# Angular-Domain Selective Channel Tracking and Doppler Compensation for High-Mobility mmWave Massive MIMO

Guanying Liu[1], An Liu[1], *Senior Member, IEEE*, Rui Zhang[2], *Fellow, IEEE*, and Min-Jian Zhao[1], *Member, IEEE*
[1]College of Information Science and Electronic Engineering, Zhejiang University
[2]Department of Electrical and Computer Engineering, National University of Singapore



*Abstract*—**In this paper, we consider a mmWave massive multiple-input multiple-output (MIMO) communication system with one static base station (BS) serving a fast-moving user, both equipped with a very large array. The transmitted signal arrives at the user through multiple paths, each with a different angle-of-arrival (AoA) and hence Doppler frequency offset (DFO), thus resulting in a fast time-varying multipath fading MIMO channel. In order to mitigate the Doppler-induced channel aging for reduced pilot overhead, we propose a new angular-domain selective channel tracking and Doppler compensation scheme at the user side. Specifically, we formulate the joint estimation of partial angular–domain channel and DFO parameters as a dynamic compressive sensing (CS) problem. Then we propose a Doppler-aware-dynamic variational Bayesian inference (DD-VBI) algorithm to solve this problem efficiently. Finally, we propose a practical DFO compensation scheme which selects the dominant paths of the fast time-varying channel for DFO compensation and thereby converts it into a slow time-varying effective channel. Compared with the existing methods, the proposed scheme can enjoy the huge array gain provided by the massive MIMO and also balance the tradeoff between the CSI signaling overhead and spatial multiplexing gain. Simulation results verify the advantages of the proposed scheme over various baseline schemes.**

*Index Terms*—**Massive MIMO, Channel Tracking, Doppler Compensation, High Mobility**

## I. Introduction

With the development of the Fifth Generation (5G) wireless communication systems, high-mobility scenarios such as high-speed rail and Vehicle-To-everything (V2x) communications have gained increasingly more interest. Due to the high-speed relative motion between the transmitter and receiver, the transmitted signal, propagating through multiple different paths, arrives at the receiver with different Doppler frequency offsets (DFOs), thus resulting in a fast time-varying multipath fading channel. In this case, the link performance such as achievable data rate will be degraded significantly due to the Doppler-induced channel aging effect [1]. To overcome this challenge, several methods have been proposed in previous works, as elaborated below.

**Direct Channel Estimation/Prediction:** For line-of-sight (LOS) channels with a single LOS path, it is relatively easy to compensate the Doppler effect and resolve the channel aging issue by estimating the DFO parameter of the LOS path. However, when there are multiple different paths due to rich scattering, it is challenging to compensate the Doppler effect because different paths with different DFOs are mixed together in the received signal. Nevertheless, some works have proposed to directly estimate the fast time-varying channels in the time/frequency domain [2], [3]. Existing channel estimators can be classified into two types. The first type approximates time-varying channels using a linearly time-varying (LTV) channel model [4], [5]. For example, a hybrid frequency/time-domain channel estimation algorithm is proposed in [4] based on the LTV model and two methods are introduced to mitigate the Doppler effect. However, this algorithm introduces a processing delay of at least one Orthogonal Frequency Division Multiplexing (OFDM) symbol. The second type of estimators adopt the basis expansion model (BEM) [6] to convert the problem of estimating the channel impulse response (CIR) to that of estimating the basis function weights [7]. For example, in [7], the channel estimation and Doppler mitigation are jointly considered by exploiting the correlations in time and frequency domains, and the basis function coefficients are estimated via the linear minimum mean squared error (LMMSE) approach. However, accurate knowledge on the maximum DFO is required to determine the minimum order of basis function and the computational burden is also heavy for multi-antenna systems [8]. Moreover, the BEM inevitably introduces approximation error to channel estimation due to the imperfect model assumed.

**Orthogonal Time Frequency Space (OTFS) Modulation:** OTFS modulation [9] is an emerging technique which is able to handle the fast time-varying channels. This method modulates transmitted symbols in the delay-Doppler domain instead of time/frequency domain as in traditional modulation techniques such as OFDM.. The idea is to transform the time-varying channel into a time-invariant channel in the delay-Doppler domain. Early works on OTFS modulation focused on the single-input single-output (SISO) systems [9]. Later, OTFS is extended to multiple-input multiple-output (MIMO) systems by transmitting consecutive impulses with proper guard time between two adjacent ones to distinguish different base station (BS) antennas [10]. However, the channel estimation method in [10] cannot be directly applied to massive MIMO system since a large number of antennas are required to be distinguished by transmitting such impulses, which will lead to large pilot overhead.

**Angular-Domain DFO Estimation and Compensation:** Since the different DFOs of multiple paths are resulted from their different angles of arrival (AoAs)/angles of departure (AoDs) at the receiver, they can be separated in the angle domain via spatial processing. As such, angular-domain DFO estimation and compensation is another popular approach to address the Doppler-induced channel aging issue [1], [11]. For MIMO systems, prior work [12] first pointed out that channel time-variation can be slowed down through beamforming with a large number of transmit/receive antennas. Motivated by this, a small-scale uniform circular antenna-array (UCA) is adopted in [13], [14] to separate multiple DFOs via array beamforming. However, the DFO compensation methods in [13], [14] only apply to scenarios with very sparse channels due to the limited spatial resolution of small-scale MIMO. Recently, some works have exploited high-spatial resolution provided by massive MIMO to address the high-mobility induced challenges [15], [16], [17], [18], [19]. For example, the authors of [8] propose to separate the DFOs in angular-domain by beamforming with a large-scale uniform linear antenna-array (ULA) at the mobile user side. After estimating and compensating the DFO in each angle, the resultant quasi time-invariant channel can be estimated more efficiently. However, the signaling overhead for the maximum likelihood (ML) based joint estimation of the massive MIMO channel matrix and DFO parameters is extremely high. Moreover, only a single data stream is transmitted from the BS to the mobile user via all possible channel directions, and thus it cannot enjoy the spatial multiplexing gain as well as the huge array signal-to-noise ratio (SNR) gain provided by massive MIMO.

The above works have focused on channel estimation. It is also possible to directly search for the best beamforming vectors without explicit channel estimation. For example, in [20], the authors propose an exhaustive search (ES) scheme, which examines all beam pairs in the codebook and determines the best pair that maximizes a given performance metric (e.g., beamforming gain). To reduce the training overhead of the ES scheme, the hierarchical search (HS) scheme proposed in [21] utilizes hierarchical codebooks and has a favorable performance at low SNRs. However, these codebook-based beam search methods suffer from the quantization error caused by the codebook and the channel aging effect.

In this paper, we consider high-mobility mmWave massive MIMO systems, where both the BS and users are equipped with massive antenna arrays to facilitate Doppler compensation, improve the spectrum and energy efficiency, as well as overcome the large path loss at high frequency band. As such, combining the mmWave and massive MIMO technologies has the potential to significantly improve the capacity and reliability of high-mobility wireless communications. However, it is very challenging to design an efficient channel estimation scheme. For example, since both ends have massive MIMO, the dimension of the channel matrix is huge and conventional downlink/uplink channel estimation or codebook-based beam search will lead to large signaling overhead. Although various pilot overhead reduction methods such as those based on compressive sensing (CS) have been proposed for the estimation of slow massive MIMO fading channels [22][23], they do not consider the Doppler effect and thus cannot be applied to the high mobility scenario. To overcome this challenge, we propose a novel angular-domain selective channel tracking and Doppler compensation scheme, which exploits the dynamic sparsity of the mmWave massive MIMO channel as well as precoded training in both downlink and uplink to significantly reduce the signaling overhead. The main new contributions of our paper are given as follows.

- **Angular-Domain Selective Channel Tracking:** We propose a selective channel tracking scheme to only estimate partial angular–domain channel parameters at the user side that are sufficient for Doppler effect compensation to significantly reduce the pilot overhead. Moreover, we propose an efficient downlink training vector design at the BS side to strike a balance between *exploitation* of most promising channel directions for array (SNR) gain and *exploration* of unknown channel directions. Compared to the conventional random training vector design, the proposed design can exploit the massive MIMO array gain to further enhance the channel tracking performance.
- **Angular-Domain Selective Doppler Compensation:** We propose an angular-domain selective DFO compensation scheme at the user side which selectively converts the dominant paths of the fast time-varying channel into a slow time-varying effective channel. Compared to the non-selective DFO compensation scheme in [18], the proposed scheme can enjoy the huge array (SNR) gain provided by the massive MIMO and also balance the tradeoff between the CSI signaling overhead reduction and spatial multiplexing gain maximization.
- **Channel Tracking Algorithm based on Dynamic VBI:** To further reduce the pilot overhead, the proposed partial channel tracking design is formulated as a dynamic CS problem with unknown DFO parameters in the measurement matrix. Then, we adopt a three-layer hierarchical Markov model to capture the dynamic sparsity of the partial angular–domain channel. The existing methods, such as Variational Bayesian inference (VBI) [24] and Sparse Bayesian Learning (SBL) [25], cannot be directly applied to this three-layer hierarchical prior. To address this challenge, we propose a Doppler-aware-dynamic Variational Bayesian inference (DD-VBI) algorithm, which combines the VBI and message-passing approaches to achieve superior channel tracking performance.

The rest of this paper is organized as follows. In Section II, we describe the system model and frame structure. In Section III, we give a brief introduction of the proposed angular-domain selective channel tracking and Doppler compensation scheme. In Sections IV and V, we present the three-layer hierarchical Markov model for partial angular channel vector and the proposed DD-VBI algorithm. The simulation results and conclusions are given in Sections VI and VII.

*Notations:* For a set of scalars $\{x_1, ..., x_N\}$ and an index set $\mathcal{S} \subseteq \{1, ..., N\}$, we use $[x_n]_{n\in\mathcal{S}}$ to denote a column vector consisting of the elements of $\{x_1, ..., x_N\}$ indexed by the set $\mathcal{S}$. Similarly, for a set of column vectors $\{\mathbf{x}_1, ..., \mathbf{x}_N\}$ with $\mathbf{x}_n \in \mathbb{C}^M$, $[\mathbf{x}_n]_{n\in\mathcal{S}}$ denotes a column vector consisting of the elements of $\{\mathbf{x}_1, ..., \mathbf{x}_N\}$ indexed by the set $\mathcal{S}$. We use $\mathbf{X}(:, j)$

| Notations | Meaning |
|---|---|
| $N_p$ | Number of downlink training vectors |
| $\mathbf{v}_t$ | Downlink training vector |
| $M(N)$ | Number of antennas at the BS (user) |
| $L_t$ | Number of propagation paths |
| $\alpha_{t,q}$ | The complex path gain of the $q$-th path |
| $f_{d,t}$ | The maximum DFO |
| $\xi_{t,q}(\vartheta_{t,q})$ | The AoD (AoA) of the $q$-th path |
| $\eta_t$ | Rotation angle of user's antenna array |
| $\theta_{T,m}(\theta_{R,m})$ | $m$-th AoD grid (AoA grid) |
| $\boldsymbol{\beta}_{T,t}(\boldsymbol{\beta}_{R,t})$ | The AoD/AoA off-grid vector |
| $N_b$ | Number of RF chains at the user |
| $\tilde{N}$ | Number of AoA grid |

Table I: The key notations used in the paper.

to denote the $j$-th column of a matrix $\mathbf{X}$. The key notations are summarized in Table I.

## II. System Model

### A. System Architecture and Frame Structure

Consider a time-division duplexing (TDD) mmWave massive MIMO system with one static BS serving a fast-moving user[1]. The BS is equipped with $M \gg 1$ antennas. The user is equipped with $N \gg 1$ antennas. The time is divided into frames, with each frame containing a downlink subframe and an uplink subframe, as illustrated in Fig. 1. Each subframe contains a large number of symbol durations.

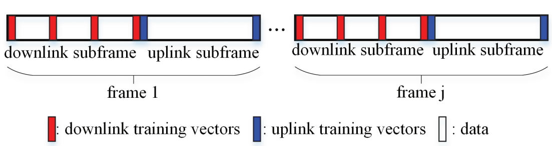


Figure 1: Illustration of frame strcture.

In the $t$-th downlink subframe, there are $N_p$ uniformly distributed training vectors, which are set to be the same vector denoted as $\mathbf{v}_t$. For convenience, we use $\mathcal{N}_p$ to denote the symbol index set for the $N_p$ training vectors. Note that inserting $N_p$ identical training vectors uniformly in the downlink subframe facilitates the estimation of AoAs and Doppler parameters at the user, as will be explained later. Based on the estimated AoAs and Doppler parameters, the user applies a Doppler compensation matrix to mitigate the Doppler effect and essentially converts the fast time-varying channel into a slow time-varying effective channel. In the uplink subframe, there are two sets of $N_p^u$ training vectors at the beginning and end of the uplink subframe, respectively. The two sets of uplink training vectors are used to estimate the slow time-varying effective channel after Doppler compensation. Specifically, the uplink transmission (e.g., beamforming and power allocation) in the $t$-th uplink subframe is optimized based on the slow time-varying effective channel estimated at the $t$-th uplink subframe. On the other hand, by making use of the channel reciprocity, the downlink transmission in the $t$-th downlink subframe is optimized based on the slow time-varying effective channel estimated at the end of the $(t-1)$-th uplink subframe. Since the effective channel after Doppler compensation changes slowly compared to the subframe duration, such a design can effectively overcome the channel aging issue caused by the Doppler effect.

[1]For clarity, we focus on a single user system. However, the proposed selective channel tracking and Doppler compensation scheme can be readily extended to multi-user systems.

