Fast Accurate Beam and Channel Tracking for Two-Dimensional Phased Antenna Arrays

YU LIU[1], **JIAHUI LI**[1], **XIUJUN ZHANG**[2], **(Member, IEEE), AND SHIDONG ZHOU**[1], **(Member, IEEE)**

[1]Department of EE, Tsinghua University, Beijing 100084, China
[2]National Laboratory for Information Science and Technology, Tsinghua University, Beijing 100084, China

Corresponding author: Yu Liu (liu_yu16@163.com)

ABSTRACT The sparsity and the severe attenuation of millimeter-wave (mmWave) channel imply that highly directional communication is needed. The narrow beam produced by large array requires accurate alignment, which is difficult to achieve when serving fast-moving users. In this article, we focus on accurate two-dimensional (2D) beam and channel tracking problem aiming at minimizing exploration overhead and tracking error. Using a typical frame structure with periodic exploration and communication, a proven minimum overhead of exploration is provided first. Then tracking algorithms are designed for three types of channels with different dynamic properties. It is proved that the algorithms for quasi-static channels and channels in Dynamic Case I are optimal in approaching the minimum Cramér-Rao lower bound (CRLB). The computational complexity of our algorithms is analyzed showing their efficiency, and simulation results verify their advantages in both tracking error and tracking speed.

INDEX TERMS Millimeter-wave mobile communication, beam and channel tracking, 2D phased antenna array, optimal exploring beam, Cramér-Rao lower bound.

I. INTRODUCTION

Millimeter-wave (mmWave) mobile communication is currently a hot topic due to its much wider bandwidth compared with the sub-6GHz spectrum. In mmWave channels, the much higher frequency leads to severe propagation loss, atmospheric absorption, penetration loss and other obstructions [1]. Fortunately, the shorter wavelength in the mmWave band allows the deployment of a larger antenna array, providing a considerable beamforming gain to compensate for the path loss [2]–[6]. For a hybrid or analog beamforming (ABF) system as cost-efficient ways to obtain this array gain, misalignment of beam direction may not only degrade the effective receiving power, increasing mutual interference, but also lead to the loss of beam observations due to the users' mobility, especially in fast-varying environments [7], [8].

The associate editor coordinating the review of this manuscript and approving it for publication was Antonino Orsino.

Therefore, accurate beam tracking is crucial for serving fast-moving users in mmWave mobile communication system.

In this article, we will focus on the problem in ABF. Since only one RF chain connected with the antennas via programmable phase shifters is available in ABF, only one set of phase shifts can be applied (forming a so-called **exploring beamforming vector** (**EBV**) in this article) and one dimension of the multiple-antenna channel can be observed at a certain time. Hence, in order to estimate the direction and the gain of the beam, the transceiver needs to try several different EBVs one by one. These EBVs can have a significant impact on tracking performance [9]–[11].

Although there already exist some beam tracking methods in [7], [12]–[16], which utilized historical exploring directions and observations to obtain current estimates, the EBVs were not optimized in those tracking algorithms. While beamforming resulting in the highest combining signal-to-noise ratio (SNR) is the best for data transmission, it is not the best for tracking accuracy [10], [11]. Optimal design of EBVs is

necessary to achieve as accurate beam alignment as possible in mmWave mobile communication.

In [9], a beam tracking algorithm was proposed, trying to optimize the EBVs, assuming that the channel gain is known. In [10], the authors started to jointly track the channel gain and the beam direction with optimal EBVs. In [11], the optimization of the EBVs is converted to a convex problem and solved by online optimization toolboxes. Despite the progress, only one-dimensional (1D) array is supported for beam tracking optimization in these works. However, in most mobile applications, two-dimensional (2D) arrays are necessary, not only for providing much higher array gain, but also for supporting both horizontal and vertical beam direction variation [17], [18]. It brings huge challenges when extending the optimal EBVs in [10], [11] to 2D arrays. The optimization of the exploring directions in [10] relies on the symmetrical property of the two exploring beams, however, keeping this symmetrical property may result in more overhead when extended to 2D tracking. As for the algorithm in [11], the objective function versus the EBVs becomes extremely complicated and quite different in 2D tracking compared with 1D tracking, leading to the failure of the previously used optimization method in 1D system. To the best of the authors' knowledge, there is no work on the design of optimal EBVs for 2D array yet.
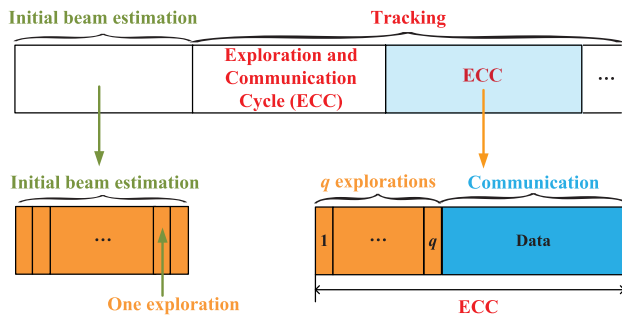

FIGURE 1. The frame structure for tracking.

In this article, we focus on the design of the optimal EBVs and the accurate single-path tracking algorithms based on 2D phased antenna array. The widely used frame structure [8], [10], [19] is adopted here. As shown in Fig. 1, the transceiver periodically works in exploration and communication mode. In the exploration stage of each **exploration and communication cycle (ECC)**, the transmitter sends a pre-defined pilot sequence for q times. At each time, the receiver forms one exploring beam pointing in one direction to observe the channel. Then the channel gain and the direction of the incoming beam are estimated with the q observations of the channel. In the communication stage of each ECC, the beam is aligned in the current estimated direction, and the current estimated channel gain will be used for the subsequent process. Based on this structure, the following questions are to be answered:

1) What is the minimum exploration overhead q in each ECC for 2D tracking?
2) How to determine the q exploring directions based on the channel observations in previous ECCs for 2D tracking?
3) How to track the 2D beam direction and the channel gain for different time-varying channels, e.g., from quasi-static channels to fast-fading channels?
4) How is the accuracy, convergence and stability of the tracking algorithm?

Following these questions, we summarize the main contributions of this article as below:

1) Based on a reasonable EBV constraint, it is proved that the minimum exploration overhead counted by the number of exploring directions is $q = 3$, for a unique solution of the 2D beam direction and the channel gain within only one ECC, while simple extension from 1D to 2D tracking will need $q = 4$.

2) Dynamic beam and channel tracking strategies for three different time-varying channels (called **Quasi-static Case**, **Dynamic Case I** and **Dynamic Case II** in this article) are proposed and optimized. The salient advantages of these tracking algorithms are given below:

i) In Quasi-static Case (channels with quasi-static beam direction and channel gain), the optimal exploration offsets are derived. Also, a joint beam direction and channel gain tracking algorithm is proposed, and the tracking error is proved to converge to the minimum **Cramér-Rao lower bound** (**CRLB**).

ii) In Dynamic Case I (channels with quasi-static beam direction and fast-fading channel gain), the Rayleigh fading channel is studied as a special case in this article. The optimal exploration offsets are obtained and an algorithm for beam (only) tracking is proposed, which is proved to converge and achieve the minimum CRLB on the beam direction.

iii) In Dynamic Case II (channels with fast-changing beam direction and channel gain), a joint tracking algorithm of the beam direction and the channel gain is proposed with faster and more accurate performance.

3) The impact of the antenna pattern on the tracking algorithms and the performance is taken into account, showing that the proposed algorithms are suitable for practical implementations.

Part of this work was presented in our conference paper [20], while the main difference and novelty of this article lies in the following four aspects: 1) tracking with a general antenna pattern rather than a simple isotropic pattern: we consider a more general direction-dependent antenna element pattern here rather than an isotropic pattern; 2) tracking for different types of time-varying channels: in addition to slow-fading channels, fast-fading channels are also studied in this article and the corresponding tracking strategy is proposed and optimized, while the prior algorithm does not support fast-fading channels; 3) the complexity analysis: we analyze the computational complexity of the proposed tracking algorithms while this was missing in our previous work; 4) more

rigorous conclusions and more complete proofs: we provide more rigorous lemmas and theorems with more complete proofs in this article, correcting the corresponding flaws in the previous work.

The remaining part of this article is organized as follows: the system model is described in Section II. In Section III, the tracking problem with some constraints is formulated. Then the minimum exploration overhead of joint 2D beam and channel tracking is given in theory in Section IV. In Section V and Section VI, the tracking problems for Quasi-static Case (Section V) and Dynamic Case I (Section VI) are studied separately. The tracking performance bounds are derived and corresponding tracking algorithms are developed with convergence and optimality analysis. In Section VII, a tracking algorithm is developed for Dynamic Case II. Then the complexity analysis of these algorithms is given in Section VIII. Section IX presents numerical results to verify the performance of our proposed algorithms.

Notations: We use lower case letters such as a and $\mathbf{a}$ to denote scalars and column vectors. Respectively, $|\mathbf{a}|$ and $\|\mathbf{a}\|_2$ represent the modulus and 2-norm of the vector $\mathbf{a}$. Upper case boldface letters, e.g., $\mathbf{A}$, are used to denote matrices. The superscript $(\bar{\cdot})$, $(\cdot)^{\mathrm{T}}$, $(\cdot)^{\mathrm{H}}$ are utilized to denote conjugate, transpose and conjugate-transpose. For a matrix $\mathbf{A}$, its inverse, pseudo-inverse and determinant are written as $\mathbf{A}^{-1}$, $\mathbf{A}^{+}$ and $|\mathbf{A}|$. The identity matrix of order q is denoted by $\mathbf{J}_q$. Let $\mathcal{CN}(\mu, \sigma^2)$ represent the symmetric complex Gaussian distribution with mean μ and variance σ^2, and $\mathcal{N}(\mu, \sigma^2)$ stand for the real Gaussian distribution with mean μ and variance σ^2. The Kronecker product is represented as $\otimes$. The statistical expectation is denoted by $\mathbb{E}[\cdot]$. The real (imaginary) part is represented as $\mathrm{Re}\{\cdot\}$ $(\mathrm{Im}\{\cdot\})$. The natural logarithm of a scalar y is obtained by $\log(\cdot)$ and the phase angle of a complex number z is written as $\angle z$. The main acronyms used in this article are summarized in TABLE 1.

TABLE 1. Summary of the main acronyms.

Cramér-Rao lower bound	CRLB
analog beamforming	ABF
exploring beamforming vector	EBV
exploration and communication cycle	ECC
direction parameter vector	DPV
exploring beamforming matrix	EBM
joint beam and channel tracking	JBCT
recursive beam tracking	RBT

II. SYSTEM MODEL

A. SYSTEM CONFIGURATION

We consider a mmWave receiver[1] equipped with a planar phased antenna array, as shown in Fig. 2. The planar array

---

[1] Note that tracking is needed at both the transmitter and the receiver. However, considering the transmitter-receiver reciprocity, the tracking of both sides have similar designs. Hence, we focus on beam and channel tracking at the receiver side.
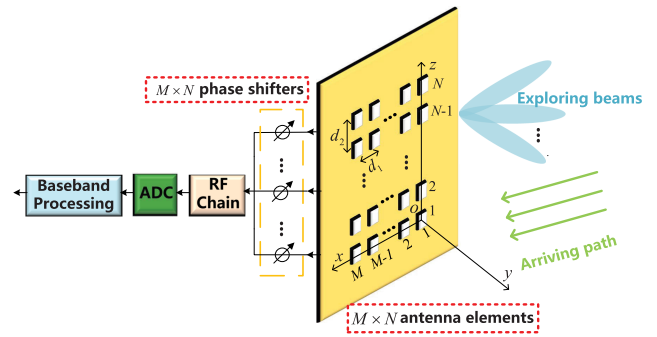

FIGURE 2. MmWave receiver with 2D phased antenna array.

consists of $M \times N$ antenna elements that are placed in a rectangular area, where $M(N)$ antenna elements are evenly distributed along x-axis (z-axis) with a distance d_1 (d_2) between neighboring elements. These antenna elements are connected to the same RF chain via programmable phase shifters.

Single RF chain of ABF makes a constraint that only one beam can be formed at any time and hence the receiver has to work alternatively in exploration and communication mode, resulting in a frame structure of periodic ECC. The angle of arrival (AoA) and the channel gain are assumed to be constant in each ECC and may change in different ECCs. In the exploration stage of one ECC, the transmitter sends a pre-defined pilot sequence $\mathbf{s}$ for q times, where $\mathbf{s} = [s_1, \cdots, s_{L_s}] \in \mathbb{C}^{1 \times L_s}$ contains L_s same symbols. At each time, the receiver forms one exploring beam pointing in one direction to observe the channel. Then the channel gain and the direction of the incoming beam are estimated according to the q observations obtained in the current and previous ECCs. In the communication stage of each ECC, the beam is aligned in current estimated direction, and the current estimated channel gain will be used for the subsequent process.

B. CHANNEL MODEL

In mmWave outdoor communication, the scattering is not rich and the number of effective propagation paths is usually limited [1], [21]. Besides, the beam formed by a large array in the mmWave system is quite narrow and the interaction between multi-path is relatively weak [22]. In other words, the incoming paths are usually sparse in space, making it possible to track each path independently. Hence, we focus on the method for tracking one path, while different paths can be tracked separately by using the same method. In k-th ECC, the direction of the incoming beam is denoted by (θ_k, ϕ_k), where $\theta_k \in [-\frac{\pi}{2}, \frac{\pi}{2})$ is the elevation AoA and $\phi_k \in [0, \pi)$ is the azimuth AoA. Then the channel vector of this path during k-th ECC is

$$\mathbf{h}_k = \eta(\theta_k, \phi_k) \beta_k^c \mathbf{a}(\mathbf{x}_k), \tag{1}$$

where $\eta(\theta, \phi)$ is the direction-dependent antenna gain (antenna pattern) of each element,[2] β_k^c is the complex channel gain, $\mathbf{x}_k \triangleq [x_{k,1}, x_{k,2}]^T = \left[\frac{Md_1 \cos(\theta_k) \cos(\phi_k)}{\lambda}, \frac{Nd_2 \sin(\theta_k)}{\lambda} \right]^T$ is the **direction parameter vector (DPV)** determined by (θ_k, ϕ_k),

$$\mathbf{a}(\mathbf{x}_k) = \mathbf{a}_1(x_{k,1}) \otimes \mathbf{a}_2(x_{k,2}) \quad (2)$$

is the 2D steering vector with

$$\mathbf{a}_1(x_{k,1}) \triangleq \left[1, e^{j2\pi \frac{x_{k,1}}{M}}, \cdots, e^{j2\pi \frac{M-1}{M} x_{k,1}} \right]^T \quad (3)$$

$$\mathbf{a}_2(x_{k,2}) \triangleq \left[1, e^{j2\pi \frac{x_{k,2}}{N}}, \cdots, e^{j2\pi \frac{N-1}{N} x_{k,2}} \right]^T, \quad (4)$$

and λ is the wavelength. For the convenience of expression, the antenna gain $\eta(\theta_k, \phi_k)$ is denoted by $\eta(\mathbf{x}_k)$ hereinafter.

Define the equivalent channel gain in k-th ECC as below:

$$\beta(\mathbf{x}_k) \triangleq \eta(\mathbf{x}_k) \beta_k^c, \quad (5)$$

then the channel vector in (1) can be rewritten as

$$\mathbf{h}_k = \beta(\mathbf{x}_k) \mathbf{a}(\mathbf{x}_k). \quad (6)$$

C. RF AND BASEBAND PREPROCESSING

Synchronization in both carrier frequency and symbol timing is necessary in mmWave wireless communications. In the initial beam estimation stage in Fig. 1, the carrier frequency synchronization information can be obtained and estimated, the residual error of which can be converted to the time-varying phase of the equivalent channel gain in (5). As for the symbol timing, since it changes much slower, it can be estimated and tracked much more easily both in the initial beam estimation stage and the tracking stage. There already exists a lot of work on the synchronization algorithms in mmWave communication systems [23]–[25]. To make the research goals more focused, we assume perfect synchronization in this article. Future work may be needed to further study the impact of residual synchronization error on the beam tracking performance.

Next, we will focus on the receiving beamforming based on the perfect synchronization assumption above. Let $\mathbf{w}_{k,i} \in \mathbb{C}^{MN \times 1}$ be the EBV for receiving the pilot sequence $\mathbf{s}$ the i-th $(i = 1, \cdots, q)$ time in k-th ECC. The entries of $\mathbf{w}_{k,i}$ are of the same amplitude with $\left| [\mathbf{w}_{k,i}]_l \right| = \frac{1}{\sqrt{MN}}$, where $[\mathbf{w}_{k,i}]_l$ denotes the l-th element of $\mathbf{w}_{k,i}$. After phase shifting and combining, the i-th received sequence in k-th ECC at the baseband output of the RF chain is given by

$$\boldsymbol{v}_{k,i} = \beta(\mathbf{x}_k) \mathbf{w}_{k,i}^H \mathbf{a}(\mathbf{x}_k) \mathbf{s} + \boldsymbol{\zeta}_{k,i}. \quad (7)$$

where $\boldsymbol{\zeta}_{k,i} \in \mathbb{C}^{1 \times L_s}$ is the receiving noise vector.

---

[2] Note that the antenna patterns of different elements in the array may not be exactly the same and need to be carefully calibrated. However, considering that it is not the focus of this article, we assume the same patterns for all the elements here.

By match filtering on the sequence $\boldsymbol{v}_{k,i}$, the i-th observation in k-th ECC is given below:

$$y_{k,i} = \boldsymbol{v}_{k,i} \frac{\mathbf{s}^H}{|\mathbf{s}|} = \beta(\mathbf{x}_k) \mathbf{w}_{k,i}^H \mathbf{a}(\mathbf{x}_k) \mathbf{s} \frac{\mathbf{s}^H}{|\mathbf{s}|} + \boldsymbol{\zeta}_{k,i} \frac{\mathbf{s}^H}{|\mathbf{s}|}$$

$$= |\mathbf{s}| \beta(\mathbf{x}_k) \mathbf{w}_{k,i}^H \mathbf{a}(\mathbf{x}_k) + z_{k,i}, \quad (8)$$

where $z_{k,i} \triangleq \boldsymbol{\zeta}_{k,i} \frac{\mathbf{s}^H}{|\mathbf{s}|}$ is an additive noise, which is modeled as i.i.d. Gaussian distributed in this article, i.e., $z_{k,i} \sim \mathcal{CN}(0, \sigma_z^2)$. This assumption is certainly held when the receiving noise vector $\boldsymbol{\zeta}_{k,i}$ is i.i.d. Gaussian distributed. Besides, even when the noise vector $\boldsymbol{\zeta}_{k,i}$ is non-Gaussian, if the real and imaginary parts of the elements in $\boldsymbol{\zeta}_{k,i}$ are i.i.d, the observation noise $z_{k,i}$ can also be regarded as i.i.d. Gaussian distributed as long as the pilot sequence length L_s is sufficiently large, according to the central limit theorem [26].

Let $\mathbf{W}_k \triangleq [\mathbf{w}_{k,1}, \ldots, \mathbf{w}_{k,q}]$, $\mathbf{z}_k \triangleq [z_{k,1}, \ldots, z_{k,q}]^T$ and $\mathbf{y}_k \triangleq [y_{k,1}, \ldots, y_{k,q}]^T$ denote the **exploring beamforming matrix (EBM)**, the noise vector and the observation vector respectively. Then we can rewrite (8) as follows:

$$\mathbf{y}_k = |\mathbf{s}| \beta(\mathbf{x}_k) \mathbf{W}_k^H \mathbf{a}(\mathbf{x}_k) + \mathbf{z}_k. \quad (9)$$

D. TRACKING LOOP

As shown in Fig. 1, an initial estimate $\hat{\beta}_0 = \hat{\beta}_0^{re} + j\hat{\beta}_0^{im}$ and $\hat{\mathbf{x}}_0 = [\hat{x}_{0,1}, \hat{x}_{0,2}]^T$ can be obtained in the beam estimation stage. It is assumed in this article that the initial beam estimator can output an estimate $\hat{\mathbf{x}}_0$ falling within the main lobe of $\mathbf{x}_0$, i.e., $\hat{\mathbf{x}}_0 \in \mathcal{B}(\mathbf{x}_0)$, where $\mathcal{B}(\mathbf{x}_t)$ denotes the main lobe of an arbitrary DPV $\mathbf{x}_t = [x_{t,1}, x_{t,2}]^T$, given by

$$\mathcal{B}(\mathbf{x}_t) \triangleq (x_{t,1} - 1, x_{t,1} + 1) \times (x_{t,2} - 1, x_{t,2} + 1). \quad (10)$$

Then our tracking starts from this initial estimate $\hat{\mathbf{x}}_0$ to find more accurate beam directions. It is worth pointing out that the main lobe in $\mathbf{x}$ domain in (10) has been normalized to a square with twice the unit length of each side and centered at the DPV $\mathbf{x}_t$ after the transformation from the angle domain to $\mathbf{x}$ domain. Hence, the main lobe size in $\mathbf{x}$ domain remains unchanged even if the antenna size M, N scale.

In the exploration stage of k-th ECC, the receiver needs to choose an EBM $\mathbf{W}_k$ based on historical observation vectors $\mathbf{y}_1, \cdots, \mathbf{y}_{k-1}$ along with the corresponding EBMs $\mathbf{W}_1, \cdots, \mathbf{W}_{k-1}$. The new observation $\mathbf{y}_k$ can be obtained by applying $\mathbf{W}_k$. Then the estimate $\hat{\boldsymbol{\psi}}_k \triangleq [\hat{\beta}_k^{re}, \hat{\beta}_k^{im}, \hat{x}_{k,1}, \hat{x}_{k,2}]^T$ of the channel parameter vector $\boldsymbol{\psi}_k \triangleq [\text{Re}\{\beta(\mathbf{x}_k)\}, \text{Im}\{\beta(\mathbf{x}_k)\}, x_{k,1}, x_{k,2}]^T$ is obtained by using all observation vectors available and the corresponding EBMs. The whole tracking loop is given in Procedure 1 and the focus of this article lies in Step 3 and Step 6.

From a control system perspective, $\boldsymbol{\psi}_k$ is the system state, $\hat{\boldsymbol{\psi}}_k$ is the estimate of the system state, the EBM $\mathbf{W}_k$ is the control action and $\mathbf{y}_k$ is a noisy observation non-linearly determined by the system state and the control action.

Procedure 1 Tracking Loop

Input: Array size M, N and the pilot sequence $\mathbf{s}$.
Output: The estimate of the channel parameter vector $\hat{\boldsymbol{\psi}}_k$.

1: Initialize $\hat{\boldsymbol{\psi}}_0 = \left[\hat{\beta}_0^{\text{re}}, \hat{\beta}_0^{\text{im}}, \hat{x}_{0,1}, \hat{x}_{0,2} \right]^{\text{T}}$;
2: **for** $k = 1, 2, \cdots$ **do**
3:      Calculate $\mathbf{W}_k$ based on $\hat{\boldsymbol{\psi}}_0$, $\mathbf{W}_1, \cdots, \mathbf{W}_{k-1}$,
     $\mathbf{y}_1, \cdots, \mathbf{y}_{k-1}$;
4:      Apply $\mathbf{W}_k$ in the exploring stage of k-th ECC;
5:      Obtain the observation vector $\mathbf{y}_k$ in k-th ECC;
6:      Estimate $\hat{\boldsymbol{\psi}}_k$ based on $\hat{\boldsymbol{\psi}}_0$, $\mathbf{W}_1, \cdots, \mathbf{W}_k$,
     $\mathbf{y}_1, \cdots, \mathbf{y}_k$;
7:      Point to $\hat{\mathbf{x}}_k$ in the communication stage of k-th ECC;
8:      use $\hat{\beta}_k$ for receiving in the communication stage of
     k-th ECC.
9: **end for**

Hence, the task of a tracking design is to find the following strategy:

$$\mathbf{W}_k = \mathbf{F}_k^c \left(\hat{\boldsymbol{\psi}}_0, \mathbf{W}_1, \cdots, \mathbf{W}_{k-1}, \mathbf{y}_1, \cdots, \mathbf{y}_{k-1} \right) \quad (11)$$

$$\hat{\boldsymbol{\psi}}_k = \mathbf{F}_k^e \left(\hat{\boldsymbol{\psi}}_0, \mathbf{W}_1, \cdots, \mathbf{W}_k, \mathbf{y}_1, \cdots, \mathbf{y}_k \right), \quad (12)$$

where $\mathbf{F}_k^c$ denotes the control function and $\mathbf{F}_k^e$ denotes the estimation function in k-th ECC.

III. PROBLEM FORMULATION

Let $\Xi_k = \{ \mathbf{F}_k^c, \mathbf{F}_k^e \}$ denote the set of beam and channel tracking schemes in k-th ECC. Then the optimal beam and channel tracking problem minimizing the mean square error (MSE) of the channel vector estimate is formulated as:

$$\min_{\Xi_k} \frac{1}{MN} \mathbb{E} \left[\left\| \hat{\mathbf{h}}_k - \mathbf{h}_k \right\|_2^2 \right] \quad (13)$$

$$\text{s.t.} \ \mathbb{E} \left[\hat{\mathbf{h}}_k \right] = \mathbf{h}_k, \quad (14)$$

$$(9), (11), (12),$$

where the constraint (14) ensures that $\hat{\mathbf{h}}_k \triangleq \hat{\beta}_k \mathbf{a}(\hat{\mathbf{x}}_k)$ is an unbiased estimate of the channel vector $\mathbf{h}_k = \beta(\mathbf{x}_k) \mathbf{a}(\mathbf{x}_k)$. It is worth explaining the following two points. First, an unbiased estimator may not be the best estimator that achieves the minimum MSE. Nevertheless, such an optimal estimator with no constraints is hard to obtain and hence we add this unbiasedness constraint. Second, we only need to guarantee the unbiasedness of $\hat{\mathbf{h}}_k$, as the objective function in (13) is the MSE of the channel vector. The estimate of the equivalent channel gain and the DPV, i.e., $\hat{\beta}_k$ and $\hat{\mathbf{x}}_k$, can be biased.

Problem (13) is challenging to be solved optimally due to the following reasons:

1) It is a partially observed Markov decision process (POMDP) which generally has not been solved optimally [27], [28].

2) There are $M \times N$ phase shifts to adjust in each EBV $\mathbf{w}_{k,i}$. This makes the optimization of the EBV too complicated due to the joint design of so many phase shifts, especially when the array size $M \times N$ grows large.

3) To obtain $\hat{\boldsymbol{\psi}}_k$ in k-th ECC, k EBMs, i.e., $\mathbf{W}_1, \cdots, \mathbf{W}_k$, need to be designed, making it difficult to optimize so many beamforming matrices simultaneously as k increases.

4) The time-varying features of the channel vector in (6) restrict the tracking algorithm and the system performance. Thus, it is hard to design an optimal tracking method for a general channel model.

These challenges above make it extremely difficult to solve this problem optimally. Hence, we add some reasonable constraints in this article to take the first step of the optimal tracking policy:

A. THE EBV CONSTRAINT

As it is complicated to obtain the optimal $M \times N$ phase shifts in general for each EBV, we use steering vectors to design the EBVs,

$$\mathbf{w}_{k,i} = \frac{1}{\sqrt{MN}} \mathbf{a}(\boldsymbol{\omega}_{k,i}), \quad (15)$$

where $\boldsymbol{\omega}_{k,i} \triangleq [\omega_{k,i1}, \omega_{k,i2}]^{\text{T}}$ denotes the i-th **exploring direction vector** in k-th ECC. This ensures that only two variables need to be designed for each EBV.

B. THE EXPLORING DIRECTION CONSTRAINT

Although the exploring direction vector $\boldsymbol{\omega}_{k,i}$ in (15) can be of any form, however, considering the tracking accuracy, it is better to make sure that $\boldsymbol{\omega}_{k,i}$ falls within the main lobe of the DPV $\mathbf{x}_k$ in (10). Thus, it is reasonable to choose exploring directions near the recently estimated direction $\hat{\mathbf{x}}_{k-1}$. For this purpose, we use such an architecture in this article. That is, the i-th exploring direction vector in k-th ECC, i.e., $\boldsymbol{\omega}_{k,i}$, is determined by the previous estimate of the DPV plus an **exploration offset** $\boldsymbol{\Delta}_{k,i}$. Considering the design of the offsets that change in different ECCs is also very complicated, we adopt fixed exploration offsets $\boldsymbol{\Delta}_i (i = 1, \cdots, q)$ in this article:

$$\boldsymbol{\omega}_{k,i} = \hat{\mathbf{x}}_{k-1} + \boldsymbol{\Delta}_i, \quad i = 1, \cdots, q. \quad (16)$$

Therefore, the EBV in (15) can be rewritten as

$$\mathbf{w}_{k,i} = \frac{1}{\sqrt{MN}} \mathbf{a}(\hat{\mathbf{x}}_{k-1} + \boldsymbol{\Delta}_i), \quad i = 1, \cdots, q. \quad (17)$$

C. THE TIME-VARYING CHANNEL CONSTRAINT

The time-varying channel vector in (1) is determined by three parts: the antenna gain $\eta(\mathbf{x}_k)$, the channel gain β_k^c and the DPV $\mathbf{x}_k$. Since the change of the antenna gain $\eta(\mathbf{x}_k)$ depends on the DPV $\mathbf{x}_k$ for a given antenna element pattern, we only consider the change of the DPV $\mathbf{x}_k$ and the channel gain β_k^c when exploring the properties of the time-varying channels. As the user motion characteristics can be quite different in various situations [29]–[31], both of the DPV $\mathbf{x}_k$ and the channel gain β_k^c may change slowly or fast. Therefore, four possible cases exist, which correspond to four different practical scenarios and can be modeled as follows:

- **Quasi-static Case**: $\mathbf{x}_k \approx \mathbf{x}, \beta_k^c \approx \beta^c$
  When both $\mathbf{x}_k$ and β_k^c change slowly, e.g., the user keeps static or quasi-static in a room, the antenna gain $\eta(\mathbf{x}_k)$ and the equivalent channel gain $\beta(\mathbf{x}_k)$ defined in (5) also change slowly. The channel in this case can be seen as approximately fixed. For the sake of convenience, we assume that $\beta(\mathbf{x}_k) = \beta = \beta^{\mathrm{re}} + j\beta^{\mathrm{im}}, \mathbf{x}_k = \mathbf{x} = [x_1, x_2]^{\mathrm{T}}$ in this case.

- **Dynamic Case**: $\mathbf{x}_k \approx \mathbf{x}, \beta_{k+1}^c \neq \beta_k^c$
  For channels that $\mathbf{x}_k$ changes slowly while β_k^c changes fast, e.g., a person walks at a fast pace in a room, the beam direction can be seen as approximately fixed, i.e., $\mathbf{x}_k = \mathbf{x}$ [29], [30]. To distinguish from other dynamic scenarios, this case is called **Dynamic Case I**.

- **Dynamic Case**: $\mathbf{x}_{k+1} \neq \mathbf{x}_k, \beta_k^c \approx \beta^c$
  This case requires that the beam direction changes fast while the channel gain keeps static or quasi-static. However, in real mmWave channels, the fast change of the beam direction usually leads to the fast change of the channel gain since the propagation paths change. This case exists only when the user rotates around the base station (BS) exactly in a circle in line of sight (LOS) channels. This is not the usual case and not studied in this article.

- **Dynamic Case**: $\mathbf{x}_{k+1} \neq \mathbf{x}_k, \beta_{k+1}^c \neq \beta_k^c$
  Both the beam direction and the channel gain in this case change fast, which happens in most fast-moving scenarios except Dynamic Case I, e.g., an unmanned aerial vehicle (UAV) flies in the sky [31]. To distinguish from Dynamic Case I, we call it **Dynamic Case II**.

It would be helpful to explain the following two aspects. First of all, extra algorithms need to be introduced to efficiently classify the channels according to the time-varying features. However, to make the research goals more focused in this article, we leave the details of this classification in future work. Second, this article only exploits the independent variation properties of β_k^c and $\mathbf{x}_k$ in these four cases to obtain theoretical results. While in real mmWave channels, the variation of the channel gain β_k^c and the DPV $\mathbf{x}_k$ might be interrelated [32], which are supposed to be jointly taken into account in future work.

With the above-mentioned **EBV constraint**, **the exploring direction constraint** and **the time-varying channel constraint**, the beam and channel tracking problem in (13) can be reformulated as:

$$\min_{\Xi} \frac{1}{MN} \mathbb{E}\left[\left\| \hat{\mathbf{h}}_k - \mathbf{h}_k \right\|_2^2 \right] \quad (18)$$
$$\text{s.t. } (9), (11), (12), (14), (17).$$

IV. HOW MANY EXPLORATIONS ARE NEEDED IN EACH ECC?
Before delving into the detailed tracking process in (18), we will first study the number of explorations needed in this section.

To estimate $\boldsymbol{\psi}_k$, sufficient measurements from different exploring directions are required. For Quasi-static Case where $\boldsymbol{\psi}_k$ remains unchanged, i.e., $\boldsymbol{\psi}_k = \boldsymbol{\psi} \triangleq [\beta^{\mathrm{re}}, \beta^{\mathrm{im}}, x_1, x_2]^{\mathrm{T}}$, one exploration in each ECC is enough since sufficient measurements are available after quite a number of ECCs. Nevertheless, in dynamic case, only using one exploration in each ECC does not work well as $\boldsymbol{\psi}_k$ may change fast. Hence, it is necessary to ensure that the estimate can be obtained even by using the explorations in a single ECC. Then the question becomes: under the condition above, how many explorations are needed in each ECC?

With the constraint in (15), two explorations in each ECC are sufficient to jointly track the equivalent channel gain and the 1D beam direction according to [10]. When tracking the 2D direction, it is straight forward that four explorations are feasible by separately using two explorations to track each dimension of the 2D direction. However, using four explorations will lower the system efficiency since it will cost time resources for each exploration. Hence, we may ask that can we reduce the times of exploration, or what is the minimum number of explorations required?

Then the following lemma is proposed to help determine the minimum exploration overhead q in each ECC:

Lemma 1: If the EBVs are of the steering vector forms, i.e., $\mathbf{w}_{k,i} = \frac{1}{\sqrt{MN}} \mathbf{a}(\boldsymbol{\omega}_{k,i})$, and the observation vector in (9) is noiseless, then

1) to obtain the unique solution of the channel parameter vector $\boldsymbol{\psi}_k$ within one ECC, the minimum exploration overhead is $q = 3$ in each ECC;

2) to obtain the unique solution of the DPV $\mathbf{x}_k$ within one ECC, the minimum exploration overhead is $q = 3$ in each ECC.

Proof: See Appendix A. ∎

Lemma 1 reveals that it is impossible to obtain the unique solution within one ECC when only using two explorations, whether we want to jointly estimate $\beta(\mathbf{x}_k)$ and $\mathbf{x}_k$ or just estimate $\mathbf{x}_k$. If we use three explorations and design three appropriate exploring directions in each ECC, then the unique solution of the channel parameter vector $\boldsymbol{\psi}_k$ can be obtained. Hence, we set $q = 3$ in this article, i.e., the EBM $\mathbf{W}_k = [\mathbf{w}_{k,1}, \mathbf{w}_{k,2}, \mathbf{w}_{k,3}]$.

V. QUASI-STATIC TRACKING: PERFORMANCE BOUND, CONVERGENCE AND OPTIMALITY
In this section, we will focus on Quasi-static Case. As mentioned in Section III, in Quasi-static Case, $\boldsymbol{\psi}_k = \boldsymbol{\psi} = [\beta^{\mathrm{re}}, \beta^{\mathrm{im}}, x_1, x_2]^{\mathrm{T}}$ and $\mathbf{h}_k = \mathbf{h} \triangleq \beta \mathbf{a}(\mathbf{x})$. For a given channel parameter vector $\boldsymbol{\psi}$ and EBM $\mathbf{W}_k$, the observation vector satisfies normal distribution with $\mathbf{y}_k \sim \mathcal{CN}(|\mathbf{s}|\beta \mathbf{W}_k^{\mathrm{H}} \mathbf{a}(\mathbf{x}), \sigma_z^2 \mathbf{J}_3)$. Hence, the conditional probability density function of $\mathbf{y}_k$ is given by

$$p_S(\mathbf{y}_k | \boldsymbol{\psi}, \mathbf{W}_k) = \frac{1}{\pi^3 \sigma_z^6} e^{-\frac{\left\| \mathbf{y}_k - |\mathbf{s}|\beta \mathbf{W}_k^{\mathrm{H}} \mathbf{a}(\mathbf{x}) \right\|_2^2}{\sigma_z^2}}. \quad (19)$$

In this section, we will first provide the lower bound of the tracking error in Quasi-static Case. Then we develop a tracking algorithm and prove it can converge to the minimum CRLB with time.

