

A Novel Beam Alignment Scheme for Mobile Millimeter-Wave Communications Based on Compressed Sensing Aided-Kalman Filter

KUN-HSIEN LIN¹ AND KUANG-HAO LIU²  (Member, IEEE)

¹Wireless Communication Department, MediaTek Inc., Hsinchu 30078, Taiwan

²Institute of Communications Engineering, National Tsing Hua University, Hsinchu 300044, Taiwan

CORRESPONDING AUTHOR: K.-H. Liu (e-mail: khliu@ee.nthu.edu.tw)

This work was supported in part by the Ministry of Science and Technology, Taiwan, under Grant MOST 111-2221-E-007-076, and in part by the Qualcomm through the Taiwan University Research Collaboration Project.

ABSTRACT In millimeter wave (mmWave) communications, fast and reliable beam alignment is crucial to the achievable data rate. The problem is even more challenging when either the transmitter or the receiver, or both move that need timely update of the best beam pair. Some existing work tracks the channel changes by Kalman filter (KF). Alternatively, compressed sensing (CS) approaches are useful to reconstruct the channel based on the sparsity of mmWave channels. However, the aforementioned methods either need a full scan over all possible beam pairs or require frequent beam training, leading to high overhead. In this paper, a novel beam alignment scheme based on an integrated KF and CS framework is proposed for the single-user mmWave channel. Since the CS performance is heavily dependent on the sparsity level, the proposed method increases the signal sparsity level by applying adaptive CS on the observation residual computed from the previous estimate of the support to predict the angles of the dominant paths, while the corresponding path gains are tracked by the reduced-order KF. To further exploit the spatial correlation of sparse signals, our approach considers the adaptation of the sensing matrix based on the previous estimate to enhance the estimation accuracy, and the weighted CS is adopted for shrinking the possible range of solutions. To balance the estimation accuracy and alignment overhead, a switching mechanism is proposed that determines the timing for switching the estimation policy between beam training and beam tracking. Simulations are performed to evaluate the performance of the proposed method subject to numerous important factors, such as signal-to-noise ratio (SNR), number of available beams, overhead for beam alignment, beam variation rate, and the dominance of line-of-sight path.

INDEX TERMS Beam alignment, beam tracking, compressed sensing, Kalman filter, millimeter wave, mobile communications.

I. INTRODUCTION

MILLIMETER Wave (mmWave) is a promising technology to fulfill the ever-growing demand of high data rate transmission. It is the core technology of the Fifth Generation (5G) cellular networks and beyond with numerous appealing use cases, such as Vehicle-to-Everything (V2X) [1], mobile Augmented Reality (AR)/Virtual Reality (VR), urgent health care, and so on [2], [3]. Compared to the sub-6 GHz counterpart, mmWave signals suffer much higher attenuation due to distance-dependent propagation loss and

blockage [4]. To remedy, the use of highly directional transmission and reception through beamforming has been widely considered. Thanks to the small wavelength of mmWave, it is feasible to install many antennas on mmWave devices so as to leverage beamforming gain provided by large antenna arrays.

To achieve reliable and high data rate using mmWave, channel state information (CSI) is necessary for designing the transmitting and the receiving beams. For mmWave communications with large antenna arrays, acquiring CSI can

be costly as the training overhead of conventional pilot-based channel estimation schemes grows with the number of antennas. To address this issue, some solutions have been proposed. For example, CSI can be obtained using beam space channel representation [5], [6] or referred to as virtual channel [7], which has a lower dimension than the antenna space channel. Alternatively, the mmWave channel can be cast into a parametric model using the angles of departure (AoD), angles of arrival (AoAs), and the path gains [8]. Since the number of paths in mmWave channels is small (i.e., sparse channel), the number of channel parameters to be estimated is not large that facilitates CSI acquisition. The sparse mmWave channel also leads to the recent development of compressed sensing (CS)-based channel estimation schemes for mmWave [9], [10].

A challenge of CSI acquisition is when the channel is time-varying due to movement, in which case CSI needs to be updated timely to ensure the beams used by the transmitter and the receiver are aligned. Unless the training rate matches the channel variation speed, beam misalignment may frequently occur that degrades achievable data rate and may even cause link failure. In typical wireless systems, such as WiFi and LTE, training signals are transmitted periodically and it is unlikely to predict the channel variation speed accurately enough. Further, simply increasing the training frequency is not desired because of high training overhead and suspended data transmission during beam training.

To acquire CSI timely in time-varying mmWave channels, efficient beam alignment strategies are of vital importance. Similar to channel tracking in time-varying multi-input and multi-output (MIMO) systems, channel statistics can be used to track the channel changes of time-varying mmWave channels. Furthermore, the sparse feature of mmWave channels suggests that it is possible to track channel changes using a small number of training signals. For mobile mmWave communications, how to accurately track the channel changes with low overhead is a great challenge, which is the focus of this work.

A. RELATED WORK

The need for efficient beam alignment methods for time-vary mmWave channels has stimulated numerous studies recently. The technique of subspace tracking used in conventional MIMO systems is applied to mmWave hybrid beamforming systems in [8], which updates the analog precoder and combiner sequentially according to temporal channel statistics in terms of correlation coefficient. One limitation of this approach is that the orthogonal basis of the subspace to be tracked needs to be updated one by one for ensuring good tracking accuracy, resulting in high tracking overhead. Tracking overhead can be reduced by exploiting the sparsity of mmWave channel as done in the compressed sensing (CS)-based beam tracking. CS-based beam tracking does not track AoAs/AoDs directly. Instead, it tracks the support set (i.e., non-zero elements) of the equivalent MIMO mmWave channel [5] or functions of AoAs/AoDs, such as the spatial

frequency [7] or the spatial direction [11], [12]. Existing CS-based beam tracking algorithms have some weaknesses as follows. (i) While the sparsity is exploited in [5], the support set is tracked using both uplink and downlink channel sounding that still results in high tracking overhead. Besides, they focused on tracking angular variations while the channel gain information is also important to mobile mmWave communications. (ii) To track the spatial frequencies or the spatial directions of the mmWave channel, the receiver needs to feed back the measured channel matrix [7], [12]. This significantly causes high feedback overhead. (iii) The work in [11] presumes a linear motion model and the presence of LoS component, which may not always be valid in practice. (iv) To improve the estimation accuracy, CS is combined with Bayesian inference (BI), leading to a so-called turbo-BI algorithm for tracking the sparse massive MIMO channel. Turbo-BI employs two modules, including a linear minimum mean square error (MMSE) estimator followed by a sparsity combiner. The output of the sparsity combiner is the posterior distribution of the beam space channel, which is used to refine the MMSE estimation and then these two modules are iteratively performed until reaching a convergence, leading to high computational complexity.

Kalman filter (KF) is another popular method to track channel variations for its simplicity. Several KF-based mmWave channel tracking algorithms have been reported [13], [14], [15], [16], [17]. In KF, the system dynamic is captured by the so-called *state*, which is estimated based on a series of measurements. The native KF assumes the linear relationship between the state and collected measurements. In the context of channel tracking, such a linearity may not hold because the measurement (i.e., the received training signal) is not linear with respect to the state (i.e., AoAs/AoDs). In this case, methods such as linear approximation [13] and Gaussian approximation [15] are widely used. However, existing KF-based tracking algorithms for mmWave channels have the following limitations. (i) Only the angular variation of the channel is tracked [13], [14] and thus, an additional mechanism is required to estimate channel gain variations. The work in [16] tracks the complex channel coefficients assuming they are governed by the correlation coefficient, which needs to be known in priori. (ii) The state is updated using full beam training, leading to the training overhead growing with the channel dimension. Besides, the sparsity of the mmWave channel is not exploited in [17].

Besides CS and KF, other beam tracking methods based on channel distributions have also been studied. In [18], a probabilistic estimation approach is proposed to track the temporal variation of the mmWave channel using the angular information embedded in the preamble. This approach saves tracking overhead because no spatial scanning is required but the received RF signal needs to be stored before baseband processing. In [19], the idea of dedicated beams is proposed to track users with different mobility. For high-mobility users, additional training symbols, referred to dedicated

beams, are used to estimate the posteriori distribution of the time-varying mmWave channel. Their beam tracking algorithm performs better than the conventional KF-based beam tracking [14] yet the posteriori channel distribution needs to be recursively calculated. Another beam tracking algorithm based on posteriori update is presented in [20] for mmWave-supported inter-unmanned aerial vehicle (UAV) communications. Compared to [19], the beam tracking algorithm [20] is more lightweight but it only tracks the angle variation.

B. PAPER CONTRIBUTION

This work proposes a novel beam alignment approach for the mobile mmWave channel. The paper has the following contributions.

- Our approach is based on [21], which uses a fixed sensing matrix to track the time-varying supports in a sparse signal. In our work, the sensing matrix is adaptively determined according to the channel variations. During beam training, the sensing matrix is formed by a subset of available analog beams at the BS and the UE instead of all of them as required in full beam scanning. To track the channel variation, the proposed beam tracking method uses a very small sensing matrix according to the tracking result in the previous slot. To improve the tracking accuracy using a small sensing matrix, we propose a weighted CS, where the weights are designed also based on the past tracking results. This is different from [21] that uses a regular CS. Moreover, we propose a switching mechanism to alternately perform beam training and beam tracking according to a hypothesis test result. A similar idea is used in [21] for detecting the change of the support set but not for switching between two different procedures. Besides, we explain how to determine the threshold used in the hypothesis test that is not specified in [21].
- Our approach can track both angular and gain variations of the sparse mmWave channel. This is achieved by using CS to find the angles of the dominant paths and a reduced-order KF to track the path gains. Notice that CS and KF are coupled in the sense that CS is performed based on KF output. Most existing approaches track the angular variations only but not the channel gains [5], [11], [13], [14], [17], [20].
- The proposed approach does not rely on a specific channel variation model as assumed in [11] and thus it is not restricted to a certain moving scenario. Besides, it updates the angles and path gains of the underlying channel by jointly using CS and KF as mentioned without recursive procedures required in [18], [19].

Notation: Upper-case and lower-case boldface letters \mathbf{A} and \mathbf{a} denote a matrix and a vector, respectively. $(\cdot)^T$, $(\cdot)^*$, $(\cdot)^H$, $(\cdot)^{-1}$, and $\|\cdot\|_p$ denote the transpose, conjugate, conjugate transpose, matrix inversion, and l_p norm operation, respectively. The $\text{diag}\{\mathbf{a}\}$ denotes the diagonal matrix whose diagonal elements consist of the elements in the vector \mathbf{v} .

