

Beamspace ESPRIT-D for Joint 3D Angle and Delay Estimation for Joint Localization and Communication at MmWave

Yun Chen[†], Nuria González-Prelcic[†], Takayuki Shimizu[‡], Hongshen Lu[‡], and Chinmay Mahabal[‡]

[†] University of California San Diego, Email: {yuc216, ngprelcic}@ucsd.edu

[‡] Toyota Motor North America, Email: {takayuki.shimizu, hongsheng.lu, chinmay.mahabal}@toyota.com

Abstract—In this paper, we address the complex task of estimating the parameters for multiple propagation paths of realistic millimeter wave (mmWave) channels at a reasonable computational complexity while maintaining high accuracy, which is required for precise positioning in joint localization and communication systems. We introduce an innovative method termed ESPRIT-D – beamspace Estimation of Signal Parameters via Rotational Invariance Techniques with a Dictionary based solution. It exploits a model for the mmWave multipath channel accounting for filtering effects, represented as a 5D tensor to extract azimuth and elevation angles of departure and arrival through beamspace ESPRIT, while retrieving the delay estimates by a greedy sparse recovery method. Our evaluation, based on realistic channels generated through ray-tracing simulations, demonstrates an average angular error below 0.01° for line-of-sight (LoS) and 0.1° for non-line-of-sight (NLoS) cases, respectively. The delay accuracy achieves an average of $3e^{-10}$ s. Compared with state-of-the-art (SOTA), our algorithm exhibits a $10\times$ improvement in estimation accuracy.

I. INTRODUCTION

Large antenna arrays and bandwidths employed by mmWave multiple-input multiple-output (MIMO) communication systems lead to the possibility of achieving remarkably precise estimations of multipath parameters, which can be exploited for high accuracy positioning [1]. This accuracy usually comes, however, at the cost of a high computational complexity in the channel estimation procedure. This limitation is especially relevant when computing high-resolution 3D time domain channel estimates [2].

Research on joint angle and delay estimation has addressed the need for manageable complexity without compromising accuracy. Some recent works have leveraged the mmWave channel sparsity [3]–[5]. The multidimensional orthogonal matching pursuit (MOMP) algorithm [3], [6], addresses the complexity issue associated with large-scale tensor multiplication by distributing tensor atoms across multiple dimensions. The algorithm proposed in [4] capitalizes on both the sparsity and low-rank nature of the channel matrix, exploiting a multi-rank aware sparse recovery strategy. The low-rank

property is also explored in [5], which formulates the channel estimation problem as a low-rank sensing problem and solves it through the generalized conditional gradient alternating minimization algorithm. Tensor decomposition techniques were also exploited for mmWave channel estimation [7]–[10]. In [7], the frequency domain received signal is modeled as a third-order tensor which admits canonical polyadic decomposition (CPD), and the channel estimates are obtained by maximizing correlations with the decomposed tensors. Beamspace estimation of signal parameters via rotational invariance techniques (ESPRIT), following tensor decomposition, is the basis for extracting channel parameters in [8]–[10]. While most of the previous work exploiting tensor decomposition assumes the use of a uniform linear array (ULA), which simplifies decomposition and reduces the complexity, the authors of [10] consider a uniform rectangular array (URA) and focus on low complexity for tensor decomposition. URAs are also considered in [11], which introduces a beamspace multiple signal classification (MUSIC) algorithm with a multi-spectral peak search method to facilitate the searching process.

All prior work presents several notable shortcomings: 1) the assumption of ULA in the system models [2], [4], [5], [7]–[9], [12], which can lead to complexity issues when transitioning to applications requiring URAs, since this transition results in an exponential dimensionality increase. 2) While certain methods, including those based on ESPRIT [4], [8]–[12], appear theoretically sound, they may not be practical, as the effect of pulse shaping and filtering is missing in the considered discrete time equivalent channel model. 3) Some recent compressed sensing based methods which account for filtering effects [3], [6] exhibit reduced complexity with respect to prior work [2], but they can still encounter complexity challenges in specific scenarios with large arrays. 4) Instead of using realistic channels, some work conducts experiments for evaluation using artificially controlled channels fulfilling specific probability distributions [4], [5], [7]–[12], which limits their applicability when considering real scenarios where these assumptions do not hold.

In this paper, we model the 3D mmWave channel accounting for pulse shaping as the summation of tensors spanning five dimensions. Thereafter, to tackle the task of estimating angles and delays for multiple propagation paths,

This material is based upon work partially supported by the National Science Foundation under grant no. 2147955 and is supported in part by funds from the federal agency and industry partners as specified in the Resilient & Intelligent NextG Systems (RINGS) program and by an unrestricted research fund from Toyota Motor North America.

we introduce an innovative method termed ESPRIT-D. This method combines the strengths of beamspace ESPRIT for the simultaneous extraction of azimuth and elevation angles with a dictionary-based sparse recovery solution that targets delay estimation. Furthermore, the algorithm's performance is assessed through ray-tracing simulations, providing a robust evaluation in the context of realistic channel conditions. The MOMP-based method exploited in [3], [6], [13]–[16] serves as the benchmark, as it closely approaches the channel estimation accuracy required for sub-meter accuracy localization while circumventing the limitations mentioned earlier when operating with moderate size arrays.

Notations: x , $\mathbf{x}$, and $\mathbf{X}$ represent a scalar, a vector, and a matrix or tensor. $[\mathbf{X}]_{i,j}$ means the element at the i -th row and j -th column of $\mathbf{X}$. Indexing a multidimensional tensor follows the same rule. $\mathbf{X}^\top$, $\bar{\mathbf{X}}$, $\mathbf{X}^*$, and $\mathbf{X}^\dagger$ are the transpose, conjugate, conjugate transpose, and pseudo inverse of $\mathbf{X}$. Operators “ $\circ$ ”, “ $\otimes$ ”, and “ $\langle \cdot, \cdot \rangle$ ” represent the vector outer product, Kronecker product, and the vector dot product operations.

