

Blind Construction of Angular Power Maps in Massive MIMO Networks

Zheng Xing[✉], *Member, IEEE*, and Junting Chen[✉], *Member, IEEE*

Abstract—Channel state information (CSI) acquisition is a challenging problem in massive multiple-input multiple-output (MIMO) networks. Radio maps provide a promising solution for radio resource management by reducing online CSI acquisition. However, conventional approaches for radio map construction require location-labeled CSI data, which is challenging in practice. This paper investigates unsupervised angular power map construction based on large timescale CSI data collected in a massive MIMO network without location labels. A hidden Markov model (HMM) is built to connect the hidden trajectory of a mobile with the CSI evolution of a massive MIMO channel. As a result, the mobile location can be estimated, enabling the construction of an angular power map. We show that under uniform rectilinear mobility with Poisson-distributed base stations (BSs), the Cramer-Rao Lower Bound (CRLB) for localization error can vanish at any signal-to-noise ratios (SNRs), whereas when BSs are confined to a limited region, the error remains nonzero even with infinite independent measurements. Based on reference signal received power (RSRP) data collected in a real multi-cell massive MIMO network, an average localization error of 18 meters can be achieved although measurements are mainly obtained from a single serving cell.

Index Terms—Angular power map, massive MIMO network, trajectory recovery, localization, CSI prediction.

I. INTRODUCTION

ACQUIRING channel state information (CSI) is essential in beamforming, resource allocation, and inter-cell

interference mitigation in multiple-input multiple-output (MIMO) networks. However, it becomes more and more challenging in estimating the CSI as the number of antennas scale up in trending massive MIMO networks [1], [2], [3]. Radio maps, which associate each mobile location with the corresponding CSI, provide a new paradigm for CSI acquisition, tracking, and prediction [4], [5], [6], [7]. For example, the work [8] enhances low-latency MIMO communications using statistical radio maps that predict and select communication parameters for reliability. Another work [9] introduces an environment-aware hybrid beamforming for mmWave massive MIMO systems, utilizing channel knowledge maps to reduce real-time training needs and improve communication rates with location accuracy flexibility.

Nonetheless, it is very challenging to construct radio maps for multi-cell MIMO communications, and one of the main challenges is the lack of location-labeled MIMO channel data collected from real scenarios. In reality, accurate location information of the mobile users is hard to obtain. First, it requires the users to continuously report the location to the network, which may not always be feasible due to privacy concerns. Second, the localization accuracy is significantly affected by non-line-of-sight (NLOS) conditions. Third, dedicated measurement campaign using drive test is costly and less timely. Nevertheless, existing approaches for radio map construction mostly require a massive amount of location-labeled CSI measurement data. For instance, the work [10] necessitated gathering a large volume of CSI with location labels to train a deep generative model. The research [11] involved collecting CSI with location labels via vehicles and employs an long short-term memory (LSTM)-based neural network to construct road radio maps. The work [12] collected a substantial amount of CSI data with location labels in ground node- unmanned aerial vehicle (UAV) networks to construct radio maps. Moreover, the study [13] utilized Kriging and covariance tapering techniques to construct radio maps in massive MIMO systems using a small amount of location-labeled CSI data. To relieve the requirement on the massive amount of location-labeled CSI data for radio map construction, the work [14] employed sparse sampling and Bayesian learning inference techniques to construct radio maps using a limited amount of location-labeled CSI data.

Is it possible to learn the radio map for MIMO communications in a completely *unsupervised* manner based on the CSI data without accurate location information? CSI captures how the signal propagation is affected by the

Received 20 December 2024; revised 7 June 2025; accepted 26 September 2025. Date of publication 14 October 2025; date of current version 5 December 2025. This work was supported in part by the NSFC under Grant 62293482; in part by the Basic Research Project through Hetao Shenzhen-HK S&T Cooperation Zone under Project HZQB-KCZY-2021067; in part by the National Key R&D Program of China under Grant 2024YFB2907500; in part by the NSFC under Grant 62171398; in part by the Guangdong Basic and Applied Basic Research Foundation under Grant 2024A1515011206; in part by the Shenzhen Science and Technology Program under Grant JCYJ20220530143804010 and Grant KJZD20230923115104009; in part by the Shenzhen Outstanding Talents Training Fund under Grant 2020002; in part by the Guangdong Research under Project 2017ZT07X152 and Project 2019CX01X104; in part by the Guangdong Provincial Key Laboratory of Future Networks of Intelligence under Grant 2022B1212010001; and in part by the Shenzhen Key Laboratory of Big Data and Artificial Intelligence under Grant ZDSYS201707251409055. The associate editor coordinating the review of this article and approving it for publication was Piya Pal. (Corresponding author: Junting Chen.)

The authors are with the School of Science and Engineering, and Shenzhen Future Network of Intelligence Institute (FNii-Shenzhen), The Chinese University of Hong Kong, Shenzhen 518172, China (e-mail: zhengxing@link.cuhk.edu.cn; juntunge@cuhk.edu.cn).

This article has supplementary downloadable material available at <https://doi.org/10.1109/TSP.2025.3620332>, provided by the authors.

Digital Object Identifier 10.1109/TSP.2025.3620332

environment [15], [16], [17], [18], [19], [20], [21]. Thus, although a single CSI realization tends to be random, there exists some geographical pattern of the CSI distribution, which is determined by the environment and cellular MIMO network topology [22], [23], [24], [25]. However, while the current cellular networks have been capturing extensive volumes of CSI from numerous mobile users for MIMO transmission and radio resource management, the instantaneous CSI is usually discarded immediately after the transmission.

This paper proposes to leverage the unlabeled instantaneous CSI to train a radio map model before the CSI data is discarded. The ultimate goal is to build a data-driven model that can describe the evolution of the CSI process. The physical intuition is that although the CSI in a massive MIMO network is of high dimension, the mobility of the corresponding user is contained in a 3D physical world, and the dominant scattering environment, such as building and vegetation, remains roughly and temporarily static. Mathematically, while the CSI process evolves in a high dimensional space, it can be embedded in a low dimensional latent space, which represents the geographic environment. Similar ideas were attempted in the channel charting literature. The work [26] proposed a channel charting approach to reduce the dimensionality of CSI data to two or three dimensions. The location labeling can be achieved by rotating the 2D or 3D data in the latent space based on a limited amount of location labels. Subsequent studies [27], [28] have introduced various channel charting methods, focusing on techniques such as auto-encoders, and Siamese networks. Yet, in the channel charting literature, the latent space does not have to represent the physical world, and hence, it is still of high interest to investigate whether it is possible to embed a geographical model to describe the low-dimensional latent space of the CSI sequence data.

More specifically, the aim of this paper is to investigate the following fundamental questions: (i) *Can we establish a radio map model that maps the CSI to a latent space that has a clear physical meaning as the geographic area*, and (ii) *Is there any theoretical guarantee in recovering the user location from the CSI*.

For elaboration purpose, a multi-cell massive MIMO network is considered, where a mobile user travels along an arbitrary trajectory in the coverage area of the network and the network keeps acquiring some partial CSI of the user. We propose a hidden Markov model (HMM) to establish the connection between the mobility of the user and the CSI evolution partially observed by the network. Under this context, to our best knowledge, the most related work appeared in [29] for indoor localization employs an HMM-based method to reconstruct trajectories from indoor received signal strength (RSS) measurements. This method, however, necessitates extensive RSS collection at each location within the indoor layout and encounters mapping failures in symmetric building configurations. Our prior work [30] employed an HMM to extract coarse user locations from measurements and recover their trajectories up to a region-level accuracy. While this approach allows for the construction of a radio map using the measurements and the estimated trajectory without any

location labels, it does not address trajectory recovery in MIMO scenarios.

To summarize, the following contributions are made:

- We develop a framework to recover the location labels from a sequence of CSI measurements using Bayesian methods. Our approach does not require the continuity of the CSI process, and hence, can be adopted in NLOS scenario.
- We establish theoretical results to show that under uniform rectilinear mobility, the Cramer-Rao Lower Bound (CRLB) of the localization error can asymptotically approach zero at any signal-to-noise ratio (SNR) if the base station (BS) topology can be modeled by a Poisson Point Process (PPP) over a large enough region. By contrast, if the BSs are only deployed in a limited region, localization error cannot approach zero even using infinite amount of independent measurement data.
- We design efficient algorithms to solve for the joint trajectory recovery and propagation parameter estimation problem, where an iterative log-transformation technique is developed to solve a nonlinear regression problem that contains coupled polynomial terms and exponential terms.
- We conduct experiments on real dataset collected from a commercial 5G massive MIMO network. A mean localization error below 18 meters from secondary synchronization block (SSB) reference signal received power (RSRP) measurements is demonstrated, although measurements from neighboring cells are mostly missing in the dataset. When richer measurements are available, a localization error of 7 meters is demonstrated from a synthetic dataset. To demonstrate the application of the angular power map, we predict the RSRP, signal-to-interference-and-noise ratio (SINR), and received signal strength indicator (RSSI) of the CSI beam using the constructed radio map. The proposed radio map-assisted method achieves the lowest errors in comparison with existing methods.

The remainder of this paper is structured as follows. Section II presents the propagation model, measurement mode, mobility model, and an HMM formulation. Section III derives the theoretical results for the CRLB of the localization error, considering both a limited region and an unlimited region for the deployment of the BS. Section IV focuses on the development of the trajectory recovery algorithm. Experimental results are provided in Section V, and the paper concludes in Section VI.

II. SYSTEM MODEL

Consider a mobile user traveling along an arbitrary trajectory as shown in Fig. 1. Denote $\mathbf{x} \in \mathbb{R}^2$ as the location of the mobile user. The communication signal of the mobile user can be acquired by Q BSs each equipped with N_t antennas, where the location of the q th BS is denoted as $\mathbf{o}_q \in \mathbb{R}^2$, $q = 1, 2, \dots, Q$. Denote $d(\mathbf{x}, \mathbf{o}_q) = \|\mathbf{x} - \mathbf{o}_q\|_2$ as the distance between the user at location $\mathbf{x}$ and the BS at $\mathbf{o}_q$. While this work adopts a topology model in 2D, the extension to 3D is straight-forward.



Fig. 1. The mobile user moves along roads in a MIMO environment. For each beam, multipath components predominantly arrive from the direction of the mobile user, while paths from other angles exhibit significantly lower probabilities.

Here, we consider a power angular map for a narrowband MIMO communication system as for easy elaboration. Note that a similar methodology may apply to constructing a radio map for other CSI statistics such as the angle of arrival (AoA) and power delay profile.

A. Propagation and Measurement Models

1) *Path Loss Model*: Let $\mathbf{h}_q(\mathbf{x}) = \sqrt{a_q(\mathbf{x})} \bar{\mathbf{h}}_q(\mathbf{x})$ be the narrowband MIMO channel between the user at position $\mathbf{x}$ and the q th BS, where $a_q(\mathbf{x})$ captures the channel gain, and $\bar{\mathbf{h}}_q(\mathbf{x})$ is the normalized channel vector that is assumed to be independent of the channel gain. We adopt a path loss model for the channel gain $a_q(\mathbf{x})$ in its logarithmic value, i.e., $10 \log_{10}(a_q(\mathbf{x}))$, as

$$[a_q(\mathbf{x})]_{\text{dB}} = \beta_q + \alpha_q \log_{10} d(\mathbf{x}, \mathbf{o}_q) + \xi' \quad (1)$$

where β_q is a constant that depends on the propagation environment surrounding the q th BS and the path loss exponent α_q characterizes the rate at which the signal power diminishes with distance. The random variable ξ' models the log-normal shadowing of BS q which is assumed to follow a zero mean Gaussian distribution. We assume independent shadowing, although incorporating shadowing correlation could improve model accuracy, it would require more parameters and increase complexity.

2) *MIMO Pattern Model*: The realization of the normalized channel vector $\bar{\mathbf{h}}_q(\mathbf{x})$ depends on the user location $\mathbf{x}$, the multipath, and the array response, and we model its statistical property via a set of *sensing vectors* $\mathbf{g}_{q,m}$ of the q th BS, where $m = 1, 2, \dots, M$ is the number beams on each BS. Specifically, consider the AoA of the paths is statistically centered around the direct path for each user position, exhibiting small variance under line-of-sight (LOS) conditions and large variance under NLOS conditions. We assume that the sensing vectors $\mathbf{g}_{q,m}$ are designed in such a way that the antenna array at the q th BS has a statistically large response for signals arriving from a reference direction $c_{q,m} \in [0, 2\pi)$, and

the array has a statistically decreasing response for signals increasingly deviating from the reference direction $c_{q,m}$, where the parameters $c_{q,m}$ is to be estimated. Mathematically, consider the transmit-side SSB beamforming gain for each BS, we use an exponential function to model such a pattern

$$10 \log_{10} \mathbb{E}\{|\mathbf{g}_{q,m}^H \bar{\mathbf{h}}_q(\mathbf{x})|^2\} = \omega_{q,m} \exp[-\eta_{q,m}(\phi(\mathbf{x}, \mathbf{o}_q) - c_{q,m})^2] + \xi'' \quad (2)$$

where the expectation is taken over the randomness due to the small-scale fading. The parameter $\omega_{q,m}$ models the beamforming gain at the reference direction $c_{q,m}$, and $\eta_{q,m}$ models the spread of the beam. The function $\phi(\mathbf{x}, \mathbf{o}_q)$ denotes the angle from BS at $\mathbf{o}_q$ to user at $\mathbf{x}$. The variable ξ'' captures the offset and the model mismatch.

A naive example of the sensing vectors $\mathbf{g}_{q,m}$ for a uniform linear array is a set of M -dimensional discrete Fourier transform (DFT) vectors applied to a sub-array of M consecutive antenna elements [31]. It is known that the received beamforming gain is maximized at a certain direction, and the gain decreases if the incident signal arrives from other directions.

The pattern model (2) captures the spatial response of the sensing vector $\mathbf{g}_{q,m}$ according to the AoA distribution of the multipaths arrived at the BS. First, if the direction of the mobile user $\phi(\mathbf{x}, \mathbf{o}_q)$ aligns with the reference direction $c_{q,m}$ of the beam, then a maximum beamforming gain $\omega_{q,m} + \xi''$ is attained. By contrast, if the user direction $\phi(\mathbf{x}, \mathbf{o}_q)$ substantially deviates from the direction $c_{q,m}$, only a small gain ξ'' is attained. The parameter ξ'' provides an offset to the gain, which can be absorbed in the parameter β_q in the path loss model (1). Second, as shown in Fig. 1, the model (2) implicitly assumes that the multipaths mainly arrive from the direction of the mobile user $\phi(\mathbf{x}, \mathbf{o}_q)$, and hence, the $\mathbf{g}_{q,m}$ pointing to $c_{q,m} = \phi(\mathbf{x}, \mathbf{o}_q)$ receives the maximum gain, whereas, the paths arriving from other angles have a substantially smaller probability. This model aligns with the scenario where BSs are placed on a high tower with only a few local scatterers.