$\mathbb{E}\{\cdot\}$ denotes the expectation. Let S denote the support set of \mathbf{x} , i.e., the indices of its non-zero elements. \mathbf{a}_S denotes the $\text{size}(S)$ length sub-vector containing the elements of \mathbf{a} corresponding to the indices in S . For a matrix \mathbf{A} , \mathbf{A}_S represents the sub-matrix obtained by extracting the columns of \mathbf{A} corresponding to the elements in S . We also use the notation \mathbf{A}_{S_1, S_2} to denote the sub-matrix of \mathbf{A} containing columns and rows corresponding to the indices in S_1 and S_2 respectively. The set operations \cup , \cap and \setminus have the usual meanings (note $S_1 \setminus S_2$ denotes elements of S_1 not in S_2). S^c denotes the complement of S . We use the notation $\mathcal{CN}(\mathbf{m}, \mathbf{R})$ to denote the complex Gaussian distribution with mean vector \mathbf{m} and covariance matrix \mathbf{R} . Finally, \mathbf{I}_N is the $N \times N$ identity matrix.

II. SYSTEM MODEL

In this work, a mmWave system with one fixed base station (BS) and a single moving user equipment (UE) is considered. The BS and the UE have N_{BS} and N_{UE} antennas, respectively, in uniform linear array (ULA). For the N -element ULA, the $N \times 1$ steering vector is given by

$$\mathbf{a}(\theta) = \frac{1}{\sqrt{N}} \left[e^{j2\pi\theta m} \right]_{m \in J(N)}, \quad (1)$$

where $J(N) = \{\frac{i-1}{N}, i = 1, \dots, N\}$ and the spatial frequency $\theta = \frac{d}{\lambda} \sin \Theta$ with $\Theta \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ representing the physical direction, λ is the signal wavelength and d is the antenna spacing.

A. CHANNEL MODEL

The BS and the UE are synchronized in time, which is divided into equal-length time slots. Denote \mathbf{H}_t the channel coefficient at slot t , which remains constant within a slot but varies in different slots. Following the widely used Rician fading model,

$$\mathbf{H}_t = \sqrt{\frac{K}{K+1}} \alpha_{0,t} \mathbf{a}_{\text{UE}}(\psi_{0,t}) \mathbf{a}_{\text{BS}}^H(\phi_{0,t}) + \sqrt{\frac{1}{K+1}} \sum_{l=1}^L \alpha_{l,t} \mathbf{a}_{\text{UE}}(\psi_{l,t}) \mathbf{a}_{\text{BS}}^H(\phi_{l,t}), \quad (2)$$

where K is the Rician factor, $\alpha_{l,t}$, $\psi_{l,t}$, and $\phi_{l,t}$ are complex path gain, the angle of arrival (AoA), and the angle of departure (AOD), respectively, of the l th path, $l = 0, 1, \dots, L$, in slot $t \geq 0$. At the mmWave frequencies, the number of resolvable paths is often small such that the channel is sparse in the angular domain. To see this, divide the spatial angle from 0 to 2π into Q spatial bins. With sufficiently large Q , it is reasonable to assume $\psi_{l,t}, \phi_{l,t} \in \{0, \frac{2\pi}{Q}, \dots, \frac{2\pi(Q-1)}{Q}\}$. This allows us to transfer the spatial-domain channel matrix \mathbf{H}_t to the angular domain as given by

$$\mathbf{H}_t = \mathbf{A}_{\text{UE}} \bar{\mathbf{H}}_t \mathbf{A}_{\text{BS}}^H, \quad (3)$$

where $\mathbf{A}_{\text{UE}} = [\mathbf{a}_{\text{UE}}(\psi_1), \dots, \mathbf{a}_{\text{UE}}(\psi_Q)] \in \mathbb{C}^{N_{\text{UE}} \times Q}$ and $\mathbf{A}_{\text{BS}} = [\mathbf{a}_{\text{BS}}(\phi_1), \dots, \mathbf{a}_{\text{BS}}(\phi_Q)] \in \mathbb{C}^{N_{\text{BS}} \times Q}$. Here $\mathbf{a}_{\text{UE}}(\psi_i)$ and $\mathbf{a}_{\text{BS}}(\phi_j)$ are the steering vectors corresponding to the i th

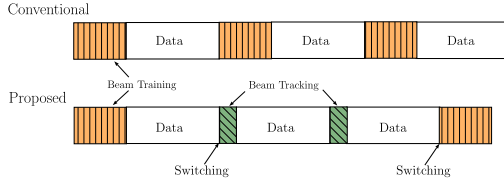


FIGURE 1. Transmission protocol: conventional and proposed beam alignment.

angular bin and the j th angular bin for AoA and AoD, respectively. Thus, the (i, j) -element in $\bar{\mathbf{H}}_t$ represents the path gain between the j -th angular bin of the BS and the i -th angular bin of the UE. When the number of paths L is small, the condition $(L + 1) \ll Q$ implies only $L + 1$ elements in $\bar{\mathbf{H}}_t$ have non-zero values and the rest are zero. The sparsity of $\bar{\mathbf{H}}_t$ suggests that the channel between the BS and the UE can be characterized by $L + 1$ dominant beam pairs, each with a specific gain and corresponding angular bins. When the UE moves, both the gains and the angles of these $L + 1$ beam pairs may change in time, leading to a time-varying mmWave channel that needs to be tracked.

B. TRANSMISSION PROTOCOL

In this work, beam alignment refers to the procedure that determines the transmit beamformer at the BS for data transmission and the receive combiner at the UE for data reception. Fig. 1 illustrates the conventional beam alignment using full beam scanning and the proposed one. Suppose the BS can steer the analog beam to a finite number of directions denoted as Q_{BS} by adjusting the phase shifters of the antenna array. If full beam scanning is used, the BS transmits the training signals using Q_{BS} analog beams where the i th transmit beam is $\mathbf{f}_i = \mathbf{a}_{BS}(\phi_i) \in \mathbb{C}^{N_{BS}}$, $i = 1, 2, \dots, Q_{BS}$. The UE receives the training signals using Q_{UE} combining vectors, each steering the analog beam to a specific direction. The j th receive beam is denoted as $\mathbf{w}_j = \mathbf{a}_{UE}(\psi_j) \in \mathbb{C}^{N_{UE}}$, $j = 1, 2, \dots, Q_{UE}$. Thus, the conventional beam alignment needs to train $Q_{BS} \times Q_{UE}$ beam pairs.

Initially, the proposed beam alignment scheme performs full beam scanning. Consider a particular beam pair composed by the i -th transmit beam and the j -th receive beam. For the sake of clarity, the slot index is dropped in the rest of this section. The received training signal can be expressed as

$$y_{ji} = \mathbf{w}_j^H \mathbf{H} \mathbf{f}_i s + \mathbf{w}_j^H n_{ji}, \quad (4)$$

where n_{ji} is the complex Gaussian noise and s is the training signal. Following [19], we let $s = 1$ for simplicity. By collecting the received signals through $Q_{BS} \times Q_{UE}$ beam pairs, we have the following matrix representation.

$$\mathbf{Y} = \mathbf{W}^H \mathbf{H} \mathbf{F} + \mathbf{N}, \quad (5)$$

where $\mathbf{F} = [\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_{Q_{BS}}]$ is the beamforming matrix of dimension $N_{BS} \times Q_{BS}$ and $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_{Q_{UE}}]$ is the combining matrix of dimension $N_{UE} \times Q_{UE}$, and $\mathbf{N} \sim \mathcal{CN}(0, \sigma_{obs}^2 \mathbf{I})$ is an $Q_{UE} \times Q_{BS}$ Gaussian noise matrix with elements being independent identically distributed (i.i.d.).

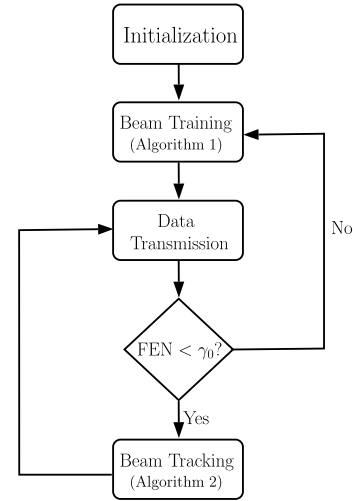


FIGURE 2. Flow chart of the proposed beam alignment scheme.

By concatenating the columns of \mathbf{Y} and letting $\mathbf{y} = \text{vec}(\mathbf{Y})$, $\mathbf{n} = \text{vec}(\mathbf{N})$, we have a vectorized expression as given by [3]

$$\begin{aligned} \mathbf{y} &= \text{vec}(\mathbf{W}^H \mathbf{H} \mathbf{F}) + \mathbf{n} \\ &= \left((\mathbf{F}^T \mathbf{A}_{BS}^*) \otimes (\mathbf{W}^H \mathbf{A}_{UE}) \right) \mathbf{x} + \mathbf{n}, \end{aligned} \quad (6)$$

where \otimes denotes the Kronecker product and $\mathbf{x} = \text{vec}(\bar{\mathbf{H}})$ with $\bar{\mathbf{H}}$ being the angular-domain channel matrix in (3) or known as the beamspace channel. The vectorized form (6) will facilitate the process of casting the beam alignment/training problem into a CS-aided KF (CS-KF) framework as will be seen in Section III. As explained in Section II-A, $\bar{\mathbf{H}}$ is sparse and thus, only a few elements in \mathbf{x} is non-zero. The indices of these non-zero elements provide the AoA and AoD information to be estimated that will be explained in the next section.

Once the initial beam pairs are acquired, the UE can decide the combining vectors for the following data reception. The UE also feeds back the indices of non-zero elements in \mathbf{x} to the BS, which uses the feedback information to decide the beamforming vectors for data transmission. After the data transmission, beams are aligned by performing either beam training or beam tracking, according to the switching mechanism presented in Section III-C. Fig. 2 illustrates the proposed beam alignment scheme with details given in the sequel.

III. PROPOSED BEAM ALIGNMENT SCHEME

We propose a beam alignment approach that jointly estimates the angular and gain variations of the time-varying mmWave channel by combining CS and KF, where the former leverages the channel sparsity and the latter deals with the temporally correlated sparsity. The proposed beam alignment scheme switches between the beam training procedure and the beam tracking procedure, each using a different sensing matrix. In this section, we elaborate these two procedures and how they are alternately triggered.