II. SYSTEM MODEL

We consider a mmWave MIMO communication system where a base station (BS) is communicating with a user for initial access. The transmitter (TX) and receiver (RX) are equipped with URAs of size $N_t = N_t^x \times N_t^y$ and $N_r = N_r^x \times N_r^y$. The d -th tap of the time domain channel is

$$\mathbf{H}_d = \sum_{\ell=1}^L \alpha_\ell p(dT_s - \tau_\ell) \dot{\mathbf{a}}(\boldsymbol{\theta}_\ell) \dot{\mathbf{a}}(\boldsymbol{\phi}_\ell)^* \in \mathbb{C}^{N_r \times N_t}, \quad (1)$$

where L is the number of paths, α_ℓ and τ_ℓ are the complex gain and the delay of the ℓ -th path, $p(\cdot)$ represents the pulse shaping and filtering effects, T_s is the sampling period, and $\boldsymbol{\theta}_\ell = [\cos \theta_\ell^y \cos \theta_\ell^x, \cos \theta_\ell^y \sin \theta_\ell^x, \sin \theta_\ell^y]^\top$ and $\boldsymbol{\phi}_\ell = [\cos \phi_\ell^y \cos \phi_\ell^x, \cos \phi_\ell^y \sin \phi_\ell^x, \sin \phi_\ell^y]^\top$ are the direction of arrival (DoA) and direction of departure (DoD) of the ℓ -th path, where θ_ℓ^x , θ_ℓ^y , ϕ_ℓ^x and ϕ_ℓ^y represent the azimuth angle-of-arrival (AoA), elevation AoA, azimuth angle-of-departure (AoD) and elevation AoD. Note that both the azimuth and elevation angles are in the range $[-\pi/2, \pi/2]$. Assuming the array is placed in the yz -plane with a half-wavelength element spacing, the receive array response is

$$\begin{aligned} & [\dot{\mathbf{a}}(\boldsymbol{\theta})]_{(n_r^x-1)N_r^y+n_r^y} = \\ & \exp\{-j\pi((n_r^x-1)[\boldsymbol{\theta}]_2 + (n_r^y-1)[\boldsymbol{\theta}]_3)\}, \end{aligned} \quad (2)$$

and $\dot{\mathbf{a}}(\boldsymbol{\phi})$ is defined in the same manner. We write $\dot{\mathbf{a}}(\cdot)$ in the form of Kronecker product as $\dot{\mathbf{a}}(\boldsymbol{\theta}_\ell) = \mathbf{a}(\theta_\ell^x) \otimes \mathbf{a}(\theta_\ell^y)$ and $\dot{\mathbf{a}}(\boldsymbol{\phi}_\ell) = \mathbf{a}(\phi_\ell^x) \otimes \mathbf{a}(\phi_\ell^y)$ for the purpose of tensor decomposition later, where

$$[\mathbf{a}(\theta_\ell^x)]_{n_r^x} = \exp\{-j\pi(n_r^x-1)[\boldsymbol{\theta}_\ell]_2\}; \quad (3)$$

$$[\mathbf{a}(\theta_\ell^y)]_{n_r^y} = \exp\{-j\pi(n_r^y-1)[\boldsymbol{\theta}_\ell]_3\}; \quad (4)$$

$$[\mathbf{a}(\phi_\ell^x)]_{n_r^x} = \exp\{-j\pi(n_r^x-1)[\boldsymbol{\phi}_\ell]_2\}; \quad (5)$$

$$[\mathbf{a}(\phi_\ell^y)]_{n_r^y} = \exp\{-j\pi(n_r^y-1)[\boldsymbol{\phi}_\ell]_3\}. \quad (6)$$

We assume the total number of taps is N_d , and the channel estimation is realized through the downlink transmission of a single training sequence of length $Q \geq N_d$. For simplicity, we assume that an analog beamforming architecture is considered at both ends, and the channel is sounded multiple times by the transmission of the training sequence through different combinations of training precoders and combiners. Assuming the analog precoder in use at the given time instant q is $\mathbf{f} \in \mathbb{C}^{N_t \times 1}$, and the analog combiner is $\mathbf{w} \in \mathbb{C}^{N_r \times 1}$, the q -th instance of the received signal is

$$y[q] = \mathbf{w}^* \sum_{d=0}^{N_d-1} \mathbf{H}_d \mathbf{f} \sqrt{P_t} s[q-d] + \mathbf{w}^* \mathbf{n}[q], \quad (7)$$

where $\sqrt{P_t}$ is the transmitted power, $s[q]$ is the q -th instance of the training signal which satisfies $\mathbb{E}[s[q]s[q]^*] = 1$, and $\mathbf{n}[q] \sim \mathcal{N}(\mathbf{0} + \mathbf{0}j, \mathcal{K}BT^{(K)}\mathbf{I}_{N_r})$ is the thermal noise at the q -th time slot, where $\mathcal{K}$ is the Boltzmann's constant, B is the system bandwidth, and $T^{(K)}$ is the absolute temperature. Note that if a hybrid architecture with several radio frequency (RF) chains is considered, it will be possible to simultaneously sound the channel with multiple pairs of precoders and combiners, and the measurement collection process can be faster. For the analog-only case, the received signal vector is

$$\mathbf{y} = \mathbf{w}^* [\mathbf{H}_0 \mathbf{f}, \mathbf{H}_1 \mathbf{f}, \dots, \mathbf{H}_{N_d-1} \mathbf{f}] \sqrt{P_t} \mathbf{S} + \mathbf{w}^* \mathbf{n}, \quad (8)$$

where $\mathbf{y} = [y[1], \dots, y[Q]]$, $\mathbf{S}$ is the transmitted signal matrix defined as

$$\mathbf{S} = \begin{bmatrix} s[1] & s[2] & \dots & s[Q] \\ 0 & s[1] & \dots & s[Q-1] \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & s[Q-(N_d-1)] \end{bmatrix}, \quad (9)$$

and $\mathbf{n} = [\mathbf{n}[1], \dots, \mathbf{n}[Q]] \in \mathbb{C}^{N_r \times Q}$. Let $\check{\mathbf{n}} = \mathbf{w}^* \mathbf{n}$, then (8) can be written as

$$\mathbf{y} = \sqrt{P_t} \mathbf{w}^* [\mathbf{H}_0, \dots, \mathbf{H}_{N_d-1}] (\mathbf{I}_{N_d} \otimes \mathbf{f}) \mathbf{S} + \check{\mathbf{n}}. \quad (10)$$

