Efficient Channel Tracking Based on Compressive Sensing for OFDM Millimeter-Wave Systems

Sota Uchimura, *Graduate Student Member, IEEE*, Koji Ishibashi, *Senior Member, IEEE*, Hiroki Iimori, *Member, IEEE*, Paulo Valente Klaine, *Member, IEEE*, and Szabolcs Malomsoky

***Abstract*—In this paper, we propose a new channel tracking method to track the time fluctuations of millimeter-wave (mmWave) channels over multiple subcarriers with few training overheads and low complexity. In essence, channel tracking is formulated as a compressive sensing problem with the proposed sensing matrix over the delay domain, which enables low complexity irrespective of the number of available subcarriers, resulting in a new tracking algorithm based on the exact top-k feature selection. To elaborate, the proposed method is based on a novel sensing matrix that exploits the time evolution models of angle of arrivals (AoAs) and angle of departures (AoDs), where the time-varying channel is effectively approximated with a few discrete AoA and AoD candidates. Moreover, the proposed tracking algorithm is refined by introducing multiple sub-sensing subsets, which further improve the tracking performance. Numerical results and complexity analyses in terms of floating operations (FLOPs) confirm that the proposed approaches achieve better trade-offs between complexity and tracking gain than the conventional Bayesian estimator approaches.**

***Index Terms*—Channel tracking, millimeter-wave (mmWave), compressive sensing.**

I. Introduction

WIRELESS cellular communication in higher frequency bands has been investigated over the last years [1]–[3], and millimeter-wave (mmWave) base station products have been introduced with the fifth-generation of mobile communication systems (5G). In particular, mmWave communication using frequency bands ranging from approximately 24 [GHz] to 300 [GHz] has recently gained significant attention as a potential solution for overcoming the spectrum shortage problem in conventional microwave communications [4]–[7]. However, despite the large available bandwidth, mmWave systems suffer from high propagation losses owing to the increased signal attenuation at higher carrier frequencies [8]–[11]. In spite of these drawbacks, shorter wavelengths enable larger antenna arrays to be equipped with more elements. Therefore, multiple-input multiple-output (MIMO) systems and beamforming utilizing multiple antenna elements to concentrate the signal power to a specific direction, are fundamental technologies for beyond 5G and the sixth-generation of mobile communication systems (6G) [12]–[15].

One of the benefits of having multiple antenna elements is that a larger antenna array improves beamforming gain by focusing beams in narrower directions. However, to maximize the beamforming gain, an accurate channel state information (CSI) must be obtained, such that it enables an effective combination of the signals from the antenna elements [16], [17]. In addition to depending on accurate CSI, highly directional beams can also be subject to impairments such as misalignment and fluctuations of channels in short duration, which result in quality of service (QoS) degradation [18]–[20].

As such, one potential solution to overcome beam misalignment is to perform a continuous beam reshaping or redesigning based on a frequently updated and accurate CSI. Accordingly, several channel estimation techniques for mmWave channels have been discussed [21]–[26]. These approaches can estimate the CSI with low complexity even for a large antenna array by exploiting the sparsity in the angle domain, which arises from the mmWave channel comprising of a few dominant paths. Although these approaches maintain low complexity, to obtain accurate estimations, a high number of training overhead is required, which increases with respect to the number of antenna elements. If CSI updates by such estimation approaches are carried out frequently, the spectral efficiency can be significantly decreased in especial for systems with a significant number of antenna elements.

To maintain an accurate CSI with lower training overhead, beam and channel tracking techniques have been investigated [27]–[41]. These techniques update the CSI by tracking a few dominant paths in mmWave channels with the temporal correlation of the channel parameters, which by its turn enables accurate CSI updates with lower training overhead than channel estimations from scratch. To name but a few, the channel-tracking algorithm proposed in [27] tracks angle of arrival (AoA) and angle of departure (AoD) via Kalman filter (KF) under the assumptions of Gaussian-distributed fluctuations for angles while assuming constant path gains over time. Considering practical mmWave channels, other algorithms that simultaneously track the time fluctuation of angles and propagation path gains via extended KF have also been proposed [28]–[30]. In addition, another low-complexity beam tracking relevant in practical communications requiring frequent CSI updates, has been proposed in [33]. Moreover, a tracking algorithm for extremely large-scale MIMO [32] and for reconfigurable intelligent surfaces [31] have been discussed. However, despite maintaining an accurate CSI with

Copyright (c) 2015 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.
(*Corresponding author: Sota Uchimura*)

S. Uchimura and K. Ishibashi are with the Advanced Wireless and Communication Research Center (AWCC), The University of Electro-Communications, Tokyo 182-8285, Japan (e-mail: uchimura@awcc.uec.ac.jp, koji@ieee.org)

H. Iimori, P. V. Klaine, and S. Malomsoky are with Ericsson Research, Ericsson Japan K.K., Yokohama 220-0012, Japan, (e-mail: {hiroki.iimori, paulo.valente.klaine, szabolcs.malomsoky}@ericsson.com)

lower training overhead and complexity in some cases, these studies [27]–[33] assume narrow band channel models, which is not an appropriate assumption for mmWave systems with large bandwidths. As such, the tracking performance of these algorithms would be limited in practical situations due to the prohibitive complexity of tracking channels over all available subcarriers.

Channel and beam tracking methods for MIMO-orthogonal frequency division multiplexing (OFDM) systems over frequency-selective mmWave channels are proposed in [34]–[39]. In particular, an algorithm based on marginalized particle filter and further extended to a generalized marginalized particle filter (GMPF) to obtain the optimal minimum mean square error (MMSE) estimator of the CSI were proposed in [35]–[37]. Lastly, in [38] the authors also extend GMPF methods to mmWave systems that can detect blockages of beams while tracking channel parameters, since mmWave signals are more susceptible to blockage caused by physical obstacles such as buildings, trees, or human bodies. These studies [35]–[38] show that channel tracking based on GMPF can obtain significant tracking gains in terms of spectral efficiency. However, despite their good performance, these algorithms require tremendous computational complexity owing to the AoAs and AoDs estimations via particle filters based on Monte Carlo sampling. In addition, such frequency-domain approaches, which estimate the path gain for all available subcarriers via KF [35]–[39], increase the complexity in cubic order with respect to the number of subcarriers. In practical situations, this computational cost hinders real-time processing and results in beam misalignment, particularly in high-mobility scenarios [40]–[43].

Therefore, we proposed a new low-complexity channel tracking framework based on compressive sensing for OFDM mmWave systems. First, the proposed approaches introduce a novel sensing matrix design that exploits the temporal evolution models of AoAs and AoDs, resulting in a sensing matrix containing only discrete angle candidates with grids designed from knowledge of these models as well as the CSI at the prior time slots. This sensing matrix converts the aforementioned estimation problem with multiple unknowns, such as path gain, AoA, and AoD, into a reconstruction problem with one path gain matrix with a unique sparse structure different from other state-of-the-art (SotA) schemes [21]–[25], [39]. Owing to the unique sparsity and limited angle candidates, the proposed formulation enables an accurate estimation of compressive sensing algorithms, while significantly reducing the complexity. Moreover, a path gain reconstruction problem in the delay domain is formulated to reduce the number of pilot symbols and achieve further complexity reductions.

To solve this reconstruction problem, an algorithm based on the exact top-k feature selection [44] is introduced. This approach realizes channel tracking with linear-order complexity with respect to the number of subcarriers that pilot symbols are allocated to, and its complexity does not depend on the number of available subcarriers. Furthermore, another channel tracking method, which leverages multiple sub-sensing matrices processing, is also proposed to improve the reconstruction accuracy of compressive sensing algorithms. This refinement yields higher tracking gain with more sub-sensing matrices and leads to accurate CSI updates even under unfavorable conditions, such as low signal-to-noise ratio (SNR) and scarce pilot symbols. Finally, we evaluate the complexity of the algorithms based on floating operations (FLOPs) and the tracking gain in terms of spectral efficiency through computer simulations. Evaluations of computational complexity and numerical results confirm that the proposed approaches exhibit better trade-off properties between the tracking gain and complexity than the conventional approach based on GMPF [37].

TABLE I: Key Quantities, Parameters, and Symbols

Notation	Meaning
$\mathbf{g}[k]$	Path gain at the k-th subcarrier for all paths
$\mathbf{g}$	Path gain for all subcarriers and all paths
$\mathbf{G}_{\mathrm{D}}$ / $\mathbf{g}_{\mathrm{D}}$	Path gain for all delay taps and all paths
$\phi_{c,r}$	AoA at the r-th ray within the c-th cluster
$\theta_{c,r}$	AoD at the r-th ray within the c-th cluster
$\boldsymbol{\phi}$	Set of AoAs for all paths
$\boldsymbol{\theta}$	Set of AoDs for all paths
$\boldsymbol{\xi}$	Union of AoA and AoD sets
$\mathbf{B}$	Sensing matrix
$x^{(n)}$	Parameter x at the n-th channel slot
$\hat{x}$	Estimated value of x
$\tilde{x}$	Candidate or approximated value of x
$\tilde{\phi}_{i,j}$ / $\tilde{\phi}_i\{j\}$	j-th candidate of the AoA at the i-th path
$\tilde{\theta}_{i,j}$ / $\tilde{\theta}_i\{j\}$	j-th candidate of the AoD at the i-th path

The following notations are used in this paper. Matrices and vectors are denoted as $\mathbf{X}$ and $\mathbf{x}$, respectively. Integers, real numbers, and complex numbers are represented using $\mathbb{N}$, $\mathbb{R}$, and $\mathbb{C}$, respectively, and the operators $(\cdot)^{\mathrm{T}}$, $(\cdot)^*$, and $(\cdot)^{\mathrm{H}}$ denote the transposition, conjugate, and complex conjugate transposition, respectively. A diagonal matrix obtained from vector $\mathbf{x}$ and a block diagonal matrix obtained from the matrices are denoted by $\mathrm{diag}(\mathbf{x})$ and $\mathrm{blkdiag}(\cdots)$, respectively. The operator $[\mathbf{X}]_i$ denotes the i-th row in the matrix $\mathbf{X}$. The N-dimensional identity matrix is denoted by $\mathbf{I}_N$. Operators $\otimes$ and $\circ$ denote Kronecker and Khatri-Rao products, respectively. Functions $\mathrm{vec}(\cdot)$, $\|\cdot\|_p$, and $\|\cdot\|_{\mathrm{F}}$ denote vectorization, the l_p norm, and Frobenius norm, respectively. The $l_{2,0}$ norm of a vector $\mathbf{x}$ calculates $\|\|\mathbf{x}\|_2\|_0$. The floor function is denoted as $\lfloor\cdot\rfloor$. The circularly symmetric complex normal distribution, normal distribution, and Laplace distribution with mean μ and variance σ^2 are denoted by $\mathcal{CN}\left(\mu,\sigma^2\right)$, $\mathcal{N}\left(\mu,\sigma^2\right)$, and $\mathcal{L}(\mu,\sigma^2)$, respectively. The major parameters are listed in Table I

II. System model

In this section, the considered channel and received signal models are introduced.

A. Communication Scenario

We consider an uplink MIMO-OFDM system in which one base station (BS) equipped with a uniform linear array (ULA)

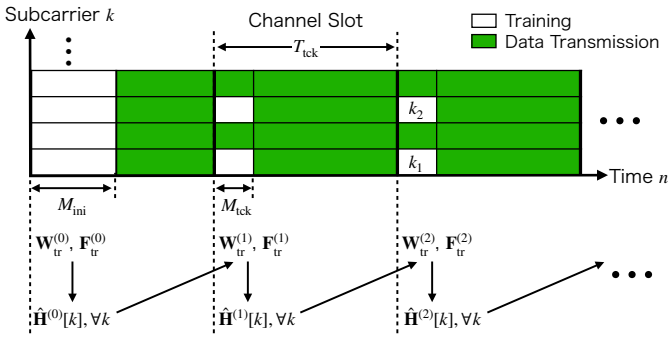


Fig. 1: Illustration of the transmit frame structure and communication scenario of an OFDM mmWave system

of $N_{\rm r}$ antenna elements and $N_{\rm r,RF}$ radio frequency (RF) chains receives time-frequency signals from one user equipment (UE) equipped with a ULA of $N_{\rm t}$ antenna elements and $N_{\rm t,RF}$ RF chains[1]. It is assumed that both ULAs at the receiver and transmitter have fully connected structures. The UE moves at a constant speed v [m/s] and its trajectory can be seen as linear within a sufficiently short period of time. The channel slot is defined as a length $T_{\rm tck} \in \mathbb{R}^+$, where the channel parameters are assumed to be constant within a timeslot and change in a timeslot-by-timeslot manner. Each channel slot is separated into two phases: 1) the CSI acquisition phase and 2) the data transmission phase, which are respectively colored in white and green as illustrated in Fig 1. In particular, in the training phase at the initial channel slot, the CSI is estimated from scratch with $M_{\rm ini} \in \mathbb{N}^+$ OFDM pilot symbols. In the subsequent channel slots, $M_{\rm tck}$ ($\ll M_{\rm ini}$) pilot symbols are allocated to the selected time-frequency resources in the training phase, where the CSI is updated to capture the time-varying nature of the channels.

B. Channel Model

Let $n \in \{0, 1, \dots, \}$ and $d \in \{0, \dots, D-1\}$ denote the channel slot and delay tap indices, respectively, where $D \in \mathbb{N}^+$ denotes the total number of delay taps. The mmWave channel at the n-th channel slot and the d-th delay tap is modeled as

$$\mathbf{H}^{(n)}[d] = \sqrt{\frac{1}{\mu}} \sum_{c=1}^{C} \sum_{r=1}^{R_c} \alpha_{c,r}^{(n)} p(dT_{\rm s} - \tau_c^{(n)}) \mathbf{a}_{N_{\rm r}}(\phi_{c,r}^{(n)}) \mathbf{a}_{N_{\rm t}}^{\rm H}(\theta_{c,r}^{(n)}), \tag{1}$$

where $C \in \mathbb{N}^+$, $R_c \in \mathbb{N}^+$, and $\mu \in \mathbb{N}^+$ denote the number of clusters, the number of rays within the c-th cluster, and the total number of paths $\mu = \sum_{c=1}^{C} R_c$, respectively.

The complex path gain at the r-th ray of the c-th cluster is denoted by $\alpha_{c,r}^{(n)} \sim \mathcal{CN}\left(0, 10^{-{\rm PL}_c/10}\right)$, where ${\rm PL}_c$ denotes the path loss component at the c-th cluster, given as [9]

$${\rm PL}_c = 20 \log_{10}\left(\frac{4\pi}{\lambda}\right) + 10\alpha \log_{10}(\gamma) + \chi_{\delta^2}, \tag{2}$$

where $\lambda \in \mathbb{R}^+$ and $\gamma \in \mathbb{R}^+$ denote the carrier wavelength and the distance between BS and UE, respectively, and $\alpha \in \mathbb{R}^+$ and $\chi_{\delta^2} \sim \mathcal{N}(0, \delta^2)$ are determined by the urban microcell street canyon closed-in model proposed in [9, Table I].

In (1), the function $p(dT_{\rm s} - \tau_c^{(n)})$ denotes the equivalent pulse shaping response at the transmitter and receiver, and is calculated from the time delay $\tau_c^{(n)} \in \mathbb{R}$ at the c-th cluster and sampling duration $T_{\rm s} \in \mathbb{R}$. Without loss of generality, $c = 1$ represents the line-of-sight (LOS) component with $R_1 = 1$ and $\tau_1^{(0)} = 0$. The vector $\mathbf{a}_{N_{\rm r}}(\phi_{c,r}^{(n)}) \in \mathbb{C}^{N_{\rm r} \times 1}$ and $\mathbf{a}_{N_{\rm t}}(\theta_{c,r}^{(n)}) \in \mathbb{C}^{N_{\rm t} \times 1}$ denote the ULA response of the AoA $\phi_{c,r}^{(n)} \in \mathbb{R}$ and AoD $\theta_{c,r}^{(n)} \in \mathbb{R}$ at the r-th ray within the c-th cluster, respectively. The ULA response with N antenna elements and half-wavelength spacing is defined as [22], [23]

$$\mathbf{a}_N(x) \triangleq \frac{1}{\sqrt{N}} \left[1, e^{j\pi \cos x}, \cdots, e^{j\pi(N-1)\cos x}\right]^{\rm T} \in \mathbb{C}^{N \times 1}. \tag{3}$$

Let $k \in \mathcal{K} \triangleq \{0, 1, \dots, K-1\}$ denote the subcarrier index where $K \in \mathbb{N}^+$ denotes the number of available subcarriers. From the Fourier transform, the mmWave channel at the n-th channel slot and the k-th subcarrier can be written as

$$\begin{aligned} \mathbf{H}^{(n)}[k] &= \sum_{d=0}^{D-1} \mathbf{H}^{(n)}[d] e^{-j\frac{2\pi k d}{K}} \\ &= \mathbf{A}_{\rm R}\left(\boldsymbol{\phi}^{(n)}\right) {\rm diag}\left\{\mathbf{g}^{(n)}[k]\right\} \mathbf{A}_{\rm T}^{\rm H}\left(\boldsymbol{\theta}^{(n)}\right), \end{aligned} \tag{4}$$

where $\boldsymbol{\phi}^{(n)} \triangleq \left\{\phi_{1,1}^{(n)}, \dots, \phi_{C,R_c}^{(n)}\right\}$ and $\boldsymbol{\theta}^{(n)} \triangleq \left\{\theta_{1,1}^{(n)}, \dots, \theta_{C,R_c}^{(n)}\right\}$ denote the set of AoAs and AoDs at the n-th channel slot, respectively[2], and the array response matrices at the receiver $\mathbf{A}_{\rm R}(\cdot)$ and transmitter $\mathbf{A}_{\rm T}(\cdot)$ are defined as

$$\mathbf{A}_{\rm R}\left(\boldsymbol{\phi}^{(n)}\right) \triangleq \left[\mathbf{a}_{N_{\rm r}}\left(\phi_{1,1}^{(n)}\right), \dots, \mathbf{a}_{N_{\rm r}}\left(\phi_{C,R_c}^{(n)}\right)\right] \in \mathbb{C}^{N_{\rm r} \times \mu}, \tag{5}$$

$$\mathbf{A}_{\rm T}\left(\boldsymbol{\theta}^{(n)}\right) \triangleq \left[\mathbf{a}_{N_{\rm t}}\left(\phi_{1,1}^{(n)}\right), \dots, \mathbf{a}_{N_{\rm t}}\left(\phi_{C,R_c}^{(n)}\right)\right] \in \mathbb{C}^{N_{\rm t} \times \mu}, \tag{6}$$

where $\mathbf{g}^{(n)}[k] \in \mathbb{C}^{\mu \times 1}$ collects the path gain at the k-th subcarrier for all propagation paths.

