# Terahertz-Band Channel and Beam Split Estimation via Array Perturbation Model

**AHMET M. ELBIR** [1] **(Senior Member, IEEE), WEI SHI** [2] **(Member, IEEE),**
**ANASTASIOS K. PAPAZAFEIROPOULOS** [3] **(Senior Member, IEEE),**
**PANDELIS KOURTESSIS** [3] **(Member, IEEE), AND SYMEON CHATZINOTAS** [1] **(Fellow, IEEE)**

[1]Interdisciplinary Centre for Security, Reliability and Trust, University of Luxembourg, 1855 Esch-sur-Alzette, Luxembourg

[2]School of Information Technology, Carleton University, Ottawa, ON K1S 5B6, Canada

[3]Communications and Intelligent Systems Research Group, University of Hertfordshire, AL10 9AB Hatfield, U.K.

CORRESPONDING AUTHOR: A. M. ELBIR (e-mail: ahmetmelbir@ieee.org)

**ABSTRACT** For the demonstration of ultra-wideband bandwidth and pencil-beamforming, the terahertz (THz)-band has been envisioned as one of the key enabling technologies for the sixth generation networks. However, the acquisition of the THz channel entails several unique challenges such as severe path loss and beam-split. Prior works usually employ ultra-massive arrays and additional hardware components comprised of time-delayers to compensate for these loses. In order to provide a cost-effective solution, this paper introduces a sparse-Bayesian-learning (SBL) technique for joint channel and beam-split estimation. Specifically, we first model the beam-split as an array perturbation inspired from array signal processing. Next, a low-complexity approach is developed by exploiting the line-of-sight-dominant feature of THz channel to reduce the computational complexity involved in the proposed SBL technique for channel estimation (SBCE). Additionally, based on federated-learning, we implement a model-free technique to the proposed model-based SBCE solution. Further to that, we examine the near-field considerations of THz channel, and introduce the range-dependent near-field beam-split. The theoretical performance bounds, i.e., Cramér-Rao lower bounds, are derived both for near- and far-field parameters, e.g., user directions, beam-split and ranges. Numerical simulations demonstrate that SBCE outperforms the existing approaches and exhibits lower hardware cost.

**INDEX TERMS** Terahertz, channel estimation, beam split, sparse Bayesian learning, near-field, federated learning.

## I. INTRODUCTION

THE DEFINITION of the THz band varies among different IEEE communities. While IEEE Terahertz Technology and Applications Committee focus on $0.3 - 3$ THz band, the IEEE Transactions on Terahertz Science and Technology Journal studies $0.3 - 10$ THz [1], [2]. On the other hand, $0.1 - 10$ THz band has been used for THz band in the recent studies on wireless communications including a large overlap with millimeter wave (mmWave), e.g., $0.03 - 0.3$ THz [1]. Furthermore, the Federal Communication Commission (FCC) has designated the frequency bands of $116 - 123$ GHz, $174.8 - 182$ GHz, $185 - 190$ GHz and $244 - 246$ GHz for the unlicensed use for the future spectrum horizon [3].

Compared to its mmWave counterpart, the THz-band channel exhibits several THz-specific challenges that should be taken into account. These unique THz features, among others, include high path loss due to spreading loss and molecular absorption, shorter transmission rage, distance-dependent bandwidth and beam-split (see, e.g., Fig. 1) [1], [4], [5]. Furthermore, the THz channels are extremely-sparse. Hence, they are usually characterized as
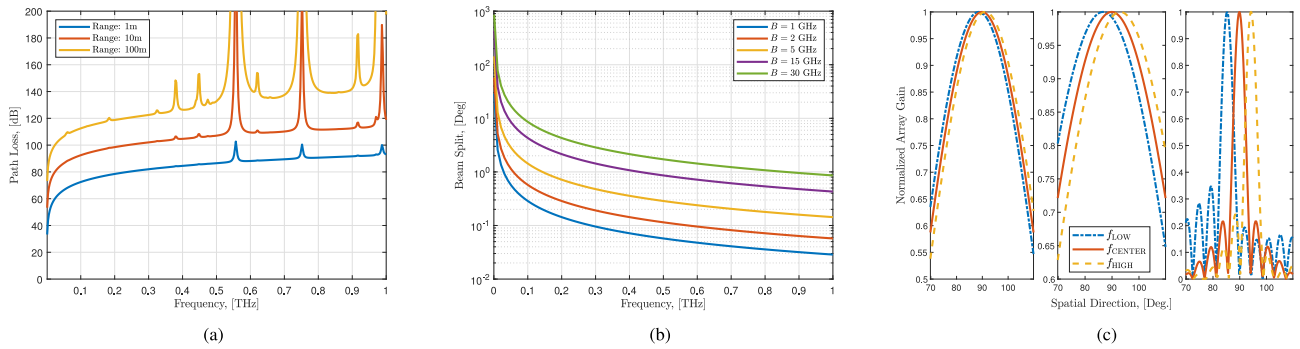
**FIGURE 1.** THz-band characteristics: (a) path loss (in dB) due to molecular absorption for various transmission ranges, (b) beam-split (in degrees) with respect to the carrier frequency $f_c \in [50, 1000]$ GHz for different bandwidths, and (c) normalized array gain with respect to spatial direction at low, center and high end subcarriers for (Left) 3.5 GHz, $B = 0.1$ GHz; (Middle) 28 GHz, $B = 2$ GHz; and (Right) 300 GHz, 30 GHz, respectively.

line-of-sight (LoS)-dominant and non-LoS (NLoS)-assisting models [1], [6], [7], [8], [9], [10], [11]. In comparison, the mmWave channel involves both LoS and NLoS paths, which are significant while LoS path is dominant in THz model [12]. In addition, to compensate the severe path loss, analogous to massive multiple-input multiple-output (MIMO) array in mmWave design, ultra-massive (UM) MIMO array configurations are proposed for THz [2], [13]. The UM-MIMO design comprises huge number of antennas which are densely-positioned (e.g., $5 \times 5$ cm$^2$) due to extremely short wavelength [5]. As a result, the usage of dedicated radio-frequency (RF) chain for each antenna element involves extreme hardware cost and labor. The hybrid design configurations, i.e., joint usage of analog and digital beamformers seems to be major possible solution as it was first envisioned for mmWave massive MIMO systems in 5G [14], [15], [16], [17], [18]. In order to exploit low cost system design, the hybrid architecture involve a few (large) number of digital (analog) beamformers. Although the digital beamformers are subcarrier-dependent (SD), the analog phase shifters are designed as subcarrier-independent (SI) components. This causes *beam-split* phenomenon, i.e., the beams generated by the analog beamformer according to a single frequency split at different subcarriers and they look at different directions (Fig. 1) because of ultra-wide bandwidth and large number of antennas [7], [19]. In mmWave, at which the subcarrier frequencies are relatively closer than THz, *beam-squint* is the term broadly used to describe the same phenomenon [12], [20], [21]. The beam-split affects the system performance and causes severe degradations in terms of spectral efficiency, normalized mean-squared-error (NMSE) and bit-error-rate (BER). For instance, the beam-split is approximately 4° (0.4°) for 0.3 THz with 30 GHz (60 GHz with 1 GHz) bandwidth, respectively for a broadside user [5]. Fig. 1b shows beam-split (in degrees) with respect to the carrier frequency for different bandwidths. Large beam-split occurs at lower frequencies since the bandwidth is relatively larger with respect to the carrier frequency, and it becomes smaller as the carrier frequency increases.

With the aforementioned THz-specific features, THz channel estimation is regarded as an even more challenging problem than that of mmWave. In prior studies, THz channel estimation has been investigated in [19], [22], [23], [24], [25], [26], [27], [28], [29], [30]. However, most of these works either ignore the effect of beam-split [22], [23], [24], [25], [31] or consider only the narrowband case [26], [27], [28], without exploiting the ultra-wide bandwidth property, the key reason of climbing to THz-band. Further to that, an machine learning (ML)-based SBL architecture used in [32] for mmWave channel estimation in the presence of beam-squint assumes SI steering vectors to construct the wideband channel. Hence, its performance is limited and, indeed, no performance improvement was observed by enhancing the grid resolution. In fact, the beam-split compensation requires certain signal processing or hardware techniques to be handled properly.

Despite the prominence of THz channel estimation, in the literature, there are only a few recent works on beam-split mitigation. The existing solutions are categorized into two classes, i.e., hardware-based techniques [29] and algorithmic methods [19], [30]. The first category of solutions consider employing time delayer (TD) networks together with phase shifters to realize virtual SD analog beamformers to mitigate beam-split. In particular, [29] devises a generalized simultaneous orthogonal matching pursuit (GSOMP) technique by exploiting the SD information collected via TD network hence achieves close to minimum MSE (MSE) performance. However, these solutions require additional hardware, i.e., each phase shifter is connected to multiple TDs, each of which consumes approximately 100 mW, which is more than that of a phase shifter (40 mW) in THz [5]. The second category of solutions do not employ additional hardware components. Instead, advanced signal processing techniques have been proposed to compensate beam-split. Specifically, an OMP-based beam-split pattern detection (BSPD) approach was proposed in [19] for the recovery of the support pattern among all subcarriers in the beamspace and construct one-to-one match between the physical and

spatial (i.e., deviated due to beam-split in the beamspace) directions. Also, [33] proposed an angular-delay rotation method, which suffers from coarse beam-split estimation and high training overhead due to the use of complete discrete Fourier transform (DFT) matrix. In [30], a beamspace support alignment (BSA) technique was introduced to align the deviated spatial beam directions among the subcarriers. Although both BSPD and BSA are based on OMP, the latter exhibits lower NMSE for THz channel estimation. Nevertheless, both methods suffer from inaccurate support detection and low precision for estimating the physical channel directions.

Due to short transmission range in THz, near-field spherical-wave models are considered for THz applications [5], [34]. In far-field model, the transmitted signal arrives at the users as plane-wave whereas the plane wave-front is spherical in the near-field if the transmission range is shorter than the Fraunhofer distance [34], [35]. Thus, the near-field effects should be taken into account for accurate channel modeling. In [34], THz near-field beamforming problem has been considered and TD network-based approach was proposed while the THz channel was assumed to be known a prior. In a different line of research, the beam-squint problem is investigated in [36] for the massive MIMO system with low resolution analog-digital converters.

In this paper, THz channel estimation in the presence of beam-split is examined. By exploiting the extreme sparsity of THz channel, we first approach the problem from the sparse recovery optimization perspective. Then, we introduce a novel approach to model the beam-split as an array perturbation as inspired by the array imperfection models, e.g., mutual coupling, gain-phase mismatch, in array signal processing context [37], [38], [39]. The array perturbation model allows us to establish a linear transformation between the nominal (constructed via physical directions) and actual (beam-split corrupted) steering vectors. Next, we propose a sparse Bayesian learning (SBL) approach to jointly estimate both THz channel and beam-split. The SBL method has been shown to be effective for sparse signal reconstruction from underdetermined observations [40], [41], [42]. Compared to other existing signal estimation techniques, e.g., *mu*ltiple *si*gnal *c*lassification (MUSIC) [43] and the $\ell_p$-norm ($0 \leq p \leq 1$) techniques such as compressed sensing (CS) [44], SBL outperforms these techniques in terms of precision and convergence [45]. Although SBL has been widely used for both channel estimation [39] and direction-of-arrival (DoA) estimation [38], [46], the proposed SBL-based channel estimation (SBCE) approach differentiates from the these studies by jointly estimating multiple hyperparameters, e.g., physical channel directions and beam-split as well as array the perturbation-based beam-split model. The proposed SBCE approach is advantageous in terms of complexity since it does not require any additional hardware components as in TD-based works, and it exhibits close to optimum performance for both THz channel and beam-split estimation. Furthermore, the estimation

of beam-split allows one to design the hybrid beamformers in massive/UM MIMO systems more accurately in a simpler way. The main contributions of this work are summarized as follows:

1) We propose a novel approach to model beam-split as an array perturbation, and design a transformation matrix between the nominal and actual steering vectors in order to estimate both THz channel and beam-split.

2) An SBL approach is devised to jointly estimate the physical channel directions and beam-split. In order to reduce the computational complexity involved in SBL iterations, we exploit the LoS-dominant feature of THz channel and design the array perturbation matrix based on a single unknown parameter, i.e., the beam-split. Thus, instead of designing the array perturbations via a full matrix with distinct elements, we consider a diagonal structure, that can be easily represented via only beam-split knowledge.

3) We also propose a model-free approach based on federated learning (FL) to ease the communication overhead while maintaining satisfactory NMSE performance. Different from our preliminary work in [30], we consider SBCE method for data labeling, hence it is called sparse Bayesian FL (SBFL).

4) We investigate the near-field considerations for THz transmission, and derive the near-field beam-split, which is both range- and direction-dependent.

5) In addition, the theoretical performance bounds for both physical channel directions and beam-split estimation are examined and the corresponding Cramér-Rao lower bounds (CRBs) have been derived.

*Paper Organization:* The rest of the paper is organized as follows. In Section II, we present the array signal model and formulate the THz channel estimation problem. Next, our proposed SBCE approach is introduced in Section III. The near-field model for beam-split and the proposed SBFL approach are given in Sections IV and V, respectively. We present extensive numerical simulations in Section VI, and finalize the paper with conclusions in Section VII.

*Notation:* $(\cdot)^{\mathsf{T}}$ and $(\cdot)^{\mathsf{H}}$ to represent the transpose and conjugate-transpose operations, respectively. For a matrix $\mathbf{A}$; $[\mathbf{A}]_{ij}$ and $[\mathbf{A}]_k$ correspond to the $(i,j)$th entry and $k$th column while $\mathbf{A}^{\dagger}$ denotes the Moore-Penrose pseudo-inverse of $\mathbf{A}$. A unit matrix of size $N$ is represented by $\mathbf{I}_N$, $\nabla$ represents the gradient operation, $\xi(a) = \frac{\sin N\pi a}{N \sin \pi a}$ is the Dirichlet sinc function, and Tr$\{\cdot\}$ stands for the trace operation. $||\cdot||_0$, $||\cdot||_1$, $||\cdot||_2$ and $||\cdot||_{\mathcal{F}}$ denote the $\ell_0$, $\ell_1$, $\ell_2$ and Frobenius norms, respectively.

## II. SIGNAL MODEL

We consider a downlink scenario for multi-user wideband MIMO system with OFDM (orthogonal frequency division multiplexing) employing $M$ subcarriers. We assume that the base station (BS) employs $N_{\mathrm{T}}$ antennas together with $N_{\mathrm{RF}}$ RF chains to communicate with $K$ single-antenna users. Hence, the data symbols, $\mathbf{s}[m] = [s_1[m], \ldots, s_K[m]]^{\mathsf{T}}$

$(m \in \mathcal{M} = \{1, \ldots, M\})$ are precoded via the SD baseband beamformer $\mathbf{F}_{BB}[m] \in \mathbb{C}^{K \times K}$. The resulting signal $\mathbf{F}_{BB}[m]\mathbf{s}[m] \in \mathbb{C}^{K}$ is constructed in the time domain by employing an $M$-point inverse DFT (IDFT). Then, the SI analog beamformers $\mathbf{F}_{RF} \in \mathbb{C}^{N_T \times N_{RF}}$ ($N_{RF} = K < N_T$) are used to process the cyclic-prefix (CP)-added signal. The analog beamformers, which are SI and realized with phase-shifters, have constant-modulus constraint, i.e., $|[\mathbf{F}_{RF}]_{i,j}| = \frac{1}{\sqrt{N_T}}$ for $i = 1, \ldots, N_{RF}$, $j = 1, \ldots, N_T$, and the constraint $\sum_{m=1}^{M} ||\mathbf{F}_{RF}\mathbf{F}_{BB}[m]||_{\mathcal{F}}^2 = MK$ enforces the transmission power. We can define the $N_T \times 1$ transmitted signal as

$$\mathbf{t}[m] = \mathbf{F}_{RF}\mathbf{F}_{BB}[m]\mathbf{s}[m], \quad (1)$$

which is then received at the $k$th user as

$$y_k[m] = \mathbf{h}_k^{\mathsf{T}}[m]\mathbf{t}[m] + w_k[m], \quad (2)$$

where $w_k[m] \in \mathbb{C}$ corresponds to the additive white Gaussian noise (AWGN), and we have $w_k[m] \sim \mathcal{CN}(0, \mu^2)$.

