CSI-based Outdoor Localization for Massive MIMO: Experiments with a Learning Approach

Alexis Decurninge, Luis García Ordóñez, Paul Ferrand, He Gaoning, Li Bojie, Zhang Wei, Maxime Guillaud

Huawei Technologies

Abstract—We report on experimental results on the use of a learning-based approach to infer the location of a mobile user of a cellular network within a cell, for a 5G-type Massive multiple input, multiple output (MIMO) system. We describe how the sample spatial covariance matrix computed from the CSI can be used as the input to a learning algorithm which attempts to relate it to user location. We discuss several learning approaches, and analyze in depth the application of extreme learning machines, for which theoretical approximate performance benchmarks are available, to the localization problem. We validate the proposed approach using experimental data collected on a Huawei 5G testbed, provide some performance and robustness benchmarks, and discuss practical issues related to the deployment of such a technique in 5G networks.

I. Introduction

Future 5G systems are expected to incorporate base stations (BSs) equipped with massive multiple-input multiple-output (MIMO) technology, which uses arrays with the number of antenna elements being some orders of magnitude larger than current state-of-the-art MIMO BSs [1]. Beyond the expected network-level performance improvements, the presence of excess antennas substantially increases the spatial resolution of the antenna array. This opens the door to localization methods of increased accuracy based on measuring the propagation characteristics of the wireless channel. In this article, we investigate to which extent the uplink channel state information (CSI), as estimated by the base station for communications purposes,can be exploited to infer the position of a user equipment (UE)[1].

There is a rich literature on localization strategies based on estimating the state of the wireless propagation channel. Most of the existing techniques can be understood as follows: the network measures or estimates some propagation parameter—e.g. channel impulse response, propagation delay, angle of arrival or received signal strength—at one or multiple positions in the network, typically BSs or access points (APs). This *fingerprint* of the user location can be then either compared to existing databases and maps [2], [3], or subsequently intepreted in a geometric sense. The assumption of the first approach is that there should be a bijective relationship between the user position and the fingerprint, although this relationship is usually too rich to be modeled and must thus be learned. A variety of learning algorithms can be applied

in this context, among which K-nearest neighbors (K-nNs), neural network (NNs), support vector machines (SVMs) and Bayesian learning are the most popular. In contrast, for the geometric approach, the fingerprints are translated into distances and angles from which the user location is triangulated or trilaterated. In both approaches, the dimension of the fingerprint is usually the parameter limiting the localization accuracy. For example, it was shown in [3] that received signal strength indicator (RSSI)-based approaches become extremely accurate for indoor environments as the number of reference measurements increases.

This limitation can be overcome through the use of multiple antennas at the APs or BSs, as proposed in [4], [5]. A similar approach has also been applied in [6] to outdoor localization in the so-called *distributed* MIMO case, in which the antenna elements belonging to the same (virtual) BS are not collocated in a single antenna array but distributed within the cell. A common way to obtain spatial fingerprints is to extract the cluster of scatterers angle and delay from the raw CSI [7]. This fingerprinting method has been studied both in [8] and [9]. In the former, the authors applied deep neural networks to the obtained fingerprints, while in the latter K-nNs type algorithms were considered for their affordable complexity.

In this article, we focus on fingerprinting at a single BS equipped with a large (massive MIMO) antenna array. We adopt a NN-based learning approach and compare it with a more classical K-nN learning algorithm. In particular, we choose a single-layer random feed-forward neural network, usually referred to as an extreme learning machine (ELM) [10], [11]. Beyond its simplicity, the ELM has been proven to exhibit universal approximation capabilities and good generalization performance for a wide range of problems. In the context of localization, the authors of [12]–[14] use RSSI measurements coupled with extreme learning machines for indoor localization. They however only consider single-antenna elements at the APs. The rest of the paper is outlined as follows. We introduce our approach in Section II; the experimental results are presented in Section III. Finally, in Section IV we discuss some general implementation issues that need to be considered before deployment.

II. Proposed CSI-Based Localization Approach

A. Input Data Pre-Processing

When multiple antennas are used in a single BS at a single cell, some additional pre-processing of the raw channel data is needed to form a useful input for any learning algorithm.

---

[1]Note that we focus on network-side localization here, i.e., the ability for the network to localize a UE, as opposed to user-side localization such as the one enabled by the GPS system, whereby the location information is directly obtained at the UE, but not necessarily shared with the network operator.