### B. Doppler Multipath Channel Model

For clarity, we focus on the case when both the BS and mobile user are equipped with a half-wavelength space ULA and the channel is flat fading. To incorporate the DFO with conventional mmWave channels, the downlink channel model for the antenna pair $\{n_t, n_r\}$ is given by [23]

$$h_{n_r,n_t t,i} = \sum_{q=1}^{L_t} \alpha_{t,q} e^{j\left[2\pi f_{d,t} i cos(\vartheta_{t,q}+\eta_t)+\psi_{n_t}(\xi_{t,q})+\psi_{n_r}(\vartheta_{t,q})\right]}, \tag{1}$$

where $t$ stands for the frame index, $i$ stands for the symbol index, $L_t$ is the total number of propagation paths, $\alpha_{t,q}$ is the random complex path gain associated with the $q$-th propagation path, $f_{d,t}$ is the maximum DFO of the $t$-th frame, $\xi_{t,q}$ and $\vartheta_{t,q}$ are the AoD and AoA of the $q$-th path, respectively, and $\eta_t$ is rotation angle of the user's antenna array with respect to the moving direction in the t-th frame. Here, $\psi_{n_t}(\xi_{t,q})$ and $\psi_{n_r}(\vartheta_{t,q})$ represent the phase shifts induced at the $n_t$-th transmit antenna and the $n_r$-th receive antenna, respectively, which depend on the antenna structure, position, and the direction of the path. Note that in (1), we have implicitly assumed that the channel parameters $L_t, \alpha_{t,q}, \xi_{t,q}, \vartheta_{t,q}, f_{d,t}, \eta_t$ are fixed within each frame but may change over different frames, which is usually true even for high-speed users [18]. However, the channel $h_{n_r,n_t,t,i}$ itself may change over different symbols at a much faster timescale due to the fast changing phase term $2\pi f_{d,t} icos(\vartheta_{t,q} + \eta_t)$ caused by the Doppler effect.

### C. Angular Domain Channel Representation

To obtain the angular domain channel representation, we introduce a uniform grid of $\tilde{M}$ AoDs and $\tilde{N}$ AoAs over $[0, 2\pi)$

$$\{\theta_{T,m}: \; sin(\theta_{T,m}) = \frac{2}{\tilde{M}}\left(m - \left\lfloor \frac{\tilde{M}-1}{2} \right\rfloor\right), m = 0, \ldots, \tilde{M}-1\},$$

$$\{\theta_{R,n}: \; sin(\theta_{R,n}) = \frac{2}{\tilde{N}}\left(n - \left\lfloor \frac{\tilde{N}-1}{2} \right\rfloor\right), n = 0, \ldots, \tilde{N}-1\}.$$

In practice, the true AoDs/AOAs usually do not lie exactly on the grid points. In this case, there will be mismatches between the true AoDs/AOAs and the nearest grid point. To overcome this issue, we introduce an off-grid basis for the angular domain channel representation, as in [25]. Specifically, let $\theta_{T,m_{t,q}}$ and $\theta_{R,n_{t,q}}$ denote the nearest grid point to $\xi_{t,q}$ and $\vartheta_{t,q}$, respectively. We introduce the AoD off-grid vector $\boldsymbol{\beta}_{T,t}=\left[\beta_{T,t,1},\beta_{T,t,2},...,\beta_{T,t,\tilde{M}}\right]^{T}$ such that

$$\beta_{T,t,m}=\begin{cases}\xi_{t,q}-\theta_{T,m_{t,q}}, & m=m_{t,q},q=1,2,...,L_{t}\\ 0, & \text{otherwise}\end{cases}.$$

Similarly, let $\boldsymbol{\beta}_{R,t}=\left[\beta_{R,t,1},\beta_{R,t,2},...,\beta_{R,t,\tilde{N}}\right]^{T}$ denote the AoA off-grid vector, such that

$$\beta_{R,t,n}=\begin{cases}\vartheta_{t,q}-\theta_{R,n_{t,q}}, & n=n_{t,q},q=1,2,...,L_{t}\\ 0, & \text{otherwise}\end{cases}.$$

For half-wavelength space ULAs, the array response vectors at the BS and user side are given by $\boldsymbol{a}_{T}(\theta)=\frac{1}{\sqrt{M}}\left[1,e^{-j\pi sin(\theta)},e^{-j2\pi sin(\theta)},\ldots,e^{-j(M-1)\pi sin(\theta)}\right]^{T}$ and $\boldsymbol{a}_{R}(\theta)=\frac{1}{\sqrt{N}}\left[1,e^{-j\pi sin(\theta)},e^{-j2\pi sin(\theta)},\ldots,e^{-j(N-1)\pi sin(\theta)}\right]^{T}$. For convenience, define two matrices $\boldsymbol{A}_{R,i}\left(\boldsymbol{\beta}_{R,t},f_{d,t},\eta_{t}\right)=\left[\tilde{\boldsymbol{a}}_{R,i}\left(\beta_{R,t,1},f_{d,t},\eta_{t}\right),\ldots,\tilde{\boldsymbol{a}}_{R,i}\left(\beta_{R,t,\tilde{N}},f_{d,t},\eta_{t}\right)\right]\in\mathbb{C}^{N\times\tilde{N}}$ and $\boldsymbol{A}_{T}(\boldsymbol{\beta}_{T,t})=\left[\boldsymbol{a}_{T}(\theta_{T,1}+\beta_{T,t,1}),...,\boldsymbol{a}_{T}(\theta_{T,\tilde{M}}+\beta_{T,t,\tilde{M}})\right]\in\mathbb{C}^{M\times\tilde{M}}$, where $\tilde{\boldsymbol{a}}_{R,i}(\beta_{R,t,n},f_{d,t},\eta_{t})=\boldsymbol{a}_{R}\left(\theta_{R,n}+\beta_{R,t,n}\right)\times e^{j2\pi f_{d,t}icos(\theta_{R,n}+\beta_{R,t,n}+\eta_{t})}$. Furthermore, define $\tilde{\boldsymbol{X}}_{t}\in\mathbb{C}^{\tilde{N}\times\tilde{M}}$ as the angular domain channel matrix with the $(n,m)$-th element given by

$$\tilde{x}_{t,n,m}=\begin{cases}\alpha_{t,q}, & (n,m)=(n_{t,q},m_{t,q}),q=1,2,...,L_{t}\\ 0, & \text{otherwise}\end{cases}.$$

Then, for given AoA off-grid, DFO parameter , rotation angle pair $\boldsymbol{\varphi}_{t}=\left\{\boldsymbol{\beta}_{R,t},f_{d,t},\eta_{t}\right\}$, and AoD off-grid vector $\boldsymbol{\beta}_{T,t}$, $\boldsymbol{H}_{t,i}$ can be expressed in a compact form as

$$\boldsymbol{H}_{t,i}\left(\boldsymbol{\varphi}_{t},\boldsymbol{\beta}_{T,t}\right)=\boldsymbol{A}_{R,i}(\boldsymbol{\varphi}_{t})\tilde{\boldsymbol{X}}_{t}\boldsymbol{A}_{T}^{H}(\boldsymbol{\beta}_{T,t}),\tag{2}$$

where t $e^{j\psi_{n_{t}}(\xi_{t,q})}$ and $e^{j\psi_{n_{r}}(\vartheta_{t,q})}$ in (1) are implicitly contained in the array response matrices $\boldsymbol{A}_{R,i}(\boldsymbol{\varphi}_{t})$ and $\boldsymbol{A}_{T}^{H}(\boldsymbol{\beta}_{T,t})$.

Note that we can also define the angular domain representation for more general 2-dimensional (2D) antenna arrays. In this case, the array response vector $\boldsymbol{a}_{T}(\theta,\phi)$ (or $\boldsymbol{a}_{R}(\theta,\phi)$) can be expressed as a function of the azimuth angle $\theta$ and elevation angle $\phi$. Please refer to [26] for the details.

## III. Angular-Domain Selective Channel Tracking and Doppler Compensation

In this section, we propose an efficient angular-domain selective channel tracking and Doppler compensation scheme at the user side. The proposed scheme can exploit both dynamic sparsity of mmWave massive MIMO channel and high resolution of AoA at multi-antenna mobile users to accurately estimate the downlink AoAs and maximum DFO. Using these estimated parameters, a Doppler compensation matrix is applied at the user to convert the fast time-varying channel into a slow time-varying effective channel, based on which efficient downlink/uplink transmissions can be achieved. The proposed scheme includes four key components, namely the Angular-Domain Selective Channel Tracking, Selective Doppler Compensation, Slow Time-Varying Effective Channel Estimation, and Downlink Training Vector Design. Fig. 2 illustrates a top-level diagram of the proposed scheme and the details of each component are elaborated below. The frame index $t$ will be omitted when there is no ambiguity.

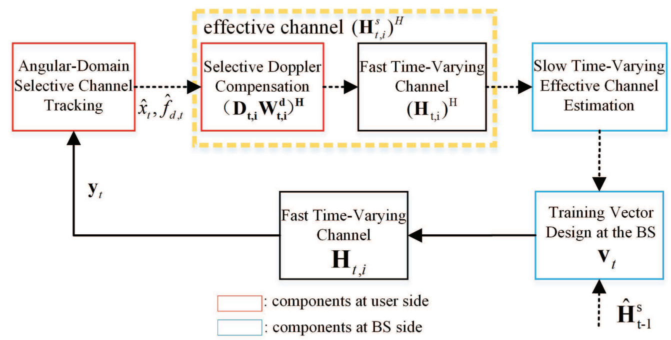


Figure 2: A top-level diagram of the proposed scheme.

### A. *Outline of Angular-Domain Selective Channel Tracking at the User*

This component is used to estimate the downlink AoAs, rotation angle and maximum DFO based on the $N_{p}$ downlink training vectors. Thanks to the high-spatial resolution provided by the large array at the user side, the user can distinguish DFOs associated with different AOAs from multiple active paths. However, since both the BS and the user are equipped with the large array, the parameter space can be very large if we attempt to estimate the full angular–domain channel parameters (i.e., the full angular domain channel matrix $\tilde{\boldsymbol{X}}$, rotation angle $\eta$ and maximum DFO $f_{d}$). Since the DFO only occurs at the mobile user side, we propose to only estimate partial angular–domain channel parameters that are just sufficient to obtain AoAs, rotation angle and maximum DFO for Doppler compensation.

Specifically, the product of channel $\boldsymbol{H}_{i}$ and downlink training vector $\mathbf{v}$ can be expressed as:

$$\begin{aligned}\boldsymbol{H}_{i}\mathbf{v}&=\sum_{n=1}^{\tilde{N}}\sum_{m=1}^{\tilde{M}}\tilde{x}_{n,m}\tilde{\boldsymbol{a}}_{R,i}(\boldsymbol{\varphi})\boldsymbol{a}_{T}^{H}(\theta_{T,m}+\beta_{T,m})\mathbf{v},\\&=\sum_{n=1}^{\tilde{N}}x_{n}\tilde{\boldsymbol{a}}_{R,i}(\boldsymbol{\varphi})=\boldsymbol{A}_{R,i}(\boldsymbol{\varphi})\boldsymbol{x},\end{aligned}\tag{3}$$

where $\boldsymbol{x}=\left[x_{1},...,x_{\tilde{N}}\right]^{T}$ with $x_{n}=\sum_{m=1}^{\tilde{M}}\tilde{x}_{n,m}\boldsymbol{a}_{T}^{H}(\theta_{T,m}+\beta_{T,m})\mathbf{v}$ are called partial angular channel coefficients since $\boldsymbol{x}$ only contain partial information about the full angular channel $\tilde{\boldsymbol{X}}$. Specifically, a non-zero $|x_n|^2$ with value larger than the noise floor indicates that there is an active path to the $n$-th AoA direction at the user side. Therefore, we only need to estimate $\tilde{N}$ partial channel parameters $\boldsymbol{x}$, the AoA off-grid vector $\boldsymbol{\beta}_R$, rotation angle $\eta$ and maximum DFO $f_d$, which are much less than the original $\tilde{N}\tilde{M}$ full channel parameters, the off-grid vector , rotation angle and maximum DFO. Note that if the $N_p$ training vectors are different, there will be $\tilde{N}N_p$ partial angular channel coefficients, leading to a larger parameter space to be estimated. Moreover, with uniformly distributed training vectors, the phase rotation due to the Doppler term $e^{j2\pi f_d icos(\theta_{R,n}+\beta_{R,t,n}+\eta_t)}$ is larger compared to the case when the $N_p$ training vectors are squeezed in the beginning of the downlink subframe, leading to a better estimation performance for the Doppler parameter $f_d$. Therefore, such a selective channel tracking design based on $N_p$ uniformly distributed and identical training vectors can significantly reduce the number of downlink training vectors $N_p$ required to achieve accurate estimation of AoAs, rotation angle and Doppler parameters.

The received baseband pilot signal is given by

$$\boldsymbol{y}_i = \boldsymbol{H}_i \mathbf{v} + \boldsymbol{n}_i, \forall i \in \mathcal{N}_p \tag{4}$$

where $\mathbf{v} \in \mathbb{C}^M$ is the training vector for downlink channel tracking, and $\mathbf{n}_i$ is the additive white Gaussian noise (AWGN) with each element having zero mean and variance $\sigma^2$, respectively. The exact choice of $\mathbf{v}$ is postponed to Section III-D.