## A. THz CHANNEL MODEL

The THz channel is modeled by combination of a single LoS path together with a few NLoS paths, which are weak due to reflection, scattering and refraction [1], [6], [7], [8]. In addition, the ray-tracing techniques for THz communication also assume the extreme sparsity of the THz channel and state that the channel is dominated by the LoS component for the graphene-based nano-transceivers [1] while the other channel models, e.g., 3GPP channel model [47], [48] are also widely used for THz transmission. Then, the THz delay-$d$ MIMO channel is given in time domain as

$$\bar{\mathbf{h}}_k(d) = \sqrt{\frac{N_T}{L}} \left( \underbrace{p(dT_s - \tau_{k,1})\alpha_k^1 \mathbf{a}(\vartheta_{k,1})}_{\text{LoS}} \right.$$

$$\left. + \underbrace{\sum_{l=2}^{L} p(dT_s - \tau_{k,l})\alpha_k^l \mathbf{a}(\vartheta_{k,l})}_{\text{NLoS}} \right), \quad (3)$$

where $\alpha_k^l \in \mathbb{C}$ denotes the complex path gain and $L$ is the total number of paths. The time delay of the $l$th path is denoted by $\tau_{k,l}$, for which $l = 1$ corresponds to the LoS path, and $p(\cdot)$ denotes the pulse-shaping function for $T_s$-signaling. Also, $\mathbf{a}(\vartheta_{k,l}) \in \mathbb{C}^{N_T}$ denotes the array steering vector corresponding to the physical direction for the $l$th path of the $k$th user $\vartheta_{k,l} = \sin\tilde{\vartheta}_{k,l}$ with $\tilde{\vartheta}_{k,l} \in [-\frac{\pi}{2}, \frac{\pi}{2}]$. In particular, for a uniform linear array (ULA),[1] the steering vector $\mathbf{a}(\vartheta_{k,l}) \in \mathbb{C}^{N_T}$ is defined as

$$\mathbf{a}(\vartheta_{k,l}) = \frac{1}{\sqrt{N_T}} \left[ 1, e^{j2\pi \frac{\bar{d}}{\lambda_c}\vartheta_{k,l}}, \ldots, e^{j2\pi \frac{\bar{d}}{\lambda_c}(N_T-1)\vartheta_{k,l}} \right]^{\mathsf{T}}, \quad (4)$$

1. Although ULA is considered in this work, the proposed approach can straightforwardly be extended for other array geometries, e.g., uniform rectangular array (URA), array-of-subarray (AoSA) [1], [13] or group-of-subarrays (GoSA) [2].

where $\lambda_c = \frac{c_0}{f_c}$, for which $c_0$ is speed of light and $f_c$ denotes the carrier frequency. Furthermore, $\bar{d} = \frac{\lambda_c}{2}$ the half-wavelength array element spacing. Performing $M$-point DFT of the delay-$d$ channel in (3) yields

$$\mathbf{h}_k[m] = \sum_{d=0}^{D-1} \bar{\mathbf{h}}_k(d)e^{-j\frac{2\pi m}{M}d},$$

$$= \sqrt{\frac{N_T}{L}} \sum_{l=1}^{L} \left( \alpha_k^l \mathbf{a}(\vartheta_{k,l}) \sum_{d=0}^{D-1} p(dT_s - \tau_{k,l})e^{-j\frac{2\pi m}{M}d} \right),$$

$$= \sqrt{\frac{N_T}{L}} \sum_{l=1}^{L} \alpha_k^l \mathbf{a}(\vartheta_{k,l})e^{-j2\pi\tau_{k,l}f_m} \quad (5)$$

where $D \leq M$ is the CP length [22], [49]. Define the $m$th subcarrier frequency as $f_m = f_c + \frac{B}{M}(m - 1 - \frac{M-1}{2})$, where $B$ is the bandwidth.

Due to large bandwidth in THz systems, the channel direction $\vartheta_{k,l}$ becomes SD in spatial domain since $f_m \neq f_c$ [7], [19], [50]. Therefore, the physical direction $\vartheta_{k,1}$ deviates to the spatial direction $\theta_{k,m,l}$, which is defined as

$$\theta_{k,m,l} = \frac{2f_m}{c_0}\bar{d}\vartheta_{k,l} = \frac{f_m}{f_c}\vartheta_{k,l} = \eta_m\vartheta_{k,l}. \quad (6)$$

Notice that $\vartheta_{k,l} = \theta_{k,m,l}$ if $f_m = f_c$, i.e., the narrowband scenario. Therefore, we define the channel vector of the $k$th user at the $m$th subcarrier in frequency domain as

$$\mathbf{h}_k[m] = \sqrt{\frac{N_T}{L}} \sum_{l=1}^{L} \alpha_k^l \mathbf{a}'(\theta_{k,m,l})e^{-j2\pi\tau_{k,l}f_m}, \quad (7)$$

where $\mathbf{a}'(\theta_{k,m,l}) \in \mathbb{C}^{N_T}$ denotes the SD array steering vector under the effect of beam-split and it is defined as

$$\mathbf{a}'(\theta_{k,m,l}) = \frac{1}{\sqrt{N_T}} \left[ 1, e^{j2\pi \frac{\bar{d}}{\lambda_m}\vartheta_{k,l}}, \ldots, e^{j2\pi \frac{\bar{d}}{\lambda_m}(N_T-1)\vartheta_{k,l}} \right]^{\mathsf{T}},$$

$$= \frac{1}{\sqrt{N_T}} \left[ 1, e^{j\pi\theta_{k,m,l}}, \ldots, e^{j\pi(N_T-1)\theta_{k,m,l}} \right]^{\mathsf{T}}. \quad (8)$$

Then, the channel model in (7) can be expressed in a compact form as

$$\mathbf{h}_k[m] = \mathbf{A}_k'[m]\tilde{\mathbf{x}}_k[m], \quad (9)$$

where $\mathbf{A}_k'[m] = [\mathbf{a}'(\theta_{k,m,1}), \ldots, \mathbf{a}'(\theta_{k,m,L})] \in \mathbb{C}^{N_T \times L}$ and we have

$$\tilde{\mathbf{x}}_k[m] = \left[ \alpha_k^1 e^{-j2\pi\tau_{k,1}f_m}, \ldots, \alpha_k^L e^{-j2\pi\tau_{k,L}f_m} \right]^{\mathsf{T}} \in \mathbb{C}^L. \quad (10)$$

In the following lemma, we show that the beam-split causes deviations in the channel directions such that maximum array gain is achieved at the beam-split deviated directions.

*Lemma 1:* Let $\mathbf{a}'(\theta_m)$ and $\mathbf{a}(\vartheta)$ be the actual and nominal steering vectors for an arbitrary physical direction $\vartheta$ and subcarrier $m \in \mathcal{M}$ as defined in (8) and (4), respectively. Then, $\mathbf{a}(\theta_m)$ achieves the maximum array gain, i.e., $A_G(\vartheta, \theta_m) = \frac{|\mathbf{a}^{\mathsf{H}}(\vartheta)\mathbf{a}'(\theta_m)|^2}{N_T^2}$, if $\theta_m = \eta_m\vartheta$.

*Proof:* Please see Appendix A. ∎

Notice that both spatial and physical directions converges as $\theta_{k,m,l} \approx \vartheta_{k,l}$ if $f_m \approx f_c$, as shown in Fig. 1(c). This allows one to design SI analog beamformers ($\mathbf{F}_{\mathrm{RF}}$) for $m \in \mathcal{M}$ in conventional wideband mmWave systems [12], [21], [51]. However, beam-split implies that the physical directions $\vartheta_{k,l}$ deviate and completely split from the spatial directions $\theta_{k,m,l}$ as the system bandwidth widens, as shown in Fig. 1(b). Hence, we define the beam-split as the difference between the spatial and physical directions as

$$\Delta_{k,l}[m] = \theta_{k,m,l} - \vartheta_{k,l} = \left( \frac{f_m}{f_c} - 1 \right) \vartheta_{k,l}, \quad (11)$$

which only depends on the frequency ratio $\frac{f_m}{f_c}$ and physical direction $\vartheta_{k,l}$.

### B. PROBLEM FORMULATION

Due to the impact of beam-split, the spatial channel directions differ at each subcarrier. Therefore, a subcarrier-wise channel estimation is required to accurately estimate the channel. In downlink, the channel estimation stage is performed simultaneously by all the users during channel training. Since the BS employs hybrid beamforming architecture, it activates only a single RF chain in each channel use to transmit the known pilot signals during channel acquisition [7], [19], [22]. Then, the BS employs $P$ beamformer vectors as $\tilde{\mathbf{F}}[m] = [\tilde{\mathbf{f}}_1[m], \dots, \tilde{\mathbf{f}}_P[m]] \in \mathbb{C}^{N_{\mathrm{T}} \times P}$ ($|\tilde{\mathbf{f}}_p[m]| = 1/\sqrt{N}$) to send $P$ orthogonal pilots, $\tilde{\mathbf{S}}[m] = \mathrm{diag}\{\tilde{s}_1[m], \dots, \tilde{s}_P[m]\} \in \mathbb{C}^{P \times P}$. Hence, the pilot signals of $P$ frames are used to reconstruct the channel. By combining the received signals at the single-antenna users for $p = 1, \dots, P$, the $P \times 1$ received signal at the $k$th user is

$$\bar{\mathbf{y}}_k[m] = \tilde{\mathbf{S}}[m]\bar{\mathbf{F}}[m]\mathbf{h}_k[m] + \mathbf{w}_k[m], \quad (12)$$

where we have $\bar{\mathbf{F}}[m] = \tilde{\mathbf{F}}^{\mathsf{T}}[m] \in \mathbb{C}^{P \times N_{\mathrm{T}}}$, $\bar{\mathbf{y}}_k[m] = [y_{k,1}[m], \dots, y_{k,P}[m]]^{\mathsf{T}}$ and $\mathbf{w}_k = [w_{k,1}[m], \dots, w_{k,P}[m]]^{\mathsf{T}}$. Defining the identical pilot signals for all subcarriers, i.e., $\mathbf{B} = \bar{\mathbf{F}}[m]$ and $\tilde{\mathbf{S}}[m] = \mathbf{I}_P$ [29], [49], [52], (12) becomes

$$\bar{\mathbf{y}}_k[m] = \mathbf{B}\mathbf{h}_k[m] + \mathbf{w}_k[m]. \quad (13)$$

Estimating the channel from (13) is performed via the traditional least squares (LS) and MMSE techniques as

$$\mathbf{h}_k^{\mathrm{LS}}[m] = \left( \mathbf{B}^{\mathsf{H}}\mathbf{B} \right)^{-1}\mathbf{B}^{\mathsf{H}}\bar{\mathbf{y}}_k[m],$$

$$\mathbf{h}_k^{\mathrm{MMSE}}[m] = \left( \mathbf{R}_k^{-1}[m] + \mathbf{B}^{\mathsf{H}}\mathbf{R}_k^{-1}[m]\mathbf{B} \right)^{-1}\mathbf{B}^{\mathsf{H}}\bar{\mathbf{y}}_k[m], \quad (14)$$

where $\mathbf{R}_k[m] = \mathbb{E}\{\mathbf{h}_k[m]\mathbf{h}_k^{\mathsf{H}}[m]\}$ is the channel covariance matrix. The estimators in (14) may need additional priori information, e.g., covariance matrix or require $P \geq N_{\mathrm{T}}$, which can be computationally prohibitive for the computation of matrix inversions as well as entailing high channel training overhead and time for large arrays. Further to that, the LS approach does not take into account the effect of beam-split, and prior information on the channel covariance is needed in MMSE estimator.

Hence, our goal is to efficiently estimate $\mathbf{h}_k[m]$ for $m \in \mathcal{M}$ in the presence of beam-split when the number of received pilot signals is small. To this end, we exploit the sparsity of the THz channel and introduce an SBL-based approach in the following.

## III. SBL FOR THz CHANNEL ESTIMATION

The SBL method has been shown to be effective for sparse signal reconstruction. In this section, we first present the sparse THz signal model and introduce the proposed array perturbation model and an efficient approach for joint THz channel and beam-split estimation.

### A. SPARSE THz CHANNEL MODEL

Due to employing UM number of antennas to compensate the significant path loss in THz frequencies, the channel is *extremely-sparse* ($L \ll N_{\mathrm{T}}$) in the beamspace [6], [13]. In order to exploit the sparsity of THz-band, the channel in (7) is expressed as

$$\mathbf{h}_k[m] = \mathbf{D}\mathbf{x}_k[m], \quad (15)$$

where $\mathbf{x}_k[m] \in \mathbb{C}^N$ is an $L$-sparse vector, i.e., $||\mathbf{x}[m]||_0 = L$ and $\mathbf{D} = [\mathbf{a}(\phi_1), \dots, \mathbf{a}(\phi_N)] \in \mathbb{C}^{N_{\mathrm{T}} \times N}$ denotes an overcomplete dictionary matrix composed of the steering vectors denoted by $\mathbf{a}(\phi_n) \in \mathbb{C}^{N_{\mathrm{T}}}$ with $\phi_n = \frac{2n-N-1}{N}$ for $n = 1, \dots, N$. Hence, the resolution of $\mathbf{D}$ is $\rho = 1/N$.

Next, we assume that the users collects only $P = N_{\mathrm{RF}}$ pilot signals, where $L \leq P \ll N_{\mathrm{T}}$. Further, we define $\mathbf{B} \in \mathbb{C}^{P \times N_{\mathrm{T}}}$ with $|[\mathbf{B}]_{i,j}| = \frac{1}{\sqrt{N_{\mathrm{T}}}}$ to represent the SI beamformer matrix at the BS. Since there is a single RF chain at the user, the received pilot data during $P$ frames are stacked in a $P \times 1$ vector as

$$\mathbf{y}_k[m] = \mathbf{B}\mathbf{F}\mathbf{x}_k[m] + \mathbf{w}_k[m]. \quad (16)$$

By exploiting the sparsity of $\mathbf{x}_k[m]$, the following $\ell_1$-norm sparse recovery (SR) problem is formulated, i.e.,

$$\hat{\mathbf{x}}_k[m] = \arg \underset{\mathbf{x}_k[m]}{\mathrm{minimize}} \ ||\mathbf{x}_k[m]||_1$$

$$\text{subject to: } ||\mathbf{y}_k[m] - \mathbf{B}\mathbf{F}\mathbf{x}_k[m]||_2^2 \leq \epsilon, \quad (17)$$

where the residual term $\epsilon$ is bounded by $\epsilon \leq \mu\sqrt{P + \kappa\sqrt{2P}}$, for which $\kappa$ corresponds to an adjustable tuning parameter in order to control the corruption due to noise [37], [44]. One can solve (17) for the $L$-sparse vector $\hat{\mathbf{x}}_k[m]$, and the channel can be reconstructed as $\hat{\mathbf{h}}_k^{\mathrm{SR}}[m] = \mathbf{D}\hat{\mathbf{x}}_k[m]$, which does not include beam-split correction.

### B. ARRAY PERTURBATION BASED BEAM-SPLIT MODEL

Due to beam-split, the true physical channel directions are observed with deviations in the spatial domain at different subcarrier frequencies since frequency of the digital beamformer changes while the frequency of the analog beamformer remains the same. Therefore, conventional channel estimation approaches fail and the effect of beam-split should be taken into account. To accurately mitigate the

effect of beam-split, we approach the problem from an array calibration point of view, wherein the beam-split is modeled as an array perturbation, e.g., mutual coupling, gain/phase mismatch [37], [38], [39]. Define $\mathbf{U}_k[m] \in \mathbb{C}^{N_T \times N_T}$ as the array perturbation matrix which maps the nominal array steering matrix, i.e., $\mathbf{A}_k$ to $\mathbf{A}'_k[m]$ which is perturbed due to SD beam-split. Then, the channel model in (7) is

$$\mathbf{h}_k[m] = \sqrt{\frac{N_T}{L}} \sum_{l=1}^{L} \alpha_k^l \mathbf{U}_k[m] \mathbf{a}(\vartheta_{k,l}) e^{-j2\pi \tau_{k,l} f_m},$$
$$= \mathbf{U}_k[m] \mathbf{A}_k \tilde{\mathbf{x}}_k[m], \quad (18)$$

where $\mathbf{A}_k = [\mathbf{a}(\vartheta_{k,1}), \ldots, \mathbf{a}(\vartheta_{k,L})] \in \mathbb{C}^{N_T \times L}$. Thus, $\mathbf{U}_k[m]$ provides a linear transformation between the steering matrices corresponding to physical and spatial directions as

$$\mathbf{A}'_k[m] = \mathbf{U}_k[m] \mathbf{A}_k. \quad (19)$$

Now, our aim is to rewrite the received signal $\mathbf{y}_k[m]$ as the linear combination of receiver output corresponding to the physical channel directions and some perturbation term. Therefore, we first write $\mathbf{U}_k[m] \mathbf{A}_k \tilde{\mathbf{x}}_k[m]$ as

$$\mathbf{A}'_k[m] \tilde{\mathbf{x}}_k[m] = \mathbf{A}_k \tilde{\mathbf{x}}_k[m] + \bar{\mathbf{Q}}_k[m] \mathbf{u}_k[m], \quad (20)$$

where $\bar{\mathbf{Q}}_k[m]$ is an $N_T \times N_T^2$ matrix with the following structure, i.e.,

$$\bar{\mathbf{Q}}_k[m] \mathbf{u}_k[m] = (\mathbf{A}'_k[m] - \mathbf{A}_k) \tilde{\mathbf{x}}_k[m], \quad (21)$$

where $\mathbf{u}_k[m] = \text{vec}\{\mathbf{U}_k[m]\} \in \mathbb{C}^{N_T^2}$, which includes the array perturbation terms in $\mathbf{U}_k[m]$. Now, we can rewrite the received signal model in (16) by using (20) and (21) as

$$\begin{aligned} \mathbf{y}_k[m] &= \mathbf{B} \mathbf{A}'_k[m] \tilde{\mathbf{x}}_k[m] + \mathbf{w}_k[m] \\ &= \mathbf{B} \mathbf{U}_k[m] \mathbf{A}_k \tilde{\mathbf{x}}_k[m] + \mathbf{w}_k[m] \\ &= \mathbf{B}(\mathbf{A}_k \tilde{\mathbf{x}}_k[m] + \bar{\mathbf{Q}}_k[m] \mathbf{u}_k[m]) + \mathbf{w}_k[m]. \quad (22) \end{aligned}$$

Now, our aim is to rewrite the received signal model with the combination of the nominal steering vectors (i.e., $\mathbf{B} \mathbf{A}_k$) and the array perturbation parameters (i.e., $\mathbf{u}$). Define $\tilde{\mathbf{P}}_k = \mathbf{B} \mathbf{A}_k \in \mathbb{C}^{P \times L}$ and $\tilde{\mathbf{Q}}_k[m] = \mathbf{B} \bar{\mathbf{Q}}_k[m] \in \mathbb{C}^{P \times N_T^2}$, then (22) becomes

$$\mathbf{y}_k[m] = \tilde{\mathbf{P}}_k \tilde{\mathbf{x}}_k[m] + \tilde{\mathbf{Q}}_k[m] \mathbf{u}_k[m] + \mathbf{w}_k[m]. \quad (23)$$

The perturbed array formulation in (23) explicitly shows the relationship between the received signal $\mathbf{y}_k[m]$ and the perturbation parameters corresponding to the beam-split $\mathbf{u}_k[m]$. Hence, one can minimize the fitting error between $\mathbf{y}_k[m]$ and $\tilde{\mathbf{P}}_k \tilde{\mathbf{x}}_k[m]$ to determine the physical channel directions as well as estimating the perturbation due to beam-split, as will be introduced in the following.

