3D On and Off-Grid Dynamic Channel Tracking for Multiple UAVs and Satellite Communications

Jiadong Yu, Xiaolan Liu, Yue Gao, *Senior Member, IEEE*, and Xuemin Shen, *Fellow, IEEE*

Abstract—The space-air-ground integrated network (SAGIN) has drawn increasing attention for its benefits, such as wide coverage, high throughput for 5G and 6G communications. As one of the links, space-air communications between multiple unmanned aerial vehicles (UAVs) and Ka-band orbiting low earth orbit (LEO) satellites face a crucial challenge in tracking the 3D dynamic channel information. This paper exploits a statistical dynamic channel model called the multi-dimensional Markov model (MD-MM), which investigates the more realistic spatial and temporal correlation in the sparse UAVs-satellite channel. Specifically, the spatial and temporal probabilistic relationships of multi-user (MU) hidden support vector, single-user (SU) joint hidden support vector, and SU hidden value vector are investigated. The specific transition probabilities that connect the SU and MU hidden support vector for both azimuth and elevation directions are defined. Moreover, based on the proposed MD-MM, we derive a novel multi-dimensional dynamic turbo approximate message passing (MD-DTAMP) algorithm for tracking the 3D dynamic channel in multiple UAVs systems. Furthermore, we also develop a gradient update scheme to recursively find the azimuth and elevation offset for 3D off-grid estimation. Numerical results verify that the proposed algorithm shows superior 3D channel tracking performance with smaller pilot overhead and comparable complexity.

Index Terms—Channel tracking, message passing, space-air communications.

I. INTRODUCTION

WITH increasing communication traffic demands for emerging applications, such as the internet of things (IoT), cloud computing, autonomous cars and so on, future networks have been expected to provide more robust communication services. Hence, the space-air-ground integrated network (SAGIN) has been widely studied for its advantages such as large coverage, high throughput and strong resilience for 5G and beyond systems [1]. SAGIN has mainly three segments: space segment consists of satellite network, air segment consists of aerial network, and ground segment consists of terrestrial network [2]. Acting as the middle segment, aerial network with unmanned aerial vehicles (UAVs) is cost–effective. It can provide a shorter line of sight (LoS) range for transmission compared with satellite-ground links [3]. UAVs are deployed as base stations in both crowded areas and non-accessible emergency scenarios for UAV-aided air-ground communications, which can provide seamless coverage. For space-air links, UAVs play a role as a relay to transmit the information received from satellite (e.g. low earth orbit (LEO) and medium earth orbit (MEO) satellites [4]) to terrestrial network with mmWave Ka-band [5], [6]. In the above two UAV-related communication links, due to the velocity of UAVs, it is undeniable that the estimation of the highly dynamic communication channels with lower overhead and good accuracy faces some challenges: the spatial and temporal variations caused by the mobility of UAVs in three dimensional (3D) space [7].

A. Related Work

Conventionally, channel state information (CSI) is estimated using techniques, such as least square (LS) and minimum mean square error (MMSE). However, for these methods, training overhead is proportional to the number of transmit antennas which is infeasible to mmWave communication with massive MIMO system [8]. To reduce the training overhead, compressive sensing (CS) based techniques, such as orthogonal matching pursuit (OMP) and compressive sampling matching pursuit (CoSaMP), have been widely studied to estimate the approximate sparse wireless channel in different domain [9].

Based on fundamental CS concepts and algorithms, some studies explored the more detailed sparse structure of the channel through spatial correlation [10]–[13] and temporal correlation [14]–[17]. Because of the physical scattering of the environment, [10] has concluded that the angular domain shows a burst sparsity structure. Hence, they proposed the burst least absolute shrinkage and selection operator (burst-LASSO) to extract the burst structure properties in the massive MIMO channels with enhanced performance. Reference [11] also exploited the clustered sparse structure of the channel

Jiadong Yu is with the Department of Electronic Engineering and Computer Science, Queen Mary University of London (QMUL), London E1 4NS, U.K. (e-mail: jiadong.yu@qmul.ac.uk).

Xiaolan Liu is with the Institute for Digital Technologies, Loughborough University London, London E20 3BS, U.K. (e-mail: xiaolan.liu@lboro.ac.uk).

Yue Gao is with the Department of Electrical and Electronic Engineering, University of Surrey, Guildford, Surrey GU2 7XH, U.K. (e-mail: yue.gao@ieee.org).

Xuemin Shen is with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: sshen@uwaterloo.ca).

support in the spatial domain through Markov prior. [12] proposed a hidden Markov model to capture the angular domain structured sparsity of the channel. Moreover, [13] exploited the double sparsity of the multipath components in both angular domain and delay domain. However, these studies only considered the angular domain structure ([13] also considered the delay domain) based on static channel environment. Effectively utilizing the prior information can reduce the training overhead and maintain a good current channel estimation performance for dynamic channel tracking. Reference [14] proposed a prior aided channel tracking scheme by taking the practical user motion model into account through a temporal variation law of the physical direction. Reference [15] modeled the channel by a Markov process and the Kalman filter to track the dynamic model. A structured matching pursuit [16] and differential orthogonal matching pursuit (D-OMP) [17] are proposed both by considering the temporal correlation that path delays change slowly over time but path gains evolve faster.

Jointly considering both spatial correlation and temporal correlation can improve the channel tracking performance [18]–[21]. Reference [18] proposed the channel tracking scheme based on the prior-aided iterative angle and Doppler shift estimation algorithm and with the help of the previously estimated channel parameters. Reference [19] decoupled the channel information into angles of arrival/departure (AoAs/AoDs) information and gain knowledge. A modified unscented Kalman filter tracks the AoAs/ AoDs, and the gain is estimated from the proposed beam training. References [20] and [21] formulated the more realistic dynamic models of the AoDs support over time through hidden Markov model in flat and 3D space, respectively.

B. Motivation and Contribution

The aforementioned channel estimation and tracking work in [10]–[12], [14]–[20] can be directly adapted to the ground links and air-ground links with mobile scenario. However, for space-air links, which consist of multiple UAVs and one satellite, there are some non-neglected features: the known orbiting satellite, the 3D UAVs moving trajectories, and the dominant LoS ray between each UAV and satellite [2].

Although [10]–[12] considered the angular domain sparsity of the channel, it's non-trivial to include the temporal information for channel tracking, especially in highly mobile space-air links with orbiting satellite and UAVs navigation. Comparatively, [14]–[17] all considered the prior channel information for time-varying channel tracking. However, they ignored the spatial angular structured sparsity. Reference [18] studied the prior information of both angle and Doppler shifts for channel tracking, the probabilistic relationship of the angle movement is not explored. Moreover, although [19] and [20] both jointly considered spatial and temporal correlation of the dynamic channel, the premise of the system were both communicated through a flat surface. Furthermore, although our previous work [21] considered a more realistic 3D space-air communications with only one UAV, however, the multiple-UAVs system can provide wider coverage in practice.

Compared with one UAV in the system, for multiple UAVs, there will have more than one structured sparse points on the quantized grids to represent the multi-user (MU) spatial correlation with sharing azimuth or elevation supports. Hence, a more complicated scenario with multiple UAVs should be further explored. The above algorithms all assumed that the true angles lie on the quantized grids. However, in the real world, the true angle is continuously distributed [22]. To solve the off-grid problem, [22] and [23] all proposed algorithms based on a two-dimensional (2D) flat environment. Hence, none of the mentioned work considered the 3D off-grid case of the channel model.

Motivated by the above literature review, we consider the multiple UAVs space-air links to recursively track the 3D off-grid dynamic channel with smaller training overhead, comparable complexity, and improved estimation performance. The main contributions are summarized as follows:

- We develop a statistical 3D on-grid and off-grid channel model called multi-dimensional Markov model (MD-MM) for space-air communication links with 3D navigation multiple UAVs and orbit LEO satellite in the system. The MD-MM captures the structured sparsity of the MU azimuth and elevation spatial domains and the probabilistic relationship between MU and SU support.
- Based on the proposed MD-MM channel model, we propose a novel multi-dimensional dynamic turbo approximate message passing (MD-DTAMP) algorithm which can recursively track the 3D on-grid dynamic channel over time.
- We further develop a gradient azimuth and elevation 3D off-grid estimation scheme for a more realistic channel model, which iteratively update the optimum 3D offset through gradient calculation.
- The proposed MD-DTAMP algorithm with and without an off-grid scheme all show superior estimation performance compared with benchmark algorithms. Moreover, the proposed algorithm can achieve solid performance with smaller pilot overhead and comparable complexity.

This paper is organized as follows. The downlink LEO satellite to multiple UAVs system model is introduced in Section II. In Section III, the 3D on-grid and off-grid MD-MM channel model are elaborated. The problem of channel tracking is formulated in Section IV. The proposed algorithm MD-DTAMP and the 3D gradient off-grid estimation scheme are presented in Section V. Numerical results and conclusions are then provided in Section VI and Section VII, respectively.

Notation: Throughout the paper, we use the following common notation. The complex numbers are denoted by $\mathbb{C}$. The transpose and the conjugate transpose are denoted by $(\cdot)^{\mathrm{T}}$, $(\cdot)^{\mathrm{H}}$ respectively. $\boldsymbol{I}$ is the identity matrix. $\mathcal{CN}$ (mean, covariance) indicates a complex Gaussian random vector with defined mean and covariance. $\|\cdot\|$ is the l_2 norm.

II. SYSTEM MODEL

In this paper, we consider the narrow-band flat fading transmission model through space-air downlink communication from LEO satellite to multiple UAVs. Since Ka-band is
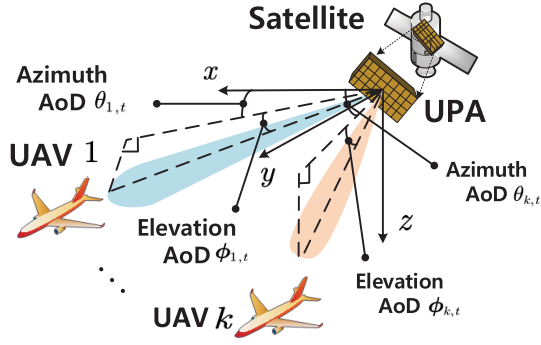
Fig. 1. The illustration of space-air communications system model at time slot t.

TABLE I
TABLE OF IMPORTANT NOTATIONS

Notations	Explanations
n_x, N_x	Index and total number of x-axis antenna/sparsity size
n_y, N_y	Index and total number of y-axis antenna/sparsity size
N_S	Total pilot number
m, M	Index and total number of the antenna and the sparse grid
t, T	Index and total number of the time slot
$\theta_{k,t}, \phi_{k,t}$	k^{th} UAV azimuth and elevation directions of AoD at t
$\tilde{\theta}_{k,t}, \tilde{\phi}_{k,t}$	k^{th} UAV on-grid azimuth and elevation directions of AoD at t
$\Delta\theta_{k,t}, \Delta\phi_{k,t}$	k^{th} UAV azimuth and elevation offset directions of AoD at t
θ_{range}	The range of azimuth directions of AoD
ϕ_{range}	The range of elevation directions of AoD
$\boldsymbol{a}(\theta_{k,t}, \phi_{k,t})$	Transmit array response vector at t
$\boldsymbol{h}_{k,t}$	k^{th} UAV frequency domain channel model at t
$\boldsymbol{S}_k$	k^{th} UAV transmit pilot symbols
$\boldsymbol{y}_{k,t}^{\text{RX}}, \boldsymbol{y}_t$	k^{th} UAV received signal at t
$\boldsymbol{y}_k^{(t)}$	k^{th} UAV total received signal for total period t
$\boldsymbol{n}_{k,t}^{\text{RX}}, \boldsymbol{n}_{k,t}$	k^{th} UAV additive white Gaussian noise at t
$\tilde{g}_{k,t}$	k^{th} UAV complex channel gain at t
$g_{k,t,m}$	k^{th} UAV channel gain element for the mth AoD at t
$\boldsymbol{g}_{k,t}$	k^{th} UAV sparse matrix complex channel gain at t
$\boldsymbol{g}_k^{(t)}, \boldsymbol{g}_k^{(T)}$	k^{th} UAV total sparse matrix complex channel gain for period t or T
$b_{k,t,m}$	k^{th} UAV hidden support element for the mth AoD at t
$\boldsymbol{b}_{k,t}$	k^{th} UAV hidden support vector at t
$\boldsymbol{b}_k^{(t)}, \boldsymbol{b}_k^{(T)}$	k^{th} UAV hidden support vector for total period t or T
$\tilde{\boldsymbol{b}}_t$	MU Hidden support vector at t
$a_{k,t,m}$	k^{th} UAV azimuth support element for the mth AoD at t
$e_{k,t,m}$	k^{th} UAV elevation support element for the mth AoD at t
$\boldsymbol{a}_{k,t}, \boldsymbol{e}_{k,t}$	k^{th} UAV azimuth and elevation support vector at t
$\boldsymbol{a}_k^{(t)}, \boldsymbol{a}_k^{(T)}$	k^{th} UAV azimuth support vector for total period t or T
$\boldsymbol{e}_k^{(t)}, \boldsymbol{e}_k^{(T)}$	k^{th} UAV elevation support vector for total period t or T
$\tilde{a}_{t,m}, \tilde{e}_{t,m}$	MU azimuth and elevation support element for the mth AoD at t
$\tilde{\boldsymbol{a}}_t, \tilde{\boldsymbol{e}}_t$	MU azimuth and elevation support vector at t
$\tilde{\boldsymbol{a}}^{(t)}, \tilde{\boldsymbol{a}}^{(T)}$	MU azimuth support vector for total period t or T
$\tilde{\boldsymbol{e}}^{(t)}, \tilde{\boldsymbol{e}}^{(T)}$	MU elevation support vector for total period t or T
$\vartheta_{k,t,m}$	k^{th} UAV hidden value element for the mth AoD at t
$\boldsymbol{\vartheta}_{k,t}$	k^{th} UAV hidden value vector at t
$\boldsymbol{\vartheta}_k^{(t)}, \boldsymbol{\vartheta}_k^{(T)}$	k^{th} UAV hidden value vector for total period t or T

weather dependent, to mitigate the high path loss, the LEO satellite is installed with uniform planar array (UPA) with $M = N_x \times N_y$ antenna elements to compensate the transmission loss [6], [24]. The $k^{th} \in \{1, \ldots, K\}$ UAV is equipped with a high gain omni-directional single antenna[1] [25], [26]. The illustration of the system model can be found in Fig. 1.

[1]For simplicity of only exploring the temporal and spatial correlation of the azimuth and elevation AoD of the satellite, the high gain omni-directional single antenna is assumed to be carried by the kth UAV. However, the proposed system model can be further extended with UAVs also installed with UPA. The main difference is that the channel model should further take the receiver array response vectors with azimuth and elevation directions of AoAs into consideration.

The important notations of the system and channel model are shown in Table I. In training procedure, as symbols $\boldsymbol{S}_k \in \mathbb{C}^{M \times N_S}$ are transmitted from satellite to UAV, the received signal at k^{th} UAV at time slot t can be expressed as

$$\boldsymbol{y}_{k,t}^{\text{RX}} = \boldsymbol{h}_{k,t}\boldsymbol{S}_k + \boldsymbol{n}^{\text{RX}}, \qquad (1)$$

with $\boldsymbol{h}_{k,t} \in \mathbb{C}^{1 \times M}$ as the frequency domain channel vector, and $\boldsymbol{n}^{\text{RX}} \sim \mathcal{CN}(0, \sigma_n^2 \boldsymbol{I})$ is the additive white Gaussian noise.

A. Channel Model

The LoS ray is the dominant path in space-air communication links [27]. Hence, the $\boldsymbol{h}_{k,t}$ denotes the frequency domain channel vector with one dominant LoS path, which is written as,

$$\boldsymbol{h}_{k,t} = \tilde{g}_{k,t}\boldsymbol{a}^H(\theta_{k,t}, \phi_{k,t}), \qquad (2)$$

where $\tilde{g}_{k,t}$ is the complex channel gain, $\boldsymbol{a}(\theta_{k,t}, \phi_{k,t}) \in \mathbb{C}^{(N_x \times N_y) \times 1}$ represents the transmit array response vectors, $\theta_{k,t}$ and $\phi_{k,t}$ as azimuth and elevation directions of AoD for k^{th} UAV at time slot t. As for UPA installed LEO satellite, the steering vector can be written as

$$\boldsymbol{a}(\theta_{k,t}, \phi_{k,t})$$
$$= \frac{1}{\sqrt{N_x \times N_y}}[1, \ldots, e^{j\frac{2\pi}{\lambda}d\sin\theta_{k,t}(n_x \sin\phi_{k,t} + n_y \cos\phi_{k,t})},$$
$$\ldots, e^{j\frac{2\pi}{\lambda}d\sin\theta_{k,t}(N_x \sin\phi_{k,t} + N_y \cos\phi_{k,t})}]^T, \qquad (3)$$

where λ is the signal wavelength, $0 \leqslant n_x < N_x$ and $0 \leqslant n_y < N_y$ are the indices of an antenna element with N_x and N_y as antenna array sizes.

B. On-Grid and Off-Grid Virtual Channel Model

Conventionally, to explore the sparse property of the channel, AoDs are assumed to lie on the quantized grids for simplicity [28]. With discretized AoDs, the channel model is called on-grid channel. However, in practice, the true angle is continuously distributed. The channel with this kind of AoDs is named as off-grid channel [23]. Hence in this subsection, we present both on-gird and off-grid cases.