We consider a system where CSI acquisition takes place in the frequency domain, consistent with the OFDM/OFDMA approach being standardized for 5G systems and we let $\mathbf{h}(s,t) \in \mathbb{C}^N$ denote the complex vector representing the instantaneous uplink propagation channel between the single-antenna user equipment (UE) and the N BS antennas on subcarrier s at time t. Even if the UE position in the horizontal plane (denoted by $\mathbf{y} \in \mathbb{R}^2$) is completely static, the instantaneous channel as measured by the baseband processor typically shows some variations due to small-scale fading and to the residual carrier frequency offset resulting from the impossibility of perfectly synchronizing the oscillators of the transmitter and the receiver. In order to reduce the noise present in the input data and simultaneously remove the common phase component due to the clock offset, we compute the BS-side channel covariance as

$$\mathbf{C}(\mathbf{y}) = \mathbb{E}\left[\mathbf{h}(s,t)\mathbf{h}(s,t)^\dagger\right] \tag{1}$$

for some fixed reference subcarrier s and where $(\cdot)^\dagger$ denotes the Hermitian transposition. We implicitly assume that the expectation is computed at time t over a small time horizon ΔT while the UE remains fixed in a given location $\mathbf{y}$, so that the fading process can be assumed stationary. In practice, the expectation in (1) will be replaced by a sample covariance estimator (see Section III for more details about our experimental set-up).

In the sequel, we consider as the input to the learning algorithm a vectorized and normalized[2] version of the covariance $\mathbf{C}(\mathbf{y})$, denoted by $\mathbf{x} \in \mathbb{R}^{N^2}$ and obtained as

$$\mathbf{x} = \left(\text{vec}_{\text{ut}}\big(\text{Re}\{\tfrac{\mathbf{C}(\mathbf{y})}{\text{Tr}(\mathbf{C}(\mathbf{y}))}\}\big)^{\text{T}}, \text{vec}_{\text{ut}}\big(\text{Im}\{\tfrac{\mathbf{C}(\mathbf{y})}{\text{Tr}(\mathbf{C}(\mathbf{y}))}\}\big)^{\text{T}}\right)^{\text{T}} \tag{2}$$

where $\text{vec}_{\text{ut}}(\cdot)$ stands for the operator stacking in a vector the upper triangular entries of a $N \times N$ matrix as long as they are non-zero and, hence, removes the inherent redundancy present in $\mathbf{C}(\mathbf{y})$. This input choice comes from the intuition that the second-order channel statistics capture most of the location-related characteristics of the CSI and, consequently, $\mathbf{C}(\mathbf{y})$ can be directly related to the UE location $\mathbf{y}$. However, the complexity of the relationship between the spatial scattering environment surrounding a user and the CSI as measured by the BS makes it extremely difficult to express or even model; therefore, we propose to infer it through statistical learning.

B. Learning approaches

It is convenient to distinguish between the following two phases of the use of a learning algorithm:

(i) *the training phase*, in which the CSI related input and the UE positions are acquired, i.e., the *training* data set $(\mathbf{x}_j, \mathbf{y}_j)_{j\in\mathcal{T}}$ consisting of $T = |\mathcal{T}|$ pairs $\mathbf{x}_j \in \mathbb{R}^M$, with $M = N^2$, as defined in (2) and the corresponding user location $\mathbf{y}_j \in \mathbb{R}^2$; this training set is used to infer the mapping relating $\mathbf{x}_j$ to $\mathbf{y}_j$, and

[2]Here we remove the pathloss information (see Section III). Note that there could be alternative approaches to taking the normalized covariance to pre-process the CSI; we did not attempt to exhaustively research this question.
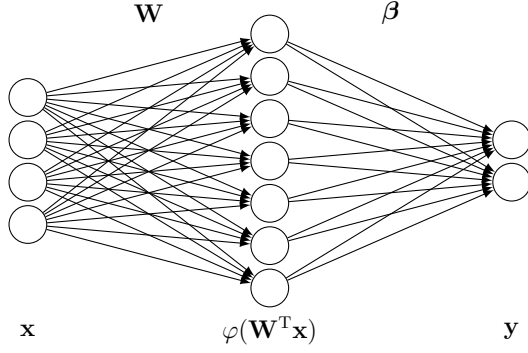
(ii) *the exploitation phase*, in which the mapping obtained in (i) is used to infer the UE position from the observed CSI; the performance of this step is evaluated on a separate *testing* data set $(\mathbf{x}_j, \mathbf{y}_j)_{j\in\mathcal{T}'}$ with $T' = |\mathcal{T}'|$ data pairs.

The adopted learning approaches are detailed below.