3) *Measurement Model*: Motivated by the models (1) and (2), we arrive at our measurement model as follows. Consider a slotted system, $t = 1, 2, \dots, T$, and denote $\mathbf{x}_t \in \mathbb{R}^2$ as the location of the user at time t . For a mobile location $\mathbf{x}_t$, denote $\tilde{y}_{t,q,m} = \mathbb{E}\{|\mathbf{g}_{q,m}^H \mathbf{h}_q(\mathbf{x}_t)|^2\}$ as the average received power from sensing vector $\mathbf{g}_{q,m}$. From (1) and (2), the logarithmic value $y_{q,m,t} = 10 \log_{10} \tilde{y}_{t,q,m}$ is modeled as

$$y_{q,m,t} = \beta_q + \alpha_q \log_{10} d(\mathbf{x}_t, \mathbf{o}_q) + \omega_{q,m} \exp[-\eta_{q,m}(\phi(\mathbf{x}_t, \mathbf{o}_q) - c_{q,m})^2] + \xi_{q,t} \quad (3)$$

where $\xi_{q,t} \sim \mathcal{N}(0, \sigma_q^2)$ models the randomness due to the shadowing and the pattern model mismatch from ξ' and ξ'' in (1) and (2) respectively, ξ'' is absorbed into the constant β_q as a general offset for each to ensure model parsimony, $\xi_{q,t}$ is assumed to be independent across BSs and different time slot t .

Thus, the probability density function (PDF) of $y_{q,m,t}$ given the mobile user location $\mathbf{x}_t$ can be given by

$$p(y_{q,m,t}|\mathbf{x}_t; \boldsymbol{\theta}_{q,m}) = \frac{1}{(2\pi)^{1/2}\sigma_q} \exp\left\{-\frac{1}{2\sigma_q^2} \left[y_{q,m,t} - \beta_q - \alpha_q \times \log_{10} d(\mathbf{x}_t, \mathbf{o}_q) - \omega_{q,m} \exp[-\eta_{q,m}(\phi(\mathbf{x}_t, \mathbf{o}_q) - c_{q,m})^2]\right]^2\right\} \quad (4)$$

where $\boldsymbol{\theta}_{q,m} = \{\alpha_q, \beta_q, \sigma_q, \omega_{q,m}, \eta_{q,m}, c_{q,m}\}$ is a collection of propagation parameters for the m th beam of the q th BS.

Assume that, given the location, measurements across beams and BSs are statistically independent, which is a standard and practically justified simplification. Denote $\mathbf{y}_t$ as the collection of measurements of all BSs q over all sensing vectors m at time slot t . The PDF of $\mathbf{y}_t$ is given by

$$p(\mathbf{y}_t|\mathbf{x}_t; \boldsymbol{\Theta}_p) = \prod_{q=1}^Q \prod_{m=1}^M p(y_{q,m,t}|\mathbf{x}_t) \quad (5)$$

where $\boldsymbol{\Theta}_p = \{\boldsymbol{\theta}_{q,m}\}$ is a collection of propagation parameters for beams m and BSs q .

In a more general case, if we only have sparse observations for a few selected sensing vectors $\mathbf{g}_{q,m}$ indexed by the set $\mathcal{M}_{q,t}$ for the q th BS, and for a few selected BSs denoted by the index set $\mathcal{Q}_t$ at time slot t , the model in (5) can be expressed as

$$p(\mathbf{y}_t|\mathbf{x}_t; \boldsymbol{\Theta}_p) = \prod_{q \in \mathcal{Q}_t} \prod_{m \in \mathcal{M}_{q,t}} p(y_{q,m,t}|\mathbf{x}_t). \quad (6)$$

It is observed that the formulation is not affected by the situation of missing data.

B. Mobility Model

We adopt the Gauss-Markov model [32], [33] for the dynamics of user mobility $\mathbf{x}_t$. Denote δ as the slot duration. The movement at the t th time slot is modeled as:

$$\mathbf{x}_t - \mathbf{x}_{t-1} = \gamma(\mathbf{x}_{t-1} - \mathbf{x}_{t-2}) + (1 - \gamma)\delta\bar{\mathbf{v}} + \sqrt{1 - \gamma^2}\delta\epsilon. \quad (7)$$

Here, the velocity $(\mathbf{x}_t - \mathbf{x}_{t-1})/\delta$ at time slot t depends on the velocity from the previous time slot following an auto-regressive model with parameter $0 < \gamma \leq 1$ and randomness $\epsilon \sim \mathcal{N}(\mathbf{0}, \sigma_v^2 \mathbf{I})$. This is to capture the fact that acceleration is bounded in practice. The parameter $\bar{\mathbf{v}}$ models the average velocity. A higher γ value indicates a stronger correlation between consecutive velocities, resulting in smoother movement. When $\gamma = 1$, the mobile user maintains a constant velocity.

C. An HMM Formulation

Denote $\mathcal{X}_t = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_t)$ and $\mathcal{Y}_t = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_t)$ as the trajectory of the mobile user and the accumulated measurements up to time t , respectively. The goal here is to estimate the trajectory $\mathcal{X}_t$ based on the measurements $\mathcal{Y}_t$.

Based on the Gauss-Markov model (7), the PDF of the location $\mathbf{x}_t$ at time slot t can be written as

$p(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{x}_{t-2}; \boldsymbol{\Theta}_m)$, where $\boldsymbol{\Theta}_m = \{\bar{\mathbf{v}}, \sigma_v^2\}$ is the collection of mobility parameters. The Bayes' rule of probability suggests that

$$\begin{aligned} p(\mathcal{Y}_t, \mathcal{X}_t) &= p(\mathbf{y}_t|\mathcal{Y}_{t-1}, \mathbf{x}_t, \mathcal{X}_{t-1})p(\mathcal{Y}_{t-1}, \mathbf{x}_t, \mathcal{X}_{t-1}) \\ &= p(\mathbf{y}_t|\mathbf{x}_t)p(\mathcal{Y}_{t-1}, \mathbf{x}_t, \mathcal{X}_{t-1}) \\ &= p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{x}_{t-2})p(\mathcal{Y}_{t-1}, \mathcal{X}_{t-1}), \end{aligned} \quad (8)$$

which is due to the fact that $\mathbf{y}_t$ is independent of $\mathcal{Y}_{t-1}$ and $\mathcal{X}_{t-1}$ given $\mathbf{x}_t$, and $\mathbf{x}_t$ is independent of $\mathcal{Y}_{t-1}$ given $\mathbf{x}_{t-1}$ and $\mathbf{x}_{t-2}$.

Recursively applying (8), one arrives at

$$p(\mathcal{Y}_T, \mathcal{X}_T) = \prod_{t=1}^T p(\mathbf{y}_t|\mathbf{x}_t) \prod_{t=3}^T p(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{x}_{t-2}).$$

Consider to maximize the log-likelihood $\log p(\mathcal{Y}_T, \mathcal{X}_T)$. We arrive at the following problem

$$\underset{\mathcal{X}_T, \boldsymbol{\Theta}_p, \boldsymbol{\Theta}_m}{\text{maximize}} \sum_{t=1}^T \log p(\mathbf{y}_t|\mathbf{x}_t) + \sum_{t=3}^T \log p(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{x}_{t-2}) \quad (9)$$

which jointly fits the parameters $\boldsymbol{\Theta}_p$ of the observation model (5) and the parameters $\boldsymbol{\Theta}_m$ of the mobility model to the data $\mathcal{Y}_T$ for the recovery of the trajectory $\mathcal{X}_T$. Due to the assumption on independent measurements, the models for multiple users are identical. Thus, it is straight-forward to apply the algorithm to the case of multiple users and multiple trajectories.

III. FUNDAMENTAL LIMITS

We study the fundamental limit on recovering the trajectory $\mathcal{X}_T$ from the power measurements under the most challenging case $N_t = 1$ with a single antenna at each BS, where we allow to scale the number of BSs to infinity. Increasing N_t enables the extraction of richer spatial features, thereby improving estimation accuracy and trajectory identifiability. Specifically, we will examine the CRLB for estimating $\mathbf{x}_t$. It is known that the CRLB provides a lower bound on the mean squared error (MSE) of any unbiased estimator of a parameter. Under certain regularity conditions, the maximum likelihood estimation (MLE) asymptotically achieves the CRLB as the number of measurements increases to infinity [34]. Thus, our goal is to understand the best possible performance for recovering the trajectory $\mathbf{x}_t$ and identify the critical parameters that affect the performance.

To understand the performance upper bound, consider the case of $\gamma = 1$ in the mobility model (7), where the Markovian mobility model degenerates to a chain of states with transition probability 1, which corresponds to the user moving at a constant speed. Consequently, the mobility $\mathbf{x}_t$ reduces to a deterministic trajectory:

$$\mathbf{x}_t = \mathbf{x} + t\mathbf{v} \quad (10)$$

which can be fully determined by the parameters $(\mathbf{x}, \mathbf{v}) \in \mathbb{R}^4$, i.e., the starting position $\mathbf{x}$ and the velocity $\mathbf{v}$. It naturally follows that decreasing γ will increase the randomness of $\mathbf{x}_t$, and more information is needed to determine $\mathbf{x}_t$. In the limiting case where $\gamma = 0$, the trajectory recovery problem

degenerates to a series of conventional Bayesian estimations for the positions $\mathbf{x}_t$ based on RSS measurements. Hence, the case $\gamma = 1$ corresponds to the best possible performance we can obtain for the RSS-based trajectory recovery problem.

Based on the constant speed mobility model (10), the function $d(\mathbf{x}_t, \mathbf{o}_q)$ for the distance between the mobile location $\mathbf{x}_t$ and the q th BS location $\mathbf{o}_q$ is simplified as $d_{t,q}(\mathbf{x}, \mathbf{v}) \triangleq \|\mathbf{l}_q(\mathbf{x}) + t\mathbf{v}\|_2$, where $\mathbf{l}_q(\mathbf{x}) = \mathbf{x} - \mathbf{o}_q$ is the direction from the q th BS to the initial position $\mathbf{x}$ of the trajectory. From the observation model (4) under $N_t = 1$ and the mobility model (7) under $\gamma = 1$, the log-likelihood function (9) becomes

$$f(\phi, \psi) = \sum_{t=1}^T \sum_{q=1}^Q \left[-\ln 2\pi\sigma_q^2 - \frac{1}{2\sigma_q^2} (y_{t,q} - \beta_q - \alpha_q \log d_{t,q}(\mathbf{x}, \mathbf{v}))^2 \right] \quad (11)$$

where $\phi = \{\alpha_q, \beta_q\}$ is the propagation parameters and $\psi = (\mathbf{x}, \mathbf{v})$ is the mobility parameters, and the term $\log p(\mathbf{x}_t | \mathbf{x}_{t-1}, \mathbf{x}_{t-2})$ in (9) disappears under constant speed mobility.

From the log-likelihood function (11), it is not surprising that once the mobility parameter ψ is available, the CRLB of the estimator for the propagation parameter ϕ approaches zero as $T \rightarrow \infty$. This is because as T approaches infinity, the number of independent measurements under the observation model (3) approaches infinity, and since equation (3) establishes a linear model on the propagation parameter ϕ with Gaussian noise, the CRLB of estimating ϕ approaches zero [34].

Hence, the focus here is to understand the fundamental limit of estimating the mobility parameter $\psi = (\mathbf{x}, \mathbf{v})$.

A. The Fisher Information Matrix

The Fisher information matrix (FIM) $\mathbf{F}_{T,\psi}$ of $\psi = (\mathbf{x}, \mathbf{v}) \in \mathbb{R}^4$ from the measurements over a duration T can be computed as

$$\begin{aligned} \mathbf{F}_{T,\psi} &\triangleq \mathbb{E}\{-\nabla_{\psi}^2 f(\phi, \psi)\} \\ &= \sum_{t,q} \frac{\alpha_q^2}{\sigma_q^2 d_{t,q}^2(\mathbf{x}, \mathbf{v})} \nabla_{\psi} d_{t,q}(\mathbf{x}, \mathbf{v}) (\nabla_{\psi} d_{t,q}(\mathbf{x}, \mathbf{v}))^T \end{aligned}$$

where the derivative $\nabla_{\psi} d_{t,q}(\mathbf{x}, \mathbf{v})$ is derived as

$$\nabla_{\psi} d_{t,q}(\mathbf{x}, \mathbf{v}) = \begin{bmatrix} \frac{\partial}{\partial \mathbf{x}} \|\mathbf{l}_q(\mathbf{x}) + t\mathbf{v}\|_2 \\ \frac{\partial}{\partial \mathbf{v}} \|\mathbf{l}_q(\mathbf{x}) + t\mathbf{v}\|_2 \end{bmatrix} = \frac{\mathbf{l}_q(\mathbf{x}) + t\mathbf{v}}{d_{t,q}(\mathbf{x}, \mathbf{v})} \begin{bmatrix} 1 \\ t \end{bmatrix}^T.$$

Thus, the FIM can be expressed as

$$\begin{aligned} \mathbf{F}_{T,\psi} &= \sum_{t,q} \frac{\alpha_q^2}{\sigma_q^2 d_{t,q}^4(\mathbf{x}, \mathbf{v})} \begin{bmatrix} 1 & t \\ t & t^2 \end{bmatrix} \\ &\quad \otimes ((\mathbf{l}_q(\mathbf{x}) + t\mathbf{v})(\mathbf{l}_q(\mathbf{x}) + t\mathbf{v})^T) \end{aligned} \quad (12)$$

in which, $\otimes$ is the Kronecker product.

For an unbiased estimator $\hat{\psi}$, the MSE is lower bounded by $\mathbb{E}\{\|\hat{\psi} - \psi\|^2\} \geq \text{tr}\{\mathbf{F}_{T,\psi}^{-1}\}$, where $\text{tr}\{\mathbf{F}_{T,\psi}^{-1}\}$ is the CRLB of

estimating $\psi = (\mathbf{x}, \mathbf{v})$. Similarly, we define the FIMs $\mathbf{F}_{T,x} = \mathbb{E}\{-\nabla_{\mathbf{x}\mathbf{x}}^2 f(\phi, \psi)\}$ and $\mathbf{F}_{T,v} = \mathbb{E}\{-\nabla_{\mathbf{v}\mathbf{v}}^2 f(\phi, \psi)\}$, which are the diagonal blocks of $\mathbf{F}_{T,\psi}$ and are associated with the CRLB $B(\mathbf{x}) = \text{tr}\{\mathbf{F}_{T,x}^{-1}\}$ and CRLB $B(\mathbf{v}) = \text{tr}\{\mathbf{F}_{T,v}^{-1}\}$ for the parameters $\mathbf{x}$ and $\mathbf{v}$, respectively.

B. BS Deployed in a Limited Region

We first investigate the case where the BSs are deployed in a limited region, but the measurement trajectory is allowed to go unbounded as T goes to infinity. Signals can always be collected by the BSs regardless of the distance. As a result, an infinite amount of measurements can be collected as $T \rightarrow \infty$.

It is observed that $\mathbf{F}_{T,\psi} \prec \mathbf{F}_{T+1,\psi}$, indicating that the Fisher information is strictly increasing. This is because each term in the summation in (12) is positive definite, provided that $\mathbf{l}_q(\mathbf{x})$ and $\mathbf{v}$ are linear independent for at least one q .

However, it is somewhat surprising that the CRLB for $\mathbf{x}$ and $\mathbf{v}$ does not decrease to zero as $T \rightarrow \infty$, despite the infinitely increasing amount of independent data.

Specifically, assume that the trajectory $\mathbf{x}_t$ does not pass any of the BS location $\mathbf{o}_q$, and hence, $d_{\min,q} = \min_t \{d_{t,q}(\mathbf{x}, \mathbf{v})\} > 0$ for all q . Define $\alpha_{\max}^2 = \max_q \{\alpha_q^2\}$, $\sigma_{\min}^2 = \min_q \{\sigma_q^2\}$.

