Multi-dimensional Sparse CSI Acquisition for Hybrid mmWave MIMO OTFS Systems

Anand Mehrotra *Graduate Student Member, IEEE*, Jitendra Singh *Graduate Student Member, IEEE*, Suraj Srivastava, *Member IEEE*, Rahul Kumar Singh, Aditya K. Jagannatham *Senior Member, IEEE*, Lajos Hanzo, *Life Fellow, IEEE*

***Abstract*—Multi-dimensional sparse channel state information (CSI) acquisition is conceived for Orthogonal time frequency space (OTFS) modulation-based millimetre wave (mmWave) multiple input and multiple output (MIMO) systems. A comprehensive end-to-end relationship is derived in the delay-Doppler (DDA) domain by additionally considering the angular parameters and a hybrid beamforming (HB) architecture. A time-domain pilot model tailored for CSI estimation (CE) in the DDA-domain is proposed, which exploits the inherent multi-dimensional (4D) sparsity that emerges in the DDA-domain during the CE process. An efficient low-complexity Bayesian learning (LC-BL) technique is conceived to fulfil the objective of CSI estimation in such systems. Subsequently, a comprehensive examination of the complexity of the algorithm under consideration is also provided. It is worth noting that the complexity of the BL scheme designed is similar to that of popular orthogonal matching pursuit (OMP), but significantly lower than that of the traditional expectation-maximization (EM) based BL technique. Moreover, a single-stage transmit precoder (TPC) and receiver combiner (RC) design is proposed. This procedure aims for maximizing the directional gain of the RF TPC/RC pair by optimizing their weights. Additionally, a series of comprehensive simulations are conducted which incorporate the use of a practical channel model and fractional Doppler shifts. In light of the inherent trade-offs between complexity and estimation algorithm performance, our proposed scheme, LC-BL, appears suitable, especially considering the substantial enhancement in the performance of CE compared to the existing benchmarks.**

***Index Terms*—OTFS, MIMO, sparsity, channel estimation, delay-Doppler-angular domain, high-mobility, mmWave, hybrid precoding.**

The work of Aditya K. Jagannatham was supported in part by the Qualcomm Innovation Fellowship; in part by the Qualcomm 6G UR Gift; and in part by the Arun Kumar Chair Professorship.
The work of S. Srivastava was supported in part by IIT Jodhpur's Research Grant No. I/RIG/SUS/20240043; and in part by Telecom Technology Development Fund (TTDF) under Grant TTDF/6G/368.
L. Hanzo would like to acknowledge the financial support of the Engineering and Physical Sciences Research Council (EPSRC) projects under grant EP/Y037243/1, EP/W016605/1, EP/X01228X/1, EP/Y026721/1, EP/W032635/1, EP/Y037243/1 and EP/X04047X/1 as well as of the European Research Council's Advanced Fellow Grant QuantCom (Grant No. 789028).

Anand Mehrotra, J.Singh and A. K. Jagannatham are with the Department of Electrical Engineering, Indian Institute of Technology Kanpur, UP 208016, India (e-mail: anandme@iitk.ac.in, jitend@iitk.ac.in, adityaj@iitk.ac.in),
R.K.Singh is with Qualcomm India Pvt. Ltd. (e-mail: sksrahulkumar9696@gmail.com)
S. Srivastava is with the Department of Electrical Engineering, Indian Institute of Technology Jodhpur, Jodhpur, Rajasthan 342030, India (e-mail: surajsri@iitj.ac.in).
L. Hanzo is with the School of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, U.K. (e-mail: lh@ecs.soton.ac.uk).

I. Introduction

Next-generation (NG) communication systems are expected to enhance the data rates and reduce the latency compared to the existing 5G systems [1]. In such scenarios, the scarcity of available bandwidth in the sub-6GHz regime naturally motivates the migration to the high-frequency mm-wave spectrum in the 30 to 300 GHz band, where wide frequency blocks are readily available. Furthermore, the 4G and 5G communication systems employ orthogonal frequency division multiplexing (OFDM). This multicarrier modulation technique efficiently handles frequency-selectivity [2] of the channel. However, the most critical assumption to simplify signal processing is the quasi-static envelope of the wireless channel's fading over each OFDM symbol duration. This assumption is however invalid, in the face of Doppler spread, namely when the receiver (Rx) and transmitter (Tx) are in relative motion with respect to each other. Thus, in high or ultra-high mobility applications [3], [4], a diverse and dynamic fading environment is a reality. In such a scenario, more Doppler-resistant modulation schemes have to be investigated. Thus, a novel technique proposed by Hadani *et al.* [5]–[7] popularly known as orthogonal time frequency space (OTFS) modulation, is ideally suited to deal with doubly selective wireless channels. This new waveform in fact, can also be realized by incorporating the symplectic fast Fourier transform (SFFT) and its inverse into an OFDM transceiver core [8], [9].

The combination of the potent mmWave and OTFS architectures taps into rich bandwidth resources, for supporting high-speed data transmission. While mmWave OTFS-based systems offer an array of benefits, the practical realization of these gains depends on finding the optimal mmWave architecture, leading to hardware constraints. Other design challenges include their high propagation losses and signal blockages [10], which must be overcome. Furthermore, higher frequencies can cause significant Doppler impairments, making it difficult to use in high-mobility scenarios. This challenge can be addressed by employing transmit and receive antenna arrays, allowing for integrating multiple antennas in a compact space due to the shorter wavelengths [11]. Thus, it is imperative to include the design of RF transmit beamformers/precoders (TPC) and receiver combiners (RC) in the overall transceiver design. Additionally, these components must have precise CE in the DD-domain. This treatise focuses on these topics, and a concise state-of-the-art available is presented in the following section.

A. State-of-the-art

In the preliminary research era of mmWave communication, scholars focused on various aspects such as channel modelling and propagation characteristics at different carrier frequencies [10], [12], [13]. Furthermore, CE is a crucial area of research considered by many authors, and they have employed various techniques. In [14], [15], compressive sensing-based CE techniques are studied wherein the sparse power angle profile and training sequence-based design optimization are investigated, respectively. A few of the latest papers on mmWave include the use of deep learning-based compressive sensing [16], where the channel is modelled using a neural network and it is trained offline using simulated data.

Alkhateeb *et al.* proposed a multi-resolution codebook design for low-complexity sparse CE [17]. Ayach *et al.* [18] exploited the sparse mmWave wireless channel structure for formulating a constrained matrix reconstruction problem for the hybrid MIMO architecture and solved it using a simultaneous OMP (SOMP)-based approach. Along similar lines, Huang *et al.* [19] proposed a quad-phase iterative sparse CE scheme. It initially employs a least squares estimation and subsequently uses sparse message-passing (MP) detection for determining the non-zero CSI positions based on the estimated sparsity ratio. The authors of [20], [21] also exploited the mmWave MIMO channel sparsity and developed an improved CE technique having a significantly reduced pilot overhead employing the OMP and sparse Bayesian learning (SBL) techniques, respectively. Furthermore Li *et al.* [22] exploited the mmWave channel characteristics such as sparsity, angular spreads over the angle of arrival(AoA), angle of departure(AoD), and elevation domains to formulate a low-rank structure, which necessitates a substantially lower number of samples for CE. Similarly, AoA and AoD estimation for a sparse system using the beam space (or virtual) model for the MIMO channel is also investigated in [23]. Along similar lines, [24] considers a 3-D structure for the AOA-AOD-delay domain cluster of the channel that is imposed by the effect of power leakage, angular spread, as well as cluster duration, and proposes an extended approximate MP technique, together with a nearest neighbour pattern learning algorithm.

The aforementioned contributions solely focus on estimating a narrowband flat-fading channel, neglecting the key fact that mmWave MIMO channels are typically frequency-selective. In order to address this limitation, Venugopal *et al.* [25] and Rodriguez *et al.* [26] propose the use of OMP and SOMP algorithms for sparse CE in both single-carrier and multicarrier-wideband mmWave MIMO systems. Their studies focus on a mmWave MIMO OFDM system and leverage the sparsity of the channel. The authors of [27] adopt the Bayesian framework to develop mmWave CE schemes. They propose a novel scheme that performs matrix factorization and subsequently employs the variational Bayesian principle. The study conducted in [28] proposes a time-domain CE technique for a wideband mmWave MIMO OFDM system. The study by Talaei *et al.* [29] focuses on resolving the issue of limited angular resolution in mmWave MMO OFDM scenarios. They propose a method for recovering the continuous angular spectrum.

However, these early contributions did not consider any mobility. Gao *et al.* [30] investigated a doubly-selective mmWave MIMO channel model that considers multipath delays and Doppler shifts. In the initial stage, a training pattern is employed to detect the sparse delay domain channel taps using an energy detector. Subsequently, a modified OMP algorithm is utilised to address beamspace sparsity and to estimate the beam direction. The authors of [31]–[33] investigated the characteristics of a doubly-selective temporally-correlated block-fading channel in wideband mmWave MIMO systems. These contributions investigated the variability of the gains of the multipath components over time to effectively track the doubly-selective channel utilizing an SBL-based Kalman filter (SBL-KF). A notable drawback of these doubly-selective CE approaches is that they assume the channel in the TF-domain to vary over the TF-grid, which calls for frequent CE and leads to excessive overheads. Moreover, their performance suffers in high-Doppler scenarios. In addition to the DD-domain sparsity, the mmWave MIMO channel is also sparse in the angular domain [10], which can be beneficially exploited for CE. This calls for the development of efficient sparse learning strategies that impose low-complexity and exhibit reliable convergence properties.

The recently introduced multicarrier modulation scheme OTFS can offer substantial benefits in the context of communication over a doubly-selective channel. Briefly, OTFS enables input data to be mapped to a delay-Doppler (DD)-domain grid [5], [43], with the transformed DD-domain channel being relatively static over an extended observation interval [5], [44]. This renders CE more tractable in comparison to other multi-carrier modulation techniques. The pertinent input-output model is elucidated in [34], [36], [45]. In the context of OTFS CE, the initial efforts involve transmitting pilots either in the form of impulse-based signals or embedding them in data over a single frame at fixed locations [46]. Although these approaches succeed in mitigating the interference between the pilots and data, they result in a significant spectral efficiency reduction. The authors in [47] establish an end-to-end DD-domain relationship for OTFS modulation employing perfect bi-orthogonal and practical rectangular pulse waveforms. This work additionally investigates the different types of interference, and explicitly determines the inter-Doppler, inter-carrier, and inter-symbol interference components that arise, when the assumption of an ideal pulse shape is relaxed. Additionally, it proposes an iterative message passing (MP) based signal detection and interference cancellation technique. The recent treatises [35], [37], [48] exploit the DD-domain sparsity for achieving a superior CE performance. A high pilot overhead is one of the downsides of DD-domain CE, and emerges because a DD-domain guard interval is required for avoiding interference with data symbols in the same OTFS frame. This approach creates significant hurdles in MIMO-OTFS systems, since numerous guard intervals must be set up at each transmit antenna. In order to reduce the pilot overhead, the authors of [38], [49] suggest a new method that utilizes TF-domain pilots by transmitting the pilot symbols on a shared TF resource block for all the transmit antennas.

TABLE I
HB-MMWAVE MIMO OTFS CE: A TABULAR LITERATURE REVIEW

	[34]	[35]	[36]	[37]	[38]	[39]	[40]	[41]	[42]	Proposed
MIMO OTFS system	✓	✓	✓				✓	✓	✓	✓
mmWave hybrid design							✓		✓	✓
Delay-Doppler-angular channel		✓						✓	✓	✓
4-Dimensional sparsity								✓		✓
Practical pulse shape			✓	✓	✓	✓		✓	✓	✓
Flexible pilot overhead					✓			✓	✓	✓
Time domain pilots							✓	✓	✓	✓
DD-domain pilots	R	R	R	R	NR	R	R	NR	NR	NR
Fractional Dopplers			✓	✓		✓	✓			✓
Bayesian learning (BL) scheme				✓	✓	✓	✓		✓	✓
Low complexity BL scheme										✓
Single-stage precoder design										✓
Requirement of DD-guards	R	R	R	R	NR	NR	NR	NR	NR	NR

R: Required, NR: Not required

While a comprehensive understanding of OTFS has been developed in the extensive body of research discussed earlier, it is important to note that these previous studies do not consider the unique features of the high-frequency mmWave spectrum. To address this issue, the performance of a mmWave SISO OTFS system has been investigated in the seminal studies [5], [50]. Additionally, in reference [42], the authors conducted an in-depth study of mmWave MIMO OTFS systems, which plays a pivotal role in attaining inreased data rates under high-mobility conditions. Their research develops SBL and block-SBL algorithms for mmWave-analog beamforming (AB) and mmWave-hybrid beamforming (HB) assisted MIMO OTFS systems, respectively. Although the framework is innovative, it does not take advantage of the angular domain sparsity, which could significantly enhance its performance. The related work [40] also suggests an EM-based block SBL scheme for CE, which considers the transmission of pilots in both the TD and DD-domain, but this substantially increases the pilot overhead. The high complexity of traditional BL schemes, such as EM-based SBL, is due to the matrix inversion required for computing the covariance term in the E-step, which presents a significant challenge. Moreover, in literature, several authors have proposed diverse alternative methodologies to mitigate complexity [51]–[54] encountered in BL approaches. However, applying these approaches to the estimation of sparse multidimensional channels is not direct and demands a substantial amount of effort. Thus, it is desirable to conceive pragmatic CE schemes with specific emphasis on their practical implementation, which forms the focus of this paper. Table I presents a comprehensive overview and highlights the differences between our contributions and the state-of-the-art in this area. The following section provides a clear itemized description of the various contributions of this paper.

B. Contributions of the paper

1) In this study, the end-to-end relationship of OTFS-based mmWave-HB MIMO systems is derived within the DD domain, which enables the transmission of several concurrent data streams for spatial multiplexing. OTFS modulation and demodulation are described at each RF chain, followed by deriving a comprehensive system model constructed for facilitating signal detection.
2) A 4D-sparse CE problem is formulated via conceptualization of the channel model in the delay-Doppler-angular (DDA) domain, wherein the characteristics of each multipath component are defined by its delay, Doppler, angle of arrival (AoA), and angle of departure (AoD) parameters.
3) A formulation based on sparse signal recovery has been conceived for estimating the complex path gains of the multipath components, including their respective delay and Doppler shifts. Subsequently, a single-step method has been devised for designing the TPC/RC waveforms. This procedure optimizes the weights of the RF TPC/RC pair to maximize their directional gains.
4) While the performance of the traditional BL scheme is significantly improved compared to the conventional OMP technique harnessed for sparse recovery, the former traditionally suffers from an excessive complexity due to matrix inversion. To overcome this, we develop a low-complexity BL (LC-BL) CE procedure, for a mmWave-HB MIMO OTFS system. This is achieved using a sequential approach based on the Type-II estimation framework and on a stochastic maximum likelihood objective.
5) A detailed analysis of the complexity of the proposed low-complexity BL (LC-BL) algorithm is also presented, which is compared to that of the conventional OMP and EM-based BL algorithms. The results demonstrate that the complexity of the propounded LC-BL method is comparable to that of the OMP algorithm, despite yielding a significantly improved performance.