Let $g_{c,r}^{(n)}[d] \triangleq \sqrt{\frac{1}{\mu}} \alpha_{c,r}^{(n)} p(dT_{\rm s} - \tau_c^{(n)})$ define the path gain at the n-th channel slot, d-th delay tap, and r-th ray within the c-th cluster. Hence, the vector $\mathbf{g}_{{\rm D},c,r}^{(n)} \triangleq [g_{c,r}^{(n)}[0], \dots, g_{c,r}^{(n)}[D-1]]^{\rm T} \in \mathbb{C}^{D \times 1}$ denotes the path gain for all delay taps at the r-th ray within the c-th cluster, and the matrix $\mathbf{G}_{\rm D}^{(n)} \triangleq [\mathbf{g}_{{\rm D},1,1}^{(n)}, \dots, \mathbf{g}_{{\rm D},C,R_c}^{(n)}] \in \mathbb{C}^{D \times \mu}$ denotes the path gain for all delay taps and propagation paths. From the partial discrete Fourier transform (DFT) matrix $\mathbf{V} \in \mathbb{C}^{K \times D}$ and matrix $\mathbf{G}_{\rm D}^{(n)}$, the path gain at the k-th subcarrier $\mathbf{g}^{(n)}[k]$ is calculated as

$$\mathbf{g}^{(n)}[k] = \left(\mathbf{p}_k^{\rm T} \mathbf{V} \mathbf{G}_{\rm D}^{(n)}\right)^{\rm T} \tag{7}$$

where $\mathbf{p}_k \in \{0, 1\}^{K \times 1}$ is the vector whose k-th element is 1 and the other elements are 0.

C. Time Evolution of Channel Parameters

Following existing works [37], [38], the timeslot-by-timeslot channel variation is modeled based on spatial auto-correlation caused by the $vT_{\rm tck}$ [m] movements of the UE in a channel slot.

[1]The proposed channel tracking framework based on compressive sensing is not limited to ULA and can be easily extended to uniform planar array (UPA).

[2]Note that the beam squint effects are neglected for the same reason as the conventional study [37], leading to constant AoAs and AoDs over subcarriers.

Let $\boldsymbol{\xi}^{(n)} \triangleq \{\boldsymbol{\theta}^{(n)}, \boldsymbol{\phi}^{(n)}\}$ denote the set of AoAs and AoDs in the n-th channel slot. Similar to [37], from the analysis for angular spread in [45], the conditional probability density function (PDF) of AoAs and AoDs $p(\boldsymbol{\xi}^{(n)} \mid \boldsymbol{\xi}^{(n-1)})$ is assumed to follow a Laplace distribution. Hence, AoAs and AoDs, $\forall c, r$ at the n-th channel slot are modeled as

$$\phi_{c,r}^{(n)} \sim \mathcal{L}\big(\phi_{c,r}^{(n-1)}, \sigma_{\mathrm{A},c,r}^2\big), \quad \theta_{c,r}^{(n)} \sim \mathcal{L}\big(\theta_{c,r}^{(n-1)}, \sigma_{\mathrm{D},c,r}^2\big), \quad (8)$$

where the standard deviations $\sigma_{\mathrm{A},c,r} \in \mathbb{R}^+$ and $\sigma_{\mathrm{D},c,r} \in \mathbb{R}^+$ depend on the environmental conditions, such as the moving speed v and time window length T_{tck}.

In turn, the temporal evolution of the path gain follows a first-order Gaussian Markov process [28], [29], [33]–[35], [37], [38], which makes the path gain of all subcarriers and propagation paths at the n-th channel slot follow the autoregressive model of

$$\mathbf{g}^{(n)} = \mathbf{R}\mathbf{g}^{(n-1)} + \big(\mathbf{I} - \mathbf{R}\mathbf{R}^{\mathrm{H}}\big)^{\frac{1}{2}} \boldsymbol{\Delta}\mathbf{g}^{(n)}, \quad (9)$$

where the vector $\mathbf{g}^{(n)} \in \mathbb{C}^{K\mu \times 1}$ is defined as $\mathbf{g}^{(n)} \triangleq [\mathbf{g}^{(n)\mathrm{T}}[0], \ldots, \mathbf{g}^{(n)\mathrm{T}}[K-1]]^{\mathrm{T}} \in \mathbb{C}^{K\mu \times 1}$, and $\mathbf{R} \in \mathbb{R}^{K\mu \times K\mu}$ denotes the correlation matrix of which elements consist of spatial auto-correlation and frequency cross-correlation calculated by [37, Eq. (27)] and parameters in [46].

Following [37], no correlations between the different paths are assumed. Furthermore, $\Delta\mathbf{g}^{(n)} \in \mathbb{C}^{K\mu \times 1}$ denotes the multivariate random variable following an independent and identical distribution as $\Delta\mathbf{g}^{(n)} \sim \mathcal{CN}(\mathbf{0}, \mathbf{C}_{\mathbf{gg}}^{(n)})$, where $\mathbf{C}_{\mathbf{gg}}^{(n)} \in \mathbb{C}^{K\mu \times K\mu}$ denotes the covariance matrix for the path gain for all subcarriers and propagation paths.

D. Pilots, Precoding, and Combining for Channel Tracking

In what follows, we only consider the training phases after the initial channel slot $n = 0$ to focus on the channel tracking. The assumptions at the initial channel estimation are explained in Section III.

The pilot symbols, precoder, and combiner are designed under the frequency-flat assumption at the training phases to update CSI efficiently. Let $\mathbf{x}_{\mathrm{tr}}^{(m,n)} \in \mathbb{C}^{N_{\mathrm{s}} \times 1}$ denote the m-th $(m = 1, 2, \ldots, M_{\mathrm{tck}})$ pilot symbol vector, which consists of, *e.g.*, normalized quadrature phase shift keying (QPSK) symbols. Besides, let $N_{\mathrm{s}} \in \mathbb{N}^+$ denote the number of data streams that may depend on the number of RF chains, given by $N_{\mathrm{s}} = \min(N_{\mathrm{t,RF}}, N_{\mathrm{r,RF}})$.

Following [37], the symbol level precoding and combining are considered to maximize the SNR by exploiting estimated CSI at a previous channel slot $\hat{\mathbf{H}}^{(n-1)}[k], \forall k \in \mathcal{K}$, as shown in Fig. 1. Let $\mathbf{F}_{\mathrm{tr}}^{(m,n)} \in \mathbb{C}^{N_{\mathrm{t}} \times N_{\mathrm{s}}}$ and $\mathbf{W}_{\mathrm{tr}}^{(m,n)} \in \mathbb{C}^{N_{\mathrm{r}} \times N_{\mathrm{s}}}$ denote the hybrid precoder and combiner for the m-th pilot symbol at the n-th channel slot, respectively.

E. Received Signal

Given above, let $k_p \in \mathcal{P} \triangleq \{k_1, k_2, \ldots, k_P\} \subseteq \mathcal{K}$ denote the index for subcarriers utilized to send reference signals, where $P(\leq K)$ denotes the total number of such pilot subcarriers. The received signal at the n-th channel slot, k_p-th subcarrier, and m-th pilot symbol can be written as

$$\mathbf{y}^{(m,n)}[k_p] = \mathbf{W}_{\mathrm{tr}}^{(m,n)\mathrm{H}}\mathbf{H}^{(n)}[k_p]\mathbf{F}_{\mathrm{tr}}^{(m,n)}\mathbf{x}_{\mathrm{tr}}^{(m,n)} + \mathbf{W}_{\mathrm{tr}}^{(m,n)\mathrm{H}}\mathbf{z}_{\mathrm{w}}^{(m,n)}[k_p], \quad (10)$$

where $\mathbf{z}_{\mathrm{w}}^{(m,n)}[k] \sim \mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I}_{N_{\mathrm{r}}})$ denotes additive white Gaussian noise (AWGN) at the k-th subcarrier and the m-th pilot symbol time interval in the n-th channel slot.

Following the property $\mathrm{vec}(\mathbf{AXB}) = (\mathbf{B}^{\mathrm{T}} \otimes \mathbf{A})\mathrm{vec}(\mathbf{X})$, the received signal can be expressed as

$$\begin{aligned}\mathbf{y}^{(m,n)}[k_p] &= \boldsymbol{\Phi}^{(m,n)}\mathrm{vec}\left\{\mathbf{H}^{(n)}[k_p]\right\} + \mathbf{z}^{(m,n)}[k_p] \\ &= \boldsymbol{\Phi}^{(m,n)}\boldsymbol{\Psi}\big(\boldsymbol{\xi}^{(n)}\big)\mathbf{g}^{(n)}[k_p] + \mathbf{z}^{(m,n)}[k_p], \quad (11)\end{aligned}$$

where $\boldsymbol{\Phi}^{(m,n)} \triangleq \big(\mathbf{x}_{\mathrm{tr}}^{(m,n)\mathrm{T}}\mathbf{F}_{\mathrm{tr}}^{(m,n)\mathrm{T}} \otimes \mathbf{W}_{\mathrm{tr}}^{(m,n)\mathrm{H}}\big) \in \mathbb{C}^{N_{\mathrm{s}}N_{\mathrm{r}} \times N_{\mathrm{t}}\mu}$, $\mathbf{z}^{(m,n)}[k] \sim \mathcal{CN}(\mathbf{0}, \mathbf{C}_{\mathbf{zz}}^{(m,n)})$ denotes the colored Gaussian noise with the covariance matrix $\mathbf{C}_{\mathbf{zz}}^{(m,n)} = \mathbf{W}_{\mathrm{tr}}^{(m,n)\mathrm{H}}\mathbf{W}_{\mathrm{tr}}^{(m,n)}\sigma^2$, and $\boldsymbol{\Psi}^{(n)}\big(\boldsymbol{\xi}^{(n)}\big) \in \mathbb{C}^{N_{\mathrm{t}}N_{\mathrm{r}} \times \mu}$ contains array responses and can be defined as

$$\boldsymbol{\Psi}^{(n)}\big(\boldsymbol{\xi}^{(n)}\big) \triangleq \mathbf{A}_{\mathrm{T}}^*\big(\boldsymbol{\theta}^{(n)}\big) \circ \mathbf{A}_{\mathrm{R}}\big(\boldsymbol{\phi}^{(n)}\big). \quad (12)$$

Since the channels for all subcarriers $\mathbf{H}^{(n)}[k], \forall k$ are constant within a channel slot, the received signal for M_{tck} pilot symbols at the k_p-th subcarrier $\mathbf{y}^{(n)}[k_p] \triangleq \big[\mathbf{y}^{(1,n)\mathrm{T}}[k_p], \ldots, \mathbf{y}^{(M_{\mathrm{tck}},n)\mathrm{T}}[k_p]\big]^{\mathrm{T}} \in \mathbb{C}^{N_{\mathrm{s}}M_{\mathrm{tck}} \times 1}$ can be written as

$$\mathbf{y}^{(n)}[k_p] = \boldsymbol{\Phi}^{(n)}\boldsymbol{\Psi}\big(\boldsymbol{\xi}^{(n)}\big)\mathbf{g}^{(n)}[k_p] + \mathbf{z}^{(n)}[k_p], \quad (13)$$

where the matrix $\boldsymbol{\Phi}^{(n)}$ and the vector $\mathbf{z}^{(n)}[k_p]$ are defined as $\boldsymbol{\Phi}^{(n)} \triangleq \big[\boldsymbol{\Phi}^{(1,n)\mathrm{T}}, \ldots, \boldsymbol{\Phi}^{(M_{\mathrm{tck}},n)\mathrm{T}}\big]^{\mathrm{T}} \in \mathbb{C}^{M_{\mathrm{tck}}N_{\mathrm{s}} \times N_{\mathrm{t}}N_{\mathrm{r}}}$ and $\mathbf{z}^{(n)}[k_p] \triangleq \big[\mathbf{z}^{(1,n)\mathrm{T}}[k_p], \ldots, \mathbf{z}^{(M_{\mathrm{tck}},n)\mathrm{T}}[k_p]\big]^{\mathrm{T}} \in \mathbb{C}^{M_{\mathrm{tck}}N_{\mathrm{s}} \times 1}$, respectively, and the covariance matrix of the colored noise $\mathbf{z}^{(n)}[k]$ is given by $\mathbf{C}_{\mathbf{zz}}^{(n)} = \mathrm{blkdiag}\big\{\mathbf{C}_{\mathbf{zz}}^{(1,n)}[k], \ldots, \mathbf{C}_{\mathbf{zz}}^{(M_{\mathrm{tck}},n)}[k]\big\} \in \mathbb{C}^{M_{\mathrm{tck}}N_{\mathrm{s}} \times M_{\mathrm{tck}}N_{\mathrm{s}}}$.

In addition, the received signal for all pilot subcarriers $\mathbf{Y}_{\mathrm{tr}}^{(n)} \triangleq \big[\mathbf{y}^{(n)}[k_1], \ldots, \mathbf{y}^{(n)}[k_P]\big] \in \mathbb{C}^{N_{\mathrm{s}}M_{\mathrm{tck}} \times P}$ can be written as

$$\mathbf{Y}_{\mathrm{tr}}^{(n)} = \boldsymbol{\Phi}^{(n)}\boldsymbol{\Psi}(\boldsymbol{\xi}^{(n)})\mathbf{G}_P^{(n)} + \mathbf{Z}_P^{(n)}, \quad (14)$$

where the matrices $\mathbf{G}_P^{(n)} \triangleq \big[\mathbf{g}^{(n)}[k_1], \ldots, \mathbf{g}^{(n)}[k_P]\big] \in \mathbb{C}^{\mu \times P}$ and $\mathbf{Z}_P^{(n)} \triangleq \big[\mathbf{z}^{(n)}[k_1], \ldots, \mathbf{z}^{(n)}[k_P]\big] \in \mathbb{C}^{N_{\mathrm{s}}M_{\mathrm{tck}} \times P}$ denote the path gain and colored noise for all pilot subcarriers, respectively.

III. Proposed Channel Approximation

In this section, we introduce novel channel approximations by leveraging the time evolution of the angle parameters so as to avoid Monte Carlo sampling and reduce computational complexity in proportion to the number of available subcarriers K. It is assumed that the $\hat{\mu} \in \mathbb{N}^+$ propagation paths and corresponding parameters (*e.g.*, AoAs, AoDs, and the covariance matrix of the path gain) are estimated at the initial channel slot $n = 0$. Let $\hat{\phi}_i^{(n)}$, $\hat{\theta}_i^{(n)}$ with $i = 1, 2, \ldots, \hat{\mu}$, and $\hat{\mathbf{C}}_{\mathbf{gg}}^{(n)} \in \mathbb{C}^{K\hat{\mu} \times K\hat{\mu}}$ denote the estimated AoAs, AoDs, and covariance matrix of the path gain at the n-th channel

slot, respectively. The correlation matrix of the path gain corresponding to the estimated paths is assumed to be available and is denoted by $\hat{\mathbf{R}} \in \mathbb{R}^{K\hat{\mu}\times K\hat{\mu}}$. In the context of channel tracking, $\hat{\mu}$ estimated paths correspond to the paths that are tracked via a tracking algorithm. AoAs, AoDs, and the path gains are updated at each channel slot $n \geq 1$.

A. Sensing Matrix Design

Updating CSI leads to estimation of the array matrix $\mathbf{\Psi}(\boldsymbol{\xi}^{(n)})$ and path gain vector $\mathbf{g}^{(n)}[k_p]$, which should be performed from the observation $\mathbf{y}^{(n)}[k_p]$. The conventional approaches, as in [36]–[38], utilize a Monte-Carlo sampling method ($i.e.$, particle filtering) to obtain the optimal MMSE CSI estimate at the cost of tremendous computational complexity. To circumvent this issue, a sensing matrix design is proposed which exploits the time evolution models of AoAs and AoDs.

As considered in [21]–[25], [39], mmWave channel estimation and tracking problems can be reformulated by introducing an angle-domain dictionary with quantized angle candidates in $[-\pi, \pi]$. The major difference from such problems is that angle-domain dictionaries need to consider the temporal evolution models of the angles and the estimated CSI at the previous channel slot(s). This leads to the need of introducing a new dictionary design ($i.e.$, sensing matrix design). With that clarified, this subsection proposes a novel dictionary design.

Let $\tilde{\phi}_{i,j}^{(n)}$ and $\tilde{\theta}_{i,j}^{(n)}$ denote the j-th candidate of AoA and AoD at the i-th tracking path and the n-th channel slot, respectively. Taking the confidence interval of the Laplace-distributed temporal evolution models in (8) into account, a reliable angle interval of $\tilde{\phi}_{i,j}^{(n)}$ and $\tilde{\theta}_{i,j}^{(n)}$ can be represented by a closed interval of $\tilde{\phi}_{i,j}^{(n)} \in \big[\hat{\phi}_i^{(n-1)} - \Phi_i\sigma_{\mathrm{A},i},\ \hat{\phi}_i^{(n-1)} + \Phi_i\sigma_{\mathrm{A},i}\big]$ and $\tilde{\theta}_{i,j}^{(n)} \in \big[\hat{\theta}_i^{(n-1)} - \Theta_i\sigma_{\mathrm{D},i},\ \hat{\theta}_i^{(n-1)} + \Theta_i\sigma_{\mathrm{D},i}\big]$, where $\Phi_i \in \mathbb{N}^+$ and $\Theta_i \in \mathbb{N}^+$ denotes the positive integer determining the confidence interval for the AoA and AoD at the i-th tracking path, respectively. Under uniform quantization, the j-th candidate of the AoA with $j \in \{0, 1, \ldots, Q_{\mathrm{R},i} - 1\}$ and AoD with $j \in \{0, 1, \ldots, Q_{\mathrm{T},i} - 1\}$ are, respectively, given as

$$\tilde{\phi}_{i,j}^{(n)} = \hat{\phi}_i^{(n-1)} + \Phi_i\sigma_{\mathrm{A},i}\frac{-(Q_{\mathrm{R},i} - 1) + 2j}{Q_{\mathrm{R},i} - 1}, \quad \forall i, \tag{15a}$$

$$\tilde{\theta}_{i,j}^{(n)} = \hat{\theta}_i^{(n-1)} + \Theta_i\sigma_{\mathrm{D},i}\frac{-(Q_{\mathrm{T},i} - 1) + 2j}{Q_{\mathrm{T},i} - 1}, \quad \forall i, \tag{15b}$$

where $Q_{\mathrm{R},i} \in \mathbb{N}^+$ and $Q_{\mathrm{T},i} \in \mathbb{N}^+$ denote the number of candidates for the AoA and AoD at the i-th tracking path, respectively.