### C. SBL

In the proposed SBL framework, we rewrite (23) in an overcomplete form. Hence, only in this part, we first drop the subscript $(\cdot)_k$ and $[m]$ for notational simplicity, and discuss the channel estimation stage for multi-user multi-subcarrier

case later. Next, we use $\mathbf{Q} \in \mathbb{C}^{P \times N_T^2}$ instead of $\tilde{\mathbf{Q}}_k[m]$, and define $\mathbf{P} = \mathbf{B} \mathbf{F} \in \mathbb{C}^{P \times N}$ as the overcomplete version of $\tilde{\mathbf{P}}_k \in \mathbb{C}^{P \times L}$ covering the whole angular domain with $\phi_n$, $n = 1, \ldots, N$. Also, we have $\mathbf{P}' = \mathbf{B} \mathbf{U} \mathbf{F} \in \mathbb{C}^{P \times N}$, i.e., the overcomplete version of the perturbed steering matrix $\mathbf{B} \mathbf{A}'_k[m] = \mathbf{B} \mathbf{U}_k[m] \mathbf{A}_k \in \mathbb{C}^{P \times L}$. Then, (23) is expressed as follows

$$\mathbf{y} = \mathbf{P} \mathbf{x} + \mathbf{Q} \mathbf{u} + \mathbf{w}, \quad (24)$$

where $\mathbf{x}$, $\mathbf{u}$ and $\mathbf{w}$ are $N \times 1$, $N_T^2 \times 1$ and $P \times 1$ vectors, respectively, as defined earlier with removed subscripts.

Now, we introduce a new set of variables $\boldsymbol{\sigma} = [\sigma_1, \ldots, \sigma_N]^{\mathsf{T}}$ as the variance of sparse vector $\mathbf{x} \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = \text{diag}\{\boldsymbol{\sigma}\}$. Then, we derive the statistical dependence of the received signal $\mathbf{y}$ on the unknown parameters, i.e., the sparse vector $\mathbf{x}$, perturbation parameter $\mathbf{u}$, and the noise variance $\mu^2$. Toward this end, we look for the maximum likelihood estimate of these parameters to reconstruct the channel vector from the support of $\mathbf{x}$ as $\hat{\mathbf{h}} = \mathbf{D} \hat{\mathbf{x}}$.

The likelihood function of the received signal in (24) is

$$\begin{aligned} p(\mathbf{y}|\mathbf{x}; \mathbf{u}, \mu^2) &= \frac{1}{\pi \mu^{2N_T}} \exp\left\{-\frac{1}{\mu^2} \|\mathbf{y} - \mathbf{P}' \mathbf{x}\|_2^2\right\} \\ &= \frac{1}{\pi \mu^{2N_T}} \exp\left\{-\frac{1}{\mu^2} \|\mathbf{y} - \mathbf{P} \mathbf{x} - \mathbf{Q} \mathbf{u}\|_2^2\right\}. \quad (25) \end{aligned}$$

By using (25), we can write the probability density function (pdf) of $\mathbf{y}$ with respect to the hyperparameters $\boldsymbol{\sigma}$, $\mathbf{u}$ and $\mu^2$ as

$$\begin{aligned} p(\mathbf{y}; \boldsymbol{\sigma}, \mathbf{u}, \mu^2) &= \int p(\mathbf{y}|\mathbf{x}; \mathbf{u}, \mu^2) \times p(\mathbf{x}; \boldsymbol{\sigma}) d\mathbf{x} \\ &= \frac{1}{|\pi \boldsymbol{\Pi}_{\mathbf{y}}|} \exp\left\{-\text{Tr}\{\boldsymbol{\Pi}_{\mathbf{y}}^{-1} \mathbf{R}_{\mathbf{y}}\}\right\}, \quad (26) \end{aligned}$$

where $\boldsymbol{\Pi}_{\mathbf{y}} \in \mathbb{C}^{P \times P}$ is the covariance of $\mathbf{y}$ as

$$\boldsymbol{\Pi}_{\mathbf{y}} = \mathbb{E}\{\mathbf{y} \mathbf{y}^{\mathsf{H}}\} = \mathbf{P}' \boldsymbol{\Sigma} \mathbf{P}'^{\mathsf{H}} + \mu^2 \mathbf{I}_P, \quad (27)$$

and $\mathbf{R}_{\mathbf{y}} = \mathbf{y} \mathbf{y}^{\mathsf{H}} \in \mathbb{C}^{P \times P}$.

Define the unknown parameters in a hyperparameter set $\mathcal{S} = \{\boldsymbol{\sigma}, \mathbf{u}, \mu^2\}$. Then, $\mathcal{S}$ is estimated by maximizing the pdf in (26), which is non-concave and computationally intractable due to the nonlinearities. Hence, we resort to the expectation-maximization (EM) algorithm, which iteratively converges to a local optimum [38], [39], [40], [41], [42]. Each EM iteration comprises two steps: E-step for *inference*, and M-step for *hyperparameter estimation* by maximizing the Bayesian expectation of the complete probability $p(\mathbf{y}, \mathbf{x}, \boldsymbol{\sigma}, \mathbf{u}, \mu^2)$. Specifically, the E-step involves the evaluation of the log-likelihood of the complete probability $p(\mathbf{y}, \mathbf{x}, \boldsymbol{\sigma}, \mathbf{u}, \mu^2)$, i.e., $\mathbb{L}\{\mathbf{y}, \mathbf{x}; \boldsymbol{\sigma}, \mathbf{u}, \mu^2\}$. Then in the M-step, it is maximized to estimate the hyperparameters $\{\boldsymbol{\sigma}, \mathbf{u}, \mu^2\}$. This is done by first computing the partial derivations of $\mathbb{L}\{\mathbf{y}, \mathbf{x}; \boldsymbol{\sigma}, \mathbf{u}, \mu^2\}$ with respect to each of the hyperparameters and equating them to zero. Finally, we obtain the closed-form expressions to update each of the hyperparameters as discussed below.

## 1) E-STEP (INFERENCE)

In the E-step of the SBL algorithm, the log-likelihood of the probability $p(\mathbf{y}, \mathbf{x}, \boldsymbol{\sigma}, \mathbf{u}, \mu^2)$ is evaluated, i.e.,

$$\mathbb{L}\left\{\mathbf{y}, \mathbf{x}; \boldsymbol{\sigma}, \mathbf{u}, \mu^2\right\} = \mathbb{E}\left\{\ln p\left(\mathbf{y}, \mathbf{x}; \boldsymbol{\sigma}, \mathbf{u}, \mu^2\right)\right\}, \quad (28)$$

which can be written as

$$\mathbb{L}\left\{\mathbf{y}, \mathbf{x}; \boldsymbol{\sigma}, \mathbf{u}, \mu^2\right\} = \mathbb{E}\left\{\ln p\left(\mathbf{y}|\mathbf{x}; \mathbf{u}, \mu^2\right) + \ln p(\mathbf{x}; \boldsymbol{\sigma})\right\}, \quad (29)$$

since $p(\mathbf{y}|\mathbf{x}; \mathbf{u}, \mu^2)$ and $p(\mathbf{x}; \boldsymbol{\sigma})$ do not depend on $\boldsymbol{\sigma}$ and $\{\mathbf{u}, \mu^2\}$, respectively [38], [39]. Then, the first term of (29) can be obtained from (25) as

$$\begin{aligned}
&\ln p\left(\mathbf{y}|\mathbf{x}; \mathbf{u}, \mu^2\right) \\
&= \ln\left(\frac{1}{\pi \mu^{2N_T}} \exp\left\{-\frac{1}{\mu^2}\|\mathbf{y} - \mathbf{Px} - \mathbf{Qu}\|_2^2\right\}\right) \\
&= -N_T \ln \mu^2 - \mu^{-2}\|\mathbf{y} - \mathbf{Px} - \mathbf{Qu}\|_2^2, \quad (30)
\end{aligned}$$

and the second term in (29) is given by

$$\ln p(\mathbf{x}; \boldsymbol{\sigma}) = -\sum_{n=1}^{N}\left(\ln \sigma_n + \frac{|x_n|^2}{\sigma_n}\right), \quad (31)$$

since $\mathbf{x} \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma})$ [53]. Combining (30) and (31), (28) becomes

$$\begin{aligned}
&\mathbb{L}\left\{\mathbf{y}, \mathbf{x}; \boldsymbol{\sigma}, \mathbf{u}, \mu^2\right\} \\
&= \mathbb{E}\left\{-N_T \ln \mu^2 - \mu^{-2}\|\mathbf{y} - \mathbf{Px} - \mathbf{Qu}\|_2^2 \right. \\
&\left. \quad - \sum_{n=1}^{N}\left(\ln \sigma_n + \frac{|x_n|^2}{\sigma_n}\right)\right\}. \quad (32)
\end{aligned}$$

## 2) M-STEP (HYPERPARAMETER ESTIMATION)

Now, we consider estimating the hyperparameter set $\mathcal{S}$ by maximizing $\mathbb{L}\{\mathbf{y}, \mathbf{x}; \boldsymbol{\sigma}, \mathbf{u}, \mu^2\}$. To this end, we first compute the partial derivatives of (32) with respect to the unknown parameters as

$$\begin{aligned}
&\frac{\partial}{\partial \mathbf{u}}\mathbb{L}\left\{\mathbf{y}, \mathbf{x}; \boldsymbol{\sigma}, \mathbf{u}, \mu^2\right\} \\
&= -2\mu^{-2}\left(\mathbb{E}\left\{\mathbf{Q}^\mathsf{H}\mathbf{Q}\right\}\mathbf{u} - \mathbb{E}\left\{\mathbf{Q}^\mathsf{H}(\mathbf{y} - \mathbf{Px})\right\}\right), \quad (33)
\end{aligned}$$

$$\frac{\partial}{\partial \mu^2}\mathbb{L}\left\{\mathbf{y}, \mathbf{x}; \boldsymbol{\sigma}, \mathbf{u}, \mu^2\right\} = -\frac{N_T}{\mu^2} + \frac{1}{\mu^4}\mathbb{E}\left\{\|\mathbf{y} - \mathbf{P}'\mathbf{x}\|_2^2\right\}, \quad (34)$$

$$\frac{\partial}{\partial \sigma_n}\mathbb{L}\left\{\mathbf{y}, \mathbf{x}; \boldsymbol{\sigma}, \mathbf{u}, \mu^2\right\} = -\frac{1}{\sigma_n} + \frac{1}{\sigma_n^2}\mathbb{E}\left\{|x_n|^2\right\}, \quad (35)$$

where $n = 1, \ldots, N$. By setting the above derivatives to zero, we can get the following expressions, i.e.,

$$\mu^2 = \frac{1}{N_T}\mathbb{E}\left\{\|\mathbf{y} - \mathbf{P}'\mathbf{x}\|_2^2\right\}, \quad (36)$$

$$\sigma_n = \mathbb{E}\left\{|x_n|^2\right\}, \quad (37)$$

$$\mathbf{u} = \mathbb{E}\{\mathbf{Q}^\mathsf{H}\mathbf{Q}\}^{-1}\mathbb{E}\{\mathbf{Q}^\mathsf{H}(\mathbf{y} - \mathbf{Px})\}. \quad (38)$$

The expression $\mathbb{E}\{\|\mathbf{y} - \mathbf{P}'\mathbf{x}\|_2^2\}$ in (36) can be written as

$$\mathbb{E}\left\{\|\mathbf{y} - \mathbf{P}'\mathbf{x}\|_2^2\right\} = \|\mathbf{y} - \mathbf{P}'\mathbf{z}\|_2^2 + \mathrm{Tr}\left\{\mathbf{P}'\boldsymbol{\Pi}\mathbf{P}'^\mathsf{H}\right\}, \quad (39)$$

by utilizing the posterior density of $\mathbf{x}$ is $p(\mathbf{x}|\mathbf{y}) \sim \mathcal{CN}(\mathbf{z}, \boldsymbol{\Pi})$, where $\mathbf{z} = \mu^{-2}\boldsymbol{\Pi}\mathbf{P}'^\mathsf{H}\mathbf{y} \in \mathbb{C}^N$ and $\boldsymbol{\Pi} = \boldsymbol{\Sigma} - \boldsymbol{\Sigma}\mathbf{P}'^\mathsf{H}\boldsymbol{\Pi}_\mathbf{y}^{-1}\mathbf{P}'\boldsymbol{\Sigma} \in \mathbb{C}^{N \times N}$. Please see Appendix B for the proof.

Furthermore, the $(i,j)$th element of the first expectation term in (38) is computed as

$$\begin{aligned}
\left[\mathbb{E}\left\{\mathbf{Q}^\mathsf{H}\mathbf{Q}\right\}\right]_{ij} &= \mathbb{E}\left\{\mathbf{x}^\mathsf{H}\mathbf{D}^\mathsf{H}\mathbf{M}_i^\mathsf{H}\mathbf{M}_j\mathbf{Dx}\right\} \\
&= \mathrm{Tr}\left\{\mathbf{M}_i^\mathsf{H}\mathbf{M}_j\mathbf{D}\mathbb{E}\left\{\mathbf{xx}^\mathsf{H}\right\}\mathbf{D}^\mathsf{H}\right\}, \quad (40)
\end{aligned}$$

where $\mathbf{M}_i \in \mathbb{C}^{P \times N_T}$ denotes the derivative of $\mathbf{BU}$ with respect to $u_i$ as

$$\mathbf{M}_i = \frac{\mathbf{B}\partial\mathbf{U}}{\partial u_i}, \quad (41)$$

for $i = 1, \ldots, N_T^2$ [46]. In other words, the $i$th column of $\mathbf{Q}$ is given by

$$[\mathbf{Q}]_i = \mathbf{M}_i\mathbf{Dx}. \quad (42)$$

Further, using (24), we can compute $\mathbb{E}\{\mathbf{xx}^\mathsf{H}\}$ in (40) as

$$\mathbb{E}\left\{\mathbf{xx}^\mathsf{H}\right\} = \boldsymbol{\Sigma} = \mathbf{zz}^\mathsf{H} + \boldsymbol{\Pi}. \quad (43)$$

Then, (40) becomes

$$\left[\mathbb{E}\left\{\mathbf{Q}^\mathsf{H}\mathbf{Q}\right\}\right]_{ij} = \mathrm{Tr}\left\{\mathbf{M}_i^\mathsf{H}\mathbf{M}_j\mathbf{D}\left(\mathbf{zz}^\mathsf{H} + \boldsymbol{\Pi}\right)\mathbf{D}^\mathsf{H}\right\}. \quad (44)$$

Also, the $i$th element of the expectation $\mathbb{E}\{\mathbf{Q}^\mathsf{H}(\mathbf{y} - \mathbf{Px})\}$ in (38) is obtained as

$$\begin{aligned}
\left[\mathbb{E}\left\{\mathbf{Q}^\mathsf{H}(\mathbf{y} - \mathbf{Px})\right\}\right]_i &= \left[\mathbb{E}\left\{\mathbf{Q}^\mathsf{H}\mathbf{y}\right\}\right]_i - \left[\mathbb{E}\left\{\mathbf{Q}^\mathsf{H}\mathbf{Px}\right\}\right]_i \\
&= \mathrm{Tr}\left\{\mathbf{M}_i^\mathsf{H}\mathbf{yz}^\mathsf{H}\mathbf{D}^\mathsf{H}\right\} - \mathrm{Tr}\left\{\mathbf{M}_i^\mathsf{H}\mathbf{P}\left(\mathbf{zz}^\mathsf{H} + \boldsymbol{\Pi}\mathbf{P}^\mathsf{H}\right)\mathbf{D}^\mathsf{H}\right\},
\end{aligned}$$
$$(45)$$

for $i = 1, \ldots, N_T^2$.