C. Organization of the work

The structure of this document is as follows:
The comprehensive model of an OTFS-based mmWave-HB MIMO system is expounded upon in Section II. This is then followed by presenting the proposed 4D-sparse CE in Section-III. This section also describes in detail our low-complexity BL scheme, referred to as the LC-BL. Subsequently, its complexity comparison with the competing OMP and EM-based BL methods is provided in Section-IV. Additionally, this section includes the convergence analysis for the proposed LC-BL.

The design pertaining to the RF TPC/ RC is elaborated upon in Section-V. The subsequent sections, namely Section-VI and Section-VII, present our simulation findings and conclude the paper, respectively.

D. Notation

Matrices are represented by uppercase boldface letters $\mathbf{A}$ and vectors by lowercase boldface letters $\mathbf{a}$. A diagonal matrix is identified by the notation $\mathrm{diag}(a_0, a_1, \cdots a_{N-1})$), with $a_0, a_1, \cdots a_{N-1}$ denoting the elements on the principal diagonal. On the other hand, the Nth-order identity matrix is represented by $\mathbf{I}_N$, while $\mathrm{vec}(\mathbf{A})$ transforms a matrix $\mathbf{A}$ into a vector via serial stacking of its columns. Conversely, $\mathrm{vec}^{-1}(\mathbf{a})$ performs the inverse operation. A noteworthy property namely $\mathrm{vec}\left(\mathbf{ABC}\right) = \left(\mathbf{C}^T \otimes \mathbf{A}\right)\mathrm{vec}\left(\mathbf{B}\right)$, is used at various places in this paper, where $\otimes$ denotes the matrix Kronecker product. Finally, $\delta(\cdot)$ represents the Dirac-delta function.

II. mmWave hybrid beamforming MIMO OTFS System Model

Consider the system model depicted in Fig.1. This system comprises of N_t transmitter antennas (TAs), N_r receiver antennas (RAs), and N_{RF} radio frequency (RF) chains. These parameters are related by obeying $N_{\text{RF}} << min(N_t, N_r)$ [10]. A transmit precoder (TPC) $\mathbf{F}_{\text{RF}} \in \mathbb{C}^{N_t \times N_{\text{RF}}}$ is utilized to map the symbols of the N_{RF} RF chains to the N_t TAs, where $N_{\text{RF}} \neq N_t$. In a similar way, the receiver employs a receiver combiner (RC) matrix $\mathbf{W}_{\text{RF}} \in \mathbb{C}^{N_r \times N_{\text{RF}}}$ to map the symbols received from the N_r RAs to the N_{RF} RF chains, with $N_{\text{RF}} \neq N_r$. The TPC and RC obey

$$|\mathbf{F}_{\text{RF}}(i,j)| = \frac{1}{\sqrt{N_t N_{RF}}}, 1 \leq i \leq N_t, 1 \leq j \leq N_{RF},$$
$$|\mathbf{W}_{\text{RF}}(k,l)| = \frac{1}{\sqrt{N_r N_{RF}}}, 1 \leq k \leq N_r, 1 \leq l \leq N_{RF}. \quad (1)$$

Within the context of this framework, the variables M represent the dimension of the symbol grid along the delay axis, and N represents the dimension along the Doppler axis. In addition, let ΔF and T denote the subcarrier spacing and the symbol duration, respectively, which satisfy the fundamental property of $T\Delta F = 1$. The following section presents the modulation and demodulation steps that are executed independently for each transmit and receive RF chain in the hybrid MIMO architecture.

A. OTFS modulation

Consider the DD-domain data input matrix at the uth transmit RF chain given by $\mathbf{X}_{\text{DD},u} \in \mathbb{C}^{M \times N}, 1 \leq u \leq N_{\text{RF}}$. The equivalent transformed TF-domain symbol matrix $\mathbf{X}_{\text{TF},u} \in \mathbb{C}^{M \times N}$ is obtained by applying the inverse symplectic finite Fourier transform (ISFFT) operation to $\mathbf{X}_{\text{DD},u}$. Consider the discrete Fourier transform (DFT) matrices of orders M and N, respectively, denoted by $\mathbf{F}_M$ and $\mathbf{F}_N$. Therefore, the expression $\mathbf{X}_{\text{TF},u}$ can be written as [47], [55]

$$\mathbf{X}_{\text{TF},u} = \mathbf{F}_M \mathbf{X}_{\text{DD},u} \mathbf{F}_N^H, \quad (2)$$

The time-domain (TD) symbol $s_u(t)$ is obtained by employing the Heisenberg transform followed by the transmit pulse shape filtering at each RF chain [5] given as

$$s_u(t) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \mathbf{X}_{\text{TF},u}(m,n) g_{\text{tx}}(t - nT) e^{j2\pi m \Delta F(t-nT)}, \quad (3)$$

where $\mathbf{X}_{\text{TF},u}(m,n)$ represents the (m,n)th element in the TF-domain matrix $\mathbf{X}_{\text{TF}}$ and g_{tx} is transmit pulse shaping filter. The TD samples can be obtained by sampling $s_u(t)$ at the Nyquist rate $\frac{M}{T}$, with the pth TD sample $s_u(p),\ 0 \leq p \leq MN-1$, described as $s_u(p) = s_u(t)|_{t=\frac{pT}{M}}$. Thus, the TD sample matrix $\mathbf{S}_u \in \mathbb{C}^{M \times N}$ and its vector representation $\mathbf{s}_u \in \mathbb{C}^{MN \times 1}$ for the uth RF chain is given by

$$\mathbf{S}_u = \mathbf{G}_{\text{tx}} \mathbf{F}_M^H \mathbf{X}_{\text{TF},u} = \mathbf{G}_{\text{tx}} \mathbf{X}_{\text{DD,u}} \mathbf{F}_N^H,$$
$$\mathbf{s}_u = \mathrm{vec}\left(\mathbf{S}_u\right) = \left(\mathbf{F}_N^H \otimes \mathbf{G}_{\text{tx}}\right) \mathbf{x}_{\text{DD,u}}, \quad (4)$$

where $\mathbf{G}_{tx} \in \mathbb{C}^{M \times M}$ is the matrix of transmit pulse samples obtained from sampling the pulse $\xi_{tx}(t)$, and defined as $\mathbf{G}_{tx} = \mathrm{diag}\left\{\xi_{tx}\left(\frac{pT}{M}\right)\right\}_{p=0}^{M-1}$. Additionally, $\mathbf{x}_{\text{DD,u}} = \mathrm{vec}\left(\mathbf{X}_{\text{DD,u}}\right)$. The system model adopted utilizes a cyclic prefix (CP)-aided approach [55], similar to OFDM systems, where a CP of length L is appended to the transmitted symbol vector $\mathbf{s_u}$ to eliminate interblock interference. This time-domain signal is communicated over the DDA-domain channel that is modeled as described in the next sub-section.

B. DDA-domain channel model for mmWave-HB MIMO OTFS systems

The wireless channel in the DDA domain, represented by the matrix $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$, can be characterized as a function of the delay (τ), Doppler shift (ν), angle of arrival (AoA) (θ), and angle of departure (AoD) (ϕ). This relationship can be expressed succinctly as [10], [56]

$$\mathbf{H}(\tau, \nu, \theta, \phi) = \sum_{p=1}^{P} \alpha_p \mathbf{a}_r\left(\theta\right) \mathbf{a}_t^H\left(\phi\right) \delta(\tau - \tau_p)\delta(\nu - \nu_p)\delta(\theta - \theta_p)\delta(\phi - \phi_p), \quad (5)$$

where P is the number of dominant multipath components, α_p denotes the path gain, ν_p is the Doppler-shift associated to the pth multipath component given by $\nu_p = \frac{k_p}{NT}$, with $k_p = \mathrm{round}(k_p) + \psi_{\nu_p}$, where $\left|\psi_{\nu_p}\right| < \frac{1}{2}$, denotes fractional-Doppler. Furthermore, as described in [36], [47], [56], [57] the delay resolution of $\Delta\tau = \frac{1}{M\Delta f}$ for a typical wide-band system is generally very fine. For instance, consider a system with parameters $M = 32$ and $\Delta f = 50$ kHz; the resultant delay resolution is $\Delta\tau = \frac{1}{M\Delta f} = 0.625\mu$s, for $M = 64$ and $\Delta f = 40$ kHz $\Delta\tau = \frac{1}{M\Delta f} = 0.39\mu$s which is reasonably small. Consequently, the delay shifts are assumed to be an integer multiple of this delay resolution, denoted as $\tau_p = l_p \Delta\tau$. Moreover, in the case of an under-spread wireless channel, it can be observed that $k_{\text{max}} = \max\left(k_p\right) << N$ and $l_{\text{max}} = \max\left(l_p\right) << M$, as discussed in [36], [47], [55]. The transmitter array steering vectors $\mathbf{a}_t\left(\phi_p\right) \in \mathbb{C}^{N_t \times 1}$ and receiver array steering vectors $\mathbf{a}_r\left(\theta_p\right) \in \mathbb{C}^{N_r \times 1}$, corresponding

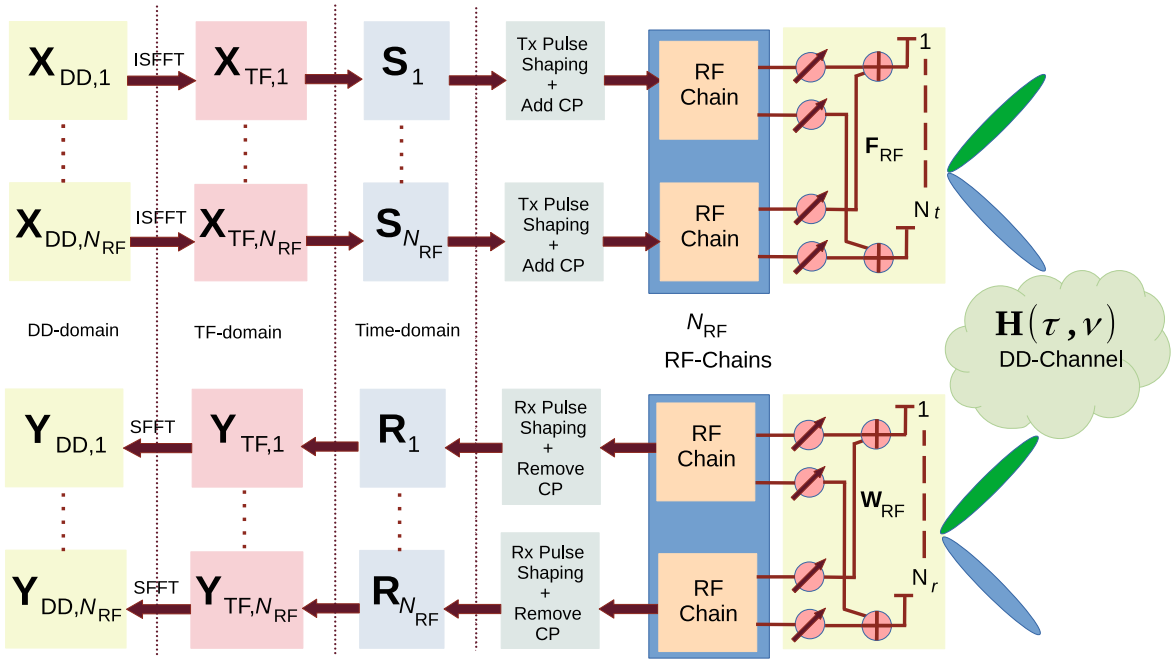


Fig. 1. Building blocks of the hybrid mmWave MIMO OTFS system.

to the AoD ϕ_p and AoA θ_p, respectively, are given as [10], [56]

$$\mathbf{a}_t(\phi_p) = \frac{1}{\sqrt{N_t}}\left[1, e^{-j\frac{2\pi}{\lambda}d_t\cos(\phi_p)}, \cdots, e^{-j\frac{2\pi}{\lambda}d_t(N_t-1)\cos(\phi_p)}\right],$$
$$\mathbf{a}_r(\theta_p) = \frac{1}{\sqrt{N_r}}\left[1, e^{-j\frac{2\pi}{\lambda}d_r\cos(\theta_p)}, \cdots, e^{-j\frac{2\pi}{\lambda}d_r(N_r-1)\cos(\theta_p)}\right]. \tag{6}$$

The symbol λ denotes the wavelength of the signal, whereas d_t and d_r represent the distances between consecutive TAs and RAs, respectively. Consequently, in response to the transmitted input signal $\mathbf{s}(t) \in \mathbb{C}^{N_{\text{RF}}\times 1}$ the output $\mathbf{r}(t) \in \mathbb{C}^{N_{\text{RF}}\times 1}$ obtained is given by

$$\mathbf{r}(t) = \sum_{p=1}^{P} \alpha_p \mathbf{W}_{\text{RF}}^H \mathbf{a}_r(\theta_p)\, \mathbf{a}_t^H(\phi_p)\, \mathbf{F}_{\text{RF}} \mathbf{s}(t-\tau) e^{j2\pi\nu_p(t-\tau_p)} + \widetilde{\mathbf{w}}(t), \tag{7}$$

where $\widetilde{\mathbf{w}}(t) = \mathbf{W}_{\text{RF}}^H \mathbf{w}(t) \in \mathbb{C}^{N_{RF}\times 1}$ and $\mathbf{w}(t) \in \mathbb{C}^{N_r\times 1}$ is the additive white Gaussian noise (AWGN).

