Unsupervised Charting of Wireless Channels

Saïd Medjkouh[1], Emre Gönültaş[1], Tom Goldstein[2], Olav Tirkkonen[3], and Christoph Studer[1]
[1]Cornell University, Ithaca, NY; [2]University of Maryland, College Park, MD; [3]Aalto University, Finland

***Abstract*—Future wireless communication systems will rely on large antenna arrays at the infrastructure base stations (BSs) to serve multiple users with high data rates in a single cell. We demonstrate that the availability of high-dimensional channel state information (CSI) acquired at such multi-antenna BSs enables one to learn a chart of the radio geometry, which captures the spatial geometry of the users so that points close in space are close in the channel chart, using no other information than wireless channels of users. Specifically, we propose a novel unsupervised framework that first extracts channel features from CSI which characterize large-scale fading effects of the channel, and then uses specialized dimensionality reduction tools to construct the channel chart. The channel chart can, for example, be used to perform (relative) user localization, predict cell hand-overs, or guide scheduling tasks, without accessing location information from global navigation satellite systems.**

I. Introduction

Fifth-generation (5G) wireless communication systems must be able to serve a large number of users within a given cell and provide higher spectral efficiency than existing communication systems [1], [2]. Massive multi-user multiple-input multiple-output (mMIMO) is widely believed to be a core technology that provides a solution for these requirements [3]–[5]. The idea of mMIMO is to deploy hundreds of antennas at the infrastructure base-stations (BSs) which enables fine-grained beamforming to transmit and receive user data at the same time and in the same frequency band. However, user mobility, sharp hand-over regions between neighboring cells, and massive device densification, mainly driven by the Internet of Things, pose severe implementation challenges of this technology.

It is widely believed that machine learning will play a critical role in unlocking the true potential of 5G technologies [6]. Concretely, one must provide the BS with information about user mobility, cell hand over, and congested areas to successfully deploy mMIMO in practice. To this end, one must lean heavily on the available high-dimensional channel state information (CSI) acquired at the multi-antenna BS. To effectively use the collected CSI, the BS has to *learn* the radio geometry in which the users are moving. What needs to be learned is a *chart* of the radio geometry, which represents *user location* and movement *related to CSI*. To automate learning and charting, to dynamically adapt to changes in the environment, to prevent extensive measurement campaigns, and to avoid human intervention, learning the radio geometry should be *unsupervised*.

The work of SM, EG, and CS was supported in part by Xilinx Inc., and by the US NSF under grants ECCS-1408006, CCF-1535897, CAREER CCF-1652065, and CNS-1717559. The work of EG research was supported in part by a fellowship from the Ministry of National Education of the Republic of Turkey. The work of TG was supported by the US NSF under grant CCF-1535902 and by the US Office of Naval Research under grant N00014-17-1-2078. The work of OT was funded in part by the Kaute foundation, Nokia Foundation, and the Academy of Finland (grant 299916). The authors thank O. Castañeda for useful comments on the paper.

A. Contributions

We show that the availability of high-dimensional CSI in mMIMO systems enables one to learn a chart of the radio geometry that captures the spatial geometry of the users so that points close in space are also close in the channel chart. The proposed method, referred to as *channel charting (CC)*, maps high-dimensional CSI of the users into a low-dimensional *channel chart* that accurately preserves the local geometry of the users' true location in space. CC first extracts suitable channel features that represent slowly varying CSI components in time and space; these features are then used to learn a channel chart via specialized dimensionality reduction methods. In contrast to existing localization methods that require vast amounts of CSI labelled with the users' true locations, e.g., obtained from global navigation satellite systems (GNSSs), CC operates in an unsupervised fashion. The unsupervised nature of CC avoids the need for extensive measurement campaigns and enables the BSs to perform cognitive and predictive tasks critical to 5G networks which are currently unavailable. We demonstrate the efficacy of CC via numerical simulations under realistic channel models and conditions.

B. Relevant Prior Art

Unsupervised charting of the radio geometry has not been addressed in the literature. Existing methods for user localization are mainly based on triangulation that use geometrical models to map a low-level descriptor of the channel, such as the received signal strength or angle-of-arrival, to a location in space [7]–[9]. For triangulation to work, cooperation of multiple BS located far from each other and line-of-sight (LoS) propagation conditions are necessary. Similarly, for channel fingerprinting [10]–[15], a channel map is generated via extensive measurement campaigns that directly associates CSI with position information from GNSSs [10], [12]–[17]. Fingerprinting methods are *fully supervised* and require training of the wireless channel at wavelength scales in space [17]. In contrast, CC is unsupervised and requires orders-of-magnitude sparser spatial sampling. Furthermore, supervision achieved by means of precise location information with application layer localization services, such as GNSS, is currently infeasible as this information is not accessible to BSs according to OSI layering principles [18].

In charting of wireless channels, we are primarily interested in preserving the local neighborhood structure of the spatial

geometry when charting the radio geometry. Tools from manifold learning [19], [20] and dimensionality reduction [21] have extensively been used to compute low-dimensional embeddings that preserve the geometry of high-dimensional datasets. Specific methods for these tasks are, among many others, multidimensional scaling (MDS) [19] or Sammon's mapping [20], which attempt to embed a high-dimensional manifold into a low-dimensional space. While such machine learning tools have been extensively used for data mining and data visualization purposes [22], dimensionality reduction has not been used in wireless communication systems. We will show how accurate channel charts can be generated using tools from manifold learning and dimensionality reduction.