A. BEAM TRAINING PROCEDURE

In the beam training procedure, CSI (including AoA/AoD and path gains) is acquired by using M_{train} number of training signals. Specifically, the BS transmits the training signal using $M_{\text{BS,train}}$ transmit beams selected from \mathbf{F} while the UE receives the training signal using $M_{\text{UE,train}}$ receive beams selected from \mathbf{W} . Overall, the number of training signals, M_{train} , in the beam training phase is equal to $M_{\text{UE,train}} \times M_{\text{BS,train}}$. To cast the beam training problem into the CS framework, we first rewrite (6) as

$$\mathbf{y}_t = \mathbf{P}_{\text{train}} \mathbf{x}_t + \mathbf{z}_t, \quad (7)$$

where \mathbf{y}_t is the observation vector at slot t , $\mathbf{P}_{\text{train}} = (\mathbf{F}_{\text{train}}^T \mathbf{A}_{\text{BS}}^*) \otimes (\mathbf{W}_{\text{train}}^H \mathbf{A}_{\text{UE}})$ is a $M_{\text{train}} \times Q_{\text{BS}} Q_{\text{UE}}$ sensing matrix, and $\mathbf{x}_t \in \mathbb{C}^{Q_{\text{BS}} Q_{\text{UE}}}$ denotes the state vector. Here $\mathbf{F}_{\text{train}} \in \mathbb{C}^{N_{\text{BS}} \times M_{\text{BS,train}}}$ is a submatrix of \mathbf{F} with $M_{\text{BS,train}}$ columns in \mathbf{F} being active. Similarly, $\mathbf{W}_{\text{train}} \in \mathbb{C}^{N_{\text{UE}} \times M_{\text{UE,train}}}$ is a submatrix of \mathbf{W} . The state is governed by the state transition equation,

$$\mathbf{x}_t = \mathbf{x}_{t-1} + \mathbf{v}_t \quad (8)$$

where \mathbf{v}_t is assumed to be a zero-mean white Gaussian noise with the covariance matrix $\mathbb{E}[\mathbf{v}_t \mathbf{v}_t^T] = \sigma_{\text{sys}}^2 \mathbf{I}_{M_{\text{train}}}$. We note that σ_{sys}^2 depends on the variation of the path gains as will be specified in Section IV.

Remark 1: The state vector \mathbf{x}_t is sparse because it is the vectorized beamspace channel as explained in Section II. Thus, only a few entries in \mathbf{x}_t have non-zero values depending on the number of resolvable paths. This implies one can estimate the path gains by estimating \mathbf{x}_t . The indices of non-zero entries indicate the path angles in the beamspace channel.

The information about path gains and angles is embedded in the support of \mathbf{x}_t denoted as S_t . For $t = 1$, the support of \mathbf{x}_1 is obtained via full beam scanning. For the rest of time, the correlation between \mathbf{x}_t and \mathbf{x}_{t-1} is exploited to estimate S_t without probing all the beam pairs. Without loss of generality, suppose the first change of the active beam pairs occurs at slot t_a . The change may require to add or remove a certain non-zero entries in S_t as explained below.

1) *Addition:* Let us first consider the case that the change corresponds to a new set Δ added to the support set of \mathbf{x}_t . To estimate the new state denoted as \mathbf{x}_{new} at $t = t_a$, we use the filtering error, which is the difference between the observation vector \mathbf{y}_t and the predicted one based on the old estimate $\hat{\mathbf{x}}_t$, as given by

$$\tilde{\mathbf{y}}_{t,f} = \mathbf{y}_t - \mathbf{P} \hat{\mathbf{x}}_t. \quad (9)$$

Since the observation can be expressed as $\mathbf{y}_t = \mathbf{P}_{S_t}(\mathbf{x}_t)_{S_t} + \mathbf{P}_{S_t^c}(\mathbf{x}_t)_{S_t^c} + \mathbf{n}_t$, (9) can be rewritten as

$$\tilde{\mathbf{y}}_{t,f} = \mathbf{P}_{S_t^c}(\mathbf{x}_t)_{S_t^c} + \mathbf{P}_{S_t}(\mathbf{x}_t - \hat{\mathbf{x}}_t)_{S_t} + \mathbf{n}_t. \quad (10)$$

Clearly, the new set Δ belongs to the complementary support set S_t^c . Thus, $\tilde{\mathbf{y}}_{t,f}$ contains the information about Δ and it is a sparse vector. This motivates us to find Δ by estimating

Algorithm 1: Beam Training Algorithm

Input: Observation \mathbf{y}_t , sensing matrix $\mathbf{P}_{\text{train}}$, current estimate $\hat{\mathbf{x}}_t$, adding/deletion threshold α .

1. **Addition.** Compute the filtering error (9)

(a) Estimate \mathbf{x}_{new} by applying CS on filtering error using DS (11).

(b) Let $\Delta_{\text{add}} = \{i : |(\mathbf{x}_{\text{new}})_i| > \alpha\}$.

The addition to the support set is $\Delta = (S_t^c)_{\Delta_{\text{add}}}$.

Update $S_t \leftarrow S_t \cup \Delta$.

(c) Let $(\mathbf{R}_{t|t-1})_{\Delta, \Delta} = \sigma_{\text{init}}^2 \mathbf{I}$.

2. **Deletion.** Let $\Delta_{\text{del}} = \{i : |(\hat{\mathbf{x}}_t)_i| < \alpha\}$.

Set $(\mathbf{R}_{t|t-1})_{:, \Delta_{\text{del}}} = \mathbf{0}$ and $S_t \leftarrow S_t \setminus \Delta_{\text{del}}$.

Output: The support set S_t and the updated state vector $\hat{\mathbf{x}}_t$.

$(\mathbf{x}_t)_{S_t^c}$ using CS. One common sparse estimator is the Dantzig selector (DS), which is an l_1 -norm minimization problem as given by [22]

$$\mathbf{x}_{\text{new}} = \min_{\mathbf{x}} \|\mathbf{x}\|_1, \quad \text{s.t.} \quad \left\| \mathbf{P}^H (\tilde{\mathbf{y}}_{t,f} - \mathbf{P} \mathbf{x}) \right\|_{\infty} < \epsilon \quad (11)$$

where ϵ is a constant depending on the noise level. Given the solution of (11), the support of \mathbf{x}_{new} , denoted as Δ_{add} is then obtained by thresholding $|(\mathbf{x}_{\text{new}})_i|$, i.e., $\Delta_{\text{add}} = \{i : |(\mathbf{x}_{\text{new}})_i| > \alpha\}$ for threshold α . Now the detected new support, namely the indices of the newly added beam pairs is given by $\Delta = (S_t^c)_{\Delta_{\text{add}}}$ and the new support of \mathbf{x}_t is updated as $S_t = S_t \cup \Delta$.

2) *Deletion:* It also happens that an active beam pair no longer matches to any path angles due to UE movement. In this case, the entries in \mathbf{x}_t corresponding to the outdated beam pairs should be removed from S_t . Also, some indices may be incorrectly added into S_t due to CS error using (11). Similar to the addition of new supports, we remove those outdated indices from S_t by thresholding. Denote the deletion set $\Delta_{\text{del}} = \{i : |(\hat{\mathbf{x}}_t)_i| < \alpha\}$, which contains the indices to be removed from S_t . Then the support set is updated as $S_t = S_t \setminus \Delta_{\text{del}}$.

A complete algorithm of the proposed beam training is presented in Algorithm 1. In each slot, we run KF including the prediction and the update steps, to capture the channel dynamics. Since the channel is sparse, it is sufficient to run a reduced-order KF on the entries of \mathbf{x}_t designated by the support. From the classical KF theory, KF prediction, also known as the time-update, performs the following steps to predict the current state $\hat{\mathbf{x}}_t$ and the estimate covariance matrix $\mathbf{R}_{t|t-1}$.

$$\begin{aligned} \hat{\mathbf{x}}_{t|t-1} &= \hat{\mathbf{x}}_{t-1} \\ (\mathbf{R}_{t|t-1})_{S_{t-1}, S_{t-1}} &= (\mathbf{R}_{t-1|t-1})_{S_{t-1}, S_{t-1}} + \sigma_{\text{sys}}^2 \mathbf{I} \end{aligned} \quad (12)$$

where $\hat{\mathbf{x}}_{t|t-1}$ represents the estimate of \mathbf{x} at slot t based on that at slot $t-1$, and $\mathbf{R}_{t|t-1}$ is the mean squared error in the estimate $\hat{\mathbf{x}}_{t|t-1}$. Initially, $\hat{\mathbf{x}}_0 = \mathbf{0}$ and $\mathbf{R}_{0|0} = \mathbf{0}$. Then the following routines are used to update system dynamics based on the current support S_{t-1} as [21]

$$\mathbf{K} = (\mathbf{R}_{t|t-1})_{S_{t-1}, S_{t-1}} \mathbf{P}_{S_{t-1}}^H \Sigma_{ie}^{-1},$$

$$\begin{aligned} \Sigma_{ie} &= \mathbf{P}_{S_{t-1}}(\mathbf{R}_{t|t-1})_{S_{t-1}, S_{t-1}} \mathbf{P}_{S_{t-1}}^H + \sigma_{\text{obs}}^2 \mathbf{I} \\ (\hat{\mathbf{x}}_t)_{S_{t-1}} &= (\hat{\mathbf{x}}_{t|t-1})_{S_{t-1}} + \mathbf{K}[\mathbf{y}_t - \mathbf{P}\hat{\mathbf{x}}_{t|t-1}] \\ (\mathbf{R}_{t|t})_{S_{t-1}, S_{t-1}} &= (\mathbf{R}_{t|t-1})_{S_{t-1}, S_{t-1}} [\mathbf{I} - \mathbf{K}\mathbf{P}_{S_{t-1}}] \end{aligned} \quad (13)$$

where \mathbf{K} is the KF gain, σ_{obs}^2 is the observation noise variance in (5), and Σ_{ie} is the covariance matrix of the innovation error, which will be defined later. Upon completing the KF prediction and update, the support S_t is updated by the addition and the deletion steps as described above. Finally, the system dynamics are refreshed using the updated support S_t .

B. BEAM TRACKING PROCEDURE

The beam training procedure presented above provides a profile of the active beam pairs to be used for data transmission. The purpose of beam tracking is to maintain the estimation accuracy of the active beam pairs using $M_{\text{track}} = M_{\text{BS,track}} \cdot M_{\text{UE,track}} \ll M_{\text{train}}$ number of training signals. This is beneficial to reduce the overhead of beam alignment but also challenging because now fewer observations are available.

