Position-agnostic Algebraic Estimation of 6G V2X MIMO Channels via Unsupervised Learning

Lorenzo Cazzella*, Dario Tagliaferri*, Marouan Mizmizi*, Matteo Matteucci*,
Damiano Badini†, Christian Mazzucco† and Umberto Spagnolini*
*Politecnico di Milano, Milan, Italy
†Huawei Technologies Italia S.r.l., Segrate, Italy
E-mails: {lorenzo.cazzella, dario.tagliaferri, marouan.mizmizi, matteo.matteucci, umberto.spagnolini}@polimi.it
{damiano.badini, christian.mazzucco}@huawei.com

***Abstract*—MIMO systems in the context of 6G Vehicle-to-Everything (V2X) will require an accurate channel knowledge to enable efficient communication. Standard channel estimation techniques, such as Unconstrained Maximum Likelihood (U-ML), are extremely noisy in massive MIMO settings, while structured approaches, e.g., compressed sensing, are sensitive to hardware impairments. We propose a novel *multi-vehicular* algebraic channel estimation method for 6G V2X based on unsupervised learning which exploits recurrent vehicle passages in typical urban settings. Multiple training sequences from different vehicle passages are clustered via K-medoids algorithm based on their *algebraic similarity* to retrieve the MIMO channel eigenmodes, which can be used to improve the channel estimates. Numerical results show the presence of an optimal number of clusters and remarkable benefits of the proposed method in terms of Mean Squared Error (MSE) compared to standard U-ML solution (15 dB less).**

***Index Terms*—Algebraic MIMO channel estimation, 6G, V2X, Unsupervised learning, Clustering, K-medoids**

I. Introduction

Millimeter Wave (mmWave) ($30-100$ GHz) and sub-THz ($100-300$ GHz) bands emerged as viable solutions for 5G and mostly 6G Vehicle-to-Everything (V2X) applications, due to the spectrum crunch at sub-6 GHz frequencies. In particular, the $24.25-52.6$ GHz band is used in 5G New Radio (NR) Frequency Range 2 (FR2), while future 6G V2X systems will exploit even larger spectrum portions in D-band [1]. The high path loss at these frequencies induces a *sparse* propagation channel, characterized by few significant paths in the Space-Time (ST) domain, i.e., Angles of Arrival/Departure (AoAs/AoDs) and delays [2]. To compensate for the path loss, Multiple-Input Multiple-Output (MIMO) antenna systems at both Transmitter (Tx) and Receiver (Rx) are used to increase antenna gain by beamforming strategies [3], requiring an accurate channel knowledge. In legacy Orthogonal Frequency Division Multiplexing/Multiple Access (OFDM/OFDMA) systems (e.g., 5G NR FR2), standard approaches are based on an Unconstrained Maximum Likelihood (U-ML) channel estimate at each training block from known pilot sequences, which however is extremely noisy in low Signal-to-Noise Ratio (SNR) conditions. Advanced approaches, such as Compressed Sensing (CS), reduce the number of unknowns exploiting the structured channel sparsity and multiple channel observations, improving the U-ML performance by orders of magnitude [4], at the price of being sensitive to system calibrations and parameters initialization [5].

Differently, algebraic LR channel estimation methods [5]–[7] operate on ensembles of training sequences from a single (or multiple) User Equipment (UE) to a fixed Base Station (BS), leveraging the algebraic (unstructured) channel sparsity and the stationarity of the ST eigenmodes (ST subspace), i.e., the invariance of AoAs/AoDs and delays, across multiple channel realizations in time or space [5], [8]. The improved LR channel estimate is retrieved by a proper modal filtering of the received training sequence onto the ST propagation subspace. LR achieves similar accuracy of CS but with an inherent robustness against hardware impairments [5]. The main drawback of the aforementioned LR implementation is that it requires either a large number of consecutive transmissions (not suited to V2X) or the knowledge of UEs position at the BS (increased control signaling).

In this work, we take advantage of the position-aware Multi-Vehicular (MV) LR approach in [8] to design a novel *position-agnostic* clustering-based MV-LR channel estimation method suited for V2X, where the channel eigenmodes are retrieved from the set of received training sequences collected from the *recurrent passages* of different vehicles in a limited urban area (radio cell), where the road constraints induce recurrences in the ST MIMO channel subspace. Instead of requiring the explicit knowledge of UEs positions, by leveraging the works in [9], [10], we frame the problem of obtaining the ensemble of received training sequences for LR as a K-medoids high-dimensional clustering problem in the ST subspace of the MIMO channel, where different received training sequences are grouped based on their algebraic similarity. Numerical simulations with ray-tracing channel data and realistic vehicle trajectories are compared with both the U-ML and the position-aware LR channel estimation [8], the latter proven to be statistically efficient attaining the theoretical MSE lower bound. The results highlight an optimal number of clusters and a Mean Squared Error (MSE) gain in the order of $\approx 15-20$ dB compared to U-ML channel estimation, achieving the same MSE performance of the position-aware one.