III. ESPRIT-D FOR JOINT 3D ANGLE AND DELAY ESTIMATION

A. Background of beamspace ESPRIT

The canonical polyadic (CP) representation of a Z -th-order tensor $\boldsymbol{\Upsilon} \in \mathbb{C}^{N_1 \times N_2 \times \dots \times N_Z}$ of rank L is written as a linear combination of the tensor outer products of rank-1 tensors $\mathbf{u}_{z,\ell}$ [17]:

$$\boldsymbol{\Upsilon} = \sum_{\ell=1}^L \gamma_\ell \mathbf{u}_{1,\ell} \circ \mathbf{u}_{2,\ell} \circ \dots \circ \mathbf{u}_{Z,\ell}, \quad (11)$$

where $\mathbf{u}_{z,\ell}$ is the rank-1 tensor spanning the z -th dimension of the ℓ -th tensor product, and γ_ℓ is the coefficient of the ℓ -th tensor product. To adapt the CP representation for our channel estimation problem, $\mathbf{u}_{z,\ell}$ is defined in the form of a steering vector as $\mathbf{u}_{z,\ell} = [e^{j0\omega_{z,\ell}}, e^{j1\omega_{z,\ell}}, \dots, e^{j(N_z-1)\omega_{z,\ell}}]^\top$, with $\omega_{z,\ell}$ being the frequency component of the z -th dimension of the ℓ -th tensor product. A tensor decomposition

method, e.g. CPD in [17], is adopted to derive the estimated rank-1 tensors, denoted as $\hat{\mathbf{u}}_{z,\ell}$, by solving the optimization problem assuming the number of summation components being N_L :

$$\min_{\hat{\mathbf{U}}_z} \left\| \mathbf{Y} - \sum_{\ell=1}^{N_L} \hat{\mathbf{u}}_{1,\ell} \circ \cdots \circ \hat{\mathbf{u}}_{Z,\ell} \right\|. \quad (12)$$

Let $\hat{\mathbf{U}}_z = [\hat{\mathbf{u}}_{z,1}, \dots, \hat{\mathbf{u}}_{z,N_L}]$, and

$$\begin{cases} \mathbf{J}_z^{(1)} = [\mathbf{I}_{N_z-1} \ \mathbf{0}_{(N_z-1) \times 1}] \\ \mathbf{J}_z^{(2)} = [\mathbf{0}_{N_z-1 \times 1} \ \mathbf{I}_{N_z-1}] \end{cases}, \quad (13)$$

the shift invariance property can be applied:

$$\mathbf{J}_z^{(1)} \hat{\mathbf{U}}_z = \mathbf{J}_z^{(2)} \hat{\mathbf{U}}_z \mathbf{\Phi}_z, \quad (14)$$

where $\mathbf{\Phi}_z$'s diagonal elements contain the estimated frequency components $\hat{\omega}_{z,\ell}$ for $1 \leq \ell \leq N_L$ for the z -th dimension, i.e., $\mathbf{\Phi}_z = \text{diag}[e^{-j\hat{\omega}_{z,1}}, \dots, e^{-j\hat{\omega}_{z,N_L}}]$.

In the context of using beamspace tensors, we define the beamspace rank-1 tensor $\mathbf{u}'_{z,\ell} = \mathbf{B}_z^* \mathbf{u}_{z,\ell} \in \mathbb{C}^{M_z \times 1}$, where $\mathbf{B}_z = [\mathbf{b}_{z,1}, \dots, \mathbf{b}_{z,M_z}] \in \mathbb{C}^{N_z \times M_z}$ contains M_z "beams", each one of them defined by a vector of length N_z . Using this identity, (11) becomes $\mathbf{Y} = \sum_{\ell=1}^L \gamma_\ell \mathbf{u}'_{1,\ell} \circ \mathbf{u}'_{2,\ell} \circ \cdots \circ \mathbf{u}'_{Z,\ell}$. Replacing $\hat{\mathbf{u}}_{z,\ell}$ with the beamspace tensor $\hat{\mathbf{u}}'_{z,\ell}$ in (12), $\hat{\mathbf{u}}'_{z,\ell}$ can be still acquired via CPD as mentioned above. We denote the concatenation of the acquired beamspace tensors of the z -th dimension as $\hat{\mathbf{U}}'_z = [\hat{\mathbf{u}}'_{z,1}, \dots, \hat{\mathbf{u}}'_{z,N_L}]$, and it turns out (14) no longer holds, since $\mathbf{J}_z^{(1)} \hat{\mathbf{U}}'_z \neq \mathbf{J}_z^{(2)} \hat{\mathbf{U}}'_z \mathbf{\Phi}_z$. A way to restore the shift invariance structure is proposed in [8]. First, we find $\mathbf{\Phi}'_z$ s.t. $\mathbf{J}_z^{(1)} \mathbf{B}_z = \mathbf{J}_z^{(2)} \mathbf{B}_z \mathbf{\Phi}'_z$ through least square (LS) estimation:

$$\mathbf{\Phi}'_z = (\mathbf{J}_z^{(2)} \mathbf{B}_z)^\dagger \mathbf{J}_z^{(1)} \mathbf{B}_z. \quad (15)$$

Then, there exists an adjustment matrix

$$\begin{aligned} \Psi_z &= \mathbf{I}_{M_z} - ([\mathbf{B}_z^*]_{:,N_z})([\mathbf{B}_z^*]_{:,N_z})^* \\ &\quad - (\mathbf{\Phi}'_z)^* [\mathbf{B}_z^*]_{:,1} [\mathbf{B}_z]_{1,:} \mathbf{\Phi}'_z \in \mathbb{C}^{M_z \times M_z}, \end{aligned} \quad (16)$$

s.t.

$$\begin{cases} \Psi_z [\mathbf{B}_z^*]_{:,N_z} = \mathbf{0}_{M_z \times 1} \\ \Psi_z (\mathbf{\Phi}'_z)^* [\mathbf{B}_z^*]_{:,1} = \mathbf{0}_{M_z \times 1} \end{cases}, \quad (17)$$

hereby $\Psi_z (\mathbf{\Phi}'_z)^* \mathbf{U}'_z = \Psi_z \mathbf{U}'_z \mathbf{\Phi}_z^*$ holds. Let $\mathbf{U}'_z = \hat{\mathbf{U}}'_z \mathbf{D}_z$, where $\mathbf{D}_z \in \mathbb{C}^{N_L \times N_L}$ is an unknown non-singular matrix. We now have $\Psi_z (\mathbf{\Phi}'_z)^* \hat{\mathbf{U}}'_z = \Psi_z \hat{\mathbf{U}}'_z \mathbf{D}_z \mathbf{\Phi}_z^* (\mathbf{D}_z)^{-1}$. The value of $\Gamma_z = \mathbf{D}_z \mathbf{\Phi}_z^* (\mathbf{D}_z)^{-1}$ can be determined by LS estimation:

$$\hat{\Gamma}_z = (\Psi_z \hat{\mathbf{U}}'_z)^\dagger \Psi_z (\mathbf{\Phi}'_z)^* \hat{\mathbf{U}}'_z. \quad (18)$$

