Channel Charting with Angle-Delay-Power-Profile Features and Earth-Mover Distance

Hanan Al-Tous[1], Parham Kazemi[1], Christoph Studer[2] and Olav Tirkkonen[1]

[1]*Department of Communications and Networking, Aalto University, Finland*
[2]*Department of Information Technology and Electrical Engineering, ETH Zurich, Switzerland*
Email: {hanan.al-tous, parham.kazemi, olav.tirkkonen}@aalto.fi, studer@ethz.ch

***Abstract*—We are interested in deducing whether two user equipments (UEs) in a cellular system are at nearby physical locations from measuring similarity of their channel state information (CSI). This becomes essential for fingerprinting localization as well as for channel charting. A channel chart is a low dimensional (e.g., 2-dimensional) radio map based on CSI measurements only, which is created using self-supervised machine learning techniques. Analyzing CSI in terms of the angle-delay power profile (ADPP) takes advantage of the uniqueness of the multipath channel between the base station and the UE over the geographical region of interest. We consider super-resolution features in the angle and delay domains in massive multiple-input multiple-output (MIMO) systems and consider the earth-mover distance (EMD) to measure the distance between two features. Simulation results based on the DeepMIMO data set show that the super-resolution ADPP features with EMD leads to a better quality channel chart as compared to other CSI features and distances from the literature.**

I. Introduction

Massive multiple-input multiple-output (mMIMO) base stations (BSs) can be used for location-based services, and the position of a user equipment (UE) can be estimated based on the high angular resolution provided by mMIMO. The wide bandwidths of 5G systems can be leveraged to further improve the accuracy of localization. With a large antenna array and a wide bandwidth, a BS can acquire higher multipath resolution in the angle and delay domains [1]. Several papers have considered joint estimation of delay, angle of arrival, and receive power for complex multipath environments [2], [3].

Localization methods based on channel state information (CSI) fingerprinting perform well under non-line-of-sight (NLoS) conditions and highly cluttered multipath environments. For example, K-nearest neighbour fingerprinting (KNN-FP) is widely used for this purpose [4]–[6]. The localization accuracy of KNN-FP highly depends on the used CSI feature and the feature distance.

Channel charting (CC) is a self supervised machine learning framework to create a radio map of the cell, constructed from CSI only, which preserves the neighborhood relations of UEs. Several CSI features, distances, and dimensionality reduction (DR) techniques have been considered for CC. Similar to KNN-FP-based localization, the CSI feature and feature distance highly affect the quality of the channel chart [7]–[9].

The simplest form of CSI feature based fingerprinting utilizes the received signal strength (RSS). In [10], a comprehensive study of distance and similarity measures for fingerprinting localization based on RSS was considered. Applying more generic CSI features for fingerprinting provides better localization performance than only relying on RSS [4]. The similarity measures devised for RSS cannot be directly used for generic CSI features based fingerprinting. CSI distances based on linear algebra, such as the Euclidean distance and correlation matrix distance, can reveal when two UEs are close to each other. Such distances, however, are unable to distinguish between whether another UE is immediately outside the nearest neighborhood of a UE or far away. Using linear algebraic CSI distances for applications like CC and KNN-FP requires a high sampling density of the radio environment. Identifying a distance measure that is able to reveal both near and far locations based on CSI is important for improving the quality of CC and for reducing the sampling density.

In this paper, we consider a super-resolution delay-angle-power profile feature and apply the earth-mover distance (EMD), which has the ability to go beyond the limitations of linear algebraic distances. We evaluate performance in terms of local topology, as well as measures of global geometry preservation and the Rajski distance measuring the mutual information between the distances of UEs in the feature domain, and distances in the spatial geometry. We use the EMD of the UEs in the cell to create the channel chart using classical DR techniques.

The remainder of this paper is organized as follows: In Section II, the system model is presented. In Section III, CSI features and linear algebraic distances are introduced. In Section IV, the super-resolution CSI feature and EMD are presented. In Section V, the channel charting and metric learning frameworks are discussed. Simulation results are presented and discussed in Section VII. Finally, conclusions are drawn in Section VIII.

II. System Model

We consider a mMIMO BS with a uniform linear array (ULA) of M antennas. The wireless signal transmitted from a UE propagates along K multipaths. The baseband channel response vector at the BS form multipath component k arising from an impulse transmitted at time 0 is

$$\mathbf{c}_k = \sqrt{\alpha_k} e^{-j\beta_k} \mathbf{a}(\phi_k)\, \delta(\tau_k), \tag{1}$$

where $\delta(\cdot)$ is the Dirac delta distribution, $\sqrt{\alpha_k}e^{-j\beta_k}$ is the complex attenuation, τ_k the propagation delay, and ϕ_k the angle of arrival (AoA) of path k. The array steering vector corresponding to AoA ϕ is

$$\mathbf{a}(\phi) = [1, e^{-j2\pi \frac{r \sin \phi_k}{\lambda_c}}, \ldots, e^{-j2\pi \frac{r(M-1) \sin \phi_k}{\lambda_c}}]^T, \quad (2)$$

where r is the antenna element spacing, and λ_c is the carrier wavelength.