1) K-nearest neighbors: As a baseline method for comparison purposes, we consider a regression mapping using K-nearest neighbors. In this case the training phase only consists in acquiring the training data set. Then, for any observed input data $\mathbf{x}$ during the exploitation phase, we find the K nearest neighbors of $\mathbf{x}$ in the training set, i.e., the K indices $j_1, \ldots, j_K \in \mathcal{T}$ such that, for some distance $d(\cdot, \cdot)$ defined on the input space (e.g., $d(\mathbf{x}_1, \mathbf{x}_2) = \|\mathbf{x}_1 - \mathbf{x}_2\|_2$), we have $d(\mathbf{x}_{j_1}, \mathbf{x}) \leq \cdots \leq d(\mathbf{x}_{j_K}, \mathbf{x}) \leq d(\mathbf{x}_j, \mathbf{x})$ for any other $j \in \mathcal{T}$. Finally, the estimated UE location $\hat{\mathbf{y}}$ is computed as the barycenter of the locations of the K nearest neighbors:

$$\hat{\mathbf{y}} = \frac{1}{K}\sum_{k=1}^{K}\mathbf{y}_{j_k}. \tag{3}$$

2) Extreme Learning Machine: As the main learning method we consider ELMs, which are feedforward neural networks consisting of a random single-hidden layer of neurons (see an example in Fig. 1). An ELM can be understood as a non-linear regression function of the form

$$f : \mathbf{x} \in \mathbb{R}^M \mapsto \sum_{i=1}^{n}\varphi(\mathbf{w}_i^{\text{T}}\mathbf{x})\boldsymbol{\beta}_i \in \mathbb{R}^2 \tag{4}$$

where $\mathbf{w}_1, \ldots, \mathbf{w}_n \in \mathbb{R}^M$ are weights of the n neurons, $\varphi(\cdot)$ is an activation function (see Table I for examples), and $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_n \in \mathbb{R}^2$ are the output weights. In ELMs the entries of the weight matrix $\mathbf{W} = (\mathbf{w}_1, \ldots \mathbf{w}_n)$ are independently drawn, e.g., from a Gaussian standard distribution and only the output weights $\boldsymbol{\beta} = (\boldsymbol{\beta}_i)_{i=1\ldots n}$ need to be optimized.

During the training phase, $\boldsymbol{\beta}$ is obtained using a simple linear ridge regression between the augmented version of the input, $(\varphi(\mathbf{w}_1^{\text{T}}\mathbf{x}), \ldots, \varphi(\mathbf{w}_n^{\text{T}}\mathbf{x}))^{\text{T}} \in \mathbb{R}^n$, and the output $\mathbf{y}$:

$$\boldsymbol{\beta} = \mathbf{Y}_{\mathcal{T}}(\boldsymbol{\Sigma}_{\mathcal{T}}^{\text{T}}\boldsymbol{\Sigma}_{\mathcal{T}} + \gamma\mathbf{I}_{\mathcal{T}})^{-1}\boldsymbol{\Sigma}_{\mathcal{T}}^{\text{T}} \tag{5}$$

where we denote $\boldsymbol{\Sigma}_{\mathcal{T}} = (\varphi(\mathbf{w}_i^{\text{T}}\mathbf{x}_j))_{i=1,\ldots n,\ j\in\mathcal{T}} \in \mathbb{R}^{n\times T}$ and $\mathbf{Y}_{\mathcal{T}} = (\mathbf{y}_j)_{j\in\mathcal{T}} \in \mathbb{R}^{2\times T}$, and we have introduced an L_2 regularization parameter γ, which makes the regression more robust without adding complexity. In the exploitation phase, the UE location for a new observed input $\mathbf{x}$ is estimated as

$$\hat{\mathbf{y}} = \sum_{i=1}^{n}\varphi(\mathbf{w}_i^{\text{T}}\mathbf{x})\boldsymbol{\beta}_i = \boldsymbol{\beta}^{\text{T}}\varphi(\mathbf{W}^{\text{T}}\mathbf{x}). \tag{6}$$

The choice of ELM as a learning approach is motivated by its theoretical capacity to learn any non-linear mapping, as well as its simplicity [10], [11] and amenability to theoretical analysis [15].

ReLu	$\varphi(t) = \max(0,t)$
Step	$\varphi(t) = \mathbb{1}_{t\geq 0}$
Sign	$\varphi(t) = \mathbb{1}_{t>0} - \mathbb{1}_{t<0}$

TABLE I: Examples of activation functions.

Fig. 1: A typical extreme learning machine with $M = 4$ inputs, $n = 8$ neurons and 2 outputs.