For on-grid case, the channel model can be further written into a sparse representation as

$$\boldsymbol{h}_{k,t} = \boldsymbol{g}_{k,t}\tilde{\boldsymbol{A}}^H(\tilde{\boldsymbol{\theta}}_{k,t}, \tilde{\boldsymbol{\phi}}_{k,t}), \qquad (4)$$

where $\boldsymbol{g}_{k,t} \in \mathbb{C}^{1 \times M}$ is a sparse vector with one non-zero elements (which means only one dominant path for the kth UAV), M is the total number of spatial grids, the quantized AoD pairs are considered, i.e., $\tilde{\boldsymbol{\theta}}_{k,t} = [0, \theta^{\text{range}}/N_x, \ldots, \theta^{\text{range}}(N_x - 1)/N_x]$ and $\tilde{\boldsymbol{\phi}}_{k,t} = [0, \phi^{\text{range}}/N_y, \ldots, \phi^{\text{range}}(N_y - 1)/N_y]$, which means the virtual k^{th} UAV's AoD pair can consist of a M uniform grid, and $\tilde{\boldsymbol{A}}(\tilde{\boldsymbol{\theta}}_{k,t}, \tilde{\boldsymbol{\phi}}_{k,t}) \in \mathbb{C}^{M \times M}$ is the dictionary matrix. For off-grid case, we define $\Delta\theta_{k,t,n_x}$ and $\Delta\phi_{k,t,n_y}$ to represent the offset angles for both azimuth and elevation directions, respectively. Hence, the true AoD azimuth and elevation angle should be the combination of the nearest on-grid angle and the relative offset angle. As an example, we demonstrate the kth UAV's AoD
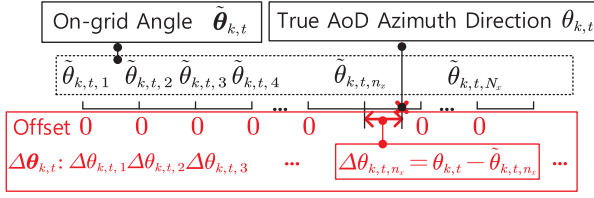
Fig. 2. Illustration of the off-grid AoD azimuth direction.

azimuth direction. The illustration can be found in Fig. 2. We assume that the true azimuth angle $\theta_{k,t}$ has the nearest on-grid angle $\tilde{\theta}_{k,t,n_x}$ and the offset angle is $\Delta\theta_{k,t,n_x} = \theta_{k,t} - \tilde{\theta}_{k,t,n_x}$. Similarly, the AoD elevation direction offset angle can be written as $\Delta\phi_{k,t,n_x} = \phi_{k,t} - \tilde{\phi}_{k,t,n_y}$. Hence, the off-grid channel can be written as

$$\boldsymbol{h}_{k,t} = \boldsymbol{g}_{k,t}\tilde{\boldsymbol{A}}^H\left(\tilde{\boldsymbol{\theta}}_{k,t} + \Delta\boldsymbol{\theta}_{k,t}, \tilde{\boldsymbol{\phi}}_{k,t} + \Delta\boldsymbol{\phi}_{k,t}\right), \quad (5)$$

where

$$\tilde{\boldsymbol{A}}\left(\tilde{\boldsymbol{\theta}}_{k,t} + \Delta\boldsymbol{\theta}_{k,t}, \tilde{\boldsymbol{\phi}}_{k,t} + \Delta\boldsymbol{\phi}_{k,t}\right)$$
$$= \left[\boldsymbol{a}\left(\tilde{\theta}_{k,t,1} + \Delta\theta_{k,t,1}, \tilde{\phi}_{k,t,1} + \Delta\phi_{k,t,1}\right),\right.$$
$$\ldots, \boldsymbol{a}\left(\tilde{\theta}_{k,t,n_x} + \Delta\theta_{k,t,n_x}, \tilde{\phi}_{k,t,n_y} + \Delta\phi_{k,t,n_y}\right),$$
$$\left.\ldots, \boldsymbol{a}\left(\tilde{\theta}_{k,t,N_x} + \Delta\theta_{k,t,N_x}, \tilde{\phi}_{k,t,N_y} + \Delta\phi_{k,t,N_y}\right)\right] \quad (6)$$

is the off-grid array response vector dictionary matrix. To be noticed, the AoD azimuth offset vector $\Delta\boldsymbol{\theta}_{k,t} = [\Delta\theta_{k,t,1}\ldots,\Delta\theta_{k,t,N_x}]^T$ with

$$\Delta\theta_{k,t,n_x} = \begin{cases} \theta_{k,t} - \tilde{\theta}_{k,t,n_x}, & \text{Active path,} \\ 0, & \text{Otherwise.} \end{cases} \quad (7)$$

The otherwise case represents the non-active virtual on-grid offsets which can be observed in Fig. 2 with the notation from $\Delta\theta_{k,t,1}$ to $\Delta\theta_{k,t,N}$ without the active path offset $\Delta\theta_{k,t,n_x}$. Similarly, the AoD elevation offset vector $\Delta\boldsymbol{\phi}_{k,t} = [\Delta\phi_{k,t,1}\ldots,\Delta\phi_{k,t,N_y}]^T$ with

$$\Delta\phi_{k,t,n_y} = \begin{cases} \phi_{k,t} - \tilde{\phi}_{k,t,n_y}, & \text{Active path,} \\ 0, & \text{Otherwise.} \end{cases} \quad (8)$$

When $\Delta\boldsymbol{\theta}_{k,t} = \boldsymbol{0}$ and $\Delta\boldsymbol{\phi}_{k,t} = \boldsymbol{0}$, the off-grid channel can be treated as the on-grid channel in (4).

III. MULTI-DIMENSIONAL MARKOV MULTIUSER CHANNEL MODEL

In our model, communication links connected between LEO satellite and multiple UAVs are considered. Because of the LEO satellite height, a certain portion of the earth's surface is covered at any given time slot. This means a constellation of LEO satellites is necessary for full coverage of the earth [29]. We consider a condition that one LEO satellite communicates with the same group of multiple UAVs simultaneously before any communication link is handed over to the adjacent LEO satellite. This means the LEO satellite would orbit in a certain

distance range, limiting the azimuth and elevation angle in fixed ranges. With assumptions above, we transfer the complex continuous relative displacement relationships between LEO satellite and UAVs into time-varying angular domain variance with transition probability similar to [19]–[21] (this is further elaborated in Section III-C).

In the MU scenario with multiple UAVs, we specially explore the relationship between the MU and SU hidden support vector. The MU azimuth $\tilde{\boldsymbol{a}}^{(T)}$ and elevation $\tilde{\boldsymbol{e}}^{(T)}$ support vector are both modeled as two dimensional Markov model (2D-MM) similarly in [21]. $\tilde{\boldsymbol{a}}^{(T)}$ and $\tilde{\boldsymbol{e}}^{(T)}$ represent the MU azimuth and elevation support vector over total time period T. During time slot t, $\tilde{\boldsymbol{a}}_t = [\tilde{a}_{t,1},\ldots,\tilde{a}_{t,N_x}] \in \{0,1\}^{N_x}$ with $\tilde{a}_{t,n_x} = a_{1,t,m} \vee a_{2,t,m} \vee \ldots \vee a_{K,t,m}$, which means MU azimuth support is the union of the SU azimuth support of total K UAVs at time slot t. Similarly, for MU elevation support $\tilde{\boldsymbol{e}}_t = [\tilde{e}_{t,1},\ldots,\tilde{e}_{t,N_x}] \in \{0,1\}^{N_x}$ with $\tilde{e}_{t,n_x} = e_{1,t,m} \vee e_{2,t,m} \vee \ldots \vee e_{K,t,m}$. Furthermore, the SU joint hidden support vector $\boldsymbol{b}_{k,t}$ is modeled as three dimensional hidden Markov model (3D-HMM) which jointly combines the single-user (SU) azimuth and elevation support.

The details of the complete multi-dimensional Markov model (MD-MM) channel model are introduced in four parts: probability model for SU channel, 2D Markov model (2D-MM) for azimuth and elevation MU hidden support vector, 3D-HMM for azimuth and elevation SU joint hidden support vector, and Gaussian Markov model for SU hidden value vector.

A. Probability Model for SU Channel

The dynamic angular domain channel for the kth UAV $\boldsymbol{g}_k^{(T)} = \{\boldsymbol{g}_{k,1},\ldots,\boldsymbol{g}_{k,T}\}$ with $\boldsymbol{g}_{k,t} = \{g_{k,t,1},\ldots,g_{k,t,M}\}$ can be modeled as a probabilistic signal model with two hidden random processes, SU joint hidden support vector $\boldsymbol{b}_k^{(T)} = \{\boldsymbol{b}_{k,1},\ldots,\boldsymbol{b}_{k,T}\}$ and SU hidden value vector $\boldsymbol{\vartheta}_k^{(T)} = \{\boldsymbol{\vartheta}_{k,1},\ldots,\boldsymbol{\vartheta}_{k,T}\}$. Here, $\boldsymbol{b}_{k,t} = [b_{k,t,1,1},\ldots,b_{k,t,N_x,N_y}] = [b_{k,t,1},\ldots,b_{k,t,M}] \in \{0,1\}^M$ represents the SU joint hidden support vector at time t, which indicates the kth UAV channel sparsity with $b_{k,t,m} \in \{1,0\}$. The SU hidden value vector $\boldsymbol{\vartheta}_{k,t} = [\vartheta_{k,t,1},\ldots,\vartheta_{k,t,M}] \in \mathbb{C}^M$ represents the temporal correlation of the kth UAV channel coefficients. The dynamic channel element for the mth AoD direction in the time t of the kth UAV can be written as

$$g_{k,t,m} = b_{k,t,m}\vartheta_{k,t,m}, \quad (9)$$

where $b_{k,t,m}$ denotes whether there is an active path or not (i.e., $b_{k,t,m} = 1$ means the mth AoD path is activated) and $\vartheta_{k,t,m}$ represents the path gain.[2]

To formulate the more realistic dynamic channel (i.e., the inactive AoD path at the current time t may be activated at time $t + 1$), a probabilistic channel model of kth UAV with channel prior distribution can be formed as

$$p\left(\boldsymbol{g}_k^{(T)}, \boldsymbol{b}_k^{(T)}, \boldsymbol{\vartheta}_k^{(T)}\right)$$
$$= p\left(\boldsymbol{g}_k^{(T)} \mid \boldsymbol{b}_k^{(T)}, \boldsymbol{\vartheta}_k^{(T)}\right) p\left(\boldsymbol{b}_k^{(T)}\right) p\left(\boldsymbol{\vartheta}_k^{(T)}\right), \quad (10)$$

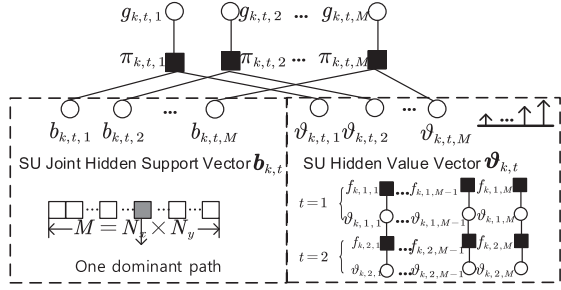[2]For paring the azimuth and elevation directions, $m = (n_x - 1) \times N_x + n_y$.

Fig. 3. Factor graph of the SU channel.

where the kth UAV channel vector conditional prior is

$$
p\left(\boldsymbol{g}_k^{(T)} \mid \boldsymbol{b}_k^{(T)}, \boldsymbol{\vartheta}_k^{(T)}\right) = \prod_{t=1}^{T} \prod_{m=1}^{M} p\left(g_{k,t,m} \mid b_{k,t,m}, \vartheta_{k,t,m}\right)
$$
$$
= \prod_{t=1}^{T} \prod_{m=1}^{M} \delta\left(g_{k,t,m} - b_{k,t,m}\vartheta_{k,t,m}\right),
$$
(11)

with $\delta(\cdot)$ is the Dirac delta function. The factor graph of the combined k^{th} UAV channel can be found in Fig. 3, with $\pi_{k,t,m}$ denotes the conditional prior $p\left(g_{k,t,m} \mid b_{k,t,m}, \vartheta_{k,t,m}\right)$. This factor graph shows that the final k^{th} UAV channel gain $\mathbf{g}_{k,t}$ is the combination of the SU joint hidden support vector $\mathbf{b}_{k,t}$ (on the left hand side of Fig. 3) and the SU hidden value vector $\vartheta_{k,t}$ (on the right hand side of Fig. 3).

B. Gaussian Markov Model for SU Hidden Value Vector

The illustration of the factor graph for the kth UAV SU hidden value vector $\vartheta_{k,t}$ and the Gaussian-Markov model can be found in Fig. 3 (on the right hand side). Due to the path gains change smoothly over time [17], the hidden value vector can be formulated as the Gauss-Markov processes [30]

$$
\vartheta_{k,t,m} = (1 - \beta_k)\left(\vartheta_{k,t-1,m} - \mu_k\right) + \beta_k \omega_{k,t,m} + \mu_k, \quad (12)
$$

where $\beta_k \in [0,1]$, $\omega_{k,t,m} \sim \mathcal{CN}(0, \zeta_k)$, $\mu_k \in \mathbb{C}$ is the mean of the process. When $\beta_k = 0$, then $\vartheta_{k,t,m} = \vartheta_{k,t-1,m}$, which indicates that $\vartheta_{k,t}$ remain unchanged over time. When $\beta_k = 1$, then $\vartheta_{k,t,m} = \beta_k \omega_{k,t,m} + \mu_k$, which means that the SU hidden value vector shows the i.i.d Gaussian distribution with mean μ_k over time. If $0 < \beta_k < 1$, the conditional probability can be written as

$$
p\left(\vartheta_{k,t,m} \mid \vartheta_{k,t-1,m}\right)
$$
$$
\sim \mathcal{CN}\left(\vartheta_{k,t,m}; (1 - \beta_k)\vartheta_{k,t-1,m} + \beta_k\mu_k, \beta_k^2\zeta_k\right). \quad (13)
$$

The distribution of the steady state of the process is $\vartheta_{k,t,m} \sim \mathcal{CN}(\mu_k, \sigma_k^2 = \frac{\beta_k \zeta_k}{2 - \beta_k})$. And the joint distribution of $\vartheta_k^{(T)}$ can be formulated as

$$
p\left(\vartheta_k^{(T)}\right) = \prod_{m=1}^{M} p\left(\vartheta_{k,1,m}\right) \prod_{t=2}^{T} p\left(\vartheta_{k,t,m} \mid \vartheta_{k,t-1,m}\right), \quad (14)
$$

where $p\left(\vartheta_{k,1,m}\right)$ and $p\left(\vartheta_{k,t,m} \mid \vartheta_{k,t-1,m}\right)$ are the transition probabilities that equal to the factor nodes $f_{k,1,m}$ and $f_{k,t,m}$ in Fig. 3.
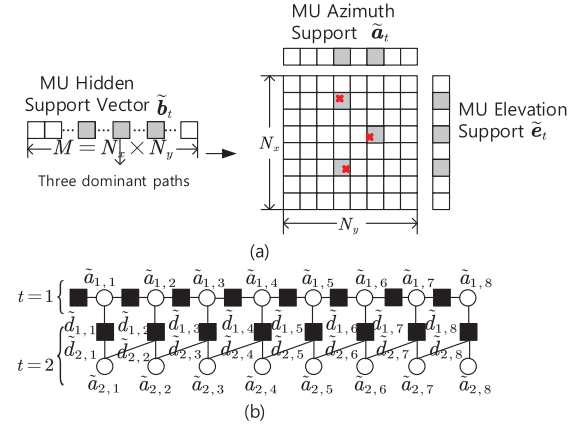

Fig. 4. (a) MU hidden support matrix illustration at time slot t with $N_x = 8$ $N_y = 8$ when three UAV users are considered with three activated paths. (b) Factor graph of 2D-MM of the MU azimuth support vectors when $T = 2$ and $N_x = 8$.

C. 2D-MM for Azimuth and Elevation MU Hidden Support Vector

The details of the azimuth and elevation MU hidden support vector are introduced as follows: To explain the 2D-MM for MU hidden support vector of azimuth and elevation directions, we take three UAV users as an example in Fig. 4(a). The relationship of the MU azimuth support $\tilde{\boldsymbol{a}}_t$ and the MU elevation support $\tilde{\boldsymbol{e}}_t$ with three activated paths are shown. To be noticed, in this example, there are two users share the same MU azimuth support. Hence, the MU azimuth support $\tilde{\boldsymbol{a}}_t$ has only two activated points at time t. For MU elevation support $\tilde{\boldsymbol{e}}_t$ has three activated points at time t.

Because of the navigation of the UAVs [31], together with the known orbital voyage of the LEO satellite, the relative displacement of the UAV and satellite is continuous. Hence, the spatial sparsity pattern will act highly related with the prior adjacent pattern, i.e., $\tilde{a}_{t+1,n_x}$ has higher probability be influenced by the adjacent $\tilde{a}_{t,n_x-1}$ in previous time slot [10]. Reference [17] has also verified that the channel support change slowly over time, which indicates $\tilde{a}_{t+1,n_x}$ has higher probability to be influenced by itself in previous time slot $\tilde{a}_{t,n_x}$. This relationship which considers both spatial correlation and temporal correlation is named as dynamic 2D sparsity. The illustration of this 2D-MM can be found in the factor graph in Fig. 4(b). The model over total time period T can be expressed as:

$$
p\left(\tilde{\boldsymbol{a}}^{(T)}\right)
$$
$$
= \underbrace{p\left(\tilde{a}_{1,1}\right)}_{q_{\tilde{a}_{1,1}}} \prod_{n_x=2}^{N_x} \underbrace{p\left(\tilde{a}_{1,n_x} \mid \tilde{a}_{1,n_x-1}\right)}_{q_{\tilde{a}_{1,n_x},\tilde{a}_{1,n_x-1}}^{(AS)}}
$$
$$
\times \prod_{t=2}^{T} \left(\underbrace{p\left(\tilde{a}_{t,1} \mid \tilde{a}_{t-1,1}\right)}_{q_{\tilde{a}_{t,1},\tilde{a}_{t-1,1}}^{(AT)}} \prod_{n_x=1}^{N_x} \underbrace{p\left(\tilde{a}_{t,n_x} \mid \tilde{a}_{t-1,n_x-1}, \tilde{a}_{t-1,n_x}\right)}_{q_{\tilde{a}_{t,n_x},\tilde{a}_{t-1,n_x-1},\tilde{a}_{t-1,n_x}}^{(A)}} \right),
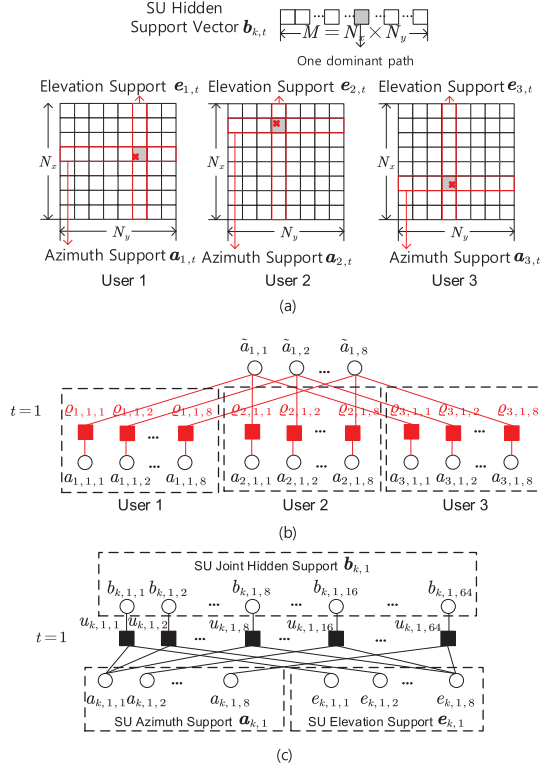$$
(15)

Fig. 5. (a) Illustration of SU joint hidden support matrix based on MU hidden support matrix in Fig.4. (b) HMM of SU azimuth support a_{k,t,n_x} with three users. (c) 3D-HMM Factor graph of the SU joint hidden support vector $b_{k,t}$ when $T = 1$, $N_x = 8$, $N_y = 8$.

where $q_{\tilde{a}_{1,1}}$, $q^{(AS)}_{\tilde{a}_{1,n_x},\tilde{a}_{1,n_x-1}}$, $q^{(AT)}_{\tilde{a}_{t,1},\tilde{a}_{t-1,1}}$ and $q^{(A)}_{\tilde{a}_{t,n_x},\tilde{a}_{t-1,n_x-1},\tilde{a}_{t-1,n_x}}$ are the transition probabilities that equal to the factor nodes $\tilde{d}_{1,1}$, $\tilde{d}_{1,n_x}$, $\tilde{d}_{t,1}$ and $\tilde{d}_{t,n_x}$ in Fig. 4(b) of the azimuth directions. To be noticed, the MU azimuth support $\tilde{a}_{t,n_x} \in \{0,1\}$. The support of the MU elevation support of AoD $p\left(\tilde{e}^{(T)}\right)$ can be modeled similarly.