A. CRAMÉR-RAO LOWER BOUND OF THE TRACKING ERROR

The Cramér-Rao lower bound theory gives the lower bound of the unbiased estimation error [33]. Based on this, we introduce the following lemma to obtain the lower bound of the tracking error in Quasi-static Case:

Lemma 2: In Quasi-static Case, given $\mathbf{W}_1, \cdots, \mathbf{W}_k$, the MSE of the channel vector estimate in (18) *is lower bounded as follows:*

$$\frac{1}{MN} \mathbb{E}\left[\left\|\hat{\mathbf{h}}_k - \mathbf{h}\right\|_2^2\right]$$
$$\geq \frac{1}{MN} \mathrm{Tr}\left\{\left(\sum_{l=1}^{k} \mathbf{I}_S(\boldsymbol{\psi}, \mathbf{W}_l)\right)^{-1}\left(\mathbf{V}^H\mathbf{V}\right)\right\}$$
$$\triangleq C_S^t\left(\boldsymbol{\psi}, \mathbf{W}_1, \cdots, \mathbf{W}_k\right), \qquad (20)$$

where $\mathbf{V}$ is the Jacobian matrix given by

$$\mathbf{V} \triangleq \frac{\partial \mathbf{h}}{\partial \boldsymbol{\psi}^T} = \left[\frac{\partial \mathbf{h}}{\partial \beta^{\mathrm{re}}}, \frac{\partial \mathbf{h}}{\partial \beta^{\mathrm{im}}}, \frac{\partial \mathbf{h}}{\partial x_1}, \frac{\partial \mathbf{h}}{\partial x_2}\right]$$
$$= \left[\mathbf{a}\left(\mathbf{x}\right), j\mathbf{a}\left(\mathbf{x}\right), \beta\frac{\partial \mathbf{a}\left(\mathbf{x}\right)}{\partial x_1}, \beta\frac{\partial \mathbf{a}\left(\mathbf{x}\right)}{\partial x_2}\right] \qquad (21)$$

and the Fisher information matrix $\mathbf{I}_S(\boldsymbol{\psi}, \mathbf{W}_l)$ is given by

$$\mathbf{I}_S(\boldsymbol{\psi}, \mathbf{W}_l) \triangleq \mathbb{E}\left[\frac{\partial \log p_S\left(\mathbf{y}_l|\boldsymbol{\psi}, \mathbf{W}_l\right)}{\partial \boldsymbol{\psi}} \cdot \frac{\partial \log p_S\left(\mathbf{y}_l|\boldsymbol{\psi}, \mathbf{W}_l\right)}{\partial \boldsymbol{\psi}^T}\right]$$
$$= \frac{2|\mathbf{s}|^2}{\sigma_z^2} \mathrm{Re}\left\{\mathbf{V}^H\mathbf{W}_l\mathbf{W}_l^H\mathbf{V}\right\}. \qquad (22)$$

Proof: See Appendix B. ∎

The CRLB in (20) is a function of the EBMs $\mathbf{W}_1, \ldots, \mathbf{W}_k$. Since it is hard to optimize so many EBMs simultaneously, we will first try to find a lower bound of the CRLB under the constraint (17), and later design a tracking algorithm approaching this lower bound.

Consider any tracking algorithm under the constraint (17) that can converge to the DPV $\mathbf{x}$, i.e.,

$$\lim_{k \to +\infty} \hat{\mathbf{x}}_k = \mathbf{x}. \qquad (23)$$

Then the EBM $\mathbf{W}_k$ also converges,

$$\lim_{k \to +\infty} \mathbf{W}_k = \mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3]^T, \qquad (24)$$

where $\mathbf{w}_i$ is given by

$$\mathbf{w}_i \triangleq \frac{1}{\sqrt{MN}}\mathbf{a}\left(\mathbf{x} + \boldsymbol{\Delta}_{S,i}\right), \quad i = 1, 2, 3 \qquad (25)$$

with $\{\boldsymbol{\Delta}_{S,1}, \boldsymbol{\Delta}_{S,2}, \boldsymbol{\Delta}_{S,3}\}$ denoting the fixed set of exploration offsets in Quasi-static Case. Hence, the normalized CRLB (by

multiplying k) converges as $k \to +\infty$:

$$\lim_{k \to +\infty} kC_S^t(\boldsymbol{\psi}, \mathbf{W}_1, \cdots, \mathbf{W}_k)$$
$$= \lim_{k \to +\infty} \frac{k}{MN} \mathrm{Tr}\left\{\left(\sum_{l=1}^{k} \mathbf{I}_S(\boldsymbol{\psi}, \mathbf{W}_l)\right)^{-1}\left(\mathbf{V}^H\mathbf{V}\right)\right\}$$
$$= \frac{1}{MN} \mathrm{Tr}\left\{\mathbf{I}_S(\boldsymbol{\psi}, \mathbf{W})^{-1}\left(\mathbf{V}^H\mathbf{V}\right)\right\}, \qquad (26)$$

which is a function of $\boldsymbol{\psi}$, $\mathbf{W}$ and will be denoted as $C_S(\boldsymbol{\psi}, \mathbf{W})$.

According to (26), for a given channel (direction and gain), there exists an optimal EBM, which leads to the minimum normalized CRLB as a function of the channel parameter vector $\boldsymbol{\psi}$:

$$C_S^{\min}(\boldsymbol{\psi}) = \min_{\mathbf{W}} C_S(\boldsymbol{\psi}, \mathbf{W}) = C_S(\boldsymbol{\psi}, \mathbf{W}_S^*). \qquad (27)$$

Solving problem (27) yields $\mathbf{W}_S^* = \left[\mathbf{w}_{S,1}^*, \mathbf{w}_{S,2}^*, \mathbf{w}_{S,3}^*\right]$ with

$$\mathbf{w}_{S,i}^* = \frac{1}{\sqrt{MN}}\mathbf{a}\left(\mathbf{x} + \boldsymbol{\Delta}_{S,i}^*\right), \quad i = 1, 2, 3, \qquad (28)$$

where $\{\boldsymbol{\Delta}_{S,1}^*, \boldsymbol{\Delta}_{S,2}^*, \boldsymbol{\Delta}_{S,3}^*\}$ denotes the optimal set of exploration offsets for a given array size and a given $\boldsymbol{\psi}$.

B. ASYMPTOTICALLY OPTIMAL SET OF EXPLORATION OFFSETS

In general, the minimum CRLB in (27) is a function of a set of system parameters including the equivalent channel gain β, the DPV $\mathbf{x}$ and the array size M, N. Hence, the optimal set of 2D exploration offsets should also be a function of these parameters. Since it is very hard to obtain the expression of this optimal set, we adopt numerical search to deal with this issue. However, as many parameters in (27) may affect the optimal result, numerical search has to be reconducted for different parameter sets, resulting in high complexity.

Fortunately, through our investigation, some useful properties of the minimum CRLB and the optimal set of exploration offsets are given to simplify the numerical search, as described in the following lemma:

Lemma 3: In Quasi-static Case, the minimum CRLB $C_S^{\min}(\boldsymbol{\psi})$ and the optimal set of exploration offsets $\{\boldsymbol{\Delta}_{S,1}^*, \boldsymbol{\Delta}_{S,2}^*, \boldsymbol{\Delta}_{S,3}^*\}$ have the following three properties:

1) $C_S^{\min}(\boldsymbol{\psi})$, $\{\boldsymbol{\Delta}_{S,1}^*, \boldsymbol{\Delta}_{S,2}^*, \boldsymbol{\Delta}_{S,3}^*\}$ are invariant to the equivalent channel gain β;

2) $C_S^{\min}(\boldsymbol{\psi})$, $\{\boldsymbol{\Delta}_{S,1}^*, \boldsymbol{\Delta}_{S,2}^*, \boldsymbol{\Delta}_{S,3}^*\}$ are invariant to the DPV $\mathbf{x}$;

3) $C_S^{\min}(\boldsymbol{\psi})$ converges as M, $N \to +\infty$ and there exists a fixed set of exploration offsets that are unrelated to the array size M, N, denoted as $\left\{\widetilde{\boldsymbol{\Delta}}_{S,1}^*, \widetilde{\boldsymbol{\Delta}}_{S,2}^*, \widetilde{\boldsymbol{\Delta}}_{S,3}^*\right\}$, such that

$$\lim_{M,N \to +\infty} C_S(\boldsymbol{\psi}, \widetilde{\mathbf{W}}_S^*) = \lim_{M,N \to +\infty} C_S^{\min}(\boldsymbol{\psi}),$$

where $\widetilde{\mathbf{W}}_S^* = [\tilde{\mathbf{w}}_{S,1}^*, \tilde{\mathbf{w}}_{S,2}^*, \tilde{\mathbf{w}}_{S,3}^*]$ is obtained with

$$\tilde{\mathbf{w}}_{S,i}^* \triangleq \frac{1}{\sqrt{MN}}\mathbf{a}\left(\mathbf{x} + \widetilde{\boldsymbol{\Delta}}_{S,i}^*\right), \quad i = 1, 2, 3. \qquad (29)$$

Proof: See Appendix C. ∎

Lemma 3 reveals that $\left\{\boldsymbol{\Delta}_{S,1}^*, \boldsymbol{\Delta}_{S,2}^*, \boldsymbol{\Delta}_{S,3}^*\right\}$ is only related to the array size M, N. Hence, the numerical search times can be reduced to one for a particular array size M, N. Numerically, we find later that even if $\left\{\widetilde{\boldsymbol{\Delta}}_{S,1}^*, \widetilde{\boldsymbol{\Delta}}_{S,2}^*, \widetilde{\boldsymbol{\Delta}}_{S,3}^*\right\}$ may change for different array sizes, $\left\{\widetilde{\boldsymbol{\Delta}}_{S,1}^*, \widetilde{\boldsymbol{\Delta}}_{S,2}^*, \widetilde{\boldsymbol{\Delta}}_{S,3}^*\right\}$ can be used to take the place of $\left\{\boldsymbol{\Delta}_{S,1}^*, \boldsymbol{\Delta}_{S,2}^*, \boldsymbol{\Delta}_{S,3}^*\right\}$ as long as M and N are sufficiently large. Therefore, the numerical search times is reduced to one in the end. As $\left\{\widetilde{\boldsymbol{\Delta}}_{S,1}^*, \widetilde{\boldsymbol{\Delta}}_{S,2}^*, \widetilde{\boldsymbol{\Delta}}_{S,3}^*\right\}$ can be used to achieve the minimum CRLB when $M, N \to +\infty$, it is called **the asymptotically optimal set of exploration offsets** in Quasi-static Case in this article.

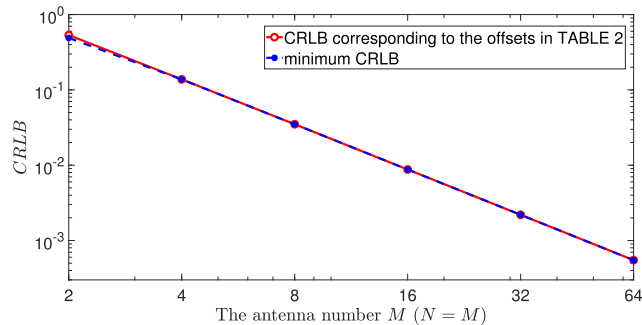**TABLE 2.** The asymptotically optimal set of exploration offsets in Quasi-static Case.

$\widetilde{\boldsymbol{\Delta}}_{S,1}^*$	$\widetilde{\boldsymbol{\Delta}}_{S,2}^*$	$\widetilde{\boldsymbol{\Delta}}_{S,3}^*$
$[-0.0963, 0.5098]^{\mathrm{T}}$	$[-0.2906, -0.2906]^{\mathrm{T}}$	$[0.5098, -0.0963]^{\mathrm{T}}$
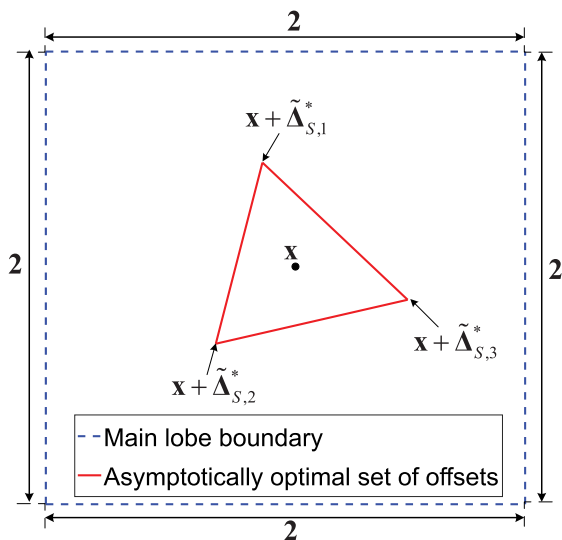
FIGURE 3. The asymptotically optimal set of exploration offsets in Quasi-static Case.

By numerical search in the main lobe in (10), we can obtain one asymptotically optimal set of exploration offsets $\left\{\widetilde{\boldsymbol{\Delta}}_{S,1}^*, \widetilde{\boldsymbol{\Delta}}_{S,2}^*, \widetilde{\boldsymbol{\Delta}}_{S,3}^*\right\}$ in TABLE 2 and Fig. 3. It can be seen that the three exploring direction vectors do not form a regular triangle as the radiation pattern produced by (25) is not isotropic from different angles. With this set in TABLE 2, a general way to generate the EBM $\widetilde{\mathbf{W}}_S^*$ is obtained by (29).

The set of exploration offsets $\left\{\widetilde{\boldsymbol{\Delta}}_{S,1}^*, \widetilde{\boldsymbol{\Delta}}_{S,2}^*, \widetilde{\boldsymbol{\Delta}}_{S,3}^*\right\}$ may become sub-optimal when the antenna size $M \times N$ is finite. To evaluate the robustness of this set of exploration offsets to finite array size, we adopt $\left\{\widetilde{\boldsymbol{\Delta}}_{S,1}^*, \widetilde{\boldsymbol{\Delta}}_{S,2}^*, \widetilde{\boldsymbol{\Delta}}_{S,3}^*\right\}$ to antenna arrays of limited size and compare the minimum CRLB with the CRLB achieved by $\left\{\widetilde{\boldsymbol{\Delta}}_{S,1}^*, \widetilde{\boldsymbol{\Delta}}_{S,2}^*, \widetilde{\boldsymbol{\Delta}}_{S,3}^*\right\}$ in TABLE 2. As illustrated in Fig. 4, when the antenna number

FIGURE 4. The performance of the offsets in TABLE 2.

$M = N \geq 8$, we can approach the minimum CRLB with a relative error less than 0.1% by using $\left\{\widetilde{\boldsymbol{\Delta}}_{S,1}^*, \widetilde{\boldsymbol{\Delta}}_{S,2}^*, \widetilde{\boldsymbol{\Delta}}_{S,3}^*\right\}$.

As a conclusion, it is practical to apply this asymptotically optimal set of exploration offsets to any antenna array with $M = N \geq 8$, any channel gain and any direction.

C. JOINT BEAM AND CHANNEL TRACKING

In the above subsections, we have provided a low-complexity numerical method to design the optimal exploration offsets and obtain the minimum CRLB, given that the DPV $\mathbf{x}$ is known. However, in a real tracking problem, the DPV $\mathbf{x}$ is unknown and the EBMs need to be adjusted dynamically. In addition, a sequence of optimal beamforming matrices only tells us what the minimum CRLB is, but it cannot tell us which tracking algorithm can achieve the minimum CRLB. In this subsection, we propose a specific tracking algorithm to approach the minimum CRLB.

The proposed tracker is motivated by the following maximum likelihood problem:

$$\max_{\mathbf{W}_k}\left\{\max_{\hat{\boldsymbol{\psi}}_k} \log p\left(\mathbf{y}_1, \cdots, \mathbf{y}_k \middle| \boldsymbol{\psi}, \mathbf{W}_1, \cdots, \mathbf{W}_k\right)\bigg|_{\boldsymbol{\psi}=\hat{\boldsymbol{\psi}}_k}\right\} \quad (30)$$

s.t. (9), (11), (12), (14), (17).

Since $\mathbf{y}_1, \cdots, \mathbf{y}_k$ are independently observed vectors, we can convert (30) as follows:

$$\max_{\mathbf{W}_k}\left\{\max_{\hat{\boldsymbol{\psi}}_k} \sum_{l=1}^{k}\left[\log p_S\left(\mathbf{y}_l \middle| \boldsymbol{\psi}, \mathbf{W}_l\right)\bigg|_{\boldsymbol{\psi}=\hat{\boldsymbol{\psi}}_k}\right]\right\} \quad (31)$$

s.t. (9), (11), (12), (14), (17).

This problem is somewhat similar to a kind of estimation problem with control in [34, Section 10.2], where a two-layer recursive algorithm is proposed and proved to converge to the minimum CRLB under several requirements. The difference is that the observation vector in [34, Section 10.2] is given by $\mathbf{y} = \mathbf{f}^o(\boldsymbol{\psi}, \mathbf{W}) + \mathbf{z}$, where $\mathbf{z}$ is an i.i.d. Gaussian noise vector and $\mathbf{f}^o(\boldsymbol{\psi}, \mathbf{W})$ is a convex function with respect to $\boldsymbol{\psi}$. However, in our problem, $\mathbf{f}^o(\boldsymbol{\psi}, \mathbf{W}) = |s|\beta\mathbf{W}^{\mathrm{H}}\mathbf{a}(\mathbf{x})$ is non-convex, leading to the failure of directly applying the algorithm and the theoretical results of [34, Section 10.2].

Despite this, we can still design a two-layer nested optimization algorithm inspired by [34, Section 10.2]. Furthermore, our proposed algorithm can be proved to converge to the minimum CRLB under some necessary requirements, as will be explained in Section V-D.

The proposed algorithm is based on iterative maximization in the inner layer and the outer layer of (31). In the *inner layer* of (31), we use the stochastic Newton's method [35] to update the estimate, given by

$$\hat{\boldsymbol{\psi}}_k = \hat{\boldsymbol{\psi}}_{k-1} + b_{S,k}\boldsymbol{\varsigma}_k, \tag{32}$$

where $b_{S,k}$ is the tracking step-size in Quasi-static Case, and $\boldsymbol{\varsigma}_k$ is the updating direction vector. This updating direction vector is a function of the observation vector $\mathbf{y}_k$ and the latest estimated value of the channel parameter vector $\hat{\boldsymbol{\psi}}_{k-1}$, and is defined as below:

$$\boldsymbol{\varsigma}_k \triangleq \mathbf{I}_S\left(\hat{\boldsymbol{\psi}}_{k-1}, \mathbf{W}_k\right)^{-1} \frac{\partial \log p_S\left(\mathbf{y}_k | \boldsymbol{\psi}, \mathbf{W}_k\right)}{\partial \boldsymbol{\psi}}\bigg|_{\boldsymbol{\psi}=\hat{\boldsymbol{\psi}}_{k-1}}. \tag{33}$$

And it is derived that

$$\boldsymbol{\varsigma}_k = \left(\text{Re}\left\{\hat{\mathbf{V}}_k^{\text{H}}\mathbf{W}_k\mathbf{W}_k^{\text{H}}\hat{\mathbf{V}}_k\right\}\right)^{-1} \begin{bmatrix} \text{Re}\left\{\mathbf{e}_k^{\text{H}}\left(\mathbf{y}_k - \hat{\mathbf{y}}_k\right)\right\} \\ \text{Im}\left\{\mathbf{e}_k^{\text{H}}\left(\mathbf{y}_k - \hat{\mathbf{y}}_k\right)\right\} \\ \text{Re}\left\{\tilde{\mathbf{e}}_{k1}^{\text{H}}\left(\mathbf{y}_k - \hat{\mathbf{y}}_k\right)\right\} \\ \text{Re}\left\{\tilde{\mathbf{e}}_{k2}^{\text{H}}\left(\mathbf{y}_k - \hat{\mathbf{y}}_k\right)\right\} \end{bmatrix}, \tag{34}$$

where $\mathbf{e}_k = \mathbf{W}_k^{\text{H}}\mathbf{a}\left(\hat{\mathbf{x}}_{k-1}\right)$, $\hat{\mathbf{y}}_k = |\mathbf{s}|\hat{\beta}_{k-1}\mathbf{W}_k^{\text{H}}\mathbf{a}\left(\hat{\mathbf{x}}_{k-1}\right)$, $\tilde{\mathbf{e}}_{k1} = \hat{\beta}_{k-1}\mathbf{W}_k^{\text{H}}\frac{\partial \mathbf{a}(\hat{\mathbf{x}}_{k-1})}{\partial x_1}$, $\tilde{\mathbf{e}}_{k2} = \hat{\beta}_{k-1}\mathbf{W}_k^{\text{H}}\frac{\partial \mathbf{a}(\hat{\mathbf{x}}_{k-1})}{\partial x_2}$ and $\hat{\mathbf{V}}_k$ is given by

$$\hat{\mathbf{V}}_k = \left[\mathbf{a}\left(\mathbf{x}\right), j\mathbf{a}\left(\mathbf{x}\right), \beta\frac{\partial \mathbf{a}\left(\mathbf{x}\right)}{\partial x_1}, \beta\frac{\partial \mathbf{a}\left(\mathbf{x}\right)}{\partial x_2}\right]\bigg|_{\boldsymbol{\psi}=\hat{\boldsymbol{\psi}}_{k-1}}. \tag{35}$$

In the outer layer of (31), assuming that the estimate of the channel parameter vector is accurate, i.e., $\boldsymbol{\psi} = \hat{\boldsymbol{\psi}}_{k-1}$, the EBM $\mathbf{W}_k = \left[\mathbf{w}_{k,1}, \mathbf{w}_{k,2}, \mathbf{w}_{k,3}\right]$ is obtained with

$$\mathbf{w}_{k,i} = \frac{1}{\sqrt{MN}}\mathbf{a}\left(\hat{\mathbf{x}}_{k-1} + \tilde{\boldsymbol{\Delta}}_{S,i}^*\right), \quad i = 1, 2, 3, \tag{36}$$

Finally, the proposed tracking algorithm is summarized in Algorithm 1.

D. ASYMPTOTIC OPTIMALITY ANALYSIS
In this subsection, the convergence and the optimality of our proposed algorithm will be discussed. Since the entire proofs are very long, we will provide the main statements and ideas or clues here, and leave the proofs in the appendices.

The convergence and optimality will be stated in three steps as follows:

i) We prove that the estimate of the proposed tracking algorithm converges to a unique point with probability one given appropriate sequences of step-sizes.

ii) We prove that if the initial estimate is within the main lobe, i.e., $\hat{\mathbf{x}}_0 \in \mathcal{B}\left(\mathbf{x}\right)$ and the step-size is appropriate, then the convergence point will be exactly the real channel parameter vector $\boldsymbol{\psi}$, with probability approaching one.

Algorithm 1 Joint Beam and Channel Tracking (JBCT) for Quasi-Static Case

1) **Exploring and Receiving** (Step 3 in Procedure 1): Transmit 3 pilot sequences in each ECC. The corresponding EBV for receiving the i-th pilot sequence in k-th ECC is given below:

$$\mathbf{w}_{k,i} = \frac{1}{\sqrt{MN}}\mathbf{a}\left(\hat{\mathbf{x}}_{k-1} + \tilde{\boldsymbol{\Delta}}_{S,i}^*\right), i = 1, 2, 3, \tag{37}$$

where $\left\{\tilde{\boldsymbol{\Delta}}_{S,1}^*, \tilde{\boldsymbol{\Delta}}_{S,2}^*, \tilde{\boldsymbol{\Delta}}_{S,3}^*\right\}$ is given by TABLE 2. After match filtering, the observation vector $\mathbf{y}_k$ is obtained via (9).

2) **Updating Estimate** (Step 6 in Procedure 1): The estimate of the channel parameter vector in k-th ECC, i.e., $\hat{\boldsymbol{\psi}}_k = \left[\hat{\beta}_k^{\text{re}}, \hat{\beta}_k^{\text{im}}, \hat{x}_{k,1}, \hat{x}_{k,2}\right]^{\text{T}}$, is updated by

$$\hat{\boldsymbol{\psi}}_k = \hat{\boldsymbol{\psi}}_{k-1} + b_{S,k}\boldsymbol{\varsigma}_k, \tag{38}$$

where $\boldsymbol{\varsigma}_k$ is the updating direction vector given by (34) and $b_{S,k}$ is the step-size that will be specified after.

---

iii) Finally, if $\hat{\boldsymbol{\psi}}_k \to \boldsymbol{\psi}$ and the step-size is appropriate, then the tracking error of our algorithm converges to the minimum CRLB.

1) CONVERGENCE TO A UNIQUE POINT
Since the observation vector $\mathbf{y}_k$ is corrupted by the Gaussian noise vector, the updating direction vector $\boldsymbol{\varsigma}_k$ in (33) is also a random vector, and can be expressed as follows:

$$\boldsymbol{\varsigma}_k = \mathbf{f}_{\boldsymbol{\psi}}\left(\hat{\boldsymbol{\psi}}_{k-1}\right) + \hat{\mathbf{z}}_k, \tag{39}$$

where $\mathbf{f}_{\boldsymbol{\psi}}\left(\hat{\boldsymbol{\psi}}_{k-1}\right)$ is the deterministic part of $\boldsymbol{\varsigma}_k$ defined as below:

$$\mathbf{f}_{\boldsymbol{\psi}}\left(\hat{\boldsymbol{\psi}}_{k-1}\right) \triangleq \mathbb{E}\left[\boldsymbol{\varsigma}_k\right], \tag{40}$$

which is a function of $\hat{\boldsymbol{\psi}}_{k-1}$ that takes $\boldsymbol{\psi}$ as a parameter vector. The zero-mean random part of $\boldsymbol{\varsigma}_k$, i.e., $\hat{\mathbf{z}}_k$, is given by

$$\hat{\mathbf{z}}_k \triangleq \boldsymbol{\varsigma}_k - \mathbf{f}_{\boldsymbol{\psi}}\left(\hat{\boldsymbol{\psi}}_{k-1}\right). \tag{41}$$

The randomness of $\boldsymbol{\varsigma}_k$ might cause the proposed algorithm to diverge. However, if we adopt the diminishing step-size as that in [34], [36], [37], i.e.,

$$b_{S,k} = \frac{\epsilon_S}{k + K_{S,0}}, \quad k = 1, 2, \cdots \tag{42}$$

where $K_{S,0} \geq 0$ and $\epsilon_S > 0$, then some convergence property can be obtained, as described in the following theorem:

Theorem 1 (Convergence to a Unique Stable Point): If we adopt the iterative method in (37), (38) and $b_{S,k}$ is given by (42) with $\epsilon_S > 0$ and $K_{S,0} \geq 0$, then $\hat{\boldsymbol{\psi}}_k$ converges to a unique stable point of $\mathbf{f}_{\boldsymbol{\psi}}(\hat{\boldsymbol{\psi}}_{k-1})$ with probability one.

A point $\hat{\boldsymbol{\psi}}_{k-1}$ is called a stable point of $\mathbf{f}_{\boldsymbol{\psi}}\left(\hat{\boldsymbol{\psi}}_{k-1}\right)$ when it satisfies two conditions: 1) $\mathbf{f}_{\boldsymbol{\psi}}\left(\hat{\boldsymbol{\psi}}_{k-1}\right) = \mathbf{0}$ and 2) $\frac{\partial \mathbf{f}_{\boldsymbol{\psi}}\left(\hat{\boldsymbol{\psi}}_{k-1}\right)}{\partial \hat{\boldsymbol{\psi}}_{k-1}^{\mathrm{T}}}$ is negative definite. Hence, the stable points set is defined as below:

$$\mathcal{S} \triangleq \left\{ \hat{\boldsymbol{\psi}}_{k-1} : \mathbf{f}_{\boldsymbol{\psi}}\left(\hat{\boldsymbol{\psi}}_{k-1}\right) = 0, \frac{\partial \mathbf{f}_{\boldsymbol{\psi}}\left(\hat{\boldsymbol{\psi}}_{k-1}\right)}{\partial \hat{\boldsymbol{\psi}}_{k-1}^{\mathrm{T}}} \prec \mathbf{0} \right\}, \quad (43)$$

where $\mathbf{A} \prec \mathbf{0}$ denotes that the matrix $\mathbf{A}$ is negative definite. In our problem, $\mathbf{f}_{\boldsymbol{\psi}}\left(\hat{\boldsymbol{\psi}}_{k-1}\right)$ defined in (40) is given by

$$\mathbf{f}_{\boldsymbol{\psi}}\left(\hat{\boldsymbol{\psi}}_{k-1}\right)$$
$$= \frac{2|\mathbf{s}|^2}{\sigma_z^2} \mathbf{I}_S\left(\hat{\boldsymbol{\psi}}_{k-1}, \mathbf{W}_k\right)^{-1} \begin{bmatrix} \mathrm{Re}\left\{ \mathbf{e}_k^{\mathrm{H}}\left(\beta \mathbf{W}_k^{\mathrm{H}} \mathbf{a}\left(\mathbf{x}\right) - \hat{\beta}_{k-1} \mathbf{e}_k\right) \right\} \\ \mathrm{Im}\left\{ \mathbf{e}_k^{\mathrm{H}}\left(\beta \mathbf{W}_k^{\mathrm{H}} \mathbf{a}\left(\mathbf{x}\right) - \hat{\beta}_{k-1} \mathbf{e}_k\right) \right\} \\ \mathrm{Re}\left\{ \tilde{\mathbf{e}}_{k1}^{\mathrm{H}}\left(\beta \mathbf{W}_k^{\mathrm{H}} \mathbf{a}\left(\mathbf{x}\right) - \hat{\beta}_{k-1} \mathbf{e}_k\right) \right\} \\ \mathrm{Re}\left\{ \tilde{\mathbf{e}}_{k2}^{\mathrm{H}}\left(\beta \mathbf{W}_k^{\mathrm{H}} \mathbf{a}\left(\mathbf{x}\right) - \hat{\beta}_{k-1} \mathbf{e}_k\right) \right\} \end{bmatrix}. \quad (44)$$

Proof of Theorem 1: See Appendix D. ∎

By Theorem 1, for the general step-size in (42), $\hat{\boldsymbol{\psi}}_k$ converges to a unique stable point in $\mathcal{S}$.

2) CONVERGENCE TO THE CHANNEL PARAMETER VECTOR $\boldsymbol{\psi}$

According to (43) and (44), it is easy to verify that the channel parameter vector $\boldsymbol{\psi}$ is a stable point by the following two points:

1) $\beta \mathbf{W}_k^{\mathrm{H}} \mathbf{a}\left(\mathbf{x}\right) = \hat{\beta}_{k-1} \mathbf{e}_k$ in (44) when $\hat{\boldsymbol{\psi}}_{k-1} = \boldsymbol{\psi}$. Hence, $\mathbf{f}_{\boldsymbol{\psi}}\left(\boldsymbol{\psi}\right) = \mathbf{0}$;

2) $\frac{\partial \mathbf{f}_{\boldsymbol{\psi}}\left(\hat{\boldsymbol{\psi}}_{k-1}\right)}{\partial \hat{\boldsymbol{\psi}}_{k-1}^{\mathrm{T}}}\Big|_{\hat{\boldsymbol{\psi}}_{k-1}=\boldsymbol{\psi}} = -\mathbf{J}_4$ by derivation, where $\mathbf{J}_4$ is the 4-order identity matrix. Thus, $\frac{\partial \mathbf{f}_{\boldsymbol{\psi}}\left(\hat{\boldsymbol{\psi}}_{k-1}\right)}{\partial \hat{\boldsymbol{\psi}}_{k-1}^{\mathrm{T}}}\Big|_{\hat{\boldsymbol{\psi}}_{k-1}=\boldsymbol{\psi}}$ is negative definite.

Therefore, $\boldsymbol{\psi}$ is a stable point, i.e., $\boldsymbol{\psi} \in \mathcal{S}$.

Other stable points in $\mathcal{S}$ correspond to the local optimal points of the beam and channel tracking problem, which are out of the main lobe $\mathcal{B}(\mathbf{x})$ in (10). Except for the channel parameter vector $\boldsymbol{\psi}$, the antenna array gain of other stable points in $\mathcal{S}$ is quite low, resulting in low tracking accuracy. Unfortunately, the estimate of the DPV $\mathbf{x}$ may jump out of the main lobe in the tracking process and converge to other local optimal points due to the existence of observation noise. Hence, one key challenge is to ensure that the tracking algorithm converges to $\boldsymbol{\psi}$ rather than other stable points. Then we develop the following theorem to deal with this challenge:

Theorem 2 (Convergence to the DPV x): If we adopt the iterative method in (37), (38) and (i) the initial estimate of $\mathbf{x}$ *is within the main lobe, i.e., $\hat{\mathbf{x}}_0 \in \mathcal{B}(\mathbf{x})$; (ii) $b_{S,k}$ is given by (42) with $\epsilon_S > 0$, then there exist some $K_{S,0} \geq 0$ and $R > 0$*

such that

$$P\left(\hat{\mathbf{x}}_k \to \mathbf{x} \mid \hat{\mathbf{x}}_0 \in \mathcal{B}\left(\mathbf{x}\right)\right) \geq 1 - 8e^{-\frac{R|\mathbf{s}|^2}{\epsilon_S^2 \sigma_z^2}}. \quad (45)$$

Proof: See Appendix E. ∎

We have assumed that the beam estimator in Fig. 1 can output an initial estimate $\hat{\mathbf{x}}_0$ within the main lobe $\mathcal{B}(\mathbf{x})$. Under the condition $\hat{\mathbf{x}}_0 \in \mathcal{B}(\mathbf{x})$, Theorem 2 tells us the probability of $\hat{\mathbf{x}}_k \to \mathbf{x}$ is related to $\frac{|\mathbf{s}|^2}{\epsilon_S^2 \sigma_z^2}$. Hence, we can reduce the step-size or increase the transmit SNR $\frac{|\mathbf{s}|^2}{\sigma_z^2}$ to make sure that $\hat{\mathbf{x}}_k \to \mathbf{x}$ approaching probability one.

According to Theorem 1, $\hat{\boldsymbol{\psi}}_k$ converges to a *unique* stable point corresponding to a local optimal point. Hence, this unique stable point will be exactly $\boldsymbol{\psi}$ when $\hat{\mathbf{x}}_k \to \mathbf{x}$, i.e., $\hat{\boldsymbol{\psi}}_k \to \boldsymbol{\psi}$.

3) CONVERGENCE WITH THE MINIMUM CRLB

Finally, the following theorem is developed to tell us the tracking error of the proposed algorithm:

Theorem 3 (Convergence to $\boldsymbol{\psi}$ With the Minimum CRLB): If we adopt the iterative method in (37), (38) and (i) $\hat{\boldsymbol{\psi}}_k \to \boldsymbol{\psi}$; (ii) $b_{S,k}$ is given by (42) with $\epsilon_S = 1$ and any $K_{S,0} \geq 0$, then $\hat{\mathbf{h}}_k - \mathbf{h}$ *is asymptotically Gaussian and*

$$\lim_{k \to +\infty} \frac{k}{MN} \mathbb{E}\left[\left\| \hat{\mathbf{h}}_k - \mathbf{h} \right\|_2^2 \Big| \hat{\boldsymbol{\psi}}_k \to \boldsymbol{\psi} \right] = C_S^{\min}(\boldsymbol{\psi}). \quad (46)$$

Proof: See Appendix F. ∎

By Theorem 1, Theorem 2 and Theorem 3, if $\hat{\mathbf{x}}_0 \in \mathcal{B}(\mathbf{x})$ and we adopt the step-size $b_{S,k}$ in (42) with $\epsilon_S = 1$ and $K_{S,0} \geq 0$, then the minimum CRLB is achieved asymptotically with high probability.

VI. RECURSIVE BEAM TRACKING FOR DYNAMIC CASE I : PERFORMANCE BOUND, CONVERGENCE AND OPTIMALITY

In Dynamic Case I, the channel gain changes fast while the beam direction changes slowly. We assume that the beam direction keeps static, i.e., $\mathbf{x}_k = \mathbf{x} = [x_1, x_2]^{\mathrm{T}}$. Hence, the antenna gain in the direction of the arriving path also keeps static, i.e., $\eta(\mathbf{x}_k) = \eta(\mathbf{x})$. When the channel gain β_k^c changes fast, it is very difficult to establish theorems of tracking the channel gain and beam direction simultaneously, as in Section V. Fortunately, acquiring the beam direction information is sufficient for alignment in mmWave mobile communication with analog beamforming. Hence, we only focus on beam direction tracking in Dynamic Case I.

The channel gains of adjacent ECCs in this section are assumed to be independent of each other. In addition, different distributions of the channel gain β_k^c can lead to different suitable tracking strategies. The tracking strategy designed for one distribution of the channel gain may deteriorate sharply when applied to other distributions. Hence, each type of channel gain distribution deserves studying, of which Rayleigh fading channel is a special case that is easier to be analyzed. This special case happens when quite a number

of rays existing in a cluster are indistinguishable. In this section, we choose Rayleigh fading channels to study for Dynamic Case I, i.e., $\beta_k^c \sim \mathcal{CN}\left(0, \left(\sigma_\beta^c\right)^2\right)$. Although the theoretical results in this section are only applicable for Rayleigh fading channels, the proposed algorithm is found robust for other types of time-varying channels according to the numerical results in Section IX-C.

When the channel gain β_k^c is Gaussian distributed, the equivalent channel gain $\beta(\mathbf{x}) = \eta(\mathbf{x}) \beta_k^c$ also satisfies Gaussian distribution with the variance given below:

$$\mathbb{E}\left[|\beta(\mathbf{x})|^2\right] = |\eta(\mathbf{x})|^2 \left(\sigma_\beta^c\right)^2 \triangleq \sigma_\beta^2. \tag{47}$$

Correspondingly, the observation vector $\mathbf{y}_k$ in (9) satisfies Gaussian distribution for a given DPV $\mathbf{x}$ and EBM $\mathbf{W}_k$, i.e., $\mathbf{y}_k \sim \mathcal{CN}\left(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{y},k}\right)$, where $\boldsymbol{\Sigma}_{\mathbf{y},k}$ is the covariance matrix of $\mathbf{y}_k$ defined as follows:

$$\boldsymbol{\Sigma}_{\mathbf{y},k} \triangleq \mathbb{E}\left[\mathbf{y}_k \mathbf{y}_k^H\right] = |\mathbf{s}|^2 \sigma_\beta^2 \mathbf{W}_k^H \mathbf{a}(\mathbf{x}) \left(\mathbf{W}_k^H \mathbf{a}(\mathbf{x})\right)^H + \sigma_z^2 \mathbf{J}_3. \tag{48}$$

According to (48), we can obtain the determinant of $\boldsymbol{\Sigma}_{\mathbf{y},k}$:

$$|\boldsymbol{\Sigma}_{\mathbf{y},k}| = \sigma_z^4 \left(\sigma_z^2 + |\mathbf{s}|^2 \sigma_\beta^2 |\mathbf{W}_k^H \mathbf{a}(\mathbf{x})|^2\right). \tag{49}$$

Then the conditional probability density function of $\mathbf{y}_k$ is given by

$$p_{DI}(\mathbf{y}_k | \mathbf{x}, \mathbf{W}_k) = \frac{1}{\pi^3 |\boldsymbol{\Sigma}_{\mathbf{y},k}|} e^{-\mathbf{y}_k^H \boldsymbol{\Sigma}_{\mathbf{y},k}^{-1} \mathbf{y}_k}, \tag{50}$$

The following structure of this section is similar to Section V: we first formulate the beam tracking problem and provide the lower bound of it. Then we develop a tracking algorithm and prove this algorithm can converge to the minimum CRLB.