Considering an orthogonal frequency-division multiplexing (OFDM) system with N subcarriers, assuming the cyclic prefix is larger than the maximum delay spread, the channel frequency response vector at the nth subcarrier becomes

$$\mathbf{h}_n = \sum_{k=0}^{K-1} \sqrt{\alpha_k} e^{-j\beta_k} \mathbf{a}(\phi_k)\, e^{-j2\pi n \frac{\upsilon_k}{N}}, \quad (3)$$

where $\upsilon_k = \lfloor \frac{\tau_k}{T} \rceil$ denotes the temporally resolvable propagation delay associated with the kth path, and T is the sample interval. The CSI matrix in the frequency-antenna domain is:

$$\mathbf{H} = [\mathbf{h}_0, \ldots, \mathbf{h}_{N-1}]^T \in \mathbb{C}^{N \times M}. \quad (4)$$

III. CSI Features and Linear Algebraic Distances

For fingerprinting based localization and CC, several CSI features have been considered [4]–[7]. The selection of the CSI feature and the distance highly affects the resulting performance. In this section, the CSI features are created from the CSI matrix using matrix operations. The distances in this section are computed based on methods of linear algebra, i.e., matrix operations as well as vector and matrix norms.

A. Channel Matrix

The channel matrix in different domains (frequency-antenna, frequency-beam, delay-antenna, and delay-beam) can be used as a CSI feature. We will consider, the frequency-antenna channel matrix $\mathbf{H}$ given by (4).

B. Covariance Matrix

The covariance matrix in different domains (antenna, beam, frequency, and delay) can be considered as a CSI feature. For example, the antenna domain covariance matrix can be estimated using S samples as

$$\mathbf{R} = \frac{1}{S\,N} \sum_{s=0}^{S-1} \mathbf{H}_s^H \mathbf{H}_s, \quad (5)$$

where $\mathbf{R} \in \mathbb{C}^{M \times M}$, and $\mathbf{H}_s \in \mathbb{C}^{N \times M}$ denotes the frequency-antenna channel matrix at time domain sample s. In $\mathbf{R}$, the effect of the frequency domain is averaged.

C. Finite Resolution Angle-Delay

The frequency-antenna channel matrix $\mathbf{H}_s$ can be transformed into an approximately sparsified matrix $\mathbf{H}_s^{'}$ in the beam-delay domain via 2D-DFT by [11]

$$\mathbf{H}_s^{'} = \mathbf{F}_{\mathrm{d}}^H \mathbf{H}_s \mathbf{F}_{\mathrm{a}}, \quad (6)$$

where $\mathbf{F}_{\mathrm{d}} \in \mathbb{C}^{N \times N}$ and $\mathbf{F}_{\mathrm{a}} \in \mathbb{C}^{M \times M}$ are DFT matrices. Due to limited spread, performing DFT on the frequency domain channel vectors can transform $\mathbf{H}_s$ into a sparse matrix in the delay domain, with only the first $N^{'} < N$ rows having distinct non-zero values. Therefore, $\mathbf{H}_s^{'}$ can be truncated to $\mathbf{H}_s^{''}$, by considering the first $N^{'}$ rows, i.e.,

$$\mathbf{H}_s^{''} = \mathbf{A}_s \mathbf{H}_s^{'}, \quad (7)$$

where $\mathbf{A} = [\mathbf{I}_{N'}, \mathbf{0}_{N' \times (N-N')}]$. The angle-delay-power matrix averaged over S samples is computed as [4]

$$\mathbf{\Omega} = \frac{1}{S} \sum_{s=0}^{S-1} \left| \mathbf{H}_s^{''} \right|^2, \quad (8)$$

where $|\cdot|^2$ is element-wise squared absolute value.

TABLE I
Notation and definitions of distances [12], [13].

Name	Notation	Expression
Euclidean	$d_{\mathrm{Euc}}\left(\mathbf{M}, \mathbf{M}^{'}\right)$	$\left\|\mathbf{M} - \mathbf{M}^{'}\right\|_{\mathrm{F}}$
CMD	$d_{\mathrm{CMD}}\left(\mathbf{M}, \mathbf{M}^{'}\right)$	$\frac{1}{2}\left\|\frac{\mathbf{M}}{\|\mathbf{M}\|_{\mathrm{F}}} - \frac{\mathbf{M}^{'}}{\|\mathbf{M}^{'}\|_{\mathrm{F}}}\right\|_{\mathrm{F}}^2$
Log Euc.	$d_{\mathrm{LogEuc}}\left(\mathbf{R}, \mathbf{R}^{'}\right)$	$\left\|\log(\mathbf{R}) - \log\left(\mathbf{R}^{'}\right)\right\|_{\mathrm{F}}$
Roots Euc.	$d_{\mathrm{RooEuc}}\left(\mathbf{R}, \mathbf{R}^{'}\right)$	$\left\|\sqrt{\mathbf{R}} - \sqrt{\mathbf{R}^{'}}\right\|_{\mathrm{F}}$
Chol. Euc.	$d_{\mathrm{CholEuc}}\left(\mathbf{R}, \mathbf{R}^{'}\right)$	$\left\|\mathrm{chol}(\mathbf{R}) - \mathrm{chol}\left(\mathbf{R}^{'}\right)\right\|_{\mathrm{F}}$
Riemannian	$d_{\mathrm{Riem}}\left(\mathbf{R}, \mathbf{R}^{'}\right)$	$\left\|\log\left(\sqrt{\mathbf{R}}\,\mathbf{R}^{'}\sqrt{\mathbf{R}}\right)\right\|_{\mathrm{F}}$
Chordal	$d_{\mathrm{Chord}}\left(\mathbf{R}, \mathbf{R}^{'}\right)$	$\left\|\mathrm{proj}_{\chi}(\mathbf{R}) - \mathrm{proj}_{\chi}\left(\mathbf{R}^{'}\right)\right\|_{\mathrm{F}}$
Procrustes	$d_{\mathrm{Proc}}\left(\mathbf{R}, \mathbf{R}^{'}\right)$	$\inf_{\mathbf{W}\mathbf{W}^H = \mathbf{I}} \left\|\sqrt{\mathbf{R}} - \sqrt{\mathbf{R}^{'}}\mathbf{W}\right\|_{\mathrm{F}}$

D. Linear Algebraic Distances

Multiple natural metrics can be derived from vector and matrix norms. The simplest distance between two matrices is the Euclidean distance (i.e., the Frobenius norm), which is invariant under unitrary transformation (including permutations), i.e., $\|\mathbf{M}\|_{\mathrm{F}} = \|\mathbf{U}_{\mathrm{L}}\mathbf{M}\mathbf{U}_{\mathrm{R}}\|_{\mathrm{F}}$ where $\mathbf{U}_{\mathrm{L}}$ and $\mathbf{U}_{\mathrm{R}}$ are unitary matrices. The correlation matrix distance (CMD) is the squared Euclidean distance between two normalized matrices, which is also unitarily invariant.