The aggregate received pilot signal (channel measurements) of all the $N_p$ downlink pilot symbols (training vectors) in the $t$-th frame can be expressed in a compact form as

$$\boldsymbol{y} = [\boldsymbol{H}_i \mathbf{v} + \boldsymbol{n}_i]_{i\in\mathcal{N}_p}\,. \tag{5}$$

Based on the received downlink training vectors, the user obtains the estimated partial channel parameters $\hat{\boldsymbol{x}}$, $\hat{\boldsymbol{\beta}}_R$, $\hat{\eta}$ and $\hat{f}_d$ using a selective channel tracking algorithm. The detailed problem formulation and algorithm design for selective channel tracking scheme are postponed to Section IV.

### B. Angular-Domain Selective Doppler Compensation at the User

This component is used to convert the fast time-varying channel into a slow time-varying effective channel after obtaining the estimated partial channel parameters $\hat{\boldsymbol{x}}$, $\hat{\boldsymbol{\beta}}_R$, $\hat{\eta}_t$ and $\hat{f}_d$. We first select a set of $N_d$ most significant AoA directions with the largest energy, where the energy of the $n$-th AoA direction $\theta_{R,n}+\beta_{R,n}$ is defined as $|\hat{x}_n|^2$. The parameter $N_d \leq N$ is used to control the tradeoff between the spatial multiplexing gain and the effective CSI signaling overhead (i.e., the CSI signaling overhead required to obtain effective channel $\boldsymbol{H}_i^s$ in (6)). Let $\mathcal{N}_d \subseteq \{1,...,N\}$ denote the index set of the selected $N_d$ most significant AoA directions. Then, in order to mitigate the Doppler effect and perform per-AoA DFO compensation for each selected AoA direction, a DFO compensation matrix $\mathbf{W}_i^d \mathbf{D}_i \in \mathbb{C}^{N\times N_d}$, which also serves as beamforming matrix, is applied at the user side. In this way, we can convert the fast time-varying channel $\boldsymbol{H}_i$ into a slow time-varying effective channel $\boldsymbol{H}_i^s \in \mathbb{C}^{N_d\times M}$ as

$$\boldsymbol{H}_i^s = \mathbf{D}_i^H (\mathbf{W}_i^d)^H \boldsymbol{H}_i, \tag{6}$$

where $\mathbf{W}_i^d = [\boldsymbol{a}_R(\theta_{R,n}+\beta_{R,n})]_{n\in\mathcal{N}_d} \in \mathbb{C}^{N\times N_d}$ and $\mathbf{D}_i = \text{Diag}\left(\left[e^{j2\pi \hat{f}_d icos(\theta_{R,n}+\beta_{R,n}+\eta_t)}\right]_{n\in\mathcal{N}_d}\right) \in \mathbb{C}^{N_d\times N_d}$.

In the following, we explain why the Doppler effect can be alleviated by applying the above DFO compensation matrix to obtain an effective channel $\boldsymbol{H}_i^s$. For half-wavelength space ULA, if there is no estimation error for the partial channel parameters, we have

$$\boldsymbol{H}_i^s = \sum_{m=1}^{\tilde{M}} [\tilde{x}_{n,m}]_{n\in\mathcal{N}_d} \boldsymbol{a}_T^H(\theta_{T,m}+\beta_{T,m}) + O\left(\frac{1}{\sqrt{N}}\right), \tag{7}$$

as $N\to\infty$ [27]. From (7), $\boldsymbol{H}_i^s$ is constant within a frame if we ignore the small order term $O\left(\frac{1}{\sqrt{N}}\right)$, i.e., the Doppler effect can be completely eliminated for sufficiently large $N$. This observation is also consistent with the results in [18].

### C. Slow Time-Varying Effective Channel Estimation at the BS

The user can simply transmit $N_d$ orthogonal pilots in the uplink training stage. Then the conventional Least Squares (LS) based channel estimation method can be used at the BS to obtain the estimated slow time-varying effective channel $\hat{\boldsymbol{H}}_i^s$. Based on $\hat{\boldsymbol{H}}_i^s$, the BS can optimize the precoder for both uplink and downlink transmissions. Note that the optimization of MIMO precoder is a standard problem and there are many existing solutions with different performance and complexity tradeoff. Then, the optimized uplink precoder is fed back to the user for uplink transmission. Since $N_d$ can be much less than $N$, the feedback overhead for the optimized uplink precoder is acceptable for practice.

### D. Training Vector Design at the BS

The training vector $\mathbf{v}_t$ at the BS is designed according to the slow time-varying effective channel $\hat{\boldsymbol{H}}_{t-1}^s$ estimated at the end of the $(t-1)$-th uplink subframe. The basic idea for training vector design is to strike a balance between *exploitation* of known channel directions (i.e., transmitting training signal over the most promising channel directions with large effective channel energy to achieve beamforming gain) and *exploration* of unknown channel directions (i.e., transmitting training signal over other channel directions to detect unknown channel directions). Since the effective channel $\boldsymbol{H}_i^s$ changes slowly, the effective channel $\hat{\boldsymbol{H}}_{t-1}^s$ estimated at the end of the $(t-1)$-th uplink subframe is expected to provide valuable information for the most promising channel directions. On the other hand, the information about the most promising channel directions extracted from $\hat{\boldsymbol{H}}_{t-1}^s$ may not be perfect due to the estimation error and CSI delay. In addition, some new direction may arise in the next frame. Therefore, the other channel directions should also be incorporated into the training vector to facilitate the detection of unknown channel directions.

Specifically, we first project the estimated effective channel $\hat{\boldsymbol{H}}_{t-1}^{s}$ onto an orthogonal basis $\mathbf{B}^{s}=[\mathbf{b}_{1}^{s},...,\mathbf{b}_{M}^{s}]\in\mathbb{C}^{M\times M}$ to obtain the effective channel energy on each basis vector (quantized channel direction) as $\lambda_{m}^{s}=\left\Vert \hat{\boldsymbol{H}}_{t-1}^{s}\mathbf{b}_{m}^{s}\right\Vert ^{2},\forall m$. The basis matrix is chosen such that the projection vector $\boldsymbol{\lambda}^{s}=[\lambda_{1}^{s},...,\lambda_{M}^{s}]^{T}$ is as sparse as possible. For half-wavelength ULAs, we can simply choose the basis matrix $\mathbf{B}^{s}$ as an $M\times M$ DFT matrix. Then we find the index set of the most promising channel directions as

$$\mathcal{M}^{*}=\text{argmin}_{\mathcal{M}}\left|\mathcal{M}\right|,\ \text{s.t.}\ \sum_{m\in\mathcal{M}}\lambda_{m}^{s}/\sum_{m=1}^{M}\lambda_{m}^{s}\geq\mu,\tag{8}$$

where $\mu$ is a threshold which is chosen to be closed to 1. In other words, the most promising channel directions contain $\mu$ fraction of the total effective channel energy. Let $N_{s}=\left|\mathcal{M}^{*}\right|$. Finally, the training vector is given by

$$\begin{aligned}\mathbf{v}_{t}= & \frac{\sqrt{\rho}}{\sqrt{N_{s}}}\sum_{m\in\mathcal{M}^{*}}e^{j\theta_{m}^{s}}\mathbf{b}_{m}^{s}\\
 & +\frac{\sqrt{1-\rho}}{\sqrt{M-N_{s}}}\sum_{m\in\{1,...,M\}\backslash\mathcal{M}^{*}}e^{j\theta_{m}^{s}}\mathbf{b}_{m}^{s}.\end{aligned}\tag{9}$$

where the first term in (9) exploites the information about the most promising $N_{s}$ channel directions extracted from $\hat{\boldsymbol{H}}_{t-1}^{s}$, $\rho$ is a system parameter which determines the proportion of transmit power used to exploit the most promising channel directions, the second term is used to detect the other unknown channel directions, $\theta_{m}^{s}$ is randomly generated from $[0,2\pi]$.

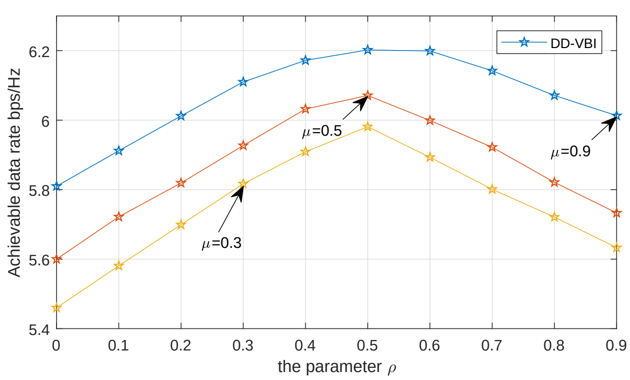


Figure 3: Achievable data rate versus the parameter $\rho$ with different values of the parameter $\mu$.

In Fig. 3, we illustrate how the achievable data rate is affected by changing the parameters $\rho$ and $\mu$ to achieve different tradeoffs between the exploration and exploitation. It can be seen that setting $\mu=0.9$ and $\rho=0.5$ can strike a good balance between *exploitation* and *exploration*.

# IV. Problem Formulation for Angular-Domain Selective Channel Tracking

## A. Three-layer Hierarchical Markov Model for Partial Angular Channel Vector

The dynamic sparsity of the partial angular channel coefficients $\boldsymbol{x}_{t}$ is captured using a three-layer hierarchical Markov model, as illustrated in Fig. 4. The first layer of random variable is the channel support vector $\boldsymbol{s}_{t}\in\{0,1\}^{\tilde{N}}$, whose $n$-th element, denoted by $s_{t,n}$, indicates whether the channel coefficient $x_{t,n}$ is active ($s_{t,n}=1$) or not ($s_{t,n}=0$). The second layer of random variable is the precision vector $\boldsymbol{\gamma}_{t}=[\gamma_{t,1},\cdots,\gamma_{t,\tilde{N}}]^{T}$, where $\gamma_{t,n}$ represents the precision (inverse of the variance) of $x_{t,n}$. The third layer of random variables are partial angular channel coefficients $\boldsymbol{x}_{t}$. For convenience, denote a time series of vectors $\{\boldsymbol{x}_{\tau}\}_{\tau=1}^{t}$ as $\boldsymbol{x}_{1:t}$ (same for $\boldsymbol{\gamma}_{1:t}$, $\boldsymbol{s}_{1:t}$, $f_{d,1:t}$,$\boldsymbol{\beta}_{R,1:t}$). Then the three-layer hierarchical Markov prior distribution (joint distribution of $\boldsymbol{x}_{1:t}$, $\boldsymbol{\gamma}_{1:t}$ and $\boldsymbol{s}_{1:t}$ ) is given by

$$p(\boldsymbol{x}_{1:t},\boldsymbol{\gamma}_{1:t},\boldsymbol{s}_{1:t})=\prod_{\tau=1}^{t}p\left(\boldsymbol{s}_{\tau}|\boldsymbol{s}_{\tau-1}\right)p(\boldsymbol{\gamma}_{\tau}|\boldsymbol{s}_{\tau})p(\boldsymbol{x}_{\tau}|\boldsymbol{\gamma}_{\tau}),\tag{10}$$

where $p\left(\boldsymbol{s}_{1}|\boldsymbol{s}_{0}\right)\triangleq p\left(\boldsymbol{s}_{1}\right)$, the conditional probability $p\left(\boldsymbol{x}_{\tau}|\boldsymbol{\gamma}_{\tau}\right)$ has a product form $p\left(\boldsymbol{x}_{\tau}|\boldsymbol{\gamma}_{\tau}\right)=\prod_{n=1}^{\tilde{N}}p\left(x_{\tau,n}|\gamma_{\tau,n}\right)$ and each is modeled as a Gaussian prior distribution

$$p\left(x_{\tau,n}|\gamma_{\tau,n}\right)=CN\left(x_{\tau,n};0,\gamma_{\tau,n}^{-1}\right),\tag{11}$$

The conditional prior of precision vector $\boldsymbol{\gamma}_{\tau}$ is given by

$$p\left(\boldsymbol{\gamma}_{\tau}|\boldsymbol{s}_{\tau}\right)=\prod_{n=1}^{\tilde{N}}\Gamma\left(\gamma_{\tau,n};a_{\tau},b_{\tau}\right)^{s_{\tau,n}}\Gamma\left(\gamma_{\tau,n};\overline{a}_{\tau},\overline{b}_{\tau}\right)^{1-s_{\tau,n}},\tag{12}$$

$\Gamma\left(\gamma;a_{\gamma},b_{\gamma}\right)$ is a Gamma hyperprior. $a_{\tau},b_{\tau}$ are the shape and rate parameters of the channel precision $\gamma_{\tau,n}$ conditioned on $s_{\tau,n}=1$ and they should be chosen such that $\frac{a_{\tau}}{b_{\tau}}=E[\gamma_{\tau,n}]=\Theta(1)$, since the variance $\gamma_{\tau,n}^{-1}$ of $x_{\tau,n}$ is $\Theta(1)$ when it is active ($s_{\tau,n}=1$). $\overline{a}_{\tau},\overline{b}_{\tau}$ are the shape and rate parameters , conditioned on the opposite event (i.e., $s_{\tau,n}=0$). In this case, the shape and rate parameters $\overline{a}_{\tau},\overline{b}_{\tau}$ of the precision $\gamma_{\tau,n}$ should be chosen such that $\frac{\overline{a}_{\tau}}{\overline{b}_{\tau}}=E[\gamma_{\tau,n}]\gg1$, since the variance $\gamma_{\tau,n}^{-1}$ of $x_{\tau,n}$ is close to zero when it is inactive.

