

Terahertz Beamforming and Group Sparse Channel Estimation Relying on Low-Resolution ADCs in MU Hybrid MIMO systems

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

Abstract—A unified beamforming and channel estimation framework relying on Bayesian learning is conceived. Recognizing the limitations imposed by low-resolution analog-to-digital converter (ADCs) and frequency-dependent propagation effects occurring in the Terahertz (THz) band, we formulate a dual-wideband channel model incorporating root raised cosine (RRC) pulse shaping. To address the non-linear distortions introduced by low-resolution ADCs, Bussgang decomposition is employed, leading to a tractable linearized inference process. By leveraging the shared sparsity inherent in a multi-user (MU) scenario of THz systems, we propose a Hierarchical Bayesian Group-sparse Regression (HBG-SR) based channel learning technique that exploits the *group-sparse* structure of THz band channels. The estimated dominant angle-of-arrival/ angle-of-departure (AoA/AoD) indices are then exploited for appropriately configuring the true-time-delay (TTD) elements in the hybrid transceiver, enabling precise beam alignment across subcarriers and the effective compensation of the beam-squint effect occurring in wideband THz systems. Extensive simulation results validate the efficiency of the proposed channel estimator and the TTD-aided beamforming architecture, highlighting their robustness and performance gains under practical wideband THz system constraints.

Index Terms—TeraHertz, dual-wideband, multi-user, Bayesian

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. The work of S. Srivastava was supported 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 and A. K. Jagannatham jointly acknowledge the funding support provided by Anusandhan National Research Foundation's Advanced Research Grant ANRF/ARG/2025/005895/ENS. The work of Lajos Hanzo is supported by Engineering and Physical Sciences Research Council (EPSRC) projects is gratefully acknowledged: Platform for Driving Ultimate Connectivity (TITAN) (EP/X04047X/1; EP/Y037243/1); Robust and Reliable Quantum Computing (RoARQ, EP/W032635/1); PerCom (EP/X012301/1); India-UK Intelligent Spectrum Innovation ICON UKRI-1859. S. Srivastava, A. K. Jagannatham and Lajos 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).

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

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

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)

learning, Cramér-Rao bound, true-time-delay

I. INTRODUCTION

The THz band, spanning from (0.1–10) THz, offers spectral windows characterized by bandwidths of 10 GHz or higher [1], thus standing out as a promising avenue of delivering ultra-high data rates and massive short-range connectivity in next-generation wireless communications. In comparison to the mmWave band that spans 30 – 100 GHz [2], the challenges in this band include having fewer spatial degrees-of-freedom (SDoF) [3], high probability of blockages and the spatial wideband effect. The short wavelength of the THz signals leads to variation in the delay across different antenna elements within the array, resulting in the deleterious spatial wideband effect [4]. Furthermore, the 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. In order to avoid this loss, it is critical to incorporate the beam-squint effect in the analog beamformer/ combiner design procedure, which poses a formidable challenge. To address this impediment, this treatise develops novel hybrid transceiver design algorithms for multi-user (MU) THz MIMO systems, while considering the specific scattering characteristics of dual-wideband THz channels [4]. It is also worth noting that accurate knowledge of the channel state information (CSI) is crucial for the design of the transmit precoder/ receive combiner (TPC/ RC) pair and for beam-squint mitigation. Therefore, the development of channel learning techniques for a spatial-wideband THz channels is a worthwhile endeavor, which is addressed in this paper. Table-I summarizes the acronyms used in this work.

To fully exploit the capabilities of the THz band, selecting a waveform tailored to THz-specific propagation and hardware constraints is essential. Effective waveform design must jointly consider performance metrics such as peak-to-average power ratio (PAPR), baseband computational complexity, and sensitivity to hardware and RF impairments [3]. In this context,

Table I
LIST OF ACRONYMS

ADC	Analog-to-digital converter
AoA/ AoD	Angle of arrival/ angle of departure
AQNM	Additive quantization noise model
AWGN	Additive white Gaussian noise
BCRLB	Bayesian Cramér Rao lower bound
BER	Bit error rate
CS	Compressive sensing
CSI	Channel state information
DPP	Delay phase precoding
EM	Expectation maximization
FFT	Fast Fourier transform
GMM	Gaussian mixture model
GSOMP	Generalized simultaneous orthogonal matching pursuit
HBG-SR	Hierarchical Bayesian Group-sparse Regression
HITRAN	High-resolution transmission molecular absorption
IFFT	Inverse fast Fourier transform
IRS	Intelligent reflecting surface
LASSO	Least absolute shrinkage and selection operator
MC	Multi carrier
MFOCUSS	MMV Focal underdetermined system solver
MMSE	Minimum mean squared error
MMV-LS	Multiple measure vector least squares
MMV-SBL/ MSBL	Multiple measure vector sparse Bayesian learning
MP	Message passing
NLoS	Non-line-of-sight
NMSE	Normalized mean square error
NOMA	Non-orthogonal multiple access
OFDM	Orthogonal frequency division multiplexing
PAPR	Peak-to-average power ratio
QoS	Quality of service
RAs	Receiving antennas
RC	Receive combiner
RRC	Root raised cosine
SBL	Sparse Bayesian learning
SC-FDE	Single carrier frequency domain equalization
SDoF	Spatial degrees-of-freedom
SE	Spectral efficiency
SMV	Single measurement vector
SNR	Signal to noise ratio
SVD	Singular value decomposition
TAs	Transmit antennas
TD	Time delay
TPC	Transmit precoder
TTD	True-time delay

the preference between SC and multi-carrier (MC) waveforms remains actively debated for THz systems. The adoption of SC modulation in the first sub-THz communication standard (IEEE 802.15.3d) [6] reinforces its suitability for this regime, as it aligns well with stringent hardware constraints. In particular, single carrier- frequency domain equalization (SC-FDE) architectures provide lower transmitter-side complexity than orthogonal frequency division multiplexing (OFDM) and inherently achieve reduced PAPR, thereby easing the burden on power amplifiers. Furthermore, in the presence of low-resolution ADCs, OFDM-based systems lose subcarrier orthogonality, which significantly degrades OFDM performance and motivates the adoption of SC transmission in this work. Moreover, SC-FDE exhibits improved resilience to THz-specific impairments such as beam misalignment and beam squint [7], making it a strong candidate for robust and energy-efficient THz communication. Table-V in [7] provides a systematic comparison between SC and MC waveforms and the results demonstrate that SC-FDE achieves a favorable balance between complexity, PAPR performance, and robustness to beam-squint impairments, further motivating its consideration as a practical waveform for THz communications. The next section will discuss the exiting state-of-the-art techniques.

A. Review of existing works

The THz band is characterized by severe propagation losses, including molecular absorption, reflection losses, and path loss, which pose significant challenges to reliable communication. In a seminal contribution, Jornet and Akyildiz [1] developed a comprehensive THz channel model that incorporates both propagation and molecular absorption effects. Their work investigates the capacity of THz systems under various molecular compositions and pulse shapes using power allocation schemes. In parallel, Piesiewicz *et al.* [8] explored the impact of reflections in indoor THz scenarios by measuring the reflection coefficients of common building materials using Kirchhoff scattering theory. Their findings highlight that scattering significantly influences both signal power and propagation behavior in the THz band. Priebe *et al.*, [9], in their pioneering treatise, model a double directional indoor THz channel, incorporating domains for the angle of AoA/AoD and time of arrival (ToA).

In their cutting-edge work, Lin and Li [10] proposed a *distance-aware* adaptive THz beamforming scheme, where long-distance users are allocated the central part of the spectrum, while short-distance users are assigned sidebands to ensure quality of service (QoS). Their hybrid beamforming technique does not account for the beam-squint effect, which is severe in the THz domain. Wan *et al.*, [11] proposed a hybrid transceiver design based on the alternating direction method of multipliers (ADMM), which compensates for the beam-squint effect in MIMO-aided OFDM systems. However, their framework is designed for the mmWave/sub-THz regime. Gao *et al.*, [12] proposed hybrid transceiver design approaches based on a virtual sub-array and true-time-delay (TTD) to compensate for the beam-squint effect. The virtual sub-array approach relies on beam broadening, while the TTD-aided approach replaces all the phase shifters with a few TTD lines. Dai *et al.*, [13], in their cutting-edge work, developed a delay phase precoding (DPP)-based beamforming architecture. This architecture, mitigates the beam-squint effect by introducing a small number of time delay (TD) elements along with the PSs, providing frequency-dependent phase shifts to alleviate the array gain loss caused by the beam-squint effect. However, none of the papers mentioned above considers a MU scenario, which is critical in practical systems. Moreover, the aforementioned works do not consider the impact of employing low-resolution ADCs. This choice makes detection more challenging at the receiver, particularly due to the beam-squint effect [4]. Furthermore, their beamformer designs are based on utilizing the true channel frequency response, which is unavailable in practical scenarios.

Omid *et al.*, [14] proposed a novel hybrid beamforming design, which incorporates the CSI uncertainty in intelligent reflecting surfaces (IRS) assisted non-orthogonal multiple access (NOMA) systems. Jafri *et al.*, [15] proposed a distinctive hybrid beamformer design in the context of a cooperative scenario for mmWave hybrid MIMO systems. Their work underscores the importance of considering realistic CSI uncertainty in cooperative scenarios, which is crucial in the beamformer design. Although, these works consider

CSI uncertainty, they do not actually estimate the channel. But again, for the design of a practical beamformer, precise channel knowledge is crucial.

Moreover, due to the highly directional nature of propagation attained by large-scale antennas, the THz band results in a sparsely populated multipath channel in the angular domain. The problem of finding sparse solutions from a single measurement vector (SMV) has received extensive attention in prior research. In the family of SMV models, sparse Bayesian learning (SBL) is one of the most popular choices due to the global convergence of its cost function [16]. Wipf and Rao [16], in their seminal work, initially introduced the SBL algorithm, specifically designed for SMV problems and subsequently expanded the scope of the SBL algorithm to multiple measurement vector (MMV) problems offering a more comprehensive solution for dealing with complex data sets. Dovelos *et al.*, [17] proposed a generalized simultaneous orthogonal matching pursuit (GSOMP) based technique for estimating the dual-wideband THz channel by considering a system with a single user (SU) relying on a single antenna. Chou *et al.*, [18] proposed an MMV least squares (MMV-LS-CV)-based CSI learning technique for estimating the dual-wideband THz channel in hybrid MIMO-OFDM systems. Although their pioneering work incorporates time selectivity, the LS based solution advocated therein yields a poor performance in comparison to SBL. Kim and Choi [19] proposed a Newtonized FCFGs-CV-based channel estimator for wideband mmWave massive MIMO systems having hybrid architectures and low-resolution ADCs. To account for the propagation delay across the antenna array, the discrete-time channel is formulated by incorporating the spatial-wideband effect. However, their formulation is based on a time-domain channel model, which leads to significant complexity in wideband scenarios. Note that, existing works [17], [18], [19] have addressed wideband effects in mmWave/ sub-THz/THz systems from different perspectives; however, their scope remains limited when compared to the present study. Specifically, both [17] and [18] consider dual-wideband channel effects, accounting for the joint frequency selectivity arising from large bandwidths and large antenna arrays. However, these works primarily focus on single-user transmission scenarios and concentrate on channel estimation or training design. On the other hand, [19] investigates spatial-wideband channel estimation under hybrid architectures with low-resolution ADCs, but is limited to mmWave systems and does not account for dual-wideband effects, MU transmission or transceiver design. In contrast, the present work jointly considers MU MIMO transmission in dual-wideband THz systems, incorporates spatial-wideband effects together with low-resolution ADC constraints, and exploits group sparsity for CSI learning, which is further leveraged for beamspace-aware hybrid transceiver design. When all these impairments are jointly considered, the conventional linear assumption becomes invalid, since these effects interact across dimensions in a nonlinear fashion rather than simply adding up.

Cui *et al.* [20] in their cutting-edge work proposed a Bayesian regression-based channel estimation framework for dense antenna systems from an information-theoretic perspec-

ive. Due to the extremely high spatial correlation inherent in dense arrays, the channel prior covariance significantly deviates from an identity matrix. Therefore, the authors designed the *observation matrix* by aligning it with the dominant subspace of the posterior kernel, thereby minimizing the estimation error through covariance-aware sensing. Importantly, the estimation itself is performed via the posterior mean, i.e., the MMSE estimator, and the core contribution lies in the covariance-driven design of the observation matrix. In contrast, our proposed framework differs fundamentally in both modeling and algorithmic structure. Firstly, instead of estimating a full covariance kernel, we impose a parameterized Gaussian prior directly on the channel coefficients, where sparsity is learned through hyperparameters. This leads to a hierarchical Bayesian inference framework that explicitly exploits structured sparsity rather than covariance alignment. Moreover, the observation matrix in our work is not designed by matching a posterior covariance subspace; instead, it is dictated by practical hybrid analog architectures by phase-only controllable combiners. Additionally, our framework is applicable to hybrid MU-THz systems and accommodates practical hardware constraints such as low-resolution ADCs and beam squint effects. Therefore, we develop a novel CSI acquisition technique that exploits the *group sparsity* naturally arising in MU THz systems, followed by the design of a TTD-based hybrid transceiver that accounts for the effects of low-resolution ADCs. Table II boldly contrasts our contributions, which are highlighted in more detail below.

B. Contributions

- 1) We commence by developing a practical MU dual-wideband THz channel model encompassing both frequency and distance-dependent characteristics including free-space, absorption and reflection losses. Furthermore, it integrates the effect of multipath delays and transmit pulse shaping, which is absent in [17], [18]. Additionally, we incorporate the modeling of diffused rays, which are present in the THz band, yet they are not considered in the existing literature [21]. Moreover, the AoA/AoDs are generated using Gaussian Mixture Model (GMM) [10], and a practical THz angle modeling technique that provides an accurate representation of the THz channel.
- 2) We propose a novel Hierarchical Bayesian Group-sparse Regression (HBG-SR)-based channel learning framework, where the BS jointly processes the signals received from all users and the channel exhibits a shared sparse structure. Owing to the shared spatial geometry across subcarriers, the channel inherently possesses a group-sparse structure that can be exploited for reliable CSI recovery. Accordingly, the proposed HBG-SR technique effectively leverages this group sparsity in the concatenated beamspace channel.
- 3) To benchmark the performance of the proposed CSI learning technique, we furthermore derive the Bayesian Cramér-Rao lower bound (BCRLB) for the MU THz scenario. This serves as an excellent benchmark for CSI estimation performance analysis.