A. PROBLEM FORMULATION

Since we only track the beam direction in Dynamic Case I, the estimation function in (12) is reformulated as follows:

$$\hat{\mathbf{x}}_k = \mathbf{F}_{DI,k}^e\left(\hat{\boldsymbol{\psi}}_0, \mathbf{W}_1, \cdots, \mathbf{W}_k, \mathbf{y}_1, \cdots, \mathbf{y}_k\right). \tag{51}$$

Let $\Xi_{DI,k} = \left\{\mathbf{F}_k^c, \mathbf{F}_{DI,k}^e\right\}$ denote a beam tracking scheme set in k-th ECC: based on historical observation vectors $\mathbf{y}_1, \cdots, \mathbf{y}_{k-1}$ along with the corresponding EBMs $\mathbf{W}_1, \cdots, \mathbf{W}_{k-1}$, choose an appropriate EBM $\mathbf{W}_k$, apply it to obtain $\mathbf{y}_k$ and make an estimation of the DPV $\mathbf{x}$ in k-th ECC by using all EBMs and observations available. Hence, in k-th ECC, the tracking problem is formulated as:

$$\min_{\Xi_{DI,k}} \mathbb{E}\left[\|\hat{\mathbf{x}}_k - \mathbf{x}\|_2^2\right] \tag{52}$$

$$\text{s.t. } \mathbb{E}\left[\hat{\mathbf{x}}_k\right] = \mathbf{x}, \tag{53}$$

$$(9), (11), (17), (51),$$

where the constraint (53) ensures that $\hat{\mathbf{x}}_k$ is an unbiased estimate of the DPV $\mathbf{x}$.

Before providing a specific tracking algorithm, we will first explore the performance bound of the problem in (52).

B. CRAMÉR-RAO LOWER BOUND OF TRACKING ERROR

We now perform some theoretical analysis on the beam tracking problem. Based on the CRLB theory in [33], we introduce the following lemma to obtain the lower bound of the tracking error:

Lemma 4: In Dynamic Case I, given $\mathbf{W}_1, \cdots, \mathbf{W}_k$, the MSE of the DPV estimate in (52) is lower bounded as follows:

$$\mathbb{E}\left[\|\hat{\mathbf{x}}_k - \mathbf{x}\|_2^2\right] \geq \text{Tr}\left\{\left(\sum_{l=1}^k \mathbf{I}_{DI}(\mathbf{x}, \mathbf{W}_l)\right)^{-1}\right\}, \tag{54}$$

where the Fisher information matrix $\mathbf{I}_{DI}(\mathbf{x}, \mathbf{W}_l)$ is given by

$$\mathbf{I}_{DI}(\mathbf{x}, \mathbf{W}_l) \triangleq \mathbb{E}\left[\frac{\partial \log p_{DI}(\mathbf{y}_k | \mathbf{x}, \mathbf{W}_l)}{\partial \mathbf{x}} \cdot \frac{\partial \log p_{DI}(\mathbf{y}_k | \mathbf{x}, \mathbf{W}_l)}{\partial \mathbf{x}^T}\right], \tag{55}$$

and the p-th row, j-th column $(p = 1, 2; j = 1, 2)$ of $\mathbf{I}_{DI}(\mathbf{x}, \mathbf{W}_l)$ is derived by (56), as shown at the bottom of the page, with $\mathbf{g}_l$, $\tilde{g}_{l,p}$ and $\mathbf{G}_{l,p}$ defined below:

$$\begin{cases} \mathbf{g}_l \triangleq \mathbf{W}_l^H \mathbf{a}(\mathbf{x}) \\ \tilde{g}_{l,p} \triangleq \dfrac{\partial |\mathbf{g}_l|^2}{\partial x_p}, p = 1, 2 \\ \mathbf{G}_{l,p} \triangleq \dfrac{\partial \mathbf{g}_l \mathbf{g}_l^H}{\partial x_p}, p = 1, 2. \end{cases} \tag{57}$$

Proof: See Appendix G. ∎

The CRLB in (54) is a function of the EBMs $\mathbf{W}_1, \ldots, \mathbf{W}_k$. Similar to that in Quasi-static Case, we consider the normalized CRLB (by multiplying k):

$$C_{DI}(\mathbf{x}, \mathbf{W}) \triangleq \text{Tr}\left\{\mathbf{I}_{DI}(\mathbf{x}, \mathbf{W})^{-1}\right\}. \tag{58}$$

By optimizing only one EBM $\mathbf{W}$, we can further get the *minimum CRLB*, given by

$$C_{DI}^{\min}(\mathbf{x}) = \min_{\mathbf{W}} C_{DI}(\mathbf{x}, \mathbf{W}) = C_{DI}(\mathbf{x}, \mathbf{W}_{DI}^*). \tag{59}$$

Solving problem (59) yields the optimal EBM $\mathbf{W}_{DI}^* = \left[\mathbf{w}_{DI,1}^*, \mathbf{w}_{DI,2}^*, \mathbf{w}_{DI,3}^*\right]$ with

$$\mathbf{w}_{DI,i}^* = \frac{1}{\sqrt{MN}} \mathbf{a}\left(\mathbf{x} + \boldsymbol{\Delta}_{DI,i}^*\right), \quad i = 1, 2, 3, \tag{60}$$

where $\left\{\boldsymbol{\Delta}_{DI,1}^*, \boldsymbol{\Delta}_{DI,2}^*, \boldsymbol{\Delta}_{DI,3}^*\right\}$ denotes the optimal set of exploration offsets in Dynamic Case I.

$$\left[\mathbf{I}_{DI}(\mathbf{x}, \mathbf{W}_l)\right]_{p,j} = \frac{\sigma_z^6 |\mathbf{s}|^6 \sigma_\beta^6}{|\boldsymbol{\Sigma}_{\mathbf{y},k}|^2}\left\{-2|\mathbf{g}_l|^2 \tilde{g}_{l,p} \tilde{g}_{l,j} + \frac{\sigma_z^2}{|\mathbf{s}|^2 \sigma_\beta^2} \text{Tr}\left\{\mathbf{G}_{l,p} \mathbf{G}_{l,j}\right\} + \mathbf{g}_l^H\left(\mathbf{G}_{l,p} \mathbf{G}_{l,j} + \mathbf{G}_{l,j} \mathbf{G}_{l,p}\right) \mathbf{g}_l\right\}, \tag{56}$$

C. ASYMPTOTICALLY OPTIMAL SET OF EXPLORATION OFFSETS

In general, the CRLB in (59) is a function of a set of system parameters including the equivalent channel gain parameter σ_β^2, the DPV $\mathbf{x}$ and the array size M, N. Hence, the optimal set of 2D exploration offsets should also be a function of these parameters. Since it is very hard to obtain the expression of this optimal set, we adopt numerical search to deal with this issue. However, as many parameters in (59) may affect the optimal result, numerical search has to be reconducted for different parameter sets, resulting in high complexity.

Fortunately, through our investigation, some useful properties of the minimum CRLB and the optimal set of exploration offsets are given to simplify the numerical search, as described in the following lemma:

Lemma 5: In Dynamic Case I, the minimum CRLB $C_{DI}^{\min}(\boldsymbol{\psi})$ and the optimal set of exploration offsets $\left\{\boldsymbol{\Delta}_{DI,1}^*, \boldsymbol{\Delta}_{DI,2}^*, \boldsymbol{\Delta}_{DI,3}^*\right\}$ have the following three properties:

1) $C_{DI}^{\min}(\boldsymbol{\psi})$ and $\left\{\boldsymbol{\Delta}_{DI,1}^*, \boldsymbol{\Delta}_{DI,2}^*, \boldsymbol{\Delta}_{DI,3}^*\right\}$ are invariant to the DPV $\mathbf{x}$;

2) $\frac{|\mathbf{s}|^2\sigma_\beta^2}{\sigma_z^2}C_{DI}^{\min}(\boldsymbol{\psi})$ converges to constant values as $\frac{|\mathbf{s}|^2\sigma_\beta^2}{\sigma_z^2} \to +\infty$;

3) $C_{DI}^{\min}(\boldsymbol{\psi})$ converges as M, $N \to +\infty$ and there exists a fixed set of exploration offsets that are unrelated to the array size and $\frac{|\mathbf{s}|^2\sigma_\beta^2}{\sigma_z^2}$, denoted as $\left\{\widetilde{\boldsymbol{\Delta}}_{DI,1}^*, \widetilde{\boldsymbol{\Delta}}_{DI,2}^*, \widetilde{\boldsymbol{\Delta}}_{DI,3}^*\right\}$, such that

$$\lim_{M,N\to+\infty} C_{DI}(\boldsymbol{\psi}, \widetilde{\mathbf{W}}_{DI}^*) = \lim_{M,N\to+\infty} C_{DI}^{\min}(\boldsymbol{\psi}),$$

where $\widetilde{\mathbf{W}}_{DI}^* = [\widetilde{\mathbf{w}}_{DI,1}^*, \widetilde{\mathbf{w}}_{DI,2}^*, \widetilde{\mathbf{w}}_{DI,3}^*]$ is obtained with

$$\widetilde{\mathbf{w}}_{DI,i}^* \triangleq \frac{1}{\sqrt{MN}}\mathbf{a}\left(\mathbf{x} + \widetilde{\boldsymbol{\Delta}}_{DI,i}^*\right), \quad i = 1, 2, 3. \tag{61}$$

Proof: See Appendix H. ∎

Lemma 5 reveals that $\left\{\boldsymbol{\Delta}_{DI,1}^*, \boldsymbol{\Delta}_{DI,2}^*, \boldsymbol{\Delta}_{DI,3}^*\right\}$ is only related to the array size M, N and $\frac{|\mathbf{s}|^2\sigma_\beta^2}{\sigma_z^2}$. Hence, the numerical search times can be reduced to one for a particular array size M, N and a particular $\frac{|\mathbf{s}|^2\sigma_\beta^2}{\sigma_z^2}$. Numerically, we find later that even if $\left\{\boldsymbol{\Delta}_{DI,1}^*, \boldsymbol{\Delta}_{DI,2}^*, \boldsymbol{\Delta}_{DI,3}^*\right\}$ may change for different array sizes and $\frac{|\mathbf{s}|^2\sigma_\beta^2}{\sigma_z^2}$, $\left\{\widetilde{\boldsymbol{\Delta}}_{DI,1}^*, \widetilde{\boldsymbol{\Delta}}_{DI,2}^*, \widetilde{\boldsymbol{\Delta}}_{DI,3}^*\right\}$ can be used to take the place of $\left\{\boldsymbol{\Delta}_{DI,1}^*, \boldsymbol{\Delta}_{DI,2}^*, \boldsymbol{\Delta}_{DI,3}^*\right\}$ as long as the antenna size M, N and $\frac{|\mathbf{s}|^2\sigma_\beta^2}{\sigma_z^2}$ are sufficiently large. Therefore, the numerical search times is reduced to one in the end. Similar to that in Quasi-static Case, $\left\{\widetilde{\boldsymbol{\Delta}}_{DI,1}^*, \widetilde{\boldsymbol{\Delta}}_{DI,2}^*, \widetilde{\boldsymbol{\Delta}}_{DI,3}^*\right\}$ is called the asymptotically optimal set of exploration offsets in Dynamic Case I in this article.

By numerical search in the main lobe in (10), one asymptotically optimal set of exploration offsets $\left\{\widetilde{\boldsymbol{\Delta}}_{DI,1}^*, \widetilde{\boldsymbol{\Delta}}_{DI,2}^*, \widetilde{\boldsymbol{\Delta}}_{DI,3}^*\right\}$ can be obtained in TABLE 3 and Fig. 5. With this set of exploration offsets, a general way to generate the EBM $\widetilde{\mathbf{W}}_{DI}^*$ is obtained by (61) to achieve the minimum CRLB.

TABLE 3. The asymptotically optimal set of exploration offsets in Dynamic Case I.

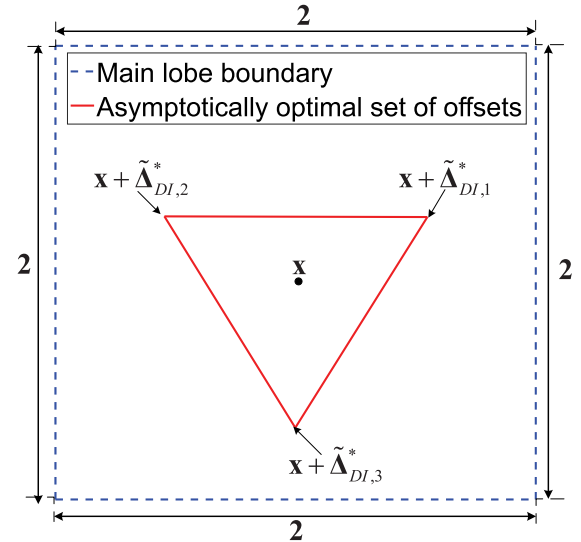| $\widetilde{\boldsymbol{\Delta}}_{DI,1}^*$	$\widetilde{\boldsymbol{\Delta}}_{DI,2}^*$	$\widetilde{\boldsymbol{\Delta}}_{DI,3}^*$
$[0.5486, 0.2451]^{\mathrm{T}}$	$[-0.5462, 0.2482]^{\mathrm{T}}$	$[-0.0012, -0.6837]^{\mathrm{T}}$

FIGURE 5. The asymptotically optimal set of exploration offsets in Dynamic Case I.
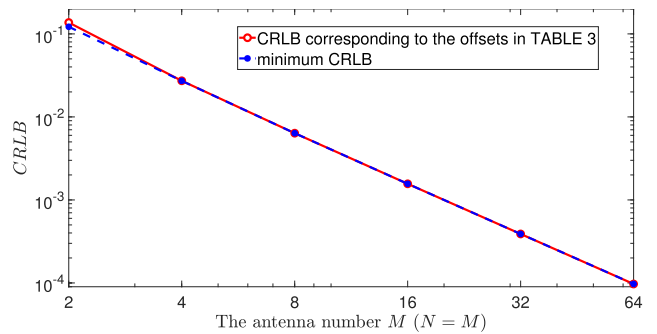
FIGURE 6. The performance of the offsets in TABLE 3 when $\frac{|\mathbf{s}|^2\sigma_\beta^2}{\sigma_z^2} = 0$ dB.

By adopting $\left\{\widetilde{\boldsymbol{\Delta}}_{DI,1}^*, \widetilde{\boldsymbol{\Delta}}_{DI,2}^*, \widetilde{\boldsymbol{\Delta}}_{DI,3}^*\right\}$ to smaller size antenna arrays when $\frac{|\mathbf{s}|^2\sigma_\beta^2}{\sigma_z^2} = 0$ dB, we compare the minimum CRLB and the CRLB achieved by $\left\{\widetilde{\boldsymbol{\Delta}}_{DI,1}^*, \widetilde{\boldsymbol{\Delta}}_{DI,2}^*, \widetilde{\boldsymbol{\Delta}}_{DI,3}^*\right\}$ in TABLE 3. As illustrated in Fig. 6, when antenna number $M = N \geq 8$, we can approach the minimum CRLB with a relative error less than 0.1% by using $\left\{\widetilde{\boldsymbol{\Delta}}_{DI,1}^*, \widetilde{\boldsymbol{\Delta}}_{DI,2}^*, \widetilde{\boldsymbol{\Delta}}_{DI,3}^*\right\}$.

By applying $\left\{\widetilde{\boldsymbol{\Delta}}_{DI,1}^*, \widetilde{\boldsymbol{\Delta}}_{DI,2}^*, \widetilde{\boldsymbol{\Delta}}_{DI,3}^*\right\}$ to different $\frac{|\mathbf{s}|^2\sigma_\beta^2}{\sigma_z^2}$ when $M = N = 8$, we compare the minimum CRLB and the CRLB achieved by $\left\{\widetilde{\boldsymbol{\Delta}}_{DI,1}^*, \widetilde{\boldsymbol{\Delta}}_{DI,2}^*, \widetilde{\boldsymbol{\Delta}}_{DI,3}^*\right\}$ in TABLE 3. As illustrated in Fig. 7, when $\frac{|\mathbf{s}|^2\sigma_\beta^2}{\sigma_z^2} \geq 0$ dB, we can approach

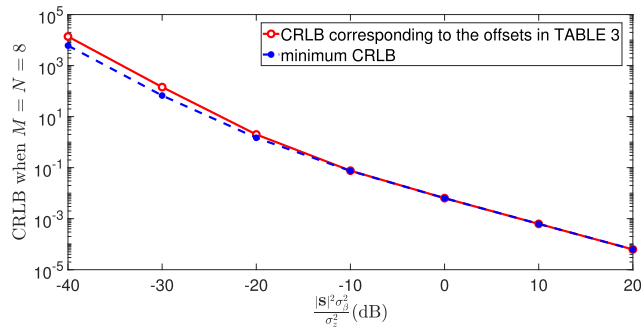**FIGURE 7.** The performance of the offsets in TABLE 3 when $M = N = 8$.

the minimum CRLB with a relative error less than 0.1% by using $\left\{ \widetilde{\boldsymbol{\Delta}}_{DI,1}^*, \widetilde{\boldsymbol{\Delta}}_{DI,2}^*, \widetilde{\boldsymbol{\Delta}}_{DI,3}^* \right\}$.

As a conclusion, it is practical to apply this asymptotically optimal set of exploration offsets to any antenna array with $M = N \geq 8$, any channel gain with $\frac{|\mathbf{s}|^2 \sigma_\beta^2}{\sigma_z^2} \geq 0$dB and any direction.

D. RECURSIVE BEAM TRACKING WITH ASYMPTOTIC OPTIMALITY ANALYSIS

For the Rayleigh fading channels, it is crucial to acquire the variance of the equivalent channel gain in (47), while it is hindered by the unknown antenna gain $\eta(\mathbf{x})$. Fortunately, the estimate of $\eta(\mathbf{x})$ and σ_β^2 can be seen as approximately accurate for a given antenna element pattern as the estimate of the DPV $\hat{\mathbf{x}}_k$ approach $\mathbf{x}$. Hence, we assume a perfectly-known σ_β^2 here to design the algorithm in Dynamic Case I. The deterioration of the tracking performance caused by the estimation error of the antenna gain will be evaluated in Section IX-D.

The proposed tracker is motivated by the following maximum likelihood problem:

$$\max_{\mathbf{W}_k} \left\{ \max_{\hat{\mathbf{x}}_k} \sum_{l=1}^{k} \left[\log p_{DI} \left(\mathbf{y}_l | \mathbf{x}, \mathbf{W}_l \right) \Big|_{\mathbf{x} = \hat{\mathbf{x}}_k} \right] \right\} \quad (62)$$
$$\text{s.t. } (9), (11), (17), (51), (53).$$

Similar to that in Section V, we propose a two-layer nested optimization algorithm to find the solution of (62). Finally, the proposed tracking algorithm is given in Algorithm 2.

We now perform the asymptotic optimality analysis. According to [34], [36], [37], the diminishing step-size is adopted as follows:

$$b_{DI,k} = \frac{\epsilon_{DI}}{k + K_{DI,0}}, \quad k = 1, 2, \cdots \quad (63)$$

where $K_{DI,0} \geq 0$ and $\epsilon_{DI} > 0$. Then we can prove that if the initial estimate $\hat{\mathbf{x}}_0$ is within the main lobe and $\epsilon_{DI} = 1$, the proposed algorithm can converge to $\mathbf{x}$ with the minimum CRLB with high probability, i.e.,

$$\lim_{k \to +\infty} k \mathbb{E} \left[\| \hat{\mathbf{x}}_k - \mathbf{x} \|_2^2 \right] = C_{DI}^{\min}(\mathbf{x}). \quad (64)$$

Algorithm 2 Recursive Beam Tracking (RBT) for Dynamic Case I

1) **Exploring and Receiving** (Step 3 in Procedure 1): Transmit 3 pilot sequences in each ECC. The corresponding EBV for receiving the i-th pilot sequence in k-th ECC is given below:

$$\mathbf{w}_{k,i} = \frac{1}{\sqrt{MN}} \mathbf{a} \left(\hat{\mathbf{x}}_{k-1} + \widetilde{\boldsymbol{\Delta}}_{DI,i}^* \right), \, i = 1, 2, 3, \quad (65)$$

where $\hat{\mathbf{x}}_k = \left[\hat{x}_{k,1}, \hat{x}_{k,2} \right]^{\mathrm{T}}$ and $\left\{ \widetilde{\boldsymbol{\Delta}}_{DI,1}^*, \widetilde{\boldsymbol{\Delta}}_{DI,2}^*, \widetilde{\boldsymbol{\Delta}}_{DI,3}^* \right\}$ is given by TABLE 3. After match filtering, the observation vector $\mathbf{y}_k$ is obtained via (9).

2) **Updating Estimate** (Step 6 in Procedure 1): The estimate $\hat{\mathbf{x}}_k = \left[\hat{x}_{k,1}, \hat{x}_{k,2} \right]^{\mathrm{T}}$ is updated by

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_{k-1} + b_{DI,k} \mathbf{I}_{DI} \left(\hat{\mathbf{x}}_{k-1}, \mathbf{W}_k \right)^{-1} \frac{\partial \log p_{DI} \left(\mathbf{y}_k | \hat{\mathbf{x}}_{k-1}, \mathbf{W}_k \right)}{\partial \hat{\mathbf{x}}_{k-1}}, \quad (66)$$

where $\mathbf{I}_{DI} \left(\hat{\mathbf{x}}_{k-1}, \mathbf{W}_k \right)$ is defined in (55) and $b_{DI,k}$ is the step size that will be specified later.

Algorithm 3 Joint Beam and Channel Tracking (JBCT) for Dynamic Case II

1) **Exploring and Receiving** (Step 3 in Procedure 1): Transmit 3 pilot sequences in each ECC. The corresponding EBV for receiving the i-th pilot sequence in k-th ECC is given below:

$$\mathbf{w}_{k,i} = \frac{1}{\sqrt{MN}} \mathbf{a} \left(\hat{\mathbf{x}}_{k-1} + \boldsymbol{\Delta}_{DII,i} \right), \, i = 1, 2, 3, \quad (67)$$

where $\hat{\mathbf{x}}_k = \left[\hat{x}_{k,1}, \hat{x}_{k,2} \right]^{\mathrm{T}}$ and $\boldsymbol{\Delta}_{DII,i} = \widetilde{\boldsymbol{\Delta}}_{S,i}^* (i = 1, 2, 3)$ are given by TABLE 2. After match filtering, the observation vector $\mathbf{y}_k$ is obtained via (9).

2) **Updating Estimate** (Step 6 in Procedure 1): The estimate of the channel parameter vector in k-th ECC, i.e., $\hat{\boldsymbol{\psi}}_k = \left[\hat{\beta}_k^{\mathrm{re}}, \hat{\beta}_k^{\mathrm{im}}, \hat{x}_{k,1}, \hat{x}_{k,2} \right]^{\mathrm{T}}$, is updated by

$$\hat{\boldsymbol{\psi}}_k = \hat{\boldsymbol{\psi}}_{k-1} + b_{DII,k} \boldsymbol{\varsigma}_k, \quad (68)$$

where $\boldsymbol{\varsigma}_k$ is the updating direction vector given by (34) and $b_{DII,k}$ is the step size for Dynamic Case II.

The proof is similar to that in Section V and the details are omitted here since nothing new is provided in the proof.

VII. JOINT BEAM AND CHANNEL TRACKING FOR DYNAMIC CASE II

In Dynamic Case II where both the channel gain β_k^c and the DPV $\mathbf{x}_k$ change fast, the observation vector $\mathbf{y}_k$ satisfies normal distribution with $\mathbf{y}_k \sim \mathcal{CN} \left(|\mathbf{s}| \beta(\mathbf{x}_k) \mathbf{W}_k^{\mathrm{H}} \mathbf{a}(\mathbf{x}_k), \sigma_z^2 \mathbf{J}_3 \right)$

for a given channel parameter vector $\boldsymbol{\psi}_k$ and EBM $\mathbf{W}_k$. Hence, the conditional probability density function of the observation vector $\mathbf{y}_k$ is given by

$$p_{DII}(\mathbf{y}_k|\boldsymbol{\psi}_k, \mathbf{W}_k) = \frac{1}{\pi^3 \sigma_z^6} e^{-\frac{\left\| \mathbf{y}_k - |s|\beta(\mathbf{x}_k)\mathbf{W}_k^H \mathbf{a}(\mathbf{x}_k) \right\|_2^2}{\sigma_z^2}}. \quad (69)$$

Establishing theorems of tracking, as in Section V and Section VI, is very difficult in Dynamic Case II. Even if the theoretical analysis is not conducted in this section, we still provide a tracking algorithm in this section.

Inspired by the asymptotically optimal tracking algorithm in Section V and Section VI, we design a similar joint beam and channel tracking algorithm in Algorithm 3.

Different from the step-size in Quasi-static Case and Dynamic Case I, we adopt constant step-size in Dynamic Case II as the diminishing step-size cannot track the fast-changing $\mathbf{x}_k$ and β_k. The constant step-size $b_{DII,k}$ will be specified later.

VIII. COMPUTATIONAL COMPLEXITY

In this section, we evaluate the computational complexity of the proposed tracking algorithms in Quasi-static Case, Dynamic Case I and Dynamic Case II. We focus on the complex arithmetic operations in the tracking stage including complex multiplication and division, while complex addition and subtraction are omitted since they require much fewer operations. It seems that Algorithm 1, Algorithm 2 and Algorithm 3 require a huge number of complex arithmetic operations due to the Fisher information matrix inversion in each ECC. However, most of these calculation work can be finished off-line, by which the complex operations are greatly reduced. The following lemma is proposed to tell us the specific computational complexity:

Lemma 6: If the number of offline complex arithmetic operations is ignored since it is much smaller than the online ones as the tracking process lasts, then

1) for Algorithm 1 in Quasi-static Case and Algorithm 3 in Dynamic Case II, 45 complex arithmetic operations are required in each ECC;

2) for Algorithm 2 in Dynamic Case I, 28 complex arithmetic operations are required in each ECC.

Proof: See Appendix I. ∎

According to Lemma 6, our algorithms can efficiently work without high complexity.

IX. NUMERICAL RESULTS

In this section, some numerical results will be provided to verify the performance of our proposed tracking algorithms for Quasi-static Case, Dynamic Case I and Dynamic Case II. Based on the model in Section II, the parameters are set as: $M = N = 8$, the antenna spacing $d_1 = d_2 = \frac{\lambda}{2}$, and the transmit SNR is $\frac{|s|^2}{\sigma_z^2} = 0$ dB. The antenna element pattern is based on the 3GPP model [38]. The vertical cut and the horizontal cut of the radiation power pattern (normalized with 0 dB at the

central direction) for each element are given below:

$$\eta_{dB}\left(\theta, \phi = \frac{\pi}{2}\right) = -\min\left\{12\left(\frac{\theta}{\theta_{3\,dB}}\right)^2, \eta_{max}\right\} \quad (70)$$

$$\eta_{dB}\left(\theta = 0, \phi\right) = -\min\left\{12\left(\frac{\phi - \frac{\pi}{2}}{\phi_{3\,dB}}\right)^2, \eta_{max}\right\}, \quad (71)$$

where $\theta_{3\,dB} = \frac{13\pi}{36}$ is the 3 dB beamwidth in the vertical direction, $\phi_{3\,dB} = \frac{13\pi}{36}$ is the 3 dB beamwidth in the horizontal direction and $\eta_{max} = 30$ dB is the maximum attenuation. The combined radiation power pattern (normalized) of each antenna element in [38] is given by

$$\eta_{dB}(\theta, \phi) = -\min\left\{-\left(\eta_{dB}\left(\theta, \phi = \frac{\pi}{2}\right) + \eta_{dB}\left(\theta = 0, \phi\right)\right), \eta_{max}\right\}. \quad (72)$$

As can be seen in Fig. 8, we define two direction regions: central direction region ($\eta_{dB}(\theta, \phi) \geq -6$ dB) and edge direction region ($\eta_{dB}(\theta, \phi) \leq -18$ dB). We will separately evaluate the tracking performance in these two direction regions afterwards.
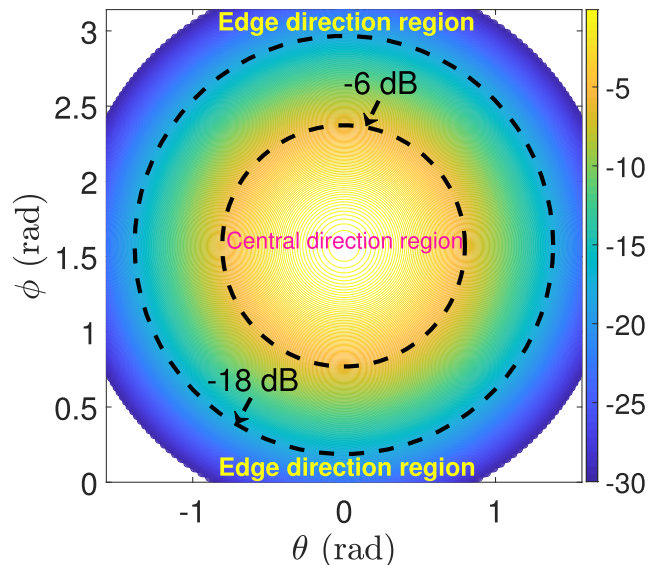
FIGURE 8. An example of the normalized antenna radiation power pattern (dB) versus the AoA θ, ϕ (rad) for each element.

A. REFERENCE ALGORITHMS

Reference algorithms include the *compressed sensing* algorithm in [14], the *3GPP New Radio (NR)* tracking algorithm in [16], the *extended Kalman filter (EKF)* algorithm in [15] and the *recursive beam and channel tracking (RBCT)* algorithm in [10]. For the compressed sensing algorithm in [14], we randomly choose phase shifts from $\{\pm 1, \pm j\}$ and then use the sparse recovery algorithm to estimate the DPV, where a discrete Fourier transform (DFT) dictionary with a size of 1024 is utilized. As for the 3GPP NR tracking algorithm, the last estimated beam direction and its adjacent beam directions are probed, then the system determines whether to

switch the estimate in current ECC according to the strength of the received signals. For the EKF algorithm, we extend the method in [15] to the 2D array and the three exploring directions used in each ECC form a regular triangle within half the main lobe of the DPV estimate. The original RBCT algorithm in [10] is designed for 1D system and cannot directly support 2D tracking. To compare with our algorithms in the 2D case, we use two symmetrical explorations to track each dimension of the 2D beam for the RBCT algorithm.

For the initial beam estimation stage in Fig. 1, an exhaustive beam sweeping is conducted. Then an initial estimate is obtained by using the orthogonal matching pursuit method in [39]. This ensures that the initial estimate of the DPV, i.e., $\hat{\mathbf{x}}_0$, is within the main lobe in (10).

In the tracking stage, three explorations are conducted in each ECC for all the algorithms. For the RBCT algorithm in [10], we use a buffer to store the received observations and update the estimate when receiving four new observations.

It is worth pointing out that the compressed sensing algorithm in [14] does not require an initial estimate. To ensure the fairness of all the algorithms, we compare the tracking performance versus the total number of explorations (i.e., assuming the compressed sensing algorithm uses the same total number of explorations) used both in the initial beam estimation stage and the tracking stage.

B. PERFORMANCE BOUND OF THE TRACKING ALGORITHMS

As revealed in Section V-A and Section VI-B, the CRLB is a function of the adopted EBMs. Since the EBMs of our algorithms and the reference algorithms are quite different, the corresponding achieved CRLBs are also different. In this subsection, we will compare the CRLBs achieved by the EBMs of these different algorithms to verify the superiority of our optimal EBMs.

The fixed AoA (θ, ϕ) in Quasi-static Case and Dynamic Case I is chosen evenly and randomly in $\theta \in \left[-\frac{\pi}{6}, \frac{\pi}{6}\right], \phi \in \left[\frac{\pi}{3}, \frac{2\pi}{3}\right]$. The corresponding antenna gain of each element varies from -5.2 dB to 0 dB via (72). For Quasi-static case, the fixed channel gain β^c is modeled as Rician fading with a K-factor $\kappa = 15$ dB, according to the channel model in [40]. For Dynamic Case I, the channel gains between different ECCs are independent of each other and modeled as Rayleigh fading with $\frac{|\mathbf{s}|^2\left(\sigma_\beta^c\right)^2}{\sigma_z^2} = 0$ dB. All simulation results in this section are averaged over 1000 random system realizations.

Fig. 9 and Fig. 10 shows the CRLBs achieved by the EBMs of different algorithms. It can be observed that our optimal EBMs both in Quasi-static Case and Dynamic Case I can achieve lower CRLB compared with the EBMs used by the reference algorithms, which results from the fact that the EBMs of our algorithms are carefully optimized and proven to be optimal in theory. Since the CRLB illustrates the performance bound, the other four algorithms cannot perform better than our algorithms in potential tracking accuracy.
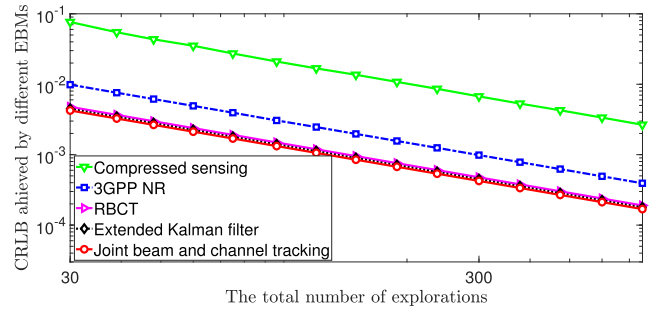
FIGURE 9. The CRLBs in Quasi-static Case.
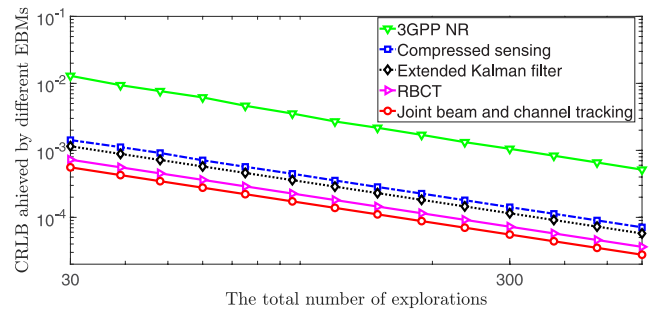
FIGURE 10. The CRLBs in Dynamic Case I.

C. RESULTS OF TRACKING ACCURACY

In this subsection, we will evaluate the tracking performance of our algorithms in the central direction region. The AoA (θ, ϕ) as defined in Section II is chosen evenly and randomly in $\theta \in \left[-\frac{\pi}{6}, \frac{\pi}{6}\right], \phi \in \left[\frac{\pi}{3}, \frac{2\pi}{3}\right]$. The corresponding antenna gain of each element varies from -5.2 dB to 0 dB via (72).

1) QUASI-STATIC CASE
The channel gain β^c is modeled as Rician fading with a K-factor $\kappa = 15$ dB, according to the channel model in [40]. The step-size is set as $b_{S,k} = \frac{1}{k}$.
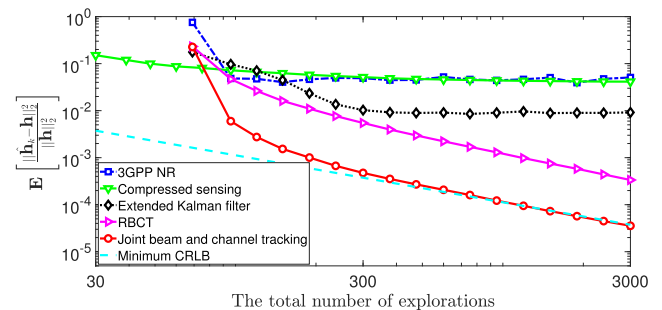
FIGURE 11. $\mathrm{MSE_h}$ in Quasi-static Case when the AoA is in the central direction region.

As can be observed in Fig. 11, the tracking accuracy of the compressed sensing algorithm and the 3GPP NR tracking algorithm gradually keeps steady with the increasing number of explorations, as both of the two algorithms are grids-of-beam based approaches and the performance is restricted by the codebook resolution. The EKF algorithm in Fig. 11

also shows a similar feature, which results from the fact that the original EKF algorithm itself cannot efficiently track static parameters [41]. Although the tracking error of the RBCT algorithm gradually reduces as the number of exploration increases, it cannot achieve the corresponding CRLB in Fig. 9. This is caused by tracking the horizontal and the vertical directions separately in the RBCT algorithm, causing loss as against joint tracking. Compared with the four reference algorithms, our proposed JBCT algorithm can approach the minimum CRLB quickly and achieve much lower tracking error.