Furthermore, the received TD signal $\mathbf{r}(t)$, it is sampled at the rate of $f_s = \frac{M}{T}$ followed by the CP removal, which leads to the following model for the received samples $\mathbf{r}(q) = \mathbf{r}(t)|_{t=\frac{qT}{M}}, 0 \le q \le MN-1$,

$$\mathbf{r}(q) = \sum_{p=1}^{P} \mathbf{W}_{\text{RF}}^H\, \mathbf{H}_p\, \mathbf{F}_{\text{RF}} \mathbf{s}\big([q-l_p]_{MN}\big) e^{j2\pi\frac{k_p(q-l_p)}{MN}} + \mathbf{w}(q), \tag{8}$$

where $\mathbf{H}_p = \alpha_p \mathbf{a}_r(\theta)\, \mathbf{a}_t^H(\phi) \in \mathbb{C}^{N_r\times N_t}$ denotes the pth dominant reflector between each TA-RA pair, while $\alpha_p = \sqrt{\frac{N_t N_r}{P}} h_p$, and $\mathbf{w}(q) = \mathbf{W}_{\text{RF}}^H \tilde{\mathbf{w}}(t)|_{t=\frac{qT}{M}}$ denotes the sampled noise vector. One can determine the complex path gain matrix $\widetilde{\mathbf{H}}_p \in \mathbb{C}^{N_{\text{RF}}\times N_{\text{RF}}}$ associated to the pth multipath component as

$$\widetilde{\mathbf{H}}_p = \mathbf{W}_{\text{RF}}^H \mathbf{H}_p \mathbf{F}_{\text{RF}}. \tag{9}$$

Moreover, the mmWave-HB MIMO OTFS system performs modulation and demodulation at each transmit and receive RF chain, respectively. The received symbol vector denoted by $\mathbf{r}_v \in \mathbb{C}^{MN\times 1}$ for the vth receive RF chain, is obtained by stacking the received symbol $\mathbf{r}_v(q)$ from (8) for $0 \le q \le MN-1$, which is given by $\mathbf{r}_v = [\mathbf{r}_v(0), \mathbf{r}_v(1), \cdots, \mathbf{r}_v(MN-1)]^T$. Thus, the received symbol vector $\mathbf{r}_v$ can be expressed as

$$\mathbf{r}_v = \sum_{u=1}^{N_{\text{RF}}} \overline{\mathbf{H}}_{v,u} \mathbf{s}_u + \mathbf{w}_v, \tag{10}$$

where $\mathbf{s}_u$ is the transmitted TD vector given by (4). The channel matrix $\overline{\mathbf{H}}_{v,u} \in \mathbb{C}^{MN\times MN}$ between the vth receive RF chain and uth transmit RF chain for $1 \le v \le N_{RF}$, $1 \le u \le N_{RF}$ can be formulated as

$$\overline{\mathbf{H}}_{v,u} = \sum_{p=1}^{P} \widetilde{h}_{p,v,u} (\mathbf{\Pi})^{l_p} \mathbf{\Delta}_p, \tag{11}$$

where $\widetilde{h}_{p,v,u}$ is obtained from (9) by using $\widetilde{h}_{p,v,u} = \widetilde{\mathbf{H}}_p(v,u)$. Furthermore, $\mathbf{\Pi}$ is a permutation matrix of size $MN \times MN$ which performs forward cyclic shift operation and $\mathbf{\Delta}_p \in \mathbb{C}^{MN\times MN}$ is a diagonal matrix defined as

$$\mathbf{\Delta}_p = \begin{cases} \text{diag}\left\{1, \omega_p, \cdots, \omega_p^{MN-l_p-1}, \omega_p^{-l_p}, \cdots, \omega_p^{-1}\right\} \\ \qquad\qquad \text{if } l_p \neq 0, \\ \text{diag}\left\{1, \omega_p, \cdots, \omega_p^{MN-1}\right\}, \text{if } l_p = 0, \end{cases} \tag{12}$$

where $\omega_p = e^{j2\pi \frac{k_p}{MN}}$.

C. OTFS demodulation

The signal $\mathbf{r}(t) \in \mathbb{C}^{N_{\text{RF}} \times 1}$ received at the output of the RC undergoes initial processing through a filter designed to match the receiver pulse $g_{\text{rx}}(t)$, of duration T. The processed signal Y(f, t) obtained can be expressed as

$$\mathbf{Y}(f,t) = \int_{t'} \mathbf{r}(t')\, g^*_{\text{rx}}(t'-t)\, e^{-j2\pi f(t'-t)} dt'. \tag{13}$$

The received signal $\mathbf{Y}(f,t)$ undergoes sampling at intervals that are integer multiples of the subcarrier spacing ΔF and symbol duration T. Thus the demodulated symbol $\mathbf{Y}_{\text{TF},v}(m,n)$ corresponding to the vth receive RF chain is formulated as $\mathbf{Y}_{\text{TF},v}(m,n) = \mathbf{Y}_v(f,t)|_{f=m\Delta F, t=nT}$, $0 \leq m \leq M-1, 0 \leq n \leq N-1$. It is important to note that the above relationship can be equivalently interpreted using the discrete Wigner transform [5], [47]. Thus, the TF-domain demodulated symbol matrix $\mathbf{Y}_{\text{TF},v} \in \mathbb{C}^{M \times N}$ at the vth receive RF chain is given by

$$\mathbf{Y}_{\text{TF},v} = \mathbf{F}_M \mathbf{G}_{\text{rx}} \mathbf{R}_v, \tag{14}$$

where similar to the transmit pulse shaping matrix, the received pulse matrix $\mathbf{G}_{rx} \in \mathbb{C}^{M \times M}$ is obtained by sampling $\xi_{rx}(t)$ and it is given as $\mathbf{G}_{rx} = \text{diag}\left\{\xi_{rx}\left(\frac{pT}{M}\right)\right\}_{p=0}^{M-1}$ as in [55] and $\mathbf{R}_v = \text{vec}^{-1}(\mathbf{r}_v) \in \mathbb{C}^{M \times N}$. Subsequently, the DD-domain demodulated signal $\mathbf{Y}_{\text{DD},v} \in \mathbb{C}^{M \times N}$ obtained by applying the SFFT transform over the TF-domain signal $\mathbf{Y}_{\text{TF},v}$, and its vector representation $\mathbf{y}_{\text{DD},v} \in \mathbb{C}^{MN \times 1}$, are given as

$$\begin{aligned} \mathbf{Y}_{\text{DD},v} &= \mathbf{F}_M^H \mathbf{Y}_{\text{TF},v} \mathbf{F}_N = \mathbf{G}_{\text{rx}} \mathbf{R}_v \mathbf{F}_N, \\ \mathbf{y}_{\text{DD},v} &= \text{vec}\left(\mathbf{Y}_{\text{DD},v}\right) = \left(\mathbf{F}_N \otimes \mathbf{G}_{\text{rx}}\right) \mathbf{r}_v. \end{aligned} \tag{15}$$

Thus, to obtain the DD-domain relationship between the transmitted signal $\mathbf{x}_{\text{DD},u}$ from the uth transmit RF chain and the received signal $\mathbf{y}_{\text{DD},v}$ of the vth receive RF chain, one can substitute $\mathbf{r}_v$ from (10), $\overline{\mathbf{H}}_{v,u}$ from (11), and $\mathbf{s}_u$ from (4) into (15), which yields

$$\begin{aligned} \mathbf{y}_{\text{DD},v} &= \left(\mathbf{F}_N \otimes \mathbf{G}_{\text{rx}}\right) \sum_{u=1}^{N_{\text{RF}}} \overline{\mathbf{H}}_{v,u} \mathbf{s}_u + \left(\mathbf{F}_N \otimes \mathbf{G}_{\text{rx}}\right) \mathbf{w}_v \\ &= \sum_{u=1}^{N_{\text{RF}}} \mathbf{H}_{\text{DD},v,u} \mathbf{x}_{\text{DD},u} + \widetilde{\mathbf{w}}_{\text{DD},v}, \end{aligned} \tag{16}$$

where $\widetilde{\mathbf{w}}_{\text{DD},v} = \left(\mathbf{F}_N \otimes \mathbf{G}_{rx}\right) \mathbf{w}_{\text{DD},v} \in \mathbb{C}^{MN \times 1}$ and the DD-domain channel matrix $\mathbf{H}_{\text{DD},v,u} \in \mathbb{C}^{MN \times MN}$ between the vth receive RF chain and uth transmit RF chain is given by

$$\mathbf{H}_{\text{DD},v,u} = \sum_{p=1}^{P} \left(\mathbf{F}_N \otimes \mathbf{G}_{rx}\right) \left[\widetilde{h}_{p,v,u} \left(\mathbf{\Pi}\right)^{l_p} \mathbf{\Delta}_p\right] \left(\mathbf{F}_N^H \otimes \mathbf{G}_{\text{tx}}\right). \tag{17}$$

Subsequently, upon stacking the demodulated signal vectors $\mathbf{y}_{\text{DD},v}$, $1 \leq v \leq N_{\text{RF}}$ from (16), one obtains $\mathbf{y}_{\text{DD}} \in \mathbb{C}^{MNN_{\text{RF}} \times 1}$. Therefore, the DD-domain end-to-end relationship can be modeled as

$$\begin{aligned} \mathbf{y}_{\text{DD}} &= \left[\mathbf{y}_{\text{DD},1}^T, \mathbf{y}_{\text{DD},2}^T, \cdots, \mathbf{y}_{\text{DD},N_{\text{RF}}}^T\right]^T \\ &= \mathbf{H}_{\text{DD}} \mathbf{x}_{\text{DD}} + \widetilde{\mathbf{w}}_{\text{DD}}. \end{aligned} \tag{18}$$

The quantities $\bar{\mathbf{x}}_{\text{DD}} \in \mathbb{C}^{MNN_{\text{RF}} \times 1}$ and $\tilde{\mathbf{w}}_{\text{DD}} \in \mathbb{C}^{MNN_{\text{RF}} \times 1}$ represent, respectively, the transmitted vector and the noise vector for all the receiver RF chains, which can be represented as

$$\overline{\mathbf{x}}_{\text{DD}} = \left[\mathbf{x}_{\text{DD},1}^T, \cdots, \mathbf{x}_{\text{DD},N_{\text{RF}}}^T\right]^T, \widetilde{\mathbf{w}}_{\text{DD}} = \left[\tilde{\mathbf{w}}_{\text{DD},1}^T, \cdots, \tilde{\mathbf{w}}_{\text{DD},N_{\text{RF}}}^T\right]^T,$$

while, $\mathbf{H}_{\text{DD}} \in \mathbb{C}^{MNN_{\text{RF}} \times MNN_{\text{RF}}}$ is the equivalent DD-domain channel matrix for the mmWave-HB MIMO OTFS system.

$\mathbf{H}_{\text{DD}}$ is defined as

$$\begin{aligned} \mathbf{H}_{\text{DD}} &= \begin{bmatrix} \mathbf{H}_{\text{DD},1,1} & \mathbf{H}_{\text{DD},1,2} & \cdots & \mathbf{H}_{\text{DD},1,N_{\text{RF}}} \\ \mathbf{H}_{\text{DD},2,1} & \mathbf{H}_{\text{DD},2,2} & \cdots & \mathbf{H}_{\text{DD},2,N_{\text{RF}}} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{H}_{\text{DD},N_{\text{RF}},1} & \mathbf{H}_{\text{DD},N_{\text{RF}},2} & \cdots & \mathbf{H}_{\text{DD},N_{\text{RF}},N_{\text{RF}}} \end{bmatrix} \\ &= \left(\mathbf{I}_{N_{\text{RF}}} \otimes \mathbf{F}_N \otimes \mathbf{G}_{rx}\right) \left[\sum_{p=1}^{P} \widetilde{\mathbf{H}}_p \otimes \left(\mathbf{\Pi}^{l_p} \mathbf{\Delta}_p\right)\right] \\ &\qquad \left(\mathbf{I}_{N_{\text{RF}}} \otimes \mathbf{F}_N^H \otimes \mathbf{G}_{\text{tx}}\right). \end{aligned} \tag{19}$$

The system model obtained in (18) can be further employed for data detection. The linear minimum mean square error (LMMSE) detector can be easily formulated for this model as

$$\mathbf{x}_{\text{DD}}^{\text{MMSE}} = \left(\mathbf{H}_{\text{DD}}^H \widetilde{\mathbf{R}}_{w,\text{DD}}^{-1} \mathbf{H}_{\text{DD}} + \mathbf{I}_{MNN_{\text{RF}}}\right)^{-1} \mathbf{H}_{\text{DD}}^H \widetilde{\mathbf{R}}_{w,\text{DD}}^{-1} \mathbf{y}_{\text{DD}}, \tag{20}$$

where the covariance matrix of the noise is $\widetilde{\mathbf{R}}_{w,\text{DD}} = \mathbb{E}\left[\widetilde{\mathbf{w}}_{\text{DD}} \widetilde{\mathbf{w}}_{\text{DD}}^H\right] = \sigma^2 \left[\left(\mathbf{W}_{\text{RF}}^H \mathbf{W}_{\text{RF}}\right) \otimes \mathbf{I}_N \otimes \left(\mathbf{G}_{\text{rx}} \mathbf{G}_{\text{rx}}^H\right)\right] \in \mathbb{C}^{MNN_{\text{RF}}}$. One of the critical phases prior to detection is the estimation of CSI, which is the focus of the subsequent section. Thus, a sparse DDA-domain CE model is developed next.