C. Notation

Lowercase and uppercase boldface letters stand for column vectors and matrices, respectively. For the matrix $\mathbf{A}$, the Hermitian conjugate is $\mathbf{A}^H$, and the kth row and ℓth column entry is $A_{n,\ell}$ or $[\mathbf{A}]_{n,\ell}$. For the vector $\mathbf{a}$, the nth entry is a_k. The Frobenius norm of $\mathbf{A}$ is denoted by $\|\mathbf{A}\|_F$. The $M \times N$ all-zeros and all-ones matrix is $\mathbf{0}_{M\times N}$ and $\mathbf{1}_{M\times N}$, respectively, and the $M \times M$ identity is $\mathbf{I}_M$. The collection of N vectors $\mathbf{a}_n$, $n = 1, \ldots, N$, is denoted by $\{\mathbf{a}_n\}_{n=1}^N$.

II. Channel Charting Principles

The objective of CC is to learn a low-dimensional *channel chart* from a large amount of high-dimensional CSI (acquired at the infrastructure BS from users at different spatial locations) that *locally* preserves the original spatial geometry, i.e., users that are nearby in physical space will be nearby in the channel chart and vice versa. CC operates in an *unsupervised* fashion from CSI only, i.e., does not require application layer location information from GNSSs. Note that practical wireless systems require the extraction of CSI anyway for reliable data transmission. Hence, CSI is readily available at the BS. We now detail the principles underlying this approach.

A. Channel Function and Assumptions

Consider a single-antenna user that is either static or moves in real space. We denote its spatial locations at discrete time instants $n = 1, \ldots, N$ by the set $\{\mathbf{x}_n\}_{n=1}^N$ with $\mathbf{x}_n \in \mathbb{R}^D$, where D is the dimensionality of the *spatial geometry* (e.g., the user's x, y, and z coordinates). At each time instant n, the user transmits pilots or information symbols $\mathbf{s}_n$, which is received at a mMIMO BS with B antennas. The received data is modeled as $\mathbf{y}_n = H(\mathbf{s}_n) + \mathbf{n}_n$, where the function $H(\cdot)$ represents the wireless channel between the transmitting user and the receiving BS; the vector $\mathbf{n}_n$ models noise.

For CC, we are not interested in the transmitted data but rather in the associated CSI. Concretely, the Rx uses the received data $\mathbf{y}_n$ to extract CSI denoted by $\mathbf{h}_n \in \mathbb{C}^M$, where M denotes the dimensionality of the acquired CSI from all antennas, frequencies, and/or delays; typically, we have $M \gg D$. We denote the mapping from spatial location $\mathbf{x}_n$ to CSI $\mathbf{h}_n$ by the channel function $\mathcal{H} : \mathbb{R}^D \to \mathbb{C}^M$, where $\mathbb{C}^M$ refers to the *radio geometry*. In practice, the CSI represented by $\mathbf{h}_n$ depends on the user's spatial location $\mathbf{x}_n$, but also on (static or moving) objects within the cell, as well as on noise and interference. CSI captures the overall geometry of the cell, including all scatterers, reflectors, and diffractors in the built and natural environment. In what follows, we will make the following assumptions:

Assumptions 1. *We assume that the statistical properties of the multi-antenna channel vary slowly across space, on a length-scale related to the macroscopic distances between scatterers in the channel, not on the small fading length-scale of wavelengths. We also assume a static channel function $\mathcal{H}$.*

Large-scale channels effects are typically created by reflection, diffraction, and scattering of the environment, whereas small-scale effects are caused by multipath propagation [23]. Accordingly, this assumption is well supported by measurements. As we will demonstrate in Section V, CC performs exceptionally well under these assumptions.

B. Channel Charting

CC first distills CSI $\mathbf{h}_n$ into *channel features* $\mathbf{f}_n \in \mathbb{R}^{M'}$ with M' denoting the feature dimension; typically, we have $M' \gg D$. Feature extraction is denoted by the function $\mathcal{F} : \mathbb{C}^M \to \mathbb{C}^{M'}$, whose purpose is to extract large-scale fading properties from CSI. CC then proceeds by using the set of N collected features $\{\mathbf{f}_n\}_{n=1}^N$ to learn the so-called *forward charting function* in an unsupervised manner. The forward charting function to be learned is denoted by $\mathcal{C} : \mathbb{C}^{M'} \to \mathbb{R}^{D'}$ and maps each channel feature $\mathbf{f}_n$ to a point $\mathbf{z}_n \in \mathbb{R}^{D'}$ in the low-dimensional *channel chart*; typically, we have $D' \approx D$. The forward charting function $\mathcal{C}$ to be learned should preserve local geometry between neighboring data points, i.e., aims at implementing the condition:

$$\text{If } d_x(\mathbf{x}, \mathbf{x}') \text{ is small, then } d_z(\mathbf{z}, \mathbf{z}') \approx d_x(\mathbf{x}, \mathbf{x}'). \quad (1)$$

Here, $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^D$ are two distinct coordinates in spatial geometry, $\mathbf{z}, \mathbf{z}' \in \mathbb{R}^{D'}$ are two distinct points in the learned channel chart, and $d_x(\mathbf{x}, \mathbf{x}')$ and $d_z(\mathbf{z}, \mathbf{z}')$ are suitably defined distance (or dissimilarity) measures. The goal of CC is to compute a channel chart $\{\mathbf{z}_n\}_{n=1}^N$ satisfying (1), solely from the set of N channel features $\{\mathbf{f}_n\}_{n=1}^N$ in an unsupervised manner, i.e., without access to the spatial locations $\{\mathbf{x}_n\}_{n=1}^N$.

Figure 1 outlines the process of CC. The users are located in spatial geometry denoted by $\mathbb{R}^D$ which represents their coordinates. The wireless channel $\mathcal{H}$ maps transmit signals into CSI in radio geometry denoted by $\mathbb{C}^M$. This non-linear mapping obfuscates the spatial relationships between the users. The purpose of feature extraction is to find a representation from which spatial geometry is easily recovered. CC learns the forward charting function $\mathcal{C}$ that maps the channel features into low-dimensional points in the channel chart $\mathbb{R}^{D'}$ such that neighboring transmit locations (in real-world coordinates) will be neighboring points in the channel chart.