Suppose beam tracking is performed at slot t , where the received training signal is given by

$$\mathbf{y}_t = \mathbf{P}_{\text{track}} \mathbf{x}_t + \mathbf{n}_t, \quad (14)$$

where $\mathbf{P}_{\text{track}} = (\mathbf{F}_{\text{track}}^T \mathbf{A}_{\text{BS}}^*) \otimes (\mathbf{W}_{\text{track}}^H \mathbf{A}_{\text{UE}}) \in \mathbb{C}^{M_{\text{track}} \times Q_{\text{BS}} Q_{\text{UE}}}$ represents the sensing matrix in the beam tracking phase. In this work, $\mathbf{F}_{\text{track}} \in \mathbb{C}^{N_{\text{BS}} \times M_{\text{BS,track}}}$ and $\mathbf{W}_{\text{track}} \in \mathbb{C}^{N_{\text{UE}} \times M_{\text{UE,track}}}$ are chosen according to the active beam pairs estimated at slot $t-1$. A reasonable choice is the active beam pairs in slot $t-1$ and their adjacent beams considering the slow channel variations within a short period. A similar idea has also been used in [19], [23] to reduce the overhead of search-based beam tracking. Specifically, [19] proposed to estimate AoD and AoA using dual beams for the time-varying mmWave channel, considering the sparsity of the mmWave channel. The two beams are selected from the codebook based on the distributions of AoD and AoA for minimizing the estimation error. When the channel varies slowly, the selected two beams are often adjacent to the active beams in the past beam training period. In [23], an analog beam is assigned with two virtual beams that form a so-called auxiliary beam pair to probe the path angles. This corresponds to insert some more columns in the beamforming and the combining matrices to enhance the angular resolution of beam training. Inspired by [19], [23], we employ the active beam pairs in the previous slot and their two neighboring beams to form the sensing matrix for beam tracking. For example, if \mathbf{f}_i is active in slot $t-1$, $\mathbf{F}_{\text{track}}$ in slot t is composed by \mathbf{f}_{i-1} and \mathbf{f}_{i+1} , for $i = 2, \dots, Q_{\text{BS}} - 1$. If \mathbf{f}_1 or $\mathbf{f}_{Q_{\text{BS}}}$ is active in slot $t-1$, only one adjacent beam will be included in $\mathbf{F}_{\text{track}}$ in slot t . The same rule applies to determine $\mathbf{W}_{\text{track}}$.

Similar to beam training, the task of estimating the active beam pairs in the beam tracking stage is equivalent to estimating the support of the state vector \mathbf{x}_t , which can be accomplished by running CS on the KF filtering error. Since

we aim to reduce the overhead for beam tracking, it is desired to use a very small number of training signals, i.e., the measurements available to estimate \mathbf{x}_t in the tracking stage is much less than in the training stage. In this case, a regular CS approach may not provide a reliable estimate. To enhance the estimation accuracy, we adopt a weighted l_1 minimization approach as given by [24]

$$\begin{aligned} \mathbf{x}_{\text{new}} &= \min_{\mathbf{x}} \|\omega_t \mathbf{x}\|_1, \\ \text{s.t. } & \left\| \mathbf{P}_{\text{track}}^H (\tilde{\mathbf{y}}_{t,f} - \mathbf{P}_{\text{track}} \mathbf{x}) \right\|_{\infty} < \epsilon, \end{aligned} \quad (15)$$

where $\omega_t \triangleq \text{diag}([\omega_{t,1}, \dots, \omega_{t, Q_{\text{BS}} Q_{\text{UE}}]})$ represents the weighting matrix and $w_{t,i}$ is the weight assigned to the i th entry of \mathbf{x}_t for $i \in [1, Q_{\text{BS}} Q_{\text{UE}}]$. The objective of norm minimization in (15) suggests a simple choice of weights being the inverse of the entries in \mathbf{x}_t (which correspond to the channel coefficients in our problem). As a result, large weights force the solution of (15) to concentrate on the indices where $\omega_{t,i}$ is small, corresponding to the indices of the active beam pairs. If the true state \mathbf{x}_t (i.e., actual channel matrix $\bar{\mathbf{H}}_t$) is k -sparse, i.e., obeying $\|\mathbf{x}_t\|_0 \leq k$, then (15) is guaranteed to find the correct solution with the choice of weights mentioned above [24]. Since the exact channel matrix $\bar{\mathbf{H}}_t$ is unknown, an alternative is to employ the previous estimate $\hat{\mathbf{x}}_{t-1}$ as the weights. This is because not only $\hat{\mathbf{x}}_{t-1}$ is sparse but also it presents the similar sparsity profile as \mathbf{x}_t when the channel varies slowly.

Based on the above explanations, we propose the following strategy to determine weights $\omega_{t,i}$ in (15). Considering the fact that small weights could be used to encourage nonzero entries, it is reasonable to assign a small weight to the i th beam pair if it is the active one in slot $t-1$. Meanwhile, the neighboring beam pairs of the active ones in slot $t-1$ are likely to be significant in slot t and thus, they should be assigned with small weights too. For the rest beam pairs, their likelihood of becoming the active beam pair could be ignored. Thus, we have the following rule.

$$\omega_{t,i} = \begin{cases} \frac{\alpha_1}{|\hat{x}_{t-1,i}|}, & \text{if } i \text{ is a active beam pair in } t-1, \\ \frac{\alpha_2}{|\hat{x}_{t-1,i}|}, & \text{if } i \text{ is a neighbor of the active beam} \\ \infty, & \text{pair in } t-1, \\ \infty, & \text{otherwise,} \end{cases} \quad (16)$$

where $\alpha_1, \alpha_2 \in (0, 1)$ and $\alpha_1 < \alpha_2$. Clearly, when $w_{t,i}$'s are all identical, WDS reduces to DS in (11). Algorithm 2 summarizes the proposed beam tracking algorithm.

C. SWITCHING MECHANISM

In practice, tracking errors might be accumulated as beam tracking relies on the previous estimate of the active beam pairs. Intuitively, one can perform beam training when the tracking error is large, but the tracking error is unknown. Besides, beam training is relatively costly due to more training signals required and thus, the switching between beam training and beam tracking should be carefully determined.

Algorithm 2: Beam Tracking Algorithm

Input:

The observation \mathbf{y}_t , sensing matrix $\mathbf{P}_{\text{track}}$, state vectors $\hat{\mathbf{x}}_{t-1}$ and $\hat{\mathbf{x}}_t$, the adding and deletion threshold α .

Initialize Set up the weighting matrix $\mathbf{\Omega}_t$ according to (16).

1. **Addition.** Compute the filtering error (9).
 - (a) Apply CS on filtering error using WDS (15).
 - (b) Let $\Delta_{\text{add}} = \{i : |(\mathbf{x}_{\text{new}})_i| > \alpha\}$.
The addition to the support set is $\Delta = (S_t^c)_{\Delta_{\text{add}}}$.
Update $S_t \leftarrow S_t \cup \Delta$.
 - (c) Let $(\mathbf{R}_{t|t-1})_{\Delta, \Delta} = \sigma_{\text{init}}^2 \mathbf{I}$.
2. **Deletion.** Let $\Delta_{\text{del}} = \{i : |(\mathbf{x}_t)_i| < \alpha\}$.
Set $(\mathbf{R}_{t|t-1})_{\cdot, \Delta_{\text{del}}} = \mathbf{0}$ and $S_t \leftarrow S_t \setminus \Delta_{\text{del}}$.

Output: The support set S_t and the updated state vector $\hat{\mathbf{x}}_t$.

Reference [21] provides a decision rule based on a hypothesis test to determine whether the estimation error is large, but it does not specify the threshold required in the test.

To find a proper switching threshold, we consider t_a where the change of beam pairs occurs that causes the estimation error, or known as the innovation error in KF, as given by

$$\begin{aligned} \tilde{\mathbf{y}}_t &\triangleq \mathbf{y}_t - \mathbf{P}\hat{\mathbf{x}}_{t|t-1} \\ &= \mathbf{P}(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1}) + \mathbf{n}_t. \end{aligned} \quad (17)$$

For $t < t_a$, S_t is perfectly known through full beam scanning performed initially. Hence, $\hat{\mathbf{x}}_{t|t-1}$ matches to \mathbf{x}_t and the distribution of $\tilde{\mathbf{y}}_t$ for $t < t_a$ is dominated by the noise, which has a zero mean. Now, consider the addition of new support Δ to S_t at t_a . The case of deletion follows the same principle and is omitted here. At $t = t_a$, the observation \mathbf{y}_t can be expressed as $\mathbf{y}_t = \mathbf{P}_{S_t}(\mathbf{x}_t)_{S_t} + \mathbf{P}_{\Delta}(\mathbf{x}_t)_{\Delta} + \mathbf{n}_t$. Then the innovation error due to the addition of Δ is

$$\begin{aligned} \tilde{\mathbf{y}}_{t_a} &= \mathbf{y}_{t_a} - \mathbf{P}\hat{\mathbf{x}}_{t_a|t_a-1} \\ &= \left[\mathbf{P}_{S_{t_a}}(\mathbf{x}_{t_a})_{S_{t_a}} + \mathbf{P}_{\Delta}(\mathbf{x}_{t_a})_{\Delta} + \mathbf{n}_{t_a} \right] - \mathbf{P}\hat{\mathbf{x}}_{t_a|t_a-1} \\ &= \mathbf{P}_{S_{t_a}}(\mathbf{x}_{t_a} - \hat{\mathbf{x}}_{t_a|t_a-1})_{S_{t_a}} + \mathbf{P}_{\Delta}(\mathbf{x}_{t_a})_{\Delta} + \mathbf{n}_{t_a} \\ &= \mathbf{P}_{\Delta}(\mathbf{x}_{t_a})_{\Delta} + \tilde{\mathbf{n}}_{t_a}, \end{aligned} \quad (18)$$

where $\tilde{\mathbf{n}}_{t_a} = [\mathbf{P}_{S_{t_a}}(\mathbf{x}_{t_a} - \hat{\mathbf{x}}_{t_a|t_a-1})_{S_{t_a}} + \mathbf{n}_{t_a}] \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Sigma}_{\text{ie}})$ and $\mathbf{\Sigma}_{\text{ie}}$ is given in (13). From (18), the new set Δ introduces non-zero mean to $\tilde{\mathbf{y}}_{t_a}$. On the other hand, $\tilde{\mathbf{y}}_t$ has a zero mean when the change of support set does not occur. This implies the decision on switching between beam training and beam tracking can be transformed into a hypothesis testing problem with two hypotheses: the null hypothesis \mathcal{H}_0 refers to the case that result of beam tracking is not trustworthy and the alternative hypothesis \mathcal{H}_1 is the case that the result of beam tracking is trustworthy. Whether the beam tracking result is trustworthy can be evaluated by filtering error norm as defined by [21]

$$\text{FEN} \triangleq \tilde{\mathbf{y}}_{t,f}^H \mathbf{\Sigma}_{\text{fe}}^{-1} \tilde{\mathbf{y}}_{t,f} \quad (19)$$

where the covariance matrix $\mathbf{\Sigma}_{\text{fe}} = [\mathbf{I} - \mathbf{P}_{S_t} \mathbf{K}] \mathbf{\Sigma}_{\text{ie}} [\mathbf{I} - \mathbf{P}_{S_t} \mathbf{K}]^H$ (see Appendix A). Then the hypothesis testing under

Algorithm 3: Beam Alignment Scheme Based on CS-KF

Input: The observation vector \mathbf{y}_t , sensing matrix \mathbf{P} , addition/deletion threshold α , switching threshold γ_0 .