Table II
CONTRASTING OUR CONTRIBUTIONS TO THE EXISTING LITERATURE

Features	[5]	[10]	[12]	[13]	[17]	[18]	[19]	[21]	[2]	This Paper
Sub-THz/ THz Band	✓	✓	✓	✓	✓	✓	✓	✓		✓
Reflection & molecular absorption losses	✓	✓			✓	✓				✓
Diffused-ray modeling	✓									✓
SC-FDE/ SC-FDMA system	✓								✓	✓
AoA/AoD with GMM		✓								✓
BCRLB	✓				✓			✓		✓
Low Resolution ADCs with spatial-wideband							✓			✓
CSI learning in THz domain	✓				✓	✓	✓	✓		✓
Transceiver design in THz domain		✓	✓	✓	✓					✓
Group sparsity based CSI learning										✓
MU-MIMO in dual-wideband THz system										✓
RRC-PSF and Rect-PSF based dual-wideband channel										✓
Beamspace-aware transceiver design with estimated angular support										✓

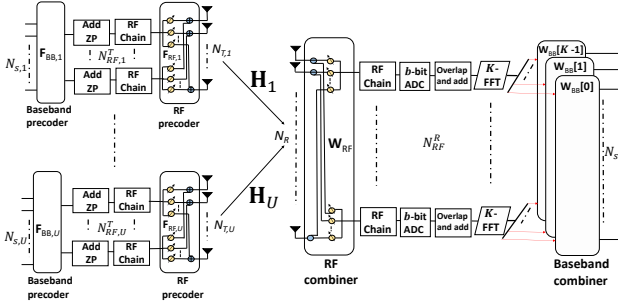


Figure 1. Schematic diagram of SC-FDE based MU THz hybrid MIMO systems with low-resolution ADCs

- 4) Moreover, the effect of low-resolution ADCs is incorporated into the MU-THz system to account for practical quantization constraints. Additionally, in the THz band, the beam-squint effect induces frequency-dependent signal variations across the antenna array. When combined with coarse quantization, this poses substantial challenges for reliable signal reconstruction and decoding. We address this challenge in this work.
- 5) While TTD-based architectures have been widely studied for beam-squint mitigation, existing studies largely focus on SU scenarios [13], [12], [17]. However, in MU systems, the shared analog TTD network must simultaneously serve users roaming in different directions, leading to different delay requirements which results in *unavoidable residual beam squint*. Moreover, in wideband THz systems using low-resolution ADCs, this residual frequency-dependent array gain directly translates into non-uniform quantization distortion, since the quantization noise power scales with the ADC input variance. Therefore, this work addresses the joint delay compensation for all the users by conditioning the wideband signal prior to quantization to mitigate the joint impact of residual beam squint and coarse quantization.
- 6) Our simulation results quantify the enhanced performance of the proposed CSI learning and the TTD based hybrid transceiver design both in terms of normalized mean square error (NMSE), bit error rate (BER), and spectral efficiency (SE).

C. Notations

Upper-case letters \mathbf{A} represent a matrix, while lower-case letters \mathbf{a} represent a vector. The quantity $\|\cdot\|_{\mathcal{F}}$ represents the Frobenius norm, while \otimes represents the Kronecker product and $\text{SVD}(\cdot)$ represent singular value decomposition. The operation $\mathcal{C}(\cdot)$ denotes the column space, and $\mathcal{R}(\cdot)$ denotes the row space of a matrix. The identity $\text{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A})\text{vec}(\mathbf{B})$ illustrates the $\text{vec}(\cdot)$ operation, and $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ represents a complex Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. The Dirichlet function is defined as $\Xi_N(x) = \frac{\sin(\frac{N\pi}{2}x)}{\sin(\frac{\pi}{2}x)}$.

II. SC-FDE BASED MU THz HYBRID MIMO SYSTEM MODEL

Consider the uplink of a SC wideband THz hybrid MIMO system, where a BS is equipped with N_{BS} receiving antennas (RAs), serving U users simultaneously. Each user is equipped with N_u transmit antennas (TAs), and the antennas of both the BS and user equipment (UEs) are arranged in a uniform linear array (ULA). Let all the users possess equal number of antennas and therefore the total number of TAs is $N_T = \sum_{u=1}^U N_u = UN_u$. Let the THz MIMO system facilitate the transmission of N_s data streams, where each user possesses $N_{s,u}$ data streams. Furthermore, in our consideration of a practical scenario, the BS is equipped with N_{RF}^B radio frequency (RF) chains, while each user is equipped with N_{RF}^u RF chains and follows the relationship of $N_s \leq \sum_{u=1}^U N_{RF}^u \leq N_{RF}^B \ll N_{BS}$ and $N_{RF}^u \ll N_u$. At the UEs, the wideband THz MIMO system comprises a frequency-flat baseband TPC $\mathbf{F}_{BB,u} \in \mathbb{C}^{N_{RF}^u \times N_{s,u}}$ coupled with the frequency-flat RF TPC $\mathbf{F}_{RF,u} \in \mathbb{C}^{N_u \times N_{RF}^u}$ for each user. Furthermore, the BS has a bank of frequency-selective hybrid combiners $\mathbf{W}_{BB}[k] \in \mathbb{C}^{N_{RF}^B \times N_s}$, $\forall 0 \leq k \leq K-1$, where K represents the number of subcarriers. This is cascaded with the frequency-flat RF RC $\mathbf{W}_{RF} \in \mathbb{C}^{N_{BS} \times N_{RF}^B}$ as shown in Fig. 1. Moreover, the elements of the RF TPC and RC matrices obey a unit-magnitude constraint and are set as $|\mathbf{F}_{RF,u}(\ell, \kappa)| = \frac{1}{\sqrt{N_u}}$, $|\mathbf{W}_{RF}(\ell, \kappa)| = \frac{1}{\sqrt{N_{BS}}} \forall \ell, \kappa$. Let us further consider a wideband THz channel consisting of D delay taps at each user, where the d -th delay tap can be expressed by a complex matrix $\mathbf{H}_{d,u} \in \mathbb{C}^{N_{BS} \times N_u}$, $\forall 0 \leq d \leq D-1$. As depicted in Fig. 2(a), each transmit frame is divided into two phases, which includes the training and data phase. The training frame



Figure 2. Frame structure utilized for wideband MU THz hybrid MIMO system using SC-FDE

consists of M blocks, where each block comprises P pilot vectors. Let $\mathbf{b}_{m,u}^{(p)} \in \mathbb{C}^{N_{RF}^u \times 1}$ denote the p -th complex pilot vector, where the relationship, $0 \leq p \leq P-1$, follows for each m -th block and u -th user. Before transmitting the $\mathbf{b}_{m,u}^{(p)}$ pilot sequences on an individual RF chain, $D-1$ zeros are appended to each block, resulting in a zero-padded (ZP) block of length $K = P + D - 1$, which is given by $\{\mathbf{g}_{m,u}^{(q)}\}_{q=0}^{K-1} = \{\mathbf{g}_{m,u}^{(0)}, \dots, \mathbf{g}_{m,u}^{(P-1)}, \mathbf{0}, \dots, \mathbf{0}\}$ [2]. In a similar vein, the THz

MIMO channel $\mathbf{H}_{d,u}$ at each user u is appended with a zero-matrix of size $N_{BS} \times N_u$ to have length K , which is given as $\{\mathbf{H}_{q,u}\}_{q=0}^{K-1} = \{\mathbf{H}_{0,u}, \dots, \mathbf{H}_{D-1,u}, \mathbf{0}, \dots, \mathbf{0}\}$. It is worth

noting that although the overall system model is decoupled into different frequency bins at the receiver, interestingly, at the transmitter, it can be viewed as a single carrier system. This is due to the fact that there is no inverse FFT (IFFT) block at the transmitter and the precoding operation is applied uniformly across the entire bandwidth, as a result, the digital precoders at the user side are naturally frequency independent, consistent with the single-carrier transmission structure. Therefore, the received signal $\mathbf{r}_m(q) \in \mathbb{C}^{N_{BS} \times 1}$ corresponding to all the U users is formulated as

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

where $\mathbf{v}_m(q)$ represents the complex additive white Gaussian noise (AWGN) following the distribution $\mathcal{CN}(\mathbf{0}_{N_{BS} \times 1}, \sigma_n^2 \mathbf{I}_{N_{BS}})$. The quantity \otimes_K represents the circular convolution of length K , is defined in Section I-C. Furthermore, the received signal $\mathbf{r}_m(q)$ is passed through the RF RC followed by a b -bit quantizer denoted as $\mathcal{Q}(\cdot)$ which is given by

$$\tilde{\mathbf{y}}_m(q) = \mathcal{Q} \left(\mathbf{W}_{RF,m}^H \sum_{u=1}^U \mathbf{H}_{q,u} \otimes_K (\mathbf{F}_{RF,m,u} \mathbf{F}_{BB,m,u} \mathbf{g}_{m,u}^{(q)}) + \mathbf{W}_{RF,m}^H \mathbf{v}_m(q) \right). \quad (1)$$

A widely adopted technique for modeling the quantization effect in the wireless system involves the use of a mid-rise quantizer [22]. However, in THz systems operating with high bandwidth and a large number of antennas, the mid-rise quantizer becomes computationally intensive and does not scale efficiently [3]. To address this, we employ the Bussgang decomposition [23], which offers a tractable and analytically

Table III
 ρ FOR DIFFERENT ADC BITS b [25]

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

convenient approximation of the quantization nonlinearity. The received signal $\tilde{\mathbf{y}}_m(q) \in \mathbb{C}^{N_{RF}^B \times 1}$ after applying the Bussgang decomposition can be approximated as

$$\tilde{\mathbf{y}}_m(q) \approx \mathbf{Q} \mathbf{W}_{RF,m} \sum_{u=1}^U \mathbf{H}_{q,u} \otimes_K \tilde{\mathbf{g}}_{m,u}^{(q)} + \mathbf{Q} \mathbf{W}_{RF,m}^H \mathbf{v}_m(q) + \tilde{\mathbf{v}}, \quad (2)$$

where $\tilde{\mathbf{g}}_{m,u}^{(q)} = \mathbf{F}_{RF,m,u} \mathbf{F}_{BB,m,u} \mathbf{g}_{m,u}^{(q)}$ represents the equivalent transmit vector. The quantization matrix $\mathbf{Q} \in \mathbb{C}^{N_{RF}^B \times N_{RF}^B}$ is diagonal in nature [24], assuming that each RF chain is quantized independently. Moreover, the matrix \mathbf{Q} is expressed as $\mathbf{Q} = \varkappa \mathbf{I}_{N_{RF}^B}$, where $\varkappa = 1 - \rho$ and ρ denotes the inverse of the signal to noise power ratio (SNR). Table-III presents the values of ρ corresponding to each quantization bit for $b \leq 5$. For $b > 5$, ρ is approximated by [25] $\frac{\pi\sqrt{3}}{2} 2^{-2b}$.

Let $\mathbb{E}\{\mathbf{g}_{m,u}^{(p)} (\mathbf{g}_{m,u}^{(p)})^H\} = \sigma_d^2 \mathbf{I}_{N_{s,u}}$, where σ_d^2 represents the pilot power. The quantized noise vector $\tilde{\mathbf{v}}$ follows the distribution $\mathcal{CN}(\mathbf{0}_{N_{RF}^B \times 1}, \mathbf{J}_m)$, where $\mathbf{J}_m \in \mathbb{C}^{N_{RF}^B \times N_{RF}^B}$ represents the quantized noise covariance matrix. Its expression is detailed in Appendix-A. For notational simplicity, let $\tilde{\boldsymbol{\eta}}_m(q) = \mathbf{Q} \mathbf{W}_{RF,m}^H \tilde{\mathbf{v}}_m(q) + \tilde{\mathbf{v}} \in \mathbb{C}^{N_{RF}^B \times 1}$ represent the equivalent noise. Therefore, the equivalent noise covariance $\mathbf{P}_{\eta\eta} = \mathbb{E}\{\tilde{\boldsymbol{\eta}}_m(q) \tilde{\boldsymbol{\eta}}_m^H(q)\} \in \mathbb{C}^{N_{RF}^B \times N_{RF}^B}$ can be formulated as $\mathbf{P}_{\eta\eta} = \varkappa^2 \sigma_n^2 \mathbf{W}_{RF,m}^H \mathbf{W}_{RF,m} + \mathbf{J}_m$. Note that in wideband MU, hybrid MIMO systems employing independent scalar ADCs, the Bussgang gain reduces to a per-RF chain attenuation and is therefore modeled as diagonal, consistent with prior works such as [26], [27]. This follows from the fact that, under the adopted scalar ADC Bussgang approximation, the gain is determined by the marginal second order statistics of each ADC input, whereas the effects of MU interference, wideband propagation, and analog combining are reflected in the pre ADC signal statistics rather than in the gain term. After applying the K -point fast Fourier transform (FFT) to the received signal $\{\tilde{\mathbf{y}}_m[k]\}_{k=0}^{K-1} = \text{FFT}(\{\tilde{\mathbf{y}}_m(q)\}_{q=0}^{K-1})$, one obtains

$$\tilde{\mathbf{y}}_m[k] \approx \mathbf{Q} \mathbf{W}_{RF,m}^H \sum_{u=1}^U \mathbf{H}_u[k] \tilde{\mathbf{g}}_{m,u}[k] + \tilde{\boldsymbol{\eta}}_m[k], \quad (3)$$

where $\{\tilde{\mathbf{g}}_{m,u}[k]\}_{k=0}^{K-1} = \text{FFT}(\{\tilde{\mathbf{g}}_{m,u}^{(q)}\}_{q=0}^{K-1})$ represents the FFT of the equivalent transmit vector, while $\{\tilde{\boldsymbol{\eta}}_m[k]\}_{k=0}^{K-1} = \text{FFT}(\{\tilde{\boldsymbol{\eta}}_m(q)\}_{q=0}^{K-1})$ represents the FFT of equivalent noise. Finally, the received signal $\mathbf{y}_m[k] \in \mathbb{C}^{N_s \times 1}$ after baseband combining $\mathbf{W}_{BB,m}[k] \in \mathbb{C}^{N_{RF}^B \times N_s}$ is given by

$$\mathbf{y}_m[k] \approx \mathbf{W}_{BB,m}^H[k] \mathbf{Q} \mathbf{W}_{RF,m}^H \mathbf{H}_U[k] \tilde{\mathbf{F}}_{RF,m} \tilde{\mathbf{F}}_{BB,m} \tilde{\mathbf{g}}_m[k] + \mathbf{W}_{BB,m}^H[k] \tilde{\boldsymbol{\eta}}_m[k], \quad (4)$$

where $\mathbf{H}_U[k] = [\mathbf{H}_1[k] \mathbf{H}_2[k] \dots \mathbf{H}_U[k]] \in \mathbb{C}^{N_{BS} \times N_T}$ represents the concatenated channel for all users. The quantities $\tilde{\mathbf{F}}_{RF,m} \in \mathbb{C}^{N_T \times N_{RF}^U}$ and $\tilde{\mathbf{F}}_{BB,m} \in \mathbb{C}^{N_{RF}^U \times N_s}$ represent the overall RF TPC and baseband TPC, respectively, for all the users and are given as