Note that the exact channel distribution is usually unknown in practice. In this case, it is reasonable to choose a prior distribution such that the derived algorithm can promote sparsity with low complexity and achieve robust performance to different channel distributions. By controlling the parameters in the Gamma distribution of $\gamma_{\tau,n}$, one can easily promote sparsity based on the knowledge of channel support $\boldsymbol{s}_{\tau}$, as explained above. Moreover, Since the Gamma distribution for $\gamma_{\tau,n}$ is the conjugate probability distribution of the Gaussian distribution for $x_{\tau,n}$, the above hierarchical prior for $x_{\tau,n}$ and $\gamma_{\tau,n}$ facilitates low-complexity VBI algorithm design with closed-form update equations [24]. Finally, the VBI-type algorithm derived from the such a hierarchical prior is well known to be insensitive to the true distribution of the sparse signals [24], [28]. As a result, similar hierarchical prior distribution has been widely adopted in sparse Bayesian learning [29].

Due to the slowly changing propagation environment, the channel supports often change slowly over time, which implies that $s_{\tau,n}$ depends on $s_{\tau-1,n}$, e.g., if $s_{\tau-1,n}=1$, then there is a higher probability that $s_{\tau,n}$ is also 1. Such dynamic sparsity of support vectors can be naturally modeled as a temporal Markov model with an initial prior distribution $p(\mathbf{s}_1)$ and a transition probability:

$$p(\mathbf{s}_\tau|\mathbf{s}_{\tau-1}) = \prod_{n=1}^{\tilde{N}} p(s_{\tau,n}|s_{\tau-1,n}), \tag{13}$$

where the transition probability is given by $p(s_{\tau,n} = 1|s_{\tau-1,n} = 0) = \rho_{0,1}$, and $p(s_{\tau,n} = 0|s_{\tau-1,n} = 1) = \rho_{1,0}$. The Markov parameters $\{\rho_{1,0}, \rho_{0,1}\}$ characterize the degree of temporal correlation of the channel support. Specifically, smaller $\rho_{1,0}$ or $\rho_{0,1}$ lead to highly correlated supports across time, which means the propagation environment between the user and BS is changing slowly. Larger $\rho_{1,0}$ or $\rho_{0,1}$ can allow support to change substantially across time, which means the propagation environment is changing significantly. Moreover, the statistic parameters $\{\rho_{1,0}, \rho_{0,1}\}$ could be automatically learned based on the EM framework during the recovery process [30], as detailed in Appendix A. The initial distribution $p(s_{1,n}), \forall n$ is set to be the steady state distribution of the Markov chain in (13), i.e.,

$$\lambda \triangleq p(s_{1,n}) = \frac{\rho_{0,1}}{\rho_{0,1}+\rho_{1,0}}.$$

This ensures that all elements of $s_{\tau,n}$ have the same marginal distribution $p(s_{\tau,n}) = \lambda^{s_{\tau,n}}(1-\lambda)^{1-s_{\tau,n}}$.

In practice, the noise precision $\kappa_\tau = \sigma_\tau^{-2}$ is usually unknown and we model it as a Gamma hyperpriror $p(\kappa_\tau) = \Gamma(\kappa_\tau; a_{\kappa,\tau}, b_{\kappa,\tau})$, where we set $a_{\kappa,\tau}, b_{\kappa,\tau} \to 0$ as in [25] so as to obtain a broad hyperprior.

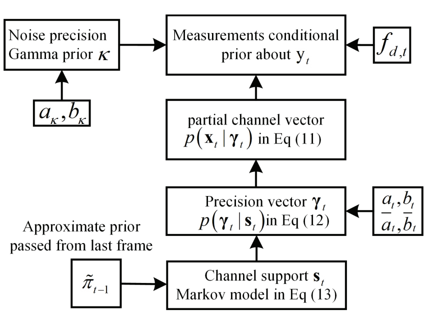


Figure 4: Three-layer hierarchical Markov model for partial angular channel coefficients.

*B. Selective Channel Tracking Formulation*

Using the angular domain channel representation, the receive signal at $t$-th frame $\boldsymbol{y}_t \in \mathbb{C}^{NN_p}$ can be rewritten as a CS model with an unknown AoA off-grid and DFO parameter pair $\boldsymbol{\varphi}_t = \{\boldsymbol{\beta}_{R,t}, \eta_t, f_{d,t}\}$ in the measurement matrix as

$$\boldsymbol{y}_t = \boldsymbol{F}_t \boldsymbol{x}_t + \boldsymbol{n}_t, \tag{14}$$

where the measurement matrix is given by $\boldsymbol{F}_t = [\boldsymbol{F}_{t,1}; ...; \boldsymbol{F}_{t,N_p}] \in C^{NN_p \times \tilde{N}}$, $\boldsymbol{F}_{t,i} = \boldsymbol{A}_{R,i}(\boldsymbol{\varphi}_t)$, $\boldsymbol{n}_t = [\boldsymbol{n}_{t,i}]_{i\in\mathcal{N}_p}$.

In each frame $t$, the user needs to estimate the partial channel parameters $\boldsymbol{x}_t$ , the AoA off-grid and DFO parameter pair $\boldsymbol{\varphi}_t = \{\boldsymbol{\beta}_{R,t}, \eta_t, f_{d,t}\}$, given the observations up to $t$ frame $\boldsymbol{y}_{1:t}$ in model (14), the estimated AoA off-grid and DFO parameter pairs $\hat{\boldsymbol{\varphi}}_{1:t-1} = \left\{\hat{\boldsymbol{\beta}}_{R,1:t-1}, \hat{\eta}_{1:t-1}, \hat{f}_{d,1:t-1}\right\}$ up to $(t-1)$ frame. In particular, for given $\boldsymbol{\varphi}_t$, we are interested in computing minimum mean-squared error (MMSE) estimates of $x_{t,n}$, $\hat{x}_{t,n} = \mathrm{E}\left[x_{t,n}|\boldsymbol{y}_{1:t}; \hat{\boldsymbol{\varphi}}_{1:t-1}, \boldsymbol{\varphi}_t\right]$, where the expectation is over the marginal posterior:

$$\begin{aligned} & p(x_{t,n}|\boldsymbol{y}_{1:t}; \hat{\boldsymbol{\varphi}}_{1:t-1}, \boldsymbol{\varphi}_t) \\ & \propto \int_{-x_{t,n}} p(\boldsymbol{y}_{1:t}, \boldsymbol{v}_t; \hat{\boldsymbol{\varphi}}_{1:t-1}, \boldsymbol{\varphi}_t), \end{aligned} \tag{15}$$

where $\boldsymbol{v}_t = \{\boldsymbol{x}_t, \boldsymbol{s}_t, \boldsymbol{\gamma}_t, \kappa_t\}$, $-x_{t,n}$ denotes the vector collections integration over $\boldsymbol{v}_t$ except for the element $x_{t,n}$ and $\propto$ denotes equality after scaling.

On the other hand, the optimal $\boldsymbol{\varphi}_t$ at the $t$-th frame is obtained by ML as follows [25]:

$$\begin{aligned} \hat{\boldsymbol{\varphi}}_t &= \arg\max_{\boldsymbol{\varphi}_t} \ln p(\boldsymbol{y}_{1:t}; \hat{\boldsymbol{\varphi}}_{1:t-1}, \boldsymbol{\varphi}_t) \\ &= \arg\max_{\boldsymbol{\varphi}_t} \ln \int_{\boldsymbol{v}_t} p(\boldsymbol{y}_{1:t}, \boldsymbol{v}_t; \hat{\boldsymbol{\varphi}}_{1:t-1}, \boldsymbol{\varphi}_t) d\boldsymbol{v}_t. \end{aligned} \tag{16}$$

Once we obtain the ML estimate of $\hat{\boldsymbol{\varphi}}_t$, and the associated conditional marginal posterior $p(x_{t,n}|\boldsymbol{y}_{1:t}; \hat{\boldsymbol{\varphi}}_{1:t-1}, \boldsymbol{\varphi}_t)$, we can obtain the MMSE estimates of $\{x_{t,n}\}$.

One challenge in computing the MMSE estimate is the calculation of the exact posterior in (15) whose factor graph has loops. In the next subsection, we propose a Doppler-aware-dynamic-VBI (DD-VBI) algorithm to approximately calculate the marginal posteriors $p(x_{t,n}|\boldsymbol{y}_{1:t}; \hat{\boldsymbol{\varphi}}_{1:t-1}, \boldsymbol{\varphi}_t)$ by combining the message passing and VBI approaches, and use the inexact majorization-minimization (MM) method (which is a generalization of the EM method) [25] to find an approximate solution for (16). The proposed DD-VBI algorithm is shown in the simulations to achieve a good performance.

## V. Doppler-Aware-Dynamic-VBI Algorithm

### *A. Decomposition and Approximation of Joint Probability Distribution*

This section is to decompose and approximate the joint probability distribution in (16) such that the joint probability distribution at the $t$-th frame only involves the probability density function (PDF) of the current hidden variables $\boldsymbol{v}_t$, the current observation $\boldsymbol{y}_t$, and the messages $\hat{p}(\boldsymbol{s}_t|\boldsymbol{y}_{1:t-1}, \hat{\boldsymbol{\varphi}}_{1:t-1})$ passed from the previous frame, based on which a more efficient algorithm can be designed.

The joint probability distribution in (16) and (15) can be written as

$$
\begin{aligned}
&p(\boldsymbol{y}_{1:t}, \boldsymbol{v}_t; \hat{\boldsymbol{\varphi}}_{1:t-1}, \boldsymbol{\varphi}_t) \\
\propto &\sum_{\boldsymbol{s}_{t-1}} p(\mathbf{s}_{t-1}|\mathbf{y}_{1:t-1}; \hat{\boldsymbol{\varphi}}_{1:t-1}) p(\boldsymbol{s}_t|\boldsymbol{s}_{t-1}) \\
&p(\boldsymbol{y}_t|\boldsymbol{x}_t, \kappa_t; \boldsymbol{\varphi}_t) p(\boldsymbol{x}_t|\boldsymbol{\gamma}_t) p(\boldsymbol{\gamma}_t|\boldsymbol{s}_t) p(\kappa_t) \\
\approx &\sum_{\boldsymbol{s}_{t-1}} q(\boldsymbol{s}_{t-1}|\boldsymbol{y}_{1:t-1}; \hat{\boldsymbol{\varphi}}_{1:t-1}) p(\boldsymbol{s}_t|\boldsymbol{s}_{t-1}) \\
&p(\boldsymbol{y}_t|\boldsymbol{x}_t, \kappa_t; \boldsymbol{\varphi}_t) p(\boldsymbol{x}_t|\boldsymbol{\gamma}_t) p(\boldsymbol{\gamma}_t|\boldsymbol{s}_t) p(\kappa_t) \\
= &\hat{p}(\boldsymbol{s}_t|\boldsymbol{y}_{1:t-1}; \hat{\boldsymbol{\varphi}}_{1:t-1}) p(\boldsymbol{y}_t|\boldsymbol{x}_t, \kappa_t; \boldsymbol{\varphi}_t) \\
&p(\boldsymbol{x}_t|\boldsymbol{\gamma}_t) p(\boldsymbol{\gamma}_t|\boldsymbol{s}_t) p(\kappa_t),
\end{aligned}
$$

where $\hat{p}(\boldsymbol{s}_t|\boldsymbol{y}_{1:t-1}; \hat{\boldsymbol{\varphi}}_{1:t-1}) = \sum_{\mathbf{s}_{t-1}} q(\boldsymbol{s}_{t-1}|\boldsymbol{y}_{1:t-1}; \hat{\boldsymbol{\varphi}}_{1:t-1}) p(\boldsymbol{s}_t|\boldsymbol{s}_{t-1})$, $q(\boldsymbol{s}_{t-1}|\boldsymbol{y}_{1:t-1}; \hat{\boldsymbol{\varphi}}_{1:t-1})$ is a tractable approximation for the posterior $p(\mathbf{s}_{t-1}|\mathbf{y}_{1:t-1}; \hat{\boldsymbol{\varphi}}_{1:t-1})$ and $p\left(\boldsymbol{y}_t|\boldsymbol{x}_t, \kappa_t; \boldsymbol{\varphi}_t\right) = \mathcal{CN}\left(\boldsymbol{y}_t; \boldsymbol{F}_t\boldsymbol{x}_t, \kappa_t^{-1}\boldsymbol{I}\right)$. Both $q(\boldsymbol{s}_{t-1}|\boldsymbol{y}_{1:t-1}; \hat{\boldsymbol{\varphi}}_{1:t-1})$ and $\hat{p}(\boldsymbol{s}_t|\boldsymbol{y}_{1:t-1}; \hat{\boldsymbol{\varphi}}_{1:t-1})$ can be calculated based on the messages passed from the previous frame. We will elaborate how to calculate $\hat{p}(\boldsymbol{s}_t|\boldsymbol{y}_{1:t-1}; \hat{\boldsymbol{\varphi}}_{1:t-1})$ later in subsection V-E. When $t = 1$, $\hat{p}(\boldsymbol{s}_t|\boldsymbol{y}_{1:t-1}; \hat{\boldsymbol{\varphi}}_{1:t-1})$ is reduced to $\hat{p}(\boldsymbol{s}_t|\boldsymbol{y}_1; \hat{\boldsymbol{\varphi}}_1) = p(\boldsymbol{s}_1)$.

For simplicity, we define

$$
\begin{aligned}
&\hat{p}(\boldsymbol{y}_{1:t}, \boldsymbol{v}_t; \hat{\boldsymbol{\varphi}}_{1:t-1}, \boldsymbol{\varphi}_t) \\
= &\hat{p}(\boldsymbol{s}_t|\boldsymbol{y}_{1:t-1}; \hat{\boldsymbol{\varphi}}_{1:t-1}) p(\boldsymbol{y}_t|\boldsymbol{x}_t, \kappa_t; \boldsymbol{\varphi}_t) \\
&p(\boldsymbol{x}_t|\boldsymbol{\gamma}_t) p(\boldsymbol{\gamma}_t|\boldsymbol{s}_t) p(\kappa_t).
\end{aligned} \tag{17}
$$

In the rest of this section, we will omit $\hat{\boldsymbol{\varphi}}_{1:t-1}$ in the PDFs when there is no ambiguity.