- 1: **Initialize** $\mathbf{R}_0 = \mathbf{0}$, $\hat{\mathbf{x}}_0 = \mathbf{0}$, $S_0 = \emptyset$. $t = 1$.
- 2: Set $S \leftarrow S_{t-1}$
- 3: Run **KF prediction** (12) with S
- 4: Run **KF update** (13) with S
- 5: Compute FEN using (19)
- 6: **if** FEN $> \gamma_0$ **then**
- 7: Run Beam Training (Algorithm 1)
- 8: **else**
- 9: Run Beam Tracking (Algorithm 2)
- 10: Run **KF update** (13) with S
- 11: Set $S_t \leftarrow S$

Output: Support set S_t and estimated state vector $\hat{\mathbf{x}}_t$
12: Set $t \leftarrow t + 1$ and go to step 2

consideration can be stated as follows.

$$\begin{cases} \mathcal{H}_0 : \text{FEN} \geq \gamma_0 \\ \mathcal{H}_1 : \text{FEN} < \gamma_0 \end{cases} \quad (20)$$

where γ_0 is the switching threshold between beam training and beam tracking. A large γ_0 avoids excessively switching to beam training, thereby reducing the false alarm probability. This, however, increases the risk of miss alarm, that is, overusing beam tracking. For a threshold γ_0 , the false alarm probability can be expressed as

$$P_{\text{fa}} = \Pr[\text{FEN} \geq \gamma_0 | \mathcal{H}_1]. \quad (21)$$

Then the switching threshold can be found as

$$\gamma_0 = F^{-1}(1 - P_{\text{fa}}). \quad (22)$$

where F^{-1} is the inverse cumulative distribution function of FEN. The following proposition is useful to determine the threshold γ_0 .

Proposition 1: For perfect beam alignment, FEN follows the chi-squared distribution, i.e., $\text{FEN} \sim \chi_{\nu}^2$, where ν is equal to $2M_{\text{train}}$ or $2M_{\text{track}}$ depending on the beam alignment scheme (beam training or tracking).

Proof: See Appendix B. ■

Accordingly, γ_0 can be computed numerically given P_{fa} and ν , and the fact that FEN is chi-squared distributed.

Now the switching mechanism can be described as follows. At slot t , suppose beam tracking is used. At slot $t+1$, the FEN is computed and compared with the threshold γ_0 . If FEN is larger than γ_0 , beam training will be adopted at slot $t+1$ for acquiring channel variations using M_{train} number of training signals. In case FEN is smaller than γ_0 , beam tracking will be performed in slot $t+1$ using M_{track} training signals. Algorithm 3 summarizes the proposed beam alignment scheme, including the beam training procedure, the beam tracking procedure, and the switching mechanism described above.

Remark 2: Sometimes the FEN may have a value very close to the threshold γ_0 that leads to too frequent switching between beam training and beam tracking. This can be avoided by running the hypothesis test multiple times

TABLE 1. Technical differences between existing schemes.

Scheme	Reference	Path angle estimation	Path gain estimation
KF	[13]	Need full beam scanning.	✗
	[14]	Need full beam scanning.	✗
	[16]	Use position, velocity, and path gain.	✓
	[17]	Use angle distribution.	✓
CS	[5]	Need full beam scanning.	✗
	[7]	Use partial beam training.	✓
	[11], [12]	Use lens array without beam scanning.	✗
	[25]	Use adaptive sensing matrix.	✓
Dual beams	[19]	Use two selected beams based on estimated angle distribution.	✓
	[23]	Use two auxiliary beams based on partial beam training.	✗

before switching. For example, we may monitor the results of hypothesis test (20) in two consecutive slots. The switching occurs if the null hypothesis is received twice. This can help to stabilize the algorithm without deteriorating the performance.

We conclude this section by summarizing the existing beam alignment schemes relevant to our work in Table 1, which highlights the key idea for path angle estimation in the related work and whether path gain estimation is included. Except [7], [16], [17], [19], [25], the related work focuses on tracking the angle variation without addressing the channel gain estimation in the time-varying sparse channel considered in our work. In [7], the path gain is estimated in an iterative manner and each iteration needs to compute the Hessian matrix of a maximum likelihood cost function. To estimate the path angle, a lengthy sounding procedure is employed and the UE needs to feed back the estimated channel matrix. The problem of costly feedback is resolved in [25] by developing an adaptive strategy for determining the sensing matrix. However, the resultant sensing matrix may not meet the restricted isometry property (RIP) and thus, sophisticated algorithms are used to recover the sparse channel matrix. The method proposed in [16] requires movement (position and velocity) and path gain information to estimate the angles. In addition to the complexity for collecting the movement information, the estimation error of path gain will propagate to the angle estimation. The method proposed in [17] tracks the angular variation for one path only and the number of training signals for gain estimation needs to be larger than that for tracking angles. References [19] and [23] use some selected beams for beam alignment for reducing the overhead as discussed in Section III-B. The similar idea is also adopted during beam tracking in our work but we do not need the angle distribution as required in [19]. The angle distribution is not required in [23] either but only the path angles are estimated while the proposed CS-KF tracks both path gains and angles using very few beams.

IV. SIMULATION RESULTS

Simulation results are presented to evaluate the performance of the proposed beam alignment scheme for the time-varying mmWave channel. The simulation scenario follows the system model described in Section II with the parameters listed in Table 2. In simulations, the path gain of the l th

path is modeled as a Gauss-Markov process as given by

$$\alpha_{l,t} = \rho\alpha_{l,t-1} + z_t, \quad l = 0, 1, \dots, L \quad (23)$$

where ρ is the correlation coefficient, and $z_t \sim \mathcal{CN}(0, \sigma_{\text{sys}}^2)$ with $\sigma_{\text{sys}}^2 = 1 - \rho^2$. As to path angles, when the AoD and AoA of the l th path, namely $\phi_{l,t}$ and $\psi_{l,t}$, in slot t match to beam pair $b \in [1, Q_{\text{BS}}Q_{\text{UE}}]$, it implies b is in the support set of \mathbf{x}_t . Thus, we model the angular variations as follows. If $b \in S_t$, $\phi_{l,t+1}$ and $\psi_{l,t+1}$ match to one of the adjacent beam pairs of b with probability p_v , while they match to b with probability $1 - p_v$. A larger p_v , referred to as the beam variation rate, renders more frequent AoA/AoD changes. The temporal change of the support set can be stated as follows. For $b \in S_t$,

$$\begin{cases} b \in S_{t+1}, & \text{with probability } 1 - p_v, \\ (b \notin S_{t+1}) \cap (b + 1 \in S_{t+1}), & \text{with probability } 0.5p_v, \\ (b \notin S_{t+1}) \cap (b - 1 \in S_{t+1}), & \text{with probability } 0.5p_v. \end{cases} \quad (24)$$

In evaluating the proposed beam alignment scheme, labeled as CS-KF, the mean square error (MSE) of the estimated vector \mathbf{x}_t at each slot t , i.e., $\mathbb{E}[\|\mathbf{x}_t - \hat{\mathbf{x}}_t\|_2^2]$, was computed by averaging 100 runs, each containing 20 slots. During the beam training stage, the even-indexed beam pairs are active. That is, $M_{\text{BS,train}} = \frac{1}{2}Q_{\text{BS}}$ and $M_{\text{UE,train}} = \frac{1}{2}Q_{\text{UE}}$. As to beam tracking, the active beam in each slot as well as their two adjacent beams are used to transmit and receive training signals. The number of active beams in the tracking stage may vary depending on the number of detected paths, i.e., non-zero entries in the solution of (15). Suppose all the $(L + 1)$ paths can be identified, the number of active beam pairs is equal to $L + 1$. Hence, $M_{\text{BS,track}} = M_{\text{UE,track}} = 3(L + 1)$.

For comparisons, the following beam alignment schemes are used as the benchmark.

- CS-only: It estimates \mathbf{x}_{new} by solving DS in (11) every slot independently. In other words, beam training is performed consistently using the same amount of training signals as the proposed beam tracking scheme.
- Genie-KF: It generates the minimal MSE estimate of \mathbf{x}_t by performing the reduced-order KF with known support set.
- Full-Extended KF (Full-EKF) [13]: It estimates AoAs and AoDs using EKF assuming path gains are fixed. In our simulations, EKF proposed in [13] is implemented with known path gains and full observations, i.e., training signals are transmitted in $Q_{\text{BS}}Q_{\text{UE}}$ beam pairs.
- Half-EKF: It is a replica of EKF except that only a quarter of beam pairs are active for training as in the proposed method.
- Beam-space: Our work is based on the beam-space channel representation (3), which is also considered in [11] for the single-antenna UE and a linear moving model. For comparisons, we use the tracking method in [11] to find the AoD (i.e., active transmit beam) of each path using Q_{BS} of training signals and assume that the receive beams and the path gains are known perfectly.

TABLE 2. Simulation parameters.