$\check{\mathbf{F}}_{\text{RF},m} = \text{blkdiag}(\mathbf{F}_{\text{RF},m,1}, \dots, \mathbf{F}_{\text{RF},m,U})$ and $\check{\mathbf{F}}_{\text{BB},m} = \text{blkdiag}(\mathbf{F}_{\text{BB},m,1}, \dots, \mathbf{F}_{\text{BB},m,U})$. Moreover, the stacked pilot vector is $\check{\mathbf{g}}_m[k] = [\mathbf{g}_{m,1}^T[k], \mathbf{g}_{m,2}^T[k], \dots, \mathbf{g}_{m,U}^T[k]] \in \mathbb{C}^{N_s \times 1}$ [28]. After applying the $\text{vec}(\cdot)$ operator as described in Section-I-C, the above equation can be re-written as

$$\mathbf{y}_m[k] \approx (\check{\mathbf{g}}_m^T[k] \check{\mathbf{F}}_{\text{BB},m}^T \check{\mathbf{F}}_{\text{RF},m}^T) \otimes (\mathbf{W}_{\text{BB},m}^H[k] \mathbf{Q} \mathbf{W}_{\text{RF},m}^H) \text{vec}(\mathbf{H}_{\mathcal{U}}[k]) + \mathbf{W}_{\text{BB},m}^H[k] \check{\boldsymbol{\eta}}_m[k], \quad (5)$$

where $\text{vec}(\mathbf{H}_{\mathcal{U}}[k]) = \mathbf{h}_{\mathcal{U}}[k] \in \mathbb{C}^{N_{BS} N_T \times 1}$ represents the concatenated vectorized channel, while $\check{\Psi}_{\mathcal{U},m}[k] = (\check{\mathbf{g}}_m^T[k] \check{\mathbf{F}}_{\text{BB},m}^T \check{\mathbf{F}}_{\text{RF},m}^T) \otimes (\mathbf{W}_{\text{BB},m}^H[k] \mathbf{Q} \mathbf{W}_{\text{RF},m}^H) \in \mathbb{C}^{N_s \times N_{BS} N_T}$ represents the concatenated sensing matrix for all users. Let the combined effective noise be $\boldsymbol{\eta}_m[k] = \mathbf{W}_{\text{BB},m}^H[k] \check{\boldsymbol{\eta}}_m[k] \in \mathbb{C}^{N_s \times 1}$, which follows the distribution $\mathcal{CN}(\mathbf{0}_{N_s \times 1}, \tilde{\mathbf{P}}_{\eta\eta})$ with $\tilde{\mathbf{P}}_{\eta\eta} = \mathbf{W}_{\text{BB},m}^H[k] \mathbf{P}_{\eta\eta} \mathbf{W}_{\text{BB},m}^H[k]$. Then, the sum spectral efficiency can be expressed as

$$\text{SE} = \frac{1}{K} \sum_{k=1}^K \log_2 |\mathbf{I}_{N_s} + \frac{1}{N_s} \tilde{\mathbf{P}}_{\eta\eta}^{-1} \mathbf{H}_{\text{eff},m}[k] \mathbf{H}_{\text{eff},m}^H[k]|, \quad (6)$$

where $\mathbf{H}_{\text{eff},m} = \mathbf{W}_m^H[k] \mathbf{H}_{\mathcal{U}}[k] \mathbf{F}_m$ and $\mathbf{W}_m[k] = \mathbf{W}_{\text{RF},m}[k] \mathbf{Q} \mathbf{W}_{\text{BB},m}$ while $\mathbf{F}_m = \check{\mathbf{F}}_{\text{RF},m} \check{\mathbf{F}}_{\text{BB},m}$. The next section will describe the formulation of dual-wideband THz MIMO channel model.

III. MU THZ DUAL-WIDEBAND CHANNEL MODEL

In general, a frequency-dependent array response vector is formulated as [2]

$$\tilde{\mathbf{a}}(\theta, f) = \frac{1}{\sqrt{N}} [1, e^{-j \frac{2\pi}{\lambda} d_c \sin \theta}, \dots, e^{-j \frac{2\pi}{\lambda} d_c (N-1) \sin \theta}]^T, \quad (7)$$

where $\lambda = \frac{c}{f}$ represents the wavelength, d_c denotes the inter-antenna spacing typically set as $d_c = \frac{c}{2f_c}$, while f_c represents the carrier frequency. The expression (7) holds for mmWave/sub-THz scenarios where the bandwidth obeys $B \ll f_c$, it does not hold for a wideband THz system. The normalized array gain, as defined in [17], is given by

$$\Upsilon = |\tilde{\mathbf{a}}^H(\theta_k, f_k) \tilde{\mathbf{a}}(\theta, f_c)|^2, \quad (8)$$

which is plotted in Fig. 2(b) for such a wideband THz MIMO system, where $f_k \triangleq f_c + (k - \frac{K+1}{2}) \frac{B}{K}$ represents the frequency of the k -th subcarrier. Observe from Eq. (8), that $\tilde{\mathbf{a}}(\theta_k, f_k) \neq \tilde{\mathbf{a}}(\theta, f_c)$ results in $\Upsilon \neq 1$, which in turn causes the beam to point in different physical directions, giving rise to *beam-squint effects*. Therefore, the modified array response vector [4] after incorporating the beam-squint effect can be expressed as

$$\tilde{\mathbf{a}}(\theta, f_k) = \frac{1}{\sqrt{N}} [1, e^{-j\pi \varrho_k \sin \theta}, \dots, e^{-j\pi \varrho_k (N-1) \sin \theta}]^T, \quad (9)$$

where $\varrho_k = \frac{f_k}{f_c}$ represents its relative frequency. Moreover, the maximum array gain is achieved only when $\varrho_k \theta_k - \theta = 0$. The derivation of this is detailed in Appendix-B. The inherent beam-squint effect of wideband THz MIMO systems, causes frequency-dependent beam misalignment and array

Table IV
NOTATION AND DESCRIPTION OF CHANNEL PARAMETERS CONSIDERED

Parameter	Description	Parameter	Description
d	delay tap	σ_n^2	noise power
K	# of FFT points	b	quantizer bit
σ_p^2	pilot power	ρ	inverse SNR of quantization bit
λ	wavelength	d_c	inter-antenna spacing
Υ	normalized array gain	ρ_k	relative frequency
α	complex-path gain	$\text{pul}(\cdot)$	pulse-shaping filter
T_s	sampling time	μ_{abs}	molecular absorption coefficient
$\theta_{(\cdot)}$	angle of arrival	γ	Fresnel reflection coefficient
N_{ray}	# of diffused rays	N_{NLoS}	# of multipath components
L_{spr}	free-space path loss		molecular absorption loss
$\tau_{(\cdot)}$	delay	Γ	first-order reflection coefficient
$\phi_{(\cdot)}$	angle of departure	φ	Rayleigh roughness factor
$Z(f_k)$	characteristic impedance	ω_i	angle of incidence
ω_r	angle of refraction	Z_0	intrinsic impedance
μ_0	free-space permeability	σ_r	standard deviation of material roughness
ϵ_0	free-space permittivity	ξ	absorption coefficient of the material
ds	transmission distance	ϖ	phase of complex path gain

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

gain degradation as discussed above. In Section-VI we will propose an optimal beamformer design. Explicitly, this will replace the conventional frequency-flat analog beamformer by a frequency-dependent one. The next subsection discusses a practical transmit pulse-shaping filter (PSF)-based dual-wideband channel formulation.

A. PSF-based dual-wideband channel

As discussed in Section-I, THz signals exhibit highly directional propagation characteristics, but they have both line-of-sight (LoS) and non-line-of-sight (NLoS) components. Consequently, the dual-wideband THz channel experienced by the k -th subcarrier corresponding to each user can be expressed as $\mathbf{H}_u[k] = \mathbf{H}_u^{\text{LoS}}[k] + \mathbf{H}_u^{\text{NLoS}}[k]$, where we have

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

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

In (11) we have $\beta_{\tau_{z,j}} = \sum_{d=0}^{K-1} \text{pul}(dT_s - \tau_{z,j}) e^{-j \frac{2\pi kd}{K}}, \forall k, d$, where $\text{pul}(\cdot)$ represents the PSF, while N_{ray} denotes the number of diffused rays within a cluster and N_{NLoS} represents the number of multipath components. The receive and transmit antenna gains are represented by \mathfrak{B}_R and $\mathfrak{B}_{T,u}$, while the quantities $\theta_{(\cdot)}$, $\phi_{(\cdot)}$, $\tau_{(\cdot)}$, $\alpha_{(\cdot)}$ represent the AoA, AoD, delay and complex path gain. Note that, $\mathfrak{B}_{T,u}/\mathfrak{B}_r$ are angle-dependent antenna gains [29], [30]. However, for analytical tractability and to remain consistent with the signal processing literature [10], [17], the explicit angle dependence expression is not shown. Nonetheless, in the simulation results, the gain is considered in 'dBi'. The complex path gain $\alpha_{(\cdot)}(f_k, ds)$ can be characterized by its magnitude and phase components $\alpha_{(\cdot)}(f_k, ds) = |\alpha_{(\cdot)}(f_k, ds)| e^{j\varpi}$. Furthermore, the magnitude $\alpha(f_k, ds)$ of the complex path gain depends on the distance ds and for a LoS component which is given by [17]

$$|\alpha(f_k, ds)|^2 = L_{\text{spr}}(f_k, ds) L_{\text{abs}}(f_k, ds), \quad (12)$$

where $L_{\text{spr}}(f_k, ds) = (\frac{c}{4\pi f_k ds})^2$ represents the free-space path loss and $L_{\text{abs}}(f_k, ds) = e^{-\mu_{\text{abs}}(f_k) ds}$ represents the molecular absorption loss. The quantity μ_{abs} represents the

molecular absorption coefficient, which can be calculated from the **HITRAN** database [31]. Similarly, the complex path gain of the NLoS components in the z -th multipath at the j -th diffused ray is given as

$$|\alpha_{z,j}(f_k, ds_{z,j})|^2 = \Gamma_{z,j}^2(f_k) L_{\text{spr}}(f_k, ds_{z,j}) L_{\text{abs}}(f_k, ds_{z,j}), \quad (13)$$

where $\Gamma_{z,j}(f_k)$ represents the first-order reflection, which is given by the product of Fresnel reflection coefficient $\gamma_{z,j}(f_k) = \frac{Z(f_k) \cos(\omega_{i,z,j}) - Z_0(\cos \omega_{r,z,j})}{Z(f_k) \cos(\omega_{i,z,j}) + Z_0(\cos \omega_{r,z,j})}$ and Rayleigh roughness factor $\varphi_{z,j}(f_k) = e^{-\frac{1}{2} \left(\frac{4\pi f_k \sigma_r \cos \omega_{i,z,j}}{c} \right)^2}$ [8]. The quantity $\omega_{i,z,j}$ denotes the angle of incidence, while the angle of refraction is given by $\omega_{r,z,j} = \sin^{-1} \left(\sin(\omega_{i,z,j}) \frac{Z(f_k)}{Z_0} \right)$, where $Z_0 = 377\Omega$ is the intrinsic impedance of the free space, and σ_r represents the standard deviation of the material. Note that the reason for explicitly considering these losses at THz frequencies is that different materials exhibit distinct electromagnetic behavior due to their frequency-dependent refractive indices and absorption characteristics. These effects significantly influence wave propagation and are modeled using the material-specific parameters [32], [8] provided in Table-IX. Moreover, the quantity $Z(f_k) = \sqrt{\frac{\mu_0}{\varepsilon_0(n^2 - (\frac{\xi c}{4\pi f_k})^2 - j \frac{2n\xi c}{4\pi f_k})}}$ represents the characteristic impedance of the reflecting medium [8], where ξ represents the absorption coefficient and n is the refractive index, μ_0 represents the free space permeability and ε_0 denotes the free space permittivity, respectively. Table-IV lists the notation used throughout the work. At this juncture, it is important to note that a THz MIMO channel typically exhibits a limited number of multipath components. Consequently, the next section explores its extended sparse representation tailored for MU THz hybrid MIMO systems.

IV. MU SPARSE CHANNEL ESTIMATION MODEL

Let $G_{T,u}$ and G_{BS} represent the transmit and receive angular bins, which follow the relationship $G_{T,u} \gg N_u, G_{BS} \gg N_{BS}$. Moreover, let Θ_{BS} and $\Phi_{T,u}$ represent the set of receive and transmit directional sine values [33]

$$\Theta_{BS} = \{\theta_{BS} : \sin(\theta_{BS}) = \frac{2}{G_{BS}}(r-1) - 1, 1 \leq r \leq G_{BS}\},$$

$$\Phi_{T,u} = \{\phi_{t,u} : \sin(\phi_{t,u}) = \frac{2}{G_{T,u}}(t-1) - 1, 1 \leq t \leq G_{T,u}\}.$$

Therefore, the extended virtual channel model [2] corresponding to the u -th user can be expressed as $\mathbf{H}_u[k] = \mathbf{A}_{BS}(\Theta_{BS}, f_k) \mathbf{H}_{b,u}[k] \mathbf{A}_{T,u}^H(\Phi_{T,u}, f_k)$. The quantities $\mathbf{A}_{BS}(\Theta_{BS}, f_k) \in \mathbb{C}^{N_{BS} \times G_{BS}}$ and $\mathbf{A}_{T,u}(\Phi_{T,u}, f_k) \in \mathbb{C}^{N_u \times G_{T,u}}$ represent the receive and transmit array manifold dictionary matrices corresponding to the u -th user, which are further given by

$$\mathbf{A}_B(\Theta_{BS}, f_k) = [\tilde{\mathbf{a}}_R(\theta_1, f_k), \tilde{\mathbf{a}}_R(\theta_2, f_k), \dots, \tilde{\mathbf{a}}_R(\theta_{G_{BS}}, f_k)],$$

$$\mathbf{A}_{T,u}(\Phi_{T,u}, f_k) = [\tilde{\mathbf{a}}_{T,u}(\phi_{1,u}, f_k), \dots, \tilde{\mathbf{a}}_{T,u}(\phi_{G_{T,u,u}}, f_k)],$$

while $\mathbf{H}_{b,u}[k] \in \mathbb{C}^{G_{BS} \times G_{T,u}}$ represents the beamspace domain sparse channel. Leveraging the property of $\text{vec}(\cdot)$ operator as defined in Section-I-C, the extended virtual channel can be re-written as

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

where $\tilde{\Delta}_u[k] \in \mathbb{C}^{N_{BS} N_u \times G_{BS} G_{T,u}}$ represents the *sparsifying-dictionary*. Furthermore, to formulate a MU dual-wideband vectorized THz hybrid MIMO channel $\mathbf{h}_u[k]$, we concatenate the sparsifying dictionary $\tilde{\Delta}_u[k]$ corresponding to each user in block-diagonal fashion, which is mathematically expressed as

$$\mathbf{h}_u[k] = \text{blkdiag} \left(\underbrace{[\tilde{\Delta}_1[k], \tilde{\Delta}_2[k], \dots, \tilde{\Delta}_U[k]]}_{\Delta_u[k]} \right) \underbrace{\begin{bmatrix} \mathbf{h}_{b,1}[k] \\ \mathbf{h}_{b,2}[k] \\ \vdots \\ \mathbf{h}_{b,U}[k] \end{bmatrix}}_{\mathbf{h}_{b,U}[k]}. \quad (14)$$

The quantity $\Delta_U[k] \in \mathbb{C}^{N_{BS} N_T \times G_{BS} G_T}$ represents the concatenated sparsifying dictionary across all the users, where we introduce $\sum_{u=1}^U G_{T,u} = G_T$ for notational simplicity while the quantity $\mathbf{h}_{b,U}[k] \in \mathbb{C}^{G_{BS} G_T \times 1}$ in (14) represents the concatenated beamspace output. Revisiting (5) again, the combined received output for the m -th block at the k -th subcarrier across all the users can be mathematically derived as

$$\mathbf{y}_m[k] \approx \mathbf{\Omega}_{u,m}[k] \mathbf{h}_{b,U}[k] + \boldsymbol{\eta}_m[k], \quad (15)$$

where $\mathbf{\Omega}_{u,m}[k] = \tilde{\Psi}_{u,m}[k] \Delta_U[k] \in \mathbb{C}^{N_s \times G_{BS} G_T}$ represents the equivalent sensing matrix. Therefore, the MU THz channel estimation model after concatenating the received pilot outputs across all the M blocks is given by

$$\underbrace{\begin{bmatrix} \mathbf{y}_1[k] \\ \mathbf{y}_2[k] \\ \vdots \\ \mathbf{y}_M[k] \end{bmatrix}}_{\mathbf{y}_U[k]} = \underbrace{\begin{bmatrix} \mathbf{\Omega}_{u,1}[k] \\ \mathbf{\Omega}_{u,2}[k] \\ \vdots \\ \mathbf{\Omega}_{u,M}[k] \end{bmatrix}}_{\mathbf{\Omega}_U[k]} \mathbf{h}_{b,U}[k] + \underbrace{\begin{bmatrix} \boldsymbol{\eta}_1[k] \\ \boldsymbol{\eta}_2[k] \\ \vdots \\ \boldsymbol{\eta}_M[k] \end{bmatrix}}_{\boldsymbol{\eta}_U[k]}, \quad (16)$$

where $\mathbf{y}_U[k] \in \mathbb{C}^{MN_s \times 1}$ denotes the stacked received output. Furthermore, $\mathbf{\Omega}_U[k] \in \mathbb{C}^{MN_s \times G_{BS} G_T}$ represents the stacked equivalent sensing matrix, while $\boldsymbol{\eta}_U[k] \in \mathbb{C}^{MN_s \times 1}$ denotes the stacked equivalent noise. Moreover, the stacked noise covariance is $\mathbf{R} = \mathbb{E}\{\boldsymbol{\eta}_U[k] \boldsymbol{\eta}_U^H[k]\} = \text{blkdiag}(\mathbf{P}_{\eta,1}, \mathbf{P}_{\eta,2}, \dots, \mathbf{P}_{\eta,M}) \in \mathbb{C}^{MN_s \times MN_s}$.

