



**HAL**  
open science

## CSI Compression using Channel Charting

Baptiste Chatelier, Vincent Corlay, Matthieu Crussière, Luc Le Magoarou

► **To cite this version:**

Baptiste Chatelier, Vincent Corlay, Matthieu Crussière, Luc Le Magoarou. CSI Compression using Channel Charting. 2024 58th Asilomar Conference on Signals, Systems, and Computers, Oct 2024, Pacific Grove, United States. <10.1109/IEEECONF60004.2024.10942729>. <hal-04835586>

**HAL Id: hal-04835586**

**<https://hal.science/hal-04835586v1>**

Submitted on 13 Dec 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

# CSI Compression using Channel Charting

Baptiste Chatelier<sup>†,‡,\*</sup>, Vincent Corlay<sup>†,‡,\*</sup>, Matthieu Crussière<sup>†,‡,\*</sup>, Luc Le Magoarou<sup>†,\*</sup>

<sup>†</sup>Univ Rennes, INSA Rennes, CNRS, IETR-UMR 6164, Rennes, France

<sup>‡</sup>Mitsubishi Electric R&D Centre Europe, Rennes, France

<sup>\*</sup>b<>com, Rennes, France

**Abstract**—Reaping the benefits of multi-antenna communication systems in frequency division duplex (FDD) requires channel state information (CSI) reporting from mobile users to the base station (BS). Over the last decades, the amount of CSI to be collected has become very challenging owing to the dramatic increase of the number of antennas at BSs. To mitigate the overhead associated with CSI reporting, compressed CSI techniques have been proposed with the idea of recovering the original CSI at the BS from its compressed version sent by the mobile users. Channel charting is an unsupervised dimensionality reduction method that consists in building a radio-environment map from CSIs. Such a method can be considered in the context of the CSI compression problem, since a chart location is, by definition, a low-dimensional representation of the CSI. In this paper, the performance of channel charting for a task-based CSI compression application is studied. A comparison of the proposed method against baselines on realistic synthetic data is proposed, showing promising results.

**Index Terms**—Channel charting, Dimensionality reduction, Machine learning, CSI compression.

## I. INTRODUCTION

THE shift towards higher operating frequencies in the Fifth Generation (5G) and Beyond 5G (B5G) communication systems has led to the widespread use of antenna arrays made up of a large number of smaller radiating elements, capable of forming high-gain directional beams. However, this increase in the BSs antenna number causes a significant overhead in the Channel State Information (CSI) reporting procedures in Frequency Division Duplex (FDD) schemes. Indeed, in FDD systems, in order to perform downlink channel estimation, User Equipments (UEs) need to periodically report one complex number per antenna and subcarrier to the BS. To mitigate this reporting overhead, CSI compression techniques have been envisioned [1]. The associated reporting procedure can be summarized as follows: in a FDD system, a BS sends pilot symbols to a UE that uses them to perform CSI estimation. The UE then compresses the estimated CSI and transmits it to the BS. Leveraging its readily available computing power, the BS is then able to reconstruct the CSI with minimal distortion, before using it to perform communication tasks such as beamforming or resource allocation. This problem has been extensively studied in the past few years, with Artificial Intelligence/Machine Learning (AI/ML) methods proposed to address it. A non-exhaustive list of relevant studies includes [1]–[6]. An in-depth review of the deep learning based CSI compression literature is provided in [7].

This paper considers the task-based CSI compression problem from a different perspective, using channel charting (CC). Since the reconstructed CSI is primarily used for beamforming, one could consider the CSI compression problem where the decoding step reconstructs a precoding matrix rather than the CSI. Channel charting [8] is an unsupervised dimensionality reduction methods

that compress CSI into chart locations. More precisely, the goal of CC is to embed high dimensional CSIs into a low dimensional space that preserves local neighborhoods: channels from close locations should be embedded into chart locations that are close one to each-other. CC has been used for beam prediction [9], [10], precoder learning [11], but also for positioning [12], [13]. In the context of CSI compression, it clearly appears that CC can be seen as a good candidate for the encoding stage, as, by definition, a chart location is a compressed version of its associated CSI. **Contributions.** This paper considers the concept of task-based CSI compression: its goal is to reduce the CSI reporting overhead while optimizing beamforming performance. The encoder consists of a CC model, while the decoder is a neural network. An end-to-end learning strategy for the encoder and decoder training is presented, with a task-oriented loss function. Furthermore, the number of learnable parameters is optimized through a subsampling approach within the encoder, resulting in a ninetyfold reduction in the number of learnable parameters, without significant performance degradation. Finally, simulations on realistic data show that the proposed approach achieves excellent beamforming performance with significantly higher compression ratios (around 1024) compared to those obtained by classical auto-encoders (around 64).