2) DYNAMIC CASE I

The channel gains between different ECCs are independent of each other and Rayleigh fading channels are adopted in each ECC with $\frac{|\mathbf{s}|^2 \left(\sigma_\beta^c \right)^2}{\sigma_z^2} = 0\,\mathrm{dB}$. The step-size is set as $b_{DI,k} = \frac{1}{k}$. As the RBCT algorithm in [10] does not support fast-fading channel tracking, it is excluded from the reference algorithms in Dynamic Case I.
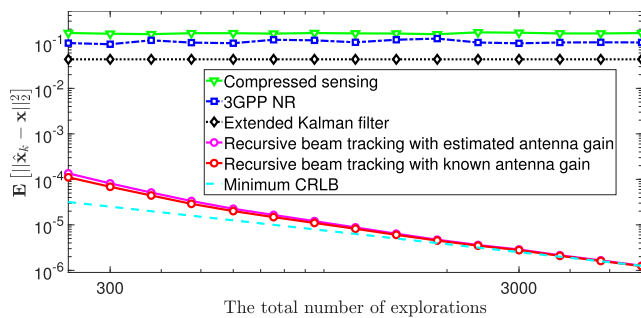
FIGURE 12. MSE$_x$ in Dynamic Case I for Rayleigh fading channel when the AoA is in the central direction region.

Fig. 12 indicates that the DPV MSE of our proposed RBT algorithm can converge to the minimum CRLB if the antenna gain can be perfectly known. Even with the estimated antenna gain for tracking, our algorithm can also converge to the minimum CRLB and achieve much lower tracking error than other algorithms. Hence, the estimation error of the antenna gain has little influence on the tracking performance of our algorithm when the AoA is in the central direction region.

Further, we evaluate the robustness of the proposed algorithm for other types of time-varying channels. The channel gains between different ECCs are still independent of each other while Rician fading channels are adopted in each ECC with a K-factor $\kappa = 15\,\mathrm{dB}$ and $\frac{|\mathbf{s}|^2 \left(\sigma_\beta^c \right)^2}{\sigma_z^2} = 0\,\mathrm{dB}$, where $\left(\sigma_\beta^c \right)^2$ denotes the average energy gain of the Rician fading channel. The algorithm designed for the Rayleigh fading channel is adopted to track the Rician fading channel here. It can be observed in Fig. 13 that our algorithm can still converge and achieve much lower tracking error than existing algorithms. These results show that the proposed algorithm in Dynamic Case I is robust to different time-varying channels as long as the variance of the channel gain is known.
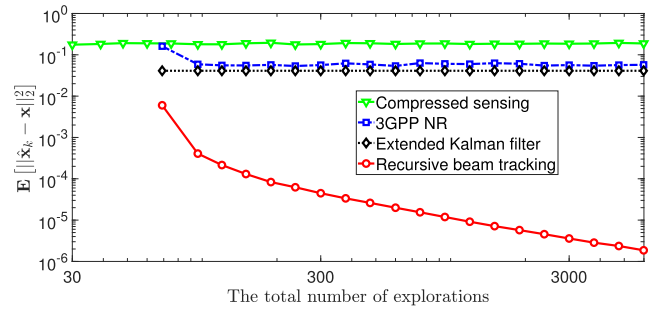
FIGURE 13. MSE$_x$ in Dynamic Case I for Rician fading channel when the AoA is in the central direction region.

3) DYNAMIC CASE II

In Dynamic Case II, the initial AoA (θ_0, ϕ_0) as defined in Section II is chosen evenly and randomly in $\theta_0 \in \left[-\frac{\pi}{6}, \frac{\pi}{6} \right]$, $\phi_0 \in \left[\frac{\pi}{3}, \frac{2\pi}{3} \right]$. The AoA (θ_k, ϕ_k) is modeled as a random walk process with return, i.e., $\theta_{k+1} = \theta_k + \varpi_k^\theta \Delta\theta_k$, $\phi_{k+1} = \phi_k + \varpi_k^\phi \Delta\phi_k$, where $\Delta\theta_k, \Delta\phi_k \sim \mathcal{N}(0, \delta_A^2)$, and $\varpi_k^\theta, \varpi_k^\phi \in \{-1, 1\}$ denote the rotation direction. The rotation direction $\varpi_k^\theta, \varpi_k^\phi$ are chosen such that θ_k varies in $\left[-\frac{\pi}{6}, \frac{\pi}{6} \right]$ and ϕ_k varies in $\left[\frac{\pi}{3}, \frac{2\pi}{3} \right]$. The channel gain is modeled as a first-order Gaussian-Markov process, i.e., $\beta_{k+1}^c = \rho \beta_k^c + \gamma_k$, where $\gamma_k \sim \mathcal{CN}(0, 1 - \rho^2)$. We adopt $\rho = 0.995$ in simulation. As for the step-size, numerical results show that when $b_{DII,k} = 0.7$, the joint beam and channel tracking algorithm can track beams with higher velocity. Hence, the step-size is set as a constant $b_{DII,k} = 0.7$.
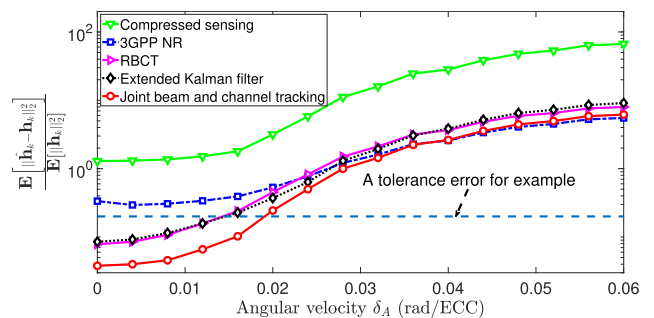
FIGURE 14. MSE$_{h_k}$ in Dynamic Case II when the AoA is in the central direction region.

Fig. 14 indicates the proposed JBCT algorithm in Dynamic Case II can achieve higher tracking accuracy than the other four algorithms. In addition, if we set a tolerance error in Fig. 14, then our algorithm can support higher angular velocities.

D. THE IMPACT OF THE ANTENNA PATTERN
In this subsection, we will evaluate the impact of the antenna pattern by setting the AoA in the edge direction region in Fig. 8. Other parameters are the same as the setting in Section IX-C.

FIGURE 15. MSE$_\mathbf{h}$ in Quasi-static Case when the AoA is in the edge direction region.

FIGURE 16. MSE$_\mathbf{x}$ in Dynamic Case I for Rayleigh fading channel when the AoA is in the edge direction region.

1) QUASI-STATIC CASE

The AoA (θ, ϕ) as defined in Section II is chosen evenly and randomly in $\theta \in \left[\frac{\pi}{2} - \frac{\pi}{60}, \frac{\pi}{2}\right]$, $\phi \in \left[\pi - \frac{\pi}{60}, \pi\right]$. The corresponding antenna gain of each element is -30 dB via (72).

In Fig. 15, it can be seen that our JBCT algorithm still outperforms existing algorithms when the AoA is in the edge direction region. Nevertheless, compared with Fig. 11, the minimum CRLB cannot be achieved any more and the performance of all algorithms deteriorate if the transmit power keeps unchanged, i.e., $\frac{|\mathbf{s}|^2}{\sigma_z^2} = 0$ dB. This can be explained by the decrease of the equivalent SNR when the AoA is in the edge direction region. If we compensate the gain loss in the edge direction region by increasing the transmit power by 30 dB, i.e., $\frac{|\mathbf{s}|^2}{\sigma_z^2} = 30$ dB, then Fig. 15 demonstrates the performance is greatly improved and the minimum CRLB can be achieved again. Furthermore, we compare the tracking performance in two cases: 1) $\frac{|\mathbf{s}|^2}{\sigma_z^2} = 30$ dB when $\theta \in \left[\frac{\pi}{2} - \frac{\pi}{60}, \frac{\pi}{2}\right]$, $\phi \in \left[\pi - \frac{\pi}{60}, \pi\right]$, where the antenna gain of each element is -30 dB; 2) $\frac{|\mathbf{s}|^2}{\sigma_z^2} = 0$ dB when $\theta \in \left[-\frac{\pi}{120}, \frac{\pi}{120}\right]$, $\phi \in \left[\frac{\pi}{2} - \frac{\pi}{120}, \frac{\pi}{2} + \frac{\pi}{120}\right]$, where the antenna gain of each element can be seen as 0 dB. It can be observed in Fig. 15 that the performance in these two cases is almost the same. This shows that the deterioration of the tracking performance in the edge direction region only results from the decrease of the equivalent SNR.

2) DYNAMIC CASE I

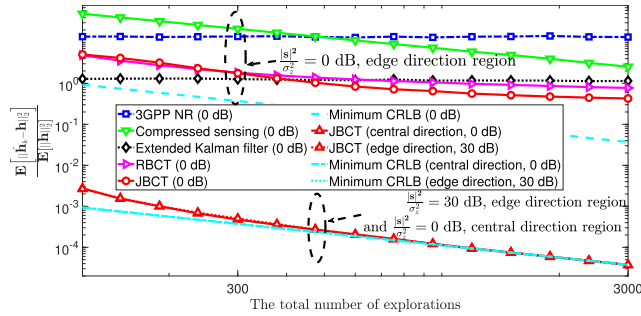The AoA (θ, ϕ) as defined in Section II is chosen evenly and randomly in $\theta \in \left[\frac{\pi}{3}, \frac{\pi}{2}\right]$, $\phi \in \left[\frac{5\pi}{6}, \pi\right]$. The corresponding antenna gain of each element varies from -30 dB to -20.4 dB via (72). As can be observed in Fig. 16, our RBT algorithm still outperforms existing algorithms for Rayleigh fading channel when the AoA is in the edge direction region. Since the equivalent SNR decreases, the proposed algorithm cannot converge to the minimum CRLB as before. If we compensate the gain loss in the edge direction region by increasing the transmit power by 30 dB, i.e., $\frac{|\mathbf{s}|^2}{\sigma_z^2} = 30$ dB, then our algorithm can still converge to the minimum CRLB
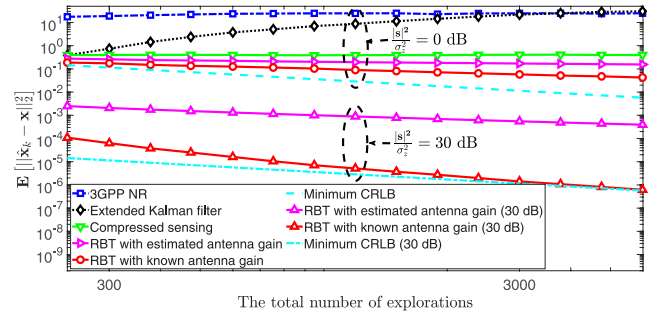
and achieve the same performance as that in Fig. 12 with the perfectly-known antenna gain.

With the estimated antenna gain, our RBT algorithm cannot converge to the minimum CRLB as before even when $\frac{|\mathbf{s}|^2}{\sigma_z^2} = 30$ dB. This is caused by the larger slope in the edge direction region compared with the central edge direction region. Hence, a small estimation error of the AoA can result in a large deviation of the estimated equivalent channel gain parameter σ_β^2 in (47), leading to the non-convergence when using estimated antenna gain in Fig. 16.
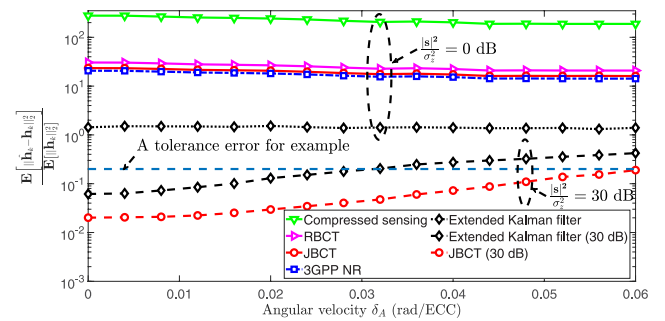
FIGURE 17. MSE$_{\mathbf{h}_k}$ in Dynamic Case II when the AoA is in the edge direction region.

3) DYNAMIC CASE II

The initial AoA (θ_0, ϕ_0) as defined in Section II is chosen evenly and randomly in $\theta_0 \in \left[\frac{\pi}{3}, \frac{\pi}{2}\right]$, $\phi_0 \in \left[\frac{5\pi}{6}, \pi\right]$. The rotation direction $\varpi_k^\theta, \varpi_k^\phi$ are chosen such that θ_k varies in $\left[\frac{\pi}{3}, \frac{\pi}{2}\right]$ and ϕ_k varies in $\left[\frac{5\pi}{6}, \pi\right]$. The corresponding antenna gain of each element varies from -30 dB to -20.4 dB via (72). As can be seen in Fig. 17, all the algorithms cannot efficiently track the channels when the AoA varies in the edge direction region, since the equivalent SNR decreases sharply. If we compensate the gain loss in the edge direction region by increasing the transmit power by 30 dB, i.e., $\frac{|\mathbf{s}|^2}{\sigma_z^2} = 30$ dB, then our JBCT algorithm can still achieve lower tracking error and faster tracking speed.

E. COMPUTATIONAL COMPLEXITY

We then evaluate the computational complexity of our proposed algorithms. As can be seen in TABLE 4, our

TABLE 4. Number of required complex operations in each ECC.

3GPP NR	Algorithm 2	Algorithm 1 and Algorithm 3	Extended Kalman filter	Compressed sensing
6	28	45	1427	129088

algorithms require fewer complex operations than other algorithms except 3GPP NR. Compared with 3GPP NR, the proposed algorithms can achieve much more accurate tracking without greatly increasing the computational complexity.

X. CONCLUSION

This article focuses on fast accurate beam and channel tracking for 2D phased antenna arrays. We first give the minimum exploration overhead of joint 2D tracking in theory. Then three tracking algorithms are developed according to different practical time-varying channel models.

In Quasi-static Case, the optimal exploration offsets are derived which are proved to a) be unrelated to the channel gain and the beam direction, b) be determined only by the array size, and c) approach constants as the array size goes to infinity. Also, a joint beam direction and channel gain tracking algorithm is proposed and the tracking error is proved to converge to the minimum CRLB.

In Dynamic Case I, an algorithm for beam only tracking is proposed, and it is proved to converge and achieve the minimum CRLB on the beam direction.

In Dynamic Case II, a joint tracking algorithm of beam direction and channel gain is proposed with faster and more accurate performance presented by simulation results.

This work is the first step to beam and channel tracking with 2D phased antenna arrays. In future work, we will further study the following problems: i) establishing the corresponding theorems in Dynamic Case II; ii) jointly tracking multipath channels; iii) tracking at both the transmitter and the receiver.

APPENDIX A
PROOF OF LEMMA 1

If the EBVs are of the steering vector forms, i.e., $\mathbf{w}_{k,i} = \frac{1}{\sqrt{MN}} \mathbf{a}(\boldsymbol{\omega}_{k,i})$, where $\boldsymbol{\omega}_{k,i} = [\omega_{k,i1}, \omega_{k,i2}]^T$ denotes the i-th exploring direction vector in k-th ECC, then the noiseless complex observation equation for the i-th observation is given

in (73), as shown at the bottom of the page, where Step (a) follows the definition of $y_a(\boldsymbol{\Delta})$:

$$y_a(\boldsymbol{\Delta}) \triangleq \frac{\sin(\pi\delta_1)}{\sin\left(\frac{\pi\delta_1}{M}\right)} \frac{\sin(\pi\delta_2)}{\sin\left(\frac{\pi\delta_2}{N}\right)} \quad (74)$$

with $\boldsymbol{\Delta} \triangleq [\delta_1, \delta_2]^T$. In our real tracking problem, the exploring direction vector $\boldsymbol{\omega}_{k,i}$ should be ensured within the main lobe of $\mathbf{x}_k$ in (10), i.e., $|\omega_{k,i1} - x_{k,1}| < 1$ and $|\omega_{k,i2} - x_{k,2}| < 1$. Hence, we have that $y_a(\boldsymbol{\omega}_{k,i} - \mathbf{x}_k) > 0$.

The complex observation equation in (73) contains two real equations, i.e., an amplitude equation and a phase angle equation. Therefore, q amplitude equations and q phase angle equations can be obtained after q observations. If we set the first observation as a reference, then we can obtain $(q-1)$ relative amplitude equations and $(q-1)$ relative phase angle equations from the remaining $(q-1)$ observations. However, these phase angle equations are not independent, as to be explained below.

From (73), we can obtain the phase angle equation:

$$\angle(y_{k,i}) = \angle\beta(\mathbf{x}_k) - \pi\left[\frac{M-1}{M}(\omega_{k,i1} - x_{k,1}) + \frac{N-1}{N}(\omega_{k,i2} - x_{k,2})\right].$$

Thus the relative phase angle equation of the i-th observation $y_{k,i}$ regarding the first observation $y_{k,1}$ ($i \neq 1$) can be obtained as below:

$$\angle(y_{k,i}) - \angle(y_{k,1})$$
$$= \pi\left[\frac{M-1}{M}(\omega_{k,11} - \omega_{k,i1}) + \frac{N-1}{N}(\omega_{k,12} - \omega_{k,i2})\right], \quad (75)$$

where $\omega_{k,11} - \omega_{k,i1}$ and $\omega_{k,12} - \omega_{k,i2}$ are determined by the exploring direction vectors and unrelated to the channel parameter vector $\boldsymbol{\psi}_k$. From (75), we can know that once the exploring directions are determined, the relative phase angles are known constants unrelated to $\boldsymbol{\psi}_k$. In other words, the relative phase angle equations as revealed in (75) cannot provide any information for estimating $\boldsymbol{\psi}_k$.

Following the conclusion above, we analyze the minimum exploration overhead in the following two cases:

1) If we want to obtain the unique solution of $\boldsymbol{\psi}_k$ within one ECC, at least 4 independent real equations with respect to $\boldsymbol{\psi}_k$ are needed since $\boldsymbol{\psi}_k$ contains four independent real

$$y_{k,i} = \frac{|\mathbf{s}|\beta(\mathbf{x}_k)}{\sqrt{MN}} \mathbf{a}(\boldsymbol{\omega}_{k,i})^H \mathbf{a}(\mathbf{x}_k) = \frac{|\mathbf{s}|\beta(\mathbf{x}_k)}{\sqrt{MN}} \sum_{m=1}^{M}\sum_{n=1}^{N} e^{-j2\pi\left[\frac{(m-1)(\omega_{k,i1} - x_{k,1})}{M} + \frac{(n-1)(\omega_{k,i2} - x_{k,2})}{N}\right]}$$

$$= \frac{|\mathbf{s}|\beta(\mathbf{x}_k)}{\sqrt{MN}} \frac{\sin[\pi(\omega_{k,i1} - x_{k,1})]}{\sin\left[\frac{\pi(\omega_{k,i1} - x_{k,1})}{M}\right]} \frac{\sin[\pi(\omega_{k,i2} - x_{k,2})]}{\sin\left[\frac{\pi(\omega_{k,i2} - x_{k,2})}{N}\right]} e^{-j\pi\left[\frac{M-1}{M}(\omega_{k,i1} - x_{k,1}) + \frac{N-1}{N}(\omega_{k,i2} - x_{k,2})\right]}$$

$$\overset{(a)}{=} \frac{|\mathbf{s}|\beta(\mathbf{x}_k)}{\sqrt{MN}} y_a(\boldsymbol{\omega}_{k,i} - \mathbf{x}_k) e^{-j\pi\left[\frac{M-1}{M}(\omega_{k,i1} - x_{k,1}) + \frac{N-1}{N}(\omega_{k,i2} - x_{k,2})\right]}, \quad (73)$$

variables (i.e., the real part $\mathrm{Re}\{\beta(\mathbf{x}_k)\}$, the imaginary part $\mathrm{Im}\{\beta(\mathbf{x}_k)\}$ of the equivalent channel gain β_k and the two direction parameters $x_{k,1}$, $x_{k,2}$). After q explorations in each ECC, we can obtain q independent amplitude equations and only 1 independent phase angle equation, which is $q+1$ independent real equations with respect to $\boldsymbol{\psi}_k$ in total. Hence, at least 3 explorations are needed to obtain 4 independent real equations and estimate 4 independent real variables of $\boldsymbol{\psi}_k$.

2) If we only want to obtain the unique solution of $\mathbf{x}_k$ within one ECC, at least 2 independent real equations with respect to $\mathbf{x}_k$ are needed since $\mathbf{x}_k$ contains two independent real variables (i.e., two direction parameters $x_{k,1}$, $x_{k,2}$). It seems that fewer explorations are sufficient. However, we cannot obtain any absolute amplitude and phase information with respect to $\mathbf{x}_k$ from one observation in (73) since $\beta(\mathbf{x}_k)$ is unknown. In addition, the relative phase angles are constants unrelated to $\mathbf{x}_k$. Thus, the phase angle equations are useless for estimating $\mathbf{x}_k$. After q explorations in each ECC, we can obtain $q-1$ independent relative amplitude equations with respect to $\mathbf{x}_k$ in total. Hence, at least 3 explorations are needed to obtain 2 independent real equations and estimate 2 independent real variables of $\mathbf{x}_k$.

Therefore, the proof is completed.

APPENDIX B
PROOF OF LEMMA 2

In problem (18), the constraint (14) ensures that $\hat{\mathbf{h}}_k$ is an unbiased estimate of $\mathbf{h}$. Consider each element of the channel vector $\mathbf{h}$, i.e., $h_{mn}(\boldsymbol{\psi}) = \beta e^{j2\pi\left(\frac{m-1}{M}x_1 + \frac{n-1}{N}x_2\right)}$. Immediately we have $\mathbb{E}\left[h_{mn}(\hat{\boldsymbol{\psi}}_k)\right] = h_{mn}(\boldsymbol{\psi})$ since $\mathbb{E}\left[\hat{\mathbf{h}}_k\right] = \mathbf{h}$. According to Section 3.8 of [33], if a function $f\left(\hat{\boldsymbol{\psi}}\right)$ is an unbiased estimate of $f(\boldsymbol{\psi})$, i.e., $\mathbb{E}\left[f(\hat{\boldsymbol{\psi}})\right] = f(\boldsymbol{\psi})$, then we can obtain that

$$\mathrm{Var}[f(\hat{\boldsymbol{\psi}})] \geq \frac{\partial f(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}^{\mathrm{T}}}\mathbf{I}(\boldsymbol{\psi})^{-1}\left(\frac{\partial f(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}^{\mathrm{T}}}\right)^{\mathrm{H}}, \qquad (76)$$

where $\mathrm{Var}[f(\hat{\boldsymbol{\psi}})]$ denotes the variance of $f(\hat{\boldsymbol{\psi}})$ and $\mathbf{I}(\boldsymbol{\psi})$ is the corresponding Fisher information matrix.

Combining (18) and (76), we have

$$\frac{1}{MN}\mathbb{E}\left[\left\|\hat{\mathbf{h}}_k - \mathbf{h}\right\|_2^2\right]$$

$$= \frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N}\mathbb{E}\left[\left|h_{mn}(\hat{\boldsymbol{\psi}}) - h_{mn}(\boldsymbol{\psi})\right|^2\right]$$

$$\stackrel{(a)}{\geq} \frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N}\left(\frac{\partial h_{mn}(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}^{\mathrm{T}}}\left(\sum_{l=1}^{k}\mathbf{I}_S(\boldsymbol{\psi}, \mathbf{W}_l)\right)^{-1}\left(\frac{\partial h_{mn}(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}^{\mathrm{T}}}\right)^{\mathrm{H}}\right)$$

$$= \frac{1}{MN}\mathrm{Tr}\left\{\left(\sum_{l=1}^{k}\mathbf{I}_S(\psi, \mathbf{W}_l)\right)^{-1}\sum_{m=1}^{M}\sum_{n=1}^{N}\left(\left(\frac{\partial h_{mn}(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}^{\mathrm{T}}}\right)^{\mathrm{H}}\frac{\partial h_{mn}(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}^{\mathrm{T}}}\right)\right\}$$

$$= \frac{1}{MN}\mathrm{Tr}\left\{\left(\sum_{l=1}^{k}\mathbf{I}_S(\psi, \mathbf{W}_l)\right)^{-1}\left(\frac{\partial \mathbf{h}}{\partial \boldsymbol{\psi}^{\mathrm{T}}}\right)^{\mathrm{H}}\frac{\partial \mathbf{h}}{\partial \boldsymbol{\psi}^{\mathrm{T}}}\right\},$$

$$\stackrel{(b)}{=} \frac{1}{MN}\mathrm{Tr}\left\{\left(\sum_{l=1}^{k}\mathbf{I}_S(\psi, \mathbf{W}_l)\right)^{-1}\mathbf{V}^{\mathrm{H}}\mathbf{V}\right\}, \qquad (77)$$

where Step (a) is obtained by substituting (76) into (77) and Step (b) is due to the definition of $\mathbf{V}$ in (21).

As for the Fisher information matrix in (22), we can obtain $\frac{\partial \log p_S(\mathbf{y}_l|\boldsymbol{\psi}, \mathbf{W}_l)}{\partial \beta^{\mathrm{re}}}$ as follows:

$$\frac{\partial \log p_S(\mathbf{y}_l|\boldsymbol{\psi}, \mathbf{W}_l)}{\partial \beta^{\mathrm{re}}} = -\frac{1}{\sigma_z^2}\left(\mathbf{y}_l - |\mathbf{s}|\mathbf{W}_l^{\mathrm{H}}\mathbf{h}\right)^{\mathrm{H}}\left(-|\mathbf{s}|\mathbf{W}_l^{\mathrm{H}}\frac{\partial \mathbf{h}}{\partial \beta^{\mathrm{re}}}\right)$$

$$+ \frac{1}{\sigma_z^2}\left(|\mathbf{s}|\mathbf{W}_l^{\mathrm{H}}\frac{\partial \mathbf{h}}{\partial \beta^{\mathrm{re}}}\right)^{\mathrm{H}}\left(\mathbf{y}_l - |\mathbf{s}|\mathbf{W}_l^{\mathrm{H}}\mathbf{h}\right)$$

$$= \frac{2|\mathbf{s}|}{\sigma_z^2}\mathrm{Re}\left\{\left(\mathbf{y}_l - |\mathbf{s}|\mathbf{W}_l^{\mathrm{H}}\mathbf{h}\right)^{\mathrm{H}}\left(\mathbf{W}_l^{\mathrm{H}}\frac{\partial \mathbf{h}}{\partial \beta^{\mathrm{re}}}\right)\right\}$$

$$= \frac{2|\mathbf{s}|}{\sigma_z^2}\mathrm{Re}\left\{\mathbf{z}_l^{\mathrm{H}}\mathbf{W}_l^{\mathrm{H}}\frac{\partial \mathbf{h}}{\partial \beta^{\mathrm{re}}}\right\}. \qquad (78)$$

Similarly, $\frac{\partial \log p_S(\mathbf{y}_l|\boldsymbol{\psi}, \mathbf{W}_l)}{\partial \beta^{\mathrm{im}}}$, $\frac{\partial \log p_S(\mathbf{y}_l|\boldsymbol{\psi}, \mathbf{W}_l)}{\partial x_1}$, and $\frac{\partial \log p_S(\mathbf{y}_l|\boldsymbol{\psi}, \mathbf{W}_l)}{\partial x_2}$ are given as

$$\begin{cases} \dfrac{\partial \log p_S(\mathbf{y}_l|\boldsymbol{\psi}, \mathbf{W}_l)}{\partial \beta^{\mathrm{im}}} = \dfrac{2|\mathbf{s}|}{\sigma_z^2}\mathrm{Re}\left\{\mathbf{z}_l^{\mathrm{H}}\mathbf{W}_l^{\mathrm{H}}\dfrac{\partial \mathbf{h}}{\partial \beta^{\mathrm{im}}}\right\} \\ \dfrac{\partial \log p_S(\mathbf{y}_l|\boldsymbol{\psi}, \mathbf{W}_l)}{\partial x_1} = \dfrac{2|\mathbf{s}|}{\sigma_z^2}\mathrm{Re}\left\{\mathbf{z}_l^{\mathrm{H}}\mathbf{W}_l^{\mathrm{H}}\dfrac{\partial \mathbf{h}}{\partial x_1}\right\} \\ \dfrac{\partial \log p_S(\mathbf{y}_l|\boldsymbol{\psi}, \mathbf{W}_l)}{\partial x_2} = \dfrac{2|\mathbf{s}|}{\sigma_z^2}\mathrm{Re}\left\{\mathbf{z}_l^{\mathrm{H}}\mathbf{W}_l^{\mathrm{H}}\dfrac{\partial \mathbf{h}}{\partial x_2}\right\}. \end{cases} \qquad (79)$$

Hence, the gradient of $\log p_S(\mathbf{y}_l|\boldsymbol{\psi}, \mathbf{W}_l)$ is obtained as follows:

$$\frac{\partial \log p_S(\mathbf{y}_l|\boldsymbol{\psi}, \mathbf{W}_l)}{\partial \boldsymbol{\psi}} = \frac{2|\mathbf{s}|}{\sigma_z^2}\mathrm{Re}\left\{\begin{bmatrix} \mathbf{z}_l^{\mathrm{H}}\mathbf{W}_l^{\mathrm{H}}\frac{\partial \mathbf{h}}{\partial \beta^{\mathrm{re}}} \\ \mathbf{z}_l^{\mathrm{H}}\mathbf{W}_l^{\mathrm{H}}\frac{\partial \mathbf{h}}{\partial \beta^{\mathrm{im}}} \\ \mathbf{z}_l^{\mathrm{H}}\mathbf{W}_l^{\mathrm{H}}\frac{\partial \mathbf{h}}{\partial x_1} \\ \mathbf{z}_l^{\mathrm{H}}\mathbf{W}_l^{\mathrm{H}}\frac{\partial \mathbf{h}}{\partial x_2} \end{bmatrix}\right\}$$

$$= \frac{2|\mathbf{s}|}{\sigma_z^2}\mathrm{Re}\left\{\left(\mathbf{z}_l^{\mathrm{H}}\mathbf{W}_l^{\mathrm{H}}\mathbf{V}\right)^{\mathrm{T}}\right\}. \qquad (80)$$

With the help of (80), we can obtain that

$$\frac{\partial \log p_S(\mathbf{y}_l|\boldsymbol{\psi}, \mathbf{W}_l)}{\partial \boldsymbol{\psi}^{\mathrm{T}}} = \left(\frac{\partial \log p_S(\mathbf{y}_l|\boldsymbol{\psi}, \mathbf{W}_l)}{\partial \boldsymbol{\psi}}\right)^{\mathrm{T}}$$

$$= \frac{2|\mathbf{s}|}{\sigma_z^2}\mathrm{Re}\left\{\mathbf{z}_l^{\mathrm{H}}\mathbf{W}_l^{\mathrm{H}}\mathbf{V}\right\}. \qquad (81)$$

Substituting (80) and (81) into (22), the Fisher information matrix is given as follows:

$$\mathbf{I}_S(\boldsymbol{\psi}, \mathbf{W}_l) \triangleq \mathbb{E}\left[\frac{\partial \log p_S(\mathbf{y}_l|\boldsymbol{\psi}, \mathbf{W}_l)}{\partial \boldsymbol{\psi}} \cdot \frac{\partial \log p_S(\mathbf{y}_l|\boldsymbol{\psi}, \mathbf{W}_l)}{\partial \boldsymbol{\psi}^{\mathrm{T}}}\right]$$

$$= \frac{4|\mathbf{s}|^2}{\sigma_z^4}\mathbb{E}\left[\mathrm{Re}\left\{\left(\mathbf{z}_l^{\mathrm{H}}\mathbf{W}_l^{\mathrm{H}}\mathbf{V}\right)^{\mathrm{T}}\right\}\mathrm{Re}\left\{\mathbf{z}_l^{\mathrm{H}}\mathbf{W}_l^{\mathrm{H}}\mathbf{V}\right\}\right]$$

$$\stackrel{(c)}{=} \frac{2|\mathbf{s}|^2}{\sigma_z^4}\mathbb{E}\left[\mathrm{Re}\left\{\left(\mathbf{z}_l^{\mathrm{H}}\mathbf{W}_l^{\mathrm{H}}\mathbf{V}\right)^{\mathrm{T}}\mathbf{z}_l^{\mathrm{H}}\mathbf{W}_l^{\mathrm{H}}\mathbf{V}\right\}\right]$$

$$+ \frac{2|\mathbf{s}|^2}{\sigma_z^4} \mathbb{E}\left[\mathrm{Re}\left\{\left(\mathbf{z}_l^{\mathrm{H}}\mathbf{W}_l^{\mathrm{H}}\mathbf{V}\right)^{\mathrm{H}}\mathbf{z}_l^{\mathrm{H}}\mathbf{W}_l^{\mathrm{H}}\mathbf{V}\right\}\right]$$

$$\overset{(d)}{=} \frac{2|\mathbf{s}|^2}{\sigma_z^4}\mathbb{E}\left[\mathrm{Re}\left\{\left(\mathbf{z}_l^{\mathrm{H}}\mathbf{W}_l^{\mathrm{H}}\mathbf{V}\right)^{\mathrm{H}}\mathbf{z}_l^{\mathrm{H}}\mathbf{W}_l^{\mathrm{H}}\mathbf{V}\right\}\right]$$

$$\overset{(e)}{=} \frac{2|\mathbf{s}|^2}{\sigma_z^2}\mathrm{Re}\left\{\mathbf{V}^{\mathrm{H}}\mathbf{W}_l\mathbf{W}_l^{\mathrm{H}}\mathbf{V}\right\}, \tag{82}$$

where in Step (c) we have used the following property of $\mathrm{Re}\{\cdot\}$:

$$\mathrm{Re}\{\mathbf{u}\}\,\mathrm{Re}\left\{\mathbf{v}^{\mathrm{T}}\right\} = \frac{1}{2}\mathrm{Re}\left\{\mathbf{u}\mathbf{v}^{\mathrm{T}}\right\} + \frac{1}{2}\mathrm{Re}\left\{\bar{\mathbf{u}}\mathbf{v}^{\mathrm{T}}\right\} \tag{83}$$

with $\mathbf{u}$, $\mathbf{v}$ denoting column vectors and $\bar{\mathbf{u}}$ denoting the conjugate of $\mathbf{u}$. Step (d) is due to the exchangeability of $\mathbb{E}[\cdot]$ and $\mathrm{Re}\{\cdot\}$:

$$\mathbb{E}\left[\mathrm{Re}\left\{\left(\mathbf{z}_l^{\mathrm{H}}\mathbf{W}_l^{\mathrm{H}}\mathbf{V}\right)^{\mathrm{T}}\mathbf{z}_l^{\mathrm{H}}\mathbf{W}_l^{\mathrm{H}}\mathbf{V}\right\}\right]$$

$$= \mathrm{Re}\left\{\mathbb{E}\left[\left(\mathbf{z}_l^{\mathrm{H}}\mathbf{W}_l^{\mathrm{H}}\mathbf{V}\right)^{\mathrm{T}}\mathbf{z}_l^{\mathrm{H}}\mathbf{W}_l^{\mathrm{H}}\mathbf{V}\right]\right\}$$

$$= \mathrm{Re}\left\{\left(\mathbf{W}_l^{\mathrm{H}}\mathbf{V}\right)^{\mathrm{T}}\mathbb{E}\left[\left(\mathbf{z}_l^{\mathrm{H}}\right)^{\mathrm{T}}\mathbf{z}_l^{\mathrm{H}}\right]\mathbf{W}_l^{\mathrm{H}}\mathbf{V}\right\}$$

$$\overset{(f)}{=} \mathbf{0}. \tag{84}$$

Step (e) is due to the *i.i.d.* circularly symmetric complex Gaussian property of each element of $\mathbf{z}_l$, which means that $\mathbb{E}\left[\mathbf{z}_l\mathbf{z}_l^{\mathrm{H}}\right] = \sigma_z^2\mathbf{J}_3$, where $\mathbf{J}_3$ is the 3-order identity matrix. Step (f) in (84) results from the property of complex Gaussian noise:

$$\mathbb{E}\left[\left(\mathbf{z}_l^{\mathrm{H}}\right)^{\mathrm{T}}\mathbf{z}_l^{\mathrm{H}}\right] = \mathbf{0}. \tag{85}$$

Therefore, the Fisher information matrix is derived in (82) and Lemma 2 is proved in the end.

APPENDIX C
PROOF OF LEMMA 3

Lemma 3 is proved in three steps:

Step 1: *We prove that $C_S^{\min}(\boldsymbol{\psi})$ and $\left\{\boldsymbol{\Delta}_{S,1}^*, \boldsymbol{\Delta}_{S,2}^*, \boldsymbol{\Delta}_{S,3}^*\right\}$ are unrelated to the equivalent channel gain β.*

The basic method is block matrix inversion. We first rewrite the Jacobian matrix $\mathbf{V}$ in (21) as follows:

$$\mathbf{V} = [\mathbf{V}_1, \beta\mathbf{V}_2], \tag{86}$$

where $\mathbf{V}_1$ and $\mathbf{V}_2$ are given by

$$\begin{cases} \mathbf{V}_1 \triangleq [\mathbf{a}(\mathbf{x}), j\mathbf{a}(\mathbf{x})] \\ \mathbf{V}_2 \triangleq \left[\dfrac{\partial\mathbf{a}(\mathbf{x})}{\partial x_1}, \dfrac{\partial\mathbf{a}(\mathbf{x})}{\partial x_2}\right]. \end{cases} \tag{87}$$

It is clear that both $\mathbf{V}_1$ and $\mathbf{V}_2$ are unrelated to β. Besides, we can obtain the following properties of $\mathbf{V}_1$:

$$\begin{cases} \mathbf{V}_1\mathbf{V}_1^{\mathrm{T}} = \mathbf{0} \\ \bar{\mathbf{V}}_1\mathbf{V}_1^{\mathrm{H}} = \mathbf{0}, \end{cases} \tag{88}$$

where $\bar{\mathbf{V}}_1$ denotes the conjugate of $\mathbf{V}_1$.