Parameter	Value	Parameter	Value
Number of antennas	$N_{BS} = N_{UE} = 32$	Number of beams	$Q_{BS} = Q_{UE} = 32$
SNR	20 dB	Rician factor	6 dB
Number of paths	3	Beam variation rate	$p_v = 0.01$
False alarm probability	$P_a = 0.05$	Correlation coefficient	$\rho = 0.995$
Number of training beams	$M_{\min} = \frac{1}{4}Q_{BS}Q_{UE}$	Weights in WDS	$\alpha_1 = 0.01, \alpha_2 = 0.5$
Tracking beams	Current active beams and ± 2 adjacent beams	Initial variance	$\sigma_{\text{init}}^2 = 1$

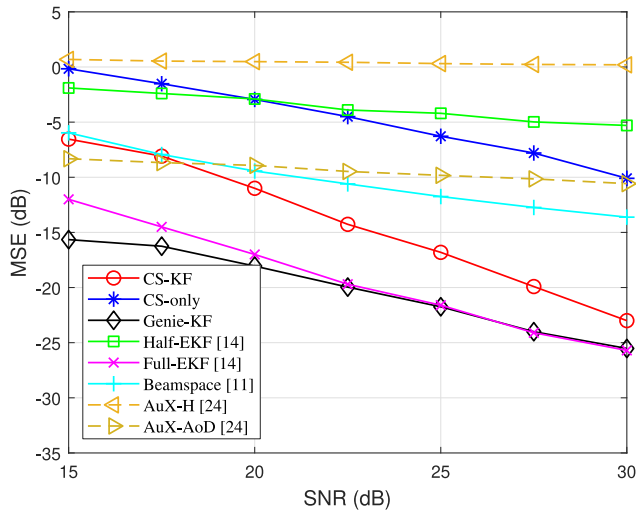


FIGURE 3. MSE vs. SNR of different algorithms.

- Auxiliary beam pair (AuX): In [23], joint AoD and AoA estimation is achieved using auxiliary beam pairs. The idea is to cover the desired angular range using analog beams at both the BS and the UE. To extend the angular coverage, each analog beam is accompanied by two virtual beams, referred to as the auxiliary beam pair. In this way, AoD and AoA can be estimated using less training signals than that by full scanning. In simulations, the number of auxiliary beam pairs at the BS is set to $N_{BS}/2 - 1$ and that at the UE is $N_{UE}/2 - 1$. As a result, the required number of training signals in AuX is close to that used in the beam training stage in the proposed CS-KF. Same as the beamspace scheme, we use AuX to estimate the AoD and assume the path gains are known perfectly. In addition to the MSE of the estimated channel matrix, labeled as AuX-H, we also evaluate the MSE of AoD estimation with the results labeled as AuX-AoD.

When solving DS in (11) and WDS in (15), we set $\epsilon = \sqrt{2} \log(N_{BS}N_{UE}) \text{trace}(\Sigma_{f_e}) / (Q_{BS}Q_{UE})$ as suggested in [21].

A. IMPACT OF SNR AND CHANNEL VARIATION RATE

We first examine the impact of observation noise to the MSE of all the considered beam alignment schemes. Given the average power of training signal equal to one, we define SNR as $1/\sigma_{\text{obs}}^2$. Fig. 3 shows the MSE performance of different schemes for varied SNRs. As expected, they all achieve better accuracy as SNR increases.

The MSE gap between the CS-KF and the genie-KF reduces with SNR, where the latter delivers the lower bound

on estimation error provided with known support. The full-EKF also performs comparably to the genie-KF using full observations. Comparing to the full-EKF, CS-KF used only 25% of the observations during the beam training stage and much fewer observations during the beam tracking stage. Clearly, the reduced observations in the CS-KF leads to certainly performance loss (10% higher MSE at SNR of 30 dB). If EKF uses the same amount of observations as the CS-KF, the result labeled as Half-EKF is much worse than CS-KF and even inferior to CS-only (i.e., performing independently beam training every slot). This reveals the major drawback of EKF that needs a large amount of observations to maintain a good beam alignment accuracy. On the other hand, the CS-only scheme also suffers a poor MSE (56% higher MSE than CS-KF at SNR of 30 dB), suggesting that using CS alone could not provide a reliable beam alignment performance. Comparing with the EKF and CS-only, CS-KF achieves a reasonable accuracy with much lower overhead. As to the beamspace scheme, its performance is close to the CS-KF when SNR is below 20 dB but it performs worst than the CS-KF at the high SNR regime (40% higher MSE than CS-KF at SNR of 30 dB). Since the AoA and the gain of each path is assumed to be perfectly known in the beamspace scheme, the only source of error is the AoD estimation. In a noise-free channel, the beamspace scheme can perfectly estimates the AoD [11], which is not the case in practice. Comparing with the CS-KF, the beamspace scheme is more sensitive to the observation noise that limits its estimation accuracy. Finally, Aux-AoD delivers a better performance than CS-only and Half-EKF using nearly the same amount of training signals while Aux-H performs the worst among all the considered schemes. This can be explained as AuX aims to estimate the physical angles of each path. The estimation error of AoD affects the phase of every antenna element and the accumulated errors result in inaccurate estimate of the channel matrix. Instead, CS-KF and the beamspace schemes estimate the active beams corresponding to the path direction (i.e., the index of elements in \mathbf{x}_t) rather than estimating the exact angles directly. Hence they do not suffer from accumulated phase errors as in AuX. The KF-based scheme can estimate the path angles reliably but this is achieved only with full beam scanning. Since full-EKF and half-EKF perform similarly to Genie-KF and CS-only, respectively, their results are omitted in the following discussions. The results of AuX-H and AuX-AoD are also excluded for brevity.

Fig. 4 shows the beam alignment performance under different beam variation rates. Here we vary the beam variation rate p_v from 0.01 to 0.15, corresponding to the mobility speed from 8.99 m/s to 134.85 m/s when the distance between the BS and UE is 10 m. As can be seen, the MSE of the proposed CS-KF increases with p_v and the same trend is observed in the genie-KF scheme. The reason is that, the required settling time of KF exceeds the coherence time of each entry in the state vector \mathbf{x}_t when the channel varies fast. As to the CS-only and the beamspace schemes, they do not track channel variation using the past estimates but performs estimation

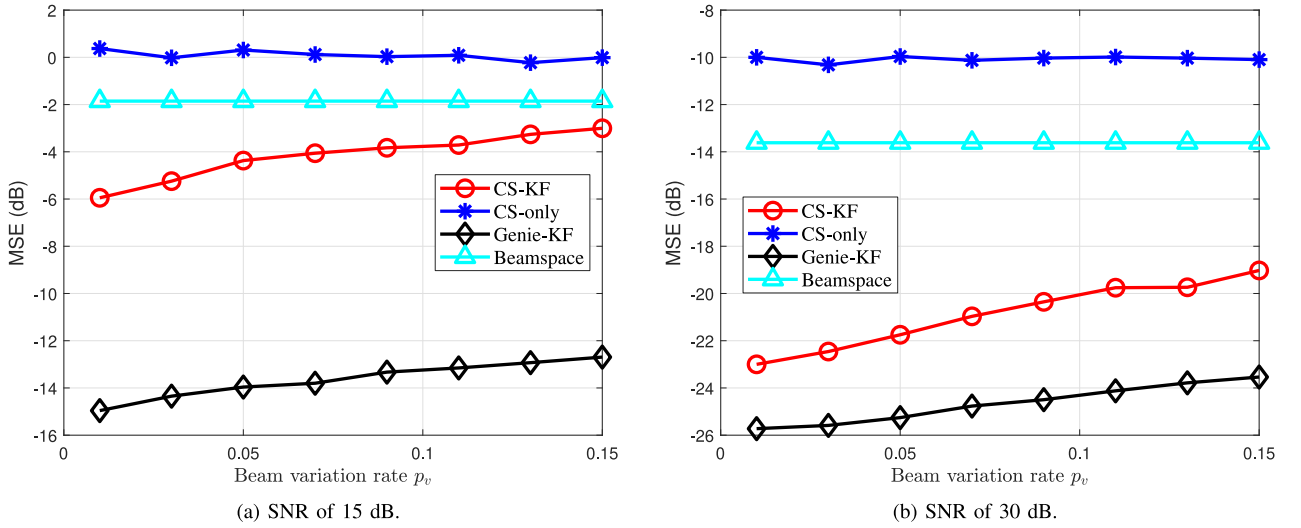


FIGURE 4. Impact of beam variation rate to the beam alignment accuracy.

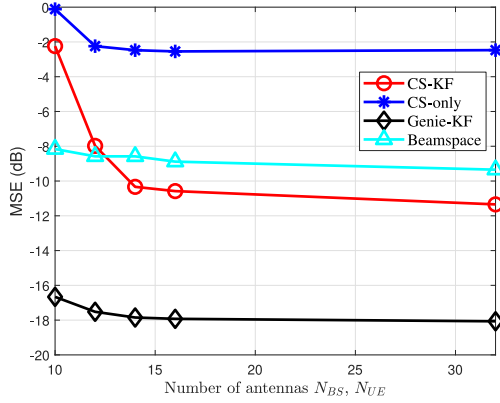


FIGURE 5. Impact of number of antennas to the beam alignment accuracy.

in each slot independently. Thus, the MSEs of CS-only and the beamspace schemes are not sensitive to p_v . In fact, the MSE of CS-only is fairly poor even at high SNR, which is the result of performing estimation using partial observations without the aid of the past estimates. Comparing the slopes of the MSE curves in Figs. 3 and 4, we see that the proposed CS-KF has a higher MSE as p_v increases but the performance degradation due to fast beam variations is less than that due to noisy observations.

B. IMPACT OF NUMBER OF ANTENNAS

Fig. 5 shows the beam alignment performance under different numbers of antennas. Here we set $N_{BS} = N_{UE}$ that vary from 10 to 32. As shown, all the considered schemes have a lower MSE as the number of antennas increases because of a higher spatial resolution for estimating the AoA and AoD. For different schemes, the MSE reduces with the number of antennas in a different rate. In the CS-only, the Genie-KF, and the beamspace schemes, the MSE improvement with more antennas is minor compared to CS-KF because their

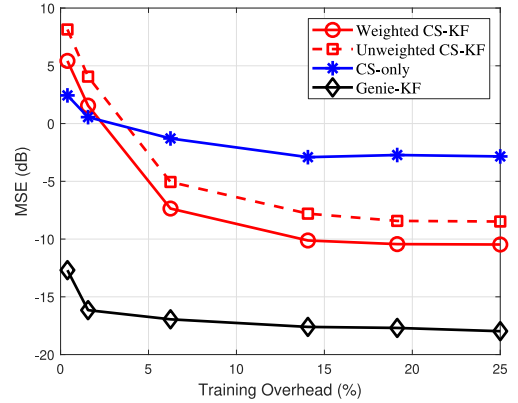
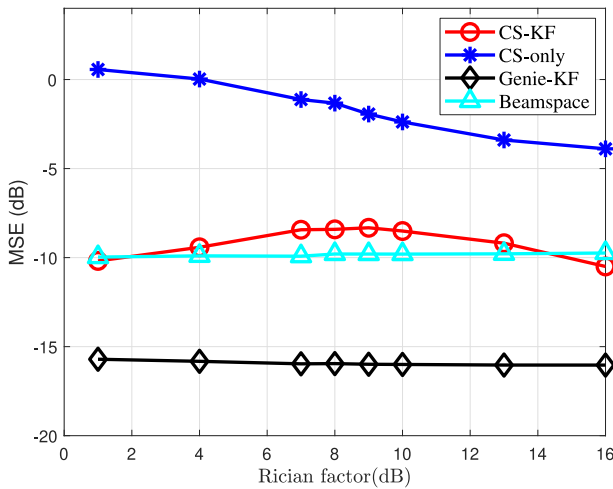


FIGURE 6. Impact of training overhead to the beam alignment accuracy.