III. Proposed 4D sparse CE

Consider the DDA-domain mmWave-HB MIMO channel given by (5). To develop a 4D-sparse representation of the channel, it is necessary to take into account the dimensions of the DDA-domain grid. These are given as M_τ for the delay axis, N_ν for the integer-spaced Doppler axis or G_ν for the fractionally spaced Doppler axis, G_r for the angle of arrival (AOA) axis, and G_t for the angle of departure (AOD) axis. For a typical under-spread wireless channel, M_τ denotes the maximum delay-spread, while N_ν, and G_ν are the maximum Doppler-spreads for integer and fractional scenarios, respectively, which satisfy $l_{\max} < M_\tau << M$, $k_{\max} < N_\nu << N$ and $G_\nu >> N_\nu$ [58]. Here $\nu_j = \frac{jN_\nu}{G_\nu NT}$, $0 \geq j \geq G_\nu - 1$, represents the Doppler-shift in Hz corresponding to the jth Doppler-grid point. More precisely, one can define the delay grid points $\mathcal{M}(\tau) = \{\tau_i : \tau_i = \frac{i}{M\Delta f}\}_{i=0}^{M_\tau - 1}$, Doppler grid points $\mathcal{M}(\nu) = \{\nu_j : \nu_j = \frac{j}{NT}\}_{j=0}^{G_\nu - 1}$, AoA grid points $\mathcal{M}(\theta) = \{\theta_k : \theta_k = k\frac{\pi}{G_r}\}_{k=1}^{G_r}$, and AoD grid points

$\mathcal{M}(\phi) = \{\phi_l : \phi_l = l\frac{\pi}{G_t}\}_{l=1}^{G_t}$. Let $\beta_{i,j,k,l}$ denote the complex-valued dominant path gain corresponding to the ith delay, jth Doppler, kth AoA and lth AoD-index, respectively. The DDA-domain channel of (5) can be expressed as

$$\mathbf{H}(\tau,\nu,\theta,\phi) = \sum_{i=0}^{M_\tau-1}\sum_{j=0}^{G_\nu-1}\sum_{k=1}^{G_r}\sum_{l=1}^{G_t}\beta_{i,j,k,l}\mathbf{a}_r(\theta)\,\mathbf{a}_t^H(\phi) \delta(\tau-\tau_i)\delta(\nu-\nu_j)\delta(\theta-\theta_k)\delta(\phi-\phi_l). \quad (21)$$

One can represent the DDA-domain channel $\mathbf{H}_{i,j,k,l} \in \mathbb{C}^{N_r\times N_t}$ between the tth TA and rth RA at the (i,j,k,l)th grid-point as

$$\mathbf{H}_{i,j,k,l} = \beta_{i,j,k,l}\mathbf{a}_r(\theta_k)\,\mathbf{a}_t^H(\phi_l). \quad (22)$$

Furthermore, the equivalent DDA-domain channel between the RF RC and TPC is given by

$$\widetilde{\mathbf{H}}_{i,j,k,l} = \mathbf{W}_{\text{RF}}^H\mathbf{H}_{i,j,k,l}\mathbf{F}_{\text{RF}} \in \mathbb{C}^{N_{\text{RF}}\times N_{\text{RF}}}. \quad (23)$$

For each of the N_{RF} RF chains, a set of N_p pilot symbols is directly transmitted within the TD. Each frame constitutes N_pN_{RF} symbols, and Q such frames are sequentially transmitted to accurately estimate the DDA-domain channel. Let $\mathbf{s}_{\mathcal{P},u,q} \in \mathbb{C}^{N_p\times 1}, 1 \le q \le Q$ denote the pilot vector transmitted by the uth transmit RF chain in the qth frame duration. The pilot output $\mathbf{r}_{\mathcal{P},v,q}$ at the vth receive RF chain for the qth frame duration after discarding the L length CP, can be formulated using (10) and(11) as

$$\mathbf{r}_{\mathcal{P},v,q} = \sum_{u=1}^{N_{\text{RF}}}\left[\sum_{p=1}^{P}\widetilde{h}_{p,v,u}\left(\overline{\mathbf{\Pi}}\right)^{i_p}\overline{\mathbf{\Delta}}_{i_p,j_p}\right]\mathbf{s}_{\mathcal{P},u,q} + \mathbf{w}_{\mathcal{P},v,q}. \quad (24)$$

The matrix $\overline{\mathbf{\Pi}}$ denotes a permutation matrix of order $N_p\times N_p$ and $\overline{\mathbf{\Delta}}_{i,j} \in \mathbb{C}^{N_p\times N_p}$ is a diagonal matrix, defined as

$$\overline{\mathbf{\Delta}}_{i,j} = \begin{cases}\text{diag}\left\{1,\bar{\omega}_j,\cdots,(\bar{\omega}_j)^{N_p-i-1},(\bar{\omega}_j)^{-i},\cdots,(\bar{\omega}_j)^{-1}\right\}, \\ \qquad\qquad\qquad\qquad \text{if } i\neq 0,\\ \text{diag}\left\{1,\bar{\omega}_j,\cdots,(\bar{\omega}_j)^{N_p-1}\right\}, \text{if } i=0,\end{cases}$$

where $\bar{\omega}_m = e^{j2\pi\frac{mN_\nu}{G_\nu MN}}$. In the above expression (24) $(i,j) = (i_p,j_p)$ represents the delay and Doppler grid points associated with pth path, while $\mathbf{w}_{\mathcal{P},v,q} \in \mathbb{C}^{N_p\times 1}$ denotes the noise vector at the vth receive RF chain. Similar to (9), $\widetilde{h}_{p,v,u}$ represents the pth multipath component corresponding to vth receive RF chain and uth transmit RF chain. The expanded version of the pilot output $\mathbf{r}_{\mathcal{P},v,q}$ can be obtained as

$$\mathbf{r}_{\mathcal{P},v,q} = \sum_{u=1}^{N_{\text{RF}}}\sum_{i=0}^{M_\tau-1}\sum_{j=0}^{G_\nu-1}\sum_{k=1}^{G_r}\sum_{l=1}^{G_t}\left[\widetilde{h}_{i,j,k,l,v,u}\left(\overline{\mathbf{\Pi}}\right)^{i}\overline{\mathbf{\Delta}}_{i,j}\right]\mathbf{s}_{\mathcal{P},u,q} + \mathbf{w}_{\mathcal{P},v,q}, \quad (25)$$

where $\widetilde{h}_{i,j,k,l,v,u} = \widetilde{\mathbf{H}}_{i,j,k,l}(v,u)$ represents the complex path gain that corresponds to the ith delay-grid point, jth Doppler-grid point, kth AOA grid point, and lth AOD grid point for the communication link between vth receive RF chain and uth transmit RF chain. Subsequently, one can develop a compact model for the pilot output from (25) as follows

$$\mathbf{r}_{\mathcal{P},v,q} = \sum_{u=1}^{N_{RF}}\mathbf{H}_{\mathcal{P},v,u}\mathbf{s}_{\mathcal{P},u,q} + \mathbf{w}_{\mathcal{P},v,q}, \quad (26)$$

where the equivalent channel matrix $\mathbf{H}_{\mathcal{P},v,u} \in \mathbb{C}^{N_p\times N_p}$ between the vth receive RF chain and the uth transmit RF chain given by

$$\mathbf{H}_{\mathcal{P},v,u} = \sum_{i,j,k,l}\widetilde{h}_{i,j,k,l,v,u}\left(\overline{\mathbf{\Pi}}\right)^{i}\overline{\mathbf{\Delta}}_{i,j}. \quad (27)$$

After stacking the received output signals from all the receive RF chains, the vectorized pilot output $\mathbf{r}_{\mathcal{P},q} \in \mathbb{C}^{N_pN_{\text{RF}}\times 1}$ can be expressed as

$$\mathbf{r}_{\mathcal{P},q} = \left[\mathbf{r}_{\mathcal{P},1,q}^T,\mathbf{r}_{\mathcal{P},2,q}^T\cdots,\mathbf{r}_{\mathcal{P},N_{\text{RF}},q}^T\right]^T.$$

Thus, the equivalent vectorized system model for the qth frame duration can be derived as

$$\mathbf{r}_{\mathcal{P},q} = \mathbf{H}_{\mathcal{P}}\,\mathbf{s}_{\mathcal{P},q} + \mathbf{w}_{\mathcal{P},q}, \quad (28)$$

where the stacked pilot and noise vectors are given as

$$\mathbf{s}_{\mathcal{P},q} = \left[\mathbf{s}_{\mathcal{P},1,q}^T,\cdots,\mathbf{s}_{\mathcal{P},N_{\text{RF}},q}^T\right]^T \in \mathbb{C}^{N_pN_{\text{RF}}\times 1},$$
$$\mathbf{w}_{\mathcal{P},q} = \left[\mathbf{w}_{\mathcal{P},1,q}^T,\cdots,\mathbf{w}_{\mathcal{P},N_{\text{RF}},q}^T\right]^T \in \mathbb{C}^{N_pN_{\text{RF}}\times 1},$$

and the equivalent channel matrix $\mathbf{H}_{\mathcal{P},q}$ can be constructed as

$$\mathbf{H}_{\mathcal{P}} = \begin{bmatrix}\mathbf{H}_{\mathcal{P},1,1} & \mathbf{H}_{\mathcal{P},1,2} & \cdots & \mathbf{H}_{\mathcal{P},1,N_{\text{RF}}}\\ \mathbf{H}_{\mathcal{P},2,1} & \mathbf{H}_{\mathcal{P},2,2} & \cdots & \mathbf{H}_{\mathcal{P},2,N_{\text{RF}}}\\ \vdots & \vdots & \ddots & \vdots\\ \mathbf{H}_{\mathcal{P},N_{\text{RF}},1} & \mathbf{H}_{\mathcal{P},N_{\text{RF}},2} & \cdots & \mathbf{H}_{\mathcal{P},N_{\text{RF}},N_{\text{RF}}}\end{bmatrix}$$
$$= \text{blkmtx}\left(\{\mathbf{H}_{\mathcal{P},v,u}\}_{v=1,u=1}^{N_{\text{RF}},N_{\text{RF}}}\right) \in \mathbb{C}^{N_pN_{\text{RF}}\times N_pN_{\text{RF}}}. \quad (29)$$

By substituting $\mathbf{H}_{\mathcal{P},v,u}$ from (27) into the above equation, and further utilizing the relationships given in (22), (23), the quantity $\mathbf{H}_{\mathcal{P}}$ can be expressed as

$$\mathbf{H}_{\mathcal{P}} = \sum_{i,j,k,l}\beta_{i,j,k,l}\left[\left(\mathbf{w}_{\text{RF}}^H\mathbf{a}_r(\theta_k)\,\mathbf{a}_t^H(\phi_l)\,\mathbf{F}_{\text{RF}}\right)\otimes\left(\bar{\mathbf{\Pi}}^i\bar{\mathbf{\Delta}}_{i,j}\right)\right]. \quad (30)$$

Thereafter, upon substituting $\mathbf{H}_{\mathcal{P}}$ into (28), we obtain

$$\mathbf{r}_{\mathcal{P},q} = \sum_{i,j,k,l}\boldsymbol{\omega}_{i,j,k,l,q}\;\beta_{i,j,k,l} + \mathbf{w}_{\mathcal{P},q}, \quad (31)$$

where the quantity $\boldsymbol{\omega}_{i,j,k,l,q} \in \mathbb{C}^{N_pN_{\text{RF}}\times 1}$ is given by

$$\boldsymbol{\omega}_{i,j,k,l,q} = \left[\left(\mathbf{w}_{\text{RF}}^H\,\mathbf{a}_r(\theta_k)\,\mathbf{a}_t^H(\phi_l)\,\mathbf{F}_{\text{RF}}\right)\otimes\left(\bar{\mathbf{\Pi}}^i\bar{\mathbf{\Delta}}_{i,j}\right)\right]\mathbf{s}_{\mathcal{P},q}.$$

The equivalent CE model can be constructed as the 4D-sparse signal recovery problem

$$\mathbf{r}_{\mathcal{P},q} = \mathbf{\Omega}_q\boldsymbol{\beta} + \mathbf{w}_{\mathcal{P},q}. \quad (32)$$

It is evident that only a few, i.e., P coefficients out of the total $M_\tau G_\nu G_r G_t$ elements are non-zero in the vector $\boldsymbol{\beta} \in \mathbb{C}^{M_\tau G_\nu G_r G_t\times 1}$. Thus $P << M_\tau G_\nu G_r G_t$, which renders $\boldsymbol{\beta}$ a sparse vector. The dictionary matrix $\mathbf{\Omega}_q \in \mathbb{C}^{N_pN_{\text{RF}}\times M_\tau G_\nu G_r G_t}$

Algorithm 1: OMP-based sparse CE

Input: $\mathbf{\Omega}$, $\mathbf{y}_{\mathcal{P}}$

Initialization: Residue $\mathbf{r}_{-1} = \mathbf{0}_{N_p N_{RF} Q \times 1}$, $\mathbf{r}_0 = \mathbf{y}_{\mathcal{P}}$, $\widehat{\boldsymbol{\beta}}_{\text{OMP}} = \mathbf{0}_{M_\tau G_\nu G_r G_t \times 1}, \mathbf{\Omega}^{\mathcal{I}} = [\,]$, $i = 0$, Set $\mathbb{S} = [\,]$

Output: $\widehat{\boldsymbol{\beta}}_{\text{OMP}}$

while $\|\mathbf{r}_i\|^2 > \sigma^2 N_p N_{RF} Q$ **do**

- $i \leftarrow i + 1$
- $l = \underset{k=1,\cdots,M_\tau G_\nu G_r G_t}{\arg\max} \left|\mathbf{\Omega}^H(:,k)\mathbf{r}_{i-1}\right|$
- $\mathbb{S} = \mathbb{S} \cup l$
- $\mathbf{\Omega}^{\mathbb{S}} = \mathbf{\Omega}(:,\mathbb{S})$
- $\widehat{\boldsymbol{\beta}}^i_{\text{LS}} = \left(\mathbf{\Omega}^{\mathbb{S}}\right)^{\dagger} \mathbf{y}_{\mathcal{P}}$
- $\mathbf{r}_i = \mathbf{y}_{\mathcal{P}} - \mathbf{\Omega}^{\mathbb{S}} \widehat{\boldsymbol{\beta}}^i_{\text{LS}}$

$\widehat{\boldsymbol{\beta}}_{\text{OMP}}(\mathcal{I}) = \widehat{\boldsymbol{\beta}}^i_{\text{LS}}$

is comprised of $\boldsymbol{\omega}_{i,j,k,l,q}$ as its column vectors. Using the end-to-end concatenated pilot outputs from all the Q frames as $[\mathbf{r}_{\mathcal{P},1}, \mathbf{r}_{\mathcal{P},2}, \cdots, \mathbf{r}_{\mathcal{P},Q}] \in \mathbb{C}^{N_p N_{\text{RF}} Q \times 1}$, one can construct the sparse CE model for the 4-D sparse mmWave-HB MIMO OTFS system as

$$\mathbf{y}_{\mathcal{P}} = \mathbf{\Omega}\boldsymbol{\beta} + \mathbf{w}_{\mathcal{P}}, \tag{33}$$

where $\mathbf{\Omega} \in \mathbb{C}^{N_p N_{\text{RF}} Q \times M_\tau G_\nu G_r G_t}$ is the concatenated dictionary matrix $\mathbf{\Omega} = \left[\mathbf{\Omega}_1^T, \mathbf{\Omega}_2^T, \cdots, \mathbf{\Omega}_Q^T\right]^T$. The aforementioned problem of sparse estimation can be effectively addressed by applying well-established methods, such as the OMP or BL methodology, where a clear perfromance vs. complexity correlation exists. Explicitly, the former design exhibits lower complexity than the latter but it also demonstrates inferior performance. This inspires us to improve the BL scheme to compete with the complexity of the OMP algorithm, while offering enhanced performance for a 4D-sparse scenario. For comparison, Algorithm 1 outlines the popular OMP method for addressing a 4D-sparse signal recovery problem. The proposed low-complexity BL (LC-BL)-based sparse recovery scheme (Algorithm 2) is discussed in the next subsection. To determine the appropriate sparsity level, an effective approach is to monitor the residual norm during the iterative process. When the residual norm falls below a certain threshold, it indicates that the estimated signal is a close approximation of the original, suggesting that the effective sparsity level has been achieved, as illustrated in Algorithm 1. Alternatively, in the case of a sequential approach like OMP or the proposed LC-BL scheme, one can choose a support set K larger than the number of non-sparse values and iterate K times, progressively refining the solution, as demonstrated in Algorithm 2. Another method involves using information-theoretic techniques [59], which maximize the mutual information between the measurements and the sparse representation providing a statistically sound basis for identifying the optimal sparsity level during recovery.

