

Fast Position-Aided MIMO Beam Training via Noisy Tensor Completion

Tzu-Hsuan Chou, Nicolò Michelusi, David J. Love, and James V. Krogmeier

Abstract—In this paper, a data-driven position-aided approach is proposed to reduce the training overhead in MIMO systems, by leveraging side information and on-the-field measurements. A data tensor is constructed by collecting beam-training measurements on a subset of positions and beams, and a hybrid noisy tensor completion (HNTC) algorithm is proposed to predict the received power across the coverage area, which exploits both the spatial smoothness and the low-rank property of MIMO channels. A recommendation algorithm based on the completed tensor, beam subset selection (BSS), is proposed to achieve fast and accurate beam-training. Besides, a grouping-based BSS algorithm is proposed to combat the detrimental effect of noisy positional information. Numerical results evaluated with the Quadriga channel simulator at 60 GHz millimeter-wave channels show that the proposed BSS recommendation algorithm in combination with HNTC achieve accurate received power predictions, enabling beam-alignment with small overhead: given power measurements on 40% of possible discretized positions, HNTC-based BSS attains a probability of correct alignment of 91%, with only 2% of trained beams, as opposed to a state-of-the-art position-aided beam-alignment scheme which achieves 54% correct alignment in the same configuration. Finally, an online HNTC method via warm-start is proposed, that alleviates the computational complexity by 50%, with no degradation in prediction accuracy.

Index Terms—Tensor completion, sparse learning, millimeter wave, position-aided, beam training, MIMO communication.

I. INTRODUCTION

Future wireless networks will be required to sustain high data rates, low latencies, and improved power efficiencies for users (UEs) in a wide variety of conditions [2]. In the last few years, millimeter-wave (mmWave) communication has been gaining more attention as a viable solution to address the throughput enhancement challenges faced by 5G and beyond wireless networks, thanks to the large bandwidth availability [3]–[5]. Nonetheless, mmWave communication suffers from severe pathloss due to the high frequency of the signal, hence it requires the use of narrow beam communication to compensate the signal attenuation. This is achieved with the use of large antenna arrays at the transmitter and receiver via massive multiple-input and multiple-output (MIMO) systems. However, traditional massive MIMO channel estimation methods are impractical, especially in mobile environments, because of the unacceptably large overhead induced by the large number of antennas.

The overhead of massive MIMO channel estimation, either operating at mmWave or sub-6GHz frequencies, can be

This work was supported in part by the National Science Foundation under grant CNS-1642982 and CCF-1816013.

A preliminary version of this paper was presented at the IEEE International Conference on Communications 2020 [1].

The authors are with the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, USA; emails: {chou59, michelus, djlove, jvk}@purdue.edu.

reduced by exploiting the spatial sparsity of the channel, resulting from few dominant clusters or paths in the angular domain and the high spatial resolution [6]. This sparsity allows to predesign a set of directional beams pointing in specific directions, so that beam training can be done directly on this predefined set, rather than on the MIMO channel matrix. The standard approach to beam training is to search in an exhaustive fashion through all possible combinations of transmit and receive beams, to determine the beam pair with maximum signal power. However, this approach incurs enormous overhead, due to the large number of possible beam combinations that need to be periodically trained. To reduce the overhead of exhaustive search, several beam-alignment schemes have been proposed in the literature, ranging from *feedback-based methods* [7]–[9], *AOAs/AODs estimation* [10]–[12], to *data-assisted schemes* [13]–[19]. *Feedback-based methods* adapt the beam-training procedure in an online fashion based on feedback collected. *AOAs/AODs estimation* leverages the spatial sparsity of the mmWave channel via compressed sensing techniques to recover AOAs/AODs and gains of the channel paths.

In a *data-assisted scheme*, the beam-training process is aided by side information from the available sensors other than mmWave communication, such as radar [13], LIDAR [14], lower-frequency communication [15], [16], or GPS position information [17]–[19]. Our work uses the positional information. Inverse fingerprinting for beam alignment is proposed in [17], using prior measurements at a given position to provide a set of candidate beam directions at the same position. Their work demonstrated that the positional information of the user can be utilized to predict a small set of candidate beams to train, based on long-term channel information collected in the database. The mmWave beam selection problem at specific locations is formulated in [18] as a machine learning classification problem using past beam training measurements with the situational awareness, which captures the environmental condition by encoding the obstacles' coordinates. In [19], a multiple-fingerprint beam alignment method which intelligently selects the fingerprint beam alignment based on the traffic density via learning is proposed.

However, all works on positional data-assisted beam-alignment techniques [17]–[19] fail to provide the channel information in the positions whose prior measurements are not available (new positions). For this reason, the existing approaches require the collection of an extremely large amount of channel measurements to cover a given service region, which may not be practical. Furthermore, the user's position acquired by the process of global navigation satellite system (GNSS)/global positioning system (GPS) estimation is possibly noisy due to the estimation error and the user's mobility [20], [21]. Positional estimation error degrades the perfor-

mance of position-aided data-assisted schemes. To address these general problems, we seek to do the channel prediction in new positions based on the channel information in observed positions, and propose a beam-recommendation scheme robust against errors in positional information.

Recently, channel cartography (channel charting) [22], [23] of the wireless network has been investigated. The prior channel measurements on positions provide helpful information for reducing the training overhead of the UE in the service area. The work [22] estimates the spatial loss field map from the measurements on few positions aided by a matrix completion technique which leverages the low-rank structure and sparsity. The channel gain between two points is modeled as the tomographic accumulation of the spatial loss field along the propagation path. The work [23] proposed a channel charting framework for locating users in multiple-antenna wireless networks, learning the relationship between channel state information and user's location with the tools from dimensionality reduction, manifold learning, and neural networks. In our work, we borrow ideas to learn the radio geometry by constructing a data tensor recording the average received power of beams on few positions. The average received power tensor is recovered by the proposed tensor completion, *hybrid noisy tensor completion* (HNTC), exploiting the low-rank structure of the channel sparsity and the smoothness from the spatial correlation.

A. Related Work

Real world data often exhibit various structural properties, which enable reconstruction from sparse samples [24], [25]. In the two-dimensional (2D) data structure (matrix), rank is a powerful factor capturing the global information. Low-rank matrix approximation has been intensely investigated [25]–[29], and various algorithmic approaches capable of estimating missing values have been developed. In [25], [26], the nuclear norm has been shown as the tightest convex envelope for the matrix rank function; and an efficient algorithm, singular value thresholding (SVT), was proposed. In [27], [28], the authors showed that the rank minimization problem can be solved by minimizing the nuclear norm under certain conditions, which justified the validity of using nuclear norm as the surrogate of the rank function theoretically. Low-rank matrix completion with noisy data is investigated in [29]. Smoothness [30], [31] is also an important feature aiding the matrix completion. The work [30] considered total variation as the objective function for the matrix completion problem. In [31], the authors exploited both the low-rank and smoothness simultaneously for the matrix completion.

In addition to the matrix approximation, there are works developed for tensors, which are higher dimensional extensions of matrices. Tensor completion has gained attention recently due to its multidimensional character in describing intricate datasets, with many applications. Most works developed the tensor completion approaches for image processing because the image color is composed of red/green/blue, which makes tensor a suitable structure. Recently, various models have been used for tensor completion including Tucker rank [24], [32]–[35], canonical decomposition/parallel factors (CANDE-

COMP/PARAFAC) rank (also known as CP rank) [35], [36], and total variation (TV) [34], [36], [37]. TV tensor completion takes advantage of the data smoothness, and the remaining models exploit the low-rank of the tensor. However, computing the rank of the tensor is an NP hard problem [38], so the best model for low-rank tensor approximation generally does not exist. A suitable solution for tensor completion depends on the application.

Tensor completion algorithms using both low-rank and smoothness are investigated in [34], [36]. The work [34] formulated the problem considering the low Tucker rank and data smoothness of the tensor model; and an algorithm, LRTV-PDS, was proposed using the primal-dual splitting approach. In [36], the authors proposed an algorithm, SPCTV, based on the low CP rank with the smoothness constraint. The works [34], [36] considered the low-rank property and the smoothness consistent on all dimensions. However, the properties of the tensor dimensions are based on the assumed structural properties. It is possible that the low-rank property only exists in some dimensions, and the data of the remaining dimensions are smooth. Moreover, the tensor completion problem with noisy data is investigated in [34]. In this work, we seek to devise a noisy tensor completion algorithm which exploits the low-rank and smoothness properties on separate dimensions for noisy measurement data.

B. Contribution

We develop a new tensor completion algorithm, HNTC, to enable fast and accurate MIMO beam training. To chart the MIMO channel conditions in the service area, we construct a data tensor by collecting the received power on a subset of positions and beams. The data tensor has the dimensions corresponding to the positional and beam information, respectively. Using the sparsity of MIMO channels [3], [4], [6], the data tensor exhibits the low-rank property in the beam dimensions. On the other hand, the spatial correlation of the channel induces smoothness across the position dimension of the data tensor. To account for noise in the received signal and random fluctuations in the channel, we formulate the data recovery as a noisy tensor completion problem considering locally low-rank and spatial smoothness in distinct dimensions, which is a convex optimization problem with a noise inequality constraint.

The contributions of the paper are detailed as follow:

- We propose a tensor completion algorithm, HNTC, that uses the low-rank and smoothness properties across the different dimensions in the data to reconstruct channel properties from noisy measurements.
- We develop an online version of HNTC using warm-start.
- We propose a position-aided beam recommendation algorithm for fast and accurate MIMO beam-training, beam subset selection (BSS), which uses the tensor completed via HNTC to recommend a small subset of beams to train.
- We propose a grouping-based BSS (G-BSS) algorithm to combat the effect of errors in positional information.

The rest of the paper is organized as follows. Section II reviews the preliminaries on tensor completion. Section III introduces a data-driven approach for learning the channel

conditions, which motivates a new noisy tensor completion problem. Section IV-A proposes the algorithm HNTC aided by alternating direction method of multipliers (ADMM) [39], and Section IV-B introduces online HNTC. Section IV-C discusses the computational complexity. Section V evaluates the performance of HNTC and position-aided beam-alignment. Section VI concludes the paper.

II. PRELIMINARIES ON TENSOR COMPLETION

Tensors are commonly used in many areas of engineering and science [40], but they have received only limited interest in communication theory. In the following, we introduce the tensor notation and terminology used throughout the rest of the paper.

A. Notations

Bold uppercase letters \mathbf{X} denote matrices, and calligraphic letters \mathcal{X} represent tensors. $\mathbf{X} \otimes \mathbf{Y}$ is the Kronecker product of \mathbf{X} and \mathbf{Y} . An M -th order tensor is defined as $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_M}$, with M being the number of dimensions (a matrix can be interpreted as a second order tensor). Given an M -th order tensor \mathcal{X} , we denote its $\mathbf{i}_{1:M} = (i_1, i_2, \dots, i_M)$ -th element as $\mathcal{X}(\mathbf{i}_{1:M}) = \mathcal{X}(i_1, i_2, \dots, i_M)$. The inner product of two tensors $\mathcal{X}, \mathcal{Y} \in \mathbb{R}^{I_1 \times \dots \times I_M}$ is defined as

$$\langle \mathcal{X}, \mathcal{Y} \rangle \triangleq \sum_{\mathbf{i}_{1:M}} \mathcal{X}(\mathbf{i}_{1:M}) \mathcal{Y}(\mathbf{i}_{1:M})$$

and the Frobenius norm of \mathcal{X} is denoted by $\|\mathcal{X}\|_F \triangleq \sqrt{\langle \mathcal{X}, \mathcal{X} \rangle}$. The mode- m unfolding (or the mode- m matricization) of a tensor \mathcal{X} is the matrix $\mathcal{X}_{(m)} \in \mathbb{R}^{I_m \times J_{M+1}}$ with entries $\mathcal{X}_{(m)}(i_m, j) = \mathcal{X}(\mathbf{i}_{1:M})$, $\forall \mathbf{i}_{1:M}$, where

$$j = 1 + \sum_{k=1, k \neq m}^M (i_k - 1) J_k,$$

with $J_k = \prod_{d=1, d \neq m}^{k-1} I_d$. The inverse operation of folding recovers the original tensor from its unfolded representation, and is denoted as $\text{fold}_m(\mathcal{X}_{(m)}) \triangleq \mathcal{X}$.