D. 3D-HMM for Azimuth and Elevation SU Joint Hidden Support Vector

To explore the relationship between MU and SU, we modeled the joint prior distribution of MU and SU azimuth or elevation support as HMM with conditional probability. Take k^{th} UAV SU azimuth support as an example:

$$p\left(\tilde{a}^{(T)}, a_k^{(T)}\right) = \prod_{t=1}^{T}\prod_{k=1}^{K} p\left(\tilde{a}_t\right) p\left(a_{k,t}|\tilde{a}_t\right)$$
$$= \prod_{t=1}^{T}\prod_{k=1}^{K}\prod_{n_x=1}^{N_x} p\left(\tilde{a}_t\right) \underbrace{p\left(a_{k,t,n_x}|\tilde{a}_{t,n_x}\right)}_{\varsigma^{(A)}_{k,\tilde{a}_{t,n_x},a_{k,t,n_x}}}. \quad (16)$$

where $\varsigma^{(A)}_{k,\tilde{a}_{t,n_x},a_{k,t,n_x}}$ is the transition probability equal to the factor node ϱ_{k,t,n_x} in Fig.5(b). Furthermore, $\tilde{a}^{(T)}$ and $a_k^{(T)}$ are the MU azimuth support vector and k^{th} UAV SU azimuth support vector over time period T, respectively. Although in

practice, multiple UAVs may share partially MU azimuth or elevation support, i.e., in Fig.5, user 2 and user 3 both have the 4th grid activated which means the same azimuth support is shared, our HMM model can still apply. To be noticed, each element in $a_k^{(T)}$ is generated by the MU azimuth support vector in the HMM with certain conditional probability, i.e., $p(a_{k,t,n_x} = 1|\tilde{a}_{t,n_x} = 0) = \varsigma^{(A)}_{k,t,0,1} = 0$ and $p(a_{k,t,n_x} = 1|\tilde{a}_{t,n_x} = 0) = \varsigma^{(A)}_{k,t,1,1}$. In another word, when the MU azimuth support element $\tilde{a}_{t,n_x} = 0$, there is no way that the kth UAV SU azimuth support element $a_{k,t,n_x} = 1$. However, when the MU azimuth support element $\tilde{a}_{t,n_x} = 1$, there is a certain probability $\varsigma^{(A)}_{k,t,1,1}$ that the n_x^{th} element a_{k,t,n_x} in the kth UAV SU azimuth support vector is 1. The kth UAV SU elevation support can be modeled similarly.

Hence, the 3D-HMM of the SU joint hidden support $b_k^{(T)}$ over total period time T which combines both azimuth and elevation direction supports can be found in Fig. 5(c). The joint conditional prior of the channel support probability of the kth UAV is given by

$$p\left(b_k^{(T)}|a_k^{(T)}, e_k^{(T)}\right)$$
$$= \prod_{t=1}^{T}\prod_{n_x=1}^{N_x}\prod_{n_y=1}^{N_y} p\left(b_{k,t,n_x,n_y}|a_{k,t,n_x}, e_{k,t,n_y}\right). \quad (17)$$

IV. 3D OFF-GRID DYNAMIC SPARSE CHANNEL TRACKING PROBLEM FORMULATION

By taking the off-grid channel (5) into consideration, the received signal at k^{th} UAV can be written as

$$y_{k,t}^{RX} = g_{k,t}\tilde{A}^H\left(\tilde{\theta}_{k,t} + \Delta\theta_{k,t}, \tilde{\phi}_{k,t} + \Delta\phi_{k,t}\right) S_k + n_{k,t}^{RX}. \quad (18)$$

with $y_{k,t}^{RX} \in \mathbb{C}^{1\times N_S}$. To convert the received signal into the standard CS model, we reshape $g_{k,t}$ from $1 \times M$ to $M \times 1$. Hence the received signal can be rewritten as

$$y_{k,t} = \Phi_{k,t}\left(\Delta\theta_{k,t}, \Delta\phi_{k,t}\right) g_{k,t} + n_{k,t}, \quad (19)$$

with the sensing matrix $\Phi_{k,t}\left(\Delta\theta_{k,t}, \Delta\phi_{k,t}\right) = S_k^T\left(\tilde{A}^H\left(\tilde{\theta}_{k,t} + \Delta\theta_{k,t}, \tilde{\phi}_{k,t} + \Delta\phi_{k,t}\right)\right)^T \in \mathbb{C}^{N_S\times M}$, $y_{k,t} \in \mathbb{C}^{N_S\times 1}$. For the training phase, the transmitted pilot S_k is designed based on the partial discrete Fourier transform (DFT) random permutation measurement matrix $\Phi_{k,t}\left(\Delta\theta_{k,t}, \Delta\phi_{k,t}\right)$ [11], [32]. Details are shown in **Appendix A**. Given the received signals of total t time period $y_k^{(t)}$, the purpose is to track the dynamic channel vector $g_{k,t}$ at time t. The channel vector can be estimated as $\hat{g}_{k,t,m} = E\left(g_{k,t,m}|y_k^{(t)}, \Delta\theta_{k,t}, \Delta\phi_{k,t}\right)$, when the expectation is over the marginal posterior

$$p\left(g_{k,t,m}|y_k^{(t)}, \Delta\theta_{k,t}, \Delta\phi_{k,t}\right)$$
$$\propto \sum_{b_k^{(t)}}\int_{g_{-(k,t,m)}^{(t)}, \vartheta_k^{(t)}} p$$
$$\times \left(g_k^{(t)}, \tilde{a}^{(t)}, \tilde{e}^{(t)}, b_k^{(t)}, \vartheta_k^{(t)}, y_k^{(t)}|\Delta\theta_{k,t}, \Delta\phi_{k,t}\right) \quad (20)$$
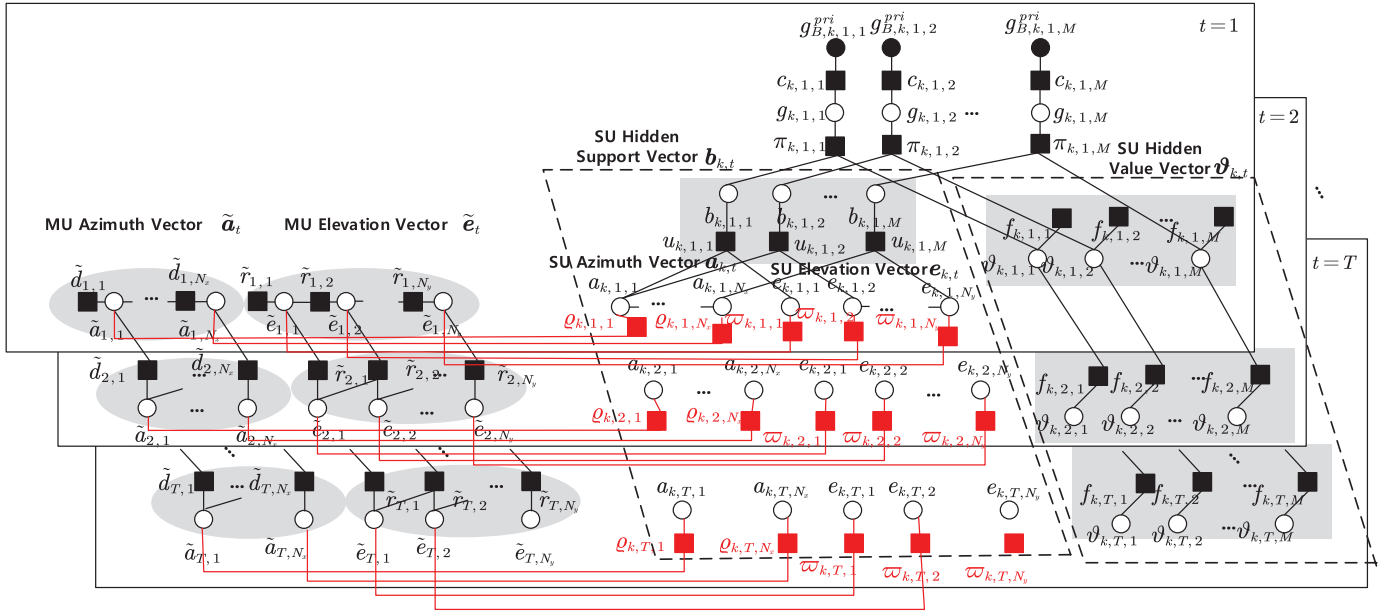
Fig. 6. Factor graph representation of the MD-MM.

with $\boldsymbol{g}_{-(k,t,m)}^{(t)}$ denotes the vector $\boldsymbol{g}_k^{(t)}$ without the element $g_{k,t,m}$. Our goal is to obtain the MMSE of the estimated $\boldsymbol{g}_k$. At the same time, the optimal off-grid azimuth and elevation offset $\Delta\boldsymbol{\theta}_{k,t}^{(opt)}$ and $\Delta\boldsymbol{\phi}_{k,t}^{(opt)}$ are obtained by maximum likelihood (ML) as follows,

$$
\begin{aligned}
&\Delta\boldsymbol{\theta}_{k,t}^{(opt)}, \Delta\boldsymbol{\phi}_{k,t}^{(opt)} \\
&= \underset{\Delta\boldsymbol{\theta}_{k,t}, \Delta\boldsymbol{\phi}_{k,t}}{\operatorname{argmax}} \ \ln p\left(\boldsymbol{y}_k^{(t)}, \Delta\boldsymbol{\theta}_{k,t}, \Delta\boldsymbol{\phi}_{k,t}\right) \\
&= \underset{\Delta\boldsymbol{\theta}_{k,t}, \Delta\boldsymbol{\phi}_{k,t}}{\operatorname{argmax}} \int_{\boldsymbol{g}_k^{(t)}} p\left(\boldsymbol{g}_k^{(t)}, \boldsymbol{y}_k^{(t)}, \Delta\boldsymbol{\theta}_{k,t}, \Delta\boldsymbol{\phi}_{k,t}\right) d\boldsymbol{g}_k^{(t)}. \quad (21)
\end{aligned}
$$

V. MULTI-DIMENSIONAL DYNAMIC TURBO APPROXIMATE MESSAGE PASSING ALGORITHM WITH 3D OFF-GRID ESTIMATION

Approximate Message Passing (AMP) based algorithms have been widely studied for user detection [33], [34] and sparse channel estimation [30], [35], [36] because of its good reconstruction accuracy and low complexity. Traditionally, AMP is applied based on a static prior information. However, the prior distribution knowledge is time-varying in our structured dynamic channel model. As can be seen from the factor graph Fig. 6, the complex structure makes it more challenging to gain the exact marginal posterior in (20). We further employ a turbo iterative framework [37] to improve the reconstruction performance. The details of the proposed algorithm is elaborated in this section.

The factor graph of the total MD-MM distribution $p(\boldsymbol{g}_{\mathcal{B},t}^{pri}, \tilde{\boldsymbol{a}}^{(t)}, \tilde{\boldsymbol{e}}^{(t)}, \boldsymbol{b}_k^{(t)}, \boldsymbol{\vartheta}_k^{(t)}, \boldsymbol{y}_k^{(t)})$ is shown in Fig. 6. This factor graph contains three sub-graphs: the MU azimuth and elevation vector $\tilde{\boldsymbol{a}}_t$ and $\tilde{\boldsymbol{e}}_t$ (on the left-hand side of Fig. 6), the SU hidden support vector $\boldsymbol{b}_{k,t}$ (in the middle of Fig. 6), and the kth UAV SU hidden value vector $\boldsymbol{\vartheta}_{k,t}$ (on the right-hand side of Fig. 6) all through time T. The MU azimuth and

elevation vector $\tilde{\boldsymbol{a}}_t$ and $\tilde{\boldsymbol{e}}_t$ are both structured as 2D-MM. The kth UAV SU hidden support vector $\boldsymbol{b}_{k,t}$ is structured as 3D-HMM which contains both SU azimuth and elevation vector $\boldsymbol{a}_{k,t}$ and $\boldsymbol{e}_{k,t}$. The details of all the necessary nodes in the factor graph are already explained in Section III. The expression of the factor nodes, distributions and functional forms are shown in Table II.

Due to the complex structure of the total factor graph of the MD-MM, it is challenging to achieve the exact marginal posterior in (20). In this section, to explore the dynamic channel structure, we explain the structural details of the proposed algorithm based on the turbo framework with the proposed MD-MM MMSE denoiser. In [37], conventional turbo compressed sensing is achieved based on a two-step iteration framework: linear minimum mean-square error (LMMSE) estimator and i.i.d. prior-based MMSE estimator. Our proposed algorithm will be introduced in four parts: Module $\mathcal{A}$ LMMSE Estimator, Module $\mathcal{B}$ MD-MM MMSE denoiser, message passing through time, and 3D off-grid estimation.

A. Module $\mathcal{A}$: LMMSE Estimator

In Module $\mathcal{A}$, the angular channel vector $\boldsymbol{g}_{k,t}$ at time t for k^{th} UAV is estimated based on the received signal $\boldsymbol{y}_{k,t}$ with the prior distribution $\mathcal{CN}\left(\boldsymbol{g}_{k,t}; \boldsymbol{g}_{\mathcal{A},k,t}^{pri}, v_{\mathcal{A},k,t}^{pri}\boldsymbol{I}\right)$. The extrinsic mean and variance are defined as [37]

$$
\boldsymbol{g}_{\mathcal{A},k,t}^{post} = \boldsymbol{g}_{\mathcal{A},k,t}^{pri} + \frac{v_{\mathcal{A},k,t}^{pri}}{v_{\mathcal{A},k,t}^{pri} + \sigma_e^2} \boldsymbol{\Phi}_{k,t}^H \left(\boldsymbol{y}_{k,t} - \boldsymbol{\Phi}_{k,t}\boldsymbol{g}_{\mathcal{A},k,t}^{pri}\right) \quad (22)
$$

and

$$
v_{\mathcal{A},k,t}^{post} = v_{\mathcal{A},k,t}^{pri} - \frac{N_S}{M} \frac{\left(v_{\mathcal{A},k,t}^{pri}\right)^2}{v_{\mathcal{A},k,t}^{pri} + \sigma_e^2}, \quad (23)
$$

TABLE II
TABLE OF FACTORS, DISTRIBUTIONS AND FUNCTIONAL FORMS ASSOCIATED WITH OUR MD-MM

Factor	Distribution	Function Form
$c_{k,t,m}\left(g_{k,t,m}, g_{\mathcal{B},k,t,m}^{pri}\right)$	$p(g_{\mathcal{B},k,t,m}^{pri}\|g_{k,t,m})$	$\mathcal{CN}\left(\boldsymbol{g}_{k,t}; \boldsymbol{g}_{\mathcal{B},k,t}^{pri}, v_{\mathcal{B},k,t}^{pri}\boldsymbol{I}\right)$
$\pi_{k,t,m}\left(g_{k,t,m}, b_{k,t,m}, \vartheta_{k,t,m}\right)$	$p\left(g_{k,t,m}\|b_{k,t,m}, \vartheta_{k,t,m}\right)$	$\delta\left(g_{k,t,m} - b_{k,t,m}\vartheta_{k,t,m}\right)$
$f_{k,1,m}\left(\vartheta_{k,1,m}\right)$	$p\left(\vartheta_{k,1,m}\right)$	$\mathcal{CN}\left(\vartheta_{k,1,m}; \mu, \sigma^2\right)$
$f_{k,t,m}\left(\vartheta_{k,t,m}, \vartheta_{k,t-1,m}\right)$	$p\left(\vartheta_{k,t,m}\|\vartheta_{k,t-1,m}\right)$	$\mathcal{CN}\left(\vartheta_{k,t,m}; (1-\beta)\vartheta_{k,t-1,m} + \beta\mu, \beta^2\zeta\right)$
$\tilde{d}_{1,1}(\tilde{a}_{1,1})$	$p(\tilde{a}_{1,1})$	$(1-\lambda)^{1-\tilde{a}_{1,1}}\lambda^{\tilde{a}_{1,1}}$
$\tilde{d}_{1,m}(\tilde{a}_{1,m}, \tilde{a}_{1,m-1})$	$p(\tilde{a}_{1,m}\|\tilde{a}_{1,m-1})$	$\begin{cases}\left(q_{0,1}^{(AS)}\right)^{\tilde{a}_{1,m}}\left(1-q_{0,1}^{(AS)}\right)^{(1-\tilde{a}_{1,m})}, \tilde{a}_{1,m-1}=0 \\ \left(1-q_{1,0}^{(AS)}\right)^{\tilde{a}_{1,m}}\left(q_{1,0}^{(AS)}\right)^{(1-\tilde{a}_{1,m})}, \tilde{a}_{1,m-1}=1\end{cases}$
$\tilde{d}_{t,1}(\tilde{a}_{t,1}, \tilde{a}_{t-1,1})$	$p(\tilde{a}_{t,1}\|\tilde{a}_{t-1,1})$	$\begin{cases}\left(q_{0,1}^{(AT)}\right)^{\tilde{a}_{t,1}}\left(1-q_{0,1}^{(AT)}\right)^{(1-\tilde{a}_{t,1})}, \tilde{a}_{t-1,1}=0 \\ \left(1-q_{1,0}^{(AT)}\right)^{\tilde{a}_{t,1}}\left(q_{1,0}^{(AT)}\right)^{(1-\tilde{a}_{t,1})}, \tilde{a}_{t-1,1}=1\end{cases}$
$\tilde{d}_{t,m}(\tilde{a}_{t,m}, \tilde{a}_{t,m-1}, \tilde{a}_{t-1,m})$	$p(\tilde{a}_{t,m}\|\tilde{a}_{t-1,m-1}, \tilde{a}_{t-1,m})$	$\begin{cases}\left(q_{\tilde{a}_{t-1,m-1},\tilde{a}_{t-1,m},1}^{(A)}\right)^{\tilde{a}_{t,m}}\left(1-q_{\tilde{a}_{t-1,m-1},\tilde{a}_{t-1,m},1}^{(A)}\right)^{(1-\tilde{a}_{t,m})} \\ \tilde{a}_{t-1,m-1},\tilde{a}_{t-1,m}\in\{0,1\}\end{cases}$
$\tilde{r}_{1,1}(\tilde{e}_{1,1})$	$p(\tilde{e}_{1,1})$	$(1-\lambda)^{1-\tilde{e}_{1,1}}\lambda^{\tilde{e}_{1,1}}$
$\tilde{r}_{1,m}(\tilde{e}_{1,m}, \tilde{e}_{1,m-1})$	$p(\tilde{e}_{1,m}\|\tilde{e}_{1,m-1})$	$\begin{cases}\left(q_{0,1}^{(ES)}\right)^{\tilde{e}_{1,m}}\left(1-q_{0,1}^{(ES)}\right)^{(1-\tilde{e}_{1,m})}, \tilde{e}_{1,m-1}=0 \\ \left(1-q_{1,0}^{(ES)}\right)^{\tilde{e}_{1,m}}\left(q_{1,0}^{(ES)}\right)^{(1-\tilde{e}_{1,m})}, \tilde{e}_{1,m-1}=1\end{cases}$
$\tilde{r}_{t,1}(\tilde{e}_{t,1}, \tilde{e}_{t-1,1})$	$p(\tilde{e}_{t,1}\|\tilde{e}_{t-1,1})$	$\begin{cases}\left(q_{0,1}^{(ET)}\right)^{\tilde{e}_{t,1}}\left(1-q_{0,1}^{(ET)}\right)^{(1-\tilde{e}_{t,1})}, \tilde{e}_{t-1,1}=0 \\ \left(1-q_{1,0}^{(ET)}\right)^{\tilde{e}_{t,1}}\left(q_{1,0}^{(ET)}\right)^{(1-\tilde{e}_{t,1})}, \tilde{e}_{t-1,1}=1\end{cases}$
$\tilde{r}_{t,m}(\tilde{e}_{t,m}, \tilde{e}_{t-1,m-1}, \tilde{e}_{t-1,m})$	$p(\tilde{e}_{t,m}\|\tilde{e}_{t-1,m-1}, \tilde{e}_{t-1,m})$	$\begin{cases}\left(q_{\tilde{e}_{t-1,m-1},\tilde{e}_{t-1,m},1}^{(E)}\right)^{\tilde{e}_{t,m}}\left(1-q_{\tilde{e}_{t-1,m-1},\tilde{e}_{t-1,m},1}^{(E)}\right)^{(1-\tilde{e}_{t,m})} \\ \tilde{e}_{t-1,m-1},\tilde{e}_{t-1,m}\in\{0,1\}\end{cases}$
$\varrho_{k,t,m}(\tilde{a}_{t,n_x}, a_{k,t,n_x})$	$p(a_{k,t,n_x}\|\tilde{a}_{t,n_x})$	$p(a_{k,t,n_x}=1\|\tilde{a}_{t,n_x}=0)=0,\ p(a_{k,t,n_x}=1\|\tilde{a}_{t,n_x}=1)=\varsigma_{k,t,1,1}^{(A)}$
$\varpi_{k,t,m}(\tilde{e}_{t,n_x}, e_{k,t,n_x})$	$p(e_{k,t,n_x}\|\tilde{e}_{t,n_x})$	$p(e_{k,t,n_x}=1\|\tilde{e}_{t,n_x}=0)=0,\ p(e_{k,t,n_x}=1\|\tilde{e}_{t,n_x}=1)=\varsigma_{k,t,1,1}^{(E)}$
$u_{k,t,m}(b_{k,t,m}, a_{k,t,n_x}, e_{k,t,n_y})$	$p(b_{k,1,m}\|a_{k,t,n_x}, e_{k,t,n_y})$	$a_{k,t,n_x}e_{k,t,n_y}(\gamma_{\boldsymbol{b}_{k,t}})^{b_{k,t,m}}(1-\gamma_{\boldsymbol{b}_{k,t}})^{(1-b_{k,t,m})} + (1-a_{k,t,n_x}e_{k,t,n_y})\delta(b_{k,t,m})$

