Learning to Localize: A 3D CNN Approach to User Positioning in Massive MIMO-OFDM Systems

Chi Wu, *Student Member, IEEE*, Xinping Yi, *Member, IEEE*, Wenjin Wang, *Member, IEEE*, Li You, *Member, IEEE*, Qing Huang, *Member, IEEE*, Xiqi Gao, *Fellow, IEEE*, and Qing Liu

***Abstract*—In this paper, we investigate user positioning in massive multiple-input multiple-output (MIMO) orthogonal frequency-division multiplexing (OFDM) systems where the base station (BS) is equipped with a uniform planar array (UPA). Taking advantage of the UPA geometry and wide bandwidth, we advocate the use of the angle-delay channel power matrix (ADCPM) as a new type of fingerprint to replace the traditional ones. The ADCPM embeds the stable and stationary multipath characteristics, e.g., delay, power, and angles in the vertical and horizontal directions, which are beneficial to positioning. We further exploit the sparsity of the ADCPM to reduce the noise contamination in the ADCPM. Taking ADCPM fingerprints as the inputs, we propose a novel three-dimensional (3D) convolution neural network (CNN) enabled learning method to localize the 3D positions of the mobile terminals (MTs). In particular, such a 3D CNN model consists of a convolution refinement module to refine the elementary feature maps from the ADCPM fingerprints, three extended Inception modules to extract the advanced feature maps, and a regression module to estimate the 3D positions. By intensive simulations, the proposed 3D CNN-enabled positioning method is demonstrated to achieve higher positioning accuracy than the traditional searching-based ones, with reduced computational complexity and storage overhead, and robust to noise contamination.**

***Index Terms*—Massive MIMO, positioning, deep learning, 3D convolution neural network, fingerprint.**

I. Introduction

As location-based applications are extensively deployed in modern society, accurate positioning has received enormous attention in both industry and academia [2]. The global positioning system (GPS) has provided real-time outdoor positioning for the mobile terminal (MT) with the accuracy within several meters [3]. However, in urban areas, the GPS positioning performance will be significantly degraded due to the blockage of buildings, cars, and pedestrians, etc.

Recently, user positioning by exploiting the rich information of multipath wireless propagation has drawn a lot of attention. Various positioning methods have been proposed in the literatures, including the geometry-based and the fingerprint-based positioning methods [4]–[18]. The geometry-based positioning is a trilateral/triangulating-to-localize method that relies on the information of wireless signals from the MTs to the base stations (BSs), e.g., angle of arrival (AOA) [4], [5], time of arrival (TOA) [6], time difference of arrival (TDOA) [7], [8], and received signal strength (RSS) [5], [9]. However, the geometry-based positioning suffers from positioning performance degradation due to the non-line of sight (NLOS) propagations, which are common in sophisticated urban environments [10]. In contrast, fingerprint-based positioning is a matching-to-localize method that consists of offline fingerprint database construction, online fingerprint matching, and location estimation [11]–[18]. In the offline phase, reference points (RPs) are selected, for which the pairs of fingerprints and corresponding positions are stored in the database. In the online phase, the position of the MT is estimated by searching the collected database and matching the input fingerprint with the stored ones. As an indicator of the surrounding environment, fingerprints have been widely adopted for user positioning in the sophisticated multipath environment [11], [12].

The most common feature used in fingerprint-based positioning is RSS [13], [18], yet there are two shortcomings of the RSS-based fingerprints. On one hand, RSS suffers from fast fading fluctuation and hardware heterogeneity and is therefore unstable for positioning. On the other hand, RSS only captures the coarsest channel information that cannot meet the demand in complex communication environment. Recently, some researchers proposed to use the channel state information (CSI) as the fingerprint [14]–[17]. Capturing more channel information than RSS, CSI has the potential to enhance the positioning accuracy. In certain scenarios (e.g., WiFi network [14]), however, due to the limited bandwidth and number of antennas, the CSI-based fingerprints are insufficient to capture the multipath characteristics because of the low resolution in the space and frequency domains.

Fortunately, such limitations can be overcome in massive multiple-input multiple-output orthogonal frequency-division multiplexing (MIMO-OFDM) systems. Thanks to the large-scale antenna array and wide bandwidth, the CSI-based fingerprints are able to capture rich multipath information including

Manuscript received March 12, 2020; revised September 3, 2020 and November 10, 2020; accepted February 7, 2021. This work was supported in part by the National Natural Science Foundation of China under Grants 61631018, 61761136016, and 61801114, the Natural Science Foundation of Jiangsu Province under Grant BK20170688, and the National Key R&D Program of China under Grant 2019YFB1803102. Part of this work was presented at IEEE International Conference on Communications (ICC), Dublin, Ireland, 2020 [1]. The associate editor coordinating the review of this article and approving it for publication was Prof. X. Cheng. *(Corresponding author: Wenjin Wang.)*

C. Wu, W. Wang, L. You, Q. Huang, and X. Q. Gao are with the National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China, and also with the Purple Mountain Laboratories, Nanjing 211100, China (e-mail: chiwu@seu.edu.cn; wangwj@seu.edu.cn; liyou@seu.edu.cn; huangqing@seu.edu.cn; xqgao@seu.edu.cn).

X. Yi is with the Department of Electrical Engineering and Electronics, University of Liverpool, L69 3BX, United Kingdom (email: xinping.yi@liverpool.ac.uk).

Q. Liu is with the Shanghai Huawei technologies Ltd. company, Shanghai 200120, China (email: patrick.liuqing@huawei.com).

powers, angles, and delays [11], [19]–[25] for positioning. Various fingerprint-based positioning techniques have been proposed for massive MIMO/MIMO-OFDM systems [11], [13], [26]. Of particular relevance is the approach proposed in [11], where a weighted k-nearest neighbor (WKNN) algorithm is applied using the angle-delay domain channel information as the fingerprints in massive MIMO-OFDM systems. Other approaches for positioning include the use of some sophisticated techniques, such as Gaussian processes regression in [13] and compressive sensing in [26], to name a few.

Most recently, deep learning has found application in user positioning, inspired by its great success in image recognition, speech signal processing, and self-driving [12], [16], [27]–[29]. As a matter of fact, the fingerprint-based positioning can be cast into an image recognition problem, in which the fingerprints can be treated as images to recognize. For indoor positioning, a two-step training deep neural network (DNN) based positioning method was proposed in [27] for the NLOS massive MIMO scenario, a single DNN classifier to determine the probabilities of the MT being on the collected RPs was used in [16], and a deep learning-based positioning method which utilizes the classic deep belief nets (DBNs) with a stack of restricted Boltzmann machines (RBMs) was proposed in [28]. For outdoor positioning, the deep convolution neural network (CNN) was utilized in [12], [29] to map the CSI into the 2D position coordinates.

However, the aforementioned methods were mainly dedicated to 2D positioning. When it comes to the uniform planar array (UPA), high angular resolution can be realized in both vertical and horizontal directions [30], which provides new opportunities for the fingerprint-based 3D positioning. To this end, we propose a novel deep learning-based 3D positioning method for the MIMO-OFDM system with the UPA, taking the angle-delay domain channel power matrix (ADCPM) as the fingerprint that contains the information of multipath angles, delays, and powers. Instead of CSI fingerprints in the space-frequency domain, we translate them into the angle-delay domain with a 2D inverse discrete Fourier transform (IDFT), by which the size of fingerprint can be reduced significantly. Further, by fully exploiting the sparsity of ADCPM, we adopt a filter with an element-wise threshold of ADCPM such that the noise contamination caused by the noisy channels in the ADCPMs can be reduced. For massive MIMO-OFDM systems, the dimension of ADCPM is still high. Then we propose a regression-oriented 3D CNN model that maps the high dimensional ADCPM fingerprints into the 3D position coordinates directly with low computational complexity. In particular, our 3D CNN model consists of four key components: (1) the convolution refinement module to refine the elementary feature maps from the ADCPM fingerprints; (2) the 3D Inception modules that are extended from the Inception modules in GoogLeNet [31]–[33]; (3) the average pooling to replace the full-connected layer at the bottom of the network using fewer parameters; (4) the batch normalization (BN) in each layer to improve the convergence speed and generalization ability of the network. Such an end-to-end positioning method works in the following way. In the offline training phase, the 3D CNN is trained by using the ADCPM fingerprints and the corresponding coordinates of RPs. In the online estimation phase, a preprocessing is applied to the ADCPM fingerprints to mitigate background noise, followed by position estimation using the trained 3D CNN model. To summarize, our contributions are as follows:

- For massive MIMO-OFDM systems, we extend the use of ADCPM as fingerprint for user positioning from uniform linear array (ULA) to UPA in a principled way. We prove that, compared with the existing fingerprints, the ADCPM fingerprints have better properties, i.e., rich and stable multipath characteristics that are of close relevance to the position information, and reduced size compared to its counterpart in the space-frequency domain.
- Exploiting the sparsity of ADCPM, we propose a simple noise reduction strategy, which is demonstrated to make the ADCPM fingerprints robust to noise contamination. This shows the superiority of using ADCPM as the positioning fingerprints.
- Taking advantage of the UPA and the delay spread effect in such a massive MIMO-OFDM system, we propose a 3D CNN based positioning method employing the sparse ADCPM as input, which achieves higher 3D positioning accuracy with reduced storage and computational overhead than the searching-based methods (e.g., WKNN).

The rest of the paper is organized as follows. In Section II, we investigate the 3D MIMO-OFDM system, and propose a new type of fingerprint extracted from the multipath channel characteristics and a noise contamination reduction method. In Section III, we introduce the design of our proposed 3D CNN-enabled positioning method, followed by the detailed network architecture in Section IV. Simulation results are presented in Section V, and conclusion is given in Section VI.

Notations: We use $\bar{\jmath} = \sqrt{-1}$ to denote the imaginary unit. Vectors and matrices are denoted in lower-case bold-faced characters and upper-case bold-faced characters, respectively, the element indices of vector and matrix start with 0. We use $[\mathbf{a}]_i$, $[\mathbf{A}]_{i,j}$, and $[\mathbf{A}]_i$ to denote the ith element of the vector $\mathbf{a}$, the (i,j)th element of matrix $\mathbf{A}$, and the ith column of matrix $\mathbf{A}$, respectively. The superscripts $(\cdot)^T$, $(\cdot)^H$, and $(\cdot)^*$ indicate the matrix transpose, conjugate-transpose, and conjugate operation. The complex number field, real number field, and integer field are represented by $\mathbb{C}$, $\mathbb{R}$, and $\mathbb{Z}$, respectively. The symbols $\odot$ and $\otimes$ denote the Hadamard and Kronecker product, respectively. $\mathrm{vec}(\cdot)$ denotes the vectorization operator, and $\mathrm{Tr}(\cdot)$ denotes the trace of a square matrix. $\mathrm{diag}(\mathbf{a})$ denotes the diagonal matrix with its diagonal elements composed of the vector $\mathbf{a}$. We use $E\{\cdot\}$ to denote the expectation of random variable and random vector variable. $\lfloor x \rfloor$ denotes the largest integer not greater than x, $\langle\cdot\rangle_N$ denotes the modulo-N operation. $\delta(\cdot)$ denotes the delta function. We employ $\mathbf{0}$ to denote the all-zeros vector or matrix and $\mathbf{0}_{W\times L}$ to represent the all-zeros matrix with a specified size of $W \times L$. We use $\mathbf{I}_L$ to denote the identity matrix of $L \times L$ dimension, and $\mathbf{I}_{W\times L}$ represents the first $W (W < L)$ columns of $\mathbf{I}_L$.

II. 3D MIMO-OFDM System and Channel Characteristics

In this section, we start with the 3D massive MIMO-OFDM channel modeling, followed by introducing a new type of fingerprint extracted from the channel in the angle-delay domain. Then the fingerprint in the angle-delay domain is compared with that in the space-frequency domain. At last, we propose a noise reduction method by exploiting the sparsity of the channel in the angle-delay domain.

A. Channel Model

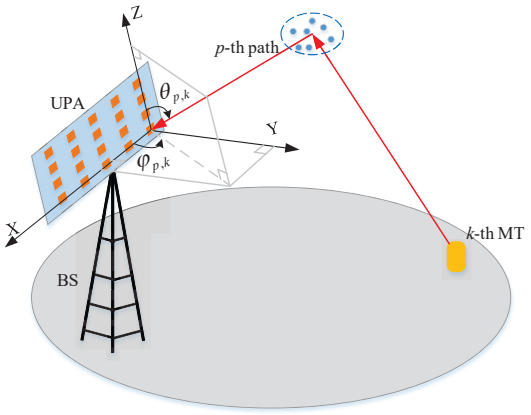


Fig. 1: AOA of the received signal at the BS antennas.