B. Preliminaries

The low-rank tensor completion is a natural extension of low-rank matrix completion to more than two dimensions. Given an M -th order incomplete tensor \mathcal{T} with Ψ as the indices of known elements, the low-rank tensor completion problem is formulated as

$$\min_{\mathcal{X}} \text{rank}(\mathcal{X}) \text{ s.t. } \mathcal{X}_{\Psi} = \mathcal{T}_{\Psi}, \quad (1)$$

where $\text{rank}(\mathcal{X})$ is the smallest number of rank-one tensors needed to sum together to construct \mathcal{X} (also known as CP rank), which is an analogue to the definition of matrix rank [40]. Here, a tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times \dots \times I_M}$ is rank-one if it can be expressed as the outer product of M vectors, $\mathbf{a}_d \in \mathbb{R}^{I_d \times 1}$, $d = 1, \dots, M$, so that $\mathcal{X}(i_1, i_2, \dots, i_M) = \prod_{d=1}^M \mathbf{a}_d(i_d)$ [40].

However, finding the rank of a specific tensor is an NP-hard problem [38], and the best rank approximation may not exist [41]. For this reason, alternative definitions of rank have been used in the literature. A well developed definition is the Tucker rank [40], defined as the rank of the unfolding matrices,

$\text{rank}(\mathcal{X}_{(n)})$, $n = 1, \dots, M$. The work [24] defined the tensor nuclear norm as a weighted sum of the Tucker ranks of all the unfolding matrices:

$$\|\mathcal{X}\|_* = \sum_{k=1}^M \alpha_k \|\mathcal{X}_{(k)}\|_*, \quad (2)$$

where the constant coefficients $\alpha_k > 0$ and $\sum_{k=1}^M \alpha_k = 1$; and then proposed the low-rank tensor completion problem as

$$\min_{\mathcal{X}} \sum_{k=1}^M \alpha_k \|\mathcal{X}_{(k)}\|_* \text{ s.t. } \mathcal{X}_{\Psi} = \mathcal{T}_{\Psi}. \quad (3)$$

The high accuracy low-rank tensor completion (HaLRTC) method [24] was proposed to solve (3) via ADMM.

Smoothness has been considered a useful property to aid matrix completion [31], especially with high ratio of missing elements. In [31], the authors proposed a combination of low-rank and smoothness minimization for the matrix completion, termed the *linear total variation approximate regularized nuclear norm minimization* problem (LTVNN). Given an incomplete matrix \mathbf{M} with Ω as the indices of known elements, the completion problem is formulated as

$$\min_{\mathbf{X}} \|\mathbf{X}\|_* + \gamma \|\mathbf{X}\|_{LTV}, \text{ s.t. } \mathbf{X}_{\Omega} = \mathbf{M}_{\Omega}, \quad (4)$$

where γ is the trade-off factor between low-rank and smoothness, and $\|\mathbf{X}\|_{LTV}$ is the linear total variation, defined as

$$\|\mathbf{X}\|_{LTV} = \sum_{i,j} \delta_1^2(\mathbf{X}(i, j)) + \delta_2^2(\mathbf{X}(i, j)), \quad (5)$$

where $\delta_1(\mathbf{X}(i, j)) \equiv \mathbf{X}(i+1, j) - \mathbf{X}(i, j)$ and $\delta_2(\mathbf{X}(i, j)) \equiv \mathbf{X}(i, j+1) - \mathbf{X}(i, j)$. This optimization problem (4) can be solved with the ADMM method [39].

For the tensor completion, there were previous works [34], [36] considering both low-rank and smoothness minimization. The work [34] proposed the algorithm considering low Tucker rank model with data smoothness, called LRTV-PDS. In [36], the tensor completion is based on the low CP rank with the smoothness on the CP components. However, the low-rank property and the smoothness are not always consistent through all dimensions. In this work, we seek to estimate the missing entries of a tensor which exhibits the low-rank property in some dimensions, and the smoothness property in the remaining dimensions.

III. MOTIVATING EXAMPLE

In this section, we provide a wireless channel learning framework which motivates a noisy tensor completion problem. The objective is to provide an approach to portray the map-based wireless channel conditions over the service area. In Section III-A, we introduce the map-based channel model [42]. In Section III-B, we introduce the cloud-based position-aided beam recommendation approach. In Section III-C, we explain the collection of channel measurements, which are post-processed and recorded along with the side information, constructing a data tensor. In Section III-D, we describe our proposed cloud-based approach for MIMO beam training. We further investigate the inherent properties of the data

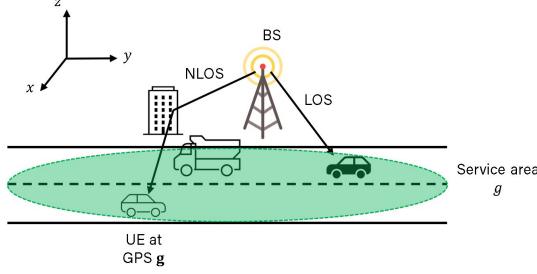


Fig. 1. Network layout of the considered scenario. A BS serves the mobile UE at GPS g in the service area \mathcal{G} .

measurements, which motivate the design of our proposed tensor completion problem, HNTC.

A. Wireless System Model and Problem Description

We consider a wireless network supporting a geographic region, as depicted in Fig. 1. A base station (BS) serves mobile UEs in a service area $\mathcal{G} \subset \mathbb{R}^2$, modeled as a compact set. We assume the BS is at a fixed position and height, and employs an antenna array with M_r antennas with fixed orientation. We consider a reference UE, whose position at time k is denoted as $\mathbf{g}(k) \in \mathcal{G}$, which employs an antenna array with M_t antennas, with time-varying orientation.

We consider an uplink system to be suitable for data collection since we expect to collect and exploit the measured data from the UE at possible positions in the service area. Due to the channel reciprocity, the collected data can also operate in the downlink; in this case, the signal strength received at the UE needs to be fed back to the BS. The UE transmits an $M_t \times 1$ signal vector $\mathbf{x}[k]$ at timeslot k , yielding the $M_r \times 1$ received signal vector

$$\mathbf{y}[k] = \mathbf{H}[k]\mathbf{x}[k] + \mathbf{n}[k], \quad (6)$$

where $\mathbf{H}[k]$ is an $M_r \times M_t$ narrowband channel matrix of the UE experienced at time k , and $\mathbf{n}[k] \sim \mathcal{CN}(\mathbf{0}, \sigma_n^2 \mathbf{I})$ is an $M_r \times 1$ noise vector. Generally, the channel $\mathbf{H}[k]$ depends on the positions of the BS and UE, the antenna setup (orientation and design) of the BS and UE, and the scattering clusters in the environment and may vary over time as a result of dynamics in the propagation environment, such as fading and mobility of clusters, mobility of the UE and changes in the orientation of the UE's antenna array. However, it exhibits some patterns which depend on the UE position $\mathbf{g}(k)$. In this paper, we are interested in developing a data-aided framework that exploits these position-dependent patterns to aid the beam-training procedure.

The BS receives the signal with the unit-norm combining vector $\mathbf{w} \in \mathcal{W}$ taking values from the BS codebook \mathcal{W} , the UE employs a unit-norm beamforming vector $\mathbf{f} \in \mathcal{F}$ taking values from the UE codebook \mathcal{F} . The transmit signal can then be expressed as $\sqrt{P_t}\mathbf{f} \cdot \mathbf{s}$, where P_t is the transmit power, and $\mathbf{s} = [s_1, \dots, s_Q]$ is the known training sequence vector with norm \sqrt{Q} . After signal combining at the receiver with the unit-norm combining vector $\mathbf{w} \in \mathcal{W}$, the received signal vector can be expressed as

$$\tilde{\mathbf{y}}[k] = \sqrt{P_t}\mathbf{w}^H\mathbf{H}[k]\mathbf{f} \cdot \mathbf{s} + \tilde{\mathbf{n}}[k], \quad (7)$$

where the received noise vector $\tilde{\mathbf{n}}[k] \sim \mathcal{CN}(0, \sigma_n^2 \mathbf{I})$. The received signal power can be expressed as

$$r[k] = |\tilde{\mathbf{y}}[k]\mathbf{s}^H|^2 = |\sqrt{P_t}\mathbf{w}^H\mathbf{H}[k]\mathbf{f} \cdot \mathbf{s}^H + \tilde{\mathbf{n}}[k]|^2. \quad (8)$$

where $\tilde{\mathbf{n}}[k] = \tilde{\mathbf{n}}[k]\mathbf{s}^H$ is a zero-mean complex Gaussian noise with variance $\sigma_n^2 \mathbf{s}^H$.

We assume the use of uniform planar arrays (UPAs) [4] at the BS in this work. However, the framework presented in this paper can be applied to any codebook design, not just UPAs. The UPA codebook can be described by the parameters $(C_y, C_z, C_\theta, C_\phi)$, where (C_y, C_z) are the number of antennas in the y and z directions with half wavelength antenna spacing, and (C_θ, C_ϕ) are the number of quantized beams along the elevation and azimuth angular directions. The array response vector representing a beam pointing in the elevation angle $\theta \in [-\pi/2, \pi/2]$ and the azimuth angle $\phi \in [-\pi/2, \pi/2]$ is denoted as

$$\mathbf{a}(\theta, \phi) = \frac{1}{\sqrt{C_y C_z}} [1 \ e^{j\Omega_z} \ \dots \ e^{j(C_z-1)\Omega_z}]^T \otimes [1 \ e^{j\Omega_y} \ \dots \ e^{j(C_y-1)\Omega_y}]^T, \quad (9)$$

with $\Omega_z = \pi \sin \theta \sin \phi$, $\Omega_y = \pi \sin \theta \cos \phi$. To construct the UPA codebook, θ_u and ϕ_v are uniformly quantized in $[-\pi/2, \pi/2]$ with resolution π/C_θ and π/C_ϕ as

$$\theta_u = -\frac{\pi}{2} + (u-1) \times \frac{\pi}{C_\theta}, \quad u = 1, \dots, C_\theta, \quad (10)$$

$$\phi_v = -\frac{\pi}{2} + (v-1) \times \frac{\pi}{C_\phi}, \quad v = 1, \dots, C_\phi. \quad (11)$$

We index the beamforming vectors in the codebook as

$$\mathcal{K} \equiv \{(u, v) : u = 1, \dots, C_\theta, v = 1, \dots, C_\phi\},$$

and we denote the beamforming vector $\mathbf{w}_{u,v}$ indexed by $(u, v) \in \mathcal{K}$.

In this work, we aim to provide a channel learning framework covering the service area in the wireless network. With the framework, we are able to provide a set of candidate beams to do the channel estimation for the UE at any possible position. To the best of our knowledge, the map-based channel model can only be acquired by real channel measurements, or by simulation via ray-tracing software which simulates the propagation environment. To attain our objective upon these facts, we should do the channel measurements or ray-tracing on every possible positions, as assumed in the state-of-the-art [17], which is impractical in a real system due to the prohibitively large overhead. To address the challenge, we propose an efficient data-driven approach to learn the channel conditions of the whole service area with limited number of channel measurements.