### *B. Outline of the Doppler-Aware-Dynamic-VBI algorithm in Frame $t$*

The basic idea of the DD-VBI algorithm is that, at every frame $t$, simultaneously approximates the marginal posterior $\{p(x_{t,n}|\boldsymbol{y}_{1:t}; \boldsymbol{\varphi}_t)\}$ and maximizes the log-likelihood $\ln p(\boldsymbol{y}_{1:t}; \boldsymbol{\varphi}_t)$ with respect to $\boldsymbol{\varphi}_t$, based on the noisy measurements of the $t$-th frame and the messages $\hat{p}(\boldsymbol{s}_t|\boldsymbol{y}_{1:t-1})$ passed from the previous frame. In summary, for every frame, the DD-VBI algorithm performs iterations between the following two major steps until convergence, as shown in Fig. 5.

- DD-VBI-E Step: Given $\boldsymbol{\varphi}_t$ at $t$-th frame and messages $\hat{p}(\boldsymbol{s}_t|\boldsymbol{y}_{1:t-1}; \hat{\boldsymbol{\varphi}}_{1:t-1})$ passed from the previous frame, calculate the approximate marginal posterior of $p(\boldsymbol{v}_t|\boldsymbol{y}_{1:t}; \boldsymbol{\varphi}_t)$, denoted as $q(\boldsymbol{v}_t|\boldsymbol{y}_{1:t}; \boldsymbol{\varphi}_t)$, using the sparse VBI approach, as elaborated in subsection V-D.
- DD-VBI-M Step: Given $q(\boldsymbol{v}_t|\boldsymbol{y}_{1:t}; \boldsymbol{\varphi}_t) \approx p(\boldsymbol{v}_t|\boldsymbol{y}_{1:t}; \boldsymbol{\varphi}_t)$, construct a surrogate function for the objective function $\ln p(\boldsymbol{y}_{1:t}; \boldsymbol{\varphi}_t)$, then maximize the surrogate function with respect to $\boldsymbol{\varphi}_t$ as elaborated in subsection V-C.

After convergence, the messages $\hat{p}(\boldsymbol{s}_{t+1}|\boldsymbol{y}_{1:t}; \hat{\boldsymbol{\varphi}}_{1:t})$ are calculated based on $q(\boldsymbol{s}_t|\boldsymbol{y}_{1:t}; \boldsymbol{\varphi}_{1:t})$ and passed to the next frame. In the following, we first elaborate the M step, which is a variation of the in-exact MM method in [25]. After that, we will elaborate how to approximately calculate the posterior $p(\boldsymbol{v}_t|\boldsymbol{y}_{1:t}; \boldsymbol{\varphi}_t) \approx q(\boldsymbol{v}_t|\boldsymbol{y}_{1:t}; \boldsymbol{\varphi}_t)$ in the E step, which is required to construct the surrogate function in the M step.

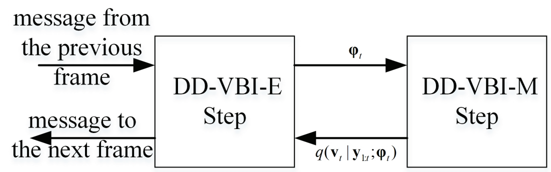


Figure 5: Iinteraction between the two modules of the DD-VBI algorithm within a frame.

### *C. DD-VBI-M Step*

It is difficult to directly maximize the log-likelihood function $\ln p(\boldsymbol{y}_{1:t}; \boldsymbol{\varphi}_t)$, because there is no closed-form expression due to the multi-dimensional integration over $\boldsymbol{v}_t$ as in (16). To make the problem tractable, in the DD-VBI-M Step, we adopt an in-exact MM method in [31], [25], which maximizes a surrogate function of $\ln p(\boldsymbol{y}_{1:t}; \boldsymbol{\varphi}_t)$ with respect to $\boldsymbol{\varphi}_t$, to find an approximate solution of (16). Specifically, let $u(\boldsymbol{\varphi}_t; \dot{\boldsymbol{\varphi}}_t)$ be the surrogate function constructed at some fixed point $\dot{\boldsymbol{\varphi}}_t$, which satisfies the following properties:

$$
\begin{aligned}
u(\boldsymbol{\varphi}_t; \dot{\boldsymbol{\varphi}}_t) &\leq \ln p(\boldsymbol{y}_{1:t}; \boldsymbol{\varphi}_t), \\
u(\dot{\boldsymbol{\varphi}}_t; \dot{\boldsymbol{\varphi}}_t) &= \ln p(\boldsymbol{y}_{1:t}; \dot{\boldsymbol{\varphi}}_t), \\
\frac{\partial u(\boldsymbol{\varphi}_t; \dot{\boldsymbol{\varphi}}_t)}{\partial \boldsymbol{\varphi}_t}|_{\boldsymbol{\varphi}_t = \dot{\boldsymbol{\varphi}}_t} &= \frac{\partial \ln p(\boldsymbol{y}_{1:t}; \boldsymbol{\varphi}_t)}{\partial \boldsymbol{\varphi}_t}|_{\boldsymbol{\varphi}_t = \dot{\boldsymbol{\varphi}}_t}.
\end{aligned} \tag{18}
$$

Inspired by the EM algorithm [31], we use the following surrogate function:

$$
u(\boldsymbol{\varphi}_t; \dot{\boldsymbol{\varphi}}_t) = \int q(\boldsymbol{v}_t|\boldsymbol{y}_{1:t}; \boldsymbol{\varphi}_t) \ln \frac{p(\boldsymbol{y}_{1:t}, \boldsymbol{v}_t; \boldsymbol{\varphi}_t)}{q(\boldsymbol{v}_t|\boldsymbol{y}_{1:t}; \boldsymbol{\varphi}_t)} d\boldsymbol{v}_t, \tag{19}
$$

where $q(\boldsymbol{v}_t|\boldsymbol{y}_{1:t}; \boldsymbol{\varphi}_t)$ is a tractable approximation of $p(\boldsymbol{v}_t|\boldsymbol{y}_{1:t}; \boldsymbol{\varphi}_t)$. In subsection V-A, we have approximated the joint probability distribution in (19) using $\hat{p}(\boldsymbol{y}_{1:t}, \boldsymbol{v}_t; \boldsymbol{\varphi}_t)$. Therefore, the surrogate function can be approximated as

$$
\hat{u}(\boldsymbol{\varphi}_t; \dot{\boldsymbol{\varphi}}_t) = \int q(\boldsymbol{v}_t|\boldsymbol{y}_{1:t}; \boldsymbol{\varphi}_t) \ln \frac{\hat{p}(\boldsymbol{y}_{1:t}, \boldsymbol{v}_t; \boldsymbol{\varphi}_t)}{q(\boldsymbol{v}_t|\boldsymbol{y}_{1:t}; \boldsymbol{\varphi}_t)} d\boldsymbol{v}_t, \tag{20}
$$

When there is no approximation error for the associated PDFs, i.e., $q(\boldsymbol{v}_t|\boldsymbol{y}_{1:t}; \boldsymbol{\varphi}_t) = p(\boldsymbol{v}_t|\boldsymbol{y}_{1:t}; \boldsymbol{\varphi}_t)$ and $\hat{p}(\boldsymbol{y}_{1:t}, \boldsymbol{v}_t; \boldsymbol{\varphi}_t) = p(\boldsymbol{y}_{1:t}, \boldsymbol{v}_t; \boldsymbol{\varphi}_t)$, it can be verified that $\hat{u}(\boldsymbol{\varphi}_t; \dot{\boldsymbol{\varphi}}_t)$ satisfies the properties in (18).

In the M step of the $j$-th iteration, we update $\boldsymbol{\varphi}_t$ as

$$
\boldsymbol{\varphi}_t^{j+1} = \arg\max_{\boldsymbol{\varphi}_t} \hat{u}(\boldsymbol{\varphi}_t; \boldsymbol{\varphi}_t^j), \tag{21}
$$

where $(\cdot)^j$ stands for the $j$-th iteration.

In our problem, $\hat{u}(\boldsymbol{\varphi}_t; \dot{\boldsymbol{\varphi}}_t)$ is a non-convex function and it is difficult to find its optimal solution. Therefore, we use a simple gradient update as in [25], i.e.

$$\varphi_t^{j+1} = \varphi_t^j + \tau^j \frac{\partial \hat{u}(\varphi_t; \dot{\varphi}_t^j)}{\partial \varphi_t}, \tag{22}$$

where $\tau^j$ is the step sizes determined by the Armijo rule [25].

The approximate posterior $q(v_t|y_{1:t};\varphi_t)$ has a factorized form as

$$\begin{aligned} & q(v_t|y_{1:t};\varphi_t) \\ & \quad = q(x_t|y_{1:t};\varphi_t)\, q(\gamma_t|y_{1:t};\varphi_t)\, q(s_t|y_{1:t};\varphi_t). \end{aligned} \tag{23}$$

Therefore, after the convergence of the DD-VBI, we not only obtain an approximate stationary solution $\hat{\varphi}_t$ of (16), but also the associated (approximate) marginal conditional posterior $q(x_t|y_{1:t};\varphi_{1:t}) \approx p(x_t|y_{1:t},\varphi_t)$.

### D. DD-VBI-E Step

DD-VBI-E Step performs the sparse VBI to approximate the conditional marginal posteriors $p(v_t|y_{1:t};\varphi_t)$ based on the following joint prior distribution:

$$\begin{aligned} & \hat{p}(y_{1:t}, v_t; \varphi_t) \\ =& \hat{p}(s_t|y_{1:t-1}) p(y_t|x_t, \kappa_t; \varphi_t) \\ & p(x_t|\gamma_t) p(\gamma_t|s_t) p(\kappa_t). \end{aligned} \tag{24}$$

The corresponding approximate posterior distributions $q(v_t)$ obtained by the sparse VBI will be given by (30)-(37).

*1) Outline of Sparse VBI within a Frame:* For convenience, we use $v_{t,n}$ to denote an individual variable in $v_t$. Let $\mathcal{H} = \{n|\forall v_{t,n} \in v_t\}$. Moreover, we use $q(v_t)$ as a simplified notation for $q(v_t|y_{1:t};\varphi_t)$ when there is no ambiguity. The approximate conditional marginal posterior $q(v_t)$ could be calculated by minimizing the Kullback-Leibler divergence (KLD) between $p(v_t|y_{1:t};\varphi_t)$ and $q(v_t)$ subject to a factorized form constraint on $q(v_t)$ as

$$\mathscr{A}_{\mathrm{VBI}}: \; q^*(v_t) = \arg\min_{q(v_t)} \int q(v_t) \ln \frac{q(v_t)}{p(v_t|y_{1:t};\varphi_t)} dv_t \tag{25}$$

$$\text{s.t.} \;\; q(v_t) = \prod_{n\in\mathcal{H}} q(v_{t,n}), \tag{26}$$

$$\int q(v_{t,n})\, dv_{t,n} = 1, \forall n \in \mathcal{H}. \tag{27}$$

Problem $\mathscr{A}_{\mathrm{VBI}}$ is non-convex, we aim at finding a stationary solution (denoted by $q^*(v_t)$) of $\mathscr{A}_{\mathrm{VBI}}$, as defined below.

**Definition 1.** [Stationary Solution]$q^*(v_t) = \prod_{n\in\mathcal{H}} q^*(v_{t,n})$ is called a stationary solution of Problem $\mathscr{A}_{\mathrm{VBI}}$ if it satisfies all the constraints in $\mathscr{A}_{\mathrm{VBI}}$ and $\forall n \in \mathcal{H}$,

$$\begin{aligned} & q^*(v_{t,n}) = \\ & \arg\min_{q(v_{t,n})} \int \prod_{l\neq n} q^*(v_{t,l})\, q(v_{t,n}) \ln \frac{\prod_{l\neq n} q^*(v_{t,l})\, q(v_{t,n})}{p(v_t|y_{1:t};\varphi_t)} dv_t. \end{aligned}$$

By finding a stationary solution $q^*(v_t)$ of $\mathscr{A}_{\mathrm{VBI}}$, we could obtain the approximate posterior $q^*(v_{t,n}) \approx p(v_{t,n}|y_{1:t};\varphi_t), \forall n \in \mathcal{H}$.

A stationary solution of $\mathscr{A}_{\mathrm{VBI}}$ can be obtained via alternately optimizing each individual density $q(v_{t,n}), n \in \mathcal{H}$. For given $q(v_{t,l}), \forall l \neq n$, the optimal $q(v_{t,n})$ that minimizes the KLD in $\mathscr{A}_{\mathrm{VBI}}$ is given by

$$q(v_{t,n}) \propto \exp\left(\langle \ln p(y_{1:t}, v_t; \varphi_t)\rangle_{\prod_{l\neq n} q(v_{t,l})}\right), \tag{28}$$

where $\langle f(x)\rangle_{q(x)} = \int f(x)\, q(x) dx$. However, $p(y_{1:t}, v_t; \varphi_t)$ is intractable. Since $p(y_{1:t}, v_t; \varphi_t) \approx \hat{p}(y_{1:t}, v_t; \varphi_t)$ in (17), (28) can be approximated as

$$q(v_{t,n}) \propto \exp\left(\langle \ln \hat{p}(y_{1:t}, v_t; \varphi_t)\rangle_{\prod_{l\neq n} q(v_{t,l})}\right). \tag{29}$$

Based on (29), the update equations of all variables are given in the subsequent subsections. The detailed derivation can be found in Appendix B. Note that the operator $\langle\cdot\rangle_{v_{t,l}}$ is equivalent to $\langle\cdot\rangle_{q(v_{t,l})}$ and the expectation $\langle f(v_{t,l})\rangle_{q(v_{t,l})}$ w.r.t. its own approximate posterior is simplified as $\langle f(v_{t,l})\rangle$.

