs $\mathbf{y}_{m,\text{MU}}[k]$ for all the M blocks into a single vector, represented as

$$\begin{bmatrix} \mathbf{y}_{1,\text{MU}}[k] \\ \mathbf{y}_{2,\text{MU}}[k] \\ \vdots \\ \mathbf{y}_{M,\text{MU}}[k] \end{bmatrix} = \begin{bmatrix} \mathbf{\Omega}_{1,\text{MU}}[k] \\ \mathbf{\Omega}_{2,\text{MU}}[k] \\ \vdots \\ \mathbf{\Omega}_{M,\text{MU}}[k] \end{bmatrix} \mathbf{h}_{b,\text{MU}}[k] + \begin{bmatrix} \mathbf{v}_1[k] \\ \mathbf{v}_2[k] \\ \vdots \\ \mathbf{v}_M[k] \end{bmatrix}, \quad (21)$$

where $\mathbf{y}_{\text{MU}}[k] \in \mathbb{C}^{MN_{RF}^R \times 1}$ denotes the stacked pilot output, $\boldsymbol{\Omega}_{\text{MU}}[k] \in \mathbb{C}^{MN_{RF}^R \times UG_R G_{T,u}}$ is the equivalent sensing matrix and $\mathbf{v}_{\text{MU}}[k] \in \mathbb{C}^{MN_{RF}^R \times 1}$ represents the equivalent noise across all the M blocks for the k -th subcarrier. We assume that the noise samples are uncorrelated across all the subcarriers and satisfy $\mathbb{E}\{\mathbf{v}_{\text{MU}}[k](\mathbf{v}_{\text{MU}}[k])^H\} = \text{blkdiag}(\mathbf{R}_{vv,1}, \mathbf{R}_{vv,2}, \dots, \mathbf{R}_{vv,M}) = \mathbf{C}_w$. The next section introduces a BGSR based algorithm for the sparse channel estimation model of (21).

V. BAYESIAN GROUP-SPARSE REGRESSION ALGORITHM FOR SPARSE CHANNEL ESTIMATION IN MU THz MIMO SYSTEMS

As discussed in Section-I, practical scenarios in the THz domain necessitate careful consideration and modeling of the beam squint effect, resulting from the variation in the effective AoA/AoDs across subcarriers. This variation further impacts the array response dictionaries [19]. Consequently, conventional MSBL algorithms, which assume a common dictionary matrix across all the subcarriers [16], do not lead to optimal performance in such scenarios. Therefore, we propose a novel BGSR based approach that jointly processes the hyperparameters across all the subcarriers and provides enhanced performance over the conventional sparse sensing techniques. Furthermore, in order to construct a group based MU sparse channel estimation model, we concatenate the output across all the subcarriers from Equation (21) to obtain $\mathbf{Y}_{\text{MU}} = [\mathbf{y}_{\text{MU}}[0] \mathbf{y}_{\text{MU}}[1] \dots \mathbf{y}_{\text{MU}}[K-1]] \in \mathbb{C}^{MN_{RF}^R \times K}$. Furthermore, the equivalent sensing matrix across all subcarriers can be given by $\boldsymbol{\Xi}_{\text{MU}}(:, :, k) = \boldsymbol{\Omega}_{\text{MU}}[k] \in \mathbb{C}^{MN_{RF}^R \times UG_R G_{T,u} \times K} \forall 0 \leq k \leq K-1$, and the corresponding noise matrix can be concatenated as $\mathbf{V}_{\text{MU}} = [\mathbf{v}_{\text{MU}}[0] \mathbf{v}_{\text{MU}}[1] \dots \mathbf{v}_{\text{MU}}[K-1]] \in \mathbb{C}^{MN_{RF}^R \times K}$. Let $\mathbf{H}_{b,\text{MU}} = [\mathbf{h}_{b,\text{MU}}[0] \mathbf{h}_{b,\text{MU}}[1] \dots \mathbf{h}_{b,\text{MU}}[K-1]] \in \mathbb{C}^{UG_R G_{T,u} \times K}$ represent the aggregate channel matrix across all the subcarriers, while $\mathbf{H}_{b,u} = [\mathbf{H}_{b,u}[0] \mathbf{H}_{b,u}[1] \dots \mathbf{H}_{b,u}[K-1]] \in \mathbb{C}^{G_R \times K G_{T,u}}$ represent the combined channel matrix across all the subcarriers of the u -th user. Moreover, we assign a parameterized Gaussian prior $f(\mathbf{H}_{b,\text{MU}}; \boldsymbol{\Gamma}_{\text{MU}})$ to the beamspace channel matrix $\mathbf{H}_{b,\text{MU}}$, which is given by

$$f(\mathbf{H}_{b,\text{MU}}; \boldsymbol{\Gamma}_{\text{MU}}) = \prod_{k=0}^{K-1} \prod_{i=1}^{UG_R G_{T,u}} (\pi \gamma_{k,i})^{-1} \exp\left(-\frac{|\mathbf{H}_{b,\text{MU}}(i, k)|^2}{\gamma_{k,i}}\right), \quad (22)$$

where $\gamma_{k,i}$ represents the i -th hyperparameter for the k -th subcarrier.

$$\begin{aligned} \widehat{\mathbf{H}}_{b,\text{MU}}(:, k) &= \left(\boldsymbol{\Xi}_{\text{MU}}^H(:, :, k) \mathbf{C}_w^{-1} \boldsymbol{\Xi}_{\text{MU}}(:, :, k) + \boldsymbol{\Gamma}_{k,\text{MU}}^{-1} \right)^{-1} \\ &\quad \boldsymbol{\Xi}_{\text{MU}}^H(:, :, k) \mathbf{C}_w^{-1} \mathbf{Y}_{\text{MU}}(:, k) \quad \forall 0 \leq k \leq K-1, \end{aligned} \quad (23)$$

$$\prod_{k=0}^{K-1} \log[f(\mathbf{Y}_{\text{MU}}(:, k); \boldsymbol{\Gamma}_{k,\text{MU}})] = \sum_{k=0}^{K-1} \left(-MN_{RF}^R \log(\pi) - \log[\det(\mathbf{C}_{w,y,k})] - \mathbf{Y}_{\text{MU}}^H(:, k) \mathbf{C}_{w,y,k}^{-1} \mathbf{Y}_{\text{MU}}(:, k) \right), \quad (24)$$

The matrix $\boldsymbol{\Gamma}_{\text{MU}}$ can be defined as $\boldsymbol{\Gamma}_{\text{MU}} = \sum_{k=0}^{K-1} \boldsymbol{\Gamma}_{k,\text{MU}} \in \mathbb{R}^{UG_R G_{T,u} \times UG_R G_{T,u}}$, where

$\boldsymbol{\Gamma}_{k,\text{MU}} = \text{diag}(\gamma_{k,1}, \gamma_{k,2}, \dots, \gamma_{k,UG_R G_{T,u}})$. In Bayesian techniques, hyperparameters are used as variance parameters that play a crucial role in *promoting sparsity*, since coefficients associated with small γ_i are automatically driven toward zero, while those with large γ_i remain active [8]. This mechanism eliminates the need to explicitly impose sparsity levels or tune penalty factors, as done in traditional approaches such as OMP or LASSO [36]. Furthermore, by iteratively updating the hyperparameters, Bayesian approaches adaptively determine the most relevant coefficients, while providing posterior distributions that quantify uncertainty in the estimates.

The MMSE estimate $\widehat{\mathbf{H}}_{b,\text{MU}} \in \mathbb{C}^{UG_R G_{T,u} \times K}$ can be derived as Eq. (23) and one can observe that estimating $\widehat{\mathbf{H}}_{b,\text{MU}}$ ultimately reduces to the estimation of the corresponding hyperparameter matrix $\boldsymbol{\Gamma}_{k,\text{MU}}$. Furthermore, the log-likelihood $\log[f(\mathbf{Y}_{\text{MU}}(:, k); \boldsymbol{\Gamma}_{k,\text{MU}})]$ of the hyperparameter matrix $\boldsymbol{\Gamma}_{k,\text{MU}}$ can be expressed as Eq. (24) where $\mathbf{C}_{w,y,k} = \mathbf{C}_w + \boldsymbol{\Xi}_{\text{MU}} \boldsymbol{\Gamma}_{k,\text{MU}} \boldsymbol{\Xi}_{\text{MU}}^H$. However, the maximization of the log-likelihood w.r.t. $\boldsymbol{\Gamma}_{k,\text{MU}}$ is mathematically intractable. As a remedy, the expectation maximization (EM) [37] algorithm offers an efficient approach for iterative maximization. Furthermore, this algorithm guarantees convergence to the local maximum.

In preparation for exploiting the EM approach, let the complete information set be defined as $\{\mathbf{Y}_{\text{MU}}, \mathbf{H}_{b,\text{MU}}\}$ and the hyperparameter matrix be represented as $\boldsymbol{\Gamma}_{\text{MU}}$. Let $\widehat{\boldsymbol{\Gamma}}_{\text{MU}}^{(j-1)}$ denote the hyperparameter matrix estimate at the $(j-1)$ st EM iteration. We now present the detailed update procedure for the hyperparameter estimate $\widehat{\boldsymbol{\Gamma}}_{\text{MU}}^{(j)}$ in the j -th EM iteration. During the E-step, we evaluate the log-likelihood of the complete information set denoted as $\mathcal{L}(\boldsymbol{\Gamma}_{\text{MU}} | \widehat{\boldsymbol{\Gamma}}_{\text{MU}}^{(j-1)})$, where

$$\mathcal{L}(\boldsymbol{\Gamma}_{\text{MU}} | \widehat{\boldsymbol{\Gamma}}_{\text{MU}}^{(j-1)}) = \mathbb{E}_{\mathbf{H}_{b,\text{MU}} | \mathbf{Y}_{\text{MU}}; \widehat{\boldsymbol{\Gamma}}_{\text{MU}}^{(j-1)}} \{\log f(\mathbf{Y}_{\text{MU}}, \mathbf{H}_{b,\text{MU}}; \boldsymbol{\Gamma}_{\text{MU}})\}. \quad (25)$$

This can be re-written as

$$\begin{aligned} \mathcal{L}(\boldsymbol{\Gamma}_{\text{MU}} | \widehat{\boldsymbol{\Gamma}}_{\text{MU}}^{(j-1)}) &= \mathbb{E}_{\mathbf{H}_{b,\text{MU}} | \mathbf{Y}_{\text{MU}}; \widehat{\boldsymbol{\Gamma}}_{\text{MU}}^{(j-1)}} \{\log f(\mathbf{Y}_{\text{MU}} | \mathbf{H}_{b,\text{MU}}) + \\ &\quad \log f(\mathbf{H}_{b,\text{MU}}; \boldsymbol{\Gamma}_{\text{MU}})\}. \end{aligned} \quad (26)$$