Notation: Bold upper- and lower-case letters describe matrices and column vectors. $(\cdot)^{\mathrm{T}}$, $(\cdot)^{\mathrm{H}}$, $(\cdot)^*$, $\|\cdot\|$, and $\mathrm{vec}(\cdot)$ denote, respectively, transpose, conjugate-transpose, conjugate,

Frobenius norm, and vectorization by columns of a matrix. $\mathrm{tr}(\cdot)$, $\mathrm{rank}(\cdot)$ extract the trace and the rank of a matrix. $\otimes$, $\diamond$ and $\odot$ denote the Kronecker, the Kathri-Rao and the element-wise product between two matrices. $\mathrm{diag}(\cdot)$ denotes either a diagonal matrix or the extraction of the diagonal of a matrix. $\mathbf{a} \sim \mathcal{CN}(\boldsymbol{\mu}, \mathbf{C})$ denotes a multi-variate circularly complex Gaussian random variable $\mathbf{a}$ with mean $\boldsymbol{\mu}$ and covariance $\mathbf{C}$. $\mathbb{E}[\cdot]$ is the expectation operator, while $\mathbb{R}$ and $\mathbb{C}$ stand for the set of real and complex numbers, respectively. Finally, δ_n is the Kronecker delta.

II. System and Channel Model

We consider a MIMO uplink system tailored to a V2I scenario, in which the Tx and the Rx are equipped with N_T and N_R antennas, respectively. At the receiving antennas, after the time and frequency synchronization and cyclic prefix removal, the Rx signal is:

$$\mathbf{y}(t) = \mathbf{H}(t) * \mathbf{x}(t) + \mathbf{n}(t), \tag{1}$$

where symbol $*$ denotes the matrix convolution between the transmitted signal $\mathbf{x}(t) \in \mathbb{C}^{N_T \times 1}$ and the frequency-selective MIMO channel $\mathbf{H}(t) \in \mathbb{C}^{N_R \times N_T}$. Vector $\mathbf{n}(t) \in \mathbb{C}^{N_R \times 1}$ denotes the additive Gaussian noise corrupting the received signal. Sampling (1) at time $t = wT$, with $T = 1/B$ being the sampling time (B the bandwidth) we obtain:

$$\mathbf{y}[w] = \mathbf{H}[w] * \mathbf{x}[w] + \mathbf{n}[w], \quad w = 1, \ldots, W \tag{2}$$

where W is the maximum number of channel taps and $\mathbf{H}[w] \equiv \mathbf{H}(wT)$ is the discrete-time MIMO channel matrix. For channel estimation purposes, the Tx signal is a known training sequence assumed to be random and uncorrelated in space and frequency as $\mathbb{E}\left[\mathbf{x}[w]\mathbf{x}[\ell]^{\mathrm{H}}\right] = \sigma_x^2 \mathbf{I}_{N_T} \delta_{w-\ell}$ (σ_x^2 is the signal power). The noise $\mathbf{n}[w]$ is instead modelled as white in time/frequency, but generally colored in space, to account for directional interference, as $\mathbb{E}\left[\mathbf{n}[w]\mathbf{n}[\ell]^{\mathrm{H}}\right] = \mathbf{Q}_n \delta_{w-\ell}$. The SNR at each antenna is

$$\mathrm{SNR} = \frac{\mathbb{E}\left[\left\| \sum_w \mathbf{H}[w] * \mathbf{x}[w] \right\|^2\right]}{\mathrm{tr}(\mathbf{Q}_n)}. \tag{3}$$

In the following, we detail the MIMO channel model.

A. MIMO Channel Model

The discrete-time wideband (frequency-selective) MIMO channel between Tx and Rx is routinely modeled as the sum of P paths as [2]

$$\begin{aligned} \mathbf{H}[w] &= \sum_{p=1}^{P} \alpha_p \, \mathbf{a}_R(\boldsymbol{\vartheta}_p) \mathbf{a}_T^{\mathrm{T}}(\boldsymbol{\psi}_p) g\left[(w-1)T - \tau_p\right] = \\ &= \mathbf{A}_R(\boldsymbol{\vartheta}) \, \boldsymbol{\Lambda}[w] \mathbf{A}_T^{\mathrm{T}}(\boldsymbol{\psi}), \end{aligned} \tag{4}$$

where:

- $\alpha_p \sim \mathcal{CN}(0, \Omega_p)$ is the complex gain of the p-th path. The paths' amplitudes $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_P]^{\mathrm{T}}$ are assumed to follow the Wide-Sense Stationary Uncorrelated Scattering (WSSUS) model, i.e., $\mathbb{E}\left[\boldsymbol{\alpha}_n \boldsymbol{\alpha}_{n+m}^{\mathrm{H}}\right] = \boldsymbol{\Omega} \, \delta_{n-m}$, in which $\boldsymbol{\Omega} = \mathrm{diag}(\Omega_1, \ldots, \Omega_P)$ contains the paths' powers and n, m are two channel realizations in time (different fading blocks) or space (different locations);
- $\mathbf{a}_T(\boldsymbol{\psi}_p) \in \mathbb{C}^{N_T \times 1}$ and $\mathbf{a}_R(\boldsymbol{\vartheta}_p) \in \mathbb{C}^{N_R \times 1}$ are the Tx and Rx array response vectors to the p-th path, respectively, function of the DoDs $\boldsymbol{\psi}_p$ and the DoAs $\boldsymbol{\vartheta}_p$;
- $g\left[(w-1)T - \tau_p\right]$ denotes the w-th sample of the cascade response of the Tx and the Rx pulse shaping filters (PSF), delayed by τ_p (p-th path delay).