We consider the uplink transmission in a wide-band massive MIMO wireless system as shown in Fig. 1. The BS is equipped with a UPA, comprising N antennas in each row and M antennas in each column, placed in the X-Z plane. Then the total number of antennas at the BS is denoted as $N_a = MN$. The MT is equipped with an omni-directional antenna. Before reaching the receiver, the signals usually undergo scattering, reflection, refraction, and diffraction, which lead to a multipath propagation scenario [30]. Assume the number of multipaths is N_p and the number of the MTs is K, the AOA of the pth path for the kth MT can be decomposed into the elevation angle $0 \leq \theta_{p,k} \leq \pi$ in the vertical direction and the azimuth angle $0 \leq \varphi_{p,k} \leq \pi$ in the horizontal direction, as shown in Fig. 1. Then, the array response vector $\mathbf{e}(\theta_{p,k}, \varphi_{p,k})$ can be written as

$$\mathbf{e}(\theta_{p,k}, \varphi_{p,k}) = \mathbf{e}^{(v)}(\theta_{p,k}) \otimes \mathbf{e}^{(h)}(\theta_{p,k}, \varphi_{p,k}), \tag{1}$$

with

$$\begin{aligned}\mathbf{e}^{(v)}(\theta_{p,k}) = [1, e^{-\bar{\jmath}2\pi \frac{d_t^{(v)}}{\lambda_c} \cos\theta_{p,k}}, \cdots, \\ e^{-\bar{\jmath}2\pi(M-1)\frac{d_t^{(v)}}{\lambda_c} \cos\theta_{p,k}}]^T,\end{aligned} \tag{2}$$

and

$$\begin{aligned}\mathbf{e}^{(h)}(\theta_{p,k}, \varphi_{p,k}) = [1, e^{-\bar{\jmath}2\pi \frac{d_t^{(h)}}{\lambda_c} \sin\theta_{p,k}\cos\varphi_{p,k}}, \cdots, \\ e^{-\bar{\jmath}2\pi(N-1)\frac{d_t^{(h)}}{\lambda_c} \sin\theta_{p,k}\cos\varphi_{p,k}}]^T,\end{aligned} \tag{3}$$

where $d_t^{(v)}$ and $d_t^{(h)}$ are the antenna spacings in the column and row, respectively, and λ_c is the carrier wavelength.

We consider OFDM modulation with N_c sub-carriers and sample interval T_s. We use $T_c = N_c T_s$ and $T_g = N_g T_s$ to denote the OFDM symbol duration and the cyclic prefix (a.k.a. guard interval) duration, respectively. We assume the cyclic prefix duration T_g is larger than the maximum channel delay of all the MTs, i.e., $T_g > \tau_{\max}$. Let $r_{p,k} = \frac{\tau_{p,k}}{T_s}$, where $\tau_{p,k}$ is the TOA of the pth path for the kth MT. The frequency of the lth sub-carrier is $f_l = \frac{l}{T_c}$. We assume different paths are wide sense stationary (WSS) and independent [34], so that the channel frequency response (CFR) associated with the kth MT and the lth sub-carrier is written as

$$\mathbf{h}_{k,l} = \sum_{p=1}^{N_p} a_{p,k}\mathbf{e}(\theta_{p,k}, \varphi_{p,k}) e^{-\bar{\jmath}2\pi \frac{l r_{p,k}}{N_c}}, \tag{4}$$

where $a_{p,k} \sim \mathcal{CN}(0, \sigma_{p,k}^2)$ denotes the complex path gain of the pth path. The space-frequency domain channel response matrix (SFCRM) of the kth MT known to the BS can be denoted by the concatenation of $\mathbf{h}_{k,l}$, i.e.,

$$\mathbf{H}_k \triangleq [\mathbf{h}_{k,0}, \mathbf{h}_{k,1}, \cdots, \mathbf{h}_{k,N_c-1}] \in \mathbb{C}^{N_a \times N_c}. \tag{5}$$

B. Fingerprint from Angle-Delay Domain

The multipath channel is determined by the surrounding scatterer environment [35]. For different positions in the 3D space, the surrounding scatterer environments are different, which leads to different channel parameters. Thus, the channel itself can be regarded as an indicator of the position. The SFCRM describes the space-frequency domain characteristics. For positioning, it needs stable channel information, so we introduce the space-frequency domain channel covariance matrix (SFCCM) of the kth MT as

$$\mathbf{T}_k \triangleq E\left\{\text{vec}\left\{\mathbf{H}_k\right\} \text{vec}^H\left\{\mathbf{H}_k\right\}\right\} \in \mathbb{C}^{N_a N_c \times N_a N_c}. \tag{6}$$

Composed of the second-order statistical information of SFCRM, the SFCCM removes the influence of small-scale fading, and remains unchanged for quite a long time period. Hence, the SFCCM can be regarded as a kind of fingerprint for massive MIMO-OFDM systems. However, the number of elements in SFCCM is proportional to the square of the number of antennas N_a^2 and the square of the number of sub-carriers N_c^2. When the SFCCM is used as the positioning fingerprint in massive MIMO-OFDM systems, the computational complexity and storage overhead are extremely large.

We now shift our focus to the channel in the angle-delay domain. In the realistic wireless communication scenario, the signal is transmitted through multipaths with a series of AOAs and TOAs to the receiver. The AOAs and TOAs vary when the MT's position changes, as shown in Fig. 1. Therefore, it is reasonable to build the correspondence between the position and the channel in the angle-delay domain.

Define $\mathbf{V}_M \in \mathbb{C}^{M \times M}$ as phase-shifted discrete Fourier transform (DFT) matrix with the (i,j)th elements given by $[\mathbf{V}_M]_{i,j} \triangleq \frac{1}{\sqrt{M}} e^{-\bar{\jmath}2\pi \frac{i(j-M/2)}{M}}$. Define $\mathbf{F}_{N_c \times N_g} \in \mathbb{C}^{N_c \times N_g}$ as the matrix composed of the first N_g columns of DFT matrix $\mathbf{F}_{N_c}$ with the (i,j)th element given by $[\mathbf{F}_{N_c \times N_g}]_{i,j} \triangleq \frac{1}{\sqrt{N_c}} e^{-\bar{\jmath}2\pi \frac{ij}{N_c}}$. Using the above definitions, we transfer the space-frequency channel into the angle-delay one, and $\mathbf{G}_k$ is

referred to as the angle-delay domain channel response matrix (ADCRM) of the kth MT, given by

$$\mathbf{G}_k \triangleq \frac{1}{\sqrt{MNN_c}} \left(\mathbf{V}_M^H \otimes \mathbf{V}_N^H\right) \mathbf{H}_k \mathbf{F}^*_{N_c \times N_g} \in \mathbb{C}^{N_a \times N_g}. \quad (7)$$

Then the angle-delay domain channel power matrix (ADCPM) of the kth MT is introduced as

$$\mathbf{\Omega}_k \triangleq E\left\{\mathbf{G}_k \odot \mathbf{G}_k^*\right\} \in \mathbb{R}^{N_a \times N_g}, \quad (8)$$

with

$$[\mathbf{\Omega}_k]_{i,j} \triangleq E\left\{\left|[\mathbf{G}_k]_{i,j}\right|^2\right\}. \quad (9)$$

With the following function

$$f_M(x) \triangleq \frac{\sin(M\pi x)}{M\sin(\pi x)}, \quad (10)$$

we present the asymptotic behavior of the elements of ADCPM as in Theorem 1.

Theorem 1: For massive MIMO-OFDM systems with UPA at the BS, when a certain dimension of the UPA approaches infinity and the number of sub-carriers $N_c \to \infty$, the ADCPM is concentrated on specific positions in the horizontal angle-delay domain or vertical angle-delay domain, namely, for $i = 0, 1, \cdots, N_a - 1$ and $j = 0, 1, \cdots, N_g - 1$,

$$\lim_{N\to\infty, N_c\to\infty} \left([\mathbf{\Omega}_k]_{i,j} - \sum_{p=1}^{N_p} \sigma_{p,k}^2 f_M^2\left(\frac{\bar{m}_{p,k} - \lfloor i/N \rfloor}{M}\right) \cdot \delta\left(\langle i \rangle_N - \bar{n}_{p,k}\right) \delta\left(j - r_{p,k}\right) \right) = 0, \quad (11)$$

$$\lim_{M\to\infty, N_c\to\infty} \left([\mathbf{\Omega}_k]_{i,j} - \sum_{p=1}^{N_p} \sigma_{p,k}^2 f_N^2\left(\frac{\bar{n}_{p,k} - \langle i \rangle_N}{N}\right) \cdot \delta\left(\lfloor i/N \rfloor - \bar{m}_{p,k}\right) \delta\left(j - r_{p,k}\right) \right) = 0, \quad (12)$$

where

$$\bar{m}_{p,k} = \frac{M}{2} + \frac{M d_t^{(v)}}{\lambda_c} \cos\theta_{p,k}, \quad (13)$$

$$\bar{n}_{p,k} = \frac{N}{2} + \frac{N d_t^{(h)}}{\lambda_c} \sin\theta_{p,k} \cos\varphi_{p,k}. \quad (14)$$

Further, in the extreme case of $M \to \infty$, $N \to \infty$, and $N_c \to \infty$, the ADCPM is concentrated on the $(\bar{m}_{p,k}N + \bar{n}_{p,k})$th angle direction and the $r_{p,k}$th delay direction, namely, for $i = 0, 1, \cdots, N_a - 1$ and $j = 0, 1, \cdots, N_g - 1$,

$$\lim_{M\to\infty, N\to\infty, N_c\to\infty} \left([\mathbf{\Omega}_k]_{i,j} - \sum_{p=1}^{N_p} \sigma_{p,k}^2 \delta\left(i - \bar{m}_{p,k}N - \bar{n}_{p,k}\right) \cdot \delta\left(j - r_{p,k}\right) \right) = 0. \quad (15)$$

Proof: See Appendix A. □

Remark 1: For 3D massive MIMO-OFDM systems, the SFCRM in the space-frequency domain is translated into the ADCPM in the angle-delay domain via using (7) and (8). Theorem 1 reveals that the (i, j)th element in the ADCPM corresponds to the average channel power of the ith AOA and the jth TOA. The ADCPM is stationary in the sense that it keeps unchanged over a relatively long period for a given location.

Remark 2: In Theorem 1, $\delta(\cdot)$ determines the sparsity and specifies the sparsity pattern. When $N_c \to \infty$, $M \to \infty$, and/or $N \to \infty$, the ADCPM is asymptotically a sparse matrix in the sense that most elements are equal to zero. For the finite N_c, M, and N, it will be shown later by numerical results that the sparsity maintains.

Remark 3: From Theorem 1, the constitution of the ADCPM of the kth MT depends on the multipath information denoted by the set $\left\{(\sigma_{p,k}^2, \tau_{p,k}, \theta_{p,k}, \varphi_{p,k})\right\}_{p=1}^{N_p}$. Given two MTs k and k' located at different positions, with high probability that the multipath components of the two MT are different due to the different scatterer environments, i.e., $\left\{(\sigma_{p,k}^2, \tau_{p,k}, \theta_{p,k}, \varphi_{p,k})\right\}_{p=1}^{N_p} \neq \left\{(\sigma_{p,k'}^2, \tau_{p,k'}, \theta_{p,k'}, \varphi_{p,k'})\right\}_{p=1}^{N_p'}$, so are the corresponding ADCPMs, i.e., $\mathbf{\Omega}_k \neq \mathbf{\Omega}_{k'}$. As such, the ADCPM can be a unique indicator to discriminate MTs from different geographical positions.

Compared with fingerprints proposed in previous literatures, e.g., AOA [4], [5], TOA [6], and RSS [5], [9], the ADCPM not only contains all these three characteristics but also provides additional multipath resolution. Besides, the ADCPM is statistical information hence can provide stable and reliable fingerprints for positioning. It turns out that, as a type of fingerprint, the ADCPM is superior to RSS, TOA, and AOA. In addition, the ADCPM fingerprint is also superior to the SFCCM, although both of them have the same asymptotic positioning capability. In Section II-C, we will demonstrate the consistency of SFCCM and ADCPM as fingerprints in the sense that both fingerprints have the same discrimination capability, but the size of the latter is substantially smaller than the former, which leads to much lower storage requirement and computational complexity. Further, compared with other fingerprints, sparsity is a unique characteristic of the ADCPM. By exploiting the sparsity of the ADCPM, we propose a noise reduction method to realize robust positioning under a noisy environment in Section II-D. And the efficacy of this method will be shown by numerical results in Section V-D. Therefore, as a highly position-related and robust feature, the ADCPM will be used as the positioning fingerprint of this paper.

C. Consistency of SFCCM and ADCPM

To evaluate the similarity between the fingerprints at different positions, we first introduce the concept of collinearity. The matrix collinearity is utilized as a similarity criterion to characterize users' channel separability [36]. Inspired by this, we employ the matrix collinearity to measure the similarity between two fingerprints. Given two ADCPM fingerprints $\mathbf{\Omega}_k$ and $\mathbf{\Omega}_{k'}$, the fingerprint collinearity is defined as [37]

$$J\left(\mathbf{\Omega}_k, \mathbf{\Omega}_{k'}\right) = \frac{\mathrm{Tr}\left\{\mathbf{\Omega}_k^H \mathbf{\Omega}_{k'}\right\}}{\|\mathbf{\Omega}_k\|_F \|\mathbf{\Omega}_{k'}\|_F}, \quad (16)$$

where $\|\cdot\|_F$ denotes the Frobenius norm. The fingerprint collinearity is a normalized metric with value between 0 and 1, where 1 is reached for collinear fingerprint matrices and 0 for orthogonal fingerprint matrices.