III. Channel Features

We start by the feature extraction stage. We show that lifting, scaling, and transforming the CSI in the angular domain yields

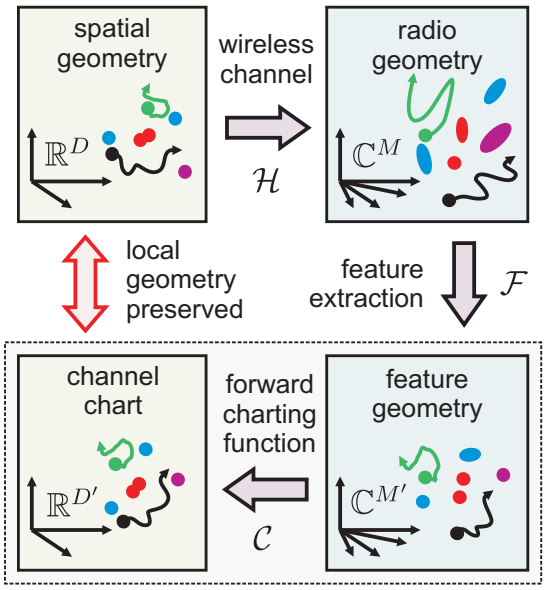


Fig. 1. Summary of CC. Users are located in spatial geometry $\mathbb{R}^D$ and a mMIMO BS extracts CSI in radio geometry $\mathbb{C}^M$. Feature extraction distills useful information into feature geometry $\mathbb{C}^{M'}$, which is used to learn the charting function $\mathcal{C}$ that maps features into a low-dimensional channel chart in $\mathbb{R}^{D'}$ that preserves the local geometry of the original spatial locations $\mathbb{R}^D$.

features that accurately represent large-scale fading properties of the channel. To limit our search for suitable channel features, we focus on Frobenius distance as dissimilarity measure on pairs of feature matrices, i.e., we use $d_f(\mathbf{F},\mathbf{F}') = \|\mathbf{F}-\mathbf{F}'\|_F$.

A. Lifted CSI Moments

CSI is affected by small-scale fading (due to multi-path propagation) and large-scale fading (due to shadowing, path loss, and other macroscopic effects). Small-scale fading effects change the phase of CSI at sub-wavelength scales and hence, are typically modeled statistically. To render our CSI features robust to such random phase shifts, we propose to compute features that are invariant to global phase shifts but contain information of the relative phase within the antenna array [24]. We compute the *raw 2nd moment* (R2M) of dimension M^2 given by $\mathbf{H} = \mathbb{E}\big[\mathbf{h}\mathbf{h}^H\big]$, where expectation is over noise and variations in CSI caused by small-scale motion during short time. In practice, we compute $\mathbf{H} = \frac{1}{T-1}\sum_{t=1}^{T}\mathbf{h}_t\mathbf{h}_t^H$, where T is the number of samples we are averaging over.

B. Adaptive Feature Scaling

One of the key aspects in the design of good channel features is to realize that CSI in radio geometry is a particularly poor representation of spatial geometry. To see this, assume that two users A and B are close to the BS, and two users C and D are further away. Due to path-loss, the CSI contained in $\mathbf{H}_\mathsf{C}$ and $\mathbf{H}_\mathsf{D}$ of users C and D appears to be weaker (i.e., has small Frobenius norm) than that of the users nearby, $\mathbf{H}_\mathsf{A}$ and $\mathbf{H}_\mathsf{B}$. If we now directly compare the Frobenius distance between C and D, their distance appears to be smaller than that between A and B because they have small overall norm, even though they should be further apart. To compensate for this "warping" phenomenon, we adaptively scale the CSI contained in the R2M so that it is more compatible with spatial geometry.

Consider a user that is separated d meters from a BS whose B antennas form a uniform linear array (ULA). Assume a narrowband line-of-sight (LoS) channel without scatterers and a 2-dimensional plane wave model (PWM), which accurately models transmitters in the far field. For this scenario, the entries h_b of the CSI vector $\mathbf{h} \in \mathbb{C}^B$ are given by [25]

$$h_b = d^{-\rho}\exp\big(-\tfrac{2\pi}{\lambda}\Delta r(b-1)\cos(\phi)\big), \tag{2}$$

for $b = 1,\dots,B$, where $\rho > 0$ is the path-loss exponent, Δr is the antenna spacing, and ϕ is the incident angle of the user to the BS. For this LoS scenario, we have the following result.

Lemma 1. *Let $\mathbf{H} = \mathbf{h}\mathbf{h}^H$ be the R2M. Assume two users A and C with the same incident angle ϕ, with distances d_A and d_C to the BS. By scaling the R2M as*

$$\tilde{\mathbf{H}} = \tfrac{B^{\beta-1}}{\|\mathbf{H}\|_F^{\beta}}\mathbf{H} \quad \textit{with} \quad \beta = 1 + 1/(2\rho), \tag{3}$$

the distance $d_h(\tilde{\mathbf{H}}_\mathsf{A},\tilde{\mathbf{H}}_\mathsf{C}) = \|\tilde{\mathbf{H}}_\mathsf{A} - \tilde{\mathbf{H}}_\mathsf{C}\|_F$ of the scaled moments $\tilde{\mathbf{H}}_\mathsf{A}$ and $\tilde{\mathbf{H}}_\mathsf{C}$ is their true distance, i.e.,

$$d_h\Big(\tilde{\mathbf{H}}_\mathsf{A},\tilde{\mathbf{H}}_\mathsf{C}\Big) = |d_\mathsf{A} - d_\mathsf{C}|. \tag{4}$$

Proof. The proof follows by inserting (3) and (2) into (4), and the fact that both users A and C are associated with the same channel vector $\mathbf{h}$ in (2) but with different path losses. □

We emphasize that adaptive scaling operation in (3) "unwraps" the radio geometry, i.e., CSI from transmitters far away is amplified and nearby CSI is attenuated. Since the path-loss exponent $\rho > 0$ is often unknown in practice, we can also use it as a tuning parameter. Note that for the special case of $\beta = 1$, the scaling in (3) simply normalizes the matrices to unit Frobenius norm, effectively discarding any path loss information. The resulting scaled CSI moments from (3) are then passed to the feature transform stage discussed next.