The Euclidean distance, while simple, is not a natural metric on the space of covariance matrices, since covariance matrices live on a non-Euclidean space. Several covariance distances based on matrix logarithms, square roots, and Cholesky decomposition have been considered [12], [13]. In addition, the affine invariant Riemannian, chordal, and Procrustes distances have been considered [14], [15]. All of these are invariant under unitary transformations except for the Cholesky distance. The *logEuclidean and Riemannian* metrics are geodesic, rotation invariant, scale invariant, and inversion invariant [14], [15]. These metrics are summarized in Table I. The notations $\mathbf{M}$ and $\mathbf{R}$ are used for a matrix and a covariance matrix, respectively. The matrix logarithm, square roots, Cholesky factorization and the projection to the subspace spanning the χ largest Eigenvectors of matrix $\mathbf{R}$ are denoted as $\log(\mathbf{R})$, $\sqrt{\mathbf{R}}$, $\mathrm{chol}(\mathbf{R})$, and $\mathrm{proj}_{\chi}(\mathbf{R})$, respectively.

IV. Super-Resolution CSI Feature and EMD

Assume that the channel has K multipath components (MPCs), represented by angle of arrivals ϕ_k, propagation delays τ_k and average powers p_k with $p_k = \mathbb{E}\{\alpha_k\}$ for $k = 1, \ldots, K$. Here, $\mathbb{E}\{\cdot\}$ denotes expectation. The angle-delay-power profile $\{\phi_k,\ \tau_k,\ p_k\}$ is the super-resolution CSI feature, which can be estimated from the channel matrix using the mutiple signal classification (MUSIC) and the space-alternating generalized expectation-maximization (SAGE) algorithms [2], [16].

It is not straightforward to measure the distance between two super-resolution CSI features. Different orderings of the MPCs may result in different distances, and handling two features with different number of MPCs is not trivial. To tackle these problems, we consider the earth-mover distance (EMD), which measures the distance between sets of MPCs, but not based on distances between individual components. As such the EMD is not affected by the ordering of MPCs, and can handle features with different numbers of MPCs.

A. Earth-Mover Distance

The EMD (also known as Wasserstein distance) is a distance defined between probability distributions on a given metric space. The EMD is computed by solving a linear program [17].

Let $\mathbf{X}_u = [\mathbf{x}_u(1), \ldots, \mathbf{x}_u(K_u)] \in \mathbb{R}^{d \times K_u}$ be the feature representation at point u in terms of K_u vectors and $\mathbf{w}_u = [w_u(1), \ldots, w_u(K_u)]$ is a weight vector representing the probability mass across the vectors $\mathbf{x}_u(k)$. To compute the EMD between the suppliers (modeled by $\mathbf{X}_1$ and $\mathbf{w}_1$) and the receivers (modeled by $\mathbf{X}_2$ and $\mathbf{w}_2$), we need first to compute the distance matrix $\mathbf{B} \in \mathbb{R}^{K_1 \times K_2}$, where $b_{i,j} = \|\mathbf{x}_1(i) - \mathbf{x}_2(j)\|_\nu$ and ν is the norm used to measure the distance between ith supplier and jth receiver. The EMD finds the flows $\{f_{i,j}\}$ moving the probability mass from the supplier points to the receiver points that minimize the overall cost, respecting the flows constraints [17]:

$$\underset{\{f_{i,j}\}}{\text{minimize}} \sum_{i=1}^{K_1} \sum_{j=1}^{K_2} f_{i,j} b_{i,j} \tag{9a}$$

$$\text{subject to } f_{i,j} \geq 0, \sum_{i=1}^{K_1} \sum_{j=1}^{K_2} f_{i,j} = 1, \tag{9b}$$

$$\sum_{j=1}^{K_2} f_{i,j} \leq w_1(i), \sum_{i=1}^{K_1} f_{i,j} \leq w_2(j). \tag{9c}$$

In (9b), the positivity constraints allow the flows to move from suppliers to receivers and not vice versa, and the equality constraint force to move the full probability mass from the suppliers to the receivers. Constraint (9c) limits the individual suppliers and receivers to their weights. Problem (9) is a LP which can be solved efficiently. Substituting the optimal flows in the objective function gives the EMD between the supplier probability mass function (PMF) and the receiver PMF.

B. Model of MPCs for EMD

We transform the super-resolution CSI feature into a form that can be used within the EMD framework. Two models are considered. For a UE at point u, the angles, delays and the powers of MPCs represented by $\{p_{u,k}\}$, $\{\tau_{u,k}\}$ and $\{\phi_{u,k}\}$ for $k = 1, \ldots, K_u$, are used to construct the feature $\mathbf{X}_u = [\mathbf{x}_u(1), \ldots, \mathbf{x}_u(K_u)] \in \mathbb{R}^{2 \times K_u}$ and the weight vector $\mathbf{w}_u = [w_u(1), \ldots, w_u(K_u)]$. In the first model, the kth component of the feature is constructed as:

$$\mathbf{x}_u(k) = [\tau_{u,k} \cos \phi_{u,k},\ \tau_{u,k} \sin \phi_{u,k}]^T, \tag{10}$$

where $\mathbf{x}_u(k) \in \mathbb{R}^{2 \times 1}$. The power of a MPC is used to create the corresponding probability weight, i.e., $w_u(k) = \frac{p_{u,k}}{\sum_{i=1}^{K_u} p_{u,i}}$. We call this model "ADPP-Prop," since the weight is proportional to the normalized power. In the second model, the kth component of the feature $\mathbf{x}_u(k)$ is the same as in (10), however, the power of a MPC is ignored and equal probability weight is assigned to each MPC, i.e., $w_u(k) = \frac{1}{K_u}$. We call this model "ADPP-Eq."