Theorem 2: For positioning in massvie MIMO-OFDM systems with UPA at the BS, when $M \to \infty$, $N \to \infty$, and $N_c \to \infty$, the similarity between the SFCCMs at different positions is consistent with that between the ADCPMs at different positions, i.e.,

$$J\left(\mathbf{T}_k, \mathbf{T}_{k'}\right) = J\left(\mathbf{\Omega}_k, \mathbf{\Omega}_{k'}\right). \tag{17}$$

Proof: See Appendix B. □

Remark 4: From Theorem 2, when $M \to \infty$, $N \to \infty$, and $N_c \to \infty$, the ADCPM and the SFCCM are asymptotically equivalent with respect to fingerprint collinearity. For the finite M, N, and N_c, it can be verified that the collinearity between two ADCPM fingerprints is approximately equivalent to the SFCCM counterpart. But compared to the SFCCM counterpart, the ADCPM reduces the size of fingerprint at a ratio of

$$r_{\text{AD-SF}} = \frac{N_a N_c^2}{N_g} \ll 1, \tag{18}$$

which leads to a significant reduction in both the computational complexity and the storage overhead for fingerprint-based positioning.

D. Noise Reduction by Exploiting the Sparsity of ADCPM

To encourage the sparsity of the ADCPM, we adopt a filter with an element-wise threshold of ADCPM such that the values below this threshold are set to 0. By doing so, the noise contamination in the ADCPMs is reduced to a certain extent.

We first define $\bar{\mathbf{G}}_k$ as the extended ADCRM, given by

$$\bar{\mathbf{G}}_k \triangleq \frac{1}{\sqrt{MNN_c}} \left(\mathbf{V}_M \otimes \mathbf{V}_N\right)^H \mathbf{H}_k \mathbf{F}_{N_c}^* \in \mathbb{C}^{N_a \times N_c}. \tag{19}$$

Given the channel estimate

$$\tilde{\mathbf{H}}_k = \mathbf{H}_k + \mathbf{Z}_k, \tag{20}$$

where $\mathbf{Z}_k \in \mathbb{C}^{N_a \times N_c}$ denotes the noise contamination and is independent of the SFCRM $\mathbf{H}_k$, we calculate the estimate of the extended ADCRM as

$$\tilde{\mathbf{G}}_k = \bar{\mathbf{G}}_k + \frac{1}{\sqrt{MNN_c}} \left(\mathbf{V}_M \otimes \mathbf{V}_N\right)^H \mathbf{Z}_k \mathbf{F}_{N_c}^*. \tag{21}$$

According to Lemma 3 in Appendix B, the extended ADCRM $\bar{\mathbf{G}}_k$ can be represented by the ADCRM $\mathbf{G}_k$ in the case of $N_c \to \infty$, given by

$$\bar{\mathbf{G}}_k = \left[\begin{array}{cc} \mathbf{G}_k & \mathbf{0}_{N_a \times (N_c - N_g)} \end{array} \right]. \tag{22}$$

Then the estimate of extended ADCPM can be divided into two parts:

$$\left[\tilde{\mathbf{\Omega}}_k\right]_j = \begin{cases} \left[\mathbf{\Omega}_k\right]_j + \left[\mathbf{D}_k\right]_j, & j = 0, 1, \cdots, N_g - 1, \\ \left[\mathbf{D}_k\right]_j, & j = N_g, N_g + 1, \cdots, N_c - 1, \end{cases} \tag{23}$$

where $\mathbf{D}_k$ is the power matrix of noise contamination, i.e.,

$$\mathbf{D}_k \triangleq E\left\{\mathbf{Z}_k \odot \mathbf{Z}_k^*\right\}. \tag{24}$$

The second part of $\tilde{\mathbf{\Omega}}_k$ is used to approximate the average power of noise contamination as

$$\bar{\sigma}_z^2 = \frac{1}{N_a \left(N_c - N_g\right)} \sum_{i=0}^{N_a - 1} \sum_{j=N_g}^{N_c - 1} \left[\tilde{\mathbf{\Omega}}_k\right]_{i,j}. \tag{25}$$

According to Theorem 1, the ADCPM is sparse with several non-zero significant elements and other zero elements. Thus we have

$$\left[\mathbf{\Omega}_k\right]_{i,j} + \left[\mathbf{D}_k\right]_{i,j} = \begin{cases} \left[\mathbf{\Omega}_k\right]_{i,j} + \left[\mathbf{D}_k\right]_{i,j}, & \text{significant elements}, \\ \left[\mathbf{D}_k\right]_{i,j}, & \text{zero elements}. \end{cases} \tag{26}$$

Then the first part of $\tilde{\mathbf{\Omega}}_k$ is exploited to get the ADCPM with noise reduction, namely, for $i = 0, 1, \cdots, N_a - 1$ and $j = 0, 1, \cdots, N_g - 1$,

$$\left[\mathbf{\Omega'}_k\right]_{i,j} = \begin{cases} \left[\tilde{\mathbf{\Omega}}_k\right]_{i,j}, & \left[\tilde{\mathbf{\Omega}}_k\right]_{i,j} > \xi, \\ 0, & \left[\tilde{\mathbf{\Omega}}_k\right]_{i,j} \leqslant \xi, \end{cases} \tag{27}$$

where $\xi = \beta \bar{\sigma}_z^2$ is the noise reduction threshold, and β is the threshold coefficient determined by the specific distribution of the noise contamination. Compared with the SFCCM counterparts which losses of sparsity, the noise power in ADCPM fingerprints is reduced without affecting much the similarity between the ADCPM fingerprints.

III. Convolution Neural Network for Positioning

Given the ADCPM as the fingerprint in Section II-B, the problem arises as to how to realize positioning by exploiting the structural properties of the ADCPM. As the ADCPM fingerprint can be seen as an image, the widely used end-to-end image recognition method - convolution neural network (CNN) empowered deep learning - can be applied here for positioning. Thanks to its characteristics of space invariant, parameter sharing, and hierarchical representations, CNN is more efficient than the traditional fully-connected networks when dealing with large dimensional inputs [38]. For 3D massive MIMO-OFDM systems, the high dimensional ADCPM has sparsity patterns, which suggests that CNN could be an ideal positioning method to extract the positioning-related features from the ADCPM and convert these features into position information with relatively low computational complexity.

A. The Sparse ADCPM as Input

As shown in Fig. 2, the asymptotic property of the ADCPM sparsity pattern presented in Theorem 1 still maintains in the practical setting with a finite number of antennas and limited bandwidth. Due to the sparsity pattern, the ADCPM fingerprint makes the difference of the channels at different positions more distinguishable. As such, the features of the channel are easier to be extracted by a neural network. In fact, the feature maps in the higher layers of CNN are also sparse because they solely focus on the discriminant structure within the input picture [39]. By using the sparse ADCPM as the input, it is easier for neural network to capture the characteristic information of the channel, thereby simplifying

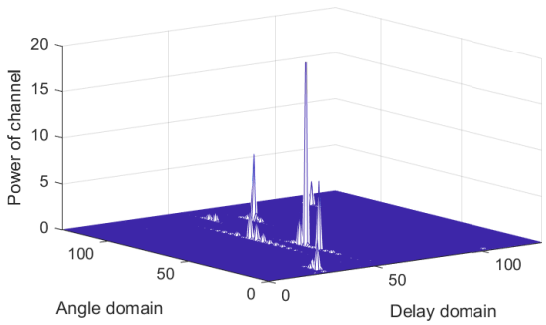


Fig. 2: The extracted ADCPM fingerprint in the case of $M = 8$, $N = 16$, and $N_g = 128$.

the neural network structure and speeding up the convergence of the neural network.

The rows and the columns of the ADCPM correspond to the angle and the delay domains, respectively. Noticing that the angle in the 3D space can be described by a pair of vertical and horizontal angles, we reshape the ADCPM into a three-dimensional tensor in the following way

$$[\mathbf{X}_k]_{m,n,j} = [\mathbf{\Omega}_k]_{mN+n,j}\,, \tag{28}$$

where $\mathbf{X}_k \in {}^{M\times N\times N_g}$ is the 3D ADCPM of the kth MT with its three dimensions indicating vertical angle, horizontal angle, and delay, respectively. We use hereafter the reshaped 3D ADCPM as the input of neural networks.

B. Regression-Oriented Positioning

The CNN-based regression-oriented positioning is actually a complex non-linear function. Denote by $f(\cdot)$ such a function that estimates the true position $\mathbf{a}_k = (x, y, z)$ according to the MT's 3D ADCPM $\mathbf{X}_k$, that is, $\hat{\mathbf{a}}_k = f(\mathbf{X}_k)$, where $\hat{\mathbf{a}}_k = (\hat{x}, \hat{y}, \hat{z})$ is the estimate by the CNN.

We use regression analysis to find the mapping function through minimizing the localization error between the true coordinate and the estimated one as

$$e_k = \|\hat{\mathbf{a}}_k - \mathbf{a}_k\|_2\,, \tag{29}$$

where $\|\cdot\|_2$ denotes the ℓ_2-norm. Note that the traditional CNN models are often used for image classification with the last layer activated by a softmax function. As a matter of fact, the CNN model itself can be regarded as a regression function if the softmax function is replaced with a fully connected layer without activation function. If we use $\mathbf{T}_0$ to denote the input of the last layer, then the estimated position produced by the last layer can be written by

$$\hat{\mathbf{a}} = \mathbf{W}\mathrm{vec}\{\mathbf{T}_0\} + \mathbf{b}, \tag{30}$$

where $\mathbf{W}$ and $\mathbf{b}$ are the parametric weight matrix and bias vector of the last fully connected layer, respectively. Let $\boldsymbol{\theta}$ be the set of trainable parameters of the neural network. The mean square error (MSE) of the training data set can be employed as the cost function, given by

$$J(\boldsymbol{\theta}) = \frac{1}{N_{\text{train}}}\sum_{i=1}^{N_{\text{train}}}\|\mathbf{a}_i - \hat{\mathbf{a}}_i\|_2^2 + \frac{\lambda}{2}\boldsymbol{\theta}^T\boldsymbol{\theta}, \tag{31}$$

where N_{train} is the number of training samples, the second term employs the L_2 regularization to avoid over-fitting, and λ is the weight factor of the L_2 regularization.

IV. Proposed 3D CNN Structure

Inspired by the 3D structure of $\mathbf{X}_k$, we propose to use the 3D CNN to realize the mapping function. Fig. 3 shows the overall network architecture of the proposed 3D CNN for fingerprint-based positioning, which is composed of a convolution refinement module, three extended 3D Inception modules, and a regression module. In addition, the max pooling is used for downsampling, and its configurable parameters are presented in the form of "size/stride". The convolution refinement module first refines the elementary feature maps from the 3D ADCPM. Then we modify and extend the Inception module into a 3D form to extract the advanced feature maps. Further, the regression module estimates the 3D position by employing a global average pooling layer and a fully connected layer without activation function.

Before proceeding, we introduce the 3D convolution-normalization-activation (CNA) layer, an elementary building block used in our 3D CNN. The 3D CNA layer consists of three parts: a 3D convolution, a BN transform, and an activation function. Consider the input feature maps $\mathbf{I} \in \mathbb{R}^{H\times W\times L\times P}$, where H, W, L, and P denote the height, width, length, and the number of channels of $\mathbf{I}$, respectively. Given the 3D convolution kernel $\mathbf{K} \in \mathbb{R}^{K_1\times K_2\times K_3\times P\times Q}$, where K_1, K_2, and K_3 are the kernel size of the height, width, and length, respectively, and Q is the number of output channels. The convolution is performed with zero-padding [38] and the strides are set to be 1. By convolving the 3D kernel with the input feature maps, the 3D convolution yields output $\mathbf{O} \in \mathbb{R}^{H\times W\times L\times Q}$ with the (h, w, l, q)th element given by

$$[\mathbf{O}]_{h,w,l,q} = \sum_{k_1,k_2,k_3}\sum_{p}[\mathbf{K}]_{k_1,k_2,k_3,p,q}\cdot[\mathbf{I}]_{h+k_1,w+k_2,l+k_3,p}. \tag{32}$$

Right after the 3D convolution, we adopt the BN operation to improve the convergence rate and generalization performance [40], followed by a rectified linear unit (ReLU) as the activation function. That is, the output feature maps of the 3D CNA layer $\mathbf{T} \in \mathbb{R}^{H\times W\times L\times Q}$ are given by

$$[\mathbf{T}]_{h,w,l,q} = \max\left(0, \mathrm{BN}\left([\mathbf{O}]_{h,w,l,q}\right)\right), \tag{33}$$

where $\mathrm{BN}(\cdot)$ represents the BN operation. Note that we discard the bias term in (32) as suggested in [40]. The individual modules used in the proposed 3D CNN are constituted by several 3D CNA layers. In what follows, we describe the individual modules in detail.