C. Localization Performance

In order to analyze the performances of the considered localization approaches, we measure the localization error as

$$e(\mathbf{y}_j, \hat{\mathbf{y}}_j) = \|\mathbf{y}_j - \hat{\mathbf{y}}_j\| \tag{7}$$

where $\hat{\mathbf{y}}_j$ is the estimated position for the input $\mathbf{x}_j$. We define the corresponding average training and testing errors as $E = \frac{1}{T} \sum_{j \in \mathcal{T}} \|\hat{\mathbf{y}}_j - \mathbf{y}_j\|$ and $E' = \frac{1}{T'} \sum_{j \in \mathcal{T}'} \|\hat{\mathbf{y}}_j - \mathbf{y}_j\|$.

In the case of the ELM approach we also introduce the training and testing mean square localization errors:

$$\text{mse}(\mathbf{W}) = \frac{1}{T} \sum_{j \in \mathcal{T}} \|\hat{\mathbf{y}}_j - \mathbf{y}_j\|^2 \tag{8}$$

$$\text{mse}'(\mathbf{W}) = \frac{1}{T'} \sum_{j \in \mathcal{T}'} \|\hat{\mathbf{y}}_j - \mathbf{y}_j\|^2 \tag{9}$$

where we have made explicit their randomness through the dependency on the random weight matrix $\mathbf{W}$. The authors of [15] show that for the regression in (6), both training and testing mean square errors almost surely converge to some deterministic limit, i.e.,

$$\text{mse}(\mathbf{W}) - \overline{\text{mse}} \xrightarrow[n,M,T\to\infty]{\text{a.s.}} 0 \tag{10}$$

$$\text{mse}'(\mathbf{W}) - \overline{\text{mse}}' \xrightarrow[n,M,T'\to\infty]{\text{a.s.}} 0 \tag{11}$$

where the expressions of the deterministic equivalents $\overline{\text{mse}}$ and $\overline{\text{mse}}'$ are given in [15, Conject. 1 and Thm. 3]. Note that $\overline{\text{mse}}$ and $\overline{\text{mse}}'$ still depend on the training and testing data sets in order to avoid assuming a generative model of the entries.

The computation of these deterministic values allow us to choose the best hyperparameters, i.e., the regularization parameter γ, the activation function and the number of neurons independently of the random matrix $\mathbf{W}$.

III. EXPERIMENTAL RESULTS

Data was collected on the Huawei 5G testbed on the campus of the University of the Chinese Academy of Science in Huairou, P.R. China. CSI data was measured for a single-antenna UE on the uplink of a massive MIMO system using a 32 dual-polarized antenna array at the BS, which is located on the
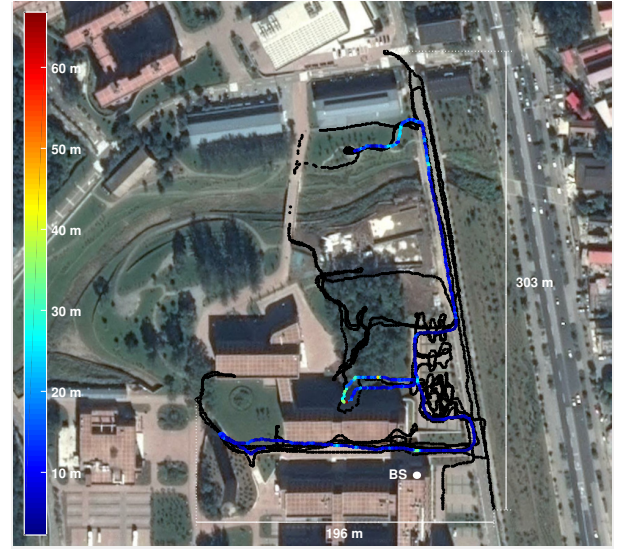

Fig. 2: Map of the measurement area with the locations of the training set (black dots) and testing data sets (colored dots) with testing location error (using the scale on the left) plotted for the true UE location. Imagery © Google, DigitalGlobe.

roof of a five-floor building. Due to some malfunctioning RF chains, the measurements from 8 antennas had to be discarded and, hence, unless otherwise stated, $N = 56$. In total, 90 minutes of outdoors measurements were acquired, covering a maximum extent of 196×303 meters, depicted on Fig. 2. Since the mobile UE was moved on a cart, the paths for which data could be recorded was limited to roads and pedestrian walkways. The scenario includes a majority of line-of-sight (LOS) locations, as well as some non-LOS locations. The UE position (latitude/longitude) was recorded using a commercial GPS with 1 Hz sampling rate; it was transformed to a local coordinate system to have dimensions in meters, and subsequently linearly interpolated in time when required. The CSI sampling rate was 10 Hz. Path-loss information was not deemed reliable due to automatic gain control, and therefore all covariance matrices were normalized as in (2). Gaps in the trace are due to the UE being out of range from the BS.