respectively. The extrinsic distribution of $\boldsymbol{g}_{k,t}$ is given by

$$\mathcal{CN}\left(\boldsymbol{g}_{k,t}; \boldsymbol{g}_{\mathcal{A},k,t}^{post}, v_{\mathcal{A},k,t}^{post}\boldsymbol{I}\right)$$
$$\propto \mathcal{CN}\left(\boldsymbol{g}_{k,t}; \boldsymbol{g}_{\mathcal{A},k,t}^{pri}, v_{\mathcal{A},k,t}^{pri}\boldsymbol{I}\right)\mathcal{CN}\left(\boldsymbol{g}_{k,t}; \boldsymbol{g}_{\mathcal{A},k,t}^{ext}, v_{\mathcal{A},k,t}^{ext}\boldsymbol{I}\right), \quad (24)$$

where the extrinsic mean and variance are respectively given by

$$\boldsymbol{g}_{\mathcal{A},k,t}^{ext} = v_{\mathcal{A},k,t}^{ext}\left(\frac{\boldsymbol{g}_{\mathcal{A},k,t}^{post}}{v_{\mathcal{A},k,t}^{post}} - \frac{\boldsymbol{g}_{\mathcal{A},k,t}^{pri}}{v_{\mathcal{A},k,t}^{pri}}\right) \quad (25)$$

and

$$v_{\mathcal{A},k,t}^{ext} = \left(\frac{1}{v_{\mathcal{A},k,t}^{post}} - \frac{1}{v_{\mathcal{A},k,t}^{pri}}\right)^{-1}. \quad (26)$$

B. Module $\mathcal{B}$: Message Passing MMSE Denoiser

Following LMMSE estimator in module $\mathcal{A}$, the main task for module $\mathcal{B}$ is to exploit the kth UAV spatial sparsity details of the proposed MD-MM in the multi-UAVs system. Hence, the extrinsic mean and variance pass through the graph, which means $\boldsymbol{g}_{\mathcal{B},k,t}^{pri} = \boldsymbol{g}_{\mathcal{A},k,t}^{ext}$ and $v_{\mathcal{B},k,t}^{pri} = v_{\mathcal{A},k,t}^{ext}$ [37]. In module $\mathcal{B}$, the purpose is to calculate the approximate posterior distributions of $p(g_{k,t,m}\|\boldsymbol{g}_{\mathcal{B},k}^{pri})$ instead of (20) using the sum-product message passing structure shown in the factor graph. For message passing algorithm, the basic assumption

is that $\boldsymbol{g}_{\mathcal{B},k,t}^{pri} = \boldsymbol{g}_{k,t} + \boldsymbol{z}_{k,t}$ where $\boldsymbol{z}_{k,t} \sim \mathcal{CN}(0, v_{\mathcal{B},k,t}^{pri}\boldsymbol{I})$ is independent of $\boldsymbol{g}_{k,t}$ [37].

The formulation details are elaborated in **Appendix B** and the message passing procedure can be summarized as follows:

- Firstly, in module $\mathcal{A}$, based on the prior information passed from the last time slot t, the message passing over the kth UAV Gaussian hidden value vector $\boldsymbol{\vartheta}_{k,t}$ are given by (52) with μ^{dyna} and $(\sigma^{dyna})^2$ as the input to update the message $v_{\vartheta_{k,t,m}\to\pi_{k,t,m}}$.
- Secondly, in module $\mathcal{B}$, $\boldsymbol{g}_{\mathcal{B},k,t}^{pri}$ and $v_{\mathcal{B},k,t}^{pri}$ from module $\mathcal{A}$ are passed over the path $g_{k,t,m} \to \pi_{t,m} \to b_{k,t,m} \to u_{k,t,m} \to a_{k,t,n_x}$ using (53) to (56).
- Thirdly, the forward-backward message passing is executed over the proposed MD-MM MU azimuth branch with three paths: $a_{k,t,n_x} \to \varrho_{k,t,n_x} \to \tilde{a}_{t,n_x}$, the 2D-MM of the MU azimuth support $\tilde{\boldsymbol{a}}_t$, and $\tilde{a}_{t,n_x} \to \varrho_{k,t,n_x} \to a_{k,t,n_x}$ using (59) to (68).
- Then the message pass through path $a_{t,n_x} \to u_{t,m} \to e_{t,n_y}$ (using (69) to (71)) to another MD-MM elevation branch with similar three paths: $e_{k,t,n_y} \to \varpi_{k,t,n_y} \to \tilde{e}_{t,n_y}$, the 2D-MM of the MU elevation support $\tilde{\boldsymbol{e}}_t$, and $\tilde{e}_{t,n_y} \to \varpi_{k,t,n_y} \to e_{k,t,n_y}$ using (71) to (82). The overall details of the MD-MM azimuth and elevation SU joint hidden support estimation can be found in **Algorithm 1**.
- The message is finally passed back over the path $e_{k,t,n_y} \to u_{k,t,m} \to b_{k,t,m} \to \pi_{k,t,m} \to g_{k,t,m}$ using (83) to (86).

Algorithm 1 MD-MM k^{th} SU Azimuth and Elevation Support Estimation

---

Input: $\tilde{\rho}^{(A)in}_{t,n_x}$, $\tilde{\rho}^{(A)dyna}_{t,n_x}$, $\tilde{\rho}^{(E)in}_{t,n_y}$, $\tilde{\rho}^{(E)dyna}_{t,n_y}$, $\rho^{(A)in}_{k,t,n_x}$, $\rho^{(E)in}_{k,t,n_y}$

Output: $\rho^{(A)out}_{k,t,n_x}$, $\rho^{(E)out}_{k,t,n_y}$

1: Calculate $\upsilon_{u_{k,t,m}\to a_{k,t,n_x}}$ as in (59) and $\upsilon_{\varrho_{k,t,n_x}\to\tilde{a}_{t,n_x}}$ as in (68)

   # HMM k^{th} SU Azimuth Support Estimation Based on 2D-MM MU Azimuth Support:

2: **if** $t = 1$ **then**

3:   **Initialization**: $\tilde{\gamma}^{(A)f}_{t,1} = \lambda$, $\tilde{\gamma}^{(A)b}_{t,N_x} = \frac{1}{2}$

4:   **for** $n_x = 2, \ldots, N_x$ **do**

5:     Use (62) to calculate $\tilde{\gamma}^{(A)f}_{t,n_x}$.

6:   **end for**

7:   **for** $n_x = 1, \ldots, N_x - 1$ **do**

8:     Use (63) to calculate $\tilde{\gamma}^{(A)b}_{t,n_x}$.

9:   **end for**

10: **else if** $t > 1$ **then**

11:   **Initialization**: $\tilde{\gamma}^f_{t,1} = q^{(T)}_{0,1}\left(1 - \tilde{\rho}^{dyna}_{t,1}\right) + \left(1 - q^{(T)}_{1,0}\right)\tilde{\rho}^{dyna}_{t,1}$, $\tilde{\gamma}^b_{t,N_x} = \frac{1}{2}$.

12:   **for** $n_x = 2, \ldots, N_x$ **do**

13:     Calculate the $\tilde{\gamma}^{(A)f}_{t,n_x}$ as in (64)

14:   **end for**

15:   **for** $n_x = 1, \ldots, N_x - 1$ **do**

16:     Calculate $\tilde{\gamma}^{(A)b}_{t,n_x} = \frac{1}{1+\tilde{\gamma}^{(A)}}$ use (65).

17:   **end for**

18: **end if**

19: Calculate the kth SU AoD azimuth support vector $\rho^{(A)out}_{k,t,n_x} = \tilde{\rho}^{out}_{\tilde{a}_{t},k,n_x}\varsigma^{(A)}_{k,t,1,1}$.

   #Message Passing Over k^{th} SU Azimuth to Elevation Support:

20: Calculate individual message $\rho^{out}_{\boldsymbol{a}_{k,t},k,t,n_x,n_y}$ using (70).

21: Calculate message $\upsilon_{a_{k,t,n_x}\to u_{k,t,m}}(a_{k,t,n_x})$ and $\upsilon_{u_{k,t,m}\to e_{k,t,n_y}}(e_{k,t,n_y})$ as in (69) and (71), respectively.

22: Calculate $\upsilon_{u_{k,t,m}\to e_{k,t,n_y}}$ as in (73) and $\upsilon_{\varpi_{k,t,n_y}\to\tilde{e}_{t,n_y}} \propto \sum_{e_{k,t,n_y}}$ as in (82).

   # HMM k^{th} SU Elevation Support Estimation Based on 2D-MM MU Elevation:

23: Similar to the procedures from step 2 to step 18 by using elevation related parameters. Details can be found from (76) to (79).

24: Calculate the kth SU AoD elevation support vector $\rho^{(E)out}_{k,t,n_y} = \tilde{\rho}^{out}_{\tilde{e}_{t},k,n_y}\varsigma^{(E)}_{k,t,1,1}$.

Based on the calculated updated messages $\upsilon_{\pi_{k,t,m}\to g_{k,t,m}}$, the posterior distributions can be written as

$$p\left(g_{k,t,m}|\boldsymbol{g}^{pri}_{\mathcal{B},k,t}\right)$$
$$\propto \upsilon_{\pi_{k,t,m}\to g_{k,t,m}}(g_{k,t,m})\,\upsilon_{c_{k,t,m}\to g_{k,t,m}}(g_{k,t,m}), \quad (27)$$

where $\upsilon_{c_{k,t,m}\to g_{k,t,m}}(g_{k,t,m}) = \mathcal{CN}\left(g_{k,t,m}; g^{pri}_{\mathcal{B},k,t,m}, v^{pri}_{\mathcal{B},k,t}\right)$. Then the posterior mean and variance of each element of $\boldsymbol{g}_{k,t}$ can be updated as

$$g^{post}_{\mathcal{B},k,t,m} = \int_{g_{k,t,m}} g_{k,t,m}p\left(g_{k,t,m}|\boldsymbol{g}^{pri}_{\mathcal{B},k,t}\right) \quad (28)$$

and

$$v^{post}_{\mathcal{B},k,t} = \frac{1}{M}\sum_{m=1}^{M}\mathrm{Var}\left(g_{k,t,m}|\boldsymbol{g}^{pri}_{\mathcal{B},k,t}\right), \quad (29)$$

where $\mathrm{Var}\left(g_{k,t,m}|\boldsymbol{g}^{pri}_{\mathcal{B},k,t}\right)$ denotes the conditional variance of $g_{k,t,m}$ given $\boldsymbol{g}^{pri}_{\mathcal{B},k,t}$. Then similar to (25) and (26), the extrinsic update mean and covariance can be written as

$$\boldsymbol{g}^{pri}_{\mathcal{A},k,t} = \boldsymbol{g}^{ext}_{\mathcal{B},k,t} = v^{pri}_{\mathcal{A},k,t}\left(\frac{\boldsymbol{g}^{post}_{\mathcal{B},k,t}}{v^{post}_{\mathcal{B},k,t}} - \frac{\boldsymbol{g}^{pri}_{\mathcal{B},k,t}}{v^{pri}_{\mathcal{B},k,t}}\right) \quad (30)$$

and

$$v^{pri}_{\mathcal{A},k,t} = v^{ext}_{\mathcal{B},k,t} = \left(\frac{1}{v^{post}_{\mathcal{B},k,t}} - \frac{1}{v^{pri}_{\mathcal{B},k,t}}\right)^{-1}. \quad (31)$$

C. Message Passing Through Time

For dynamic scenario, prior information which is also treated as the temporal correlation can be considered for better reconstruction of the time-varying channel model. Hence, we further consider the message passing through time for channel tracking. There are mainly three branches passing through time in Fig. 6 the factor graph of the MD-MM: the path of the kth UAV Gaussian hidden value vector $\boldsymbol{\vartheta}_{k,t}$, the path of MU azimuth support $\tilde{\boldsymbol{a}}_t$ and the path of MU elevation support $\tilde{\boldsymbol{e}}_t$.

1) The kth UAV Gaussian Hidden Value Vector $\boldsymbol{\vartheta}_{k,t}$ Passing Through Time: The message passing from factor node $\pi_{k,t,m}$ to variable node $\vartheta_{k,t,m}$ should be

$$\upsilon^{true}_{\pi_{k,t,m}\to\vartheta_{k,t,m}}(\vartheta_{k,t,m})$$
$$= \rho^{out}_{k,t,m}\mathcal{CN}\left(\vartheta_{k,t,m}; \boldsymbol{g}^{pri}_{\mathcal{B},k,t}, v^{pri}_{\mathcal{B},k,t}\right)$$
$$+ \left(1 - \rho^{out}_{k,t,m}\right)\mathcal{CN}\left(0; \boldsymbol{g}^{pri}_{\mathcal{B},k,t}, v^{pri}_{\mathcal{B},k,t}\right). \quad (32)$$

To be noticed, if $b_{k,t,m} = 0$, $g_{k,t,m}$ would be 0 in (9), which makes $\vartheta_{k,t,m}$ unobservable. Hence, similarly in [30], a threshold which is slightly less than 1 is introduced. The modified message passing is given by

$$\upsilon_{\pi_{k,t,m}\to\vartheta_{k,t,m}}(\vartheta_{k,t,m}) = \mathcal{CN}\left(\vartheta_{k,t,m}; \mu^{in}_{k,t,m}, (\sigma^{in}_{k,t,m})^2\right), \quad (33)$$

with

$$\left(\mu^{in}_{k,t,m}, (\sigma^{in}_{k,t,m})^2\right) = \begin{cases} \left(\frac{1}{\epsilon}g^{pri}_{\mathcal{B},k,t}, \frac{1}{\epsilon^2}v^{pri}_{\mathcal{B},k,t}\right) & , \rho^{out}_{k,t,m} \leqslant \mathrm{Th}, \\ \left(g^{pri}_{\mathcal{B},k,t}, v^{pri}_{\mathcal{B},k,t}\right) & , \rho^{out}_{k,t,m} > \mathrm{Th}, \end{cases} \quad (34)$$

where ϵ is set as a small positive value close to 0. Then the message passing from factor node $f_{k,t+1,m}$ to variable node $\vartheta_{k,t+1,m}$ is formed as

$$\upsilon_{f_{k,t+1,m}\to\vartheta_{k,t+1,m}}(\vartheta_{k,t+1,m})$$
$$= \mathcal{CN}\left(\vartheta_{k,t+1,m}; \mu^{dyna}_{k,t+1,m}, \left(\sigma^{dyna}_{k,t+1,m}\right)^2\right), \quad (35)$$

where

$$
\mu_{k,t+1,m}^{dyna} = (1-\beta) \left(\frac{\left(\sigma_{k,t,m}^{dyna}\right)^2 \left(\sigma_{k,t,m}^{in}\right)^2}{\left(\sigma_{k,t,m}^{dyna}\right)^2 + \left(\sigma_{k,t,m}^{in}\right)^2} \right)
$$

$$
\times \left(\frac{\mu_{k,t,m}^{dyna}}{\left(\sigma_{k,t,m}^{dyna}\right)^2} + \frac{\mu_{k,t,m}^{in}}{\left(\sigma_{k,t,m}^{in}\right)^2} \right) + \beta\mu \quad (36)
$$

and

$$
\left(\sigma_{k,t+1,m}^{dyna}\right)^2 = (1-\beta)^2 \left(\frac{\left(\sigma_{k,t,m}^{dyna}\right)^2 \left(\sigma_{k,t,m}^{in}\right)^2}{\left(\sigma_{k,t,m}^{dyna}\right)^2 + \left(\sigma_{k,t,m}^{in}\right)^2} \right) + \beta^2\zeta. \quad (37)
$$

2) The MU Azimuth Support $\tilde{a}_t$ Passing Through Time: The azimuth support prior information passing to the next time $\upsilon_{\tilde{a}_{t,n_x} \to \tilde{d}_{t+1,n_x}}(\tilde{a}_{t,n_x})$ can be given as

$$
\upsilon_{\tilde{a}_{t,n_x} \to \tilde{d}_{t+1,n_x}}(\tilde{a}_{t,n_x})
$$
$$
= \tilde{\rho}_{t+1,n_x}^{(A)dyna} \delta\left(\tilde{a}_{t,n_x} - 1\right) + \left(1 - \tilde{\rho}_{t+1,n_x}^{(A)dyna}\right) \delta\left(\tilde{a}_{t,n_x}\right), \quad (38)
$$

with

$$
\tilde{\rho}_{t+1,n_x}^{(A)dyna}
$$
$$
= \frac{\gamma_{t,n_x}^{(A)f} \gamma_{t,n_x}^{(A)b} \tilde{\rho}_{t,n_x}^{(A)in}}{\gamma_{t,n_x}^{(A)f} \gamma_{t,n_x}^{(A)b} \tilde{\rho}_{t,n_x}^{(A)in} + (1-\gamma_{t,n_x}^{(A)f})(1-\gamma_{t,n_x}^{(A)b})(1-\tilde{\rho}_{t,n_x}^{(A)in})}. \quad (39)
$$