B. Cloud-Based Position-Aided Beam Alignment

The idea of this approach is to provide a set of candidate beams at a give UE position. Since the overhead of the conventional beam-sweeping approach is unacceptable (it scales up with $|\mathcal{W}| \cdot |\mathcal{F}|$ and is typically very large), our objective is to design a learning algorithm that recommends a small subset $\mathcal{S} \subset \mathcal{W}$ of beams to train at the BS, which is likely to contain the best BS beam (the one with the largest received

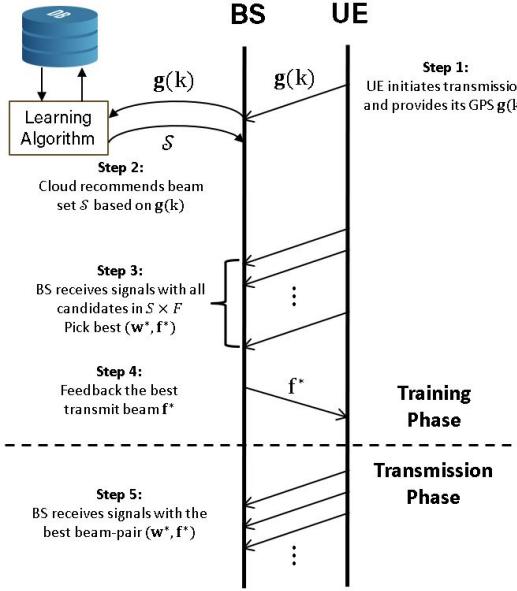


Fig. 2. Cloud-based position-aided beam alignment protocol.

signal power). Thus, the training overhead can be reduced to $|\mathcal{S}| \cdot |\mathcal{F}|$.

In Fig. 2, we introduce the flow diagram of the cloud-based position-aided beam alignment. In Step 1, the UE initiates the uplink transmission request with its current GPS coordinate $g(k)$ to the BS using sub-6 GHz control channels. This information is available via a suite of sensors such as GPS [20], [21]. In Step 2, the BS forwards the GPS coordinate $g(k)$ to the cloud, which processes the learning algorithm to provide the recommended beam set \mathcal{S} to train at the BS. Note that, in order to provide robustness against different UE antenna designs (such as UPAs with different number of antennas and configurations) and temporal dynamics in UE antenna orientation, our proposed algorithm recommends the beam-training set \mathcal{S} only on the BS side (whose antenna has a fixed position and orientation), but does not recommend a beam-training set on the UE side, so that the UE is required to scan all possible beamforming vectors in \mathcal{F} to acquire the best one to be used for data communication. Feedback-based beam-training schemes such as [7]–[9], or schemes that leverage the mobility of the UE such as [43], can be used to further reduce the training overhead, but the analysis of this case is out of the scope of this paper. In Step 3, the UE transmits a sequence of $|\mathcal{S}| \cdot |\mathcal{F}|$ known signals, where the BS receives the signals with the beam pairs $\{(\mathbf{w}, \mathbf{f}) : \mathbf{w} \in \mathcal{S}, \mathbf{f} \in \mathcal{F}\}$. Then, the BS selects the best $(\mathbf{w}^*, \mathbf{f}^*)$ among all candidate beam pairs ranked by the received power. In Step 4, the BS feeds back the selected UE beamforming vector \mathbf{f}^* to the UE. In Step 5, the UE communicates with the BS using the selected beam pair $(\mathbf{w}^*, \mathbf{f}^*)$ for the subsequent uplink or downlink data transmission.

C. Data Model

Here, we describe how we collect the channel measurements, and store the information in the database. Let $\bar{\mathcal{G}} = [X_0, X_{end}] \times [Y_0, Y_{end}]$ be a rectangular region containing the service area \mathcal{G} . We discretize $\bar{\mathcal{G}}$ with resolution Δ_s , thus

defining the discrete GPS coordinates $\mathbf{g} = (g_x, g_y) \in \bar{\mathcal{G}}$. Then, we define the position labels $\mathbf{p} = (p_x, p_y)$ with $p_x \in \{1, \dots, L_x\}$ and $p_y \in \{1, \dots, L_y\}$, where $L_x = \lceil \frac{X_{end}-X_0}{\Delta_s} \rceil$ and $L_y = \lceil \frac{Y_{end}-Y_0}{\Delta_s} \rceil$ are the number of points in the x and y axes, and $\lceil x \rceil$ is the ceiling function. The function $\rho(\mathbf{g})$ maps the coordinate $\mathbf{g} \in \bar{\mathcal{G}}$ to the closest discretized position label \mathbf{p} as

$$\mathbf{p} = \rho(\mathbf{g}) = \left(1 + \left\lceil \frac{g_x - X_0}{\Delta_s} \right\rceil, 1 + \left\lceil \frac{g_y - Y_0}{\Delta_s} \right\rceil \right), \quad (12)$$

where $\lfloor x \rfloor$ denotes the nearest integer to x .

During the data collection, the BS measures the received power, using the UE's current position and beam as side information. To do so, the BS receives the signal using an arbitrary combining vector $\mathbf{w}_{u,v} \in \mathcal{W}$ indexed by $(u, v) \in \mathcal{K}$, whereas the UE performs the beam-training exhaustively over the UE beamforming set \mathcal{F} ; with the UE in the discretized position $\mathbf{p} = \rho(\mathbf{g}(k))$, the strongest received signal power is then denoted as

$$r_k^{(\mathbf{p}, u, v)} = \max_{\mathbf{f}_i \in \mathcal{F}} \sum_{i=0, \dots, |\mathcal{F}|-1} |\sqrt{P_i} \mathbf{w}_{u,v}^H \mathbf{H}[k+i] \mathbf{f}_i + \hat{n}[k+i]|^2, \quad (13)$$

where $\mathbf{H}[k]$ is the uplink MIMO channel of the UE at timeslot k , and $\hat{n}[k]$ is the received noise at timeslot k . We assume that, during this process, the discretized UE position does not change, i.e. $\rho(\mathbf{g}(k+i))$ remains fixed $\forall i = 0, \dots, |\mathcal{F}|-1$. The received signal power $r_k^{(\mathbf{p}, u, v)}$ at timeslot k , along with the side information \mathbf{p} (UE discretized position) and (u, v) (BS beamforming index), is then recorded in the database, with the procedure described below.

The BS might collect multiple measurements on a given combination of position \mathbf{p} and combining vector $\mathbf{w}_{u,v}$. Therefore, the cloud database stores an average of the measurements collected. Moreover, to enable adaptation in non-stationary environments, it computes a weighted average which discounts past measurements. To this end, let $\chi_k^{(\mathbf{p}, u, v)} \in \{0, 1\}$ be an indicator variable, equal to one if and only if a measurement is collected at time k in the discretized position \mathbf{p} on the combining vector $\mathbf{w}_{u,v}$. Let $\bar{r}_k^{(\mathbf{p}, u, v)}$ be the corresponding weighted received power stored in the database at time k (initially, $\bar{r}_0^{(\mathbf{p}, u, v)} = 0$). This is computed as

$$\bar{r}_k^{(\mathbf{p}, u, v)} = \frac{\sum_{\ell=0}^k \alpha^\ell \chi_{k-\ell}^{(\mathbf{p}, u, v)} r_{k-\ell}^{(\mathbf{p}, u, v)}}{\sum_{\ell=0}^k \alpha^\ell \chi_{k-\ell}^{(\mathbf{p}, u, v)}}, \quad (14)$$

where $\alpha \in (0, 1]$ is a discount factor which enables adaptation to non-stationary environments. In the stationary-case, one can set $\alpha = 1$ and (14) yields the sample average of past measurements. In addition to $\bar{r}_k^{(\mathbf{p}, u, v)}$, the database stores also the weighted number of past measurements collected on a given position and beam, denoted as $\bar{N}_k^{(\mathbf{p}, u, v)}$ and given by

$$\bar{N}_k^{(\mathbf{p}, u, v)} = \sum_{\ell=0}^k \alpha^\ell \chi_{k-\ell}^{(\mathbf{p}, u, v)}, \quad (15)$$

which yields the total number of measurements collected up to time k in the stationary case ($\alpha = 1$). In fact, the average

received power $\bar{r}_k^{(\mathbf{p}, u, v)}$ is a noisy estimate of the expected signal power, and the variance of this measurement typically decreases with the number of measurements. In case of N i.i.d. measurements, the variance decreases by a factor $1/N$. Therefore, the term $\bar{N}_k^{(\mathbf{p}, u, v)}$ will be used to track the accuracy of the received power estimates stored in the database, and will be used in the noisy tensor completion algorithm developed in the next section.

Note that (14) and (15) can be computed in an online fashion, without requiring to store all past measurements. In fact, upon collecting a new measurement $r_{k+1}^{(\mathbf{p}, u, v)}$ in position \mathbf{p} using the combiner $\mathbf{w}_{u, v}$ at timeslot $k+1$, (14) and (15) can be updated as

$$\bar{N}_{k+1}^{(\mathbf{p}, u, v)} = \alpha \bar{N}_k^{(\mathbf{p}, u, v)} + \chi_{k+1}^{(\mathbf{p}, u, v)}, \quad (16)$$

$$\bar{r}_{k+1}^{(\mathbf{p}, u, v)} = \bar{r}_k^{(\mathbf{p}, u, v)} + \frac{\chi_{k+1}^{(\mathbf{p}, u, v)}}{\bar{N}_{k+1}^{(\mathbf{p}, u, v)}} (r_{k+1}^{(\mathbf{p}, u, v)} - \bar{r}_k^{(\mathbf{p}, u, v)}), \quad (17)$$

which are updated in the database along with the side information, including the UE position \mathbf{p} and the index of BS combining vector (u, v) , as in TABLE I.

TABLE I
DATABASE FORM

p_x	p_y	u	v	$\bar{r}^{(\mathbf{p}, u, v)}$	$\bar{N}^{(\mathbf{p}, u, v)}$
1	1	1	4	5.2	3
1	2	4	5	6.1	1
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots

We represent the average receive power as a fourth order tensor in $\mathbb{R}^{L_x \times L_y \times C_\theta^r \times C_\phi^r}$:

$$\mathcal{T}(\mathbf{p}, u, v) = \begin{cases} \bar{r}^{(\mathbf{p}, u, v)}, & (\mathbf{p}, u, v) \in \Psi, \\ 0, & \text{otherwise,} \end{cases} \quad (18)$$

where Ψ is the set of observed combinations of positions and BS codewords stored in the database and the unobserved entries $(\mathbf{p}, u, v) \notin \Psi$ are set to zero. Moreover, the weighted number of measurements for all positions and BS codewords are recorded as a fourth order tensor in $\mathbb{R}^{L_x \times L_y \times C_\theta^r \times C_\phi^r}$:

$$\mathcal{V}(\mathbf{p}, u, v) = \bar{N}^{(\mathbf{p}, u, v)}, \quad (19)$$

which can be updated in an online fashion as new measurements are collected, as described previously. It is impractical to collect the measurements with all combinations of positions and BS codewords into the database due to the limited sampling resource and non-stationarity of the propagation environment which causes past measurements to become outdated. Therefore, some positions possibly have no representation in the database. Even in the observed positions, there might be only a limited number of BS codewords' information recorded. Thus, the data tensor \mathcal{T} may be highly incomplete.

D. What do we do with the data?

Our goal is to recommend a set of N_{tr} candidate beams for the BS to train based on UE position. If the UE is in a position \mathbf{p} represented in the database, and the measurements with all beams available, then this task can be easily accomplished by recommending the N_{tr} beams with highest average received

power in the given position. Otherwise, we are required to design an algorithm based on tensor completion that employs the knowledge of limited number of beams' information at UE's neighboring positions to provide the beam recommendation.