With the help of the Jacobian matrix $\mathbf{V}$ in (86), the Fisher information matrix in (22) can be divided into four 2×2 matrices as follows:

$$\mathbf{I}_S(\boldsymbol{\psi}, \mathbf{W}) = \frac{2|\mathbf{s}|^2}{\sigma_z^2}\mathrm{Re}\left\{\mathbf{V}^{\mathrm{H}}\mathbf{W}\mathbf{W}^{\mathrm{H}}\mathbf{V}\right\}$$

$$= \frac{2|\mathbf{s}|^2}{\sigma_z^2}\begin{bmatrix} \mathrm{Re}\left\{\mathbf{V}_1^{\mathrm{H}}\mathbf{W}\mathbf{W}^{\mathrm{H}}\mathbf{V}_1\right\} & \mathrm{Re}\left\{\beta\mathbf{V}_1^{\mathrm{H}}\mathbf{W}\mathbf{W}^{\mathrm{H}}\mathbf{V}_2\right\} \\ \mathrm{Re}\left\{\bar{\beta}\mathbf{V}_2^{\mathrm{H}}\mathbf{W}\mathbf{W}^{\mathrm{H}}\mathbf{V}_1\right\} & |\beta|^2\mathrm{Re}\left\{\mathbf{V}_2^{\mathrm{H}}\mathbf{W}\mathbf{W}^{\mathrm{H}}\mathbf{V}_2\right\} \end{bmatrix}$$

$$= \frac{2|\mathbf{s}|^2}{\sigma_z^2}\begin{bmatrix} \mathbf{A} & \mathrm{Re}\{\beta\mathbf{B}\} \\ \mathrm{Re}\left\{\bar{\beta}\mathbf{B}^{\mathrm{H}}\right\} & |\beta|^2\mathbf{D} \end{bmatrix}, \tag{89}$$

where $\bar{\beta}$ denotes the conjugate of β and $\mathbf{A}$, $\mathbf{B}$, $\mathbf{D}$ are defined as:

$$\begin{cases} \mathbf{A} \triangleq \mathrm{Re}\left\{\mathbf{V}_1^{\mathrm{H}}\mathbf{W}\mathbf{W}^{\mathrm{H}}\mathbf{V}_1\right\} = \left\|\mathbf{W}^{\mathrm{H}}\mathbf{a}(\mathbf{x})\right\|_2^2\mathbf{J}_2 \\ \mathbf{B} \triangleq \mathbf{V}_1^{\mathrm{H}}\mathbf{W}\mathbf{W}^{\mathrm{H}}\mathbf{V}_2. \\ \mathbf{D} \triangleq \mathrm{Re}\left\{\mathbf{V}_2^{\mathrm{H}}\mathbf{W}\mathbf{W}^{\mathrm{H}}\mathbf{V}_2\right\} \end{cases} \tag{90}$$

with $\mathbf{J}_2$ denoting the 2-order identity matrix. By combining (88) and (90), we can obtain the properties of $\mathbf{B}$:

$$\begin{cases} \mathbf{B}^{\mathrm{H}}\bar{\mathbf{B}} = \mathbf{0} \\ \mathbf{B}^{\mathrm{T}}\mathbf{B} = \mathbf{0} \\ \mathbf{B}^{\mathrm{H}}\mathbf{V}_1^{\mathrm{T}} = \mathbf{V}_1\bar{\mathbf{B}} = \mathbf{0} \\ \mathbf{B}^{\mathrm{T}}\mathbf{V}_1^{\mathrm{H}} = \bar{\mathbf{V}}_1\mathbf{B} = \mathbf{0}. \end{cases} \tag{91}$$

By using the block matrix inversion method, the inverse of the Fisher information matrix in (89) is given by

$$\mathbf{I}_S(\boldsymbol{\psi}, \mathbf{W})^{-1} = \frac{\sigma_z^2}{2|\mathbf{s}|^2}\left\{\mathbf{I}_{ip_1} + \mathbf{I}_{ip_2}(\beta)\right\}, \tag{92}$$

where $\mathbf{I}_{ip_1}$ and $\mathbf{I}_{ip_2}(\beta)$ are defined in (93) and (94) [as shown at the bottom of the next page]:

$$\mathbf{I}_{ip_1} \triangleq \begin{bmatrix} \mathbf{A}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \tag{93}$$

The middle part of $\mathbf{I}_{ip_2}$, i.e., $(|\beta|^2\mathbf{D} - \mathrm{Re}\left\{\bar{\beta}\mathbf{B}^{\mathrm{H}}\right\}\mathbf{A}^{-1}\mathrm{Re}\{\beta\mathbf{B}\})$, can be rewritten as follows:

$$|\beta|^2\mathbf{D} - \mathrm{Re}\left\{\bar{\beta}\mathbf{B}^{\mathrm{H}}\right\}\mathbf{A}^{-1}\mathrm{Re}\{\beta\mathbf{B}\}$$

$$= |\beta|^2\mathbf{D} - \frac{\bar{\beta}\mathbf{B}^{\mathrm{H}} + \beta\mathbf{B}^{\mathrm{T}}}{2}\mathbf{A}^{-1}\frac{\beta\mathbf{B} + \bar{\beta}\bar{\mathbf{B}}}{2}$$

$$\overset{(a)}{=} |\beta|^2\mathbf{D} - \frac{\bar{\beta}\mathbf{B}^{\mathrm{H}}\mathbf{A}^{-1}\beta\mathbf{B} + \beta\mathbf{B}^{\mathrm{T}}\mathbf{A}^{-1}\bar{\beta}\bar{\mathbf{B}}}{4}$$

$$\overset{(b)}{=} |\beta|^2\mathbf{D} - \frac{\mathrm{Re}\left\{\bar{\beta}\mathbf{B}^{\mathrm{H}}\mathbf{A}^{-1}\beta\mathbf{B}\right\}}{2}$$

$$= |\beta|^2\left(\mathbf{D} - \frac{\mathrm{Re}\left\{\mathbf{B}^{\mathrm{H}}\mathbf{A}^{-1}\mathbf{B}\right\}}{2}\right)$$

$$\overset{(c)}{=} |\beta|^2\mathbf{I}_s, \tag{95}$$

where Step (a) results from the properties of $\mathbf{B}$ in (91) and the definition of $\mathbf{A}$ in (90), Step (b) is due to the fact that $\mathbf{A}$ is a real matrix and Step (c) is due to the definition of $\mathbf{I}_s$:

$$\mathbf{I}_s \triangleq \mathbf{D} - \frac{\mathrm{Re}\left\{\mathbf{B}^{\mathrm{H}}\mathbf{A}^{-1}\mathbf{B}\right\}}{2}. \tag{96}$$

Therefore, we can rewrite $\mathbf{I}_{ip_2}$ in (94) as follows:

$$\mathbf{I}_{ip_2}(\beta) = \begin{bmatrix} \mathbf{A}^{-1}\mathrm{Re}\left\{\beta\mathbf{B}\right\} \\ -\mathbf{J}_2 \end{bmatrix} \left(|\beta|^2\mathbf{I}_s\right)^{-1}\left[\mathrm{Re}\left\{\bar{\beta}\mathbf{B}^{\mathrm{H}}\right\}\mathbf{A}^{-1} -\mathbf{J}_2\right]. \tag{97}$$

By combining (26) and (92), we can obtain that

$$C_S(\boldsymbol{\psi}, \mathbf{W}) = \frac{1}{MN}\mathrm{Tr}\left\{(\mathbf{I}_S(\boldsymbol{\psi}, \mathbf{W}))^{-1}\mathbf{V}^{\mathrm{H}}\mathbf{V}\right\}$$
$$= \frac{1}{MN}\frac{\sigma_z^2}{2|\mathbf{s}|^2}\left(\mathrm{Tr}\left\{\mathbf{I}_{ip_1}\mathbf{V}^{\mathrm{H}}\mathbf{V}\right\} + \mathrm{Tr}\left\{\mathbf{I}_{ip_2}(\beta)\mathbf{V}^{\mathrm{H}}\mathbf{V}\right\}\right)$$
$$\overset{(d)}{=} \frac{1}{MN}\frac{\sigma_z^2}{2|\mathbf{s}|^2}\left(\mathrm{Tr}\left\{\mathbf{A}^{-1}\mathbf{V}_1^{\mathrm{H}}\mathbf{V}_1\right\} + \mathrm{Tr}\left\{\mathbf{I}_{ip_2}(\beta)\mathbf{V}^{\mathrm{H}}\mathbf{V}\right\}\right), \tag{98}$$

where Step (d) is by substituting (86) and (93) into (98). Since both $\mathbf{V}_1$ in (87) and $\mathbf{A}$ in (90) are unrelated to the equivalent channel gain β, the first part of (98), i.e., $\mathrm{Tr}\left\{\mathbf{A}^{-1}\mathbf{V}_1^{\mathrm{H}}\mathbf{V}_1\right\}$ are unrelated to β. By substituting (86) and (97), we can obtain the second part of (98), i.e., $\mathrm{Tr}\left\{\mathbf{I}_{ip_2}(\beta)\mathbf{V}^{\mathrm{H}}\mathbf{V}\right\}$ in (99), as shown at the next page, where Step (e) and Step (f) follow the properties of the $\mathbf{B}$ in (91) and the definition of $\mathbf{A}$ in (90). It is clear that $\mathrm{Tr}\left\{\mathbf{I}_{ip_2}(\beta)\mathbf{V}^{\mathrm{H}}\mathbf{V}\right\}$ is also unrelated to β because none of the matrix $\mathbf{A}, \mathbf{B}, \mathbf{V}_1, \mathbf{V}_2$ is related to β. Hence, $C_S(\boldsymbol{\psi}, \mathbf{W})$ is unrelated to β.

Since $C_S(\boldsymbol{\psi}, \mathbf{W})$ is unrelated to β, the minimum CRLB $C_S^{\min}(\boldsymbol{\psi})$ in (27) and the optimal EBM $\mathbf{W}_S^*$ are also unrelated to β. Hence, the optimal set of exploration offsets $\left\{\boldsymbol{\Delta}_{S,1}^*, \boldsymbol{\Delta}_{S,2}^*, \boldsymbol{\Delta}_{S,3}^*\right\}$ is unrelated to the equivalent channel gain β.

Step 2: *We prove that $C_S^{\min}(\boldsymbol{\psi})$ and $\left\{\boldsymbol{\Delta}_{S,1}^*, \boldsymbol{\Delta}_{S,2}^*, \boldsymbol{\Delta}_{S,3}^*\right\}$ are unrelated to the DPV $\mathbf{x}$.*

Consider the CRLB in (26). we will first prove that the Fisher information matrix $\mathbf{I}_S(\boldsymbol{\psi}, \mathbf{W})$ is unrelated to the DPV $\mathbf{x}$. Next, we will prove that $\mathbf{V}^{\mathrm{H}}\mathbf{V}$ is also unrelated to $\mathbf{x}$. Then it is clear that the minimum CRLB and the optimal set of exploration offsets $\left\{\boldsymbol{\Delta}_{S,1}^*, \boldsymbol{\Delta}_{S,2}^*, \boldsymbol{\Delta}_{S,3}^*\right\}$ are unrelated to $\mathbf{x}$.

The Fisher information matrix in (89) tells us that only $\mathbf{W}^{\mathrm{H}}\mathbf{V}$ may be related to $\mathbf{x}$, which is given by

$$\mathbf{W}^{\mathrm{H}}\mathbf{V} = \left[\mathbf{W}^{\mathrm{H}}\mathbf{a}(\mathbf{x}), j\mathbf{W}^{\mathrm{H}}\mathbf{a}(\mathbf{x}), \beta\mathbf{W}^{\mathrm{H}}\frac{\partial\mathbf{a}(\mathbf{x})}{\partial x_1}, \beta\mathbf{W}^{\mathrm{H}}\frac{\partial\mathbf{a}(\mathbf{x})}{\partial x_2}\right] \tag{100}$$

with $\mathbf{W}^{\mathrm{H}}\mathbf{a}(\mathbf{x})$, $\mathbf{W}^{\mathrm{H}}\frac{\partial\mathbf{a}(\mathbf{x})}{\partial x_1}$ and $\mathbf{W}^{\mathrm{H}}\frac{\partial\mathbf{a}(\mathbf{x})}{\partial x_2}$ expanded as follows:

$$\begin{cases} \mathbf{W}^{\mathrm{H}}\mathbf{a}(\mathbf{x}) = \left[\mathbf{w}_1^{\mathrm{H}}\mathbf{a}(\mathbf{x}), \mathbf{w}_2^{\mathrm{H}}\mathbf{a}(\mathbf{x}), \mathbf{w}_3^{\mathrm{H}}\mathbf{a}(\mathbf{x})\right]^{\mathrm{T}} \\ \mathbf{W}^{\mathrm{H}}\frac{\partial\mathbf{a}(\mathbf{x})}{\partial x_1} = \left[\mathbf{w}_1^{\mathrm{H}}\frac{\partial\mathbf{a}(\mathbf{x})}{\partial x_1}, \mathbf{w}_2^{\mathrm{H}}\frac{\partial\mathbf{a}(\mathbf{x})}{\partial x_1}, \mathbf{w}_3^{\mathrm{H}}\frac{\partial\mathbf{a}(\mathbf{x})}{\partial x_1}\right]^{\mathrm{T}} \\ \mathbf{W}^{\mathrm{H}}\frac{\partial\mathbf{a}(\mathbf{x})}{\partial x_2} = \left[\mathbf{w}_1^{\mathrm{H}}\frac{\partial\mathbf{a}(\mathbf{x})}{\partial x_2}, \mathbf{w}_2^{\mathrm{H}}\frac{\partial\mathbf{a}(\mathbf{x})}{\partial x_2}, \mathbf{w}_3^{\mathrm{H}}\frac{\partial\mathbf{a}(\mathbf{x})}{\partial x_2}\right]^{\mathrm{T}}. \end{cases} \tag{101}$$

Since the EBVs are of the steering vector forms, i.e., $\mathbf{w}_i = \frac{1}{\sqrt{MN}}\mathbf{a}(\mathbf{x} + \boldsymbol{\Delta}_i)$, where $\boldsymbol{\Delta}_i = [\delta_{i1}, \delta_{i2}]^{\mathrm{T}}$ denotes the i-th exploration offset, the elements of $\mathbf{W}^{\mathrm{H}}\mathbf{a}(\mathbf{x})$ and $\mathbf{W}^{\mathrm{H}}\frac{\partial\mathbf{a}(\mathbf{x})}{\partial x_1}$ can be written in (102) and (103) [as shown at the next page].

$$\mathbf{w}_i^{\mathrm{H}}\mathbf{a}(\mathbf{x}) = \frac{1}{\sqrt{MN}}\mathbf{a}(\mathbf{x} + \boldsymbol{\Delta}_i)^{\mathrm{H}}\mathbf{a}(\mathbf{x})$$
$$= \frac{1}{\sqrt{MN}}\sum_{m=1}^{M}\sum_{n=1}^{N}e^{-j2\pi\left[\frac{(m-1)\delta_{i1}}{M} + \frac{(n-1)\delta_{i2}}{N}\right]}$$
$$= \frac{1}{\sqrt{MN}}\frac{\sin(\pi\delta_{i1})}{\sin\left(\frac{\pi\delta_{i1}}{M}\right)}\frac{\sin(\pi\delta_{i2})}{\sin\left(\frac{\pi\delta_{i2}}{N}\right)}e^{-j\pi\left(\frac{M-1}{M}\delta_{i1} + \frac{N-1}{N}\delta_{i2}\right)}, \tag{102}$$

As shown in (102) and (103), both $\mathbf{w}_i^{\mathrm{H}}\mathbf{a}(\mathbf{x})$ and $\mathbf{w}_i^{\mathrm{H}}\frac{\partial\mathbf{a}(\mathbf{x})}{\partial x_1}$ are unrelated to the DPV $\mathbf{x}$. Similarly, $\mathbf{w}_i^{\mathrm{H}}\frac{\partial\mathbf{a}(\mathbf{x})}{\partial x_2}$ is also unrelated to $\mathbf{x}$. Therefore, $\mathbf{W}^{\mathrm{H}}\mathbf{V}$ in (100) is unrelated to $\mathbf{x}$. Hence, the whole Fisher information matrix in (89) is invariant to $\mathbf{x}$.

As for $\mathbf{V}^{\mathrm{H}}\mathbf{V}$, we write it in (104), as shown at the next page, which indicates that $\mathbf{V}^{\mathrm{H}}\mathbf{V}$ is unrelated to $\mathbf{x}$.

Now it is clear that the CRLB in (26), i.e., $C_S(\boldsymbol{\psi}, \mathbf{W})$, is unrelated to $\mathbf{x}$ because both the Fisher information matrix $\mathbf{I}_S(\boldsymbol{\psi}, \mathbf{W})$ and $\mathbf{V}^{\mathrm{H}}\mathbf{V}$ are unrelated to $\mathbf{x}$. Therefore, the minimum CRLB in (27) and the optimal set of exploration offsets $\left\{\boldsymbol{\Delta}_{S,1}^*, \boldsymbol{\Delta}_{S,2}^*, \boldsymbol{\Delta}_{S,3}^*\right\}$ are invariant to the DPV $\mathbf{x}$.

Step 3: *We prove that $C_S^{\min}(\boldsymbol{\psi})$ converges as $M, N \to +\infty$ and*

$$\lim_{M,N\to+\infty}C_S(\boldsymbol{\psi}, \widetilde{\mathbf{W}}_S^*) = \lim_{M,N\to+\infty}C_S^{\min}(\boldsymbol{\psi}),$$

Let us go into the asymptotic features of (26). According to (102) and (103), when the antenna number $M, N \to +\infty$, the limit of the i-th ($i = 1, 2, 3$) element of $\mathbf{W}^{\mathrm{H}}\mathbf{a}(\mathbf{x})$, $\mathbf{W}^{\mathrm{H}}\frac{\partial\mathbf{a}(\mathbf{x})}{\partial x_1}$

$$\mathbf{I}_{ip_2}(\beta) \triangleq \begin{bmatrix} \mathbf{A}^{-1}\mathrm{Re}\left\{\beta\mathbf{B}\right\} \\ -\mathbf{J}_2 \end{bmatrix} \left(|\beta|^2\mathbf{D} - \mathrm{Re}\left\{\bar{\beta}\mathbf{B}^{\mathrm{H}}\right\}\mathbf{A}^{-1}\mathrm{Re}\left\{\beta\mathbf{B}\right\}\right)^{-1}\left[\mathrm{Re}\left\{\bar{\beta}\mathbf{B}^{\mathrm{H}}\right\}\mathbf{A}^{-1} -\mathbf{J}_2\right]. \tag{94}$$

$$\begin{aligned}
\mathrm{Tr}\left\{\mathbf{I}_{ip_2}\left(\beta\right)\mathbf{V}^{\mathrm{H}}\mathbf{V}\right\} &= \mathrm{Tr}\left\{\begin{bmatrix}\mathbf{A}^{-1}\mathrm{Re}\left\{\beta\mathbf{B}\right\}\\ -\mathbf{J}_2\end{bmatrix}\left(|\beta|^2\mathbf{I}_s\right)^{-1}\begin{bmatrix}\mathrm{Re}\left\{\bar{\beta}\mathbf{B}^{\mathrm{H}}\right\}\mathbf{A}^{-1} & -\mathbf{J}_2\end{bmatrix}\begin{bmatrix}\mathbf{V}_1^{\mathrm{H}}\mathbf{V}_1 & \beta\mathbf{V}_1^{\mathrm{H}}\mathbf{V}_2\\ \bar{\beta}\mathbf{V}_2^{\mathrm{H}}\mathbf{V}_1 & |\beta|^2\mathbf{V}_2^{\mathrm{H}}\mathbf{V}_2\end{bmatrix}\right\}\\[2mm]
&= \mathrm{Tr}\left\{\mathbf{A}^{-1}\mathrm{Re}\left\{\beta\mathbf{B}\right\}\left(|\beta|^2\mathbf{I}_s\right)^{-1}\left(\mathrm{Re}\left\{\bar{\beta}\mathbf{B}^{\mathrm{H}}\right\}\mathbf{A}^{-1}\mathbf{V}_1^{\mathrm{H}}\mathbf{V}_1 - \bar{\beta}\mathbf{V}_2^{\mathrm{H}}\mathbf{V}_1\right)\right\}\\
&\quad + \mathrm{Tr}\left\{\left(|\beta|^2\mathbf{I}_s\right)^{-1}\left(|\beta|^2\mathbf{V}_2^{\mathrm{H}}\mathbf{V}_2 - \mathrm{Re}\left\{\bar{\beta}\mathbf{B}^{\mathrm{H}}\right\}\mathbf{A}^{-1}\beta\mathbf{V}_1^{\mathrm{H}}\mathbf{V}_2\right)\right\}\\[2mm]
&= \mathrm{Tr}\left\{\mathbf{A}^{-1}\frac{\beta\mathbf{B}+\bar{\beta}\bar{\mathbf{B}}}{2}\left(|\beta|^2\mathbf{I}_s\right)^{-1}\left(\frac{\bar{\beta}\mathbf{B}^{\mathrm{H}}+\beta\mathbf{B}^{\mathrm{T}}}{2}\mathbf{A}^{-1}\mathbf{V}_1^{\mathrm{H}}\mathbf{V}_1 - \bar{\beta}\mathbf{V}_2^{\mathrm{H}}\mathbf{V}_1\right)\right\}\\
&\quad + \mathrm{Tr}\left\{\left(|\beta|^2\mathbf{I}_s\right)^{-1}\left(|\beta|^2\mathbf{V}_2^{\mathrm{H}}\mathbf{V}_2 - \frac{\bar{\beta}\mathbf{B}^{\mathrm{H}}+\beta\mathbf{B}^{\mathrm{T}}}{2}\mathbf{A}^{-1}\beta\mathbf{V}_1^{\mathrm{H}}\mathbf{V}_2\right)\right\}\\[2mm]
&\overset{(e)}{=} \mathrm{Tr}\left\{\mathbf{A}^{-1}\frac{\beta\mathbf{B}+\bar{\beta}\bar{\mathbf{B}}}{2}\left(|\beta|^2\mathbf{I}_s\right)^{-1}\left(\frac{\bar{\beta}\mathbf{B}^{\mathrm{H}}}{2}\mathbf{A}^{-1}\mathbf{V}_1^{\mathrm{H}}\mathbf{V}_1 - \bar{\beta}\mathbf{V}_2^{\mathrm{H}}\mathbf{V}_1\right)\right\}\\
&\quad + \mathrm{Tr}\left\{\left(|\beta|^2\mathbf{I}_s\right)^{-1}\left(|\beta|^2\mathbf{V}_2^{\mathrm{H}}\mathbf{V}_2 - \frac{\bar{\beta}\mathbf{B}^{\mathrm{H}}}{2}\mathbf{A}^{-1}\beta\mathbf{V}_1^{\mathrm{H}}\mathbf{V}_2\right)\right\}\\[2mm]
&= \mathrm{Tr}\left\{\left(\frac{\bar{\beta}\mathbf{B}^{\mathrm{H}}}{2}\mathbf{A}^{-1}\mathbf{V}_1^{\mathrm{H}}\mathbf{V}_1 - \bar{\beta}\mathbf{V}_2^{\mathrm{H}}\mathbf{V}_1\right)\mathbf{A}^{-1}\frac{\beta\mathbf{B}+\bar{\beta}\bar{\mathbf{B}}}{2}\left(|\beta|^2\mathbf{I}_s\right)^{-1}\right\}\\
&\quad + \mathrm{Tr}\left\{\mathbf{I}_s^{-1}\left(\mathbf{V}_2^{\mathrm{H}}\mathbf{V}_2 - \frac{\mathbf{B}^{\mathrm{H}}\mathbf{A}^{-1}\mathbf{V}_1^{\mathrm{H}}\mathbf{V}_2}{2}\right)\right\}\\[2mm]
&\overset{(f)}{=} \mathrm{Tr}\left\{\left(\frac{\bar{\beta}\mathbf{B}^{\mathrm{H}}}{2}\mathbf{A}^{-1}\mathbf{V}_1^{\mathrm{H}}\mathbf{V}_1 - \bar{\beta}\mathbf{V}_2^{\mathrm{H}}\mathbf{V}_1\right)\mathbf{A}^{-1}\frac{\beta\mathbf{B}}{2}\left(|\beta|^2\mathbf{I}_s\right)^{-1}\right\}\\
&\quad + \mathrm{Tr}\left\{\mathbf{I}_s^{-1}\left(\mathbf{V}_2^{\mathrm{H}}\mathbf{V}_2 - \frac{\mathbf{B}^{\mathrm{H}}\mathbf{A}^{-1}\mathbf{V}_1^{\mathrm{H}}\mathbf{V}_2}{2}\right)\right\}\\[2mm]
&= \mathrm{Tr}\left\{\left(\frac{\mathbf{B}^{\mathrm{H}}}{2}\mathbf{A}^{-1}\mathbf{V}_1^{\mathrm{H}}\mathbf{V}_1 - \mathbf{V}_2^{\mathrm{H}}\mathbf{V}_1\right)\mathbf{A}^{-1}\frac{\mathbf{B}}{2}\mathbf{I}_s^{-1}\right\}\\
&\quad + \mathrm{Tr}\left\{\mathbf{I}_s^{-1}\left(\mathbf{V}_2^{\mathrm{H}}\mathbf{V}_2 - \frac{\mathbf{B}^{\mathrm{H}}\mathbf{A}^{-1}\mathbf{V}_1^{\mathrm{H}}\mathbf{V}_2}{2}\right)\right\}\\[2mm]
&= \mathrm{Tr}\left\{\mathbf{I}_s^{-1}\left(\frac{\mathbf{B}^{\mathrm{H}}\mathbf{A}^{-1}\mathbf{V}_1^{\mathrm{H}}\mathbf{V}_1\mathbf{A}^{-1}\mathbf{B}}{4} + \mathbf{V}_2^{\mathrm{H}}\mathbf{V}_2 - \frac{\mathbf{B}^{\mathrm{H}}\mathbf{A}^{-1}\mathbf{V}_1^{\mathrm{H}}\mathbf{V}_2 + \mathbf{V}_2^{\mathrm{H}}\mathbf{V}_1\mathbf{A}^{-1}\mathbf{B}}{2}\right)\right\},
\end{aligned}\tag{99}$$

$$\begin{aligned}
\mathbf{w}_i^{\mathrm{H}}\frac{\partial\mathbf{a}\left(\mathbf{x}\right)}{\partial x_1} &= \frac{1}{\sqrt{MN}}\mathbf{a}\left(\mathbf{x}+\boldsymbol{\Delta}_i\right)^{\mathrm{H}}\frac{\partial\mathbf{a}(\mathbf{x})}{\partial x_1} = \frac{1}{\sqrt{MN}}\left(\sum_{m=1}^{M}\sum_{n=1}^{N}j2\pi\frac{m-1}{M}e^{-j2\pi\left[\frac{(m-1)\delta_{i1}}{M}+\frac{(n-1)\delta_{i2}}{N}\right]}\right)\\
&= \frac{j2\pi}{M\sqrt{MN}}\left(\frac{\sin(\pi\delta_{i2})}{\sin\left(\frac{\pi\delta_{i2}}{N}\right)}e^{-j\pi\frac{N-1}{N}\delta_{i2}}\frac{(M-1)e^{-j2\pi\delta_{i1}}-Me^{-j2\pi\frac{M-1}{M}\delta_{i1}}+1}{\left[1-e^{-j2\pi\frac{\delta_{i1}}{M}}\right]^2}e^{-j2\pi\frac{\delta_{i1}}{M}}\right).
\end{aligned}\tag{103}$$

$$\begin{aligned}
\mathbf{V}^{\mathrm{H}}\mathbf{V} &= \begin{bmatrix}\mathbf{a}\left(\mathbf{x}\right)^{\mathrm{H}}\\ -j\mathbf{a}\left(\mathbf{x}\right)^{\mathrm{H}}\\ \bar{\beta}\frac{\partial\mathbf{a}(\mathbf{x})^{\mathrm{H}}}{\partial x_1}\\ \bar{\beta}\frac{\partial\mathbf{a}(\mathbf{x})^{\mathrm{H}}}{\partial x_2}\end{bmatrix}\begin{bmatrix}\mathbf{a}\left(\mathbf{x}\right), & j\mathbf{a}\left(\mathbf{x}\right), & \beta\frac{\partial\mathbf{a}\left(\mathbf{x}\right)}{\partial x_1}, & \beta\frac{\partial\mathbf{a}\left(\mathbf{x}\right)}{\partial x_2}\end{bmatrix}\\[2mm]
&= MN\begin{bmatrix}1 & j & j\pi\beta\frac{M-1}{M} & j\pi\beta\frac{N-1}{N}\\ -j & 1 & \pi\beta\frac{M-1}{M} & \pi\beta\frac{N-1}{N}\\ -j\pi\bar{\beta}\frac{M-1}{M} & \pi\bar{\beta}\frac{M-1}{M} & \frac{2}{3}\pi^2|\beta|^2\frac{(M-1)(2M-1)}{M^2} & \pi^2|\beta|^2\frac{(M-1)(N-1)}{MN}\\ -j\pi\bar{\beta}\frac{N-1}{N} & \pi\bar{\beta}\frac{N-1}{N} & \pi^2|\beta|^2\frac{(M-1)(N-1)}{MN} & \frac{2}{3}\pi^2|\beta|^2M\frac{(N-1)(2N-1)}{N^2}\end{bmatrix},
\end{aligned}\tag{104}$$

and $\mathbf{W}^{\mathrm{H}} \frac{\partial \mathbf{a}(\mathbf{x})}{\partial x_2}$ in (101) are given as follows:

$$
\begin{cases}
\lim\limits_{M,N \to +\infty} \dfrac{\mathbf{w}_i^{\mathrm{H}} \mathbf{a}(\mathbf{x})}{\sqrt{MN}} = \mathrm{Sa}\left[\pi \delta_{i1}\right] \mathrm{Sa}[\pi \delta_{i2}] e^{-j\pi(\delta_{i1}+\delta_{i2})} \\[2mm]
\lim\limits_{M,N \to +\infty} \dfrac{\mathbf{w}_i^{\mathrm{H}} \frac{\partial \mathbf{a}(\mathbf{x})}{\partial x_1}}{\sqrt{MN}} = j2\pi\, \mathrm{Sa}[\pi \delta_{i2}] \\[2mm]
\quad \times e^{-j\pi \delta_{i2}} \dfrac{e^{-j2\pi \delta_{i1}}(1+j2\pi \delta_{i1})-1}{(2\pi \delta_{i1})^2} \\[2mm]
\lim\limits_{M,N \to +\infty} \dfrac{\mathbf{w}_i^{\mathrm{H}} \frac{\partial \mathbf{a}(\mathbf{x})}{\partial x_2}}{\sqrt{MN}} = j2\pi\, \mathrm{Sa}[\pi \delta_{i1}] \\[2mm]
\quad \times e^{-j\pi \delta_{i1}} \dfrac{e^{-j2\pi \delta_{i2}}(1+j2\pi \delta_{i2})-1}{(2\pi \delta_{i2})^2}
\end{cases}
\tag{105}
$$

where $\mathrm{Sa}[t] \triangleq \frac{\sin t}{t}$. Hence, each element of $\mathbf{W}^{\mathrm{H}}\mathbf{V}/\sqrt{MN}$ in (100) converges when $M, N \to +\infty$, which results in that $\mathbf{I}_S(\boldsymbol{\psi}, \mathbf{W})/MN$ in (89) also converges. The limit is defined as follows:

$$
\mathbf{I}_l(\boldsymbol{\psi}, \mathbf{W}) \triangleq \lim_{M,N \to +\infty} \frac{1}{MN} \mathbf{I}_S(\boldsymbol{\psi}, \mathbf{W}).
\tag{106}
$$

The limit of $\mathbf{V}^{\mathrm{H}}\mathbf{V}$ in (104) is given by

$$
\lim_{M,N \to +\infty} \frac{1}{MN} \mathbf{V}^{\mathrm{H}}\mathbf{V}
$$

$$
= \begin{bmatrix}
1 & j & j\pi\beta & j\pi\beta \\
-j & 1 & \pi\beta & \pi\beta \\
-j\pi\bar{\beta} & \pi\bar{\beta} & \frac{4}{3}\pi^2|\beta|^2 & \pi^2|\beta|^2 \\
-j\pi\bar{\beta} & \pi\bar{\beta} & \pi^2|\beta|^2 & \frac{4}{3}\pi^2|\beta|^2
\end{bmatrix}
$$

$$
\triangleq \mathbf{H}_l.
\tag{107}
$$

By combining (106) and (107), we obtain the limit of $C_S(\boldsymbol{\psi}, \mathbf{W})$ in (26) as $M, N \to +\infty$:

$$
\lim_{M,N \to +\infty} (MN \times C_S(\boldsymbol{\psi}, \mathbf{W}))
$$

$$
= \lim_{M,N \to +\infty} \mathrm{Tr}\left\{ (\mathbf{I}_S(\boldsymbol{\psi}, \mathbf{W}))^{-1} \mathbf{V}^{\mathrm{H}}\mathbf{V} \right\}
$$

$$
= \lim_{M,N \to +\infty} \mathrm{Tr}\left\{ (MN\mathbf{I}_l(\boldsymbol{\psi}, \mathbf{W}))^{-1} \mathbf{V}^{\mathrm{H}}\mathbf{V} \right\}
$$

$$
= \lim_{M,N \to +\infty} \mathrm{Tr}\left\{ (\mathbf{I}_l(\boldsymbol{\psi}, \mathbf{W}))^{-1} \frac{1}{MN} \mathbf{V}^{\mathrm{H}}\mathbf{V} \right\}
$$

$$
= \mathrm{Tr}\left\{ (\mathbf{I}_l(\boldsymbol{\psi}, \mathbf{W}))^{-1} \mathbf{H}_l \right\},
\tag{108}
$$

which reveals that the CRLB in (27), i.e., $C_S(\boldsymbol{\psi}, \mathbf{W})$, converges. Hence, the minimum CRLB $C_S^{\min}(\boldsymbol{\psi})$ also converges.

Let

$$
\widetilde{\mathbf{W}}_S^* = \arg\min_{\mathbf{W}} \left(\lim_{M,N \to +\infty} C_S(\boldsymbol{\psi}, \mathbf{W}) \right).
\tag{109}
$$

Then we have

$$
\lim_{M,N \to +\infty} C_S^{\min}(\boldsymbol{\psi}) = \lim_{M,N \to +\infty} C_S(\boldsymbol{\psi}, \mathbf{W}^*)
\tag{110}
$$

$$
\stackrel{(g)}{\geq} \lim_{M,N \to +\infty} C_S(\boldsymbol{\psi}, \widetilde{\mathbf{W}}_S^*),
\tag{111}
$$

where Step (g) results from (109). On the other hand, we have

$$
\lim_{M,N \to +\infty} C_S^{\min}(\boldsymbol{\psi}) \stackrel{(h)}{\leq} \lim_{M,N \to +\infty} \left(C_S(\boldsymbol{\psi}, \widetilde{\mathbf{W}}_S^*) \right),
\tag{112}
$$

where Step (h) results from (27). Hence, we can obtain that

$$
\lim_{M,N \to +\infty} C_S^{\min}(\boldsymbol{\psi}) = \lim_{M,N \to +\infty} C_S(\boldsymbol{\psi}, \widetilde{\mathbf{W}}_S^*),
$$

Therefore, Lemma 3 gets proved.

APPENDIX D
PROOF OF THEOREM 1

According to (32) and (39), the tracking procedure can be rewritten as

$$
\hat{\boldsymbol{\psi}}_k = \hat{\boldsymbol{\psi}}_{k-1} + b_{S,k}\left(\mathbf{f}_{\boldsymbol{\psi}}\left(\hat{\boldsymbol{\psi}}_{k-1}\right) + \hat{\mathbf{z}}_k \right).
\tag{113}
$$

And it can be derived in (41) that

$$
\hat{\mathbf{z}}_k = \frac{2|\mathbf{s}|}{\sigma_z^2} \mathbf{I}_S\left(\hat{\boldsymbol{\psi}}_{k-1}, \mathbf{W}_k\right)^{-1}
\begin{bmatrix}
\mathrm{Re}\left\{\mathbf{e}_k^{\mathrm{H}}\mathbf{z}_k\right\} \\
\mathrm{Im}\left\{\mathbf{e}_k^{\mathrm{H}}\mathbf{z}_k\right\} \\
\mathrm{Re}\left\{\bar{\mathbf{e}}_{k1}^{\mathrm{H}}\mathbf{z}_k\right\} \\
\mathrm{Re}\left\{\bar{\mathbf{e}}_{k2}^{\mathrm{H}}\mathbf{z}_k\right\}
\end{bmatrix}.
\tag{114}
$$

Since $\hat{\mathbf{z}}_k \triangleq [\hat{z}_{k,1}, \hat{z}_{k,2}, \hat{z}_{k,3}]$ in (114) is composed of three *i.i.d.* circularly symmetric complex Gaussian random variables, the expectation of $\hat{\mathbf{z}}_k$ is $\mathbb{E}[\hat{\mathbf{z}}_k] = \mathbf{0}$ and the covariance matrix is given in (115), as shown at the bottom of the next page, where Step (a) is the result of (80) and Step (b) follows the definition of the Fisher information matrix in (22).