*2) Initialization of Sparse VBI:* In order to trigger the alternating optimization (AO) algorithm, we use the following initializations for the distribution functions $q(x_t), q(\gamma_t)$ in the first iteration of every frame and $q(s_1)$ in the first iteration of first frame. In the rest iterations, we initialize $q(x_t)$,$q(s_t)$,$q(\gamma_t)$ to the (approximate) posterior calculated in the previous frame.

- Initialize $q(s_1) = \hat{p}(s_1|y_1;\varphi_t) = \prod_{n=1}^{\tilde{N}} q(s_{1,n})$ with $q(s_{1,n}) = (\tilde{\pi}_{1,n})^{s_{1,n}} (1-\tilde{\pi}_{1,n})^{1-s_{1,n}}$.
- For given $\hat{p}(s_t|y_{1:t};\varphi_t) = \prod_{n=1}^{\tilde{N}} (\tilde{\pi}_{t,n})^{s_{t,n}} (1-\tilde{\pi}_{t,n})^{1-s_{t,n}}$, initialize a gamma distribution for $\gamma_t$: $q(\gamma_t) = \prod_{n=1}^{\tilde{N}} \Gamma\left(\gamma_{t,n}; \tilde{a}_{\gamma,t,n}, \tilde{b}_{\gamma,t,n}\right)$, where $\tilde{a}_{\gamma,t,n} = \tilde{\pi}_{t,n} a_t + (1-\tilde{\pi}_{t,n})\, \overline{a}_t$, $\tilde{b}_{\gamma,t,n} = \tilde{\pi}_{t,n} b_t + (1-\tilde{\pi}_{t,n})\, \overline{b}_t$.
- Initialize a Gaussian distribution for $x_t$: $q(x_t) = \mathcal{CN}(x_t; \mu_t, \Sigma_t)$, where $\Sigma_t = \left(\mathrm{diag}(\langle\gamma_t\rangle) + (F_t)^H F_t\right)^{-1}$, $\mu_t = \Sigma_t (F_t)^H y_t$.

*3) Update for $q(\kappa_t)$:* From (29), $q(\kappa_t)$ can be derived as

$$q(\kappa_t) = \Gamma(\kappa_t; \tilde{a}_{\kappa,t}, \tilde{b}_{\kappa,t}). \tag{30}$$

where $\tilde{a}_{\kappa,t} = a_\kappa + NN_p$, $\tilde{b}_{\kappa,t} = b_\kappa + \left\langle \|y_t - F_t x_t\|^2 \right\rangle_{x_t} = b_\kappa + \|y_t - F_t \mu_t\|^2 + \mathrm{tr}\left(F_t \Sigma_t (F_t)^H\right)$.

*4) Update for $q(x_t)$:* $q(x_t)$ can be derived as

$$q(x_t) = \mathcal{CN}(x_t; \mu_t, \Sigma_t). \tag{31}$$

$\mu_t$ and $\Sigma_t$ can be calculated through

$$\begin{aligned} \Sigma_t &= \left(\mathrm{diag}(\langle\gamma_t\rangle) + \langle\kappa_t\rangle (F_t)^H F_t\right)^{-1}, \\ &= \mathbf{R} - \upsilon \mathbf{R} (F_t)^H \left(\mathbf{I} + \upsilon F_t \mathbf{R} (F_t)^H\right)^{-1} F_t \mathbf{R}. \end{aligned} \tag{32}$$

$$\mu_t = \langle\kappa_t\rangle \Sigma_t (F_t)^H y_t. \tag{33}$$

where $\langle \gamma_t \rangle = \frac{\tilde{a}_{\gamma,t,n}}{\tilde{b}_{\gamma,t,n}}$, $\langle \kappa_t \rangle = \frac{\tilde{a}_{\kappa,t}}{\tilde{b}_{\kappa,t}}$, $\mathbf{R} = \mathrm{diag}\left(\left[\frac{\tilde{b}_{\gamma,t,1}}{\tilde{a}_{\gamma,t,1}}, \cdots, \frac{\tilde{b}_{\gamma,t,\tilde{N}}}{\tilde{a}_{\gamma,t,\tilde{N}}}\right]\right)$, $\upsilon = \langle \kappa_t \rangle = \frac{\tilde{a}_{\kappa,t}}{\tilde{b}_{\kappa,t}}$.

*5) Update for $q(\boldsymbol{\gamma}_t)$:* $q(\boldsymbol{\gamma}_t)$ can be derived as

$$q(\boldsymbol{\gamma}_t) = \prod_{n=1}^{\tilde{N}} \Gamma\left(\gamma_{t,n}; \tilde{a}_{\gamma,t,n}, \tilde{b}_{\gamma,t,n}\right), \tag{34}$$

where $\tilde{a}_{\gamma,t,n}, \tilde{b}_{\gamma,t,n}$ are given by:

$$\tilde{a}_{\gamma,t,n} = \langle s_{t,n} \rangle a_t + \langle 1 - s_{t,n} \rangle \overline{a}_t + 1, \tag{35}$$

$$\tilde{b}_{\gamma,t,n} = \langle s_{t,n} \rangle b_t + \langle 1 - s_{t,n} \rangle \overline{b}_t + \left\langle |x_{t,n}|^2 \right\rangle. \tag{36}$$

where $\langle s_{t,n} \rangle = \tilde{\pi}_{t,n}, \langle 1 - s_{t,n} \rangle = 1 - \tilde{\pi}_{t,q}$, $\left\langle |x_{t,n}|^2 \right\rangle = |\mu_{t,n}|^2 + \Sigma_{t,n}$, $\mu_{t,n}$ is the $n$-th element of $\boldsymbol{\mu}_t$, $\Sigma_{t,n}$ is the $n$-th diagonal element of $\boldsymbol{\Sigma}_t$.

*6) Update for $q(\boldsymbol{s}_t)$:* $q(\boldsymbol{s}_t)$ can be derived as

$$q(\boldsymbol{s}_t) = \prod_{n=1}^{\tilde{N}} (\pi_{t,n})^{s_{t.n}} (1 - \pi_{t,n})^{1 - s_{t,n}}, \tag{37}$$

where $\pi_{t,n}$ is given by

$$\pi_{t,n} = \frac{1}{C} \frac{\tilde{\pi}_{t,n} b_t^{a_t}}{\Gamma(a_t)} e^{(a_t - 1)\langle \ln \gamma_{t,n} \rangle - b_t \langle \gamma_{t,n} \rangle}, \tag{38}$$

and $C$ is the normalization constant, given by $C = \frac{\tilde{\pi}_{t,n} b_t^{a_t}}{\Gamma(a_t)} e^{(a_t-1)\langle \ln \gamma_{t,n} \rangle - b_t \langle \gamma_{t,n} \rangle} + \frac{(1-\tilde{\pi}_{t,n})\overline{b}_t^{\overline{a}_t}}{\Gamma(\overline{a}_t)} e^{(\overline{a}_t - 1)\langle \ln \gamma_{t,n} \rangle - \overline{b}_t \langle \gamma_{t,n} \rangle}$, $\langle \ln \gamma_{t,n} \rangle = \psi(\tilde{a}_{\gamma,t,n}) - \ln\left(\tilde{b}_{\gamma,t,n}\right)$, $\psi(x) = \frac{d}{dx} \ln(\Gamma(x))$ is the digamma function, defined as the logarithmic derivative of the gamma function.

### *E. Messages $\hat{p}(\boldsymbol{s}_{t+1} | \boldsymbol{y}_{1:t}; \hat{\boldsymbol{\varphi}}_{1:t})$ Passed to the Next Frame*

After the convergence of the DD-VBI iterations in frame $t$, let $\hat{f}_{d,t}$ and $q(\boldsymbol{s}_t | \boldsymbol{y}_{1:t}, ; \hat{\boldsymbol{\varphi}}_{1:t}) = \prod_{n=1}^{\tilde{N}} (\pi_{t,n})^{s_{t.n}} (1 - \pi_{t,n})^{1 - s_{t,n}}$ denote the converged Doppler parameter and the associated approximate posterior for $\boldsymbol{s}_t$, respectively. Then, the messages $\hat{p}(\boldsymbol{s}_{t+1} | \boldsymbol{y}_{1:t}; ; \hat{\boldsymbol{\varphi}}_{1:t})$ is calculated as

$$\begin{aligned} &\hat{p}(\boldsymbol{s}_{t+1} | \boldsymbol{y}_{1:t}; \hat{\boldsymbol{\varphi}}_{1:t}) \\ &= \prod_{n=1}^{\tilde{N}} \sum_{\mathbf{s}_{t,n}} q(s_{t,n} | \boldsymbol{y}_{1:t}, \hat{\boldsymbol{\varphi}}_{1:t}) p(s_{t+1,n} | s_{t,n}) \qquad (39) \\ &= \prod_{n=1}^{\tilde{N}} (\tilde{\pi}_{t+1,n})^{s_{t+1,n}} (1 - \tilde{\pi}_{t+1,n})^{1 - s_{t+1,n}}. \qquad (40) \end{aligned}$$

where

$$\tilde{\pi}_{t+1,n} = (1 - \pi_{t,n}) \rho_{0,1} + \pi_{t,n}(1 - \rho_{1,0}).$$

**Algorithm 1** Doppler-Aware-Dynamic-VBI algorithm

**for** $t = 1, 2, ...$ **do**
  Initialize the distribution functions according to Section V-D2.
  **while** not converge **do**
  **while** not converge **do**
    **%DD-VBI-E Step:**
    Update $q(\kappa_t)$ using (30).
    Update $q(\boldsymbol{x}_t)$ using (31).
    Update $q(\boldsymbol{\gamma}_t)$ using (34).
    Update $q(\boldsymbol{s}_t)$ using (37) .
  **end while**
    **%DD-VBI-M Step:**
    Construct the surrogate function $\hat{u}$ in (20) using the output of DD-VBI-E Step $q(\boldsymbol{v}_t | \boldsymbol{y}_{1:t}; \boldsymbol{\varphi}_t)$.
    Update $\boldsymbol{\varphi}_t$ using (21).
  **end while**
  Let $\hat{f}_{d,t}$ denote the converged maximum DFO for frame $t$. Calculate $\hat{p}(\boldsymbol{s}_{t+1} | \boldsymbol{y}_{1:t}; \hat{\boldsymbol{\varphi}}_{1:t})$ using (40).
  Pass messages $\hat{p}(\boldsymbol{s}_{t+1} | \boldsymbol{y}_{1:t}; \hat{\boldsymbol{\varphi}}_{1:t})$ to frame $t+1$.
  Estimate $x_{t,n}$ using (33).
**end for**

Finally, the messages $\hat{p}(\boldsymbol{s}_{t+1} | \boldsymbol{y}_{1:t}; \hat{\boldsymbol{\varphi}}_{1:t})$, as the prior to the channel support, is passed to the next frame.

The overall DD-VBI algorithm is summarized in Algorithm 1. Note that in the $t$-th frame of DD-VBI, the contribution of the previous observations $\boldsymbol{y}_{1:t-1}$ on the estimation of $\boldsymbol{v}_t$ and $\boldsymbol{\varphi}_t$ is summarized in the messages $\hat{p}(\boldsymbol{s}_t | \boldsymbol{y}_{1:t-1}; \hat{\boldsymbol{\varphi}}_{1:t-1})$ passed from frame $t-1$. Therefore, in the $t$-th frame, there is no need to store all the observations $\boldsymbol{y}_{1:t-1}$ up to frame $t-1$.

*Remark* 2. In practical mmWave massive MIMO systems, the number of RF chains can be less than the number of antennas at both the BS and user sides to reduce the hardware cost and power consumption. The proposed scheme can be easily extended to the case with limited RF chains. For example, suppose there are only $M_b < M$ RF chains at the BS and $N_b < N$ RF chains at the user. In this case, when the BS transmits the training vector $\mathbf{v} = \mathbf{gF} \in \mathbb{C}^M$ for downlink channel tracking in the $i$-th symbol duration, the user employs $\mathbf{U}_i = \mathbf{W}_i \mathbf{G}_i \in \mathbb{C}^{N \times N_b}$ as a combining matrix to combine the received signal into $N_b$ baseband channel measurements, where $\mathbf{F} \in \mathbb{C}^{M \times M_b}$ and $\mathbf{g} \in \mathbb{C}^{M_b}$ are the RF training matrix and baseband training vector at the BS, respectively, and $\mathbf{W}_i \in \mathbb{C}^{N \times N_b}$ and $\mathbf{G}_i \in \mathbb{C}^{N_b \times N_b}$ are the RF and baseband combining matrix at the mobile user in the $i$-th symbol duration, respectively. We can still write the received pilot signal as a CS model as in (14), but with $\boldsymbol{F}_t = [\boldsymbol{F}_{t,1}; ...; \boldsymbol{F}_{t,N_p}] \in C^{N_b N_p \times \tilde{N}}$, $\boldsymbol{F}_{t,i} = \boldsymbol{U}_{t,i}^H \boldsymbol{A}_{R,i}(\boldsymbol{\varphi}_t)$, $\boldsymbol{n}_t = [\boldsymbol{U}_{t,i}^H \boldsymbol{n}_{t,i}]_{i \in \mathcal{N}_p}$.