To address the issue, given the data tensor \mathcal{T} (18) and the measurement tensor \mathcal{V} (19), we formulate the noisy tensor completion problem as

$$\arg \min_{\mathcal{X}} \sum_{p_x, p_y} \|\mathcal{X}(p_x, p_y, :, :) \|_* + \gamma \sum_{u, v} S(\mathcal{X}(:, :, u, v)) \quad (20)$$

$$\text{s.t. } \sum_{\psi} \mathcal{W}(\psi) (\mathcal{T}(\psi) - \mathcal{X}(\psi))^2 \leq \eta, \quad (21)$$

where γ is the trade-off parameter between the smoothness property across the positional dimensions and the low-rank property across the beam dimensions. The smoothness function $S(\mathcal{X})$ measures the difference between the values of adjacent elements in \mathcal{X} . The minimizer tensor \mathcal{X}^* is the reconstructed tensor. The tensor \mathcal{W} is the weighting tensor, with elements $\mathcal{W}(\psi) = \mathcal{V}(\psi) / \sum_{\psi'} \mathcal{V}(\psi')$ associated to the entry ψ . For an unobserved element ψ , $\mathcal{W}(\psi) = 0$, so that the corresponding entry has no contribution to the error term (21). The inequality constraint (21) originates from the noisy measurements, and induces an upper bound η on the weighted sum of the mean square error between \mathcal{X} and \mathcal{T} , which provides additional flexibility to the observed elements. The noiseless tensor completion is a special case when $\eta = 0$. Note that tensor entries associated to more measurements (larger $\mathcal{V}(\psi)$) are associated a larger weight ($\mathcal{W}(\psi)$), so they contribute more to the squared error term (21). In fact, this is a desirable feature since the variance of the sample averaged receive power decreases with the number of measurements.

The first term of the objective function (20) represents the low-rank property of the measurement data. For each position, the channel sparsity of mmWave propagation dictates that only few combinations of beams lead to significant signal measurements, which imposes the low-rank property to the collected measurement data tensor in the beam dimensions. However, the tensor completion with low-rank alone often fails to provide predictions on unobserved positions due to the dependence on the sampling set, especially in the case with high ratio of missing elements. The smoothness can be considered to support the data recovery. Regarding the second term of the objective function (20), it quantifies the smoothness of the measurement data. The spatial coherence of the channels in adjacent positions can be exploited to learn the channel conditions. The received power measured using a certain beam exhibits spatial correlation, so that the corresponding data tensor exhibits smoothness across the positional dimensions.

The low-rank property across the beam dimensions and the smoothness property across the positional dimensions will be exploited to develop the HNTC algorithm in the next section.

IV. HYBRID NOISY TENSOR COMPLETION

In the previous section, we introduced a data-driven approach to capture the channel conditions by constructing a data tensor as in (18), containing the position and BS codeword as the side information. In a wireless network, the UEs are possibly equipped with different antenna designs. To enable

our proposed approach on a variety of UE antenna designs, we construct the data tensor by collecting the received signal power under the best UE beam. Besides, the channel is time-varying in the propagation environment, so we record the weighted average received power in the data tensor to learn some structures preserved over time. However, due to the limited sampling resources, the collected data tensor might be highly incomplete, so it is crucial to devise an algorithm for the tensor completion problem that exploits both low-rank and smoothness properties across separate dimensions. The noise effect is also an important factor to be considered since the signal in wireless systems is inevitably contaminated.

Section IV-A formulates a noisy tensor completion problem, and then proposes the algorithm, HNTC, to solve the tensor completion problem using the ADMM method. Section IV-B proposes the online HNTC, and Section IV-C discusses the computational complexity.

A. Proposed Noisy Tensor Completion

We propose a completion problem for recovering an incomplete data tensor, using the assumption that a portion of the dimensions have the smoothness property and the remaining dimensions are low-rank. As motivated in the previous section, we consider a tensor $\mathcal{T} \in \mathbb{R}^{I_1^s \times \dots \times I_{n_1}^s \times I_1^\ell \times \dots \times I_{n_2}^\ell}$ which separates its dimensions into two partitions. The first n_1 dimensions are smooth and the last n_2 dimensions have the low-rank property. It means that the n_1 -th order subtensor obtained by fixing the indices of the last n_2 dimensions, $\mathcal{T}(:, \dots, :, \mathbf{i}_{1:n_2}) \in \mathbb{R}^{I_1^s \times \dots \times I_{n_1}^s}$, is a smooth tensor; and the n_2 -th order subtensor obtained by fixing the indices of the first n_1 dimensions, $\mathcal{T}(\mathbf{i}_{1:n_1}, :, \dots, :) \in \mathbb{R}^{I_1^\ell \times \dots \times I_{n_2}^\ell}$, is low-rank. We use the following smoothness metric.

Definition 1. The linear tensor total variation (LTTV) of the tensor $\mathcal{B} \in \mathbb{R}^{I_1 \times \dots \times I_M}$, quantifying the smoothness of the tensor \mathcal{B} , is defined as

$$\text{LTTV}(\mathcal{B}) = \sum_{\mathbf{i}_{1:m}} \sum_{d=1}^M [\delta_d(\mathcal{B}(\mathbf{i}_{1:m}))]^2,$$

where $\delta_d(\mathcal{B}(\mathbf{i}_{1:M})) = \mathcal{B}(\mathbf{i}_{1:M} + \mathbf{e}_d) - \mathcal{B}(\mathbf{i}_{1:M})$, \mathbf{e}_d is the vector of zeros except in the d -th position in which it is equal to 1.

Note that $\text{LTTV}(\mathcal{B})$ is a natural extension of linear total variation (LTV) [31] to tensors, which accumulates the element squared differences of \mathcal{B} along every dimension.

Assuming an incomplete tensor \mathcal{T} and the associated tensor \mathcal{V} representing the number of measurements collected on each entry, the goal is to estimate the missing elements in \mathcal{T} by solving

$$\begin{aligned} \arg \min_{\mathcal{X}} & \sum_{\mathbf{i}_{1:n_1}} \sum_{k=1}^{n_2} \alpha_k \left\| \{\mathcal{X}(\mathbf{i}_{1:n_1}, :, \dots, :) \}_{(k)} \right\|_* \\ & + \gamma \sum_{\mathbf{i}_{1:n_2}} \text{LTTV}(\mathcal{X}(:, \dots, :, \mathbf{i}_{1:n_2})) \\ \text{s.t. } & \sum_{\psi} \mathcal{W}(\psi) (\mathcal{T}(\psi) - \mathcal{X}(\psi))^2 \leq \eta, \end{aligned} \quad (23)$$

where γ is the trade-off parameter between the smoothness property across the first n_1 dimensions and the low-rank property across the remaining n_2 dimensions, and the constraint (23) accounts for the noise in the tensor \mathcal{T} . Note that this formulation employs the tensor nuclear norm defined in (2), and the term $\{\mathcal{X}(\mathbf{i}_{1:n_1}, :, \dots, :) \}_{(k)}$ is the mode- k unfolding of $\mathcal{X}(\mathbf{i}_{1:n_1}, :, \dots, :) \in \mathbb{R}^{I_1^\ell \times \dots \times I_{n_2}^\ell}$. The tensor \mathcal{W} is the weighting tensor, with elements $\mathcal{W}(\psi) = \mathcal{V}(\psi) / \sum_{\psi'} \mathcal{V}(\psi')$ associated to the entry ψ . The minimizer tensor \mathcal{X}^* is the reconstructed tensor. The proposed HNTC optimization problem is convex, since it is a linear combination of the tensor nuclear norm (convex, see (3)) and the LTTV (quadratic and convex, see **Definition 1**), and the inequality constraint is also a convex quadratic function. The problem (22) is a general form of the tensor completion problem in (20), where the positional information are associated with the smooth dimensions, and the beam information are associated with the low-rank dimensions. Note that the proposed HNTC problem is a generalization of other state-of-the-art tensor completion problems. By setting $n_1 = 0$, we obtain the low-rank tensor completion [24]; by setting $n_2 = 0$, we obtain the linear tensor total variation minimization of [31].

We use the ADMM technique [39] to solve (22). The optimization problem is reformulated as

$$\arg \min_{\mathcal{X}, \{\mathcal{Y}_k\}_{k=1}^{n_2}} \sum_{\mathbf{i}_{1:n_1}} \sum_{k=1}^{n_2} \alpha_k \left\| \{\mathcal{Y}_k(\mathbf{i}_{1:n_1}, :, \dots, :) \}_{(k)} \right\|_* \quad (24)$$

$$\begin{aligned} & + \gamma \sum_{\mathbf{i}_{1:n_2}} \text{LTTV}(\mathcal{X}(:, \dots, :, \mathbf{i}_{1:n_2})) + \sum_{k=1}^{n_2} \frac{\lambda}{2} \|\mathcal{Y}_k - \mathcal{X}\|_F^2 \\ \text{s.t. } & \sum_{\psi} \mathcal{W}(\psi) (\mathcal{T}(\psi) - \mathcal{X}(\psi))^2 \leq \eta, \end{aligned} \quad (25)$$

$$\mathcal{X} = \mathcal{Y}_k, \quad k = 1, \dots, n_2, \quad (26)$$

where λ is a small fixed positive parameter. With ADMM, we introduce the variables $\{\mathcal{Y}_k\}_{k=1}^{n_2}$ to separate the smooth and low-rank dimensions of the tensor. The additional equality constraints $\mathcal{X} = \mathcal{Y}_k$, $k = 1, \dots, n_2$ guarantee that (24) is equivalent to the problem (22).

We introduce the Lagrangian multiplier \mathcal{Z}_k associated with the k -th equality constraint (26), and μ associated with the inequality constraint (25). The corresponding augmented Lagrangian function is expressed as

$$\begin{aligned} & L\left(\mathcal{X}, \{\mathcal{Y}_k\}_{k=1}^{n_2}, \{\mathcal{Z}_k\}_{k=1}^{n_2}, \mu\right) \\ & = \sum_{k=1}^{n_2} \left\{ \sum_{\mathbf{i}_{1:n_1}} \alpha_k \left\| \{\mathcal{Y}_k(\mathbf{i}_{1:n_1}, :, \dots, :) \}_{(k)} \right\|_* \right. \\ & \quad \left. + \left\langle \mathcal{Z}_k, \mathcal{Y}_k - \mathcal{X} \right\rangle + \frac{\lambda}{2} \|\mathcal{Y}_k - \mathcal{X}\|_F^2 \right\} \\ & + \gamma \sum_{\mathbf{i}_{1:n_2}} \text{LTTV}(\mathcal{X}(:, \dots, :, \mathbf{i}_{1:n_2})) \\ & + \mu \left(\sum_{\psi} \mathcal{W}(\psi) (\mathcal{T}(\psi) - \mathcal{X}(\psi))^2 - \eta \right). \end{aligned} \quad (27)$$

The ADMM algorithm is implemented by minimizing L over $\mathcal{X}, \mathcal{Y}_k$ and then updating the Lagrangian multipliers \mathcal{Z}_k and μ in an iterative fashion as

$$\mathcal{X}_{t+1} = \arg \min_{\mathcal{X}} L(\mathcal{X}, \{\mathcal{Y}_{k,t}\}_{k=1}^{n_2}, \{\mathcal{Z}_{k,t}\}_{k=1}^{n_2}, \mu_t); \quad (28)$$

$$\begin{aligned} \mathcal{Y}_{k,t+1} &= \arg \min_{\mathcal{Y}_k} L(\mathcal{X}_{t+1}, \{\mathcal{Y}_k\}_{k=1}^{n_2}, \{\mathcal{Z}_{k,t}\}_{k=1}^{n_2}, \mu_t), \\ \forall k &= 1, \dots, n_2; \end{aligned} \quad (29)$$

$$\mathcal{Z}_{k,t+1} = \mathcal{Z}_{k,t} + \beta_1 (\mathcal{Y}_{k,t+1} - \mathcal{X}_{t+1}), \quad \forall k = 1, \dots, n_2; \quad (30)$$

$$\mu_{t+1} = \left(\mu_t + \beta_2 \left(\sum_{\psi} \mathcal{W}(\psi) (\mathcal{T}(\psi) - \mathcal{X}_{t+1}(\psi))^2 - \eta \right) \right)^+; \quad (31)$$

where β_1 is the step-size for updating \mathcal{Z}_k , β_2 is the step-size for updating μ , and $(\mu)^+ = \max(\mu, 0)$ is the projection of μ onto \mathbb{R}_+ .