Algorithm 2: LC-BL-based sparse CE

Input: $\mathbf{\Omega}$, $\mathbf{y}_{\mathcal{P}}$, K

Initialization: $\hat{\lambda} = \mathbf{0}$, $\mathbf{\Sigma}_{\mathbf{y}}^{-1} = \frac{1}{\sigma^2}\mathbf{I}$, $i = 0$, Set $\mathbb{S} = \emptyset$

Output: $\widehat{\boldsymbol{\beta}}_{\text{LC-BL}} = \widehat{\boldsymbol{\mu}}_{\beta}$

while $i < K$ **do**

- $i = i + 1$
- Compute a_i and $b_i, \forall i \notin \mathbb{S}$
- $l = \max\left\{\frac{|a_i|^2}{b_i}, 1\right\}$,
- $\widehat{\lambda}_l = \max\left\{\frac{|a_l|^2 - b_l}{b_l^2}, 0\right\}$; Rank 1 update $\mathbf{\Sigma}_{\mathbf{y}}^{-1}$
- $\mathbb{S} = \mathbb{S} \cup l$

Compute posterior mean, $\hat{\boldsymbol{\mu}}_{\beta}$, and covariance, $\hat{\mathbf{\Sigma}}_{\beta}$

A. LC-BL based sparse CE

Let us consider the sparse estimation problem described by equation (33). The BL procedure commences by assigning the Gaussian prior to the channel coefficients $\boldsymbol{\beta}$ follows as

$$f(\boldsymbol{\beta}; \boldsymbol{\lambda}) = \prod_{i=0}^{M_\tau G_\nu G_t G_r - 1} \frac{1}{(2\pi\lambda_i)^{1/2}} \exp\left(-\frac{|\boldsymbol{\beta}(i)|^2}{2\lambda_i}\right). \tag{34}$$

The unknown hyperparameter for the i-th component of the vector $\boldsymbol{\beta}$ is denoted as λ_i. By stacking all these hyperparameters, the vector is expressed as $\boldsymbol{\lambda} = [\lambda_1, \lambda_2, \ldots, \lambda_{M_\tau G_\nu G_t G_r}]^T$. The corresponding hyperparameter matrix is $\mathbf{\Lambda} = \text{diag}[\boldsymbol{\lambda}] \in \mathbb{R}^{M_\tau G_\nu G_t G_r \times M_\tau G_\nu G_t G_r}$ [8], [58], where the diag operator converts the vector into a diagonal matrix by placing its elements along the diagonal. The parameter $\boldsymbol{\lambda}$ can be estimated using the maximum likelihood framework formulated as

$$\begin{aligned}
\widehat{\boldsymbol{\lambda}} &= \arg\max_{\boldsymbol{\lambda}} \log p(\mathbf{y}_{\mathcal{P}}; \boldsymbol{\lambda}, \sigma^2) \\
&= \arg\max_{\boldsymbol{\lambda}} \log\left(\frac{1}{(2\pi)^{\frac{M_\tau G_\nu G_t G_r}{2}} |\mathbf{\Sigma}_{\mathbf{y}}|^{\frac{1}{2}}} \exp\left(-\frac{\mathbf{y}_{\mathcal{P}}^H \mathbf{\Sigma}_{\mathbf{y}}^{-1} \mathbf{y}_{\mathcal{P}}}{2}\right)\right) \\
&= \arg\max_{\boldsymbol{\lambda}} \left(-\frac{\mathbf{y}_{\mathcal{P}}^H \mathbf{\Sigma}_{\mathbf{y}}^{-1} \mathbf{y}_{\mathcal{P}}}{2} - \frac{1}{2}\log|\mathbf{\Sigma}_{\mathbf{y}}| - \frac{M_\tau G_\nu G_t G_r}{2}\log 2\pi\right) \\
&= \arg\max_{\boldsymbol{\lambda}} \mathcal{L}(\lambda),
\end{aligned} \tag{35}$$

where we have $\mathbf{\Sigma}_{\mathbf{y}} = \mathbf{\Omega}\mathbf{\Lambda}\mathbf{\Omega}^H + \sigma^2 \mathbf{I}_{N_p N_{\text{RF}} Q}$. The above problem can be expressed as

$$\widehat{\boldsymbol{\lambda}} = \arg\min_{\boldsymbol{\lambda}} \left(\mathbf{y}_{\mathcal{P}}^H \mathbf{\Sigma}_{\mathbf{y}}^{-1} \mathbf{y}_{\mathcal{P}} + \log|\mathbf{\Sigma}_{\mathbf{y}}|\right). \tag{36}$$

The second term $\log|\mathbf{\Sigma}_{\mathbf{y}}|$ above is a non-convex function in $\mathbf{\Lambda}$, which makes the hyperparamter estimation problem intractable. Thus, the likelihood maximization above can be achieved via the iterative EM algorithm. However, the complexity of using such a method is excessive.

Thus a low-complexity variant can be obtained by decomposing $\mathbf{\Sigma_y}$ in terms of the hyperparameter λ_i i.e., by separating the contribution of ith column to the cost function as

$$\begin{aligned}\mathbf{\Sigma_y} &= \mathbf{\Omega\Lambda\Omega}^H + \sigma^2\mathbf{I}_{N_pN_{\text{RF}}Q} \\ &= \sigma^2\mathbf{I}_{N_pN_{RF}N_Q} + \sum_{m\neq i}\lambda_m^{-1}\mathbf{\Xi}_m\mathbf{\Xi}_m^H + \lambda_i^{-1}\mathbf{\Xi}_i\mathbf{\Xi}_i^H \\ &= \mathbf{\Sigma}_{\mathbf{y},m\neq i} + \lambda_i^{-1}\mathbf{\Xi}_i\mathbf{\Xi}_i^H. \end{aligned} \tag{37}$$

where $m = [1, 2, \cdots, M_\tau G_\nu G_t G_r]$ and $\mathbf{\Omega} = [\mathbf{\Xi}_1, \mathbf{\Xi}_2 \cdots, \mathbf{\Xi}_{M_\tau G_\nu G_t G_r}]$. The quantities $|\mathbf{\Sigma_y}|$ and $\mathbf{\Sigma}_\mathbf{y}^{-1}$ can be computed from (37) as

$$\begin{aligned}|\mathbf{\Sigma_y}| &= |\mathbf{\Sigma}_{\mathbf{y},m\neq i}|\left|1 + \lambda_i^{-1}\mathbf{\Xi}_i^H\mathbf{\Sigma}_{\mathbf{y},m\neq i}^{-1}\mathbf{\Xi}_i\right|, \\ \mathbf{\Sigma}_\mathbf{y}^{-1} &= \mathbf{\Sigma}_{\mathbf{y},m\neq i}^{-1} - \frac{\mathbf{\Sigma}_{\mathbf{y},m\neq i}^{-1}\mathbf{\Xi}_i\mathbf{\Xi}_i^H\mathbf{\Sigma}_{\mathbf{y},m\neq i}^{-1}}{\lambda_i^{-1} + \mathbf{\Xi}_i^H\mathbf{\Sigma}_{\mathbf{y},m\neq i}^{-1}\mathbf{\Xi}_i}. \end{aligned} \tag{38}$$

Substituting $|\mathbf{\Sigma_y}|$ and $\mathbf{\Sigma}_\mathbf{y}^{-1}$ in (35) yields (39) as shown at the top of the next page. Here, we calculate the marginal likelihood by excluding the contribution of $\mathbf{\Xi}_i$ given by $\mathcal{L}(\boldsymbol{\lambda}_{m\neq i})$, and the contribution of $\mathbf{\Xi}_i$ is given by $\ell(\lambda_i)$. Furthermore, $\ell(\lambda_i)$ in (39) can be simplified as

$$\ell(\lambda_i) = -\frac{1}{2}\left[-\frac{|a_i|^2}{\lambda_i + b_i} + \log\left(1 + \lambda_i^{-1}b_i\right)\right], \tag{40}$$

where a_i and b_i are given as

$$\begin{aligned}a_i &= \mathbf{\Xi}_i^H\mathbf{\Sigma}_{\mathbf{y},m\neq i}^{-1}\mathbf{y}_\mathcal{P}, \\ b_i &= \mathbf{\Xi}_i^H\mathbf{\Sigma}_{\mathbf{y},m\neq i}^{-1}\mathbf{\Xi}_i. \end{aligned} \tag{41}$$

Minimizing $\ell(\lambda_i)$ with respect to λ_i yields

$$\lambda_i^{\text{opt}} = \max\left\{\frac{|a_i|^2 - b_i}{b_i^2}, 0\right\}. \tag{42}$$

Upon substituting λ_i^{opt} into (40), yields

$$\ell\left(\lambda_i^{\text{opt}}\right) = \begin{cases} \log\frac{|a_i|^2}{b_i} - \frac{|a_i|^2}{b_i} + 1 & \text{if } |a_i|^2 > b_i \\ 0 & \text{if } |a_i|^2 \leq b_i. \end{cases} \tag{43}$$

Note that the above function $\ell\left(\lambda_i^{\text{opt}}\right)$ is monotonic non-increasing. Moreover, since $\lambda_{\text{prev},i} = 0$, it follows that $\Delta\ell(\lambda_i, \lambda_{\text{prev},i}) = \ell(\lambda_i) - \ell(\lambda_{\text{prev},i}) = \ell(\lambda_i)\ \forall i \notin \mathbb{S}$. Thus, in order to determine the column to be added in $\mathbb{S}$, one computes

$$\begin{aligned} l &= \arg\min_{i\notin\mathbb{S}}\min_{\lambda_i\geq 0}\ell(\lambda_i) \\ &= \arg\min_{i\notin\mathbb{S}}\min_{\lambda_i\geq 0}\Delta\ell(\lambda_i, \lambda_{\text{prev},i}) \\ &= \max\left\{\frac{|a_i|^2}{b_i}, 1\right\}. \end{aligned} \tag{44}$$

Updates of a_i and b_i:
In order to reduce the complexity, we take advantage of the fact that each iteration adds one column of $\mathbf{\Omega}$ and updates $\mathbf{\Sigma}_\mathbf{y}^{-1}$ using the column selected in $\mathbb{S}$. Let $l[k]$ represent the index of the column of $\mathbf{\Omega}$ that will be added in the kth iteration given by $\mathbf{\Xi}_{l[k]}$. The quantity $\left(\mathbf{\Sigma}_\mathbf{y}^{[k+1]}\right)^{-1}$ can be obtained using (38) as

$$\left(\mathbf{\Sigma}_\mathbf{y}^{[k+1]}\right)^{-1} = \left(\mathbf{\Sigma}_\mathbf{y}^{[k]}\right)^{-1} - \frac{\mathbf{c}^{[k]}\left(\mathbf{c}^{[k]}\right)^H}{\widehat{\lambda}_{l[k]}^{-1} + b_{l[k]}^{[k]}}, \tag{45}$$

where $\mathbf{\Sigma}_\mathbf{y}^{[k]}$ includes the contribution of all the columns in $\mathbb{S}$, except for the column corresponding to index $l[k+1]$ of $\mathbf{\Omega}$, while $\widehat{\lambda}_{l[k]} = \max\left\{\frac{\left|a_{l[k]}^{[k]}\right|^2 - b_{l[k]}}{(b_{l[k]}^{[k]})^2}, 0\right\}$, $\mathbf{c}^{[k]}$ and $b_{l[k]}^{[k]}$ are defined as

$$\begin{aligned}\mathbf{c}^{[k]} &= \left(\mathbf{\Sigma}_\mathbf{y}^{[k]}\right)^{-1}\mathbf{\Xi}_{l[k]}, \\ b_{l[k]}^{[k]} &= \mathbf{\Xi}_{l[k]}^H\left(\mathbf{\Sigma}_\mathbf{y}^{[k]}\right)^{-1}\mathbf{\Xi}_{l[k]}. \end{aligned} \tag{46}$$

Thus, one can update $\left(a_i^{[k]}, b_i^{[k]}\right)$ using (41) and (45) to obtain $\left(a_i^{[k+1]}, b_i^{[k+1]}\right)$ as

$$\begin{aligned} a_i^{[k+1]} &= \mathbf{\Xi}_i^H\left(\mathbf{\Sigma}_\mathbf{y}^{[k+1]}\right)^{-1}\mathbf{y}_\mathcal{P} \\ &= a_i^{[k]} - \frac{\mathbf{\Xi}_i^H\mathbf{c}^{[k]}}{\widehat{\lambda}_{l^{[k]}}^{-1} + b_{l[k]}^{[k]}}a_{l[k]}^{[k]}, \end{aligned} \tag{47}$$

$$\begin{aligned} b_i^{[k+1]} &= \mathbf{\Xi}_i^H\left(\mathbf{\Sigma}_\mathbf{y}^{[k+1]}\right)^{-1}\mathbf{\Xi}_i \\ &= b_i^{[k]} - \frac{\left|\mathbf{\Xi}_i^H\mathbf{c}^{[k]}\right|^2}{\widehat{\lambda}_{l^{[k]}}^{-1} + b_{l[k]}^{[k]}}. \end{aligned} \tag{48}$$

Thus, the posterior mean $\widehat{\boldsymbol{\mu}}_\beta$ and covariance matrix $\widehat{\mathbf{\Sigma}}_\beta$ can be estimated as

$$\widehat{\boldsymbol{\mu}}_\beta = \widehat{\mathbf{\Lambda}}\mathbf{\Omega}^H\left(\mathbf{\Omega}\widehat{\mathbf{\Lambda}}\mathbf{\Omega}^H + \sigma^2\mathbf{I}\right)^{-1}\mathbf{y}_\mathcal{P} = \widehat{\mathbf{\Lambda}}\mathbf{a}^{[K+1]}, \tag{49}$$

where $\mathbf{a}^{[K+1]} = \left[a_1^{[K+1]}, \ldots, a_{M_\tau G_\nu G_r G_t}^{[K+1]}\right]^T$. Note that we have $\widehat{\lambda}_i \neq 0$ only for $i \in \mathbb{S}$, demonstrating that the above equation provides a sparse solution. Also, the diagonal entries of $\widehat{\mathbf{\Sigma}}_\beta$ can be readily obtained as follows

$$\begin{aligned}\left[\widehat{\mathbf{\Sigma}}_\beta\right]_{i,i} &= \widehat{\lambda}_i - \widehat{\lambda}_i^2\mathbf{\Xi}_i^H\left(\mathbf{\Omega}\widehat{\mathbf{\Lambda}}\mathbf{\Omega}^H + \sigma^2\mathbf{I}\right)^{-1}\mathbf{\Xi}_i \\ &= \widehat{\lambda}_i - \widehat{\lambda}_i^2 b_i^{[K+1]}, \end{aligned} \tag{50}$$

which is similarly non-zero only for $i \in \mathbb{S}$.