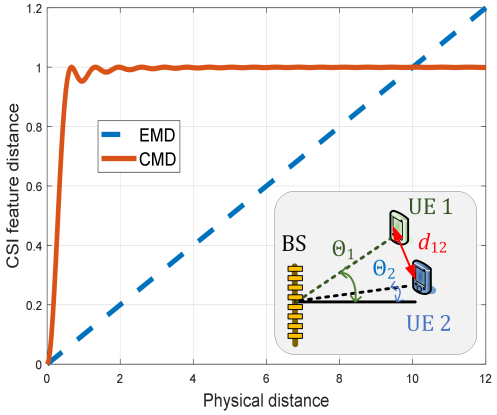

Fig. 1. The feature distance versus physical distance for covariance features using CMD, and super-resolution features using EMD. The layout is shown in the lower right corner of the figure.

C. Illustrative Example

To understand the difference between linear algebraic distances and EMD, we consider a simple scenario, where the BS is equipped with ULA consisting of $M = 32$ elements. Two UEs are in the radio environment, each equipped with a single antenna. The received signal at the BS of each UE has one MPC as illustrated in the lower right corner of Figure 1. The AoA of UE i is denoted by Θ_i. We assume that the location of the first UE is fixed, and the second UE moves from that location along a circle centered at the BS. We consider two CSI features and distances: (i) the covariance with CMD distance (other algebraic distances can be used). (ii) The super-resolution (i.e., AoA) with EMD. Figure 1 shows the CSI feature distances (i.e., CMD and EMD) versus the physical distance. We can see that the covariance feature with CMD can resolve only small physical distances, as long as the other user stays within the beam width. In contrast, the super-resolution feature with EMD is able to accurately reveal larger physical distances.

V. Channel Charting and Metric Learning

The underlying assumption of CC is that there exits a continuous mapping from the spatial location $\boldsymbol{z}_u$ of UE u

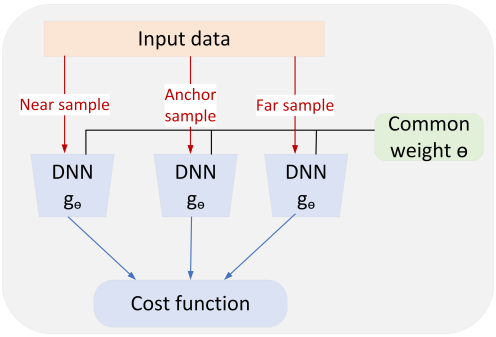

Fig. 2. Triplet network architecture. The three DNN blocks are identical and use common weights [9], [19].

to the CSI feature $\mathbf{y}_u$ at the BS:

$$\mathcal{H} : \mathbb{R}^{\rho} \to \mathbb{C}^{F}; \quad \mathcal{H}(\boldsymbol{z}_u) = \mathbf{y}_u. \tag{11}$$

Here, ρ is the spatial dimension which is either 2 or 3, and F is the feature dimension. CC starts by measuring the distance of the CSI features between pairs of UEs as seen at the BS. Next, based on the distance matrix, low dimensional channel chart is found in a *self-supervised manner*, providing chart locations for the set of UEs, such that UEs that are neighbors in the physical space will be neighbors in the channel chart. Several DR techniques have been considered for CC. We apply t-distributed stochastic neighbor embedding (t-SNE) for CC, as it has been shown to perform well in [8], [18].

We consider metric learning to benchmark the performance of CSI features and distances. For this, the physical locations are assumed to be known and used to label the CSI features. The data set $\mathcal{T}$ consists of CSI features $\{\mathbf{y}_i\}$ and the corresponding true physical locations $\{\mathbf{z}_i\}$. The data set $\mathcal{T}$ is arranged in the form of triplets of samples $(i, j, k) \in \mathcal{T}_{\mathrm{a}}$, for which it is known that the respective true distances

$$\|\mathbf{z}_i - \mathbf{z}_j\|_2 < \|\mathbf{z}_i - \mathbf{z}_k\|_2 \,, \tag{12}$$

i.e., point j is closer to the reference point i than than point k. Distance learning aims to find a deep neural network (DNN) which maps CSI features to channel chart points with a function $g_{\boldsymbol{\theta}}(.)$ parameterized with $\boldsymbol{\theta}$ such that

$$d_{\boldsymbol{\theta}}(\mathbf{y}_i, \mathbf{y}_j) < d_{\boldsymbol{\theta}}(\mathbf{y}_i, \mathbf{y}_k), \tag{13}$$

where $d_{\boldsymbol{\theta}}(\mathbf{y}_i, \mathbf{y}_j) = \|g_{\boldsymbol{\theta}}(\mathbf{y}_i) - g_{\boldsymbol{\theta}}(\mathbf{y}_j)\|$. The cost function is defined as [9]:

$$\frac{1}{|\mathcal{T}_{\mathrm{a}}|} \sum_{(i,j,k)\in\mathcal{T}_{\mathrm{a}}} \left(d_{\boldsymbol{\theta}}(\mathbf{y}_i, \mathbf{y}_j) - d_{\boldsymbol{\theta}}(\mathbf{y}_i, \mathbf{y}_k) + \mu\right)^{+} , \tag{14}$$

where $(t)^{+} = \max(t, 0)$ and $\mu > 0$ is a margin to avoid trivial solutions. Figure 2 shows the triplet network structure: the three DNN blocks share common weights and biases.