C. Transformation into Beamspace

We are now ready to transform the scaled R2M $\tilde{\mathbf{H}}$ into channel features. A straightforward choice would be to directly set the feature to the scaled CSI moments $\mathbf{F} = \tilde{\mathbf{H}}$; for simplicity, we now allow features to be matrices and denote this feature by "$\mathbb{C}\{\cdot\}$". It turns out, however, that applying certain well-designed nonlinear transforms to the scaled CSI moments significantly improves the feature quality. In particular, we transform $\tilde{\mathbf{H}}$ into $\mathbf{D}\tilde{\mathbf{H}}\mathbf{D}^H$ where $\mathbf{D}$ is the $M \times M$ discrete Fourier transform (DFT) matrix. This approach converts the scaled R2M from the antenna domain into the so-called beamspace domain, which represents the incident angles of the user and potential scatterers to the BS array in a concise way [26]. To see this, it is key to realize that the vector $\mathbf{h}$ for the LoS PWM model in (2) resembles a basis function of the DFT. As a final step, we take the entry-wise absolute value of the scaled and transformed R2M, i.e., our features are given by $\mathbf{F} = |\mathbf{D}\tilde{\mathbf{H}}\mathbf{D}^H|$. This last step mitigates noise that is caused by residual phase fluctuations.

IV. Channel Charting Algorithms

We now propose two dimensionality reduction algorithms that have been specialized for the purpose of channel charting. Corresponding simulation results are provided in Section V.

A. Sammon's Mapping

Sammon's mapping (SM) [20] is a classical nonlinear method that maps a high-dimensional point set into a point set of lower dimensionality with the goal of retaining small pairwise distances between both point sets as in (1). We next detail how SM can be used for CC, explain an efficient algorithm to compute the channel chart, and propose a specialized version that takes into account side information that is typically available in wireless communication systems.

1) SM Basics: First, we compute a pairwise distance matrix $\mathbf{D}$ between all channel features, i.e.,

$$D_{n,\ell} = d_f(\mathbf{F}_n, \mathbf{F}_\ell),\ n = 1, \ldots, N,\ \ell = 1, \ldots, N,$$

using the Frobenius distance. SM tries to find a low-dimensional channel chart, i.e., a point set $\{\mathbf{z}_n\}_{n=1}^N$, that results from the following optimization problem:

$$\text{(SM)} \begin{cases} \underset{\substack{\mathbf{z}_n \in \mathbb{R}^{D'} \\ n=1,\ldots,N}}{\text{minimize}} \displaystyle\sum_{\substack{n=2,\ldots,N \\ \ell=1,\ldots,n-1}} D_{n,\ell}^{-1}(D_{n,\ell} - \|\mathbf{z}_n - \mathbf{z}_\ell\|_2)^2 \\ \text{subject to} \displaystyle\sum_{n=1,\ldots,N} \mathbf{z}_n = \mathbf{0}_{D' \times 1}, \end{cases}$$

where we omit pairs of points for which $D_{n,\ell} = 0$. The objective function of SM promotes channel charts for which the Euclidean distance of pairs of nearby points in $\mathbb{R}^{D'}$ agrees with the feature distance. Points for which $D_{n,\ell}^{-1}$ is small (i.e., points that are dissimilar in feature geometry) are discounted; this ensures that SM ignores relationships of points that are far apart in feature geometry. Since the objective function is invariant to global translations, we enforce the channel chart to be centered in each of the coordinates in $\mathbb{R}^{D'}$.

2) Forward-Backward Splitting for SM: While the problem (SM) is non-convex, we next detail an efficient first-order method that enables us to include side information that is available for CC; see Section IV-B. Concretely, we use an accelerated forward-backward splitting (FBS) procedure [27], [28] that solves a class of convex optimization problems of the following general form: $\text{minimize}\ f(\mathbf{Z}) + g(\mathbf{Z})$, where the function $f(\mathbf{Z}) = \sum_{n=1}^N f_n(\mathbf{z}_n)$ should be convex and smooth and g should be convex, but does not need to be smooth or bounded. FBS consists of the iteration

$$\mathbf{Z}^{(t+1)} = \text{prox}_g\big(\mathbf{Z}^{(t)} - \tau^{(t)} \nabla f(\mathbf{Z}^{(k)}), \tau^{(t)}\big)$$

for $t = 1, \ldots, T_{\text{max}}$. Here, $\nabla f(\mathbf{Z})$ is the gradient of the smooth function f, and the proximal operator for the nonsmooth function g is [29]

$$\text{prox}_g(\mathbf{Z}, \tau) = \arg\min_{\mathbf{V}} \big\{\tau g(\mathbf{V}) + \tfrac{1}{2}\|\mathbf{V} - \mathbf{Z}\|_F^2\big\}.$$

The sequence $\{\tau^{(t)} > 0\}$ contains carefully selected step-size parameters that ensure convergence of FBS.