Let $\tilde{\boldsymbol{\phi}}_i^{(n)} \triangleq \big\{\tilde{\phi}_{i,j}^{(n)}\big\}_{j=0}^{Q_{\mathrm{R},i}-1}$, $\tilde{\boldsymbol{\theta}}_i^{(n)} \triangleq \big\{\tilde{\theta}_{i,j}^{(n)}\big\}_{j=0}^{Q_{\mathrm{T},i}-1}$, and $\tilde{\boldsymbol{\xi}}_i^{(n)} \triangleq \big\{\tilde{\boldsymbol{\phi}}_i^{(n)}, \tilde{\boldsymbol{\theta}}_i^{(n)}\big\}$ be the set of candidates for the AoA, AoD, and the union of the AoA and AoD sets at the i-th tracking path, respectively. In light of the above, all the possible combinations of the array responses of any AoA-AoD pair in $\tilde{\boldsymbol{\xi}}_i^{(n)}$ can be expressed as

$$\tilde{\mathbf{A}}\big(\tilde{\boldsymbol{\xi}}_i^{(n)}\big) \triangleq \Big(\tilde{\mathbf{A}}_{\mathrm{T}}^*\Big(\tilde{\boldsymbol{\theta}}_i^{(n)}\Big) \otimes \tilde{\mathbf{A}}_{\mathrm{R}}\Big(\tilde{\boldsymbol{\phi}}_i^{(n)}\Big)\Big) \in \mathbb{C}^{N_{\mathrm{t}}N_{\mathrm{r}}\times Q_{\mathrm{T},i}Q_{\mathrm{R},i}}. \tag{16}$$

Since the total number of such AoA-AoD combinations can be written as $E = \sum_{i=1}^{\hat{\mu}} Q_{\mathrm{R},i}Q_{\mathrm{T},i}$ and $\tilde{\boldsymbol{\xi}}^{(n)} \triangleq \big\{\tilde{\boldsymbol{\xi}}_i^{(n)}\big\}_{i=1}^{\hat{\mu}}$, concatenating $\tilde{\mathbf{A}}\big(\tilde{\boldsymbol{\xi}}_i^{(n)}\big)$ over all the tracking paths yields the array dictionary $\tilde{\mathbf{\Psi}}(\tilde{\boldsymbol{\xi}}^{(n)}) \triangleq \big[\tilde{\mathbf{A}}\big(\tilde{\boldsymbol{\xi}}_1^{(n)}\big), \ldots, \tilde{\mathbf{A}}\big(\tilde{\boldsymbol{\xi}}_{\hat{\mu}}^{(n)}\big)\big] \in \mathbb{C}^{N_{\mathrm{t}}N_{\mathrm{r}}\times E}$, which enables us to rewrite the vectorized channel as

$$\mathrm{vec}\big\{\mathbf{H}^{(n)}[k]\big\} \approx \tilde{\mathbf{\Psi}}\big(\tilde{\boldsymbol{\xi}}^{(n)}\big)\tilde{\mathbf{g}}^{(n)}[k], \tag{17}$$

where the vector $\tilde{\mathbf{g}}^{(n)}[k] \in \mathbb{C}^{E\times 1}$ is a sparse representation of the path gain vector with the e-th element being the path gain at the k-th subcarrier and AoA-AoD pair corresponding to the e-th column of the array dictionary $\tilde{\mathbf{\Psi}}\big(\tilde{\boldsymbol{\xi}}^{(n)}\big)$.

Hence, the received signal at the k_p-th subcarrier and n-th channel slot can be approximated as

$$\mathbf{y}^{(n)}[k_p] \approx \mathbf{\Phi}^{(n)}\tilde{\mathbf{\Psi}}\big(\tilde{\boldsymbol{\xi}}^{(n)}\big)\tilde{\mathbf{g}}^{(n)}[k_p] + \mathbf{z}^{(n)}[k_p]. \tag{18}$$

Under this approximation, the sparse path gain vector $\tilde{\mathbf{g}}^{(n)}[k_p]$ is the only unknown variable, and reconstructing $\tilde{\mathbf{g}}^{(n)}[k_p]$ is equivalent to estimating AoAs, AoDs, and the path gains jointly. Specifically, the angle information can be retrieved from the position of non-zeros of $\tilde{\mathbf{g}}^{(n)}[k_p]$ and those non-zeros correspond to the path gains. Note that the sensing matrix in (18) consists of the pilot symbols, precoders, combiners, and the array dictionary, which are all known a priori. For the sake of simplicity, we hereafter consider $\mathbf{B}^{(n)} \triangleq \mathbf{\Phi}^{(n)}\tilde{\mathbf{\Psi}}\big(\tilde{\boldsymbol{\xi}}^{(n)}\big) \in \mathbb{C}^{M_{\mathrm{tck}}N_{\mathrm{s}}\times E}$.

B. Received Signal with Delay Domain Path Gain

With the frequency-independent sensing matrix $\mathbf{B}^{(n)}$, the received signal for all pilot subcarriers in (14) can be approximated as

$$\mathbf{Y}_{\mathrm{tr}}^{(n)} \approx \mathbf{B}^{(n)}\tilde{\mathbf{G}}_P^{(n)} + \mathbf{Z}_P^{(n)}, \tag{19}$$

where $\tilde{\mathbf{G}}_P^{(n)} \triangleq \big[\tilde{\mathbf{g}}^{(n)}[k_1], \ldots, \tilde{\mathbf{g}}^{(n)}[k_P]\big] \in \mathbb{C}^{E\times P}$ denotes a row-sparse matrix collecting all $\tilde{\mathbf{g}}^{(n)}[k_p], \forall k_p$ owing to the fact that the AoAs and AoDs are considered to be shared over subcarriers.

As in the conventional studies such as [35]–[38], the frequency-domain representation of the received signal in (19) requires allocating pilot sequences to all subcarriers in order to estimate the CSI of all subcarriers, which increases computational complexity in proportion to the number of available subcarriers K. In order to bypass this issue, we consider the delay domain counterpart and interpolation via the Fourier transform [47]. To be specific, the path gain at the k-th subcarrier $\mathbf{g}^{(n)}[k]$ can be represented in the delay domain as in (7), which yields the following received signal model

$$\mathbf{Y}_{\mathrm{tr}}^{(n)} = \mathbf{\Phi}^{(n)}\mathbf{\Psi}(\boldsymbol{\xi}^{(n)})\mathbf{G}_{\mathrm{D}}^{(n)\mathrm{T}}(\mathbf{PV})^{\mathrm{T}} + \mathbf{Z}_P^{(n)}, \tag{20}$$

where $\mathbf{P} \triangleq [\mathbf{p}_{k_1}, \mathbf{p}_{k_2}, \ldots, \mathbf{p}_{k_P}]^{\mathrm{T}} \in \{0, 1\}^{P\times K}$.

The relationship between delay and frequency domain is immutable under channel approximation using the array dictionary $\tilde{\mathbf{\Psi}}\big(\tilde{\boldsymbol{\xi}}^{(n)}\big)$. Hence, the approximated received signal with the sensing matrix and the path gain in the delay domain can be expressed as

$$\mathbf{Y}_{\mathrm{tr}}^{(n)} \approx \mathbf{B}^{(n)}\tilde{\mathbf{G}}_{\mathrm{D}}^{(n)\mathrm{T}}(\mathbf{PV})^{\mathrm{T}} + \mathbf{Z}_P^{(n)}, \tag{21}$$

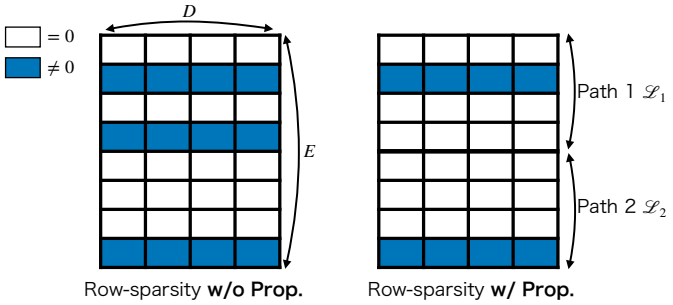


Fig. 2: Row-sparse structures of the transpose matrix $\tilde{\mathbf{G}}_{\mathrm{D}}^{(n)\mathrm{T}}$ without and with the proposed sensing matrix $\mathbf{B}^{(n)}$ under $\hat{\mu} = 2$ tracking paths

where $\tilde{\mathbf{G}}_{\mathrm{D}}^{(n)} \in \mathbb{C}^{D\times E}$ denotes the approximated path gain matrix in the delay domain, each row of which corresponds to the path gain at each delay tap.

C. Sparsity of Path Gain Matrix

The transpose matrix $\tilde{\mathbf{G}}_{\mathrm{D}}^{(n)\mathrm{T}} \in \mathbb{C}^{E\times D}$ has a row-sparse structure due to a larger number of candidates $E \gg \hat{\mu}$ and the fact that AoAs and AoDs are constant over the delay taps. Moreover, the position and number of non-zero rows of this matrix are subject to the proposed sensing matrix design.

As shown in Fig. 2, the aforementioned sensing matrix $\mathbf{B}^{(n)}$ can reveal the position of rows of the transpose matrix $\tilde{\mathbf{G}}_{\mathrm{D}}^{(n)\mathrm{T}}$, which corresponds to the path gain at the i-th tracking path. In addition, only one AoA (AoD) candidate belonging to set $\tilde{\boldsymbol{\phi}}_i^{(n)}$ ($\tilde{\boldsymbol{\theta}}_i^{(n)}$) is selected as the estimated AoA (AoD) for the i-th tracking path, as highlighted in blue in Fig. 2. Therefore, the sparsity of the path gain matrix $\tilde{\mathbf{G}}_{\mathrm{D}}^{(n)\mathrm{T}}$ can be expressed as

$$\sum_{\ell\in\mathcal{L}_i} \left\| \left[\tilde{\mathbf{G}}_{\mathrm{D}}^{(n)\mathrm{T}}\right]_{\ell} \right\|_{2,0} = 1,\ i = 1, 2, \ldots, \hat{\mu} \tag{22}$$

where $\mathcal{L}_i \triangleq \{\sum_{i'=1}^{i} Q_{\mathrm{R},i'-1}Q_{\mathrm{T},i'-1} + 1, \sum_{i'=1}^{i} Q_{\mathrm{R},i'-1}Q_{\mathrm{T},i'-1} + 2, \ldots, \sum_{i'=1}^{i} Q_{\mathrm{R},i'}Q_{\mathrm{T},i'}\}$ denotes the index set of rows corresponding to all AoA and AoD candidates for the i-th tracking path $\tilde{\boldsymbol{\xi}}_i^{(n)}$ for all delay taps, where $Q_{\mathrm{R},0} = Q_{\mathrm{T},0} = 0$, and $|\mathcal{L}_i| = Q_{\mathrm{R},i}Q_{\mathrm{T},i}$.

Note that the approximated path gain matrix in the frequency domain $\tilde{\mathbf{G}}_P^{(n)}$ also has the same sparse structure in (22), and therefore the proposed algorithm presented in the following sections is not limited to the delay domain.

IV. Proposed Channel Tracking

In this section, we propose a channel tracking algorithm based on compressive sensing.

A. Compressive Sensing for Channel Tracking

From all the above, channel tracking can be formulated as a multiple measurement vector (MMV) reconstruction problem with $l_{2,0}$ norm constraint given as

$$\underset{\tilde{\mathbf{G}}_{\mathrm{D}}^{(n)}}{\text{minimize}}\ \frac{1}{2}\left\|\mathbf{B}^{(n)}\tilde{\mathbf{G}}_{\mathrm{D}}^{(n)\mathrm{T}}(\mathbf{PV})^{\mathrm{T}} - \mathbf{Y}_{\mathrm{tr}}^{(n)}\right\|_{\mathrm{F}}^2, \tag{23a}$$

$$\text{subject to}\ \sum_{\ell\in\mathcal{L}_i}\left\|\left[\tilde{\mathbf{G}}_{\mathrm{D}}^{(n)\mathrm{T}}\right]_{\ell}\right\|_{2,0} = 1, \forall i. \tag{23b}$$

Assuming that the number of nonzero elements in the sparse signal is known, the sparse signal reconstruction problem with l_0 norm constraints can be efficiently solved using the exact top-k feature selection as in [44]. In this study, the exact top-$\hat{\mu}$ feature selection is proposed, which obtains $\hat{\mu}$ nonzero elements exploiting the unique sparse structure in the path gain matrix caused by the sensing matrix design.

First, we introduce an auxiliary variable $\mathbf{L} \in \mathbb{C}^{E\times D}$ with the same $l_{2,0}$-norm constraint as $\tilde{\mathbf{G}}_{\mathrm{D}}^{(n)\mathrm{T}}$, which transforms constraint (23b) into the constraint $\left\|\tilde{\mathbf{G}}_{\mathrm{D}}^{(n)\mathrm{T}} - \mathbf{L}\right\|_{\mathrm{F}}^2 = 0$. Under this transformation, the optimization problem (23) can be rewritten as an unconstrained problem via the augmented Lagrangian method, such that

$$\underset{\tilde{\mathbf{G}}_{\mathrm{D}}^{(n)\mathrm{T}},\mathbf{L}}{\text{minimize}}\ \left\|\mathbf{B}^{(n)}\tilde{\mathbf{G}}_{\mathrm{D}}^{(n)\mathrm{T}}(\mathbf{PV})^{\mathrm{T}} - \mathbf{Y}_{\mathrm{tr}}^{(n)}\right\|_{\mathrm{F}}^2 + \left\|\tilde{\mathbf{G}}_{\mathrm{D}}^{(n)\mathrm{T}} - \mathbf{L} + \boldsymbol{\Lambda}/\beta\right\|_{\mathrm{F}}^2, \tag{24}$$

where the scalar $\beta \in \mathbb{R}^+$ and matrix $\boldsymbol{\Lambda} \in \mathbb{C}^{E\times K}$ denote the penalty coefficient and Lagrangian multiplier, respectively, and note that this unconstrained optimization problem can be efficiently solved by iteratively updating $\tilde{\mathbf{G}}_{\mathrm{D}}^{(n)\mathrm{T}}$, $\mathbf{L}$, $\boldsymbol{\Lambda}$, and β.

B. Update of Path Gain Matrix

Let us first derive the updated expression of the approximated path gain matrix $\tilde{\mathbf{G}}_{\mathrm{D}}^{(n)\mathrm{T}}$ for a fixed $\mathbf{L}$. In (24), the optimization problem for the matrix $\tilde{\mathbf{G}}_{\mathrm{D}}^{(n)\mathrm{T}}$ is a Frobenius norm minimization problem. Hence, the optimal solution can be obtained at $\frac{\partial}{\partial\tilde{\mathbf{G}}_{\mathrm{D}}^{(n)\mathrm{T}}} f\left(\tilde{\mathbf{G}}_{\mathrm{D}}^{(n)\mathrm{T}}\right) = \mathbf{0}$, where function $f(\tilde{\mathbf{G}}_{\mathrm{D}}^{(n)\mathrm{T}})$ denotes the objective function in (24). Following the Wirtinger derivative and the fact that $(\mathbf{PV})^{\mathrm{T}}\left\{(\mathbf{PV})^{\mathrm{T}}\right\}^{\mathrm{H}} = \frac{P}{K}\mathbf{I}_D$, the closed-form solution can be written as

$$\tilde{\mathbf{G}}_{\mathrm{D}}^{(n)\mathrm{T}} = \left(\frac{P}{K}\mathbf{B}^{(n)\mathrm{H}}\mathbf{B}^{(n)} + \mathbf{I}_E\right)^{-1}\left(\mathbf{L} - \frac{\boldsymbol{\Lambda}}{\beta} + \mathbf{B}^{(n)\mathrm{H}}\mathbf{Y}_{\mathrm{tr}}^{(n)}\mathbf{PV}^*\right). \tag{25}$$

C. Update of Auxiliary Variable

For a given $\tilde{\mathbf{G}}_{\mathrm{D}}^{(n)\mathrm{T}}$, the update expression of the auxiliary variable $\mathbf{L}$ is given in what follows. The optimization problem with respect to the auxiliary variable $\mathbf{L}$ is also a Frobenius norm minimization problem, given by

$$\underset{\mathbf{L}}{\text{minimize}}\ \left\|\mathbf{L} - (\tilde{\mathbf{G}}_{\mathrm{D}}^{(n)\mathrm{T}} + \boldsymbol{\Lambda}/\beta)\right\|_{\mathrm{F}}^2. \tag{26}$$

Since the domain of the auxiliary variable $\mathbf{L}$ is subject to the sparse structure described in Section III-C, the e-th row with $e \in \mathcal{L}_i$ in the auxiliary variable $\mathbf{L}$ is determined by

$$[\mathbf{L}]_e = \begin{cases} [\mathbf{X}]_e, \text{ if } \left\|[\mathbf{X}]_e\right\|_2 = \max\limits_{\forall\ell\in\mathcal{L}_i}\left\|[\mathbf{X}]_\ell\right\|_2 \\ \mathbf{0}, \text{ otherwise} \end{cases}, \tag{27}$$

where $\mathbf{X} \triangleq (\tilde{\mathbf{G}}_{\mathrm{D}}^{(n)\mathrm{T}} + \boldsymbol{\Lambda}/\beta)$.