The expression within the first expectation operator can be simplified as seen in Eq. (27) given at the top of the next page. One can observe that Eq. (27) is independent of $\boldsymbol{\Gamma}_{\text{MU}}$, which can be ignored in the subsequent M-step. Thus, the equivalent optimization problem can be formulated as

$$\widehat{\boldsymbol{\Gamma}}_{\text{MU}}^{(j)} = \arg \max_{\boldsymbol{\Gamma}_{\text{MU}}} \mathbb{E}_{\mathbf{H}_{b,\text{MU}} | \mathbf{Y}_{\text{MU}}; \widehat{\boldsymbol{\Gamma}}_{\text{MU}}^{(j-1)}} \{\log f(\mathbf{H}_{b,\text{MU}}; \boldsymbol{\Gamma}_{\text{MU}})\}. \quad (28)$$

Substituting $f(\mathbf{H}_{b,\text{MU}}; \boldsymbol{\Gamma}_{\text{MU}})$ from Equation (22), one can observe that the maximization problem can be separated w.r.t the hyperparameters and one obtains

$$\begin{aligned} \widehat{\boldsymbol{\Gamma}}_{\text{MU}}^{(j)} &= \arg \max_{\boldsymbol{\Gamma}_{\text{MU}}} \sum_{i=1}^{UG_R G_{T,u}} \left(- \sum_{k=0}^{K-1} \log(\gamma_{k,i}) - \right. \\ &\quad \left. \sum_{k=0}^{K-1} \frac{1}{\gamma_{k,i}} \mathbb{E}\{|\mathbf{H}_{b,\text{MU}}(i, k)|^2\} \right). \end{aligned} \quad (29)$$

To evaluate the hyperparameter estimate, it is necessary to compute a stationary point of the above cost function. This

$$\begin{aligned} \mathbb{E}_{\mathbf{H}_{b,\text{MU}}|\mathbf{Y}_{\text{MU}}; \widehat{\boldsymbol{\Gamma}}_{\text{MU}}^{(j-1)}} \{ \log f(\mathbf{Y}_{\text{MU}}|\mathbf{H}_{b,\text{MU}}) \} &= \sum_{k=0}^{K-1} \left(-MN_{RF}^R \log(\pi) - \log[\det(\mathbf{R}_{y,k})] - \right. \\ &\quad \left. (\mathbf{Y}_{\text{MU}}(:,k) - \boldsymbol{\Xi}_{\text{MU}}(:, :, k) \mathbf{H}_{b,\text{MU}}(:,k))^H \mathbf{C}_w^{-1} (\mathbf{Y}_{\text{MU}}(:,k) - \boldsymbol{\Xi}_{\text{MU}}(:, :, k) \mathbf{H}_{b,\text{MU}}(:,k)) \right), \end{aligned} \quad (27)$$

can be determined by differentiating the cost function above with respect to $\gamma_{k,i}$ and setting the derivative to zero. This leads to the following update equation for the hyperparameter estimate $\widehat{\gamma}_{k,i}^{(j)}$ in the j -th EM iteration

$$\widehat{\gamma}_{k,i}^{(j)} = \sum_{k=0}^{K-1} \mathbb{E}_{\mathbf{H}_{b,\text{MU}}|\mathbf{Y}_{\text{MU}}; \widehat{\boldsymbol{\Gamma}}_{\text{MU}}^{(j-1)}} \{ |\mathbf{H}_{b,\text{MU}}(i,k)|^2 \}, \quad (30)$$

where the conditional expectation in the above equation can be evaluated using the *a posteriori* pdf of $\mathbf{H}_{b,\text{MU}}$ given by $f(\mathbf{H}_{b,\text{MU}}|\mathbf{Y}_{\text{MU}}; \widehat{\boldsymbol{\Gamma}}_{\text{MU}}^{(j-1)}) = \mathcal{CN}(\widehat{\mathbf{H}}_{b,\text{MU}}^{(j)}, \boldsymbol{\Sigma}_{\text{MU}}^{(j)})$. Therefore, the *a posteriori* mean $\widehat{\mathbf{H}}_{b,\text{MU}} \in \mathbb{C}^{UG_R G_{T,u} \times K}$ and *a posteriori* covariance $\boldsymbol{\Sigma}_{\text{MU}} \in \mathbb{C}^{UG_R G_{T,u} \times UG_R G_{T,u} \times K}$ for the MU scenario are derived as Eq. (31) and (32) respectively. Note that we adopt a Gaussian prior for the beampoint channel coefficients. This choice is essential for two main reasons. Firstly, the observation model for the received pilots is Gaussian due to AWGN noise, and using a *Gaussian prior ensures conjugacy with the likelihood*. This leads to closed-form expressions for the posterior mean and covariance of the channel coefficients, as given by Eqs. (31) and (32) which are directly exploited in the E-step of the EM algorithm. Secondly, by associating each Gaussian prior with a variance hyperparameter that is iteratively refined in the M-step, the proposed framework is able to automatically enforce sparsity. As discussed above, the coefficients with small estimated variances are effectively pruned, while only the dominant beampoint coefficients are retained, thereby yielding accurate sparse channel recovery. Thus, the Gaussian prior is not an arbitrary choice, but a key enabler that provides both computational tractability within EM and the sparsity-inducing behavior that underpins the BGSR algorithm.

$$\widehat{\mathbf{H}}_{b,\text{BGSR}}(:,k) = \boldsymbol{\Sigma}_{\text{MU}}(:, :, k) \boldsymbol{\Xi}_{\text{MU}}^H(:, :, k) \mathbf{C}_w^{-1} \mathbf{Y}_{\text{MU}}(:,k) \quad \forall 0 \leq k \leq K-1, \quad (31)$$

$$\boldsymbol{\Sigma}_{\text{MU}}(:, :, k) = (\boldsymbol{\Xi}_{\text{MU}}^H(:, :, k) \mathbf{C}_w^{-1} \boldsymbol{\Xi}_{\text{MU}}(:, :, k) + \boldsymbol{\Gamma}_{k,\text{MU}}^{-1})^{-1} \quad \forall 0 \leq k \leq K-1. \quad (32)$$

The complexity of the proposed BGSR algorithm is mainly dominated by computing the *a posteriori* covariance matrix as it involves the inverse operation. The inversion in the update has $\mathcal{O}(U^3 G_R^3 G_{T,u}^3)$ complexity. In order to reduce the complexity, we adopt the **Woodbury matrix identity** $(\mathbf{A} + \mathbf{U}\mathbf{C}\mathbf{V})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{U}(\mathbf{C}^{-1} + \mathbf{V}\mathbf{A}^{-1}\mathbf{U})^{-1}\mathbf{V}\mathbf{A}^{-1}$ and the lemma $(\mathbf{I} + \mathbf{P})^{-1} = \mathbf{I} - (\mathbf{I} + \mathbf{P})^{-1}\mathbf{P}$, where the *a posteriori* covariance and mean can be re-expressed as Eq. (33) and (34) respectively. The inversion required for the new updates of the *a posteriori* covariance matrix has a complexity of $\mathcal{O}(U^2 G_R^2 G_{T,u}^2 M N_{RF}^R + M^3 (N_{RF}^R)^3)$, which reduces the overall computational load as $UG_R G_{T,u} \gg MN_{RF}^R$. Determining the *a posteriori* PDF from the mean of (31)

and covariance of (32), followed by substitution in (30), the hyperparameter update finally simplifies to

$$\widehat{\boldsymbol{\Gamma}}_{k,\text{MU}}^{(j)} = \text{diag} \left(\boldsymbol{\Sigma}_{\text{MU}}^{(j)}(:, :, k) + |\mathbf{H}_{b,\text{MU}}(:,k)|^2 \right). \quad (35)$$

To refine the hyperparameter estimates further, one can average these parameters across all the subcarriers to glean the processing gain resulting from multiple observations. Thus, the modified hyperparameter update equation is given by

$$\widehat{\boldsymbol{\Gamma}}_{\text{MU}}^{(j)} = \frac{1}{K} \sum_{k=0}^{K-1} \widehat{\boldsymbol{\Gamma}}_{k,\text{MU}}^{(j)}. \quad (36)$$

Note that upon forcing the dictionaries to be subcarrier-dependent due to beam squint, the underlying sparsity pattern of the channel is forced to remain consistent across subcarriers. This allows the hyperparameters to be averaged, since the non-zero coefficient locations do not change with subcarrier. Mathematically, the averaging can be interpreted as approximating the expectation of the hyperparameters across subcarriers, which reduces the variance of the estimates and suppresses the influence of noise or outliers from any individual subcarrier. Thus, the subcarrier-dependent dictionaries capture frequency variation, while hyperparameter averaging improves robustness without contradicting the problem formulation. This trade-off ensures that the estimates are not biased toward any single subcarrier, but instead they reflect the joint statistics across all subcarriers. The MU equivalent sensing matrix $\boldsymbol{\Xi}_{\text{MU}}$, equivalent received output \mathbf{Y}_{MU} , noise covariance \mathbf{C}_w , total subcarriers K and stopping parameters are given by inputs to Algorithm 1 to obtain the group sparse output across all the users. Let the concatenated MU transmit dictionary $\mathbf{A}_{T,\text{MU}}(\Theta_T, f_{0:K-1}) \in \mathbb{C}^{UN_{T,u} \times UG_{T,u} \times K}$ be formulated as

$$\mathbf{A}_{T,\text{MU}}(\Theta_T, f_{0:K-1}) = \text{blkdiag} \left(\mathbf{A}_{T,1}(\Theta_T, f_{0:K-1}), \mathbf{A}_{T,2}(\Theta_T, f_{0:K-1}), \dots, \mathbf{A}_{T,U}(\Theta_T, f_{0:K-1}) \right). \quad (37)$$

The estimated channel obtained through the BGSR framework can now be determined as

$$\widehat{\mathbf{H}}_{\text{BGSR}} = \mathbf{A}_R(\Phi_R, f_{0:K-1}) \text{vec}^{-1}(\widehat{\mathbf{H}}_{b,\text{BGSR}}) \mathbf{A}_{T,\text{MU}}^H(\Theta_T, f_{0:K-1}). \quad (38)$$