A. Convolution Refinement Module

The convolution refinement module is designed to extract features from the input 3D ADCPM. When designing the

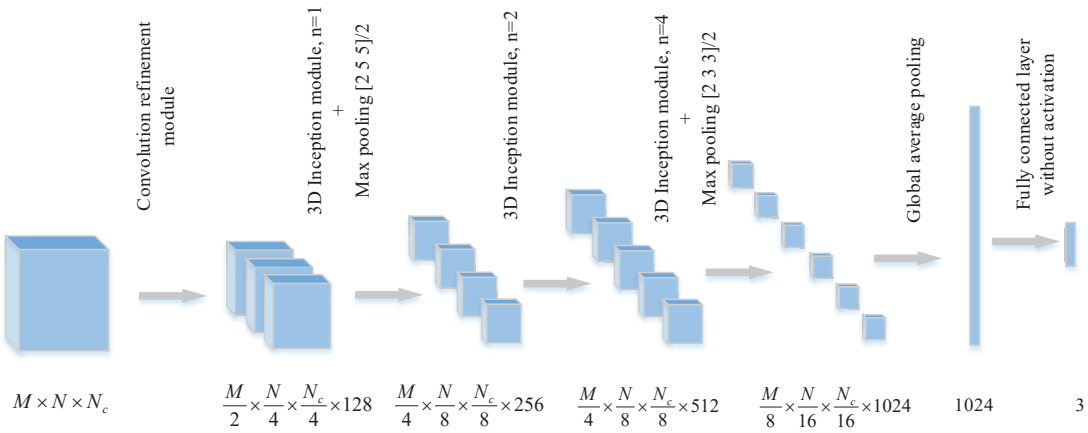


Fig. 3: The network architecture of the proposed 3D CNN for fingerprint-based positioning.

module, we take into account the corresponding physical meaning of each dimension of the 3D ADCPM. While the existing designs use the symmetric kernel, for which the size of each dimension is equal, we propose to use the asymmetric convolution kernel, based on the intuition that the size of kernel should reflect the correlation statistics of the corresponding dimension. As a consequence, the size of the subsequent max pooling should also change accordingly. From Fig. 2, we observe that the input units are more concentrated in the angular dimension and relatively dispersed in the delay dimension. It suggests that the size of the corresponding convolution kernel size of the delay dimension can be larger than that of the angle dimension. At the same time, it is worth noting that the vertical and horizontal angles are somehow correlated, while the delay dimension is independent of the other two. Therefore, the delay-vertical angle and the delay-horizontal angle domains demand special treatments, compared with the vertical-horizontal angle domain. This motivates our design of the convolution refinement module.

We utilize the case of $M = 8$, $N = 16$, and $N_g = 128$ as an example to describe the convolution refinement module. The structure of the convolution refinement module is shown in Fig. 4(a). The expression in the box is in the form of "size*number" for 3D CNA's convolution kernel and "size/stride" for max pooling. As one of the key designs, we build two branches in this module, each of which consists of two 3D CNA layers and one max pooling layer. The left branch aims to extract the information from the delay-horizontal angle domain, while the right branch is dedicated to the information from the delay-vertical angle domain. It is worth noting that both branches use asymmetrical convolution kernels and pooling size. The outputs of two branches are combined using kernel-wise concatenation. After that, we employ a 3D CNA layer with the symmetric kernel to extract the information from all three dimensions, followed by a max pooling to downsample the feature maps.

B. 3D Inception Module

In the deeper layers, we use the 3D Inception module to extract more precise features. The Inception module is a combination of 2D convolutions with different kernel sizes, e.g., 1×1, 3×3, and 5×5, with a parallel average-pooling operation [31]. We extend the Inception module into a 3D form by replacing the 2D convolution with the 3D convolution. The structure has four parallel branches, whose outputs are concatenated into a single output vector that servers as the input of the next stage, as shown in Fig. 4(b), where n is the factor of kernel number. The $5 \times 5 \times 5$ kernel is replaced by two cascaded $3 \times 3 \times 3$ kernels to reduce the computational overhead, as in [31]. Since the correlation statistics in the deeper layers are unknown, the usage of 3D convolution kernels with different sizes can avoid the loss of important features.

C. Regression Module

The regression module includes a global average pooling and a fully connected layer without activation function. While existing designs tend to add several layers of fully connected layers before the output layer of the network, we use a global average pooling to replace them. The global average pooling takes the average of each feature map [41], and reshapes the output of the convolution layers for the final fully connected layer. By replacing the middle fully connected layers with the global average pooling, we can avoid the huge parameters brought from the middle fully connected layers. Thus, the global pooling improves convergence rate while reducing overfitting. The output of global average pooling in vector form is fed into the fully connected layer indicated by (30) directly. The output of the regression module is the estimated coordinates of the 3D position.

V. Simulation Results

In this section, we first introduce the simulation setup, followed by the evaluation of the positioning accuracy, time, and storage overhead of the proposed 3D CNN-enabled positioning method. Next, we demonstrate the influence of antenna geometry on our proposed 3D CNN-enabled positioning method. At last, we evaluate the robustness of our positioning method with the noise-contaminated inputs.

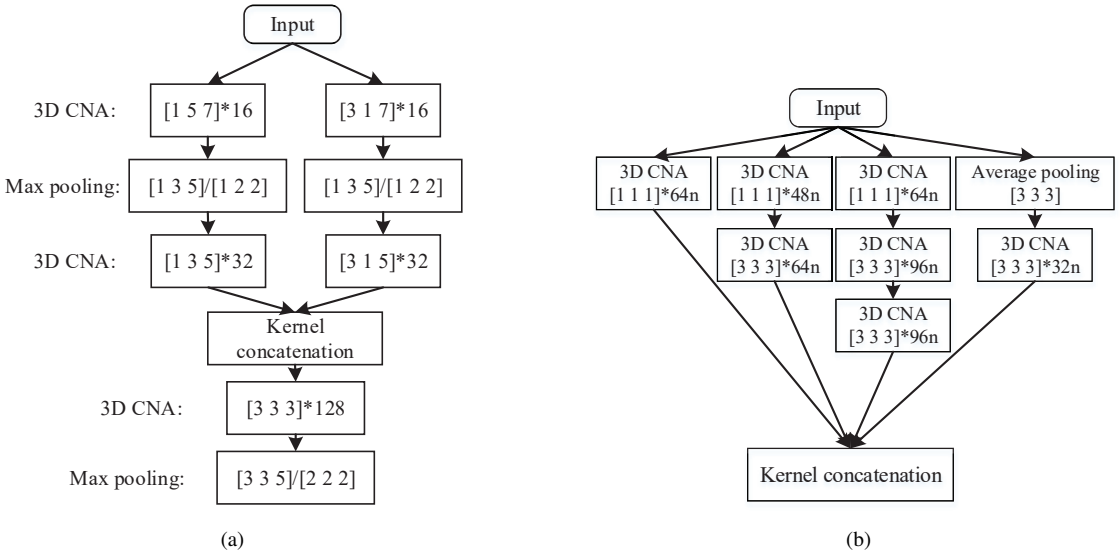


Fig. 4: The proposed modules. (a) The proposed convolution refinement module with two branches of CNA layers taking special care of the delay-vertical angle and delay-horizontal angle domains; (b) the modified 3D Inception module.

A. Simulation Setup

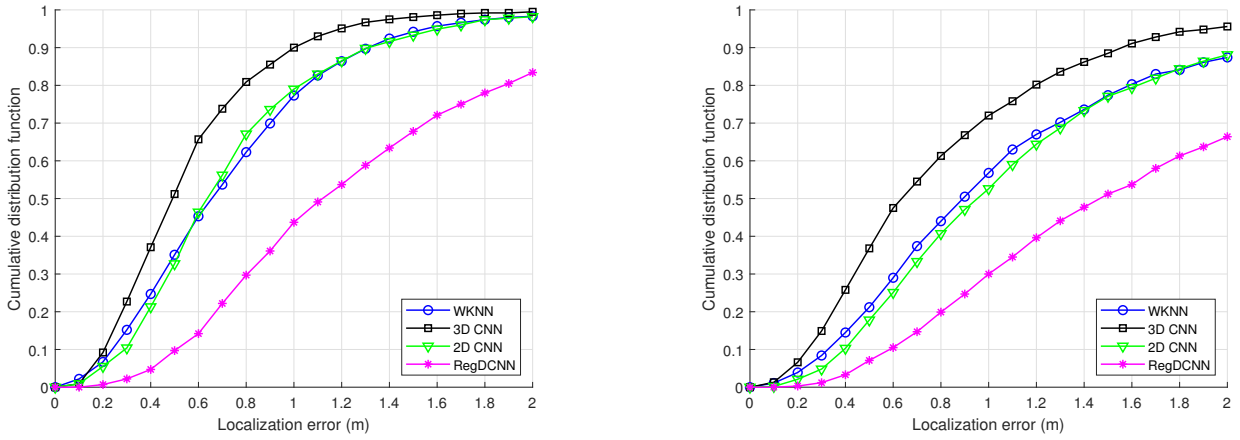
Our simulation is based on the QuaDRiGa model, which can emulate the real-world scenario accurately and has been validated by actual measurements [42]. In the QuaDRiGa model, a stochastic part first generates the channel coefficients based on measurement data and calculates the 3D positions of the scattering clusters. Then a time evolution technique proposed in [43] and developed in [42] is used to simulate the position-related channels at different spatial positions. We consider the 3GPP urban macro (UMa) NLOS and indoor NLOS scenarios [44] with the carrier frequency 2GHz. We choose the NLOS scenario rather than the line of sight (LOS) one because the positioning in the NLOS scenario is more challenging due to a lot of occlusions and reflections. The size of the positioning area is 30m $\times$ 30m $\times$ 9m, with the center point of its bottom surface coincides with the origin. The BS is equipped with a UPA at $(-100, 0, 25)$m, and the antenna plane is perpendicular to the ground, facing the positioning area. The MT is equipped with an omni-directional antenna. Unless otherwise specified, the transmission bandwidth is 20MHz, and the configuration of UPA is $M = 8$, $N = 16$. For simplicity but without much loss of generality, we divide the cube positioning area into three planes with heights 1.5m, 4.5m, and 7.5m, respectively. In the offline phase, the training points are uniformly selected on these three planes with an interval of 1m, by which the fingerprints and the corresponding 3D positions are collected. In the online phase, the test points are randomly distributed on these three planes with a total number of $N_{\text{test}} = 1000$, for which their positions are inferred by putting the fingerprint into the trained 3D CNN.

The fingerprints and the corresponding 3D positions of the training points are generated and saved using MATLAB 2018b, and the 3D CNN training and testing are processed using TensorFlow 1.9. Our simulation is carried out on a workstation equipped with two E5 2643v3 CPUs and one Titan X Pascal 12GB GPU. The time overhead mentioned below only refers to the run time on TensorFlow 1.9.

B. Comparison with Other Positioning Algorithms

To evaluate the performance of the proposed 3D CNN-enabled positioning method, we use the following three methods as the benchmarks.

- WKNN [11]: To verify the benefits of the neural network-enabled positioning, we simulate the WKNN positioning method in which the fingerprint collinearity in (16) is used to find the $\bar{K}$th nearest fingerprints and the corresponding reference positions. $\bar{K} = 4$ is adopted in our simulation.
- Regression-based deep CNN (RegDCNN) [12]: To demonstrate the benefits of the deliberated design of the 3D CNN model, we simulate the RegDCNN positioning method which only employs conventional symmetric convolution and pooling.
- Downgraded 2D CNN: To justify the motivation and the superiority of our 3D design, we propose a downgraded 2D CNN-enabled positioning method with the ADCPM presented by (8) as the input. In that case, the ADCPM is regarded as an image with sparse highlights (supports), where the vertical and horizontal angles are collocated in the same dimension. The overall structure of proposed 2D CNN consists of one 2D convolution refinement module, three 2D Inception modules, and one 2D regression module. The 2D convolution refinement module is a cascade of one layer of CNA, one layer of max pooling, two layers of CNA, and one layer of max pooling. The parameters of the 2D convolution kernels in three CNA layers are [15 15]*32, [7 7]*64, and [5 5]*128, respectively. The parameters of the two max pooling are both [5 5]/2. The 2D Inception module is modified from the 3D Inception module in Fig. 4(b) by replacing the 3D convolution



(a) (b)

Fig. 5: Comparisons of the localization error with different positioning methods in different scenarios: (a) UMa; (b) indoor.