As $\Gamma_z \mathbf{D}_z = \mathbf{D}_z \mathbf{\Phi}_z^*$, the eigenvalues of $\hat{\Gamma}_z$, denoted as $\text{eig}(\hat{\Gamma}_z) \in \mathbb{C}^{N_L \times 1}$, are the estimation of the diagonal elements of $\mathbf{\Phi}_z^* = \text{diag}[e^{j\hat{\omega}_{z,1}}, \dots, e^{j\hat{\omega}_{z,N_L}}]$.

B. Beamspace ESPRIT-D for Joint 3D Angle and Delay Estimation

We first denote the delay response vector w.r.t delay τ_ℓ as

$$\mathbf{p}(\tau_\ell) = [p(0 \cdot T_s - \tau_\ell), \dots, p((N_d - 1)T_s - \tau_\ell)]^\top \in \mathbb{R}^{N_d \times 1}, \quad (19)$$

and rearrange the elements in (10) so that it can be reformatted as

$$\mathbf{y} = \sum_{\ell=1}^L \alpha_\ell \mathbf{w}^* \hat{\mathbf{a}}(\theta_\ell) \hat{\mathbf{a}}(\phi_\ell)^* \mathbf{f} \mathbf{p}(\tau_\ell)^\top \sqrt{P_t} \mathbf{S} + \check{\mathbf{n}}. \quad (20)$$

Assuming M_r combiners (the m_r -th of which is denoted as $\mathbf{w}_{m_r}$) and M_t precoders (the m_t -th of which is denoted as $\mathbf{f}_{m_t}$) are used to collect $M = M_r M_t$ measurements, a tensor containing all the measurements can be written in the outer product format as

$$\mathbf{Y} = \sum_{\ell=1}^L \mathbf{W}^* \hat{\mathbf{a}}(\theta_\ell) \circ \mathbf{F}^\top \hat{\mathbf{a}}(\phi_\ell) \circ \alpha_\ell \sqrt{P_t} \mathbf{S}^\top \mathbf{p}(\tau_\ell) + \check{\mathbf{N}}, \quad (21)$$

where $\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_{M_r}]$, $\mathbf{F} = [\mathbf{f}_1, \dots, \mathbf{f}_{M_t}]$, and $[\mathbf{Y}]_{m_r, m_t, q}$ and $[\check{\mathbf{N}}]_{m_r, m_t, q}$ are the q -th instance of the received signal and the noise acquired with $\mathbf{f}_{m_t}$ and $\mathbf{w}_{m_r}$. We introduce $\hat{\mathbf{H}}$ as the spatial representation of the channel including the precoder/combiner effects, whose ℓ -th component is

$$\hat{\mathbf{H}}_\ell = \mathbf{W}^* \hat{\mathbf{a}}(\theta_\ell) \circ \mathbf{F}^\top \hat{\mathbf{a}}(\phi_\ell), \quad (22)$$

which can be further extended to a 4D tensor as

$$\hat{\mathbf{H}}_\ell = \mathbf{W}_x^* \hat{\mathbf{a}}(\theta_\ell^x) \circ \mathbf{W}_y^* \hat{\mathbf{a}}(\theta_\ell^y) \circ \mathbf{F}_x^\top \hat{\mathbf{a}}(\theta_\ell^x) \circ \mathbf{F}_y^\top \hat{\mathbf{a}}(\theta_\ell^y), \quad (23)$$

where $\mathbf{W}_x = [\mathbf{w}_x^{(1)}, \dots, \mathbf{w}_x^{(M_r^x)}] \in \mathbb{C}^{N_r^x \times M_r^x}$ and $\mathbf{W}_y = [\mathbf{w}_y^{(1)}, \dots, \mathbf{w}_y^{(M_r^y)}] \in \mathbb{C}^{N_r^y \times M_r^y}$ are used to build $\mathbf{W}$ as $\mathbf{W}_x \otimes \mathbf{W}_y$, and $M_r = M_r^x M_r^y$. Similarly, $\mathbf{F}_x \otimes \mathbf{F}_y = \mathbf{F}$, where $\mathbf{F}_x = [\mathbf{f}_x^{(1)}, \dots, \mathbf{f}_x^{(M_t^x)}]$ and $\mathbf{F}_y = [\mathbf{f}_y^{(1)}, \dots, \mathbf{f}_y^{(M_t^y)}]$. Hereby $\mathbf{Y}$ can be transformed into the form of the outer product of tensors spreading 5 dimensions:

$$\begin{aligned} \mathbf{Y}_{5D} &= \sum_{\ell=1}^L \mathbf{W}_x^* \hat{\mathbf{a}}(\theta_\ell^x) \circ \mathbf{W}_y^* \hat{\mathbf{a}}(\theta_\ell^y) \\ &\quad \circ \mathbf{F}_x^\top \hat{\mathbf{a}}(\theta_\ell^x) \circ \mathbf{F}_y^\top \hat{\mathbf{a}}(\theta_\ell^y) \\ &\quad \circ \alpha_\ell \sqrt{P_t} \mathbf{S}^\top \mathbf{p}(\tau_\ell) + \check{\mathbf{N}}, \end{aligned} \quad (24)$$

and the relationship between $\mathbf{Y}$ and $\mathbf{Y}_{5D}$ can be acquired as $[\mathbf{Y}_{5D}]_{m_r^x, m_r^y, m_t^x, m_t^y, q} = [\mathbf{Y}]_{(m_r^x-1)M_r^y + m_r^y, (m_t^x-1)M_t^y + m_t^y, q}$. Now we solve the tensor decomposition problem so that $\mathbf{Y}_{5D}$ can be approximated by the summation of N_L vector outer products, i.e.,