### *F. Complexity and Signaling Overhead Comparison*

The computational complexity of the proposed algorithm is dominated by the update of $q(\boldsymbol{x}_t)$. Assuming the arithmetic with individual elements has complexity $\mathcal{O}(1)$, the computational complexity of matrix inversion in (32) is $\mathcal{O}(N_b^3 N_p^3)$

| Algorithms | Complexity order | Typical complexity order |
| --- | --- | --- |
| proposed | $\mathcal{O}\left(I(N_bN_p)\tilde{N}^2\right)$ | $\mathcal{O}\left(IL_D^2N\right)-\mathcal{O}\left(IN^3\right)$ |
| ML | $\mathcal{O}\left(\tilde{N}^3\right)$ | $\mathcal{O}\left(N^3\right)$ |
| ES | $\mathcal{O}\left(L_D^2K_E^4\right)$ | $\mathcal{O}\left(L_D^2N^4\right)$ |
| HS | $\mathcal{O}\left(SL_D^4K_H^2\right)$ | $\mathcal{O}\left(\log_{K_H}(N/L_D)L_D^4K_H^2\right)$ |

Table II: Complexity orders for different schemes.

| Algorithms | Signaling overhead | Typical signaling overhead |
| --- | --- | --- |
| proposed | $\mathcal{O}(L_D/N_b)+L_D$ | $4+L_D$ |
| ML | $N$ | $N$ |
| HS | $K_H^2\left(L_D\right)^3S$ | $K_H^2\left(L_D\right)^3\log_{K_H}(N/L_D)$ |
| ES | $K_E^2$ | $N^2$ |

Table III: Signaling overhead for different schemes.

and the total number of multiplications to update $q(\boldsymbol{x}_t)$ is $3\tilde{N}+2N_bN_p\tilde{N}^2+2(N_bN_p)^2\tilde{N}+(N_bN_p)^2$. Supposing the algorithm executes $I$ iterations, the total complexity order of the proposed method is $\mathcal{O}\left(I(N_bN_p)^2\tilde{N}\right)$, considering that $\tilde{N}$ is usually larger than $N_bN_p$. In Table II and III, we assume the mmWave channel has $L_D$ dominant paths and compare the complexity and signaling overhead of the proposed algorithm with the following baseline algorithms:

- **Baseline 1** (ML) [8]: This is the ML based joint DFO and channel estimation algorithm proposed in [8].
- **Baseline 2** (ES) [20]: This is the codebook based exhaustive beam search algorithm in [20]. $K_E$ denotes the number of beamforming vectors of the codebook in ES.
- **Baseline 3** (HS) [21]: This is the codebook based hierarchical beam search algorithm in [21]. $S$ denotes the total level of hierarchical codebook. $K_H$ denotes the number of beamforming vectors of each codebook level in HS.

As seen from Table II, the complexity order of the proposed algorithm is similar to the baselines. For example, $N_bN_p$ can range from $\mathcal{O}\left(L_D\right)$ to $\mathcal{O}\left(N\right)$ to achieve different tradeoff between performance and complexity, and we usually have $\tilde{N}=\mathcal{O}\left(N\right)$, $K_E=\mathcal{O}\left(N\right)$. For a resolution $\mathcal{O}(2\pi/N)$ of the HS scheme, we usually have $\mathcal{S}=\mathcal{O}\left(\log_{K_H}(N/L_D)\right)$ [21]. In this typical case, the complexity of the proposed scheme and the baseline schemes are shown in the third column of Table II. On the other hand, the proposed algorithm has the lowest signaling overhead for both the general case and the typical case, as shown in Table III.

## VI. Simulation Results

In this section, we compare the performance of the proposed algorithm with the baseline algorithms described in Section V-F. For the ML baseline, we also consider the case when only partial channel parameters are estimated as in the proposed scheme (ML-Partial). For the proposed scheme, we also consider the case when the training vector is generated randomly (DD-VBI-Random). The channel parameters are based on the millimeter-wave statistical spatial channel model (mm-SSCM) as specified in [32], which was developed according to the 28- and 73-GHz ultrawideband propagation measurements in New York City. The signal bandwidth is $50MHz$ and the frame duration is set as $T_b=0.5ms$. The carrier frequency is 28GHz . The mobile user employs a ULA of $M=128$ antennas and the inter antenna spacing is $\lambda/2$, and the BS also employs a ULA. The user velocity is assumed to be 380km/h, which translates to $f_{d,t}\approx 10$KHz.

In the simulation, we will consider both cases when the user is equipped with a full set of RF chains and limited RF chains. For the case with limited RF chains, the number of RF chains is set to be $N_b=16$. The MSE for DFO estimation and the uplink achievable data rate are adopted as the performance metrics. The frequency MSE and the channel MSE is defined as $\frac{\left\|\hat{f}_{d,t}-f_{d,t}\right\|^2}{\left\|f_{d,t}\right\|^2}$ and $\frac{\left\|\hat{\boldsymbol{x}}_t-\boldsymbol{x}_t\right\|^2}{\left\|\boldsymbol{x}_t\right\|^2}$, respectively. The parameter $N_d$ is chosen to be equal to the number of dominant AoA directions at the user side (an AoA direction is called a dominant AoA direction if its energy is no less than 10% of the most significant AoA direction) and $N_d$ data streams are transmitted over the $N_d$ dominant AoA directions in the uplink with equal power allocation. Since the uplink transmission is only designed based on the dominant AoA directions which are known at the user, there is no need for the BS to feed back the slow time-varying effective channel.

### A. Doppler Frequency and Channel MSE Performance

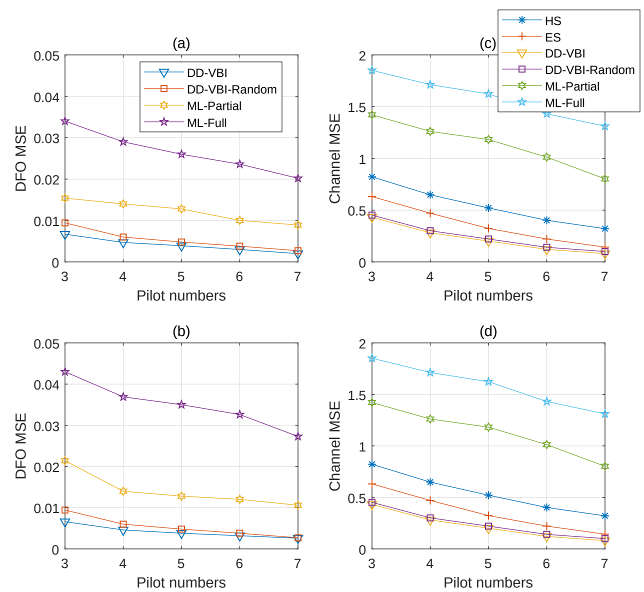


Figure 6: Frequency MSE and channel MSE performance versus the pilot number. Set SNR= 0 dB. (a) full RF chains with $N=128$. (b) limited RF chains with $N=128$, $N_b=16$.

The Doppler frequency and channel MSE performance of different algorithms versus the pilot number, SNR and number of BS antennas are shown in Fig. 6 and Fig. 7. It can be seen that the proposed DD-VBI algorithm achieves large performance gain over all the baseline algorithms, under both full RF chains and limited RF chains. By using the proposed training vector design to strike a balance between *exploitation* of known channel directions and *exploration* of unknown channel directions, the DD-VBI algorithm could further improve the MSE performance compared to the case with a random training vector. This demonstrates that the proposed algorithm can effectively estimate the maximum DFO by selective channel tracking and efficient training vector design. Note that as the number of BS antennas increases, the parameters to be estimated increase significantly and the spatial resolution and array gain will also increase. Since the ML-Full method does not exploit the selective channel estimation method or channel sparsity to reduce the number of free parameters, its performance may degrade with the number of BS antennas. However, the performance gain of the proposed algorithm improves because in this case, the number of parameters to be estimated does not increase with the number of BS antennas. This demonstrates that the proposed algorithm is a powerful method for accurate DFO estimation, even when both BS and mobile user are equipped with massive MIMO, and the number of RF chains at the user side is limited.

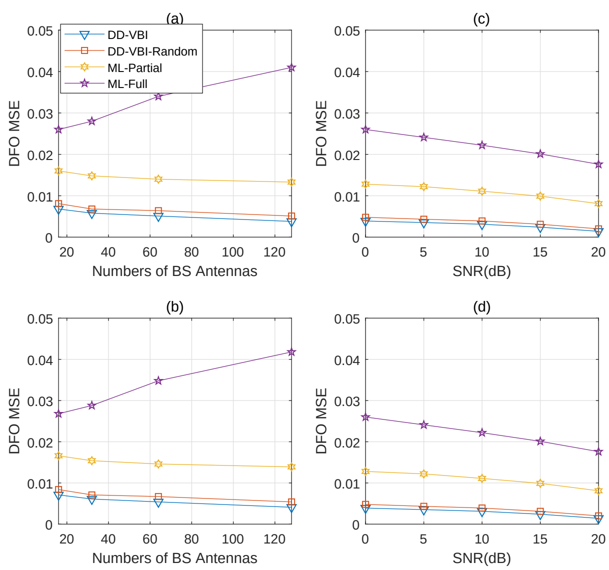


Figure 7: Frequency MSE performance versus the number of BS antennas and SNR. (a) Set SNR= 0 dB. Full RF chains with $N = 128$. (b) Set SNR= 0 dB. Limited RF chains with $N = 128$, $N_b = 16$. (c) The number of pilots is fixed as 5. Full RF chains with $N = 128$. (d) The number of pilots is fixed as 5. Limited RF chains with $N = 128$, $N_b = 16$.

### B. Achievable Data Rate Performance

The achievable data rates of different algorithms versus the pilot number, SNR and number of BS antennas are shown in Fig. 8. It can be observed that the performance of all algorithms increases with the number of pilots, SNR and number of BS antennas. The proposed DD-VBI algorithm can achieve large performance gain over various baselines under

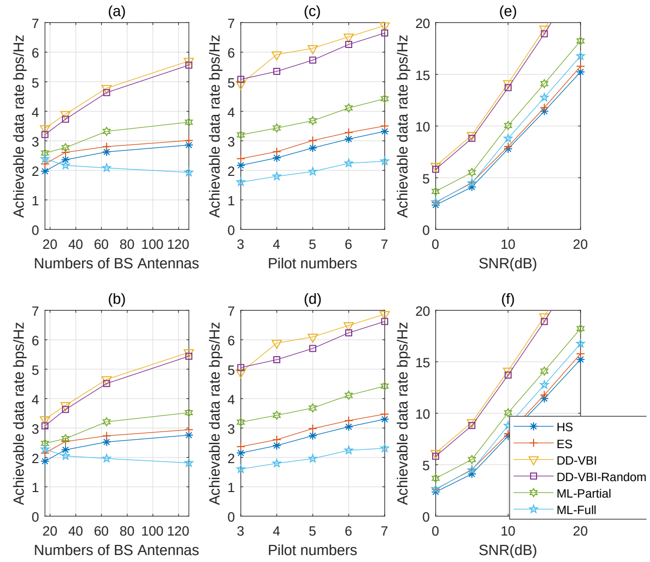


Figure 8: Achievable data rate versus the number of BS antennas, pilot number and SNR. (a) and (c) Set SNR= 0 dB. Full RF chains with $N = 128$. (b) and (d) Set SNR= 0 dB. Limited RF chains with $N = 128$, $N_b = 16$. (e) The number of pilots is fixed as 5. Full RF chains with $N = 128$. (f) The number of pilots is fixed as 5. Limited RF chains with $N = 128$, $N_b = 16$.

both full RF chains and limited RF chains. Moreover, by using the proposed training vector design, the proposed DD-VBI algorithm could further improve the achievable data rate. This verifies that the proposed selective channel tracking and Doppler compensation scheme can also enhance the achievable data rate with low pilot overhead, under different SNRs and numbers of BS antennas.

### C. Complexity versus Realized Gain

In practice, we can control the tradeoff between the complexity and realized gain of the proposed algorithm by adjusting the number of iterations. In Fig. 9, we plot the achievable rate versus the CPU time. It can be seen that the proposed algorithm can achieve a better performance than the baseline algorithms for the same CPU time. Moreover, the performance gain increases with the CPU time. Therefore, the proposed algorithm provides a better and more flexible tradeoff between the performance and computational power.

## VII. Conclusion

We propose an angular-domain selective channel tracking and Doppler compensation scheme for high-mobility massive MIMO systems. Firstly, we propose a selective channel tracking scheme and the associated Doppler-aware-dynamic-VBI algorithm to accurately estimate the DFO and partial angular–domain channel parameters with reduced pilot overhead. Then, we propose an angular-domain selective DFO compensation

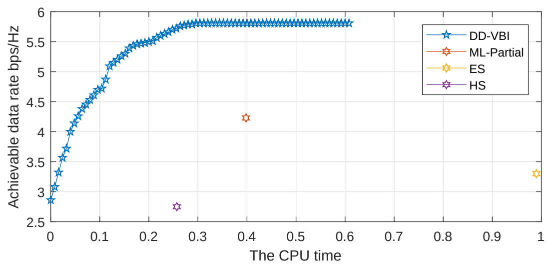


Figure 9: The tradeoff between the complexity and realized gain of the proposed algorithm.

scheme to convert the dominant paths of the fast time-varying channel into a slow time-varying effective channel, based on which efficient uplink and downlink transmissions can be achieved. Simulations verify that the proposed scheme not only can mitigate the Doppler and channel aging effect with much less pilots than existing schemes, but also can achieve a good tradeoff between the CSI signaling overhead and spatial multiplexing/array gain.