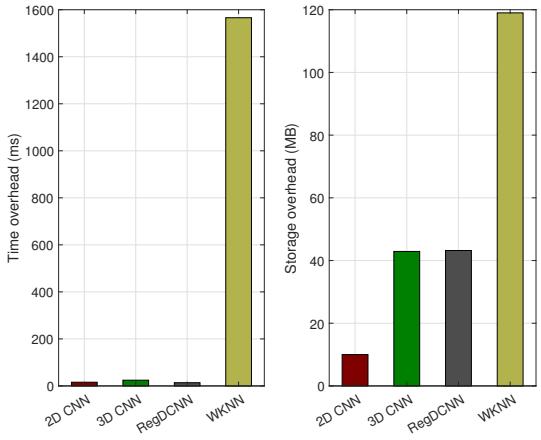


Fig. 6: The time overhead (left) and the storage overhead (right) with different positioning methods.

with the 2D convolution. The parameters of 2D Inception modules are n =1, 2, and 4, respectively. The first and the third Inception module are followed by max pooling with parameters [5 5]/2 and [3 3]/2, respectively.

We first compare the positioning accuracy of the four methods in both UMa and indoor scenarios. The cumulative distribution function (CDF) of the localization error using different methods is illustrated in Fig. 5(a) and Fig. 5(b). In the UMa scenario, the 3D CNN-enabled positioning method reaches the highest positioning accuracy with 90% of localization errors within 1m. Compared with the 3D counterpart, the 2D CNN method realizes the inferior positioning accuracy with only 80%, but is still superior to the WKNN positioning method with 78%. The RegDCNN has the worst positioning accuracy with only 44% within 1m, which does not contain the any use of the key designs in our paper. The observation of the simulation results in the indoor scenario is similar to that in the UMa scenario.

We focus on the computational complexity and storage overhead in the online phase while those in the training phase are ignored because the training phase is not needed often when the positioning environment remains unchanged. The computational complexity is measured by the run time overhead, which is defined as the time spent per user positioning in the online phase. And the storage overhead refers to the allocated storage resource for online positioning. Then we compare the time and storage overhead of the four positioning methods in Fig. 6. The RegDCNN requires the time overhead of 13.72ms and storage overhead of 43.2MB. The 2D CNN-enabled positioning method requires the least time overhead, e.g., 15.78ms, and the least storage overhead, e.g., 10MB. Compared with the 2D counterpart, the 3D CNN method gains higher positioning accuracy at the cost of higher time overhead (24.42ms running time) and storage overhead (42.9MB). The WKNN positioning method has the highest computational complexity (1566ms) and storage requirement (119MB), as it needs to store the database collected in the positioning area and then search through the positioning area to find the nearest positions. Thus the time and storage overhead of WKNN positioning method will increase with the expansion of the positioning area. In contrast, the 2D/3D CNN-enabled and RegDCNN positioning methods only need to store the trained parameters and infer the position directly through the trained regression networks, by which the time and storage overhead is independent of the size of the positioning area. Meanwhile, through deliberate design, the 3D CNN can realize much smaller localization error with nearly equal storage overhead compared with the RegDCNN.

C. Different Configurations of Antenna Array Geometry

To inspect how vertical and horizontal angles interact with each other, we compare the different antenna array geometry. Fig. 7 shows the CDF of localization error using different antenna array configurations, where the case of 1×128 corresponds to the uniform linear array (ULA). To make a fair comparison, we maintain the same number of antenna elements. It can be observed that the two UPAs perform

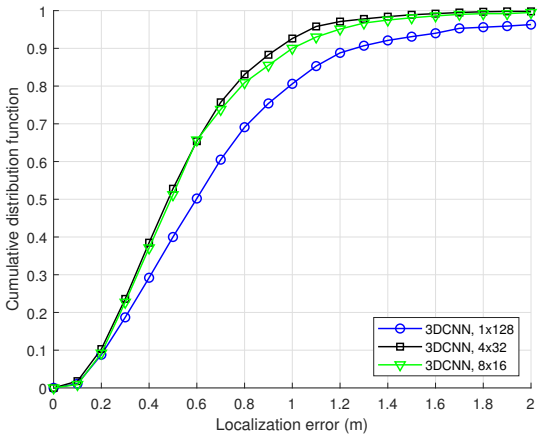


Fig. 7: The CDF of localization error with different number of BS antennas in the column and row of UPA.

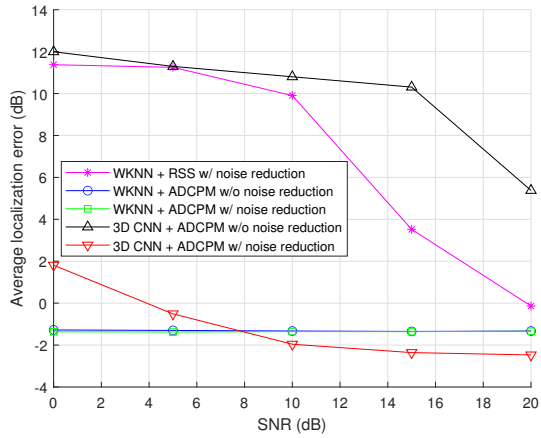


Fig. 8: The average localization error versus the SNR using different fingerprints with or without noise reduction.

better with 90% (8×16) and 93% (4×32) reliability for 1m positioning accuracy, in contrast to 80% by the ULA. Thus, providing additional angle resolution in the vertical direction, the UPA with a 3D CNN positioning method is more suitable for the 3D positioning than the ULA.

D. Positioning Robustness

In this subsection, we evaluate the robustness of our fingerprint extraction and 3D CNN-enabled positioning method against noise contamination. In the training phase, the artificial noiseless channel matrices are generated by Monte-Carlo method to train the model for 3D CNN or to construct the look-up table for WKNN. In the estimation phase, noisy channels are used for positioning. We adopt the noise reduction method proposed in Section II-D to reduce the noise contamination in the input ADCPM fingerprints, and use the average localization error in dB to evaluate the positioning performance under noisy environment, given as

$$\bar{e} = 10\log(\frac{1}{k}\sum_{k=1}^{N_{\text{test}}} e_k). \tag{34}$$

Fig. 8 presents the average localization error versus the SNR using different fingerprints with or without noise reduction. We observe that (1) our proposed 3D CNN using ADCPM fingerprint with noise reduction achieves the best performance in the medium and high SNR regime, while there is some performance degradation if noise reduction is not applied; (2) WKNN with ADCPM fingerprints is insensitive to the noise contamination in the fingerprints and therefore performs identically with or without noise reduction; and (3) Other fingerprints, such as RSS in [13] (i.e., a special case of SFCCM fingerprint), performs worse than the ADCPM fingerprint due to lack of sparsity when applying WKNN to both of them. As SNR grows, the performance of the learning-based method (i.e., 3D CNN) exceeds that of the searching-based one (i.e., WKNN) with considerably reduced computational complexity and storage requirement. By taking advantage of its sparse structure, it turns out that user positioning using ADCPM fingerprints can be resistant to noisy inputs. As such, with the noise reduction preprocessing, the proposed 3D CNN-enabled positioning method which employs the ADCPM fingerprint can realize robust positioning in noisy environment.

VI. Conclusion

In this paper, we advocated a learning-to-localize strategy for user positioning for 3D massive MIMO-OFDM systems, by proposing a deep learning model to exploit the sparsity properties of channel statistics in the 3D angle-delay domain. In particular, we employed the 3D ADCPM as the positioning fingerprint which contains stable and stationary multipath characteristics, e.g., delay, power, and angle in the vertical and horizontal directions. By casting user positioning as a 3D image recognition problem, we proposed a novel 3D CNN architecture with the aim of regression to realize the fingerprint-based positioning. The simulation results demonstrated that the proposed method can achieve higher user positioning accuracy with reduced computational complexity and storage overhead, and is robust to the noise contamination of the fingerprints.

Appendix A
Proof of Theorem 1

To prove Theorem 1, we provide the following preliminary results stated as Lemma 1. We define the following vectors

$$\tilde{\mathbf{e}}_{p,k}^{(v)} = \frac{1}{\sqrt{M}}\mathbf{V}_M^H \mathbf{e}^{(v)}(\theta_{p,k}), \tag{35}$$

$$\tilde{\mathbf{e}}_{p,k}^{(h)} = \frac{1}{\sqrt{N}}\mathbf{V}_N^H \mathbf{e}^{(h)}(\theta_{p,k}, \varphi_{p,k}). \tag{36}$$

Lemma 1: When $M \to \infty$ and $N \to \infty$, the AOA is extracted from array response vector explicitly via the DFT operation, given by

$$\lim_{M\to\infty} \tilde{\mathbf{e}}_{p,k}^{(v)} = \boldsymbol{\alpha}_M^{(\bar{m}_{p,k})}, \tag{37}$$

$$\lim_{N\to\infty} \tilde{\mathbf{e}}_{p,k}^{(h)} = \boldsymbol{\alpha}_N^{(\bar{n}_{p,k})}, \tag{38}$$

where

$$\boldsymbol{\alpha}_i^{(j)} = [\underbrace{0, \cdots 0}_{j}, 1, \underbrace{0, \cdots, 0}_{i-j-1}]^T. \tag{39}$$

Proof: The ith element of $\tilde{\mathbf{e}}_{p,k}^{(v)}$ is calculated as

$$\begin{aligned}[\tilde{\mathbf{e}}_{p,k}^{(v)}]_i &= \frac{1}{\sqrt{M}}\sum_{m=0}^{M-1}\frac{1}{\sqrt{M}}e^{\bar{\jmath}2\pi\frac{m(i-M/2)}{M}}e^{-\bar{\jmath}2\pi m\frac{d_t^{(v)}}{\lambda_c}\cos\theta_{p,k}}\\ &= \frac{1}{M}e^{-\bar{\jmath}\pi(M-1)\left(\frac{d_t^{(v)}}{\lambda_c}\cos\theta_{p,k}-\frac{i}{M}+\frac{1}{2}\right)}\\ &\quad\cdot\frac{\sin\left(M\pi\left(\frac{d_t^{(v)}}{\lambda_c}\cos\theta_{p,k}-\frac{i}{M}+\frac{1}{2}\right)\right)}{\sin\left(\pi\left(\frac{d_t^{(v)}}{\lambda_c}\cos\theta_{p,k}-\frac{i}{M}+\frac{1}{2}\right)\right)}\\ &= e^{-\bar{\jmath}\pi(M-1)\left(\frac{d_t^{(v)}}{\lambda_c}\cos\theta_{p,k}-\frac{i}{M}+\frac{1}{2}\right)}\\ &\quad\cdot f_M\left(\frac{d_t^{(v)}}{\lambda_c}\cos\theta_{p,k}-\frac{i}{M}+\frac{1}{2}\right),\end{aligned} \tag{40}$$

where $f_M(\cdot)$ is defined in (10). From (40), $[\tilde{\mathbf{e}}_{p,k}^{(v)}]_i \neq 0$ if and only if $f_M\left(\frac{d_t^{(v)}}{\lambda_c}\cos\theta_{p,k}-\frac{i}{M}+\frac{1}{2}\right)\neq 0$. Further, $\lim_{M\to\infty} f_M(x)\neq 0$ if and only if $\sin(\pi x)=0$, and in this case we have

$$\lim_{M\to\infty} f_M(x)|_{x=0,\pm1,\cdots} = \frac{\cos(M\pi x)}{\cos(\pi x)}. \tag{41}$$

Hence, $\lim_{M\to\infty} f_M\left(\frac{d_t^{(v)}}{\lambda_c}\cos\theta_{p,k}-\frac{i}{M}+\frac{1}{2}\right)\neq 0$ if and only if the following constraint is satisfied

$$\frac{d_t^{(v)}}{\lambda_c}\cos\theta_{p,k}-\frac{i}{M}+\frac{1}{2}=n,\quad n=0,\pm1,\cdots. \tag{42}$$

Owing to $0\leqslant i\leqslant M-1$, we derive that $\lim_{M\to\infty}[\tilde{\mathbf{e}}_{p,k}^{(v)}]_i\neq 0$ if and only if

$$i=\frac{M}{2}+\frac{Md_t^{(v)}}{\lambda_c}\cos\theta_{p,k}=\bar{m}_{p,k}, \tag{43}$$

where in the case of $M\to\infty$, $\left(\frac{M}{2}+\frac{Md_t^{(v)}}{\lambda_c}\cos\theta_{p,k}\right)$ can be arbitrarily close to an integer. Thus, we derive from (40) and (41) that

$$\lim_{M\to\infty}\tilde{\mathbf{e}}_{p,k}^{(v)}=\boldsymbol{\alpha}_M^{(\bar{m}_{p,k})}. \tag{44}$$