To optimize \mathcal{X} , we minimize L with fixed $\{\mathcal{Y}_{k,t}\}_{k=1}^{n_2}$, $\{\mathcal{Z}_{k,t}\}_{k=1}^{n_2}$, and μ_t , yielding

$$\begin{aligned} \mathcal{X}_{t+1} &= \arg \min_{\mathcal{X}} \sum_{k=1}^{n_2} \left\{ \frac{\lambda}{2} \|\mathcal{Y}_{k,t} - \mathcal{X}\|_F^2 + \langle \mathcal{Z}_{k,t}, \mathcal{Y}_{k,t} - \mathcal{X} \rangle \right\} \\ &\quad + \gamma \sum_{\mathbf{i}_{1:n_2}} \text{LTTV}(\mathcal{X}(:, \dots, :, \mathbf{i}_{1:n_2})) \\ &\quad + \mu_t \sum_{\psi} \mathcal{W}(\psi) (\mathcal{T}(\psi) - \mathcal{X}(\psi))^2. \end{aligned} \quad (32)$$

To solve (32), we consider the optimization problem for each $\mathbf{i}_{1:n_2}$, separately. For a given $\mathbf{i}_{1:n_2}$, we define the subtensor $\hat{\mathcal{X}} = \mathcal{X}(:, \dots, :, \mathbf{i}_{1:n_2})$, $\hat{\mathcal{Y}}_{k,t} = \mathcal{Y}_{k,t}(:, \dots, :, \mathbf{i}_{1:n_2})$, $\hat{\mathcal{Z}}_{k,t} = \mathcal{Z}_{k,t}(:, \dots, :, \mathbf{i}_{1:n_2})$. For the data tensor, we define the subtensor $\hat{\mathcal{T}} = \mathcal{T}(:, \dots, :, \mathbf{i}_{1:n_2})$, $\hat{\mathcal{W}} = \mathcal{W}(:, \dots, :, \mathbf{i}_{1:n_2})$. Then, the optimization problem can be reformulated as

$$\begin{aligned} \arg \min_{\hat{\mathcal{X}}} \sum_{k=1}^{n_2} \left\{ \frac{\lambda}{2} \|\hat{\mathcal{Y}}_{k,t} - \hat{\mathcal{X}}\|_F^2 + \langle \hat{\mathcal{Z}}_{k,t}, \hat{\mathcal{Y}}_{k,t} - \hat{\mathcal{X}} \rangle \right\} \\ + \gamma \text{LTTV}(\hat{\mathcal{X}}) + \mu_t \sum_{\psi} \hat{\mathcal{W}}(\psi) (\hat{\mathcal{T}}(\psi) - \hat{\mathcal{X}}(\psi))^2. \end{aligned} \quad (33)$$

Since the objective function is a quadratic function of $\hat{\mathcal{X}}$, we compute the derivative of (33) with respect to each element in $\hat{\mathcal{X}}$ and set it equal to zero, yielding the system of equations

$$\begin{aligned} \langle \mathcal{A}^{\mathbf{i}_{1:n_1}}, \hat{\mathcal{X}} \rangle + 2\mu_t \hat{\mathcal{W}}(\mathbf{i}_{1:n_1}) (\hat{\mathcal{X}}(\mathbf{i}_{1:n_1}) - \hat{\mathcal{T}}(\mathbf{i}_{1:n_1})) \\ = \sum_{k=1}^{n_2} \left\{ \lambda \hat{\mathcal{Y}}_{k,t}(\mathbf{i}_{1:n_1}) + \hat{\mathcal{Z}}_{k,t}(\mathbf{i}_{1:n_1}) \right\}, \quad \forall \mathbf{i}_{1:n_1}, \end{aligned} \quad (34)$$

where $\mathcal{A}^{\mathbf{i}_{1:n_1}} \in \mathbb{R}^{I_1^s \times \dots \times I_{n_1}^s}$ is defined in (35). There are $\prod_{d=1}^{n_1} I_d^s$ unknowns and $\prod_{d=1}^{n_1} I_d^s$ linear equation, so that the subtensor $\hat{\mathcal{X}}$ can be found by solving the linear system in (34). Note that the linear independence of the system of equations

depends on the measurement data and may not be guaranteed, so that the solution may not be unique (in this case, it can be found using the Moore-Penrose pseudo-inverse).

The minimization of L over \mathcal{Y}_k with fixed $\mathcal{X}_{t+1}, \{\mathcal{Z}_{k,t}\}_{k=1}^{n_2}$, and μ_t can be formulated as

$$\begin{aligned} \mathcal{Y}_{k,t+1} &= \arg \min_{\mathcal{Y}_k} \sum_{\mathbf{i}_{1:n_1}} \left\{ \alpha_k \left\| \{\mathcal{Y}_k(\mathbf{i}_{1:n_1}, :, \dots, :)\}_{(k)} \right\|_* \right. \\ &\quad \left. + \frac{\lambda}{2} \left\| \left\{ \left\{ \mathcal{Y}_k - \left(\mathcal{X}_{t+1} - \frac{\mathcal{Z}_{k,t}}{\lambda} \right) \right\}(\mathbf{i}_{1:n_1}, :, \dots, :) \right\}_{(k)} \right\|_F^2 \right\}. \end{aligned}$$

For a given $\mathbf{i}_{1:n_1}$, we define $\tilde{\mathcal{X}}_{t+1} = \mathcal{X}_{t+1}(\mathbf{i}_{1:n_1}, :, \dots, :)$, $\tilde{\mathcal{Y}}_k = \mathcal{Y}_k(\mathbf{i}_{1:n_1}, :, \dots, :)$, and $\tilde{\mathcal{Z}}_{k,t} = \mathcal{Z}_{k,t}(\mathbf{i}_{1:n_1}, :, \dots, :)$. Thus, we reformulate the problem as

$$\arg \min_{\tilde{\mathcal{Y}}_k} \alpha_k \left\| \{\tilde{\mathcal{Y}}_k\}_{(k)} \right\|_* + \frac{\lambda}{2} \left\| \left\{ \tilde{\mathcal{Y}}_k - \left(\tilde{\mathcal{X}}_{t+1} - \frac{\tilde{\mathcal{Z}}_{k,t}}{\lambda} \right) \right\}_{(k)} \right\|_F^2.$$

This problem is shown to be strictly convex [25], and the solution is given by singular value thresholding. The update can be written as

$$\tilde{\mathcal{Y}}_k = \text{fold}_k \left(\mathcal{D}_{\frac{\alpha_k}{\lambda}} \left(\left\{ \tilde{\mathcal{X}}_{t+1} - \frac{\tilde{\mathcal{Z}}_{k,t}}{\lambda} \right\}_{(k)} \right) \right), \quad (36)$$

where \mathcal{D}_τ is the soft-thresholding operator. For a matrix \mathbf{A} with singular value decomposition (SVD) $\mathbf{A} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^H$, where $\mathbf{\Sigma} = \text{diag}(\sigma_1, \dots, \sigma_r)$, this operation is defined as $\mathcal{D}_\tau(\mathbf{A}) = \mathbf{U} \mathcal{D}_\tau(\mathbf{\Sigma}) \mathbf{V}^H$, $\mathcal{D}_\tau(\mathbf{\Sigma}) = \text{diag}(\{\max\{\sigma_i - \tau, 0\}\})$. With the update of $\tilde{\mathcal{Y}}_k = \mathcal{Y}_k(\mathbf{i}_{1:n_1}, :, \dots, :)$ for each $\mathbf{i}_{1:n_1}$, the updated $\mathcal{Y}_{k,t+1}$ can thus be completed.

Then, we update the Lagrangian multipliers \mathcal{Z}_k and μ as in (30) and (31). With the convergence of ADMM [39], the iteration approaches the primal feasibilities in (25) and (26), and we set the stop criteria as

$$\sum_{k=1}^{n_2} \|\mathcal{X}_t - \mathcal{Y}_{k,t}\|_F < \epsilon, \quad (37)$$

where ϵ is a constant threshold, and

$$\sum_{\psi} \mathcal{W}(\psi) (\mathcal{T}(\psi) - \mathcal{X}_t(\psi))^2 - \eta \leq 0. \quad (38)$$

The ADMM algorithm iteratively updates $\mathcal{X}, \{\mathcal{Y}_k\}_{k=1}^{n_2}, \{\mathcal{Z}_k\}_{k=1}^{n_2}$, and μ until the stop criteria (37) and (38) are satisfied. It follows that $\|\mathcal{Z}_{k,t+1} - \mathcal{Z}_{k,t}\|_F = \beta_1 \|\mathcal{Y}_{k,t+1} - \mathcal{X}_{t+1}\|_F \rightarrow 0$, which guarantees the convergence of \mathcal{Z}_k . The non-positive inequality constraint $\sum_{\psi} \mathcal{W}(\psi) (\mathcal{T}(\psi) - \mathcal{X}_t(\psi))^2 - \eta \leq 0$, and the projection $(\cdot)^+$ guarantee the convergence of μ . HNTC is shown in **Algorithm 1**.

$$\mathcal{A}^{\mathbf{i}_{1:n_1}}(\hat{\mathbf{i}}_{1:n_1}) = \begin{cases} n_2 \cdot \lambda + 2\gamma \sum_{k=1}^{n_1} \{\delta(i_k > 1) + \delta(i_k < I_k^s)\} & , \text{ if } \hat{\mathbf{i}}_{1:n_1} = \mathbf{i}_{1:n_1}, \\ -2\gamma & , \text{ if } \|\hat{\mathbf{i}}_{1:n_1} - \mathbf{i}_{1:n_1}\|_2 = 1, \\ 0 & , \text{ otherwise.} \end{cases} \quad (35)$$

Algorithm 1 Hybrid Noisy Tensor Completion (HNTC)

Input: incomplete data tensor \mathcal{T} , and weighting tensor \mathcal{W}
Output: \mathcal{T}_c

```

1: Initialization  $\mathcal{X}_t = \{\mathcal{Y}_{k,t}\}_{k=1}^{n_2} = \mathcal{T}$ ,  $\{\mathcal{Z}_{k,t}\}_{k=1}^{n_2} = 0$ ,  

    $\mu_t = 0$ ,  $\varepsilon_t = \iota_t = \infty$ 
2: while  $\varepsilon_t > \epsilon$  or  $\iota_t > 0$  do
3:   for  $\mathbf{i}_{1:n_2} \in [1, I_1^\ell] \times \cdots \times [1, I_{n_2}^\ell]$  do
4:      $\mathcal{X}_{k+1}(:, \dots, :, \mathbf{i}_{1:n_2}) = \hat{\mathcal{X}}$ ;  

5:     ( $\hat{\mathcal{X}}$  is obtained by solving (34))
6:   end for
7:   for  $\mathbf{i}_{1:n_1} \in [1, I_1^s] \times \cdots \times [1, I_{n_1}^s]$  do
8:      $\mathcal{Y}_{k,t+1}(\mathbf{i}_{1:n_1}, :, \dots, :) = \tilde{\mathcal{Y}}_k$ ,  $k = 1, \dots, n_2$ ;  

9:     ( $\tilde{\mathcal{Y}}_k$  is obtained by (36))
10:  end for
11:  Update  $\mathcal{Z}_{k,t+1}$ ,  $k = 1, \dots, n_2$ , by (30);
12:  Update  $\mu_{t+1}$  by (31);
13:   $\varepsilon_{t+1} = \sum_{k=1}^{n_2} \|\mathcal{X}_{t+1} - \mathcal{Y}_{k,t+1}\|_F$ ;
14:   $\iota_{t+1} = \sum_{\psi} \mathcal{W}(\psi) (\mathcal{T}(\psi) - \mathcal{X}_{t+1}(\psi))^2 - \eta$ ;  

15:   $t := t + 1$ ;
16: end while
17:  $\mathcal{T}_c = \mathcal{X}_t$ 

```

B. Online Hybrid Noisy Tensor Completion

Since the tensor completion problem (22) is a very large-scale convex optimization problem, the computing overhead for solving the problem would be quite high, which is challenging in a real-time system. For the first channel estimation aided by HNTC, the high computing overhead is not harmful because we can do the offline learning, which means that there is a predefined period of time for building and completing the data tensor before it is applied for the efficient channel training. However, for the subsequent channel estimation using HNTC, the data tensor necessitates doing the tensor completion again to provide new predictions exploiting the updated data. Thus, the large computing overhead of tensor completion would be impractical in real-time system.