A significant implication of the proposed BGSR framework is the sparsity of the distribution across distinct dictionaries. This implies that when we use a common dictionary across all subcarriers, the sparse locations will vary for each subcarrier. However, when we employ a dictionary corresponding to each subcarrier, it fixes the sparse locations, resulting in group sparsity in the channel. In essence, the sparse locations remain consistent across different subcarrier dictionaries, despite each subcarrier having its unique dictionary matrix. This characteristic further enhances the robustness of the algorithm, which makes it novel and can be used in practical scenarios. A popular choice for embedding the non-linearity caused by low-resolution ADCs includes Bayesian inference [38], which corresponds to a discrete likelihood. However, in a dual-wideband

$$\Sigma_{\text{MU}}(:, :, k) = \Gamma_{k, \text{MU}} - \Gamma_{k, \text{MU}} \Xi_{\text{MU}}^H(:, :, k) (\mathbf{C}_w^{-1} + \Xi_{\text{MU}}(:, :, k) \Gamma_{k, \text{MU}} \Xi_{\text{MU}}^H(:, :, k))^{-1} \Xi_{\text{MU}}(:, :, k) \Gamma_{k, \text{MU}} \forall 0 \leq k \leq K-1, \quad (33)$$

$$\hat{\mathbf{H}}_{b, \text{BGSR}}(:, k) = \Gamma_{k, \text{MU}} \Xi_{\text{MU}}^H(:, :, k) \mathbf{C}_w^{-1} \left(\mathbf{I}_{M N_{\text{RF}}^R} + \Xi_{\text{MU}}(:, :, k) \Gamma_{k, \text{MU}} \Xi_{\text{MU}}^H(:, :, k) \mathbf{C}_w^{-1} \right)^{-1} \mathbf{Y}(:, k) \forall 0 \leq k \leq K-1, \quad (34)$$

Algorithm 1: BGSR based sparse channel estimation for dual-wideband MU THz MIMO systems

Input: Ξ_{MU} , \mathbf{Y}_{MU} , \mathbf{C}_w , K , $\mathbf{A}_R(\Phi_R, f_{0:K-1})$ and $\mathbf{A}_{T, \text{MU}}(\Theta_T, f_{0:K-1})$, ϵ and K_{max}
Output: $\hat{\mathbf{H}}_{b, \text{BGSR}} = \mathbf{A}_R(\Phi_R, f_{0:K-1}) \text{vec}^{-1}(\hat{\mathbf{H}}_{b, \text{BGSR}}) \mathbf{A}_{T, \text{MU}}^H(\Theta_T, f_{0:K-1})$

1 Initialization: $\hat{\Gamma}_{\text{MU}}^{(1)} = \mathbf{I}_{U G_R G_{T,u}}$, $\hat{\Gamma}_{\text{MU}}^{(0)} = \mathbf{0}_{U G_R G_{T,u}}$ and counter $j = 1$

2 while ($\| \hat{\Gamma}_{\text{MU}}^{(j)} - \hat{\Gamma}_{\text{MU}}^{(j-1)} \|_F^2 > \epsilon \& \& j < K_{\text{max}}$) **do**

E-Step: Compute the *a posteriori* covariance and mean as

for $k = 0, \dots, K-1$ **do**

 Compute the covariance using Eq. (33)

 Compute the mean using Eq. (34)

M-Step: Update the estimates of the hyperparameters as

for $k = 0, \dots, K-1$ **do**

 Update the hyperparameters using Eq. (35).

 Averaging hyperparameter using Eq. (36)

$j \leftarrow j + 1$

12 return: $\hat{\mathbf{H}}_{b, \text{BGSR}}$

THz system associated with a large number of antennas and subcarriers, this discrete likelihood must be recalculated iteratively for each subcarrier, leading to substantial memory usage and computational overhead. As a result, maintaining a Bayesian inference loop under these conditions becomes prohibitively demanding, making closed-form iterative updates elusive, unless additional approximations or dimensionality reduction techniques are introduced. The next subsection briefly describes the GSMP-based sparse channel estimation method, originally presented in [19] for THz communication and later extended for the MU-MIMO system model with low-resolution ADCs exploiting cross-subcarrier sparsity.

A. Group Structure Matching Pursuit based sparse channel estimation

In order to effectively compare our proposed MU BGSR approach, this treatise also introduces a MU group structure matching pursuit (GSMP)-based channel learning technique, which is described in the following section. The concatenated group sparse based channel model is detailed in Section V. Moreover, the MU-equivalent sensing matrix Ξ_{MU} , the received output \mathbf{Y}_{MU} , and the stopping parameter are fed into Algorithm-2 to obtain the estimated sparse channel for all the users. The *greedy-iterative* procedure consists of three distinct

phases which includes *Identification* (Steps 5, 6, 7), *Augmentation* (Steps 9 and 10) and the *Residual Updation* phase. Step 5 performs summed correlation across all subcarriers, rather than independently selecting the strongest correlation for each subcarrier, as it is done in standard OMP [9]. Step 6 then updates the residual in a grouped manner. While the existing MMV-based OMP approach [29] also processes the subcarriers jointly, they use a common dictionary for computing a single correlation metric aggregated both across all measurement vectors in each iteration. This typically selects one index at a time based on the aggregated correlation. By contrast, our GSMP variant in Algorithm-2 performs a more refined, group-based selection (as shown in Step 5 and 6), where multiple subcarrier indices are chosen to refine the group structure in each iteration. Additionally, Step 9 augments the empty matrix with columns obtained from the index set of Step 7, while Step 10 computes the intermediate LS solution from the matrix Ξ_{MU}^A . Finally, Step 11 updates the residual vector by subtracting the LS solution obtained from the original output vector, whereas Step 12 averages the residual vector at each iteration. The estimated channel can be reconstructed using Eq. (38) corresponding to all the users. Note that the GSMP algorithm is introduced in this work to demonstrate the *trade-off* between computational complexity and estimation accuracy in comparison to the BGSR framework and it is also one of the pair of algorithms proposed.

Remark 1: In order to relax the grid-based dictionary constraint, we also adopt the continuous basis pursuit [39] based dictionary, which is approximated by the first order Taylor series interpolation of

$$\tilde{\mathbf{a}}(\phi, f_k) = \tilde{\mathbf{a}}(\phi_r, f_k) + (\phi - \phi_r) \frac{\partial \tilde{\mathbf{a}}(\phi, f_k)}{\partial \phi} \Big|_{\phi_r} + \mathcal{O}((\phi - \phi_r)^2), \quad (39)$$

where ϕ represents the continuous AoA/ AoD, while $\phi_r = 2\pi \frac{(r-1)}{G_R}$ represents the grid point having the minimal distance from ϕ . Therefore, the new dictionary will contain the grid points from the quantized array factor along with its derivatives, which can be further formulated as Eq. (40)

$$\begin{aligned} \mathfrak{T}_R(\phi_r, f_k) = & [\tilde{\mathbf{a}}(\phi_1, f_k), \dots, \tilde{\mathbf{a}}(\phi_{G_R}, f_k), \\ & \tilde{\mathbf{b}}(\phi_1, f_k), \dots, \tilde{\mathbf{b}}(\phi_{G_R}, f_k)], \end{aligned} \quad (40)$$

where $\tilde{\mathbf{b}}(\phi_r, f_k) = \frac{\partial \tilde{\mathbf{a}}(\phi, f_k)}{\partial \phi} \Big|_{\phi_r}$. Furthermore, to incorporate the offset angle wrt the grid, an interpolator is defined as

$$\mathbf{t}_R = \left[\underbrace{1, \dots, 1}_{G_R}, \underbrace{\Delta\phi, \dots, \Delta\phi}_{G_R} \right], \quad (41)$$

where $|\Delta\phi| \leq \frac{\pi}{G_R}$. Finally, the Taylor series based off-grid dictionary (TBoD) is given by Eq. (42). Note that, a similar

$$\tilde{\mathbf{A}}_R(\Phi_R, f_k) = \mathfrak{T}_R(\phi_r, f_k) \mathbf{t}_R = [\tilde{\mathbf{a}}_R(\phi_1, f_k), \dots, \tilde{\mathbf{a}}_R(\phi_{G_R}, f_k), \Delta\phi \tilde{\mathbf{b}}_R(\phi_1, f_k), \dots, \Delta\phi \tilde{\mathbf{b}}_R(\phi_{G_R}, f_k)]. \quad (42)$$

$$\tilde{\mathbf{A}}_{T,u}(\Theta_{T,u}, f_k) = [\tilde{\mathbf{a}}_{T,u}(\theta_1, f_k), \dots, \tilde{\mathbf{a}}_{T,u}(\theta_{G_{T,u}}, f_k), \Delta\theta \tilde{\mathbf{b}}_{T,u}(\theta_1, f_k), \dots, \Delta\theta \tilde{\mathbf{b}}_{T,u}(\theta_{G_{T,u}}, f_k)]. \quad (43)$$

Algorithm 2: GSMP based sparse channel estimation for dual-wideband MU THz MIMO systems

Input: Ξ_{MU} , \mathbf{Y}_{MU} , $\mathbf{A}_R(\Phi_R, f_{0:K-1})$ and $\mathbf{A}_{T,\text{MU}}(\Theta_T, f_{0:K-1})$, ϵ_0
Output: \mathbf{H}_{GSMP} = $\mathbf{A}_R(\Phi_R, f_{0:K-1}) \text{vec}^{-1}(\hat{\mathbf{H}}_{b,\text{GSMP}}) \mathbf{A}_{T,\text{MU}}^H(\Theta_T, f_{0:K-1})$

1 Initialization: Index set $\mathcal{A} = []$, Empty matrix $\Xi_{\text{MU}}^{\mathcal{A}} = []$, $\hat{\mathbf{H}}_{\text{LS},\text{MU}} = []$, Residue vector $\mathbf{T}_0 = \mathbf{0}_{MN_{\text{RF}}^R \times K}$, $\mathbf{T}_1 = \mathbf{Y}_{\text{MU}}$, Estimated beamspace channel $\hat{\mathbf{H}}_{b,\text{GSMP}} = \mathbf{0}_{UG_R G_{T,u} \times K}$, counter $j = 1$

2 while $(\|\mathbf{T}_{j-1}\|_F^2 - \|\mathbf{T}_j\|_F^2) \geq \epsilon_0$ **do**

3 $\mathbf{Z}_{\text{MU}} = []$

4 **for** $k = 0, 1, \dots, K-1$ **do**

5 $\mathbf{Z}_{\text{MU}} = \mathbf{Z}_{\text{MU}} \cup (\Xi_{\text{MU}}^{\mathcal{A}}(:, :, k) \mathbf{T}_j(:, k))$

6 $\xi = \arg \max_k \sum_k |\mathbf{Z}_{\text{MU}}(:, k)|^2$

7 $\mathcal{A} = \mathcal{A} \cup \xi$

8 **for** $k = 0, 1, \dots, K-1$ **do**

9 $\Xi_{\text{MU}}^{\mathcal{A}} = \Xi_{\text{MU}}(:, \mathcal{A}, k)$

10 $\hat{\mathbf{H}}_{\text{LS},\text{MU}}(:, k) = (\Xi_{\text{MU}}^{\mathcal{A}})^\dagger \mathbf{Y}_{\text{MU}}(:, k)$

11 $\mathbf{T}_j(:, k) = \mathbf{Y}_{\text{MU}}(:, k) - \Xi_{\text{MU}}^{\mathcal{A}} \hat{\mathbf{H}}_{\text{LS},\text{MU}}(:, k)$

12 $\mathbf{T}_j = \frac{\mathbf{T}_j}{K}$

13 $j \leftarrow j + 1$

14 $\hat{\mathbf{H}}_{b,\text{GSMP}}(\mathcal{A}, 0 : K-1) = \hat{\mathbf{H}}_{\text{LS},\text{MU}}$

15 return: $\hat{\mathbf{H}}_{b,\text{GSMP}}$

TBoD matrix is formed at the transmitter, which is given by Eq. (43). The next section presents the computational complexity of the proposed algorithm along with that of GSMP algorithm.