In (4), matrices $\mathbf{A}_T(\boldsymbol{\psi}) = [\mathbf{a}_T(\boldsymbol{\psi}_1), \ldots, \mathbf{a}_T(\boldsymbol{\psi}_P)] \in \mathbb{C}^{N_T \times P}$ and $\mathbf{A}_R(\boldsymbol{\vartheta}) = [\mathbf{a}_R(\boldsymbol{\vartheta}_1), \ldots, \mathbf{a}_R(\boldsymbol{\vartheta}_P)] \in \mathbb{C}^{N_R \times P}$ identify the Tx and Rx frequency-independent beam spaces, respectively, while $\boldsymbol{\Lambda}[w] = \mathrm{diag}(\alpha_1 \, g[(w-1)T - \tau_1], \ldots, \alpha_P \, g[(w-1)T - \tau_P]) \in \mathbb{C}^{P \times P}$ is a diagonal matrix collecting all the channel amplitudes scaled by the w-th tap of the PSF. Matrices $\mathbf{A}_T(\boldsymbol{\psi})$ and $\mathbf{A}_R(\boldsymbol{\vartheta})$ define the Tx and Rx diversity orders of channel $\mathbf{H}[w]$ as

$$r_{\mathrm{S}}^{\mathrm{TX}} = \mathrm{rank}(\mathbf{A}_T(\boldsymbol{\psi})) \leq \min(N_T, P), \tag{5}$$

$$r_{\mathrm{S}}^{\mathrm{RX}} = \mathrm{rank}(\mathbf{A}_R(\boldsymbol{\vartheta})) \leq \min(N_R, P), \tag{6}$$

i.e., the number of resolvable spatial paths given the number of Tx and Rx antennas, respectively. By rearranging channel (4), we can isolate the temporal (delays) features of the channel as:

$$\boldsymbol{\mathcal{H}} = \boldsymbol{\mathcal{A}}(\boldsymbol{\vartheta}, \boldsymbol{\psi}) \, \mathbf{D} \, \mathbf{G}^{\mathrm{T}}(\boldsymbol{\tau}), \tag{7}$$

where: *(i)* $\boldsymbol{\mathcal{H}} = [\mathrm{vec}(\mathbf{H}[1]), \ldots, \mathrm{vec}(\mathbf{H}[W])]$; *(ii)* $\boldsymbol{\mathcal{A}}(\boldsymbol{\vartheta}, \boldsymbol{\psi}) = \mathbf{A}_T(\boldsymbol{\psi}) \diamond \mathbf{A}_R(\boldsymbol{\vartheta}) \in \mathbb{C}^{N_T N_R \times P}$ span the *joint* Tx and Rx beam space; *(iii)* $\mathbf{D} = \mathrm{diag}(\alpha_1, \ldots, \alpha_P)$ and *(iv)* $\mathbf{G}(\boldsymbol{\tau}) = [\mathbf{g}(\tau_1), \ldots, \mathbf{g}(\tau_P)]$ embed the temporal features τ. Vector $\mathbf{g}(\tau_p) \in \mathbb{R}^{W \times 1} = [g[-\tau_p], \ldots, g[(W-1)T - \tau_p]]^{\mathrm{T}}$ collects PSF samples delayed by τ_p. With this channel formulation, the temporal diversity order is:

$$r_{\mathrm{T}} = \mathrm{rank}(\mathbf{G}(\boldsymbol{\tau})) \leq \min(W, P), \tag{8}$$

namely the number of temporally-distinguishable channel paths.

III. MV-LR MIMO Channel Estimation

To overcome the limitations of U-ML channel estimation, we adapt here the LR method to high-mobility V2X systems by exploiting the MV concept proposed in [7], [8]. The BS estimates the ST eigenmodes of channel $\widetilde{\mathbf{H}}[w]$ from the ensemble of L received training sequences $\{\mathbf{y}_\ell[w]\}_{\ell=1}^{\ell=L}$, $w = 1, \ldots, W$, collected from *recurrent vehicle passages* in the radio cell, such that $\{\mathbf{y}_\ell[w]\}_{\ell=1}^{\ell=L}$ share the same ST propagation subspace (i.e., each UE experiences the same AoDs/AoAs and delays in communicating with the BS, and different fading amplitudes).

There are two possible implementations of the method, based on the available degree of cooperation between the UEs and the infrastructure (BS). In both cases, the notable advantage of MV-LR is the possibility, for the BS, to store the ST eigenmodes list, in order to avoid repeating the training procedure for each vehicle, minimizing the computations. Notice that a new training is requested only when macroscopic changes in the propagation environment occur.

Position-aware approach: The ST eigenmodes of the MIMO channel are explicitly associated to the physical UE position in the cell. The BS collects the L received training sequences $\{\mathbf{y}_\ell[w]\}_{\ell=1}^{\ell=L}$, $w = 1, \ldots, W$, for each location in the cell by relating them with the estimated physical UEs positions, obtained through either a suitable signaling or other localization techniques. The UEs are requested to cooperate with the infrastructure to build the database of ST channel eigenmodes, and the LR estimation performance depends on the positioning accuracy, which can be in the order of few meters in urban scenarios. As shown in [8], the position-aware MV-LR approach can attain the theoretical MSE lower bound only when the positioning accuracy is within ≈ 1 m, for typical hardware settings.