To address this issue, we introduce the warm-start method [39], which initializes the iterative method using the solution obtained from the previous iteration. In HNTC, we apply an iterative method to solve the completion problem. The rate of convergence for this iterative method highly depends on the initial points of the variables, \mathcal{X} , \mathcal{Y}_k , \mathcal{Z}_k and μ , in the Lagrange function (27). The warm-start method aims to select the initial points of the variables \mathcal{X} , \mathcal{Y}_k , \mathcal{Z}_k , μ close to the convergence point based on the prior knowledge. For tensor \mathcal{X} , the initial point could be arbitrarily chosen since it is first updated in the iterative algorithm and irrelevant to the initial point of \mathcal{X} . For tensors \mathcal{Y}_k , \mathcal{Z}_k and μ , we choose their initial points as $\mathcal{Y}_{k,old}$, $\mathcal{Z}_{k,old}$ and μ_{old} , which are obtained in the previous iteration. Since the data tensor records the average received power which captures the long-term channel conditions, the recorded data are not expected to change drastically for each data update. Therefore, the previous ADMM iteration often provides an acceptable guess which leads to fewer iterations than the one with random initialization. The performance of the proposed online HNTC is evaluated in Section V-C.

C. Complexity Analysis

The computational complexity of HNTC is dominated by the update of \mathcal{X} and \mathcal{Y}_k . The update of \mathcal{X} requires matrix inversions to solve the linear equations in (34), leading to a complexity of $\mathcal{O}\left(\left(\prod_{j=1}^{n_2} I_j^\ell\right) \left(\prod_{m=1}^{n_1} I_m^s\right)^3\right)$. For the update of \mathcal{Y}_k , the SVD is used to perform the soft-thresholding operation, with the complexity as $\mathcal{O}\left(\left(\prod_{m=1}^{n_1} I_m^s\right) \max_k \left((I_k^\ell)^2 \cdot \prod_{j=1, j \neq k}^{n_2} I_j^\ell\right)\right)$ [44].

V. NUMERICAL RESULTS AND EXAMPLE APPLICATIONS

We evaluate the performance of HNTC with data generated by Quadriga [45]. We consider an uplink MIMO scenario with carrier frequency f_c using UPAs (as in (9)) having M_r antennas at the BS, and M_t antennas at the UE. The scenario *mmMAGIC_UMi_NLOS* is selected [45]. The simulation parameters are given in Table II. The network layout is depicted in Fig. 1, containing one BS in position $(0, 0)$ m at height 10m UEs in the rectangular area $\mathcal{G} = [10\text{m}, 60\text{m}] \times [-25\text{m}, 25\text{m}]$ at height 1.5m. We consider $51 \times 51 = 2601$ reference GPS coordinates uniformly located in the area \mathcal{G} . In each of the reference GPS coordinate, we collect the MIMO channel as the ground truth data. The position labels are derived as in (12) within the rectangular region $\bar{\mathcal{G}} = \mathcal{G}$, with resolution $\Delta_s = 5$ m, so that the number of discretized positions in the x and y coordinates are $L_x = L_y = 11$. Thus, for each position label, we have around $\frac{2601}{11 \times 11} \approx 21$ channel measurements.

TABLE II
COMMON SIMULATION PARAMETERS

Parameter	Symbol	Value
Carrier frequency	f_c	58.68 GHz
BS antenna number	$M_r (C_y^r, C_z^r)$	256 (16, 16)
UE antenna number	$M_t (C_y^t, C_z^t)$	16 (4, 4)
BS codebook size	$ \mathcal{W} (C_\theta^r, C_\phi^r)$	256 (16, 16)
UE codebook size	$ \mathcal{F} (C_\theta^t, C_\phi^t)$	16 (4, 4)

The ground truth data representing the noiseless average received power on all combinations of positions and BS codewords is collected in a fourth order tensor $\mathcal{T}_{avg} \in \mathbb{R}^{L_x \times L_y \times C_\theta^r \times C_\phi^r}$. We define the incomplete data tensor \mathcal{T} containing the measured data, along with the tensor \mathcal{V} denoting the number of measurements collected. In our numerical evaluations, we vary the ratio of observed positions $K_{op} = C_{op}/(L_x L_y)$, where C_{op} denotes the number of observed positions. Regarding the incomplete data tensor \mathcal{T} , we make **Assumptions 1, 2** for the experiment setting.

Assumption 1. *The observed positions \mathbf{p} are randomly chosen from $L_x \times L_y$ grid. For an observed position \mathbf{p}' , the measurements of the reference GPS coordinates corresponding to position \mathbf{p}' , $\{\mathbf{g} : \rho(\mathbf{g}) = \mathbf{p}', \mathbf{g} \in \bar{\mathcal{G}}\}$, are observed.*

Assumption 2. *For each observed GPS coordinate \mathbf{g} , only the measurements of the top 10% beams (ranked by received signal power) are stored in the database.*

With these two assumptions, \mathcal{T} is incomplete in both positions' and beams' dimensions. We consider two kinds

of collected measurement data: noise-free data and noisy data. For noisy data, we assume that channel measurements are collected with a beamforming signal-to-noise ratio at the receiver $\text{SNR}_r = 20$ dB. Given a MIMO channel \mathbf{H} , SNR_r is defined as

$$\text{SNR}_r = 10 \log_{10} \frac{P_t \|\mathbf{H}\|_2^2}{\sigma_n^2}, \quad (39)$$

where P_t is the transmit power, and σ_n^2 is the noise variance.

Note that the data of \mathcal{T} may be unavailable in some positions, which means no reconstruction on unknown positions is possible if we only consider the low-rank property. To show the advantage of HNTC, we compare it with existing tensor completion methods that use both low-rank and smoothness during reconstruction, LRTV-PDS [34] and SPCTV [36]. LRTV-PDS considers the low Tucker rank and smoothness during tensor reconstruction; SPCTV is based on the low CP rank prior with a smoothness constraint. Since these approaches [34], [36] are originally designed for image reconstruction, they consider the data tensor model with low-rank and smoothness consistent through all dimensions, and they assume noiseless measurements in SPCTV or noisy measurements in LRTV-PDS; in contrast, HNTC considers smoothness and low rank on the distinct dimensions, and is designed for noisy measurements, by taking into account the number of measurements as weighting contribution to the squared-error term.

A. Prediction Accuracy Comparison

In Fig. 3, we evaluate the relative square error (RSE) of the reconstructed tensor versus the observed position ratio K_{op} . The RSE is defined as

$$\text{RSE} = \frac{\|\mathcal{T}_c - \mathcal{T}_{avg}\|_F}{\|\mathcal{T}_{avg}\|_F}, \quad (40)$$

where \mathcal{T}_c is the reconstructed tensor and \mathcal{T}_{avg} is the ground-truth data tensor. In Fig. 3, the trend of RSE is monotonically decreasing with K_{op} . In fact, with more measurements recorded in the database, the tensor completion algorithms provide better reconstruction. We observe that HNTC outperforms both LRTV-PDS and SPCTV in RSE, which means that HNTC provides a better tensor approximation to \mathcal{T}_{avg} . Given noise-free measurements on only 40% of any possible positions, HNTC attains $\text{RSE} = 0.57$, as opposed to LRTV-PDS with $\text{RSE} = 0.64$, and SPCTV with $\text{RSE} = 0.68$ in the same configuration. For the noisy measurement data, the RSE degrades drastically for all approaches. Given the noisy measurement tensor on 40% positions, the prediction accuracy attains $\text{RSE} = 0.87$ for HNTC, $\text{RSE} = 1.17$ for LRTV-PDS, and $\text{RSE} = 1.2$ for SPCTV. However, HNTC is more robust to noise than the state-of-the-art schemes, since it accounts for noisy measurements as seen in (23).

B. Position-Aided Beam Recommendation

1) *Recommendation Algorithm*: For the position-aided beam-recommendation in Section III-B, we illustrate the beam subset selection (BSS) algorithm. With the completed tensor \mathcal{T}_c , we have the estimated received power of all receive beams

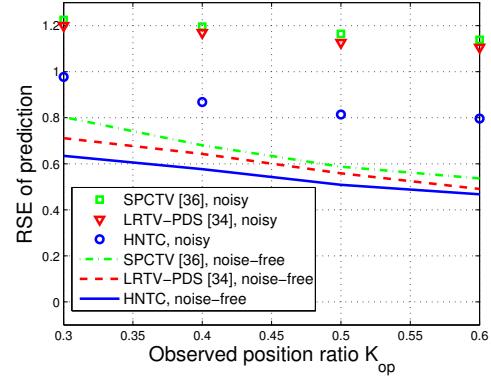


Fig. 3. The RSE of prediction versus the ratio of observed positions (K_{op}).

at UE position \mathbf{p} . If the number of trained beams is N_{tr} , the construction of the recommended beam set is a subset selection problem, that selects the N_{tr} beams with largest estimated received power in the completed tensor \mathcal{T}_c , as shown in **Algorithm 2**.

Algorithm 2 Beam Subset Selection (BSS)

Input: completed tensor \mathcal{T}_c , number of trained beams N_{tr} , BS codebook \mathcal{W} with indices \mathcal{K} , UE GPS coordinate \mathbf{g}

Output: recommended beam subset $\mathcal{S}_{N_{tr}}$

- 1: **Initialization** $\mathcal{S}_0 \leftarrow \emptyset$
- 2: **for** $n = 1 : N_{tr}$ **do**
- 3: $(u^*, v^*) = \arg \max_{(u, v) \in \mathcal{K} \setminus \mathcal{S}_{n-1}} \mathcal{T}_c(\rho(\mathbf{g}), u, v)$
- 4: $\mathcal{S}_n \leftarrow \mathcal{S}_{n-1} \cup (u^*, v^*)$
- 5: **end for**

2) *Performance of Proposed Beam-Alignment*: Here, we evaluate the performance of the position-aided beam recommendation with HNTC, compared with the ones with LRTV-PDS and SPCTV. Our formulation with tensor completion allows prediction for unknown positions by exploiting spatial correlation. However, the state-of-art approach [17] for position-aided beam alignment uses only the prior knowledge available at a given position, and is thus unable to make predictions in positions where the measurements are not available. For comparison, we consider the type-B fingerprinting method [17] by providing the recommended beam set based on the closest position having available prior knowledge if the prior measurements of UE position are not given. Moreover, we consider the position-aided beam recommendation aided by BSS with \mathcal{T}_{avg} , called *Genie-Aided*. The Genie-Aided approach is expected to have the best recommendation performance since it uses ground truth data at all positions and beams.