Interestingly, in wideband THz hybrid MIMO systems, the channel vectors across different subcarriers under the frequency-dependent sparsifying dictionary $\mathbf{\Omega}_U[k]$ of Eq. (16) share a common spatial scattering structure, thereby exhibiting *simultaneous sparsity* that can be further exploited for enhancing the channel estimation accuracy. In this regard, we formulate a MMV model by concatenating the pilot output vectors across all K subcarriers, represented as $\mathbf{Y} = [\mathbf{y}_U[1] \mathbf{y}_U[2] \dots \mathbf{y}_U[K]] \in \mathbb{C}^{MN_s \times K}$. Moreover, the MU concatenated beamspace channel is denoted as $\mathbf{\Lambda}_b =$

$[\mathbf{h}_{b,\mathcal{U}}[1] \mathbf{h}_{b,\mathcal{U}}[2] \cdots \mathbf{h}_{b,\mathcal{U}}[K]] \in \mathbb{C}^{G_{BS}G_T \times K}$. One can readily observe from the simultaneously sparse formulation that the dominant spatial scattering paths share a common support across all the subcarriers, which naturally motivates recasting the problem into a group-sparse recovery framework for efficient joint channel estimation over the entire frequency band. Therefore, the next section will discuss the framework of group sparse Bayesian regression.

V. HIERARCHICAL BAYESIAN GROUP SPARSE REGRESSION BASED CHANNEL ESTIMATION

In this section, we introduce a novel CSI learning technique based on hierarchical Bayesian group-sparse regression, specifically designed for MU uplink THz hybrid MIMO systems, where the BS jointly processes the signal received from all the users and the MU CSI exhibits a shared sparse structure. The proposed HBG-SR technique exploits the *group sparse* structure of the concatenated beamspace channel $\mathbf{\Lambda}_b$, whose vectorized form is given by $\mathbf{h}_b = \text{vec}(\mathbf{\Lambda}_b^T)$, and exhibits a structured group-wise representation. Let the g -th group in \mathbf{h}_b be denoted as $\mathbf{h}_b^g \in \mathbb{C}^{K \times 1}$, $\forall 1 \leq g \leq G_{BS}G_T$, which can be further expressed as $\mathbf{h}_b^g = \mathbf{\Lambda}_b(g, :)$. Let the beamspace channel matrix corresponding to each user u , concatenated over all subcarriers K , be defined as $\mathbf{H}_{b,u} = [\mathbf{h}_{b,u}[0] \mathbf{h}_{b,u}[1] \cdots \mathbf{h}_{b,u}[K-1]] \in \mathbb{C}^{G_{BS}G_T \times K}$. The prior corresponding to the overall beamspace matrix $\mathbf{\Lambda}_b$ is assumed to be

$$f(\mathbf{\Lambda}_b; \mathbf{\Gamma}_{\mathcal{U}}) = \prod_{u=1}^U \prod_{g=1}^{G_{BS}G_u} f(\mathbf{H}_{b,u}(g, :); \gamma_{g,u}), \quad (17)$$

where the prior distribution $f(\mathbf{H}_{b,u}(g, :); \gamma_{g,u})$ associated with the g -th group of user u is given by

$$f(\mathbf{H}_{b,u}(g, :); \gamma_{g,u}) = \frac{1}{(\pi\gamma_{g,u})^K} \exp\left(-\frac{(\mathbf{H}_{b,u}(g, :))^H \mathbf{H}_{b,u}(g, :)}{\gamma_{g,u}}\right).$$

Still referring to (17), $\gamma_{g,u}$ denotes the hyperparameter corresponding to the g -th group of the u -th user, which is assumed to be unknown. Let the hyperparameter matrix $\mathbf{\Gamma}_{\mathcal{U}}$ be defined as $\mathbf{\Gamma}_{\mathcal{U}} = \text{blkdiag}(\mathbf{\Gamma}_1 \mathbf{\Gamma}_2 \cdots \mathbf{\Gamma}_U) \in \mathbb{R}^{G_{BS}G_T \times G_{BS}G_T}$, where $\mathbf{\Gamma}_u = \text{diag}(\gamma_{1,u} \gamma_{2,u} \cdots \gamma_{G_{BS}G_u}) \in \mathbb{R}^{G_{BS}G_u \times G_{BS}G_u}$, and $\mathbf{H}_{b,u}(g, :) \in \mathbb{C}^{K \times 1}$ represents the g -th group corresponding to the u -th user. Note that, if the well-known Bernoulli-Gaussian prior [34] is applied within the Bayesian learning framework, the inference becomes intractable and requires additional approximations [35] (as derived in Appendix A). Moreover, adopting a Laplacian prior instead of a Gaussian prior fundamentally alters the inference framework, as the resulting posterior does not admit closed-form expressions [36]. Therefore, we assume a Gaussian prior, because the likelihood and prior are conjugates of each other.

Let the vector of hyperparameters corresponding to the concatenated beamspace channel $\mathbf{\Lambda}_b$ be given by $[\tilde{\gamma}_1 \tilde{\gamma}_2 \cdots \tilde{\gamma}_{G_{BS}G_T}] = \text{diag}(\mathbf{\Gamma}_{\mathcal{U}})$. Using this, the prior distribution corresponding to the complete beamspace channel matrix $\mathbf{\Lambda}_b$, can be rewritten from Eq. (17) as

$$f(\mathbf{\Lambda}_b; \mathbf{\Gamma}_{\mathcal{U}}) = \prod_{g=1}^{G_{BS}G_T} \frac{1}{(\pi\tilde{\gamma}_g)^K} \exp\left(-\frac{\mathbf{\Lambda}_b^H(g, :)\mathbf{\Lambda}_b(g, :)}{\tilde{\gamma}_g}\right), \quad (18)$$

where $\tilde{\gamma}_g$ denotes the hyperparameter associated with the g -th group. Therefore, to maximize the Bayesian evidence $\log f(\mathbf{Y}; \mathbf{\Gamma}_{\mathcal{U}})$, we employ the expectation-maximization (EM) algorithm, which guarantees local convergence and it is used for estimating the hyperparameter matrix $\mathbf{\Gamma}_{\mathcal{U}}$.

Let the hyperparameter estimate at the $(j-1)$ -st EM iteration be denoted by $\hat{\mathbf{\Gamma}}_{\mathcal{U}}^{(j-1)}$. The E-step of the j -th EM iteration computes the expected value of the log-likelihood function $\mathcal{L}(\mathbf{\Gamma}_{\mathcal{U}}|\hat{\mathbf{\Gamma}}_{\mathcal{U}}^{(j-1)})$ for the complete dataset $(\mathbf{y}_{\mathcal{U}}, \mathbf{h}_b)$, where we have $\mathbf{y}_{\mathcal{U}} = \text{vec}(\mathbf{Y}^T)$, yielding

$$\begin{aligned} \mathcal{L}(\mathbf{\Gamma}_{\mathcal{U}}|\hat{\mathbf{\Gamma}}_{\mathcal{U}}^{(j-1)}) &= \mathbb{E}_{\mathbf{h}_b|\mathbf{y}_{\mathcal{U}};\hat{\mathbf{\Gamma}}_{\mathcal{U}}^{(j-1)}} \{\log f(\mathbf{y}_{\mathcal{U}}, \mathbf{h}_b; \mathbf{\Gamma}_{\mathcal{U}})\}, \\ &= \mathbb{E}_{\mathbf{h}_b|\mathbf{y}_{\mathcal{U}};\hat{\mathbf{\Gamma}}_{\mathcal{U}}^{(j-1)}} \{\log [f(\mathbf{y}_{\mathcal{U}}|\mathbf{h}_b)] + \mathbb{E}_{\mathbf{h}_b|\mathbf{y}_{\mathcal{U}};\hat{\mathbf{\Gamma}}_{\mathcal{U}}^{(j-1)}} \{\log f(\mathbf{h}_b; \mathbf{\Gamma}_{\mathcal{U}})\}\}. \end{aligned} \quad (19)$$

By applying Bayes' rule to Eq. (19), the term $\mathbb{E}_{\mathbf{h}_b|\mathbf{y}_{\mathcal{U}};\hat{\mathbf{\Gamma}}_{\mathcal{U}}^{(j-1)}} \{\log f(\mathbf{y}_{\mathcal{U}}|\mathbf{h}_b)\}$ becomes independent of the hyperparameter matrix $\mathbf{\Gamma}_{\mathcal{U}}$, and can thus be omitted in the subsequent M-step. As a result, the equivalent optimization problem that maximizes $\mathcal{L}(\mathbf{\Gamma}_{\mathcal{U}}|\hat{\mathbf{\Gamma}}_{\mathcal{U}}^{(j-1)})$ with respect to $\mathbf{\Gamma}_{\mathcal{U}}$ can be formulated as

$$\hat{\mathbf{\Gamma}}_{\mathcal{U}}^{(j)} = \arg \max_{\mathbf{\Gamma}_{\mathcal{U}}} \mathbb{E}_{\mathbf{h}_b|\mathbf{y}_{\mathcal{U}};\hat{\mathbf{\Gamma}}_{\mathcal{U}}^{(j-1)}} \{\log f(\mathbf{h}_b; \mathbf{\Gamma}_{\mathcal{U}})\}. \quad (20)$$

From Eq. (17)-(18), it readily follows that the M-step update for the g -th hyperparameter in the j -th iteration is given by

$$\begin{aligned} \hat{\gamma}_{g,u}^{(j)} &= \arg \max_{\gamma_{g,u}} \mathbb{E}_{\mathbf{h}_b|\mathbf{y}_{\mathcal{U}};\hat{\mathbf{\Gamma}}_{\mathcal{U}}^{(j-1)}} \{\log f(\mathbf{H}_{b,u}(g, :); \gamma_{g,u})\}, \\ &= \arg \max_{\gamma_{g,u}} \mathbb{E}_{\mathbf{h}_b|\mathbf{y}_{\mathcal{U}};\hat{\mathbf{\Gamma}}_{\mathcal{U}}^{(j-1)}} \left\{ -K \log(\gamma_{g,u}) - \right. \\ &\quad \left. \frac{1}{\gamma_{g,u}} \text{Tr}(\mathbf{H}_{b,u}(g, :)\mathbf{H}_{b,u}^H(g, :)) \right\}. \end{aligned} \quad (21)$$

Upon differentiating the above equation with respect to $\gamma_{g,u}$ and setting the result equal to zero, we obtain

$$\hat{\gamma}_{g,u} = \frac{1}{K} \text{Tr}(\mathbb{E}(\mathbf{H}_{b,u}(g, :)\mathbf{H}_{b,u}^H(g, :))). \quad (22)$$

Moreover, the *a posteriori* pdf is $f(\mathbf{\Lambda}_b|\mathbf{y}_{\mathcal{U}};\hat{\mathbf{\Gamma}}_{\mathcal{U}}^{(j-1)}) = \mathcal{CN}(\mathbf{H}_{b,\mathcal{U}}^{(j)}, \mathbf{\Sigma}^{(j)})$, where $\hat{\mathbf{H}}_{b,\mathcal{U}}^{(j)} \in \mathbb{C}^{G_{BS}G_T \times K}$ represents the *a posteriori* mean and $\tilde{\mathbf{\Sigma}}_k^{(j)} \in \mathbb{C}^{G_{BS}G_T \times G_{BS}G_T}$ denotes the *a posteriori* covariance matrix, which are given by

$$\begin{aligned} \tilde{\mathbf{\Sigma}}_k^{(j)} &= [(\hat{\mathbf{\Gamma}}_{\mathcal{U}}^{(j-1)})^{-1} + \mathbf{\Omega}_{\mathcal{U}}^H[k]\mathbf{R}^{-1}\mathbf{\Omega}_{\mathcal{U}}[k]]^{-1}, \\ \hat{\mathbf{H}}_{b,\mathcal{U}}^{(j)}(:, k) &= \tilde{\mathbf{\Sigma}}_k^{(j)} \mathbf{\Omega}_{\mathcal{U}}^H[k]\mathbf{R}^{-1}\mathbf{Y}(:, k). \end{aligned} \quad (23)$$

Moreover, the *a posteriori* mean $\hat{\mathbf{h}}_{g,u} \in \mathbb{C}^{K \times 1}$ and the *a posteriori* covariance $\mathbf{\Sigma}_{g,u} \in \mathbb{C}^{K \times K}$ corresponding to the g -th group of each user are given as

$$\begin{aligned} \mathbf{h}_{g,u}^{(j)} &= [\hat{\mathbf{H}}_{b,\mathcal{U}}^{(j)}(i_{g,u}, 1), \cdots, \hat{\mathbf{H}}_{b,\mathcal{U}}^{(j)}(i_{g,u}, K)], \\ \tilde{\mathbf{\Sigma}}_{g,u}^{(j)} &= \text{diag}(\tilde{\mathbf{\Sigma}}_k^{(j)}(i_{g,u}, i_{g,u}; 1), \cdots, \tilde{\mathbf{\Sigma}}_k^{(j)}(i_{g,u}, i_{g,u}; K)), \end{aligned} \quad (24)$$

where $i_{g,u} = \tilde{S}_{u-1} + g$, $\forall \tilde{S}_{u-1} < g \leq \tilde{S}_u$ represents the global index, and $S_u = \sum_{m=1}^u G_{BS}G_m$. Upon substituting Eq. (24) into (22), we obtain the M-step as

$$\hat{\gamma}_{g,u}^{(j)} = \frac{1}{K} \text{Tr}[\tilde{\mathbf{\Sigma}}_{g,u}^{(j)} + \mathbf{h}_{g,u}^{(j)}(\mathbf{h}_{g,u}^{(j)})^H], \quad (25)$$

Algorithm 1: HBG-SR based MU THz channel estimation

Input: Pilot output \mathbf{Y} , Equivalent sensing matrix Ξ , noise covariance \mathbf{R} , total subcarriers K , ϵ and K_{\max}