**Related work.** Inferring a precoder from a chart location has been studied for Cell Free massive MIMO systems in [14]. However, that study did not optimize the encoder complexity and only trained the decoder network. Similar studies have been carried out with precoders constrained in a codebook in [9], [10]. Jointly optimizing CSI compression and beamforming performance has been studied in [15]–[17]. However, none of these studies considered channel charting for the encoding stage. Finally, CSI compression using AI/ML methods have been extensively studied in [1]–[6]. Such approaches obtain good performance in CSI reconstruction but often present poor compression ratio compared to our approach as they are not specialized to a given task.

## II. PROBLEM FORMULATION

This paper considers a multi-user Multiple Input Multiple Output (MU-MIMO) system comprised of a BS equipped with  $N_a$  antennas and  $N_u$  mono-antenna UEs. The BS transmits symbols to the UEs on  $N_s$  subcarriers.

**Reporting procedure.** A BS sends pilot symbols to a set of UEs. The UEs use the pilots to estimate their downlink channel  $\mathbf{h}_k \in \mathbb{C}^D$ . Note that vectorized channels are considered, so that  $D = N_a N_s$ . Then, UEs perform CSI compression: the downlink channel is encoded into a low dimensional vector  $\mathbf{z}_k \in \mathbb{R}^d$ . Finally, the BS receives the encoded channels from all UEs and decode them into a precoding matrix  $\mathbf{W}$ , used to perform spatial multiplexing.

**Objective.** The goal of this paper is to design an efficient CSI compression method that solves the following sum-rate (SR) optimization problem in a parallel transmission to  $K \leq N_u$  UEs:

$$\begin{aligned} \mathbf{W} \in \mathbb{C}^{D \times K} \quad & \max \sum_{k=1}^K \log_2 \left( 1 + \frac{|\mathbf{h}_k^H \mathbf{w}_k|^2}{\sigma_k^2 + \sum_{j \neq k} |\mathbf{h}_k^H \mathbf{w}_j|^2} \right), \\ \text{s.t.} \quad & \mathcal{C}_p : \forall k \in [1, K], \|\mathbf{w}_k\|_2 = \frac{1}{\sqrt{K}} \end{aligned} \quad (1)$$

where  $\mathbf{w}_k$  is a precoding vector decoded from a compressed representation of the channel  $\mathbf{h}_k$  and  $\sigma_k^2$  is the associated noise variance. The power constraint  $\mathcal{C}_p$  considers uniform power allocation between each UE.

The foundation of the proposed method is formally based on defining a well suited encoding function  $\mathcal{E}$  and decoding function  $\mathcal{D}$ . The role of the encoding function is to project the CSI into a low-dimensional space:

$$\begin{aligned} \mathcal{E} : \mathbb{C}^D &\longrightarrow \mathbb{R}^d \\ \mathbf{h} &\longrightarrow \mathbf{z} \triangleq \mathcal{E}(\mathbf{h}), \end{aligned} \quad (2)$$

where  $d \ll D$ . Conversely, the decoding function maps the low-dimensional channel charts to a precoding matrix:

$$\begin{aligned} \mathcal{D} : \mathbb{R}^d \times \dots \times \mathbb{R}^d &\longrightarrow \mathbb{C}^{D \times K} \\ \{\mathbf{z}_k\}_{k=1}^K &\longrightarrow \mathbf{W} \triangleq \mathcal{D}(\{\mathbf{z}_k\}_{k=1}^K). \end{aligned} \quad (3)$$

**Performance measure.** The performance of the proposed method is measured in different configurations: mono and multi-UE scenarios. In a mono-UE scenario, it has been shown in [11] that the transmission capacity is linked to the squared cosine similarity between the channel  $\mathbf{h}_k$  and the associated inferred precoder  $\mathbf{v}_k$ :

$$\rho_k = \frac{|\mathbf{v}_k^H \mathbf{h}_k|^2}{\|\mathbf{v}_k\|_2^2 \|\mathbf{h}_k\|_2^2}. \quad (4)$$

This metric is normalized within the  $[0, 1]$  range. 0 means that the precoder is orthogonal to the channel, leading to no transmission, while 1 means that the precoder is perfectly collinear with the channel. The mean and median correlations over the UE set will be considered.

In a multi-UE scenario, the performance measure is the sum-rate averaged across distinct groups of UEs in the location space. Considering the partition of the UE set  $\mathcal{T}_u$  into  $B$  groups of  $K$  UEs:  $\mathcal{T}_u = \bigcup_{b=1}^B \mathcal{T}_{u,b}$  with  $|\mathcal{T}_{u,b}| = K$  and  $\forall m \neq n, \mathcal{T}_{u,m} \cap \mathcal{T}_{u,n} = \emptyset$ , the sum-rate can be defined as:

$$\mathring{\mathcal{R}} = \frac{1}{B} \sum_{b=1}^B \sum_k \log_2 \left( 1 + \frac{|\mathbf{h}_{k,b}^H \mathbf{w}_{k,b}|^2}{\sigma_{k,b}^2 + \sum_{j \neq k} |\mathbf{h}_{k,b}^H \mathbf{w}_{j,b}|^2} \right), \quad (5)$$

where  $\mathbf{h}_{i,b} \in \mathcal{T}_{u,b}$  if the channel for the  $i$ th UE of the  $b$ th group and  $\mathbf{w}_{i,b}$  is the associated decoded precoder.