B. Computational Complexity

This subsection derives the computational complexities of the BGSR and GSMP based THz channel learning techniques. Table-IV list the per-subcarrier and per-iteration complexity of the GSMP-framework along with various key computational steps. The overall computational complexity of the GSMP-framework is on the order of $(KUG_R G_{T,u} MN_{\text{RF}}^R + K(\frac{2}{3}j^3 + j^2 MN_{\text{RF}}^R + 3j MN_{\text{RF}}^R))$, where j represents the current iteration. The worst-case complexity order is $\mathcal{O}(K(M^3 N_{\text{RF}}^R)^3)$, which is attributed to the necessity of an intermediate LS estimate at each iteration. Table-V lists the per-subcarrier and per-iteration complexity of the BGSR-framework along with various key computational steps. The worst-case complexity of the BGSR technique for each subcarrier is on the order of $\mathcal{O}(2(MN_{\text{RF}}^R)^3 + 2UG_R G_{T,u} (MN_{\text{RF}}^R)^2 + 3UG_R G_{T,u} MN_{\text{RF}}^R + (UG_R G_{T,u})^2 MN_{\text{RF}}^R)$, which is dominated by the matrix-inversion of \mathbf{C}_w . Moreover, as discussed above, $UG_R G_{T,u} \gg$

MN_{RF}^R ; therefore, the approximated complexity of the BGSR algorithm for each subcarrier is $\mathcal{O}(U^2 G_R^2 G_{T,u}^2 MN_{\text{RF}}^R + M^3 (N_{\text{RF}}^R)^3)$. The worst-case computational complexity for all the subcarriers is $\mathcal{O}(K(U^2 G_R^2 G_{T,u}^2 MN_{\text{RF}}^R + M^3 (N_{\text{RF}}^R)^3))$, implying that the computational complexity increases linearly with the number of subcarriers and also higher than GSMP framework. However, as shown in Section-VI, BGSR indicates superior performance. Hence, there is a clear trade-off between computational complexity and estimation performance. The BCRB characterizing the estimation of the MU channel is derived in detail next.

C. MU Bayesian Cramér-Rao bound

Consider the parameterized Gaussian prior associated with the beamspace channel matrix $\mathbf{H}_{b,\text{MU}}$ as

$$f(\mathbf{H}_{b,\text{MU}}; \boldsymbol{\Gamma}_{\text{MU}}) = \prod_{k=0}^{K-1} \frac{1}{(\pi)^{UG_R G_{T,u}} \det(\boldsymbol{\Gamma}_{k,\text{MU}})} \exp(-\mathbf{H}_{b,\text{MU}}^H(:, k) \boldsymbol{\Gamma}_{k,\text{MU}}^{-1} \mathbf{H}_{b,\text{MU}}(:, k)). \quad (44)$$

The log-likelihood of the beamspace channel $\mathbf{H}_{b,\text{MU}}$ is given as

$$\mathcal{L}(f(\mathbf{H}_{b,\text{MU}}; \boldsymbol{\Gamma}_{\text{MU}})) = \sum_{k=0}^{K-1} \left(-UG_R G_{T,u} \log(\pi) - \log[\det(\boldsymbol{\Gamma}_{k,\text{MU}})] - \mathbf{H}_{b,\text{MU}}^H(:, k) \boldsymbol{\Gamma}_{k,\text{MU}}^{-1} \mathbf{H}_{b,\text{MU}}(:, k) \right). \quad (45)$$

The conditional PDF of the pilot output vector \mathbf{Y}_{MU} , is represented by the complex normal distribution $\prod_{k=0}^{K-1} \mathcal{CN}(\Xi_{\text{MU}}(:, :, k) \mathbf{H}_{b,\text{MU}}(:, k), \mathbf{C}_w)$. Therefore, the conditional PDF is given by Equation (49), where $\Upsilon = \frac{1}{(\pi)^{MN_{\text{RF}}^R} \det(\mathbf{C}_w)}$. Taking the logarithm on both sides of Equation (49) yields Equation (50). Let $\mathbf{I}_{\text{MU}} \in \mathbb{C}^{UG_R G_{T,u} \times UG_R G_{T,u}}$ denote the Bayesian Fisher information matrix (FIM), formulated as

$$\mathbf{I}_{\text{MU}} = \mathbf{I}_{y,\text{MU}} + \mathbf{I}_{b,\text{MU}}, \quad (46)$$

where $\mathbf{I}_{y,\text{MU}} \in \mathbb{C}^{UG_R G_{T,u} \times UG_R G_{T,u}}$ denotes the FIM associated with the combined received signal \mathbf{Y}_{MU} , while $\mathbf{I}_{b,\text{MU}} \in \mathbb{C}^{UG_R G_{T,u} \times UG_R G_{T,u}}$ quantifies the FIM arising from the *a priori* knowledge about the beamspace CSI matrix $\mathbf{H}_{b,\text{MU}}$. The following equations represent these quantities mathematically

$$\mathbf{I}_{y,\text{MU}} = -\mathbb{E}_{\mathbf{Y}_{\text{MU}}, \mathbf{H}_{b,\text{MU}}} \left\{ \frac{\partial^2 \mathcal{L}(\mathbf{Y}_{\text{MU}} | \mathbf{H}_{b,\text{MU}})}{\partial \mathbf{H}_{b,\text{MU}} \partial \mathbf{H}_{b,\text{MU}}^H} \right\}, \quad (47)$$

$$\mathbf{I}_{b,\text{MU}} = -\mathbb{E}_{\mathbf{H}_{b,\text{MU}}} \left\{ \frac{\partial^2 \mathcal{L}(\mathbf{H}_{b,\text{MU}}; \boldsymbol{\Gamma}_{\text{MU}})}{\partial \mathbf{H}_{b,\text{MU}} \partial \mathbf{H}_{b,\text{MU}}^H} \right\}. \quad (48)$$

Using the conditional PDF as derived in Equation (50), the FIM $\mathbf{I}_{y,\text{MU}}$ can be modeled as

$$\mathbf{I}_{y,\text{MU}} = \sum_{k=0}^{K-1} \Xi_{\text{MU}}^H(:, :, k) \mathbf{C}_w^{-1} \Xi_{\text{MU}}(:, :, k). \quad (51)$$

Table IV
COMPUTATION COMPLEXITY OF GSMP SCHEME, PER-SUBCARRIER IN THE j -TH ITERATION

Operation	Complex multiplications	Complex additions
$\Xi_{\text{MU}}^H(:, :, k)\mathbf{T}(:, k)$	$UG_R G_{T,u} M N_{\text{RF}}^R$	$UG_R G_{T,u} (M N_{\text{RF}}^R - 1)$
$\widehat{\mathbf{H}}_{\text{LS},\text{MU}}(:, k) = (\Xi_{\text{MU}}^A)^\dagger \mathbf{Y}_{\text{MU}}(:, k)$	$\frac{2}{3}j^3 + j^2 M N_{\text{RF}}^R + 2j M N_{\text{RF}}^R$	$\frac{2}{3}j^3 + j^2 (M N_{\text{RF}}^R - 1) + j (M N_{\text{RF}}^R - 1)$
$\mathbf{Y}_{\text{MU}}(:, k) - \Xi_{\text{MU}}^A \widehat{\mathbf{H}}_{\text{LS},\text{MU}}(:, k)$	$j M N_{\text{RF}}^R$	$j M N_{\text{RF}}^R$

Table V
COMPUTATION COMPLEXITY OF BGSR FRAMEWORK, PER-SUBCARRIER, PER-EM ITERATION

Operation	Complex multiplication	Complex additions
$\Gamma_{k,\text{MU}} \Xi_{\text{MU}}^H(:, :, k)$	$UG_R G_{T,u} M N_{\text{RF}}^R$	0
\mathbf{C}_w^{-1}	$\frac{(M N_{\text{RF}}^R)^3}{2} + \frac{3(M N_{\text{RF}}^R)^2}{2}$	$\frac{(M N_{\text{RF}}^R)^3}{2} - \frac{3(M N_{\text{RF}}^R)^2}{2}$
$\Xi_{\text{MU}}(:, :, k) \Gamma_{k,\text{MU}} \Xi_{\text{MU}}^H(:, :, k)$	$(M N_{\text{RF}}^R)^2 UG_R G_{T,u} + M N_{\text{RF}}^R UG_R G_{T,u}$	$(M N_{\text{RF}}^R)^2 (UG_R G_{T,u} - 1) + (M N_{\text{RF}}^R) (UG_R G_{T,u} - 1)$
$(\mathbf{C}_w^{-1} + \Xi_{\text{MU}}(:, :, k))^{-1}$	$\frac{(M N_{\text{RF}}^R)^3}{2} + \frac{3 UG_R G_{T,u} (M N_{\text{RF}}^R)^2}{2} + UG_R G_{T,u} M N_{\text{RF}}^R$	$\frac{(M N_{\text{RF}}^R)^3}{2} - \frac{3 UG_R G_{T,u} (M N_{\text{RF}}^R)^2}{2} - UG_R G_{T,u} M N_{\text{RF}}^R$

$$f(\mathbf{Y}_{\text{MU}} | \mathbf{H}_{b,\text{MU}}) = \prod_{k=0}^{K-1} \Upsilon \exp \left(- (\mathbf{Y}_{\text{MU}}(:, k) - \Xi_{\text{MU}}(:, :, k) \mathbf{H}_{b,\text{MU}}(:, k))^H \mathbf{C}_w^{-1} (\mathbf{Y}_{\text{MU}}(:, k) - \Xi_{\text{MU}}(:, :, k) \mathbf{H}_{b,\text{MU}}(:, k)) \right), \quad (49)$$

$$\mathcal{L}(\mathbf{Y}_{\text{MU}} | \mathbf{H}_{b,\text{MU}}) = \sum_{k=0}^{K-1} \left(- M N_{\text{RF}}^R \log(\pi) - \log[\det(\mathbf{C}_w)] - (\mathbf{Y}_{\text{MU}}(:, k) - \Xi_{\text{MU}}(:, :, k) \mathbf{H}_{b,\text{MU}}(:, k))^H \mathbf{C}_w^{-1} (\mathbf{Y}_{\text{MU}}(:, k) - \Xi_{\text{MU}}(:, :, k) \mathbf{H}_{b,\text{MU}}(:, k)) \right). \quad (50)$$