$$\min_{\hat{\mathbf{U}}'_1, \dots, \hat{\mathbf{U}}'_5} \left\| \mathbf{Y}_{5D} - \sum_{\ell=1}^{N_L} \gamma_\ell (\circ_{z=1}^5 \hat{\mathbf{u}}'_{z,\ell}) \right\|, \quad (25)$$

where the first 4 dimensions will be associated with $\mathbf{W}_x^* \hat{\mathbf{a}}(\theta_\ell^x)$, $\mathbf{W}_y^* \hat{\mathbf{a}}(\theta_\ell^y)$, $\mathbf{F}_x^\top \hat{\mathbf{a}}(\theta_\ell^x)$, and $\mathbf{F}_y^\top \hat{\mathbf{a}}(\theta_\ell^y)$ respectively, and the last dimension corresponds to the delay domain. Let the tensor decomposition result be $\hat{\mathbf{U}}'_z = [\hat{\mathbf{u}}'_{z,1}, \dots, \hat{\mathbf{u}}'_{z,N_L}]$

Algorithm 1 Beamspase ESPRIT-D

- 1: **Inputs:** The received signal in the form of a 5D tensor $\mathbf{Y}_{5D}$ as in (24); Transmitted signal matrix $\mathbf{S}$; The combiner matrices $\mathbf{W}_x$ and $\mathbf{W}_y$ that can construct $\mathbf{W}$; The precoder matrices $\mathbf{F}_x$ and $\mathbf{F}_y$ that can construct $\mathbf{F}$; The number of taps N_d ; The delay dictionary resolution G_t ; The number of channel paths N_L to be estimated.
 - 2: **Initialize system parameters:**
 The rotation matrices $\mathbf{J}_z^{(1)}$ and $\mathbf{J}_z^{(2)}$ as defined in (13);
 The matrices: $\mathbf{B}_1 = \mathbf{W}_x$, $\mathbf{B}_2 = \mathbf{W}_y$, $\mathbf{B}_3 = \bar{\mathbf{F}}_x$, $\mathbf{B}_4 = \bar{\mathbf{F}}_y$, and $\mathbf{B}_5 = \bar{\mathbf{S}}$;
 The number of atoms N_z in $\mathbf{u}'_z$ for each dimension: $N_1 = M_r^x$, $N_2 = M_r^y$, $N_3 = M_t^x$, $N_4 = M_t^y$, and $N_5 = Q$;
 The delay dictionary $\mathbf{P}$ as in (29);
 The tensor decomposition results by solving (25): $\hat{\mathbf{U}}'_z = [\hat{\mathbf{u}}'_{z,1}, \dots, \hat{\mathbf{u}}'_{z,\ell}, \dots, \hat{\mathbf{u}}'_{z,N_L}]$.
 - 3: **% Angle estimation**
 - 4: **for** $z = [1 : 4]$ **do**
 - 5: Calculate the diagonal matrix Φ'_z as in (15);
 - 6: Calculate the adjustment matrix Ψ_z as in (16);
 - 7: Acquire $\hat{\Gamma}_z$ as in (18), and its eigenvalues $\text{eig}(\hat{\Gamma}_z)$;
 - 8: Estimated frequency components: $\hat{\omega}_z = -j \ln(\text{eig}(\hat{\Gamma}_z))$;
 - 9: **if** $z \in \{1, 2\}$ **then**
 - 10: Estimated AoAs: $\begin{cases} \hat{\theta}_\ell^y = [\sin^{-1}(\frac{\hat{\omega}_2}{\pi})]_\ell; \\ \hat{\theta}_\ell^x = \sin^{-1}(\frac{[\hat{\omega}_1]_\ell}{-\pi \cos \hat{\theta}_\ell^y}) \end{cases}$.
 - 11: **else**
 - 12: Estimated AoDs: $\begin{cases} \hat{\phi}_\ell^y = [\sin^{-1}(\frac{\hat{\omega}_4}{\pi})]_\ell; \\ \hat{\phi}_\ell^x = \sin^{-1}(\frac{[\hat{\omega}_3]_\ell}{\pi \cos \hat{\phi}_\ell^y}) \end{cases}$.
 - 13: **end if**
 - 14: **end for**
 - 15: **% Delay estimation**
 - 16: **for** $\ell = [1 : N_L]$ **do**
 - 17: Acquire the estimation of $\eta_\ell \mathbf{p}(\tau_\ell)$ as in (31);
 - 18: Calculate the distributed correlation $\mathbf{c}$ as in (32);
 - 19: The index of the estimated delay in $\mathbf{P}$: $i_\ell = \arg \max_{i_\ell} \mathbf{c}$;
 - 20: The estimated delay: $\hat{\tau}_\ell = \ddot{t}_{i_\ell}$;
 - 21: **end for**
 - 22: Calculate gain estimations γ as in (34);
-

following that in Sec. III-A, then the beamspase ESPRIT can be performed to extract the angular information, while the delay information needs to be acquired by solving the sparse recovery problem through a dictionary based method.

Angle estimation: To apply the beamspase model, we determine $\mathbf{B}_z$ for $1 \leq z \leq 4$ as $\mathbf{B}_1 = \mathbf{W}_x$, $\mathbf{B}_2 = \mathbf{W}_y$, $\mathbf{B}_3 = \bar{\mathbf{F}}_x$, and $\mathbf{B}_4 = \bar{\mathbf{F}}_y$. The frequency component of the steering vector in each dimension is defined as:

$$\begin{cases} \omega_{1,\ell} = -\pi \cos \theta_\ell^y \sin \theta_\ell^x \\ \omega_{2,\ell} = -\pi \sin \theta_\ell^y \\ \omega_{3,\ell} = \pi \cos \phi_\ell^y \sin \phi_\ell^x \\ \omega_{4,\ell} = \pi \sin \phi_\ell^y \end{cases}, \quad (26)$$

so that $\mathbf{u}_{1,\ell} = \mathbf{a}(\theta_\ell^x)$, $\mathbf{u}_{2,\ell} = \mathbf{a}(\theta_\ell^y)$, $\mathbf{u}_{3,\ell} = \bar{\mathbf{a}}(\phi_\ell^x)$ and $\mathbf{u}_{4,\ell} = \bar{\mathbf{a}}(\phi_\ell^y)$. Hereby $\hat{\mathbf{H}}_\ell$ defined in (23) becomes

$$\hat{\mathbf{H}}_\ell = \circ_{z=1}^4 \mathbf{B}_z^* \mathbf{u}_{z,\ell} \approx \circ_{z=1}^4 \hat{\mathbf{u}}'_{z,\ell}. \quad (27)$$