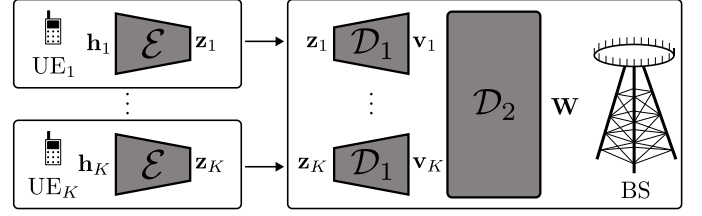


Fig. 1: Overview of the proposed approach: encoder (left), decoder (right)

### III. PROPOSED APPROACH

This section presents the chosen models for the encoding and decoding functions, as well as the learning strategy and the learnable parameter reduction method. The CC-based encoder and decoder models are similar to those proposed in [10], [14], [18]. An overview of the proposed approach is presented in Fig. 1.

**Encoder.** As in [10], [14], [18], the encoding function is done through CC with a neural model presented in Fig. 2. This model is initialized through matrices  $\mathbf{D}$  and  $\mathbf{Z}$ . Matrix  $\mathbf{D} = \{\mathbf{h}_i\} \in \mathbb{C}^{D \times N_c}$  consists of the concatenation of  $N_c$  calibration channel vectors, while  $\mathbf{Z}$  is initialized through the ISOMAP algorithm. This algorithm embeds the calibration channels into chart locations  $\mathbf{z}_k \in \mathbb{R}^d$  ( $d \ll D$ ). Note that the phase-insensitive distance of [19] is used in the ISOMAP algorithm. The chart locations are then concatenated into matrix  $\mathbf{Z} = \{\mathbf{z}_i\}_{i=1}^{N_c} \in \mathbb{R}^{d \times N_c}$ . As many other dimensionality reduction methods, ISOMAP is known to be computationally intensive in out-of-sample scenarios, which reduces the efficiency of the encoding stage for channels not included in the calibration set. The presented CC model aims to address this issue by mapping an unseen channel  $\mathbf{h}_j$  into a wisely chosen convex combination of calibration chart locations.

**Decoder.** It is proposed to split the decoder into two parts  $\mathcal{D}_1$  and  $\mathcal{D}_2$ , so that  $\mathcal{D} = \mathcal{D}_2 \circ \mathcal{D}_1$ , as presented in Fig. 2. The first stage,  $\mathcal{D}_1$ , allows to map the chart locations to mono-UE precoding vectors. Then, the second stage,  $\mathcal{D}_2$ , is designed to concatenate and process the mono-UE precoding vectors into a multi-UE precoding matrix. The mapping of chart locations to precoding vectors is done through a neural network. Note that such mapping can present relatively fast variations in the embedding space. Indeed, the optimal precoders may vary rapidly in some regions of the location space. As channels originating from nearby locations are expected to be embedded to close chart locations through CC, this fast variation behavior is expected to be translated in the embedding space. As a result, it is proposed to use a random Fourier features (RFF) [20] network for the chart-location-to-precoder mapping learning. Indeed, classical neural architectures such as MLPs have been shown to be biased towards learning low frequency

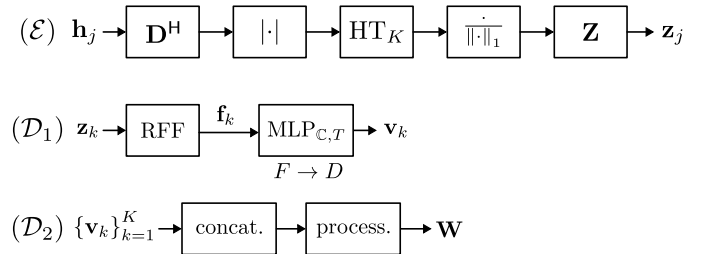


Fig. 2: Encoder [14], [18] and decoder architectures

functions [21]–[23]: RFF networks have been proposed as a solution to this spectral bias issue [24]. The RFF block is defined as:

$$\mathbf{f}_k = \begin{bmatrix} \cos(2\pi \mathbf{B} \mathbf{z}_k) \\ \sin(2\pi \mathbf{B} \mathbf{z}_k) \end{bmatrix}, \quad (6)$$

where  $\mathbf{B} \in \mathbb{C}^{F \times d}$  is a frequency matrix initialized as  $\mathbf{B} \sim \mathcal{N}(\mathbf{0}_F, \sigma_B^2 \text{Id}_F)$ . Note that  $\sigma_B$  is an hyperparameter linked to the variation intensity in the chart location space: a higher  $\sigma_B$  facilitates the decoder network’s ability to learn rapid variations in the target mapping. The  $\text{MLP}_{\mathbb{C}, T}$  block denotes a three-layer MLP of size  $T$ , with  $\text{ReLU}_{\mathbb{C}}^1$  activation functions and complex weights and biases.