Assume $\{\mathcal{G}_k : k \geq 0\}$ is an increasing sequence of σ-fields of $\{\hat{\boldsymbol{\psi}}_0, \hat{\boldsymbol{\psi}}_1, \hat{\boldsymbol{\psi}}_2, \ldots\}$, i.e., $\mathcal{G}_{k-1} \subset \mathcal{G}_k$, where $\mathcal{G}_0 \stackrel{\Delta}{=} \sigma(\hat{\boldsymbol{\psi}}_0)$ and $\mathcal{G}_k \stackrel{\Delta}{=} \sigma(\hat{\boldsymbol{\psi}}_0, \hat{\mathbf{z}}_1, \ldots, \hat{\mathbf{z}}_k)$ for $k \geq 1$. Because the $\hat{\mathbf{z}}_k$'s are composed of *i.i.d.* circularly symmetric complex Gaussian random variables with zero mean, $\hat{\mathbf{z}}_k$ is independent of $\mathcal{G}_{k-1}$, and $\hat{\boldsymbol{\psi}}_{k-1} \in \mathcal{G}_{k-1}$. Hence, we have

$$
\mathbb{E}\left[\mathbf{f}_{\boldsymbol{\psi}}\left(\hat{\boldsymbol{\psi}}_{k-1}\right) + \hat{\mathbf{z}}_k \,\Big|\, \mathcal{G}_{k-1} \right]
$$

$$
= \mathbb{E}\left[\mathbf{f}_{\boldsymbol{\psi}}\left(\hat{\boldsymbol{\psi}}_{k-1}\right) \Big| \mathcal{G}_{k-1} \right] + \mathbb{E}\left[\hat{\mathbf{z}}_k \,\big|\, \mathcal{G}_{k-1} \right]
$$

$$
= \mathbf{f}_{\boldsymbol{\psi}}\left(\hat{\boldsymbol{\psi}}_{k-1}\right),
\tag{116}
$$

for $k \geq 1$ and $\boldsymbol{\varsigma}_k = \mathbf{f}_{\boldsymbol{\psi}}\left(\hat{\boldsymbol{\psi}}_{k-1}\right) + \hat{\mathbf{z}}_k$ is also independent of $\mathcal{G}_{k-1}$.

Theorem 5.2.1 in [37, Section 5.2.1] gives the conditions that ensure $\hat{\boldsymbol{\psi}}_k$ converges to a unique point with probability one when there are several stable points. Next, we will prove that if the step-size $b_{S,k}$ is given by (42) with any $\varepsilon_S > 0$ and $K_{S,0} \geq 0$, then the joint beam and channel tracking algorithm in (38) satisfies the corresponding conditions below:

1) Step-size requirements:

$$
\begin{cases}
b_{S,k} = \dfrac{\varepsilon_S}{k + K_{S,0}} \to 0, \\[2mm]
\displaystyle\sum_{k=1}^{+\infty} b_{S,k} = \sum_{k=1}^{+\infty} \frac{\varepsilon_S}{k + K_{S,0}} = +\infty, \\[2mm]
\displaystyle\sum_{k=1}^{+\infty} b_{S,k}^2 = \sum_{k=1}^{+\infty} \frac{\varepsilon_S^2}{(k + K_{S,0})^2} \le \sum_{l=1}^{+\infty} \frac{\varepsilon_S^2}{l^2} < +\infty.
\end{cases}
\tag{117}
$$

2) It is necessary to prove that

$$
\sup_k \mathbb{E}\left[\left\| \mathbf{f}_\psi\left(\hat{\boldsymbol{\psi}}_{k-1}\right) + \hat{\mathbf{z}}_k \right\|_2^2 \right] < +\infty.
$$

From (113) and (115), we have

$$
\mathbb{E}\left[\left\| \mathbf{f}_\psi\left(\hat{\boldsymbol{\psi}}_{k-1}\right) + \hat{\mathbf{z}}_k \right\|_2^2 \right]
$$
$$
= \mathbb{E}\left[\left\| \mathbf{f}_\psi\left(\hat{\boldsymbol{\psi}}_{k-1}\right) \right\|_2^2 + 2 \mathbf{f}_\psi\left(\hat{\boldsymbol{\psi}}_{k-1}\right)^{\mathrm{H}} \hat{\mathbf{z}}_k + \left\| \hat{\mathbf{z}}_k \right\|_2^2 \right]
$$
$$
\stackrel{(c)}{=} \mathbb{E}\left[\left\| \mathbf{f}_\psi\left(\hat{\boldsymbol{\psi}}_{k-1}\right) \right\|_2^2 \right] + \mathrm{Tr}\left\{ \mathbf{I}_S(\hat{\boldsymbol{\psi}}_{k-1}, \mathbf{W}_k)^{-1} \right\},
\tag{118}
$$

where Step (*c*) is due to (115) and that $\hat{\mathbf{z}}_k$ is independent of $\mathbf{f}_\psi\left(\hat{\boldsymbol{\psi}}_{k-1}\right)$.

From (44), we have

$$
\left\| \mathbf{f}_\psi\left(\hat{\boldsymbol{\psi}}_{k-1}\right) \right\|_2^2
$$
$$
\le \left\| \mathbf{I}_S(\hat{\boldsymbol{\psi}}_{k-1}, \mathbf{W}_k)^{-1} \right\|_{\mathrm{F}}^2
$$

$$
\cdot \left\| \frac{2|\mathbf{s}|^2}{\sigma_z^2}
\begin{bmatrix}
\mathrm{Re}\left\{ \mathbf{e}_k^{\mathrm{H}} \left(\beta \mathbf{W}_k^{\mathrm{H}} \mathbf{a}(\mathbf{x}) - \hat{\beta}_{k-1} \mathbf{e}_k \right) \right\} \\
\mathrm{Im}\left\{ \mathbf{e}_k^{\mathrm{H}} \left(\beta \mathbf{W}_k^{\mathrm{H}} \mathbf{a}(\mathbf{x}) - \hat{\beta}_{k-1} \mathbf{e}_k \right) \right\} \\
\mathrm{Re}\left\{ \tilde{\mathbf{e}}_{k1}^{\mathrm{H}} \left(\beta \mathbf{W}_k^{\mathrm{H}} \mathbf{a}(\mathbf{x}) - \hat{\beta}_{k-1} \mathbf{e}_k \right) \right\} \\
\mathrm{Re}\left\{ \tilde{\mathbf{e}}_{k2}^{\mathrm{H}} \left(\beta \mathbf{W}_k^{\mathrm{H}} \mathbf{a}(\mathbf{x}) - \hat{\beta}_{k-1} \mathbf{e}_k \right) \right\}
\end{bmatrix}
\right\|_2^2 .
\tag{119}
$$

As the Fisher information matrix is invertible, we get

$$
\left\| \mathbf{I}_S(\hat{\boldsymbol{\psi}}_{k-1}, \mathbf{W}_k)^{-1} \right\|_{\mathrm{F}}^2 < +\infty.
\tag{120}
$$

Besides, $\mathbf{W}_k = [\mathbf{w}_{k,1}, \mathbf{w}_{k,2}, \mathbf{w}_{k,3}]$, $\mathbf{e}_k = \mathbf{W}_k^{\mathrm{H}} \mathbf{a}(\hat{\mathbf{x}}_{k-1})$, $\tilde{\mathbf{e}}_{k1} = \hat{\beta}_{k-1} \mathbf{W}_k^{\mathrm{H}} \frac{\partial \mathbf{a}(\hat{\mathbf{x}}_{k-1})}{\partial x_1}$, $\tilde{\mathbf{e}}_{k2} = \hat{\beta}_{k-1} \mathbf{W}_k^{\mathrm{H}} \frac{\partial \mathbf{a}(\hat{\mathbf{x}}_{k-1})}{\partial x_2}$, hence we have

$$
\left| \mathbf{w}_{k,i}^H \mathbf{a}(\mathbf{x}) \right|
$$
$$
= \left| \frac{1}{\sqrt{MN}} \sum_{m=1}^{M} \sum_{n=1}^{N} e^{-j2\pi\left(\frac{(m-1)\delta_{k,i1}}{M} + \frac{(n-1)\delta_{k,i2}}{N} \right)} \right|
$$
$$
\le \frac{1}{\sqrt{MN}} \sum_{m=1}^{M} \sum_{n=1}^{N} \left| e^{-j2\pi\left(\frac{(m-1)\delta_{k,i1}}{M} + \frac{(n-1)\delta_{k,i2}}{N} \right)} \right|
$$
$$
= \sqrt{MN} < +\infty,
\tag{121}
$$
$$
\left| \mathbf{w}_{k,i}^H \frac{\partial \mathbf{a}(\mathbf{x})}{\partial x_1} \right|
$$
$$
= \left| \frac{1}{\sqrt{MN}} \sum_{m=1}^{M} \sum_{n=1}^{N} j2\pi \frac{m-1}{M} e^{-j2\pi\left(\frac{(m-1)\delta_{k,i1}}{M} + \frac{(n-1)\delta_{k,i2}}{N} \right)} \right|
$$

$$
\mathbb{E}\left[\left(\hat{\mathbf{z}}_k - \mathbb{E}\left[\hat{\mathbf{z}}_k \right] \right) \left(\hat{\mathbf{z}}_k - \mathbb{E}\left[\hat{\mathbf{z}}_k \right] \right)^{\mathrm{T}} \right]
$$
$$
= \frac{4|\mathbf{s}|^2}{\sigma_z^4} \mathbf{I}_S\left(\hat{\boldsymbol{\psi}}_{k-1}, \mathbf{W}_k\right)^{-1} \mathbb{E}\left\{
\begin{bmatrix}
\mathrm{Re}\{\mathbf{e}_k^{\mathrm{H}} \mathbf{z}_k\} \\
\mathrm{Im}\{\mathbf{e}_k^{\mathrm{H}} \mathbf{z}_k\} \\
\mathrm{Re}\{\tilde{\mathbf{e}}_{k1}^{\mathrm{H}} \mathbf{z}_k\} \\
\mathrm{Re}\{\tilde{\mathbf{e}}_{k2}^{\mathrm{H}} \mathbf{z}_k\}
\end{bmatrix}
\cdot
\begin{bmatrix}
\mathrm{Re}\{\mathbf{e}_k^{\mathrm{H}} \mathbf{z}_k\} \\
\mathrm{Im}\{\mathbf{e}_k^{\mathrm{H}} \mathbf{z}_k\} \\
\mathrm{Re}\{\tilde{\mathbf{e}}_{k1}^{\mathrm{H}} \mathbf{z}_k\} \\
\mathrm{Re}\{\tilde{\mathbf{e}}_{k2}^{\mathrm{H}} \mathbf{z}_k\}
\end{bmatrix}^{\mathrm{T}}
\right\} \mathbf{I}_S\left(\hat{\boldsymbol{\psi}}_{k-1}, \mathbf{W}_k\right)^{-1}
$$
$$
= \frac{4|\mathbf{s}|^2}{\sigma_z^4} \mathbf{I}_S\left(\hat{\boldsymbol{\psi}}_{k-1}, \mathbf{W}_k\right)^{-1} \mathbb{E}\left\{
\mathrm{Re}\left\{
\begin{bmatrix}
\mathbf{z}_k^{\mathrm{H}} \mathbf{W}_k^{\mathrm{H}} \frac{\partial \hat{\mathbf{h}}_{k-1}}{\partial \hat{\beta}_{k-1}^{\mathrm{re}}} \\
\mathbf{z}_k^{\mathrm{H}} \mathbf{W}_k^{\mathrm{H}} \frac{\partial \hat{\mathbf{h}}_{k-1}}{\partial \hat{\beta}_{k-1}^{\mathrm{im}}} \\
\mathbf{z}_k^{\mathrm{H}} \mathbf{W}_k^{\mathrm{H}} \frac{\partial \hat{\mathbf{h}}_{k-1}}{\partial \hat{x}_{k-1,1}} \\
\mathbf{z}_k^{\mathrm{H}} \mathbf{W}_k^{\mathrm{H}} \frac{\partial \hat{\mathbf{h}}_{k-1}}{\partial \hat{x}_{k-1,2}}
\end{bmatrix}
\right\}
\mathrm{Re}\left\{
\begin{bmatrix}
\mathbf{z}_k^{\mathrm{H}} \mathbf{W}_k^{\mathrm{H}} \frac{\partial \hat{\mathbf{h}}_{k-1}}{\partial \hat{\beta}_{k-1}^{\mathrm{re}}} \\
\mathbf{z}_k^{\mathrm{H}} \mathbf{W}_k^{\mathrm{H}} \frac{\partial \hat{\mathbf{h}}_{k-1}}{\partial \hat{\beta}_{k-1}^{\mathrm{im}}} \\
\mathbf{z}_k^{\mathrm{H}} \mathbf{W}_k^{\mathrm{H}} \frac{\partial \hat{\mathbf{h}}_{k-1}}{\partial \hat{x}_{k-1,1}} \\
\mathbf{z}_k^{\mathrm{H}} \mathbf{W}_k^{\mathrm{H}} \frac{\partial \hat{\mathbf{h}}_{k-1}}{\partial \hat{x}_{k-1,2}}
\end{bmatrix}^{\mathrm{T}}
\right\}
\right\} \mathbf{I}_S\left(\hat{\boldsymbol{\psi}}_{k-1}, \mathbf{W}_k\right)^{-1}
$$
$$
\stackrel{(a)}{=} \frac{4|\mathbf{s}|^2}{\sigma_z^4} \mathbf{I}_S\left(\hat{\boldsymbol{\psi}}_{k-1}, \mathbf{W}_k\right)^{-1} \mathbb{E}\left\{ \frac{\sigma_z^2}{2|\mathbf{s}|} \frac{\partial \log p_S\left(\mathbf{y}_k | \hat{\boldsymbol{\psi}}_{k-1}, \mathbf{W}_k\right)}{\partial \hat{\boldsymbol{\psi}}_{k-1}} \frac{\sigma_z^2}{2|\mathbf{s}|} \frac{\partial \log p_S\left(\mathbf{y}_k | \hat{\boldsymbol{\psi}}_{k-1}, \mathbf{W}_k\right)}{\partial \hat{\boldsymbol{\psi}}_{k-1}^{\mathrm{T}}} \right\} \mathbf{I}_S\left(\hat{\boldsymbol{\psi}}_{k-1}, \mathbf{W}_k\right)^{-1}
$$
$$
\stackrel{(b)}{=} \mathbf{I}_S\left(\hat{\boldsymbol{\psi}}_{k-1}, \mathbf{W}_k\right)^{-1} \mathbf{I}_S\left(\hat{\boldsymbol{\psi}}_{k-1}, \mathbf{W}_k\right) \mathbf{I}_S\left(\hat{\boldsymbol{\psi}}_{k-1}, \mathbf{W}_k\right)^{-1}
$$
$$
= \mathbf{I}_S\left(\hat{\boldsymbol{\psi}}_{k-1}, \mathbf{W}_k\right)^{-1},
\tag{115}
$$

$$\leq \frac{2\pi}{M\sqrt{MN}} \sum_{m=1}^{M} \sum_{n=1}^{N} (m-1) \left| e^{-j2\pi \left(\frac{(m-1)\delta_{k,i1}}{M} + \frac{(n-1)\delta_{k,i2}}{N} \right)} \right|$$
$$= \sqrt{MN} \, (M-1) < +\infty, \tag{122}$$

and

$$\left| \mathbf{w}_{k,i}^{H} \frac{\partial \mathbf{a}(\mathbf{x})}{\partial x_2} \right|$$

$$= \left| \frac{1}{\sqrt{MN}} \sum_{m=1}^{M} \sum_{n=1}^{N} j2\pi \frac{n-1}{N} e^{-j2\pi \left(\frac{(m-1)\delta_{k,i1}}{M} + \frac{(n-1)\delta_{k,i2}}{N} \right)} \right|$$

$$\leq \frac{2\pi}{N\sqrt{MN}} \sum_{m=1}^{M} \sum_{n=1}^{N} (n-1) \left| e^{-j2\pi \left(\frac{(m-1)\delta_{k,i1}}{M} + \frac{(n-1)\delta_{k,i2}}{N} \right)} \right|$$

$$= \sqrt{MN} \, (N-1) < +\infty, \tag{123}$$

for $i = 1, 2, 3$ and all possible $\mathbf{w}_{k,i}$ and $\mathbf{x}$, where $\left[\delta_{k,i1}, \delta_{k,i2} \right]^{\mathrm{T}} = \boldsymbol{\omega}_{k,i} - \mathbf{x}$. Thus we can get

$$\left\| \frac{2|\mathbf{s}|^2}{\sigma_z^2} \begin{bmatrix} \mathrm{Re} \left\{ \mathbf{e}_k^{\mathrm{H}} \left(\beta \mathbf{W}_k^{\mathrm{H}} \mathbf{a}(\mathbf{x}) - \hat{\beta}_{k-1} \mathbf{e}_k \right) \right\} \\ \mathrm{Im} \left\{ \mathbf{e}_k^{\mathrm{H}} \left(\beta \mathbf{W}_k^{\mathrm{H}} \mathbf{a}(\mathbf{x}) - \hat{\beta}_{k-1} \mathbf{e}_k \right) \right\} \\ \mathrm{Re} \left\{ \tilde{\mathbf{e}}_{k1}^{\mathrm{H}} \left(\beta \mathbf{W}_k^{\mathrm{H}} \mathbf{a}(\mathbf{x}) - \hat{\beta}_{k-1} \mathbf{e}_k \right) \right\} \\ \mathrm{Re} \left\{ \tilde{\mathbf{e}}_{k2}^{\mathrm{H}} \left(\beta \mathbf{W}_k^{\mathrm{H}} \mathbf{a}(\mathbf{x}) - \hat{\beta}_{k-1} \mathbf{e}_k \right) \right\} \end{bmatrix} \right\|_2^2 < +\infty. \tag{124}$$

Combining (120) and (124), we have

$$\mathbb{E} \left[\left\| \mathbf{f}_{\boldsymbol{\psi}} \left(\hat{\boldsymbol{\psi}}_{k-1} \right) \right\|_2^2 \right] < +\infty. \tag{125}$$

According to (120), it is clear that $\mathrm{Tr} \left\{ \mathbf{I}(\hat{\boldsymbol{\psi}}_{k-1}, \mathbf{W}_k)^{-1} \right\} < +\infty$. Then, we can get that

$$\sup_k \mathbb{E} \left[\left\| \mathbf{f}_{\boldsymbol{\psi}} \left(\hat{\boldsymbol{\psi}}_{k-1} \right) + \hat{\mathbf{z}}_k \right\|_2^2 \right] < +\infty. \tag{126}$$

3) The function $\mathbf{f}_{\boldsymbol{\psi}} \left(\hat{\boldsymbol{\psi}}_{k-1} \right)$ should be continuous with respect to $\hat{\boldsymbol{\psi}}_{k-1}$.

By using (44), we know that each element of $\mathbf{f}_{\boldsymbol{\psi}} \left(\hat{\boldsymbol{\psi}}_{k-1} \right)$ is continuous with respect to $\hat{\boldsymbol{\psi}}_{k-1} = \left[\hat{\beta}_{k-1}^{\mathrm{re}}, \hat{\beta}_{k-1}^{\mathrm{im}}, \hat{x}_{k-1,1}, \hat{x}_{k-1,2} \right]^{\mathrm{T}}$. Therefore, $\mathbf{f}_{\boldsymbol{\psi}} \left(\hat{\boldsymbol{\psi}}_{k-1} \right)$ is continuous with respect to $\hat{\boldsymbol{\psi}}_{k-1}$.

4) Let $\boldsymbol{\mu}_k = \mathbb{E} \left[\mathbf{f}_{\boldsymbol{\psi}} \left(\hat{\boldsymbol{\psi}}_{k-1} \right) + \hat{\mathbf{z}}_k \middle| \mathcal{G}_{k-1} \right] - \mathbf{f} \left(\hat{\boldsymbol{\psi}}_{k-1}, \boldsymbol{\psi} \right)$. We need to prove that $\sum_{k=1}^{+\infty} \| b_{S,k} \boldsymbol{\mu}_k \|_2 < +\infty$ with probability one.

From (116), we get $\boldsymbol{\mu}_k = \mathbf{0}$ for all $k \geq 1$. So we have $\sum_{k=1}^{+\infty} \| b_{S,k} \boldsymbol{\mu}_k \|_2 = 0 < +\infty$ with probability one.

By Theorem 5.2.1 in [37], $\hat{\boldsymbol{\psi}}_k$ converges to a unique stable point within the stable points set with probability one.

APPENDIX E
PROOF OF THEOREM 2

Theorem 2 is proven in three steps:

Step 1: Two continuous processes based on the discrete process $\hat{\boldsymbol{\psi}}_k = [\hat{\beta}_k^{re}, \hat{\beta}_k^{im}, \hat{x}_{k,1}, \hat{x}_{k,2}]^{\mathrm{T}}$ are established, i.e., $\bar{\boldsymbol{\psi}}(t) \triangleq [\bar{\beta}^{re}(t), \bar{\beta}^{im}(t), \bar{x}_1(t), \bar{x}_2(t)]^{\mathrm{T}}$ and $\tilde{\boldsymbol{\psi}}^k(t) \triangleq [\tilde{\beta}^{re,k}(t), \tilde{\beta}^{im,k}(t), \tilde{x}_1^k(t), \tilde{x}_2^k(t)]^{\mathrm{T}}$.

The discrete time parameters are defined as: $t_0 \triangleq 0$, $t_k \triangleq \sum_{l=1}^{k} b_{S,l}$, $k \geq 1$. The first continuous process $\bar{\boldsymbol{\psi}}(t)$, $t \geq 0$ is constructed as the linear interpolation of the sequence $\hat{\boldsymbol{\psi}}_k$, $k \geq 0$, where $\bar{\boldsymbol{\psi}}(t_k) = \hat{\boldsymbol{\psi}}_k$, $k \geq 0$. Therefore, $\bar{\boldsymbol{\psi}}(t)$ is given by

$$\bar{\boldsymbol{\psi}}(t) = \bar{\boldsymbol{\psi}}(t_k) + \frac{(t - t_k)}{b_{S,k+1}} \left[\bar{\boldsymbol{\psi}}(t_{k+1}) - \bar{\boldsymbol{\psi}}(t_k) \right], \quad t \in [t_k, t_{k+1}]. \tag{127}$$

The second continuous process $\tilde{\boldsymbol{\psi}}^k(t)$ is the solution of the following ordinary differential equation (ODE):

$$\frac{d\tilde{\boldsymbol{\psi}}^k(t)}{dt} = \mathbf{f}_{\boldsymbol{\psi}} \left(\tilde{\boldsymbol{\psi}}^k(t) \right), \tag{128}$$

for $t \in [t_k, \infty)$, where $\tilde{\boldsymbol{\psi}}^k(t_k) = \bar{\boldsymbol{\psi}}(t_k) = \hat{\boldsymbol{\psi}}_k$, $k \geq 0$. Thus, $\tilde{\boldsymbol{\psi}}^k(t)$ can be given as

$$\tilde{\boldsymbol{\psi}}^k(t) = \bar{\boldsymbol{\psi}}(t_k) + \int_{t_k}^{t} \mathbf{f}_{\boldsymbol{\psi}} \left(\tilde{\boldsymbol{\psi}}^k(v) \right) dv, \quad t \geq t_k. \tag{129}$$

Step 2: By using the two continuous processes $\bar{\boldsymbol{\psi}}(t)$ and $\tilde{\boldsymbol{\psi}}^k(t)$ constructed in Step 1, a sufficient condition for the convergence of the discrete process $\hat{\mathbf{x}}_k$ is provided here.

We first construct a time-invariant set $\mathcal{I}$ that includes the DPV $\mathbf{x}$ within the main lobe, i.e., $\mathbf{x} \in \mathcal{I} \subset \mathcal{B}(\mathbf{x})$. Define $\tilde{\mathbf{x}}^0(t) \triangleq \left[\tilde{x}_1^0(t), \tilde{x}_2^0(t) \right]^{\mathrm{T}}$ and denote $\hat{\mathbf{x}}_{\mathrm{b}} = \tilde{\mathbf{x}}^0(t_{\mathrm{b}})$ as the beam direction of the process $\tilde{\boldsymbol{\psi}}^0(t)$ that is closest to the boundary of the main lobe, which is given by[3]

$$\inf_{\mathbf{v} \in \partial \mathcal{B}(x), t \geq 0} \left\| \mathbf{v} - \tilde{\mathbf{x}}^0(t) \right\|_2 = \inf_{\mathbf{v} \in \partial \mathcal{B}(x)} \left\| \mathbf{v} - \hat{\mathbf{x}}_{\mathrm{b}} \right\|_2 > 0. \tag{130}$$

Then we pick δ such that

$$\min \left\{ \inf_{\mathbf{v} \in \partial \mathcal{B}(x)} \left\| \mathbf{v} - \hat{\mathbf{x}}_{\mathrm{b}} \right\|_{-\infty}, \left\| \hat{\mathbf{x}}_b - \mathbf{x} \right\|_{-\infty} \right\} > \delta > 0, \tag{131}$$

where $\| \mathbf{u} \|_{-\infty} = \min_{l=1,2} [\mathbf{u}]_l$ denotes the minimum element of $\mathbf{u}$. Note that when $t \geq t_b$, the solution $\tilde{\boldsymbol{\psi}}^0(t)$ of the ODE (128) will approach the real equivalent channel gain β and DPV $\mathbf{x}$ monotonically as time t increases. Hence, we construct the invariant set $\mathcal{I}$ in (132), as shown at the bottom of the page. An example of the invariant set $\mathcal{I}$ is shown in Fig. 18.

---

[3] The boundary of the set $\mathcal{B}(\mathbf{x})$ is denoted by $\partial \mathcal{B}(\mathbf{x})$.

$$\mathcal{I} = \left(x_1 - |x_1 - \hat{x}_{1,\mathrm{b}}| - \delta, \; x_1 + |x_1 - \hat{x}_{1,\mathrm{b}}| + \delta \right) \times \left(x_2 - |x_2 - \hat{x}_{2,\mathrm{b}}| - \delta, \; x_2 + |x_2 - \hat{x}_{2,\mathrm{b}}| + \delta \right) \subset \mathcal{B}(\mathbf{x}). \tag{132}$$
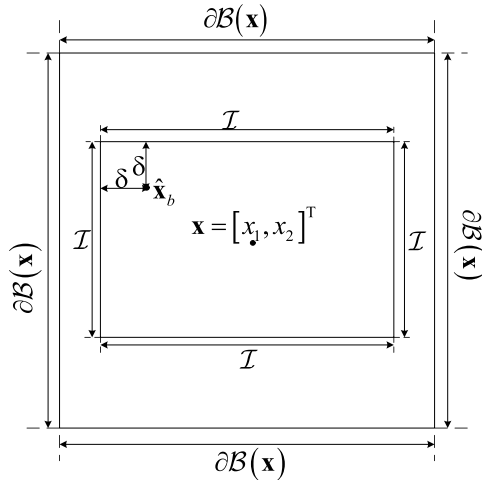
FIGURE 18. An illustration of the invariant set $\mathcal{I}$.

Then, a sufficient condition will be established in Lemma 7 that ensures $\hat{\mathbf{x}}_k \in \mathcal{I}$ for $k \geq 0$, and hence from Corollary 2.5 in [36], we can obtain that $\hat{\mathbf{x}}_k$ converges to $\mathbf{x}$. Before giving Lemma 7, let us provide some definitions first:

- Pick $T > 0$ such that the solution $\tilde{\boldsymbol{\psi}}^0(t), t \geq 0$ of the ODE (128) with $\tilde{\boldsymbol{\psi}}^0(0) = \left[\hat{\beta}_0^{\mathrm{re}}, \hat{\beta}_0^{\mathrm{im}}, \hat{x}_{0,1}, \hat{x}_{0,2}\right]^{\mathrm{T}}$ satisfies $\inf_{\mathbf{v} \in \partial \mathcal{B}} \left|\mathbf{v} - \tilde{\mathbf{x}}^0(t)\right| \geq 2\delta$ for $t \geq T$. Since when $t \geq t_b$, $\tilde{\mathbf{x}}^0(t)$ will approach the DPV $\mathbf{x}$ monotonically as time t increases, one possible T is given by

$$T = \arg \min_{t \in [t_b, +\infty)} \left| \left[\left[\int_{t_b}^{t} \mathbf{f}_{\boldsymbol{\psi}} \left(\tilde{\boldsymbol{\psi}}^0(v)\right) dv \right]_3 \right] - \delta \right|,$$
(133)

where $[\cdot]_i$ denotes the i-th element of the vector.

- Let $T_0 \triangleq 0$ and $T_{l+1} \triangleq \min \{t_i : t_i \geq T_l + T, i \geq 0\}$ for $l \geq 0$. Then $T_{l+1} - T_l \in [T, T + b_{S,1}]$ and $T_l = t_{\tilde{k}(l)}$ for some $\tilde{k}(l) \uparrow +\infty$, where $\tilde{k}(0) = 0$. Let $\tilde{\boldsymbol{\psi}}^{\tilde{k}(l)}(t)$ denote the solution of ODE (128) for $t \in I_l \triangleq [T_l, T_{l+1}]$ with $\tilde{\boldsymbol{\psi}}^{\tilde{k}(l)}(T_l) = \bar{\boldsymbol{\psi}}(T_l), l \geq 0$.

Hence, we can obtain the following lemma:

Lemma 7: If $\sup_{t \in I_l} \left\| \bar{\mathbf{x}}(t) - \tilde{\mathbf{x}}^{\tilde{k}(l)}(t) \right\|_2 \leq \delta$ for all $l \geq 0$, then $\hat{\mathbf{x}}_k \in \mathcal{I}$ for all $k \geq 0$.

Proof: If $\sup_{t \in I_l} \left\| \bar{\mathbf{x}}(t) - \tilde{\mathbf{x}}^{\tilde{k}(l)}(t) \right\|_2 \leq \delta$ for all $l \geq 0$, then $\sup_{t \in I_l} \left| \bar{x}_1(t) - \tilde{x}_1^{\tilde{k}(l)}(t) \right| \leq \delta$ and $\sup_{t \in I_l} \left| \bar{x}_2(t) - \tilde{x}_2^{\tilde{k}(l)}(t) \right| \leq \delta$. According to Lemma 1 in [42], $\hat{x}_{k,1} \in \mathcal{I}$ for all $k \geq 0$ and $\hat{x}_{k,2} \in \mathcal{I}$ for all $k \geq 0$. Hence, $\hat{\mathbf{x}}_k \in \mathcal{I}$ for all $k \geq 0$. ∎

Step 3: *We will derive the probability lower bound for the condition in Lemma 7, which is also a lower bound for $P\left(\hat{\mathbf{x}}_k \to \mathbf{x} \mid \hat{\mathbf{x}}_0 \in \mathcal{B}(\mathbf{x})\right).$*

We will derive the probability lower bound for the condition in Lemma 7, which results in the following lemma:

Lemma 8: If (i) the initial point satisfies $\hat{\mathbf{x}}_0 \in \mathcal{B}(\mathbf{x})$, (ii) $b_{S,k}$ is given by (42) with any $\epsilon_S > 0$, then there exist $K_{S,0} \geq 0$ and $R > 0$ such that

$$P\left(\hat{\mathbf{x}}_k \in \mathcal{I}, \forall k \geq 0\right) \geq 1 - 8e^{-\frac{R|\mathbf{s}|^2}{\epsilon_S^2 \sigma_z^2}}.$$
(134)

Proof: See Appendix J. ∎

Finally, by applying Lemma 8 and Corollary 2.5 in [36], we can obtain

$$P\left(\hat{\mathbf{x}}_k \to \mathbf{x} \mid \hat{\mathbf{x}}_0 \in \mathcal{B}\right) \geq P\left(\hat{\mathbf{x}}_k \in \mathcal{I}, \forall k \geq 0\right) \geq 1 - 8e^{-\frac{R|\mathbf{s}|^2}{\epsilon_S^2 \sigma_z^2}},$$

which completes the proof of Theorem 2.

APPENDIX F
PROOF OF THEOREM 3

If the step-size $b_{S,k}$ is given by (42) with any $\varepsilon_S > 0$ and $K_{S,0} \geq 0$, the sufficient conditions are provided by Theorem 6.6.1 [34, Section 6.6] to prove the asymptotic normality of $\sqrt{k}\left(\hat{\mathbf{x}}_k - \mathbf{x}\right)$, i.e., $\sqrt{k}\left(\hat{\mathbf{x}}_k - \mathbf{x}\right) \xrightarrow{d} \mathcal{N}\left(0, \Sigma_{\mathbf{x}}\right)$. With the condition that $\hat{\boldsymbol{\psi}}_k \to \boldsymbol{\psi}$, we can prove that the beam and channel tracking algorithm satisfies the conditions above and obtain the variance $\Sigma_{\mathbf{x}}$ as follows:

1) Equation (113) is supposed to satisfy: (i) there exists an increasing sequence of σ-fields $\{\mathcal{F}_k : k \geq 0\}$ such that $\mathcal{F}_l \subset \mathcal{F}_k$ for $l < k$, and (ii) the random noise $\hat{\mathbf{z}}_k$ is $\mathcal{F}_k$-measurable and independent of $\mathcal{F}_{k-1}$.

   As is shown in Appendix X, there exists an increasing sequence of σ-fields $\{\mathcal{G}_k : k \geq 0\}$, where ς_k is measurable with respect to $\mathcal{G}_{k-1}$ and independent of $\mathcal{G}_{k-1}$.

2) $\hat{\mathbf{x}}_k$ should converge to $\mathbf{x}$ almost surely as $k \to +\infty$. We assume that $\hat{\boldsymbol{\psi}}_k \to \boldsymbol{\psi}$, hence $\hat{\mathbf{x}}_k$ converges to $\mathbf{x}$ almost surely when $k \to +\infty$.

3) The stable condition:
   In (44), we rewrite $\mathbf{f}_{\boldsymbol{\psi}}\left(\hat{\boldsymbol{\psi}}_{k-1}\right)$ as follows:

$$\mathbf{f}_{\boldsymbol{\psi}}\left(\hat{\boldsymbol{\psi}}_{k-1}\right) = \mathbf{D}_1 \left(\hat{\boldsymbol{\psi}}_{k-1} - \boldsymbol{\psi}\right) + \begin{bmatrix} o(\|\hat{\boldsymbol{\psi}}_{k-1} - \boldsymbol{\psi}\|_2) \\ o(\|\hat{\boldsymbol{\psi}}_{k-1} - \boldsymbol{\psi}\|_2) \\ o(\|\hat{\boldsymbol{\psi}}_{k-1} - \boldsymbol{\psi}\|_2) \\ o(\|\hat{\boldsymbol{\psi}}_{k-1} - \boldsymbol{\psi}\|_2) \end{bmatrix},$$
(135)

where $\mathbf{D}_1$ is given by

$$\mathbf{D}_1 = \left. \frac{\partial \mathbf{f}_{\boldsymbol{\psi}}\left(\hat{\boldsymbol{\psi}}_{k-1}\right)}{\partial \hat{\boldsymbol{\psi}}_{k-1}^{\mathrm{T}}} \right|_{\hat{\boldsymbol{\psi}}_{k-1} = \boldsymbol{\psi}} = - \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

(136)

Then the stable condition is obtained that:

$$\mathbf{E} = \mathbf{D}_1 \cdot \varepsilon_S + \frac{1}{2}$$

$$= \begin{bmatrix} \frac{1}{2} - \varepsilon_S & 0 & 0 & 0 \\ 0 & \frac{1}{2} - \varepsilon_S & 0 & 0 \\ 0 & 0 & \frac{1}{2} - \varepsilon_S & 0 \\ 0 & 0 & 0 & \frac{1}{2} - \varepsilon_S \end{bmatrix} \prec 0,$$

which leads to $\varepsilon_S > \frac{1}{2}$.

4) The noise vector $\hat{\mathbf{z}}_k$ satisfies:

$$\mathbb{E}\left[\|\hat{\mathbf{z}}_k\|_2^2\right] = \text{tr}\left\{\mathbf{I}_S(\hat{\boldsymbol{\psi}}_{k-1}, \mathbf{W}_k)^{-1}\right\} < +\infty,$$

(137)

and

$$\lim_{v \to +\infty} \sup_{k \ge 1} \int_{\|\hat{z}_k\|_2 > v} \|\hat{\mathbf{z}}_k\|_2^2 \, p(\hat{\mathbf{z}}_k) d\hat{\mathbf{z}}_k = 0. \quad (138)$$

Let

$$\mathbf{F} = \lim_{\substack{k \to +\infty \\ \hat{\boldsymbol{\psi}}_k \to \boldsymbol{\psi}}} \mathbb{E}\left[\hat{\mathbf{z}}_k \hat{\mathbf{z}}_k^{\mathsf{T}}\right] \overset{(a)}{=} \lim_{\substack{k \to +\infty \\ \hat{\boldsymbol{\psi}}_k \to \boldsymbol{\psi}}} \mathbf{I}_S(\hat{\boldsymbol{\psi}}_{k-1}, \mathbf{W}_k)^{-1}$$

$$= \mathbf{I}_S(\boldsymbol{\psi}, \tilde{\mathbf{W}}_S^*)^{-1}, \quad (139)$$

where Step (a) is obtained from (115).