- 1 **Initialization:** $\hat{\Gamma}_{\mathcal{U}}^{(0)} = \mathbf{I}_{G_{BS}G_T}$, $\hat{\Gamma}_{\mathcal{U}}^{(-1)} = \mathbf{0}_{G_{BS}G_T}$ and counter $j = 1$
- 2 **while** ($\|\hat{\Gamma}_{\mathcal{U}}^{(j)} - \hat{\Gamma}_{\mathcal{U}}^{(j-1)}\|_F^2 > \epsilon$ && $j < K_{\max}$) **do**
- 3 **E-Step:** Compute the *a posteriori* covariance and mean as
- 4 **for** $k = 1, \dots, K$ **do**
- 5 $\tilde{\Sigma}_k^{(j)} = [(\hat{\Gamma}_{\mathcal{U}}^{(j-1)})^{-1} + \Omega_{\mathcal{U}}^H[k]\mathbf{R}^{-1}\Omega_{\mathcal{U}}[k]]^{-1}$
- 6 $\hat{\mathbf{H}}_{b,\mathcal{U}}^{(j)}(:, k) = \tilde{\Sigma}_k^{(j)}\Omega_{\mathcal{U}}^H[k]\mathbf{R}^{-1}\mathbf{Y}(:, k)$
- 7 **M-Step:** Update hyperparameters estimates using
- 8 $\hat{\gamma}_g^{(j)} = \frac{1}{K} \sum_{k=1}^K (\tilde{\Sigma}^{(j)}(i_g, i_g; k) + |\hat{\mathbf{H}}_{b,\mathcal{U}}^{(j)}(i_g, k)|^2)$
- 9 $j \leftarrow j + 1$

10 **return:** $\hat{\mathbf{H}}_{b,\mathcal{U}}$

which can be re-written as

$$\hat{\gamma}_g^{(j)} = \frac{1}{K} \sum_{k=1}^K \left(\tilde{\Sigma}^{(j)}(i_g, i_g; k) + |\hat{\mathbf{H}}_{b,\mathcal{U}}^{(j)}(i_g, k)|^2 \right), \quad (26)$$

where $i_g = \tilde{S}_{u(g)-1} + g$ and \tilde{S}_u remain the same as defined earlier. Algorithm-1 summarizes the proposed sparse estimation technique. The procedure is repeated until convergence. Let the concatenated MU transmit dictionary $\mathbf{A}_T(\Phi_T, f_{0:K-1}) \in \mathbb{C}^{N_T \times G_T \times K}$ be formulated as

$$\mathbf{A}_T(\Phi_T, f_{0:K-1}) = \text{blkdiag}[\mathbf{A}_{T,1}(\Phi_{T,1}, f_{0:K-1}), \dots, \mathbf{A}_{T,U}(\Phi_{T,U}, f_{0:K-1})]. \quad (27)$$

Therefore, the estimated MU THz MIMO channel using the HBG-SR algorithm is given as

$$\hat{\mathbf{H}}_{\text{HBG}} = \mathbf{A}_B(\Theta_{BS}, f_{0:K-1}) \text{vec}^{-1}(\hat{\mathbf{H}}_{b,\mathcal{U}}) \mathbf{A}_T^H(\Phi_T, f_{0:K-1}).$$

Note that, conventional MSBL algorithms, which assume a common dictionary matrix across all the subcarriers [37], do not lead to optimal performance in dual-wideband scenarios. 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.

As detailed in Section-I-A, the design of hybrid beamformers having TTD elements necessitates accurate beam-steering angles for each RF chain for counteracting the frequency-dependent delays that arise in wideband THz MIMO systems. These delays effectively mitigate the beam-squint by leveraging the angular-domain information embedded in the estimated beamspace channel. Specifically, the proposed HBG-SR inherently promotes sparsity in the channel estimates, wherein the

Algorithm 2: Determination of optimal physical angles from the estimated beamspace

Input: $\hat{\mathbf{H}}_{b,\mathcal{U}}, N_{RF}^u, N_{RF}^B$

- 1 **Initialization:** $\tilde{\Phi}_T = [\]$, $\tilde{\Theta}_{BS} = [\]$, $E = [\]$
- 2 **for** $u = 1 : U$ **do**
- 3 $\hat{\mathbf{h}}_u = \hat{\mathbf{H}}_{b,\mathcal{U}}[:, (u-1)G_{T,u} + 1 : uG_{T,u}]$
- 4 $[\sim, \tilde{\mathbf{I}}] = \text{sort}(|\hat{\mathbf{h}}_u|, \text{descend})$
- 5 $[\sim, \tilde{\Phi}_T] = \text{ind2sub}(G_{BS}, G_T, \tilde{\mathbf{I}}(1 : N_{RF}^u))$
- 6 $[\tilde{\Theta}_{BS}, \sim] = \text{ind2sub}(G_{BS}, G_T, \tilde{\mathbf{I}}(1 : \frac{N_{RF}^B}{U}))$
- 7 $E_u = \mathbf{I}(1 : N_{RF}^u)$
- 8 $\tilde{\Phi}_T = [\tilde{\Phi}_T \ \tilde{\Phi}_T]; \tilde{\Theta}_{BS} = [\tilde{\Theta}_{BS} \ \tilde{\Theta}_{BS}]; E = [E \ E_u]$

9 **Output:** $\tilde{\Phi}_T, \tilde{\Theta}_{BS}, E$

significant non-zero coefficients directly correspond to dominant propagation paths, implicitly encoding precise directional information. Consequently, by prioritizing these coefficients and mapping their indices to the corresponding AoA/AoD pairs, the steering angles required can be efficiently extracted without additional pilot overhead.

Towards this, Algorithm-2 summarizes the procedure for extracting the dominant AoD/AoA pairs directly from the estimated beamspace channel. Specifically, Step-3 extracts the estimated hyperparameter matrix for each user, and Step-4 transforms this matrix into a vector by selecting its diagonal elements. Step-5 then sorts these hyperparameter values in descending order for ranking their angular significance. Subsequently, Step-6 maps the dominant indices of N_{RF}^u hyperparameters onto their corresponding AoD bins for configuring the user-side analog TPC, while Step-7 maps the dominant $\frac{N_{RF}^B}{U}$ hyperparameters onto AoA bins for configuring the BS-side analog RC. Finally, Step-8 aggregates and returns the selected AoA and AoD indices, representing the dominant transmit and receive angular directions for each user. Consequently, the algorithm circumvents the need to store the complete CSI at the receiver or to fully feed it back to the transmitter, thereby reducing complexity and making it well-suited for practical implementations.

A. MU Bayesian Cramér-Rao lower bound

Let us consider the parameterized Gaussian prior of the beamspace matrix Λ_b , as defined in Eq. (18). Upon taking the natural logarithm on both sides of the Eq. (18), we obtain

$$\log\{f(\Lambda_b; \mathbf{\Gamma}_u)\} = \sum_{g=1}^{G_{BS}G_T} \left(-K \log(\pi\tilde{\gamma}_g) - \frac{1}{\tilde{\gamma}_g} \Lambda_b^H(g, :) \Lambda_b(g, :) \right). \quad (28)$$

To express the log-likelihood of the complete beamspace matrix Λ_b , Eq. (28) can be rewritten as

$$\log\{f(\Lambda_b; \mathbf{\Gamma}_u)\} = -K \sum_{g=1}^{G_{BS}G_T} \log(\tilde{\gamma}_g) - \sum_{g=1}^{G_{BS}G_T} \frac{\Lambda_b^H(g, :) \Lambda_b(g, :)}{\tilde{\gamma}_g} - GK \log(\pi). \quad (29)$$

Let $\sum_{g=1}^{G_{BS}G_T} \tilde{\gamma}_g^{-1} \mathbf{\Lambda}_b^H(g, \cdot) \mathbf{\Lambda}(g, \cdot) = \text{Tr}(\mathbf{\Lambda}_b^H \mathbf{\Gamma}_U^{-1} \mathbf{\Lambda}_b)$. Therefore, Eq. (29) can be re-written as

$$\log f(\mathbf{\Lambda}_b; \mathbf{\Gamma}_U) = -K \log \det(\mathbf{\Gamma}_U) - \text{Tr}(\mathbf{\Lambda}_b^H \mathbf{\Gamma}_U^{-1} \mathbf{\Lambda}_b) + \text{cons.} \quad (30)$$

Thus, the conditional PDF $f(\mathbf{Y}|\mathbf{\Lambda}_b)$ corresponding to the received output \mathbf{Y} is given by $\prod_{k=1}^K \mathcal{CN}(\mathbf{y}_U[k]; \mathbf{\Omega}_U[k] \mathbf{h}_{b,U}[k], \mathbf{R})$. Moreover, the log-likelihood of the conditional PDF $f(\mathbf{Y}|\mathbf{\Lambda}_b)$ is given by

$$\log\{f(\mathbf{Y}|\mathbf{\Lambda}_b)\} = -\sum_{k=1}^K \left[\log \det(\mathbf{R}) + (\mathbf{y}_U[k] - \mathbf{\Omega}_U \mathbf{h}_b)^H \times \mathbf{R}^{-1} (\mathbf{y}_U[k] - \mathbf{\Omega}_U \mathbf{h}_b) \right]. \quad (31)$$

Let $\mathbf{J} \in \mathbb{C}^{G_{BS}G_T K \times G_{BS}G_T K}$ denote the Bayesian Fisher Information Matrix (FIM) formulated as

$$\mathbf{J} = \mathbf{J}_y + \mathbf{J}_b, \quad (32)$$

where $\mathbf{J}_y \in \mathbb{C}^{G_{BS}G_T K \times G_{BS}G_T K}$ denotes the FIM associated with the combined received output \mathbf{Y} , while $\mathbf{J}_b \in \mathbb{C}^{G_{BS}G_T K \times G_{BS}G_T K}$ represents the FIM corresponding to the *a priori* information of the beamspace CSI $\mathbf{\Lambda}_b$. Thus, the quantities \mathbf{J}_y and \mathbf{J}_b [37] can be defined as

$$\mathbf{J}_y = -\mathbb{E}_{\mathbf{Y}, \mathbf{\Lambda}_b} \left\{ \frac{\partial^2 \mathcal{L}(\mathbf{Y}|\mathbf{\Lambda}_b)}{\partial \mathbf{\Lambda}_b \partial \mathbf{\Lambda}_b^H} \right\}, \mathbf{J}_b = -\mathbb{E}_{\mathbf{\Lambda}_b} \left\{ \frac{\partial^2 \mathcal{L}(\mathbf{\Lambda}_b; \mathbf{\Gamma}_{MU})}{\partial \mathbf{\Lambda}_b \partial \mathbf{\Lambda}_b^H} \right\},$$

where $\mathcal{L}(\cdot)$ represents the $\log(f(\cdot))$. By solving Eqs. (29) and (31), one can obtain the corresponding FIM as

$$\begin{aligned} \mathbf{J}_y &= \text{blkdiag}(\mathbf{\Omega}_U^H[1] \mathbf{R}^{-1} \mathbf{\Omega}_U[1], \dots, \mathbf{\Omega}_U^H[K] \mathbf{R}^{-1} \mathbf{\Omega}_U[K]), \\ \mathbf{J}_b &= \hat{\mathbf{\Gamma}}_U^{-1} \otimes \mathbf{I}_K. \end{aligned} \quad (33)$$

Using the expressions for the received output FIM and the beamspace FIM from equation (33), the Bayesian FIM can therefore be obtained as

$$\mathbf{J} = \text{blkdiag}(\mathbf{\Omega}_U^H[1] \mathbf{R}^{-1} \mathbf{\Omega}_U[1], \dots, \mathbf{\Omega}_U^H[K] \mathbf{R}^{-1} \mathbf{\Omega}_U[K]) + \hat{\mathbf{\Gamma}}_U^{-1} \otimes \mathbf{I}_K. \quad (34)$$

Let the concatenated equivalent sparsifying dictionary be defined as $\mathbf{S} = \text{blkdiag}(\mathbf{\Delta}_U[1] \mathbf{\Delta}_U[2] \dots \mathbf{\Delta}_U[K]) \in \mathbb{C}^{N_{BS} N_T K \times G_{BS} G_T K}$. Finally, the BCRLB for the MSE estimate of $\hat{\mathbf{\Lambda}}_b$ can be formulated as

$$\text{MSE}\{\hat{\mathbf{\Lambda}}_b\} \geq \text{Tr}\{\mathbf{S} \mathbf{J}^{-1} \mathbf{S}^H\}. \quad (35)$$

B. Computational complexity

This subsection derives the computational complexities of the HBG-SR, GSOMP and MMV-LS based THz channel learning techniques. Table-V list the computational complexity of the MMV-LS scheme. As described in [38], the MMV-LS scheme possesses a worst-case computational complexity of $\mathcal{O}(N_{BS}^3 N_T^3)$ flops, primarily arising from the inversion and multiplication operations involved in the LS estimation process. Table-VI list the per-subcarrier and per-iteration complexity of the GSOMP framework along with various key computational steps. Note that, the GSOMP algorithm follows the same algorithmic structure as done in our earlier work [39]. Moreover, it applies to a generic MMV formulation and is not

restated explicitly to avoid redundancy. The associated computational complexity is, however, included to enable comparison and complexity performance trade-offs. The overall computational complexity of the GSOMP framework is on the order of $(K G_{BS} G_T M N_s + K(\frac{2}{3} j^3 + j^2 M N_s + 3j M N_s))$, where j represents the current iteration. The worst-case complexity order is $\mathcal{O}(K M^3 N_s^3)$, which is attributed to the necessity of an intermediate LS estimate at each iteration. Table-VII lists the per-subcarrier and per-iteration complexity of the HBG-SR framework along with various key computational steps. The worst-case computational complexity for all the subcarriers is $\mathcal{O}(K(G_{BS}^2 G_T^2 M N_s + M^3 N_s^3))$, implying that the computational complexity increases linearly with the number of subcarriers and also higher than GSOMP and MMV-LS framework. However, as shown in Section-VII, HBG-SR indicates superior performance. Hence, there is a trade-off between computational complexity and estimation performance.

VI. TTD BASED HYBRID COMBINER DESIGN

In this section, we develop a beam-squint mitigation strategy in which frequency-dependent TTD elements perform coarse delay compensation across the signal bandwidth, followed by a bank of frequency-flat PSs that refine the residual steering error. Consequently, the overall beamformer design is structured into two stages; the first stage involves identifying the optimal frequency-independent beam steering angles, while the second stage utilizes these angles to construct frequency-dependent RF TPCs and RCs. Furthermore, to determine the angles in the first stage, we leverage the estimated beamspace AoA/AoD pairs obtained from Algorithm-2 to configure the TTD elements.