The second decoding stage consists of a concatenation and a processing block. The role of the processing block is to introduce inter-UE interference cancellation in the concatenated channel matrix. This could be achieved using a neural network; however, the proposed approach leverages classical linear precoding techniques such as Maximum Ratio Transmitting (MRT) or Minimum Mean Squared Error (MMSE). These methods are grounded in well-established statistical criteria, offering robustness and interpretability. Thus, the processing block of  $\mathcal{D}_2$  only consists of a denormalization step, linear matrix operations, and a normalization step to respect the uniform power allocation constraint. A more comprehensive discussion on this matter is presented below.

**Learning strategy.** It is proposed to train the proposed model using an end-to-end approach and a task-oriented loss function. As the objective is to maximize the sum-rate, one could use the negative sum-rate of Eq. (5) as a loss function. However, Eq. (5) depends on the UEs noise variance and on the number of parallel UEs. This implies that, when using Eq. (5) as a loss function, the trained encoder and decoder would be optimal only for a fixed noise variance and a fixed number of parallel UEs, limiting the practical interest of the proposed method. It is suggested to use a heuristic learning strategy that considers the following loss function:

$$\mathcal{L} = 1 - \frac{1}{|\mathcal{B}|} \sum_{\mathbf{h} \in \mathcal{B}} \frac{|\mathbf{v}^H \mathbf{h}|^2}{\|\mathbf{h}\|_2^2}, \quad (7)$$

where  $\mathcal{B}$  denotes the batch channel set and  $\mathbf{v} = (\mathcal{D}_1 \circ \mathcal{E})(\mathbf{h})$ . One can show that Eq. (7) is minimized when the precoders are set to  $\mathbf{v}^* = \mathbf{h} e^{j\phi} / \|\mathbf{h}\|_2$ . Thus, under the constraint of effective loss convergence, the trained decoder  $\mathcal{D}_1$  provides the normalized channel affected by a phase shift. Through denormalization, one can then access a close approximation of the channel  $\mathbf{h}$ . Such channel can then be concatenated into a matrix, and used to form a precoding matrix, following the classical formulas of linear precoders. Note that, when considering this heuristic approach, the final decoding stage  $\mathcal{D}_2$  does not present any learnable parameters so that the learning is done on the parameters of the  $\mathcal{D}_1$  decoding stage.

**Parameter optimization.** The learnable parameters are conveyed by matrices  $\mathbf{D} \in \mathbb{C}^{D \times N_c}$  and  $\mathbf{Z} \in \mathbb{C}^{d \times N_c}$ , frequency matrix  $\mathbf{F} \in \mathbb{C}^{F \times d}$ , in addition to the weights and biases of the decoder’s MLP. As  $d \ll D$ , the main parameter complexity comes from the encoder. Indeed,  $N_c$  needs to be high in order to have a good initialization in the ISOMAP algorithm, resulting in a large matrix

<sup>1</sup>Note that  $\forall z_1 \in \mathbb{C}, \text{ReLU}_{\mathbb{C}}(z_1) = \text{ReLU}(\Re(z_1)) + j\text{ReLU}(\Im(z_1))$ .

---

### Algorithm 1 Similarity subsampling [25]

---

**Initialize:** Let  $\tilde{\mathbf{D}} \in \mathbb{C}^{D \times \tilde{N}}$ ,  $\tilde{\mathbf{Z}} \in \mathbb{R}^{d \times \tilde{N}}$ ,  $\mathbf{D}^c \in \mathbb{C}^{D \times (N_c - \tilde{N})}$ ,  $\mathbf{Z}^c \in \mathbb{R}^{d \times (N_c - \tilde{N})}$ , s.t.  $\mathbf{D} = \tilde{\mathbf{D}} \cap \mathbf{D}^c$ ,  $\mathbf{Z} = \tilde{\mathbf{Z}} \cap \mathbf{Z}^c$ .