Theorem 1: The CRLB of $\mathbf{x}$ satisfies $B(\mathbf{x}) = \text{tr}\{\mathbf{F}_{T,x}^{-1}\} \geq \bar{\Delta}_{T,x}$, where equality can be achieved when $\sigma_{\min}^2 = \sigma_q^2$ and $\alpha_{\max}^2 = \alpha_q^2$ for all q . In addition, $\bar{\Delta}_{T,x}$ is strictly decreasing in T , provided that at least two vectors in $\{\mathbf{l}_1(\mathbf{x}), \mathbf{l}_2(\mathbf{x}), \dots, \mathbf{l}_Q(\mathbf{x}), \mathbf{v}\}$ are linear independent, but $\bar{\Delta}_{T,x}$ converges to a strictly positive number as $T \rightarrow \infty$.

Proof: See Appendix A. $\square$

Theorem 1 suggests that the CRLB of $\mathbf{x}$ cannot decrease to zero even when we estimate only two parameters for the initial location $\mathbf{x} \in \mathbb{R}^2$ based on *infinite* measurement samples collected over an infinite geographical horizon as $T \rightarrow \infty$.

Through the development of the proof, a physical interpretation of Theorem 1 can be given as follows. As T increases, the distances $d_{t,q}(\mathbf{x}, \mathbf{v}) = \|\mathbf{x}_t - \mathbf{o}_q\|_2$ grow larger because the user moves away from the BSs. For a position $\mathbf{x}_t$ at a sufficiently large distance, the term $\mathbf{x}_t - \mathbf{o}_q$ approximates to $\mathbf{x}_t$ since $\|\mathbf{x}_t\| \gg \|\mathbf{o}_q\|$. Consequently, all measurements point to approximately the same direction independent of the BS locations $\mathbf{o}_q$, i.e., the incremental information provided by each new measurement diminishes rapidly. Therefore, although the FIM $\mathbf{F}_{T,x}$ strictly increases with T , the increment $\mathbf{F}_{T+1,x} - \mathbf{F}_{T,x}$ decreases quickly. As a result, the CRLBs strictly decreases but only approaches a non-zero lower bound, similar to how the series $\sum_{t=1}^{\infty} 1/t^r$ converges for $r > 1$.

While it is not possible to perfectly estimate the starting point $\mathbf{x}$ under $T = \infty$ with a finite number of BSs, one might expect that estimating the velocity $\mathbf{v}$ could be more promising. However, we have the following result.

Theorem 2: The CRLB of $\mathbf{v}$ satisfies $B(\mathbf{v}) = \text{tr}\{\mathbf{F}_{T,v}^{-1}\} \geq \bar{\Delta}_{T,v}$ with equality achieved when $\sigma_{\min}^2 = \sigma_q^2$, $\alpha_{\max}^2 = \alpha_q^2$ for all q . In addition,

$$\bar{\Delta}_{T,v} \rightarrow C_v = \left(\frac{\alpha_{\max}^2}{\sigma_{\min}^2} \sum_{q=1}^Q s_{\infty,q}^{(2)} \|\mathbf{P}_v^\perp \mathbf{l}_q(\mathbf{x})\|^2 \right)^{-1}$$

as $T \rightarrow \infty$, where

$$s_{\infty,q}^{(2)} = \lim_{T \rightarrow \infty} \sum_{t=1}^T \frac{t^2}{d_{t,q}^4(\mathbf{x}, \mathbf{v})}, \quad \mathbf{P}_v^\perp = \mathbf{I} - \mathbf{v}\mathbf{v}^T / \|\mathbf{v}\|^2$$

in which, the parameter $s_{\infty,q}^{(2)}$ is upper bounded by $1/\rho^4 \lim_{T \rightarrow \infty} \sum_{t=1}^T 1/t^2 \approx \pi^2/(6\rho^4)$, where $\rho > 0$ is sufficiently small such that $d_{t,q}(\mathbf{x}, \mathbf{v}) > \rho t$ for all $t \geq 1$.

Proof: See Appendix B. $\square$

While traditional work emphasizes BS geometric diversity for single-point localization, Theorems 1 and 2 suggest that, under a finite number of BSs in a limited region, neither the initial position $\mathbf{x}$ nor the velocity $\mathbf{v}$ can be perfectly estimated by merely increasing the observation time T .

Theorem 2 implies that the fundamental limit to the estimation accuracy is affected by the spatial distribution of the BSs and the nature of the RSS measurements. Specifically, the non-diminishing error lower bound C_v in estimating the velocity $\mathbf{v} \in \mathbb{R}^2$ is inversely affected by the norm of $\mathbf{P}_v^\perp \mathbf{l}_q(\mathbf{x})$, i.e., the vector $\mathbf{l}_q(\mathbf{x}) = \mathbf{x} - \mathbf{o}_q$ projected on the orthogonal direction of $\mathbf{v}$. The more spread of the BSs in the orthogonal direction of the moving direction $\mathbf{v}$, the lower the estimation error of $\mathbf{v}$.

Another surprising observation that differentiates the trajectory recovery problem (9) from a conventional RSS-based localization problem is that a large path loss exponent $|\alpha_q|$ is expected to decrease the CRLB for recovering the trajectory. Recall that in the empirical path loss model, $\alpha_q = -20$ corresponds to free space propagation, and $\alpha_q < -20$ usually corresponds to propagations in the NLOS scenarios. Hence, for the same shadowing deviation σ , a “deep fade” that empirically leads to a large $|\alpha_q|$ seems to be preferred for a small CRLB in trajectory recovery, or more rigorously, a large ratio $|\alpha_q/\sigma|$ is preferred. This phenomenon can be interpreted by the property that a large $|\alpha_q|$ can better differentiate the movement distance along $\mathbf{x}_t$ based on the RSS measurements in the model (3).

C. BS Deployed in an Unlimited Region

We now study the case where the BSs follow a PPP with density κ in an unlimited region, but the users can only connect with a subset of BSs within a radius of R . As a result, the number of connected BSs is still finite. We investigate the CRLBs as the user trajectory goes unbounded as T goes to infinity.

Specifically, on average, measurements from $\bar{Q} = \kappa\pi R^2$ BSs can be collected in each time slot. It turns out that, in such a scenario, although we still have a limited number of active BSs, the estimation lower bound now can decrease to zero. Denote $\alpha_{\min}^2 = \min_q \{\alpha_q^2\}$, $\sigma_{\max}^2 = \max_q \{\sigma_q^2\}$.

Theorem 3: Assume that the minimum distance to the nearest BS is greater than r_0 along the trajectory.¹ The CRLB of $\mathbf{x}$ satisfies $B(\mathbf{x}) = \text{tr}\{\mathbf{F}_{T,x}^{-1}\} \leq \tilde{\Delta}_{T,x}$ and as $T \rightarrow \infty$

$$T\tilde{\Delta}_{T,x} \rightarrow \frac{2\sigma_{\max}^2}{\alpha_{\min}^2 \kappa\pi \ln(R/r_0)}.$$

¹In practice, the parameter r_0 can be understood as the height of the antenna. More rigorously, we should employ a 3D model to compute the distance $d_{t,q}$, but the asymptotic result would be the same.

Proof: See Appendix C in [35]. $\square$

Theorem 4: Assume that the minimum distance to the nearest BS is greater than r_0 along the trajectory. The CRLB of $\mathbf{v}$ satisfies $B(\mathbf{v}) = \text{tr}\{\mathbf{F}_{T,v}^{-1}\} \leq \tilde{\Delta}_{T,v}$ and as $T \rightarrow \infty$

$$T(T+1)(2T+1)\tilde{\Delta}_{T,v} \rightarrow \frac{12\sigma_{\max}^2}{\alpha_{\min}^2 \kappa\pi \ln(R/r_0)}.$$

Proof: See Appendix D in [35]. $\square$

It is observed from the above theorems that the CRLB of $\mathbf{x}$ decreases as $\mathcal{O}(1/T)$ and the CRLB of $\mathbf{v}$ decreases as $\mathcal{O}(1/T^3)$. Estimating the velocity $\mathbf{v}$ is significantly easier than estimating the initial location $\mathbf{x}$. Furthermore, a longer measurement range R enhances the accuracy of the estimation. Additionally, a higher density of BSs κ within the radius also leads to improved estimation performance.

To summarize, recall that $1/\sigma_{\max}^2$ is proportional to the SNR. Our results show that it is possible to perfectly recover a full trajectory from RSS measurements without any location labels under all SNR conditions. However, a larger $\sigma_{\max}^2$ leads to a slower decrease in the CRLB of $\mathbf{x}$ and $\mathbf{v}$ as T increases.

IV. ALGORITHM DESIGN

To solve the joint trajectory recovery and parameter estimation problem (9), it is observed that given $\mathcal{X}_T$, the variables Θ_p and Θ_m are decoupled, because the first term in (9) only depends on Θ_p and the second term in (9) only depends on Θ_m . As a result, Θ_p and Θ_m can be solved by two parallel subproblems from (9) as follows

$$\begin{aligned} \text{(P1)} : & \underset{\Theta_m}{\text{maximize}} \quad \sum_{t=3}^T \log p(\mathbf{x}_t | \mathbf{x}_{t-1}, \mathbf{x}_{t-2}; \Theta_m), \\ \text{(P2)} : & \underset{\Theta_p}{\text{maximize}} \quad \sum_{t=1}^T \log p(\mathbf{y}_t | \mathbf{x}_t; \Theta_p). \end{aligned}$$

On the other hand, given the variables $\hat{\Theta}_p$ and $\hat{\Theta}_m$ as the solutions to (P1) and (P2), respectively, $\mathcal{X}_T$ can be solved by

$$\begin{aligned} \text{(P3)} : & \underset{\mathcal{X}_T}{\text{maximize}} \quad \sum_{t=1}^T \log p(\mathbf{y}_t | \mathbf{x}_t; \hat{\Theta}_p) \\ & + \sum_{t=3}^T \log p(\mathbf{x}_t | \mathbf{x}_{t-1}, \mathbf{x}_{t-2}; \hat{\Theta}_m). \end{aligned}$$

This naturally leads to an alternating optimization strategy, which solves for $\mathcal{X}_T$ from problem (P3), and then for Θ_p and Θ_m from (P1) and (P2), in an iterative manner. Since the corresponding iterations never decrease the objective (9) which is bounded above, the iteration must converge.

The solutions to these subproblems are derived as follows.

A. Solution to (P1) for the Mobility Model

According to the mobility model in (7), the PDF of the location $\mathbf{x}_t$ at time slot t is given by

$$p(\mathbf{x}_t | \mathbf{x}_{t-1}, \mathbf{x}_{t-2}; \Theta_m) = \frac{1}{2\pi\sigma_v \sqrt{1 - \gamma^2}}$$

$$\times \exp \left\{ -\frac{\|\mathbf{x}_t - (1 + \gamma)\mathbf{x}_{t-1} + \gamma\mathbf{x}_{t-2} - (1 - \gamma)\delta\bar{\mathbf{v}}\|_2^2}{2(1 - \gamma^2)\delta^2\sigma_v^2} \right\} \quad (13)$$

for $\gamma \neq 1$.

From (13), the objective function of (P1) can be expressed as

$$-\sum_{t=3}^T \log(2\pi\sigma_v\sqrt{1-\gamma^2}) \quad (14)$$

$$-\sum_{t=3}^T \frac{\|\mathbf{x}_t - (1 + \gamma)\mathbf{x}_{t-1} + \gamma\mathbf{x}_{t-2} - (1 - \gamma)\delta\bar{\mathbf{v}}\|_2^2}{2(1 - \gamma^2)\delta^2\sigma_v^2}.$$

Setting the derivative of (14) with respect to (w.r.t.) $(\bar{\mathbf{v}}, \sigma_v^2)$ to zero, we find that the corresponding solution

$$\bar{\mathbf{v}} = \frac{\sum_{t=3}^T (\mathbf{x}_t - (1 + \gamma)\mathbf{x}_{t-1} + \gamma\mathbf{x}_{t-2})}{(T - 2)(1 - \gamma)\delta} \quad (15)$$

$$\sigma_v^2 = \frac{\sum_{t=3}^T \|\mathbf{x}_t - (1 + \gamma)\mathbf{x}_{t-1} + \gamma\mathbf{x}_{t-2} - (1 - \gamma)\delta\bar{\mathbf{v}}\|_2^2}{2(T - 2)\delta^2} \quad (16)$$

is unique. Since (P1) is an unconstrained optimization problem, (15)–(16) give the optimal solution to (P1).

B. Solution to (P2) via Separable Regression With Log-Transformation

Solving (P2) is very challenging because the regression model (4) and (5) for (P2) is highly non-linear containing both polynomial terms and exponential terms. Therefore, it is important to investigate whether (P2) can be separated into easier subproblems.

It is observed that there are two groups of parameters, where one group of parameters $\{\alpha_q, \beta_q, \sigma_q\}$ describe the path loss model for each BS q , and the other group $\{w_{q,m}, \eta_{q,m}, c_{q,m}\}$ describe the spatial pattern for each index m of each BS q .

1) *Separability*: We first investigate the separability of the two groups of parameters.

Denote

$$y'_{q,m,t} = y_{q,m,t} - w_{q,m} \exp[-\eta_{q,m}(\phi(\mathbf{x}_t, \mathbf{o}_q) - c_{q,m})^2]. \quad (17)$$

Using (4)–(5) and given the group of pattern parameters $\{w_{q,m}, \eta_{q,m}, c_{q,m}\}$, problem (P2) can be written into Q parallel subproblems for $q = 1, 2, \dots, Q$,

$$\begin{aligned} &\underset{\alpha_q, \beta_q, \sigma_q}{\text{minimize}} \quad \frac{1}{2\sigma_q^2} \sum_{m,t} (y'_{q,m,t} - \alpha_q \tilde{d}_{q,t} - \beta_q)^2 \\ &\quad + \frac{QMT}{2} \ln(2\pi\sigma_q^2) \end{aligned} \quad (18)$$

where $\tilde{d}_{q,t} = \log_{10} d(\mathbf{x}_t, \mathbf{o}_q)$ is the log-distance between the q th BS and the user at $\mathbf{x}_t$. It follows that the solutions to (18) depends on the pattern parameters $\{w_{q,m}, \eta_{q,m}, c_{q,m}\}$ via the variables $y'_{q,m,t}$ in (17). Denote the solution to (18) as $\alpha_q^{(1)}$, $\beta_q^{(1)}$, and $\sigma_q^{(1)}$.

Consider an *auxiliary* problem that does not explicitly depends on the pattern parameters $\{w_{q,m}, \eta_{q,m}, c_{q,m}\}$

$$\underset{\alpha_q, \beta_q, \sigma_q}{\text{minimize}} \quad \frac{1}{2\sigma_q^2} \sum_t (\bar{y}_{q,t} - \alpha_q \tilde{d}_{q,t} - \beta_q)^2 + \frac{QMT}{2} \ln(2\pi\sigma_q^2) \quad (19)$$

where $\bar{y}_{q,t} = \sum_m y_{q,m,t}$, which only depends on the aggregated values from $y_{q,m,t}$ regardless of the parameters $\{w_{q,m}, \eta_{q,m}, c_{q,m}\}$ under some conditions to be specified later. This is because while different beams have different energy distribution in space, it is possible that the aggregate energy over all beams can be uniform over the coverage area of a BS. Denote the corresponding solution as $\alpha_q^{(2)}$, $\beta_q^{(2)}$, and $\sigma_q^{(2)}$.