Clearly, when the UE is moving, $\mathbf{h}(s,t)$ can not be considered to be stationary and the statistics cannot be related to the position. This makes it difficult to evaluate the expectation when computing in practice the covariance as defined in (1). In order to escape this apparent contradiction and to be able to process the experimental data acquired for a moving user, we apply a sample covariance estimator over a 1-second horizon,[3] where the statistics can be assumed approximately constant. In the experiments presented here, a single subcarrier (always the same) is used, since including multiple ones have shown little or no gain, even when the selected subcarriers were spaced apart within the available 10 MHz band. We speculate that

---

[3]Note that at pedestrian speeds, 1 s translates into approximately 1 m, which is small with respect to the size of the considered area, and roughly commensurate with the expected accuracy of the GPS location data.
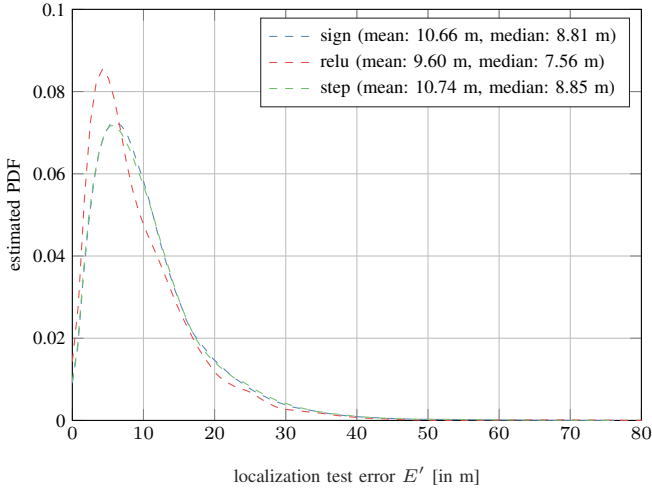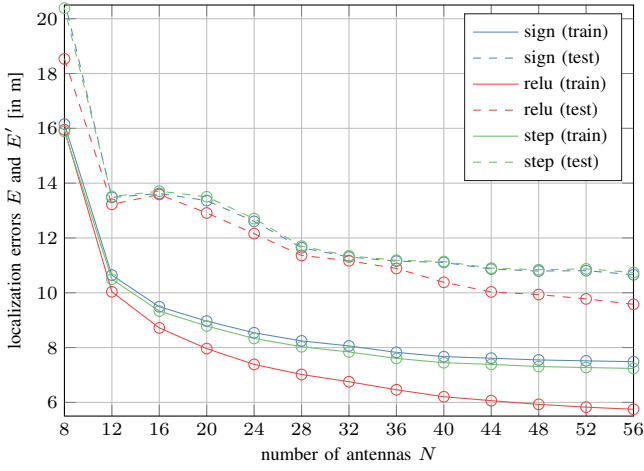
Fig. 3: Localization error distribution.

Fig. 4: Average localization error vs. number of antennas

this is due to the lack of channel frequency selectivity in the considered scenario.

In Fig. 2 we present the resulting localization error as defined in (7) for the ELM approach with $n = 10^4$, a ReLu activation function, and $\mathbf{W}$ with i.i.d. standard normal entries. The black dots depict the training data set of size $T = 3500$, while the colored dots show the localization error for each location within the testing set of size $T' = 550$, using the scale on the left. Observe that the accuracy lies mostly below 10 m for most locations, while some loss of accuracy is noticeable for locations far away from the BS (where the CSI estimates are more noisy due to a lower SNR) and for a set of NLOS locations where the BS is shadowed by a tall (six floors) building. This can be further confirmed by the testing error distribution estimated over 1000 realizations of $\mathbf{W}$ in Fig. 3.

A. Influence of the number of antennas

In order to evaluate the sensitivity of the localization error to the number of antennas, we ran simulations with varying number of BS antennas by discarding part of the data. For each value of N, the subset of antenna elements in use was chosen

so as to heuristically maximize the inter-element spacing, and the polarization diversity. The result is shown in Fig. 4. Our experiments indicate that the localization error is low and exhibits little sensitivity to N as long as $N \geq 12$, while for $N \leq 8$ the localization accuracy decreases significantly. This indicates that the potential applications of the proposed method are not limited to the massive MIMO regime, but also extends to the typical number of antennas found in 4.5G MIMO systems.

B. Optimization of the ELM hyperparameters

The theoretical framework of [15] was used to optimize certain parameters of the ELM (number of neurons n, activation function, and regularization parameter γ) based on the training and testing data sets. The results are presented in Fig. 5, where it can be noticed that the ReLu activation function uniformly outperforms the sign and step functions.