D. Overall Algorithm

Finally, the Lagrangian multiplier and penalty coefficient are updated as $\mathbf{\Lambda} = \mathbf{\Lambda} + (\tilde{\mathbf{G}}_{\mathrm{D}}^{(n)\mathrm{T}} - \mathbf{L})$ and $\beta = \rho \cdot \beta$, where $\rho \in \mathbb{R}^{+}$ is a hyperparameter. Thus, $\tilde{\mathbf{G}}_{\mathrm{D}}^{(n)\mathrm{T}}$ and $\mathbf{L}$ are solved using the same procedure as in IV-B and IV-C, which is then iterated until convergence. Note that the convergence behavior is determined by the hyperparameter ρ, and convergence analyses are provided by [44].

At convergence, the index of non-zeros of the matrix $\tilde{\mathbf{G}}_{\mathrm{D}}^{(n)\mathrm{T}}$ belonging to the set of the i-th tracking path $\ell^{\star} \in \mathcal{L}_i$ is finally extracted as

$$\ell^{\star} = \underset{\ell, \forall \ell \in \mathcal{L}_i}{\operatorname{argmax}} \left\| \left[\tilde{\mathbf{G}}_{\mathrm{D}}^{(n)\mathrm{T}}\right]_{\ell} \right\|_2 = \underset{\ell, \forall \ell \in \mathcal{L}_i}{\operatorname{argmax}} \left\| [\mathbf{L}]_{\ell} \right\|_2. \tag{28}$$

Then, the estimated AoA, AoD, and the path gain at the i-th tracking path and the n-th channel slot are determined by

$$\hat{\phi}_i^{(n)} = \tilde{\boldsymbol{\phi}}_i^{(n)}\{(\ell^{\star} - 1) \mod Q_{\mathrm{T},i} Q_{\mathrm{R},i}\}, \tag{29}$$

$$\hat{\theta}_i^{(n)} = \tilde{\boldsymbol{\theta}}_i^{(n)}\left\{ \left\lfloor \tfrac{\ell^{\star} - 1}{Q_{\mathrm{T},i} Q_{\mathrm{R},i}} \right\rfloor \right\}, \tag{30}$$

$$\hat{\mathbf{g}}_{\mathrm{D},i}^{(n)\mathrm{T}} = \left\| \left[\tilde{\mathbf{G}}_{\mathrm{D}}^{(n)\mathrm{T}}\right]_{\ell^{\star}} \right\|_2, \tag{31}$$

where $\tilde{\boldsymbol{\phi}}_i^{(n)}\{j\}$ and $\tilde{\boldsymbol{\theta}}_i^{(n)}\{j\}$ denote the j-th candidates for the AoA $\tilde{\phi}_{i,j}^{(n)}$ and AoD $\tilde{\theta}_{i,j}^{(n)}$ at the i-th tracking path and the n-th channel slot, respectively, and the estimated path gain for all delay taps and tracking paths at the n-th channel slot is expressed by the matrix $\hat{\mathbf{G}}_{\mathrm{D}}^{(n)} \triangleq \left[\hat{\mathbf{g}}_{\mathrm{D},1}^{(n)}, \ldots, \hat{\mathbf{g}}_{\mathrm{D},\hat{\mu}}^{(n)}\right] \in \mathbb{C}^{D \times \hat{\mu}}$.

Channel tracking based on the exact top-$\hat{\mu}$ feature selection is summarized in the pseudocode in Algorithm 1, where $\hat{\boldsymbol{\phi}}^{(n)} \triangleq \{\hat{\phi}_i^{(n)}\}_{i=1}^{\hat{\mu}}$ and $\hat{\boldsymbol{\theta}}^{(n)} \triangleq \{\hat{\theta}_i^{(n)}\}_{i=1}^{\hat{\mu}}$ denote the set of estimated AoAs and AoDs at the n-th channel slot, respectively. Unlike particle filter (PF) [37], [38] which requires PDF evaluations for all AoA and AoD candidates (*i.e.*, particles) in Monte Carlo sampling, the proposed compressive sensing algorithm only requires one least squares (LS) solution in (25) and norm calculation for the matrix $(\tilde{\mathbf{G}}_{\mathrm{D}}^{(n)\mathrm{T}} + \mathbf{\Lambda}/\beta)$ in (27) in each iteration, resulting in lower complexity.

Algorithm 1 Exact top-$\hat{\mu}$ feature selection for channel tracking

Input: $\mathbf{Y}_{\mathrm{tr}}^{(n)}$, $\mathbf{\Phi}^{(n)}$, ρ, $\hat{\boldsymbol{\phi}}^{(n-1)}$, $\hat{\boldsymbol{\theta}}^{(n-1)}$, $\sigma_{\mathrm{A},i}$, $\sigma_{\mathrm{D},i}$, T_{top}
Output: $\hat{\mathbf{H}}^{(n)}[k], \forall k$

Generate candidates $\tilde{\boldsymbol{\phi}}^{(n)}$, $\tilde{\boldsymbol{\theta}}^{(n)}$ following (15a), (15b)
Design sensing matrix $\mathbf{B}^{(n)} = \mathbf{\Phi}^{(n)} \tilde{\mathbf{\Psi}}(\tilde{\boldsymbol{\xi}}^{(n)})$
Augmented Lagrangian
Initialize $\mathbf{\Lambda}(0) = \mathbf{0}$, $\beta(0) = 1$, $\mathbf{L} = \mathbf{0}$
for $t = 1, 2, \ldots, T_{\mathrm{top}}$ **do**
    Solve $\tilde{\mathbf{G}}_{\mathrm{D}}^{(n)\mathrm{T}}$ following (25)
    Initialize $\mathbf{L} = \mathbf{0}$
    for $i = 1, 2, \ldots, \hat{\mu}$ **do**
        Determine $[\mathbf{L}]_e$, $\forall e \in \mathcal{L}_i$ following (27)
    end for
    Update $\mathbf{\Lambda}(t+1) = \mathbf{\Lambda}(t) + (\tilde{\mathbf{G}}_{\mathrm{D}}^{(n)\mathrm{T}} - \mathbf{L})$
    Update $\beta(t+1) = \rho\beta(t)$
end for
Obtain channel parameters
for $i = 1, 2, \ldots, \hat{\mu}$ **do**
    Obtain $\hat{\phi}_i^{(n)}$, $\hat{\theta}_i^{(n)}$, $\hat{\mathbf{g}}_{\mathrm{D},i}^{(n)}$ following (29), (30), (31)
end for
$\hat{\mathbf{H}}^{(n)}[k] = \mathbf{A}_{\mathrm{R}}(\hat{\boldsymbol{\phi}}^{(n)}) \mathrm{diag}\{(\mathbf{p}_k^{\mathrm{T}} \mathbf{V} \hat{\mathbf{G}}_{\mathrm{D}}^{(n)})^{\mathrm{T}}\} \mathbf{A}_{\mathrm{T}}^{\mathrm{H}}(\hat{\boldsymbol{\theta}}^{(n)}), \forall k$

V. Algorithm Refinement and Sub-Sensing Matrices

In this section, the previously proposed algorithm is refined to achieve further reconstruction accuracy and robustness against unfavorable conditions, such as low SNR and scarce measurements $M_{\mathrm{tck}} N_{\mathrm{s}} \to 1$.

A. Reformulation

Although the proposed channel tracking in Algorithm 1 can avoid tremendous computational costs of Monte Carlo sampling and available subcarriers K, the temporal evolution model and previous path gain estimates cannot be fully exploited. In order to achieve further performance improvements, instead of (31), the estimated path gain $\hat{\mathbf{G}}_{\mathrm{D}}^{(n)}$ is optimized using MMSE fashion by exploiting prior knowledge of path gain and the AoAs and AoDs obtained from equations (29) and (30).

This refinement implies that minimizing both the Frobenius norm in (23a) and the mean square error of the path gain estimation is needed to estimate all the parameters. A reasonable option to deal with two objectives is considering the linear combination of these [48] so that the channel tracking is rewritten by

$$\underset{\tilde{\mathbf{G}}_{\mathrm{D}}^{(n)\mathrm{T}}, \hat{\mathbf{G}}_{\mathrm{D}}^{(n)}}{\text{minimize}} \frac{1}{2} \left\| \mathbf{B}^{(n)} \tilde{\mathbf{G}}_{\mathrm{D}}^{(n)\mathrm{T}} (\mathbf{P}\mathbf{V})^{\mathrm{T}} - \mathbf{Y}_{\mathrm{tr}}^{(n)} \right\|_{\mathrm{F}}^2 + \mathbb{E}\left[\left\| \mathbf{G}_{\hat{\mu}}^{(n)} - \hat{\mathbf{G}}_{\mathrm{D}}^{(n)} \right\|_{\mathrm{F}}^2 \right], \tag{32a}$$

$$\text{subject to } \sum_{\ell \in \mathcal{L}_i} \left\| \left[\tilde{\mathbf{G}}_{\mathrm{D}}^{(n)\mathrm{T}}\right]_{\ell} \right\|_{2,0} = 1, \forall i, \tag{32b}$$

where $\mathbf{G}_{\hat{\mu}}^{(n)} \in \mathbb{C}^{D \times \hat{\mu}}$ satisfies $\mathbf{\Psi}(\hat{\boldsymbol{\xi}}^{(n)}) \mathbf{G}_{\hat{\mu}}^{(n)\mathrm{T}} \mathbf{V}^{\mathrm{T}} \mathbf{p}_k = \mathrm{vec}\{\mathbf{H}^{(n)}[k]\}$ for $\forall k \in \mathcal{K}$ with $\hat{\boldsymbol{\xi}}^{(n)} \triangleq \{\hat{\boldsymbol{\phi}}^{(n)}, \hat{\boldsymbol{\theta}}^{(n)}\}$ denoting the set of estimated AoAs and AoDs, and note that the MMSE approach with actual path gain $\mathbf{G}_{\hat{\mu}}^{(n)}$ that cannot be explicitly expressed is explained in detail in Section V-D.

B. Multiple Sub-sensing Matrices

When it comes to the AoAs and AoDs estimation, the first term of the objective function (32a) suffices while fixing $\hat{\mathbf{G}}_{\mathrm{D}}^{(n)}$, which can be performed by Algorithm 1. However, Algorithm 1 does not take into account the mutual coherence of the sensing matrix $\mathbf{B}^{(n)}$; in other words, any pair of the columns of $\mathbf{B}^{(n)}$ may be highly correlated, leading to high mutual coherence and performance limitation [49].

The mutual coherence of the sensing matrix consisting of the array dictionary depends on the minimum angle difference between AoA and AoD candidates. In the proposed design, the limitations of quantization intervals of the angle candidates may lead to higher mutual coherence than that in conventional estimation methods [21]–[25], [39]. In particular, tracking channel fluctuations in the order of less than milliseconds

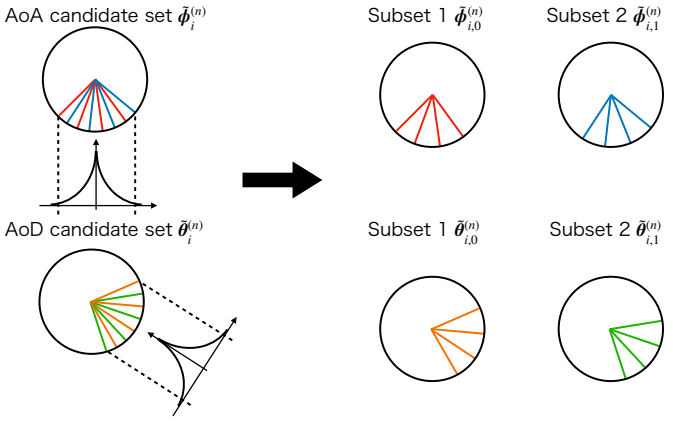


Fig. 3: Example of subsets for angle candidates.

are of interest in the considered channel tracking context, which can result in rather high mutual coherence as the angle difference may be minuscule. Moreover, mutual coherence affects the reconstruction performance more remarkably under a lower SNR and/or smaller measurement dimensions $M_{\mathrm{tck}}N_{\mathrm{s}}$ relative to the variable dimensions $N_{\mathrm{r}}N_{\mathrm{t}}$.

To accurately update CSI, even under such conditions, the concept of multiple sub-sensing matrices by splitting the sensing matrix $\mathbf{B}^{(n)}$ for mutual coherence reduction is introduced. Note that, in the proposed sensing matrix design, the estimated AoA and AoD at the i-th tracking path are selected from only within the sets $\tilde{\boldsymbol{\phi}}_i^{(n)}$ and $\tilde{\boldsymbol{\theta}}_i^{(n)}$, respectively, and the effective mutual coherence is determined by the minimum angle difference within the sets $\tilde{\boldsymbol{\phi}}_i^{(n)}$ and $\tilde{\boldsymbol{\theta}}_i^{(n)}$ for $\forall i$. Therefore, the sets $\tilde{\boldsymbol{\phi}}_i^{(n)}$ and $\tilde{\boldsymbol{\theta}}_i^{(n)}$, $\forall i$ are split into several subsets to maximize the minimum angle difference within the subsets, as shown in Fig. 3, which is an example that a candidate set is split into two subsets.

Specifically, for the i-th tracking path, $Q_{\mathrm{R},i} = U_{\mathrm{R},i} \cdot q_{\mathrm{R},i}$ AoA and $Q_{\mathrm{T},i} = U_{\mathrm{T},i} \cdot q_{\mathrm{T},i}$ AoD candidates are considered and split into $U_{\mathrm{R},i} \in \mathbb{N}^+$ and $U_{\mathrm{T},i} \in \mathbb{N}^+$ subsets, respectively, which are defined as

$$\tilde{\boldsymbol{\phi}}_{i,u}^{(n)} \triangleq \left\{\tilde{\phi}_{i,u}^{(n)}, \tilde{\phi}_{i,u+U_{\mathrm{R},i}}^{(n)}, \tilde{\phi}_{i,u+2U_{\mathrm{R},i}}^{(n)}, \ldots, \tilde{\phi}_{i,u+(q_{\mathrm{R},i}-1)U_{\mathrm{R},i}}^{(n)}\right\}, \tag{33}$$

$$\tilde{\boldsymbol{\theta}}_{i,u}^{(n)} \triangleq \left\{\tilde{\theta}_{i,u}^{(n)}, \tilde{\theta}_{i,u+U_{\mathrm{T},i}}^{(n)}, \tilde{\theta}_{i,u+2U_{\mathrm{T},i}}^{(n)}, \ldots, \tilde{\theta}_{i,u+(q_{\mathrm{T},i}-1)U_{\mathrm{T},i}}^{(n)}\right\}, \tag{34}$$

where $u \in \mathbb{N}$ denotes the index of the subsets, and note that the sets of AoA and AoD candidates for the i-th tracking path can be interpreted as

$$\tilde{\boldsymbol{\phi}}_i^{(n)} \triangleq \left\{\tilde{\boldsymbol{\phi}}_{i,0}^{(n)}, \tilde{\boldsymbol{\phi}}_{i,1}^{(n)}, \ldots, \tilde{\boldsymbol{\phi}}_{i,U_{\mathrm{R},i}-1}^{(n)}\right\}, \quad \left|\tilde{\boldsymbol{\phi}}_i^{(n)}\right| = U_{\mathrm{R},i}, \tag{35}$$

$$\tilde{\boldsymbol{\theta}}_i^{(n)} \triangleq \left\{\tilde{\boldsymbol{\theta}}_{i,0}^{(n)}, \tilde{\boldsymbol{\theta}}_{i,1}^{(n)}, \ldots, \tilde{\boldsymbol{\theta}}_{i,U_{\mathrm{T},i}-1}^{(n)}\right\}, \quad \left|\tilde{\boldsymbol{\theta}}_i^{(n)}\right| = U_{\mathrm{T},i}, \tag{36}$$

where each element corresponds to the subset, meaning $|\tilde{\boldsymbol{\phi}}_{i,u}^{(n)}| = q_{\mathrm{R},i}$ and $|\tilde{\boldsymbol{\theta}}_{i,u}^{(n)}| = q_{\mathrm{T},i}$.

Then, the union of the u-th AoA subset and the u'-th AoD subset $\tilde{\boldsymbol{\xi}}_{i,u,u'}^{(n)} \triangleq \left\{\tilde{\boldsymbol{\phi}}_{i,u}^{(n)}, \tilde{\boldsymbol{\theta}}_{i,u'}^{(n)}\right\}$ is one of the possible AoA-AoD subset pairs for the i-th tracking path. Hence, the direct product of the sets $\tilde{\boldsymbol{\phi}}_i^{(n)} \times \tilde{\boldsymbol{\theta}}_i^{(n)}$ expresses all the possible AoA-AoD subset pairs for the i-th tracking path $\tilde{\boldsymbol{\phi}}_i^{(n)} \times \tilde{\boldsymbol{\theta}}_i^{(n)} \triangleq \left\{\tilde{\boldsymbol{\xi}}_{i,0,0}^{(n)}, \ldots, \tilde{\boldsymbol{\xi}}_{i,U_{\mathrm{R},i}-1,U_{\mathrm{T},i}-1}^{(n)}\right\}$. Considering all tracking paths, the direct product of all the subsets for AoA and AoD candidates can express all the possible AoA-AoD subset pairs, defined as

$$\begin{aligned}\tilde{\boldsymbol{\xi}}_{\mathrm{all}}^{(n)} &\triangleq \prod_{i=1}^{\hat{\mu}} \tilde{\boldsymbol{\phi}}_i^{(n)} \times \tilde{\boldsymbol{\theta}}_i^{(n)} \\ &= \Big\{ \underbrace{\left\{\tilde{\boldsymbol{\xi}}_{1,0,0}^{(n)}, \ldots, \tilde{\boldsymbol{\xi}}_{\hat{\mu},0,0}^{(n)}\right\}}_{\triangleq \left\{\tilde{\boldsymbol{\xi}}_{\mathrm{all}}^{(n)}\right\}_0}, \ldots, \\ &\qquad \underbrace{\left\{\tilde{\boldsymbol{\xi}}_{1,U_{\mathrm{R},i}-1,U_{\mathrm{T},i}-1}^{(n)}, \cdots, \tilde{\boldsymbol{\xi}}_{\hat{\mu},U_{\mathrm{R},i}-1,U_{\mathrm{T},i}-1}^{(n)}\right\}}_{\triangleq \left\{\tilde{\boldsymbol{\xi}}_{\mathrm{all}}^{(n)}\right\}_{S-1}} \Big\},\end{aligned} \tag{37}$$

whose cardinality is $\left|\tilde{\boldsymbol{\xi}}_{\mathrm{all}}^{(n)}\right| = \prod_{i=1}^{\hat{\mu}} U_{\mathrm{R},i}U_{\mathrm{T},i}$.