Based on beamspase ESPRIT mentioned in (13)-(18), the frequency components of each dimension are estimated re-

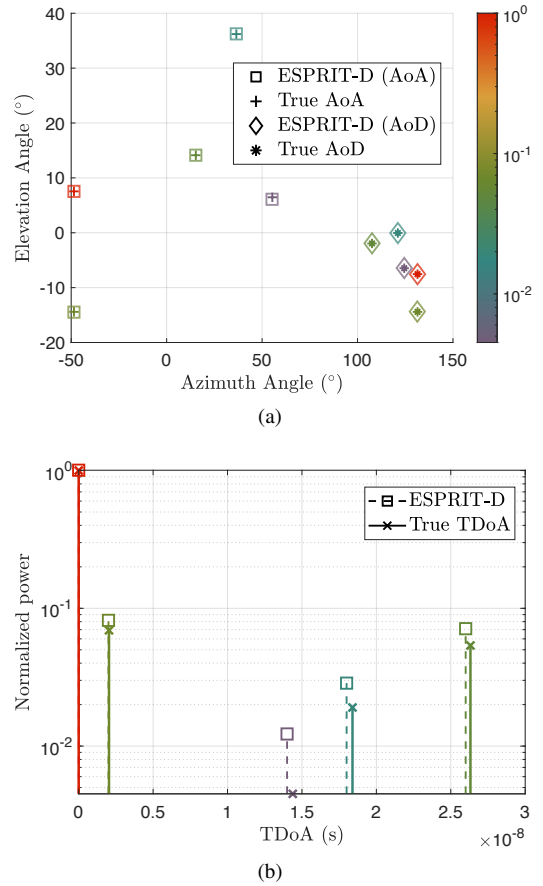


Fig. 1: An example of channel estimation results of the 5 strongest paths using ESPRIT-D, with the colorbar indicating the normalized power strength. DoA estimation error $\leq 0.38^\circ$; DoD estimation error $\leq 0.11^\circ$; TDoA estimation error $\leq 3.8e^{-10}$ s.

ferring to $\text{eig}(\hat{\Gamma}_z)$

$$\hat{\omega}_{z,\ell} = -j \ln([\text{eig}(\hat{\Gamma}_z)]_\ell), \quad (28)$$

and the angle estimates are retrieved using (26).

Delay estimation: Previous work [8]–[11] neglecting the pulse shaping and filtering effects express the channel at the k -th subcarrier as $\mathbf{H}[k] = \sum_{\ell=1}^L \alpha_\ell \dot{\mathbf{a}}(\theta_\ell) \dot{\mathbf{a}}(\phi_\ell)^* e^{-j2\pi k \Delta f \tau_\ell}$, where Δf denotes the subcarrier spacing. In such cases, beamspase ESPRIT remains applicable considering $\mathbf{B}_5 = \mathbf{I}_K$ with K being the number of subcarriers, and $\omega_{5,\ell} = -2\pi \Delta f \tau_\ell$. In our model, however, the delay response $\mathbf{p}$ does not follow the formulation of a steering vector, and the decomposed vectors of the 5th dimension should be converted into delay estimates. Consequently, a dictionary based method emerges as the most viable solution to estimate the delay through sparse recovery. We first construct the delay dictionary $\mathbf{P}$ as:

$$\mathbf{P} = [\mathbf{p}(\ddot{t}_1), \dots, \mathbf{p}(\ddot{t}_{G_t})], \quad (29)$$

where G_t determines the dictionary resolution, and the g -th column is the delay response w.r.t time $\ddot{t}_g = (g-1) \cdot \frac{N_d T_s}{G_t}$.

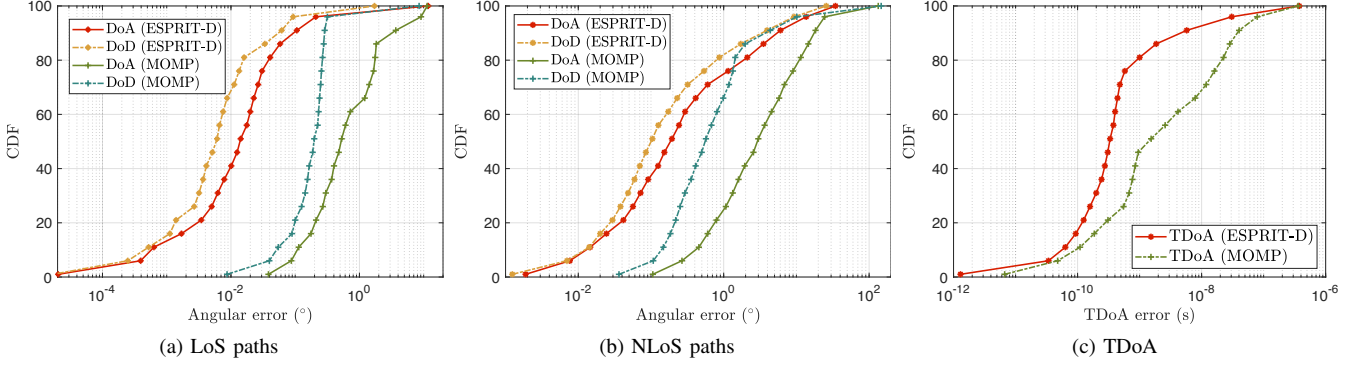


Fig. 2: CDF of estimation errors based on a total of 200 channels: (a) CDF of angular error for LoS paths. (b) CDF of angular error for NLoS paths; (c) CDF of error in the TDoA.