A. Generation of frequency dependent beamformers

Let each RF chain be connected to S TD elements, which are further connected to $P = \frac{N}{S}$ PSs. Upon revisiting Eq. (7) and dividing the array response vector into S equal number of elements, we obtain

$$\tilde{\mathbf{a}}_{1,k}^l = \tilde{\mathbf{a}}_1^l e^{-j2\pi f_k t_1^l}, \dots, \tilde{\mathbf{a}}_{S,k}^l = \tilde{\mathbf{a}}_S^l e^{-j2\pi f_k t_S^l}, \quad (36)$$

where $\tilde{\mathbf{a}}_{(\cdot)}^T \in \mathbb{C}^{P \times 1}$ and $\tilde{\mathbf{a}}(\theta_k, f_k) = [\tilde{\mathbf{a}}_1^T, \dots, \tilde{\mathbf{a}}_S^T]^T \in \mathbb{C}^{N \times 1}$. Let $\mathbf{t}^l = [t_1^l, \dots, t_S^l]^T \in \mathbb{C}^{S \times 1}$ represent the delays corresponding to each RF chain. Therefore, the overall frequency-dependent PS corresponding to the k -th subcarrier can be expressed as

$$\tilde{\mathbf{a}}_k^l = \text{blkdiag}((\tilde{\mathbf{a}}_1^l)^T, \dots, (\tilde{\mathbf{a}}_S^l)^T) e^{-j2\pi f_k \mathbf{t}^l}. \quad (37)$$

Let $\mathbf{p}_k^l = e^{-j2\pi f_k \mathbf{t}^l}$ represent the frequency-dependent PS. To maintain the directivity, one can generate a frequency-dependent beam to share the same form as the array response vector $\tilde{\mathbf{a}}_k$ expressed as

$$\mathbf{p}_k^l = [1, e^{-j\pi v_k^l}, \dots, e^{-j\pi(S-1)v_k^l}]^T, \quad (38)$$

where $v_k^l \in [-1, 1]$ represents the directional rotation vector. Therefore, by adjusting v_k^l , the beams generated by $\tilde{\mathbf{a}}_k^l$ can be aligned with the target physical direction θ^l corresponding to

Table V
COMPUTATION COMPLEXITY OF MMV-LS

Operation	Complex multiplications	Complex additions
$\Psi_U^H[k]\Psi_U[k]$	$\frac{MN_s N_{BS} N_T (N_{BS} N_T + 1)}{2}$	$\frac{(MN_s - 1) N_{BS} N_T (N_{BS} N_T + 1)}{2}$
$(\Psi_U^H[k]\Psi_U[k])^{-1}$	$\frac{N_{BS}^3 N_T^3}{2} + \frac{3N_{BS}^2 N_T^2}{2}$	$\frac{N_{BS}^3 N_T^3}{2} - \frac{3N_{BS}^2 N_T^2}{2}$
$(\Psi_U^H[k]\Psi_U[k])^{-1}\Psi_U^H[k]$	$N_{BS}^2 N_T^2 M N_s$	$N_{BS} N_T M N_s (N_{BS} N_T - 1)$
$(\Psi_U^H[k]\Psi_U[k])^{-1}\Psi_U^H[k]\mathbf{Y}_{MU}(:, k)$	$N_{BS} N_T M N_s$	$N_{BS} N_T (M N_s - 1)$

Table VI
COMPUTATION COMPLEXITY OF GSOMP SCHEME, PER-SUBCARRIER IN THE j -TH ITERATION

Operation	Complex multiplications	Complex additions
$\Omega_U^H[k]\mathbf{T}(:, k)$	$G_{BS} G_T M N_s$	$G_{BS} G_T (M N_s - 1)$
$\hat{\mathbf{H}}_{LS, U}(:, k) = (\Omega_U^A)^\dagger \mathbf{Y}_U(:, k)$	$\frac{2}{3}j^3 + j^2 M N_s + 2j M N_s$	$\frac{2}{3}j^3 + j^2 (M N_s - 1) + j (M N_s - 1)$
$\mathbf{Y}_U(:, k) - \Omega_U^A \hat{\mathbf{H}}_{LS, U}(:, k)$	$j M N_s$	$j M N_s$

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

Operation	Complex multiplication	Complex additions
$(\hat{\Gamma}_U^{(j-1)})^{-1}$	$\frac{G_{BS}^3 G_T^3}{2} + \frac{3G_{BS}^2 G_T^2}{2}$	$\frac{G_{BS}^3 G_T^3}{2} - \frac{3G_{BS}^2 G_T^2}{2}$
\mathbf{R}^{-1}	$\frac{M^3 N_s^3}{2} + \frac{3M^2 N_s^2}{2}$	$\frac{M^3 N_s^3}{2} - \frac{3M^2 N_s^2}{2}$
$\Omega_U^H[k]\mathbf{R}^{-1}$	$G_{BS} G_T M^2 N_s^2$	$G_{BS} G_T M^2 N_s^2 - G_{BS} G_T M N_s$
$\Omega_U^H[k]\mathbf{R}^{-1}\Omega_U[k]$	$G_{BS}^2 G_T^2 M N_s$	$G_{BS}^2 G_T^2 M N_s - G_{BS}^2 G_T^2$
$[(\hat{\Gamma}_U^{(j-1)})^{-1} + \Omega_U^H[k]\mathbf{R}^{-1}\Omega_U[k]]^{-1}$	$\frac{G_{BS}^3 G_T^3}{2} + \frac{3G_{BS}^2 G_T^2}{2}$	$\frac{G_{BS}^3 G_T^3}{2} - \frac{3G_{BS}^2 G_T^2}{2}$
$\Omega_U^H[k]\mathbf{R}^{-1}\mathbf{Y}(:, k)$	$G_B G_T M N_s$	$G_B G_T (M N_s - 1)$
$\hat{\Sigma}_k^{(j)} \Omega_U^H[k]\mathbf{R}^{-1}\mathbf{Y}(:, k)$	$G_B^2 G_T^2$	$G_B G_T (G_B G_T - 1)$

all the subcarriers k . Moreover, the maximum array gain is achieved at an angle ω_{opt} , when

$$\omega_{\text{opt}} = \arg \max_{\theta^l} |(\tilde{\mathbf{a}}(\theta^l, f_c))^H \tilde{\mathbf{a}}(\theta_k^l, f_k)|, \quad (39)$$

which further reduces to

$$\omega_{\text{opt}} = \frac{\theta^l}{\varrho_k} - \frac{v_k^l}{P \varrho_k}, \quad (40)$$

as detailed in Appendix-C. Thus, aligning the optimal beam steering angle with the actual physical direction yields $v_k = (1 - \varrho_k)\theta P$, implying that the beam-squint effect can be largely compensated using the TD elements. Additionally, it can be observed from Eq. (38) that the phase differences between adjacent TTD elements are equal and implying that the TD vector \mathbf{t} should satisfy the following form

$$\mathbf{t}^l = [0, nT_c, \dots, (S-1)nT_c]^T, \quad (41)$$

where $T_c = \frac{1}{f_c}$ represents the time-period of the carrier frequency and n represents number of periods corresponding to each RF chain.

B. Calculation of optimal time delays

Upon substituting the TD value in the expression of the frequency-dependent PS, one obtains

$$\mathbf{p}_k^l = [1, e^{-j2\pi f_k n T_c}, \dots, e^{-j2\pi f_k (S-1)n T_c}]^T. \quad (42)$$

When comparing it to Eq. (38), the time period obtained becomes $n = \frac{P\theta^l}{2\varrho_k}(\varrho_k - 1)$. In addition, one can observe that the number of periods depends not only on the fixed PS P and the physical direction θ , but also on the relative frequency ϱ_k . Furthermore, substituting the value of n into the TD vector, one arrives at

$$t_s^l = (s-1)nT_c = (s-1)\frac{P\theta^l}{2}T_c - (s-1)\frac{P\theta^l}{2\varrho_k}T_c, \quad (43)$$

which represents the first term to be fixed as $\tilde{t}_s^l = (s-1)\frac{P\theta^l}{2}T_c$, $\forall 1 < s < S$, for all subcarriers, while the second term can be realized by adding an extra phase shift. Revisiting Eq. (37) and substituting the value of TD vector into it, the overall array response vector can be expressed as (44) where the actual time delay obeys $\tilde{\mathbf{t}}^l = [\tilde{t}_1^l, \tilde{t}_2^l, \dots, \tilde{t}_S^l]$. Therefore, the generalized time delay corresponding to each RF chain can be obtained as

$$\tilde{t}_{l,s} = \begin{cases} (s-1)\frac{P\theta_l}{2}T_c, & \theta_l \geq 0, \\ (S-1)\left|\frac{P\theta_l}{2}\right|T_c + (s-1)\frac{P\theta_l}{2}T_c, & \theta_l < 0. \end{cases} \quad (45)$$

where $t_{s,l} \in [0, \frac{N}{2}T_c]$ with $\theta_l \in [-1, 1]$ and $1 \leq s \leq S$. For a representative configuration with $f_c = 0.65$ THz, $N_{BS} = 64$ and $S = 2$, the corresponding delay range is between 0 to 50 ps. The delay values are practically achievable by artificial transmission-line based TTDs and delay-line based TTD architectures [40], [41]. These designs support delay ranges and resolutions that are consistent with the values considered in this work. In our analysis, we assume sufficiently fine-grained and practically realizable delay resolution in order to focus on the fundamental signal-processing behavior. At the same time, we do not impose a specific hardware-dependent delay quantization or implementation constraint, as the objective of this work is not to optimize a particular TTD circuit design but to study the system-level interaction between delay compensation and low-resolution ADCs. Algorithm-3 summarizes the design procedure of a frequency-dependent RC. A similar procedure can be applied for the frequency-dependent TPC, where the optimal precoder angle Φ_T is used as input; the detailed procedure is omitted here due to space constraints.

$$\begin{aligned}\tilde{\mathbf{a}}_k^l &= \text{blkdiag}((\tilde{\mathbf{a}}_1^l)^T, (\tilde{\mathbf{a}}_2^l)^T, \dots, (\tilde{\mathbf{a}}_S^l)^T) e^{-j2\pi f_k t^l}, \\ &= \text{blkdiag}((\tilde{\mathbf{a}}_1^l)^T, (\tilde{\mathbf{a}}_2^l)^T e^{j\pi P \theta^l}, \dots, (\tilde{\mathbf{a}}_S^l)^T e^{j\pi(S-1)P \theta^l}) e^{-j2\pi f_k t^l},\end{aligned}\quad (44)$$

Algorithm 3: TTD based frequency-dependent combiner design

Input: $\tilde{\mathbf{H}}_u[k]$, $\mathbf{A}_B(\Theta_{BS}, f_c)$, $\tilde{\Theta}_{BS}$, E

- 1 **for** $i = 1, 2, \dots, N_{RF}^B$ **do**
- 2 $\tilde{\mathbf{a}}^i = \mathbf{A}_B(\Theta_{BS}, f_c)[:, E(i)]$;
- 3 $\theta^i = \Theta_{BS}(:, \tilde{\Theta}_{BS}(i))$
- 4 **for** $s = 1, 2, \dots, S$ **do**
- 5 $\tilde{\mathbf{a}}_s^i = \tilde{\mathbf{a}}^i[1 + (s-1)P : sP] e^{j\pi(S-1)P \theta^i}$
- 6 Calculate TD \tilde{t}_s^i using equation (45)
- 7 $\tilde{\mathbf{a}}_{s,k}^i = \tilde{\mathbf{a}}_s^i e^{-j2\pi f_k \tilde{t}_s^i}$
- 8 $\mathbf{W}_{RF}[k](:, i) = [(\tilde{\mathbf{a}}_{1,k}^i)^T, \dots, (\tilde{\mathbf{a}}_{S,k}^i)^T]^T$
- 9 $\tilde{\mathbf{H}}_{eq}[k] = \mathbf{H}_u^H[k] \mathbf{W}_{RF}[k]$
- 10 $[\mathbf{U}_{eq}[k] \Sigma_{eq}[k] \mathbf{V}_{eq}^H[k]] = \text{SVD}(\tilde{\mathbf{H}}_{eq}[k])$
- 11 $\mathbf{W}_{BB,u}[k] = \mathbf{V}_{eq}[:, 1:N_u^s][k]$
- 12 **return:** $\mathbf{W}_{RF}[k]$, $\mathbf{W}_{BB,u}[k]$

C. Minimum number of TD elements required to mitigate the beam-squint effect

To effectively mitigate the beam-squint effect, it is further essential to determine the minimum number of TD elements required per subarray, ensuring frequency-independent beam alignment across the entire operational bandwidth. Revisiting, the directional rotation vector which includes, $-1 \leq v_k^l \leq 1$, and substituting the optimal value, we get

$$-1 \leq (1 - \varrho_k) P \theta^l \leq 1. \quad (46)$$

Moreover, to design the worst-case design constraint, we consider $\theta = 1$. Under this condition, the minimum number of TD elements required is given by

$$S \geq N(1 - \varrho_k), \quad (47)$$

which represents the theoretical lower bound on M . In practical implementations, this bound must be rounded up to the nearest integer to ensure realizability, i.e., the effective number of TD elements is given by $\lceil N(1 - \varrho_k) \rceil$.

VII. SIMULATION RESULTS

This section provides the performance analysis of the proposed TTD based hybrid beamformer and the HBG-SR based channel learning algorithm. Table-VIII outlines the key numerical parameters used in our simulation study. Note that, for simplicity, we consider only one diffused ray within each cluster. However, the framework can be readily extended to accommodate multiple diffused rays. Additionally, the inter-antenna spacing between the AEs is set to be $\frac{\lambda}{2}$. The number of channel paths is $N_{\text{LoS}} + N_{\text{NLoS}} = 4$. The path positions are assumed to be independent and identically distributed, with their delay $\tau_{(\cdot)}$ chosen uniformly from $[0, (D-1)T_s]$ with $T_s = \frac{1}{B}$ [2]. We consider an indoor office environment with a

Table VIII
SIMULATION PARAMETERS FOR THE HYBRID TRANSCEIVER DESIGN AND SPARSE ESTIMATION

Parameter	Value
# of TAs at UE (N_u)	4
# of RAs (N_{BS})	64
# of TAs RF chains (N_{RF}^u)	2
# of RAs RF chains (N_{RF}^B)	8
# of time delay per RF chain (S)	2
# of subcarriers (K)	128
# of pilot blocks (M)	20
# of data vectors (N_d)	100
# of Users (U)	3
# of NLoS components/ clusters (N_{NLoS})	3
# of diffused rays (N_{ray})	1
# of delay taps (D)	6
Molecular Absorption loss (K_{abs})	0.015 m^{-1}
Transmit Antenna Gain at each UE ($\mathfrak{B}_{T,u}$)	8 dBi
Receive Antenna Gain at each subarray (\mathfrak{B}_R)	28 dBi
Transmit angular grid size at UE ($G_{T,u}$)	8
Receive angular grid size (G_{BS})	128
Operating frequency (f_c)	0.65 THz
Transmission distance (ds)	15m
Bandwidth (B)	5 GHz
Angle quantization parameter (N_Q)	4
ADC Resolution (ρ)	3-bit
Constellation	8-PSK
Roll-off factor for RRC-PSF	0.80
Upsampling factor	20
HBG-SR Threshold ($\varepsilon, N_{\text{max}}$)	$10^{-4}, 30$

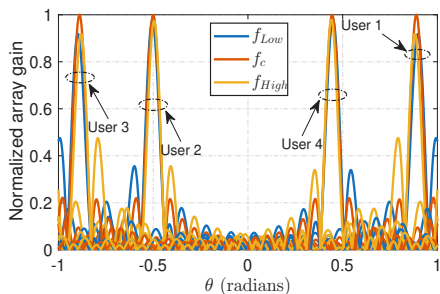
Table IX
LIST OF MATERIALS USED FOR THE SIMULATION ENVIRONMENT [8], [32]

Material Type	σ_r (in mm)	ξ (in cm^{-1})	n
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