In a similar vein, the FIM corresponding to the $\mathbf{H}_{b,\text{MU}}$ can be derived as

$$\mathbf{I}_{b,\text{MU}} = \widehat{\Gamma}_{\text{MU}}^{-1}. \quad (52)$$

The complete Bayesian FIM is now obtained by combining Equation (51) and (52), which is formulated as

$$\mathbf{I}_{\text{MU}} = \sum_{k=0}^{K-1} (\Xi_{\text{MU}}^H(:, :, k) \mathbf{C}_w^{-1} \Xi_{\text{MU}}(:, :, k)) + \widehat{\Gamma}_{\text{MU}}^{-1}. \quad (53)$$

We further define the equivalent sparsifying dictionary across all subcarriers as $\Delta_{\text{MU}} = [\Psi_{\text{MU}}[0] \Psi_{\text{MU}}[1] \dots \Psi_{\text{MU}}[K-1]] \in \mathbb{C}^{UN_R N_{T,u} \times UG_R G_{T,u} \times K}$. Finally, the BCRB for the MSE of the estimated CSI is expressed as

$$\text{MSE}(\widehat{\mathbf{H}}_{\text{MU}}) \geq \text{Tr} (\Delta_{\text{MU},0:K-1} \mathbf{I}_{\text{MU}}^{-1} \Delta_{\text{MU},0:K-1}^H). \quad (54)$$

VI. SIMULATION RESULTS

This section presents our results characterizing the performance of the proposed channel learning technique, namely the BGSR designed for MU THz hybrid MIMO systems. Table-VI provides the numerical parameters considered for the simulation study. Furthermore, the root raised cosine and rectangular pulse shapes denoted by RRC-PSF and Rect-PSF, respectively, are considered for performance evaluation. Note that, Rect-PSF represents the dual-wideband channel when a rectangular filter is used instead of Root Raised Cosine filter in Equation (12). The quantity \mathfrak{B}_T in Table-VI represents the combined transmit gain for all the users, i.e., $\sum_{u=1}^U \mathfrak{B}_{T,u} = \mathfrak{B}_T$. To generate the amplitudes of the LoS and NLoS components, for

both the dual wideband formulations, we utilize Equations (13) and (15) respectively. The phase shifts corresponding to (13) and (15) are generated as independent samples of a uniformly distributed random variable over the interval $(-\pi, \pi]$. Both the transmit and receive antennas maintain inter-antenna spacings of $d_t = d_r = \frac{\lambda}{2}$. The MA coefficient appearing in Equations (13) and (15) is calculated using the HITRAN database [1]. We consider an indoor office scenario with a molecular composition of 20.9% oxygen, 78.1% nitrogen and 1% water vapour, where the temperature and pressure are set to 298 K and 1 atm, respectively. We consider 6 scatters that can be present in a general office scenario, which are detailed in Table-VII. Furthermore, the data symbols are randomly generated with an average power of unity. Therefore, the SNR in decibels (dB) is calculated as $\text{SNR}(\text{dB}) = 10 \log_{10} \left(\frac{1}{\sigma^2} \right)$. The receiving and transmitting array response matrices are generated using (18) and (37) respectively, and the grid-sizes are considered to be $G_R \geq 2N_R$ and $G_{T,u} \geq 2N_{T,u}$ [35]. The frequency-independent phase shifters as shown in Eqs. (2) are modeled as [41]

$$\mathbf{F}_{\text{RF},m,u}(\kappa, \ell) = \frac{1}{\sqrt{N_{T,u}}} \exp(j\mathcal{O}_{\kappa,\ell}), \quad (55)$$

$$\mathbf{W}_{\text{RF},m}(\kappa, \ell) = \frac{1}{\sqrt{N_R}} \exp(j\vartheta_{\kappa,\ell}), \quad (56)$$

where the phases $\mathcal{O}_{\kappa,\ell}, \vartheta_{\kappa,\ell}$ are randomly sampled with uniform probability over $\mathcal{T} = \left\{ 0, \frac{2\pi}{2^{N_Q}}, \dots, \frac{(2^{N_Q}-1)}{2^{N_Q}} \right\}$. For performance evaluation, we employ the widely used NMSE

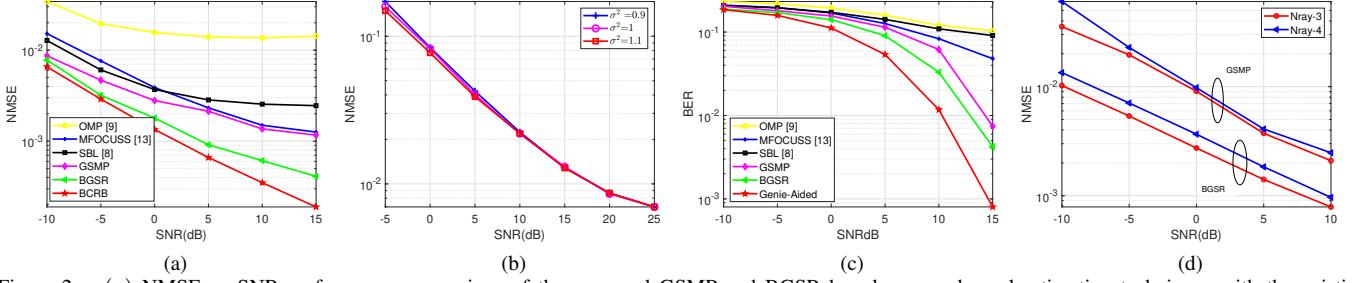


Figure 2. (a) NMSE vs SNR performance comparison of the proposed GSMP and BGSR based sparse channel estimation techniques with the existing state-of-the-art techniques (b) NMSE vs SNR comparison under varying noise variance (c) BER vs SNR performance comparison of the proposed GSMP and BGSR based sparse channel estimation techniques with the existing state-of-the-art techniques (d) Effect of increasing diffused rays for MU scenarios.

Table VI
SIMULATION PARAMETERS FOR THE DUAL-WIDEBAND THz HYBRID MIMO SYSTEM

Parameter	Value
# of TAs (N_T)	4
# of RAs (N_R)	48
# of TAs RF chains (N_{RF}^T)	2
# of RAs RF chains (N_{RF}^R)	8
# of subcarriers (K)	64
# of pilot blocks (M)	20
# of data vectors (N_d)	100
# of Users (U)	3
# of NLoS components (N_{NLoS})	3
# of delay taps (L)	3
Transmit Antenna Gain (\mathfrak{B}_T)	31 dB
Receive Antenna Gain (\mathfrak{B}_R)	31 dB
Transmit angular grid size ($G_{T,u}$)	8
Receive angular grid size (G_R)	96
Operating frequency (f_c)	0.65 THz
Transmission distance (d)	15m
Bandwidth (B)	5 GHz
Angle quantization parameter (N_Q)	4
ADC Resolution	3-bit
Constellation	8-PSK
Roll-off factor for RRC-PSF	0.80
Upsampling factor	20
GSMP Threshold (ϵ_0)	2
BGSR Thresholds (ϵ, K_{\max})	1, 20

Table VII
LIST OF MATERIALS USED FOR SIMULATION ENVIRONMENT [34], [40]

Material Type	σ_r (in mm)	\varkappa (in cm^{-1})	η
Polycarbonate (PC)	0	23	1.52
Polystyrene (PS)	0.002	2	1.6
Polyvinyl chloride (PVC)	0.028	19	1.68
Plaster s1	0.05	10	2
Gypsum plaster	0.13	38	1.4
Plaster s2	0.15	10	2

metric given

$$\text{NMSE} = \frac{\sum_{k=0}^{K-1} \|\hat{\mathbf{H}}[k] - \mathbf{H}[k]\|_F^2}{\sum_{k=0}^{K-1} \|\mathbf{H}[k]\|_F^2}. \quad (57)$$

A. GMM based MU AoA/ AoD generation

We harness a GMM for generating the AoAs/AoDs, as it provides an accurate representation of the angular characteristics in the THz band [7]. To generate angles corresponding to distinct users, the mean angle for each user is drawn from a uniform distribution, $\tilde{\theta}_u \sim \mathcal{U}[-180^\circ, 180^\circ]$, where $\tilde{\theta}_u$ denotes the mean angle of the u -th user. Around each mean, the angular spread is modeled by a Gaussian kernel of variance $\check{\sigma}^2$, where $\check{\sigma}^2$ is chosen according to the measurements reported in [28] (Table III). Thus, the AoA/AoD distribution using a two-component GMM can be expressed as

$$\text{GMM}(x_u) = \sum_{i=1}^2 a_{u,i} \mathcal{N}(x_u | \tilde{\theta}_{u,i}, \check{\sigma}^2), \quad (58)$$

where $a_{u,i}$ represent the mixture weight. They are randomly drawn from a uniform distribution $\{a_{u,1}, a_{u,2}\} \sim \mathcal{U}(0, 1)$ satisfying $a_{u,1} + a_{u,2} = 1$, ensuring a valid convex combination. Furthermore, to avoid angular overlap among different users, we impose a *minimum separation constraint* of

$$|\tilde{\theta}_u - \tilde{\theta}_v| \geq d_{\min} \quad \forall u \neq v \quad (59)$$

where d_{\min} is chosen to guarantee sufficiently distinct spatial signatures. For our simulations we have considered $d_{\min} \approx 5^\circ$. This ensures that although all users share the same angular spread parameter σ , their mean directions remain sufficiently separated.