For CC, the matrix $\mathbf{Z} = [\mathbf{z}_1, \ldots, \mathbf{z}_N]$ contains all points in the channel chart. The function f is chosen to be

$$f(\mathbf{Z}) = \sum_{\substack{n=2,\ldots,N \\ \ell=1,\ldots,n-1}} D_{n,\ell}^{-1}(D_{n,\ell} - \|\mathbf{z}_n - \mathbf{z}_\ell\|_2)^2, \tag{5}$$

and the nth column of the gradient of f is

$$[\nabla f(\mathbf{Z})]_n = 2 \sum_{\substack{\ell=1,\ldots,n \\ \ell \neq n}} D_{n,\ell}^{-1}(D_{n,\ell} - \|\mathbf{z}_n - \mathbf{z}_\ell\|_2) \frac{\mathbf{z}_n - \mathbf{z}_\ell}{\|\mathbf{z}_n - \mathbf{z}_\ell\|_2}.$$

The centering constraint of (SM) is enforced by the function $g(\mathbf{Z}) = \chi(\sum_{n=1}^N \mathbf{z}_n)$ which is zero when its argument $\sum_{n=1}^N \mathbf{z}_n$ is zero and infinity otherwise. The proximal operator for this function is a re-projection onto the centering constraint:

$$\text{prox}_g(\mathbf{Z}, \tau) = \mathbf{Z} - \frac{1}{N}\mathbf{Z}\mathbf{1}_{N\times 1}\mathbf{1}_{N\times 1}^T.$$

Note that since f is nonconvex, FBS is not guaranteed to find a global minimizer. Nevertheless, as shown in Section V, FBS initialized with a solution from principal component analysis (PCA) combined with the adaptive step-size procedure proposed in [28] yields high-quality channel charts.

B. Sammon's Mapping with Side-Information

In practice, one often collects a large number of CSI vectors from a single user over time. For such situations, the channel features for a given user u form a time series $\{\mathbf{f}_n\}_{n\in\mathcal{N}_u}$, where $\mathcal{N}_u$ contains the temporally ordered indices associated with user u. Since users move with finite velocity, we know that temporally adjacent CSI vectors from the same user must be close in the channel chart. To exploit such side information, we include an elastic-net penalty in the objective function that keeps temporally adjacent points in $\mathcal{N}_u$ nearby in the channel chart. Concretely, for each user u, we add

$$f_u(\mathbf{Z}) = \alpha_u \textstyle\sum_{n\in\mathcal{N}_u} \|\mathbf{z}_n - \mathbf{z}_{n+1}\|_2^2$$

to the objective of (SM), where the parameter $\alpha_u > 0$ determines the spatial smoothness of user u in the channel chart. In what follows, we refer to this CC algorithm as Sammon's mapping *plus* (SM+).

V. Results

We now demonstrate the efficacy of CC with our channel features and charting algorithms for realistic channel models.

A. Quality Measures of Channel Charts

To characterize the quality of the generated channel charts, we need a performance metric of how well points in the channel chart preserve the spatial geometry of the true user locations. To this end, we borrow two metrics commonly used to measure the quality of mappings in dimensionality reduction tasks, namely *continuity* (CT) and *trustworthiness* (TW) [22], [30], [31]. Both of these quality measures can be described in the context of two abstract sets of data points with cardinality N, i.e., $\{\mathbf{u}_n\}_{n=1}^N$ from an *original space* and $\{\mathbf{v}_n\}_{n=1}^N$ from a *representation* of the original space; the

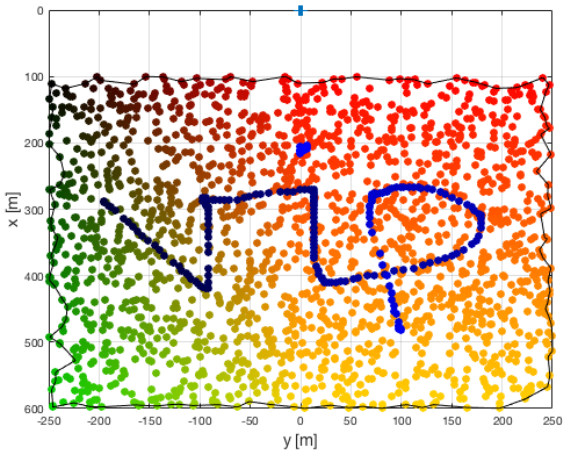


Fig. 2. Top view of the considered channel scenario: a massive MIMO BS is equipped with a $B = 32$ antenna uniform linear array at location $(x, y, z) = (0, 0, 10)$ meters. The BS acquires CSI from 2048 randomly placed points in space; the points forming the "VIP" curve have been added to simplify visualization of the channel charts in Figure 3.

point $\mathbf{v}_n$ is said to represent $\mathbf{u}_n$. In the CC context, the original space would be spatial geometry and the representation space is the channel chart (cf. Figure 1). We define the *K-neighborhood* of a point $\mathbf{u}$ as the set containing its K nearest neighbors in terms of the chosen distance $d_u(\mathbf{u}, \mathbf{u}')$; the neighborhood of $\mathbf{v}$ is defined analogously with $d_v(\mathbf{u}, \mathbf{u}')$.

1) Continuity (CT): Neighbors in spatial geometry (original space) can be far away (or dissimilar) in the channel chart (representation space). For such situations, the representation space does not preserve the *continuity* of the original point set. To measure such situations, we first define the *point-wise continuity* for K neighbors of the data point $\mathbf{u}_i$. Let $\mathcal{V}_K(\mathbf{u}_i)$ be the K-neighborhood of point $\mathbf{u}_i$ in the original space. Also, let $\hat{r}(i, j)$ be the ranking (in terms of dissimilarity) of point $\mathbf{v}_j$ among the neighbors of point $\mathbf{v}_i$. Then, the point-wise continuity of the representation $\mathbf{v}_i$ of the point $\mathbf{u}_i$ is

$$\mathrm{CT}_i(K) = 1 - \frac{2}{K(2N - 3K - 1)} \sum_{j \in \mathcal{V}_K(\mathbf{u}_i)} (\hat{r}(i, j) - K).$$

The (global) *continuity* between the sets $\{\mathbf{u}_n\}_{n=1}^N$ and $\{\mathbf{v}_n\}_{n=1}^N$ is the average over the point-wise continuities. Continuity values close to zero indicate that points similar in spatial geometry are dissimilar in the channel chart; continuity values close to one indicate neighbor-preserving mappings.