Computation of $\mathbf{c}^{[k]}$:
Substituting $\left(\mathbf{\Sigma}_\mathbf{y}^{[k]}\right)^{-1}$ from (45) into (46) and expressing it in terms of $\left(\mathbf{\Sigma}_\mathbf{y}^{[k-1]}\right)^{-1}$ one obtains

$$\mathbf{c}^{[k]} = \left(\mathbf{\Sigma}_\mathbf{y}^{[k-1]}\right)^{-1}\mathbf{\Xi}_{l[k]} - \frac{\mathbf{c}^{[k-1]}\left(\mathbf{c}^{[k-1]}\right)^H\mathbf{\Xi}_{l[k]}}{\lambda_{l[k-1]}^{-1} + b_{l[k-1]}^{[k-1]}} \tag{51}$$

$$= \left(\mathbf{\Sigma}_\mathbf{y}^{[1]}\right)^{-1}\mathbf{\Xi}_{l^{[k]}} - \sum_{\widetilde{k}=1}^{k-1}\frac{\mathbf{c}^{[\widetilde{k}]}\left(\mathbf{c}^{[\widetilde{k}]}\right)^H\mathbf{\Xi}_{l[k]}}{\lambda_{l^{[\widetilde{k}]}}^{-1} + b_{l^{[\widetilde{k}]}}^{[\widetilde{k}]}}, \tag{52}$$

$$
\begin{aligned}
\mathcal{L}(\boldsymbol{\lambda}) &= -\frac{\mathbf{y}_{\mathcal{P}}^H \boldsymbol{\Sigma}_{\mathbf{y}}^{-1} \mathbf{y}_{\mathcal{P}}}{2} - \frac{1}{2}\log|\boldsymbol{\Sigma}_{\mathbf{y}}| - \frac{M_\tau G_\nu G_t G_r}{2}\log 2\pi \\
&= -\frac{1}{2}\left[\mathbf{y}_{\mathcal{P}}^H \boldsymbol{\Sigma}_{\mathbf{y},m\neq i}^{-1}\mathbf{y}_{\mathcal{P}} - \mathbf{y}_{\mathcal{P}}^H \frac{\boldsymbol{\Sigma}_{\mathbf{y},m\neq i}^{-1}\boldsymbol{\Xi}_i\boldsymbol{\Xi}_i^H\boldsymbol{\Sigma}_{\mathbf{y},m\neq i}^{-1}}{\lambda_i^{-1}+\boldsymbol{\Xi}_i^H\boldsymbol{\Sigma}_{\mathbf{y},m\neq i}^{-1}\boldsymbol{\Xi}_i}\mathbf{y}_{\mathcal{P}} + \log\left|\boldsymbol{\Sigma}_{\mathbf{y},m\neq i}^{-1}\right| + \log\left(1+\lambda_i^{-1}\boldsymbol{\Xi}_i^T\boldsymbol{\Sigma}_{\mathbf{y},m\neq i}^{-1}\boldsymbol{\Xi}_i\right) + M_\tau G_\nu G_t G_r \log 2\pi\right] \\
&= \underbrace{-\frac{\mathbf{y}_{\mathcal{P}}^H \boldsymbol{\Sigma}_{\mathbf{y},m\neq i}^{-1}\mathbf{y}_{\mathcal{P}}}{2} - \frac{1}{2}\log\left|\boldsymbol{\Sigma}_{\mathbf{y},m\neq i}^{-1}\right| - \frac{M_\tau G_\nu G_t G_r}{2}\log 2\pi}_{\mathcal{L}(\boldsymbol{\lambda})_{m\neq i}} - \frac{1}{2}\left[-\frac{|\boldsymbol{\Xi}_i^H\boldsymbol{\Sigma}_{\mathbf{y},m\neq i}^{-1}\mathbf{y}_{\mathcal{P}}|^2}{\lambda_i^{-1}+\boldsymbol{\Xi}_i^H\boldsymbol{\Sigma}_{\mathbf{y},m\neq i}^{-1}\boldsymbol{\Xi}_i} + \log\left(1+\lambda_i^{-1}\boldsymbol{\Xi}_i^H\boldsymbol{\Sigma}_{\mathbf{y},m\neq i}^{-1}\boldsymbol{\Xi}_i\right)\right] \\
&= \mathcal{L}\left(\boldsymbol{\lambda}_{m\neq i}\right) + \ell\left(\lambda_i\right). \qquad (39)
\end{aligned}
$$

where $\left(\boldsymbol{\Sigma}_{\mathbf{y}}^{[1]}\right)^{-1} = 1/\sigma^2\mathbf{I}$. Algorithm 2 succinctly presents the sequence of steps that constitute the LC-BL scheme. The next section delves into the computational complexity of the proposed scheme in comparison to both that of the OMP scheme in Algorithm 1 and to the conventional EM-based BL approach [42].

IV. Computational Complexity and Convergence Analysis

The assessment of complexity for these algorithms is based on the number of multiplication as well as division operations, and the step-wise complexity analysis is presented in Table 2. From the table, it can be observed that the overall complexity of the proposed LC-BL algorithm is on the order of $\mathcal{O}(N_p N_{RF} Q M_\tau G_\nu G_r G_t)$, which is comparable to the complexity of OMP algorithm and is much lower compared to the conventional EM-based BL algorithm complexity which is of order $\mathcal{O}(M_\tau^3 G_\nu^3 G_r^3 G_t^3)$. This analysis reveals that the complexity of LC-BL is similar to that of the OMP scheme, while its performance is significantly superior to the conventional OMP scheme, as it will be shown by the simulation results of Section VI. Another important aspect of the proposed LC-BL algorithm is its convergence analysis, which is presented below.

The proposed sequential algorithm iteratively optimizes the likelihood function by selecting a column or basis vector from the dictionary that maximizes the marginal-likelihood at each step. This is done by isolating the contribution of column i as $l(\lambda_i)$ from the overall likelihood function $\mathcal{L}(\lambda)$, as shown in equation (40). Notably $\mathcal{L}(\lambda)$ has a unique maximum with respect to λ_i. The selection of the column l is based on a closed-form expression derived in equation (44), where the $l(\lambda_i^{\text{opt}})$ is defined in equation (43). Consequently, the algorithm either increases or maintains the log-likelihood value upon each iteration, making it a non-decreasing function. This, in turn, results in a non-increasing negative log-likelihood, ensuring convergence to a stationary point for the optimal set of hyperparameters. It is important to emphasize that the algorithm is guaranteed to increase the marginal likelihood at each step until it reaches a local maximum. Although it may seem that basis vectors are being "added" upon each iteration, in reality, the algorithm simultaneously maintains posterior statistics for all basis vectors (all elements of $\boldsymbol{\Sigma}$ and $\boldsymbol{\mu}$ corresponding to 'out of model' basis vectors are trivially zero).

The proposed LC-BL algorithm differs from the EM-BL algorithm in the following aspects. EM-BL operates by simultaneously updating all hyperparameters γ_j across the entire dictionary matrix $\boldsymbol{\Phi}$. This comprehensive approach ensures that all components of the model are adjusted in concert, which can lead to improved convergence to the global minimum of the negative log-marginal likelihood. However, the simultaneous update strategy is computationally expensive due to the need for repeated inversion of large matrices.

LC-BL, on the other hand, updates the hyperparameters λ_j sequentially, focusing on a single component at a time. This is done by exploring the properties of the likelihood function $\mathcal{L}(\lambda)$. To elaborate, the proposed sequential algorithm iteratively optimizes the likelihood function by selecting a column or basis vector from the dictionary matrix $\boldsymbol{\Phi}$ that maximizes the marginal-likelihood at each step. While this significantly reduces the computational complexity, it can potentially lead to suboptimal updates in certain iterations, as the algorithm may get trapped in local minima. The sequential update may not fully capture the interactions between different components of the model, which can lead to a less accurate estimation of the hyperparameters compared to EM-BL. However, the convergence to local minima also ensures a sparse solution [51], at a lower complexity.

It is important to emphasize that the proposed scheme is particularly relevant for MIMO systems because the dictionary size increases with the number of transmit and receive antennas in the MIMO scenario. This significantly exacerbates the complexity of the problem, especially when using standard approaches like the EM algorithm, which involves the inversion of high-dimensional matrices. In the proposed LC-BL algorithm the presence of a MIMO system impacts both the formulation of the dictionary matrix, which depends on the number of transmit and receive antennas, and the selection of hyperparameters. Since LC-BL does not optimize all the hyperparameters simultaneously, the complexity is lower. However, in each iteration of this sequential approach, a hyperparameter is chosen from a set of size $(M_\tau G_\nu G_t G_r)$ using equation (43) to minimize the cost function and maximise the marginal likelihood. The process is iteratively repeated, focusing on the remaining hyperparameters, until the number of iterations matches the size of the support set. This support

TABLE II
COMPUTATIONAL COMPLEXITY COMPARSION

Steps	LC-BL	Steps	OMP	Steps	EM-BL [42] [58]
Iteration $i+1$	$N_pN_{RF}QM_\tau G_\nu G_r G_t + 2(N_pN_{RF}Q)p+$ $M_\tau G_\nu G_r G_t + N_pN_{RF}Q+$ $2(M_\tau G_\nu G_r G_t - i) + (M_\tau G_\nu G_r G_t - i)$	Projection step	$N_pN_{RF}QM_\tau G_\nu G_r G_t$	$\boldsymbol{\Sigma}^{(i)}$	$\frac{M_\tau^3 G_\nu^3 G_r^3 G_t^3}{2} - \frac{3M_\tau^2 G_\nu^2 G_r^2 G_t^2}{2} +$ $M_\tau^2 G_\nu^2 G_r^2 G_t^2 N_p N_{RF} Q$
Solution	$N_pN_{RF}Q(3K-1)+2K$	LS step	$\frac{1}{2}i^3 + \frac{5}{2}i^2 + i^2 N_p N_{RF} Q +$ $i N_p N_{RF} Q$	$\boldsymbol{\mu}^{(i)}$	$M_\tau^2 G_\nu^2 G_r^2 G_t^2 +$ $M_\tau G_\nu G_r G_t N_p N_{RF} Q$
		Residue step	$iN_pN_{RF}Q$	$\gamma^{(i)}$	$M_\tau G_\nu G_r G_t$
Complexity	$\mathcal{O}(N_pN_{RF}QM_\tau G_\nu G_r G_t)$		$\mathcal{O}(N_pN_{RF}QM_\tau G_\nu G_r G_t$		$\mathcal{O}(M_\tau^3 G_\nu^3 G_r^3 G_t^3)$

set is chosen to have a size larger than the number of non-zero values or can be determined using sparse estimation techniques such as the OMP. The next section presents the proposed RF TPC/RC design.

V. RF PRECODER/ COMBINER DESIGN FOR MMWAVE-HB MIMO OTFS SYSTEMS

This treatise proposes a single-stage RF TPC $\mathbf{F}_{\text{RF}} \in \mathbb{C}^{N_t \times N_{\text{RF}}}$ and RC $\mathbf{W}_{\text{RF}} \in \mathbb{C}^{N_r \times N_{\text{RF}}}$strategy, which is based on selecting the dominant path gains from the estimated channel and using these gains to construct the RF TPC and RC. Hence our design maximizes the directional beamforming gain. The strategy leverages the correlation between the received pilots and the array response vectors corresponding to the dominant paths. After the sparse CE, the array response vector corresponding to the N_{RF} dominant paths yields the optimal precoder and combiner. These correspond to the N_{RF} indices of the combining beams $a_r(\theta_k)$ in the codebook $\mathbf{A}_R$ that exhibit strong correlation with the received pilots $\mathbf{y}_{\mathcal{P},q}$. Consider the set $\mathcal{B}$ be constructed as $\mathcal{B} = \{b_1, b_2, \cdots, b_{N_{\text{RF}}}\}$, where $b_1, b_2, \cdots, b_{N_{\text{RF}}}$ denote the positions corresponding to the N_{RF} dominant gains, and let the AOA set $\mathcal{K}$ and AOD set $\mathcal{L}$ corresponding to the aforementioned position set $\mathcal{B}$ be defined as

$$\mathcal{K} = \{k_1, k_2, \cdots, k_{N_{\text{RF}}}\}, \mathcal{L} = \{l_1, l_2, \cdots, l_{N_{\text{RF}}}\},$$

where the elements of the sets $\mathcal{K}$ and $\mathcal{L}$ represent the Doppler and delay indices corresponding to the N_{RF} dominant path-gain positions given in the set $\mathcal{B}$. The optimal RF RC $\mathbf{W}_{\text{RF,opt}}$ and TPC $\mathbf{F}_{\text{RF,opt}}$ are therefore given by

$$\mathbf{W}_{\text{RF,opt}} = \mathbf{A}_R(:, \mathcal{K}) \in \mathbb{C}^{N_r \times N_{\text{RF}}}, \tag{53}$$

$$\mathbf{F}_{\text{RF,opt}} = \mathbf{A}_T(:, \mathcal{L}) \in \mathbb{C}^{N_t \times N_{\text{RF}}}. \tag{54}$$

It is important to note that the proposed methodology is not optimal in terms of the sumrate, since it considers equal power allocation. The subsequent section characterizes the performance of the system relying on the proposed algorithm.

VI. SIMULATION RESULTS

This section presents an empirical evaluation of the multi-dimensional (4D) sparse CSI acquisition methodology within the context of HB-mmWave OTFS systems. In this framework, three distinct systems, termed System 1, System 2 and System 3 are considered. For System 1, the channel is constructed using a uniform power delay profile. By contrast, for System 2, the 3GPP-EVA (Extended Vehicular A) channel model [60] is employed, considering a maximum velocity of 500 Km/h. The delay profile for the EVA channel model is given in Table IV. The details of the parameters used for System 1, 2 and 3 are presented in Table III for reference.

TABLE III
SYSTEM 1,2,3 DESIGN SPECIFICATIONS

Parameters	System1	System 2	System 3
f_c in GHz	24	32	32
Δf in KHz	25	50	40
$M \times N$	32×16	32×32	64×32
Max. Doppler spread M_τ	8	10	12
Max. delay spread N_ν	8	10	12
# of RF chains N_{RF}	3	3	3
$N_r \times N_t$	4×4	4×4	4×4
$G_r \times G_t$	6×6	6×6	6×6
# of pilots N_p	128	256	256
# of frames Q	$1, 2$	$1, 2$	$1, 2$
# of clusters Cl	-	-	2
# of dominant scatterers P	5	9	10 per Cl
Modulation scheme	8-PSK	8-PSK	BPSK
Pulse-shape	Rectangular	Rectangular	Rectangular

The fractional Doppler indices considered in System 1 have a maximum Doppler spread of $N_\nu = 8$, where the integer part is generated for each path using a random integer between 0 and $N_\nu - 1$, while the fractional part having two decimal places is generated using a uniform distribution over the interval $[0, 0.5)$. The delay parameters considered are generated via the permutation of the P indices between 0 to $M_\tau - 1$. Note that these values change for every Monte Carlo iteration and these scenarios allow us to comprehensively investigate the performance of our proposed 4D sparse CE technique under various conditions and configurations. The metrics chosen for evaluating the performance of the two systems under investigation are the normalized mean square error (NMSE) and the symbol error rate (SER).