B. MU performance analysis

In this section, we evaluate the performance of the proposed BGSR based CSI estimation technique. The parameters considered for simulation are shown in Table VI. Fig. 2(a) portrays the improved performance of the proposed BGSR technique in contrast to several orthodox sparse estimation techniques, such as SBL [8], OMP [9], MFOCUSS [13], GSMP for THz hybrid MIMO systems. The performance of the OMP algorithm is significantly affected by the choice of the stopping criterion and the dictionary matrix, as well as error propagation. Additionally, convergence-related issues are experienced with the MFOCUSS algorithm, which exhibits high sensitivity to variations in the regularization parameter. The SMV-based SBL does not exploit group sparsity, as it estimates the channel separately on individual subcarriers, which

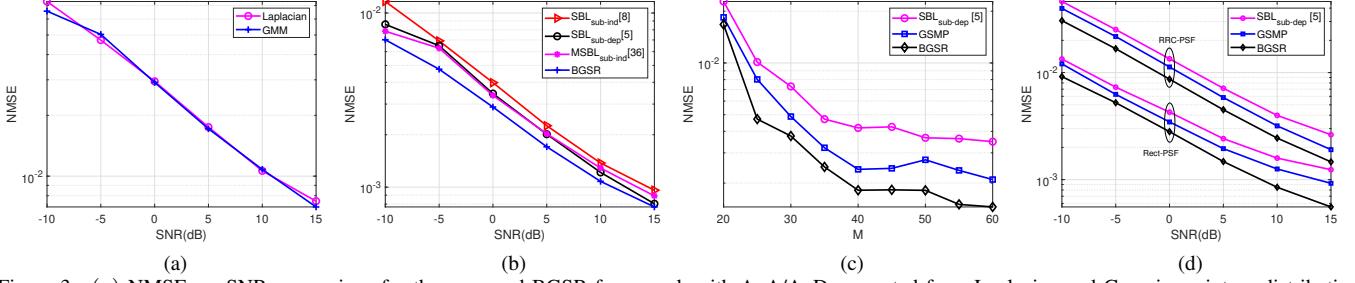


Figure 3. (a) NMSE vs. SNR comparison for the proposed BGSR framework with AoA/AoD generated from Laplacian and Gaussian mixture distributions (b) NMSE vs SNR comparison for subcarrier-independent and subcarrier-dependent Bayesian frameworks (c) NMSE vs training frames for MU scenario (d) NMSE vs SNR performance comparison for the GSMP and BGSR based sparse estimation techniques with RRC-PSF and Rect-PSF based dual-wideband channel for MU scenario

degrades its performance. The GSMP framework demonstrates noteworthy performance improvement, since it leverages the group sparsity via the estimation of the channel over all the subcarriers. However, similar to the OMP algorithm, the GSMP framework is sensitive to the stopping criterion and to the choice of the dictionary matrix. By contrast, the proposed BGSR framework leverages EM-based hyperparameter estimation, eliminating the need for manual parameter adjustments, while ensuring robust convergence. Furthermore, the proposed BGSR framework is unaffected by the selection of a dictionary matrix and stopping threshold, which vindicates the efficacy of the proposed algorithm. Moreover, observe from the figure that the performance of the proposed BGSR framework closely tracks the BCRB. This is of significant importance, as the BCRB is derived under the assumption of perfect knowledge of the AoA/ AoDs, which is an ideal scenario, while the proposed BGSR framework does not require any prior information. Note that we have assumed \mathbf{C}_w to be perfectly known. However, as depicted in Fig. 2(b), if the noise covariance is estimated, the NMSE performance remains almost identical, illustrating that the proposed BGSR algorithm is robust to noise mismatches.

Fig. 2(c) shows the accuracy of the data detection achieved using the MMSE receiver for the BGSR and other traditional sparse estimation based approaches along with that of the Genie-aided detector. It is evident from the figure that the resultant BER improves upon increasing the data SNR and yields the best performance for the proposed BGSR based technique. This is in line with the trend for the NMSE of the estimated CSI. Furthermore, the BERs of the BGSR and GSMP techniques closely approximate the hypothetical Genie-aided detector, demonstrating the efficacy of the proposed schemes. Fig. 2(d) illustrates the impact of a diffused multipath channel on the proposed BGSR scheme for $N_{ray} = \{3, 4\}$, with a comparison to the GSMP approach. As the number of diffused rays increases, the performance of both schemes degrades. This behavior arises from the GMM adopted for generating AoAs/AoDs. In the GMM framework, additional rays correspond to extra Gaussian components, which broaden the angular spread around each user's mean direction. A larger spread reduces the sparsity of the beamspace channel, as more angular bins become active. Consequently, the dictionary atoms become less distinguishable, which diminishes

the efficacy of sparse recovery and leads to degraded NMSE performance. Note that, in practical THz scenarios, the true angular statistics may deviate from the Gaussian assumption. As shown in Fig. 3(a), the BGSR framework exhibits almost identical NMSE performance under both the Gaussian and non-Gaussian distributions. This robustness arises because the Bayesian framework primarily exploits the inherent sparsity of the beamspace channel, rather than relying on the *exact functional form of the angular distribution*. While the Gaussian prior is adopted in our formulation due to its analytical tractability and closed-form EM updates, the results confirm that the algorithm remains resilient to moderate distributional mismatches. Performance degradation is observed only when the variance of rays becomes significantly larger or when the number of clusters increase significantly, as this reduces sparsity and leads to overlapping dictionary atoms, as depicted in Fig. 2(d). These results demonstrate that the BGSR framework is not limited to Gaussian angular models and generalizes well to other practical distributions.

Fig. 3(b) compares the NMSE vs SNR performance of subcarrier-independent and subcarrier-dependent dictionaries-based sparse recovery frameworks. The relatively poor performance of the subcarrier-independent dictionary-based SMV-SBL [8] is primarily due to its reliance on a dictionary constructed at the central subcarrier, which cannot accurately capture the frequency-dependent variations introduced by the beam-squint effect. This mismatch leads to degraded channel estimation performance across a wideband. When a subcarrier-dependent dictionary is incorporated into the SMV-SBL [5], the dictionary better reflects the frequency-selective characteristics, resulting in performance improvement. However, the single-measurement formulation remains limited, since it does not exploit the joint sparsity across users. By contrast, the MMV-SBL framework substantially enhances the performance by jointly estimating all user channels under a shared sparsity structure. The subcarrier-independent dictionary-based MMV-SBL [36] provides noticeable improvement compared to the SMV variants, and the proposed BGSR achieves the best performance, as it simultaneously leverages joint sparsity and accurately models the beam-squint effect, leading to the lowest NMSE across all SNR values. Fig. 3(c) depicts the NMSE versus the number of pilot blocks M for the proposed BGSR and GSMP techniques in comparison with traditional

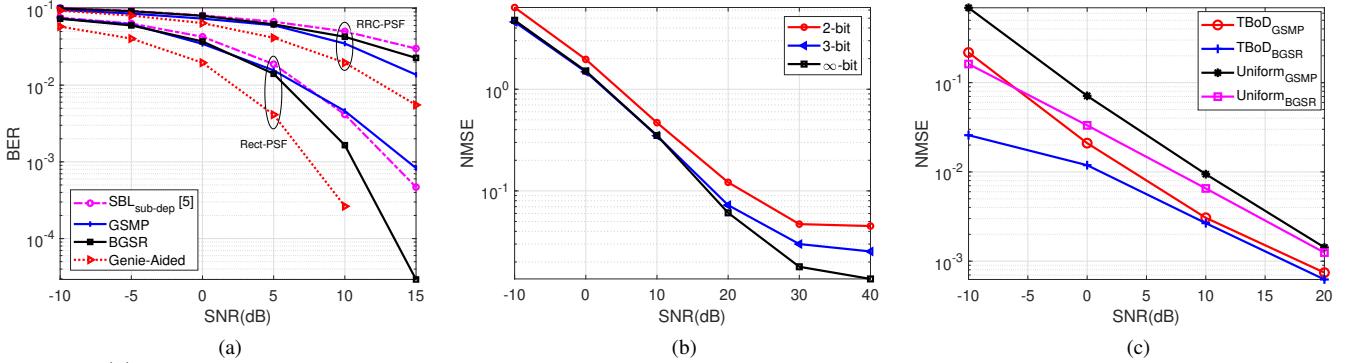


Figure 4. (a) BER vs SNR performance comparison for the GSMP and BGSR based sparse estimation techniques with RRC-PSF and Rect-PSF based dual-wideband channel for MU scenario (b) Effect of low resolution ADCs for the BGSR based technique for MU scenario. (c) NMSE vs SNR performance comparison for uniform and TBoD based channel estimation

Bayesian methods. The NMSE decreases with increasing number of pilot blocks, as a larger number of pilots provides richer observations and thereby improves channel estimation accuracy. Furthermore, BGSR consistently outperforms the other schemes by leveraging simultaneous sparsity within the Bayesian framework, while GSMP achieves moderate performance and SBL performs the worst due to its limited ability to exploit the joint sparsity structure. These results further validate the robustness of the proposed BGSR and GSMP frameworks.

Fig. 3(d) compares the performance of the BGSR and GSMP based channel estimation approaches by considering both the RRC-PSF and Rect-PSF based dual-wideband channels. The Rect-PSF based dual-wideband channel suffers from spectral leakage, which leads to a higher level of interference between the adjacent channels, which may erode the overall performance of the communication system. On the other hand, the RRC-PSF based dual wideband channel introduces the problem of signal strength reduction at the edges of the frequency band. This attenuation may adversely affect signal recovery at the receiver end, as the diminished signal may fall below the receiver's detection threshold. Thus, there is a clear trade off between both the models and the consideration of the filter depends upon the specific application. Fig. 4(a) compares the performance of the RRC-PSF and Rect-PSF based dual-wideband channels in terms of its BER. The improved accuracy of GSMP and BGSR based CSI learning techniques is clearly evident from the figure. Moreover, the BER is compared to that of a Genie-aided hypothetical scenario, which assumes perfect knowledge of the THz hybrid MIMO channel. It is evident that the BGSR technique learns the dual-wideband channel effectively and performs close to the ideal genie receiver. Thus, the improved performance makes the BGSR technique eminently suitable for practical THz system implementation.