Position-agnostic approach: The ST eigenmodes of the MIMO channel are *not* related to a given physical UE position but rather are *subspace-dependent*. A huge dataset of N received training sequences $\{\mathbf{y}_\ell[w]\}_{\ell=1}^{\ell=N}$, $N \gg L$, not explicitly related to physical positions, is clustered at the BS with an unsupervised learning approach to *automatically* devise the algebraic similarity (subspace similarity) in the dataset. The cooperation between UEs and BS is minimal (exchange of training sequences, already in place for communication), and the performance of the system depends on the number K of chosen clusters, on the dataset (cardinality, data diversity), and on the selected similarity metric.

In the following, we briefly outline the algebraic background for the MV-LR channel estimation from L different training sequences $\{\mathbf{y}_\ell[w]\}_{\ell=1}^{\ell=L}$, $w = 1, \ldots, W$ (obtained from different vehicular UEs), assumed to share the same ST propagation subspace. The complete analytical treatment can be found in [6]. The MV-LR-estimated channel is retrieved through the application of a *training sequence-specific* matrix $\mathbf{T}_\ell$ and an *ensemble-specific* matrix $\mathbf{\Pi}(L)$ on single received training signal $\mathbf{y}_\ell = \left[\mathbf{y}_\ell^{\mathrm{T}}[1], \ldots, \mathbf{y}_\ell^{\mathrm{T}}[W]\right]^{\mathrm{T}} \in \mathbb{C}^{WN_R \times 1}$ as:

$$\widehat{\mathbf{h}}_\ell = \mathbf{\Pi}(L)\, \mathbf{T}_\ell\, \mathbf{y}_\ell = \mathbf{\Pi}(L)\, \overline{\mathbf{y}}_\ell, \tag{9}$$

where: *(i)* $\widehat{\mathbf{h}}_\ell \in \mathbb{C}^{WN_RN_T \times 1}$ is the MV-LR-estimated channel vector, that can be rearranged to obtain either $\widehat{\mathbf{H}}_\ell[w]$ and *(ii)* $\overline{\mathbf{y}}_\ell = \mathbf{T}_\ell\, \mathbf{y}_\ell \in \mathbb{C}^{WN_RN_T \times 1}$ is the U-ML channel estimate, analytically detailed in [5].

The ensemble-specific linear processing in (9) is designed as [6]:

$$\mathbf{\Pi}(L) = \widehat{\mathbf{C}}^{\frac{\mathrm{H}}{2}}\, \widehat{\mathbf{\Pi}}\, \widehat{\mathbf{C}}^{-\frac{\mathrm{H}}{2}}, \tag{10}$$

where *(i)* $\widehat{\mathbf{C}} \approx (1/\sigma_x^2)\, (\mathbf{I}_W \otimes \mathbf{I}_{N_T} \otimes \mathbf{Q}_n^{\mathrm{T}})$ is the estimated covariance matrix of $\overline{\mathbf{y}}_\ell$, needed to handle spatial/temporal noise correlations (e.g., interfering users) and *(ii)* $\widehat{\mathbf{\Pi}} = \widehat{\mathbf{U}}_{\mathrm{T}}^{*} \widehat{\mathbf{U}}_{\mathrm{T}}^{\mathrm{T}} \otimes \widehat{\mathbf{U}}_{\mathrm{S}}^{\mathrm{Tx},*} \widehat{\mathbf{U}}_{\mathrm{S}}^{\mathrm{Tx},\mathrm{T}} \otimes \widehat{\mathbf{U}}_{\mathrm{S}}^{\mathrm{Rx}} \widehat{\mathbf{U}}_{\mathrm{S}}^{\mathrm{Rx},\mathrm{H}}$ is the projection matrix onto the ST propagation subspace associated to the separable basis $\widehat{\mathbf{U}} = \widehat{\mathbf{U}}_{\mathrm{T}}^{*} \otimes \widehat{\mathbf{U}}_{\mathrm{S}}^{\mathrm{Tx},*} \otimes \widehat{\mathbf{U}}_{\mathrm{S}}^{\mathrm{Rx}}$. Orthonormal bases $\widehat{\mathbf{U}}_{\mathrm{T}} \in \mathbb{C}^{W \times r_{\mathrm{T}}}$, $\widehat{\mathbf{U}}_{\mathrm{S}}^{\mathrm{Tx}} \in \mathbb{C}^{N_T \times r_{\mathrm{S}}^{\mathrm{Tx}}}$ and $\widehat{\mathbf{U}}_{\mathrm{S}}^{\mathrm{Rx}} \in \mathbb{C}^{N_T \times r_{\mathrm{S}}^{\mathrm{Rx}}}$ are retrieved as the $r_{\mathrm{S}}^{\mathrm{Tx}}$, $r_{\mathrm{S}}^{\mathrm{Rx}}$ and r_{T} leading eigenvectors of the spatial (Tx and Rx) and temporal sample correlation matrices over L *whitened* U-ML channel estimates $\overline{\overline{\mathbf{y}}}_\ell = \widehat{\mathbf{C}}^{-\frac{\mathrm{H}}{2}} \overline{\mathbf{y}}_\ell$. Therefore, matrix $\mathbf{\Pi}(L)$ operates a noise-aware modal filtering on the U-ML channel estimate, whose effectiveness is proportional to the *unstructured sparsity degree* of the channel. It can be demonstrated that, if at least one of the following conditions holds:

$$r_{\mathrm{S}}^{\mathrm{Tx}} < N_T, \quad r_{\mathrm{S}}^{\mathrm{Rx}} < N_R, \quad r_{\mathrm{T}} < W, \tag{11}$$

the LR method outperforms the U-ML one. Asymptotically ($L \to \infty$), the MV-LR attains the maximum performance. The value of L for the asymptotic convergence depends on N_T, N_R and W as well as on the SNR. For the MIMO settings and bandwidths considered in Section V, $L \approx 100$ guarantees the convergence. Notice that $\mathbf{\Pi}(L)$ can be explicitly position-dependent (position-aware approach) or not (position-agnostic approach).

IV. Clustering-based MV-LR Channel Estimation

In this section, we describe the clustering algorithm used for the position-agnostic MV-LR implementation in V2X urban settings. Let us consider a large number, N, of training sequences $\{\overline{\overline{\mathbf{y}}}_i\}_{i=1}^{i=N}$, collected at the BS over the whole radio cell, already pre-processed by matrix $\mathbf{T}_i$ (U-ML channel estimates) and whitened. We aim at clustering them in order to *(i)* identify few representative received training sequences with markedly different ST features, allowing to define a finite set of K comprehensive ST patterns (*clusters*) easy to discriminate in a noisy setting; *(ii)* compute the LR orthonormal sets $\widehat{\mathbf{U}}_{\mathrm{T}}$, $\widehat{\mathbf{U}}_{\mathrm{S}}^{\mathrm{Tx}}$ and $\widehat{\mathbf{U}}_{\mathrm{S}}^{\mathrm{Rx}}$ for each cluster to efficiently apply MV-LR.

The proposed goals can be modelled in the framework of the K-medoids problem. With respect to the well-known K-means algorithm, K-medoids does not require the computation of a mean—which is meaningless for received training sequences belonging to different locations in space—, and it is more resilient to outliers and noise. Given a set of data points $X = \{x_j\}$, $j = 1, \cdots, N$, K-medoids clustering aims at selecting K elements m_i—called medoids—among them such that the sum of dissimilarities

$$D = \sum_{k=1}^{K} \sum_{x_n \in C_k} d(x_n, m_k) \tag{12}$$

is minimized, where C_k is the cluster represented by medoid m_k, and d is an arbitrary dissimilarity measure between two data points. A medoid m_k minimizes the intra-cluster sum of dissimilarities:

$$m_k = \underset{x_n \in C_k}{\operatorname{argmin}} \sum_{x_t \in C_k} d(x_n, x_t). \tag{13}$$

After a random initialization, the clusters are defined by assigning, according to the utilized dissimilarity measure, each dataset point to the nearest medoid, which can be considered a representative element of the cluster. In this work, to solve the K-medoids problem, the Partitioning Around Medoids (PAM) [11] algorithm has been used.

TABLE I
SIMULATION PARAMETERS

Simulation parameter	Symbol	Value
Carrier frequency	f_0	28 GHz
Bandwidth	B	1, 50 MHz
BS height from the ground	-	6 m
Number of BS antennas	N_R	64 (8×8)
Number of UE antennas	N_T	16 (4×4)
Training dataset size	N	$5 \cdot 10^4$ samples
Number of clusters	K	$2 - 30$
Signal to Noise Ratio	SNR	0 dB

For grouping ST-similar received training sequences, we take advantage of the subspace correlation index proposed in [12], deriving the following similarity metric:

$$\eta_{i,j} = \frac{\mathrm{tr}[\mathbf{R}_i \mathbf{R}_j^H]}{\sqrt{\mathrm{tr}[\mathbf{R}_i \mathbf{R}_i^H]\mathrm{tr}[\mathbf{R}_j \mathbf{R}_j^H]}} = \frac{1}{1 + d_{i,j}} \tag{14}$$

for $i, j \in 1, \cdots, N$, where $\mathbf{R}_i = \overline{\overline{\mathbf{y}}}_i \overline{\overline{\mathbf{y}}}_i^{\mathrm{H}}$, and $\eta_{i,j} \in (0, 1]$. The distance measure $d_{i,j} \in [0, \infty)$ is able to capture the *dissimilarity* of two received sequences $\overline{\overline{\mathbf{y}}}_i$, $\overline{\overline{\mathbf{y}}}_j$ in the ST domain, as shown in Section V. The proposed method can be summarized by the following steps: (1) collection at the BS of training sequences $\{\overline{\overline{\mathbf{y}}}_i\}_{i=1}^{i=N}$, transmitted by UEs crossing the radio cell; (2) clustering of the collected training sequences within the ST domain into K clusters by means of the PAM algorithm, using the dissimilarity metric $d_{i,j}$ in (14); (3) computation of the MV-LR ST orthonormal bases $\{\widehat{\mathbf{U}}_{\mathrm{T}}, \widehat{\mathbf{U}}_{\mathrm{S}}^{\mathrm{Tx}}, \widehat{\mathbf{U}}_{\mathrm{S}}^{\mathrm{Rx}}\}_k$, $k = 1, \cdots, K$, by using the corresponding clustered received training sequences; (4) filtering of the ℓ-th new received sequence $\overline{\overline{\mathbf{y}}}_\ell$ by using the set of LR orthonormal bases corresponding to medoid m_k, $k = 1, \cdots, K$, nearest to $\overline{\overline{\mathbf{y}}}_\ell$ with respect to dissimilarity $d_{i,j}$. Since the convergence of the MV-LR algorithm is affected by the number of available received training sequences per cluster, we adopt the silhouette method [13] to determine the clustering quality and to select a suitable number of clusters K, searching for: *(i)* an even distribution of the training points among clusters to ensure the convergence of the MV-LR algorithm for each of them, *(ii)* a high intra-cluster cohesion, and *(iii)* a low inter-cluster similarity.