3) The MU Elevation Support $\tilde{e}_t$ Passing Through Time: Similarly, the elevation support prior information passing to the next time can be written as

$$
\upsilon_{\tilde{e}_{t,n_y} \to \tilde{r}_{t+1,n_y}}(\tilde{e}_{t,n_y}) = \tilde{\rho}_{t+1,n_y}^{(E)dyna} \delta\left(\tilde{e}_{t,n_y} - 1\right)
$$
$$
+ \left(1 - \tilde{\rho}_{t+1,n_y}^{(E)dyna}\right) \delta\left(\tilde{e}_{t,n_y}\right), \quad (40)
$$

with

$$
\tilde{\rho}_{t+1,n_y}^{(E)dyna} = \frac{\gamma_{t,n_y}^{(E)f} \gamma_{t,n_y}^{(E)b} \tilde{\rho}_{t,n_y}^{(E)in}}{\gamma_{t,n_y}} \gamma_{t,n_y}^{(E)b} \tilde{\rho}_{t,n_y}^{(E)in}
$$
$$
+ (1-\gamma_{t,n_y}^{(E)f})(1-\gamma_{t,n_y}^{(E)b})(1-\tilde{\rho}_{t,n_y}^{(E)in}). \quad (41)
$$

D. 3D Off-Grid Estimation

As there is no closed-form expression due to the complex MD-MM structure of $g_k^{(t)}$, it's hard to directly maximize the log-likelihood function $\ln p\left(y_k^{(t)}, \Delta\theta_{k,t}, \Delta\phi_{k,t}\right)$ in (21). We extend the proposed in-exact majorization-minimization (MM) algorithm in [38] to help us solve the off-grid problem. As MM is the generalization of the expectation maximization method, the convergence of expectation maximization algorithm can prove that the in-exact MM can converge to a stationary solution [39]. Firstly, we iteratively construct the lower bound for $\ln p\left(y_k^{(t)}, \Delta\theta_{k,t}, \Delta\phi_{k,t}\right)$ which can be named as surrogate function, i.e., for any fixed azimuth and

elevation offset pairs $\left(\Delta\dot{\theta}_{k,t}, \Delta\dot{\phi}_{k,t}\right)$, the surrogate function can be defined as [38]

$$
\mathcal{L}\left(\Delta\theta_{k,t}, \Delta\phi_{k,t} | \Delta\dot{\theta}_{k,t}, \Delta\dot{\phi}_{k,t}\right)
$$
$$
= \int p\left(g_{k,t} | y_k^{(t)}, \Delta\dot{\theta}_{k,t}, \Delta\dot{\phi}_{k,t}\right)
$$
$$
\times \ln \frac{p\left(g_{k,t} y_k^{(t)}, \Delta\dot{\theta}_{k,t}, \Delta\dot{\phi}_{k,t}\right)}{p\left(g_{k,t} | y_k^{(t)}, \Delta\dot{\theta}_{k,t}, \Delta\dot{\phi}_{k,t}\right)} dg_{k,t}, \quad (42)
$$

which satisfies the following properties:

$$
\mathcal{L}\left(\Delta\theta_{k,t}, \Delta\phi_{k,t} | \Delta\dot{\theta}_{k,t}, \Delta\dot{\phi}_{k,t}\right)
$$
$$
\leqslant \ln p\left(y_k^{(t)}, \Delta\theta_{k,t}, \Delta\phi_{k,t}\right), \forall \Delta\theta_{k,t}, \Delta\phi_{k,t}, \quad (43)
$$
$$
\mathcal{L}\left(\Delta\dot{\theta}_{k,t}, \Delta\dot{\phi}_{k,t} | \Delta\dot{\theta}_{k,t}, \Delta\dot{\phi}_{k,t}\right)
$$
$$
= \ln p\left(y_k^{(t)}, \Delta\dot{\theta}_{k,t}, \Delta\dot{\phi}_{k,t}\right), \quad (44)
$$
$$
\left. \frac{\partial\mathcal{L}(\Delta\theta_{k,t}, \Delta\dot{\phi}_{k,t} | \Delta\dot{\theta}_{k,t}, \Delta\dot{\phi}_{k,t})}{\partial\Delta\theta_{k,t}} \right|_{\Delta\theta_{k,t}=\Delta\dot{\theta}_{k,t}}
$$
$$
= \left. \frac{\partial\ln p\left(y_k^{(t)}, \Delta\theta_{k,t}, \Delta\dot{\phi}_{k,t}\right)}{\partial\Delta\theta_{k,t}} \right|_{\Delta\theta_{k,t}=\Delta\dot{\theta}_{k,t}}, \quad (45)
$$
$$
\left. \frac{\partial\mathcal{L}(\Delta\dot{\theta}_{k,t}, \Delta\phi_{k,t} | \Delta\dot{\theta}_{k,t}, \Delta\dot{\phi}_{k,t})}{\partial\Delta\phi_{k,t}} \right|_{\Delta\phi_{k,t}=\Delta\dot{\phi}_{k,t}}
$$
$$
= \left. \frac{\partial\ln p\left(y_k^{(t)}, \Delta\dot{\theta}_{k,t}, \Delta\phi_{k,t}\right)}{\partial\Delta\phi_{k,t}} \right|_{\Delta\phi_{k,t}=\Delta\dot{\phi}_{k,t}}. \quad (46)
$$

Then $\Delta\theta_{k,t}$ and $\Delta\phi_{k,t}$ can be updated at each iteration as

$$
\Delta\theta_{k,t}^{(i+1)} = \underset{\theta_{k,t}}{\operatorname{argmax}} \mathcal{L}\left(\Delta\theta_{k,t}, \Delta\phi_{k,t}^{(i)} | \Delta\theta_{k,t}^{(i)}, \Delta\phi_{k,t}^{(i)}\right), \quad (47)
$$
$$
\Delta\phi_{k,t}^{(i+1)} = \underset{\phi_{k,t}}{\operatorname{argmax}} \mathcal{L}\left(\Delta\theta_{k,t}^{(i+1)}, \Delta\phi_{k,t} | \Delta\theta_{k,t}^{(i+1)}, \Delta\phi_{k,t}^{(i)}\right), \quad (48)
$$

with $(\cdot)^{(i)}$ represents for the $i^{(th)}$ iteration. To be noticed, for the $(i+1)^{th}$ iteration elevation update, we use the updated $(i+1)^{th}$ azimuth offset as fixed known knowledge in (48). However, the problems in (47) and (48) are non-convex. Hence the gradient update can be applied as follows:

$$
\Delta\theta_{k,t}^{(i+1)}
$$
$$
= \Delta\theta_{k,t}^{(i)}
$$
$$
+ \Delta_\theta^{(i)} \left. \frac{\partial\mathcal{L}\left(\Delta\theta_{k,t}, \Delta\phi_{k,t}^{(i)} | \Delta\theta_{k,t}^{(i)}, \Delta\phi_{k,t}^{(i)}\right)}{\partial\Delta\theta_{k,t}} \right|_{\Delta\theta_{k,t}=\Delta\theta_{k,t}^{(i)}} \quad (49)
$$

Algorithm 2 3D Off-Grid Estimation

Input: $g_{\mathcal{B},k,t,m}^{post}$, $v_{\mathcal{B},k,t}^{post}$, $\Delta\boldsymbol{\theta}_{k,t}^{(0)} = \mathbf{0}$, $\Delta\boldsymbol{\phi}_{k,t}^{(0)} = \mathbf{0}$, I.

Output: $\Delta\boldsymbol{\theta}_{k,t}^{(opt)}$, $\Delta\boldsymbol{\phi}_{k,t}^{(opt)}$

1: **for** $i = 1 : I$ **do**
2:   Azimuth offset $\Delta\boldsymbol{\theta}_{k,t}^{(i+1)}$ can be updated use gradient update (49) and (88) with known i^{th} iteration elevation offset $\Delta\boldsymbol{\phi}_{k,t}^{(i)}$.
3:   Elevation offset $\Delta\boldsymbol{\phi}_{k,t}^{(i+1)}$ can be updated similarly with known $(i+1)^{th}$ iteration azimuth offset $\Delta\boldsymbol{\theta}_{k,t}^{(i+1)}$.
4: **end for**
5: $\Delta\boldsymbol{\theta}_{k,t}^{(opt)} = \Delta\boldsymbol{\theta}_{k,t}^{(I)}, \Delta\boldsymbol{\phi}_{k,t}^{(opt)} = \Delta\boldsymbol{\phi}_{k,t}^{(I)}$

---

and

$$
\Delta\phi_{k,t}^{(i+1)} = \Delta\phi_{k,t}^{(i)} + \Delta_{\phi}^{(i)} \frac{\partial\mathcal{L}\left(\Delta\boldsymbol{\theta}_{k,t}^{(i+1)}, \Delta\boldsymbol{\phi}_{k,t}|\Delta\boldsymbol{\theta}_{k,t}^{(i+1)}, \Delta\boldsymbol{\phi}_{k,t}^{(i)}\right)}{\partial\Delta\phi_{k,t}}\Bigg|_{\Delta\phi_{k,t}=\Delta\phi_{k,t}^{(i)}},
$$
(50)

where $\Delta_{\theta}^{(i)}$ and $\Delta_{\phi}^{(i)}$ are the stepsizes. The off-grid estimation algorithm is summarized in **Algorithm 2** and the detailed equation can be found in **Appendix C**. Finally, the overall proposed MD-DTAMP can be summarized in **Algorithm 3**.

VI. NUMERICAL RESULTS

In this section, the complexity and the estimation performance of the proposed algorithm MD-DTAMP with and without 3D off-grid estimation are evaluated. We assume that the proposed algorithm is performed in the distributed UAVs. For the space-air link of each single UAV, the sparsity ratio for both azimuth and elevation directions is $\lambda = \frac{1}{N_x} = \frac{1}{N_y}$ which indicates only one dominant LoS ray for each UAV-satellite link. The parameters are set as follows[3]: We assume that there are total 3 UAVs in the system. The MU azimuth and elevation spatial correlation parameters $\{q_{0,1}^{(AS)}, q_{1,0}^{(AS)}, q_{0,1}^{(ES)}, q_{1,0}^{(ES)}\} = \{0.05, 0.05, 0.025, 0.025\}$, the MU azimuth and elevation temporal correlation parameters $\{q_{0,1}^{(AT)}, q_{1,0}^{(AT)}, q_{0,1}^{(ET)}, q_{1,0}^{(ET)}\} = \{0.05, 0.05, 0.025, 0.175\}$, the MU azimuth and elevation joint parameters $\{q_{1,1,1}^{(A)}, q_{0,0,1}^{(A)}, q_{0,1,1}^{(A)}, q_{1,0,1}^{(A)}, q_{1,1,1}^{(E)}, q_{0,0,1}^{(E)}, q_{0,1,1}^{(E)}, q_{1,0,1}^{(E)}\} = \{0.9958, 0.0013, 0.3276, 0.3936, 0.99546, 0.0007, 0.5, 0.1078\}$, the hidden value vector parameters for k^{th} UAV $\{\mu_k, \sigma_k^2, \beta_k, \zeta_k, \varsigma_{k,t,1,1}^{(A)}, \varsigma_{k,t,1,1}^{(E)}\} = \{0, \frac{1}{3}, 0.5, 1, 0.9375, 0.9375\}$, Th $= 1 - 10^{-2}$, $\epsilon = 10^{-7}$, $\gamma_{\boldsymbol{b}_{k,t}} = 1$, the azimuth range $\theta^{range} = \pi$ and the elevation range $\phi^{range} = \pi/4$, 3D Off-grid azimuth and elevation update step sizes are $\Delta_{\theta}^{(i)} = \pi/(100 \times M)$ and $\Delta_{\phi}^{(i)} = \pi/(4 \times 100 \times M)$.

---

[3]The parameters can be trained through expectation maximization algorithm which can refer to [30] for details. The MU azimuth and elevation related parameters for multiple UAVs system in this paper are referred 2D settings in [20].

Algorithm 3 MD-DTAMP Dynamic Channel Tracking

Input: $\boldsymbol{y}_{k,1}, \ldots, \boldsymbol{y}_{k,T}$, sensing matrix $\boldsymbol{\Phi}(\mathbf{0}, \mathbf{0})$

Output: $\hat{\boldsymbol{g}}_{k,1}, \ldots, \hat{\boldsymbol{g}}_{k,T}$

1: **for** t=1,...,T **do**
2:   **Initialization**: $\boldsymbol{g}_{\mathcal{A},k,t}^{pri} = \mathbf{0}$, $v_{\mathcal{A},k,t}^{pri} = \sigma^2$.
3:   **while** not converge **do**
    #Module $\mathcal{A}$ LMMSE estimator:
4:     Update $\boldsymbol{g}_{\mathcal{A},k,t}^{post}$ and $v_{\mathcal{A},k,t}^{post}$ in (22) and (23), respectively.
5:     Calculate $\boldsymbol{g}_{\mathcal{B},k,t}^{pri} = \boldsymbol{g}_{\mathcal{A},k,t}^{ext}$, $v_{\mathcal{B},k,t}^{pri} = v_{\mathcal{A},k,t}^{ext}$ using (25) and (26).
    #Module $\mathcal{B}$ MD-MM MMSE denoiser:
6:     Message passing over Gaussian hidden value vector $\boldsymbol{\vartheta}_{k,t}$ using (52).
7:     Message Passing Over $g_{k,t,m} \rightarrow \pi_{k,t,m} \rightarrow b_{k,t,m} \rightarrow u_{k,t,m} \rightarrow a_{k,t,n_x}$ using (53)-(58)
8:     AoD SU azimuth support $\boldsymbol{a}_{k,t}$ and SU elevation support $\boldsymbol{e}_{k,t}$ estimation in **Algorithm 1**.
9:     Message Passing Over the Path $e_{k,t,n_y} \rightarrow u_{k,t,m} \rightarrow b_{k,t,m} \rightarrow \pi_{k,t,m} \rightarrow g_{k,t,m}$ using (83)-(86)
10:    Calculate the approximate posterior distributions $p\left(g_{k,t,m}|\boldsymbol{g}_{\mathcal{B},k,t}^{pri}\right)$ using (27).
11:    Update $\boldsymbol{g}_{\mathcal{B},k,t,m}^{post}$ and $v_{\mathcal{B},k,t}^{post}$ using (28) and (29).
12:    Update $\boldsymbol{g}_{\mathcal{A},k,t}^{pri} = \boldsymbol{g}_{\mathcal{B},k,t}^{ext}$ and $v_{\mathcal{A},k,t}^{pri} = v_{\mathcal{B},k,t}^{ext}$ using (30) and (31).
13:    Repeat Module $\mathcal{A}$ and Module $\mathcal{B}$ until convergence.
14:    Hence $\hat{p}\left(\boldsymbol{g}_{k,t}|\boldsymbol{y}_k^{(t)}, \Delta\boldsymbol{\theta}_{k,t}^{(opt)}, \Delta\boldsymbol{\phi}_{k,t}^{(opt)}\right) = \mathcal{CN}\left(\boldsymbol{g}_{\mathcal{B},k,t}^{post}, v_{\mathcal{B},k,t}^{post}\boldsymbol{I}\right)$.
    #3D Off-grid estimation:
15:    Update azimuth offset $\Delta\boldsymbol{\theta}_{k,t}^{(opt)}$ and elevation offset $\Delta\boldsymbol{\phi}_{k,t}^{(opt)}$ use **Algorithm 2**.
16:   **end while**
17:   Update dynamic messages to the next time slot $(t+1)$: $v_{a_{k,t,n_x}\rightarrow d_{k,t+1,n_x}}$, $v_{e_{k,t,n_y}\rightarrow r_{k,t+1,n_y}}$ and $v_{f_{k,t+1,m}\rightarrow\vartheta_{k,t+1,m}}$ using (38), (40) and (35).
18: **end for**

To be noticed, in our simulation, our proposed algorithm MD-DTAMP with 3D off-grid estimation is presented in **Algorithm 3**. MD-DTAMP on grid which is **Algorithm 3** without step 15. The benchmark algorithms which can be employed in our MU 3D channel model are sparse Bayesian learning (SBL) and D-OMP proposed in [38] and [17], respectively. All simulations are average of 500 times.

A. Complexity Analysis

The computation complexity of the LMMSE estimator in module $\mathcal{A}$ and MD-MM MMSE denoiser in module $\mathcal{B}$ are $\mathcal{O}(MN_S)$ and $\mathcal{O}(N_x + N_y)$, respectively. Hence, the total complexity of the proposed MD-DTAMP without 3D off-grid estimation over time T is $\mathcal{O}(T(MN_S + N_x + N_y))$ for each iteration. For off-grid estimation, with defined stepsize $\Delta_{\theta}^{(i)}$, the complexity for update azimuth offset $\Delta\boldsymbol{\theta}_{k,t}^{(i+1)}$ is $\mathcal{O}(N_S M^2)$ per iteration. The reason why the complexity is not $\mathcal{O}(N_S N_x^2)$ is that the original dimension has to
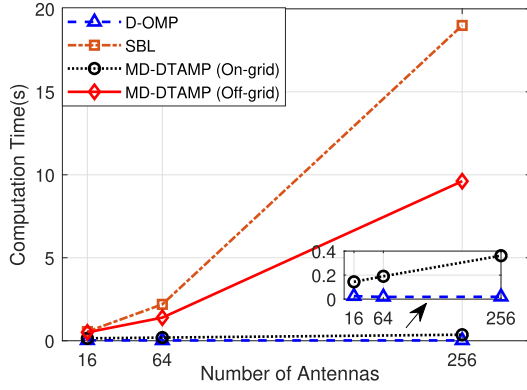
Fig. 7.	Computation time of various algorithms versus the number of antennas when SNR $= 10$dB, $N_s = 10$, $T = 50$, $M = 16, 64, 256$ and $N_x = N_y = 4, 8, 16$.

be extended for paring the azimuth and elevation angles. Similarly, the complexity for update elevation offset $\Delta\phi_{k,t}^{(i+1)}$ is also $\mathcal{O}\left(N_S M^2\right)$ per iteration.[4]

Fig. 7 shows the computation time of various algorithms versus the number of antennas. It can be observed that the proposed on-gird MD-DTAMP shows comparable complexity compared with D-OMP. Furthermore, compared with SBL algorithm which requires matrix multiplication and inversion with high computation complexity, the proposed MD-DTAMP with off-grid estimation consumes less computation time for different number of antennas.

B. Simulation Results

We choose time-averaged normalized mean square error (TNMSE) as the performance metric as [30]

$$\text{TNMSE} = \frac{1}{T} \sum_{t}^{T} \frac{\|\boldsymbol{g}_t - \hat{\boldsymbol{g}}_t\|^2}{\|\boldsymbol{g}_t\|^2}, \tag{51}$$

where $\hat{\boldsymbol{g}}_t$ is the estimated result of $\boldsymbol{g}_t$ in time t. This performance metric is to evaluate the recovery performance of various algorithms.