Finally, we write the expectation-free expressions for the EM algorithm in closed-form as

$$\mu^2 = \frac{1}{N_T}\|\mathbf{y} - \mathbf{P}'\mathbf{z}\|_2^2 + \mathrm{Tr}\left\{\mathbf{P}'\boldsymbol{\Pi}\mathbf{P}'^\mathsf{H}\right\}, \quad (46)$$

$$\sigma_n = |z_n|^2, \quad (47)$$

$$\begin{aligned}
u_i = \mathrm{Tr}&\left\{\mathbf{M}_i^\mathsf{H}\mathbf{M}_j\mathbf{D}\left(\mathbf{zz}^\mathsf{H} + \boldsymbol{\Pi}\right)\mathbf{D}^\mathsf{H}\right\}\left(\mathrm{Tr}\left\{\mathbf{M}_i^\mathsf{H}\mathbf{yz}^\mathsf{H}\mathbf{D}^\mathsf{H}\right\}\right. \\
&\left. - \mathrm{Tr}\left\{\mathbf{M}_i^\mathsf{H}\mathbf{P}\left(\mathbf{zz}^\mathsf{H} + \boldsymbol{\Pi}\mathbf{P}^\mathsf{H}\right)\mathbf{D}^\mathsf{H}\right\}\right), \quad (48)
\end{aligned}$$

for $n = 1, \ldots, N$ and $i, j = 1, \ldots, N_T^2$.

By computing (46)-(48) iteratively, one can estimate the parameters $\boldsymbol{\sigma}$, $\mathbf{u}$ and $\mu^2$. Since the EM algorithm is proved to be convergent [38], [39], [40], [41], most of the terms in $\boldsymbol{\sigma}$ and $\mathbf{x}$ converge to 0, yielding to a sparse profile.

## D. LOW-COMPLEXITY APPROACH FOR ARRAY PERTURBATION UPDATE

Updating $\mathbf{u}$ involves the computation of $(\mathbf{Q}^{\mathsf{H}}\mathbf{Q})^{-1} \in \mathbb{C}^{N_{\mathrm{T}}^2 \times N_{\mathrm{T}}^2}$ and $\mathbf{Q}(\mathbf{y} - \mathbf{Px}) \in \mathbb{C}^{N_{\mathrm{T}}^2}$ in (38) which are computationally prohibitive due to large number of unknown parameters, i.e., $N_{\mathrm{T}}^2$, especially when the number of antennas is high. Therefore, we propose a low-complexity approach in the following.

Instead of employing an $N_{\mathrm{T}} \times N_{\mathrm{T}}$ full matrix $\mathbf{U}_k[m]$ in (18) to represent the array perturbations, we exploit the LoS-dominant feature of the THz channel, we assume that $L = 1$ by neglecting the NLoS paths, which are approximately 10 dB weaker than the LoS paths [1], [8]. This allows us to model $\mathbf{U}_k[m]$ as a diagonal matrix $\mathbf{C}_k[m] \in \mathbb{C}^{N_{\mathrm{T}} \times N_{\mathrm{T}}}$ instead of a full matrix. Rewrite (19) for $L = 1$ as

$$\mathbf{a}'_k[m] = \mathbf{C}_k[m]\mathbf{a}_k, \tag{49}$$

where $\mathbf{a}'_k[m] \in \mathbb{C}^{N_{\mathrm{T}}}$ and $\mathbf{a}_k \in \mathbb{C}^{N_{\mathrm{T}}}$ correspond to the perturbed and nominal array steering vectors for the LoS paths, respectively. In (49), $\mathbf{C}_k[m] = \mathrm{diag}(\mathbf{c}_k[m])$ is an $N_{\mathrm{T}} \times N_{\mathrm{T}}$ diagonal transformation matrix. $\mathbf{c}_k[m]$ corresponds to the perturbation due to beam-split and its $i$th entry is defined as

$$[\mathbf{c}_k[m]]_i = e^{j\pi(i-1)\Delta_k[m]}, \tag{50}$$

for $i = 2, \ldots, N_{\mathrm{T}}$ and $[\mathbf{c}_k[m]]_1 = 1$. Notice that this approach involves only a single unknown parameter, i.e., $\Delta_k[m]$ to perform the update from $\mathbf{a}_k[m]$ to $\mathbf{a}'_k[m]$ whereas $N_{\mathrm{T}}^2$ parameters are involved in (38). Furthermore, the transformation in (49) also allows us to accurately estimate the beam-split as presented in the following lemma.

*Lemma 2:* Let $\theta_{k,m}$ be the spatial channel direction, then the beam-split introduced at the $m$th subcarrier is uniquely recovered as

$$\Delta_k[m] = \frac{1}{N_{\mathrm{T}} - 1} \sum_{i=2}^{N_{\mathrm{T}}} \frac{\angle[\mathbf{c}_k[m]]_i}{\pi(i-1)}, \tag{51}$$

where $\angle[\mathbf{c}_k[m]]_i$ is the angle of $[\mathbf{c}_k[m]]_i$.

*Proof:* Consider the steering vector $\mathbf{a}(\theta_{k,m})$. Then, using (11), the $i$th entry of $\mathbf{a}(\theta_{k,m})$ is given by

$$\big[\mathbf{a}(\theta_{k,m})\big]_i = e^{j\pi(i-1)\theta_{k,m}}. \tag{52}$$

Now, we find the angle of $[\mathbf{a}(\theta_{k,m})]_i$ as $\Omega_{k,i}[m] = \angle[\mathbf{a}(\theta_{k,m})]_i$ for $i = 1, \ldots, N_{\mathrm{T}}$. Then, we compute the unwrapped angles as $\boldsymbol{\Omega}_k[m] = \mathrm{unwrap}\{[\Omega_{k,1}[m], \ldots, \Omega_{k,N_{\mathrm{T}}}[m]]^{\mathsf{T}}\}$ to ensure not losing information due to the $[-\pi, \pi]$ periodicity of the exponential.[2] Using $\boldsymbol{\Omega}_k[m]$ and (11), the $i$th element of the steering vector corresponding to the physical direction $\vartheta_k = \frac{f_c}{f_m}\theta_{k,m}$ becomes

$$[\mathbf{a}(\vartheta_k)]_i = e^{j\Omega_{k,i}[m]\frac{f_c}{f_m}}. \tag{53}$$

Substituting (53) and (52) into (49) yields $[\mathbf{a}(\theta_{k,m})]_i = c_{k,i}[m][\mathbf{a}(\vartheta_k)]_i$, where the angle of $[\mathbf{c}_k[m]]_i$ is given by

---

2. The unwrapped angle information is easily obtained via the `unwrap` command in MATLAB.

---

**Algorithm 1 SBCE**

**Input:** $\mathbf{D}$, $\mathbf{B}$, $\epsilon_{\mathrm{TH}}$, $f_c$, $\mathbf{y}_k[m]$, $f_m$ for $m \in \mathcal{M}$.
**Output:** $\hat{\mathbf{h}}_k[m]$, $\hat{\vartheta}_k$, $\hat{\Delta}_k[m]$.
1: **for** $k \in \mathcal{K}$
2:    **for** $m \in \mathcal{M}$
3:      Initialize: $t=0$, $\boldsymbol{\sigma}^{(t)} = \mathbf{1}_N$, $\mathbf{C}_k^{(t)}[m] = \mathbf{I}_{N_{\mathrm{T}}}$ and $\mu^{2^{(t)}} = 0$.
4:      $\mathbf{y} = \mathbf{y}_k[m]$, $\mathbf{C}^{(t)} = \mathbf{C}_k^{(t)}[m]$, $\mathbf{P}'^{(t)} = \mathbf{BC}^{(t)}\mathbf{D}$.
5:      flag = true.
6:      **while** flag = true
7:         $\boldsymbol{\Sigma}^{(t)} = \mathrm{diag}\{\boldsymbol{\sigma}^{(t)}\}$.
8:         $\boldsymbol{\Pi}_{\mathbf{y}}^{(t)} = \mathbf{P}'^{(t)}\boldsymbol{\Sigma}^{(t)}\mathbf{P}'^{(t)\mathsf{H}} + \mu^{2^{(t)}}\mathbf{I}_P$.
9:         $\boldsymbol{\Pi}_{\mathbf{x}}^{(t)} = (\boldsymbol{\Sigma}^{(t)} - \boldsymbol{\Sigma}^{(t)}\mathbf{P}'^{(t)\mathsf{H}}\boldsymbol{\Pi}_{\mathbf{y}}^{(t)-1}\mathbf{P}'^{(t)}\boldsymbol{\Sigma}^{(t)})^{-1}$.
10:        $t \leftarrow t + 1$.
11:        $\mathbf{z}^{(t)} = \mu^{2^{(t)}}\boldsymbol{\Pi}_{\mathbf{x}}^{(t)}\mathbf{P}'^{(t)\mathsf{H}}\mathbf{y}$.
12:        Update $\sigma_n^{(t)} = |z_n^{(t)}|^2$.
13:        Find $\phi_{n^\star}^{(t)}$ from $n^\star = \mathrm{argmax}_n|z_n^{(t)}|^2$.
14:        Construct $\mathbf{a}(\phi_{n^\star}^{(t)})$.
15:        Update $\mathbf{C}^{(t)}$ as $\mathbf{a}(\frac{f_m}{f_c}\phi_{n^\star}^{(t)}) = \mathbf{C}^{(t)}\mathbf{a}(\phi_{n^\star}^{(t)})$.
16:        Update $\mathbf{P}'^{(t)} = \mathbf{BC}^{(t)}\mathbf{D}$.
17:        Update $\mu^{2^{(t)}} = \frac{1}{N_{\mathrm{T}}}||\mathbf{y} - \mathbf{P}'^{(t)\mathsf{H}}\mathbf{z}^{(t)}||_2^2 +$ $\mathrm{Tr}\{\mathbf{P}'^{(t)}\boldsymbol{\Pi}^{(t)}\mathbf{P}'^{(t)\mathsf{H}}\}$.
18:        **if** $||\boldsymbol{\sigma}^{(t)} - \boldsymbol{\sigma}^{(t-1)}||_2/||\boldsymbol{\sigma}^{(t)}||_2 < \epsilon_{\mathrm{TH}}$
19:           flag = false.
20:        **end if**
21:      **end while**
22:      $\mathbf{y}_k[m] = \mathbf{y}$, $\mathbf{C}_k[m] = \mathbf{C}^{(t)}$, $\bar{\vartheta}_k = \phi_{n^\star}^{(t)}$.
23:      Insert $\bar{\vartheta}_k$ into (55)-(60), find the refined directions $\hat{\vartheta}_k$.
24:      Beam-split estimate: $\hat{\Delta}_k[m] = \frac{1}{N_{\mathrm{T}}-1}\sum_{i=2}^{N_{\mathrm{T}}}\{\frac{\angle[\mathbf{c}_k[m]]_i}{\pi(i-1)}\}$.
25:      Support estimate $\hat{x}_k[m] = (\mathbf{Ba}(\hat{\vartheta}_k))^\dagger\mathbf{y}_k[m]$.
26:      Channel estimate: $\hat{\mathbf{h}}_k[m] = \mathbf{a}(\hat{\vartheta}_k)\hat{x}_k[m]$.
27:    **end for**
28: **end for**

---

$$\begin{aligned}
\angle[\mathbf{c}_k[m]]_i &= \Omega_{k,i}[m]\left(1 - \frac{f_c}{f_m}\right) \\
&= \pi(i-1)\theta_{k,m}\left(\frac{\theta_{k,m} - \vartheta_k}{\theta_{k,m}}\right) \\
&= \pi(i-1)\Delta_k[m]. \tag{54}
\end{aligned}$$

Then, taking average of $\angle[\mathbf{c}_k[m]]_i$ for $i = 2, \ldots, N_{\mathrm{T}}$ and leaving alone $\Delta_k[m]$ leads to (51). ∎

Lemma 2 allows us to implement the SBL iterations with significantly lower complexity. We present the algorithmic steps of SBCE in Algorithm 1. In particular, the SBCE is initialized with $(t = 0)$ $\boldsymbol{\sigma}^{(t)} = \mathbf{1}_N$, $\mathbf{C}_k^{(t)}[m] = \mathbf{I}_{N_{\mathrm{T}}}$ and $\mu^{2^{(t)}} = 0$. Then, the covariances $\boldsymbol{\Pi}_{\mathbf{y}}^{(t)}$ and $\boldsymbol{\Pi}_{\mathbf{x}}^{(t)}$ are computed. As the SBCE iterates, $\mathbf{C}_k^{(t)}[m]$ involves the array perturbations due to beam-split and providing a linear transformation between the nominal and actual steering vectors as $\mathbf{a}(\frac{f_m}{f_c}\phi_{n^\star}^{(t)}) = \mathbf{C}^{(t)}\mathbf{a}(\phi_{n^\star}^{(t)})$. Once the SBCE converges, the beam-split and physical direction $\hat{\vartheta}_k$ are estimated. Then, the channel support from (15) is reconstructed as $\hat{x}_k[m] = (\mathbf{Ba}(\hat{\vartheta}_k))^\dagger\mathbf{y}_k[m]$ for $L = 1$, and the estimated channel is obtained as $\hat{\mathbf{h}}_k[m] = \mathbf{a}(\hat{\vartheta}_k)\hat{x}_k[m]$. In the following, we discuss refining the estimated physical directions.