Let $s \in \{0, 1, \ldots, S-1\}$ denote the index of the element for the set $\tilde{\boldsymbol{\xi}}_{\mathrm{all}}^{(n)}$, where S is the cardinality of the set $S \triangleq \left|\tilde{\boldsymbol{\xi}}_{\mathrm{all}}^{(n)}\right|$. The s-th element of the set $\left\{\tilde{\boldsymbol{\xi}}_{\mathrm{all}}^{(n)}\right\}_s$ is also the set consisting of AoA and AoD subset pairs for all tracking paths $\tilde{\boldsymbol{\xi}}_{i,u,u'}^{(n)}$, whose cardinality is $\left|\left\{\tilde{\boldsymbol{\xi}}_{\mathrm{all}}^{(n)}\right\}_s\right| = \mu$. Since the union of AoA and AoD subsets $\tilde{\boldsymbol{\xi}}_{i,u,u'}^{(n)}$ consists of AoA and AoD candidates in subsets in (33) and (34), AoA and/or AoD candidates of all or some tracking paths in the s'-th set $\left\{\tilde{\boldsymbol{\xi}}_{\mathrm{all}}^{(n)}\right\}_{s'} (s' \neq s)$ differ from those in the s-th set, as shown in Fig. 4. In this figure, all the sets with $Q_{\mathrm{T},i} = Q_{\mathrm{R},i} = 8, \forall i$ candidates are split into $U_{\mathrm{R},i} = U_{\mathrm{T},i} = 2, \forall i$ subsets with $q_{\mathrm{T},i} = q_{\mathrm{R},i} = 4, \forall i$ candidates to track $\hat{\mu} = 2$ paths, resulting in $S = \prod_{i=1}^{2} U_{\mathrm{R},i}U_{\mathrm{T},i} = 16$ sub-array dictionaries and sub-sensing matrices. The $s = 0$-th set includes the $u = 0$-th subset for AoA at the $i = 1$-st tracking path, whereas the $s = 1$-st set includes the $u = 1$-st one.

The sub-array dictionary in the s-th element can be defined as $\tilde{\boldsymbol{\Psi}}\left(\left\{\tilde{\boldsymbol{\xi}}_{\mathrm{all}}^{(n)}\right\}_s\right) \in \mathbb{C}^{N_{\mathrm{t}}N_{\mathrm{r}} \times E_{\mathrm{sub}}}$, where $E_{\mathrm{sub}} \in \mathbb{N}^+$ denotes the total number of AoA-AoD combinations in the s-th element, which is given as $E_{\mathrm{sub}} = \sum_{i=1}^{\hat{\mu}} q_{\mathrm{R},i}q_{\mathrm{T},i}$. With sub-array dictionaries $\tilde{\boldsymbol{\Psi}}\left(\left\{\tilde{\boldsymbol{\xi}}_{\mathrm{all}}^{(n)}\right\}_s\right), \forall s$, sensing matrix $\mathbf{B}^{(n)}$ can be split into S sub-sensing matrices $\mathbf{B}_s^{(n)} \triangleq \boldsymbol{\Phi}^{(n)}\tilde{\boldsymbol{\Psi}}\left(\left\{\tilde{\boldsymbol{\xi}}_{\mathrm{all}}^{(n)}\right\}_s\right) \in \mathbb{C}^{M_{\mathrm{tck}}N_{\mathrm{s}} \times E_{\mathrm{sub}}}, \forall s$, whose mutual coherence is less than that of the original.

Using sub-sensing matrices, the optimization problem in (32) with a fixed $\hat{\mathbf{G}}_{\mathrm{D}}^{(n)}$ is reduced to S sub-problems, each given as

$$\underset{\tilde{\mathbf{G}}_{\mathrm{D},s}^{(n)\mathrm{T}}}{\text{minimize}} \ \frac{1}{2}\left\|\mathbf{B}_s^{(n)}\tilde{\mathbf{G}}_{\mathrm{D},s}^{(n)\mathrm{T}}(\mathbf{PV})^{\mathrm{T}} - \mathbf{Y}_{\mathrm{tr}}^{(n)}\right\|_{\mathrm{F}}^2 \tag{38a}$$

$$\text{subject to} \ \sum_{\ell \in \mathcal{J}_i} \left\|\left[\tilde{\mathbf{G}}_{\mathrm{D},s}^{(n)\mathrm{T}}\right]_\ell\right\|_{2,0} = 1, \forall i, \tag{38b}$$

where $\tilde{\mathbf{G}}_{\mathrm{D},s}^{(n)\mathrm{T}} \in \mathbb{C}^{E_{\mathrm{sub}} \times D}$ denotes the row-sparse path gain matrix corresponding to the s-th sub-sensing matrix, and $\mathcal{J}_i \triangleq \{\sum_{i'=1}^{i} q_{\mathrm{R},i'-1}q_{\mathrm{T},i'-1} + 1, \sum_{i'=1}^{i} q_{\mathrm{R},i'-1}q_{\mathrm{T},i'-1} + 2, \ldots, \sum_{i'=1}^{i} q_{\mathrm{R},i'}q_{\mathrm{T},i'}\}$ denotes the index set of rows corresponding to all AoA and AoD candidates at the i-th tracking path, where $q_{\mathrm{R},0} = q_{\mathrm{T},0} = 0$.

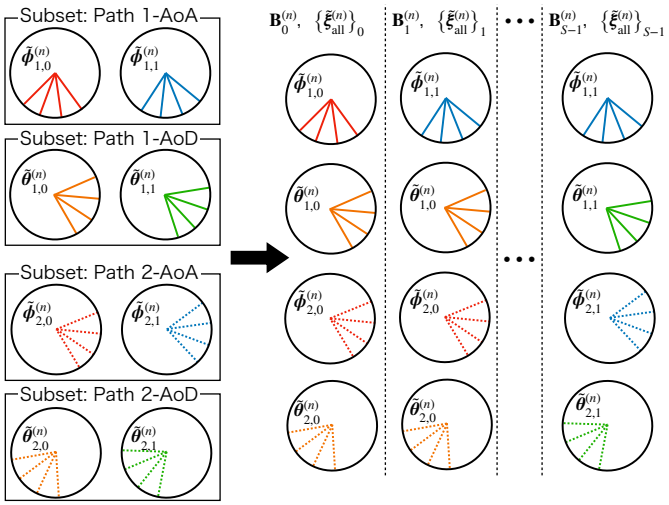


Fig. 4: Example of sub-sensing matrices with $\hat{\mu} = 2$ tracking paths.

C. AoA and AoD Estimation via Consensus

Each sub-problem in (38) can be solved using Algorithm 1. In the exact top-$\hat{\mu}$ feature selection, the indices of the nonzero elements in the matrix $\tilde{\mathbf{G}}_{\mathrm{D},s}^{(n)\mathrm{T}}$ depend on the first iteration of the algorithm. Thus, to obtain only the estimated AoAs and AoDs, it is sufficient to determine the nonzero elements using (25) and (27) with $\mathbf{B}_s^{(n)}$ and $\tilde{\mathbf{G}}_{\mathrm{D},s}^{(n)\mathrm{T}}$ without iterations.

Let $\hat{\phi}_{i,s}^{(n)} \in \mathbb{R}$ and $\hat{\theta}_{i,s}^{(n)} \in \mathbb{R}$ denote the estimated AoA and AoD at the i-th tracking path and the n-th channel slot obtained from the s-th sub-problem, respectively. The estimated AoA and AoD for the channel at the n-th channel slot can be obtained as

$$\hat{\phi}_i^{(n)} = \sum_{s=1}^{S} \omega_s^{(n)} \hat{\phi}_{i,s}^{(n)}, \qquad \hat{\theta}_i^{(n)} = \sum_{s=1}^{S} \omega_s^{(n)} \hat{\theta}_{i,s}^{(n)}, \tag{39}$$

where $\omega_i^{(n)} \in \mathbb{R}^+$ denotes a weight of the s-th estimator.

From (39) it can be seen that the AoA and AoD estimation accuracy depends on both the consensus processes (*i.e.*, weight calculation) and the reconstruction performance of each sub-problem.

In what follows, a weight calculation based on the conditional PDF of the received signal is introduced as one possible solution to this problem. To calculate the PDF easier, the vectorized received signal $\mathbf{y}_{\mathrm{tr}}^{(n)} \triangleq \mathrm{vec}\{\mathbf{Y}_{\mathrm{tr}}^{(n)}\} \in \mathbb{C}^{PM_{\mathrm{tck}}N_{\mathrm{s}} \times 1}$ is considered and given as

$$\mathbf{y}_{\mathrm{tr}}^{(n)} = \left(\mathbf{I}_P \otimes \mathbf{\Phi}^{(n)} \mathbf{\Psi}\left(\boldsymbol{\xi}^{(n)}\right)\right)(\mathbf{P}\mathbf{V} \otimes \mathbf{I}_\mu)\mathbf{g}_{\mathrm{D}}^{(n)} + \mathbf{z}_P^{(n)}, \tag{40}$$

where $\mathbf{g}_{\mathrm{D}}^{(n)} \triangleq \mathrm{vec}\{\mathbf{G}_{\mathrm{D}}^{(n)\mathrm{T}}\} \in \mathbb{C}^{D\mu \times 1}$ and $\mathbf{z}_P^{(n)} \triangleq \mathrm{vec}\{\mathbf{Z}_P^{(n)}\} \in \mathbb{C}^{PM_{\mathrm{tck}}N_{\mathrm{s}} \times 1}$. Note that if the pilot symbols are allocated for all subcarriers $\mathcal{P} = \mathcal{K}$, $\mathbf{P} = \mathbf{I}_K$ and $(\mathbf{V} \otimes \mathbf{I}_\mu)\mathbf{g}_{\mathrm{D}}^{(n)} = \mathbf{g}^{(n)}$ holds.

Let $\hat{\mathcal{A}}_{\mathrm{s}}^{(n)} \triangleq \{\hat{\boldsymbol{\xi}}_s^{(n')}\}_{n'=0}^{n}$ and $\mathcal{Y}^{(n)} \triangleq \{\mathbf{y}_{\mathrm{tr}}^{(n')}\}_{n'=0}^{n}$ denote the set of angle estimates at the s-th sub-problem and received signals over the channel slots, respectively, where the set of angle estimates at the n-th channel slot is defined as $\hat{\boldsymbol{\xi}}_s^{(n)} \triangleq \{\hat{\phi}_{1,s}^{(n)}, \ldots, \hat{\phi}_{\hat{\mu},s}^{(n)}, \hat{\theta}_{1,s}^{(n)}, \ldots, \hat{\theta}_{\hat{\mu},s}^{(n)}\}$. The conditional PDF of the observation over the channel slots given AoA and AoD estimates from the s-th sub-problem corresponds to the reliability of the estimates. Hence, the weight for the s-th estimator can be calculated as

$$\omega_s^{(n)} = \frac{p(\mathbf{y}_{\mathrm{tr}}^{(n)} \mid \hat{\mathcal{A}}_s^{(n)}, \mathcal{Y}^{(n-1)})}{\sum_{s=1}^{S} p(\mathbf{y}_{\mathrm{tr}}^{(n)} \mid \hat{\mathcal{A}}_s^{(n)}, \mathcal{Y}^{(n-1)})}. \tag{41}$$

Since the time evolution of the path gain follows a Gaussian distribution, the conditional PDF of the estimated path gain is given by

$$p(\hat{\mathbf{g}}_{\mathrm{D}}^{(n)} \mid \mathcal{G}_{\mathrm{D}}^{(n-1)}) = \mathcal{CN}(\hat{\mathbf{g}}_{\mathrm{D}}^{(n|n-1)}, \hat{\mathbf{C}}_{\mathrm{D},\mathbf{gg}}^{(n|n-1)}),$$

where $\mathcal{G}_{\mathrm{D}}^{(n)} \triangleq \{\mathbf{g}_{\mathrm{D}}^{(n')}\}_{n'=0}^{n}$ denotes the set of path gain estimates over channel slots; the vector $\hat{\mathbf{g}}_{\mathrm{D}}^{(n|n-1)} \in \mathbb{C}^{D\hat{\mu}\times 1}$ and matrix $\hat{\mathbf{C}}_{\mathrm{D},\mathbf{gg}}^{(n|n-1)} \in \mathbb{C}^{D\hat{\mu}\times D\hat{\mu}}$ denote the prediction value of the path gain and covariance matrix of the prediction errors at the n-th channel slot, respectively, calculated as

$$\hat{\mathbf{g}}_{\mathrm{D}}^{(n|n-1)} = \hat{\mathbf{R}}_{\mathrm{D}} \hat{\mathbf{g}}_{\mathrm{D}}^{(n-1)}, \tag{42}$$

$$\hat{\mathbf{C}}_{\mathrm{D},\mathbf{gg}}^{(n|n-1)} = \hat{\mathbf{R}}_{\mathrm{D}} \hat{\mathbf{C}}_{\mathrm{D},\mathbf{gg}}^{(n-1)} \hat{\mathbf{R}}_{\mathrm{D}}^{\mathrm{H}} + \left(\mathbf{I}_{\hat{\mu}D} - \hat{\mathbf{R}}_{\mathrm{D}} \hat{\mathbf{R}}_{\mathrm{D}}^{\mathrm{H}}\right) \hat{\mathbf{C}}_{\mathrm{D},\mathbf{gg}}^{(0)}, \tag{43}$$

where $\hat{\mathbf{R}}_{\mathrm{D}} \in \mathbb{C}^{D\hat{\mu}\times D\hat{\mu}}$ and $\hat{\mathbf{C}}_{\mathrm{D},\mathbf{gg}}^{(n)} \in \mathbb{C}^{D\hat{\mu}\times D\hat{\mu}}$ denotes the correlation matrix and covariance matrix of the path gain in the delay domain, respectively, given as $\hat{\mathbf{R}}_{\mathrm{D}} = (\mathbf{V}^{\mathrm{H}} \otimes \mathbf{I}_{\hat{\mu}})\hat{\mathbf{R}}(\mathbf{V} \otimes \mathbf{I}_{\hat{\mu}})$ and $\hat{\mathbf{C}}_{\mathrm{D},\mathbf{gg}}^{(n)} = (\mathbf{V}^{\mathrm{H}} \otimes \mathbf{I}_{\hat{\mu}})\hat{\mathbf{C}}_{\mathbf{gg}}^{(n)}(\mathbf{V} \otimes \mathbf{I}_{\hat{\mu}})$.

Given estimated angles $\hat{\boldsymbol{\xi}}_s^{(n)}$, the matrix $\hat{\mathbf{T}}_{\mathrm{D},s}^{(n)} \triangleq \left(\mathbf{I}_P \otimes \mathbf{\Phi}^{(n)} \mathbf{\Psi}\left(\hat{\boldsymbol{\xi}}_s^{(n)}\right)\right)(\mathbf{P}\mathbf{V} \otimes \mathbf{I}_{\hat{\mu}}) \in \mathbb{C}^{PM_{\mathrm{tck}}N_{\mathrm{s}} \times D\hat{\mu}}$ can be considered a transfer function for the received signal $\mathbf{y}_{\mathrm{tr}}^{(n)}$ and the path gain. Therefore, the conditional PDF $p(\mathbf{y}_{\mathrm{tr}}^{(n)} \mid \hat{\mathcal{A}}_s^{(n)}, \mathcal{Y}^{(n-1)})$ is given by

$$p(\mathbf{y}_{\mathrm{tr}}^{(n)} \mid \hat{\mathcal{A}}_s^{(n)}, \mathcal{Y}^{(n-1)}) \tag{44}$$
$$= \mathcal{CN}\left(\hat{\mathbf{T}}_{\mathrm{D},s}^{(n)} \hat{\mathbf{g}}_{\mathrm{D}}^{(n|n-1)}, \hat{\mathbf{T}}_{\mathrm{D},s}^{(n)} \hat{\mathbf{C}}_{\mathrm{D},\mathbf{gg}}^{(n|n-1)} \hat{\mathbf{T}}_{\mathrm{D},s}^{(n)\mathrm{H}} + \mathbf{I}_P \otimes \mathbf{C}_{\mathbf{zz}}^{(n)}\right).$$

D. Path Gain Estimation via Kalman Filter

After obtaining the estimated AoAs and AoDs $\hat{\boldsymbol{\xi}}^{(n)}$, what is left is the estimation of the path gain $\hat{\mathbf{G}}_{\mathrm{D}}$, which can be obtained by solving the optimization problem in (32) for a fixed $\tilde{\mathbf{G}}_{\mathrm{D}}^{(n)\mathrm{T}}$, namely,

$$\underset{\hat{\mathbf{g}}_{\mathrm{D}}^{(n)}}{\text{minimize}}\ \mathbb{E}\left[\left\|\mathbf{g}_{\hat{\mu}}^{(n)} - \hat{\mathbf{g}}_{\mathrm{D}}^{(n)}\right\|_{\mathrm{F}}^2\right], \tag{45}$$

where $\mathbf{g}_{\hat{\mu}}^{(n)} \triangleq \mathrm{vec}\{\mathbf{G}_{\hat{\mu}}^{(n)\mathrm{T}}\}$ and $\hat{\mathbf{g}}_D^{(n)} \triangleq \mathrm{vec}\{\mathbf{G}_{\mathrm{D}}^{(n)\mathrm{T}}\}$, and note that well-known MMSE estimation methods cannot solve the optimization problem (45) due to the actual path gain.