In Fig. 8, we further compare the TNMSE performance metric versus SNR among various algorithms with antenna array elements set as $M = 64, 256$. With SNR increases, all four algorithms show better recovery performance. To be noticed, our proposed MD-DTAMP on-grid and off-grid tracking algorithms always show superior performance compared to benchmark algorithms. Moreover, the off-grid can further improve the performance through gradient update of the 3D offset.

In Fig. 9, we further consider the relationship between the tracking performance and the training pilot overhead. It is easy to be noticed that our proposed algorithm with or without off-grid estimation can maintain a good performance with smaller pilot overhead. This proves that our proposed algorithms can achieve a solid performance by taking the dynamic channel spatial and temporal correlation into consideration.

[4]To be notice, empirically, the proposed algorithm normally converges within 30 iterations.
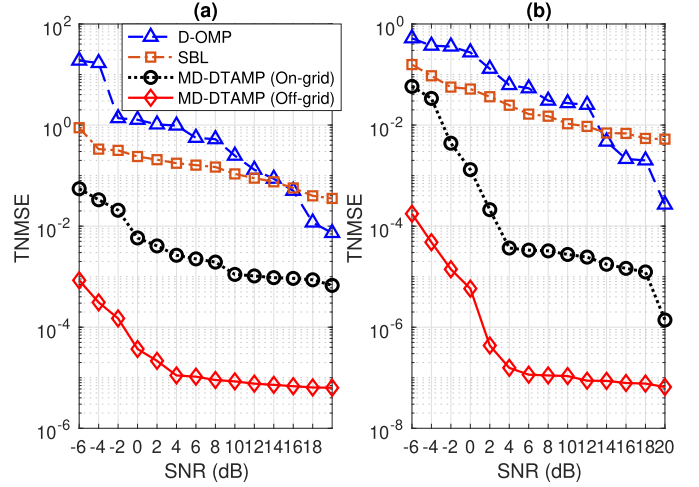

Fig. 8.	TNMSE versus SNR when $T = 35$. (a) $M = 64$, $N_x = N_y = 8$, $N_s = 20$. (b) $M = 256$, $N_x = N_y = 16$, $N_s = 26$.

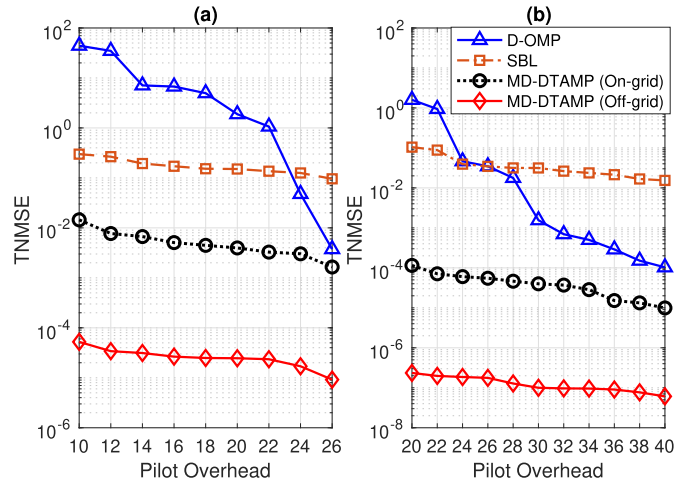Fig. 9.	TNMSE versus total pilot overhead when $T = 35$. (a) $M = 64$, $N_x = N_y = 8$, SNR $= 0$dB. (b) $M = 256$, $N_x = N_y = 16$, SNR $= 5$dB.
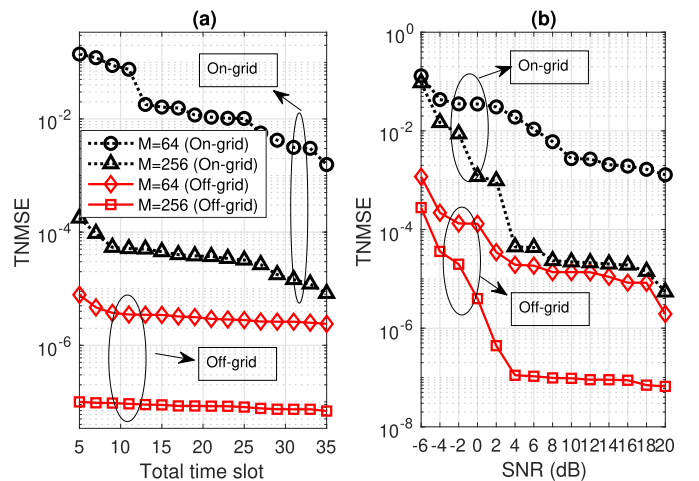

Fig. 10.	TNMSE versus total time slot and SNR when (a) $N_s = 26$, SNR $= 20$dB, (b) $T = 35$, $N_s = 22$.

The proposed on-grid and off-grid algorithms are compared in Fig. 10. The off-grid estimation achieves sufficient performance gain over the on-grid algorithm. In Fig. 10 (a),

the performance over continuous total time slot is compared. It is obvious that, with time increases, the performance show superior gain for both on-grid and off-grid algorithms. Moreover, it can be observed from Fig. 10 (a) and (b) that with M increases from 64 to 256, the tracking performance also improves. This is because with a larger antenna array and denser virtual gird, and the on-grid 3D tracking angle can be estimated with a narrower grid and closer to the true angle.

VII. Conclusion

This paper exploits the on-grid and off-grid 3D channel modelling and tracking for space-air communications with multiple UAVs and an LEO satellite in the system. The relative displacement between the orbit movement of the LEO satellite and the 3D trajectory of multiple UAVs was modelled based on the statistical dynamic channel model called MD-MM. The proposed MD-MM was composed of the insight probability relationship of the spatial and temporal correlation of the MU hidden support vector, SU joint hidden support vector, and SU hidden value vector. We further developed a novel MD-DTAMP algorithm to track the dynamic channel recursively. Furthermore, the gradient update for the 3D off-grid estimation scheme was proposed, which can iteratively find the azimuth and elevation offset. In addition, we compared the computation complexity between benchmark algorithms and the proposed algorithm with and without the off-grid scheme. Simulations verified that the proposed algorithm derived from the MD-MM dynamic channel model can better exploit the more realistic space-air 3D dynamic channel to achieve more accurate recovery performance over other benchmark algorithms with smaller pilot overhead.

Appendix A
Training Pilot Matrix Design

It has been proved that a partial DFT sensing matrix outperforms the independent identically distributed sensing matrix in solving the CS problem with AMP algorithm [11], [32]. Hence, in our proposed model, the sensing matrix $\boldsymbol{\Phi}_{k,t}(\Delta\boldsymbol{\theta}_{k,t}, \Delta\boldsymbol{\phi}_{k,t})$ is referred to be a partial DFT random permutation matrix as $\boldsymbol{\Phi}_{k,t}(\Delta\boldsymbol{\theta}_{k,t}, \Delta\boldsymbol{\phi}_{k,t}) \approx \boldsymbol{S}_{\text{SEL}}\boldsymbol{S}_{\text{DFT}} \boldsymbol{S}_{\text{RAND}}$, where $\boldsymbol{S}_{\text{SEL}}$ is a selection matrix consisting of randomly selected and reordered rows of $N_S \times M$ identity matrix, $\boldsymbol{S}_{\text{DFT}}$ is the $M \times M$ DFT matrix, and $\boldsymbol{S}_{\text{RAND}}$ is random permutation matrix generated by a randomly reordered $M \times M$ identity matrix. For large M antenna elements array, $\boldsymbol{S}_{\text{SEL}}\boldsymbol{S}_{\text{DFT}}\boldsymbol{S}_{\text{RAND}} = \boldsymbol{S}_k^T \Big(\tilde{\boldsymbol{A}}^H\big(\tilde{\boldsymbol{\theta}}_{k,t} + \Delta\boldsymbol{\theta}_{k,t}, \tilde{\boldsymbol{\phi}}_{k,t} + \Delta\boldsymbol{\phi}_{k,t}\big)\Big)^T$. Hence, for simplicity with fixed on-grid and approximate partial orthogonal, the pilot matrix $\boldsymbol{S}_k^T = \boldsymbol{S}_{\text{SEL}}\boldsymbol{S}_{\text{DFT}}\boldsymbol{S}_{\text{RAND}}\left(\big(\tilde{\boldsymbol{A}}^H\big(\tilde{\boldsymbol{\theta}}_{k,t}, \tilde{\boldsymbol{\phi}}_{k,t}\big)\big)^T\right)^{-1}$.

Appendix B
MD-MM Message Passing in Module $\mathcal{B}$

1) Message Passing Over Gaussian Hidden Value Vector $\boldsymbol{\vartheta}_{k,t}$: The message $\upsilon_{f_{k,t,m}\to\vartheta_{k,t,m}}$ passed from the last time

slot can be formed as $\mathcal{CN}(\vartheta_{k,t,m}; \mu_{k,t,m}^{dyna}, (\sigma_{k,t,m}^{dyna})^2)$. The message passed from variable node $\vartheta_{k,t,m}$ to factor node $\pi_{k,t,m}$ can be formed as

$$\upsilon_{\vartheta_{k,t,m}\to\pi_{k,t,m}} = \mathcal{CN}(\vartheta_{k,t,m}; \mu_{k,t,m}^{out}, (\sigma_{k,t,m}^{out})^2), \quad (52)$$

with $(\mu_{k,t,m}^{out}, (\sigma_{k,t,m}^{out})^2) = (\mu_{k,t,m}^{dyna}, (\sigma_{k,t,m}^{dyna})^2)$. When $t = 1$, we set $(\mu_{1,m}^{dyna}, (\sigma_{1,m}^{dyna})^2) = (\mu, \sigma^2)$

2) Message Passing Over $g_{k,t,m} \to \pi_{k,t,m} \to b_{k,t,m} \to u_{k,t,m} \to a_{k,t,n_x}$: The message from variable node $g_{k,t,m}$ to factor node $\pi_{k,t,m}$ is

$$\upsilon_{g_{k,t,m}\to\pi_{k,t,m}}(g_{k,t,m}) = \mathcal{CN}\Big(g_{k,t,m}; g_{\mathcal{B},k,t,m}^{pri}, v_{\mathcal{B},k,t,m}^{pri}\Big). \quad (53)$$

The message from factor node $\pi_{k,t,m}$ to variable node $b_{k,t,m}$ is

$$\upsilon_{\pi_{k,t,m}\to b_{k,t,m}}(b_{k,t,m})$$
$$= \rho_{b,k,t,m}^{in}\delta(b_{k,t,m} - 1) + \big(1 - \rho_{b,k,t,m}^{in}\big)\delta(b_{k,t,m}), \quad (54)$$

where

$$\rho_{\boldsymbol{b}_{k,t},k,t,m}^{in}$$
$$= \left(1 + \frac{\mathcal{CN}(0; g_{\mathcal{B},k,t,m}^{pri}, v_{\mathcal{B},k,t}^{pri})}{\mathcal{CN}(0; g_{\mathcal{B},k,t,m}^{pri} - \mu_{k,t,m}^{out}, v_{\mathcal{B},k,t}^{pri} + (\sigma_{k,t,m}^{out})^2)}\right)^{-1}. \quad (55)$$

The message $\upsilon_{b_{k,t,m}\to u_{k,t,m}}$ from variable node $b_{k,t,m}$ to factor node $u_{k,t,m}$ is the same as $\upsilon_{\pi_{k,t,m}\to b_{k,t,m}}(b_{k,t,m})$. The message from factor node $u_{k,t,m}$ to variable node a_{k,t,n_x} is

$$\upsilon_{u_{k,t,m}\to a_{k,t,n_x}}(a_{k,t,n_x}) = \rho_{\boldsymbol{a}_{k,t},k,t,n_x,n_y}^{in}\delta(a_{k,t,n_x} - 1)$$
$$+ \Big(1 - \rho_{\boldsymbol{a}_{k,t},k,t,n_x,n_y}^{in}\Big)\delta(a_{k,t,n_x}), \quad (56)$$

where

$$\rho_{\boldsymbol{a}_{k,t},k,t,n_x,n_y}^{in} = \frac{\rho_{\boldsymbol{a}_{k,t}}}{1 + \rho_{\boldsymbol{a}_{k,t}} - \rho_{\boldsymbol{b}_{k,t},k,t,m}^{in}}, \quad (57)$$

with

$$\rho_{\boldsymbol{a}_{k,t}} = \gamma_{\boldsymbol{b}_{k,t}}\rho_{\boldsymbol{b}_{k,t},k,t,m}^{in}\rho_{\boldsymbol{e}_{k,t},k,t,n_x,n_y}^{out}$$
$$+ \big(1 - \gamma_{\boldsymbol{b}_{k,t}}\big)\big(1 - \rho_{\boldsymbol{b}_{k,t},k,t,m}^{in}\big)\rho_{\boldsymbol{e}_{k,t},k,t,n_x,n_y}^{out}$$
$$+ \Big(1 - \rho_{\boldsymbol{b}_{k,t},k,t,m}^{in}\Big)\Big(1 - \rho_{\boldsymbol{e}_{k,t},k,t,n_x,n_y}^{out}\Big) \quad (58)$$

and $\rho_{\boldsymbol{e}_t,k,t,n_x,n_y}^{out} = \upsilon_{e_{k,t,n_y}\to u_{k,t,m}}(e_{k,t,n_y} = 1)$.

3) Message Passing Over $a_{k,t,n_x} \to \varrho_{k,t,n_x} \to \tilde{a}_{t,n_x}$, the 2D-MM of the MU Azimuth Support $\tilde{\boldsymbol{a}}_{t,n_x}$, and $\tilde{a}_{t,n_x} \to \varrho_{k,t,n_x} \to a_{k,t,n_x}$: The forward-backward message passing is performed over the HMM $\boldsymbol{a}_{k,t}$ for given input messages $\upsilon_{u_{k,t,n_x,n_y}\to a_{k,t,n_x}}$.

We firstly introduce the details of the MU azimuth support estimation. The condition can be divided into two section, one is when $t = 1$, the other is when $t > 1$. To be notice, the message from factor node $u_{t,m}$ to variable

node a_{t,n_x} is

$$
\begin{aligned}
v_{u_{k,t,m} \to a_{k,t,n_x}} &= \rho_{k,t,n_x}^{(A)in} \\
&= \frac{\prod_{n_y} \rho_{\boldsymbol{a}_{k,t},k,t,n_x,n_y}^{in}}{\prod_{n_y} \rho_{\boldsymbol{a}_{k,t},k,t,n_x,n_y}^{in} + \prod_{n_y} \left(1 - \rho_{\boldsymbol{a}_{k,t},k,t,n_x,n_y}^{in}\right)},
\end{aligned} \tag{59}
$$

which is derivated based on the matched dimension M from the total gain, n_x from the azimuth support and n_y from the elevation support.

The message from factor node ϱ_{k,t,n_x} to variable node $\tilde{a}_{t,n_x}$ is

$$
\begin{aligned}
v_{\varrho_{k,t,n_x} \to \tilde{a}_{t,n_x}} &\propto \sum_{a_{k,t,n_x}} \prod_{n_x} v_{u_{k,t,m} \to a_{k,t,n_x}}(a_{k,t,n_x}) \varsigma_{\tilde{a}_{t,n_x},a_{k,t,n_x}}^{(A)} \\
&= \rho_{\tilde{\boldsymbol{a}}_t,k,n_x}^{in} \delta(\tilde{a}_{t,n_x} - 1) + (1 - \rho_{\tilde{\boldsymbol{a}}_t,k,n_x}^{in}) \tilde{a}_{t,n_x},
\end{aligned} \tag{60}
$$

where

$$
\begin{aligned}
&\rho_{\tilde{\boldsymbol{a}}_t,k,n_x}^{in} \\
&= \frac{\varsigma_{k,t,1,1}^{(A)} \rho_{k,t,n_x}^{(A)in} + (1 - \varsigma_{k,t,1,1}^{(A)})(1 - \rho_{k,t,n_x}^{(A)in})}{\varsigma_{k,t,1,1}^{(A)} \rho_{k,t,n_x}^{(A)in} + (1 - \varsigma_{k,t,1,1}^{(A)})(1 - \rho_{k,t,n_x}^{(A)in}) + (1 - \rho_{k,t,n_x}^{(A)in})}.
\end{aligned} \tag{61}
$$

When $t = 1$, the MU azimuth forward parameter is

$$
\begin{aligned}
&\tilde{\gamma}_{t,n_x}^{(A)f} \\
&= \frac{q_{0,1}^{(AS)} \left(1 - \tilde{\rho}_{t,n_x-1}^{(A)in}\right) \left(1 - \tilde{\gamma}_{t,n_x-1}^{(A)f}\right) + q_{1,1}^{(AS)} \tilde{\rho}_{t,n_x-1}^{(A)in} \tilde{\gamma}_{t,n_x-1}^{(A)f}}{\left(1 - \tilde{\rho}_{t,n_x-1}^{(A)in}\right) \left(1 - \tilde{\gamma}_{t,n_x-1}^{(A)f}\right) + \tilde{\rho}_{t,n_x-1}^{(A)in} \tilde{\gamma}_{t,n_x-1}^{(A)f}},
\end{aligned} \tag{62}
$$

where $\tilde{\rho}_{t,n_x}^{(A)in} = \frac{\prod_k \rho_{\tilde{\boldsymbol{a}}_t,k,n_x}^{in}}{\prod_k \rho_{\tilde{\boldsymbol{a}}_t,k,n_x}^{in} + \prod_k (1 - \rho_{\tilde{\boldsymbol{a}}_t,k,n_x}^{in})}$. The azimuth backward parameter $\tilde{\gamma}_{t,n_x}^{(A)b}$ can be found in (63), shown at the bottom of the page. When $t > 1$, the azimuth forward

parameter $\tilde{\gamma}_{t,n_x}^{(A)f}$ can be found as

$$
\begin{aligned}
&\tilde{\gamma}_{t,n_x}^{(A)f} \\
&= \frac{q_{1,1,1}^{(A)} \tilde{\rho}_{t,n_x}^{dyna} + q_{1,0,1}^{(A)} \left(1 - \tilde{\rho}_{t,n_x}^{(A)dyna}\right)}{1 + \left(\left(\tilde{\rho}_{t,n_x-1}^{(A)in}\right)^{-1} - 1\right) \left(\left(\tilde{\gamma}_{t,n_x-1}^{(A)f}\right)^{-1} - 1\right)} \\
&\quad + \frac{q_{0,1,1}^{(A)} \tilde{\rho}_{t,n_x}^{(A)dyna} + q_{0,0,1}^{(A)} \left(1 - \tilde{\rho}_{t,n_x}^{(A)dyna}\right)}{\left(\left(\tilde{\rho}_{t,n_x-1}^{(A)in}\right)^{-1} - 1\right)^{-1} \left(\left(\left(\tilde{\gamma}_{t,n_x-1}^{(A)f}\right)^{-1} - 1\right)\right)^{-1} + 1}.
\end{aligned} \tag{64}
$$

The azimuth backward parameter $\tilde{\gamma}_{t,n_x}^{(A)b} = \frac{1}{1 + \tilde{\gamma}^{(A)}}$ is given in (65), shown at the bottom of the page. The final MU azimuth message which represents the non-zero probability can be written as

$$
v_{\tilde{a}_{t,n_x} \to \varrho_{k,t,n_x}} = \tilde{\rho}_{\tilde{\boldsymbol{a}}_t,k,n_x}^{out} \delta(\tilde{a}_{t,n_x} - 1) + (1 - \tilde{\rho}_{\tilde{\boldsymbol{a}}_t,k,n_x}^{out}) \delta(\tilde{a}_{t,n_x}),
\tag{66}
$$

where $\tilde{\rho}_{\tilde{\boldsymbol{a}}_t,k,n_x}^{out}$ can be seen in (67), shown at the bottom of the page. The message from factor node ϱ_{k,t,n_x} to variable node a_{k,t,n_x} is

$$
\begin{aligned}
&v_{\varrho_{k,t,n_x} \to a_{k,t,n_x}}(a_{k,t,n_x}) \\
&= \rho_{k,t,n_x}^{(A)out} \delta(a_{k,t,n_x} - 1) + (1 - \rho_{k,t,n_x}^{(A)out}) \delta(a_{k,t,n_x}),
\end{aligned} \tag{68}
$$

where $\rho_{k,t,n_x}^{(A)out} = \tilde{\rho}_{\tilde{\boldsymbol{a}}_t,k,n_x}^{out} \varsigma_{k,t,1,1}^{(A)}$.