## E. REFINED DIRECTION ESTIMATION

The estimation accuracy of the direction estimates obtained via the support of $\mathbf{x}$ is subject to the angular resolution of the

selected grid, i.e., $\rho = \frac{1}{N}$. Hence, once the EM algorithm terminates, we use the resulting estimates and search for the refined directions by employing a finer grid [38], [39]. Define the finer grid for the $l$th physical direction $\vartheta_l$ as $\Phi_l = [\vartheta_l - \varphi, \vartheta_l + \varphi]$, where the grid interval is $2\varphi$. Then, we rewrite the covariance of the received signal $\boldsymbol{\Pi_y}$ as

$$\boldsymbol{\Pi_y} = \boldsymbol{\Pi_{y\backslash l}} + \eta_l \mathbf{g}'(\vartheta_l)\mathbf{g}'^{\mathsf{H}}(\vartheta_l), \tag{55}$$

where $\mathbf{g}'(\vartheta_l) = \mathbf{BCa}(\vartheta_l) \in \mathbb{C}^P$ and $\eta_l$ is the power of the $l$th signal. $\boldsymbol{\Pi_{y\backslash l}}$ denotes to the covariance matrix in (27) by excluding the columns of $\mathbf{P}'$ corresponding to $\vartheta_l$. By substituting (55) into (26), we can obtain the refined directions by maximizing (25), wherein we substitute $\eta_l$ and $\vartheta_l$ as $\{\hat{\vartheta}_l, \hat{\eta}_l\} = \arg\max_{\eta_l, \vartheta_l} f(\eta_l, \vartheta_l)$, where $f(\eta_l, \vartheta_l)$ is

$$f(\eta_l, \vartheta_l) = \frac{1}{\left|\pi\left[\boldsymbol{\Pi_{y\backslash l}} + \eta_l \mathbf{g}'(\vartheta_l)\mathbf{g}'^{\mathsf{H}}(\vartheta_l)\right]\right|}$$
$$\times \exp\left\{-\mathrm{Tr}\left\{\left[\boldsymbol{\Pi_{y\backslash l}} + \eta_l \mathbf{g}'(\vartheta_l)\mathbf{g}'^{\mathsf{H}}(\vartheta_l)\right]^{-1}\mathbf{R_y}\right\}\right\}, \tag{56}$$

for which the equivalent minimization problem is given by $\{\hat{\vartheta}_l, \hat{\eta}_l\} = \arg\min_{\eta_l, \vartheta_l} \bar{f}(\eta_l, \vartheta_l)$, where

$$\bar{f}(\eta_l, \vartheta_l) = \ln\left|\pi\left[\boldsymbol{\Pi_{y\backslash l}} + \eta_l \mathbf{g}'(\vartheta_l)\mathbf{g}'^{\mathsf{H}}(\vartheta_l)\right]\right|$$
$$+ \mathrm{Tr}\left\{\left[\boldsymbol{\Pi_{y\backslash l}} + \eta_l \mathbf{g}'(\vartheta_l)\mathbf{g}'^{\mathsf{H}}(\vartheta_l)\right]^{-1}\mathbf{R_y}\right\}, \tag{57}$$

To simplify (57), we first set the derivative $\frac{\partial \bar{f}(\eta_l, \vartheta_l)}{\partial \eta_l}$ to zero, which yields

$$\hat{\eta}_l = \frac{\mathbf{g}'^{\mathsf{H}}(\vartheta_l)\boldsymbol{\Pi_{y\backslash l}^{-1}}(\mathbf{R_y} - \boldsymbol{\Pi_{y\backslash l}})\boldsymbol{\Pi_{y\backslash l}^{-1}}\mathbf{g}'(\vartheta_l)}{\left[\mathbf{g}'^{\mathsf{H}}(\vartheta_l)\boldsymbol{\Pi_{y\backslash l}^{-1}}\mathbf{g}'(\vartheta_l)\right]}, \tag{58}$$

and insert (58) into $\frac{\partial \bar{f}(\eta_l, \vartheta_l)}{\partial \vartheta_l} = 0$. Then, we get

$$\mathrm{Re}\left\{\mathbf{g}'^{\mathsf{H}}(\vartheta_l)\boldsymbol{\Pi_{y\backslash l}^{-1}}\left[\mathbf{g}'(\vartheta_l)\mathbf{g}'^{\mathsf{H}}(\vartheta_l)\boldsymbol{\Pi_{y\backslash l}^{-1}}\mathbf{R_y}\right.\right.$$
$$\left.\left. - \mathbf{R_y}\boldsymbol{\Pi_{y\backslash l}^{-1}}\mathbf{g}'(\vartheta_l)\mathbf{g}'^{\mathsf{H}}(\vartheta_l)\right]\boldsymbol{\Pi_{y\backslash l}^{-1}}\dot{\mathbf{g}}'(\vartheta_l)\right\} = 0, \tag{59}$$

where $\dot{\mathbf{g}}'(\vartheta_l) = \frac{\partial \mathbf{g}'(\vartheta_l)}{\partial \vartheta_l}$. Using (59), we can write the refined direction estimation problem as

$$\hat{\vartheta}_l = \arg\max_{\vartheta_l}\left|\mathrm{Re}\left\{\mathbf{g}'^{\mathsf{H}}(\vartheta_l)\boldsymbol{\Pi_{y\backslash l}^{-1}}\left[\mathbf{g}'(\vartheta_l)\mathbf{g}'^{\mathsf{H}}(\vartheta_l)\boldsymbol{\Pi_{y\backslash l}^{-1}}\mathbf{R_y}\right.\right.\right.$$
$$\left.\left.\left. - \mathbf{R_y}\boldsymbol{\Pi_{y\backslash l}^{-1}}\mathbf{g}'(\vartheta_l)\mathbf{g}'^{\mathsf{H}}(\vartheta_l)\right]\boldsymbol{\Pi_{y\backslash l}^{-1}}\dot{\mathbf{g}}'(\vartheta_l)\right\}\right|^{-1}. \tag{60}$$

### F. COMPUTATIONAL COMPLEXITY
The complexity of SBCE is mainly due to the matrix computations in Algorithm 1, which are $O(PN^2 + NP^2)$ (Step 8), $O(2N^2P + NP^2 + 2N^3 + P^3)$ (Step 9), $O(N^2P + NP)$ (Step 11), $O(P^2N_T^3N)$ (Step 16) and $O(NP + PN^2 + P^2N)$ (Step 17), respectively. Hence, the overall complexity order is $O(T[PN(5N^2 + 3P^2 + PN_T^3 + 2) + 2P^3 + 3N^3])$. In addition, the complexity of the direction refinement process in Section III-E is proportional to the number of grid points to compute (60).

## IV. NEAR-FIELD CONSIDERATIONS
Due to operating at high frequencies as well as employing extremely small array aperture, THz-band transmission may encounter near-field phenomenon for close-proximity users. Specifically, the far-field model involves the reception of the transmitted signal at the users as plane-wave whereas the plane wavefront is spherical in the near-field if the transmission range is shorter than the Fraunhofer distance, i.e., $R = \frac{2G^2 f_c}{c_0}$ where $G$ is the array aperture [4], [54], [55]. For ULA, we have $G = (N_T - 1)\frac{c_0}{2f_c}$, for which the Fraunhofer distance becomes $R \approx \frac{N_T^2 c_0}{2f_c}$ for large $N_T$.

Compared to its far-field counterpart, the near-field model involves additional range parameter. Define $r$ as the distance between the user and the array center at the BS. Then, the $i$th element of the steering vectors for ULA in near- and far-field are given by [34]

$$[\mathbf{a}_{\mathrm{NF}}(\vartheta, r)]_i = e^{-j2\pi\frac{f_m}{c_0}\left(r^{(i)} - r\right)} \tag{61}$$

$$[\mathbf{a}_{\mathrm{FF}}(\vartheta)]_i = e^{j2\pi\bar{d}\frac{f_m}{c_0}(i-1)\vartheta}, \tag{62}$$

where $r^{(i)}$ is the distance between the user and the $i$th BS antenna as

$$r^{(i)} = r\sqrt{1 + \left(\frac{(i-1)\bar{d}}{r}\right)^2 - \frac{2(i-1)\bar{d}\vartheta}{r}}. \tag{63}$$

Applying Taylor series expansion to the right hand side of (63) [35], [54], we get

$$r^{(i)} = r - (i-1)\bar{d}\vartheta + \frac{\left((i-1)\bar{d}\cos\tilde{\vartheta}\right)^2}{2r}, \tag{64}$$

where the last term tends to 0 in the far-field scenario, i.e., $e^{-j2\pi\frac{f_m}{c_0}\left((r - (i-1)\bar{d}\vartheta) - r\right)} = e^{j\pi\frac{f_m}{f_c}(i-1)\vartheta}$. Now, we calculate beam-split in near-field. Using (64) and $\bar{d} = \frac{c_0}{2f_c}$, we rewrite (61) as

$$[\mathbf{a}_{\mathrm{NF}}(\vartheta, r)]_i = e^{-j2\pi\frac{f_m}{c_0}\left(-(i-1)\bar{d}\vartheta + \frac{\left((i-1)\bar{d}\cos\tilde{\vartheta}\right)^2}{2r}\right)}$$
$$= e^{j\left(\pi\frac{f_m}{f_c}(i-1)\left(\vartheta - \frac{c_0(i-1)\cos^2\tilde{\vartheta}}{4f_c r}\right)\right)}$$
$$= e^{j\pi(i-1)\theta_{r,m,i}}, \tag{65}$$

where $\theta_{r,m,i}$ denotes the distance-dependent near-field spatial channel direction as

$$\theta_{r,m,i} = \frac{f_m}{f_c}\vartheta - \frac{f_m c_0(i-1)\cos^2\tilde{\vartheta}}{4f_c^2 r}. \tag{66}$$

Then, the near-field beam-split is defined as

$$\Delta_{r,i}[m] = \theta_{r,m,i} - \vartheta$$
$$= \left(\frac{f_m}{f_c} - 1\right)\vartheta - \frac{f_m c_0(i-1)\cos^2\tilde{\vartheta}}{4f_c^2 r}. \tag{67}$$

Notice that the near-field beam-split also depends on the antenna index $i$ with an additional distance-dependent term $-\frac{f_m c_0(i-1)\cos^2\tilde{\vartheta}}{4f_c^2 r}$ while it converges to far-field beam-split,

i.e., $\Delta_{r,i}[m] \approx \Delta[m]$ as the user distance increases since $\lim_{r\to\inf} \frac{f_m c_0 (i-1)\cos^2 \tilde{\vartheta}}{4f_c^2 r} = 0$.

## V. SBFL FOR THz CHANNEL ESTIMATION

Since the THz system involves the UM number of antennas, the channel estimation stage may be computationally complex, for which the data-driven techniques, such as ML and FL can be employed to ease the computational burden. Hence, we propose learning-based approach, i.e., SBFL, for THz channel estimation. While traditional ML method can also be considered, FL is communication-efficient since it does not involve dataset transmission [56]. Therefore, we consider an FL-based training scheme to provide a communication-efficient solution for THz channel estimation.

Denote $\boldsymbol{\xi} \in \mathbb{R}^Q$ as the vector of real-valued model parameters, and define $\mathcal{D}_k$ as the dataset of the $k$th user. By training the model $\boldsymbol{\xi}$ on $\mathcal{D}_k$ allows us to construct the nonlinear function $f(\boldsymbol{\xi})$ as $\mathcal{Y}_k^{(i)} = f(\boldsymbol{\xi})\mathcal{X}_k^{(i)}$, $i = 1, \ldots, \mathsf{D}_k$. Here, $\mathsf{D}_k = |\mathcal{D}_k|$ stands for the number of samples for the $k$th dataset. Also, $\mathcal{X}_k^{(i)}$ and $\mathcal{Y}_k^{(i)}$, respectively, denote the input and output data for the $i$th sample of the $k$th dataset. Thus, we have $\mathcal{D}_k^{(i)} = (\mathcal{X}_k^{(i)}, \mathcal{Y}_k^{(i)})$.

### A. TRAINING

To begin with, we construct the input-output data used for training. We define the input data by $\mathcal{X}_k \in \mathbb{R}^{P\times 3}$, which involves the real, imaginary parts and angle of the received pilot signal $\mathbf{y}_k[m]$. Note that we use a *three-channel* input data for enhanced feature extraction performance [56], [57]. As a result, we have $[\mathcal{X}_k]_1 = \text{Re}\{\mathbf{y}_k[m]\}$, $[\mathcal{X}_k]_2 = \text{Im}\{\mathbf{y}_k[m]\}$ and $[\mathcal{X}_k]_3 = \angle\{\mathbf{y}_k[m]\}$. In addition, the channel estimate obtained via SBCE, i.e., $\hat{\mathbf{h}}_k[m]$, is used to construct the output data. Thus, we define the output as $\mathcal{Y}_k = [\text{Re}\{\hat{\mathbf{h}}_k[m]\}^\mathsf{T}, \text{Im}\{\hat{\mathbf{h}}_k[m]\}^\mathsf{T}]^\mathsf{T} \in \mathbb{R}^{2N_\mathrm{T}}$.

Now, we state the FL-based training model as

$$\underset{\boldsymbol{\xi}}{\text{minimize}} \quad \frac{1}{K}\sum_{k=1}^{K}\mathcal{L}_k(\boldsymbol{\xi})$$
$$\text{subject to: } f\left(\mathcal{X}_k^{(i)}|\boldsymbol{\xi}\right) = \mathcal{Y}_k^{(i)}, \quad (68)$$

where $i = 1, \ldots, \mathsf{D}_k$ and $k = 1 \ldots, K$. In (68), $\mathcal{L}_k(\boldsymbol{\xi}) = \frac{1}{\mathsf{D}_k}\sum_{i=1}^{\mathsf{D}_k}||f(\mathcal{X}_k^{(i)}|\boldsymbol{\xi}) - \mathcal{Y}_k^{(i)}||_\mathcal{F}^2$ denotes to the loss function at the $k$th user and the wireless channel estimate at the $k$th user is given by $f(\mathcal{X}_k^{(i)}|\boldsymbol{\xi})$ when the input is $\mathcal{X}_k^{(i)}$. The FL problem in (68) is effectively solved via stochastic gradient descent algorithms, for which the learning model parameters $\boldsymbol{\xi}$ is iteratively computed by each user, and then aggregated at the cloud server via BS.[3] Therefore, we consider the model

---

3. In a general setting, the computational resources of all users are assumed to be satisfactory. Otherwise, hybrid federated/centralized learning strategies can be followed [58]. Then, the dataset of the users who are not able to compute the model parameters are sent to the BS, which performs computation on behalf of them.

---

update rule for the $j$th iteration as

$$\boldsymbol{\xi}_{j+1} = \boldsymbol{\xi}_j - \varepsilon\frac{1}{K}\sum_{k=1}^{K}\boldsymbol{\beta}_k(\boldsymbol{\xi}_j), \quad (69)$$

for $j = 1, \ldots, J$. Here, $\boldsymbol{\beta}_k(\boldsymbol{\xi}_j) = \nabla\mathcal{L}_k(\boldsymbol{\xi}_j) \in \mathbb{R}^Q$ is composed of the gradients of the model parameter $\boldsymbol{\xi}_j$ and $\varepsilon$ is the learning rate.

### B. COMMUNICATION OVERHEAD

Similar to the previous works, the communication overhead is defined as the amount of transmitted data during model training [57], [59], [60]. In particular, the overhead of CL mainly due to the dataset transmission of $K$ users whereas the overhead of FL involves the model parameter transmission of $K$ users for $J$ iterations. As a result, we define the overhead of CL and FL as $\mathcal{T}_\mathrm{FL} = 2QJK$ and $\mathcal{T}_\mathrm{CL} = \sum_{k=1}^{K}\mathsf{D}_k(3P+2N_\mathrm{T})$, respectively.

## VI. NUMERICAL EXPERIMENTS

In this section, the performance of our SBCE technique is evaluated and compared with the state-of-the-art techniques such as LS and MMSE in (14), GSOMP [29], OMP, BSPD [19] and BSA [30]. Note that MMSE is selected as benchmark since it relies on the channel covariance matrix at each subcarrier, hence it presents beam-split-free performance. The NMSE is computed as NMSE $= \frac{1}{J_\mathrm{T}KM}\sum_{i=1,k=1,m=1}^{J_\mathrm{T},K,M}\frac{||\mathbf{h}_k[m]-\hat{\mathbf{h}}_k^{(i)}[m]||_2^2}{||\mathbf{h}_k[m]||_2^2}$, where $J_\mathrm{T} = 100$ is the number of trials. The simulation parameters for the THz channel scenario are $f_c = 300$ GHz, $B = 30$ GHz, $M = 128$, $N_\mathrm{T} = 256$, $N_\mathrm{RF} = P = 32$, $L = 1$, $K = 8$ and $\tilde{\vartheta}_k \in \text{unif}[-\frac{\pi}{2}, \frac{\pi}{2}]$ [1], [2], [19]. $\mathbf{D}$ is formed with $N = 8N_\mathrm{T}$, and $[\mathbf{B}]_{i,j} = \frac{1}{\sqrt{N_\mathrm{T}}}e^{j\tilde{\varphi}}$, where $\tilde{\varphi} \sim \text{unif}(-1, 1)$. The proposed SBCE algorithm is initialized with $\boldsymbol{\sigma}^{(0)} = \mathbf{1}_N$, $\mathbf{C}_k^{(0)}[m] = \mathbf{I}_{N_\mathrm{T}}$, $\mu^{2(0)} = 0$, and run as presented in Algorithm 1. The termination parameter is selected as $\epsilon_\mathrm{TH} = 0.001$. It is observed that the SBCE algorithm converges in approximately $T = 50$ iterations.