By Theorem 6.6.1 [34, Section 6.6], we have

$$\sqrt{k + K_{S,0}} \left(\hat{\boldsymbol{\psi}}_k - \boldsymbol{\psi}\right) \overset{d}{\to} \mathcal{N}(0, \boldsymbol{\Sigma}_{\mathbf{x}}),$$

where

$$\boldsymbol{\Sigma}_{\mathbf{x}} = \varepsilon_S^2 \cdot \int_0^\infty e^{\mathbf{E}v} \mathbf{F} e^{\mathbf{E}^{\mathsf{H}} v} dv = \frac{\varepsilon_S^2}{2\varepsilon_S - 1} \mathbf{I}_S(\boldsymbol{\psi}, \tilde{\mathbf{W}}^*)^{-1}. \quad (140)$$

Due to that $\lim_{k \to +\infty} \sqrt{(k + K_{S,0})/k} = 1$, we have

$$\sqrt{k}\left(\hat{\boldsymbol{\psi}}_k - \boldsymbol{\psi}\right) \to \sqrt{k} \cdot \sqrt{\frac{k + K_{S,0}}{k}} \left(\hat{\boldsymbol{\psi}}_k - \boldsymbol{\psi}\right) \overset{d}{\to} \mathcal{N}(0, \boldsymbol{\Sigma}_{\mathbf{x}}),$$

if $k \to +\infty$. Thus, we can get

$$\sqrt{k}\left(\hat{\boldsymbol{\psi}}_k - \boldsymbol{\psi}\right) \overset{d}{\to} \mathcal{N}(0, \boldsymbol{\Sigma}_{\mathbf{x}}). \quad (141)$$

By adopting $\epsilon_S = 1$ in (140), we can obtain

$$\sqrt{k}\left(\hat{\boldsymbol{\psi}}_k - \boldsymbol{\psi}\right) \overset{d}{\to} \mathcal{N}\left(0, \mathbf{I}_S(\boldsymbol{\psi}, \tilde{\mathbf{W}}^*)^{-1}\right). \quad (142)$$

Since $\hat{\boldsymbol{\psi}}_k \to \boldsymbol{\psi}$ as $k \to +\infty$, $\hat{\mathbf{h}}_k - \mathbf{h}$ is linear to $\hat{\boldsymbol{\psi}}_k - \boldsymbol{\psi}$. Hence, $\hat{\mathbf{h}}_k - \mathbf{h}$ is also asymptotically Gaussian.

Combining (76), (142) and (27), we can conclude that

$$\lim_{k \to +\infty} \frac{k}{MN} \mathbb{E}\left[\left\|\hat{\mathbf{h}}_k - \mathbf{h}\right\|_2^2 \Big| \hat{\boldsymbol{\psi}}_k \to \boldsymbol{\psi}\right] = C_S^{\min}(\boldsymbol{\psi}). \quad (143)$$

APPENDIX G
PROOF OF LEMMA 4

In problem (52), the constraint (53) ensures that $\hat{\mathbf{x}}_k$ is an unbiased estimate of $\mathbf{x}$. According to Section 3.7 of [33], if $\hat{\mathbf{x}}$ is an unbiased estimate of $\mathbf{x}$, then we can obtain that

$$\text{Cov}\left(\hat{\mathbf{x}}\right) - \mathbf{I}^{-1}(\mathbf{x}) \succeq \mathbf{0}, \quad (144)$$

where $\text{Cov}\left(\hat{\mathbf{x}}\right)$ denotes the covariance matrix of $\hat{\mathbf{x}}$, $\mathbf{I}(\mathbf{x})$ is the corresponding Fisher information matrix and $\mathbf{A} \succeq \mathbf{0}$ means that the matrix $\mathbf{A}$ is nonnegative definite. From (144), we can get that

$$\text{Cov}\left(\hat{\mathbf{x}}_k\right) - \left(\sum_{l=1}^k \mathbf{I}_{DI}(\mathbf{x}, \mathbf{W}_l)\right) \succeq \mathbf{0}, \quad (145)$$

which implies that the diagonal elements of the matrix on the left side of '$\succeq$' are nonnegative because all matrices are 2×2 in (145). Therefore, we obtain that

$$\text{Tr}\left\{\text{Cov}\left(\hat{\mathbf{x}}_k\right)\right\} - \text{Tr}\left\{\left(\sum_{l=1}^k \mathbf{I}(\mathbf{x}, \mathbf{W}_l)\right)\right\} \ge 0, \quad (146)$$

i.e.,

$$\mathbb{E}\left[\|\hat{\mathbf{x}}_k - \mathbf{x}\|_2^2\right] - \text{Tr}\left\{\left(\sum_{l=1}^k \mathbf{I}_{DI}(\mathbf{x}, \mathbf{W}_l)\right)\right\} \ge 0, \quad (147)$$

which yields the result of (54).

Now we try to obtain the Fisher information matrix in (55). According to (48), the determinant and the inverse of the covariance matrix can be written as follows:

$$\begin{cases} |\boldsymbol{\Sigma}_{\mathbf{y},k}| = \sigma_z^4 \left(|\mathbf{s}|^2 \sigma_\beta^2 \left(\mathbf{W}_k^{\mathsf{H}} \mathbf{a}(\mathbf{x})\right)^{\mathsf{H}} \mathbf{W}_k^{\mathsf{H}} \mathbf{a}(\mathbf{x}) + \sigma_z^2\right) \\ \boldsymbol{\Sigma}_{\mathbf{y},k}^{-1} = \dfrac{\mathbf{J}_3}{\sigma_z^2} - \dfrac{\sigma_z^2 |\mathbf{s}|^2 \sigma_\beta^2 \mathbf{W}_k^{\mathsf{H}} \mathbf{a}(\mathbf{x}) \left(\mathbf{W}_k^{\mathsf{H}} \mathbf{a}(\mathbf{x})\right)^{\mathsf{H}}}{|\boldsymbol{\Sigma}_{\mathbf{y},k}|}. \end{cases}$$

(148)

Based on the definition in (57), the determinant and the inverse of the covariance matrix in (148) can be rewritten as

$$\begin{cases} |\boldsymbol{\Sigma}_{\mathbf{y},k}| = \sigma_z^4 \left(|\mathbf{s}|^2 \sigma_\beta^2 \mathbf{g}_k^{\mathsf{H}} \mathbf{g}_k + \sigma_z^2\right) \\ \boldsymbol{\Sigma}_{\mathbf{y},k}^{-1} = \dfrac{\mathbf{J}_3}{\sigma_z^2} - \dfrac{\sigma_z^2 |\mathbf{s}|^2 \sigma_\beta^2 \mathbf{g}_k \mathbf{g}_k^{\mathsf{H}}}{|\boldsymbol{\Sigma}_{\mathbf{y},k}|}. \end{cases} \quad (149)$$

In addition, with the help of (50), we can obtain that

$$\frac{\partial \log p_{DI}(\mathbf{y}_k | \mathbf{x}, \mathbf{W}_k)}{\partial x_p} = -\frac{1}{|\boldsymbol{\Sigma}_{\mathbf{y},k}|} \frac{\partial |\boldsymbol{\Sigma}_{\mathbf{y},k}|}{\partial x_p} - \mathbf{y}_k^{\mathsf{H}} \frac{\partial \boldsymbol{\Sigma}_{\mathbf{y},k}^{-1}}{\partial x_p} \mathbf{y}_k, \quad (150)$$

where $\frac{|\boldsymbol{\Sigma}_{\mathbf{y},k}|}{\partial x_p}$ and $\frac{\partial \boldsymbol{\Sigma}_{\mathbf{y},k}^{-1}}{\partial x_p}$ are given by (151) according to (149):

$$\begin{cases} \dfrac{\partial |\boldsymbol{\Sigma}_{\mathbf{y},k}|}{\partial x_p} = \sigma_z^4 |\mathbf{s}|^2 \sigma_\beta^2 \dfrac{\partial \mathbf{g}_k^{\mathsf{H}} \mathbf{g}_k}{\partial x_p} \\ \dfrac{\partial \boldsymbol{\Sigma}_{\mathbf{y},k}^{-1}}{\partial x_p} = -\sigma_z^2 |\mathbf{s}|^2 \sigma_\beta^2 \dfrac{\frac{\partial \mathbf{g}_k \mathbf{g}_k^{\mathsf{H}}}{\partial x_p} |\boldsymbol{\Sigma}_{\mathbf{y},k}| - \mathbf{g}_k \mathbf{g}_k^{\mathsf{H}} \frac{|\boldsymbol{\Sigma}_{\mathbf{y},k}|}{\partial x_p}}{|\boldsymbol{\Sigma}_{\mathbf{y},k}|^2}. \end{cases} \quad (151)$$

By combining (48), (50), (55) and (150), we can obtain the p-th row, j-th column element of the Fisher information below:

$$
\begin{aligned}
& [\mathbf{I}_{DI}(\mathbf{x}, \mathbf{W}_k)]_{p,j} \\
&= \mathbb{E}\left[\frac{\partial \log p_{DI}(\mathbf{y}_k|\mathbf{x}, \mathbf{W}_k)}{\partial x_p} \frac{\partial \log p_{DI}(\mathbf{y}_k|\mathbf{x}, \mathbf{W}_k)}{\partial x_j}\right] \\
&= -\frac{1}{|\mathbf{\Sigma}_{\mathbf{y},k}|} \frac{\partial |\mathbf{\Sigma}_{\mathbf{y},k}|}{\partial x_p} \frac{\partial |\mathbf{\Sigma}_{\mathbf{y},k}|}{\partial x_j} + 2|\mathbf{s}|^4 \sigma_\beta^4 \mathbf{g}_k^H \frac{\partial \mathbf{\Sigma}_{\mathbf{y},k}^{-1}}{\partial x_p} \mathbf{g}_k \mathbf{g}_k^H \frac{\partial \mathbf{\Sigma}_{\mathbf{y},k}^{-1}}{\partial x_j} \mathbf{g}_k \\
&\quad + \sigma_z^2 |\mathbf{s}|^2 \sigma_\beta^2 \mathbf{g}_k^H \frac{\partial \mathbf{\Sigma}_{\mathbf{y},k}^{-1}}{\partial x_p} \mathbf{g}_k \operatorname{Tr}\left\{\frac{\partial \mathbf{\Sigma}_{\mathbf{y},k}^{-1}}{\partial x_j}\right\} \\
&\quad + \sigma_z^2 |\mathbf{s}|^2 \sigma_\beta^2 \mathbf{g}_k^H \frac{\partial \mathbf{\Sigma}_{\mathbf{y},k}^{-1}}{\partial x_j} \mathbf{g}_k \operatorname{Tr}\left\{\frac{\partial \mathbf{\Sigma}_{\mathbf{y},k}^{-1}}{\partial x_p}\right\} \\
&\quad + \sigma_z^4 \operatorname{Tr}\left\{\frac{\partial \mathbf{\Sigma}_{\mathbf{y},k}^{-1}}{\partial x_p}\right\} \operatorname{Tr}\left\{\frac{\partial \mathbf{\Sigma}_{\mathbf{y},k}^{-1}}{\partial x_j}\right\} + \sigma_z^4 \operatorname{Tr}\left\{\frac{\partial \mathbf{\Sigma}_{\mathbf{y},k}^{-1}}{\partial x_p} \frac{\partial \mathbf{\Sigma}_{\mathbf{y},k}^{-1}}{\partial x_j}\right\} \\
&\quad + \sigma_z^2 |\mathbf{s}|^2 \sigma_\beta^2 \mathbf{g}_k^H \left(\frac{\partial \mathbf{\Sigma}_{\mathbf{y},k}^{-1}}{\partial x_p} \frac{\partial \mathbf{\Sigma}_{\mathbf{y},k}^{-1}}{\partial x_j} + \frac{\partial \mathbf{\Sigma}_{\mathbf{y},k}^{-1}}{\partial x_j} \frac{\partial \mathbf{\Sigma}_{\mathbf{y},k}^{-1}}{\partial x_p}\right) \mathbf{g}_k.
\end{aligned}
\tag{152}
$$

Then we substitute (48), (49), (57), (149), (151) into (152), which yields the result of (56).

Finally, the proof of Lemma 4 is completed.

APPENDIX H
PROOF OF LEMMA 5
The proof of property 1) in Lemma 5 is similar to the proof of that in Lemma 3. Hence, we focus on the proof of property 2) and property 3) in Lemma 5.

Consider the p-th row, j-th column element of the Fisher information matrix $\mathbf{I}_{DI}(\mathbf{x}, \mathbf{W}_k)$ in (56). We can rewrite it in (153), as shown at the bottom of the page, where Step (a) is obtained by substituting (148) into (153). When $\frac{|\mathbf{s}|^2 \sigma_\beta^2}{\sigma_z^2} \to +\infty$, we can obtain the element of $\mathbf{I}_{DI}(\mathbf{x}, \mathbf{W}_k)$ in (154), as shown at the bottom of the page, which reveals that $\frac{\sigma_z^2}{|\mathbf{s}|^2 \sigma_\beta^2} \mathbf{I}_{DI}(\mathbf{x}, \mathbf{W}_k)$ converges as $\frac{|\mathbf{s}|^2 \sigma_\beta^2}{\sigma_z^2} \to +\infty$. Then the property 2) of Lemma 5 is proved.

Let us see the property 3) in Lemma 5. Similar to Step 3 in Appendix X, we can obtain that $C_{DI}^{\min}(\boldsymbol{\psi})$ converge as $M, N \to +\infty$ and

$$
\lim_{M,N\to+\infty} C_{DI}(\boldsymbol{\psi}, \widetilde{\mathbf{W}}_{DI}^*) = \lim_{M,N\to+\infty} C_{DI}^{\min}(\boldsymbol{\psi}). \tag{155}
$$

According to (57) and (105), $\mathbf{g}_k$ is $\Theta\left(\sqrt{MN}\right)$ while $\tilde{g}_{k,p}$ and $\mathbf{G}_{k,p}$ are $\Theta(MN)$. Hence, $\sigma_z^2 \operatorname{Tr}\left\{\mathbf{G}_{k,p}\mathbf{G}_{k,j}\right\}$ can be omitted since it is $\Theta\left((MN)^2\right)$ while other parts are $\Theta\left((MN)^{\frac{5}{2}}\right)$. Then the p-th row, j-th column element of the Fisher information matrix in (153) can be rewritten in (156), as shown at the bottom of the page, which reveals that $\left\{\widetilde{\boldsymbol{\Delta}}_{DI,1}^*, \widetilde{\boldsymbol{\Delta}}_{DI,2}^*, \widetilde{\boldsymbol{\Delta}}_{DI,3}^*\right\}$ is unrelated to $\frac{|\mathbf{s}|^2 \sigma_\beta^2}{\sigma_z^2}$.

Finally, the proof is completed.

APPENDIX I
PROOF OF LEMMA 6
We first analyze the computational complexity of Algorithm 1, which is composed of three steps:

$$
\begin{aligned}
[\mathbf{I}_{DI}(\mathbf{x}, \mathbf{W}_k)]_{p,j} &= \frac{\sigma_z^6 |\mathbf{s}|^6 \sigma_\beta^6}{|\mathbf{\Sigma}_{\mathbf{y},k}|^2}\left\{-2|\mathbf{g}_k|^2 \tilde{g}_{k,p}\tilde{g}_{k,j} + \frac{\sigma_z^2}{|\mathbf{s}|^2 \sigma_\beta^2}\operatorname{Tr}\{\mathbf{G}_{k,p}\mathbf{G}_{k,j}\} + \mathbf{g}_k^H\left(\mathbf{G}_{k,p}\mathbf{G}_{k,j} + \mathbf{G}_{k,j}\mathbf{G}_{k,p}\right)\mathbf{g}_k\right\} \\
&\stackrel{(a)}{=} \frac{\sigma_z^6 |\mathbf{s}|^6 \sigma_\beta^6}{\sigma_z^8\left(|\mathbf{s}|^2 \sigma_\beta^2 \mathbf{g}_k^H \mathbf{g}_k + \sigma_z^2\right)^2}\left\{-2|\mathbf{g}_k|^2 \tilde{g}_{k,p}\tilde{g}_{k,j} + \frac{\sigma_z^2}{|\mathbf{s}|^2 \sigma_\beta^2}\operatorname{Tr}\{\mathbf{G}_{k,p}\mathbf{G}_{k,j}\} + \mathbf{g}_k^H\left(\mathbf{G}_{k,p}\mathbf{G}_{k,j} + \mathbf{G}_{k,j}\mathbf{G}_{k,p}\right)\mathbf{g}_k\right\} \\
&= \frac{|\mathbf{s}|^2 \sigma_\beta^2}{\sigma_z^2\left(\mathbf{g}_k^H \mathbf{g}_k + \frac{\sigma_z^2}{|\mathbf{s}|^2 \sigma_\beta^2}\right)^2}\left\{-2|\mathbf{g}_k|^2 \tilde{g}_{k,p}\tilde{g}_{k,j} + \frac{\sigma_z^2}{|\mathbf{s}|^2 \sigma_\beta^2}\operatorname{Tr}\{\mathbf{G}_{k,p}\mathbf{G}_{k,j}\} + \mathbf{g}_k^H\left(\mathbf{G}_{k,p}\mathbf{G}_{k,j} + \mathbf{G}_{k,j}\mathbf{G}_{k,p}\right)\mathbf{g}_k\right\},
\end{aligned}
\tag{153}
$$

$$
\lim_{\frac{|\mathbf{s}|^2 \sigma_\beta^2}{\sigma_z^2}\to+\infty} \frac{\sigma_z^2}{|\mathbf{s}|^2 \sigma_\beta^2}[\mathbf{I}_{DI}(\mathbf{x}, \mathbf{W}_k)]_{p,j} = \frac{1}{\left(\mathbf{g}_k^H \mathbf{g}_k\right)^2}\left\{-2|\mathbf{g}_k|^2 \tilde{g}_{k,p}\tilde{g}_{k,j} + \mathbf{g}_k^H\left(\mathbf{G}_{k,p}\mathbf{G}_{k,j} + \mathbf{G}_{k,j}\mathbf{G}_{k,p}\right)\mathbf{g}_k\right\},
\tag{154}
$$

$$
\begin{aligned}
\lim_{M,N\to+\infty} \frac{[\mathbf{I}_{DI}(\mathbf{x}, \mathbf{W}_k)]_{p,j}}{(MN)^{5/2}} &= \frac{\sigma_z^6 |\mathbf{s}|^4 \sigma_\beta^4}{|\mathbf{\Sigma}_{\mathbf{y},k}|^2}\left\{-2|\mathbf{s}|^2 \sigma_\beta^2 \frac{|\mathbf{g}_k|^2 \tilde{g}_{k,p}\tilde{g}_{k,j}}{(MN)^{5/2}} + |\mathbf{s}|^2 \sigma_\beta^2 \frac{\mathbf{g}_k^H\left(\mathbf{G}_{k,p}\mathbf{G}_{k,j} + \mathbf{G}_{k,j}\mathbf{G}_{k,p}\right)\mathbf{g}_k}{(MN)^{5/2}}\right\} \\
&= \frac{\sigma_z^6 |\mathbf{s}|^6 \sigma_\beta^6}{|\mathbf{\Sigma}_{\mathbf{y},k}|^2}\left\{-2\frac{|\mathbf{g}_k|^2 \tilde{g}_{k,p}\tilde{g}_{k,j}}{(MN)^{5/2}} + \frac{\mathbf{g}_k^H\left(\mathbf{G}_{k,p}\mathbf{G}_{k,j} + \mathbf{G}_{k,j}\mathbf{G}_{k,p}\right)\mathbf{g}_k}{(MN)^{5/2}}\right\},
\end{aligned}
\tag{156}
$$

Step 1: *We evaluate the computational arithmetic operations of the Fisher information matrix inversion.*

The Fisher information matrix is obtained as follows:

$$\mathbf{I}_S\left(\hat{\boldsymbol{\psi}}_{k-1}, \mathbf{W}_k\right) = \frac{2|\mathbf{s}|^2}{\sigma_z^2} \mathrm{Re}\left\{\mathbf{V}_k^{\mathrm{H}} \mathbf{W}_k \mathbf{W}_k^{\mathrm{H}} \mathbf{V}_k\right\}, \quad (157)$$

where $\mathbf{V}_k$ is given by

$$\mathbf{V}_k = \left[\mathbf{a}\left(\hat{\mathbf{x}}_{k-1}\right), j\mathbf{a}\left(\hat{\mathbf{x}}_{k-1}\right), \hat{\beta}_{k-1}\frac{\partial\mathbf{a}\left(\hat{\mathbf{x}}_{k-1}\right)}{\partial x_1}, \hat{\beta}_{k-1}\frac{\partial\mathbf{a}\left(\hat{\mathbf{x}}_{k-1}\right)}{\partial x_2}\right]$$

$$\overset{(a)}{=} \left[\mathbf{V}_k^1, \hat{\beta}_{k-1}\mathbf{V}_k^2\right] \quad (158)$$

with Step (a) resulting from the definition of $\mathbf{V}_k^1$ and $\mathbf{V}_k^2$:

$$\mathbf{V}_k^1 \triangleq \left[\mathbf{a}\left(\hat{\mathbf{x}}_{k-1}\right), j\mathbf{a}\left(\hat{\mathbf{x}}_{k-1}\right)\right] \quad (159)$$

$$\mathbf{V}_k^2 \triangleq \left[\frac{\partial\mathbf{a}\left(\hat{\mathbf{x}}_{k-1}\right)}{\partial x_1}, \frac{\partial\mathbf{a}\left(\hat{\mathbf{x}}_{k-1}\right)}{\partial x_2}\right]. \quad (160)$$

By combining (37), (102) and (103), we can obtain that $\mathbf{W}_k^{\mathrm{H}}\mathbf{V}_k^1$ and $\mathbf{W}_k^{\mathrm{H}}\mathbf{V}_k^2$ are determined matrices that remain unchanged for different ECCs, given by

$$\mathbf{U}_1 = \mathbf{W}_k^{\mathrm{H}}\mathbf{V}_k^1 \quad (161)$$

$$\mathbf{U}_2 = \mathbf{W}_k^{\mathrm{H}}\mathbf{V}_k^2, \quad (162)$$

where both $\mathbf{U}_1$ and $\mathbf{U}_2$ can be obtained by offline calculation. Hence, we can rewrite the Fisher information matrix in (157) as:

$$\mathbf{I}_S\left(\hat{\boldsymbol{\psi}}_{k-1}, \mathbf{W}_k\right)$$

$$= \frac{2|\mathbf{s}|^2}{\sigma_z^2}\begin{bmatrix} \mathrm{Re}\left\{\mathbf{U}_1^{\mathrm{H}}\mathbf{U}_1\right\} & \mathrm{Re}\left\{\tilde{\beta}_{k-1}\mathbf{U}_1^{\mathrm{H}}\mathbf{U}_2\right\} \\ \mathrm{Re}\left\{\bar{\beta}_{k-1}\mathbf{U}_2^{\mathrm{H}}\mathbf{U}_1\right\} & |\tilde{\beta}_{k-1}|^2\mathrm{Re}\left\{\mathbf{U}_2^{\mathrm{H}}\mathbf{U}_2\right\} \end{bmatrix}$$

$$= \frac{2|\mathbf{s}|^2}{\sigma_z^2}\begin{bmatrix} \widetilde{\mathbf{A}} & \mathrm{Re}\left\{\tilde{\beta}_{k-1}\widetilde{\mathbf{B}}\right\} \\ \mathrm{Re}\left\{\bar{\beta}_{k-1}\widetilde{\mathbf{B}}^{\mathrm{H}}\right\} & |\tilde{\beta}_{k-1}|^2\widetilde{\mathbf{D}} \end{bmatrix}, \quad (163)$$

where $\bar{\beta}_{k-1}$ denotes the conjugate of $\hat{\beta}_{k-1}$ and $\widetilde{\mathbf{A}}, \widetilde{\mathbf{B}}, \widetilde{\mathbf{D}}$ are defined as:

$$\begin{cases} \widetilde{\mathbf{A}} \triangleq \mathrm{Re}\left\{\mathbf{U}_1^{\mathrm{H}}\mathbf{U}_1\right\} \\ \widetilde{\mathbf{B}} \triangleq \mathbf{U}_1^{\mathrm{H}}\mathbf{U}_2. \\ \widetilde{\mathbf{D}} \triangleq \mathrm{Re}\left\{\mathbf{U}_2^{\mathrm{H}}\mathbf{U}_2\right\} \end{cases} \quad (164)$$

Note that $\widetilde{\mathbf{A}}$ is a diagonal matrix with the same diagonal elements and the block matrices $\widetilde{\mathbf{A}}, \widetilde{\mathbf{B}}, \widetilde{\mathbf{D}}$ can all be obtained by offline calculation.

Similar to the derivation in (92), the inverse of the Fisher information matrix in (163) can be calculated by using the block matrix inversion method, given by

$$\mathbf{I}_S\left(\hat{\boldsymbol{\psi}}_{k-1}, \mathbf{W}_k\right)^{-1} = \frac{\sigma_z^2}{2|\mathbf{s}|^2}\left\{\widetilde{\mathbf{I}}_{ip_1} + \widetilde{\mathbf{I}}_{ip_2}\left(\hat{\beta}_{k-1}\right)\right\}, \quad (165)$$

where $\widetilde{\mathbf{I}}_{ip_1}$ and $\widetilde{\mathbf{I}}_{ip_2}\left(\hat{\beta}_{k-1}\right)$ are defined in (166) and (167) [as shown at the bottom of the page]

$$\widetilde{\mathbf{I}}_{ip_1} \triangleq \begin{bmatrix} \widetilde{\mathbf{A}}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \quad (166)$$

with $\widetilde{\mathbf{I}}_s$ defined as follows:

$$\widetilde{\mathbf{I}}_s \triangleq \widetilde{\mathbf{D}} - \frac{\mathrm{Re}\left\{\widetilde{\mathbf{B}}^{\mathrm{H}}\widetilde{\mathbf{A}}^{-1}\widetilde{\mathbf{B}}\right\}}{2}. \quad (168)$$

Since $\widetilde{\mathbf{A}}^{-1}$ in (166) can be obtained by offline calculation, $\widetilde{\mathbf{I}}_{ip_1}$ requires none online complex arithmetic operations.

As for $\widetilde{\mathbf{I}}_{ip_2}$, we can further rewrite it as a block matrix:

$$\widetilde{\mathbf{I}}_{ip_2}\left(\hat{\beta}_{k-1}\right) = \begin{bmatrix} \widetilde{\mathbf{I}}_{ip_2}^{11}\left(\hat{\beta}_{k-1}\right) & \widetilde{\mathbf{I}}_{ip_2}^{12}\left(\hat{\beta}_{k-1}\right) \\ \widetilde{\mathbf{I}}_{ip_2}^{21}\left(\hat{\beta}_{k-1}\right) & \widetilde{\mathbf{I}}_{ip_2}^{22}\left(\hat{\beta}_{k-1}\right) \end{bmatrix}, \quad (169)$$

where the four block matrices are given by (170), as shown at the bottom of the page.

$$\widetilde{\mathbf{I}}_{ip_2}\left(\hat{\beta}_{k-1}\right) \triangleq \begin{bmatrix} \widetilde{\mathbf{A}}^{-1}\mathrm{Re}\left\{\hat{\beta}_{k-1}\widetilde{\mathbf{B}}\right\} \\ -\mathbf{J}_2 \end{bmatrix}\left(|\hat{\beta}_{k-1}|^2\widetilde{\mathbf{I}}_s\right)^{-1}\left[\mathrm{Re}\left\{\bar{\beta}_{k-1}\widetilde{\mathbf{B}}^{\mathrm{H}}\right\}\widetilde{\mathbf{A}}^{-1} \; -\mathbf{J}_2\right] \quad (167)$$

$$\begin{cases} \widetilde{\mathbf{I}}_{ip_2}^{11}\left(\hat{\beta}_{k-1}\right) = \widetilde{\mathbf{A}}^{-1}\mathrm{Re}\left\{\hat{\beta}_{k-1}\widetilde{\mathbf{B}}\right\}\left(|\hat{\beta}_{k-1}|^2\widetilde{\mathbf{I}}_s\right)^{-1}\mathrm{Re}\left\{\bar{\beta}_{k-1}\widetilde{\mathbf{B}}^{\mathrm{H}}\right\}\widetilde{\mathbf{A}}^{-1} \\ \widetilde{\mathbf{I}}_{ip_2}^{12}\left(\hat{\beta}_{k-1}\right) = -\mathrm{Re}\left\{\hat{\beta}_{k-1}|\hat{\beta}_{k-1}|^{-2}\widetilde{\mathbf{A}}^{-1}\widetilde{\mathbf{B}}\widetilde{\mathbf{I}}_s^{-1}\right\} \\ \widetilde{\mathbf{I}}_{ip_2}^{21}\left(\hat{\beta}_{k-1}\right) = -\mathrm{Re}\left\{\bar{\beta}_{k-1}|\hat{\beta}_{k-1}|^{-2}\widetilde{\mathbf{I}}_s^{-1}\widetilde{\mathbf{B}}^{\mathrm{H}}\widetilde{\mathbf{A}}^{-1}\right\} = \left(\widetilde{\mathbf{I}}_{ip_2}^{12}\left(\hat{\beta}_{k-1}\right)\right)^{\mathrm{H}} \\ \widetilde{\mathbf{I}}_{ip_2}^{22}\left(\hat{\beta}_{k-1}\right) = \left(|\hat{\beta}_{k-1}|^{-2}\right)\widetilde{\mathbf{I}}_s^{-1}. \end{cases} \quad (170)$$

$$\widetilde{\mathbf{I}}_{ip_2}^{11}\left(\hat{\beta}_{k-1}\right) \triangleq \widetilde{\mathbf{A}}^{-1}\mathrm{Re}\left\{\hat{\beta}_{k-1}\widetilde{\mathbf{B}}\right\}\left(|\hat{\beta}_{k-1}|^2\widetilde{\mathbf{I}}_s\right)^{-1}\mathrm{Re}\left\{\bar{\beta}_{k-1}\widetilde{\mathbf{B}}^{\mathrm{H}}\right\}\widetilde{\mathbf{A}}^{-1}$$

$$= \widetilde{\mathbf{A}}^{-1}\left\{\frac{\hat{\beta}_{k-1}\widetilde{\mathbf{B}} + \bar{\beta}_{k-1}\bar{\mathbf{B}}}{2}\left(|\hat{\beta}_{k-1}|^2\widetilde{\mathbf{I}}_s\right)^{-1}\frac{\bar{\beta}_{k-1}\widetilde{\mathbf{B}}^{\mathrm{H}} + \hat{\beta}_{k-1}\widetilde{\mathbf{B}}^{\mathrm{T}}}{2}\right\}\widetilde{\mathbf{A}}^{-1}$$

$$= \widetilde{\mathbf{A}}^{-1}\frac{\mathrm{Re}\left\{\widetilde{\mathbf{B}}\widetilde{\mathbf{I}}_s^{-1}\widetilde{\mathbf{B}}^{\mathrm{H}}\right\}}{2}\widetilde{\mathbf{A}}^{-1}. \quad (171)$$

Since $\widetilde{\mathbf{A}}, \widetilde{\mathbf{B}}, \widetilde{\mathbf{D}}, \widetilde{\mathbf{I}}_s$ can all be obtained by offline calculation, $\widetilde{\mathbf{I}}_{ip_2}^{12}\left(\hat{\beta}_{k-1}\right)$ and $\widetilde{\mathbf{I}}_{ip_2}^{22}\left(\hat{\beta}_{k-1}\right)$ only require 6 online complex arithmetic operations. In addition, $\widetilde{\mathbf{I}}_{ip_2}^{21}\left(\hat{\beta}_{k-1}\right)$ requires none online complex arithmetic operations as it can be obtained directly from $\widetilde{\mathbf{I}}_{ip_2}^{12}\left(\hat{\beta}_{k-1}\right)$. As for $\widetilde{\mathbf{I}}_{ip_2}^{11}\left(\hat{\beta}_{k-1}\right)$, we can convert it to (171), as shown at the bottom of the previous page. Finally, $\widetilde{\mathbf{I}}_{ip_2}^{11}\left(\hat{\beta}_{k-1}\right)$ requires none online complex arithmetic operations. Hence, the calculation of $\widetilde{\mathbf{I}}_{ip_2}\left(\hat{\beta}_{k-1}\right)$ in (170) requires none online complex arithmetic operations for $\widetilde{\mathbf{I}}_{ip_2}^{11}\left(\hat{\beta}_{k-1}\right)$, 6 online complex arithmetic operations for $\widetilde{\mathbf{I}}_{ip_2}^{12}$, none online complex arithmetic operations for $\widetilde{\mathbf{I}}_{ip_2}^{21}\left(\hat{\beta}_{k-1}\right)$, and 5 online complex arithmetic operations for $\widetilde{\mathbf{I}}_{ip_2}^{11}\left(\hat{\beta}_{k-1}\right)$, which are 11 complex arithmetic operations in total.

In the end, the calculation of $\mathbf{I}_S\left(\hat{\boldsymbol{\psi}}_{k-1}, \mathbf{W}_k\right)^{-1}$ in (165) requires 11 online complex arithmetic operations.

Step 2: *We evaluate the computational arithmetic operations of* $\frac{\partial \log p_S(\mathbf{y}_k|\boldsymbol{\psi}, \mathbf{W}_k)}{\partial \boldsymbol{\psi}}\bigg|_{\boldsymbol{\psi}=\hat{\boldsymbol{\psi}}_{k-1}}$.

We write the $\frac{\partial \log p_S(\mathbf{y}_k|\boldsymbol{\psi}, \mathbf{W}_k)}{\partial \boldsymbol{\psi}}\bigg|_{\boldsymbol{\psi}=\hat{\boldsymbol{\psi}}_{k-1}}$ as follows:

$$
\frac{\partial \log p_S\left(\mathbf{y}_k|\boldsymbol{\psi}, \mathbf{W}_k\right)}{\partial \boldsymbol{\psi}}\bigg|_{\boldsymbol{\psi}=\hat{\boldsymbol{\psi}}_{k-1}} = \begin{bmatrix} \mathrm{Re}\left\{\mathbf{e}_k^{\mathrm{H}}\left(\mathbf{y}_k - \hat{\mathbf{y}}_k\right)\right\} \\ \mathrm{Im}\left\{\mathbf{e}_k^{\mathrm{H}}\left(\mathbf{y}_k - \hat{\mathbf{y}}_k\right)\right\} \\ \mathrm{Re}\left\{\tilde{\mathbf{e}}_{k1}^{\mathrm{H}}\left(\mathbf{y}_k - \hat{\mathbf{y}}_k\right)\right\} \\ \mathrm{Re}\left\{\tilde{\mathbf{e}}_{k2}^{\mathrm{H}}\left(\mathbf{y}_k - \hat{\mathbf{y}}_k\right)\right\} \end{bmatrix},
$$

(172)

where $\mathbf{e}_k = \mathbf{W}_k^{\mathrm{H}}\mathbf{a}\left(\hat{\mathbf{x}}_{k-1}\right)$, $\hat{\mathbf{y}}_k = |\mathbf{s}|\hat{\beta}_{k-1}\mathbf{W}_k^{\mathrm{H}}\mathbf{a}\left(\hat{\mathbf{x}}_{k-1}\right)$, $\tilde{\mathbf{e}}_{k1} = \hat{\beta}_{k-1}\mathbf{W}_k^{\mathrm{H}}\frac{\partial \mathbf{a}(\hat{\mathbf{x}}_{k-1})}{\partial x_1}$, $\tilde{\mathbf{e}}_{k2} = \hat{\beta}_{k-1}\mathbf{W}_k^{\mathrm{H}}\frac{\partial \mathbf{a}(\hat{\mathbf{x}}_{k-1})}{\partial x_2}$. Since $\mathbf{W}_k^{\mathrm{H}}\mathbf{a}\left(\hat{\mathbf{x}}_{k-1}\right)$, $\mathbf{W}_k^{\mathrm{H}}\frac{\partial \mathbf{a}(\hat{\mathbf{x}}_{k-1})}{\partial x_1}$, $\mathbf{W}_k^{\mathrm{H}}\frac{\partial \mathbf{a}(\hat{\mathbf{x}}_{k-1})}{\partial x_2}$ can all be obtained by offline calculation, $\hat{\mathbf{y}}_k$ requires 3 online complex arithmetic operations, $\mathbf{y}_k - \hat{\mathbf{y}}_k$ requires none complex arithmetic operations and $\tilde{\mathbf{e}}_{k1}$, $\tilde{\mathbf{e}}_{k2}$ both require 3 complex arithmetic operations. Together with the inner-product calculation in (172), the final number of online complex arithmetic operations is 18.

Step 3: *We evaluate the total computational arithmetic operations.*

Considering the multiplication of $\mathbf{I}_S\left(\hat{\boldsymbol{\psi}}_{k-1}, \mathbf{W}_k\right)^{-1}$ and $\frac{\partial \log p_S(\mathbf{y}_k|\boldsymbol{\psi}, \mathbf{W}_k)}{\partial \boldsymbol{\psi}}\bigg|_{\boldsymbol{\psi}=\hat{\boldsymbol{\psi}}_{k-1}}$ (16 complex arithmetic operations), and the updating direction vector plus the previous estimate (none complex arithmetic operation), the final number of online complex arithmetic operations is 45 in each ECC.

By using a similar method, the total number of complex computational arithmetic operations for Algorithm 2 (Algorithm 3) is 28 (45) in each ECC.

Therefore, Lemma 6 gets proved.

APPENDIX J
PROOF OF LEMMA 8
The following lemmas are introduced to prove Lemma 8.