First, we evaluate the power loss probability $P_{pl}(\mathcal{S}_p)$ versus the percentage of trained beams $K_{tr} = N_{tr}/|\mathcal{K}|$, where N_{tr} is the number of trained BS beams and $|\mathcal{K}|$ is the total number of BS beams. The set \mathcal{S}_p is the recommended beam subset at position \mathbf{p} using BSS. To measure the beam alignment accuracy for the recommended set \mathcal{S}_p , we define the power

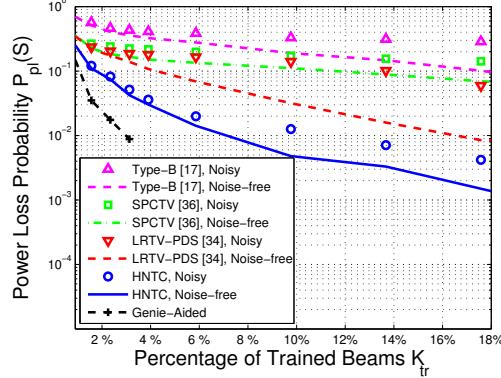


Fig. 4. Power loss probability ($P_{pl}(\mathcal{S}_p)$) versus the percentage of trained beams (K_{tr}), with the observed position ratio $K_{op} = 40\%$.

loss probability metric as

$$P_{pl}(\mathcal{S}_p) = \mathbb{P} \left(\max_{(u,v) \in \mathcal{S}_p} r^{(\mathbf{p},u,v)} < \max_{(u,v) \in \mathcal{K}} r^{(\mathbf{p},u,v)} \right),$$

where $r^{(\mathbf{p},u,v)}$ is the received power defined in (13). The power loss probability is averaged over the channels at the GPS coordinates corresponding to all positions. In Fig. 4, we evaluate the power loss probability versus the percentage of trained beams with the observed position ratio $K_{op} = 40\%$. For the Genie-Aided approach, it attains $P_{pl}(\mathcal{S}_p) = 10\%$ with $K_{tr} = 1.1\%$. With $K_{tr} > 3\%$ of trained beams, the Genie-Aided approach can always recommend the beam set including the best receive beam. For the noise-free measurement data tensor, only 1.5% of the trained beams is required for HNTC to attain $P_{pl}(\mathcal{S}_p) = 10\%$, as opposed to 4% for LRTV-PDS, 12% for SPC-TV, and 18% for Type-B. The position-aided beam alignment approach supported by tensor completion (HNTC, LRTV-PDS, or SPC-TV) outperforms the state-of-the-art method (Type-B) since the tensor completion provide accurate power prediction on the unavailable positions. With the noisy measurement data tensor, the $P_{pl}(\mathcal{S}_p)$ supported by our proposed approach (HNTC) is more robust than the ones using LRTV-PDS, SPC-TV, and Type-B. This behavior is in line with the improved performance of HNTC observed in Fig. 3.

3) *Spectral Efficiency*: Next, we evaluate the spectral efficiency versus the receive SNR. We define the achievable transmission rate as

$$R = B \log_2 \left(1 + \text{SNR}_r \frac{\|\mathbf{w}^H \mathbf{H} \mathbf{f}\|^2}{\|\mathbf{H}\|_2^2} \right), \quad (41)$$

where B is the bandwidth; SNR_r is the receive beamforming SNR defined in (39); \mathbf{w}/\mathbf{f} is the BS/UE beamforming vector; \mathbf{H} is the MIMO channel. The selected BS/UE beamforming vector $(\mathbf{w}^*, \mathbf{f}^*)$ is the best beam pair (ranked by received power) chosen from $\mathcal{S} \times \mathcal{F}$, where \mathcal{S} is the BS beam set recommended by BSS, and \mathcal{F} is the UE codebook. Since the set of beam pairs $\mathcal{S} \times \mathcal{F}$ is scanned exhaustively, the resulting overhead is $T_{train} = (|\mathcal{S}| \cdot |\mathcal{F}| + 1)\delta_S$, where the microslot duration $\delta_S = 10 \mu\text{s}$ is the time required to scan a single beam. The fraction of time used for data transmission is $f_{comm} = \frac{T_{frame} - T_{train}}{T_{frame}}$, where $T_{frame} = 10 \text{ ms}$ is the

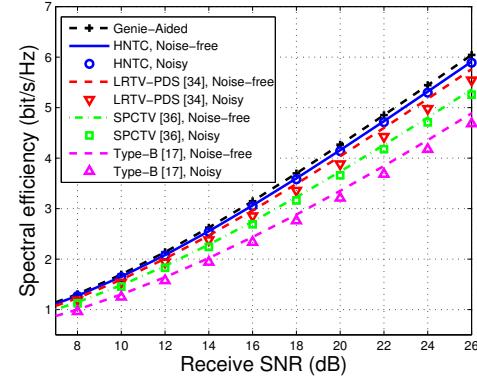


Fig. 5. Spectral efficiency ($\frac{f_{comm} \times R}{B}$) versus the receive SNR (SNR_r) with $(K_{op}, K_{tr}) = (40\%, 2\%)$.

frame duration. The spectral efficiency is defined as $\frac{f_{comm} \times R}{B}$, which accounts for the loss due to the training overhead. The proposed recommendation-based method is much more efficient than a conventional exhaustive search method, where all BS and UE beam pairs $\mathcal{W} \times \mathcal{F}$ are scanned. In fact, the conventional *exhaustive search* cannot be implemented in our considered scenario since its training overhead exceeds the frame duration, $T_{train} = (|\mathcal{W}| \cdot |\mathcal{F}| + 1)\delta_S > T_{frame}$, leaving no time for data transmission.

In Fig. 5, we evaluate the spectral efficiency ($\frac{f_{comm} \times R}{B}$) versus SNR_r with $(K_{op}, K_{tr}) = (40\%, 2\%)$. The Genie-Aided approach attains the largest spectral efficiency because it has the smallest power loss probability $P_{pl}(\mathcal{S}_p) = 2.5\%$ with the support of available information on all possible positions. For the noise-free measurement data and a reference $\text{SNR}_r = 20 \text{ dB}$, the spectral efficiency of HNTC is 4.16 bits/s/Hz, which is better than the state-of-the-art approach (Type-B) with 3.35 bits/s/Hz. Besides, compared with the beam recommendation aided by other tensor completion approaches, the spectral efficiency of HNTC outperforms the one with LRTV-PDS by 0.13 bits/s/Hz, and the one with SPC-TV by 0.43 bits/s/Hz. For the compared approaches, the spectral efficiency with the noisy measurement data is worse than the one with the noise-free measurement data. However, our proposed HNTC with noisy measurement data has almost the same spectral efficiency as HNTC with noise-free measurement data. HNTC is more robust because $P_{pl}(\mathcal{S}_p)$ with noisy measurement data is still fairly low ($< 10\%$) in this configuration (as shown in Fig. 4), which is quite similar to $P_{pl}(\mathcal{S}_p)$ with noise-free measurement data.

C. Evaluation of Online Hybrid Noisy Tensor Completion

We evaluate the performance of the online HNTC proposed in Section IV-B **with the warm start**, compared with the one **without warm start**. For the prediction accuracy, we observe the RSE between the reconstructed tensor and \mathcal{T}_{avg} and also the number of iterations of ADMM for the convergence rate. The noise-free measurement data tensor \mathcal{T} is considered. For the online updating scenario, at update instant 0, we consider the initial ratio of observed positions K_{op} as $K_{ini} = 30\%$. We assume that the measurements of $N_{upd} = 5$ new positions are updated to the data tensor \mathcal{T} in each subsequent update instant.

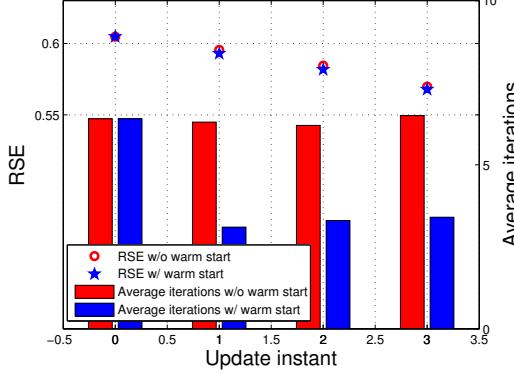


Fig. 6. RSE and average number of iterations comparison between HNTC **with warm start** and HNTC **without warm start**, $(K_{ini}, N_{upd}) = (30\%, 5)$.

For HNTC **with warm start**, we initialize the variables in ADMM as the ones retrieved from the previous update instant.

In Fig. 6, for the RSE comparison, the HNTC **with warm start** is close to the HNTC **without warm start**. In update instant 0, the average number of iterations of both cases is around 6.3 because there is no available prior information. For the subsequent update instants, HNTC **with warm start** converges in around 3.2 iterations, while HNTC **without warm start** requires 6.3 iterations. The warm start method reduces the computational complexity by converging in 50% fewer iterations without compromising on the prediction accuracy.

D. Noisy Positional Information

Here, we investigate the influence of noisy positional information on the position-aided beam recommendation. In a practical setting, the positional information is acquired via the process of GNSS/GPS estimation [20], [21]. Due to the mobility and estimation error, the obtained positional information may be noisy, which would impair the performance of the position-aided channel estimation. To model the positional error, let $E(d) = \{\mathbf{x} \in \mathbb{R}^2 : \|\mathbf{x}\|_2 \leq d\}$ be a two dimensional closed disk centered at the origin with radius d ; we then model the random error as $\mathbf{e} \in \mathbb{R}^2$ uniformly distributed in $E(d)$ [21]. The noisy spatial coordinate is defined as $\mathbf{g}_r = \mathbf{g} + \mathbf{e}$, where \mathbf{g} is the ground truth.

To alleviate the impairment of the noisy spatial coordinates, we propose the grouping-based beam subset selection (G-BSS) in **Algorithm 3**. Given the received noisy spatial coordinate \mathbf{g}_r and error radius d , we collect all possible positions as a set $\mathcal{P} = \{\rho(\mathbf{g}) : \|\mathbf{g} - \mathbf{g}_r\|_2 \leq \zeta d\}$, where ζ is a constant coefficient. Note that the performance is influenced by the selection of ζ , and we choose $\zeta = 0.4$ in this work. Then, we derive the subtensor $\bar{\mathcal{R}} = \frac{1}{|\mathcal{P}|} \sum_{\mathbf{p} \in \mathcal{P}} \mathcal{T}_c(\mathbf{p}, :, :)$, which contains the predicted received power of each beam by averaging over all possible positions in the set \mathcal{P} . Finally, we use the tensor $\bar{\mathcal{R}}$ to provide the beam recommendations for the UE.