Therefore, (37) is obtained. (38) can be derived in a similar way. □

Similar to (40), the jth column of $\frac{1}{\sqrt{N_c}}\mathbf{H}_k\mathbf{F}^*_{N_c\times N_g}$ is calculated as

$$\begin{aligned}&\frac{1}{\sqrt{N_c}}[\mathbf{H}_k\mathbf{F}^*_{N_c\times N_g}]_j\\ &= \frac{1}{\sqrt{N_c}}\sum_{m=0}^{N_c-1}\sum_{p=1}^{N_p}a_{p,k}\mathbf{e}(\theta_{p,k},\varphi_{p,k})e^{-\bar{\jmath}2\pi\frac{mr_{p,k}}{N_c}}\frac{1}{\sqrt{N_c}}e^{\bar{\jmath}2\pi\frac{mj}{N_c}}\\ &= \sum_{p=1}^{N_p}a_{p,k}\mathbf{e}\left(\theta_{p,k},\varphi_{p,k}\right)e^{-\bar{\jmath}\pi(N_c-1)\frac{r_{p,k}-j}{N_c}}f_{N_c}\left(\frac{r_{p,k}-j}{N_c}\right),\end{aligned} \tag{45}$$

where $f_{N_c}(\cdot)$ is defined by substituting N_c for M in (10). When $N_c\to\infty$, $\lim_{N_c\to\infty} f_{N_c}\left(\frac{1}{N_c}\left(r_{p,k}-j\right)\right)\neq 0$ if and only if $j=r_{p,k}$. In this case, from (41), we have

$$\frac{1}{\sqrt{N_c}}[\mathbf{H}_k\mathbf{F}^*_{N_c\times N_g}]_j=\sum_{p=1}^{N_p}a_{p,k}\mathbf{e}(\theta_{p,k},\varphi_{p,k})\delta(j-r_{p,k}). \tag{46}$$

Therefore, we have

$$\lim_{N_c\to\infty}\frac{1}{\sqrt{N_c}}\mathbf{H}_k\mathbf{F}^*_{N_c\times N_g}=\sum_{p=1}^{N_p}a_{p,k}\mathbf{e}(\theta_{p,k},\varphi_{p,k})\left[\alpha_{N_g}^{(r_{p,k})}\right]^T. \tag{47}$$

Substituting (1), (35), (36), and (47) into (7), we have

$$\begin{aligned}&\lim_{N_c\to\infty}\mathbf{G}_k\\ &= \frac{1}{\sqrt{MN}}\sum_{p=1}^{N_p}a_{p,k}\left(\mathbf{V}_M^H\otimes\mathbf{V}_N^H\right)\mathbf{e}(\theta_{p,k},\varphi_{p,k})\left[\alpha_{N_g}^{(r_{p,k})}\right]^T\\ &= \frac{1}{\sqrt{MN}}\sum_{p=1}^{N_p}a_{p,k}\left(\mathbf{V}_M^H\otimes\mathbf{V}_N^H\right)\\ &\quad\cdot\left(\mathbf{e}^{(v)}(\theta_{p,k})\otimes\mathbf{e}^{(h)}(\theta_{p,k},\varphi_{p,k})\right)\left[\alpha_{N_g}^{(r_{p,k})}\right]^T\\ &= \sum_{p=1}^{N_p}a_{p,k}\left(\frac{1}{\sqrt{M}}\mathbf{V}_M^H\mathbf{e}^{(v)}(\theta_{p,k})\right)\\ &\quad\otimes\left(\frac{1}{\sqrt{N}}\mathbf{V}_N^H\mathbf{e}^{(h)}(\theta_{p,k},\varphi_{p,k})\right)\left[\alpha_{N_g}^{(r_{p,k})}\right]^T\\ &= \sum_{p=1}^{N_p}a_{p,k}\left(\tilde{\mathbf{e}}_{p,k}^{(v)}\otimes\tilde{\mathbf{e}}_{p,k}^{(h)}\right)\left[\alpha_{N_g}^{(r_{p,k})}\right]^T.\end{aligned} \tag{48}$$

When $N\to\infty$ and $N_c\to\infty$, substituting (38) into (48), we have

$$\mathbf{G}_k=\sum_{p=1}^{N_p}a_{p,k}\left(\tilde{\mathbf{e}}_{p,k}^{(v)}\otimes\boldsymbol{\alpha}_N^{(\bar{n}_{p,k})}\right)\left[\boldsymbol{\alpha}_{N_g}^{(r_{p,k})}\right]^T. \tag{49}$$

Notice that $\left[\tilde{\mathbf{e}}_{p,k}^{(v)}\otimes\boldsymbol{\alpha}_N^{(\bar{n}_{p,k})}\right]_i=\left[\tilde{\mathbf{e}}_{p,k}^{(v)}\right]_{\lfloor i/N\rfloor}\cdot\left[\boldsymbol{\alpha}_N^{(\bar{n}_{p,k})}\right]_{\langle i\rangle_N}$. From (13) and (40), the (i,j)th element of $\mathbf{G}_k$ can be written as

$$\begin{aligned}[\mathbf{G}_k]_{i,j} =& \sum_{p=1}^{N_p}a_{p,k}e^{-\bar{\jmath}\pi(M-1)\left(\frac{\bar{m}_{p,k}-\lfloor i/N\rfloor}{M}\right)}\\ &\cdot f_M\left(\frac{\bar{m}_{p,k}-\lfloor i/N\rfloor}{M}\right)\delta\left(\langle i\rangle_N-\bar{n}_{p,k}\right)\delta\left(j-r_{p,k}\right).\end{aligned} \tag{50}$$

Under the assumption that different paths are WSS, independent [34], and $a_{p,k}\sim\mathcal{CN}(0,\sigma_{p,k}^2)$, i.e.,

$$E\left\{a_{p,k}a_{p',k}^*\right\}=\begin{cases}\sigma_{p,k}^2, & p=p',\\ 0, & p\neq p',\end{cases} \tag{51}$$

the average channel power in angle-delay domain is calculated as

$$\begin{aligned}E\left\{\left|[\mathbf{G}_k]_{i,j}\right|^2\right\} =& \sum_{p=1}^{N_p}\sigma_{p,k}^2f_M^2\left(\frac{\bar{m}_{p,k}-\lfloor i/N\rfloor}{M}\right)\\ &\cdot\delta\left(\langle i\rangle_N-\bar{n}_{p,k}\right)\delta\left(j-r_{p,k}\right).\end{aligned} \tag{52}$$

Substituting (9) into (52), we can obtain (11). When $M \to \infty$ and $N_c \to \infty$, (12) can be derived by using the similar method.

When $M \to \infty$, $N \to \infty$ and $N_c \to \infty$, substituting (37) and (38) into (48), we have

$$\mathbf{G}_k = \sum_{p=1}^{N_p} a_{p,k} \left(\boldsymbol{\alpha}_M^{(\bar{m}_{p,k})} \otimes \boldsymbol{\alpha}_N^{(\bar{n}_{p,k})} \right) \left[\boldsymbol{\alpha}_{N_g}^{(r_{p,k})} \right]^T, \tag{53}$$

and the (i,j)th element of $\mathbf{G}_k$ satisfies

$$[\mathbf{G}_k]_{i,j} = \sum_{p=1}^{N_p} a_{p,k} \delta(i - \bar{m}_{p,k} N - \bar{n}_{p,k}) \delta(j - r_{p,k}). \tag{54}$$

From (51), the average channel power in angle-delay domain is calculated as

$$E\left\{ \left| [\mathbf{G}_k]_{i,j} \right|^2 \right\} = \sum_{p=1}^{N_p} \sigma_{p,k}^2 \delta\left(i - \bar{m}_{p,k} N - \bar{n}_{p,k}\right) \delta\left(j - r_{p,k}\right). \tag{55}$$

Substituting (9) into (55), we get (15). This completes the proof.

Appendix B
Proof of Theorem 2

To prove Theorem 2, we provide the following preliminary results stated as Lemma 2-3. Similar to (6), we define the angle-delay domain channel covariance matrix (ADCCM) as

$$\mathbf{B}_k \triangleq E\left\{ \text{vec}\{\mathbf{G}_k\} \text{vec}^H\{\mathbf{G}_k\} \right\} \in \mathbb{C}^{N_a N_g \times N_a N_g}. \tag{56}$$

Lemma 2: For massive MIMO-OFDM systems with UPA at the BS, when $M \to \infty$, $N \to \infty$, and $N_c \to \infty$, the ADCCM $\mathbf{B}_k$ tends to diag $\{\text{vec}\{\boldsymbol{\Omega}_k\}\}$ in the sense that, $\forall i, j = 0, 1, \cdots, N_a N_g - 1$,

$$\lim_{M\to\infty, N\to\infty, N_c\to\infty} \left[\mathbf{B}_k \text{ - diag} \left\{ \text{vec}\{\boldsymbol{\Omega}_k\} \right\} \right]_{i,j} = 0. \tag{57}$$

Proof: We first introduce some auxiliary symbols to simplify the notations. Define $i_C \triangleq \lfloor i/N_a \rfloor$ and $i_R \triangleq \langle i \rangle_{N_a}$ given an arbitrary non-negative integer i. Since the element indices of vector and matrix start with 0 in our paper, the correspondences between the elements of vec$\{\mathbf{G}_k\}$ and $\mathbf{G}_k$, vec$\{\boldsymbol{\Omega}_k\}$ and $\boldsymbol{\Omega}_k$ are given as

$$\left[\text{vec}\{\mathbf{G}_k\} \right]_i = \left[\mathbf{G}_k \right]_{i_R, i_C}, \tag{58}$$

$$\left[\text{vec}\{\boldsymbol{\Omega}_k\} \right]_i = \left[\boldsymbol{\Omega}_k \right]_{i_R, i_C}. \tag{59}$$

Employing (56), (58), and (59), we derive

$$\begin{aligned}
&\left[\mathbf{B}_k \text{ - diag} \left\{ \text{vec}\{\boldsymbol{\Omega}_k\} \right\} \right]_{i,j} \\
&= \left(E\left\{ \left[\text{vec}\{\mathbf{G}_k\} \right]_i \left[\text{vec}\{\mathbf{G}_k\} \right]_j^* \right\} - \left[\text{diag} \left\{ \text{vec}\{\boldsymbol{\Omega}_k\} \right\} \right]_{i,j} \right) \\
&= \left(E\left\{ \left[\mathbf{G}_k \right]_{i_R, i_C} \left[\mathbf{G}_k \right]_{j_R, j_C}^* \right\} - \left[\text{diag} \left\{ \text{vec}\{\boldsymbol{\Omega}_k\} \right\} \right]_{i,j} \right) \\
&= \begin{cases} E\left\{ \left[\mathbf{G}_k \right]_{i_R, i_C} \left[\mathbf{G}_k \right]_{i_R, i_C}^* \right\} - \left[\boldsymbol{\Omega}_k \right]_{i_R, i_C}, & i = j, \\ E\left\{ \left[\mathbf{G}_k \right]_{i_R, i_C} \left[\mathbf{G}_k \right]_{j_R, j_C}^* \right\}, & i \neq j. \end{cases}
\end{aligned} \tag{60}$$

When $M \to \infty$, $N \to \infty$ and $N_c \to \infty$, according to (51) and (54), we derive

$$\begin{aligned}
&\lim_{M\to\infty, N\to\infty, N_c\to\infty} E\left\{ \left[\mathbf{G}_k \right]_{i,j} \left[\mathbf{G}_k \right]_{i',j'}^* \right\} \\
&= \begin{cases} \sum\limits_{p=1}^{N_p} \sigma_{p,k}^2 \delta(i - \bar{m}_{p,k} N - \bar{n}_{p,k}) \delta(j - r_{p,k}), & i = i', j = j', \\ 0, & \text{others.} \end{cases}
\end{aligned} \tag{61}$$

Substituting (61) and (9) into (60), we get (57). □

To assist the subsequent proof process, we recall the definition of the extended ADCRM $\bar{\mathbf{G}}_k$ in (19). Then we investigate the property of $\bar{\mathbf{G}}_k$ as illustrated in Lemma 3.

Lemma 3: For massvie MIMO-OFDM system, when the number of sub-carriers $N_c \to \infty$, the extended ADCRM $\bar{\mathbf{G}}_k$ tends to $\mathbf{G}_k \mathbf{I}_{N_c \times N_g}^T$ in the sense that

$$\lim_{N_c\to\infty} \left(\bar{\mathbf{G}}_k - \mathbf{G}_k \mathbf{I}_{N_c \times N_g}^T \right) = \mathbf{0}. \tag{62}$$

Proof: Employing the following representation

$$\mathbf{F}_{N_c \times N_g}^* \mathbf{I}_{N_c \times N_g}^T = \left[\begin{array}{cc} \mathbf{F}_{N_c \times N_g}^* & \mathbf{0}_{N_c \times (N_c - N_g)} \end{array} \right], \tag{63}$$

we derive

$$\begin{aligned}
&\lim_{N_c\to\infty} \left(\bar{\mathbf{G}}_k - \mathbf{G}_k \mathbf{I}_{N_c \times N_g}^T \right) \\
&= \lim_{N_c\to\infty} \frac{1}{\sqrt{MNN_c}} \left(\mathbf{V}_M \otimes \mathbf{V}_N \right)^H \mathbf{H}_k \left(\mathbf{F}_{N_c}^* - \mathbf{F}_{N_c \times N_g}^* \mathbf{I}_{N_c \times N_g}^T \right) \\
&= \left[\begin{array}{cc} \mathbf{0}_{N_c \times N_g} & \lim\limits_{N_c\to\infty} \frac{1}{\sqrt{MNN_c}} \left(\mathbf{V}_M \otimes \mathbf{V}_N \right)^H \mathbf{H}_k \ddot{\mathbf{F}}_{N_c \times (N_c - N_g)}^* \end{array} \right],
\end{aligned} \tag{64}$$

where $\ddot{\mathbf{F}}_{N_c \times (N_c - N_g)} \in \mathbb{C}^{N_c \times (N_c - N_g)}$ denotes the matrix composed of the last $(N_c - N_g)$ columns of $\mathbf{F}_{N_c}$ with the (i,j)th element $\left[\ddot{\mathbf{F}}_{N_c \times (N_c - N_g)} \right]_{i,j} \triangleq \frac{1}{\sqrt{N_c}} e^{-\bar{\jmath} 2\pi \frac{i(j+N_g)}{N_c}}$.