The complexity of the proposed method can be effectively measured by the number of requested training sequences N; the complexity of PAM algorithm scales $\propto N^2$, which is affordable for the considered dataset (see Section V), but still inherently limited for very large datasets. A valid alternative is CLARA (Clustering for Large Applications) [14], which runs PAM multiple times on small subsamples of the original dataset.

V. NUMERICAL RESULTS

We analyse the performance of the proposed channel estimation method in the communication scenario (radio cell) depicted in Fig. 1. The BS, located at a height of 6 m from the ground, is equipped with a planar array of $N_R = 64$ antennas (8×8), while each UE with a planar array of $N_T = 16$ antennas (4×4). We select 28 GHz as the carrier frequency (compliant to 5G NR FR2) and two communication bandwidths: *(i)* $B = 1$ MHz, for which the MIMO channel is frequency-flat ($W = 1$); *(ii)* $B = 50$ MHz, producing a frequency-selective ($W = 7$ taps) MIMO channel. The set of simulation parameters is reported in Table I.

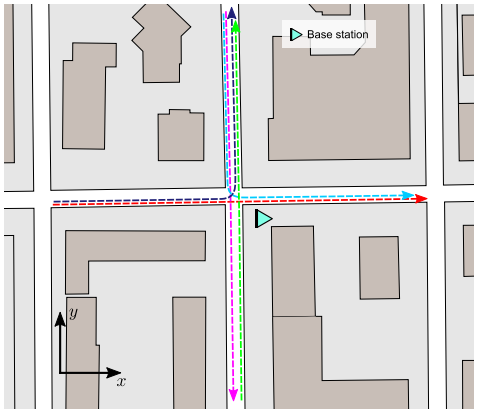

Fig. 1. Selected urban scenario and representation of the considered vehicular trajectories.

The recurrent vehicle passages in the cell are generated using SUMO (Simulation of Urban MObility) software [15], providing position, velocity and heading of vehicles over time for different realistic trajectories, exemplified in Fig. 1. The MIMO channel data over the trajectories is generated with the Altair WinProp ray-tracing software [16]. The algorithm has been trained using a dataset of $N = 5 \cdot 10^4$ received training sequences, sampled over the vehicular trajectories at 0 dB of SNR. The clustering-based MV-LR performance has been evaluated in terms of Normalized Mean Squared Error (NMSE), defined as:

$$\mathrm{NMSE} = \frac{\mathbb{E}[\|\mathbf{h}_\ell - \widehat{\mathbf{h}}_\ell\|^2]}{\mathbb{E}[\|\mathbf{h}_\ell\|^2]}, \tag{15}$$

where the channel estimate $\widehat{\mathbf{h}}_\ell$ can be U-ML, position-aware MV-LR or position-agnostic MV-LR. The position-aware MV-LR approach has been analytically proved in [5] to attain the theoretical MSE lower bound for $L \to \infty$ (in practice, $L > 100$); therefore, in this work we use it as benchmark to test the effectiveness of the proposed position-agnostic method. As we deal with complete vehicles' trajectories, the NMSE of the reference position-aware MV-LR method is averaged over the whole trajectory length.

Fig. 2 summarizes the results. Figs. 2a and 2b represent the optimal extracted clusters—depicted with different colors—over the geographical map of the urban scenario, and the selected trajectory realization, not comprised in the training dataset (dashed arrow). It is worth noticing that the colored clustered points on the map are not necessarily representative of UEs positions; they depict the invariance regions of the channel estimates in the ST domain for the retrieved clustering, each with a different spatial configuration. By using the