We find the conditions when the path loss parameters estimated from the separated problem (19) are identical to the original problem (18).

Proposition 1: [Separability] Suppose that the measurement model (3) satisfies

$$\sum_{m=1}^M w_{q,m} \exp[-\eta_{q,m}(\phi(\mathbf{x}_t, \mathbf{o}_q) - c_{q,m})^2] = \bar{C}_q \quad (20)$$

for all $\mathbf{x}_t$ and some constant $\bar{C}_q$. Then, the solutions to the path loss parameters satisfy $\alpha_q^{(1)} = \alpha_q^{(2)}$ and $\beta_q^{(1)} = \beta_q^{(2)} - \bar{C}_q$.

Proof: See Appendix E in [35]. $\square$

The condition (20) requires that the aggregated beamforming gain over all beams is identical for all locations $\mathbf{x}_t$. This implicitly requires that the BS distributes the beams uniformly within the coverage range such that the aggregated beamforming gain is uniform at all possible directions.

Proposition 1 delivers two important messages. First, while the condition (20) may be challenging to be met in practice, Proposition 1 suggests that distributing the beams as uniform as possible can simplify the parameter estimation, since the estimation problem (P2) can be approximately decomposed into a subproblem of estimating the path loss parameters separately. Second, when there is beamforming gain, solving the separated path loss estimation problem (19) tends to over estimate the path gain by $\bar{C}_q$, which equals to the aggregated beamforming gain, but the estimation of the path loss exponent α_q is not affected and accurate.

Motivated by Proposition 1, one can easily initialize the path loss parameters by solving the separated problem (19), and then, one fine tunes the estimate by iteratively estimating the pattern parameters and the path loss parameters, leading to the proposed strategy described as follows.

2) *Path Loss Parameters*: To solve problem (19), denote $\bar{\mathbf{y}}_q \in \mathbb{R}^T$ as the collection of the variables $\bar{y}_{q,t}$ for the q th BS along the trajectory $\mathbf{x}_t$, $\mathbf{d}_q \in \mathbb{R}^T$ as the collection of all the log-distances $\tilde{d}_{q,t}$, $\mathbf{D}_q = [\mathbf{d}_q, \mathbf{1}] \in \mathbb{R}^{T \times 2}$ where $\mathbf{1}$ is a all-one vector with T elements, and $\boldsymbol{\theta}_q = [\alpha_q, \beta_q]^T$. The first term of (19) can be written into the matrix form as $\frac{1}{2\sigma_q^2} \|\bar{\mathbf{y}}_q - \mathbf{D}_q \boldsymbol{\theta}_q\|^2$, and the second term of (19) does not depend on $\boldsymbol{\theta}_q$. Such a least-squares problem (19) has the solution in the matrix form

$$\hat{\boldsymbol{\theta}}_q = (\mathbf{D}_q^T \mathbf{D}_q)^{-1} \mathbf{D}_q^T \bar{\mathbf{y}}_q. \quad (21)$$

The solution to problem (18) can be obtained in a similar way and is found as

$$\hat{\theta}_q = (\tilde{\mathbf{D}}_q^T \tilde{\mathbf{D}}_q)^{-1} \tilde{\mathbf{D}}_q^T \mathbf{y}'_q \quad (22)$$

where $\mathbf{y}'_q \in \mathbb{R}^{MT}$ is the collection of variable $y'_{q,m,t}$ and $\tilde{\mathbf{D}}_q = \mathbf{D}_q \otimes \mathbf{1}$, in which, $\mathbf{1}$ is an all-one vector with M elements.

The solution to σ_q can be found as setting the derivative of the objective function in problem (18) to zero, and is found as

$$\hat{\sigma}_q = \sqrt{\frac{1}{MT} \sum_{m,t} (y'_{q,m,t} - \alpha_q d_{q,t} - \beta_q)^2}. \quad (23)$$

3) *Pattern Parameters via Log-transformation:* To estimate the pattern parameters $\{w_{q,m}, \eta_{q,m}, c_{q,m}\}$, denote $y''_{q,m,t} = y_{q,m,t} - \alpha_q d_{q,t} - \beta_q$. Given the path loss parameters $\{\alpha_q, \beta_q, \sigma_q\}$, problem (P2) can be equivalently written as QM parallel subproblems for $q = 1, 2, \dots, Q$ and $m = 1, 2, \dots, M$,

$$\begin{aligned} & \underset{w_{q,m}, \eta_{q,m}, c_{q,m}}{\text{minimize}} \quad \sum_t \left(y''_{q,m,t} - w_{q,m} \right. \\ & \quad \left. \exp[-\eta_{q,m}(\phi_{q,t} - c_{q,m})^2] \right)^2 \end{aligned} \quad (24)$$

where $\phi_{q,t} = \phi(\mathbf{x}_t, \mathbf{o}_q)$ captures the direction from the q th BS to the user at $\mathbf{x}_t$.

It is observed that even by removing the path loss components, the regression problem (24) is still difficult to solve as the regression model contains both exponential and polynomial terms. Here, we propose to linearize the problem via log-transformation.

Note that the regression model $w_{q,m} \exp[-\eta_{q,m}(\phi_{q,t} - c_{q,m})^2]$ in (24) is positive, but it is possible that $y''_{q,m,t}$ is negative due to the estimation error from the path loss parameters $\{\alpha_q, \beta_q, \sigma_q\}$. In addition, Proposition 1 suggests that the path gain is overestimated by solving the separated subproblem (19), i.e., $\beta^{(2)} = \beta_q^{(1)} + \bar{C}_q > \beta_q^{(1)}$, resulting in a negative bias for all $y''_{q,m,t}$. As a result, it is more reliable to focus on the set of data $\mathcal{T}_{q,m}^\epsilon = \{t: y''_{q,m,t} > \epsilon\}$ for the estimation of the pattern parameters $\{w_{q,m}, \eta_{q,m}, c_{q,m}\}$, where $y''_{q,m,t}$ is greater than some positive threshold ϵ . A positive threshold ϵ is used to exclude noisy, low-power measurements and ensure robust parameter estimation from reliable, high-SNR data.

We perform nonlinear regression estimation using a linear approximation [36]. Specifically, for $t \in \mathcal{T}_{q,m}^\epsilon$, we take the natural logarithm on both sides of the model, resulting in the following auxiliary weighted linear regression problem:

$$\begin{aligned} & \underset{w_{q,m}, \eta_{q,m}, c_{q,m}}{\text{minimize}} \quad \sum_{t \in \mathcal{T}_{q,m}^\epsilon} \lambda_t \left(\ln y''_{q,m,t} - \ln(w_{q,m} \right. \\ & \quad \left. \times \exp[-\eta_{q,m}(\phi_{q,t} - c_{q,m})^2]) \right)^2 \end{aligned} \quad (25)$$

where $\lambda_t > 0$ is a weighting factor introduced to ensure equivalence with the original pattern estimation problem (24).

Observing that $\ln(w_{q,m} \exp[-\eta_{q,m}(\phi_{q,t} - c_{q,m})^2]) = \ln w_{q,m} - \eta_{q,m}(\phi_{q,t} - c_{q,m})^2$, denote a set of variables

$b_1 = -\eta_{q,m}$, $b_2 = 2\eta_{q,m}c_{q,m}$, and $b_3 = \ln w_{q,m} - \eta_{q,m}c_{q,m}^2$. The pattern estimation problem (25) can be rewritten as

$$\underset{\mathbf{b}}{\text{minimize}} \quad \sum_{t \in \mathcal{T}_{q,m}^\epsilon} \lambda_t (\ln y''_{q,m,t} - (\phi_{q,t}^2 b_1 + \phi_{q,t} b_2 + b_3))^2 \quad (26)$$

and we have the following result.

Proposition 2: [Equivalence Condition] Suppose that $\mathcal{T}_{q,m}^\epsilon = \{1, 2, \dots, T\}$. Denote $B(\boldsymbol{\vartheta}; \phi_{q,t}) = w_{q,m} \exp[-\eta_{q,m}(\phi_{q,t} - c_{q,m})^2]$, where $\boldsymbol{\vartheta} = (w_{q,m}, \eta_{q,m}, c_{q,m})$. If the weights λ_t satisfy

$$\lambda_t = \frac{y''_{q,m,t} - B(\boldsymbol{\vartheta}; \phi_{q,t})}{\ln y''_{q,m,t} - \ln B(\boldsymbol{\vartheta}; \phi_{q,t})} B(\boldsymbol{\vartheta}; \phi_{q,t}) \quad (27)$$

then, the auxiliary problem (26) is equivalent to the pattern estimation problem (24). Specifically, if $\mathbf{b}$ is the solution to (26), then

$$w_{q,m} = \exp\left(b_3 - \frac{b_2^2}{4b_1}\right), \eta_{q,m} = -b_1, c_{q,m} = -\frac{b_2}{2b_1} \quad (28)$$

is the solution to (24).

Proof: See Appendix F in [35]. $\square$

Proposition 2 imposes a stronger, sufficient (but not necessary) condition by requiring the subset $\mathcal{T}_{q,m}^\epsilon$ to include all samples; while this may not always hold initially, the equivalence becomes a good approximation as optimization progresses and parameter estimates improve, especially under high SNR or near the optimum.

It is observed that a small value $y''_{q,m,t} \ll 1$ tends to receive a small weight λ_t , because the term $\ln y''_{q,m,t}$ in the denominator of (27) has a large magnitude. As a result, the auxiliary problem tends to focus more on the data with a large $y''_{q,m,t}$, i.e., the center of the beam, which aligns with the goal of (P2).

Proposition 2 provides a convenience way to solve the pattern estimation problem (24), because the auxiliary problem (26) is a weighted least-squares linear regression problem, and the solution is given by

$$\mathbf{b} = \left(\sum_{t \in \mathcal{T}_{q,m}^\epsilon} \lambda_t \phi_{q,t} \phi_{q,t}^T \right)^{-1} \left(\sum_{t \in \mathcal{T}_{q,m}^\epsilon} (\lambda_t \ln y''_{q,m,t}) \phi_{q,t} \right) \quad (29)$$

which is obtained by setting the derivative of (26) w.r.t. $\mathbf{b}$ to 0, where $\phi_{q,t} = [\phi_{q,t}^2, \phi_{q,t}, 1]^T$. In addition, the solution $\boldsymbol{\vartheta} = (w_{q,m}, \eta_{q,m}, c_{q,m})$ is obtained from (28).

Finally, as λ_t depends on the solution $\boldsymbol{\vartheta}$, an iterative approach can be applied. At first, all weights are initialized to 1, and solution of $\boldsymbol{\vartheta}$ is obtained from (28) and (29). Then, λ_t is updated according to (27), followed by an update of $\boldsymbol{\vartheta}$ from (28) and (29), and iteration goes on.

The overall algorithm is summarized in Algorithm 1.

C. Solution to (P3) for Trajectory Optimization

1) *Problem Discretization:* Solving problem (P3) is challenging due to its non-convex nature and the high-dimensional solution space. To address this, we propose

Algorithm 1 An alternating optimization procedure for propagation parameter estimation.

Initialize the path loss parameter $\{\alpha_q^{(0)}, \beta_q^{(0)}\}$ using (21) and $\sigma_q^{(0)}$ using (23).

Loop for the $(i + 1)$ th iteration:

- Initialize all weights $\lambda_t^{(0)}$ to 1.
 - Loop for the $(j + 1)$ th iteration:
 - Update $\{w_{q,m}^{(j+1)}, \eta_{q,m}^{(j+1)}, c_{q,m}^{(j+1)}\}$ from (28) and (29).
 - Update $\lambda_t^{(j+1)}$ using (27).
 - Repeat Until $w_{q,m}^{(j+1)} = w_{q,m}^{(j)}, \eta_{q,m}^{(j+1)} = \eta_{q,m}^{(j)},$ and $c_{q,m}^{(j+1)} = c_{q,m}^{(j)}$.
 - Update $\{\alpha_q^{(i+1)}, \beta_q^{(i+1)}\}$ using (22) and $\sigma_q^{(i+1)}$ using (23).
- Repeat Until $\alpha_q^{(i+1)} = \alpha_q^{(i)}, \beta_q^{(i+1)} = \beta_q^{(i)},$ and $\sigma_q^{(i+1)} = \sigma_q^{(i)}$.

a discretization strategy that approximates the solution to problem (P3) with linear complexity within a discrete domain. Specifically, we discretize the area of interest into grid locations spaced equally by τ meters, thereby constructing the location set $\mathcal{V}$. The area is then represented as a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where edges in the set $\mathcal{E}$ connect locations that are reachable within a maximum of K hops. The parameter K is calculated as $K = \lceil \delta v_{\max} / \tau \rceil$ where $v_{\max}$ is the maximum speed attainable by the mobile user.

We first discretized $p(\mathbf{x}_t | \mathbf{x}_{t-1}, \mathbf{x}_{t-2}; \hat{\Theta}_m)$ in (P3) as

$$\begin{aligned} & \mathbb{P}(\mathbf{x}_t | \mathbf{x}_{t-1}, \mathbf{x}_{t-2}; \hat{\Theta}_m) \\ &= \frac{p(\mathbf{x}_t | \mathbf{x}_{t-1}, \mathbf{x}_{t-2}; \hat{\Theta}_m)}{\sum_{\mathbf{v} \in \{\mathbf{v} | (\mathbf{x}_{t-1}, \mathbf{v}) \in \mathcal{E}, \mathbf{v} \in \mathcal{V}\}} p(\mathbf{v} | \mathbf{x}_{t-1}, \mathbf{x}_{t-2}; \hat{\Theta}_m)}, \end{aligned}$$

ensuring that the sum of transition probabilities at each time slot equals 1.

Then, problem (P3) can be discretized to

$$\begin{aligned} \text{(P3.1): maximize} \quad & \sum_{t=1}^T \left\{ \log p(\mathbf{y}_t | \mathbf{x}_t; \hat{\Theta}_p) \right. \\ & \left. + \mathbb{I}(t > 2) \log \mathbb{P}(\mathbf{x}_t | \mathbf{x}_{t-1}, \mathbf{x}_{t-2}; \hat{\Theta}_m) \right\} \\ \text{subject to} \quad & \mathbf{x}_t \in \mathcal{V}, (\mathbf{x}_t, \mathbf{x}_{t-1}) \in \mathcal{E} \end{aligned}$$

Our goal is to find a trajectory within a discrete space that maximizes the log-likelihood $\log p(\mathcal{Y}_T, \mathcal{X}_T)$ given the signal propagation parameters $\hat{\Theta}_p$ and mobility model parameters $\hat{\Theta}_m$.

2) *Algorithm and Complexity*: Problem (P3.1) follows a classical HMM optimization form, with the distinction that the current state depends on the previous two states. Problem (P3.1) can be efficiently solved using a modified version of the Viterbi algorithm with globally optimal guarantee.

At each step, there are $|\mathcal{V}|$ candidate locations considered, but states with very low probabilities $p(\mathbf{y}_t | \mathbf{x}_t; \hat{\Theta}_p)$ are highly unlikely to contribute to the optimal path. To improve efficiency, states with probabilities below a threshold ζ are pruned. Mathematically, this corresponds to retaining only the top $n_t(\zeta)$ most probable locations at time slot t , where $n_t(\zeta)$ is the number of elements in the set $\{\mathbf{x}_t | p(\mathbf{y}_t | \mathbf{x}_t; \hat{\Theta}_p) > \zeta, \mathbf{x}_t \in \mathcal{V}\}$. Denote the maximum number of element in the set $n_{\max}(\zeta) = \max_t \{n_t(\zeta)\}$.