A. NMSE performance analysis

The NMSE versus SNR trend is a crucial metric characteristic that quantifies the accuracy of CE and signal recovery in our hybrid mmWave OTFS system. A lower NMSE indicates better estimation and signal recovery performance, reflecting

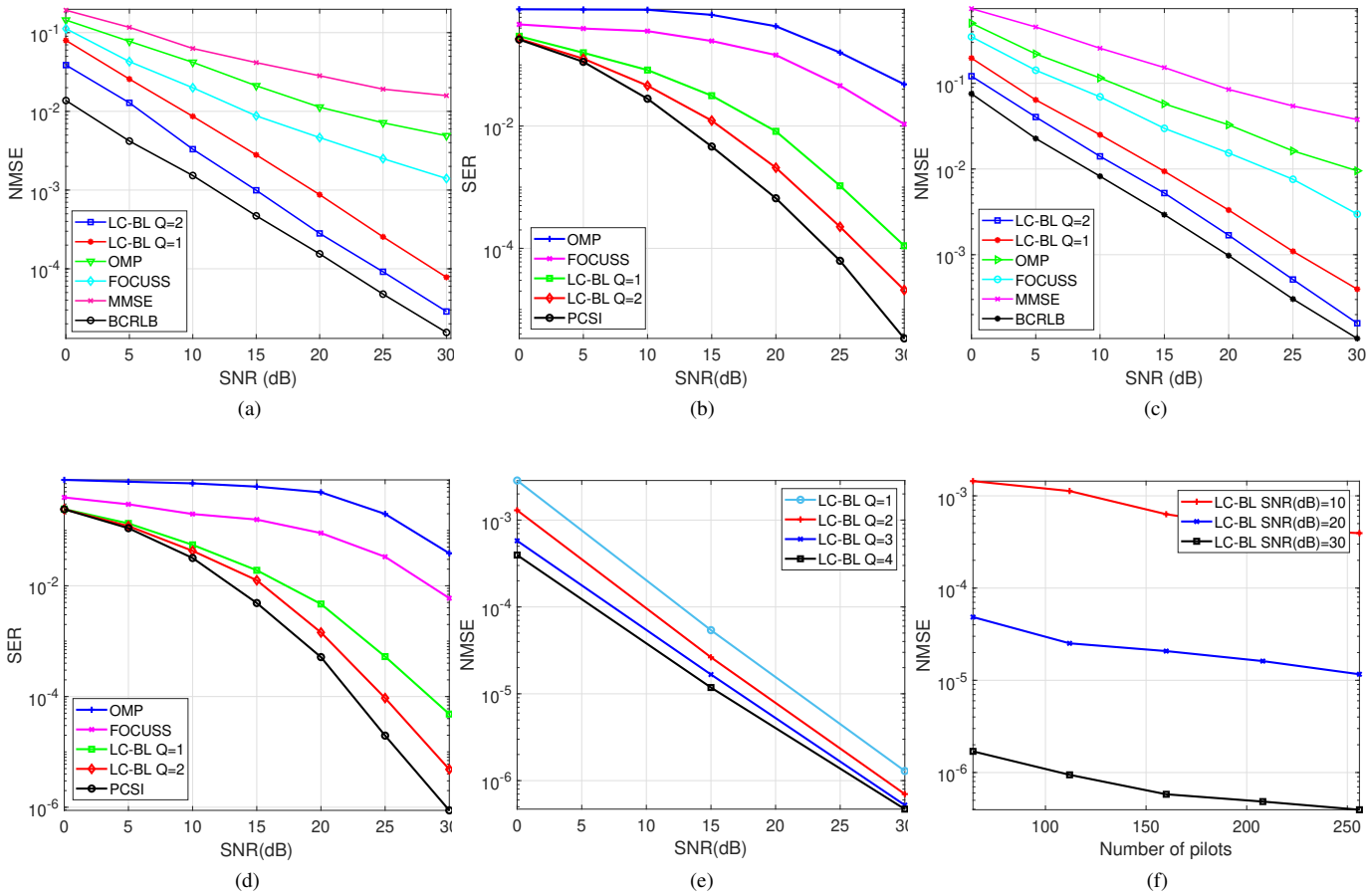


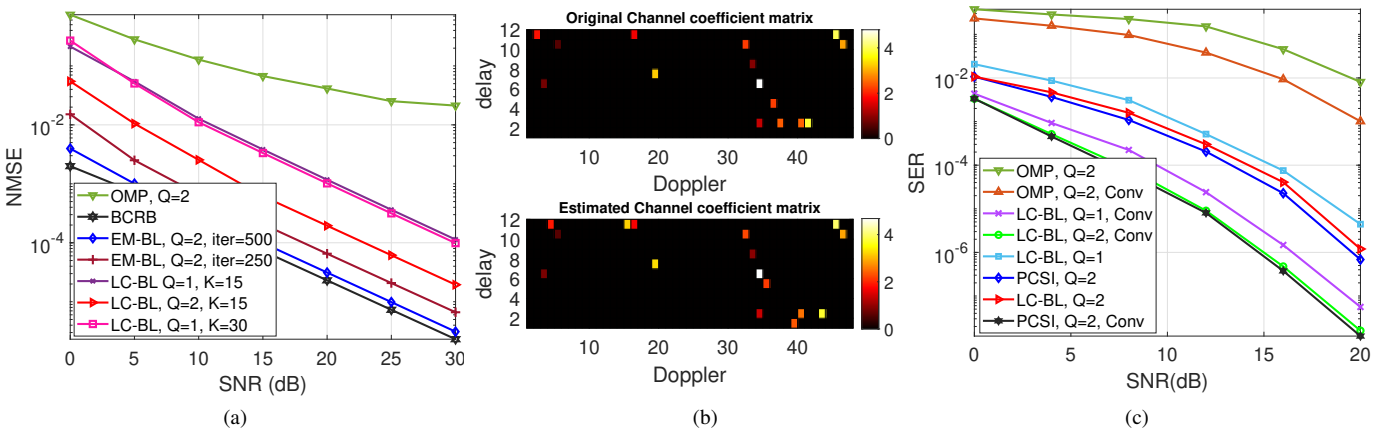
Fig. 2. Performance analysis of 4D-sparse HB-mmWave MIMO OTFS modulated System 1 and 2
(a) NMSE vs. SNR for System 1 with $M = 32$, $N = 16$, $N_p = 128$ and using 8-PSK (b) SER vs. SNR for System 1 using 8-PSK
(c) NMSE vs. SNR for System 2 with $M = 32$, $N = 32$, $N_p = 256$ and using 8-PSK (d) SER vs. SNR for System 2 using 8-PSK
(e) NMSE vs. SNR for $M = 64$, $N = 64$, $N_p = 64$ with different values of Q using BPSk
(f) NMSE vs. Number of pilots for $M = 64$, $N = 64$, $Q = 1$ using BPSK.

the ability to mitigate the effects of noise and channel impairments as the SNR improves. The NMSE is defined as

$$\text{NMSE} = \frac{||\widehat{\mathbf{H}}_{\text{DD}} - \mathbf{H}_{\text{DD}}||_F^2}{||\mathbf{H}_{\text{DD}}||_F^2}. \tag{55}$$

The NMSE performance of System 1 is depicted in Figure 2(a), while that of System 2 is shown in Figure 2(c). This study is focused on assessing the efficiency of the proposed LC-BL scheme through a comparison with established methods, including OMP [61], FOCUSS [62], and MMSE. In all of these methods, we set the number of frames to $Q = 2$. Additionally, we also showcase the performance of the LC-BL scheme when $Q = 1$.

The results presented in the figures demonstrate that the LC-BL estimation performance surpasses that of other estimation schemes, namely of OMP, FOCUSS, and MMSE, in our HB-mmWave MIMO OTFS system. Furthermore, it is worth noting that the performance of the LC-BL, when utilising a single frame ($Q = 1$), surpasses that of other schemes employing two frames ($Q = 2$). The performance of LC-BL for $Q = 2$ is superior to that of LC-BL with $Q = 1$ due to the larger size of the dictionary matrix, because the increased number of pilots results in a more accurate CE. The suboptimal performance of the OMP method can be primarily attributed to its inherent reliance on the stopping parameter, which is prone to vulnerabilities. Conversely, it is imperative to underscore that the efficacy of FOCUSS [62] is notably encumbered by issues related to convergence, thus impeding its ability to deliver consistent results due to its susceptibility to the choice of the regularization parameter. The conventional MMSE technique fails to exploit the inherent sparsity of the CSI in the DDA-domain. Hence it exhibits the poorest NMSE performance. Consequently, it is evident that non-Bayesian sparse estimation methodologies, exemplified by OMP and FOCUSS, exhibit an inferior performance compared to the BL approach. Again, this can be attributed to the above deficiencies. The LC-BL method shows a remarkably improved performance in comparison to the other sparse estimation schemes. Additionally, the performance of LC-BL is close to that of the Bayesian Cramer-Rao lower bound (BCRLB). Furthermore, it is worth noting that it does not require prior knowledge regarding



(a) (b) (c)

Fig. 3. Performance analysis of 4D-sparse HB-mmWave MIMO OTFS modulated System 3
(a) NMSE vs. SNR for System 3 with $M = 64$, $N = 32$, $N_p = 256$ and using BPSK
(b) Delay vs. Doppler 'Heatmap' representation for original channel coefficient matrix and estimated channel coefficient matrix with Pilot $SNR = 20dB$
(c) SER vs. SNR for System 3 using BPSK.

the sparse channel. Thus, the LC-BL exhibits convenient compatibility with practical OTFS systems, particularly in scenarios wherein no prior information is available.

Additionally, Fig. 2(e) shows the NMSE vs. SNR plot for different values of the parameter Q, considering the LC-BL scheme, where NMSE is computed using matrix $\mathbf{H}_{\mathcal{P}}$ in (30) The figures clearly demonstrate that as the value of Q increases, the NMSE decreases. This trend can be attributed to the transmission of a higher number of pilots, which directly stems from an augmentation in the number of frames. Furthermore, this observation supports our perception that increasing Q beyond the maximum number of frames $Q = 2$ yields negligible NMSE performance improvement. In Fig. 2(f), we delve into the impact of the number of pilots in each frame on the NMSE for different SNR values in the set $\{10, 20, 30\}$dB. Once again, we observe similar results, reinforcing the conclusion that increasing the number of pilots tends to decrease the NMSE. Moreover, it is essential to highlight that the trend observed in this context is consistent with the previous results. The following subsection illustrates the SER performance of the various schemes discussed before.

B. SER performance analysis for System 1 and 2

The SER performance of our technique conceived for HB-mmWave MIMO OTFS systems is characterized by Figures 2(b) and 2(d). The various simulation settings for System 1 and System 2 are consistent with those provided in Table III. The SER is evaluated for receivers using the CSI acquired from the aforementioned estimation schemes. We assess the SER performance by comparing it to an ideal receiver having perfect CSI. Notably, the BL-based schemes LC-BL of $Q = 1$ and $Q = 2$, outperform the non-BL schemes, such as OMP and FOCUSS. This can be attributed to the improved accuracy of CE, as supported by the NMSE plots shown in Figures 1(a) and 1(b). Furthermore, it is evident that the LC-BL scheme using $Q = 2$ consistently exhibits superior performance, with its SER closely approaching that of the hypothetical detector having perfect CSI.

C. NMSE and SER performance analysis for System 3

In System 3, a clustered channel model is under consideration having $f_c = 32GHz$, $\Delta f = 40KHz$, $M = 64$, $N = 32$, $N_p = 256$, $M_\tau = 12$, $N_\nu = 12$, $G_\nu = 48$ and assuming the presence of two scattering clusters within the angular range of $[-90°, 90°]$. Each cluster consists of 10 sub-paths, with an angular spread of $\pm 5°$. The delays and Doppler shifts are uniformly distributed.

Figure 3(a) shows the NMSE versus SNR performance for System 3, while Fig. 3(b) presents a 'heatmap' of the channel coefficient matrix, visually comparing the estimated CSI obtained using the LC-BL method with perfect CSI (PCSI) in the presence of fractional Doppler. This plot is intended to assess the accuracy of the system's delay and Doppler estimate. Fig. 3(c) illustrates the SER versus SNR performance with convolutional channel coding having Constraint Length $C = 7$, Generator Polynomials $G1 = 171$, $G2 = 133$. The results in Fig. 3(a) and 3(c) align with the performance of Systems 1 and 2; also, Fig 3(b) indicates that the proposed LC-BL scheme accurately estimates both the delay and Doppler coefficients reinforcing the efficiency of the proposed algorithm. The proposed algorithm is well-suited for practical scenarios, including the popular EVA and CDL-A/B models. While practical models like EVA and CDL-A/B are only approximately sparse, the proposed LC-BL continues to perform satisfactorily regardless of the specific channel model employed, with only a slight degradation in performance. Even when incorporating the cluster model, the LC-BL consistently delivers superior performance. This highlights the efficiency of our LC-BL-based channel estimation in the 4D-sparse delay-Doppler-angular (DDA) domain mmWave MIMO OTFS system under consideration.

TABLE IV
THE EVA CHANNEL MODELS DELAY PROFILE

Excess tap delay (ns)	0	30	150	310	370	710	1090	1730	2510
Relative power (dB)	0.0	-1.5	-1.4	-3.6	-0.6	-9.1	-7.0	-12.0	-16.9

VII. SUMMARY AND CONCLUSION

We established the input-output relationship for mmWave-HB MIMO systems based on OTFS modulation within the DD-domain. The CE problem was reformulated as a sparse estimation challenge, accommodating the inherent 4D-sparse structure. In this approach, a channel model was conceived in the Delay-Doppler-Angular (DDA) domain, based on characterizing each multipath component by its delay, Doppler, angle of arrival (AoA), and angle of departure (AoD) components. To address CE we developed a low-complexity BL CSI estimation scheme, termed LC-BL, for a mmWave-HB MIMO OTFS system. Furthermore, a detailed complexity analysis of the proposed LC-BL algorithm was presented, and compared to the conventional OMP and EM-based BL algorithms. The results demonstrated that even though the complexity of the method conceived is low and comparable to that of the OMP algorithm, it exhibited a superior performance, similar to that of the BL technique. Moreover, a single-step method has been devised to design RF TPC/RC waveforms, where the weights of the RF TPC/RC pair are optimized for maximizing the directional gains. Finally, the judicious balance between the complexity and performance of our proposed LC-BL scheme makes it feasible for practical implementation.