- 1: **for each**  $\mathbf{h}_i \in \mathbf{D}^c$  **do**
- 2:    $s \leftarrow \max_{\mathbf{h}_j \in \tilde{\mathbf{D}}} \text{sim}(\mathbf{h}_i, \mathbf{h}_j)$
- 3:    $\{k, l\} \leftarrow \arg \max_{\substack{(\mathbf{h}_m, \mathbf{h}_n) \in \tilde{\mathbf{D}} \\ 1 \leq m \neq n \leq \tilde{N}}} \text{sim}(\mathbf{h}_m, \mathbf{h}_n)$
- 4:   **if**  $s < \text{sim}(\mathbf{h}_k, \mathbf{h}_l)$  **then**
- 5:     With proba.  $p : r \leftarrow k$ , otherwise  $r \leftarrow l$
- 6:      $\tilde{\mathbf{D}} \leftarrow (\tilde{\mathbf{D}} \setminus \mathbf{h}_r) \cup \mathbf{h}_i$
- 7:      $\tilde{\mathbf{Z}} \leftarrow (\tilde{\mathbf{Z}} \setminus \mathbf{z}_r) \cup \mathbf{z}_i$
- 8:   **end if**
- 9: **end for**

**Output:**  $\tilde{\mathbf{D}}$  (Subs. channels),  $\tilde{\mathbf{Z}}$  (Subs. chart locations)

---

**D.** Optimizing the encoder parameter complexity is of paramount importance as the encoding stage is done at the UE level. Once trained, the BS needs to send the learned encoder parameters to the UEs: reducing the number of parameters in the encoder is then equivalent to reducing the encoder update overhead.

In order to reduce this parameter complexity, it is proposed to introduce an intelligent subsampling of matrices  $\mathbf{D}$  and  $\mathbf{Z}$ . Firstly,  $\mathbf{D}$  and  $\mathbf{Z}$  are initialized using ISOMAP as aforementioned. Then the similarity subsampling algorithm proposed in [25], and recalled in Algorithm 1, is applied. Its objective is to discard channels that produce chart locations that are too similar. Indeed, similar chart locations can be considered as redundant, thereby providing no additional useful information. Theoretically, removing this redundancy from the calibration dataset allows to reduce the number of learnable parameters without impacting the global performance. More formally, Algorithm 1 aims to solve the following minimax problem:

$$\min_{(\mathbf{h}_i, \mathbf{h}_j) \in \tilde{\mathbf{D}}} \max_{(\mathbf{h}_i, \mathbf{h}_j) \in \tilde{\mathbf{D}}} \text{sim}(\mathbf{h}_i, \mathbf{h}_j), \quad (8)$$

where  $\tilde{\mathbf{D}}$  is the subsampled channel matrix and  $\text{sim}(\mathbf{h}_i, \mathbf{h}_j)$  is a similarity metric defined as the cosine similarity of the embedded channels:

$$\text{sim}(\mathbf{h}_i, \mathbf{h}_j) = \frac{|\mathcal{E}(\mathbf{h}_i)^H \mathcal{E}(\mathbf{h}_j)|}{\|\mathcal{E}(\mathbf{h}_i)\|_2 \|\mathcal{E}(\mathbf{h}_j)\|_2}. \quad (9)$$

Note that, when using Algorithm 1, the maximum similarity in  $\tilde{\mathbf{D}}$  is not hard-thresholded but rather indirectly controlled by the hyperparameter  $\tilde{N}$ .

## IV. EXPERIMENTS

**Simulation settings.** Two separate datasets are considered: a DeepMIMO [26] dataset with the urban outdoor ‘01’ scenario where the downlink central frequency is set to 3.5GHz, and a Sionna [27] dataset generated with the *Etoile* scene in Paris at 28GHz. Both datasets considers  $N_a = 64$  antennas and  $N_s = 16$  subcarriers over a 20 MHz bandwidth,  $N_c = 5k$  calibration channels, 3k training channels and 2k test channels.

Note that, for all following simulations, the precoders are learned for only one subcarrier, e.g. the central subcarrier.

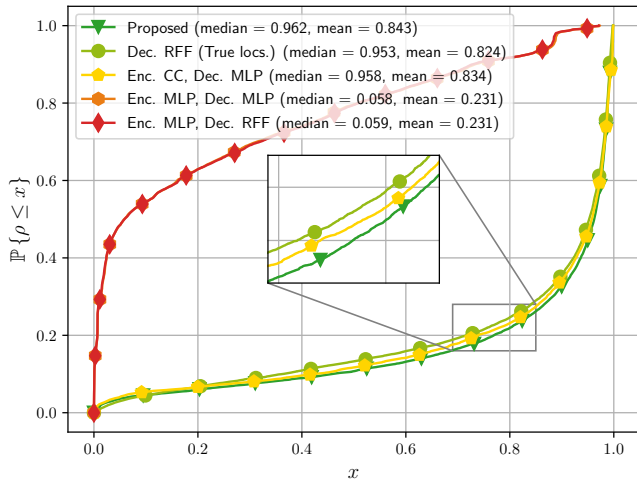


Fig. 3: CDF of the squared cosine similarities: Sionna dataset

Although it is possible to extend this approach to all subcarriers, this will be addressed in future work.