Algorithm 2 An alternating optimization algorithm for trajectory recovery.

Initialize the parameter $\Theta_p^{(0)}, \Theta_m^{(0)}$ randomly.

Loop for the $(i + 1)$ th iteration:

- Update $\mathcal{X}_T^{(i+1)}$ using the gradient descent method, initialized with the output of the Viterbi algorithm.
 - Update $\Theta_p^{(i+1)}$ using Algorithm 1.
 - Update $\Theta_m^{(i+1)}$ using (15)-(16)
- Until $\mathcal{X}_T^{(i+1)} = \mathcal{X}_T^{(i)}$.

Considering the number of candidate previous states for the current state, which is constrained by the graph structure, it is of the order $\mathcal{O}(K^2)$ for a square region and $\mathcal{O}(K)$ for an unbranched road network. Thus, we have the following result.

Proposition 3: The computational complexity of solving problem (P3.1) is upper bounded by $\mathcal{O}(T n_{\max}(\zeta) (\delta v_{\max} / \tau)^2)$ and lower bounded by $\mathcal{O}(T n_{\max}(\zeta) (\delta v_{\max} / \tau))$.

We can solve problem (P3.1) with linear complexity, as stated in Proposition 3. Given the solution $\mathcal{X}_T^{(0)}$ of (P3.1), we employ the gradient descent method with a learning rate l_r to obtain a convergent solution to problem (P3), initialized with $\mathcal{X}_T^{(0)}$. The overall algorithm is summarized in Algorithm 2. We first initialize the propagation parameter Θ_p and the mobility parameter Θ_m randomly and then begin the alternating update of $\mathcal{X}_T, \Theta_p$ and Θ_m alternatively until convergence. Since each iteration of this procedure never decreases the objective function, which is bounded above, the iterative process is therefore guaranteed to converge. Moreover, as problem (9) is inherently non-convex, the algorithm is not guaranteed to find a unique global optimum, and may yield different solutions depending on the initialization.

V. NUMERICAL RESULTS

A. Datasets

This paper validates the proposed algorithm using three datasets:

- **Synthetic Single-Antenna Dataset I**: We simulate a trajectory of length 1×10^5 meters using the mobility model defined in (7), parameterized by $\gamma = 1, \mathbf{v} = [10, 0]^T$ m/s, $\mathbf{x} = [0, 0]^T$ m, and $\delta = 0.5$ seconds. We consider BS at a height of 50 meters and a mobile user at a height of 2 meters equipped with a single antenna. The parameters in the signal propagation model (3) are $N_t = 1, \beta_q = 5, \alpha_q = -20$, and $\sigma_q = 0.2, 0.5, 1$. Two scenarios are considered: in Scenario 1 (BS deployed in a limited region, c.f., Section III-B), the number of BSs surrounding the trajectory is fixed at $Q = 4, 8, 12, 16, 20$; in Scenario 2 (BS deployed in an unlimited region, c.f., Section III-C), the BSs in the target area follow a PPP with densities $\kappa = 1.02 \times 10^{-3}, 3.02 \times 10^{-3}, 5.02 \times 10^{-3}, 7.02 \times 10^{-3}, 9.02 \times 10^{-3}$, and 1.02×10^{-2} units per m^2 . The mobile user can only detect BSs within a radius of $R = 50, 100, 200, 300, 400, 500$ meters.

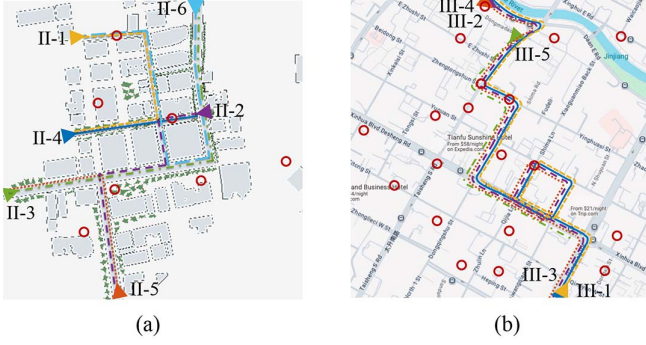


Fig. 2. The data collection environment of (a) synthetic MIMO dataset and (b) real MIMO dataset. The signal from BSs (gray points) is measured along the trajectory (distinct line styles and colors, begin with triangles).

- **Synthetic MIMO Dataset II:** We utilized Wireless Insite[®] to simulate a environment encompassing a 700 m × 700 m area in San Francisco, USA, featuring building heights ranging from 12 m to 204 m. As illustrated in Fig. 2(a), seven BSs with a height of 55 meters, each were manually deployed on selected rooftops to ensure comprehensive coverage of the area of interest. Each BS is equipped with a 64-antenna dual-polarized MIMO array and configured with $M = 7$ beams. The antenna orientation of each BS spans approximately 120° , and the transmit power is set to 0 dBm. The sensing vectors are constructed using the Kronecker product of the array responses and are utilized to compute the SSB RSRP by evaluating signal strengths across various receivers equipped with isotropic antennas along the route. We recorded the SSB RSRP in receivers positioned at a height of 2 meters long six predefined trajectories with lengths 675 m, 792 m, 1085 m, 355 m, 627 m, and 1182 m, called II-1~6. All measurements are conducted at speeds of 6, 12, 18, 24, 30 m/s with a sampling interval of $\delta = 0.5$ s.

- **Real MIMO Dataset III:** We conducted a driving procedure in an urban area of China, covering a 1350 m × 1350 m region, where we collected global positioning system (GPS)-reported location data, as well as the RSRP, RSSI, and SINR of 32 CSI beams from the serving cell. Additionally, we measured the RSRP of $M = 8$ SSB beams both the serving and neighboring cells using a 5G-enabled smartphone. The receiver reliably acquired 8 beam values from the BS in the serving cell, while BSs in neighboring cells provided only 0 to 14 beam values due to device limitations and signal propagation interference. The vehicle traversed five distinct trajectories, labeled III-1 through III-5, and measured signals from 39 surrounding BSs. Specifically, Trajectory III-1 followed a 2652 m path with speeds between 0 and 13.1 m/s and a sampling interval of $\delta = 0.5$ s, yielding 692 samples. Trajectory III-2 retraced III-1 in reverse, III-3 doubled the average speed on III-1, III-4 doubled the speed while reversing III-1, and III-5 covered a 1253 m segment of III-1.

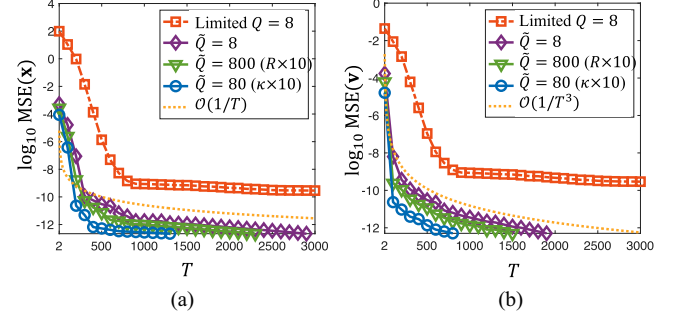


Fig. 3. MSE of (a) $\mathbf{x}$ and (b) $\mathbf{v}$ with different sample number T , the number of BS Q , radius R , and density κ .

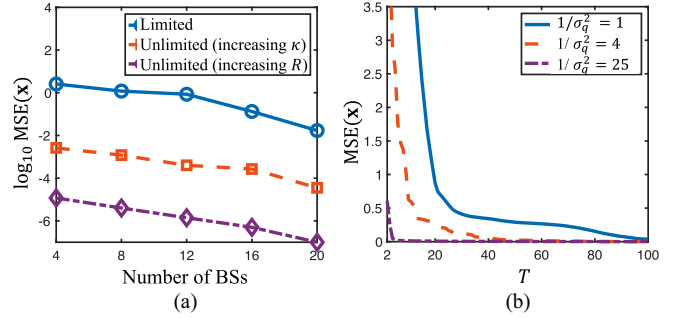


Fig. 4. (a) The relationship between $\text{MSE}(\mathbf{x})$ and the number of BSs. (b) $\text{MSE}(\mathbf{x})$ under different noise σ_q .

B. Numerical Validation of the Theoretical Results

Fig. 3 illustrates the MSE defined as $\text{MSE}(\mathbf{x}) = \|\mathbf{x} - \hat{\mathbf{x}}\|_2^2$ m² and $\text{MSE}(\mathbf{v}) = \|\mathbf{v} - \hat{\mathbf{v}}\|_2^2$ on the synthetic single-antenna dataset with the parameter $\beta_q = 5$, $\alpha_q = -20$, and $\sigma_q = 0.1$, where $\hat{\mathbf{x}}$ and $\hat{\mathbf{v}}$ are the outputs of the proposed algorithm.

In Scenario 1 of the synthetic single-antenna dataset with $Q = 8$, the MSE of $\mathbf{x}$ and $\mathbf{v}$ decreases as T increases within a limited region but does not converge to zero even when $T = 20000$ in our experiments. This behavior is consistent with Theorem 1 and Theorem 2.

In Scenario 2 of the synthetic single-antenna dataset, we set $R = 50$ m and $\kappa = 1.02 \times 10^{-3}$ units per m², resulting in $\tilde{Q} \approx 8$. As T increases, the rate at which $\text{MSE}(\mathbf{x})$ decreases follows $\mathcal{O}(1/T)$, and the rate at which $\text{MSE}(\mathbf{v})$ decreases follows $\mathcal{O}(1/T^3)$, which is consistent with Theorem 3 and Theorem 4. The $\text{MSE}(\mathbf{x})$ and $\text{MSE}(\mathbf{v})$ for the curves $\tilde{Q} = 8$ and $\tilde{Q} = 800$ in Fig. 3 both reach zero when $T > 3200$. In addition, we found that increasing the radius R from 50 to 500 meters results in a lower MSE, and increasing the density κ from 1.02×10^{-3} to 1.02×10^{-2} also yields a lower MSE.

We investigate the effect of the number of BSs, Q and $\tilde{Q}$, on $\text{MSE}(\mathbf{x})$ and $\text{MSE}(\mathbf{v})$. The trajectory length is set to 100 m, with $\beta_q = 5$, $\alpha_q = -5$, and $\sigma_q = 0.1$. For Scenario 2, we consider a fixed radius $R = 50$ m and incrementally increase the density κ to achieve $\tilde{Q} = 4, 8, \dots, 20$. Additionally, we also examine a fixed $\kappa = 1.02 \times 10^{-3}$ units per m² while increasing R to obtain $\tilde{Q} = 4, 8, \dots, 20$. As shown in Fig. 4(a),

TABLE I
COMPARISON OF AVERAGE LOCALIZATION ERROR ON SYNTHETIC MIMO DATASET II (SIX TRAJECTORIES) AND REAL MIMO DATASET III (FIVE TRAJECTORIES)

	without a road map					use a road map to constrain the trajectory				
	MaR [37]	WCL [38]	Proposed ($M=1$)	Proposed	GMA	MaR [37]	WCL [38]	Proposed ($M=1$)	Proposed	GMA
Dataset II	47.6	41.4	10.7	<u>7.2</u>	7.0	42.3	34.8	9.1	<u>7.0</u>	6.3
Dataset III	167.8	124.5	22.7	<u>18.7</u>	17.8	102.2	86.3	19.2	<u>15.9</u>	14.9

TABLE II
AVERAGE LOCALIZATION ERROR (E_1) ON THE SIX TRAJECTORIES SEPARATELY IN SYNTHETIC MIMO DATASET WITH THE SPEED 6 M/S UNDER MAP KNOWLEDGE FREE

Trajectory	II-4	II-5	II-3	II-1	II-2	II-6
Length [m]	355	627	1085	675	792	1182
No. of turns	0	1	1	2	3	3
E_1 [m]	4.39	7.41	6.67	7.80	8.03	7.88

the unlimited scenario consistently achieves a lower MSE compared to the limited scenario under the same number of BSs. This observation is also validated in Fig. 3, where Scenario 2 with $\bar{Q} \approx 8$ achieves a lower MSE compared to Scenario 1 with $Q = 8$. Furthermore, we found that increasing R results in a lower MSE than increasing κ under the same number of BSs. This is because $\hat{\Delta}_{T,x}$ and $\hat{\Delta}_{T,v}$ in Theorem 3 and Theorem 4 is related to $\mathcal{O}(1/(\kappa \ln R))$.

We also investigate the effect of the noise variance σ_q^2 under the unlimited scenario, with $R = 50$ m, $\kappa = 1.02 \times 10^{-3}$ per m^2 , $\beta_q = 5$, $\alpha_q = -5$, and a trajectory length of 500 m. We consider $1/\sigma_q^2 = 1, 4, 25$ for all BSs. As shown in Fig. 4(b), a larger $1/\sigma_q^2$ results in a faster convergence rate. Recall that $1/\sigma_{\max}^2$ is proportional to the SNR. Thus, a smaller $1/\sigma_q^2$ leads to a slower decrease in the CRLB of $\mathbf{x}$ and $\mathbf{v}$ as T increases, as stated in Theorem 3 and Theorem 4.

C. Trajectory Recovery Performance

We use the average localization error $E_1 = \frac{1}{T} \sum_{t=1}^T \|\mathbf{x}_t - \hat{\mathbf{x}}_t\|_2$ to evaluate the trajectory recovery performance of the proposed method. Here, $\mathbf{x}_t$ is the data collection location at time slot t , and $\hat{\mathbf{x}}_t$ is the t th location in the output trajectory of the proposed algorithm. Four baselines are designed for comparison:

- Max-RSS (MaR): At time t , the strongest RSS among the surrounding BSs is selected, and the estimated position is determined by the location of the BS with the strongest RSS.
- Weighted Centroid Localization (WCL): This method computes a weighted location $\hat{\mathbf{p}}_t = \sum_{q=1}^Q w_{t,q} \mathbf{o}_q$, where $w_{t,q} = \sum_{j=1}^M 10^{y_{q,m,t}/20} / [\sum_{j=1}^M \sum_{l=1}^Q 10^{y_{q,j,l}/20}]$.
- Proposed ($M=1$) [30]: This method uses only the max RSRP among the M beams of a BS as the RSS for that BS, and considers only the path-loss model (1) in the signal propagation probability.
- Genius-aided map-assisted (GMA): This method utilizes the true location information available in the real MIMO

dataset III (via GPS) to fit the propagation models described by equation (3) and assumes that the propagation model parameters are known. Only the mobility model parameters and the trajectory are updated alternately until convergence. The GMA method serves as an upper bound for performance comparison.

- Direct AI positioning: We adopt the ‘fingerprinting based on channel observation’ method, as defined in [39], for direct AI positioning. A multi-layer perceptron (MLP) is constructed with fully connected layers of dimensions 49, 128, 64, 16, and 2. This method is supervised, requiring user location labels for training.

For the comparison, if the road map knowledge is given, the output location of the comparison is projected to the nearest road as stated in [40] to utilize the map knowledge. For the proposed method, the target area will be the road network space if the map information is given, and the whole target area otherwise. We set $\tau = 1$ m, $v_{\max} = 120$ km/h, $\zeta = 0.8$, $l_r = 0.01$, $\epsilon = 0.01$, and $\gamma = 0.9$ for the proposed method.