Fig. 4(b) demonstrates the influence of using low-resolution ADCs on the estimated CSI performance for the BGSR framework. One can readily observe from the figure that the proposed BGSR technique with 3-bit ADC resolution closely approaches the performance of an ∞ -bit resolution transceiver. This observation is of considerable importance as it highlights the practicality of the BGSR scheme, which

often necessitates low-resolution ADCs to efficiently handle its substantial bandwidth at a nominal power consumption. Furthermore, none of the existing studies has addressed the influence of low-resolution ADCs in the face of dual-wideband effects. Moreover, it can be observed from Table-II that the quantization noise-to-signal power ratio v increases with decreasing the ADC resolution b . Therefore, for ADCs with fewer than 3 bits ($b < 3$), the higher quantization noise leads to significant performance degradation, as illustrated in Fig. 4(b). Thus, the proposed model not only conserves power but also leads to a reduction in hardware costs and energy consumption, thereby highlighting the practical advantages of the proposed approach. Fig 4 (c) compares the performance of the existing on-grid dictionary and of the TBoD matrix. The TBoD outperforms the on-grid dictionary based performance for both the proposed techniques due to its ability to more accurately represent continuous angular domains. Unlike the on-grid approach, which is constrained by discretized grid points and suffers from basis mismatch errors, the off-grid dictionary eliminates such limitations by allowing a finer and more flexible representation of the angular information. This enhanced adaptability leads to improved estimation accuracy and to enhanced overall performance. Fig. 5(a) depicts the no. of the convergence of the proposed BGSR technique as a function of EM iterations for multiple users. To achieve a target NMSE of approximately 0.1, the algorithm requires about 8 EM iterations for a single user, but roughly 14 iterations for three users. Although the required no. of EM iterations increases from approximately 8 to 14, this growth is not strictly linear with respect to the number of users. The fact that the increment is less than a factor of **three** indicates that the dictionary-based algorithm benefits from the shared angular structure of multiple users, preventing the computational complexity from growing linearly. Hence, while the MU extension naturally expands the sparse channel space, the iterative process still leverages inter-user correlations, preventing an unmanageable increase in complexity. This sub-linear scaling highlights that the MU scenario remains computationally tractable, enabling the framework to effectively accommodate additional users without requiring an unbounded increase in iteration count. Fig. 5(b) compares the NMSE performance for different numbers of users. As the

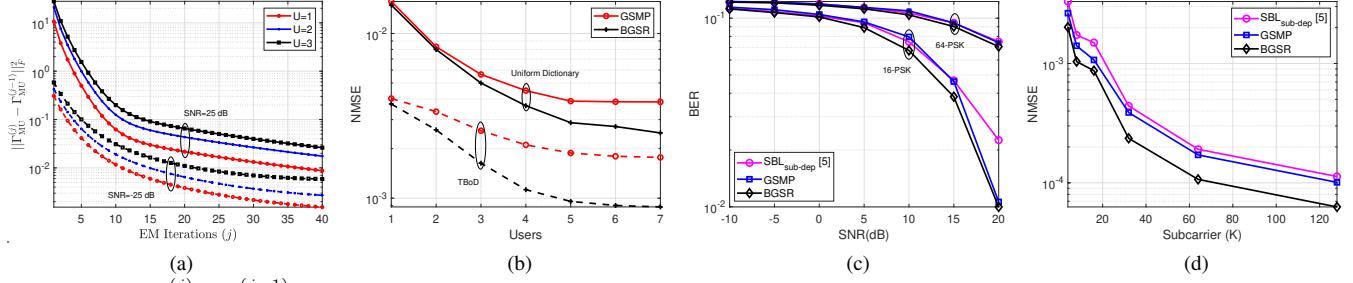


Figure 5. (a) $\|\mathbf{\Gamma}_{\text{MU}}^{(j)} - \mathbf{\Gamma}_{\text{MU}}^{(j-1)}\|_{\mathcal{F}}^2$ vs. number of EM iterations by varying the users (b) NMSE vs users for both the proposed algorithms under uniform and TBoD based dictionary constraints. (c) BER vs SNR performance comparison for the proposed GSMP and BGSR techniques with different modulation for SU scenario (d) NMSE vs subcarrier performance comparison for the GSMP and BGSR based approaches for SU scenario.

number of users increases, the NMSE performance improves due to the availability of additional measurement vectors, which enhances the estimation accuracy. Moreover, the TBoD-based dictionary outperforms the uniform dictionary, as it provides a more accurate representation of the continuous angular domains. This improvement highlights the advantage of structured dictionary learning in capturing the underlying sparsity of the THz channel, leading to more efficient channel estimation. Therefore, this sub-linear scaling ensures that the framework remains computationally feasible, while efficiently supporting MU in high-dimensional THz systems.

C. Trade-off between communication resources and computational complexity

As a special case for MU, Fig. 5(c) and (d) compares the performance for an SU scenario. All the other parameters are identical to the ones in Table-VI. Fig. 5(c) compares the BER performance of the proposed channel estimation techniques for {16, 64}-PSK modulation. It is apparent from the figure that the proposed technique outperforms the other sparse techniques for higher-order modulation as well. Moreover, the increased constellation size of say 64-PSK mandates a higher SNR due to its tighter spacing between the constellation points, leading to a higher BER. Fig. 5(d) demonstrates our performance comparison between the NMSE and the number of subcarriers. Observe that by increasing the number of subcarriers, a large number of pilot symbols can be embedded within the signal, which further helps in improving the channel estimation performance. Moreover, given the precise channel knowledge at each subcarrier, the receiver can accurately recover the data, thus leading to improved performance. Furthermore, the improved performance of the proposed algorithms is due to the fact that they leverage the unique sparsity of the channel within a group framework, enabling group channel estimation across all the subcarriers. Therefore, as the number of subcarriers K increases, the number of group sparse vectors also increases, enhancing the overall estimation performance. Additionally, the pilots are loaded onto subcarriers, meaning that the number of pilot blocks effectively increases with the number of subcarriers. Furthermore, as the number of pilot blocks increases, the NMSE performance also improves.

A clear trade-off exists between SU and MU channel estimation. SU estimation is computationally efficient due to its lower dimensionality; however, it requires sequential

processing of user channels, often over different time slots, and relies on separate pilots for each user. As a result, by the time later users are estimated, the CSI of earlier users may already be outdated, especially in dynamic environments. By contrast, joint estimation is more resource-intensive but leverages a common communication sequence and shared pilot transmission over the same time-frequency band to simultaneously estimate all user channels. This joint processing eliminates sequential delays, ensures consistency across users, and avoids performance loss due to outdated CSI. Moreover, MU estimation leverages the inherent group sparsity and inter-user correlations in THz channels, which cannot be fully exploited by independent SU estimation. Consequently, despite the higher computational cost, joint estimation achieves more reliable, up-to-date, and spectrally efficient CSI acquisition, making it highly advantageous in practical MU systems.

VII. CONCLUSIONS

New channel learning techniques have been conceived by exploiting the angular domain sparsity in practical THz hybrid MIMO systems. The THz channel was rigorously modeled considering the absorption coefficient (obtained through the HITRAN database), reflection losses, diffused ray modeling including the dual wideband effect. Furthermore, we also considered the effect of low resolution ADCs along with the practical implementation of THz hybrid MIMO systems. The i/o model was linearized using the popular Bussgang decomposition. Moreover, an SC-FDE based system model was conceived to convert the TD processing into its equivalent FD counterpart, and to provide frame-wise processing for attaining significant performance improvements. Furthermore, a novel Bayesian learning based channel learning framework i.e., BGSR, was proposed whose performance was also characterized by deriving the MU-BCRB. Moreover, the performance of the RRC-PSF and Rect-PSF based dual wideband channels was explicitly compared. Our simulation results offer empirical evidence of the enhanced performance achieved by the BGSR scheme in comparison to other sparse estimation techniques.

APPENDIX A CALCULATION OF NOISE COVARIANCE

The received signal after performing analog combining and before passing through the low-resolution ADC can be

expressed as

$$\mathbf{y}_{m,\text{MU}}(q) = \mathbf{W}_{\text{RF},m}^H \sum_{u=1}^U \mathbf{H}_{q,u} \otimes (\mathbf{F}_{\text{RF},m,u} \mathbf{a}_{m,u}^{(q)}) + \mathbf{W}_{\text{RF},m}^H \tilde{\mathbf{v}}_m(q), \quad (60)$$

Let $\mathbf{x}_{m,u}(q) = \mathbf{F}_{\text{RF},m,u} \mathbf{a}_{m,u}^{(q)}$ for notational simplicity. Moreover, $\mathbb{E}(\mathbf{x}_{m,u}(q) \mathbf{x}_{m,u}^H(q)) = \sigma_b^2 \mathbf{F}_{\text{RF},m,u} \mathbf{F}_{\text{RF},m,u}^H = \mathbf{R}_{xx,m}$. Expanding the circular convolution as defined in Section-I-C, Eq. (60) can be re-written as

$$\begin{aligned} \tilde{\mathbf{y}}_m(q) = \mathbf{W}_{\text{RF},m}^H \sum_{u=1}^U & [\mathbf{H}_u(0) \mathbf{x}_{m,u}(q) + \mathbf{H}_u(1) \mathbf{x}_{m,u}(q-1) + \\ & \cdots + \mathbf{H}_u(K-1) \mathbf{x}_{m,u}(q-K+1)] + \mathbf{W}_{\text{RF},m}^H \tilde{\mathbf{v}}_m(q). \end{aligned} \quad (61)$$

Let $\tilde{\mathbf{C}}_m \in \mathbb{C}^{N_{\text{RF}}^R \times N_{\text{RF}}^R}$ represents the covariance matrix which is defined as $\tilde{\mathbf{C}}_m = \mathbb{E}\{\tilde{\mathbf{y}}_m(q) \tilde{\mathbf{y}}_m^H(q)\}$. After substituting $\tilde{\mathbf{y}}_m(q)$ into the expression of the covariance matrix, we obtain

$$\begin{aligned} \tilde{\mathbf{C}}_m = \mathbf{W}_{\text{RF},m}^H \sum_{u=1}^U & [\mathbf{H}_u(0) \mathbf{R}_{xx,m} \mathbf{H}_u^H(0) + \cdots + \\ & \mathbf{H}_u(K-1) \mathbf{R}_{xx,m} \mathbf{H}_u^H(K-1)] \mathbf{W}_{\text{RF},m}^H + \sigma_n^2 \mathbf{W}_{\text{RF},m}^H \mathbf{W}_{\text{RF},m}. \end{aligned} \quad (62)$$

Therefore, the received signal covariance matrix before passing through low-resolution ADCs can be expressed as $\tilde{\mathbf{C}}_m = \mathbf{W}_{\text{RF},m}^H \mathbf{Q}_m \mathbf{W}_{\text{RF},m} + \sigma_n^2 \mathbf{W}_{\text{RF},m}^H \mathbf{W}_{\text{RF},m}$ where $\mathbf{Q}_m = \sum_{u=1}^U \sum_{n=0}^{K-1} \mathbf{H}_u(n) \mathbf{R}_{xx,m} \mathbf{H}_u^H(n)$. Additionally, the noise covariance matrix after passing through low-resolution ADC is given as $\mathbf{C}_m = \varepsilon(1-\varepsilon)\text{diag}(\tilde{\mathbf{C}}_m)$.