## Appendix

### A. EM Estimation of the Parameters $\rho_{1,0}$ and $\rho_{0,1}$

Denote the $\boldsymbol{s}_t$ as the channel support vector, as defined in SubSec.IV-A. We apply the EM method in [31] to obtain an estimate of the parameter $\rho_{1,0}$ by perform the following iterations until convergence:

$$\rho_{1,0}^{i+1} = \arg\max_{\rho_{1,0}} E_{\hat{p}^i(\boldsymbol{s}_{1:t})}\left\{\ln p(\boldsymbol{s}_{1:t}, \boldsymbol{y}_{1:t}; \rho_{1,0}) | \rho_{1,0}^i\right\}, \quad (41)$$

where $\rho_{1,0}^i$ stands for the $i$-th iteration, $E_{\hat{p}^i(\boldsymbol{s}_{1:t})}\{\cdot\}$ denotes expectation over the posterior distribution $\hat{p}^i(\boldsymbol{s}_{1:t}) = p(\boldsymbol{s}_{1:t}|\boldsymbol{y}_{1:t}, \rho_{1,0}^i)$ conditioned on $\boldsymbol{y}_{1:t}$ and $\rho_{1,0}^i$. Solving the problem (41), we can get a closed-form expression for updating the $\rho_{1,0}$ as

$$\rho_{1,0}^{i+1} = \frac{\sum_{\tilde{t}=2}^{t}\sum_{n=1}^{\tilde{N}} E_{\hat{p}^i(\boldsymbol{s}_{1:t})}\left[s_{\tilde{t},n}\right] - E_{\hat{p}^i(\boldsymbol{s}_{1:t})}\left[s_{\tilde{t}-1,n}s_{\tilde{t},n}\right]}{\sum_{\tilde{t}=2}^{t}\sum_{n=1}^{\tilde{N}} E_{\hat{p}^i(\boldsymbol{s}_{1:t})}\left[s_{\tilde{t}-1,n}\right]} \quad (42)$$

Since the channel support vector $\boldsymbol{s}_{1:t}$ forms a Markov chain, the posterior distribution $\hat{p}^i(\boldsymbol{s}_{1:t})$ can be approximately calculated using the sum-product message passing over the factor graph of the associated Markov chain with the priors of $s_{\tilde{t},n}$'s given by the output of the DD-VBI algorithm at each frame $\tilde{t}$ [30]. After the iterations in (42) convergence, we can obtain the EM estimation of the parameter $\rho_{1,0}$ in the t-th frame. The EM estimation of the parameter $\rho_{0,1}$ can be obtained in a similar way.

### B. Derivation of (30)-(37)

From (29), $q(\kappa_t)$ in (30) can be obtained a

$$\begin{aligned}
&\ln q(\kappa_t) \propto \langle \ln p(\boldsymbol{y}_t|\boldsymbol{x}_t, \kappa_t; \boldsymbol{\varphi}_t)\rangle_{\boldsymbol{x}_t} + \ln p(\kappa_t) \\
&\propto -\kappa_t \left\langle \left\|\boldsymbol{y}_t - \boldsymbol{F}_t\boldsymbol{x}_t\right\|^2 \right\rangle_{\boldsymbol{x}_t} \\
&\quad + NN_p \ln \kappa_{t,i} + (a_\kappa - 1)\ln \kappa_t - b_\kappa \kappa_t \\
&\propto (\tilde{a}_\kappa - 1)\ln \kappa - \tilde{b}_\kappa \kappa.
\end{aligned}$$

$q(\boldsymbol{x}_t)$ in (31) can be obtained as

$$\begin{aligned}
&\ln q(\boldsymbol{x}_t) \propto \langle \ln p(\boldsymbol{y}_t|\boldsymbol{x}_t, \kappa_t; \boldsymbol{\varphi}_t)\rangle_{\kappa_t} + \langle \ln p(\boldsymbol{x}_t|\boldsymbol{\gamma}_t)\rangle_{\boldsymbol{\gamma}_t} \\
&\propto -\langle \kappa_t\rangle \left\|\boldsymbol{y}_t - \boldsymbol{F}_t\boldsymbol{x}_t\right\|^2 - \boldsymbol{x}_t^H \mathrm{diag}(\langle \boldsymbol{\gamma}_t\rangle)\boldsymbol{x}_t \\
&\propto -(\boldsymbol{x}_t - \boldsymbol{\mu}_t)^H (\boldsymbol{\Sigma}_t)^{-1}(\boldsymbol{x}_t - \boldsymbol{\mu}_t).
\end{aligned}$$

$q(\boldsymbol{\gamma}_t)$ in (34) can be obtained as

$$\begin{aligned}
&\ln q(\boldsymbol{\gamma}_t) \propto \langle \ln p(\boldsymbol{x}_t|\boldsymbol{\gamma}_t)\rangle_{\boldsymbol{x}_t} + \langle \ln p(\boldsymbol{\gamma}_t|\boldsymbol{s}_t)\rangle_{\boldsymbol{s}_t} \\
&\propto \sum_n (\tilde{a}_{\gamma,t,n} - 1)\ln \boldsymbol{\gamma}_{t,n} - \tilde{b}_{\gamma,t,n}\boldsymbol{\gamma}_{t,n}.
\end{aligned}$$

$q(\boldsymbol{s}_t)$ in (37) can be obtained as

$$\begin{aligned}
&\propto \ln q(\boldsymbol{s}_t) \propto \langle \ln p(\boldsymbol{\gamma}_t|\boldsymbol{s}_t)\rangle_{\boldsymbol{\gamma}_t} + \ln \hat{p}(\boldsymbol{s}_t) \\
&\propto \sum_n s_t \left(\ln \frac{b_t^{a_t}}{\Gamma(a_t)} + (a_t - 1)\left\langle \ln \boldsymbol{\gamma}_{t,n}\right\rangle\right. \\
&\quad \left. - b_t \left\langle \boldsymbol{\gamma}_{t,n}\right\rangle + \ln \tilde{\pi}_{t,n}\right) + (1 - s_t)\left(\ln \frac{\bar{b}_t^{\bar{a}_t}}{\Gamma(\bar{a}_t)} + (\bar{a}_t - 1)\right. \\
&\quad \left. \times \left\langle \ln \boldsymbol{\gamma}_{t,n}\right\rangle - \bar{b}_t \left\langle \boldsymbol{\gamma}_{t,n}\right\rangle + \ln(1 - \tilde{\pi}_{t,n})\right) \\
&\propto \ln \prod_{n=1}^{\tilde{N}} (\pi_{t,n})^{s_{t,n}} (1 - \pi_{t,n})^{1 - s_{t,n}}.
\end{aligned}$$

## References


[1] M. Souden, S. Affes, J. Benesty, and R. Bahroun, "Robust Doppler Spread Estimation in the Presence of a Residual Carrier Frequency Offset," *IEEE Transactions on Signal Processing*, vol. 57, no. 10, pp. 4148–4153, 2009.

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

[3] X. Wang, G. Wang, R. Fan, and B. Ai, "Channel Estimation with Expectation Maximization and Historical Information based Basis Expansion Model for Wireless Communication Systems on High Speed Railways," *IEEE Access*, vol. 6, no. 99, pp. 72–80, 2018.

[4] Y. Mostofi and D. C. Cox, "ICI Mitigation for Pilot-aided OFDM Mobile Systems," *IEEE Transactions on Wireless Communications*, vol. 4, no. 2, pp. 765–774, 2005.

[5] T. L. Liu, W. H. Chung, S. Y. Yuan, and S. Y. Kuo, "On the Banded Approximation of the Channel Matrix for Mobile OFDM Systems," *Vehicular Technology IEEE Transactions on*, vol. 64, no. 8, pp. 3526–3535, 2015.

[6] W. Xin, X. X. Yang, and Y. Z. Wang, "Study on Channel Estimation in Time-Varing Amplify and Forward Relay Networks," *Applied Mechanics Materials*, vol. 519-520, pp. 934–938, 2014.

[7] T. Y. Al-Naffouri, K. M. Z. Islam, N. Al-Dhahir, and S. Lu, "A Model Reduction Approach for OFDM Channel Estimation Under High Mobility Conditions," *IEEE Transactions on Signal Processing*, vol. 58, no. 4, pp. 2181–2193, 2010.

[8] Y. Ge, W. Zhang, F. Gao, and H. Minn, "Angle-Domain Approach for Parameter Estimation in High-Mobility OFDM with Fully/Partly Calibrated Massive ULA," *IEEE Transactions on Wireless Communications*, pp. 1–1, 2018.

[9] A. Monk, R. Hadani, M. Tsatsanis, and S. Rakib, "OTFS-Orthogonal Time Frequency Space," 2016.

[10] M. K. Ramachandran and A. Chockalingam, "MIMO-OTFS in High-Doppler Fading Channels: Signal Detection and Channel Estimation," 2018.

[11] F. Bellili and S. Affes, "A Low-cost and Robust Maximum Likelihood Doppler Spread Estimator," in *Global Communications Conference*, 2014, pp. 4325–4330.

[12] D. Chizhik, "Slowing the Time-fluctuating MIMO Channel," *IEEE Transactions on Wireless Communications*, vol. 3, no. 5, pp. 1554–1565, 2004.

[13] Y. Zhang, Q. Yin, P. Mu, and L. Bai, "Multiple Doppler Shifts Compensation and ICI Elimination by Beamforming in High-mobility OFDM systems," in *International ICST Conference on Communications and NETWORKING in China*, 2012, pp. 170–175.

[14] W. Guo, P. Mu, Q. Yin, and H. M. Wang, "Multiple Doppler frequency offsets compensation technique for high-mobility OFDM uplink," in *IEEE International Conference on Signal Processing, Communication and Computing*, 2013, pp. 1–5.

[15] Z. Liu, J. Wu, and P. Fan, "On the Uplink Capacity of High Speed Railway Communications with Massive MIMO Systems," in *Vehicular Technology Conference*, 2014, pp. 1–5.

[16] X. Chen, J. Lu, T. Li, P. Fan, and K. B. Letaief, "Directivity-Beamwidth Tradeoff of Massive MIMO Uplink Beamforming for High Speed Train Communication," *IEEE Access*, vol. PP, no. 99, pp. 1–1, 2017.

[17] W. Guo, W. Zhang, P. Mu, and F. Gao, "High-Mobility OFDM Downlink Transmission with Large-Scale Antenna Array," *IEEE Transactions on Vehicular Technology*, vol. PP, no. 99, pp. 1–1, 2017.

[18] W. Guo, W. Zhang, P. Mu, F. Gao, and H. Lin, "High-Mobility Wideband Massive MIMO Communications: Doppler Compensation, Analysis and Scaling Law," 2018.

[19] W. Guo, W. Zhang, P. Mu, F. Gao, and B. Yao, "Angle-Domain Doppler Pre-Compensation for High-Mobility OFDM Uplink with Massive ULA," in *GLOBECOM 2017 - 2017 IEEE Global Communications Conference*, 2018, pp. 1–6.

[20] J. Wang, L. Zhou, C. W. Pyo, T. Baykas, and S. Kato, "Beam Codebook Based Beamforming Protocol for Multi-Gbps Millimeter-Wave WPAN Systems," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 8, pp. 1390–1399, 2009.

[21] A. Alkhateeb, O. E. Ayach, G. Leus, and R. W. H. Jr, "Channel Estimation and Hybrid Precoding for Millimeter Wave Cellular Systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 831–846, 2017.

[22] Z. Gao, L. Dai, H. Chen, and Z. Wang, "Channel Estimation for Millimeter-Wave Massive MIMO with Hybrid Precoding over Frequency-Selective Fading Channels," *IEEE Communications Letters*, vol. 20, no. 6, pp. 1259–1262, 2016.

[23] W. U. Bajwa, J. Haupt, A. M. Sayeed, and R. Nowak, "Compressed Channel Sensing: A New Approach to Estimating Sparse Multipath Channels," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1058–1076, 2010.

[24] D. G. Tzikas, C. L. Likas, and N. P. Galatsanos, "The Variational Approximation for Bayesian Inference," *IEEE Signal Processing Magazine*, vol. 25, no. 6, pp. 131–146, 2008.

[25] J. Dai, A. Liu, and V. K. N. Lau, "Fdd massive mimo channel estimation with arbitrary 2d-array geometry," *IEEE Transactions on Signal Processing*, vol. PP, no. 99, 2018.

[26] C. B. Dietrich, "Adaptive Arrays and Diversity Antenna Configurations for Handheld Wireless Communication Terminals," *Dissertation Abstracts International, Volume: 63-08, Section: B, page: 3841.;Chair: Warren L. Stutz*, 2000.

[27] A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire, "Joint Spatial Division and Multiplexing - The Large-Scale Array Regime," vol. 59, no. 10, pp. 6441–6463, Oct. 2013.

[28] S. Ji, Y. Xue, and L. Carin, "Bayesian Compressive Sensing," *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2346–2356, 2008.

[29] D. P. Wipf and B. D. Rao, "Bayesian Learning for Sparse Signal Reconstruction," in *IEEE International Conference on Acoustics*, 2003.

[30] J. Ziniel and P. Schniter, "Dynamic Compressive Sensing of Time-Varying Signals via Approximate Message Passing," *IEEE Transactions on Signal Processing*, vol. 61, no. 21, pp. 5270–5284, 2013.

[31] A. P. Dempster, "Maximum Likelihood Estimation from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society*, vol. 39, no. 1, pp. 1–38, 1977.

[32] M. K. Samimi and T. S. Rappaport, "3-D Millimeter-Wave Statistical Channel Model for 5G Wireless System Design," *IEEE Transactions on Microwave Theory Techniques*, vol. 64, no. 7, pp. 1–19, 2016.