Lemma 9 (Lemma 3 [42]): Given T by (133) and

$$
k_T \triangleq \inf\left\{i \in \mathbb{Z}: t_{k+i} \geq t_k + T\right\}. \tag{173}
$$

If there exists a constant $C > 0$, which satisfies

$$
\left\|\bar{\boldsymbol{\psi}}\left(t_{k+l}\right) - \tilde{\boldsymbol{\psi}}^k\left(t_{k+l}\right)\right\|_2
$$

$$
\leq L \sum_{i=1}^{l} b_{S,k+i}\left\|\bar{\boldsymbol{\psi}}\left(t_{k+i-1}\right) - \tilde{\boldsymbol{\psi}}^k\left(t_{k+i-1}\right)\right\|_2 + C, \tag{174}
$$

for all $k \geq 0$ and $1 \leq l \leq k_T$, then

$$
\sup_{t \in [t_k, t_{k+k_T}]}\left\|\bar{\boldsymbol{\psi}}(t) - \tilde{\boldsymbol{\psi}}^k(t)\right\|_2 \leq \frac{C_{\mathbf{f}}b_{S,k+1}}{2} + Ce^{L(T+b_{S,1})}, \tag{175}
$$

where L and $C_{\mathbf{f}}$ are defined in (180) and (181) separately.

Lemma 10 (Lemma 4 [43]): If $\{M_i : i = 1, 2, \ldots\}$ satisfies that: (i) M_i is Gaussian distributed with zero mean, and (ii) M_i is a martingale in i, then

$$
P\left(\sup_{0 \leq i \leq k}|M_i| > \eta\right) \leq 2\exp\left\{-\frac{\eta^2}{2\,\mathrm{Var}[M_k]}\right\}, \tag{176}
$$

for any $\eta > 0$.

Lemma 11 (Lemma 5 [43]): If given a constant $C > 0$, then

$$
G(v) = \frac{1}{v}\exp\left[-\frac{C}{v}\right], \tag{177}
$$

is increasing for all $0 < v < C$.

Let $\boldsymbol{\xi}_0 \triangleq \mathbf{0}$ and $\boldsymbol{\xi}_k \triangleq \sum_{l=1}^{k}b_{S,l}\hat{\mathbf{z}}_l$, $k \geq 1$, where $\hat{\mathbf{z}}_l$ is given in (114). With (127) and (129), we have for t_{k+l}, $1 \leq l \leq k_T$,

$$
\bar{\boldsymbol{\psi}}\left(t_{k+l}\right) = \bar{\boldsymbol{\psi}}\left(t_k\right) + \sum_{i=1}^{l}b_{S,k+i}\mathbf{f}_{\boldsymbol{\psi}}\left(\bar{\boldsymbol{\psi}}\left(t_{n+i-1}\right)\right) + \left(\boldsymbol{\xi}_{k+l} - \boldsymbol{\xi}_k\right), \tag{178}
$$

and

$$
\tilde{\boldsymbol{\psi}}^n\left(t_{k+l}\right) = \tilde{\boldsymbol{\psi}}^k\left(t_k\right) + \int_{t_k}^{t_{k+l}}\mathbf{f}_{\boldsymbol{\psi}}\left(\tilde{\boldsymbol{\psi}}^k(v)\right)dv
$$

$$
= \tilde{\boldsymbol{\psi}}^k\left(t_k\right) + \sum_{i=1}^{l}b_{S,k+i}\mathbf{f}_{\boldsymbol{\psi}}\left(\tilde{\boldsymbol{\psi}}^k\left(t_{k+i-1}\right)\right)
$$

$$
+ \int_{t_k}^{t_{k+l}}\left[\mathbf{f}_{\boldsymbol{\psi}}\left(\tilde{\boldsymbol{\psi}}^k(v)\right) - \mathbf{f}_{\boldsymbol{\psi}}\left(\tilde{\boldsymbol{\psi}}^k(\underline{v})\right)\right]dv, \tag{179}
$$

where $\underline{v} \triangleq \max\{t_k : t_k \leq v, k \geq 0\}$ for $v \geq 0$.

To bound $\int_{t_k}^{t_{k+l}}\left[\mathbf{f}_{\boldsymbol{\psi}}\left(\tilde{\boldsymbol{\psi}}^k(v)\right) - \mathbf{f}_{\boldsymbol{\psi}}\left(\tilde{\boldsymbol{\psi}}^k(\underline{v})\right)\right]dv$ on the RHS of (179), we obtain the Lipschitz constant of function $\mathbf{f}_{\boldsymbol{\psi}}(\mathbf{v})$ considering the first variable $\mathbf{v}$, given by

$$
L \triangleq \sup_{\mathbf{v}_1 \neq \mathbf{v}_2}\frac{\left\|\mathbf{f}_{\boldsymbol{\psi}}(\mathbf{v}_1) - \mathbf{f}_{\boldsymbol{\psi}}(\mathbf{v}_2)\right\|_2}{\left\|\mathbf{v}_1 - \mathbf{v}_2\right\|_2}. \tag{180}
$$

Similar to (119), for any $t \geq t_k$, we can obtain that there exists a constant $0 < C_{\mathbf{f}} < +\infty$ such that

$$\left\| \mathbf{f}_{\boldsymbol{\psi}}\left(\tilde{\boldsymbol{\psi}}^k(t) \right) \right\|_2 \leq C_{\mathbf{f}}. \tag{181}$$

Hence, we have

$$
\begin{aligned}
& \left\| \int_{t_k}^{t_{k+m}} \left[\mathbf{f}_{\boldsymbol{\psi}}\left(\tilde{\boldsymbol{\psi}}^k(v) \right) - \mathbf{f}_{\boldsymbol{\psi}}\left(\tilde{\boldsymbol{\psi}}^k(\underline{v}) \right) \right] dv \right\|_2 \\
& \leq \int_{t_k}^{t_{k+l}} \left\| \mathbf{f}_{\boldsymbol{\psi}}\left(\tilde{\boldsymbol{\psi}}^k(v) \right) - \mathbf{f}_{\boldsymbol{\psi}}\left(\tilde{\boldsymbol{\psi}}^k(\underline{v}) \right) \right\|_2 dv \\
& \overset{(a)}{\leq} \int_{t_k}^{t_{k+l}} L \left\| \tilde{\boldsymbol{\psi}}^k(v) - \tilde{\boldsymbol{\psi}}^k(\underline{v}) \right\|_2 dv \\
& \overset{(b)}{\leq} \int_{t_k}^{t_{k+l}} L \left\| \int_{\underline{v}}^{v} \mathbf{f}_{\boldsymbol{\psi}}\left(\tilde{\boldsymbol{\psi}}^k(s) \right) ds \right\|_2 dv \\
& \leq \int_{t_k}^{t_{k+l}} \int_{\underline{v}}^{v} L \left\| \mathbf{f}_{\boldsymbol{\psi}}\left(\tilde{\boldsymbol{\psi}}^k(s) \right) \right\|_2 ds dv \\
& \overset{(c)}{\leq} \int_{t_k}^{t_{k+l}} \int_{\underline{v}}^{v} C_{\mathbf{f}} L \, ds dv = \int_{t_k}^{t_{k+l}} C_{\mathbf{f}} L (v - \underline{v}) dv \\
& = \sum_{i=1}^{l} \int_{t_{k+i-1}}^{t_{k+i}} C_{\mathbf{f}} L (v - t_{k+i-1}) dv \\
& = \sum_{i=1}^{l} \frac{C_{\mathbf{f}} L (t_{k+i} - t_{k+i-1})^2}{2} = \frac{C_{\mathbf{f}} L}{2} \sum_{i=1}^{l} b_{S,k+i}^2, \tag{182}
\end{aligned}
$$

where Step (a) is due to (180), Step (b) is due to the definition in (129), and Step (c) is due to (181). Then, by subtracting $\tilde{\boldsymbol{\psi}}^k(t_{k+l})$ in (179) from $\bar{\boldsymbol{\psi}}(t_{k+l})$ in (178) and taking norms, the following inequality can be obtained from (180) and (182) for $k \geq 0, 1 \leq l \leq k_T$:

$$
\begin{aligned}
& \left\| \bar{\boldsymbol{\psi}}(t_{k+l}) - \tilde{\boldsymbol{\psi}}^k(t_{k+l}) \right\|_2 \\
& \leq L \sum_{i=1}^{l} b_{S,k+i} \left\| \bar{\boldsymbol{\psi}}(t_{k+i-1}) - \tilde{\boldsymbol{\psi}}^k(t_{k+i-1}) \right\|_2 \\
& \quad + \frac{C_{\mathbf{f}} L}{2} \sum_{i=1}^{l} b_{S,k+i}^2 + \left\| \boldsymbol{\xi}_{k+l} - \boldsymbol{\xi}_k \right\|_2 \\
& \leq L \sum_{i=1}^{l} b_{S,k+i} \left\| \bar{\boldsymbol{\psi}}(t_{k+i-1}) - \tilde{\boldsymbol{\psi}}^k(t_{k+i-1}) \right\|_2 \\
& \quad + \frac{C_{\mathbf{f}} L}{2} \sum_{i=1}^{k_T} b_{S,k+i}^2 + \sup_{1 \leq l \leq k_T} \left\| \boldsymbol{\xi}_{k+l} - \boldsymbol{\xi}_k \right\|_2. \tag{183}
\end{aligned}
$$

Applying Lemma 9 to (183) and letting

$$C = \frac{C_{\mathbf{f}} L}{2} \sum_{i=1}^{k_T} b_{S,k+i}^2 + \sup_{1 \leq l \leq k_T} \left\| \boldsymbol{\xi}_{k+l} - \boldsymbol{\xi}_k \right\|_2,$$

yields

$$
\begin{aligned}
& \sup_{t \in [t_k, t_{k+k_T}]} \left\| \bar{\boldsymbol{\psi}}(t) - \tilde{\boldsymbol{\psi}}^k(t) \right\|_2 \\
& \leq C_e \left\{ \frac{C_{\mathbf{f}} L}{2} \left[c(k) - c(k+k_T) \right] + \sup_{1 \leq l \leq k_T} \left\| \boldsymbol{\xi}_{k+l} - \boldsymbol{\xi}_k \right\|_2 \right\} \\
& \quad + \frac{C_{\mathbf{f}} c_{k+1}}{2}, \tag{184}
\end{aligned}
$$

where $C_e \triangleq e^{L(T+b_{S,1})}$, and $c(k) \triangleq \sum_{i>k} b_{S,i}^2$. Letting $k = \tilde{k}(l)$ in (184), we have $k + k_T = \tilde{k}(l+1)$ due to the definition of $T_{l+1} = t_{\tilde{k}(l+1)}$ in *Step 2* of Appendix X and

$$
\begin{aligned}
& \sup_{t \in I_l} \left\| \bar{\boldsymbol{\psi}}(t) - \tilde{\boldsymbol{\psi}}^{\tilde{k}(l)}(t) \right\|_2 \\
& \leq C_e \left\{ \frac{C_{\mathbf{f}} L}{2} \left[c(\tilde{k}(l)) - c(\tilde{k}(l+1)) \right] + \sup_{\tilde{k}(l) \leq p \leq \tilde{k}(l+1)} \left\| \boldsymbol{\xi}_p - \boldsymbol{\xi}_{\tilde{k}(l)} \right\|_2 \right\} \\
& \quad + \frac{C_{\mathbf{f}} b_{S,\tilde{k}(l)+1}}{2}. \tag{185}
\end{aligned}
$$

Suppose that the step size $\{b_{S,k} : k > 0\}$ satisfies

$$C_e \frac{C_{\mathbf{f}} L}{2} \left[c(\tilde{k}(l)) - c(\tilde{k}(l+1)) \right] + \frac{C_{\mathbf{f}} b_{S,\tilde{k}(l)+1}}{2} < \frac{\delta}{2}, \tag{186}$$

for $l \geq 0$. Given $\sup_{t \in I_l} \left\| \bar{\mathbf{x}}(t) - \tilde{\mathbf{x}}^{\tilde{k}(l)}(t) \right\| > \delta$, we can obtain from (185) and (186) that

$$
\begin{aligned}
& \sup_{\tilde{k}(l) \leq p \leq \tilde{k}(l+1)} \left\| \boldsymbol{\xi}_p - \boldsymbol{\xi}_{\tilde{k}(l)} \right\|_2 \\
& \geq \frac{1}{C_e} \left(\sup_{t \in I_l} \left\| \bar{\boldsymbol{\psi}}(t) - \tilde{\boldsymbol{\psi}}^{\tilde{k}(l)}(t) \right\|_2 \right. \\
& \qquad \left. - \frac{C_{\mathbf{f}} L}{2} \left[c(\tilde{k}(l)) - c(\tilde{k}(l+1)) \right] - \frac{C_{\mathbf{f}} b_{S,\tilde{k}(l)+1}}{2} \right) \\
& > \frac{1}{C_e} \left(\sup_{t \in I_l} \left| \bar{\mathbf{x}}(t) - \tilde{\mathbf{x}}^{\tilde{k}(l)}(t) \right| - \frac{\delta}{2} \right) \\
& > \frac{\delta}{2C_e}.
\end{aligned}
$$

Then, we get

$$
\begin{aligned}
& P \left(\sup_{t \in I_l} \left\| \bar{\mathbf{x}}(t) - \tilde{\mathbf{x}}^{\tilde{k}(l)}(t) \right\| > \delta \right| \\
& \qquad \times \sup_{t \in I_i} \left\| \bar{\mathbf{x}}(t) - \tilde{\mathbf{x}}^{\tilde{k}(i)}(t) \right\| \leq \delta, 0 \leq i < l \right) \\
& \leq P \left(\sup_{\tilde{k}(l) \leq p \leq \tilde{k}(l+1)} \left\| \boldsymbol{\xi}_p - \boldsymbol{\xi}_{\tilde{k}(l)} \right\|_2 > \frac{\delta}{2C_e} \right| \\
& \qquad \times \sup_{t \in I_i} \left\| \bar{\mathbf{x}}(t) - \tilde{\mathbf{x}}^{\tilde{k}(i)}(t) \right\| \leq \delta, 0 \leq i < l \right) \\
& \overset{(d)}{=} P \left(\sup_{\tilde{k}(l) \leq p \leq \tilde{k}(l+1)} \left\| \boldsymbol{\xi}_p - \boldsymbol{\xi}_{\tilde{k}(l)} \right\|_2 > \frac{\delta}{2C_e} \right), \tag{187}
\end{aligned}
$$

where Step (d) is due to the independence of noise, i.e., $\boldsymbol{\xi}_p - \boldsymbol{\xi}_{\tilde{k}(l)}, \tilde{k}(l) \leq p \leq \tilde{k}(l+1)$ are independent of $\hat{\mathbf{x}}_k, 0 \leq k \leq \tilde{k}(l)$.

The lower bound of the probability that the sequence $\{\hat{\mathbf{x}}_k : k \geq 0\}$ remains in the invariant set $\mathcal{I}$ is given by

$$
\begin{aligned}
& P \left(\hat{\mathbf{x}}_k \in \mathcal{I}, \forall k \geq 0 \right) \\
& \overset{(e)}{\geq} P \left(\sup_{t \in I_l} \left\| \bar{\mathbf{x}}(t) - \tilde{\mathbf{x}}^{\tilde{k}(l)}(t) \right\| \leq \delta, \forall l \geq 0 \right)
\end{aligned}
$$

$$\overset{(f)}{\geq} 1 - \sum_{l \geq 0} P\left(\sup_{t \in I_l} \left\|\bar{\mathbf{x}}(t) - \tilde{\mathbf{x}}^{\tilde{k}(l)}(t)\right\| > \delta\right|$$

$$\times \sup_{t \in I_i} \left\|\bar{\mathbf{x}}(t) - \tilde{\mathbf{x}}^{\tilde{k}(i)}(t)\right\| \leq \delta, 0 \leq i < l\right)$$

$$\overset{(g)}{\geq} 1 - \sum_{l \geq 0} P\left(\sup_{\tilde{k}(l) \leq p \leq \tilde{k}(l+1)} \left\|\boldsymbol{\xi}_p - \boldsymbol{\xi}_{\tilde{k}(l)}\right\|_2 > \frac{\delta}{2C_e}\right),$$

(188)

where Step (*e*) is due to Lemma 7, Step (*f*) is due to Lemma 4.2 in [36], and Step (*g*) is due to (187). Let $\|\cdot\|_\infty$ denote the max-norm, i.e., $\|\mathbf{u}\|_\infty = \max_l |[\mathbf{u}]_l|$. Note that for $\mathbf{u} \in \mathbb{R}^D$, $\|\mathbf{u}\|_2 \leq \sqrt{D} \|\mathbf{u}\|_\infty$. Hence we have

$$P\left(\sup_{\tilde{k}(l) \leq p \leq \tilde{k}(l+1)} \left\|\boldsymbol{\xi}_p - \boldsymbol{\xi}_{\tilde{k}(l)}\right\|_2 > \frac{\delta}{2C_e}\right)$$

$$\leq P\left(\sup_{\tilde{k}(l) \leq p \leq \tilde{k}(l+1)} \left\|\boldsymbol{\xi}_p - \boldsymbol{\xi}_{\tilde{k}(l)}\right\|_\infty > \frac{\delta}{4C_e}\right)$$

$$= P\left(\sup_{\tilde{k}(l) \leq p \leq \tilde{k}(l+1)} \max_{1 \leq j \leq 4} \left|[\boldsymbol{\xi}_p]_j - [\boldsymbol{\xi}_{\tilde{k}(l)}]_j\right| > \frac{\delta}{4C_e}\right)$$

$$= P\left(\max_{1 \leq j \leq 4} \sup_{\tilde{k}(l) \leq p \leq \tilde{k}(l+1)} \left|[\boldsymbol{\xi}_p]_j - [\boldsymbol{\xi}_{\tilde{k}(l)}]_j\right| > \frac{\delta}{4C_e}\right)$$

$$\leq \sum_{j=1}^{4} P\left(\sup_{\tilde{k}(l) \leq p \leq \tilde{k}(l+1)} \left|[\boldsymbol{\xi}_p]_j - [\boldsymbol{\xi}_{\tilde{k}(l)}]_j\right| > \frac{\delta}{4C_e}\right).$$

(189)

With the increasing σ-fields $\{\mathcal{G}_k : k \geq 0\}$ defined in Appendix X, we have for $k \geq 0$,

1) $\boldsymbol{\xi}_k = \sum_{l=1}^{k} b_{S,l}\hat{\mathbf{z}}_l \sim \mathcal{N}(0, \sum_{l=1}^{k} b_{S,k}^2 \mathbf{I}_S(\hat{\boldsymbol{\psi}}_{l-1}, \mathbf{W}_l)^{-1})$,
2) $\boldsymbol{\xi}_k$ is $\mathcal{G}_k$-measurable, i.e., $\mathbb{E}[\boldsymbol{\xi}_k | \mathcal{G}_k] = \boldsymbol{\xi}_k$,
3) $\mathbb{E}[\|\boldsymbol{\xi}_k\|_2^2] = \sum_{l=1}^{k} b_{S,k}^2 \operatorname{tr}\{\mathbf{I}_S(\hat{\boldsymbol{\psi}}_{l-1}, \mathbf{W}_l)^{-1}\} < +\infty$,
4) $\mathbb{E}[\boldsymbol{\xi}_k | \mathcal{G}_l] = \boldsymbol{\xi}_l$ for all $0 \leq l < k$.

Therefore, $[\boldsymbol{\xi}_k]_j, j = 1, 2, 3, 4$ is a Gaussian martingale with respect to $\mathcal{G}_k$, and satisfies

$$\operatorname{Var}\left[[\boldsymbol{\xi}_{k+l}]_j - [\boldsymbol{\xi}_k]_j\right] = \sum_{i=k+1}^{k+l} b_{S,i}^2 \left[\mathbf{I}_S(\hat{\boldsymbol{\psi}}_{i-1}, \mathbf{W}_i)^{-1}\right]_{j,j}$$

$$\leq \sum_{i=k+1}^{k+l} b_{S,i}^2 \frac{C_{\mathbf{I}}\sigma_z^2}{|\mathbf{s}|^2}$$

$$= \frac{C_{\mathbf{I}}\sigma_z^2}{|\mathbf{s}|^2}[c(k) - c(k+l)], \quad (190)$$

where $C_{\mathbf{I}} \overset{\Delta}{=} \max_s \max_{i \geq 1} \frac{|\mathbf{s}|^2}{\sigma_z^2}[\mathbf{I}(\hat{\boldsymbol{\psi}}_{i-1}, \mathbf{W}_i)^{-1}]_{j,j}$. Let $\eta = \frac{\delta}{4C_e}$, $M_i = [\boldsymbol{\xi}_{\tilde{k}(l)+i}]_j - [\boldsymbol{\xi}_{\tilde{k}(l)}]_j, j = 1, 2, 3, 4$ and $p = \tilde{k}(l+1) - \tilde{k}(l)$ in Lemma 10, then from (189) and (190), we

can obtain

$$P\left(\sup_{\tilde{k}(l) \leq p \leq \tilde{k}(l+1)} \left|[\boldsymbol{\xi}_p]_j - [\boldsymbol{\xi}_{\tilde{k}(l)}]_j\right| > \frac{\delta}{4C_e}\right)$$

$$\leq 2\exp\left\{-\frac{\delta^2}{32C_e^2 \operatorname{Var}\left[[\boldsymbol{\xi}_{\tilde{k}(l)+i}]_j - [\boldsymbol{\xi}_{\tilde{k}(l)}]_j\right]}\right\}$$

$$\leq 2\exp\left\{-\frac{\delta^2|\mathbf{s}|^2}{32C_{\mathbf{I}}C_e^2[c(\tilde{k}(l)) - c(\tilde{k}(l+1))]\sigma_z^2}\right\}.$$

(191)

Combining (188), (189) and (191), we have

$$P\left(\hat{\mathbf{x}}_k \in \mathcal{I}, \forall k \geq 0\right)$$

$$\geq 1 - 8\sum_{l \geq 0} \exp\left\{-\frac{\delta^2|\mathbf{s}|^2}{32C_{\mathbf{I}}C_e^2[c(\tilde{k}(l)) - c(\tilde{k}(l+1))]\sigma_z^2}\right\}.$$

(192)

To use Lemma 11, we assume that the step-size $b_{S,k}$ satisfies

$$c(0) = \sum_{i>0} b_{S,i}^2 \leq \frac{\delta^2|\mathbf{s}|^2}{32C_{\mathbf{I}}C_e^2\sigma_z^2}. \quad (193)$$

Then, from Lemma 11, we can obtain

$$\frac{\exp\left\{-\frac{\delta^2|\mathbf{s}|^2}{32C_{\mathbf{I}}C_e^2[c(\tilde{k}(l))-c(\tilde{k}(l+1))]\sigma_z^2}\right\}}{c(\tilde{k}(l)) - c(\tilde{k}(l+1))} \leq \frac{\exp\left\{-\frac{\delta^2|\mathbf{s}|^2}{32C_{\mathbf{I}}C_e^2 c(0)\sigma_z^2}\right\}}{c(0)}$$

for $c(\tilde{k}(l)) - c(\tilde{k}(l+1)) < c(\tilde{k}(l)) \leq c(0)$. Hence, we have

$$\sum_{l \geq 0} \exp\left\{-\frac{\delta^2|\mathbf{s}|^2}{32C_{\mathbf{I}}C_e^2[c(\tilde{k}(l)) - c(\tilde{k}(l+1))]\sigma_z^2}\right\}$$

$$\leq \sum_{l \geq 0}\left[c(\tilde{k}(l)) - c(\tilde{k}(l+1))\right] \cdot \frac{\exp\left\{-\frac{\delta^2|\mathbf{s}|^2}{32C_{\mathbf{I}}C_e^2 c(0)\sigma_z^2}\right\}}{c(0)}$$

$$= c(0) \cdot \frac{\exp\left\{-\frac{\delta^2|\mathbf{s}|^2}{32C_{\mathbf{I}}C_e^2 c(0)\sigma_z^2}\right\}}{c(0)} = \exp\left\{-\frac{\delta^2|\mathbf{s}|^2}{32C_{\mathbf{I}}C_e^2 c(0)\sigma_z^2}\right\}.$$

(194)

As $C_e = e^{L(T+b_{S,1})}$, $c(0) = \sum_{i>0} b_{S,i}^2$, and $b_{S,k}, T, L$ are given by (42), (133), (180) respectively, we can obtain

$$\frac{\delta^2|\mathbf{s}|^2}{32C_{\mathbf{I}}C_e^2 c(0)\sigma_z^2} = \frac{\delta^2|\mathbf{s}|^2}{32C_{\mathbf{I}}e^{2L(T+\frac{\epsilon_S}{K_{S,0}+1})}\sigma_z^2 \sum_{i \geq 1}\frac{\epsilon_S^2}{(i+K_{S,0})^2}}$$

$$= \frac{\delta^2}{\sum_{i \geq 1}\frac{32C_{\mathbf{I}}e^{2L(T+\frac{\epsilon_S}{K_{S,0}+1})}}{(i+K_{S,0})^2}} \cdot \frac{|\mathbf{s}|^2}{\epsilon_S^2\sigma_z^2}. \quad (195)$$

In (195), $0 < \delta < \inf_{\mathbf{v} \in \partial\mathcal{B}} \|\mathbf{v} - \hat{\mathbf{x}}_b\|$, (186) and (193) should be satisfied, where a sufficiently large $K_{S,0} \geq 0$ can make both (186) and (193) true.

To ensure that $\hat{\mathbf{x}}_0 + b_{S,1}\left[\mathbf{f}_{\boldsymbol{\psi}}\left(\hat{\boldsymbol{\psi}}_0\right)\right]_{3,4}$ does not exceed the main lobe $\mathcal{B}(\mathbf{x})$, i.e., the first step-size $b_{S,1}$ satisfies

$$\left|\hat{x}_{0,1} + b_{S,1}\left[\mathbf{f}_{\boldsymbol{\psi}}\left(\hat{\boldsymbol{\psi}}_0\right)\right]_3 - x_1\right| < 1$$

$$\left|\hat{x}_{0,2} + b_{S,1}\left[\mathbf{f}_{\boldsymbol{\psi}}\left(\hat{\boldsymbol{\psi}}_0\right)\right]_4 - x_2\right| < 1,$$

we can obtain the maximum ϵ_S as follows

$$\epsilon_{S,\max} = \min \frac{(K_{S,0}+1)}{\left|\left[\mathbf{f}_{\boldsymbol{\psi}}\left(\hat{\boldsymbol{\psi}}_0\right)\right]_3\right|}\left\{1-|x_1-\hat{x}_{0,1}|,\ 1-|x_2-\hat{x}_{0,2}|\right\}$$

$$\leq \frac{(K_{S,0}+1)}{\left|\left[\mathbf{f}_{\boldsymbol{\psi}}\left(\hat{\boldsymbol{\psi}}_0\right)\right]_3\right|} \triangleq \epsilon_b. \tag{196}$$

Hence, from (195), we have

$$\frac{\delta^2|\mathbf{s}|^2}{32C_{\mathbf{I}}C_e^2 c(0)\sigma_z^2} \cdot \frac{\epsilon_S^2\sigma_z^2}{|\mathbf{s}|^2} \geq \frac{\delta^2}{\sum\limits_{i\geq 1}\frac{32C_{\mathbf{I}}e^{2L\left(T+\frac{\epsilon_b}{K_{S,0}+1}\right)}}{(i+K_{S,0})^2}} \triangleq R. \tag{197}$$

Combining (192), (194) and (197), yields

$$P\left(\hat{\mathbf{x}}_k \in \mathcal{I}, \forall k \geq 0\right) \geq 1 - 8e^{-\frac{R|\mathbf{s}|^2}{\epsilon_S^2\sigma_z^2}},$$

which completes the proof.

REFERENCES

[1] M. Xiao, S. Mumtaz, Y. Huang, L. Dai, Y. Li, M. Matthaiou, G. K. Karagiannidis, E. Björnson, K. Yang, C.-L. I, and A. Ghosh, "Millimeter wave communications for future mobile networks," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1909–1935, Sep. 2017.

[2] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 101–107, Jun. 2011.

[3] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[4] S. Han, C.-L. I, Z. Xu, and C. Rowell, "Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G," *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 186–194, Jan. 2015.

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

[6] A. F. Molisch, V. V. Ratnam, S. Han, Z. Li, S. L. H. Nguyen, L. Li, and H. Katsuyuki, "Hybrid beamforming for massive MIMO: A survey," *IEEE Commun. Mag.*, vol. 55, no. 9, pp. 134–141, Sep. 2017.

[7] 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.

[8] D. Zhang, A. Li, M. Shirvanimoghaddam, P. Cheng, Y. Li, and B. Vucetic, "Codebook-based training beam sequence design for millimeter-wave tracking systems," *IEEE Trans. Wireless Commun.*, vol. 18, no. 11, pp. 5333–5349, Nov. 2019.

[9] J. Li, Y. Sun, L. Xiao, S. Zhou, and C. E. Koksal, "Analog beam tracking in linear antenna arrays: Convergence, optimality, and performance," in *Proc. 51st Asilomar Conf. Signals, Syst., Comput.*, Oct. 2017, pp. 1193–1198.

[10] J. Li, Y. Sun, L. Xiao, S. Zhou, and A. Sabharwal, "How to mobilize mmwave: A joint beam and channel tracking approach," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 3624–3628.

[11] N. Garcia, H. Wymeersch, and D. T. M. Slock, "Optimal precoders for tracking the AoD and AoA of a mmWave path," *IEEE Trans. Signal Process.*, vol. 66, no. 21, pp. 5718–5729, Nov. 2018.

[12] A. Alkhateeb, G. Leus, and R. W. Heath, Jr., "Compressed sensing based multi-user millimeter wave systems: How many measurements are needed?" in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2015, pp. 2909–2913.

[13] *IEEE 802.11ad WLAN Enhancements for Very High Throughput in the 60 GHz Band*, Standard IEEE 802.11ad, Dec. 2012.

[14] R. Mendez-Rial, C. Rusu, N. Gonzalez-Prelcic, A. Alkhateeb, and R. W. Heath, Jr., "Hybrid MIMO architectures for millimeter wave communications: Phase shifters or switches?" *IEEE Access*, vol. 4, pp. 247–267, 2016.

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

[16] *NR—Physical Layer Procedures for Data—Release 15*, 3GPP Standard TS 38.214, Jun. 2019.

[17] T. S. Rappaport, F. Gutierrez, E. Ben-Dor, J. N. Murdock, Y. Qiao, and J. I. Tamir, "Broadband millimeter-wave propagation measurements and models using adaptive-beam antennas for outdoor urban cellular communications," *IEEE Trans. Antennas Propag.*, vol. 61, no. 4, pp. 1850–1859, Apr. 2013.

[18] G. Brown, O. Koymen, and M. Branda, "The promise of 5G mmWave-How do we make it mobile?" Qualcomm Technol., San Diego, CA, USA, Tech. Rep., Jun. 2016. [Online]. Available: https://www.qualcomm.com/media/documents/files/the-promise-of-5g-mmwave-how-do-we-make-it-mobile.pdf

[19] V. Boljanovic, H. Yan, and D. Cabric, "Tracking sparse mmWave channel under time varying multipath scatterers : (Invited Paper)," in *Proc. 52nd Asilomar Conf. Signals, Syst., Comput.*, Oct. 2018, pp. 1274–1279.

[20] Y. Liu, J. Li, Y. Sun, and S. Zhou, "Joint beam and channel tracking for two-dimensional phased antenna arrays," in *Proc. 2nd ACM Workshop Millim. Wave Netw. Sens. Syst. (mmNets)*, vol. 1, 2018, pp. 15–20.

[21] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1164–1179, Jun. 2014.

[22] D. Zhu, J. Choi, and R. W. Heath, Jr., "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.

[23] X. Meng, X. Gao, and X.-G. Xia, "Omnidirectional precoding and combining based synchronization for millimeter wave massive MIMO systems," *IEEE Trans. Commun.*, vol. 66, no. 3, pp. 1013–1026, Mar. 2018.

[24] D. Zhu, R. Bendlin, S. Akoum, A. Ghosh, and R. W. Heath, Jr., "Directional frame timing synchronization in wideband millimeter-wave systems with low-resolution ADCs," *IEEE Trans. Wireless Commun.*, vol. 18, no. 11, pp. 5350–5366, Nov. 2019.

[25] D. Zhu, R. Bendlin, S. Akoum, A. Ghosh, and R. W. Heath, Jr., "Double-sequence frequency synchronization for wideband millimeter-wave systems with few-bit ADCs," *IEEE Trans. Wireless Commun.*, vol. 19, no. 2, pp. 1357–1372, Feb. 2020.

[26] D. V. Lindley, B. V. Gnedenko, A. N. Kolmogorov, and K. L. Chung, "Limit distributions for sums of independent random variables," *J. Roy. Stat. Soc.*, vol. 39, no. 330, pp. 342–343, Dec. 1955.

[27] W. S. Lovejoy, "Computationally feasible bounds for partially observed Markov decision processes," *Oper. Res.*, vol. 39, no. 1, pp. 162–175, Feb. 1991.

[28] M. Hauskrecht, "Value-function approximations for partially observable Markov decision processes," *J. Artif. Intell. Res.*, vol. 13, no. 1, pp. 33–94, Aug. 2000.

[29] Z. Xiao, X.-G. Xia, D. Jin, and N. Ge, "Iterative eigenvalue decomposition and multipath-grouping Tx/Rx joint beamformings for millimeter-wave communications," *IEEE Trans. Wireless Commun.*, vol. 14, no. 3, pp. 1595–1607, Mar. 2015.

[30] N. Ishikawa, R. Rajashekar, C. Xu, M. El-Hajjar, S. Sugiura, L.-L. Yang, and L. Hanzo, "Differential-detection aided large-scale generalized spatial modulation is capable of operating in high-mobility millimeter-wave channels," *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 6, pp. 1360–1374, Oct. 2019.

[31] Y. Ke, H. Gao, W. Xu, L. Li, L. Guo, and Z. Feng, "Position prediction based fast beam tracking scheme for multi-user UAV-mmWave communications," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2019, pp. 1–7.

[32] M. Shafi, J. Zhang, H. Tataria, A. F. Molisch, S. Sun, T. S. Rappaport, F. Tufvesson, S. Wu, and K. Kitao, "Microwave vs. millimeter-wave propagation channels: Key differences and impact on 5G cellular systems," *IEEE Commun. Mag.*, vol. 56, no. 12, pp. 14–20, Dec. 2018.

[33] S. K. Sengijpta, "Fundamentals of statistical signal processing: Estimation theory," *Technometrics*, vol. 37, no. 4, pp. 465–466, Nov. 1995.

[34] M. B. Nevel'son and R. Z. Has'minskii, *Stochastic Approximation and Recursive Estimation*. Providence, RI, USA: American Mathematical Society, 1973.

[35] J. C. Spall, *Introduction to Stochastic Search and Optimization*. New York, NY, USA: Wiley, 2003.

[36] V. S. Borkar, *Stochastic Approximation: A Dynamical Systems Viewpoint*. Cambridge, U.K.: Cambridge Univ. Press, 2008.

[37] H. Kushner and G. G. Yin, *Stochastic Approximation and Recursive Algorithms and Applications*. New York, NY, USA: Springer, 2003.

[38] *NR—Study on Channel Model for Frequencies From 0.5 to 100 GHz—Release 15*, 3GPP Standard TS 38.901, Sep. 2019.

[39] T. T. Cai and L. Wang, "Orthogonal matching pursuit for sparse signal recovery with noise," *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4680–4688, Jul. 2011.

[40] M. K. Samimi, G. R. MacCartney, S. Sun, and T. S. Rappaport, "28 GHz millimeter-wave ultrawideband small-scale fading models in wireless channels," in *Proc. IEEE 83rd Veh. Technol. Conf. (VTC Spring)*, May 2016, pp. 1–6.

[41] D. Simon, *Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches*. New York, NY, USA: Wiley, 2006.

[42] J. Li, Y. Sun, L. Xiao, S. Zhou, and A. Sabharwal, "How to mobilize mmWave: A joint beam and channel tracking approach," 2018, *arXiv:1802.02125*. [Online]. Available: http://arxiv.org/abs/1802.02125

[43] J. Li, Y. Sun, L. Xiao, S. Zhou, and C. E. Koksal, "Fast analog beam tracking in phased antenna arrays: Theory and performance," 2017, *arXiv:1710.07873*. [Online]. Available: http://arxiv.org/abs/1710.07873

YU LIU received the B.E. degree in electronic engineering from Tsinghua University, Beijing, in 2016, where he is currently pursuing the Ph.D. degree with the Wireless Communication Center, Department of Electronic Engineering. His research interests include millimeter-wave mobile communication, channel estimation, and channel tracking.
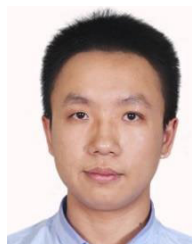
JIAHUI LI received the B.E. degree in electronic information science and technology from Tsinghua University, China, in 2013, where he is currently pursuing the Ph.D. degree in information and communication engineering. His research interests include channel estimation, beamforming, and resource allocation of massive MIMO and mmWave systems.

XIUJUN ZHANG (Member, IEEE) received the B.E. and Ph.D. degrees in electronic engineering from Tsinghua University, Beijing, China, in 2001 and 2013, respectively. She is currently an Assistant Researcher with the Beijing National Research Center for Information Science and Technology. Her research interests include signal processing and wireless communications.

SHIDONG ZHOU (Member, IEEE) received the B.S. and M.S. degrees from Southeast University, Nanjing, China, in 1991 and 1994, respectively, and the Ph.D. degree from Tsinghua University, Beijing, China, in 1998. He is currently a Professor with the Department of Electronic Engineering, Tsinghua University. He was involved in Program for New Century Excellent Talents in University 2005 (Ministry of Education). His research interests include wireless transmission techniques, including distributed wireless communication systems, channel sounding and modeling, coordination of communication, control and computing, and application in future mobile communications. He received the Special Award of 2016 National Prize for Progress in Science and Technology.

• • •