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*, pp. 3211–3221, 2011.
- [2] H. Li, M. Li, Q. Liu, and A. L. Swindlehurst, "Dynamic hybrid beamforming with low-resolution PSs for wideband mmWave MIMO-OFDM systems," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 9, pp. 2168–2181, 2020.
- [3] H. Sarieddeen, M.-S. Alouini, and T. Y. Al-Naffouri, "An overview of signal processing techniques for terahertz communications," *Proceedings of the IEEE*, vol. 109, no. 10, pp. 1628–1665, 2021.
- [4] 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*, vol. 66, no. 13, pp. 3393–3406, 2018.
- [5] A. Garg, S. Srivastava, N. Yadav, A. K. Jagannatham, and L. Hanzo, "Angularly Sparse Channel Estimation in Dual-Wideband Tera-Hertz (THz) Hybrid MIMO Systems Relying on Bayesian Learning," *IEEE Transactions on Communications*, 2024.
- [6] D. P. Wipf and B. D. Rao, "Sparse Bayesian learning for basis selection," *IEEE Transactions on Signal Processing*, vol. 52, pp. 2153–2164, 2004.
- [7] C. Lin and G. Y. Li, "Adaptive beamforming with resource allocation for distance-aware multi-user indoor TeraHertz communications," *IEEE Transactions on Communications*, vol. 63, no. 8, pp. 2985–2995, 2015.
- [8] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *Journal of Machine Learning Research*, pp. 211–244, 2001.
- [9] Y.-C. Cheng, P.-Y. Tsai, and M.-H. Huang, "Matrix-inversion-free compressed sensing with variable orthogonal multi-matching pursuit based on prior information for ECG signals," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 10, no. 4, pp. 864–873, 2016.
- [10] Y. Huang, Y. He, Q. Luo, L. Shi, and Y. Wu, "Channel estimation in MIMO-OFDM systems based on a new adaptive greedy algorithm," *IEEE Wireless Communications Letters*, vol. 8, no. 1, pp. 29–32, 2018.
- [11] H. Du, T. Ratnarajah, M. Sellathurai, and J. Chambers, "Mixed norm minimization for MIMO cellular interference channel," in *2015 IEEE 16th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*. IEEE, 2015, pp. 635–639.
- [12] K. A. Chethan and C. R. Murthy, "An iterative re-weighted minimization framework for resource allocation in the single-cell relay-enhanced OFDMA network," in *2016 IEEE 17th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2016.
- [13] S. F. Cotter, B. D. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with measurement vectors," *IEEE Transactions on Signal Processing*, vol. 53, no. 7, pp. 2477–2488, 2005.
- [14] S. Srivastava and A. K. Jagannatham, "MSBL-based simultaneous sparse channel estimation in SC wideband mmWave hybrid MIMO systems," in *IEEE Global Communications Conference*. IEEE, 2020, pp. 1–6.
- [15] G. Joseph, C. R. Murthy, R. Prasad, and B. D. Rao, "Online recovery of temporally correlated sparse signals using multiple measurement vectors," in *2015 IEEE GLOBECOM*. IEEE, 2015, pp. 1–6.
- [16] S. Srivastava, J. Nath, and A. K. Jagannatham, "Data aided quasistatic and doubly-selective CSI estimation using affine-precoded superimposed pilots in millimeter wave MIMO-OFDM systems," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 7, pp. 6983–6998, 2021.
- [17] S. Srivastava, A. Tripathi, N. Varshney, A. K. Jagannatham, and L. Hanzo, "Hybrid transceiver design for TeraHertz MIMO systems relying on Bayesian learning aided sparse channel estimation," *IEEE Transactions on Wireless Communications*, 2022.
- [18] Z. Sha and Z. Wang, "Channel estimation and equalization for TeraHertz receiver with RF impairments," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 6, pp. 1621–1635, 2021.
- [19] K. Dovelos, M. Matthaiou, H. Q. Ngo, and B. Bellalta, "Channel estimation and hybrid combining for wideband TeraHertz massive MIMO systems," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 6, pp. 1604–1620, 2021.
- [20] T.-H. Chou, N. Michelusi, D. J. Love, and J. V. Krogmeier, "Compressed Training for Dual-Wideband Time-Varying Sub-Terahertz Massive MIMO," *IEEE Transactions on Communications*, 2023.
- [21] Y. Li and A. Madhukumar, "Hybrid near-and far-field thz um-mimo channel estimation: a sparsifying matrix learning-aided bayesian approach," *IEEE Transactions on Wireless Communications*, 2024.
- [22] Y. Zhang, D. Li, D. Qiao, and L. Zhang, "Analysis of indoor THz communication systems with finite-bit DACs and ADCs," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 1, pp. 375–390, 2021.
- [23] R. Nikbakht and A. Lozano, "Terahertz transmit beamforming with 1-bit DACs and ADCs," in *2021 29th European Signal Processing Conference (EUSIPCO)*. IEEE, 2021, pp. 826–830.
- [24] K. Wang, Z. Gao, S. Chen, B. Ning, G. Chen, Y. Su, Z. Wang, and H. V. Poor, "Knowledge and Data Dual-Driven Channel Estimation and Feedback for Ultra-Massive MIMO Systems under Hybrid Field Beam Squint Effect," *IEEE Transactions on Wireless Communications*, 2024.
- [25] I.-S. Kim and J. Choi, "Spatial wideband channel estimation for mmWave massive MIMO systems with hybrid architectures and low-resolution ADCs," *IEEE Transactions on Wireless Communications*, vol. 20, no. 6, pp. 4016–4029, 2021.
- [26] Y. Ding, S.-E. Chiu, and B. D. Rao, "Bayesian channel estimation algorithms for massive MIMO systems with hybrid analog-digital processing and low-resolution ADCs," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 3, pp. 499–513, 2018.
- [27] C. Mollen, J. Choi, E. G. Larsson, and R. W. Heath, "One-bit ADCs in wideband massive MIMO systems with OFDM transmission," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2016, pp. 3386–3390.
- [28] S. Priebe, M. Jacob, and T. Kuerner, "AoA, AoD and ToA characteristics of scattered multipath clusters for THz indoor channel modeling," in *17th European Wireless -Sustainable Wireless Technologies*, 2011, pp. 1–9.
- [29] J. P. González-Coma, J. Rodriguez-Fernandez, N. González-Prelcic, L. Castedo, and R. W. Heath, "Channel estimation and hybrid precoding for frequency selective multiuser mmWave MIMO systems," *IEEE Journal of Selected Topics in Signal Processing*, pp. 353–367, 2018.
- [30] L. S. Rothman, I. E. Gordon, A. Barbe, D. C. Benner, P. F. Bernath, M. Birk, V. Boudon, L. R. Brown, A. Campargue, J.-P. Champion *et al.*, "The HITRAN 2008 molecular spectroscopic database," *Journal of Quantitative Spectroscopy and Radiative Transfer*, pp. 533–572, 2009.
- [31] L. Fan, S. Jin, C.-K. Wen, and H. Zhang, "Uplink achievable rate for massive MIMO systems with low-resolution ADC," *IEEE Communications Letters*, vol. 19, no. 12, pp. 2186–2189, 2015.
- [32] A. Mezghani and J. A. Nossek, "Capacity lower bound of MIMO channels with output quantization and correlated noise," in *Proc. IEEE Int. Symp. Inf. Theory*, 2012, pp. 1–5.
- [33] B. Liu, W. Tan, H. Hu, and H. Zhu, "Hybrid beamforming for mmWave MIMO-OFDM system with beam squint," in *2018 IEEE 29th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*. IEEE, 2018, pp. 1422–1426.

- [34] R. Piesiewicz, C. Jansen, D. Mittleman, T. Kleine-Ostmann, M. Koch, and T. Kürner, "Scattering analysis for the modeling of THz communication systems," *IEEE Transactions on Antennas and Propagation*, vol. 55, no. 11, pp. 3002–3009, 2007.
- [35] J. Rodríguez-Fernández, N. González-Prelicic, K. Venugopal, and R. W. Heath, "Frequency-domain compressive channel estimation for frequency-selective hybrid millimeter wave MIMO systems," *IEEE Transactions on Wireless Communications*, pp. 2946–2960, 2018.
- [36] D. P. Wipf and B. D. Rao, "An empirical bayesian strategy for solving the simultaneous sparse approximation problem," *IEEE Transactions on Signal Processing*, vol. 55, no. 7, pp. 3704–3716, 2007.
- [37] R. Prasad, C. R. Murthy, and B. D. Rao, "Joint channel estimation and data detection in MIMO-OFDM systems: A sparse Bayesian learning approach," *IEEE Transactions on Signal Processing*, 2015.
- [38] J. Mo, P. Schniter, and R. W. Heath, "Channel estimation in broadband millimeter wave MIMO systems with few-bit ADCs," *IEEE Transactions on Signal Processing*, vol. 66, no. 5, pp. 1141–1154, 2017.
- [39] C. Ekanadham, D. Tranchina, and E. P. Simoncelli, "Recovery of sparse translation-invariant signals with continuous basis pursuit," *IEEE transactions on signal processing*, vol. 59, no. 10, pp. 4735–4744, 2011.
- [40] R. Piesiewicz, C. Jansen, S. Wietzke, D. Mittleman, M. Koch, and T. Kürner, "Properties of building and plastic materials in the THz range," *International Journal of Infrared and Millimeter Waves*, 2007.
- [41] K. Venugopal, A. Alkhateeb, N. G. Prelicic, and R. W. Heath, "Channel estimation for hybrid architecture-based wideband millimeter wave systems," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 9, pp. 1996–2009, 2017.