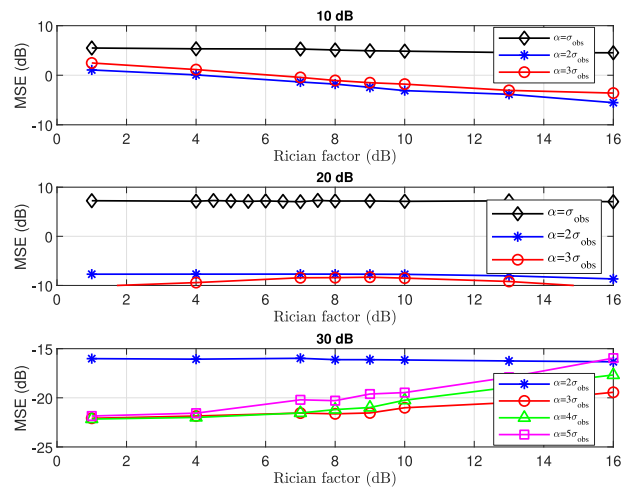
accuracy is dominated by the observation noise that is invariant of the number of antennas. While the observation noise also affects the alignment accuracy of CS-KF, some estimation errors could be corrected by the addition and deletion steps. This allows CS-KF to benefit more from the increased spatial resolution with a larger number of antennas. The beamspace scheme performs better than the CS-KF when $N_{BS} = N_{UE} < 12$ because the former performs full beam scanning in each slot. However, a 3-dB performance loss is incurred as the number of antennas increases. As explained above, the beamspace scheme is very sensitive to the observation noise, which can not be mitigated by adding more antennas.

C. IMPACT OF TRAINING OVERHEAD

Fig. 6 shows the MSE versus the average training overhead, which is the average between the number of training signals $M_{BS,train} \cdot M_{UE,train}$ to the total number of available beams (equal to 1,024). We plot the training overhead from 5% to 50% because the two benchmarks, namely CS-only and



(a) Comparison of different algorithms.



(b) MSE of CS-KF under different threshold α .

FIGURE 7. Impact of Rician factor to the beam alignment accuracy.

Genie-KF reach their performance limit within this range. Here the beamspace scheme is not considered because it requires full beam scanning to obtain a reasonable accuracy. For the proposed CS-KF, we consider two options to estimate the support in the tracking stage: (i) use (16) to determine the weights in WDS with the results labeled as “Weighted CS-KF (weighted)” and (ii) set al. weights equal to one with the results labeled as “Unweighted CS-KF”. One can see that Weighted CS-KF has a smaller MSE than Unweighted CS-KF. The MSE gain of Weighted CS-KF is around 2 ~ 2.7 dB, confirming the effectiveness of the proposed weighting rule (16) in improving the estimation accuracy of the support set. The improvement is invariant to the training overhead because the weighting rule (16) is used in the beam tracking stage where the number of training signals is fixed. When the training overhead is very small ($< 3\%$), the CS-KF performs worse than the CS-only. In this region, the underlying system is highly underdetermined and thus CS performs poorly. Also, KF fails to reliably predict \mathbf{x}_t based on $\hat{\mathbf{x}}_{t-1}$ when the number of training signal is too small. With the increased training overhead, the estimation accuracy of both CS and KF gets improved while the estimation accuracy of CS-only is limited because it relies on partial observations only without the aid of past estimates as in KF. proposed CS-KF.

D. IMPACT OF RICIAN FACTOR

The MSE performance subject to different Rician factors is illustrated in Fig. 7(a). As shown, the trends of MSE curves are quite different for different schemes. For CS-KF, the MSE first increases because the NLoS paths become weaker as the Rician factor increases and thus, the corresponding beam pairs are more likely to be removed incorrectly in the deletion step. The MSE curve reaches a maximum around the Rician factor of 9 dB. As the Rician factor further increases,

the trend reverses because the LoS component is more dominant, implying that the channel is more sparse. The sparser channel helps to improve the estimation accuracy of the CS-based algorithm. This also explains the reduced MSE of CS-only as the Rician factor increases. Since the deletion step is not used in CS-only, the MSE decreases monotonically. For the Genie-KF, it assumes perfect knowledge of the support set and thus the MSE is not sensitive to the Rician factor. As to the beamspace scheme, its MSE performance is quite stable within the considered range of the Rician factor. This is because when the Rician factor is large, the estimation error is primarily dependent on the quality of the LoS path, whose strength is determined by the SNR.

Since the addition and deletion steps play important roles in the proposed CS-KF, we investigate the impact of the addition/deletion threshold α as a function of the observation noise variance to the MSE in Fig. 7(b). The threshold α is related to the detection sensitivity of beam update. With a small α , addition and deletion of the support set occur more frequently. While this increases the detection sensitivity of beam changes due to channel variations, the chance of incorrect addition and deletion also increases. The choice of α depends on the noise variance. From Fig. 7(b), $\alpha = 2\sigma_{\text{obs}}$ achieves the lowest MSE for SNR of 10 dB while $\alpha = 3\sigma_{\text{obs}}$ is the best for the SNR of 20 and 30 dB. This can be explained because the SNR is inversely proportional to σ_{obs} . Therefore a higher α should be used to improve the effectiveness of the addition and deletion steps.

E. PERFORMANCE OF SWITCHING MECHANISM

In the proposed beam alignment scheme, beam training and beam tracking are adaptively performed according to the tracking error. To demonstrate the merit of such a dynamic operation, we compare the proposed beam alignment scheme and the periodic beam training scheme without adaptation, which performs beam training in a fixed period from 3 to

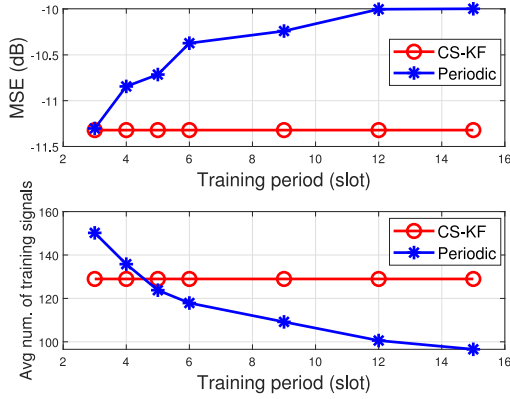


FIGURE 8. Comparison of the proposed beam alignment scheme and the periodic beam training scheme.

15 slots in Fig. 8. Here we plot the MSE and the average number of used training signals. From the figure, a longer training period rapidly increases the MSE despite a certain saving in terms of the training signals used. Unfortunately, it is not straightforward to determine a proper training period as it depends on the observation quality, the channel variation speed, and the available spatial resolution. This difficulty is unlocked by the proposed switching mechanism as explained in Section III-C.

The effectiveness of the dynamic switching mechanism is evaluated in Fig. 9 using two metrics: detection rate and false alarm rate. The former measures the percentage of correctly accepting the null hypothesis \mathcal{H}_0 (i.e., the switching from beam tracking to beam training) and the latter measure the percentage of falsely rejecting \mathcal{H}_0 . Clearly, there is a tradeoff between the detection rate and the false alarm rate, as seen from the figure. A high detection rate of the proposed beam alignment scheme suggests that it can avoid error accumulation by properly switching between beam training and beam tracking. For the SNR between 15 dB and 30 dB, CS-KF maintains the detection rate of 99%. On the other hand, a low false alarm rate avoids unnecessary switching from beam tracking to beam training, thereby reducing the training overhead. As SNR increases, the false alarm rate reduces from 23% to 7%. The above results indicate that the proposed beam alignment scheme achieves a good balance between the training overhead and the detection sensitivity of beam changes.

V. CONCLUSION

A novel beam alignment scheme for a single-user time-varying mmWave channel was proposed in this paper. We showed that the proposed CS-KF can accurately track the path gains and angles using very low overhead. This is achieved by exploiting the integrated CS and KF framework previously proposed for sparse signal reconstruction. Simulation results revealed that CS-KF has a much lower MSE than using KF and CS alone given the same amount of training signals. Besides, CS-KF can approach to the

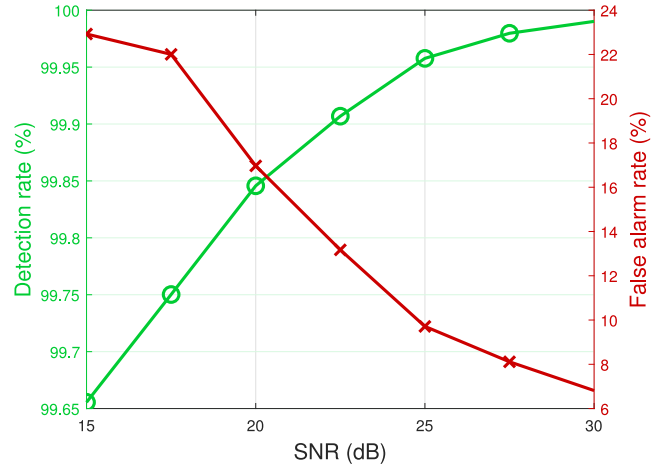


FIGURE 9. Detection rate and false alarm rate in different SNR values.

estimation lower bound at high SNR using much fewer training signals. The advantage of low overhead comes from using previous estimates obtained by KF for estimating the sparse channel using CS. Besides, the dynamic switching between beam training and beam tracking is important. The former consumes more training signals than the latter and thus, it should be used only when the estimation error of beam tracking is large. A simple switching mechanism based on hypothesis testing is employed to determine the timing of switching between beam training and beam tracking. This strategy performs better than the conventional mmWave beam alignment approach that performs beam training with a fixed period. The merits of high alignment accuracy and low overhead make the proposed beam alignment scheme attractive to mmWave communications over the time-varying channel.