From (24) and (25), we know that $\alpha_\ell \sqrt{P_t} \mathbf{S}^T \mathbf{p}(\tau_\ell)$ corresponds to $\gamma_\ell \hat{\mathbf{u}}_5'$, and $\|\gamma_\ell\| \propto \|\alpha_\ell \sqrt{P_t} \mathbf{S}^T \mathbf{p}(\tau_\ell)\|$. As we now focus on delay estimation, we treat $\frac{\alpha_\ell \sqrt{P_t} \mathbf{S}^T \mathbf{p}(\tau_\ell)}{\gamma_\ell}$ as $\eta_\ell \mathbf{S}^T \mathbf{p}(\tau_\ell)$, which can be approximated by $\hat{\mathbf{u}}_5'$. The delay information for the ℓ -th path can be estimated by solving the sparse recovery problem:

$$\min_{\mathbf{x}_\ell} \|\mathbf{P} \mathbf{x}_\ell - \eta_\ell \mathbf{p}(\tau_\ell)\|, \quad (30)$$

where $\eta_\ell \mathbf{p}(\tau_\ell)$ is obtained by LS estimation:

$$\widehat{\eta_\ell \mathbf{p}(\tau_\ell)} = (\mathbf{S} \mathbf{S}^T)^\dagger \mathbf{S} \hat{\mathbf{u}}_{5,\ell}. \quad (31)$$

We calculate the distributed correlation of $\mathbf{P}$ and $\eta_\ell \mathbf{p}(\tau_\ell)$ as

$$\mathbf{c} = \mathbf{P}^* \widehat{\eta_\ell \mathbf{p}(\tau_\ell)}, \quad (32)$$

then the maximum projection gives the index of the support of the sparse vector $\mathbf{x}_\ell$ as

$$i_\ell = \arg \max_{i_\ell} \mathbf{c}, \quad (33)$$

so the delay estimation is given by $\hat{\tau}_\ell = \hat{t}_{i_\ell}$.

Finally, let $\boldsymbol{\gamma} = [\gamma_1, \dots, \gamma_{N_L}]^T$, then $\boldsymbol{\gamma}$ can be retrieved by solving (34) via LS estimation:

$$\min_{\boldsymbol{\gamma}} \|\text{vec}(\mathbf{Y}_{5D}) - [\text{vec}(\mathbf{o}_{z=1}^5 \hat{\mathbf{u}}_{z,1}), \dots, \text{vec}(\mathbf{o}_{z=N_L}^5 \hat{\mathbf{u}}_{z,N_L})] \boldsymbol{\gamma}\|. \quad (34)$$

The ESPRIT-D algorithm, as outlined in Algorithm 1, exhibits spatial complexity of $\mathcal{O}\left(\sum_{z=1}^4 N_z M_z + N_5 G\right)$, where each N_z corresponds to the response vector length, M_z relates to number of beams, and G depends on the resolution of the dictionary for the delay. In comparison to the time-domain dictionary based channel estimation methods [2], which suffer from a spatial complexity of $\mathcal{O}\left(\prod_{z=1}^5 N_z G_z\right)$ with each G_z linked to dictionary resolution along the dimension, resulting in exponential complexity growth with finer resolutions, our proposed method offers a much more practical and manageable computational approach. Our implementation of ESPRIT-D in MATLAB can be found in [18].

IV. SIMULATION RESULTS

In this section, we present the 5D channel estimation results based on realistic ray-tracing channels simulated by *Wireless Insite* [19]. The MOMP algorithm without DoA then retrieving DoA [14], [16] is taken as the baseline, since it exhibits lower complexity than the original MOMP [3], [6]. The MATLAB code developed to obtain the results shown in this section can be found in [18].

Ray-tracing setup: We consider a realistic outdoor environment for vehicular communication, where an active car equipped with 4 communication arrays at a height of 1.6 m, is communicating with a roadside unit (RSU). The 4 arrays on the vehicle are facing front, back, right, and left, respectively, to make sure at least one of them receives signals with sufficient power strength for channel estimation. The RSU array is located at a known location of $[120, -21.0034, 5]$. The ray-tracing simulation operates at a carrier frequency of 73 GHz, with other parameters chosen as in [20]. A total of 200 scenes are generated, where the active car is placed randomly within the rectangular area from $[5, -12.75]$ to $[220, -1.75]$. We set the number of output paths for each channel to 25.

Communication system setup: We consider the 4 arrays of size $N_r = 8 \times 8$ on the active car, and the array of size $N_t = 16 \times 16$ on the RSU. The transmitted pilot length is $Q = 50$ and the transmitted power is set to $P_t = 40$ dBm. The system operates at a carrier frequency of $f_c = 73$ GHz using the bandwidth of $B = 1$ GHz. We use a raised cosine filter for pulse shaping with the roll-off factor of 0.2. The number of delay taps is fixed to $N_d = 48$. Random precoding/combining is adopted, and a total of 64 measurements are collected at the RX under the temperature of $T^{(K)} = 300$ K (noise power: -84 dBm). In terms of the delay dictionary for delay estimation, $G_t = N_d \times 4 = 320$ is set for sufficiently fine resolutions, i.e., $\hat{t}_{g+1} - \hat{t}_g = 2.5e^{-10}$ s. For all later results, we assume the number of paths to be estimated is $N_L = 5$.

Results: We first show an example of channel estimation results for the 5 strongest paths using ESPRIT-D, as depicted in Fig. 1, where the square boxes overlap with “+”

markers, and the diamond boxes overlap with “*” markers, representing accurate estimations for both AoAs and AoDs. The average estimation errors for AoA and AoD are 0.06° and 0.02° , respectively. The 3D angular estimation error is calculated by $\epsilon(\hat{\varphi}) = \cos^{-1}(\langle \hat{\varphi}, \varphi \rangle)$, which are 0.10° and 0.04° in average for DoA and DoD estimations in the instance. In addition, as we assume an unknown clock offset, the delay estimation results are represented by time difference of arrival (TDoA) (s) w.r.t the shortest path in the channel, and the estimation error is $2.84e^{-10}$ s on average.

The cumulative distribution function (CDF) plots in Fig. 2 illustrate the estimation errors of angular and TDoA estimations based on the 200 channels, where the LoS paths and NLoS paths are analyzed separately. ESPRIT-D achieves an angular accuracy of 0.04° for 90% of the LoS cases, and an accuracy of 2.35° (AoAs) / 1.20° (AoDs) for 90% of the NLoS cases. AoD estimations exhibit a higher accuracy compared to AoA estimations, primarily due to the larger array employed at the RSU. TDoA estimation with ESPRIT-D keeps the errors mostly $\leq 4.57e^{-9}$ s. This level of precision is particularly valuable for applications such as vehicle positioning. Compared with results using MOMP [16], the proposed method brings an improvement of more than $10\times$ on average for both the angular and delay estimations. Note that outliers with significant deviations from the ground truth persist, possibly attributed to the intricacies of real-world outdoor channels, such as large power attenuation (≥ 30 dB) for NLoS paths within a LoS channel [16].