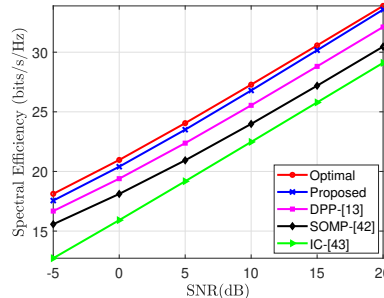
molecular composition of 20.9% oxygen, 78.1% nitrogen and 1% water vapor. The material parameters of various scatterers are listed in Table-IX. Note that in the THz band, the refractive index exhibits frequency-dependent behavior, which causes these materials to interact differently with electromagnetic waves across the frequency spectrum. The SNR in decibels (dB) is calculated as $\text{SNR} = 10 \log_{10}(\frac{1}{\sigma_n^2})$. The transmitting and receiving grid-sizes are considered to be $G_{T,u} \geq 2N_u$ and $G_{BS} \geq 2N_{BS}$ [28]. The frequency-independent PSs are modeled similar to [5] given as

$$\mathbf{F}_{RF,m,u}(a, b) = \frac{1}{\sqrt{N_{T,u}}} e^{jv_{a,b}}, \mathbf{W}_{RF,m} = \frac{1}{\sqrt{N_R}} e^{j\zeta_{a,b}}, \quad (48)$$

where the phases $v_{a,b}$ and $\zeta_{a,b}$ are randomly sampled with uniform probability as $\mathcal{F} = \left\{0, \frac{2\pi}{2^{N_Q}}, \dots, \frac{(2^{N_Q}-1)}{2^{N_Q}}\right\}$. Moreover, the AoAs and AoDs are generated using GMM to provide a more accurate spatial representation [10], with their second-order statistics [9] given by



(a)



(b)

Figure 3. (a) Normalized array gain vs physical direction corresponding to considered scenario (b) Spectral efficiency in (bits/sec/Hz) vs SNR for the proposed and existing state-of-the-art approaches

$$\text{GMM}(r) = \frac{p_1}{\sqrt{2\pi\nu_1}} e^{-\frac{1}{2}\left(\frac{r-r_1}{\nu_1}\right)^2} + \frac{p_2}{\sqrt{2\pi\nu_2}} e^{-\frac{1}{2}\left(\frac{r-r_2}{\nu_2}\right)^2}, \quad (49)$$

where r_1 and r_2 represent the actual physical approximation corresponding to the AoA/AoD. The quantity $p_{(\cdot)}$ represents the mixing weights uniformly distributed in the range $[0, 1]$, and following $p_2 = 1 - p_1$, while $\nu_{(\cdot)}$ represents the ray variance. The values of this are defined in Table-X. The motivation

Table X
GMM PARAMETERS FOR AO/AO D STATISTICS [9]

Parameter	AoA (θ)	AoD (ϕ)
p_1	0.724	0.429
p_2	2.198	1.811
ν_1	0.276	0.571
ν_2	7.297	12.201

for choosing the GMM over the Laplacian distribution [5] lies in its ability to provide finer angular resolution and to capture a larger number of multipath components. Moreover, GMM can accommodate arbitrary reflection orders, which is particularly important in the THz band [9]. The quantity $p_{(\cdot)}$ represents the mixing weights uniformly distributed in the range $[0, 1]$, and following $p_2 = 1 - p_1$, while $\nu_{(\cdot)}$ represents the ray variance.

The NMSE is given by $\text{NMSE} = \frac{\sum_{k=0}^{K-1} \|\hat{\mathbf{H}}[k] - \mathbf{H}[k]\|_{\mathcal{F}}^2}{\sum_{k=0}^{K-1} \|\mathbf{H}[k]\|_{\mathcal{F}}^2}$. Note that for all simulations, we employ the RRC-PSF-based dual-wideband channel model, unless stated otherwise. A trade-off analysis between the two channel formulations is also presented in Section-VII-B. Furthermore, to calculate the optimal digital precoders and combiners we follow the following procedure. Let the singular value decomposition (SVD) of the u -th user channel response $\mathbf{H}_u[k]$ can be obtained as

$$\mathbf{H}_u[k] = \mathbf{U}[k]\mathbf{\Sigma}[k]\mathbf{V}^H[k], \quad (50)$$

and the optimal transmit precoder $\mathbf{F}_u^{\text{opt}}[k] \in \mathbb{C}^{N_u \times N_{s,u}}$ is comprised of the dominant $N_{s,u}$ columns of the right-singular matrix $\mathbf{V}[k]$ [33]. Let $\mathbf{H}_{\text{eq},u}[k] = \mathbf{H}_u[k]\mathbf{F}_u^{\text{opt}}[k] \in \mathbb{C}^{N_{\text{BS}} \times N_{s,u}}$ represent the equivalent channel response. One can now formulate the optimal MMSE combiner $\mathbf{W}_{\text{opt}}[k] \in \mathbb{C}^{N_{\text{BS}} \times N_{s,u}}$

$$\mathbf{W}_{\text{opt},u}[k] = \mathbf{H}_{\text{eq},u}[k] (\mathbf{H}_{\text{eq},u}^H[k]\mathbf{H}_{\text{eq},u}[k] + \sigma^2 N_{s,u} \mathbf{I}_{N_{s,u}})^{-1}.$$

A. Analysis of hybrid transceiver

Fig. 3(a) shows the normalized array gain for the first and third users as a function of their physical steering directions.

It is evident that after applying the proposed TTD-based hybrid transceiver design, the array responses are frequency-aligned across all subcarriers. This demonstrates the ability of the TTD elements to effectively compensate for beam-squint by introducing frequency-dependent delays, thereby aligning the beams precisely with the user's physical directions at $\theta_1 \approx 27.85^\circ$ and $\theta_3 \approx 34.37^\circ$. As a result, consistent beam coverage is achieved over the entire bandwidth, validating the proposed design's efficiency in mitigating beam-squint effects in wideband THz systems.

Fig. 3(b) depicts a substantial improvement in the achievable data rate when compared to existing state-of-the-art beamforming schemes, including DPP [13], SOMP [42] and interference cancellation (IC) [43] across varying SNR levels. The performance of SOMP and IC degrades significantly in the presence of beam-squint, since neither of these approaches incorporate a delay-compensation mechanism. Without delay elements their beamformers become frequency-dependent, leading to beam misalignment and reduced array gain across the bandwidth. The DPP method also suffers from performance limitations due to the absence of low-resolution modeling, making it less effective under coarse quantization, which is common in practical THz systems. By contrast, the proposed framework directly leverages the angular information embedded in the estimated beamspace, thereby eliminating the need for separate angle recovery and facilitating an integrated TTD-based hybrid transceiver design. This not only reduces computational complexity but also improves angular alignment across subcarriers, leading to significantly enhanced performance in wideband THz systems.

B. Analysis of channel estimation

This section presents the performance analysis of the proposed HBGR framework and compares it to the existing state-of-the-art techniques. Fig. 4(a) illustrates the NMSE performance as a function of SNR. The poor performance of the LS [38] method is attributed to its inability to incorporate sparsity in the estimation process, which is a significant limitation compared to compressive sensing (CS) based techniques. Furthermore, the OMP algorithm [2] suffers from both structural and convergence errors. Specifically, a low stopping threshold leads to many iterations and significant nonzero components, causing structural errors. Conversely, a high stopping threshold

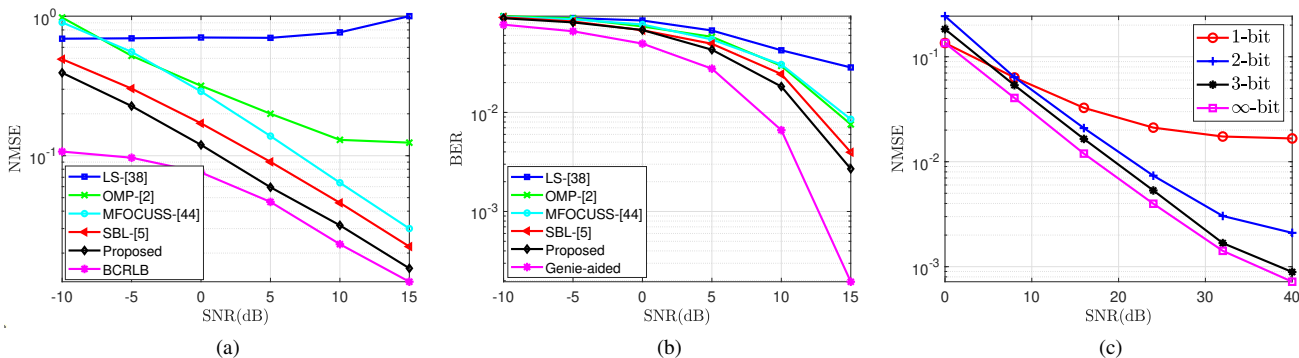


Figure 4. (a) NMSE vs SNR comparison for the proposed and existing state-of-the-art approaches (b) BER vs SNR comparison for the proposed and existing state-of-the-art approaches (c) Effect of low-resolution ADCs on the proposed HBG-SR channel learning technique

reduces iterations, but fails to capture all dominant channel components, potentially resulting in convergence errors. Moreover, the MMV Focal Underdetermined System Solver (MFOCUSS) algorithm [44] often converges to suboptimal local minima, resulting in convergence errors. The Bayesian framework leverages the EM algorithm for hyperparameter learning, eliminating the need for manual tuning or regularization and turns out to be robust the choice of the dictionary matrix. The conventional Bayesian learning technique [5] estimates the channel support independently for each subcarrier. In wideband systems affected by beam-squint, the effective steering vectors change with frequency, causing the support to vary across subcarriers. Since the exact supports differ, these variations cause SBL to perform poorly in this scenario. By contrast, the proposed HBG-SR approach enforces joint sparsity by associating a shared group of hyperparameters with each angle bin across all subcarriers. This leverages the underlying group-sparse structure of the channel, facilitating more consistent support detection and producing significantly lower NMSE, as observed in the simulation results. Moreover, the performance of the proposed method closely approaches the BCRLB, which is particularly noteworthy given that the BCRLB is derived under the idealized assumption of perfect AoA/AoD knowledge. Note that the BCRLB characterizes a fundamental lower bound on the achievable MSE under the assumed probabilistic channel model and prior, but does not by itself imply that the bound is always tight or exactly attainable by practical estimators. Prior studies [45], [46] have shown that Cramér Rao type bounds are asymptotically tight in regimes, where the effective support is correctly identified and the SNR is sufficiently high. In particular, for sparse parameters associated with maximal support, the CRB coincides with the oracle bound and it is achievable by the maximum likelihood estimator in the high-SNR limit, as also reflected by the simulation results of Fig. 4(a). However, in the THz hybrid MIMO setting considered, finite-SNR operation, model mismatch, and hardware impairments generally prevent exact achievability of the BCRLB. Therefore, the bound should be interpreted as a benchmark for best-case performance.

Fig 4(b) depicts the accuracy of data detection achieved by employing the Minimum Mean Squared Error (MMSE) receiver with the proposed HBG-SR technique, along with

other conventional sparse estimation based techniques and the Genie-aided detector as benchmarks. As illustrated, the resultant BER decreases consistently upon increasing the SNR, which closely aligns with the NMSE results presented previously, underscoring the high quality of the CSI estimated. Additionally, the BER of the proposed channel learning approach closely matches the hypothetical Genie-aided detector, highlighting its efficacy and robustness in realistic scenarios. Fig 4(c) evaluates the effect of employing low-resolution ADCs on the CSI estimation accuracy. It can be observed that even with 3-bit ADC resolution, the proposed algorithm attains a performance close to an ideal ∞ -resolution transceiver scenario. This result holds significant practical value, demonstrating the robustness of our channel learning technique in scenarios, where low-resolution ADCs are essential for managing high bandwidth with limited power budgets.

C. Performance comparison for RRC-PSF and Rect-PSF based dual-wideband formulations

To establish a baseline for comparison and to highlight the performance trade-offs introduced by practical pulse shaping, we also consider the Rect-PSF based dual-wideband channel model, as adopted in [17], [18]. Fig. 5(a) compares the performance of the HBG-SR based channel learning approach 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 symbol duration. 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. 5(b) compares the BER performance of the proposed HBG-SR-based channel learning framework over dual-wideband THz channels modeled using the RRC-PSF and Rect-PSF filters. In both cases, it is evident that the HBG-SR-based approach consistently outperforms the conventional SBL and MFOCUSS techniques across the entire SNR range.

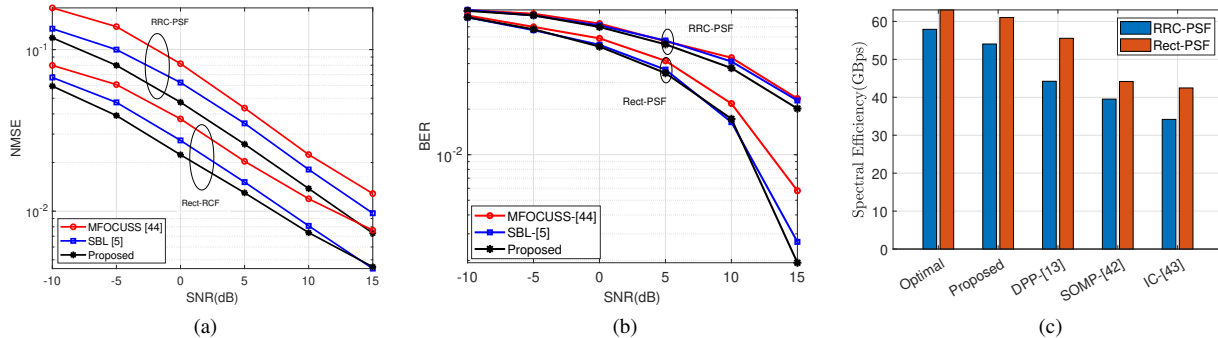


Figure 5. Performance comparison for the proposed HBG-SR technique with RRC-PSF and Rect-PSF based dual-wideband channel formulations (a) NMSE vs SNR (b) BER vs SNR (c) Spectral efficiency with different transceiver schemes.

Notably, the improvement is more prominent in the Rect-PSF-based dual-wideband channel due to its flat response for the entire symbol duration, which exacerbates beam-squint and frequency-dependent distortions. The robustness of the proposed method under both pulse-shaping filters highlights its practical applicability and makes it a strong candidate.

Fig. 5(c) compares the achievable spectral efficiency (in Gbps) of the proposed TTD-based hybrid beamforming scheme to several baseline methods, including the SOMP and IC technique, under both the RRC-PSF and Rect-PSF based dual-wideband channel models. The proposed approach consistently achieves higher throughput for both pulse-shaping filters and closely approaches the performance of the fully digital optimal benchmark. This gain is attributed to the TTD elements, which effectively compensate for frequency-dependent phase misalignments and mitigates the beam-squint effect across wideband subcarriers. Notably, all simulations are conducted under practical hardware constraints with 3-bit ADC resolution, highlighting the robustness and efficiency of the proposed design in low-resolution scenarios. These results validate the framework's capability to jointly address wideband beamforming and channel estimation challenges in THz systems with dual-wideband impairments. Potential future research includes the incorporation of blockage effects and molecular absorption noise to further investigate the performance of the proposed framework under realistic THz propagation conditions.