The dependency on the number of neurons n is shown in Fig. 5a. Both deterministic equivalents for the training and testing MSEs decrease when n grows. The sharp drop in the training MSE that can be observed for $n \geq 10^4$ corresponds to $n \geq T$. Note, however, that no adverse over-fitting effect is observed for large n (the testing MSEs keep monotonically decreasing towards the asymptotic values of $\overline{\mathsf{mse}}'$ plotted in dash-dotted lines). The curves of Fig. 5a have been generated by computing the optimum γ minimizing $\overline{\mathsf{mse}}'$ for each n.

Fig. 5b shows the influence of the regularization parameter γ on the localization MSE, for the three considered activation functions. Here, we fixed $n = 10^4$ since it provides a good tradeoff between complexity and performance. The dots correspond to the experimental values of $\mathbb{E}_{\mathbf{W}}\{\mathsf{mse}(\mathbf{W})\}$ and $\mathbb{E}_{\mathbf{W}}\{\mathsf{mse}'(\mathbf{W})\}$ directly computed from running the trained ELM on the training and testing data for different realizations of $\mathbf{W}$, while the solid and dashed lines show the corresponding deterministic equivalents (see (10) and (11)). It can be seen that there is an excellent match between the deterministic equivalents and the experimental values.

C. Comparison of ELM with K-nN approach

In order to benchmark our results, we compare in Table II the ELM learning approach with the K-nN approach presented in Section II-B1 wherein the chosen distance in the input space is the 2-norm. As we see in Table II, K-nN outperforms the ELM-based approach in terms of the average accuracy, while at the same time being more prone to catastrophic errors. However, at this stage we refrain from drawing conclusions from this comparison, since the set-up is unfavorable to the ELM approach due to the small size of the training dataset (with respect to the dimension of the data). Furthermore, the dataset is very structured in the sense that it consists only of pathways, and this guarantees that for every data point in the testing data set there are always a couple of "good" nearest neighbors, which results in the best situation for the K-nN algorithm. Finally, we would like also to remark that the computational complexity for the K-nN localization strategy grows with the size of the training data set, whereas in the

ELM case is only determined by the number of neurons. This can render the K-nN approach unfeasible is practice.

IV. Practical implementation issues

Here we discuss the practical issues related to the implementation of a learning-based localization approach in a cellular network. In particular, the availability of training data, and potentially the necessity to periodically update this training, are important questions in the context of the proposed application.

A. Potential training approaches

The method outlined in Section II relies on the availability of a training dataset for which CSI has been obtained for a large number of reference positions, i.e., we need to assume that a reference position information is available for this data set.

Since the mapping between CSI and location is strongly dependent on the geometric properties of the scattering environment which surrounds the BS (i.e. shape and location of buildings and other macroscopic obstacles), it is expected that a dedicated training must be performed for each BS site. This also indicates that additional training might be required if significant changes occur in the scattering environment (new buildings...). Therefore, it is important to identify practically feasible strategies to deal with the acquisition (and potentially, with the updates) of the training data set. Among the possible approaches, we distinguish between operator-based and crowdsourcing-based training.

a) Operator-based training: In this strategy, the training data is acquired through dedicated measurements performed by the operator. The location of the reference UE is determined either through GPS, or through other kinds of geometric measures (distance to a reference point in the horizontal plane, elevation based on the floor number in case of 3-D localization in indoor environments...); this operation can be performed by a human or robotic operator. Although it is clearly a costly approach, it comes with the guarantee that the training data set is accurate and exhaustive.

b) Crowdsourcing-based training: In this approach, the location data is collected by a subset of the end users of the cellular network. Practically, this can take the form of an operator-provided application which runs on the users' mobile device and relies on its existing localization service (GPS-based or other) to determine its position, and feed it back to the operator; the operator can provide a reward (monetary, or through a discount or additional services) in exchange for the voluntary participation of the user to this crowdsourcing data collection. For a given user participating in the crowdsourcing, this data collection does not have to be permanent, so that the associated impact on power and data

consumption is minimized. In that scenario, the operator has no control over the geographical area covered by the training data, since it merely follows the end user's behavior; thus, it can not be ensured that the training data covers uniformly and/or exhaustively the geographical area covered by the cell. On the other hand, it can be argued that, if sufficiently many users can be enrolled in the crowdsourcing data collection, the distribution of user locations in the training data will match the real user distribution, ensuring a faithful location estimate during the exploitation phase.

B. Training Update

In the context of the considered application, the mapping between CSI and localization within the cell is clearly a function of the scattering environment, which can not always be assumed constant over long periods of time (due e.g. to new buildings...). Hardware updates, or the aging of the electronic components can also make this mapping vary over time. It is therefore necessary to take into account these slow-scale time variations to properly assess the accuracy of learning-based localization. This calls for updates of the learned function over time, and thus the collection of new data.