APPENDIX A DERIVATION OF Σ_{FE}

Since $\Delta \in S_t^c$, (18) can be written as

$$\tilde{\mathbf{y}}_t = \mathbf{P}_{S_t^c}(\mathbf{x}_t)_{S_t^c} + \tilde{\mathbf{n}}_t \quad (\text{A.1})$$

where $\tilde{\mathbf{n}}_t$ is the noise of the innovation error in (18). Replace $\mathbf{P}_{S_t^c}(\mathbf{x}_t)_{S_t^c}$ in (10) by $\tilde{\mathbf{y}}_t - \tilde{\mathbf{n}}_t$, the filtering error can be expressed as

$$\tilde{\mathbf{y}}_{t,f} = \tilde{\mathbf{y}}_t - \tilde{\mathbf{n}}_t + \mathbf{P}_{S_t}(\mathbf{x}_t - \hat{\mathbf{x}}_t)_{S_t} + \mathbf{n}_t, \quad (\text{A.2})$$

Since $\tilde{\mathbf{n}}_t = \mathbf{P}_{S_t}(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1})_{S_t} + \mathbf{n}_t$, (A.2) can be shown as

$$\tilde{\mathbf{y}}_{t,f} = \tilde{\mathbf{y}}_t - \mathbf{P}_{S_t}(\hat{\mathbf{x}}_t - \hat{\mathbf{x}}_{t|t-1})_{S_t}. \quad (\text{A.3})$$

From (13), $(\hat{\mathbf{x}}_t - \hat{\mathbf{x}}_{t-1})_{S_t}$ is a function of the innovation error, i.e., $\hat{\mathbf{x}}_t - \hat{\mathbf{x}}_{t-1} = \mathbf{K}\tilde{\mathbf{y}}_t$. Thus, the filtering error is also a function of the innovation error as can be seen from rewritten (A.3) as

$$\tilde{\mathbf{y}}_{t,f} = [\mathbf{I} - \mathbf{P}_{S_t}\mathbf{K}]\tilde{\mathbf{y}}_t. \quad (\text{A.4})$$

As $\tilde{\mathbf{y}}_t$ is noisy, $\tilde{\mathbf{y}}_{t,f}$ is also noisy with the noise

$$\tilde{\mathbf{n}}_{t,f} = [\mathbf{I} - \mathbf{P}_{S_t}\mathbf{K}]\tilde{\mathbf{n}}_t, \quad (\text{A.5})$$

which is obtained by substituting (A.1) into (A.4). Since $\tilde{\mathbf{n}}_t$ is Gaussian distributed with zero mean, $\tilde{\mathbf{n}}_{t,f}$ is also Gaussian distributed with zero mean and covariance matrix

$$\begin{aligned}\Sigma_{fe} &= \mathbb{E}[\tilde{\mathbf{n}}_{t,f}\tilde{\mathbf{n}}_{t,f}^H] \\ &= (\mathbf{I} - \mathbf{P}_{S_t}\mathbf{K})\Sigma_{ie}(\mathbf{I} - \mathbf{P}_{S_t}\mathbf{K})^H\end{aligned}\quad (\text{A.6})$$

where Σ_{ie} is given in (13). Comparing $\tilde{\mathbf{y}}_{t,f}$ and $\tilde{\mathbf{y}}_t$, the former is less noisy [21] because (i) the change of states, $\hat{\mathbf{x}}_t - \hat{\mathbf{x}}_{t-1}$ has been subtracted as seen from (A.3), and (ii) the new component, $\mathbf{P}_{S_t^c}(\mathbf{x}_t)_{S_t^c}$, is suppressed by $\mathbf{P}_{S_t}\mathbf{K}$ as seen from (A.4).

APPENDIX B PROOF OF PROPOSITION 1

Recall that the filter error norm is defined as $\text{FEN} = \tilde{\mathbf{y}}_{t,f}^H \Sigma_{fe}^{-1} \tilde{\mathbf{y}}_{t,f}$ where $\tilde{\mathbf{y}}_{t,f} = \mathbf{y}_t - \mathbf{P}\hat{\mathbf{x}}_t$. Without loss of generality, let us drop the time index and consider the case of beam tracking. With accurate beam alignment, $\hat{\mathbf{x}} = \mathbf{x}$ and thus, $\tilde{\mathbf{y}}_f \approx \mathbf{z}$. i.e., the filtering error is dominated by the M_{track} -dimensional observation noise vector. Thus, the covariance matrix of $\tilde{\mathbf{y}}_f$ reduces to $\Sigma_{fe} = \sigma_{\text{obs}}^2 \mathbf{I}_{M_{\text{track}}}$. Consequently,

$$\begin{aligned}\text{FEN} &\approx \mathbf{z}^H \text{diag}(1/\sigma_{\text{obs}}^2) \mathbf{z} \\ &= \sum_{m=1}^{M_{\text{track}}} \frac{|z_m|^2}{\sigma_{\text{obs}}^2}\end{aligned}\quad (\text{A.7})$$

where $z_m \sim \mathcal{CN}(0, \sigma_{\text{obs}}^2)$ is the m th entry in \mathbf{z} . Since $\frac{z_m}{\sigma_{\text{obs}}^2}$ is a standard complex normal random variable, FEN is the squared sum of $2M_{\text{track}}$ standard normal random variables, which follows the chi-squared distribution with $2M_{\text{track}}$ degree of freedom. Following the same manner, FEN in the beam training stage is chi-squared distributed with $2M_{\text{train}}$ degree of freedom.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their insightful comments, which helped to improve the quality of the paper.

REFERENCES

- [1] K. Sakaguchi *et al.*, "Towards mmWave V2X in 5G and beyond to support automated driving," *IEICE Trans. Commun.*, vol. E104.B, no. 6, pp. 587–603, 2021.
- [2] "Understanding millimeter wave spectrum for 5G networks," 5G Americas, Bellevue, WA, USA, White Paper, Dec. 2020.
- [3] J. W. Choi, B. Shim, Y. Ding, B. Rao, and D. I. Kim, "Compressed sensing for wireless communications: Useful tips and tricks," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 3, pp. 1527–1550, 3rd Quart., 2017.
- [4] T. S. Rappaport, J. George, R. MacCartney, M. K. Samimi, and S. Sun, "Wideband millimeter-wave propagation measurements and channel models for future wireless communication system design," *IEEE Trans. Commun.*, vol. 63, no. 9, pp. 3029–3056, Sep. 2015.
- [5] Q. Duan and T. Kim, "AoD and AoA tracking with directional soundingbeam design for millimeter wave MIMO systems," in *Proc. IEEE Int. Symp. Personal Indoor Mobile Radio Commun. (PIMRC)*, Aug./Sep. 2015, pp. 2271–2276.
- [6] S. H. Mirfarshbafan, A. Gallyas-Sanhueza, R. Ghods, and C. Studer, "Beamspace channel estimation for massive MIMO mmwave systems: Algorithm and VLSI design," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 67, no. 12, pp. 5482–5495, Dec. 2020.
- [7] Z. Marzi, D. Ramasamy, and U. Madhoo, "Compressive channel estimation and tracking for large arrays in mm-Wave picocells," *IEEE J. Sel. Areas Commun.*, vol. 10, no. 3, pp. 514–527, Apr. 2016.
- [8] J. He, T. Kim, H. Ghauch, K. Liu, and G. Wang, "Millimeter wave MIMO channel tracking systems," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Austin, TX, USA, Dec. 2014, pp. 416–421.
- [9] Y. Han and J. Lee, "Two-stage compressed sensing for millimeter wave channel estimation," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Barcelona, Spain, 2016, pp. 860–864.
- [10] S. Sun and T. S. Rappaport, "Millimeter wave MIMO channel estimation based on adaptive compressed sensing," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, Paris, France, 2017, pp. 47–53.
- [11] 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.
- [12] P.-Y. Lai and K.-H. Liu, "Group-based multi-user tracking in mobile millimeter-wave networks," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, 2020, pp. 1–6.
- [13] C. Zhang, D. Guo, and P. Fan, "Tracking angles of departure and arrival in a mobile millimeter wave channel," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Kuala Lumpur, Malaysia, May 2016, pp. 1–6.
- [14] V. Va, H. Vikalo, and R. W. Heath, "Beam tracking for mobile millimeter wave communication systems," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Dec. 2016, pp. 743–747.
- [15] S. G. Larew and D. J. Love, "Adaptive beam tracking with the unscented Kalman filter for millimeter wave communication," *IEEE Signal Process. Lett.*, vol. 26, no. 11, pp. 1658–1662, Nov. 2019.
- [16] S. Shaham, M. Ding, M. Kokshoorn, Z. Lin, S. Dang, and R. Abbas, "Fast channel estimation and beam tracking for millimeter wave vehicular communications," *IEEE Access*, vol. 7, pp. 141104–141118, 2019.
- [17] J. Zhao, F. Gao, W. Jia, 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.
- [18] J. Palacios, D. D. Donno, and J. Widmer, "Tracking mm-Wave channel dynamics: Fast beam training strategies under mobility," in *Proc. IEEE INFOCOM*, Atlanta, GA, USA, May 2017, pp. 1–9.
- [19] S. H. Lim, S. Kim, B. Shim, and J. W. Choi, "Efficient beam training and sparse channel estimation for millimeter wave communications under mobility," *IEEE Trans. Commun.*, vol. 68, no. 10, pp. 6583–6596, Oct. 2020.
- [20] N. Ronquillo and T. Javidi, "Active beam tracking under stochastic mobility," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Montreal, QC, Canada, Jun. 2021, pp. 1–7.
- [21] N. Vaswani, "Kalman filtered compressed sensing," in *Proc. IEEE Int. Conf. Image Processing*, San Diego, CA, USA, Oct. 2008, pp. 893–896.
- [22] E. Candes and T. Tao, "The dantzig selector: Statistical estimation when p is much larger than n ," *Ann. Stat.*, vol. 35, no. 6, pp. 2313–2351, Dec. 2007.
- [23] D. Zhu, J. Choi, and R. W. Heath, "Auxiliary beam pair enabled AoD and AoA estimation in closed-loop large-scale millimeter-wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 7, pp. 4770–4785, Jul. 2017.
- [24] E. J. Candes, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted l^1 minimization," *J. Fourier Anal. Appl.*, vol. 14, nos. 5–6, pp. 877–905, 2008.
- [25] Y. Teng, L. Jia, A. Liu, and V. K. N. Lau, "A hybrid pilot beamforming and channel tracking scheme for massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 20, no. 9, pp. 6078–6092, Sep. 2021.