In Fig. 7, we evaluate the power loss probability versus the percentage of trained beams, under the scenario with noisy positional information. We compare the performance of G-BSS with that of BSS, which neglects the error in the positional information. The completed tensor \mathcal{T}_c is reconstructed by HNTC,

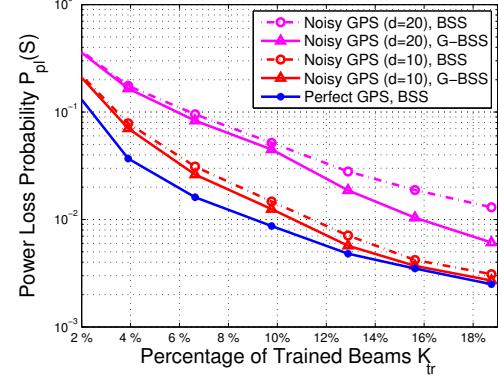


Fig. 7. Power loss probability versus the percentage of trained beams in the scenario with noisy positional information.

and the noise-free measurement data is considered. With the BSS in **Algorithm 2**, the power loss probability increases when the positional error radius d increases. The performance of the position-aided beam recommendation is impaired by the noisy positional information. However, G-BSS is more robust than BSS against these impairments. Given $K_{tr} = 10\%$, the BSS with perfect GPS attains $P_{pl}(\mathcal{S}_p) = 0.87\%$. For the scenario with positional error $d = 10$ m, the power loss probability is improved by G-BSS to be $P_{pl}(\mathcal{S}_p) = 1.2\%$, compared with $P_{pl}(\mathcal{S}_p) = 1.5\%$ by BSS. If the positional error increases to $d = 20$ m, the power loss probability is improved with the support of G-BSS to be $P_{pl}(\mathcal{S}_p) = 4.5\%$, as opposed to $P_{pl}(\mathcal{S}_p) = 5\%$ by BSS.

Algorithm 3 Grouping-based Beam Subset Selection (G-BSS)

Input: completed tensor \mathcal{T}_c , number of trained beams N_{tr} , BS codebook \mathcal{W} with indices \mathcal{K} , UE noisy GPS \mathbf{g}_r , positional error radius d

Output: recommended beam subset $\mathcal{S}_{N_{tr}}$

- 1: **Initialization** $\mathcal{S}_0 \leftarrow \emptyset$
- 2: $\mathcal{P} = \{\rho(\mathbf{g}) : \|\mathbf{g} - \mathbf{g}_r\|_2 \leq \zeta d\}$
- 3: $\bar{\mathcal{R}} = \frac{1}{|\mathcal{P}|} \sum_{\mathbf{p} \in \mathcal{P}} \mathcal{T}_c(\mathbf{p}, :, :)$
- 4: **for** $n = 1 : N_{tr}$ **do**
- 5: $(u^*, v^*) = \arg \max_{(u, v) \in \mathcal{K} \setminus \mathcal{S}_{n-1}} \bar{\mathcal{R}}(u, v)$
- 6: $\mathcal{S}_n \leftarrow \mathcal{S}_{n-1} \cup (u^*, v^*)$
- 7: **end for**

VI. CONCLUSIONS

In this paper, we proposed a learning framework to perform data-assisted beamforming in MIMO communication over a fixed service area with noisy power measurements on a small subset of possible positions. In our model, the received power and side information (e.g., user positions and receive beams) were collected into a data tensor. We developed a noisy tensor completion, HNTC, exploiting the low-rank and smoothness properties of the channel data. The numerical results showed that HNTC provides more accurate received power prediction than the state-of-the-art tensor completion method [34], [36] utilizing both the low-rank and the smoothness of the data. Furthermore, the beam recommendation aided by HNTC was

shown to improve the performance of beam alignment over the state-of-the-art data-assisted beam alignment approach [17], by improving the prediction accuracy and reducing the beam training overhead.

REFERENCES

- [1] T.-H. Chou, N. Michelusi, D. J. Love, and J. V. Krogmeier, "Millimeter wave beam recommendation via tensor completion," to appear at IEEE ICC 2020.
- [2] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. Soong, and J. C. Zhang, "What will 5G be?" *IEEE Journal on selected areas in communications*, vol. 32, no. 6, pp. 1065–1082, 2014.
- [3] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE access*, vol. 1, pp. 335–349, 2013.
- [4] R. W. Heath, N. González-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE journal of selected topics in signal processing*, vol. 10, no. 3, pp. 436–453, 2016.
- [5] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 74–80, 2014.
- [6] H. Xie, F. Gao, and S. Jin, "An overview of low-rank channel estimation for massive MIMO systems," *IEEE Access*, vol. 4, pp. 7313–7321, 2016.
- [7] S. Hur, T. Kim, D. J. Love, J. V. Krogmeier, T. A. Thomas, and A. Ghosh, "Millimeter wave beamforming for wireless backhaul and access in small cell networks," *IEEE Transactions on Communications*, vol. 61, no. 10, pp. 4391–4403, October 2013.
- [8] M. Hussain and N. Michelusi, "Energy-efficient interactive beam alignment for millimeter-wave networks," *IEEE Transactions on Wireless Communications*, vol. 18, no. 2, pp. 838–851, 2018.
- [9] M. Hussain and N. Michelusi, "Second-best beam-alignment via bayesian multi-armed bandits," in *IEEE Global Communications Conference (GLOBECOM)*, 2019, pp. 1–6.
- [10] A. Alkhateeb, G. Leus, and R. W. Heath, "Compressed sensing based multi-user millimeter wave systems: How many measurements are needed?" in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2015, pp. 2909–2913.
- [11] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 831–846, 2014.
- [12] Z. Marzi, D. Ramasamy, and U. Madhow, "Compressive channel estimation and tracking for large arrays in mm-Wave picocells," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 514–527, 2016.
- [13] N. González-Prelcic, R. Méndez-Rial, and R. W. Heath, "Radar aided beam alignment in mmWave V2I communications supporting antenna diversity," in *2016 Information Theory and Applications Workshop (ITA)*. IEEE, 2016, pp. 1–7.
- [14] A. Klautau, N. González-Prelcic, and R. W. Heath, "Lidar data for deep learning-based mmWave beam-selection," *IEEE Wireless Communications Letters*, vol. 8, no. 3, pp. 909–912, June 2019.
- [15] M. Hashemi, C. E. Koksal, and N. B. Shroff, "Out-of-band millimeter wave beamforming and communications to achieve low latency and high energy efficiency in 5G systems," *IEEE Transactions on Communications*, vol. 66, no. 2, pp. 875–888, Feb 2018.
- [16] N. González-Prelcic, A. Ali, V. Va, and R. W. Heath, "Millimeter-wave communication with out-of-band information," *IEEE Communications Magazine*, vol. 55, no. 12, pp. 140–146, Dec 2017.
- [17] V. Va, J. Choi, T. Shimizu, G. Bansal, and R. W. Heath, "Inverse multipath fingerprinting for millimeter wave V2I beam alignment," *IEEE Trans. Veh. Technol.*, vol. 67, no. 5, pp. 4042–4058, May 2018.
- [18] Y. Wang, A. Klautau, M. Ribeiro, A. C. K. Soong, and R. W. Heath, "MmWave vehicular beam selection with situational awareness using machine learning," *IEEE Access*, vol. 7, pp. 87479–87493, 2019.
- [19] K. Satyanarayana, M. El-Hajjar, A. A. M. Mourad, and L. Hanzo, "Deep learning aided fingerprint-based beam alignment for mmWave vehicular communication," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 11, pp. 10858–10871, Nov 2019.
- [20] N. Roy, H. Wang, and R. Roy Choudhury, "I am a smartphone and I can tell my user's walking direction," in *Proceedings of the 12th annual international conference on Mobile systems, applications, and services*, 2014, pp. 329–342.
- [21] F. Maschietti, D. Gesbert, P. de Kerret, and H. Wymeersch, "Robust location-aided beam alignment in millimeter wave massive MIMO," in *GLOBECOM 2017-2017 IEEE Global Communications Conference*. IEEE, 2017, pp. 1–6.
- [22] D. Lee, S. Kim, and G. B. Giannakis, "Channel gain cartography for cognitive radios leveraging low rank and sparsity," *IEEE Transactions on Wireless Communications*, vol. 16, no. 9, pp. 5953–5966, Sep. 2017.
- [23] 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.
- [24] J. Liu, P. Musialski, P. Wonka, and J. Ye, "Tensor completion for estimating missing values in visual data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 208–220, 2013.
- [25] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM Journal on optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.
- [26] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM review*, vol. 52, no. 3, pp. 471–501, 2010.
- [27] E. J. Candès and T. Tao, "The power of convex relaxation: Near-optimal matrix completion," *IEEE Transactions on Information Theory*, vol. 56, no. 5, pp. 2053–2080, 2010.
- [28] E. Candès and B. Recht, "Exact matrix completion via convex optimization," *Commun. ACM*, vol. 55, no. 6, p. 111–119, Jun. 2012.
- [29] E. J. Candès and Y. Plan, "Matrix completion with noise," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 925–936, June 2010.
- [30] Q. Dai and W. Sha, "The physics of compressive sensing and the gradient-based recovery algorithms," *arXiv preprint arXiv:0906.1487*, 2009.
- [31] X. Han, J. Wu, L. Wang, Y. Chen, L. Senhadji, and H. Shu, "Linear total variation approximate regularized nuclear norm optimization for matrix completion," in *Abstract and Applied Analysis*, vol. 2014. Hindawi, 2014.
- [32] R. Tomioka, K. Hayashi, and H. Kashima, "Estimation of low-rank tensors via convex optimization," *arXiv preprint arXiv:1010.0789*, 2010.
- [33] S. Gandy, B. Recht, and I. Yamada, "Tensor completion and low-n-rank tensor recovery via convex optimization," *Inverse Problems*, vol. 27, no. 2, p. 025010, 2011.
- [34] T. Yokota and H. Hontani, "Simultaneous tensor completion and denoising by noise inequality constrained convex optimization," *IEEE Access*, vol. 7, pp. 15669–15682, 2019.
- [35] Y. Liu, Z. Long, H. Huang, and C. Zhu, "Low CP rank and Tucker rank tensor completion for estimating missing components in image data," *IEEE Transactions on Circuits and Systems for Video Technology*, 2019.
- [36] T. Yokota, Q. Zhao, and A. Cichocki, "Smooth PARAFAC decomposition for tensor completion," *IEEE Transactions on Signal Processing*, vol. 64, no. 20, pp. 5423–5436, 2016.
- [37] X. Li, Y. Ye, and X. Xu, "Low-rank tensor completion with total variation for visual data inpainting," in *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.
- [38] C. J. Hillar and L.-H. Lim, "Most tensor problems are NP-hard," *Journal of the ACM (JACM)*, vol. 60, no. 6, pp. 1–39, 2013.
- [39] S. Boyd, N. Parikh, and E. Chu, *Distributed optimization and statistical learning via the alternating direction method of multipliers*. Now Publishers Inc, 2011.
- [40] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM review*, vol. 51, no. 3, pp. 455–500, 2009.
- [41] T. G. Kolda, "Orthogonal tensor decompositions," *SIAM Journal on Matrix Analysis and Applications*, vol. 23, no. 1, pp. 243–255, 2001.
- [42] 3GPP, "Study on channel model for frequencies from 0.5 to 100 GHz," 3rd Generation Partnership Project (3GPP), Technical Report (TR) 38.901, Oct. 2019, version 16.0.0.
- [43] M. Hussain, M. Scalabrin, M. Rossi, and N. Michelusi, "Mobility and blockage-aware communications in millimeter-wave vehicular networks," 2020, submitted to the IEEE Transactions on Vehicular Technology.
- [44] G. H. Golub and C. F. van Loan, *Matrix Computations*, 4th ed. JHU Press, 2013. [Online]. Available: <http://www.cs.cornell.edu/cv/GVL4/golubandvanloan.htm>
- [45] S. Jaekel, L. Raschkowski, K. Börner, and L. Thiele, "Quadriga: A 3-D multi-cell channel model with time evolution for enabling virtual field trials," *IEEE Transactions on Antennas and Propagation*, vol. 62, no. 6, pp. 3242–3256, 2014.