VI. Performance Evaluation

We assume the original space is the physical location space, and the representation space is the CSI feature space. The distance in the original space is based on Euclidean distance, while, the distance in the representation space is based on CSI features distance.

To measure the degree of preserving the local topology, we consider the continuity (CT) and trustworthiness (TW). For a data set of U points, these can be computed by considering a neighborhood of J points, denoted as $V_J(\mathbf{z}_i)$, around locations $\{\mathbf{z}_i\}_{i=1}^{U}$ in the original space, and the J-neighborhood denoted as $V'_J(\mathbf{y}_i)$, around the corresponding points $\{\mathbf{y}_i\}_{i=1}^{U}$ in the representation space. The equations to compute the average values are given as

$$\mathrm{CT}(J) = 1 - a \sum_{i} \sum_{\substack{j\in V_J(\mathbf{z}_i) \\ j\notin V'_J(\mathbf{y}_i)}} \left(r(i,j) - J\right), \tag{15}$$

$$\mathrm{TW}(J) = 1 - a \sum_{i} \sum_{\substack{j\notin V_J(\mathbf{z}_i) \\ j\in V'_J(\mathbf{y}_i)}} \left(r^{'}(i,j) - J\right), \tag{16}$$

where $r(i,j)$ is the rank of a point $\mathbf{z}_i$ in terms of its distance from a point $\mathbf{z}_j$ in original space, $r^{'}(i,j)$ is the rank of a point $\mathbf{y}_i$ in terms of its distance from a point $\mathbf{y}_j$ in the representation space and $a = \frac{2}{UJ(2U-3J-1)}$ is a normalization factor.

Global geometry preservation is measured by the Kruskal Stress (KS). It is computed by comparing pairwise distance matrix $\bar{\mathbf{D}}$ of the points in original space $\{\mathbf{z}_i\}_{i=1}^{U}$ with pairwise distance matrix $\mathbf{D}$ of points in the representation space $\{\mathbf{y}_i\}_{i=1}^{U}$ using a distance scaling factor λ as [20]:

$$\mathrm{KS} = \min_{\lambda} \sqrt{\frac{\sum_{i,j} \left(d_{i,j} - \lambda \bar{d}_{i,j}\right)^2}{\sum_{i,j} d_{i,j}^2}}, \tag{17}$$

where $\bar{d}_{i,j} = \|\mathbf{z}_i - \mathbf{z}_j\|_2$ and $d_{i,j} = d(\mathbf{y}_i, \mathbf{y}_j)$.

In addition, we consider the Rajski distance to measure the mutual information between the quantized pairwise distances in the original space and the quantized pairwise distances in the representation space, which is computed as [21]:

$$d_{\mathrm{RD}}(V, Q) = 1 - \frac{I(V,Q)}{H(V,Q)}, \text{ for } H(V,Q) \neq 0, \tag{18}$$

where V and Q are two discrete random variables,

$$I(V,Q) = \sum_{v\in V,\, q\in Q} P_{V,Q}(v,q) \, \log_2 \frac{P_{V,Q}(v,q)}{P_V(v) P_Q(q)}, \tag{19}$$

is the mutual information, which measures the dependence between the two distributions,

$$H(V,Q) = -\sum_{v\in V,\, q\in Q} P_{V,Q}(v,q) \, \log_2 P_{V,Q}(v,q), \tag{20}$$

is the joint entropy information of V and Q, $P_{V,Q}(v,q)$ is the joint probability distribution, and $P_V(v)$ and $P_Q(q)$ are the marginal distributions of the quantized pairwise distances in the original space and the quantized pairwise distances in the representation space, respectively.

All four metrics are in the range $[0, 1]$ with the optimal value being 1 for TW and CT and 0 for KS and RD.

TABLE II
SIMULATION PARAMETERS.

Parameter	Value	Parameter	Value
Center Freq.	3.5 GHz	Subcarriers	4086
Scenario	O1-3p5	BW	100 MHz
BS Location	[287.5, 489.5] m	UE Locations	Vert. str.
BS Height	6	UE Height	2 m
BS Array	32 ULA	UE Array	1

VII. SIMULATION

We use the DeepMIMO data set to generate the CSI [22]. Channels are constructed based on ray-tracing data obtained from the Remcom Wireless InSite [23] channel emulator. Figure 3 shows the simulation layout. One BS is considered, marked with a red star in the layout. The UEs are located in the southern part of the vertical street. The streets have buildings on both sides. Buildings are 60 m×60 m or 60 m×30 m, with the height written on the layout. A distance of 1.2 m between adjacent UEs samples is considered. Table II summarizes the simulation parameters. Some location in the street are blocked. The data set consists of 4100 UE locations, after filtering out the blocked locations. For each location six samples with 0.2 m spacing are used to compute the covariance matrix and the finite resolution CSI features. The super-resolution features are taken directly from the data set.

We evaluate the performance of several CSI features and distances in terms of CT, TW, KS, and RD. We consider $J = 100$ neighbours (which corresponds to 2.4% of the data set) to compute the CT and TW. To compute the Rajski distance, the pairwise distances are divided into 20 bins. Performance results are summarized in Table III.