2) Trustworthiness (TW): In dimensionality reduction, it may happen that the representation mapping introduces *new* neighbor relations that were absent in the original space. Trustworthiness measures how well the feature mapping avoids introducing such false relationships. We first define the point-wise *trustworthiness* for a K-neighborhood of $\mathbf{v}_i$. Let $\mathcal{U}_K(\mathbf{v}_i)$ be the set of "false neighbors" that are in the K-neighborhood of $\mathbf{v}_i$, but not of $\mathbf{u}_i$ in the original space. Also, let $r(i, j)$ be the ranking (in terms of dissimilarity) of point $\mathbf{u}_j$ in the neighborhood of point $\mathbf{u}_i$. The point-wise trustworthiness of the representation at $\mathbf{u}_i$ is then

$$\mathrm{TW}_i(K) = 1 - \frac{2}{K(2N - 3K - 1)} \sum_{j \in \mathcal{U}_K(\mathbf{v}_i)} (r(i, j) - K).$$

The (global) *trustworthiness* is simply the average over all the point-wise trustworthiness values. Trustworthiness values close to zero indicate situations in which most data points that appear similar in the channel chart are actually dissimilar in spatial geometry; trustworthiness values close to one indicate that data points nearby in the channel chart are also nearby in spatial geometry,

Remark 1. *We set K to 5% of the total number of points N, i.e., $K = 0.05N$, as we are interested in preserving* local *geometry; this is a common choice in the literature [30].*

B. Simulation Settings

Due to space constraints, we focus on a single scenario as depicted in Figure 2 with a narrowband non-LoS channel generated from the state-of-the-art Quadriga channel model [32] with the following parameters. We simulate the Berlin NLoS scenario (UMa) at a carrier frequency of 2.0 GHz with a bandwidth of 312.5 KHz. The mMIMO BS is located at coordinate $(x, y, z) = (0, 0, 10)$ meters and consists of a $B = 32$ antenna array arranged as a ULA with $\Delta r = \lambda/2$ antenna spacing. We record CSI of $N = 2048$ randomly selected spatial locations (with the exception of the "VIP" curve) within an area of 500 m $\times$ 500 m; the median distance between nearest neighbors is approximately 7.86 meters, i.e., we sample the space at roughly 53 wavelengths. We acquire CSI at an SNR of 0 dB, i.e., consider channel charting under adversarial conditions, use $T = 10$, and set $\rho = 16$.

C. Feature Comparison

As briefly mentioned in Section III, applying certain nonlinear transforms to the scaled CSI moments $\tilde{\mathbf{H}}$ can significantly improve the feature quality. To identify suitable candidate features, we also considered taking the entry-wise real part (denoted by "$\Re\{\cdot\}$"), imaginary part (denoted by "$\Im\{\cdot\}$"), angle (denoted by "$\angle(\cdot)$"), or absolute value (denoted by "$|\cdot|$") of the scaled CSI moments. We furthermore say that all these channel features were taken in the *antenna domain*. We also consider the case in which we take the scaled CSI vectors and transform then into the *beamspace domain* followed by applying one of the above nonlinearities.

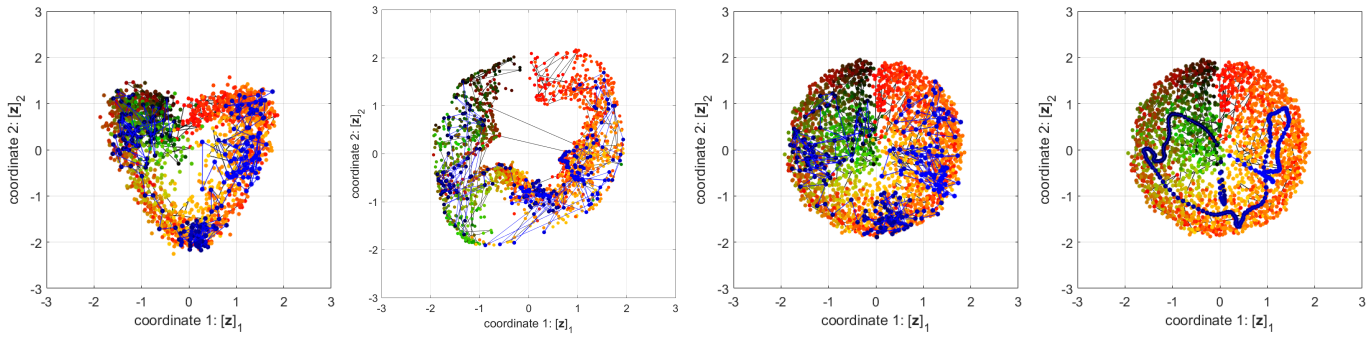
Table I compares the TW and CT measured between the true locations and between the channel features for all the combinations of features for the scenario depicted in Figure 2. This comparison confirms that taking the absolute value in the beamspace domain, as detailed in Section III, significantly outperforms all other methods in terms of TW and CT. Hence, in the remainder of the paper, we only consider on the absolute value in the beamspace domain, denoted by "beamspace, $|\cdot|$".

D. Channel Charting Results

In what follows, CT and TW is measured between the true spatial locations and the associated points in the channel chart.

TABLE I
COMPARISON OF R2M CHANNEL FEATURES IN TERMS OF TRUSTWORTHINESS (TW) AND CONTINUITY (CT) FOR $K = 0.05N$.