VIII. CONCLUSION

A novel HBG-SR-based channel estimation framework for wideband THz MIMO systems was proposed. To accurately capture the wideband propagation characteristics, we formulated a dual-wideband channel model incorporating RRC-PSF. The Bussgang decomposition was employed to linearize the nonlinear signal model arising from low-resolution ADCs, facilitating tractable inference under coarse quantization. Leveraging the estimated beamspace information, a hybrid transceiver design based on TTD was presented to mitigate the beam-squint effect and enable frequency-flat beam alignment across subcarriers. Simulation results demonstrated that the proposed approach consistently outperforms existing techniques in terms of channel estimation accuracy, BER and achievable data rate, while remaining efficient under low-resolution hardware constraints. Overall, the joint design offers

a scalable, robust, and hardware-aware solution for next-generation wideband THz communication systems.

APPENDIX A

CALCULATION OF NOISE COVARIANCE

Revisiting Eq. (1), the received signal before passing through a low-resolution ADC can be expressed as

$$\hat{y}_m(q) = \mathbf{W}_{\text{RF},m}^H \sum_{u=1}^U \mathbf{H}_{q,u} \otimes_k \tilde{\mathbf{g}}_{m,u}^{(q)} + \mathbf{W}_{\text{RF},m}^H \mathbf{v}_m(q). \quad (51)$$

where $\mathbb{E}\{\tilde{\mathbf{g}}_{m,u}^{(p)} \tilde{\mathbf{g}}_{m,u}^{(q)H}\} = \mathbf{T}_{\tilde{g}\tilde{g},m} = \sigma_d^2 \mathbf{F}_{\text{RF},m,u} \mathbf{F}_{\text{BB},m,u} \mathbf{F}_{\text{BB},m,u}^H \mathbf{F}_{\text{RF},m,u}^H$. Expanding the circular convolution, as defined in Section-I-C, the above equation can be rewritten as

$$\hat{y}_m(q) = \mathbf{W}_{\text{RF},m}^H \sum_{u=1}^U [\mathbf{H}_u(0) \tilde{\mathbf{g}}_{m,u}(q) + \mathbf{H}_u(1) \tilde{\mathbf{g}}_{m,u}(q-1) + \dots + \mathbf{H}_u(K-1) \tilde{\mathbf{g}}_{m,u}(q-K+1)] + \mathbf{W}_{\text{RF},m}^H \mathbf{v}_m(q). \quad (52)$$

Let the received signal's covariance matrix $\tilde{\mathbf{J}}_m \in \mathbb{C}^{N_{\text{RF}}^B \times N_{\text{RF}}^B}$ be defined as $\tilde{\mathbf{J}}_m = \mathbb{E}\{\hat{y}_m(q) \hat{y}_m^H(q)\}$. After substituting Eq. (52) in $\tilde{\mathbf{J}}_m$, one obtains

$$\tilde{\mathbf{J}}_m = \mathbf{W}_{\text{RF},m}^H \sum_{u=1}^U [\mathbf{H}_u(0) \mathbf{T}_{\tilde{g}\tilde{g},m} \mathbf{H}_u^H(0) + \dots + \mathbf{H}_u(K-1) \mathbf{T}_{\tilde{g}\tilde{g},m} \mathbf{H}_u^H(K-1)] \mathbf{W}_{\text{RF},m}^H + \sigma_n^2 \mathbf{W}_{\text{RF},m}^H \mathbf{W}_{\text{RF},m}^H,$$

which can be compactly written as $\tilde{\mathbf{J}}_m = \mathbf{W}_{\text{RF},m}^H \mathbf{L}_m \mathbf{W}_{\text{RF},m}^H + \sigma_n^2 \mathbf{W}_{\text{RF},m}^H \mathbf{W}_{\text{RF},m}^H$, where we have $\mathbf{L}_m = \sum_{u=1}^U \sum_{n=0}^{K-1} \mathbf{H}_u(n) \mathbf{T}_{\tilde{g}\tilde{g},m} \mathbf{H}_u^H(n)$. Therefore, the noise covariance matrix $\mathbf{J}_m \in \mathbb{C}^{N_{\text{RF}}^B \times N_{\text{RF}}^B}$ after passing through a low-resolution ADC [24] can be expressed as $\mathbf{J}_m = \varkappa(1 - \varkappa) \text{diag}(\tilde{\mathbf{J}}_m)$.

APPENDIX B

CALCULATION OF MAXIMUM ARRAY GAIN

Substituting Eq. (9) into (8), one obtains

$$\Upsilon = \frac{1}{N} \left| \left[\begin{array}{c} 1 \\ e^{-j\pi \varrho_k \sin \theta_k} \\ \vdots \\ e^{-j\pi(N-1)\varrho_k \sin \theta_k} \end{array} \right]^H \left[\begin{array}{c} 1 \\ e^{-j\pi \sin \theta} \\ \vdots \\ e^{-j\pi(N-1) \sin \theta} \end{array} \right] \right|. \quad (53)$$

Further simplifying the above equation and assuming θ is small, we obtain

$$\Upsilon = \frac{1}{N} \left| \sum_{n=0}^{N-1} e^{jn\pi(\varrho_k\theta_k - \theta)} \right|. \quad (54)$$

Therefore, in order to achieve maximum array gain, $\theta_k = \frac{\theta}{\varrho_k}$.

APPENDIX C CALCULATION OF OPTIMAL ANGLE

Revisiting Eq. (8) and substituting the array response vector as given in Eq. (37) into the normalized array gain expression of Eq. (8), one obtains

$$\Upsilon_k = |(\tilde{\mathbf{a}}(\theta, f_c))^H \tilde{\mathbf{a}}_k(\theta_k, f_k)|, \quad (55)$$

which can be further simplified to Eq. (56)-(58).

$$\Upsilon_k = \frac{1}{N} \left| \sum_{s=1}^S e^{j\pi(s-1)\{P\theta - \varrho_k P\theta_k - v_k\}} \sum_{p=1}^P e^{-j\pi(p-1)\{-\theta + \varrho_k\theta_k\}} \right|. \quad (58)$$

Therefore, the normalized array gain can be represented in the form of the Dirichlet sinc function, as

$$\Upsilon_k = \frac{1}{N} \left| \Xi_S(P\theta - \varrho_k P\theta_k - v_k) \Xi_P(-\theta + \varrho_k\theta_k) \right|. \quad (59)$$

Thus, the optimal angle at which the array gain is maximum can be derived by substituting $P\theta - \varrho_k P\theta_k - v_k = 0$ into Eq. (59) which finally yields Eq. (40).

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] K. Venugopal, A. Alkhatieb, N. G. Prelcic, 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.
- [3] H. Sarridedden, 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, G. Y. Li, S. Sun, and T. S. Rappaport, "Spatial-wideband effect in massive MIMO with application in mmWave systems," *IEEE Communications Magazine*, pp. 134–141, 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*, vol. 72, no. 7, pp. 4384–4400, 2024.
- [6] I. S. . 3d 2017, "IEEE standard for high data rate wireless multimedia networks—amendment 2: 100 Gb/s wireless switched point-to-point physical layer (standard)," 2017.
- [7] S. Tarboush, H. Sarridedden, M.-S. Alouini, and T. Y. Al-Naffouri, "Single-versus multicarrier terahertz-band communications: A comparative study," *IEEE Open Journal of the Communications Society*, vol. 3, pp. 1466–1486, 2022.
- [8] R. Piesiewicz, C. Jansen, D. Mittleman, T. Kleine-Ostmann, M. Koch, and T. Kurner, "Scattering analysis for the modeling of THz communication systems," *IEEE Transactions on Antennas and Propagation*, vol. 55, no. 11, pp. 3002–3009, 2007.
- [9] 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 2011-Sustainable Wireless Technologies*, 2011.
- [10] 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.
- [11] Q. Wan, J. Fang, Z. Chen, and H. Li, "Hybrid precoding and combining for millimeter wave/sub-THz MIMO-OFDM systems with beam squint effects," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 8, pp. 8314–8319, 2021.
- [12] F. Gao, B. Wang, C. Xing, J. An, and G. Y. Li, "Wideband beamforming for hybrid massive MIMO THz communications," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 6, pp. 1725–1740, 2021.
- [13] L. Dai, J. Tan, Z. Chen, and H. V. Poor, "Delay-phase precoding for wideband THz massive MIMO," *IEEE Transactions on Wireless Communications*, vol. 21, no. 9, pp. 7271–7286, 2022.
- [14] Y. Omid, S. M. Shahabi, C. Pan, Y. Deng, and A. Nallanathan, "Robust beamforming design for an IRS-aided NOMA communication system with CSI uncertainty," *IEEE Transactions on Wireless Communications*, vol. 23, no. 2, pp. 874–889, 2023.
- [15] M. Jafri, A. Anand, S. Srivastava, A. K. Jagannatham, and L. Hanzo, "Robust Distributed Hybrid Beamforming in Coordinated Multi-User Multi-Cell mmWave MIMO Systems Relying on Imperfect CSI," *IEEE Transactions on Communications*, vol. 70, no. 12, pp. 8123–8137, 2022.
- [16] D. P. Wipf and B. D. Rao, "Sparse bayesian learning for basis selection," *IEEE Transactions on Signal processing*, pp. 2153–2164, 2004.
- [17] 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.
- [18] 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.
- [19] 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.
- [20] M. Cui, Z. Zhang, L. Dai, and K. Huang, "Ice-filling: Near-optimal channel estimation for dense array systems," *IEEE Transactions on Wireless Communications*, 2025.
- [21] 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.
- [22] 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.
- [23] O. T. Demir and E. Bjornson, "The bussgang decomposition of nonlinear systems: Basic theory and mimo extensions [lecture notes]," *IEEE Signal Processing Magazine*, vol. 38, no. 1, pp. 131–136, 2020.
- [24] S. Jacobsson, G. Durisi, M. Coldrey, U. Gustavsson, and C. Studer, "Throughput analysis of massive MIMO uplink with low-resolution ADCs," *IEEE Transactions on Wireless Communications*, vol. 16, no. 6, pp. 4038–4051, 2017.
- [25] 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.
- [26] Y. Wang, W. Xu, H. Zhang, and X. You, "Wideband mmWave channel estimation for hybrid massive MIMO with low-precision ADCs," *IEEE Wireless Communications Letters*, vol. 8, no. 1, pp. 285–288, 2018.
- [27] M. Ma, N. T. Nguyen, I. Atzeni, and M. Juntti, "Hybrid receiver design for massive MIMO-OFDM with low-resolution ADCs and oversampling," in *2025 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE, 2025, pp. 1–6.
- [28] J. P. González-Coma, J. Rodríguez-Fernandez, N. Gonzalez-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*, vol. 12, no. 2, pp. 353–367, 2018.
- [29] W. Tang, M. Z. Chen, X. Chen, J. Y. Dai, Y. Han, M. Di Renzo, Y. Zeng, S. Jin, Q. Cheng, and T. J. Cui, "Wireless communications with reconfigurable intelligent surface: Path loss modeling and experimental measurement," *IEEE Transactions on Wireless Communications*, vol. 20, no. 1, pp. 421–439, 2020.
- [30] C. A. Balanis, *Antenna Theory: Analysis and Design*. John Wiley & sons, 2015.
- [31] C. Hill. (1999) HITRAN Online Interface to the HITRAN Database. [Online]. Available: <https://hitran.org/>
- [32] R. Piesiewicz, C. Jansen, S. Wietzke, D. Mittleman, M. Koch, and T. Kurner, "Properties of building and plastic materials in the THz range," *International Journal of Infrared and Millimeter Waves*, vol. 28, pp. 363–371, 2007.
- [33] S. Srivastava, P. Sharma, S. Dwivedi, A. K. Jagannatham, and L. Hanzo, "Fast block LMS based estimation of angularly sparse channels for single-carrier wideband millimeter wave hybrid MIMO systems," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 1, pp. 666–681, 2021.

$$\Upsilon_k = \left| \frac{1}{N} \left\{ (1 + e^{j\pi\theta} e^{-j\pi\varrho_k\theta_k} + \dots + e^{j\pi(P-1)\theta} e^{-j\pi(P-1)\varrho_k\theta_k}) + e^{-j\pi v_k} (1 + e^{j\pi P\theta} e^{-j\pi P\varrho_k\theta_k} + \dots + e^{j\pi(2P-1)\theta} e^{-j\pi(2P-1)\varrho_k\theta_k}) \right. \right. \\ \left. \left. + \dots + e^{-j\pi(S-1)v_k} (1 + e^{j\pi(S-1)P\theta} e^{-j\pi(S-1)P\varrho_k\theta_k} + \dots + e^{j\pi(SP-1)\theta} e^{-j\pi(SP-1)\varrho_k\theta_k}) \right\} \right|, \quad (56)$$

$$\Upsilon_k = \frac{1}{N} \left| \sum_{m=1}^S \sum_{p=1}^P e^{j\pi[(s-1)P+(s-1)]\theta} e^{-j\pi\varrho_k[(s-1)P+(p-1)]\theta_k} e^{-j\pi(s-1)v_k} \right|, \quad (57)$$

- [34] L. Liu, C. Huang, Y. Chi, C. Yuen, Y. L. Guan, and Y. Li, "Sparse vector recovery: Bernoulli-Gaussian message passing," in *GLOBECOM 2017-2017 IEEE Global Communications Conference*. IEEE, 2017.
- [35] A. Rajoriya and R. Budhiraja, "Joint AMP-SBL algorithms for device activity detection and channel estimation in massive MIMO mMTC systems," *IEEE Transactions on Communications*, vol. 71, no. 4, pp. 2136–2152, 2023.
- [36] T. Park and G. Casella, "The Bayesian lasso," *Journal of the american statistical association*, vol. 103, no. 482, pp. 681–686, 2008.
- [37] 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.
- [38] S. M. Kay, "Statistical signal processing: Estimation Theory," *Prentice Hall*, vol. 1, pp. Chapter–3, 1993.
- [39] A. Garg, A. Kumar, S. Srivastava, N. Yadav, A. K. Jagannatham, and L. Hanzo, "Bayesian Learning Aided Simultaneous Sparse Estimation of Dual-Wideband THz Channels in Multi-User Hybrid MIMO Systems," *IEEE Transactions on Vehicular Technology*, pp. 1–17, 2025.
- [40] X. Meng, C. Chen, and B. Chi, "A compact 3–20-GHz 106-ps passive true-time delay circuit in 65-nm CMOS," *IEEE Microwave and Wireless Technology Letters*, vol. 34, no. 5, pp. 493–496, 2024.
- [41] M. Jung and B.-W. Min, "A compact 3–30-GHz 68.5-ps CMOS true-time delay for wideband phased array systems," *IEEE Transactions on Microwave Theory and Techniques*, vol. 68, no. 12, 2020.
- [42] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in mmwave MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 13, no. 3, pp. 1499–1513, 2014.
- [43] J. Li, L. Xiao, X. Xu, and S. Zhou, "Robust and low complexity hybrid beamforming for uplink multiuser mmWave MIMO systems," *IEEE Communications Letters*, vol. 20, no. 6, pp. 1140–1143, 2016.
- [44] 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.
- [45] R. Niazadeh, M. Babaie-Zadeh, and C. Jutten, "On the achievability of Cramér–Rao bound in noisy compressed sensing," *IEEE Transactions on Signal Processing*, vol. 60, no. 1, pp. 518–526, 2011.
- [46] Z. Ben-Haim and Y. C. Eldar, "The Cramér–Rao bound for estimating a sparse parameter vector," *IEEE Transactions on Signal Processing*, vol. 58, no. 6, pp. 3384–3389, 2010.