REFERENCES

[1] C.-X. Wang, X. You, X. Gao, X. Zhu, Z. Li, C. Zhang, H. Wang, Y. Huang, Y. Chen, H. Haas, J. S. Thompson, E. G. Larsson, M. D. Renzo, W. Tong, P. Zhu, X. Shen, H. V. Poor, and L. Hanzo, "On the road to 6G: Visions, requirements, key technologies, and testbeds," *IEEE Communications Surveys & Tutorials*, vol. 25, no. 2, pp. 905–974, 2023.

[2] L. Hanzo, Y. Akhtman, J. Akhtman, L. Wang, and M. Jiang, *MIMO-OFDM for LTE, WiFi and WiMAX: Coherent versus non-coherent and cooperative turbo transceivers*. John Wiley & Sons, 2011.

[3] F. Hasegawa, A. Taira, G. Noh, B. Hui, H. Nishimoto, A. Okazaki, A. Okamura, J. Lee, and I. Kim, "High-speed train communications standardization in 3GPP 5G NR," *IEEE Communications Standards Magazine*, vol. 2, no. 1, pp. 44–52, 2018.

[4] Y. Liu, C.-X. Wang, and J. Huang, "Recent developments and future challenges in channel measurements and models for 5G and beyond high-speed train communication systems," *IEEE Communications Magazine*, vol. 57, no. 9, pp. 50–56, 2019.

[5] R. Hadani, S. Rakib, M. Tsatsanis, A. Monk, A. J. Goldsmith, A. F. Molisch, and R. Calderbank, "Orthogonal time frequency space modulation," in *2017 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE, 2017, pp. 1–6.

[6] R. Hadani, S. Rakib, S. Kons, M. Tsatsanis, A. Monk, C. Ibars, J. Delfeld, Y. Hebron, A. J. Goldsmith, A. F. Molisch *et al.*, "Orthogonal time frequency space modulation," *ArXiv preprint, ArXiv:1808.00519*, 2018.

[7] R. Hadani and A. Monk, "OTFS: A new generation of modulation addressing the challenges of 5G," *ArXiv preprint, ArXiv:1802.02623*, 2018.

[8] A. Mehrotra, S. Srivastava, S. Asifa, A. K. Jagannatham, and L. Hanzo, "Online bayesian learning aided sparse CSI estimation in OTFS modulated MIMO systems for ultra-high-doppler scenarios," *IEEE Transactions on Communications*, pp. 1–1, 2023.

[9] C. Xu, L. Xiang, S. Sugiura, R. G. Maunder, L.-L. Yang, D. Niyato, G. Y. Li, R. Schober, and L. Hanzo, "Noncoherent orthogonal time frequency space modulation," *IEEE Transactions on Wireless Communications*, pp. 1–1, 2024.

[10] R. W. Heath, N. Gonzalez-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 436–453, 2016.

[11] J. Lorca, M. Hunukumbure, and Y. Wang, "On overcoming the impact of doppler spectrum in millimeter-wave V2I communications," in *2017 IEEE Globecom Workshops (GC Wkshps)*, 2017, pp. 1–6.

[12] T. S. Rappaport, R. W. Heath Jr, R. C. Daniels, and J. N. Murdock, *Millimeter wave wireless communications*. Pearson Education, 2015.

[13] I. A. Hemadeh, K. Satyanarayana, M. El-Hajjar, and L. Hanzo, "Millimeter-wave communications: Physical channel models, design considerations, antenna constructions, and link-budget," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 2, pp. 870–913, 2017.

[14] D. E. Berraki, S. M. D. Armour, and A. R. Nix, "Application of compressive sensing in sparse spatial channel recovery for beamforming in mmwave outdoor systems," in *2014 IEEE Wireless Communications and Networking Conference (WCNC)*, 2014, pp. 887–892.

[15] X. Ma, F. Yang, S. Liu, J. Song, and Z. Han, "Design and optimization on training sequence for mmwave communications: A new approach for sparse channel estimation in massive MIMO," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 7, pp. 1486–1497, 2017.

[16] W. Ma, C. Qi, Z. Zhang, and J. Cheng, "Sparse channel estimation and hybrid precoding using deep learning for millimeter wave massive MIMO," *IEEE Transactions on Communications*, vol. 68, no. 5, pp. 2838–2849, 2020.

[17] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 831–846, 2014.

[18] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 13, no. 3, pp. 1499–1513, 2014.

[19] C. Huang, L. Liu, C. Yuen, and S. Sun, "Iterative channel estimation using LSE and sparse message passing for mmwave MIMO systems," *IEEE Transactions on Signal Processing*, vol. 67, no. 1, pp. 245–259, 2019.

[20] J. Lee, G.-T. Gil, and Y. H. Lee, "Channel estimation via orthogonal matching pursuit for hybrid MIMO systems in millimeter wave communications," *IEEE Transactions on Communications*, vol. 64, no. 6, pp. 2370–2386, 2016.

[21] 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 Transactions on Signal Processing*, vol. 67, no. 5, pp. 1251–1266, 2018.

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

[23] T. Kim and D. J. Love, "Virtual AoA and AoD estimation for sparse millimeter wave MIMO channels," in *2015 IEEE 16th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2015, pp. 146–150.

[24] X. Lin, S. Wu, C. Jiang, L. Kuang, J. Yan, and L. Hanzo, "Estimation of broadband multiuser millimeter wave massive MIMO-OFDM channels by exploiting their sparse structure," *IEEE Transactions on Wireless Communications*, vol. 17, no. 6, pp. 3959–3973, 2018.

[25] K. Venugopal, A. Alkhateeb, N. G. Prelcic, and R. W. Heath, "Channel estimation for hybrid architecture-based wideband millimeter wave systems," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 9, pp. 1996–2009, 2017.

[26] J. Rodríguez-Fernández, N. González-Prelcic, K. Venugopal, and R. W. Heath, "Frequency-domain compressive channel estimation for frequency-selective hybrid millimeter wave MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 17, no. 5, pp. 2946–2960, 2018.

[27] K. Liu, X. Li, J. Fang, and H. Li, "Bayesian mmwave channel estimation via exploiting joint sparse and low-rank structures," *IEEE Access*, vol. 7, pp. 48 961–48 970, 2019.

[28] Y. Wang, C. Qi, P. Li, Z. Lu, and P. Lu, "Channel estimation for wideband mmwave MIMO OFDM system exploiting block sparsity," *IEEE Communications Letters*, pp. 1–1, 2022.

[29] F. Talaei and X. Dong, "Hybrid mmwave MIMO-OFDM channel estimation based on the multi-band sparse structure of channel," *IEEE Transactions on Communications*, vol. 67, no. 2, pp. 1018–1030, 2019.

[30] S. Gao, X. Cheng, and L. Yang, "Estimating doubly-selective channels for hybrid mmwave massive MIMO systems: A doubly-sparse approach," *IEEE Transactions on Wireless Communications*, vol. 19, no. 9, pp. 5703–5715, 2020.

[31] S. Srivastava, C. S. K. Patro, A. K. Jagannatham, and L. Hanzo, "Sparse, group-sparse and online Bayesian learning aided channel estimation for doubly-selective mmwave hybrid MIMO OFDM systems," *IEEE Transactions on Communications*, 2021.

[32] S. Srivastava, P. Singh, A. K. Jagannatham, A. Karandikar, and L. Hanzo, "Bayesian learning-based doubly-selective sparse channel estimation for millimeter wave hybrid MIMO-FBMC-OQAM systems," *IEEE Transactions on Communications*, vol. 69, no. 1, pp. 529–543, 2020.

[33] S. Srivastava, J. Nath, and A. K. Jagannatham, "Data aided quasistatic and doubly-selective CSI estimation using affine-precoded superimposed pilots in millimeter wave MIMO-OFDM systems," *IEEE Transactions on Vehicular Technology*, 2021.

[34] M. K. Ramachandran and A. Chockalingam, "MIMO-OTFS in high-Doppler fading channels: Signal detection and channel estimation," in *2018 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2018, pp. 206–212.

[35] W. Shen, L. Dai, J. An, P. Fan, and R. W. Heath, "Channel estimation for orthogonal time frequency space (OTFS) massive MIMO," *IEEE Transactions on Signal Processing*, vol. 67, no. 16, pp. 4204–4217, 2019.

[36] P. Raviteja, K. T. Phan, and Y. Hong, "Embedded pilot-aided channel estimation for OTFS in delay-Doppler channels," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 5, pp. 4906–4917, 2019.

[37] L. Zhao, W. Gao, and W. Guo, "Sparse Bayesian learning of delay-Doppler channel for OTFS system," *IEEE Communications Letters*, 2020.

[38] S. Srivastava, R. K. Singh, A. K. Jagannatham, and L. Hanzo, "Bayesian learning aided sparse channel estimation for orthogonal time frequency space modulated systems," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 8, pp. 8343–8348, 2021.

[39] Z. Wei, W. Yuan, S. Li, J. Yuan, and D. W. K. Ng, "Off-grid channel estimation with sparse Bayesian learning for OTFS systems," *IEEE Transactions on Wireless Communications*, vol. 21, no. 9, pp. 7407–7426, 2022.

[40] Y. Yan, C. Shan, J. Zhang, and H. Zhao, "Off-grid channel estimation for OTFS-based mmwave hybrid beamforming systems," *IEEE Communications Letters*, vol. 27, no. 8, pp. 2167–2171, 2023.

[41] S. Srivastava, R. K. Singh, A. K. Jagannatham, and L. Hanzo, "Delay-doppler and angular domain 4D-sparse CSI estimation in OTFS aided MIMO systems," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 12, pp. 13 447–13 452, 2022.

[42] S. Srivastava, R. K. Singh, A. K. Jagannatham, A. Chockalingam, and L. Hanzo, "OTFS transceiver design and sparse doubly-selective CSI estimation in analog and hybrid beamforming aided mmwave MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 21, no. 12, pp. 10 902–10 917, 2022.

[43] C. Heil, *A discrete ZAK transform*. MITRE Corporation, Washington C31 Division, 1989.

[44] M. Ramachandran, G. Surabhi, and A. Chockalingam, "OTFS: A new modulation scheme for high-mobility use cases," *Journal of the Indian Institute of Science*, pp. 1–22, 2020.

[45] K. Murali and A. Chockalingam, "On OTFS modulation for high-Doppler fading channels," in *2018 Information Theory and Applications Workshop (ITA)*. IEEE, 2018, pp. 1–10.

[46] A. Mehrotra, R. K. Singh, S. Srivastava, and A. K. Jagannatham, "Channel estimation techniques for CP-aided OTFS systems relying on practical pulse shapes," in *2022 IEEE International Conference on Signal Processing and Communications (SPCOM)*, 2022, pp. 1–5.

[47] P. Raviteja, K. T. Phan, Y. Hong, and E. Viterbo, "Interference cancellation and iterative detection for orthogonal time frequency space modulation," *IEEE Transactions on Wireless Communications*, vol. 17, no. 10, pp. 6501–6515, 2018.

[48] O. K. Rasheed, G. Surabhi, and A. Chockalingam, "Sparse delay-Doppler channel estimation in rapidly time-varying channels for multiuser OTFS on the uplink," in *2020 IEEE 91st Vehicular Technology Conference (VTC-Spring)*. IEEE, 2020, pp. 1–5.

[49] S. Srivastava, R. K. Singh, A. K. Jagannatham, and L. Hanzo, "Bayesian learning aided simultaneous row and group sparse channel estimation in orthogonal time frequency space modulated MIMO systems," *IEEE Transactions on Communications*, 2021.

[50] G. Surabhi, M. K. Ramachandran, and A. Chockalingam, "OTFS modulation with phase noise in mmWave communications," in *2019 IEEE 89th Vehicular Technology Conference (VTC2019-Spring)*. IEEE, 2019, pp. 1–5.

[51] M. E. Tipping and A. C. Faul, "Fast marginal likelihood maximisation for sparse bayesian models," in *Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics*, vol. R4, Jan 2003, pp. 276–283.

[52] S. Cotter, B. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Transactions on Signal Processing*, vol. 53, no. 7, pp. 2477–2488, 2005.

[53] M. Al-Shoukairi, P. Schniter, and B. D. Rao, "A GAMP-based low complexity sparse bayesian learning algorithm," *IEEE Transactions on Signal Processing*, vol. 66, no. 2, pp. 294–308, 2018.

[54] R. R. Pote and B. D. Rao, "Light-weight sequential SBL algorithm: An alternative to OMP," in *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023, pp. 1–5.

[55] P. Raviteja, Y. Hong, E. Viterbo, and E. Biglieri, "Practical pulse-shaping waveforms for reduced-cyclic-prefix OTFS," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 1, pp. 957–961, 2018.

[56] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.

[57] A. Mehrotra, S. Srivastava, S. Reddy, A. K. Jagannatham, and L. Hanzo, "Sparse channel estimation for MIMO OTFS/OTSM systems using finite-resolution ADCs," *IEEE Transactions on Communications*, pp. 1–1, 2024.

[58] A. Mehrotra, S. Srivastava, A. K. Jagannatham, and L. Hanzo, "Data-aided CSI estimation using affine-precoded superimposed pilots in orthogonal time frequency space modulated MIMO systems," *IEEE Transactions on Communications*, vol. 71, no. 8, pp. 4482–4498, 2023.

[59] I. Y. Chun and B. Adcock, "Optimal sparse recovery for multi-sensor measurements," in *2016 IEEE Information Theory Workshop (ITW)*. IEEE, Sep. 2016, p. 270–274. [Online]. Available: http://dx.doi.org/10.1109/ITW.2016.7606838

[60] 3rd Generation Partnership Project; Technical Specification Group Radio Access Network, "Evolved Universal Terrestrial Radio Access (E-UTRA); User Equipment (UE) Radio Transmission and Reception," 3GPP, Tech. Rep., 2011. [Online]. Available: https://www.3gpp.org

[61] T. T. Cai and L. Wang, "Orthogonal matching pursuit for sparse signal recovery with noise," *IEEE Transactions on Information Theory*, vol. 57, no. 7, pp. 4680–4688, 2011.

[62] I. Gorodnitsky and B. Rao, "Sparse signal reconstruction from limited data using FOCUSS: a re-weighted minimum norm algorithm," *IEEE Transactions on Signal Processing*, vol. 45, no. 3, pp. 600–616, 1997.