For SBFL, the FL toolbox in MATLAB is employed for data generation and training. The learning model is a CNN with 12 layers and $Q = 603,648$ learnable parameters [30], [57]. Specifically, the first layer has the size of $P \times 3$ for input. The {2, 4, 6, 8}th layers are convolutional layers, and they are constructed with 128 kernels of size $3\times 3$. Each convolutional layer is followed by a normalization layer in the {3, 5, 7, 9}th layers. There is a fully-connected layer in the 10th layer of size 1024, which is followed by a dropout layer. The last 12th layer is the output layer of size $2N_\mathrm{T} \times 1$. During data generation, 100 channel realizations are generated. In order to provide robust performance against the imperfections in the receive data, we added AWGN on the input received pilot data for three signal-to-noise-ratio (SNR) levels, i.e., SNR = {15, 20, 25} dB and 50 noisy realizations [57]. The CNN is trained for $J = 100$ with the rate of $\varepsilon = 0.001$.
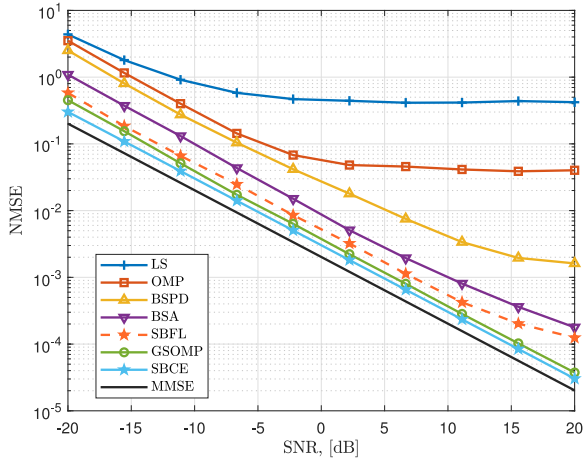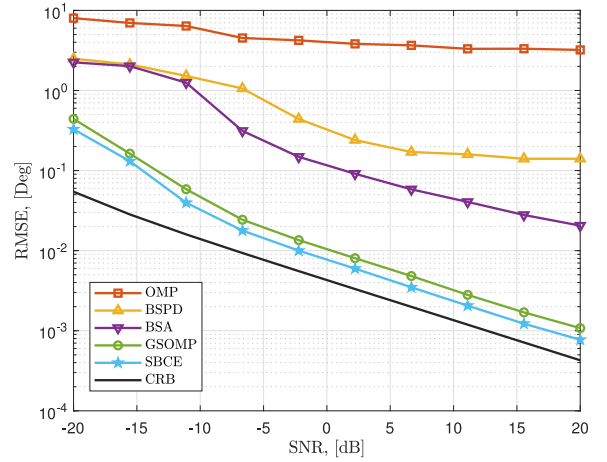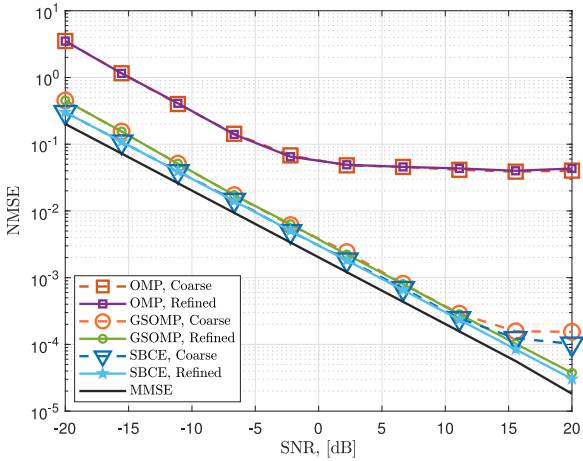
**FIGURE 2.** Channel estimation NMSE versus SNR.



**FIGURE 4.** Physical channel direction RMSE versus SNR.



**FIGURE 3.** Channel estimation NMSE versus SNR for coarse and refined dictionaries.

Fig. 2 shows the NMSE performance of the competing algorithms with respect to SNR. We observe that the proposed SBCE approach attains better NMSE as compared to other algorithmic-based approaches, i.e., GSOMP, BSA and BSPD. The direct application of both OMP and LS yields poor NMSE since they do not involve any specific mechanism to mitigate beam-split. While BSPD and BSA are beam-split-aware techniques, they suffer from inaccurate support detection, which leads to precision loss as SNR increases. Although GSOMP employs SD dictionary matrices with TDs, SBCE exhibits lower NMSE than that of GSOMP. This is mainly due to two reasons that SBCE performs better support recovery than GSOMP, which is an OMP-based algorithm [45], [53], [61], and GSOMP does not include refined direction estimation stage. In order to assess impact of refined direction estimation, Fig. 3 shows the channel estimation performance with coarse and refined direction estimates. During simulations the same coarse dictionary is employed for OMP, GSOMP and SBCE. Then, refined direction estimation method in Section III-E is run based on the estimate obtained from each algorithm. We can see that no significant improvement is obtained for OMP

algorithm since it does not take into account the impact of beam-split. However, improved NMSE is achieved, especially in high SNR, for GSOMP and SBCE. Again, we obtain similar observations as in Fig. 2, that is, SBCE has better performance than that of GSOMP in both coarse and refined direction scenario thanks to its accurate support recovery.

In order to demonstrate the physical direction estimation performance, RMSE of the estimated path directions is presented in Fig. 4. While the proposed SBCE approach follows the CRB closely, the remaining methods suffer from grid error and yield approximately $0.1°$ RMSE.

Comparing SBCE and SBFL, we observe that the latter has slight performance loss as well as it suffers from low NMSE in high SNR. This is because of the neural network losing precision due to decentralize learning and noisy dataset to prove robustness against imperfect data. Nevertheless, the main advantage of FL is the communication-efficiency, i.e., model training with less data/model transmission overhead. According to the analysis in Section V-B, the communication overhead of CL is computed as $\mathcal{T}_{CL} = 8 \cdot D_k \cdot (3 \cdot 32 + 2 \cdot 128) = 10.8 \times 10^9$, where the number of input-output tuples in the dataset is $D_k = 3MVG = 3 \cdot 128 \cdot 100 \cdot 100 = 3.07 \times 10^6$. On the other hand, the communication overhead of FL is $\mathcal{T}_{FL} = 2 \cdot 603,648 \cdot 100 \cdot 8 = 0.96 \times 10^9$, which exhibits approximately 11 times lower than that of CL. Thus, the proposed SBFL approach is an efficient tool especially for THz-related datasets, which usually involve large number of antennas.

In Fig. 5, beam-split estimation RMSE is shown with respect to SNR. As it is seen, the proposed approach effectively estimates the beam-split and follows the statistical lower bound CRB, which is derived in Appendix C, with a slight performance gap (e.g., $\sim 0.01°$ in 0 dB). The channel estimation techniques, i.e., GSOMP, BSPD and BSA, however, do not have the capability to obtain the beam-split.

Fig. 6 shows the channel estimation NMSE against bandwidth for the interval of [1, 30] GHz at 300 GHz carrier frequency. We can see that all of the algorithms provide approximately 0.02 NMSE for the bandwidth $B < 7$ GHz.
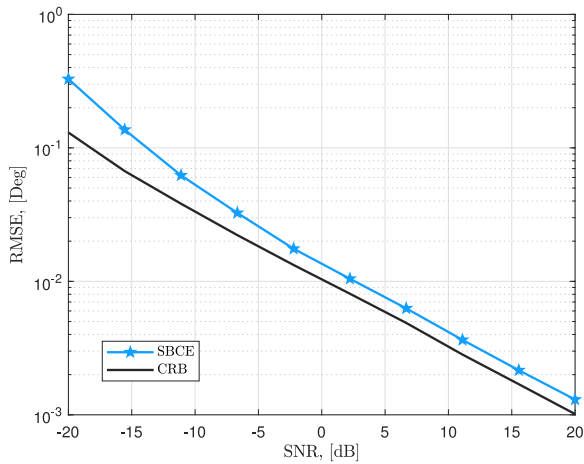
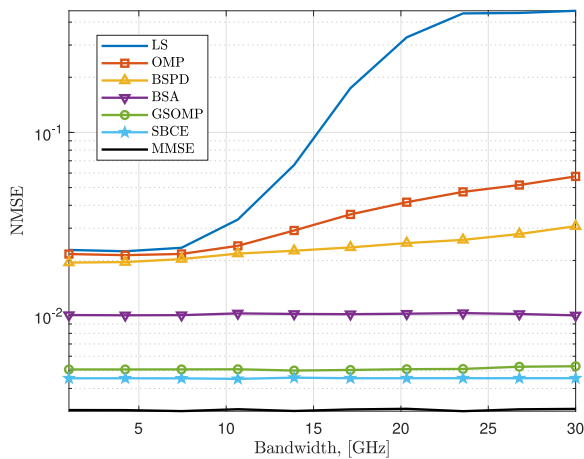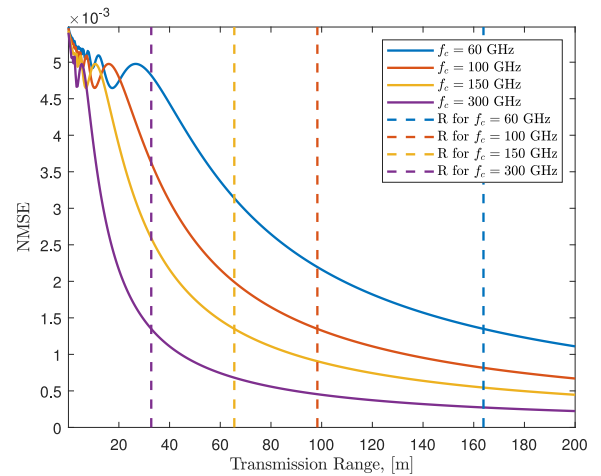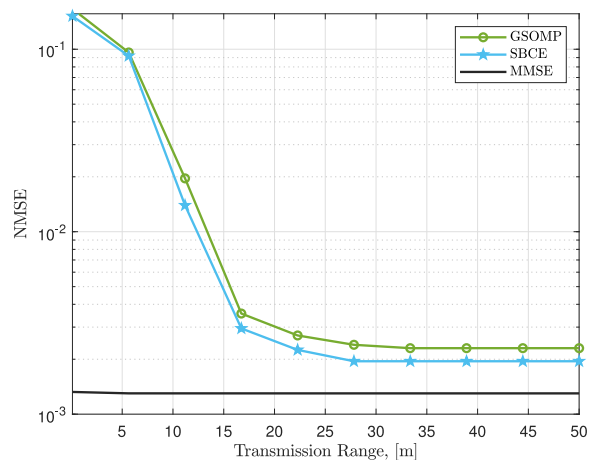**FIGURE 5.** Beam-split estimation RMSE versus SNR.



(a)



**FIGURE 6.** Channel estimation NMSE against bandwidth.



(b)

**FIGURE 7.** (a) NMSE between near- and far-field steering vectors, and (b) channel estimation NMSE for the usage of far-field model with respect to transmission range.

However, the performance of LS and OMP degrades as the bandwidth widens since these method do not involve beam-split mitigation. Also, BSPD yields a slight NMSE loss at wide bandwidth while BSA has robust performance. On the other hand, the proposed SBCE method enjoys robustness against the increase of the bandwidth.

Finally, we investigate the deviation from far-field model in terms of transmission range and frequency. Fig. 7(a) shows the normalized difference between near- and far-field steering vectors, i.e., $\frac{\|\mathbf{a}_{\mathrm{NF}} - \mathbf{a}_{\mathrm{FF}}\|_2^2}{\|\mathbf{a}_{\mathrm{FF}}\|_2^2}$, with respect to transmission range for various frequencies and corresponding Fraunhofer distances. As it is seen, the NMSE curve crosses the Fraunhofer distance approximately at 0.0013 for all frequencies at different ranges. Specifically, NMSE crosses the Fraunhofer distance at 32 m for $f_c = 300$ GHz, which shows that the near-field model should be considered for $r < 32$ m when $N_{\mathrm{T}} = 256$. Note that this distance is smaller if rectangular arrays are used. For instance, a $16 \times 16$ URA has the array aperture of $G = 16\sqrt{2}\bar{d}$ leading to $R = 0.256$ m, which yields much safer use of far-field model in shorter distance. Fig. 7(b) illustrates the channel estimation NMSE with respect to transmission range when

far-field model is used in GSOMP and SBCE. We observe that relatively large error for $r < R$ due to mismatch between the far- and near-field steering vectors. Although fixing the direction information may yield less error [29], which only due to range-mismatch, the mismatch in the range also causes inaccurate direction estimates.

## VII. CONCLUSION

In this work, we proposed an SBL-based approach for joint THz channel and beam-split estimation. The proposed method is based on modeling the beam-split as an array perturbation. We have shown that the proposed SBCE approach effectively estimates THz channel as compared to existing methods without requiring additional hardware components such as TD networks. The SBCE technique is also capable of effectively estimating the beam-split. This is particularly helpful to design the hybrid beamformers more easily with the knowledge of beam-split. Also, a model-free technique, SBFL, is introduced to realize the problem from ML perspective for computational- and communication-efficiency.

Furthermore, we examined the near-field considerations of the THz channel and evaluate the performance of SBCE with far-field model. The estimation of the near-field THz channel is one particular problem that we reserve to study for future work.

## APPENDIX A
## PROOF OF LEMMA 1

The array gain varies across the whole bandwidth as

$$A_G(\vartheta, \theta_m) = \frac{\left| \mathbf{a}^{\mathsf{H}}(\vartheta)\mathbf{a}'(\theta_m) \right|^2}{N_{\mathrm{T}}^2}. \tag{70}$$

By using (8), (70) is rewritten as

$$
\begin{aligned}
A_G(\vartheta, \theta_m) &= \frac{1}{N_{\mathrm{T}}^2} \left| \sum_{n_1=1}^{N_{\mathrm{T}}} \sum_{n_2=1}^{N_{\mathrm{T}}} e^{-\mathrm{j}\pi \left( (n_1-1)\theta_m - (n_2-1)\frac{\lambda_c \vartheta}{\lambda_m} \right)} \right|^2 \\
&= \frac{1}{N_{\mathrm{T}}^2} \left| \sum_{n=0}^{N_{\mathrm{T}}-1} e^{-\mathrm{j}2\pi nd \left( \frac{\theta_m}{\lambda_c} - \frac{\vartheta}{\lambda_m} \right)} \right|^2 \\
&= \frac{1}{N_{\mathrm{T}}^2} \left| \sum_{n=0}^{N_{\mathrm{T}}-1} e^{-\mathrm{j}2\pi nd \frac{(f_c \theta_m - f_m \vartheta)}{c_0}} \right|^2 \\
&= \frac{1}{N_{\mathrm{T}}^2} \left| \frac{1 - e^{-\mathrm{j}2\pi N_{\mathrm{T}} d \frac{(f_c \theta_m - f_m \vartheta)}{c_0}}}{1 - e^{-\mathrm{j}2\pi d \frac{(f_c \theta_m - f_m \vartheta)}{c_0}}} \right|^2 \\
&= \frac{1}{N_{\mathrm{T}}^2} \left| \frac{\sin(\pi N_{\mathrm{T}} \mu_m)}{\sin(\pi \gamma_m)} \right|^2 \\
&= \left| \left| \xi \left( d \frac{(f_c \theta_m - f_m \vartheta)}{c_0} \right) \right| \right|^2. \tag{71}
\end{aligned}
$$

The array gain in (71) implies that most of the power is focused only on a small portion of the beamspace due to the power-focusing capability of $\xi(a)$, which substantially reduces across the subcarriers as $|f_m - f_c|$ increases. Furthermore, $|\xi(\mu_m)|^2$ gives peak when $\mu_m = 0$, i.e., $f_c \theta_m - f_m \vartheta = 0$. Thus, we have $\theta_m = \eta_m \vartheta$, which completes the proof.