First, we consider the frequency-antenna domain channel matrix given by (4) with the Euclidean distance and CMD. The average of CT and TW, i.e., $(\mathrm{CT} + \mathrm{TW})/2$, is considered to characterize topology preservation. The results are inconclusive; Euclidean distance outperforms CMD in terms of all measures except KS. The covariance matrix feature (5) is considered with several linear algebraic distances. Overall, the covariance feature outperforms the channel matrix feature in terms of all performance measures. For the covariance feature, the Log-Euclidean distance outperforms CMD, Euclidean, square-root Euclidean, Cholesky Euclidean, affine invariant Riemannian, chordal and Procrustes distances in terms of all measures except KS, for which CMD is better.

For the finite resolution angle-delay (FRAD) feature in (8), $N^{'} = 10$ is used. The Euclidean distance and CMD are considered with this feature. Overall, the covariance feature is better than FRAD, except for KS. As for the covariance feature, for FRAD the Euclidean metric is better at preserving topology, while CMD preserves global geometry better and provides more mutual information.

The super-resolution feature with EMD outperforms all of the other features and distances in terms of all considered performance measures with a wide margin. Using weights

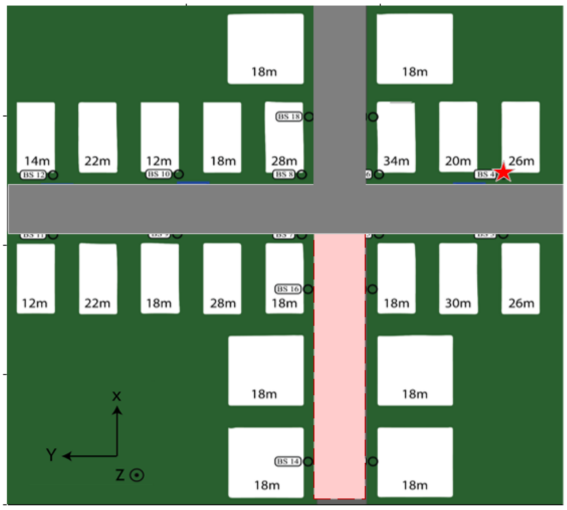

Fig. 3. Simulation layout. The selected street segment is in pink colour. The BS location is indicated by a red star.

based on the normalized power (i.e., ADPP-Prop) slightly improves performance.

A DNN composed of six fully connected layers and five batch normalization layers is trained to learn the mapping function $g_{\boldsymbol{\theta}}$ by optimizing the cost function (14) using the Adam optimizer. Each fully connected layer except the last one applies a ReLU nonlinearity [9]. We consider the following CSI features to train the DNN: (i) The covariance matrix. (ii) The covariance matrix normalized by its Frobenius norm. (iii) The logarithm of the covariance matrix. (iv) The finite-resolution angle delay profile. (v) The super-resolution angle-delay-power-profile. For the first three features, the real and imaginary parts are stacked in a vector to create a data entry. With metric learning, the super-resolution feature outperforms all other features. Interestingly, EMD outperforms metric learning in terms of all performance measures. Figure 4 shows the average of the CT and TW measures as a function of the number of neighbours J (in percentage of the number of points in the data set). The covariance matrix feature with CMD and LogEuc is compared to the super-resolution feature with EMD and metric leaning. The super-resolution feature with EMD performs best, with metric learning performing worse for small J, but approaching EMD when J increases. The gains of the super-resolution based methods over the linear algebraic distances grows with the number of neighbours. This confirms the observation made in Section IV-C, i.e., the super-resolution feature with EMD preserves larger distances than the linear algebraic distances. Simulation results indicate that metric learning is capable of preserving this property of the super-resolution feature, albeit slightly worse than EMD.

Figure 5 shows the Rajski distance as a function of the number of bins for the same features and metrics as in the previous figure. Super-resolution features consistently outperform covariance features with linear algebraic distances.

Finally, we create 2D channel charts for several features and

TABLE III
PERFORMANCE MEASURES FOR DIFFERENT CSI FEATURES AND DISTANCES. ALL MEASURES ARE IN $[0, 1]$; ↑ LARGE IS BETTER; ↓ SMALL IS BETTER.

CSI Feature	Distance	CT↑	TW↑	KS↓	RD↓
Channel	Euclidean	0.73	0.57	0.63	0.97
	CMD	0.70	0.50	0.56	0.98
Covariance	Euclidean	0.93	0.92	0.81	0.97
	CMD	0.72	0.91	0.55	0.96
	LogEuc	0.93	0.93	0.60	0.96
FRAD	Euclidean	0.92	0.87	0.81	0.97
	CMD	0.79	0.89	0.51	0.96
ADPP-Prop	EMD	**0.99**	**0.96**	**0.25**	**0.84**
ADPP-Eq		0.98	0.94	0.27	0.85
Covariance	Met. Lea.	0.81	0.81	0.83	0.98
Nor. Cov.		0.73	0.73	0.84	0.98
Log Cov.		0.92	0.85	0.45	0.95
FRAD		0.92	0.89	0.47	0.96
ADPPP		0.96	0.95	0.30	0.88
CC Cov. CMD	Euclidean	0.74	0.86	0.54	0.96
CC Cov. LogEuc		0.93	0.92	0.59	0.96
CC FRAD CMD		0.73	0.86	0.53	0.97
CC ADPPP EMD		**0.97**	**0.95**	**0.22**	**0.83**

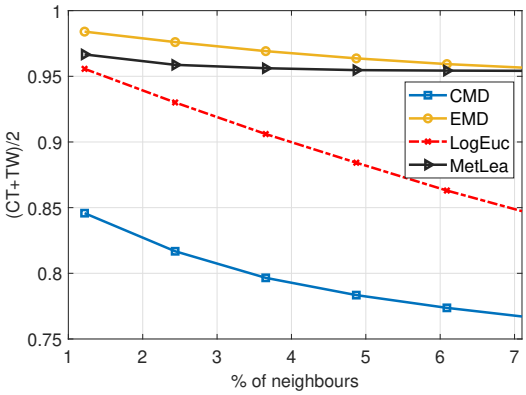

Fig. 4. $(\mathrm{TW} + \mathrm{CT})/2$ as a function of the number of neighbours J for covariance with CMD and LogEuc, super-resolution with EMD and super-resolution with metric learning.