1) *Synthetic MIMO Dataset*: We first evaluate the performance of the proposed method on a synthetic MIMO dataset. As shown in Table I, first, MaR exhibits the poorest performance, demonstrating that relying solely on proximity to BSs is ineffective for accurate mobile user positioning. While the WCL method improves estimation by applying weights to the distances from surrounding BSs, it only marginally outperforms the MaR method. Second, although the proposed ($M=1$) method outperforms MaR and WCL, its performance is slightly inferior to the proposed method that accounts for beam effects due to the absence of beam consideration, underscoring the importance of incorporating angular domain considerations in modeling signal propagation. Third, the proposed method consistently outperforms the comparison methods MaR and WCL, including the $M=1$ variant. Additionally, the proposed method performs slightly worse than GMA. This discrepancy arises because GMA and the proposed method solve the same optimization problem; however, GMA assumes a known real signal propagation model, thereby establishing an upper bound for the proposed method. Nevertheless, the gap between them is only less than 1 meter. Finally, compared to the method with map information, the method without map exhibits a slight increase in localization error, suggesting that all methods perform better when road network information is available.

We investigate the trajectory recovery performance by varying the trajectory length and the complexity of the trajectory in terms of the number of turns. As shown in Table II,

TABLE III
AVERAGE LOCALIZATION ERROR (E_1) ON ALL SIX TRAJECTORIES IN SYNTHETIC MIMO DATASET WITH DIFFERENT SPEEDS UNDER MAP KNOWLEDGE FREE

Speed m/s	6	12	18	24	30
E_1 [m]	7.29	8.12	8.36	8.77	9.49

TABLE IV
COMPARISON OF AVERAGE LOCALIZATION ERROR (E_1) [M]) PERFORMANCE BETWEEN THE PROPOSED METHOD AND THE SUPERVISED METHOD ON SYNTHETIC MIMO DATASET II

	Training error (m)	Test error (m)		
		II-4	II-5	II-6
Direct AI positioning	6.24	8.35	10.23	9.59
Proposed (w/o road map)	(no training)	4.39	7.41	7.88

firstly, increasing the trajectory length indeeds decreases the localization error. This can be seen by comparing the MSE of trace II-3 and II-5, and the MSE of trace II-2 and II-6. Second, for roughly the same trajectory length, making a turn can slightly decreases the localization performance. As seen from traces II-1, II-2, II-4, II-5, the more turns, the higher the MSE.

We also investigate the effect of moving speed on trajectory recovery performance by varying the speed from 6 m/s to 30 m/s. As shown in Table III, the error gradually increases with increasing speed. For a fixed trajectory length, increasing the speed results in a reduced number of samples. As stated in Theorems 1 ~ 4, the CRLB of the MSE decreases as T increases.

For comparison with the Direct AI method, we use datasets II-1, II-2, and II-3 for training, and datasets II-4, II-5, and II-6 for testing. Note that, the trajectories in the test datasets are subsets of that in the training datasets. As shown in Table IV, the Direct AI method is able to achieve a small training error, but the test performance is poor. the proposed method (w/o road map) outperforms the Direct AI approach. This is primarily because our method directly estimates the entire trajectory with mobility model constraints, whereas the Direct AI approach predicts each location independently at each time slot. Additionally, our method incorporates a BS-specific channel model for more accurate RSRP fitting, while the Direct AI approach does not leverage any underlying physical models.

2) *Real MIMO Dataset*: Similar to the performance observed in the synthetic MIMO dataset, the proposed method still achieves the lowest average localization error on real MIMO datasets compared with MaR, WCL, and Proposed ($M = 1$), and shows a very small gap to the GMA method. However, we observed that the performance on the real MIMO dataset is inferior compared to the synthetic MIMO dataset. To investigate the source of this discrepancy, we conducted the following analysis.

Although the synthetic dataset comprises only seven BSs compared to 39 BSs in the real MIMO dataset, in the real MIMO dataset, an average of only 16 values can be measured for each beam in the whole trajectory, whereas the synthetic

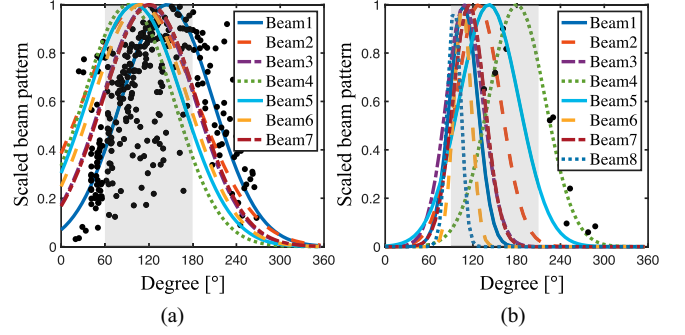


Fig. 5. (a) The seven scaled patterns of the q th BS in synthetic MIMO dataset with measurements (black points) belonging to beam 1. (b) The eight scaled patterns of the q th BS in real MIMO dataset with measurements (black points) belonging to beam 4.

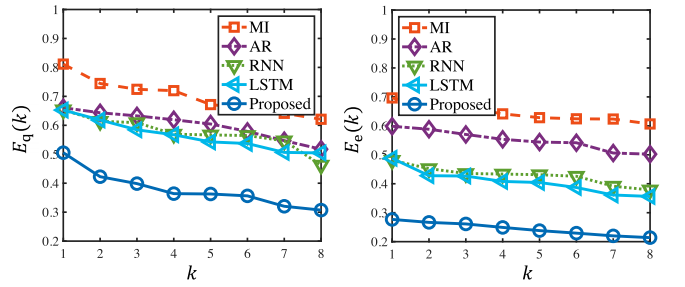


Fig. 6. Performance of CSI beam RSRP prediction.

MIMO dataset allows for the measurement of approximately 1050 values for each beam. The sparsity observed in the real MIMO dataset arises because the mobile device can measure beam values from only six neighboring BSs and records only the strongest beam value for each neighboring BS. As illustrated in Fig. 5, the amount of data used to fit beam 1 in the synthetic MIMO dataset is significantly larger than that used to fit beam 4 in the real MIMO dataset. This larger dataset facilitates a more accurate estimation of the beam pattern in the synthetic scenario.

Fig. 5(b) demonstrates the beam pattern fitting, where the standard deviation σ_q is found to be 0.25. The reconstructed beam patterns align with the expectations communicated by the network operator: eight beams point in different directions for spatial multiplexing. The beam coverage spans approximately 120 degrees; for example, beams are concentrated between 60° and 180° in the synthetic MIMO dataset and between 90° and 210° in the real MIMO dataset. The estimated channel model parameters vary across different BSs in Dataset III. This variation reflects the diversity of propagation environments experienced by each BS.

D. Application: CSI Prediction

When the trajectories are recovered based on unlabeled RSRP measurements of the SSB beam, a *radio map* can be constructed by pairing the recovered location labels with the CSI data. Note that such a construction can be done in an accumulative way to keep improving the accuracy of the radio map. When CSI prediction is needed [41], a

TABLE V
PERFORMANCE OF CSI BEAM SINR AND RSSI PREDICTION

		SINR					RSSI				
		MI	AR	CNN	LSTM	Proposed	MI	AR	CNN	LSTM	Proposed
Average quantity deviation $E_q(k)$	$k = 1$	0.86	0.68	0.61	0.59	0.42	0.96	0.70	0.68	0.69	0.43
	$k = 8$	0.70	0.53	0.52	0.42	0.34	0.70	0.50	0.46	0.47	0.35
	$k = 16$	0.47	0.27	0.22	0.19	0.13	0.56	0.37	0.31	0.25	0.16
Average relative energy deviation $E_e(k)$	$k = 1$	0.63	0.56	0.46	0.41	0.23	0.76	0.50	0.49	0.46	0.31
	$k = 8$	0.60	0.49	0.39	0.30	0.28	0.61	0.46	0.40	0.38	0.27
	$k = 16$	0.41	0.32	0.27	0.21	0.10	0.54	0.38	0.29	0.22	0.12

sequence of *sparse* and *coarse* SSB CSI data is observed over consecutive $L + 1$ time slots, and then, the full CSI data can be recovered from the radio map using a finger-printing approach; consequently, applications such as CSI tracking and quality-of-service prediction may follow.

Consider a constructed radio map includes signal propagation parameters, RSRP of SSB beams, and RSRP, RSSI and SINR of CSI beams, along with the location label. Given the current SSB RSRP measurement $\mathbf{y}_t$ and its L -length history $\{\mathbf{y}_i\}_{i=t-L}^{t-1}$, our objective is to predict the RSRP, RSSI and SINR of CSI beams for the next time slot $t + 1$.

We begin by addressing the estimation of SSB RSRP $\mathbf{y}_{t+1}$ at time slot $t + 1$ through maximizing $\sum_{i=t-L}^t \log p(\mathbf{y}_i|\mathbf{x}_i) + \log p(\mathbf{y}_{t+1}|\bar{\mathbf{x}}_{t+1}) + \sum_{j=t-L_p+2}^t \log p(\mathbf{x}_j|\mathbf{x}_{j-1}, \mathbf{x}_{j-2})$ w.r.t. $\bar{\mathbf{x}}_{t+1}$, where the location for the next time slot is estimated as $\bar{\mathbf{x}}_{t+1} = (1 + \gamma)\mathbf{x}_t - \gamma\mathbf{x}_{t-1} + (1 - \gamma)\delta\bar{\mathbf{v}}$. This optimization problem is solved by alternately updating $\{\mathbf{x}_i\}_{i=t-L_p}^t$ and $\{\bar{\mathbf{v}}, \sigma_v^2\}$ based on Algorithm 2. Subsequently, the estimated location $\bar{\mathbf{x}}_{t+1}$ is calculated, and the estimated SSB RSRP $\hat{\mathbf{y}}_{t+1}$ is obtained using the measurement model in (3). Finally, the RSRP, RSSI and SINR of CSI beams at time slot $t + 1$ are predicted by matching $\hat{\mathbf{y}}_{t+1}$ with the constructed radio map through a nearest SSB RSRP search.

We construct the radio map using III-1 in the real MIMO dataset and evaluate CSI prediction performance on III-2 to III-5. We compare our method with four baselines: Mean Inference (MI), which computes the average beam variation over a window of length L and adds this average to the current measurement $\mathbf{y}_t$ to estimate $\hat{\mathbf{y}}_{t+1}$; AutoRegressive (AR), a linear order- L model trained via gradient descent; Convolutional Neural Network (CNN), featuring three 3×3 convolutional layers with ReLU activations and a fully connected layer; and Long Short-Term Memory Network (LSTM), comprising three LSTM layers with dropout and a fully connected output. All models input sequences of past normalized SSB RSRP measurements to predict the SSB RSRP of the next time slot, with CNN and LSTM trained using the Adam optimizer for up to 1000 epochs at a learning rate of 0.001, and L set to 12 for all methods.

We employ three metrics to evaluate the accuracy and reliability of prediction models in estimating future CSI values based on current and historical measurements. Let $\{\hat{\mathbf{h}}_n\}_{n=1}^N$ represent the predicted CSI beam information, and $\{\mathbf{h}_n\}_{n=1}^N$ denote the real measurements. The first metric, *average quantity deviation error* on the k -strongest beams, $E_q(k) = \frac{1}{NQ} \sum_{n=1}^N \sum_{q=1}^Q \left(1 - |\hat{\mathcal{B}}_{q,n,k} \cap \mathcal{B}_{q,n,k}|/k\right)$,

TABLE VI
PERFORMANCE FOR CSI BEAM MAXIMUM RSRP PREDICTION

MAE (dB)	MI	AR	CNN	LSTM	Proposed
E_a	11.16	8.93	8.80	7.49	4.44

measures the average mismatch in top- k beam selection across all base stations and time indices, where $\hat{\mathcal{B}}_{q,n,k}$ denotes the set of indices corresponding to the k strongest beams in $\hat{\mathbf{h}}_n$ for BS q , and $\mathcal{B}_{q,n,k}$ denotes the same for $\mathbf{h}_n$. The quantity $|\hat{\mathcal{B}}_{q,n,k} \cap \mathcal{B}_{q,n,k}|$ represents the number of beams correctly predicted within the top- k set for BS q . Lower values of $E_q(k)$ indicate more accurate beam predictions. The second metric, *average relative energy deviation* on the k -strongest beams, $E_e(k) = \frac{1}{NQ} \sum_{n=1}^N \sum_{q=1}^Q |e_{q,n,k} - \hat{e}_{q,n,k}|/e_{q,n,k}$, assesses the average relative difference in the total energy of the predicted versus actual top- k beams, where $\hat{e}_{q,n,k}$ and $e_{q,n,k}$ denote the total energy of the k strongest beams in the prediction and actual measurements, respectively, for BS q . Lower values of $E_e(k)$ reflect higher accuracy in the energy allocation of the predictions. The third metric is the mean absolute error (MAE) of the predicted maximum beam CSI, defined as $E_a = \frac{1}{NQ} \sum_{n=1}^N \sum_{q=1}^Q |\nu_{n,q} - \hat{\nu}_{n,q}|$ where $\hat{\nu}_{n,q}$ and $\nu_{n,q}$ denote the predicted and real CSI measurements of the strongest beam, respectively for BS q .

Fig. 6 presents the CSI beam RSRP prediction performance of the proposed radio map-assisted method in comparison with the baseline methods. The proposed radio map-assisted method consistently demonstrates superior performance over the baselines. As the parameter k increases, both $E_q(k)$ and $E_e(k)$ decrease, indicating enhanced prediction accuracy. The proposed radio map-assisted method achieves average energy deviation errors of 21% and average quantity deviation error of 31% for predicting the eight strongest RSRP of 32 CSI beams. Table V presents the E_a performance for CSI beam maximum RSRP prediction. The proposed method achieves a prediction error of less than 4.44 dB.

Table V displays the prediction performance for CSI beam SINR and RSSI. The proposed radio map-assisted method always outperforms the comparison methods. Specifically, the proposed radio map-assisted method achieves average energy deviation errors of 28%, 27% and average quantity deviation error of 34%, 35% for predicting the eight strongest RSSI and SINR beams of CSI, respectively. For the strongest $k = 16$ beams prediction, the E_q of the proposed radio map-assisted method is only 0.13 and 0.16 for SINR and RSSI, respectively, which are reductions of 0.07 and 0.09 compared

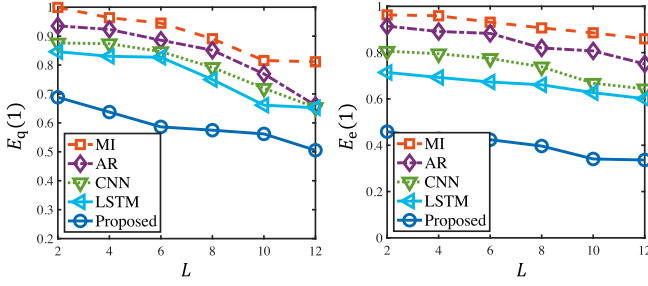


Fig. 7. Performance of CSI beam RSRP prediction with different L .

TABLE VII
PERFORMANCE ($E_q(8)$) OF CSI BEAM RSRP
PREDICTION ON DIFFERENT TESTING
TRAJECTORIES IN THE REAL MIMO DATASET

	III-2	III-3	III-4	III-5
MI	0.51	0.61	0.81	0.46
AR	0.42	0.51	0.68	0.36
CNN	0.38	0.45	0.60	0.32
LSTM	0.41	0.49	0.66	0.35
Proposed	0.18	0.35	0.44	0.16

to the LSTM method. Additionally, the $E_e(16)$ of the proposed method is 0.10 and 0.12 for SINR and RSSI, respectively, which are decreases of 0.11 and 0.10 relative to the LSTM method.