Similar to (45), the jth column of $\frac{1}{\sqrt{N_c}} \mathbf{H}_k \ddot{\mathbf{F}}_{N_c \times (N_c - N_g)}^*$ is calculated as

$$\begin{aligned}
&\left[\frac{1}{\sqrt{N_c}} \mathbf{H}_k \ddot{\mathbf{F}}_{N_c \times (N_c - N_g)}^* \right]_j \\
&= \sum_{p=1}^{N_p} a_{p,k} \mathbf{e}\left(\theta_{p,k}, \varphi_{p,k}\right) e^{-\bar{\jmath}\pi(N_c - 1)} f_{N_c} \left(\frac{r_{p,k} - j - N_g}{N_c} \right).
\end{aligned} \tag{65}$$

When $N_c \to \infty$, $\lim\limits_{N_c\to\infty} f_{N_c} \left(\frac{r_{p,k} - j - N_g}{N_c} \right) \neq 0$ if and only if $j = r_{p,k} - N_g$. Note that the maximum delay is smaller than T_g, namely, $r_{p,k} < N_g$ holds for arbitrary pth path. Therefore, $j = r_{p,k} - N_g$ can not be satisfied for arbitrary non-negative integer j which leads to

$$\lim_{N_c\to\infty} \frac{1}{\sqrt{N_c}} \mathbf{H}_k \ddot{\mathbf{F}}_{N_c \times (N_c - N_g)}^* = \mathbf{0}_{N_a \times (N_c - N_g)}. \tag{66}$$

Substituting (66) into (64), we obtain the proof of (62). □

According to the definition of $\bar{\mathbf{G}}_k$ in (19), we derive

$$\begin{aligned}
\text{vec}\left\{ \bar{\mathbf{G}}_k \right\} &= \frac{1}{\sqrt{MNN_c}} \left(\mathbf{F}_{N_c}^H \otimes \left(\mathbf{V}_M \otimes \mathbf{V}_N \right)^H \right) \text{vec}\left(\mathbf{H}_k \right) \\
&= \frac{1}{\sqrt{MNN_c}} \left(\mathbf{F}_{N_c} \otimes \left(\mathbf{V}_M \otimes \mathbf{V}_N \right) \right)^H \text{vec}\left(\mathbf{H}_k \right),
\end{aligned} \tag{67}$$

through employing the properties of "vec" operation and

Kronecker product $\text{vec}\{\mathbf{AXB}\} = \left(\mathbf{B}^T \otimes \mathbf{A}\right)\text{vec}\{\mathbf{X}\}$ and $\mathbf{A}^H \otimes \mathbf{B}^H = (\mathbf{A} \otimes \mathbf{B})^H$ [45]. Similar to (56), we define the extended ADCCM as

$$\bar{\mathbf{B}}_k \triangleq E\left\{\text{vec}\left\{\bar{\mathbf{G}}_k\right\}\text{vec}^H\left\{\bar{\mathbf{G}}_k\right\}\right\} \in \mathbb{C}^{N_a N_c \times N_a N_c}. \quad (68)$$

Substituting (67) and (6) into (68), we obtain

$$\begin{aligned}\bar{\mathbf{B}}_k = \frac{1}{MNN_c}&(\mathbf{F}_{N_c} \otimes (\mathbf{V}_M \otimes \mathbf{V}_N))^H \mathbf{T}_k \\ &\cdot (\mathbf{F}_{N_c} \otimes (\mathbf{V}_M \otimes \mathbf{V}_N))\end{aligned}, \quad (69)$$

and

$$\begin{aligned}\text{vec}\left\{\bar{\mathbf{B}}_k\right\} &= \frac{1}{MNN_c}\Big((\mathbf{F}_{N_c} \otimes (\mathbf{V}_M \otimes \mathbf{V}_N))^T \\ &\quad \otimes (\mathbf{F}_{N_c} \otimes (\mathbf{V}_M \otimes \mathbf{V}_N))^H\Big)\text{vec}\{\mathbf{T}_k\} \\ &= \frac{1}{MNN_c}\left(\mathbf{U}^T \otimes \mathbf{U}^H\right)\text{vec}\{\mathbf{T}_k\},\end{aligned} \quad (70)$$

where $\mathbf{U} \triangleq \mathbf{F}_{N_c} \otimes (\mathbf{V}_M \otimes \mathbf{V}_N) \in \mathbb{C}^{N_a N_c \times N_a N_c}$ is used to simplify the representation with

$$\begin{aligned}\mathbf{U}\mathbf{U}^H &= (\mathbf{F}_{N_c} \otimes (\mathbf{V}_M \otimes \mathbf{V}_N))(\mathbf{F}_{N_c} \otimes (\mathbf{V}_M \otimes \mathbf{V}_N))^H \\ &= \left(\mathbf{F}_{N_c}\mathbf{F}_{N_c}^H\right) \otimes \Big((\mathbf{V}_M \otimes \mathbf{V}_N)(\mathbf{V}_M \otimes \mathbf{V}_N)^H\Big) \\ &= \mathbf{I}_{N_c} \otimes \left(\left(\mathbf{V}_M\mathbf{V}_M^H\right) \otimes \left(\mathbf{V}_N\mathbf{V}_N^H\right)\right) = \mathbf{I}_{N_a N_c}.\end{aligned} \quad (71)$$

Then it can be shown that

$$\begin{aligned}&\text{Tr}\left\{\bar{\mathbf{B}}_k^H\bar{\mathbf{B}}_{k'}\right\} = \text{vec}^H\left\{\bar{\mathbf{B}}_k\right\}\text{vec}\left\{\bar{\mathbf{B}}_k\right\} \\ &= \frac{1}{(MNN_c)^2}\text{vec}^H\{\mathbf{T}_k\}\left(\mathbf{U}^T \otimes \mathbf{U}^H\right)^H\left(\mathbf{U}^T \otimes \mathbf{U}^H\right) \\ &\quad \cdot \text{vec}\{\mathbf{T}_k\} \\ &= \frac{1}{(MNN_c)^2}\text{vec}^H\{\mathbf{T}_k\}\left(\mathbf{U}^*\mathbf{U}^T\right) \otimes \left(\mathbf{U}\mathbf{U}^H\right)\text{vec}\{\mathbf{T}_k\} \\ &= \frac{1}{(MNN_c)^2}\text{vec}^H\{\mathbf{T}_k\}\text{vec}\{\mathbf{T}_k\} \\ &= \frac{1}{(MNN_c)^2}\text{Tr}\left\{\mathbf{T}_k^H\mathbf{T}_{k'}\right\}.\end{aligned} \quad (72)$$

From Lemma 3, the extended ADCRM $\bar{\mathbf{G}}_k$ can be represented by the ADCRM $\mathbf{G}_k$ in the case of $N_c \to \infty$, given by

$$\bar{\mathbf{G}}_k = \mathbf{G}_k\mathbf{I}_{N_c \times N_g}^T, \quad (73)$$

and further

$$\begin{aligned}\text{vec}\left\{\bar{\mathbf{G}}_k\right\} &= \text{vec}\left\{\mathbf{G}_k\mathbf{I}_{N_c \times N_g}^T\right\} = \left(\mathbf{I}_{N_c \times N_g} \otimes \mathbf{I}_{N_a}\right)\text{vec}\{\mathbf{G}_k\} \\ &= \mathbf{I}_{N_a N_c \times N_a N_g} \cdot \text{vec}\{\mathbf{G}_k\}.\end{aligned} \quad (74)$$

Then substituting (56) and (74) into (68), we derive the formula between the extended ADCCM $\bar{\mathbf{B}}_k$ and the ADCCM $\mathbf{B}_k$, given by

$$\bar{\mathbf{B}}_k = \mathbf{I}_{N_a N_c \times N_a N_g}\mathbf{B}_k\mathbf{I}_{N_a N_c \times N_a N_g}^T, \quad (75)$$

and further

$$\text{vec}\left\{\bar{\mathbf{B}}_k\right\} = \left(\mathbf{I}_{N_a N_c \times N_a N_g} \otimes \mathbf{I}_{N_a N_c \times N_a N_g}\right)\text{vec}\{\mathbf{B}_k\}. \quad (76)$$

Then it can be shown that

$$\begin{aligned}\text{Tr}\left\{\bar{\mathbf{B}}_k^H\bar{\mathbf{B}}_{k'}\right\} &= \text{vec}^H\left\{\bar{\mathbf{B}}_k\right\}\text{vec}\left\{\bar{\mathbf{B}}_k\right\} \\ &= \text{vec}^H\{\mathbf{B}_k\}\left(\mathbf{I}_{N_a N_c \times N_a N_g}^T\mathbf{I}_{N_a N_c \times N_a N_g}\right) \\ &\quad \otimes \left(\mathbf{I}_{N_a N_c \times N_a N_g}^T\mathbf{I}_{N_a N_c \times N_a N_g}\right)\text{vec}\{\mathbf{B}_k\} \\ &= \text{vec}^H\{\mathbf{B}_k\}\text{vec}\{\mathbf{B}_k\} = \text{Tr}\left\{\mathbf{B}_k^H\mathbf{B}_{k'}\right\}.\end{aligned} \quad (77)$$

Combining (72) and (77), we can obtain

$$\text{Tr}\left\{\mathbf{B}_k^H\mathbf{B}_{k'}\right\} = \frac{1}{(MNN_c)^2}\text{Tr}\left(\mathbf{T}_k^H\mathbf{T}_{k'}\right). \quad (78)$$

According to Lemma 2, in the case of $M \to \infty$ and $N \to \infty$, we have

$$\mathbf{B}_k = \text{diag}\{\text{vec}\{\mathbf{\Omega}_k\}\}. \quad (79)$$

Hence

$$\begin{aligned}\text{Tr}\left\{\mathbf{B}_k^H\mathbf{B}_{k'}\right\} &= \text{Tr}\left\{\text{diag}^H\{\text{vec}\{\mathbf{\Omega}_k\}\}\,\text{diag}\{\text{vec}\{\mathbf{\Omega}_{k'}\}\}\right\} \\ &= \text{vec}^H\{\mathbf{\Omega}_k\}\text{vec}\{\mathbf{\Omega}_{k'}\} = \text{Tr}\left\{\mathbf{\Omega}_k^H\mathbf{\Omega}_{k'}\right\}.\end{aligned} \quad (80)$$

By comparing (78) and (80), it follows that

$$\text{Tr}\left\{\mathbf{\Omega}_k^H\mathbf{\Omega}_{k'}\right\} = \frac{1}{(MNN_c)^2}\text{Tr}\left(\mathbf{T}_k^H\mathbf{T}_{k'}\right). \quad (81)$$

Substituting (81) into (16), finally we obtain (17). This completes the proof.

REFERENCES

[1] C. Wu, X. Yi, W. Wang, Q. Huang, and X. Gao, "3D CNN-enabled positioning in 3D massive MIMO-OFDM systems," in *Proc. IEEE ICC*, Dublin, Ireland, Jun. 2020, pp. 1–6.

[2] W. H. Chin, Z. Fan, and R. Haines, "Emerging technologies and research challenges for 5G wireless networks," *IEEE Wireless Commun.*, vol. 21, no. 2, pp. 106–112, Apr. 2014.

[3] B. A. Renfro, M. Stein, and N. Boeker, *An analysis of global positioning system (GPS) standard positioning service (SPS) performance for 2018*, Space Geophys. Lab. Appl. Res. Lab., Univ. Texas Austin, Tech. Rep. TR-SGL-19-02, Mar. 2018. [Online]. Available: https://www.gps.gov/systems/gps/performance/

[4] M. H. Bergen, X. Jin, D. Guerrero, H. A. L. F. Chaves, N. V. Fredeen, and J. F. Holzman, "Design and implementation of an optical receiver for angle-of-arrival-based positioning," *J. Light. Technol*, vol. 35, no. 18, pp. 3877–3885, Sep. 2017.

[5] S. Tomic, M. Beko, and R. Dinis, "3-D target localization in wireless sensor networks using RSS and AoA measurements," *IEEE Trans. Veh. Technol.*, vol. 66, no. 4, pp. 3197–3210, Apr. 2017.

[6] A. Y. Z. Xu, E. K. S. Au, A. K. S. Wong, and Q. Wang, "A novel threshold-based coherent TOA estimation for IR-UWB systems," *IEEE Trans. Veh. Technol.*, vol. 58, no. 8, pp. 4675–4681, Oct. 2009.