distances. We consider covariance with CMD, covariance with LogEuc, finite resolution angle-delay profile with CMD and super-resolution with EMD using t-SNE DR technique. The CC based on super-resolution with metric learning is obtained directly from the mapping function $g_{\boldsymbol{\theta}}(.)$. The channel charts are shown in Figure 6. The CC locations are marked by colors corresponding to the physical locations shown in (a). Note that areas of blockage are removed from the image. We evaluate the CCs in terms of CT, TW, KS, and RD, comparing the pairwise distances of physical and CC locations. The CCs performance measures are shown in lower part of Table III. Metric learning

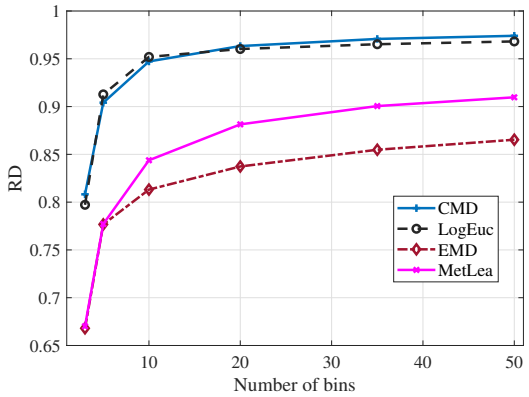

Fig. 5. The Rajski distance as a function of the number of bins for covariance with CMD and LogEuc, super-resolution with EMD and super-resolution with metric learning.

directly learns a metric in 2D, thus for metric learning the results in Table III represent DR. The KS of the CC based on super-resolution CSI with EMD is 0.22, super-resolution CSI with metric learning gives 0.30 while covariance features with Log-Euclidean distance give 0.59. The channel chart based on the super-resolution CSI with EMD is best at preserving global geometry. Note that KS is slightly improved in dimensionality reduction from feature dimension to 2D.

VIII. CONCLUSIONS

In this paper, we studied super-resolution angle-delay-power profile channel state information (CSI) features. The discrete nature of these features makes linear algebraic distances not particularly suitable to measure the distance between two features. In this regard, we considered the earth-mover distance (EMD). We compared the performance of the super-resolution features and EMD with other CSI features and linear algebraic distances in terms of local topology preservation, global geometry preservation, and mutual information measures. Our simulation results showed that the super-resolution features with EMD outperformed other features and distances in all considered performance measures. To benchmark performance, we also considered a metric learning approach, in which a triplet neural network was trained based on full knowledge of the true physical distances. Super-resolution features with EMD outperformed metric learning in terms of all performance measures, even when super-resolution features were used for metric learning. This highlights the importance of domain knowledge when applying machine learning to CSI. Two-dimensional channel chart based on the super-resolution features with EMD outperformed the channel chart based on other CSI features and distances as well as the channel chart based on a triplet neural network.

ACKNOWLEDGMENT

This work was funded in part by the Academy of Finland (grant 319484), and by the WINDMILL project funded by the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 813999.