## APPENDIX B
## DERIVATION OF (39)

In order to obtain (39), we utilize the Bayes rule to obtain the posterior density of $\mathbf{x}$ as $p(\mathbf{x}|\mathbf{y}) \sim \mathcal{CN}(\mathbf{z}, \mathbf{\Pi})$ where the mean and the variance are

$$\mathbf{z} = \mu^{-2} \mathbf{\Pi} \mathbf{P}'^{\mathsf{H}} \mathbf{y}, \tag{72}$$

and

$$\mathbf{\Pi} = \left[ \mathbf{\Sigma}^{-1} + \mu^{-2} \mathbf{P}'^{\mathsf{H}} \mathbf{P}' \right]^{-1}, \tag{73}$$

respectively [38], [62]. Now, $\mathbb{E}\{\|\mathbf{y} - \mathbf{P}'\mathbf{x}\|_2^2\}$ is rewritten as

$$\mathbb{E}\left\{ \left\| \mathbf{y} - \mathbf{P}'\mathbf{x} \right\|_2^2 \right\} = \mathbb{E}\left\{ \left\| \mathbf{y} - \mathbf{P}'\mathbf{z} \right\|_2^2 + \left\| \mathbf{P}'(\mathbf{x} - \mathbf{z}) \right\|_2^2 \right\}, \tag{74}$$

where

$$\mathbb{E}\left\{ \left\| \mathbf{P}'(\mathbf{x} - \mathbf{z}) \right\|_2^2 \right\} = \mathbb{E}\left\{ \mathrm{Tr}\left\{ (\mathbf{x} - \mathbf{z})^{\mathsf{H}} \mathbf{P}'^{\mathsf{H}} \mathbf{P}'(\mathbf{x} - \mathbf{z}) \right\} \right\}$$

$$
\begin{aligned}
&= \mathbb{E}\left\{ \mathrm{Tr}\left\{ (\mathbf{x} - \mathbf{z})^{\mathsf{H}} \mathbf{P}'^{\mathsf{H}} \mathbf{P}'(\mathbf{x} - \mathbf{z}) \right\} \right\} \\
&= \mathbb{E}\left\{ \mathrm{Tr}\left\{ \mathbf{P}'(\mathbf{x} - \mathbf{z})(\mathbf{x} - \mathbf{z})^{\mathsf{H}} \mathbf{P}'^{\mathsf{H}} \right\} \right\}, \\
&= \mathrm{Tr}\left\{ \mathbf{P}' \mathbb{E}\left\{ (\mathbf{x} - \mathbf{z})(\mathbf{x} - \mathbf{z})^{\mathsf{H}} \right\} \mathbf{P}'^{\mathsf{H}} \right\}. \tag{75}
\end{aligned}
$$

In order to compute $\mathbb{E}\{(\mathbf{x} - \mathbf{z})(\mathbf{x} - \mathbf{z})^{\mathsf{H}}\}$, we first obtain $\mathbb{E}\{\mathbf{z}\mathbf{z}^{\mathsf{H}}\}$ by utilizing (72) as

$$\mathbf{z} = \mu^{-2} \mathbf{\Pi} \mathbf{P}'^{\mathsf{H}} \mathbf{y}, \tag{76}$$

where $\mathbf{\Pi}$ from (73) can be written via applying matrix inversion lemma as

$$\mathbf{\Pi} = \mathbf{\Sigma} - \mathbf{\Sigma}\mathbf{P}'^{\mathsf{H}} \mathbf{\Pi}_{\mathbf{y}}^{-1} \mathbf{P}' \mathbf{\Sigma}, \tag{77}$$

for which we have

$$
\begin{aligned}
\mu^{-2} \mathbf{\Pi} \mathbf{P}'^{\mathsf{H}} \mathbf{y} &= \mu^{-2} \mathbf{\Sigma} \mathbf{P}'^{\mathsf{H}} \mathbf{y} - \mu^{-2} \mathbf{\Sigma} \mathbf{P}'^{\mathsf{H}} \mathbf{\Pi}_{\mathbf{y}}^{-1} \mathbf{P}' \mathbf{\Sigma} \mathbf{P}'^{\mathsf{H}} \mathbf{y} \\
&= \mu^{-2} \mathbf{\Sigma} \mathbf{P}'^{\mathsf{H}} \underbrace{\left[ \mathbf{I}_P - \mathbf{\Pi}_{\mathbf{y}}^{-1} \mathbf{P}' \mathbf{\Sigma} \mathbf{P}'^{\mathsf{H}} \right]}_{\mu^2 \mathbf{\Pi}_{\mathbf{y}}^{-1}} \mathbf{y} \\
&= \mathbf{\Sigma} \mathbf{P}'^{\mathsf{H}} \mathbf{\Pi}_{\mathbf{y}}^{-1} \mathbf{y} = \mathbf{z}, \tag{78}
\end{aligned}
$$

for which the second line in the right hand side involves

$$
\begin{aligned}
\mathbf{\Pi}_{\mathbf{y}} &= \mathbf{P}' \mathbf{\Sigma} \mathbf{P}'^{\mathsf{H}} + \mu^2 \mathbf{I}_P \\
\mathbf{I}_P &= \mathbf{\Pi}_{\mathbf{y}}^{-1} \mathbf{P}' \mathbf{\Sigma} \mathbf{P}'^{\mathsf{H}} + \mu^2 \mathbf{\Pi}_{\mathbf{y}}^{-1} \mu^2 \\
\mathbf{\Pi}_{\mathbf{y}}^{-1} &= \mathbf{I}_P - \mathbf{\Pi}_{\mathbf{y}}^{-1} \mathbf{P}' \mathbf{\Sigma} \mathbf{P}'^{\mathsf{H}}. \tag{79}
\end{aligned}
$$

Then, we have $\mathbb{E}\{(\mathbf{x} - \mathbf{z})(\mathbf{x} - \mathbf{z})^{\mathsf{H}}\} = \mathbb{E}\{\mathbf{x}\mathbf{x}^{\mathsf{H}}\} + \mathbb{E}\{\mathbf{z}\mathbf{z}^{\mathsf{H}}\}$ due to their independence, and combining with (78) yields

$$
\begin{aligned}
\mathbb{E}\left\{ \mathbf{x}\mathbf{x}^{\mathsf{H}} \right\} + \mathbb{E}\left\{ \mathbf{z}\mathbf{z}^{\mathsf{H}} \right\} &= \mathbf{\Sigma} + \mathbf{\Sigma} \mathbf{P}'^{\mathsf{H}} \mathbf{\Pi}_{\mathbf{y}}^{-1} \mathbb{E}\left\{ \mathbf{y}\mathbf{y}^{\mathsf{H}} \right\} \mathbf{\Pi}_{\mathbf{y}}^{-1} \mathbf{P}' \mathbf{\Sigma}^{\mathsf{H}} \\
&= \mathbf{\Sigma} + \mathbf{\Sigma} \mathbf{P}'^{\mathsf{H}} \mathbf{\Pi}_{\mathbf{y}}^{-1} \mathbf{\Pi}_{\mathbf{y}} \mathbf{\Pi}_{\mathbf{y}}^{-1} \mathbf{P}' \mathbf{\Sigma}^{\mathsf{H}} \\
&= \mathbf{\Sigma} + \mathbf{\Sigma} \mathbf{P}'^{\mathsf{H}} \mathbf{\Pi}_{\mathbf{y}}^{-1} \mathbf{P}' \mathbf{\Sigma}^{\mathsf{H}} \\
&= \mathbf{\Pi}, \tag{80}
\end{aligned}
$$

due to (77). Finally, combining (80) and (75), we get

$$\mathrm{Tr}\left\{ \mathbf{P}' \mathbb{E}\left\{ (\mathbf{x} - \mathbf{z})(\mathbf{x} - \mathbf{z})^{\mathsf{H}} \right\} \mathbf{P}'^{\mathsf{H}} \right\} = \mathrm{Tr}\left\{ \mathbf{P}' \mathbf{\Pi} \mathbf{P}'^{\mathsf{H}} \right\}. \tag{81}$$

Thus, (74) becomes

$$\mathbb{E}\left\{ \|\mathbf{y} - \mathbf{P}'\mathbf{x}\|_2^2 \right\} = \|\mathbf{y} - \mathbf{P}'\mathbf{z}\|_2^2 + \mathrm{Tr}\left\{ \mathbf{P}' \mathbf{\Pi} \mathbf{P}'^{\mathsf{H}} \right\}. \tag{82}$$

## APPENDIX C
## CRAMÉR-RAO LOWER BOUND

In order to derive the CRB, we assume a single user case, i.e., $K = 1$ without loss of generality. The CRB expressions can easily be extended for the $K > 1$ case by using the decoupledness of the unknown parameters of multiple users. Define the unknown parameter vector as the physical directions and beam-splits corresponding to $L$ paths. Hence, we have a $2L \times 1$ unknown vector as

$$\mathbf{v} = [\vartheta_1, \ldots, \vartheta_L, \Delta_1, \ldots, \Delta_L]^{\mathsf{T}} \in \mathbb{C}^{2L}. \tag{83}$$

Consider the log-likelihood function of the joint pdf in (26) with respect to $\mathbf{v}$ as

$$\mathcal{L}(\mathbf{v}) = \ln\{p(\mathbf{y}|\mathbf{v})\} = -\ln|\mathbf{\Pi_y}| - \mathrm{Tr}\left\{\mathbf{\Pi_y}^{-1}\mathbf{R_y}\right\}. \quad (84)$$

In order to obtain the CRB, we first need to find the Fisher information matrix (FIM), which measures the amount of information contained for the unknown variables. Let $\mathbf{FIM} \in \mathbb{R}^{2L \times 2L}$ be the FIM. Then, the $(i,j)$th entry of $\mathbf{FIM}$ is calculated as the second derivative of $\mathcal{L}(\mathbf{v})$ as

$$[\mathbf{FIM}]_{ij} = -\mathbb{E}\left\{\frac{\partial^2 \mathcal{L}(\mathbf{v})}{\partial v_i \partial v_j}\right\}, \quad (85)$$

for $i,j = 1, \ldots, 2L$. Calculate the first derivatives of $\mathcal{L}(\mathbf{v})$ by using

$$\frac{\partial \ln\{|\mathbf{\Pi_y}|\}}{\partial v_i} = \mathrm{Tr}\left\{\mathbf{\Pi_y}^{-1}\frac{\partial \mathbf{\Pi_y}}{\partial v_i}\right\}$$

$$\frac{\partial \mathbf{\Pi_y}^{-1}}{\partial v_i} = -\mathbf{\Pi_y}^{-1}\frac{\partial \mathbf{\Pi_y}}{\partial v_i}\mathbf{\Pi_y}^{-1}, \quad (86)$$

we get the first derivative of $\mathcal{L}(\mathbf{v})$ as

$$\frac{\partial \mathcal{L}(\mathbf{v})}{\partial v_i} = -\mathrm{Tr}\left\{\mathbf{\Pi_y}^{-1}\frac{\partial \mathbf{\Pi_y}}{\partial v_i}\right\} + \mathbf{\Pi_y}^{-1}\frac{\partial \mathbf{\Pi_y}}{\partial v_i}\mathbf{\Pi_y}^{-1}\mathbf{R_y}$$

$$= \mathrm{Tr}\left\{\left(\mathbf{\Pi_y}^{-1}\frac{\partial \mathbf{\Pi_y}}{\partial v_i}\right)\left(\mathbf{\Pi_y}^{-1}\mathbf{R_y} - \mathbf{I}_P\right)\right\}, \quad (87)$$

and the second derivative becomes

$$\frac{\partial^2 \mathcal{L}(\mathbf{v})}{\partial v_i \partial v_j} = \mathrm{Tr}\left\{\frac{\partial}{\partial v_j}\left(\mathbf{\Pi_y}^{-1}\frac{\partial \mathbf{\Pi_y}}{\partial v_i}\right)\left(\mathbf{\Pi_y}^{-1}\mathbf{R_y} - \mathbf{I}_P\right)\right.$$

$$\left. + \left(\mathbf{\Pi_y}^{-1}\frac{\partial \mathbf{\Pi_y}}{\partial v_i}\right)\frac{\partial}{\partial v_j}\left(\mathbf{\Pi_y}^{-1}\mathbf{R_y} - \mathbf{I}_P\right)\right\}$$

$$= \mathrm{Tr}\left\{\frac{\partial}{\partial v_j}\left(\mathbf{\Pi_y}^{-1}\frac{\partial \mathbf{\Pi_y}}{\partial v_i}\right)\left(\mathbf{\Pi_y}^{-1}\mathbf{R_y} - \mathbf{I}_P\right)\right.$$

$$\left. + \left(\mathbf{\Pi_y}^{-1}\frac{\partial \mathbf{\Pi_y}}{\partial v_i}\right)\frac{\partial}{\partial v_j}\mathbf{\Pi_y}^{-1}\frac{\partial \mathbf{\Pi_y}}{\partial v_j}\mathbf{\Pi_y}^{-1}\mathbf{R_y}\right\}. \quad (88)$$

Since $\mathbb{E}\{\mathbf{R_y}\} = \mathbf{\Pi_y}$, (85) is rewritten as

$$[\mathbf{FIM}]_{ij} = -\mathbb{E}\left\{\frac{\partial^2 \mathcal{L}(\mathbf{v})}{\partial v_i \partial v_j}\right\} = \mathrm{Tr}\left\{\frac{\partial}{\partial v_i}\mathbf{\Pi_y}^{-1}\frac{\partial \mathbf{\Pi_y}}{\partial v_j}\mathbf{\Pi_y}^{-1}\mathbf{R_y}\right\}, \quad (89)$$

which is the general FIM expression for Gaussian signals. Then, by following the steps in [37], [63], the CRB expressions corresponding to the $(i,j)$th entries of the FIM is given by

$$\mathrm{CRB}_{ij} = \frac{1}{\frac{2}{\mu^2}\mathrm{Tr}\{\mathbf{MK}_{ij}\}}, \quad (90)$$

where $\mathbf{M} = \overline{\mathbf{\Sigma}}^{\mathsf{H}}\mathbf{A'}^{\mathsf{H}}\mathbf{\Pi_y}^{-1}\mathbf{A'}\overline{\mathbf{\Sigma}} \in \mathbb{C}^{L \times L}$, where $\overline{\mathbf{\Sigma}}$ is an $L \times L$ matrix comprised of signal powers. $\mathbf{K}_{ij} \in \mathbb{C}^{L \times L}$ includes the derivative of the actual steering matrix $\mathbf{A'}$ with respect to the unknown parameters, and it is given by

$$\mathbf{K}_{ij} = \left[\frac{\partial \mathbf{A'}}{\partial v_i}\right]^{\mathsf{H}}\left(\mathbf{I}_{N_{\mathrm{T}}} - \mathbf{A'}\mathbf{A'}^{\dagger}\right)\left[\frac{\partial \mathbf{A'}}{\partial v_j}\right]. \quad (91)$$

In particular, the $i$th entry of the derivative of $\mathbf{a'}(\vartheta_l)$, i.e., the $l$th column of $\mathbf{A'}$ with respect to the physical direction $\tilde{\vartheta}_l$ and beam-split $\Delta_l$ are respectively given by

$$\frac{\partial[\mathbf{a'}]_i}{\partial \tilde{\vartheta}_l} = j\pi(i-1)\frac{f_m}{f_c}\cos\tilde{\vartheta}_l[\mathbf{a'}]_i, \quad (92)$$

$$\frac{\partial[\mathbf{a'}]_i}{\partial \Delta_l} = j\pi(i-1)[\mathbf{a'}]_i, \quad (93)$$

where we have via (8) and (11) that $[\mathbf{a'}]_i = e^{j\pi(i-1)\frac{f_m}{f_c}\sin\tilde{\vartheta}_l} = e^{j\pi(i-1)(\Delta_l + \sin\tilde{\vartheta}_l)}$.

In case of near-field scenario, the range parameters are also added to the unknown vector defined in (83), hence its size becomes $3L \times 1$. By following the steps from (84) to (90), the CRB for the near-field scenario can be derived, wherein the derivatives of the steering vector element in (65), i.e., $[\mathbf{a'}_{\mathrm{NF}}]_i = e^{j(\pi\frac{f_m}{f_c}(i-1)(\vartheta_l - \frac{c_0(i-1)\cos^2\tilde{\vartheta}_l}{4f_c r}))} = e^{j\pi(i-1)\theta_{r,m,i}} = e^{j\pi(i-1)(\Delta_{r,l} + \sin\tilde{\vartheta}_l)}$ with respect to $\tilde{\vartheta}_l$, $r_l$ and near-field beam-split $\Delta_{r,l}$ are

$$\frac{\partial[\mathbf{a'}_{\mathrm{NF}}]_i}{\partial \tilde{\vartheta}_l} = -j\pi(i-1)\frac{f_m}{f_c}\left(\cos\tilde{\vartheta}_l + \frac{c_0(i-1)\sin 2\tilde{\vartheta}_l}{4f_c r}\right)[\mathbf{a'}_{\mathrm{NF}}]_i, \quad (94)$$

$$\frac{\partial[\mathbf{a'}_{\mathrm{NF}}]_i}{\partial r_l} = \frac{j\pi f_m c_0(i-1)^2\cos^2\tilde{\vartheta}_l}{4f_c^2 r^2}[\mathbf{a'}_{\mathrm{NF}}]_i, \quad (95)$$

$$\frac{\partial[\mathbf{a'}_{\mathrm{NF}}]_i}{\partial \Delta_{r,l}} = j\pi(i-1)[\mathbf{a'}_{\mathrm{NF}}]_i. \quad (96)$$

## REFERENCES

[1] H. Sarieddeen, M.-S. Alouini, and T. Y. Al-Naffouri, "An overview of signal processing techniques for terahertz communications," *Proc. IEEE*, vol. 109, no. 10, pp. 1628–1665, Oct. 2021.

[2] A. M. Elbir, K. V. Mishra, and S. Chatzinotas, "Terahertz-band joint ultra-massive MIMO radar-communications: Model-based and model-free hybrid beamforming," *IEEE J. Sel. Topics Signal Process.*, vol. 15, no. 6, pp. 1468–1483, Nov. 2021.

[3] "Spectrum horizons, RM-11795, FCC 19-19." Mar. 2019. Accessed: Nov. 22, 2022. [Online]. Available: https://www.federalregister.gov/documents/2019/06/04/2019-10925/spectrum-horizons

[4] I. F. Akyildiz, C. Han, Z. Hu, S. Nie, and J. M. Jornet, "Terahertz band communication: An old problem revisited and research directions for the next decade," *IEEE Trans. Commun.*, vol. 70, no. 6, pp. 4250–4285, Jun. 2022.

[5] A. M. Elbir, K. V. Mishra, S. Chatzinotas, and M. Bennis, "Terahertz-band integrated sensing and communications: Challenges and opportunities," Aug. 2022, *arXiv:2208.01235*.

[6] H. Yuan, N. Yang, K. Yang, C. Han, and J. An, "Hybrid beamforming for terahertz multi-carrier systems over frequency selective fading," *IEEE Trans. Commun.*, vol. 68, no. 10, pp. 6186–6199, Oct. 2020.

[7] J. Tan and L. Dai, "Wideband beam tracking in THz massive MIMO systems," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 6, pp. 1693–1710, Jun. 2021.