Domain		$\mathbb{C}\{\cdot\}$	$\Re\{\cdot\}$	$\Im\{\cdot\}$	$\angle(\cdot)$	$\lvert\cdot\rvert$
Antenna	TW	0.76 ($\pm$0.11)	0.62 ($\pm$0.12)	0.70 ($\pm$0.09)	0.67 ($\pm$0.09)	0.54 ($\pm$0.07)
	CT	0.76 ($\pm$0.07)	0.71 ($\pm$0.07)	0.69 ($\pm$0.08)	0.63 ($\pm$0.08)	0.56 ($\pm$0.09)
Beamspace	TW	see TW above	0.76 ($\pm$0.12)	0.56 ($\pm$0.08)	0.55 ($\pm$0.07)	**0.81** (**$\pm$0.13**)
	CT	see CT above	0.74 ($\pm$0.07)	0.52 ($\pm$0.06)	0.53 ($\pm$0.09)	**0.84** (**$\pm$0.09**)



(a) PCA (0.92, 0.85) (b) AE (0.91, 0.86) (c) SM (0.93, 0.85) (d) SM+ (0.93, 0.85)

Fig. 3. $D' = 2$ dimensional channel charts for various CC algorithms under the Quadriga non-LoS channel model for the scenario depicted in Figure 2. We compare PCA, autoencoder (AE), Sammon's mapping (SM), and Sammon's mapping with temporal continuity (SM+). The values in the brackets denote continuity (CT) and thrustworthiness (TW). We see that AE, SM, and SM+ achieve the highest CT and TW, whereas SM+ also delivers visually pleasing results.

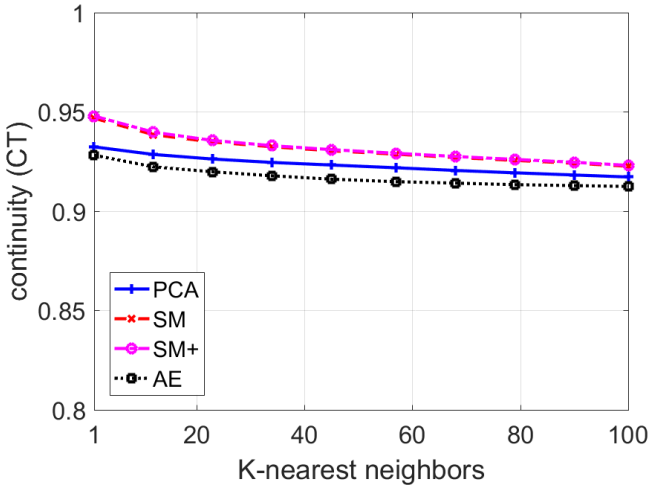
1) Channel Charts: We learn channel charts for SM and SM+ as detailed in Section IV, as well as PCA and a deep autoencoder (AE), two commonly used methods in the dimensionality reduction literature [21]. Figure 3 shows the learned channel charts. For all algorithms, we obtain CT values ranging between 0.91 and 0.93. This implies that the neighborhood of a point in spatial geometry is well-preserved in the learned channel charts. The TW values range between 0.85 and 0.86; this indicates that most neighbors of a point in the channel charts are also neighbors in spatial geometry. We can also visually inspect the quality of the obtained results by comparing (i) the color gradient in Figure 3 with that of the scenario in Figure 2 or (ii) the "VIP" curve in spatial geometry and in the channel chart.

Figure 3(a) shows that PCA yields surprisingly high CT and TW values, and provides a visually accurate embedding of spatial geometry. We address this behavior to the fact that we use carefully engineered channel features that well-represent spatial geometry. Figure 3(b) shows that the AE yields even higher CT and TW values, comparable to those of SM/SM+. The AE channel charts are less visually pleasing than, for example, those of SM+. Figure 3(c) shows that SM yields high CT and TW and provides excellent preservation of the color gradients. Figure 3(c) shows that SM+ is able to exploit temporal side information, while the CT and TW is comparable to AE and SE. Nevertheless, SM+ provides extremely well-preserved embeddings of the channel geometry. In fact, one can even identify the "VIP" curve in the learned channel chart.

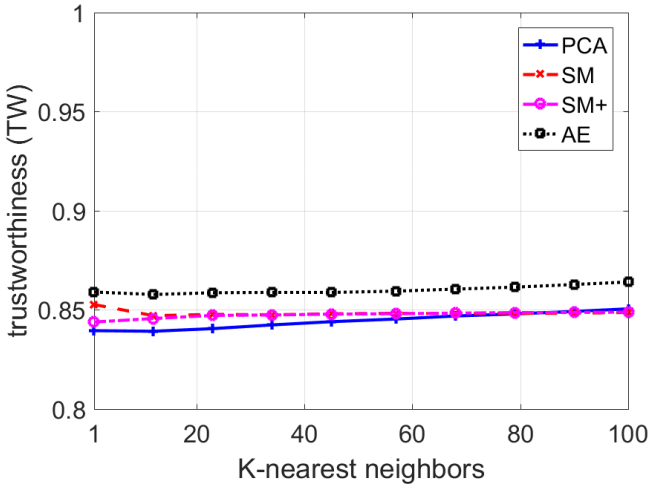
2) CT and TW Measures: To gain additional insight into the quality of the learned channel charts, Figure 4 shows the CT and TW values for different neighborhood sizes, i.e., K ranges from 1 to 100. We see that, for the challenging Quadriga non-LoS channel, which models complex scattering and multipath behavior, SM and SM+ perform best, followed by PCA. Evidently, the AE struggles in achieving high CT but has the advantage over SM and SM+ of providing a parametric mapping, i.e., given a new CSI vector, we can use the existing AE to directly compute the location in the channel chart.

VI. CONCLUSIONS

We have proposed a novel application of dimensionality reduction to multi-antenna wireless systems. More specifically, we have developed *channel charting* (CC), an unsupervised framework to learn a map between channel-state information (CSI) acquired at a single base-station (BS) and the relative user locations. Our method relies on the extraction of carefully designed channel features from large amounts of high-dimensional CSI acquired at a massive MIMO BS, followed by CC algorithms that borrow ideas from dimensionality reduction and manifold learning. We have developed new CC algorithms with varying complexity, flexibility, and accuracy that produce charts that preserve the local geometry of the transmitter locations for realistic channel models. Since channel charting is unsupervised, i.e., does not require knowledge of the true user locations, the proposed framework finds use in numerous applications relevant to 5G wireless networks, such as rate adaptation, network planning, user scheduling, hand-over, cell search, user tracking, beam finding in millimeter-wave systems, and other cognitive tasks that rely on CSI and the user movement relative to the BS.