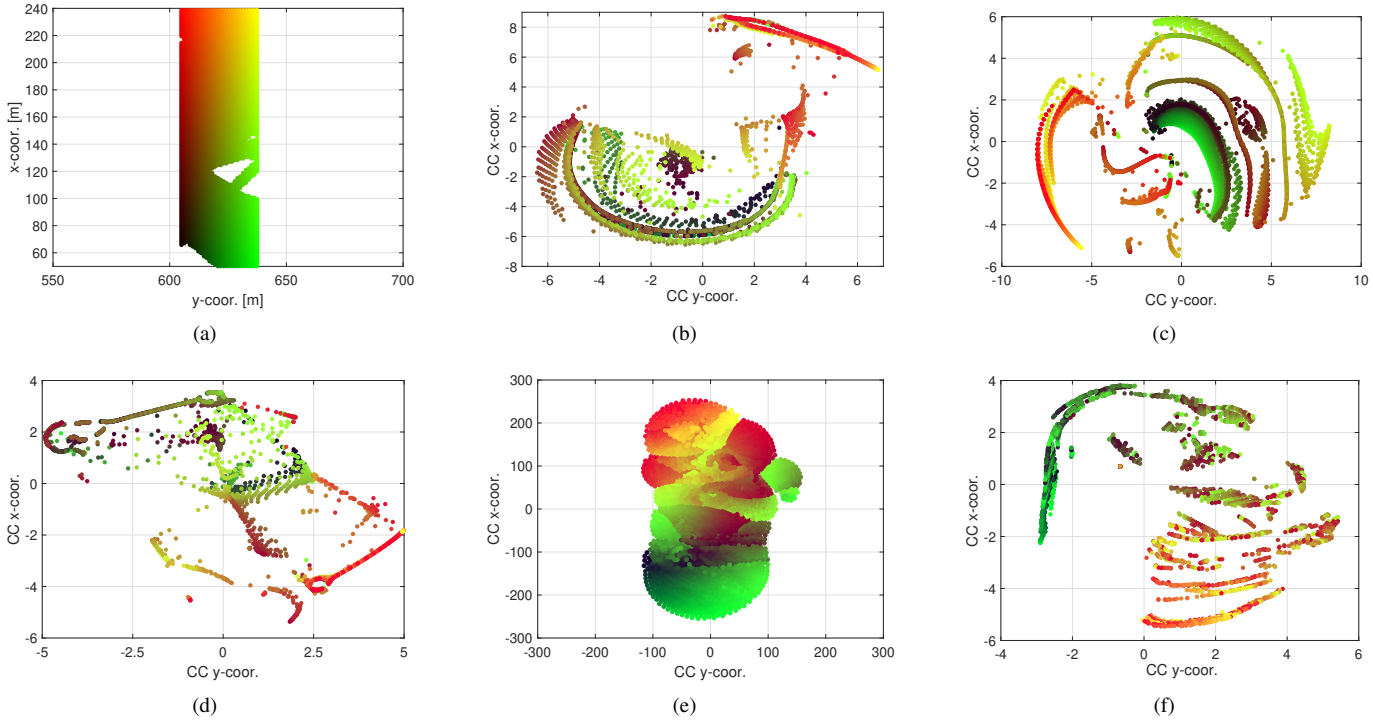

Fig. 6. (a); UE physical locations. (b); CC based on covariance with CMD using t-SNE. (c); CC based on covariance with LogEuc using t-SNE. (d); CC based on FRAD with CMD using t-SNE. (e); CC based on super-resolution with EMD using t-SNE. (f); CC based on super-resolution using DNN. CC locations are marked by colours corresponding to the physical locations in (a).

REFERENCES

[1] D. Burghal, A. T. Ravi, V. Rao, A. Alghafis, and A. F. Molisch, "A comprehensive survey of machine learning based localization with wireless signals," *ArXiv*, vol. abs/2012.11171, 2020.

[2] H. Xu, Y. Zhang, B. Ba, D. Wang, and X. Li, "Fast joint estimation of time of arrival and angle of arrival in complex multipath environment using OFDM," *IEEE Access*, vol. 6, pp. 60 613–60 621, 2018.

[3] J. Du, W. Cui, B. Ba, C. Jian, and H. Li, "Fast joint estimation for time delay and angle of arrival based on smooth preprocessing with orthogonal frequency division multiplexing," *IET Radar, Sonar & Navigation*, pp. 1–12, 2022.

[4] 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, 2018.

[5] A. Sobehy, Éric Renault, and P. Muhlethaler, "CSI-MIMO: K-nearest neighbor applied to indoor localization," in *Proc. ICC*, 2020, pp. 1–6.

[6] E. Gonultas, E. Lei, J. Langerman, H. Huang, and C. Studer, "CSI-Based multi-antenna and multi-point indoor positioning using probability fusion," *IEEE Trans. Wireless Commun.*, vol. 21, no. 4, pp. 2162–2176, 2022.

[7] C. Studer, S. Medjkouh, E. Gonultaş, T. Goldstein, and O. Tirkkonen, "Channel charting: Locating users within the radio environment using channel state information," *IEEE Access*, vol. 6, pp. 47 682–47 698, 2018.

[8] J. Deng, S. Medjkouh, N. Malm, O. Tirkkonen, and C. Studer, "Multipoint channel charting for wireless networks," in *Proc. Asilomar Conf. Sign., Syst., Comput.*, Oct. 2018, pp. 286–290.

[9] P. Ferrand, A. Decurninge, L. G. Ordonez, and M. Guillaud, "Triplet-based wireless channel charting: Architecture and experiments," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 8, pp. 2361–2373, 2021.

[10] J. Torres-Sospedra, R. Montoliu, S. Trilles, Óscar Belmonte, and J. Huerta, "Comprehensive analysis of distance and similarity measures for Wi-Fi fingerprinting indoor positioning systems," *Expert Syst. Appl.*, vol. 42, no. 23, pp. 9263–9278, 2015.

[11] T. Wang, C.-K. Wen, S. Jin, and G. Y. Li, "Deep learning-based CSI feedback approach for time-varying massive MIMO channels," *IEEE Wireless Commun. Lett.*, vol. 8, no. 2, pp. 416–419, 2019.

[12] I. L. Dryden, A. Koloydenko, and D. Zhou, "Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging," *Ann Appl Stat*, vol. 3, no. 3, pp. 1102 – 1123, 2009.

[13] S. Kaltenstadler, S. Nakajima, K.-R. Muller, and W. Samek, "Wasserstein stationary subspace analysis," *IEEE J. Sel. Top. Signal Process.*, vol. 12, no. 6, pp. 1213–1223, 2018.

[14] Y. Thanwerdas and X. Pennec, "O(n)-invariant Riemannian metrics on SPD matrices," Sep. 2021.

[15] R. Vemulapalli and D. W. Jacobs, "Riemannian metric learning for symmetric positive definite matrices," *ArXiv*, vol. abs/1501.02393, 2015.

[16] B. Fleury, M. Tschudin, R. Heddergott, D. Dahlhaus, and K. I. Pedersen, "Channel parameter estimation in mobile radio environments using the SAGE algorithm," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 3, pp. 434–450, 1999.

[17] Y. Rubner, C. Tomasi, and L. Guibas, "The earth mover's distance as a metric for image retrieval," *Int. J. Comput. Vis.*, vol. 40, no. 2, pp. 99–121, 2010.

[18] L. van der Maaten, "Learning a parametric embedding by preserving local structure," in *AISTATS*, 2009.

[19] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *Proc. CVPR*, 2015, pp. 815–823.

[20] J. B. Kruskal, "Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis," *Psychometrika*, vol. 29, pp. 1–27, 1964.

[21] C. Rajski, "A metric space of discrete probability distributions," *Inf. Control*, vol. 4, no. 4, pp. 371–377, 1961.

[22] A. Alkhateeb, "DeepMIMO: A generic deep learning dataset for millimeter wave and massive MIMO applications," in *Proc. ITA*, Feb. 2019, pp. 1–8.

[23] Remcom, "Wireless InSite," http://www.remcom.com/wireless-insite.