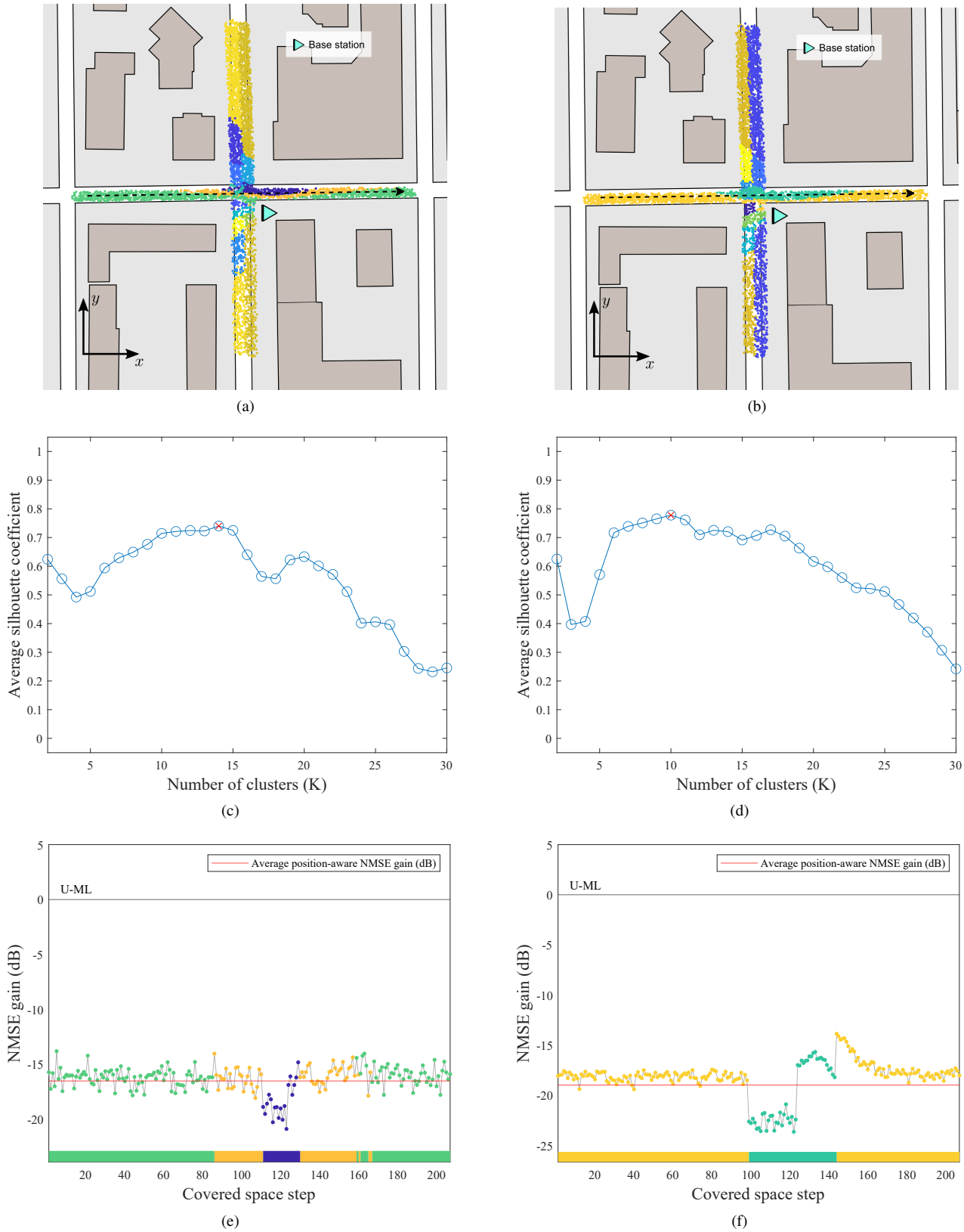

Fig. 2. Performance of the proposed algorithm at 1 MHz channel bandwidth (left) for $K = 14$, and 50 MHz channel bandwidth (right) for $K = 10$: spatial representation of clusters and reference trajectory (a,b); average silhouette coefficients for $K = 2$ to $K = 30$ clusters (c,d); NMSE performance comparison on a reference trajectory realization (e,f).

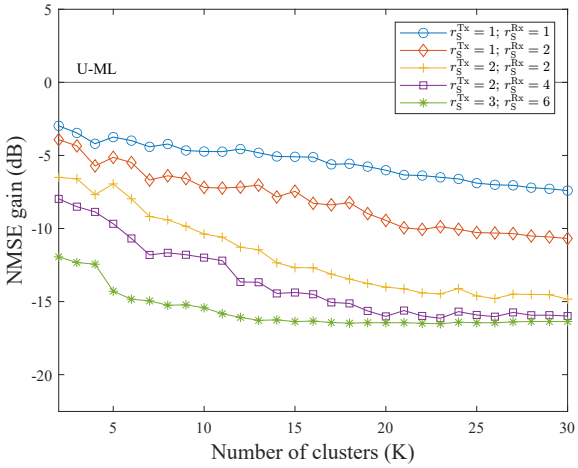

Fig. 3. Position-agnostic NMSE gain with respect to U-ML varying the number of clusters K, as function of spatial channel ranks, for $B = 1$ MHz bandwidth