4) Message Passing Over the Path $a_{k,t,n_x} \to u_{k,t,m} \to e_{k,t,n_y}$: The message from factor node a_{k,t,n_x} to variable node $u_{k,t,m}$ is

$$
\begin{aligned}
v_{a_{k,t,n_x} \to u_{k,t,m}}(a_{k,t,n_x}) &= \rho_{\boldsymbol{a}_{k,t},k,t,n_x,n_y}^{out} \delta(a_{k,t,n_x} - 1) \\
&\quad + (1 - \rho_{\boldsymbol{a}_{k,t},k,t,n_x,n_y}^{out}) \delta(a_{k,t,n_x}),
\end{aligned} \tag{69}
$$

where $\rho_{\boldsymbol{a}_{k,t},k,t,n_x,n_y}^{out}$ can be found in (70), shown at the bottom of the next page. The message from factor node $u_{k,t,m}$ to variable node e_{k,t,n_y} is

$$
\begin{aligned}
v_{u_{k,t,m} \to e_{k,t,n_y}}(e_{k,t,n_y}) &= \rho_{\boldsymbol{e}_{k,t},k,t,n_x,n_y}^{in} \delta\left(e_{k,t,n_y} - 1\right) \\
&\quad + \left(1 - \rho_{\boldsymbol{e}_{k,t},k,t,n_x,n_y}^{in}\right) \delta\left(e_{k,t,n_y}\right),
\end{aligned} \tag{71}
$$

$$
\tilde{\gamma}_{t,n_x}^{(A)b} = \frac{q_{1,0}^{(AS)} \left(1 - \tilde{\rho}_{t,n_x+1}^{(A)in}\right) \left(1 - \tilde{\gamma}_{t,n_x+1}^{(A)b}\right) + \left(1 - q_{1,0}^{(S)}\right) \tilde{\rho}_{t,n_x+1}^{(A)in} \tilde{\gamma}_{t,n_x+1}^{(A)b}}{\left(q_{0,0}^{(AS)} + q_{1,0}^{(S)}\right) \left(1 - \tilde{\rho}_{t,n_x+1}^{(A)in}\right) \left(1 - \tilde{\gamma}_{t,n_x+1}^{(A)b}\right) + \left(q_{1,1}^{(AS)} + q_{0,1}^{(AS)}\right) \tilde{\rho}_{t,n_x+1}^{(A)in} \tilde{\gamma}_{t,n_x+1}^{(A)b}}
\tag{63}
$$

$$
\tilde{\gamma}^{(A)} = \frac{\left(1 - \tilde{\rho}_{t,n_x+1}^{(A)in} - \tilde{\gamma}_{t,n_x+1}^{(A)b}\right) \left[q_{0,1,0}^{(A)} \tilde{\rho}_{t,n_x+1}^{(A)dyna} + q_{0,0,0}^{(A)} \left(1 - \tilde{\rho}_{t,n_x+1}^{(A)dyna}\right)\right] + \tilde{\rho}_{t,n_x+1}^{(A)in} \tilde{\gamma}_{t,n_x+1}^{(A)b}}{\left(1 - \tilde{\rho}_{t,n_x+1}^{(A)in} - \tilde{\gamma}_{t,n_x+1}^{(A)b}\right) \left[q_{1,1,0}^{(A)} \tilde{\rho}_{t,n_x+1}^{(A)dyna} + q_{1,0,0}^{(A)} \left(1 - \tilde{\rho}_{t,n_x+1}^{(A)dyna}\right)\right] + \tilde{\rho}_{t,n_x+1}^{(A)in} \tilde{\gamma}_{t,n_x+1}^{(A)b}}
\tag{65}
$$

$$
\tilde{\rho}_{\tilde{\boldsymbol{a}}_t,k,n_x}^{out} = \frac{\tilde{\gamma}_{t,n_x}^{(A)f} \tilde{\gamma}_{t,n_x}^{(A)b}}{\tilde{\gamma}_{t,n_x}^{(A)f} \tilde{\gamma}_{t,n_x}^{(A)b} + \prod_{k' \neq k} \left((\rho_{\tilde{\boldsymbol{a}}_t,k',n_x}^{in})^{-1} - 1\right) \left(1 - \tilde{\gamma}_{t,n_x}^{(A)f}\right) \left(1 - \tilde{\gamma}_{t,n_x}^{(A)b}\right)}.
\tag{67}
$$

where $\rho^{in}_{\boldsymbol{e}_{k,t},k,t,n_x,n_y}$ can be found in (72), shown at the bottom of the page.

5) Message Passing Over $e_{k,t,n_y} \rightarrow \varpi_{k,t,n_y} \rightarrow \tilde{e}_{t,n_y}$, *the 2D-MM of the MU Elevation Support* $\tilde{\boldsymbol{e}}_{t,n_y}$, *and* $\tilde{e}_{t,n_y} \rightarrow \varpi_{k,t,n_y} \rightarrow e_{k,t,n_y}$: The details can be derived similarly as the message passing process of the MU azimuth support $\tilde{\boldsymbol{a}}_t$ estimation. The message from factor node $u_{k,t,m}$ to variable node e_{k,t,n_y} is

$$
\begin{aligned}
\upsilon_{u_{k,t,m} \rightarrow e_{k,t,n_y}} &= \rho^{(E)in}_{k,t,n_y} \\
&= \frac{\prod_{n_x} \rho^{in}_{\boldsymbol{e}_{k,t},k,t,n_x,n_y}}{\prod_{n_x} \rho^{in}_{\boldsymbol{e}_{k,t},k,t,n_x,n_y} + \prod_{n_x}\left(1 - \rho^{in}_{\boldsymbol{e}_{k,t},k,t,n_x,n_y}\right)},
\end{aligned} \tag{73}
$$

which is derivated based on the matched dimension M from the total gain, n_x from the azimuth support and n_y from the elevation support.

The message from factor node ϖ_{k,t,n_y} to variable node $\tilde{e}_{t,n_y}$ is

$$
\begin{aligned}
\upsilon_{\varpi_{k,t,n_y} \rightarrow \tilde{e}_{t,n_y}} &\propto \sum_{e_{k,t,n_y}} \prod_{n_y} \upsilon_{u_{k,t,m} \rightarrow e_{k,t,n_y}}(e_{k,t,n_y}) \varsigma^{(E)}_{\tilde{e}_{t,n_y},e_{k,t,n_y}} \\
&= \rho^{in}_{\tilde{\boldsymbol{e}}_t,k,n_y} \delta(\tilde{e}_{t,n_y} - 1) + (1 - \rho^{in}_{\tilde{\boldsymbol{e}}_t,k,n_y})\tilde{e}_{t,n_y},
\end{aligned} \tag{74}
$$

where

$$
\begin{aligned}
&\rho^{in}_{\tilde{\boldsymbol{e}}_t,k,n_y} \\
&= \frac{\varsigma^{(E)}_{k,t,1,1} \rho^{(E)in}_{k,t,n_y} + (1 - \varsigma^{(E)}_{k,t,1,1})(1 - \rho^{(E)in}_{k,t,n_y})}{\varsigma^{(E)}_{k,t,1,1} \rho^{(E)in}_{k,t,n_y} + (1 - \varsigma^{(E)}_{k,t,1,1})(1 - \rho^{(E)in}_{k,t,n_y}) + (1 - \rho^{(E)in}_{k,t,n_y})}.
\end{aligned} \tag{75}
$$

When $t = 1$, the MU elevation forward parameter is

$$
\begin{aligned}
&\tilde{\gamma}^{(E)f}_{t,n_y} \\
&= \frac{q^{(ES)}_{0,1}\left(1 - \tilde{\rho}^{(E)in}_{t,n_y-1}\right)\left(1 - \tilde{\gamma}^{(E)f}_{t,n_y-1}\right) + q^{(ES)}_{1,1}\tilde{\rho}^{(E)in}_{t,n_y-1}\tilde{\gamma}^{(E)f}_{t,n_y-1}}{\left(1 - \tilde{\rho}^{(E)in}_{t,n_y-1}\right)\left(1 - \tilde{\gamma}^{(E)f}_{t,n_y-1}\right) + \tilde{\rho}^{(E)in}_{t,n_y-1}\tilde{\gamma}^{(E)f}_{t,n_y-1}},
\end{aligned} \tag{76}
$$

where $\tilde{\rho}^{(E)in}_{t,n_y} = \frac{\prod_k \rho^{in}_{\tilde{\boldsymbol{e}}_t,k,n_y}}{\prod_k \rho^{in}_{\tilde{\boldsymbol{e}}_t,k,n_y} + \prod_k(1 - \rho^{in}_{\tilde{\boldsymbol{e}}_t,k,n_y})}$. The MU elevation backward parameter $\tilde{\gamma}^{(E)b}_{t,n_y}$ can be found in (77), shown at the bottom of the page. When $t > 1$, the azimuth forward parameter $\tilde{\gamma}^{(E)f}_{t,n_x}$ can be found as

$$
\begin{aligned}
&\tilde{\gamma}^{(E)f}_{t,n_y} \\
&= \frac{q^{(E)}_{1,1,1}\tilde{\rho}^{dyna}_{t,n_y} + q^{(E)}_{1,0,1}\left(1 - \tilde{\rho}^{(E)dyna}_{t,n_y}\right)}{1 + \left(\left(\tilde{\rho}^{(E)in}_{t,n_y-1}\right)^{-1} - 1\right)\left(\left(\tilde{\gamma}^{(E)f}_{t,n_y-1}\right)^{-1} - 1\right)} \\
&\quad + \frac{q^{(E)}_{0,1,1}\tilde{\rho}^{(E)dyna}_{t,n_y} + q^{(E)}_{0,0,1}\left(1 - \tilde{\rho}^{(E)dyna}_{t,n_y}\right)}{\left(\left(\tilde{\rho}^{(E)in}_{t,n_y-1}\right)^{-1} - 1\right)^{-1}\left(\left(\left(\tilde{\gamma}^{(E)f}_{t,n_y-1}\right)^{-1} - 1\right)\right)^{-1} + 1}.
\end{aligned} \tag{78}
$$

The azimuth backward parameter $\tilde{\gamma}^{(E)b}_{t,n_y} = \frac{1}{1+\tilde{\gamma}^{(E)}}$ is given in (79), shown at the bottom of the page.

The final MU elevation message which represents the non-zero probability can be written as

$$
\begin{aligned}
&\upsilon_{\tilde{e}_{t,n_y} \rightarrow \varpi_{k,t,n_y}} \\
&= \tilde{\rho}^{out}_{\tilde{\boldsymbol{e}}_t,k,n_y}\delta(\tilde{e}_{t,n_y} - 1) + (1 - \tilde{\rho}^{out}_{\tilde{\boldsymbol{e}}_t,k,n_y})\delta(\tilde{e}_{t,n_y}),
\end{aligned} \tag{80}
$$

where $\tilde{\rho}^{out}_{\tilde{\boldsymbol{e}}_t,k,n_y}$ can be seen in (81), shown at the top of the next page.

$$
\rho^{out}_{\boldsymbol{a}_{k,t},k,t,n_x,n_y} = \frac{\rho^{(A)out}_{k,t,n_x}}{\rho^{(A)out}_{k,t,n_x} + \left(1 - \rho^{(A)out}_{k,t,n_x}\right)\prod_{n'_y \neq n_y}\left(\left(\rho^{in}_{\boldsymbol{a}_{k,t},k,t,n_x,n'_y}\right)^{-1} - 1\right)}. \tag{70}
$$

$$
\begin{aligned}
&\rho^{in}_{\boldsymbol{e}_{k,t},k,t,n_x,n_y} \\
&= \frac{\gamma_{\boldsymbol{b}_{k,t}}\rho^{in}_{\boldsymbol{b}_{k,t},k,t,m}\rho^{out}_{\boldsymbol{a}_{k,t},k,t,n_x,n_y} + (1 - \gamma_{\boldsymbol{b}_{k,t}})(1 - \rho^{in}_{\boldsymbol{b}_{k,t},k,t,m})\rho^{out}_{\boldsymbol{a}_{k,t},k,t,n_x,n_y} + (1 - \rho^{in}_{\boldsymbol{b}_{k,t},k,t,m})(1 - \rho^{out}_{\boldsymbol{a}_{k,t},k,t,n_x,n_y})}{\gamma_{\boldsymbol{b}_{k,t}}\rho^{in}_{\boldsymbol{b}_{k,t},k,t,m}\rho^{out}_{\boldsymbol{a}_{k,t},k,t,n_x,n_y} + (1 - \gamma_{\boldsymbol{b}_{k,t}})(1 - \rho^{in}_{\boldsymbol{b}_{k,t},k,t,m})\rho^{out}_{\boldsymbol{a}_{k,t},k,t,n_x,n_y} + (1 - \rho^{in}_{\boldsymbol{b}_{k,t},k,t,m})(1 - \rho^{out}_{\boldsymbol{a}_{k,t},k,t,n_x,n_y}) + 1 - \rho^{in}_{\boldsymbol{b}_{k,t},k,t,m}}
\end{aligned} \tag{72}
$$

$$
\tilde{\gamma}^{(E)b}_{t,n_y} = \frac{q^{(ES)}_{1,0}\left(1 - \tilde{\rho}^{(E)in}_{t,n_y+1}\right)\left(1 - \tilde{\gamma}^{(E)b}_{t,n_y+1}\right) + \left(1 - q^{(S)}_{1,0}\right)\tilde{\rho}^{(E)in}_{t,n_y+1}\tilde{\gamma}^{(E)b}_{t,n_y+1}}{\left(q^{(ES)}_{0,0} + q^{(S)}_{1,0}\right)\left(1 - \tilde{\rho}^{(E)in}_{t,n_y+1}\right)\left(1 - \tilde{\gamma}^{(E)b}_{t,n_y+1}\right) + \left(q^{(ES)}_{1,1} + q^{(ES)}_{0,1}\right)\tilde{\rho}^{(E)in}_{t,n_y+1}\tilde{\gamma}^{(E)b}_{t,n_y+1}} \tag{77}
$$

$$
\tilde{\gamma}^{(E)} = \frac{\left(1 - \tilde{\rho}^{(E)in}_{t,n_y+1} - \tilde{\gamma}^{(E)b}_{t,n_y+1}\right)\left[q^{(E)}_{0,1,0}\tilde{\rho}^{(E)dyna}_{t,n_y+1} + q^{(E)}_{0,0,0}\left(1 - \tilde{\rho}^{(E)dyna}_{t,n_y+1}\right)\right] + \tilde{\rho}^{(E)in}_{t,n_y+1}\tilde{\gamma}^{(E)b}_{t,n_y+1}}{\left(1 - \tilde{\rho}^{(E)in}_{t,n_y+1} - \tilde{\gamma}^{(E)b}_{t,n_y+1}\right)\left[q^{(E)}_{1,1,0}\tilde{\rho}^{(E)dyna}_{t,n_y+1} + q^{(E)}_{1,0,0}\left(1 - \tilde{\rho}^{(E)dyna}_{t,n_y+1}\right)\right] + \tilde{\rho}^{(E)in}_{t,n_y+1}\tilde{\gamma}^{(E)b}_{t,n_y+1}} \tag{79}
$$

$$\tilde{\rho}^{out}_{\tilde{e}_t,k,n_y} = \frac{\tilde{\gamma}^{(E)f}_{t,n_y} \tilde{\gamma}^{(E)b}_{t,n_y}}{\tilde{\gamma}^{(E)f}_{t,n_y} \tilde{\gamma}^{(E)b}_{t,n_y} + \prod_{k' \neq k} \left((\rho^{in}_{\tilde{e}_t,k',n_y})^{-1} - 1 \right) \left(1 - \tilde{\gamma}^{(E)f}_{t,n_y} \right) \left(1 - \tilde{\gamma}^{(E)b}_{t,n_y} \right)}. \tag{81}$$

$$\rho^{out}_{e_{k,t},k,t,n_x,n_y} = \frac{\rho^{(E)out}_{k,t,n_y}}{\rho^{(E)out}_{k,t,n_y} + \left(1 - \rho^{(E)out}_{k,t,n_y} \right) \prod_{n'_x \neq n_x} \left(\left(\rho^{in}_{e_{k,t},k,t,n'_x,n_y} \right)^{-1} - 1 \right)}. \tag{84}$$

$$\Lambda^{(i)}(\Delta\theta_{k,t,m}) = 2\text{Re}\left(A'\left(\tilde{\theta}_{k,t,m} + \Delta\theta_{k,t,m}, \tilde{\phi}_{k,t,m} + \Delta\phi^{(i)}_{k,t,m} \right)^H S_k S_k^H A\left(\tilde{\theta}_{k,t,m} + \Delta\theta_{k,t,m}, \tilde{\phi}_{k,t,m} + \Delta\phi^{(i)}_{k,t,m} \right) \right) \Theta^{(i)}_1$$
$$+ 2\text{Re}\left(A'\left(\tilde{\theta}_{k,t,m} + \Delta\theta_{k,t,m}, \tilde{\phi}_{k,t,m} + \Delta\phi^{(i)}_{k,t,m} \right)^H S_k \Theta^{(i)}_2 \right), \tag{88}$$

The message from factor node ϖ_{k,t,n_y} to variable node e_{k,t,n_y} is

$$
\begin{aligned}
&v_{\varpi_{k,t,n_y} \to e_{k,t,n_y}}(e_{k,t,n_y}) \\
&= \rho^{(E)out}_{k,t,n_y} \delta(e_{k,t,n_y} - 1) + (1 - \rho^{(E)out}_{k,t,n_y})\delta(e_{k,t,n_y}),
\end{aligned} \tag{82}
$$

where $\rho^{(E)out}_{k,t,n_y} = \tilde{\rho}^{out}_{\tilde{e}_t,k,n_y} \varsigma^{(E)}_{k,t,1,1}$.