(a) Continuity (CT)



(b) Trustworthiness (TW)

Fig. 4. Comparison of continuity (CT) and trustworthiness (TW) for various CC algorithms under the Quadriga non-LoS channel model. We observe that Sammon's mapping (SM) and its extension (SM+) outperform PCA and autoencoders (AEs) in terms of CT, while the AE slightly outperforms the other methods in terms of the TW. PCA yields surprisingly good results and performs close to that of SM and SM+ in terms of CT and TW.

REFERENCES

[1] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. Soong, and J. Zhang, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

[2] A. Osseiran, J. Monserrat, and P. Marsch, eds., *5G Mobile and Wireless Communications Technology*. Cambridge Univ. Press, 2016.

[3] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[4] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[5] E. Larsson, O. Edfors, F. Tufvesson, and T. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[6] T. J. O'Shea and J. Hoydis, "An introduction to machine learning communications systems," *arXiv preprint: 1702.00832*, July 2017.

[7] F. Gustafsson and F. Gunnarsson, "Mobile positioning using wireless networks," *IEEE Signal Process. Mag.*, pp. 41–53, July 2005.

[8] S. Kumar, S. Gil, D. Katabi, and D. Rus, "Accurate indoor localization with zero start-up cost," in *Proc. 20th Annu. Int. Conf. Mobile Comput.*, Sep. 2014, pp. 483–494.

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

[10] M. Ibrahim and M. Youssef, "Cellsense: An accurate energy-efficient GSM positioning system," *IEEE Trans. Veh. Technol.*, vol. 61, no. 1, pp. 286–296, Jan. 2012.

[11] A. Prasad, O. Tirkkonen, P. Lunden, O. N. C. Yilmaz, L. Dalsgaard, and C. Wijting, "Energy-efficient inter-frequency small cell discovery techniques for LTE-advanced heterogeneous network deployments," *IEEE Commun. Mag.*, vol. 51, no. 5, pp. 72–81, May 2013.

[12] Y. Chapre, A. Ignjatovic, A. Seneviratne, and S. Jha, "CSI-MIMO: An efficient Wi-Fi fingerprinting using channel state information with MIMO," *Pervasive and Mobile Comput.*, vol. 23, pp. 89–103, Oct. 2015.

[13] X. Wang, L. Gao, S. Mao, and S. Pandey, "DeepFi: Deep learning for indoor fingerprinting using channel state information," in *IEEE Wireless Commun. Netw. Conf. (WCNC)*, Mar. 2015, pp. 1666–1671.

[14] L. Gao, "Channel state information fingerprinting based indoor localization: a deep learning approach," Auburn University M.S. Thesis, Aug. 2015.

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

[16] V. Savic and E. G. Larsson, "Fingerprinting-based positioning in distributed massive MIMO systems," in *IEEE Veh. Technol. Conf.*, Sep. 2015, pp. 1–5.

[17] J. Vieira, E. Leitinger, M. Sarajlic, X. Li, and F. Tufvesson, "Deep convolutional neural networks for massive MIMO fingerprint-based positioning," in *IEEE 28th Annu. Int. Symp. Personal, Indoor, and Mobile Radio Commun. (PIMRC)*, Oct. 2017.

[18] E. Dahlman, S. Parkvall, and J. Sköld, *4G: LTE/LTE-Advanced for Mobile Broadband*. Elsevier, Mar. 2011.

[19] J. Kruskal, "Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis," *Psychometrika*, vol. 29, no. 1, pp. 1–27, Mar. 1964.

[20] J. W. Sammon, "A nonlinear mapping for data structure analysis," *IEEE Trans. Comput.*, vol. 100, no. 5, pp. 401–409, May 1969.

[21] L. Van Der Maaten, E. Postma, and J. Van den Herik, "Dimensionality reduction: a comparative review," Tilburg University, TiCC-TR 2009-005, 2009.

[22] S. Kaski, J. Nikkilä, M. Oja, J. Venna, P. Törönen, and E. Castrén, "Trustworthiness and metrics in visualizing similarity of gene expression," *BMC Bioinformatics*, vol. 4, no. 1, p. 48, Oct. 2003.

[23] T. S. Rappaport, *Wireless Communications: Principles and Practice*. Prentice Hall, 1996.

[24] D.-S. Shiu, G. J. Foschini, M. J. Gans, and J. M. Kahn, "Fading correlation and its effect on the capacity of multielement antenna systems," *IEEE Trans. Commun.*, vol. 48, no. 3, pp. 502–513, Mar. 2000.

[25] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. Cambridge Univ. Press, 2005.

[26] J. Brady, N. Behdad, and A. M. Sayeed, "Beamspace MIMO for millimeter-wave communications: System architecture, modeling, analysis, and measurements," *IEEE Trans. Antennas Propag.*, vol. 61, no. 7, pp. 3814–3827, July 2013.

[27] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, Jan. 2009.

[28] T. Goldstein, C. Studer, and R. G. Baraniuk, "A field guide to forward-backward splitting with a FASTA implementation," *arXiv preprint: 1411.3406*, Nov. 2014. [Online]. Available: http://arxiv.org/abs/1411.3406

[29] N. Parikh and S. Boyd, "Proximal algorithms," *Foundations and Trends Optimization*, vol. 1, no. 3, pp. 123–231, Jan. 2013.

[30] J. Venna and S. Kaski, "Neighborhood preservation in nonlinear projection methods: An experimental study," in *Int. Conf. on Artificial Neural Networks*. Springer, 2001, pp. 485–491.

[31] Á. Vathy-Fogarassy and J. Abonyi, *Graph-based clustering and data visualization algorithms*. Springer, 2013.

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