V. CONCLUSION

In this paper, we have formulated the 3D time-domain mmWave channel, accounting for pulse shaping, as a 5D tensor. Then, our proposed ESPRIT-D algorithm combines beamspace ESPRIT with a dictionary-based sparse recovery approach to enable joint 3D angle and delay estimations for multiple signal paths. This method effectively mitigates the complexity issue and outperforms the MOMP algorithm described in [16], which approximates the accuracy requirements for vehicle localization but employs a two-stage strategy that does not estimate all parameters simultaneously due to complexity concerns. According to statistical results acquired based on realistic simulated channels, our algorithm attains angular accuracy of 0.04° for 90% of the LoS cases, with angles of arrival (AoAs) and angles of departure (AoDs) at 2.35° and 1.20° , respectively, for 90% of the NLoS cases. Additionally, the delay estimation brings an accuracy at the level of $4.57e^{-9}$ s. Notably, our algorithm achieves accuracy levels $10\times$ higher than the SOTA algorithms.

REFERENCES

- [1] F. Jiang, Y. Ge, M. Zhu, and H. Wymeersch, “High-dimensional channel estimation for simultaneous localization and communications,” in *2021 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE, 2021, pp. 1–6.
- [2] K. Venugopal, A. Alkhateeb, N. G. Prelcic, and R. W. Heath, “Channel estimation for hybrid architecture-based wideband millimeter wave systems,” *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 9, pp. 1996–2009, 2017.
- [3] J. Palacios, N. González-Prelcic, and C. Rusu, “Multidimensional orthogonal matching pursuit: theory and application to high accuracy joint localization and communication at mmWave,” *arXiv preprint arXiv:2208.11600*, 2022.
- [4] X. Shao, X. Chen, C. Zhong, and Z. Zhang, “Exploiting simultaneous low-rank and sparsity in delay-angular domain for millimeter-wave/terahertz wideband massive access,” *IEEE Transactions on Wireless Communications*, vol. 21, no. 4, pp. 2336–2351, 2021.
- [5] K. F. Masood, J. Tong, J. Xi, J. Yuan, Q. Guo, and Y. Yu, “Low-rank matrix sensing-based channel estimation for mmWave and THz hybrid MIMO systems,” *IEEE Journal of Selected Topics in Signal Processing*, 2023.
- [6] J. Palacios, N. González-Prelcic, and C. Rusu, “Low complexity joint position and channel estimation at millimeter wave based on multidimensional orthogonal matching pursuit,” in *2022 30th European Signal Processing Conference (EUSIPCO)*, 2022, pp. 1002–1006.
- [7] R. Zhang, L. Cheng, S. Wang, Y. Lou, W. Wu, and D. W. K. Ng, “Tensor decomposition-based channel estimation for hybrid mmWave massive MIMO in high-mobility scenarios,” *IEEE Transactions on Communications*, vol. 70, no. 9, pp. 6325–6340, 2022.
- [8] F. Jiang, Y. Ge, M. Zhu, H. Wymeersch, and F. Tufvesson, “Low-complexity channel estimation and localization with random beamspace observations,” in *ICC 2023-IEEE International Conference on Communications*. IEEE, 2023, pp. 5985–5990.
- [9] J. Zhang, D. Rakhimov, and M. Haardt, “Gridless channel estimation for hybrid mmWave MIMO systems via tensor-ESPRIT algorithms in DFT beamspace,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, no. 3, pp. 816–831, 2021.
- [10] X. Gong, W. Chen, L. Sun, J. Chen, and B. Ai, “An ESPRIT-based supervised channel estimation method using tensor train decomposition for mmWave 3-D MIMO-OFDM systems,” *IEEE Transactions on Signal Processing*, vol. 71, pp. 555–570, 2023.
- [11] P. Zhu, H. Lin, J. Li, D. Wang, and X. You, “High-performance channel estimation for mmWave wideband systems with hybrid structures,” *IEEE Transactions on Communications*, vol. 71, no. 4, pp. 2503–2516, 2023.
- [12] B. Hadji, A. Aïssa-El-Bey, L. Fergani, and M. Djedjou, “Joint hybrid precoding and combining design based multi-stage compressed sensing approach for mmWave MIMO channel estimation,” *IEEE Access*, 2023.
- [13] Y. Chen, N. González-Prelcic, T. Shimizu, H. Lu, and C. Mahabal, “Sparse recovery with attention: A hybrid data/model driven solution for high accuracy position and channel tracking at mmWave,” in *IEEE 24th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2023, pp. 491–495.
- [14] Y. Chen, J. Palacios, N. González-Prelcic, T. Shimizu, and H. Lu, “Joint initial access and localization in millimeter wave vehicular networks: a hybrid model/data driven approach,” in *IEEE 12th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, 2022, pp. 355–359.
- [15] M. Bayraktar, J. Palacios, N. González-Prelcic, and C. J. Zhang, “Multidimensional orthogonal matching pursuit-based RIS-aided joint localization and channel estimation at mmWave,” in *2022 IEEE 23rd International Workshop on Signal Processing Advances in Wireless Communication (SPAWC)*, 2022, pp. 1–5.
- [16] Y. Chen, N. González-Prelcic, T. Shimizu, and H. Lu, “Learning to localize with attention: from sparse mmWave channel estimates from a single bs to high accuracy 3D location,” *arXiv preprint arXiv:2307.00167*, 2023.
- [17] I. Kisl, G. G. Calvi, B. Scalzo Dees, and D. P. Mandic, “Tensor decompositions and practical applications: A hands-on tutorial,” in *Recent Trends in Learning From Data: Tutorials from the INNS Big Data and Deep Learning Conference (INNSBDDL2019)*. Springer, 2020, pp. 69–97.
- [18] Y. Chen, “ESPRIT-D,” Dec. 2023. [Online]. Available: <https://github.com/WiSeCom-Lab/ESPRIT-D.git>
- [19] “Wireless Insite,” <http://www.remcom.com/wireless-insite>.
- [20] A. Ali, N. González-Prelcic, and A. Ghosh, “Passive radar at the roadside unit to configure millimeter wave vehicle-to-infrastructure links,” *IEEE Transactions on Vehicular Technology*, vol. 69, no. 12, pp. 14903–14917, 2020.