

Cooperative Deep-Learning Positioning in mmWave 5G-Advanced Networks

Bernardo Camajori Tedeschini¹, Graduate Student Member, IEEE, and Monica Nicoli², Senior Member, IEEE

Abstract—In application verticals that rely on mission-critical control, such as cooperative intelligent transport systems (C-ITS), 5G-Advanced networks must be able to provide dynamic positioning with accuracy down to the centimeter level. To achieve this level of precision, technology enablers, such as massive multiple-input multiple-output (mMIMO), millimeter waves (mmWave), machine learning and cooperation are of paramount importance. In this paper, we propose a cooperative deep learning (DL)-based positioning methodology that combines these key technologies into a new promising solution for precise 5G positioning. Sparse channel impulse response (CIR) data are used by the positioning infrastructure to extract position-dependent features. We model the problem as a joint task composed of non-line-of-sight (NLOS) identification and position estimation which permits to suitably handle geometrical location measurements and channel fingerprints. The network of base stations (BSs) automatically steers between egocentric (in case of NLOS) and cooperative (for LOS) positioning mode. We perform extensive standard-compliant simulations in a 5G urban micro (UMi) vehicular scenario obtained by ray-tracing and simulation of urban mobility (SUMO) software. Results show that the proposed cooperative DL architecture is able to outperform conventional geometrical positioning algorithms operating in LOS by 47%, achieving a median error of 71 cm on unseen trajectories.

Index Terms—Cooperative positioning, deep learning, 5G, channel impulse response, cooperative intelligent transport systems.

I. INTRODUCTION

THE currently deployed release of 5th generation (5G) cellular networks, 3rd generation partnership project (3GPP) Release 16, lays the foundations for future location awareness systems that are foreseen to extend the location services (LCS) beyond regulatory use cases (i.e., emergency and lawful interception) to include roaming and commercial capabilities [1], [2], [3], [4]. These systems are made possible by the use of higher frequencies, i.e., millimeter waves (mmWave), with enlarged bandwidth, and massive multiple-input multiple-output (mMIMO) technologies [5], which enhance radio access technology (RAT)-dependent

Manuscript received 15 February 2023; revised 13 August 2023; accepted 13 August 2023. Date of publication 9 October 2023; date of current version 22 November 2023. This work was supported in part by a Ph.D. Grant from the Ministry of the Italian Government Ministero dell’Istruzione, dell’Università e della Ricerca (MIUR), and in part by the project Centro Nazionale per la Mobilità Sostenibile (MOST), funded by the Italian Ministry of University and Research under the PNRR funding program. (Corresponding author: Bernardo Camajori Tedeschini.)

Bernardo Camajori Tedeschini is with Dipartimento di Elettronica, Informazione e Bioingegneria (DEIB), Politecnico di Milano, 20133 Milan, Italy (e-mail: bernardo.camajori@polimi.it).

Monica Nicoli is with Dipartimento di Ingegneria Gestionale (DIG), Politecnico di Milano, 20133 Milan, Italy (e-mail: monica.nicoli@polimi.it).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/JSAC.2023.3322795>.

Digital Object Identifier 10.1109/JSAC.2023.3322795

measurements in accordance with 3GPP standards [6], [7]. However, future releases of 5G and beyond, such as 5G-Advanced from Release 18, will face challenges in achieving the centimeter-level absolute accuracy required by the most demanding 5G use cases due to higher path loss and frequent blockages [8], [9].

Legacy solutions such as least squares (LS) multi-lateration/angulation may struggle to effectively handle situations with low signal-to-noise ratio (SNR), multipath ambiguities, and particularly non-line-of-sight (NLOS) conditions [10], [11]. A potential solution is already foreseen by a novel paradigm called integrated sensing and communication (ISAC) [12], [13], [14]. Specifically, 5G base stations (BSs) (also known as gNodeBs) can natively support ISAC through the use of a joint signal processing framework, allowing for a more efficient utilization of spectrum resources. The integration of communication and sensing features on the same hardware platform is, at present time, not yet been commercialized. Nevertheless, synthetic datasets are emerging [15], [16], [17] with the clear objective of permitting the design of novel artificial intelligence (AI) and machine learning (ML) algorithms which will play a fundamental role in next-generation networks [18], [19], [20], [21], [22].

In the context of cooperative intelligent transport systems (C-ITS), connected automated vehicles (CAV) rely on ML, and more specifically deep learning (DL), for various functions, including identifying and segmenting objects within images, controlling the vehicle and avoiding collisions, and determining the most efficient route [23], [24]. Because of the high complexity of such tasks (including precise positioning), urban areas may install computing units, namely roadside units (RSUs), on busy roads that CAV will be able to use to offload part of the computing activities [25], [26]. Cooperation between nearby RSUs, here referred to as BSs, is of paramount importance for enabling network-based precise localization [27], [28]. A key aspect is that, in 5G industrial applications such as automated driving, BSs often have access to a large amount of historical channel state information (CSI) data that is continuously received from geolocalized vehicles [29], [30].