However, the variable $\hat{\mathbf{g}}_{\mathrm{D}}^{(n)}$, its time fluctuation, and measurement noise $\mathbf{z}_P^{(n)}$ follow Gaussian distributions. Moreover, given the estimated AoAs and AoDs, the relationship between the received signal $\mathbf{y}_{\mathrm{tr}}^{(n)}$ and path gain $\mathbf{g}_{\hat{\mu}}^{(n)}$ is linear with respect to the transfer function $\hat{\mathbf{T}}_{\mathrm{D}}^{(n)} \triangleq (\mathbf{I}_P \otimes \mathbf{\Phi}^{(n)} \mathbf{\Psi}(\hat{\boldsymbol{\xi}}^{(n)}))(\mathbf{P}\mathbf{V} \otimes \mathbf{I}_{\hat{\mu}})$. Therefore, the optimal MMSE estimator of the path gain can be obtained using KF with optimal

Algorithm 2 Channel tracking with sub-sensing matrices

Input: $\mathbf{Y}_{\rm tr}^{(n)}, \mathbf{\Phi}^{(n)}, \hat{\boldsymbol{\xi}}^{(n-1)}, \sigma_{{\rm A},\hat{\nu}}^2 \sigma_{{\rm D},\hat{\nu}}^2 \hat{\mathbf{g}}_{\rm D}^{(n-1)}, \hat{\mathbf{C}}_{{\rm D},\mathbf{gg}}^{(n-1)}, \hat{\mathbf{R}}_{\rm D}$
Output: $\hat{\mathbf{H}}^{(n)}[k], \forall k$

Generate candidates $\tilde{\boldsymbol{\phi}}^{(n)}$, $\tilde{\boldsymbol{\theta}}^{(n)}$ following (15a), (15b)
KF prediction following (42), (43)
for $s = 1, 2, \ldots, S$ **do**
  Design sensing matrix $\mathbf{B}_s^{(n)} \triangleq \mathbf{\Phi}^{(n)} \tilde{\mathbf{\Psi}}\big(\{\tilde{\boldsymbol{\xi}}_{\rm all}^{(n)}\}_s\big)$
  Initialize $\mathbf{L}_s = \mathbf{0} \in \mathbb{C}^{E_{\rm sub} \times D}$
  Solve $\tilde{\mathbf{G}}_{{\rm D},s}^{(n){\rm T}}$ following (25) with $\mathbf{B}_s^{(n)}$
  for $i = 1, 2, \ldots, \hat{\mu}$ **do**
    Determine $[\mathbf{L}_s]_e$, $\forall e \in \mathcal{J}_i$ following (27)
    Obtain $\ell^\star = \underset{\ell, \forall \ell \in \mathcal{J}_i}{\rm argmax} \big\| [\mathbf{L}_s]_\ell \big\|$, $\forall \ell \in \mathcal{J}_i$
    Obtain $\hat{\phi}_{i,s}^{(n)}$ following (29) with $\{\tilde{\boldsymbol{\xi}}_{\rm all}^{(n)}\}_s$, $q_{{\rm R},i}$, $q_{{\rm T},i}$
    Obtain $\hat{\theta}_{i,s}^{(n)}$ following (30) with $\{\tilde{\boldsymbol{\xi}}_{\rm all}^{(n)}\}_s$, $q_{{\rm R},i}$, $q_{{\rm T},i}$
  end for
  Calculate the PDF $p\big(\mathbf{y}_{\rm tr}^{(n)} \mid \hat{\mathcal{A}}_s^{(n)}, \mathcal{Y}_{\rm tr}^{(n-1)}\big)$
end for
Calculate weights $\omega_s^{(n)}$, $i = 1, \ldots, S$ following (41)
Determine the estimated angles $\hat{\boldsymbol{\xi}}^{(n)}$ following (39).
KF measurement update $\hat{\mathbf{g}}_{\rm D}^{(n)}$ following (46), (47), (48)
Estimated CSI $\hat{\mathbf{h}}^{(n)} = \big(\mathbf{I}_P \otimes \hat{\mathbf{\Psi}}\big(\hat{\boldsymbol{\xi}}^{(n)}\big)\big)(\mathbf{V} \otimes \mathbf{I}_{\hat{\mu}}) \hat{\mathbf{g}}_{\rm D}^{(n)}$
Reshape $\big[\hat{\mathbf{H}}^{(n)}[0], \ldots, \hat{\mathbf{H}}^{(n)}[K-1]\big] = {\rm unvec}\big\{\hat{\mathbf{h}}^{(n)}\big\}, \forall k$

Kalman gain and without the explicit expression of $\mathbf{g}_{\hat{\mu}}^{(n)}$, and the estimation procedures are expressed as

$$\mathbf{K}_{\rm D}^{(n)} = \hat{\mathbf{C}}_{{\rm D},\mathbf{gg}}^{(n|n-1)} \hat{\mathbf{T}}_{\rm D}^{(n){\rm H}} \big(\hat{\mathbf{T}}_{\rm D}^{(n)} \hat{\mathbf{C}}_{{\rm D},\mathbf{gg}}^{(n|n-1)} \hat{\mathbf{T}}_{\rm D}^{(n){\rm H}} + \mathbf{I}_P \otimes \mathbf{C}_{\mathbf{zz}}^{(n)}\big)^{-1}, \tag{46}$$

$$\hat{\mathbf{g}}_{\rm D}^{(n)} = \hat{\mathbf{g}}_{\rm D}^{(n|n-1)} + \mathbf{K}_{\rm D}^{(n)} \big(\mathbf{y}^{(n)} - \hat{\mathbf{T}}_{\rm D}^{(n)} \hat{\mathbf{g}}_{\rm D}^{(n|n-1)}\big), \tag{47}$$

$$\hat{\mathbf{C}}_{{\rm D},\mathbf{gg}}^{(n)} = \hat{\mathbf{C}}_{{\rm D},\mathbf{gg}}^{(n|n-1)} - \mathbf{K}_{\rm D}^{(n)} \hat{\mathbf{T}}_{\rm D}^{(n)} \hat{\mathbf{C}}_{{\rm D},\mathbf{gg}}^{(n|n-1)}, \tag{48}$$

where $\mathbf{K}_{\rm D}^{(n)} \in \mathbb{C}^{D\hat{\mu} \times DM_{\rm tck}N_{\rm s}}$ denotes the Kalman gain.

Combining multiple sub-sensing matrices and Kalman filtering with Algorithm 1 yields the pseudocode in Algorithm 2, where the previously estimated path gain $\hat{\mathbf{g}}_{\rm D}^{(n-1)}$ is set to zero at the $n = 1$-th channel slot, that is, $\hat{\mathbf{g}}_{\rm D}^{(n-1)} = \hat{\mathbf{g}}_{\rm D}^{(0)} = \mathbf{0}$.

VI. Performance Assessment

In this section, thorough performance assessments are provided for the proposed methods.

A. Floating Operations

In order to assess the effect of the proposed complexity reductions quantitatively, the computational complexity of the channel tracking algorithms is evaluated by FLOPs, including the number of multiplications.

1) State-of-the-Art Method: For the conventional algorithm based on GMPF proposed in [37], the evaluations of particles for AoAs and AoDs via PF and path gain estimation via KF are the most computationally expensive. The complexity of these operations are proportional to the cube of the number of available subcarriers K and training overhead $M_{\rm tck}N_{\rm s}$. Unlike the proposed algorithms, this algorithm estimates the path gain for all subcarriers and estimates it for all resampled particles. Let $N_{\rm PF} \in \mathbb{N}^+$ and $N'_{\rm PF} (\leq N_{\rm PF})$ denote the numbers of particles and resampled particles, respectively. The total number of FLOPs required for the conventional algorithm is approximately given as

$$\begin{aligned} F_{\rm conv} \approx & N_{\rm PF} \big\{K^3 M_{\rm tck}^2 N_{\rm s}^2 (\hat{\mu} + 2M_{\rm tck}N_{\rm s}) + M_{\rm tck}N_{\rm s}N_{\rm t}N_{\rm r}\hat{\mu}\big\} \\ & + N'_{\rm PF} \big\{K^3 \hat{\mu} (M_{\rm tck}^2 N_{\rm s}^2 + \hat{\mu} M_{\rm tck}N_{\rm s} + 2\hat{\mu}^2)\big\}, \end{aligned} \tag{49}$$

where $M_{\rm tck}N_{\rm s}N_{\rm t}N_{\rm r}\hat{\mu}$ denotes the number of FLOPs required to calculate the transfer function in the KF update step. In this algorithm, the number of resampled particles $N'_{\rm PF}$ is not known in advance. Thus, the total number of FLOPs is calculated assuming $N'_{\rm PF} = 1$ to be the lower bound through performance evaluations.

2) One sensing matrix: The proposed sensing matrix design requires $F_{\rm sens} = M_{\rm tck}N_{\rm s}N_{\rm t}N_{\rm r}E$ FLOPs. For Algorithm 1, an E-dimensional inverse matrix must be calculated to solve the sparse matrix $\tilde{\mathbf{G}}_{\rm D}^{(n){\rm T}}$ following (25), once before iterations. Each iteration requires one multiplication in (25) and l_2 norm calculations for all rows in the E-by-D matrix $\mathbf{X}$ in (27). Hence, the total number of FLOPs required for Algorithm 1 is given by

$$\begin{aligned} F_{\rm top} = & F_{\rm sens} + E^3 + E^2 (M_{\rm tck}N_{\rm s} + T_{\rm top}D) \\ & + E\{3T_{\rm top}D + P(D + M_{\rm tck}N_{\rm s})\} + 1. \end{aligned} \tag{50}$$

This algorithm achieves linear-order computational complexity with respect to both the number of pilot symbols $M_{\rm tck}$ and pilot subcarriers P. Although the complexity is proportional to the cube of the total number of candidates E, it can be kept small $E^3 < K^3 M_{\rm tck}^3 N_{\rm s}^3$ because of the limitations of the quantization intervals for the angle candidates in the proposed sensing matrix design.

3) Multiple sub-sensing matrices: The design of the sub-sensing matrices requires $F_{\rm sub} = SM_{\rm tck}N_{\rm s}N_{\rm t}N_{\rm r}E_{\rm sub}$ FLOPs. For Algorithm 2, weight calculations in (41) and path gain estimation via KF are the most computationally expensive. In contrast to the conventional algorithm, KF estimates the path gain for all delay taps only once. Therefore, the total number of FLOPs required for Algorithm 2 is given by

$$\begin{aligned} F_{\rm multi} \approx & F_{\rm sub} + SE_{\rm sub}^3 + (2S+1)P^3 M_{\rm tck}^3 N_{\rm s}^3 + 2D^3\hat{\mu}^3 \\ & + M_{\rm tck}N_{\rm s}N_{\rm t}N_{\rm r}\hat{\mu} + (S+2)P^2 D M_{\rm tck}^2 N_{\rm s}^2 \hat{\mu} \\ & + (S + 1 + \hat{\mu}) P D^2 M_{\rm tck}N_{\rm s}\hat{\mu}. \end{aligned} \tag{51}$$

It is evident from (50) and (51) that the total number of FLOPs for the proposed algorithms is proportional to the number of pilot subcarriers P and not to the available subcarriers K.

B. Other Compressive Sensing Algorithms

Well-known algorithms, such as orthogonal matching pursuit (OMP) [21]–[23] and sparse Bayesian learning (SBL) [25], can be applied to solve the proposed compressive sensing problem in (23).

However, in OMP, to ensure orthogonality, the constraint for the unique sparsity in (22) should be relaxed into the constraint for the standard row-sparsity

$$\left\|\tilde{\mathbf{G}}_{\mathrm{D}}^{(n)\mathrm{T}}\right\|_{2,0} = \hat{\mu}. \tag{52}$$

The effectiveness of the proposed sensing matrix design is vanished by this relaxation. In other words, the channel tracking based on OMP cannot ensure the updates of the path gain, AoAs, and AoDs for all tracking paths, leading to poor or no tracking gain.

For SBL, although the unique sparse structure in (22) can be considered, multiple inverse matrix calculations are required to update unknown variables. Specifically, in the considered system, block SBL [25], which is an iterative algorithm to estimate the block-sparse signals, requires inverse matrices with dimensions E and $M_{\mathrm{tck}}N_{\mathrm{s}}$ for each iteration. Therefore, if the number of iterations required by SBL becomes large, the complexity of channel tracking based on compressive sensing is comparable with or higher than that based on GMPF [37].

In contrast to SBL, the proposed approaches ensure a lower complexity than GMPF [37] if the number of sub-sensing matrix S is less than the number of particles N_{PF}. From all the above, in the following subsections, only the channel tracking methods based on the proposed algorithms and GMPF are evaluated to focus on the trade-off in terms of channel tracking gain and complexity. The choice of the optimal compressive sensing algorithm remains as our future work.

C. Simulation Setup

In what follows, the tracking gains for the conventional and proposed channel tracking methods are evaluated via computer simulations.

Consider an uplink cellular system with a cell radius of 100 [m] operating at a carrier frequency of $f_{\mathrm{c}} = 60$ [GHz] with $K = 36$ subcarriers and $W = 960$ [kHz] subcarrier spacing[3]. The sampling duration T_{s} is calculated as $T_{\mathrm{s}} = 1/(WK) = 28.9$ [ns]. The time delays are generated following $\tau_c^{(n)} \sim \mathcal{U}_{\mathrm{cont}}[0, (D-1)T_{\mathrm{s}}]$ as in [22], [23], where $D = K/4 = 9$ and $\mathcal{U}_{\mathrm{cont}}$ denotes a continuous uniform distribution. The cyclic prefix length is also set to D, leading to the OFDM symbol duration $T_{\mathrm{sys}} = (K + D)T_{\mathrm{s}} = 1.30$ [μs]. The number of OFDM symbols M_{total} in each channel slot, including both pilot and data symbols, is assumed to be $M_{\mathrm{total}} = 110$. The AWGN variation at each subcarrier is modeled as

$$10\log_{10}(\sigma_u^2) = 10\log_{10}(1000\kappa T) + 10\log_{10}(W) + \mathrm{NF}, \tag{53}$$

where κ denotes the Boltzmann constant, $T = 293.15$ [K] is the physical temperature, and $\mathrm{NF} = 5$ [dB] is the noise figure.

At the initial channel slot $n = 0$, it is assumed that the composed path from each cluster is estimated by a certain mmWave channel estimation method with $M_{\mathrm{ini}} = 100$ pilot symbols. Note that the initial mmWave channel estimation can be performed by developed estimation methods in the literature such as [23], [24]. Thus, the estimated AoAs and AoDs in the initial channel slot are $\hat{\phi}_c^{(0)} = \frac{1}{R_c}\sum_{r=1}^{R_c}\phi_{c,r}^{(0)}$ and $\hat{\theta}_c^{(0)} = \frac{1}{R_c}\sum_{r=1}^{R_c}\theta_{c,r}^{(0)}$, respectively. In turn, the estimated covariance matrix $\hat{\mathbf{C}}_{\mathrm{gg}}^{(0)}$ for the path gain in the frequency domain is given as $\hat{\mathbf{C}}_{\mathrm{gg}}^{(0)} = \mathbf{V}\mathbf{V}^* \otimes \frac{1}{\mu}\mathrm{diag}\{\mathbf{c}_{\mathrm{PL}}^{\mathrm{T}}\}$, where the vector $\mathbf{c}_{\mathrm{PL}}$ contains the path loss components for composite paths defined by $\mathbf{c}_{\mathrm{PL}} \triangleq [R_1\mathrm{PL}_1, \ldots, R_C\mathrm{PL}_C]^{\mathrm{T}} \in \mathbb{R}^{C\times 1}$. Consequently, the number of tracking paths is $\hat{\mu} = C = 3$.

The other parameters and assumptions for the time fluctuations, channel parameters, available antenna elements, and transmit power are similar to those in [37], and are shown in Table II, where $\mathcal{U}$ denotes a discrete uniform distribution. The root-raised cosine roll-off filter is used for pulse shaping similar to [37]. Then, the equivalent pulse shaping response $p(t)$ is given by the raised cosine roll-off filter whose expression is

$$p(t) = \mathrm{sinc}\left(\tfrac{t}{T_{\mathrm{sys}}}\right)\frac{\cos\left(\frac{\pi\alpha}{T_{\mathrm{sys}}}t\right)}{1-\left(\frac{2\alpha}{T_{\mathrm{sys}}}t\right)^2}, \tag{54}$$

where $\mathrm{sinc}(\cdot)$ denotes the normalized sinc function, and the roll-off factor $\alpha \in [0, 1]$ is set to 0.3.

The tracking gain is evaluated in terms of the sum spectral efficiency over subcarriers. At each channel slot ($n \geq 1$), the sum spectral efficiency is defined as

$$\mathrm{SE}^{(n)} \triangleq \sum_{k=0}^{K-1}\eta[k]\log_2\left|\mathbf{I}_{N_{\mathrm{s}}} + \mathbf{H}_{\mathrm{eff}}^{(n)}[k]\mathbf{H}_{\mathrm{eff}}^{(n)\mathrm{H}}[k]{\mathbf{C}_{\mathbf{ww}}^{(n)}}^{-1}[k]\right|, \tag{55}$$

where $\mathbf{C}_{\mathbf{ww}}^{(n)}[k] \triangleq \sigma^2\mathbf{W}^{(n)\mathrm{H}}[k]\mathbf{W}^{(n)}[k]$ and $\mathbf{H}_{\mathrm{eff}}^{(n)}[k] \triangleq \mathbf{W}^{(n)\mathrm{H}}[k]\mathbf{H}^{(n)}[k]\mathbf{F}^{(n)}[k]$. In the data transmission, block level singular value decomposition (SVD) precoding $\mathbf{F}^{(n)}[k] \in \mathbb{C}^{N_{\mathrm{t}}\times N_{\mathrm{s,data}}}$ and combining $\mathbf{W}^{(n)}[k] \in \mathbb{C}^{N_{\mathrm{r}}\times N_{\mathrm{s,data}}}$ based on the estimated CSI $\hat{\mathbf{H}}^{(n)}[k], \forall k$ are performed for each subcarrier, where $N_{\mathrm{s,data}} \in \mathbb{N}^+$ denotes the number of data streams in the data transmission. In the computer simulations, the number of data streams is assumed to be $N_{\mathrm{s,data}} = 3$. For precoding, the power allocation is based on water filling. Scalar $\eta \in \mathbb{R}^+$ denotes the data transmission efficiency provided by

$$\eta[k] = \begin{cases} \frac{1}{M_{\mathrm{total}}}(M_{\mathrm{total}} - M_{\mathrm{tck}}), & \text{if } k \in \mathcal{P} \\ 1, & \text{otherwise} \end{cases}. \tag{56}$$

D. Comparison to State-of-the-Art Methods

The conventional approach based on GMPF with $N_{\mathrm{PF}} = 50$ particles [37] is evaluated to offer the tracking gain achieved by the MMSE estimator of the channel.