**Hyperparameters and baselines.** When considering encoder parameter subsampling, the number of kept channels is set to  $\tilde{N} = 10$  unless stated otherwise. Similarly, the embedding space dimension is set to  $d = 2$ . For the decoder,  $F = 200$  frequencies are considered, and  $\sigma_B$  is optimized for each dataset. Finally, the MLP layer size is set to  $T = 128$ .

Baselines consist of an MLP autoencoder where the embedding layer has dimension  $d$  as well as variants between the proposed approach and this MLP autoencoder: MLP encoder/RFF decoder and CC encoder/MLP decoder. Additionally, the RFF decoder with true locations as inputs is considered.

**Performance against baselines.** It clearly appears in Fig. 3, that CC, as a model-based encoder, has a significant role in the mono-UE model performance: every model with a MLP encoder presents poor performance. Additionally, it can be seen that the RFF decoder architecture provides a slight increase in performance. As the location manifold is inherently bidimensional, it is not surprising that bidimensional chart locations perform well in the proposed approach. This holds true under the continuity hypothesis of the underlying location-to-channel mapping [28], [29]. Additionally, note that the learned bidimensional channel charts carry more meaningful informations than true locations, as the proposed architecture outperforms the RFF decoder used with true locations.

**Encoder learning and subsampling impact.** The different networks are denoted as follows: a. No enc. learning, no subsampling, b. Enc. learning, no subsampling, c. Enc. learning, subsampling ( $\tilde{N} = 100$ ), d. Enc. learning, subsampling ( $\tilde{N} = 10$ ). It is shown in Table I and Fig. 4 that allowing the encoder to be optimized during training is beneficial: both the median and mean normalized correlations are increased. Additionally, it is shown that the subsampling procedure does not introduce a significant performance drop. Conversely, when considering  $\tilde{N} = 100$ , a slight performance increase is observed. The minor performance drop when considering  $\tilde{N} = 10$  can be attributed to the possibility of reduced scene coverage with such limited number of calibration channels. Indeed, as new chart locations are inferred from a convex combination of

Network		a	b	c	d
Params.		101k	10.3M	306k	121k
Sionna	Median $\rho$	0.88	<b>0.94</b>	<b>0.94</b>	<b>0.94</b>
	Mean $\rho$	0.79	0.86	<b>0.88</b>	<b>0.88</b>
DeepMIMO	Median $\rho$	0.91	0.96	<b>0.97</b>	0.96
	Mean $\rho$	0.73	0.85	<b>0.87</b>	0.84

TABLE I: Mono-UE precoder regression performance

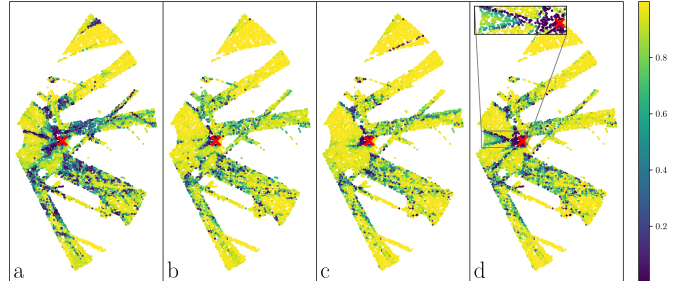


Fig. 4: Squared cosine similarity maps: Sionna dataset

the calibration chart locations, considering too few calibration channels may result in insufficient coverage for some regions of the chart location space. This can be seen in the region immediately on the left of the BS (denoted by a red cross) in Fig. 4. Notwithstanding, with the considered hyperparameters, the subsampling approach with  $\tilde{N} = 10$  allows almost a ninetyfold learnable parameter reduction with almost no performance drop. **Compression ratio impact.** The compression ratio can be defined as:

$$\gamma = \frac{2N_a N_s}{d + 1}, \quad (10)$$

as downlink channel vectors are composed of  $N_a N_s$  complex coefficients. The  $+1$  term at the denominator arises from the requirement of channel norm transfer to the BS, as a denormalization step is performed.

Fig. 5 presents the evolution of the median squared cosine similarity of the proposed approach with respect to the compression ratio. This is achieved by varying the embedding space dimension  $d$ . It can be seen that increasing  $d$  slightly increases performance. It should also be noted that it is possible to reach good performance with  $d = 2 \Rightarrow \gamma \simeq 1024$ , which significantly exceeds compression ratios reached with classical autoencoders (around 64). This confirms the interest of the proposed model-based approach.