While having a perfect knowledge of the overall CSI is unfeasible, e.g., accurate delays, angle of arrivals (AoAs) and power gains, the usage of estimates of the channel impulse response (CIR) can be adopted by ML approaches to learn relevant information about the environment and its propagation characteristics [31]. It is important to observe that not only LOS but also NLOS CIRs embed significant location information, though embedded in distinct peculiar distributions. Therefore, CIRs can be exploited in both cases

for positioning goals. Distinguishing between LOS/NLOS conditions is extremely important since the derived measurements are fundamentally different. In case of LOS condition, the extracted features can be combined by a cooperative set of BSs as in conventional geometric methods in order to enhance satellite positioning. On the contrary, NLOS features represent a real challenge for the system since they are related to a specific environment, acting as area-specific fingerprints. A complete solution for both situations has yet to be found. Therefore, in this paper, we propose a cooperative DL solution for the joint problem of NLOS identification and 3D position estimation that takes as input the high-dimensional sparse CIRs. We model the problem of NLOS position estimation as an egocentric system at each BS, while a cooperative architecture is proposed for LOS conditions.

Table I and II list the main abbreviations and notations used throughout the paper, respectively.

A. Paper Organization

The remainder of this article is structured as follows. Sec. II presents an overview of the state of the art on wireless localization with ML and DL. Sec. III describes the system model, the MIMO-orthogonal frequency division multiplexing (OFDM) channel and related angle-delay channel power matrix (ADCPM) adopted as input for ML-based positioning. In Sec. IV, we discuss the proposed supervised ML setting and the DL model stored in each BS for the joint NLOS detection-positioning task prediction. Sec. V extends the model to be compliant with a cooperative architecture based on a set of collaborative BSs for positioning purposes. Section VI provides information on the simulated 5G scenario and compares the proposed method with conventional positioning algorithms. Finally, in Section VII, we draw the conclusions.

II. RELATED WORKS AND CONTRIBUTIONS

A. Early Works on NLOS Detection and Positioning

First works on NLOS identification and localization with ML were applied to ultra wide-band (UWB) systems. They typically used hand-crafted features of the channel, such as energy, maximum power, rise time, mean excess delay, root-mean-square delay spread, and kurtosis, as inputs [32], [33], [34], [35], [36]. The most commonly used ML models in these studies were Gaussian processes (GPs), support vector machines (SVMs), and relevance vector machines (RVMs). While achieving good results, these methods could not fully exploit the ML potential as they heavily relied on a pre-defined and limited set of features that could not express the whole location information enclosed in the CSI.

Other methods, mainly based on received signal strength (RSS) fingerprinting, were proposed for precise positioning using Wi-Fi technology [37], [38]. With the advent of MIMO-OFDM systems in IEEE 802.11a/n protocol, allowing for the extraction of CSI from commercial Wi-Fi devices, there has been significant research on wireless positioning, behavioral awareness, and target tracking using CSI. The access to channel information over multiple carriers and antennas gave the possibility to extract detailed information about the propagation of the radio signal, and to learn not only the

position of the user [39], [40], [41] but also information about the environment that shaped such propagation [42]. DL approaches were employed to directly learn the optimal non-linear combination of features and produce the desired NLOS classification or position estimation as output. Studies in this field can be found in [43], [44], [45], and [46], adopting convolutional neural networks (CNN) for feature extraction.

B. DL for High-Precision Positioning

In outdoor conditions, DL-based positioning is a relatively new concept, but with a great potential of achieving high levels of accuracy. Authors in [16] achieve a mean positioning error of 1.5 m by using a cell-specific neural network (NN) in 5G Rel. 15 networks, but considering LOS environments only. NLOS prediction is handled through statistical tests [47] or directly included into the model's prediction [48], [49]. A recent study [50] adopted a variational autoencoder (VAE) to extract features and impose a Gaussian distribution on the latent features for the purpose of binary classification of samples as LOS or NLOS. While the use of an autoencoder (AE) to obtain a compact representation of the channel can yield good results, the reliance on sampling-based methods for prediction makes this approach not suitable for real-time applications.

The employment of full CIR data, especially stacked into image-shapes, has emerged recently as a promising approach. Authors in [51] adopted both CIR (i.e., path power gain, phase and time of flight (ToF)) as well as geographical information (i.e., AoA and angle of departure (AoD)) to predict the user equipment (UE) location. However, they assumed perfect CIR knowledge by ray-tracing, which is hard to achieve under practical conditions. Authors in [52] employed as input the channel frequency response (CFR) matrix computed by practical channel estimation and augmented with additive noise at training time. Despite achieving good results, the CFR matrix does not explicitly express AoA nor ToF of each path and thus may add complexity in feature extraction. A recent work [53] adopted a 3D CNN with inception modules to directly predict the position of a UE from an ADCPM. This approach, however, highly relies on the fingerprint sampling-distance and it is not able to distinguish between LOS and NLOS conditions, treating each position as equal. In other words, there are no distinctions between geometrical features, useful in LOS environments, and merely NLOS position-dependent fingerprints. Furthermore, no works are available in the literature on DL-based location estimation using data collected by multiple cells, i.e., by the cooperation of BS's. Works mainly rely on centralized processing [54] or vehicle-to-vehicle (V2V) communications [55], [56].