For the proposed approaches, in the sensing matrix design, the number of candidates for each tracking path is set to $Q_{\mathrm{R},1} = Q_{\mathrm{T},i} = 8, \forall i$ and $Q_{\mathrm{R},i} = 1, (i = 2, \ldots, \hat{\mu})$, meaning $\sigma_{\mathrm{A},i} = 0$ and $\tilde{\phi}_{i,0}^{(n)} = \hat{\phi}_i^{(0)}$, $(i = 2, \ldots, \hat{\mu})$. Further parameters are $\Phi_i = \Theta_i = 4, \forall i$, which lead to a 99% confidence interval in the Laplace distribution. The pilot allocation is considered as $\mathcal{K}_P = \{1, 3, 5, \ldots, K-1\}$, $P = K/2 = 18$.

For the exact top-$\hat{\mu}$ feature selection in Algorithm 1, the hyperparameter $\rho = 1.02$ and the maximum number of iterations $T_{\mathrm{top}} = 100$ are considered.

[3]We assume only 36 subcarriers to consider the balance between the simulation time and reasonable setup. Note that because the total number of candidates E is independent of the number of subcarriers, the effectiveness of the complexity reduction by the proposed algorithms is more remarkable in practical situations employing more subcarriers.

For the proposed approach with multiple sub-sensing matrices in Algorithm 2, considering the trade-off between performance and complexity, only AoA and AoD candidates for the LOS component are split into two subsets, meaning $U_{\mathrm{R},1} = U_{\mathrm{T},1} = 2$ and $U_{\mathrm{R},i} = U_{\mathrm{T},i} = 1, (i = 2, 3)$. Consequently, $S = \prod_{i=1}^{\hat{\mu}} U_{\mathrm{R},i} \cdot U_{\mathrm{T},i} = 4$ sub-sensing matrices with $E_{\mathrm{sub}} = \sum_{i=1}^{\hat{\mu}} q_{\mathrm{R},i} q_{\mathrm{T},i} = 32$ candidate combinations are used for each sub-problem in (38).

The proposed approaches in the frequency domain are also evaluated to clarify the difference between compressive sensing and particle filter [37] and the effectiveness of delay-domain approaches. Specifically, frequency-domain channel tracking based on the top-$\hat{\mu}$ feature selection is performed by solving the optimization problem

$$\underset{\tilde{\mathbf{G}}_P^{(n)}}{\text{minimize}} \ \frac{1}{2}\left\|\mathbf{B}^{(n)}\tilde{\mathbf{G}}_P^{(n)} - \mathbf{Y}_{\mathrm{tr}}^{(n)}\right\|_{\mathrm{F}}^2, \tag{57a}$$

$$\text{subject to} \ \sum_{\ell \in \mathcal{L}_i} \left\|\left[\tilde{\mathbf{G}}_P^{(n)}\right]_{\ell}\right\|_{2,0} = 1, \forall i, \tag{57b}$$

with $\mathcal{K}_P = \mathcal{K}$ and $P = K$.

For the refined version of the proposed approach in the frequency domain, channel tracking is formulated as

$$\underset{\tilde{\mathbf{G}}_P^{(n)}, \hat{\mathbf{G}}_P^{(n)}}{\text{minimize}} \ \frac{1}{2}\left\|\mathbf{B}^{(n)}\tilde{\mathbf{G}}_P^{(n)} - \mathbf{Y}_{\mathrm{tr}}^{(n)}\right\|_{\mathrm{F}}^2 + \mathbb{E}\left[\left\|\mathbf{G}_{\hat{\mu},\mathrm{F}}^{(n)} - \hat{\mathbf{G}}_P^{(n)}\right\|_{\mathrm{F}}^2\right], \tag{58a}$$

$$\text{subject to} \ \sum_{\ell \in \mathcal{L}_i} \left\|\left[\tilde{\mathbf{G}}_P^{(n)}\right]_{\ell}\right\|_{2,0} = 1, \forall i, \tag{58b}$$

where $\mathbf{G}_{\hat{\mu},\mathrm{F}}^{(n)} \in \mathbb{C}^{\hat{\mu} \times K}$ and $\hat{\mathbf{G}}_P^{(n)} \in \mathbb{C}^{\hat{\mu} \times P}$ denote actual and estimated path gain matrices in the frequency domain, respectively.

In addition, to confirm the effect of the angle estimation for both conventional and proposed approaches, the path gain tracking, which tracks only the path gain for all subcarriers via KF and assumes constant AoAs and AoDs, as estimated in the first slot, is evaluated and formulated as

$$\underset{\hat{\mathbf{G}}_P^{(n)}}{\text{minimize}} \ \mathbb{E}\left[\left\|\mathbf{G}_{\hat{\mu},\mathrm{F}}^{(n)} - \hat{\mathbf{G}}_P^{(n)}\right\|_{\mathrm{F}}^2\right], \tag{59a}$$

$$\text{subject to} \ \hat{\phi}_i^{(n)} = \hat{\phi}_i^{(0)}, (i = 1, \ldots, \hat{\mu}), \tag{59b}$$

$$\hat{\phi}_i^{(n)} = \hat{\theta}_i^{(0)}, (i = 1, \ldots, \hat{\mu}). \tag{59c}$$

This approach is denoted as “Only Gain Tracking” in the following subsections.

E. Spectral Efficiency vs. Channel Slot

First, we evaluate the variation in the sum spectral efficiency with time, as shown in Fig. 5. In this article, solid lines with plus “+” markers and the solid line without markers show the performances for the cases of perfect tracking and without tracking, respectively. The frequency and delay domain channel tracking are denoted by lines with and without markers, respectively. Thus, the spectral efficiency difference between the standard solid line and these lines indicates the tracking gain. In this figure, the spectral efficiency at the initial channel slot $n = 0$ is lower than that at the others owing to the channel estimation with larger training overheads compared with channel tracking $M_{\mathrm{ini}} \ll M_{\mathrm{tck}}$. The number of pilot symbols is $M_{\mathrm{tck}} = 6$.

TABLE II: Simulation Parameters

Symbol	Parameter	Value
C	Num. of clusters	3
$R_c, c \neq 1$	Num. of rays within a cluster	$\mathcal{U} \sim [6, 20]$
T_{tck}	Length of each channel slot	0.1577 [ms]
v	Velocity of UE	30 [m/s]
$\sigma_{\mathrm{A},1,1}$ (LOS)	Time fluctuation of the AoA	0.99 [rad]
$\sigma_{\mathrm{A},c,r}$	Time fluctuation of AoAs	0 [rad]
$\sigma_{\mathrm{D},c,r}$	Time fluctuation of all AoDs	0.99 [rad]
M_{total}	Num. of OFDM symbols	110
M_{ini}	Num. of pilot symbols	100
(a, b, c)	Parameters to calculate $\mathbf{R}$	$\left(0.91, \frac{0.1}{\lambda}, 0\right)$
N_{r}	Num. of received antennas	64
N_{t}	Num. of transmit antennas	32
$N_{\mathrm{t,RF}}$	Num. of transmit RF chains	4
$N_{\mathrm{r,RF}}$	Num. of received RF chains	4
P	Transmit power	35 [dBm]

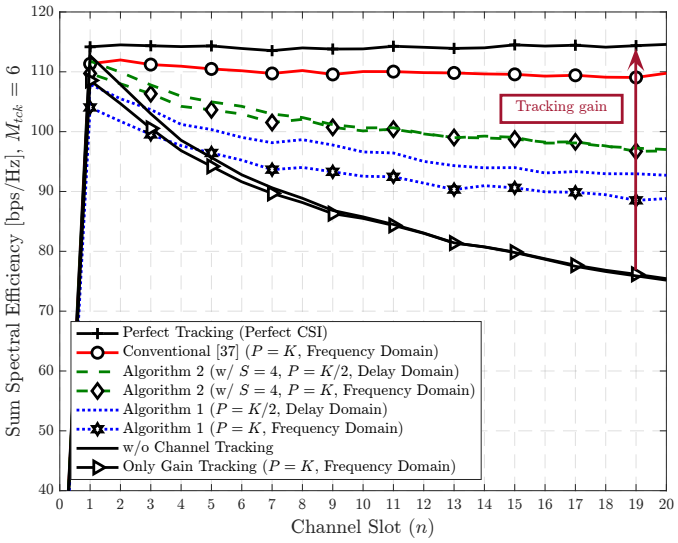


Fig. 5: Sum spectral efficiency as a function of channel slots

From the figure, the delay-domain channel tracking obtains a tracking gain comparable to that of the frequency domain, while reducing the number of FLOPs $K^3 \rightarrow P^3$, P^2D or PD^2. The comparisons between “Only Gain Tracking”, which is denoted by a solid line with triangles, and the proposed approach denoted by the dotted line and in Algorithm 1, confirm the effectiveness of AoAs and AoDs tracking by compressive sensing. Moreover, the proposed approach can estimate the path gain without KF and requires $F_{\mathrm{top}} \approx 1.09 \times 10^7$ FLOPs, which are about 100 times lower than that of the path gain tracking via KF, given as 8.20×10^8.

The conventional approach of [37], denoted by a solid line with circles, achieves the sum spectral efficiency closest to the perfect tracking case by AoA and AoD estimations based

on the particle filter. However, the total number of FLOPs required for this algorithm is $F_{\rm conv} \approx 6.86 \times 10^{10}$, which is 5×10^3 times higher than that of the proposed approach in Algorithm 1.

The proposed approach in Algorithm 2 with $S = 4$ sub-sensing matrices obtains higher tracking gain than the single-sensing approaches while maintaining 100 times lower complexity than the conventional approach $F_{\rm multi} \approx 7.59 \times 10^8 < F_{\rm conv} \approx 6.86 \times 10^{10}$. The results confirm the effectiveness of the proposed sub-sensing approach, consensus process, and incorporation of KF.

F. Spectral Efficiency vs. Training Overhead

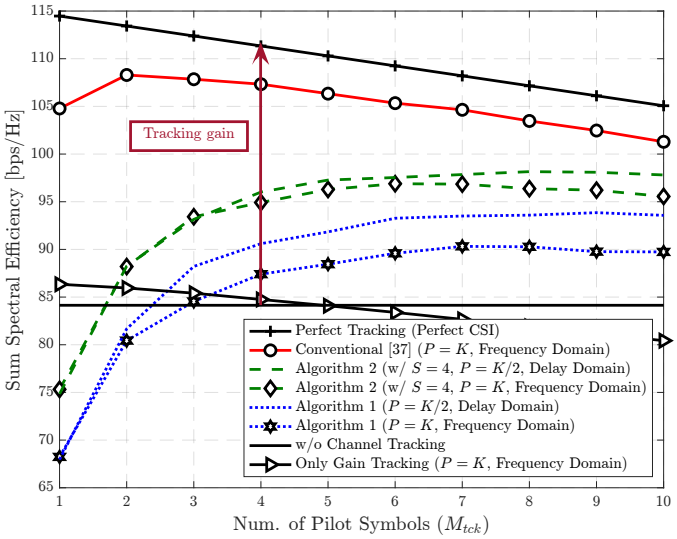
Next, we evaluate the relationship between the training overhead and tracking gain, as shown in Fig. 6. To confirm the robustness of Algorithm 2 against low SNR, transmit power is $P = 20$ [dBm] in Fig. 6(b). The total number of FLOPs as a function of the number of pilot symbols $M_{\rm tck}$ for each algorithm is shown in Fig. 7. The spectral efficiencies plotted in Fig. 6 are the mean values over 21 channel slots given as $\sum_{n=0}^{20} {\rm SE}^{(n)}/21$.

In Fig. 6(a), all the proposed approaches achieve higher spectral efficiencies with more pilot symbols $M_{\rm tck}$ through $M_{\rm tck} = 1$ to $M_{\rm tck} = 6$, whereas the proposed approaches cannot obtain tracking gain in $M_{\rm tck} = 1$. These results indicate that compressive sensing is sensitive to the size of the measurement dimensions $M_{\rm tck}N_{\rm s}$ relative to the estimation dimensions $N_{\rm t}N_{\rm r}$. Under small measurements, especially with the number of pilot symbols $M_{\rm tck} = 2$, the proposed coherence reduction by introducing multiple sub-sensing matrices is more effective. The conventional approach is less dependent on the measurement dimensions. Hence, the overhead increase is more pronounced than the channel estimation accuracy improvement in the spectral efficiency at more than $M_{\rm tck} = 2$ pilot symbols. Although the conventional approach with $M_{\rm tck} = 2$ pilot symbols obtains the highest tracking gain, the total number of FLOPs is more than 100 times larger than the proposed approach of Algorithm 1 with $M_{\rm tck} = 6$ pilot symbols from Fig. 7.

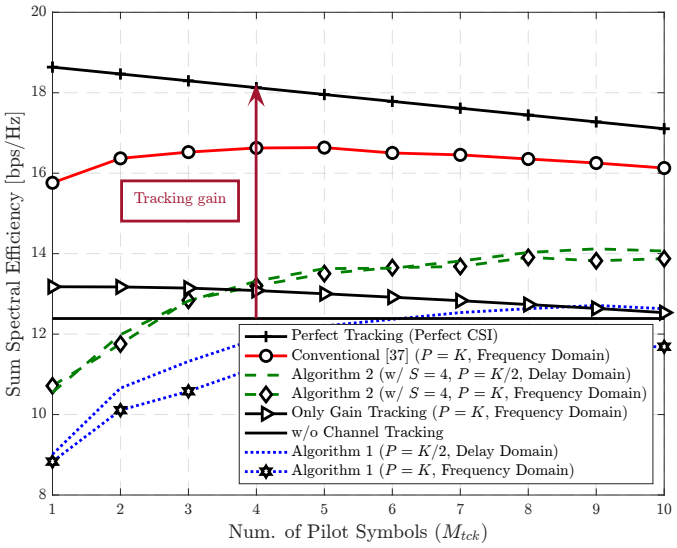
From Fig. 6(b), in the low SNR region, although channel tracking based on Algorithm 1 cannot obtain tracking gain, that based on Algorithm 2 can obtain tracking gain with more than $M_{\rm tck} = 3$ pilot symbols. This result confirms that the mutual coherence reduction by sub-sensing matrices and the consensus process utilizing KF filter in the proposed algorithm refinement bring robustness against low SNR to channel tracking based on compressive sensing.

G. Trade-off between Tracking Gain and Complexity

In practical situations, the computational complexity should be adjusted considering real-time processing by tuning the algorithm parameters, such as the number of particles $N_{\rm PF}$, candidates $Q_{{\rm R},i}, Q_{{\rm T},i}$, and sub-sensing matrices S. Therefore, we evaluate the trade-off between the tracking gain and complexity based on the sum spectral efficiency as a function of those parameters, as shown in Fig. 8. Since the relationship between the frequency and delay domains in the proposed



(a) Transmit power: $P = 35$ [dBm]



(b) Transmit power: $P = 20$ [dBm]

Fig. 6: Sum spectral efficiency as a function of the number of pilot symbols $M_{\rm tck}$

channel tracking is already revealed in previous subsections, only the delay domain channel tracking is considered for the proposed approaches in Fig. 8. In addition, the number of pilot symbols is $M_{\rm tck} = 6$.

In this figure, the horizontal axis corresponds to the number of particles $N_{\rm PF}$ and candidates $Q_{{\rm R},1} = Q_{{\rm T},i}$ for the conventional and proposed approaches with a sensing matrix, respectively. For the proposed approach with multiple sub-sensing matrices, since tracking accuracy depends on the number of sub-sensing matrices S rather than that of candidates $Q_{{\rm R},i}$ and $Q_{{\rm T},i}$, the algorithms with $S = 4$ and $S = 16$ sub-sensing matrices under the $Q_{{\rm R},1} = Q_{{\rm T},i} = 8$ candidates are evaluated. Especially in the case of $S = 16$ sub-sensing matrices, the AoA and AoD candidates for the LOS component are split into four subsets, meaning $U_{{\rm R},1} = U_{{\rm T},1} = 4$, $U_{{\rm R},i} = U_{{\rm T},i} = 1, (i \neq 1)$, and $S = \prod_{i=1}^{3} = U_{{\rm R},i} \cdot U_{{\rm T},i} = 16$. The vertical axis shows the mean value of the sum spectral

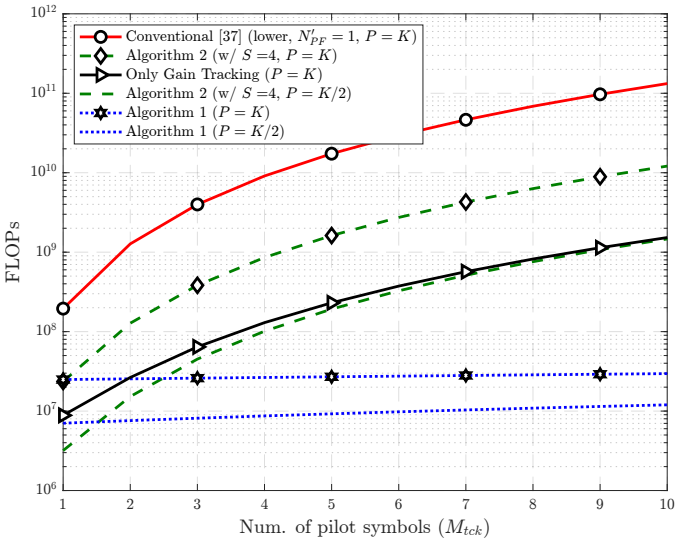


Fig. 7: The number of FLOPs as a function of the number of pilot symbols M_{tck}

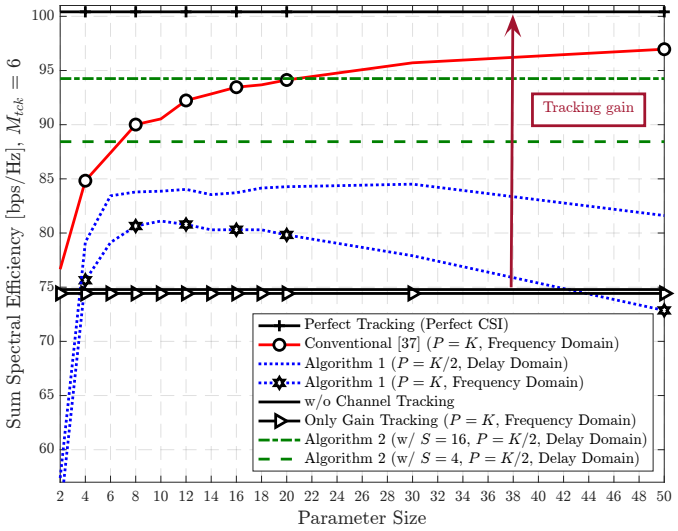


Fig. 8: Sum spectral efficiency as a function of algorithm parameters (N_{PF} or $Q_{\mathrm{R},1}=Q_{\mathrm{T},i}$)

efficiency, which is similar to that in Fig. 6.