If the old training data is not completely stale, it might be useful to re-train the neural network using the merged old and new training datasets. In order to reduce the complexity of this re-training, update mechanisms (such as [16] for ELMs) can readily be used to provide an online solution.

The question of when to trigger an update (through the collection of new training data) can be answered by using the theoretical framework of [15], e.g., by checking how good the new training data can be predicted by the old one. In the case of crowdsourcing, the system can include on-demand acquisition of training data through a mechanism whereby data collection is enabled if the measured CSI for a crowdsourcing user can not be accurately matched to its location.

### C. Channel frequency selectivity

Modern cellular networks rely on some form of frequency-division multiple access (FDMA) to multiplex the users in a cell. As a consequence, for a given user, the channel estimate is available only for a fraction of the spectrum (typically, a group of OFDM subcarriers, or *resource block group*). The consequence is that the channel covariance matrix associated to a given user can only be estimated for a fraction of the spectrum. If the total system bandwidth is large, this covariance matrix can not be assumed constant in the frequency domain. This complicates the learning process, since the mapping between channel covariance and UE location will depend on the considered frequency. One approach is to fractionate the system bandwidth into subbands in which the channel covariance can be assumed constant, and repeat the learning process independently for each subband. Another solution is to add a pre-processing step consisting in transposing (in the frequency domain) the measured covariance matrix to a reference frequency (see e.g. [17]–[20]), and to let the neural network operate solely on the reference frequency.

Statistic	ELM (ReLu)	3-nN
Average localization error [in m]	9.60	8.16
Median localization error [in m]	7.56	4.29
Maximum localization error [in m]	73.26	108.25

TABLE II: Comparison of statistics computed on the testing set on localization error between 3-nNs and ELM.
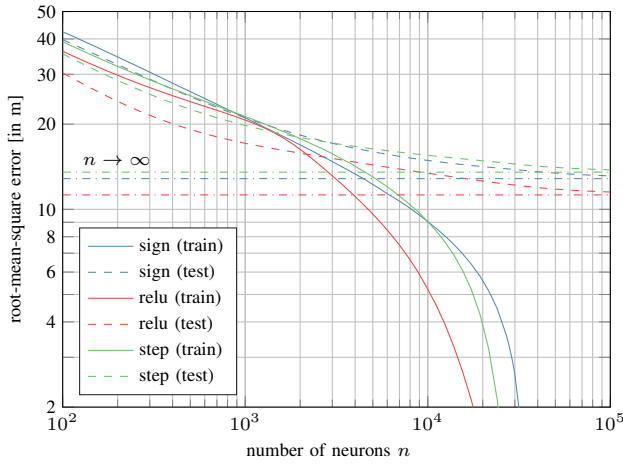
(a) $\sqrt{\mathrm{mse}}$ (solid) and $\sqrt{\mathrm{mse}'}$ (dashed) vs number of neurons.
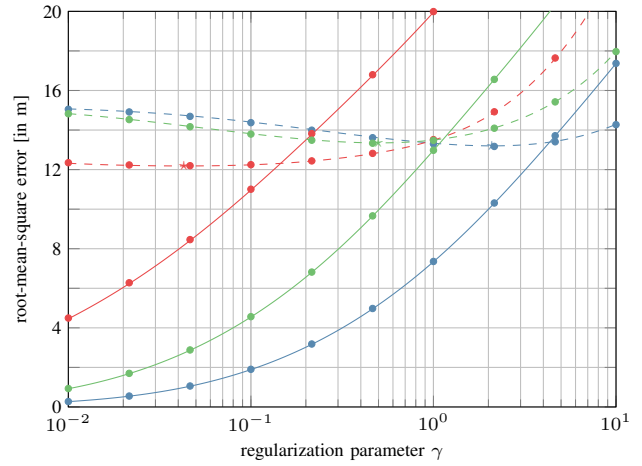
(b) $\sqrt{\mathrm{mse}}$ (solid) and $\sqrt{\mathrm{mse}'}$ (dashed) vs. regularization parameter.

Fig. 5: Optimization of the neural network parameters based on the deterministic equivalent analysis of the ELM.

D. Channel covariance estimation

A full estimate of the uplink channel matrix is not always available at the BS; in particular this is the case for multiple-antenna UEs when precoded reference signals are in use—in this case, only the product between the precoder and the instantaneous channel can be estimated. Another case is when channel estimation is performed for multiple users simultaneously using non-orthogonal reference symbols [21]. However, even in these cases where the complete instantaneous CSI matrix is not estimated, it is generally possible to obtain unbiased estimates of the covariance matrix of each user, at the cost of more complex processing [22].