We investigate the effect of L on the CSI prediction performance. As shown in Fig. 7, as L increases, the proposed radio map-assisted method exhibits a decreasing trend in $E_e(1)$ and $E_q(1)$. This is because incorporating more historical RSRP measurements enables more accurate estimation of the current location, resulting in improved CSI prediction accuracy. Furthermore, the proposed radio map-assisted method consistently outperforms the comparisons.

We investigate the performance of the proposed radio map-assisted method on different testing trajectories as shown in Table VII. First, we compare the performance on Trajectory III-5 with that on Trajectories III-2 to III-4. The proposed radio map-assisted method, along with the baseline methods, achieves the best performance on Trajectory III-5. This superior performance is attributed to the fact that Trajectory III-5 is included in the training set, resulting in prediction errors that are closely aligned with the training error. Second, we assess the performance on Trajectory III-2 in comparison with Trajectories III-3 and III-4. The proposed radio map-assisted method and the baseline methods demonstrate better performance on Trajectory III-2. The only difference between Trajectory III-2 and the training trajectory is the direction of movement, with Trajectory III-2 moving in the opposite direction. Although reversing the direction introduces some increase in prediction error, the increase remains limited. Third, we examine the performance on Trajectories III-3 and III-4 relative to Trajectories III-2 and III-5. We observe that doubling the speed results in an increase in prediction error significantly.

VI. CONCLUSION

This paper presents a method for constructing an angular power map without the need for location labels. We begin by proposing a mobility model for the mobile user and

modeling the signal propagation of each beam. Subsequently, we introduce a novel HMM-based RSRP embedding technique to recover the data collection trajectory of CSI sequences in massive MIMO networks. As a result, an angular power map is constructed without requiring calibration efforts. We establish theoretical results demonstrating that under uniform rectilinear mobility and PPP BSs, the CRLB of the localization error can asymptotically approach zero at any SNR. Moreover, if the BS are deployed only within a limited region, the localization error cannot approach zero even with an infinite amount of independent measurement data. Experiments conducted in a real commercial 5G network confirm the effectiveness of our method, achieving a mean localization error below 18 meters based on sparse SSB RSRP measurements.

APPENDIX A PROOF OF THEOREM 1

From (12), the FIM $\mathbf{F}_{T,x}$, as the upper diagonal block in $\mathbf{F}_{T,\psi}$, can be expressed as

$$\mathbf{F}_{T,x} = \sum_{t=1}^T \sum_{q=1}^Q \frac{\alpha_q^2}{\sigma_q^2 d_{t,q}^4(\mathbf{x}, \mathbf{v})} (\mathbf{l}_q(\mathbf{x}) + t\mathbf{v})(\mathbf{l}_q(\mathbf{x}) + t\mathbf{v})^T. \quad (30)$$

In the following text, we simplify the notation by writing $\mathbf{l}_q(\mathbf{x})$ as $\mathbf{l}_q$ and $d_{t,q}(\mathbf{x}, \mathbf{v})$ as $d_{t,q}$. Denoting $\alpha_{\max}^2 = \max_q \{\alpha_q^2\} \geq \alpha_q^2$, $\sigma_{\min}^2 = \min_q \{\sigma_q^2\} \leq \sigma_q^2$, we have

$$\begin{aligned} \mathbf{F}_{T,x} &\preceq \frac{\alpha_{\max}^2}{\sigma_{\min}^2} \sum_{t=1}^T \sum_{q=1}^Q \frac{1}{d_{t,q}^4} (\mathbf{l}_q + t\mathbf{v})(\mathbf{l}_q + t\mathbf{v})^T \\ &= C_0 \left[\sum_{q=1}^Q s_{T,q}^{(0)} \mathbf{l}_q \mathbf{l}_q^T + \sum_{q=1}^Q s_{T,q}^{(1)} (\mathbf{l}_q \mathbf{v}^T + \mathbf{v} \mathbf{l}_q^T) \right. \\ &\quad \left. + \sum_{q=1}^Q s_{T,q}^{(2)} \mathbf{v} \mathbf{v}^T \right] = C_0 \mathbf{A}_{T,x} \end{aligned} \quad (31)$$

where $C_0 = \frac{\alpha_{\max}^2}{\sigma_{\min}^2}$, $s_{T,q}^{(n)} = \sum_{t=1}^T \frac{t^n}{d_{t,q}^4}$, and

$$\mathbf{A}_{T,x} = \sum_{q=1}^Q s_{T,q}^{(0)} \mathbf{l}_q \mathbf{l}_q^T + \sum_{q=1}^Q s_{T,q}^{(1)} (\mathbf{l}_q \mathbf{v}^T + \mathbf{v} \mathbf{l}_q^T) + \sum_{q=1}^Q s_{T,q}^{(2)} \mathbf{v} \mathbf{v}^T \quad (32)$$

and equality (31) can be achieved when $\sigma_{\min}^2 = \sigma_q^2$ and $\alpha_{\max}^2 = \alpha_q^2$.

It is observed that (31) $\mathbf{A}_{T,x} \prec \mathbf{A}_{T+1,x}$ because each component $\sum_q \frac{1}{d_{t,q}^4} (\mathbf{l}_q + t\mathbf{v})(\mathbf{l}_q + t\mathbf{v})^T$ in (31) is a positive definite matrix since at least two vectors in $\{\mathbf{l}_1, \mathbf{l}_2, \dots, \mathbf{l}_Q, \mathbf{v}\}$ are linear independent. Therefore, $\text{tr}\{\mathbf{F}_{T,x}^{-1}\} \geq \Delta_{T,x} \triangleq \text{tr}\{(C_0 \mathbf{A}_{T,x})^{-1}\}$ is strictly decreasing in T .

Lemma 5: Suppose $d_{\min,q} > 0$. The sequence $s_{T,q}^{(n)}$ is bounded for $n < 3$ and divergent as $s_{T,q}^{(n)} \rightarrow \infty$ as $T \rightarrow \infty$ for $n \geq 3$. In addition, $s_{T,q}^{(n+1)} / s_{T,q}^{(n)} \rightarrow \infty$ as $T \rightarrow \infty$ for $n > 3$.

Proof: Recall $d_{t,q} = \|l_q + tv\|_2$ which is asymptotically a linear function in t for large t . Then, there exists a sufficiently small $\rho_1 > 0$, such that $d_{t,q} > \rho_1 t$ for all $t \geq 1$. As a result,

$$0 < s_{T,q}^{(n)} = \sum_{t=1}^T \frac{t^n}{d_{t,q}^4} \leq \sum_{t=1}^T \frac{t^n}{(\rho_1 t)^4} = \frac{1}{\rho_1^4} \sum_{t=1}^T \frac{1}{t^{4-n}} < \infty$$

for $n < 3$. This is because $\sum_{t=1}^T \frac{1}{t^r}$ is convergent for all $r > 1$.

In addition, there exists a large enough $\rho_2 < \infty$, such that $d_{t,q} < \rho_2 t$ for all $t \geq 1$. As a result, $s_{T,q}^{(n)} = \sum_{t=1}^T \frac{t^n}{d_{t,q}^4} \geq \sum_{t=1}^T \frac{t^n}{(\rho_2 t)^4} = \frac{1}{\rho_2^4} \sum_{t=1}^T \frac{1}{t^{4-n}} \rightarrow \infty$ as $T \rightarrow \infty$ if $4 - n \leq 1$, i.e., $n \geq 3$.

Moreover, for $n \geq 3$, we have

$$\frac{s_{T,q}^{(n+1)}}{s_{T,q}^{(n)}} \geq \frac{\sum_{t=1}^T \frac{t^{n+1}}{(\rho_2 t)^4}}{\sum_{t=1}^T \frac{t^n}{(\rho_1 t)^4}} = \frac{\frac{1}{\rho_2^4} \sum_{t=1}^T \frac{1}{t^{4-n-1}}}{\frac{1}{\rho_1^4} \sum_{t=1}^T \frac{1}{t^{4-n}}} \rightarrow \infty$$

as $T \rightarrow \infty$. $\square$

Using Lemma 5, since $s_{T,q}^{(n)}$ are bounded for $n < 3$ and Q is finite, we have $\mathbf{A}_{T,x}$ bounded. Thus, $\bar{\Delta}_{T,x}$ converges to a strictly positive number as $T \rightarrow \infty$.

APPENDIX B PROOF OF THEOREM 2

From (12), the FIM $\mathbf{F}_{T,v}$, as the lower diagonal block in $\mathbf{F}_{T,\psi}$ can be expressed as

$$\mathbf{F}_{T,v} = \sum_{t=1}^T \sum_{q=1}^Q \frac{\alpha_q^2 t^2}{\sigma_q^2 d_{t,q}^4} (l_q + tv)(l_q + tv)^T. \quad (33)$$

Similar to (31), we have,

$$\begin{aligned} \mathbf{F}_{T,v} &\preceq \sum_{t,q} \frac{\alpha_{\max}^2}{\sigma_{\min}^2} \frac{t^2}{d_{t,q}^4} (l_q + tv)(l_q + tv)^T \\ &= C_0 \cdot \left[\sum_{q=1}^Q s_{T,q}^{(2)} l_q l_q^T + \sum_{q=1}^Q s_{T,q}^{(3)} (l_q \mathbf{v}^T + \mathbf{v} l_q^T) \right. \\ &\quad \left. + \sum_{q=1}^Q s_{T,q}^{(4)} \mathbf{v} \mathbf{v}^T \right] = C_0 \mathbf{A}_{T,v} \end{aligned} \quad (34)$$

where $\mathbf{A}_{T,v} = \sum_{q=1}^Q s_{T,q}^{(2)} l_q l_q^T + \sum_{q=1}^Q s_{T,q}^{(3)} (l_q \mathbf{v}^T + \mathbf{v} l_q^T) + \sum_{q=1}^Q s_{T,q}^{(4)} \mathbf{v} \mathbf{v}^T$ and the equality is achieved when $\alpha_{\max}^2 = \alpha_q^2$, $\sigma_{\min}^2 = \sigma_q^2$.

Lemma 6: The eigenvalues of $\mathbf{A}_{T,v}$ satisfies $\lambda_{\min}(\mathbf{A}_{T,v}) \rightarrow \sum_{q=1}^Q s_{T,q}^{(2)} \mathbf{v}_{\perp}^T l_q l_q^T \mathbf{v}_{\perp}$ and $\lambda_{\max}(\mathbf{A}_{T,v}) \rightarrow \sum_{q=1}^Q s_{T,q}^{(4)} \|\mathbf{v}\|^2$ as $T \rightarrow \infty$.

Proof: Using Lemma 5, $s_{T,q}^{(2)}$ is bounded, and $s_{T,q}^{(3)}$ and $s_{T,q}^{(4)}$ are divergent, where $s_{T,q}^{(4)}$ dominates $s_{T,q}^{(3)}$ for asymptotically large T . Therefore, the term $\sum_{q=1}^Q s_{T,q}^{(4)} \mathbf{v} \mathbf{v}^T$ dominates $\mathbf{A}_{T,v}$ for a sufficiently large T . Thus, for a sufficiently large T , the larger eigenvalue satisfies

$$\begin{aligned} \lambda_{\max}(\mathbf{A}_{T,v}) &= \max_{\|\mathbf{u}\|=1} \mathbf{u}^T \mathbf{A}_{T,v} \mathbf{u} \\ &= \max_{\|\mathbf{u}\|=1} \mathbf{u}^T \left[\sum_{q=1}^Q s_{T,q}^{(2)} l_q l_q^T \right. \end{aligned}$$

$$\begin{aligned} &+ \sum_{q=1}^Q s_{T,q}^{(3)} (l_q \mathbf{v}^T + \mathbf{v} l_q^T) + \sum_{q=1}^Q s_{T,q}^{(4)} \mathbf{v} \mathbf{v}^T \Big] \mathbf{u} \\ &\approx \max_{\|\mathbf{u}\|=1} \sum_{q=1}^Q s_{T,q}^{(4)} \cdot \mathbf{u}^T (\mathbf{v} \mathbf{v}^T) \mathbf{u} \end{aligned} \quad (35)$$

where the solution to (35) is $\mathbf{u} = \mathbf{v}/\|\mathbf{v}\|$, and $\lambda_{\max}(\mathbf{A}_{T,v}) \rightarrow \sum_{q=1}^Q s_{T,q}^{(4)} \|\mathbf{v}\|^2$. As a result, asymptotically, the larger eigenvector is $\mathbf{v}/\|\mathbf{v}\| \in \mathbb{R}^2$, and hence, the smaller eigenvector is denoted as $\mathbf{v}_{\perp}$, which satisfies $\mathbf{v}_{\perp}^T \mathbf{v} = 0$.

Consequently, we have

$$\begin{aligned} \lambda_{\min}(\mathbf{A}_{T,v}) &= \mathbf{v}_{\perp}^T \left[\sum_{q=1}^Q s_{T,q}^{(2)} l_q l_q^T + \sum_{q=1}^Q s_{T,q}^{(3)} (l_q \mathbf{v}^T + \mathbf{v} l_q^T) \right. \\ &\quad \left. + \sum_{q=1}^Q s_{T,q}^{(4)} \mathbf{v} \mathbf{v}^T \right] \mathbf{v}_{\perp} \\ &= \mathbf{v}_{\perp}^T \sum_{q=1}^Q s_{T,q}^{(2)} l_q l_q^T \mathbf{v}_{\perp} + \mathbf{v}_{\perp}^T \left(\sum_{q=1}^Q s_{T,q}^{(3)} (l_q \mathbf{v}^T \right. \\ &\quad \left. + \mathbf{v} l_q^T) \right) \mathbf{v}_{\perp} + \mathbf{v}_{\perp}^T \sum_{q=1}^Q s_{T,q}^{(4)} \mathbf{v} \mathbf{v}^T \mathbf{v}_{\perp} \\ &= \sum_{q=1}^Q s_{T,q}^{(2)} \mathbf{v}_{\perp}^T l_q l_q^T \mathbf{v}_{\perp} \end{aligned}$$

$\square$

From $\mathbf{F}_{T,v} \preceq C_0 \mathbf{A}_{T,v}$, since both $\mathbf{F}_{T,v}$ and $\mathbf{A}_{T,v}$ are P.S.D., we have

$$\lambda_{\min}(\mathbf{F}_{T,v}) \leq C_0 \lambda_{\min}(\mathbf{A}_{T,v}), \lambda_{\max}(\mathbf{F}_{T,v}) \leq C_0 \lambda_{\max}(\mathbf{A}_{T,v}). \quad (36)$$

Denoting the Eigen Value Decomposition (EVD) of $\mathbf{F}_{T,v}$ as $\mathbf{F}_{T,v} = \mathbf{u}_{T,v} \mathbf{A}_{T,v} \mathbf{u}_{T,v}^{-1}$, where

$$\mathbf{A}_{T,v} = \begin{bmatrix} \lambda_{\max}(\mathbf{A}_{T,v}) & 0 \\ 0 & \lambda_{\min}(\mathbf{A}_{T,v}) \end{bmatrix},$$

we have

$$\begin{aligned} \text{tr}\{\mathbf{F}_{T,v}^{-1}\} &= \text{tr}\{(\mathbf{u}_{T,v} \mathbf{A}_{T,v} \mathbf{u}_{T,v}^{-1})^{-1}\} = \text{tr}\{\mathbf{u}_{T,v} \mathbf{A}_{T,v}^{-1} \mathbf{u}_{T,v}^{-1}\} \\ &= \lambda_{\max}^{-1}(\mathbf{F}_{T,v}) + \lambda_{\min}^{-1}(\mathbf{F}_{T,v}) \\ &\geq \frac{1}{C_0} \lambda_{\max}^{-1}(\mathbf{A}_{T,v}) + \frac{1}{C_0} \lambda_{\min}^{-1}(\mathbf{A}_{T,v}) \end{aligned} \quad (37)$$

$$\geq \frac{1}{C_0} \lambda_{\min}^{-1}(\mathbf{A}_{T,v}) \triangleq \bar{\Delta}_{T,v} \quad (38)$$

where (37) is due to (36) with equality achieved when $\alpha_{\max}^2 = \alpha_q^2$, $\sigma_{\min}^2 = \sigma_q^2$ and (38) is due to the fact that $C_0 \lambda_{\max}(\mathbf{A}_{T,v}) > 0$ and equality can be asymptotically achieved at large T as $\lambda_{\max}^{-1}(\mathbf{A}_{T,v}) \rightarrow 1/(\sum_{q=1}^Q s_{T,q}^{(4)} \|\mathbf{v}\|^2)$ which converges to zero.