6) Message Passing Over the Path $e_{k,t,n_y} \to u_{k,t,m} \to b_{k,t,m} \to \pi_{k,t,m} \to g_{k,t,m}$: The message from factor node e_{k,t,n_y} to variable node $u_{k,t,m}$ is

$$
\begin{aligned}
v_{e_{k,t,n_y} \to u_{k,t,m}}(e_{k,t,n_y}) &= \rho^{out}_{e_{k,t},k,t,n_x,n_y} \delta\left(e_{k,t,n_y} - 1\right) \\
&+ \left(1 - \rho^{out}_{e_{k,t},k,t,n_x,n_y}\right) \delta\left(e_{k,t,n_y}\right),
\end{aligned} \tag{83}
$$

where $\rho^{out}_{e_{k,t},k,t,n_x,n_y}$ can be found in (84), shown at the top of the page. The message from factor node $u_{k,t,m}$ to variable node $b_{k,t,m}$ is

$$
\begin{aligned}
v_{u_{k,t,m} \to b_{k,t,m}}(b_{k,t,m}) &= \rho^{out}_{b_{k,t},k,t,m} \delta(b_{k,t,m} - 1) \\
&+ \left(1 - \rho^{out}_{b_{k,t},k,t,m}\right) \delta(b_{k,t,m}),
\end{aligned} \tag{85}
$$

where $\rho^{out}_{b_{k,t},k,t,m} = \rho^{out}_{a_{k,t},k,t,n_x,n_y} \rho^{out}_{e_{k,t},k,t,n_x,n_y} \gamma_{b_{k,t}}$. The message $v_{b_{k,t,m} \to \pi_{k,t,m}}$ from variable node $b_{k,t,m}$ to factor node $\pi_{k,t,m}$ is the same as $v_{u_{k,t,m} \to b_{k,t,m}}(b_{k,t,m})$. The message from factor node $\pi_{k,t,m}$ back to variable node $g_{k,t,m}$ is

$$
\begin{aligned}
&v_{\pi_{k,t,m} \to g_{k,t,m}}(g_{k,t,m}) \\
&= \rho^{out}_{b_{k,t},k,t,m} \mathcal{CN}\left(g_{k,t,m}; \mu^{out}_{k,t,m}, (\sigma^{out}_{k,t,m})^2\right) \\
&+ \left(1 - \rho^{out}_{b_{k,t},k,t,m}\right) \delta(g_{k,t,m}).
\end{aligned} \tag{86}
$$

APPENDIX C
GRADIENT UPDATE FOR 3D OFF-GRID ESTIMATION

As the 3D off-grid estimation step is following the step 15 in **Algorithm 3** which represents the posterior estimation at the ith iteration with $\hat{p}\left(g_{k,t}|y^{(t)}_k, \Delta\theta^{(i)}_{k,t}, \Delta\phi^{(i)}_{k,t}\right) = \mathcal{CN}\left(g^{post}_{\mathcal{B},k,t}, v^{post}_{\mathcal{B},k,t}I\right)$. The surrogate function can be

calculated as

$$
\begin{aligned}
&\hat{\mathcal{L}}\left(\Delta\theta_{k,t}, \Delta\phi_{k,t}|\Delta\theta^{(i)}_{k,t}, \Delta\phi^{(i)}_{k,t}\right) \\
&\propto E_{\hat{p}\left(g_{k,t}|y^{(t)}_k, \Delta\theta^{(i)}_{k,t}, \Delta\phi^{(i)}_{k,t}\right)} \\
&\quad \times \left[-\frac{1}{\sigma^2_e} \| y_{k,t} - \Phi_{k,t}\left(\Delta\theta_{k,t}, \Delta\phi_{k,t}\right) g_{k,t} \|^2 \right] \\
&\propto -\frac{1}{\sigma^2_e}\left(\| y_{k,t} - \Phi_{k,t}\left(\Delta\theta_{k,t}, \Delta\phi_{k,t}\right) g^{post}_{\mathcal{B},k,t} \|^2 \right. \\
&\quad \left. + v^{post}_{\mathcal{B},k,t}\text{tr}\left(\Phi_{k,t}\left(\Delta\theta_{k,t}, \Delta\phi_{k,t}\right) \Phi_{k,t}\left(\Delta\theta_{k,t}, \Delta\phi_{k,t}\right)^H \right) \right).
\end{aligned} \tag{87}
$$

The azimuth offset[5] $\Delta\theta^{(i+1)}_{k,t}$ can be updated use (49). The derivative $\left. \frac{\partial\mathcal{L}\left(\Delta\theta_{k,t}, \Delta\phi^{(i)}_{k,t}|\Delta\theta^{(i)}_{k,t}, \Delta\phi^{(i)}_{k,t}\right)}{\partial\Delta\theta_{k,t}} \right|_{\Delta\theta_{k,t}=\Delta\theta^{(i)}_{k,t}}$ with previous iteration updated elevation offset $\Delta\phi^{(i)}_{k,t}$ can be written as $\Lambda^{(i)}_{\Delta\theta_{k,t}} = [\Lambda^{(i)}(\Delta\theta_{k,t,1}), \ldots, \Lambda^{(i)}(\Delta\theta_{k,t,M})]^T$, with $\Lambda^{(i)}(\Delta\theta_{k,t,m})$ as in (88), shown at the top of the page, where $\Theta^{(i)}_1 = -\frac{1}{\sigma^2_e}\left(g^{post}_{\mathcal{B},k,t,m} + v^{post}_{\mathcal{B},k,t} \right)$, $\Theta^{(i)}_2 = \frac{1}{\sigma^2_e}\left(g^{post}_{\mathcal{B},k,t,m} \right) * y_{k,t,-m}$, $y_{k,t,-m} = y_{k,t} - S^H_k \sum_{j \neq m} A\left(\tilde{\theta}_{k,t,j} + \Delta\theta_{k,t,j}, \tilde{\phi}_{k,t,j} + \Delta\phi^{(i)}_{k,t,j} \right) g^{post}_{\mathcal{B},k,t,j}$, and

$$
\begin{aligned}
&A'\left(\tilde{\theta}_{k,t,m} + \Delta\theta_{k,t,m}, \tilde{\phi}_{k,t,m} + \Delta\phi^{(i)}_{k,t,m} \right) \\
&= dA'\left(\tilde{\theta}_{k,t,m} + \Delta\theta_{k,t,m}, \tilde{\phi}_{k,t,m} + \Delta\phi^{(i)}_{k,t,m} \right) / d\Delta\theta_{k,t,m}.
\end{aligned} \tag{89}
$$

REFERENCES

[1] Z. Li *et al.*, "Energy efficient resource allocation for UAV-assisted space-air-ground internet of remote things networks," *IEEE Access*, vol. 7, pp. 145348–145362, 2019.

[2] J. Liu, Y. Shi, Z. M. Fadlullah, and N. Kato, "Space-air-ground integrated network: A survey," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 4, pp. 2714–2741, 4th Quart., 2018.

[3] Y. Zeng, R. Zhang, and T. J. Lim, "Wireless communications with unmanned aerial vehicles: Opportunities and challenges," *IEEE Commun. Mag.*, vol. 54, no. 5, pp. 36–42, May 2016.

[5]To be noticed, in order to have azimuth and elevation one-one correspondence, we extend the original N_x dimension of offset into M dimension with azimuth and elevation paring. The elevation offset can be written in the similar way.

[4] N. Cheng *et al.*, "Space/aerial-assisted computing offloading for IoT applications: A learning-based approach," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 5, pp. 1117–1129, May 2019.

[5] H. Dai, H. Bian, C. Li, and B. Wang, "UAV-aided wireless communication design with energy constraint in space-air-ground integrated green IoT networks," *IEEE Access*, vol. 8, pp. 86251–86261, 2020.

[6] F. Rusek *et al.*, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[7] A. A. Khuwaja, Y. Chen, N. Zhao, M.-S. Alouini, and P. Dobbins, "A survey of channel modeling for UAV communications," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 4, pp. 2804–2821, 4th Quart., 2018.

[8] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 264–273, Mar. 2013.

[9] C. R. Berger, Z. Wang, J. Huang, and S. Zhou, "Application of compressive sensing to sparse channel estimation," *IEEE Commun. Mag.*, vol. 48, no. 11, pp. 164–174, Nov. 2010.

[10] A. Liu, V. K. N. Lau, and W. Dai, "Exploiting burst-sparsity in massive MIMO with partial channel support information," *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7820–7830, Nov. 2016.

[11] L. Chen, A. Liu, and X. Yuan, "Structured turbo compressed sensing for massive MIMO channel estimation using a Markov prior," *IEEE Trans. Veh. Technol.*, vol. 67, no. 5, pp. 4635–4639, May 2018.

[12] A. Liu, L. Lian, V. K. N. Lau, and X. Yuan, "Downlink channel estimation in multiuser massive MIMO with hidden Markovian sparsity," *IEEE Trans. Signal Process.*, vol. 66, no. 18, pp. 4796–4810, Sep. 2018.

[13] A. Liao, Z. Gao, H. Wang, S. Chen, M.-S. Alouini, and H. Yin, "Closed-loop sparse channel estimation for wideband millimeter-wave full-dimensional MIMO systems," *IEEE Trans. Commun.*, vol. 67, no. 12, pp. 8329–8345, Dec. 2019.

[14] X. Gao, L. Dai, Y. Zhang, T. Xie, X. Dai, and Z. Wang, "Fast channel tracking for terahertz beamspace massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 66, no. 7, pp. 5689–5696, Jul. 2017.

[15] O. Mehanna and N. D. Sidiropoulos, "Channel tracking and transmit beamforming with frugal feedback," *IEEE Trans. Signal Process.*, vol. 62, no. 24, pp. 6402–6413, Dec. 2014.

[16] X. Zhu, L. Dai, W. Dai, Z. Wang, and M. Moonen, "Tracking a dynamic sparse channel via differential orthogonal matching pursuit," in *Proc. IEEE Mil. Commun. Conf. (MILCOM)*, Oct. 2015, pp. 792–797.

[17] X. Zhu, L. Dai, G. Gui, W. Dai, Z. Wang, and F. Adachi, "Structured matching pursuit for reconstruction of dynamic sparse channels," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2014, pp. 1–5.

[18] A. Liao *et al.*, "Terahertz ultra-massive MIMO-based aeronautical communications in space-air-ground integrated networks," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 6, pp. 1741–1767, Jun. 2021.

[19] J. Zhao, F. Gao, W. Jia, S. Zhang, S. Jin, and H. Lin, "Angle domain hybrid precoding and channel tracking for millimeter wave massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 10, pp. 6868–6880, Oct. 2017.

[20] L. Lian, A. Liu, and V. K. N. Lau, "Exploiting dynamic sparsity for downlink FDD-massive MIMO channel tracking," *IEEE Trans. Signal Process.*, vol. 67, no. 8, pp. 2007–2021, Apr. 2019.

[21] J. Yu, X. Liu, Y. Gao, and X. Shen, "3D channel tracking for UAV-satellite communications in space-air-ground integrated networks," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 12, pp. 2810–2823, Dec. 2020.

[22] B. Qi, W. Wang, and B. Wang, "Off-grid compressive channel estimation for mm-wave massive MIMO with hybrid precoding," *IEEE Commun. Lett.*, vol. 23, no. 1, pp. 108–111, Jan. 2019.

[23] C. K. Anjinappa, A. C. Gurbuz, Y. Yapici, and I. Guvenc, "Off-grid aware channel and covariance estimation in mmWave networks," *IEEE Trans. Commun.*, vol. 68, no. 6, pp. 3908–3921, Jun. 2020.

[24] H. Fenech, S. Amos, and T. Waterfield, "The role of array antennas in commercial telecommunication satellites," in *Proc. 10th Eur. Conf. Antennas Propag. (EuCAP)*, Apr. 2016, pp. 1–4.

[25] Y. Xu and C. Ruan, "A novel design of circularly polarized omnidirectional antenna for Ka band," in *Proc. Global Symp. Millim. Waves*, Apr. 2008, pp. 378–379.

[26] I. Gungor and A. Unal, "Design of a vertically polarized omnidirectional antenna at Ka-band," in *Proc. IEEE Int. Symp. Antennas Propag. (APSURSI)*, Jun. 2016, pp. 129–130.

[27] J. Zhao, F. Gao, Q. Wu, S. Jin, Y. Wu, and W. Jia, "Beam tracking for UAV mounted SatCom on-the-move with massive antenna array," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 2, pp. 363–375, Feb. 2018.

[28] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 831–846, Oct. 2014.

[29] Y. Wu, G. Hu, F. Jin, and J. Zu, "A satellite handover strategy based on the potential game in LEO satellite networks," *IEEE Access*, vol. 7, pp. 133641–133652, 2019.

[30] J. Ziniel and P. Schniter, "Dynamic compressive sensing of time-varying signals via approximate message passing," *IEEE Trans. Signal Process.*, vol. 61, no. 21, pp. 5270–5284, Nov. 2013.

[31] W. Shi *et al.*, "Multi-drone 3D trajectory planning and scheduling in drone assisted radio access networks," 2019, *arXiv:1906.00777*. [Online]. Available: http://arxiv.org/abs/1906.00777

[32] J. Ma, X. Yuan, and L. Ping, "On the performance of turbo signal recovery with partial DFT sensing matrices," *IEEE Trans. Signal Process.*, vol. 22, no. 10, pp. 1580–1584, Oct. 2015.

[33] M. Ke, Z. Gao, Y. Wu, X. Gao, and K.-K. Wong, "Massive access in cell-free massive MIMO-based Internet of Things: Cloud computing and edge computing paradigms," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 3, pp. 756–772, Mar. 2021.

[34] M. Ke, Z. Gao, Y. Wu, X. Gao, and R. Schober, "Compressive sensing-based adaptive active user detection and channel estimation: Massive access meets massive MIMO," *IEEE Trans. Signal Process.*, vol. 68, pp. 764–779, 2020.

[35] S. Wu, H. Yao, C. Jiang, X. Chen, L. Kuang, and L. Hanzo, "Downlink channel estimation for massive MIMO systems relying on vector approximate message passing," *IEEE Trans. Veh. Technol.*, vol. 68, no. 5, pp. 5145–5148, May 2019.

[36] N. Dandanov, K. Tonchev, V. Poulkov, and P. Koleva, "Approximate message passing for downlink sparse channel estimation in FDD massive MIMO," in *Proc. 42nd Int. Conf. Telecommun. Signal Process. (TSP)*, Jul. 2019, pp. 487–491.

[37] Z. Xue, J. Ma, and X. Yuan, "Denoising-based turbo compressed sensing," *IEEE Access*, vol. 5, pp. 7193–7204, 2017.

[38] J. Dai, A. Liu, and V. K. N. Lau, "FDD massive MIMO channel estimation with arbitrary 2D-array geometry," *IEEE Trans. Signal Process.*, vol. 66, no. 10, pp. 2584–2599, May 2018.

[39] C. F. J. Wu, "On the convergence properties of the EM algorithm," *Ann. Statist.*, vol. 11, no. 1, pp. 95–103, Mar. 1983.

Jiadong Yu received the B.S. degree in communication engineering from Dalian Maritime University and the M.S. degree (Hons.) in wireless communication from the University of Southampton. She is currently pursuing the Ph.D. degree with the Department of Electronic Engineering and Computer Science, Queen Mary University of London. Her research interests include machine learning for millimeter wave communications and the Internet of Things.

Xiaolan Liu received the Ph.D. degree from the Queen Mary University of London, U.K., in 2021. She was a Research Assistant with King's College London from 2020 to 2021. Since October 2021, she has been with the Institute for Digital Technologies, Loughborough University London, U.K., where she is currently a Lecturer. Her current research interests include wireless distributed learning, multi-agent reinforcement learning for edge computing, and machine learning for wireless communication optimization.

Yue Gao (Senior Member, IEEE) received the Ph.D. degree from the Queen Mary University of London (QMUL), U.K., in 2007.

He was a Lecturer, a Senior Lecturer, and a Reader in antennas and signal processing with QMUL. He is currently a Professor of wireless communications with the Institute for Communication Systems, University of Surrey, U.K. He also leads a team developing fundamental research into practice in the interdisciplinary area of smart antennas, signal processing, spectrum sharing, and millimetre-wave and the Internet of Things technologies in mobile and satellite systems. He has published over 180 peer-reviewed journal articles and conference papers, three patents, one book and five book chapters, and three best paper awards. He is an Engineering and Physical Sciences Research Council Fellow from 2018 to 2023. He was a co-recipient of the EU Horizon Prize Award on Collaborative Spectrum Sharing in 2016 and shortlisted for the Newton Prize on the IoT systems for smart farming in 2019. He served as the Signal Processing for Communications Symposium Co-Chair for IEEE ICCC 2016, the Publicity Co-Chair for the IEEE GLOBECOM 2016, the Cognitive Radio Symposium Co-Chair for the IEEE GLOBECOM 2017, and the General Chair for the IEEE WoWMoM and iWEM 2017. He is the Chair of the IEEE Technical Committee on Cognitive Networks and the Secretary of the IEEE ComSoc Technical Committee Wireless Communication. He is an Editor for the IEEE INTERNET OF THINGS JOURNAL, IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, and IEEE TRANSACTIONS ON COGNITIVE NETWORKS. He is an IEEE Distinguished Lecturer of the Vehicular Technology Society.

Xuemin (Sherman) Shen (Fellow, IEEE) received the Ph.D. degree in electrical engineering from Rutgers University, New Brunswick, NJ, USA, in 1990.

He is currently a University Professor with the Department of Electrical and Computer Engineering, University of Waterloo, Canada. His research focuses on network resource management, wireless network security, the Internet of Things, 5G and beyond, and vehicular *ad hoc* and sensor networks. He is a Registered Professional Engineer of ON, Canada, an Engineering Institute of Canada Fellow, a Canadian Academy of Engineering Fellow, a Royal Society of Canada Fellow, and a Chinese Academy of Engineering Foreign Fellow. He is elected as the IEEE Communications Society Vice President for Technical and Educational Activities, the Vice President for Publications, a Member-at-Large on the Board of Governors, the Chair of the Distinguished Lecturer Selection Committee, and a member of the IEEE Fellow Selection Committee. He received the R. A. Fessenden Award in 2019 from IEEE, Canada, the Award of Merit from the Federation of Chinese Canadian Professionals (Ontario) presents in 2019, the James Evans Avant Garde Award in 2018 from the IEEE Vehicular Technology Society, the Joseph LoCicero Award in 2015 and the Education Award in 2017 from the IEEE Communications Society, and the Technical Recognition Award from the Wireless Communications Technical Committee in 2019 and the AHSN Technical Committee in 2013. He has also received the Excellent Graduate Supervision Award in 2006 from the University of Waterloo and the Premier's Research Excellence Award (PREA) in 2003 from ON, Canada. He served as the Technical Program Committee Chair/the Co-Chair for the IEEE Globecom 2016, the IEEE Infocom 2014, the IEEE VTC 2010 Fall, the IEEE Globecom 2007, the Symposia Chair for the IEEE ICC 2010, and the Chair for the IEEE Communications Society Technical Committee on Wireless Communications. He was/is the Editor-in-Chief of the IEEE INTERNET OF THINGS JOURNAL, IEEE Network, IET Communications, and Peer-to-Peer Networking and Applications. He is a Distinguished Lecturer of the IEEE Vehicular Technology Society and Communications Society.