silhouette method, detailed in [13], we selected the number of clusters leading to the largest average silhouette coefficient. Figs. 2c and 2d show the average silhouette coefficients over the whole scenario varying the number of clusters K from 2 to 30. For the considered urban scenario, a number of clusters $K = 14$ for $B = 1$ MHz and $K = 10$ for $B = 50$ MHz yields the maximum average silhouette coefficient and a substantial MSE improvement over U-ML channel estimates. From Fig. 3, we notice that the average NMSE gain decreases with K and with the selected channel spatial diversity orders (ranks) (for $B = 1$ MHz); this is due to the increasing cohesion in the ST domain of the the channel estimates grouped within each cluster. However, the analysis of the average silhouette coefficients provides a measure of the clusters' *relevance* for each K. As reported in Section IV, a clustering that suitably distributes the dataset points among clusters, as here, leads to a uniform NMSE performance within each cluster, allowing to match the position-aware MV-LR method. Therefore, a proper selection of K according to the silhouette coefficient allows to avoid performance unfairness across clusters. Finally, Figs. 2e and 2f show the position-agnostic MV-LR NMSE performance, normalized to the U-ML one, over the selected trajectory sampled over the covered space by 0.5 m steps. As can be seen, position-agnostic MV-LR outperforms U-ML by achieving ≈ 15 dB less ($B = 1$ MHz) and ≈ 20 dB less ($B = 50$ MHz) of MSE, attaining the same performance of the position-aware one. A similar behavior has been observed for all the other testing trajectories, not reported here, confirming the effectiveness of the proposed clustering-based MV-LR channel estimation method.

VI. Conclusion

This paper proposes a novel clustering-based MV-LR channel estimation method for 6G V2X. By clustering, through a K-medoids approach, a dataset of received training sequences from multiple UEs, the BS learns, in a completely unsupervised way, to aggregate training sequences sharing similar ST subspaces, to estimate the cluster-specific ST MIMO channel eigenmodes without the knowledge of UEs' geographical positions. For an optimal number of clusters selected by means of the silhouette method, numerical results show remarkable benefits in terms of NMSE, with an average reduction of $\approx 15 - 20$ dB with respect to the U-ML channel estimates, attaining the position-aware MV-LR performance. Future investigations will extend the proposed method to hybrid MIMO systems and to propagation affected by blockage, as well as to dynamic scenarios.

Acknowledgements

The research has been carried out in the framework of the Joint Lab between Huawei and Politecnico di Milano.

References

[1] C. De Lima *et al.*, "Convergent communication, sensing and localization in 6g systems: An overview of technologies, opportunities and challenges," *IEEE Access*, vol. 9, pp. 26 902–26 925, 2021.

[2] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1164–1179, 2014.

[3] S. Yang and L. Hanzo, "Fifty years of mimo detection: The road to large-scale mimos," *IEEE Communications Surveys Tutorials*, vol. 17, no. 4, pp. 1941–1988, 2015.

[4] W. U. Bajwa, J. Haupt, A. M. Sayeed, and R. Nowak, "Compressed channel sensing: A new approach to estimating sparse multipath channels," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1058–1076, 2010.

[5] A. Brighente, M. Cerutti, M. Nicoli, S. Tomasin, and U. Spagnolini, "Estimation of wideband dynamic mmwave and thz channels for 5g systems and beyond," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 9, pp. 2026–2040, 2020.

[6] M. Nicoli, O. Simeone, and U. Spagnolini, "Multislot estimation of fast-varying space-time communication channels," *IEEE Transactions on Signal Processing*, vol. 51, no. 5, pp. 1184–1195, 2003.

[7] M. Brambilla, D. Pardo, and M. Nicoli, "Location-assisted subspace-based beam alignment in los/nlos mm-wave v2x communications," in *ICC 2020 - 2020 IEEE International Conference on Communications (ICC)*, 2020, pp. 1–6.

[8] M. Mizmizi, D. Tagliaferri, D. Badini, C. Mazzucco, and U. Spagnolini, "Channel Estimation for 6G V2X Hybrid Systems Using Multi-Vehicular Learning," *IEEE Access*, vol. 9, pp. 95 775–95 790, 2021.

[9] K. Jung and H. Wang, "Pilotless channel estimation scheme using clustering-based unsupervised learning," in *2018 15th International Symposium on Wireless Communication Systems (ISWCS)*. IEEE, 2018, pp. 1–5.

[10] M. J. Azizipour and K. Mohamed-Pour, "Channel estimation for fdd multi-user massive mimo systems: a greedy approach based on user clustering," *IET Signal Processing*, vol. 13, no. 9, pp. 778–786, 2019.

[11] L. Kaufman and P. J. Rousseeuw, "Partitioning around medoids (program pam)," *Finding groups in data: an introduction to cluster analysis*, vol. 344, pp. 68–125, 1990.

[12] R. Bosisio and U. Spagnolini, "Enhanced broadcast opportunistic scheme based on spatial covariance feedback," in *Proc. International ITG-IEEE Workshop on Smart Antennas (WSA 2006)*, 2006, pp. 1–7.

[13] P. J. Rousseeuw, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," *Journal of computational and applied mathematics*, vol. 20, pp. 53–65, 1987.

[14] L. Kaufman and P. J. Rousseeuw, "Clustering large data sets," in *Pattern Recognition in Practice*, E. S. GELSEMA and L. N. KANAL, Eds. Amsterdam: Elsevier, 1986, pp. 425–437.

[15] P. A. Lopez, M. Behrisch, L. Bieker-Walz, J. Erdmann, Y.-P. Flötteröd, R. Hilbrich, L. Lücken, J. Rummel, P. Wagner, and E. Wießner, "Microscopic traffic simulation using sumo," in *The 21st IEEE International Conference on Intelligent Transportation Systems*. IEEE, 2018. [Online]. Available: https://elib.dlr.de/124092/

[16] Altair Engineering Inc., "Altair Feko," https://www.altair.com, 2020.