Using Lemma 6, as $T \rightarrow \infty$, we have $\bar{\Delta}_{T,v} \rightarrow C_v = (C_0 \sum_{q=1}^Q s_{T,q}^{(2)} \mathbf{v}_{\perp}^T l_q l_q^T \mathbf{v}_{\perp})^{-1}$, which is strictly positive. Since the orthogonal projection $\mathbf{v}_{\perp}^T l_q = l_q - (l_q^T \mathbf{v}/\|\mathbf{v}\|) \mathbf{v} = (\mathbf{I} - \mathbf{v} \mathbf{v}^T / \|\mathbf{v}\|^2) l_q$, we have $\sum_{q=1}^Q s_{T,q}^{(2)} \mathbf{v}_{\perp}^T l_q l_q^T \mathbf{v}_{\perp} = \sum_{q=1}^Q s_{T,q}^{(2)} \|\mathbf{v}_{\perp}\|^2 \|\mathbf{l}_q\|^2 = \sum_{q=1}^Q s_{T,q}^{(2)} \|\mathbf{l}_q\|^2 \|\mathbf{v}_{\perp}\|^2 = \sum_{q=1}^Q s_{T,q}^{(2)} \|\mathbf{P}_{\mathbf{v}_{\perp}}^{\perp} l_q\|^2$, where $\mathbf{P}_{\mathbf{v}_{\perp}}^{\perp} = \mathbf{I} - \mathbf{v} \mathbf{v}^T / \|\mathbf{v}\|^2$ is

orthogonal projector, and $\mathbf{P}_v^\perp \mathbf{l}_q$ is to project the vector $\mathbf{l}_q$ onto the null space spanned by $\mathbf{v}_\perp$ of $\mathbf{v}$.

Thus, we have $C_v = (C_0 \sum_{q=1}^Q s_{T,q}^{(2)} \|\mathbf{P}_v^\perp \mathbf{l}_q\|^2)^{-1}$, where $s_{T,q}^{(2)}$ is bounded as stated in Lemma 5. Recall $\rho > 0$ is sufficiently small such that $d_{t,q}(\mathbf{x}, \mathbf{v}) > \rho t$ for all $t \geq 1$, we have $s_{\infty,q}^{(2)} = \lim_{T \rightarrow \infty} \sum_{t=1}^T \frac{t^2}{d_{t,q}^4} < \lim_{T \rightarrow \infty} \sum_{t=1}^T \frac{t^2}{(\rho t)^4} = \frac{1}{\rho^4} \lim_{T \rightarrow \infty} \sum_{t=1}^T \frac{1}{t^2} \approx \frac{\pi^2}{6\rho^4}$. Thus, the element $s_{\infty,q}^{(2)}$ is upper bounded by $\frac{\pi^2}{6\rho^4}$.

REFERENCES

- [1] H. Du, Y. Deng, J. Xue, D. Meng, Q. Zhao, and Z. Xu, "Robust online CSI estimation in a complex environment," *IEEE Trans. Wireless Commun.*, vol. 21, no. 10, pp. 8322–8336, Oct. 2022.
- [2] Z. Liu, A. Bhandari, and B. Clerckx, "λ-MIMO: Massive MIMO via modulo sampling," *IEEE Trans. Commun.*, vol. 71, no. 11, pp. 6301–6315, Nov. 2023.
- [3] Z. Xing et al., "Spectrum efficiency prediction for real-world 5G networks based on drive testing data," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, 2022, pp. 2136–2141.
- [4] Y. Zeng et al., "A tutorial on environment-aware communications via channel knowledge map for 6G," *IEEE Commun. Surv. Tut.*, vol. 26, no. 3, pp. 1478–1519, 3rd Quart. 2024.
- [5] Z. Xing and J. Chen, "Constructing indoor region-based radio map without location labels," *IEEE Trans. Signal Process.*, vol. 72, pp. 2512–2526, 2024.
- [6] Z. Xing, J. Chen, and Y. Tang, "Integrated segmentation and subspace clustering for RSS-based localization under blind calibration," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, 2022, pp. 5360–5365.
- [7] Z. Xing, W. Liu, B. Li, J. Tian, M. Chu, and J. Chen, "HMM-based CSI embedding for trajectory recovery via feature engineering on MIMO-OFDM channels in LOS/NLOS regions," in *Proc. IEEE/CIC Int. Conf. Commun. China (ICCC)*, 2025, pp. 1–6.
- [8] T. Kallehauge, A. E. Kalør, P. Ramírez-Espinoza, M. Guillaud, and P. Popovski, "Delivering ultra-reliable low-latency communications via statistical radio maps," *IEEE Wireless Commun.*, vol. 30, no. 2, pp. 14–20, Apr. 2023.
- [9] D. Wu, Y. Zeng, S. Jin, and R. Zhang, "Environment-aware hybrid beamforming by leveraging channel knowledge map," *IEEE Trans. Wireless Commun.*, vol. 23, no. 5, pp. 4990–5005, May 2024.
- [10] S. Timilsina, S. Shrestha, and X. Fu, "Quantized radio map estimation using tensor and deep generative models," *IEEE Trans. Signal Process.*, vol. 72, pp. 173–189, 2024.
- [11] S. Roger, M. Brambilla, B. C. Tedeschini, C. Botella-Mascarell, M. Cobos, and M. Nicoli, "Deep-learning-based radio map reconstruction for V2X communications," *IEEE Trans. Veh. Technol.*, vol. 73, no. 3, pp. 3863–3871, Mar. 2023.
- [12] W. Liu and J. Chen, "UAV-aided radio map construction exploiting environment semantics," *IEEE Trans. Wireless Commun.*, vol. 22, no. 9, pp. 6341–6355, Sep. 2023.
- [13] W. B. Chikha, M. Masson, Z. Altman, and S. B. Jemaa, "Radio environment map based inter-cell interference coordination for massive-MIMO systems," *IEEE Trans. Mob. Comput.*, vol. 23, no. 1, pp. 785–796, Jan. 2022.
- [14] J. Wang et al., "Sparse Bayesian learning-based hierarchical construction for 3D radio environment maps incorporating channel shadowing," *IEEE Trans. Wireless Commun.*, vol. 23, no. 10, pp. 14560–14574, Oct. 2024.
- [15] Y. Wang, H. Zhao, T. Ohtsuki, H. Sari, and G. Gui, "Regularized multi-label learning empowered joint activity recognition and indoor localization with CSI fingerprints," *IEEE Trans. Wireless Commun.*, no. 11, pp. 16865–16874, Nov. 2024.
- [16] X. Zhou, L. Chen, Y. Ruan, and R. Chen, "Indoor localization with multi-beam of 5G new radio signals," *IEEE Trans. Wireless Commun.*, vol. 23, no. 9, pp. 11260–11275, Sep. 2024.
- [17] Z. Xing and W. Zhao, "Calibration-free indoor positioning via regional channel tracing," *IEEE Internet Things J.*, vol. 12, no. 5, pp. 5449–5461, Mar. 2025.
- [18] K. Vuckovic, S. Hosseini, F. Hejazi, and N. Rahnavard, "A CSI-based data-driven localization framework using small-scale training datasets in single-site MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 23, no. 11, pp. 16346–16358, Nov. 2024.
- [19] Z. Xing and W. Zhao, "Trajectory map-matching in urban road networks based on RSS measurements," *IEEE Trans. Intell. Transp. Syst.*, vol. 26, no. 4, pp. 4647–4660, Apr. 2025.
- [20] Z. Xing and J. Chen, "Constructing angular power maps in massive MIMO networks using measurements without location labels," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2025, pp. 2309–2314.
- [21] J. Guo, Y. Lv, C.-K. Wen, X. Li, and S. Jin, "Learning-based integrated CSI feedback and localization in massive MIMO," *IEEE Trans. Wireless Commun.*, no. 10, pp. 14988–15001, Oct. 2024.
- [22] K. Gao, H. Wang, H. Lv, and W. Liu, "Localization-oriented digital twinning in 6G: A new indoor-positioning paradigm and proof-of-concept," *IEEE Trans. Wireless Commun.*, vol. 23, no. 8, pp. 10473–10486, Aug. 2024.
- [23] Z. Xing and W. Zhao, "Block-diagonal guided DBSCAN clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 36, no. 11, pp. 5709–5722, Nov. 2024.
- [24] L. Zhang, X. Lei, T. Ma, H. Niu, and C. Yuen, "Joint user localization, channel estimation, and pilot optimization for RIS-ISAC," *IEEE Trans. Wireless Commun.*, vol. 23, no. 12, pp. 19302–19316, Dec. 2024.
- [25] Z. Xing and W. Zhao, "Unsupervised action segmentation via fast learning of semantically consistent actoms," in *Proc. AAAI Conf. Artif. Intell.*, 2024, pp. 6270–6278.
- [26] C. Studer, S. Medjkouh, E. Gonultas, T. Goldstein, and O. Tirkkonen, "Channel charting: Locating users within the radio environment using channel state information," *IEEE Access*, vol. 6, pp. 47682–47698, 2018.
- [27] P. Ferrand, A. Decurminge, L. G. Ordonez, and M. Guillaud, "Triplet-based wireless channel charting: Architecture and experiments," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 8, pp. 2361–2373, Aug. 2021.
- [28] P. Ferrand, M. Guillaud, C. Studer, and O. Tirkkonen, "Wireless channel charting: Theory, practice, and applications," *IEEE Commun. Mag.*, vol. 61, no. 6, pp. 124–130, Jun. 2023.
- [29] S.-h. Jung, B.-c. Moon, and D. Han, "Unsupervised learning for crowdsourced indoor localization in wireless networks," *IEEE Trans. Mob. Comput.*, vol. 15, no. 11, pp. 2892–2906, Nov. 2015.
- [30] Z. Xing and J. Chen, "HMM-based CSI embedding for trajectory recovery from RSS measurements of non-cooperative devices," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, (ICASSP), 2024, pp. 7060–7064.
- [31] J. Lota, S. Sun, T. S. Rappaport, and A. Demosthenous, "5G Uniform linear arrays with beamforming and spatial multiplexing at 28, 37, 64, and 71 GHz for outdoor urban communication: A two-level approach," *IEEE Trans. Veh. Technol.*, vol. 66, no. 11, pp. 9972–9985, Nov. 2017.
- [32] B. Liang and Z. J. Haas, "Predictive distance-based mobility management for PCS networks," in *Proc. IEEE Int. Conf. Comput. Commun. (INFOCOM)*, vol. 3, 1999, pp. 1377–1384.
- [33] R. He, B. Ai, G. L. Stüber, and Z. Zhong, "Mobility model-based non-stationary mobile-to-mobile channel modeling," *IEEE Trans. Wireless Commun.*, vol. 17, no. 7, pp. 4388–4400, Jul. 2018.
- [34] S. M. Kay, *Fundamentals of Statistical Signal Processing: estimation Theory*. Englewood Cliffs, NJ, USA: Prentice-Hall, Inc., 1993.
- [35] Z. Xing and J. Chen, "Blind construction of angular power maps in massive MIMO networks," 2025. [Online]. Available: <https://arxiv.org/abs/2510.07071>
- [36] D. M. Bates and D. G. Watts, *Nonlinear Regression Analysis and its Applications*, vol. 2. Hoboken, NJ, USA: Wiley, 1988.
- [37] M. S. Iqbal, C. F. Choudhury, P. Wang, and M. C. González, "Development of origin–destination matrices using mobile phone call data," *Transp. Res. Part C Emerg. Technol.*, vol. 40, no. 0968–090X, pp. 63–74, 2014.
- [38] K. Magow, A. Giorgetti, S. Kandeepan, and X. Yu, "Accurate analysis of weighted centroid localization," *IEEE Trans. Cogn. Commun. Netw.*, vol. 5, no. 1, pp. 153–164, Mar. 2018.
- [39] 3GPP TR 38.843, "Study on Artificial Intelligence (AI)/Machine Learning (ML) for NR Air Interface," p. 3GPP, Tech. Rep. V18.0.0, Dec. 2023.
- [40] H. Hu et al., "AMM: An adaptive online map matching algorithm," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 5, pp. 5039–5051, 2023.
- [41] Z. Liu, G. Singh, C. Xu, and D. Vasisht, "FIRE: Enabling reciprocity for FDD MIMO systems," in *Proc. 27th Annu. Int. Conf. Mobile Comput. Netw.*, 2021, pp. 628–641.



Zheng Xing (Member, IEEE) received the B.S. degree from Ocean University of China, Qingdao, China, in 2017, the M.S. degree from Beihang University, Beijing, China, in 2020, and the Ph.D. degree from The Chinese University of Hong Kong, Shenzhen, China, in 2025. He is currently an Assistant Professor with the College of Computer Science and Software Engineering, Shenzhen University. His research interests encompass optimization, mobile computing, and machine learning.



Junting Chen (Member, IEEE) received the B.Sc. degree in electronic engineering from Nanjing University, Nanjing, China, in 2009, and the Ph.D. degree in electronic and computer engineering from Hong Kong University of Science and Technology (HKUST), Hong Kong SAR China, in 2015. From 2014 to 2015, he was a visiting student with the Wireless Information and Network Sciences Laboratory, MIT, Cambridge, MA, USA. He is an Assistant Professor with the School of Science and Engineering and the Future Network of Intelligence

Institute (FNii), The Chinese University of Hong Kong, Shenzhen (CUHK–Shenzhen), Guangdong, China. Prior to joining CUHK–Shenzhen, he was a Postdoctoral Research Associate with Ming Hsieh Department of Electrical Engineering, University of Southern California (USC), Los Angeles, CA, USA, from 2016 to 2018, and with the Communication Systems Department of EURECOM, Sophia–Antipolis, France, from 2015 to 2016. His research interests include channel estimation, MIMO beamforming, machine learning, and optimization for wireless communications and localization. His current research focuses on radio map sensing, construction, and application for wireless communications. He was a recipient of the HKTIIT Post-Graduate Excellence Scholarships in 2012. He was nominated as the Exemplary Reviewer of IEEE WIRELESS COMMUNICATIONS LETTERS in 2018. His paper received the Charles Kao Best Paper Award from WOCC 2022.