**Muti-UE beamforming performance.** It can be seen in Fig. 6 that the learned MRT precoding matrix exhibits the same behavior as the actual MRT precoder. Indeed, the learned MRT can be seen as the concatenation of mono-UE MRT precoders affected by a phase shift. Such phase shift introduces the observed performance degradation. Additionally, it can be seen that, at high SNR, the learned MMSE precoder, obtained by considering the inferred precoding matrix  $\mathbf{W}$  as a channel matrix, outperforms the actual MRT precoder. This outlines the potential of the proposed approach, as it is possible to perform inter-UE interference cancellation with the proposed heuristic learning procedure that only considers individual UEs and compressed channel representations. Additionally, considering a lower compression ratio increases the mono-UE precoder learning

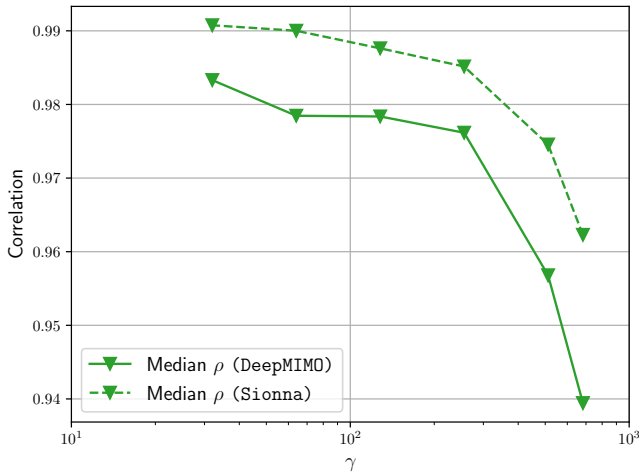


Fig. 5: Median  $\rho$  evolution with the compression ratio

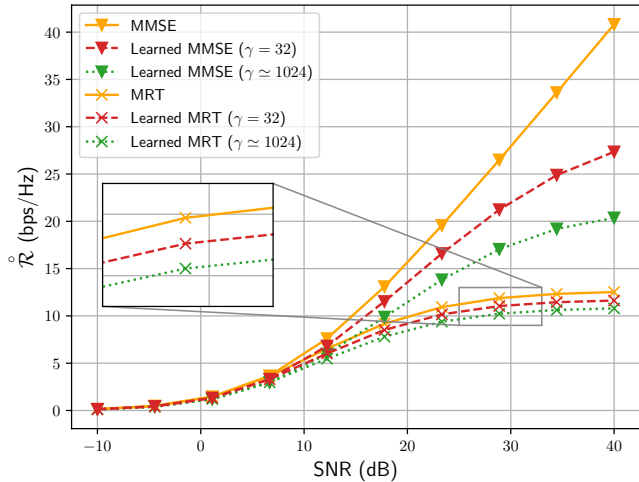


Fig. 6: SR performance: Sionna dataset

performance, as observed in Fig. 5, which translates into the multi-UE setting with better inter-UE interference cancellation.

## V. CONCLUSION AND PERSPECTIVES

In this paper, the problem of task-oriented CSI compression was studied. A model-based neural architecture, composed of a CC encoder and a RFF decoder, was proposed with the objective of maximizing the SR in a MU-MIMO system. An end-to-end heuristic learning strategy was introduced, and the number of learnable parameters was optimized through a subsampling method. The proposed method performance have been evaluated on realistic data, showing its superior performance against several baselines at high compression ratios. Additionally, this study outlined the performance of CC as a channel compression method for other tasks than localization. Future work will explore SR optimization through the introduction of a UE grouping partition policy and a learnable power allocation strategy.

## REFERENCES

- [1] C.-K. Wen, W.-T. Shih, and S. Jin, "Deep learning for massive MIMO CSI feedback," *IEEE Wireless Commun. Lett.*, vol. 7, no. 5, pp. 748–751, 2018.
- [2] J. Guo, C.-K. Wen, S. Jin, and G. Y. Li, "Convolutional neural network-based multiple-rate compressive sensing for massive MIMO CSI feedback: Design, simulation, and analysis," *IEEE Trans. on Wireless Commun.*, vol. 19, no. 4, pp. 2827–2840, 2020.