[8] S. Tarboush et al., "TeraMIMO: A channel simulator for wideband ultra-massive MIMO terahertz communications," Apr. 2021, *arXiv:2104.11054*.

[9] H. Sarieddeen, M.-S. Alouini, and T. Y. Al-Naffouri, "Terahertz-band ultra-massive spatial modulation MIMO," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 9, pp. 2040–2052, Sep. 2019.

[10] A. Faisal, H. Sarieddeen, H. Dahrouj, T. Y. Al-Naffouri, and M.-S. Alouini, "Ultramassive MIMO systems at terahertz bands: Prospects and challenges," *IEEE Veh. Technol. Mag.*, vol. 15, no. 4, pp. 33–42, Dec. 2020.

[11] B. Ning, Z. Chen, W. Chen, Y. Du, and J. Fang, "Terahertz multi-user massive MIMO with intelligent reflecting surface: Beam training and hybrid beamforming," *IEEE Trans. Veh. Technol.*, vol. 70, no. 2, pp. 1376–1393, Feb. 2021.

[12] B. Wang, M. Jian, F. Gao, G. Y. Li, and H. Lin, "Beam squint and channel estimation for wideband mmWave massive MIMO-OFDM systems," *IEEE Trans. Signal Process.*, vol. 67, no. 23, pp. 5893–5908, Dec. 2019.

[13] C. Lin and G. Y. L. Li, "Terahertz communications: An array-of-subarrays solution," *IEEE Commun. Mag.*, vol. 54, no. 12, pp. 124–131, Dec. 2016.

[14] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.

[15] R. W. Heath, N. González-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 436–453, Apr. 2016.

[16] R. Méndez-Rial, C. Rusu, A. Alkhateeb, N. González-Prelcic, and R. W. Heath, "Channel estimation and hybrid combining for mmWave: Phase shifters or switches?" in *Proc. IEEE Inf. Theory Appl. Workshop*, 2015, pp. 90–97.

[17] A. Alkhateeb, G. Leus, and R. W. Heath, "Limited feedback hybrid precoding for multi-user millimeter wave systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 11, pp. 6481–6494, Nov. 2015.

[18] A. M. Elbir, K. V. Mishra, S. A. Vorobyov, and R. W. Heath, "Twenty-five years of advances in beamforming: From convex and nonconvex optimization to learning techniques," Nov. 2022, *arXiv:2211.02165*.

[19] J. Tan and L. Dai, "Wideband channel estimation for THz massive MIMO," *China Commun.*, vol. 18, no. 5, pp. 66–80, May 2021.

[20] K. Wu, W. Ni, T. Su, R. P. Liu, and Y. J. Guo, "Exploiting spatial-wideband effect for fast AoA estimation at lens antenna array," *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 5, pp. 902–917, Sep. 2019.

[21] M. Wang, F. Gao, N. Shlezinger, M. F. Flanagan, and Y. C. Eldar, "A block sparsity based estimator for mmWave massive MIMO channels with beam squint," *IEEE Trans. Signal Process.*, vol. 68, pp. 49–64, Nov. 2019, doi: 10.1109/TSP.2019.2956677.

[22] A. Brighente, M. Cerutti, M. Nicoli, S. Tomasin, and U. Spagnolini, "Estimation of wideband dynamic mmWave and THz channels for 5G systems and beyond," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 9, pp. 2026–2040, Sep. 2020.

[23] S. Srivastava, A. Tripathi, N. Varshney, A. K. Jagannatham, and L. Hanzo, "Hybrid transceiver design for tera-hertz MIMO systems relying on Bayesian learning aided sparse channel estimation," Sep. 2021, *arXiv:2109.09664*.

[24] W. Yu, Y. Shen, H. He, X. Yu, J. Zhang, and K. B. Letaief, "Hybrid Far- and near-field channel estimation for THz ultra-massive MIMO via fixed point networks," May 2022, *arXiv:2205.04944*.

[25] E. Balevi and J. G. Andrews, "Wideband channel estimation with a generative adversarial network," *IEEE Trans. Wireless Commun.*, vol. 20, no. 5, pp. 3049–3060, May 2021.

[26] B. Peng, S. Wesemann, K. Guan, W. Templ, and T. Kürner, "Precoding and detection for broadband single carrier terahertz massive MIMO systems using LSQR algorithm," *IEEE Trans. Wireless Commun.*, vol. 18, no. 2, pp. 1026–1040, Feb. 2019.

[27] S. Nie and I. F. Akyildiz, "Deep kernel learning-based channel estimation in ultra-massive MIMO communications at 0.06-10 THz," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Dec. 2019, pp. 1–6.

[28] Y. Chen and C. Han, "Deep CNN-based spherical-wave channel estimation for terahertz ultra-massive MIMO systems," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2020, pp. 1–6.

[29] K. Dovelos, M. Matthaiou, H. Q. Ngo, and B. Bellalta, "Channel estimation and hybrid combining for wideband terahertz massive MIMO systems," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 6, pp. 1604–1620, Jun. 2021.

[30] A. M. Elbir, W. Shi, K. V. Mishra, and S. Chatzinotas, "Federated multi-task learning for THz wideband channel and DoA estimation," Jul. 2022, *arXiv:2207.06017*.

[31] Z. Sha and Z. Wang, "Channel estimation and equalization for terahertz receiver with RF impairments," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 6, pp. 1621–1635, Jun. 2021.

[32] J. Gao, C. Zhong, G. Y. Li, J. B. Soriaga, and A. Behboodi, "Deep learning-based channel estimation for wideband hybrid mmWave massive MIMO," May 2022, *arXiv:2205.05202*.

[33] B. Wang, F. Gao, S. Jin, H. Lin, and G. Y. Li, "Spatial- and frequency-wideband effects in millimeter-wave massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 66, no. 13, pp. 3393–3406, Jul. 2018.

[34] M. Cui, L. Dai, R. Schober, and L. Hanzo, "Near-field wideband beamforming for extremely large antenna arrays," Sep. 2021, *arXiv:2109.10054*.

[35] J.-J. Jiang, F.-J. Duan, J. Chen, Y.-C. Li, and X.-N. Hua, "Mixed near-field and far-field sources localization using the uniform linear sensor array," *IEEE Sensors J.*, vol. 13, no. 8, pp. 3136–3143, Aug. 2013.

[36] I.-S. Kim and J. Choi, "Spatial wideband channel estimation for mmWave massive MIMO systems with hybrid architectures and low-resolution ADCs," *IEEE Trans. Wireless Commun.*, vol. 20, no. 6, pp. 4016–4029, Jun. 2021.

[37] A. M. Elbir and T. E. Tuncer, "2-D DOA and mutual coupling coefficient estimation for arbitrary array structures with single and multiple snapshots," *Digit. Signal Process.*, vol. 54, pp. 75–86, Jul. 2016.

[38] Z.-M. Liu, Z.-T. Huang, and Y.-Y. Zhou, "An efficient maximum likelihood method for direction-of-arrival estimation via sparse Bayesian learning," *IEEE Trans. Wireless Commun.*, vol. 11, no. 10, pp. 1–11, Oct. 2012.

[39] H. Tang, J. Wang, and L. He, "Off-grid sparse Bayesian learning-based channel estimation for mmWave massive MIMO uplink," *IEEE Wireless Commun. Lett.*, vol. 8, no. 1, pp. 45–48, Feb. 2019.

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

[41] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Stat. Soc. Ser. B, Methodol.*, vol. 39, no. 1, pp. 1–22, Sep. 1977.

[42] B. M. Radich and K. M. Buckley, "Single-snapshot DOA estimation and source number detection," *IEEE Signal Process. Lett.*, vol. 4, no. 4, pp. 109–111, Apr. 1997.

[43] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. AP-34, no. 3, pp. 276–280, Mar. 1986.

[44] E. J. Candes and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.

[45] X. Tan and J. Li, "Compressed sensing via sparse Bayesian learning and Gibbs sampling," in *Proc. IEEE 13th Digit. Signal Process. Workshop 5th IEEE Signal Process. Educ. Workshop*, Jan. 2009, pp. 690–695.

[46] Z.-M. Liu and Y.-Y. Zhou, "A unified framework and sparse Bayesian perspective for direction-of-arrival estimation in the presence of array imperfections," *IEEE Trans. Signal Process.*, vol. 61, no. 15, pp. 3786–3798, Aug. 2013.

[47] K. M. S. Huq, S. A. Busari, J. Rodriguez, V. Frascolla, W. Bazzi, and D. C. Sicker, "Terahertz-enabled wireless system for beyond-5G ultra-fast networks: A brief survey," *IEEE Netw.*, vol. 33, no. 4, pp. 89–95, Jul./Aug. 2019.

[48] Y. Lu, M. Hao, and R. Mackenzie, "Reconfigurable intelligent surface based hybrid precoding for THz communications," *Intell. Converg. Netw.*, vol. 3, no. 1, pp. 103–118, 2022.

[49] A. Alkhateeb and R. W. Heath, "Frequency selective hybrid precoding for limited feedback millimeter wave systems," *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 1801–1818, May 2016.

[50] A. M. Elbir and S. Chatzinotas, "BSA-OMP: Beam-split-aware orthogonal matching pursuit for THz channel estimation," *IEEE Wireless Commun. Lett.*, early access, Feb. 7, 2023, doi: 10.1109/LWC.2023.3242699.

[51] B. Wang et al., "Spatial-wideband effect in massive MIMO with application in mmWave systems," *IEEE Commun. Mag.*, vol. 56, no. 12, pp. 134–141, Dec. 2018.

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

[53] X. Wu, W.-P. Zhu, and J. Yan, "Direction of arrival estimation for off-grid signals based on sparse Bayesian learning," *IEEE Sensors J.*, vol. 16, no. 7, pp. 2004–2016, Apr. 2016.

[54] A. M. Elbir and T. E. Tuncer, "Far-field DOA estimation and near-field localization for multipath signals," *Radio Sci.*, vol. 49, no. 9, pp. 765–776, Sep. 2014.

[55] A. M. Elbir, K. V. Mishra, and S. Chatzinotas, "NBA-OMP: Near-field beam-split-aware orthogonal matching pursuit for wideband THz channel estimation," Feb. 2023, *arXiv:2302.01682*.

[56] A. M. Elbir, A. K. Papazafeiropoulos, and S. Chatzinotas, "Federated learning for physical layer design," *IEEE Commun. Mag.*, vol. 59, no. 11, pp. 81–87, Nov. 2021.

[57] A. M. Elbir and S. Coleri, "Federated learning for channel estimation in conventional and RIS-assisted massive MIMO," *IEEE Trans. Wireless Commun.*, vol. 21, no. 6, pp. 4255–4268, Jun. 2022.

[58] A. M. Elbir, S. Coleri, A. K. Papazafeiropoulos, P. Kourtessis, and S. Chatzinotas, "A hybrid architecture for federated and centralized learning," *IEEE Trans. Cogn. Commun. Netw.*, vol. 8, no. 3, pp. 1529–1542, Sep. 2022.

[59] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. Y. Arcas, "Communication-efficient learning of deep networks from decentralized data," Feb. 2016, *arXiv:1602.05629*.

[60] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions," *IEEE Signal Process. Mag.*, vol. 37, no. 3, pp. 50–60, May 2020.

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

[62] J. Dai and H. C. So, "Sparse Bayesian learning approach for outlier-resistant direction-of-arrival estimation," *IEEE Trans. Signal Process.*, vol. 66, no. 3, pp. 744–756, Feb. 2018.

[63] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and Cramér–Rao bound: Further results and comparisons," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 12, pp. 2140–2150, Dec. 1990.

**ANASTASIOS K. PAPAZAFEIROPOULOS** (Senior Member, IEEE) received the B.Sc. degree (Hons.) in physics, the M.Sc. degree (Hons.) in electronics and computers science, and the Ph.D. degree from the University of Patras, Greece, in 2003, 2005, and 2010, respectively. From 2011 to 2012 and from 2016 to 2017, he was a Postdoctoral Research Fellow with the Institute for Digital Communications, University of Edinburgh, Edinburgh, U.K. From 2012 to 2014, he was a Research Fellow with Imperial College London, London, U.K. He is currently a Vice-Chancellor Fellow with the University of Hertfordshire, Hertfordshire, U.K., and also a Visiting Research Fellow with the SnT University of Luxembourg, Luxembourg. He has been involved in several EPSRC and EU FP7 Projects, including HIATUS and HARP. His research interests include machine learning for wireless communications, intelligent reflecting surfaces, massive MIMO, heterogeneous networks, 5G and beyond wireless networks, full-duplex radio, mm-wave communications, random matrix theory, hardware-constrained communications, and performance analysis of fading channels. He was the recipient of a Marie Curie Fellowship (IEF-IAWICOM).

**PANDELIS KOURTESSIS** (Member, IEEE) is currently the Director of the Centre for Engineering Research and a Reader in Communication Networks, University of Hertfordshire, U.K., leading the activities of the Networks Engineering Research Group into Communications and Information Engineering, including next generation passive optical networks, optical and wireless MAC protocols, 5G RANs, software-defined network & network virtualization 5G and satellite networks and more recently machine learning for next generation networks. His funding ID includes EU COST, FP7, H2020, European Space Agency (ESA), UKRI, and industrially funded projects. He has published more than 80 papers at peer-reviewed journals, peer-reviewed conference proceedings, and international conferences. His research has received coverage at scientific journals, magazines, white papers, and international workshops. He has served as the general chair, the co-chair, a technical program committee member, and at the scientific committees and expert groups for IEEE workshops and conferences, European technology platforms and European networks of excellence. He has been the Co-Editor of a Springer book and a Chapter Editor of an IET book on softwarization for 5G.

**AHMET M. ELBIR** (Senior Member, IEEE) received the B.S. degree (Hons.) in electrical engineering from Firat University in 2009, and the Ph.D. degree in electrical engineering from Middle East Technical University (METU) in 2016. He held Visiting Postdoctoral Researcher positions with Koc University, Istanbul, Turkey, from 2020 to 2021, and Carleton University, Canada, from 2022 to 2023. From 2016 to 2022, he was a Senior Researcher with Duzce University, Turkey. He is currently a Research Fellow with the University of Luxembourg, Luxembourg, and the University of Hertfordshire, U.K. His research interests include array signal processing, sparsity-driven convex optimization, signal processing for communications, and deep learning for array signal processing. He is the recipient of 2016 METU Best Ph.D. Thesis Award for his doctoral studies, and the IET Radar, Sonar and Navigation Best Paper Award in 2022. He has been serving as an Associate Editor for IEEE ACCESS since 2018 and the Lead Guest Editor for the IEEE JOURNAL OF SELECTED TOPICS ON SIGNAL PROCESSING and IEEE WIRELESS COMMUNICATIONS on Near-field Communications and Signal Processing.

**SYMEON CHATZINOTAS** (Fellow, IEEE) received the M.Eng. degree in telecommunications from the Aristotle University of Thessaloniki, Thessaloniki, Greece, in 2003, and the M.Sc. and Ph.D. degrees in electronic engineering from the University of Surrey, Guildford, U.K., in 2006 and 2009, respectively. He is currently a Full Professor and the Deputy Head of the SIGCOM Research Group, Interdisciplinary Centre for Security, Reliability, and Trust, University of Luxembourg, Esch-sur-Alzette, Luxembourg, and a Visiting Professor with the University of Parma, Parma, Italy. He has coauthored more than 400 technical papers in refereed international journals, conferences and scientific books. He has been involved in numerous research and development projects with the Institute of Informatics Telecommunications, National Center for Scientific Research Demokritos, Institute of Telematics and Informatics, Center of Research and Technology Hellas, and Mobile Communications Research Group, Center of Communication Systems Research, University of Surrey. His research interests include multiuser information theory, cooperative/cognitive communications, and wireless network optimization. He was the co-recipient of the 2014 IEEE Distinguished Contributions to Satellite Communications Award, the CROWNCOM 2015 Best Paper Award, and the 2018 EURASIP JWCN Best Paper Award. He is currently on the editorial board of the IEEE OPEN JOURNAL OF VEHICULAR TECHNOLOGY and the *International Journal of Satellite Communications and Networking*.

**WEI SHI** (Member, IEEE) received the Bachelor of Computer Engineering degree from the Harbin Institute of Technology and the master's and Ph.D. degrees in computer science from Carleton University, Ottawa, ON, Canada, where she is currently an Associate Professor with the School of Information Technology, cross appointed to the Department of Systems and Computer Engineering, Faculty of Engineering and Design. She is also a Professional Engineer licensed in Canada. She is specialized in the design and analysis of fault tolerance algorithms addressing security issues in distributed environments, such as data-center networks, clouds, mobile agents and actuator systems, wireless sensor networks, as well as critical infrastructures, such as power grids and smart cities. She has also been conducting research in data privacy and big data analytics.