[7] Y. T. Chan and K. C. Ho, "A simple and efficient estimator for hyperbolic location," *IEEE Trans. Signal Process.*, vol. 42, no. 8, pp. 1905–1915, Aug. 1994.

[8] Y. Yuan, S. Hou, and Q. Zhao, "An improved TDOA localization algorithm based on wavelet transform," in *Proc. IEEE ICEIEC*, Macau, China, Jul. 2017, pp. 111–114.

[9] Z. Wang, H. Zhang, T. Lu, and T. A. Gulliver, "Cooperative RSS-based localization in wireless sensor networks using relative error estimation and semidefinite programming," *IEEE Trans. Veh. Technol.*, vol. 68, no. 1, pp. 483–497, Jan. 2019.

[10] H. Miao, K. Yu, and M. J. Juntti, "Positioning for NLOS propagation: Algorithm derivations and Cramer-Rao bounds," *IEEE Trans. Veh. Technol.*, vol. 56, no. 5, pp. 2568–2580, Sep. 2007.

[11] X. Sun, X. Gao, G. Y. Li, and W. Han, "Single-site localization based on a new type of fingerprint for massive MIMO-OFDM systems," *IEEE Trans. Veh. Technol.*, vol. 67, no. 7, pp. 6134–6145, Jul. 2018.

[12] X. Sun, C. Wu, X. Gao, and G. Y. Li, "Fingerprint-based localization for massive MIMO-OFDM system with deep convolutional neural networks," *IEEE Trans. Veh. Technol.*, vol. 68, no. 11, pp. 10 846–10 857, Nov. 2019.

[13] K. N. R. S. V. Prasad, E. Hossain, and V. K. Bhargava, "Machine learning methods for RSS-based user positioning in distributed massive MIMO," *IEEE Trans. Wireless Commun.*, vol. 17, no. 12, pp. 8402–8417, Dec. 2018.

[14] H. Chen, Y. Zhang, W. Li, X. Tao, and P. Zhang, "ConFi: Convolutional neural networks based indoor Wi-Fi localization using channel state information," *IEEE Access*, vol. 5, pp. 18 066–18 074, Sep. 2017.

[15] X. Wang, L. Gao, and S. Mao, "CSI phase fingerprinting for indoor localization with a deep learning approach," *IEEE Internet of Things Journal*, vol. 3, no. 6, pp. 1113–1123, Dec. 2016.

[16] G. Wu and P. Tseng, "A deep neural network-based indoor positioning method using channel state information," in *Proc. ICNC*, Maui, HI, USA, Mar. 2018, pp. 290–294.

[17] A. Decurninge, L. G. Ordonez, P. Ferrand, H. Gaoning, L. Bojie, Z. Wei, and M. Guillaud, "CSI-based outdoor localization for massive MIMO: Experiments with a learning approach," in *Proc. ISWCS*, Lisbon, Portugal, Aug. 2018, pp. 1–6.

[18] W. Sun, M. Xue, H. Yu, H. Tang, and A. Lin, "Augmentation of fingerprints for indoor WiFi localization based on Gaussian process regression," *IEEE Trans. Veh. Technol.*, vol. 67, no. 11, pp. 10 896–10 905, Nov. 2018.

[19] L. You, X. Gao, X. Xia, N. Ma, and Y. Peng, "Pilot reuse for massive MIMO transmission over spatially correlated Rayleigh fading channels," *IEEE Trans. Wireless Commun.*, vol. 14, no. 6, pp. 3352–3366, Jun. 2015.

[20] C. Sun, X. Gao, S. Jin, M. Matthaiou, Z. Ding, and C. Xiao, "Beam division multiple access transmission for massive MIMO communications," *IEEE Trans. Commun.*, vol. 63, no. 6, pp. 2170–2184, Jun. 2015.

[21] L. You, X. Gao, A. L. Swindlehurst, and W. Zhong, "Channel acquisition for massive MIMO-OFDM with adjustable phase shift pilots," *IEEE Trans. Signal Process.*, vol. 64, no. 6, pp. 1461–1476, Mar. 2016.

[22] X. Li, S. Jin, H. A. Suraweera, J. Hou, and X. Gao, "Statistical 3-D beamforming for large-scale MIMO downlink systems over Rician fading channels," *IEEE Trans. Commun.*, vol. 64, no. 4, pp. 1529–1543, Apr. 2016.

[23] L. You, X. Gao, G. Y. Li, X. Xia, and N. Ma, "BDMA for millimeter-wave/terahertz massive MIMO transmission with per-beam synchronization," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 7, pp. 1550–1563, Jul. 2017.

[24] B. Wang, F. Gao, S. Jin, H. Lin, and G. Y. Li, "Spatial- and frequency-wideband effects in millimeter-wave massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 66, no. 13, pp. 3393–3406, Jul. 2018.

[25] L. You, K.-X. Li, J. Wang, X. Gao, X.-G. Xia, and B. Ottersten, "Massive MIMO transmission for LEO satellite communications," Feb. 2020. [Online]. Available: https://arxiv.org/abs/2002.08148

[26] N. Garcia, H. Wymeersch, E. G. Larsson, A. M. Haimovich, and M. Coulon, "Direct localization for massive MIMO," *IEEE Trans. Signal Process.*, vol. 65, no. 10, pp. 2475–2487, May. 2017.

[27] M. Arnold, S. Dorner, S. Cammerer, and S. Ten Brink, "On deep learning-based massive MIMO indoor user localization," in *Proc. IEEE SPAWC*, Kalamata, India, Jun. 2018, pp. 1–5.

[28] X. Wang, L. Gao, S. Mao, and S. Pandey, "CSI-based fingerprinting for indoor localization: A deep learning approach," *IEEE Trans. Veh. Technol.*, vol. 66, no. 1, pp. 763–776, Jan. 2017.

[29] J. Vieira, E. Leitinger, M. Sarajlic, X. Li, and F. Tufvesson, "Deep convolutional neural networks for massive MIMO fingerprint-based positioning," in *Proc. IEEE PIMRC*, Montreal, QC, Canada, Oct. 2017, pp. 1–6.

[30] R. Shafin, L. Liu, Y. Li, A. Wang, and J. Zhang, "Angle and delay estimation for 3-D massive MIMO/FD-MIMO systems based on parametric channel modeling," *IEEE Trans. Wireless Commun.*, vol. 16, no. 8, pp. 5370–5383, Aug. 2017.

[31] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE CVPR*, Boston, MA, USA, Jun. 2015, pp. 1–9.

[32] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception architecture for computer vision," in *Proc. IEEE CVPR*, Las Vegas, NV, USA, Jun. 2016, pp. 2818–2826.

[33] C. Szegedy, S. Ioffe, and V. Vanhoucke, "Inception-v4, Inception-ResNet and the impact of residual connections on learning," *CoRR*, vol. abs/1602.07261, 2016. [Online]. Available: http://arxiv.org/abs/1602.07261

[34] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005. [Online]. Available: https://doi.org/10.1017/CBO9780511807213

[35] L. Liu, C. Oestges, J. Poutanen, K. Haneda, P. Vainikainen, F. Quitin, F. Tufvesson, and P. D. Doncker, "The COST 2100 MIMO channel model," *IEEE Wireless Communications*, vol. 19, no. 6, pp. 92–99, Dec. 2012.

[36] X. Yi and E. K. S. Au, "User scheduling for heterogeneous multiuser MIMO systems: A subspace viewpoint," *IEEE Trans. Veh. Technol.*, vol. 60, no. 8, pp. 4004–4013, Oct. 2011.

[37] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. London, Uk.: The Johns Hopkins University Press, 1996.

[38] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, http://www.deeplearningbook.org.

[39] M. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Computer Vision, ECCV 2014 - 13th European Conference, Proceedings*, vol. 8689 LNCS, no. PART 1. Springer Verlag, 2014, pp. 818–833.

[40] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *CoRR*, vol. abs/1502.03167, 2015. [Online]. Available: http://arxiv.org/abs/1502.03167

[41] M. Lin, Q. Chen, and S. Yan, "Network in network," *CoRR*, vol. abs/1312.4400, 2013. [Online]. Available: http://arxiv.org/abs/1312.4400

[42] S. Jaeckel, L. Raschkowski, K. Borner, and L. Thiele, "QuaDRiGa: A 3-D multi-cell channel model with time evolution for enabling virtual field trials," *IEEE Trans. Antennas Propag.*, vol. 62, no. 6, pp. 3242–3256, Jun. 2014.

[43] P. K. et al., *WINNER II Channel Models, Version 1.1.*, Sep. 2007. [Online]. Available: http://www.ist-winner.org/

[44] *Study on channel model for frequencies from 0.5 to 100 GHz, Version 15.0.0*, document 3GPP T.R. 38.901 , Jun. 2018.

[45] K. B. Petersen and M. S. Pedersen, "The matrix cookbook," Nov. 2012, version 20121115. [Online]. Available: http://localhost/pubdb/p.php?3274

Chi Wu (Student Member, IEEE) received the B.E. degree in information science and engineering from Southeast University, Nanjing, China, in 2018. He is currently pursuing the M.E. degree with the National Mobile Communications Research Laboratory, Southeast University, Nanjing, China. From August 2019 to November 2019, he was a visiting student at the Department of Electrical Engineering and Electronics, University of Liverpool, U.K. His research interests include massive MIMO wireless localization, channel prediction, and their integration with machine learning.

Xinping Yi (Member, IEEE) received the Ph.D. degree in electronics and communications from Télécom ParisTech, Paris, France, in 2015. He is currently a Lecturer (Assistant Professor) with the Department of Electrical Engineering and Electronics, University of Liverpool, U.K. Prior to Liverpool, he was a Research Associate with Technische Universität Berlin, Berlin, Germany, from 2014 to 2017, a Research Assistant with EURECOM, Sophia Antipolis, France, from 2011 to 2014, and a Research Engineer with Huawei Technologies, Shenzhen, China, from 2009 to 2011. His main research interests include information theory, graph theory, and machine learning, and their applications in wireless communications and artificial intelligence.

Wenjin Wang (Member, IEEE) received the Ph.D. degree in communication and information systems from Southeast University, Nanjing, China, in 2011. From 2010 to 2014, he was with the School of System Engineering, University of Reading, Reading, U.K. He is currently an Associate Professor with the National Mobile Communications Research Laboratory, Southeast University, Nanjing. His research interests include advanced signal processing for future wireless communications and satellite communications. He was a recipient of the Best Paper Award at IEEE WCSP 2009 and the first grade Technological Invention Award of the State Education Ministry of China, in 2009.

Li You (Member, IEEE) received the B.E. and M.E. degrees from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2009 and 2012, respectively, and the Ph.D. degree from Southeast University, Nanjing, in 2016, all in electrical engineering.

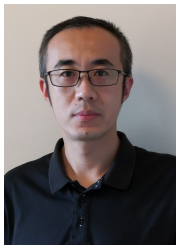
From 2014 to 2015, he conducted Visiting Research at the Center for Pervasive Communications and Computing, University of California Irvine, Irvine, CA, USA. Since 2016, he has been with the Faculty of the National Mobile Communications Research Laboratory, Southeast University. His research interests lie in the general areas of communications, signal processing, and information theory with the current emphasis on massive MIMO communications.

Qing Huang (Member, IEEE) received the B.E degree from the Department of Radio Engineering, Nanjing Institute of Technology, Nanjing, China, in 1982. Since 1999, he has been a professor with the School of Information Science and Engineering, Southeast University, China. He joined the the National Mobile Communications Research Laboratory, Southeast University, in 2002. His research interests include multi-carrier mobile communication and modern signal processing.

Xiqi Gao (Fellow, IEEE) received the Ph.D. degree in electrical engineering from Southeast University, Nanjing, China, in 1997.

He joined the Department of Radio Engineering, Southeast University, in April 1992. Since May 2001, he has been a professor of information systems and communications. From September 1999 to August 2000, he was a visiting scholar at Massachusetts Institute of Technology, Cambridge, and Boston University, Boston, MA. From August 2007 to July 2008, he visited the Darmstadt University of Technology, Darmstadt, Germany, as a Humboldt scholar. His current research interests include broadband multi-carrier communications, MIMO wireless communications, channel estimation and turbo equalization, and multi-rate signal processing for wireless communications. From 2007 to 2012, he served as an Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS. From 2009 to 2013, he served as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING. From 2015 to 2017, he served as an Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS.

Dr. Gao received the Science and Technology Awards of the State Education Ministry of China in 1998, 2006 and 2009, the National Technological Invention Award of China in 2011, and the 2011 IEEE Communications Society Stephen O. Rice Prize Paper Award in the field of communications theory.



Qing Liu received the B.S. and M.S. degrees, from the department of Information Science & Electronic Engineering, Zhejiang University, in 2004 and 2006, respectively. Since 2006, he joined Huawei Technology Co., Ltd and worked as a senior engineer. His research interests include both L1 and L2 algorithm design, like advanced receiver, architecture of baseband algorithm, AI in wireless system and so on.