- [3] 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.
- [4] Z. Lu, J. Wang, and J. Song, "Multi-resolution CSI feedback with deep learning in massive MIMO system," in *IEEE Int. Conf. on Commun.*, 2020, pp. 1–6.
- [5] X. Chen, C. Deng, B. Zhou *et al.*, "High-accuracy CSI feedback with super-resolution network for massive MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 11, no. 1, pp. 141–145, 2022.
- [6] Y. Cui, A. Guo, and C. Song, "Transnet: Full attention network for CSI feedback in FDD massive MIMO system," *IEEE Wireless Commun. Lett.*, vol. 11, no. 5, pp. 903–907, 2022.
- [7] J. Guo, C.-K. Wen, S. Jin, and G. Y. Li, "Overview of deep learning-based CSI feedback in massive MIMO systems," *IEEE Trans. on Commun.*, vol. 70, no. 12, pp. 8017–8045, 2022.
- [8] C. Studer, S. Medjkouh, E. Gonultaş *et al.*, "Channel charting: Locating users within the radio environment using channel state information," *IEEE Access*, vol. 6, pp. 47 682–47 698, 2018.
- [9] T. Ponnada, P. Kazemi, H. Al-Tous *et al.*, "Best beam prediction in non-standalone mm wave systems," in *Joint European Conf. on Netw. and Commun./6G Summit*, 2021, pp. 532–537.
- [10] T. Yassine, B. Chatelier, V. Corlay *et al.*, "Model-based deep learning for beam prediction based on a channel chart," in *2023 57th Asilomar Conf. on Signals, Syst., and Comput.*, 2023, pp. 1636–1640.
- [11] L. Le Magoarou, T. Yassine, S. Paquelet, and M. Crussière, "Channel charting based beamforming," in *2022 56th Asilomar Conf. on Signals, Syst., and Comput.*, 2022, pp. 1185–1189.
- [12] F. Euchner, P. Stephan, and S. t. Brink, "Augmenting channel charting with classical wireless source localization techniques," in *2023 57th Asilomar Conf. on Signals, Syst., and Comput.*, 2023, pp. 1641–1647.
- [13] S. Taner, V. Palhares, and C. Studer, "Channel charting in real-world coordinates with distributed MIMO," 2024.
- [14] L. Le Magoarou, T. Yassine, S. Paquelet, and M. Crussière, "Channel charting based beamforming," in *2022 56th Asilomar Conf. on Signals, Syst., and Comput.*, 2022, pp. 1185–1189.
- [15] J. Guo, C.-K. Wen, and S. Jin, "Deep learning-based CSI feedback for beamforming in single- and multi-cell massive MIMO systems," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 7, pp. 1872–1884, 2021.
- [16] Q. Sun, H. Zhao, J. Wang, and W. Chen, "Deep learning-based joint CSI feedback and hybrid precoding in FDD mmWave massive MIMO systems," *Entropy*, vol. 24, no. 4, 2022.
- [17] Q. Xue, C. Dong, X. Li *et al.*, "Integrated deep implicit CSI feedback and beamforming design for FDD mmWave massive MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 12, no. 1, pp. 119–123, 2023.
- [18] T. Yassine, L. Le Magoarou, S. Paquelet, and M. Crussière, "Leveraging triplet loss and nonlinear dimensionality reduction for on-the-fly channel charting," in *23rd Int. Workshop on Signal Proc. Advances in Wireless Commun.*, 2022, pp. 1–5.
- [19] L. Le Magoarou, "Efficient channel charting via phase-insensitive distance computation," *IEEE Wireless Commun. Lett.*, vol. 10, no. 12, pp. 2634–2638, 2021.
- [20] A. Rahimi and B. Recht, "Random features for large-scale kernel machines," in *Adv. Neural Inf. Process.*, vol. 20, 2007.
- [21] N. Rahaman, A. Baratin, D. Arpit *et al.*, "On the spectral bias of neural networks," in *Int. Conf. on Mach. Learn.*, 2019, pp. 5301–5310.
- [22] Y. Cao, Z. Fang, Y. Wu *et al.*, "Towards understanding the spectral bias of deep learning," in *Proc. of the Thirtieth Int. Joint Conf. on Artif. Intell.*, 2021, pp. 2205–2211.
- [23] Z.-Q. John Xu, Y. Zhang, T. Luo *et al.*, "Frequency principle: Fourier analysis sheds light on deep neural networks," *Commun. in Comput. Phys.*, vol. 28, no. 5, pp. 1746–1767, 2020.
- [24] M. Tancik, P. Srinivasan, B. Mildenhall *et al.*, "Fourier features let networks learn high frequency functions in low dimensional domains," *Advances in Neural Inf. Process. Syst.*, vol. 33, pp. 7537–7547, 2020.
- [25] S. Taner, M. Guillaud, O. Tirkkonen, and C. Studer, "Channel charting for streaming CSI data," in *2023 57th Asilomar Conf. on Signals, Syst., and Comput.*, 2023, pp. 1648–1653.
- [26] A. Alkhateeb, "DeepMIMO: A generic deep learning dataset for millimeter wave and massive MIMO applications," in *Proc. of Inf. Theory and Appl. Workshop*, San Diego, CA, 2019, pp. 1–8.
- [27] J. Hoydis, S. Cammerer, F. Ait Aoudia *et al.*, "Sionna: An open-source library for next-generation physical layer research," *arXiv preprint*, 2022.
- [28] B. Chatelier, L. L. Magoarou, V. Corlay, and M. Crussière, "Model-based learning for location-to-channel mapping," in *IEEE Int. Conf. on Acoust., Speech and Signal Process.*, 2024, pp. 12 836–12 840.
- [29] B. Chatelier, V. Corlay, M. Crussière, and L. Le Magoarou, "Model-based learning for multi-antenna multi-frequency location-to-channel mapping," *arXiv preprint*, 2024.