REFERENCES

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

[2] M. Triki, D. T. M. Slock, V. Rigal, and P. Francois, "Mobile terminal positioning via power delay profile fingerprinting: Reproducible validation simulations," in *Proc. IEEE Vehicular Technology Conference (VTC)*, Sept 2006.

[3] A. Khalajmehrabadi, N. Gatsis, and D. Akopian, "Modern WLAN fingerprinting indoor positioning methods and deployment challenges," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 3, 2017.

[4] S. Dayekh, S. Affes, N. Kandil, and C. Nerguizian, "Cost-effective localization in underground mines using new SIMO/MIMO-like fingerprints and artificial neural networks," in *ICC Workshop on Communications in Underground and Confined Environments*, 2014.

[5] Y. Chapre, A. Ignjatovic, A. Seneviratne, and S. Jha, "CSI-MIMO: Indoor Wi-Fi fingerprinting system," in *IEEE Conference on Local Computer Networks*, 2014.

[6] V. Savic and E. G. Larsson, "Fingerprinting-based positioning in distributed massive MIMO systems," in *IEEE Vehicular Technology Conference (VTC Fall)*, 2015.

[7] D. Shutin, "Clustering wireless channel impulse responses in angular-delay domain," in *IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2004.

[8] J. Vieira, E. Leitinger, M. Sarajlic, X. Li, and F. Tufvesson, "Deep convolutional neural networks for massive MIMO fingerprint-based positioning," in *Proc. IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, 2017.

[9] X. Sun, X. Gao, Y. G. Li, and W. Han, "Single-site localization based on a new type of fingerprint for massive MIMO-OFDM systems," *IEEE Transactions on Vehicular Technology*, 2018.

[10] G. Huang, Q. Zhu, and C. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, no. 1–3, pp. 489–501, 2006.

[11] G. Huang, G.-B. Huang, S. Song, and K. You, "Trends in extreme learning machines: A review," *Neural Networks*, vol. 61, pp. 32–48, 2015.

[12] H. Zou, B. Huang, X. Lu, H. Jiang, and L. Xie, "A robust indoor positioning system based on the Procrustes analysis and weighted extreme learning machine," *IEEE Transactions on Wireless Communications*, vol. 15, no. 2, Feb. 2016.

[13] J. Zhang, W. Xiao, S. Zhang, and S. Huang, "Device-free localization via an extreme learning machine with parameterized geometrical feature extraction," *Sensors*, vol. 17, no. 4, Apr. 2017.

[14] X. Lu, H. Zou, H. Zhou, L. Xie, and G.-B. Huang, "Robust extreme learning machine with its application to indoor positioning," *IEEE Transactions on Cybernetics*, vol. 46, no. 1, Jan. 2016.

[15] C. Louart, Z. Liao, and R. Couillet, "A random matrix approach to neural networks," *Annals of Applied Probability*, 2017.

[16] N.-Y. Liang, G.-B. Huang, P. Saratchandran, and N. Sundararajan, "A fast and accurate online sequential learning algorithm for feedforward networks," *IEEE Transactions on Neural Networks*, vol. 17, no. 6, pp. 1411–1423, Nov. 2006.

[17] G. D. Yantao Han, Jiqing Ni, "The potential approaches to achieve channel reciprocity in FDD system with frequency correction algorithms," in *Proc. CHINACOM*, 2010.

[18] A. Decurninge, M. Guillaud, and D. Slock, "Channel covariance estimation in massive MIMO frequency division duplex systems," in *Proc. IEEE Global Telecommunications Conference (GLOBECOM)*, San Diego, CA, USA, 2015.

[19] S. Haghighatshoar, M. B. Khalilsarai, and G. Caire, "Multi-band covariance interpolation with applications in massive MIMO," 2018, submitted to ISIT 2018. [Online]. Available: https://arxiv.org/abs/1801.03714

[20] L. Miretti, R. L. Cavalcante, and S. Stańczak, "FDD massive MIMO channel spatial covariance conversion using projection methods," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2018.

[21] P. Ferrand, A. Decurninge, M. Guillaud, and L. G. Ordóñez, "Efficient channel state information acquisition in massive MIMO systems using non-orthogonal pilots," in *Proc. Workshop on Smart Antennas (WSA)*, Mar. 2017.

[22] A. Decurninge and M. Guillaud, "Covariance estimation with projected data: Applications to CSI covariance acquisition and tracking," in *Proc. European Signal Processing Conference (EUSIPCO)*, 2017.