A larger number of AoA and AoD candidates leads to higher mutual coherence of the sensing matrix. Hence, unlike the conventional approach of [37], the proposed approach with more candidates may obtain less tracking gain than that with fewer candidates. The proposed sensing matrix design shows its worth with a few candidates. For example, the proposed approaches are shown in Fig. 8 to be effective even with the limited number of $Q_{\mathrm{R},1}=Q_{\mathrm{T},i}=4$ candidates, while compressive sensing methods with the conventional sensing matrix design, $e.g.$, [24], under the same number of 4 AoAs and AoDs candidates in the range of $[0,\pi]$ result in limited tracking performance.

The proposed approach summarized in Algorithm 1 with 4 candidates obtains a higher tracking gain than the conventional approach of [37] with $N_{\mathrm{PF}}=2$ particles, with approximately 10^3 times lower FLOPs $F_{\mathrm{top}}\approx 1.88\times10^6 \ll F_{\mathrm{conv}}\approx 2.83\times10^9$. It should be noted that this FLOPs $F_{\mathrm{top}}\approx 1.88\times10^6$ is lower than that of the conventional approach with $M_{\mathrm{tck}}=2$ pilot symbols, which achieves the maximum spectral efficiency in Fig. 6(a), even if the number of particles is one $N_{\mathrm{PF}}=1$. Moreover, the proposed approaches with $S=4$ and $S=16$ sub-sensing matrices can obtain higher tracking gain to the conventional approaches with $N_{\mathrm{PF}}=6$ and $N_{\mathrm{PF}}=20$ particles, respectively. In these cases, the proposed approaches achieve more than 10 times lower FLOPs than the conventional ones, $i.e.$, $F_{\mathrm{multi}}\approx 7.59\times10^8 < F_{\mathrm{conv}}\approx 8.32\times10^9$ and $F_{\mathrm{multi}}\approx 2.79\times10^9 < F_{\mathrm{conv}}\approx 2.75\times10^{10}$.

Overall, the proposed approach represents a better trade-off between the tracking gain and computational complexity.

VII. Conclusion

We proposed a novel low-complexity channel tracking method based on compressive sensing to track the time fluctuations of mmWave channels with low training overhead. The proposed approach considers sparse path gain reconstruction, formulating a compressive sensing problem in the delay domain with the sensing matrix design exploiting time evolution models for AoAs and AoDs. This problem is then solved using the exact top-$\hat{\mu}$ feature selection approach. To improve the tracking accuracy, mutual coherence reduction in the sensing matrix is realized by splitting it into multiple sub-sensing matrices. Unlike the conventional approach based on GMPF, the proposed approaches avoid Monte Carlo sampling and cubic-order complexity for the number of available subcarriers. The computational complexity in terms of the total number of FLOPs required for each algorithm and the numerical results confirm that the proposed approach obtains higher tracking gain than the conventional approach at lower computational complexity.

As future work, more sophisticated compressive sensing algorithms to capture the unique sparsity caused by the proposed sensing matrix and mathematical analysis of the relationship between the number of pilot symbols, pilot subcarriers, improvement of estimation accuracy, and overhead increase should be considered.

References

[1] M. Xiao *et al.*, "Millimeter wave communications for future mobile networks," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1909–1935, Sep. 2017.

[2] T. S. Rappaport *et al.*, "Wireless communications and applications above 100 GHz: Opportunities and challenges for 6G and beyond," *IEEE Access*, vol. 7, pp. 78 729–78 757, June 2019.

[3] I. F. Akyildiz, A. Kak, and S. Nie, "6G and beyond: The future of wireless communications systems," *IEEE Access*, vol. 8, pp. 133 995–134 030, Jul. 2020.

[4] S. Uchimura, H. Iimori, G. T. F. de Abreu, and K. Ishibashi, "Outage-minimization coordinated multi-point for millimeter-wave OFDM with random blockages," *IEEE Trans. Veh. Technol.*, vol. 72, no. 7, pp. 8783–8796, Jul. 2023.

[5] S. Uchimura, G. T. F. De Abreu, and K. Ishibashi, "Blockage-robust hybrid beamforming enabling high sum rate for millimeter-wave OFDM systems," *IEEE Trans. Wireless Commun.*, pp. 1–15, 2023, Early Access.

[6] W. Hong *et al.*, "The role of millimeter-wave technologies in 5G/6G wireless communications," *IEEE J. Microw.*, vol. 1, no. 1, pp. 101–122, Jan. 2021.

[7] H. Iimori, G. T. F. de Abreu, O. Taghizadeh, R.-A. Stoica, T. Hara, and K. Ishibashi, "Stochastic learning robust beamforming for millimeter-wave systems with path blockage," *IEEE Wireless Commun. Lett.*, vol. 9, no. 9, pp. 1557–1561, Sep. 2020.

[8] T. S. Rappaport, G. R. MacCartney, M. K. Samimi, and S. Sun, "Wideband millimeter-wave propagation measurements and channel models for future wireless communication system design," *IEEE Trans. Commun.*, vol. 63, no. 9, pp. 3029–3056, Sep. 2015.

[9] S. Sun *et al.*, "Propagation path loss models for 5G urban micro- and macro-cellular scenarios," in *Proc IEEE 83rd Veh. Technol. Conf. (VTC Spring)*, May 2016, pp. 1–6.

[10] T. S. Rappaport *et al.*, "Overview of millimeter wave communications for fifth-generation 5G wireless networks—with a focus on propagation models," *IEEE Trans. Antennas Propag.*, vol. 65, no. 12, pp. 6213–6230, Dec. 2017.

[11] J. G. Andrews, T. Bai, M. N. Kulkarni, A. Alkhateeb, A. K. Gupta, and R. W. Heath, "Modeling and analyzing millimeter wave cellular systems," *IEEE Trans. Commun.*, vol. 65, no. 1, pp. 403–430, Jan. 2017.

[12] W. Roh *et al.*, "Millimeter-wave beamforming as an enabling technology for 5G cellular communications: theoretical feasibility and prototype results," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 106–113, Feb. 2014.

[13] F. A. Pereira de Figueiredo, "An overview of massive MIMO for 5G and 6G," *IEEE Latin America Trans.*, vol. 20, no. 6, pp. 931–940, Jan. 2022.

[14] H. Iimori, G. T. F. de Abreu, O. Taghizadeh, R.-A. Stoica, T. Hara, and K. Ishibashi, "A stochastic gradient descent approach for hybrid MmWave beamforming with blockage and CSI-error robustness," *IEEE Access*, vol. 9, pp. 74 471–74 487, May 2021.

[15] T. Kebede *et al.*, "Precoding and beamforming techniques in mmWave-massive MIMO: Performance assessment," *IEEE Access*, vol. 10, pp. 16 365–16 387, Feb. 2022.

[16] C. Pradhan, A. Li, L. Zhuo, Y. Li, and B. Vucetic, "Hybrid-precoding for mmWave multi-user communications in the presence of beam-misalignment," *IEEE Trans. Wireless Commun.*, vol. 19, no. 9, pp. 6083–6099, Jun. 2020.

[17] W. Attaoui, K. Bouraqia, and E. Sabir, "Initial access & beam alignment for mmWave and terahertz communications," *IEEE Access*, vol. 10, pp. 35 363–35 397, Mar. 2022.

[18] J. Lee, J. Liang, M.-D. Kim, and J.-J. Park, "Millimeter-wave beam misalignment analysis based on 28 and 38 GHz urban measurements," in *Proc IEEE 86th Veh. Technol. Conf. (VTC-Fall)*, Sep. 2017, pp. 1–5.

[19] K. Rikkinen, P. Kyosti, M. E. Leinonen, M. Berg, and A. Parssinen, "THz radio communication: Link budget analysis toward 6G," *IEEE Commun. Mag.*, vol. 58, no. 11, pp. 22–27, Nov. 2020.

[20] M. S. Zia, D. M. Blough, and M. A. Weitnauer, "Effects of SNR-dependent beam alignment errors on millimeter-wave cellular networks," *IEEE Trans. Veh. Technol.*, vol. 71, no. 5, pp. 5216–5230, Mar. 2022.

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

[22] K. Venugopal *et al.*, "Channel estimation for hybrid architecture-based wideband millimeter wave systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1996–2009, Sep. 2017.

[23] 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 Trans. Wireless Commun.*, vol. 17, no. 5, pp. 2946–2960, Mar. 2018.

[24] R.-A. Stoica *et al.*, "Frame theory and fractional programming for sparse recovery-based mmWave channel estimation," *IEEE Access*, vol. 7, pp. 150 757–150 774, Oct. 2019.

[25] X. Wu, S. Ma, X. Yang, and G. Yang, "Clustered sparse bayesian learning based channel estimation for millimeter-wave massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 71, no. 12, pp. 12 749–12 764, Dec. 2022.

[26] K. Arai and K. Ishibashi, "Self-calibration for channel estimation in hybrid millimeter-wave MIMO systems," in *Proc IEEE 98th Veh. Technol. Conf. (VTC2023-Fall)*, 2023, pp. 1–5.

[27] C. Zhang, D. Guo, and P. Fan, "Tracking angles of departure and arrival in a mobile millimeter wave channel," in *Proc IEEE Int. Conf. Commun. (ICC)*, May 2016, pp. 1–6.

[28] V. Va, H. Vikalo, and R. W. Heath, "Beam tracking for mobile millimeter wave communication systems," in *Proc Global Conf. Signal Inf. Process. (GlobalSIP)*, Dec. 2016, pp. 743–747.

[29] S. Jayaprakasam, X. Ma, J. W. Choi, and S. Kim, "Robust beam-tracking for mmWave mobile communications," *IEEE Commun. Lett.*, vol. 21, no. 12, pp. 2654–2657, Dec. 2017.

[30] L. Chen, S. Zhou, and W. Wang, "Mmwave beam tracking with spatial information based on extended Kalman filter," *IEEE Wireless Commun. Lett.*, vol. 12, no. 4, pp. 615–619, Apr. 2023.

[31] Y. Liu, M. Chen, C. Pan, Y. Pan, Y. Wang, Y. Huang, T. Cao, and J. Wang, "Channel tracking for RIS-aided mmWave communications under high mobility scenarios," *IEEE Commun. Lett.*, vol. 27, no. 5, pp. 1397–1401, May 2023.

[32] K. Chen, C. Qi, C.-X. Wang, and G. Y. Li, "Beam training and tracking for extremely large-scale MIMO communications," *IEEE Trans. Wireless Commun.*, pp. 1–15, 2023, early Access.

[33] S. Kim, H. Han, N. Kim, and H. Park, "Robust beam tracking algorithm for mmWave MIMO systems in mobile environments," in *Proc IEEE 90th Veh. Technol. Conf. (VTC2019-Fall)*, Sep. 2019, pp. 1–5.

[34] H. Yan, S. Chaudhari, and D. Cabric, "Wideband channel tracking for mmWave MIMO system with hybrid beamforming architecture: (Invited Paper)," in *Proc Int. Workshop Comput. Advances Multi-Sensor Adaptive Process. (CAMSAP)*, Mar. 2017, pp. 1–5.

[35] J. Rodriguez-Fernandez *et al.*, "Position-aided compressive channel estimation and tracking for millimeter wave multi-user MIMO air-to-air communications," in *Proc IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, May 2018, pp. 1–6.

[36] ——, "Position-aided compressive channel tracking for wideband millimeter wave multi-user communication," in *Proc Int. Conf. Commun. (ICC)*, July 2019, pp. 1–6.

[37] N. González-Prelcic, H. Xie, J. Palacios, and T. Shimizu, "Wideband channel tracking and hybrid precoding for mmWave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 20, no. 4, pp. 2161–2174, Oct. 2021.

[38] H. Xie, N. González-Prelcic, and T. Shimizu, "Blockage detection and channel tracking in wideband mmWave MIMO systems," in *Proc IEEE Int. Conf. Commun. (ICC)*, Aug. 2021, pp. 1–6.

[39] S.-H. Wu and G.-Y. Lu, "Compressive beam and channel tracking with reconfigurable hybrid beamforming in mmWave MIMO OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 22, no. 2, pp. 1145–1160, Sep. 2023.

[40] Q. Xue, C. Ji, S. Ma, J. Guo, Y. Xu, Q. Chen, and W. Zhang, "A survey of beam management for mmWave and THz communications towards 6G," 2023. [Online]. Available: https://arxiv.org/abs/2308.02135

[41] M. Qurratulain Khan, A. Gaber, P. Schulz, and G. Fettweis, "Machine learning for millimeter wave and terahertz beam management: A survey and open challenges," *IEEE Access*, vol. 11, pp. 11 880–11 902, Feb. 2023.

[42] V. Petrov, D. Moltchanov, Y. Koucheryavy, and J. M. Jornet, "Capacity and outage of terahertz communications with user micro-mobility and beam misalignment," *IEEE Trans. Veh. Technol.*, vol. 69, no. 6, pp. 6822–6827, Apr. 2020.

[43] J. Li *et al.*, "Mobility support for millimeter wave communications: Opportunities and challenges," *IEEE Commun. Surv. Tut.*, vol. 24, no. 3, pp. 1816–1842, May 2022.

[44] X. Cai, F. Nie, and H. Huang, "Exact top-k feature selection via l2,0-norm constraint," in *Proc the Twenty-Third Int. Joint Conf. on Artif. Intell.*, ser. IJCAI '13. AAAI Press, 2013, pp. 1240–1246.

[45] J. Ko, Y.-J. Cho, S. Hur, T. Kim, J. Park, A. F. Molisch, K. Haneda, M. Peter, D.-J. Park, and D.-H. Cho, "Millimeter-wave channel measurements and analysis for statistical spatial channel model in in-building and urban environments at 28 GHz," *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 5853–5868, Dec. 2017.

[46] M. K. Samimi *et al.*, "28 GHz millimeter-wave ultrawideband small-scale fading models in wireless channels," in *Proc IEEE 83rd Veh. Technol. Conf. (VTC Spring)*, May 2016, pp. 1–6.

[47] T. Andersen and A. Enmark, *Fourier Transforms and Interpolation*. New York, NY: Springer New York, 2011, pp. 45–76. [Online]. Available: https://doi.org/10.1007/978-1-4614-0149-0_4

[48] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[49] Y. C. Eldar and G. Kutyniok, *Compressed Sensing: Theory and Applications*. Cambridge: Cambridge University Press, 2012.



Sota Uchimura (Graduate Student Member, IEEE) received the B.E. and M.E. degree in engineering from The University of Electro-Communications, Tokyo, Japan in 2021 and 2023, respectively. His current research interests include millimeter-wave communications and signal processing. He was the recipient of IEICE RCS Active Research Award in 2022 and IEICE SITA Young Researcher Award in 2022.



Koji Ishibashi (Senior Member, IEEE) received the B.E. and M.E. degrees in engineering from The University of Electro-Communications, Tokyo, Japan, in 2002 and 2004, respectively, and the Ph.D. degree in engineering from Yokohama National University, Yokohama, Japan, in 2007. From 2007 to 2012, he was an Assistant Professor at the Department of Electrical and Electronic Engineering, Shizuoka University, Hamamatsu, Japan. Since April 2012, he has been with the Advanced Wireless and Communication Research Center (AWCC), The University of Electro-Communications, Tokyo, Japan where he is currently a Professor. From 2010 to 2012, he was a Visiting Scholar at the School of Engineering and Applied Sciences, Harvard University, Cambridge, MA. He is a senior member of IEICE and IEEE. He is a recipient of Takayanagi Research Encouragement Award in 2009 and KDDI Foundation Award in 2023. He was certified as an Exemplary Reviewer of IEEE Communications Wireless Letters in 2015 and awarded by the Telecommunication Technology Committee (TTC) for his devotion to standardization activities in 2020. He in the past served as an Associate Editor for IEICE Transactions on Communications, an Associate Editor for IEEE Journal on Selected Areas in Communications (JSAC) and a Guest Editor for IEEE Open Journal of the Communications Society. His current research interests are robust beamforming, grant-free access, energy-harvesting, compressed sensing, coding, cell-free network, and MIMO technologies.



Hiroki Iimori (S'18–M'22) received his Ph.D. with special distinction from Jacobs University Bremen, Bremen, Germany, in 2022, after receiving his B.Eng. and M.Eng. (Hons.) degrees in electrical and electronic engineering from Ritsumeikan University, Kyoto, Japan, in 2017 and 2019. In 2020, he was a visiting scholar at the Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON, Canada. In 2021, he was a research intern at the Ericsson Radio S&R Research Laboratory, Yokohama, Japan, where he currently holds the position of Experienced Researcher. His research interests include optimization theory, wireless communications, and signal processing. He was awarded the YKK Doctoral Fellowship by the Yoshida Scholarship Foundation, the IEICE Young Researcher of the Year Award by the IEICE Smart Radio Committee in 2020, among others.



Paulo Valente Klaine received his B. Eng. degree in electrical and electronic engineering from the Federal University of Technology - Parana (UTFPR), Brazil in 2014, the MSc. degree from the University of Surrey, U.K., in mobile communications systems in 2015, and the PhD degree in electrical and Electronics engineering from the University of Glasgow, U.K., in 2019. He received the IEEE ICEICT ' 21 Best Paper award. His research interests include self organizing networks, machine learning in wireless networks,V2X communications, wireless blockchain and massive MIMO. Since February of 2022 he has been working as an experienced researcher at Ericsson K. K. Japan.



Szabolcs Malomsoky is the head of Ericsson Research Japan. Earlier he was leading research units in Hungary, Sweden and the US. He received a Ph.D. degree from the Budapest University of Technology and Economics in 2003. Szabolcs worked with strategy setting and technical leadership in research areas including wireless technologies, network analytics, cloud computing, network management and programmable networks.