C. Contributions

In this paper, we address the problem of precise cooperative positioning in urban environments covered by 5G mm-wave networks and we propose a new DL-based approach that allows to exploit the full potential of wideband space-time CIR for localization. The main contributions are as follows.

- We propose a new method for the extraction of location-related features from the CIR of 5G mmWave

TABLE I
LIST OF ABBREVIATIONS

ACRONYM	DEFINITION
ADCPM	Angle-delay channel power matrix
ADCRM	Angle-delay channel response matrix
AoA	Angle of arrival
AoD	Angle of departure
BS	Base station
CDF	Cumulative density function
CFR	Channel frequency response
CIR	Channel impulse response
CNN	Convolutional neural networks
CSI	Channel state information
C-ITS	Cooperative intelligent transport systems
DL	Deep learning
GDoP	Geometrical dilution of precision
ISD	Inter-site distance
ML	Machine learning
MLP	Multi-layer perceptron
mMIMO	Massive multiple-input multiple-output
NN	Neural network
NLOS/LOS	Non-/Line-of-sight
NLS	Non-linear least squares
SFCRM	Space-frequency channel response matrix
TDoF	Time difference of flight
ToF	Time of flight
UE	User equipment
UPA	Uniform planar array
VAE/AE	Variational/ Autoencoder

TABLE II
LIST OF NOTATIONS

NOTATION	DEFINITION
K	N. of UEs
$\mathcal{S}_{BS,i}$	Set of BSs detecting a same UE at time i
$\mathcal{N}_i^{(j)}$	Set of neighboring BSs for BS j at time i
λ_c	Carrier wavelength
N, M	N. of antennas in vertical and horizontal directions
$d^{(v)}, d^{(h)}$	Antenna spacing in vertical and horizontal directions
N_k	N. of paths between the BS and user k
$\theta_{k,p}, \phi_{k,p}$	Zenith and azimuth AoA for UE k and path p
T_s, T_c, T_g	Sampling, symbol and cyclic prefix interval
N_c	N. of sub-carriers
N_g	Cyclic prefix length in sampling intervals
$r_{k,p}$	Resolvable propagation delay of UE k and path p
$\tau_{k,p}$	Propagation delay for UE k and path p
$\alpha_{k,p}$	Complex channel gain for UE k and path p
$\bar{\alpha}_{k,p}$	Equivalent gain in frequency domain for UE k and path p
$\nu_{k,p}$	Doppler frequency shift for UE k and path p
$\mathbf{q}_k(\tau)$	Baseband CIR of UE k
$\mathbf{h}_{k,\ell}$	CFR for UE k at the sub-carrier ℓ
$\mathbf{H}_k, \mathbf{G}_k, \mathbf{P}_k$	SFCRM, ADCRM and ADCPM, respectively
$\mathbf{x}_i^{(j)}, \hat{\mathbf{x}}_i^{(j)}$	ADCPM input and related estimate at time i for BS j
$\mathbf{z}_i^{(j)}$	Latent features of sample at time i and BS j
$\mathbf{u}_i, \hat{\mathbf{u}}_i$	Real and estimated UE position at time i
$\hat{p}_{s,i}^{(j)}$	LOS probability at time i and BS j
\mathbf{W}	NN parameters
N_b	N. of samples per batch

OFDM MIMO links. The channel is parameterized in terms of ADCPM matrices, typically sparse, that can be treated as images and used to capture the location-dependent features. The sparse ADCPM directly includes the AoAs and ToFs and is fully compliant with the 3GPP Release 16 standard, providing CSI data that is representative of real 5G communication.

- We model the problem of NLOS identification and position estimation as a joint task, proposing a novel loss function to simultaneously learn a compact representation of the channel, i.e., latent features, and maximize the log-likelihood of the joint task.
- We develop a DL model for network-based localization that automatically steers between ego-positioning in NLOS environments and cooperative-positioning in LOS situations, exploiting neighbors BSs predictions to enhance the location accuracy.
- We simulate a realistic C-ITS scenario in an urban environment. The simulated network is fully compliant with the 5G standard [57] and provides realistic outdoor conditions through the use of Matlab ray-tracing software. We simulate multiple trajectories of vehicles, i.e., UEs, created with simulation of urban mobility (SUMO) software [58]. Simulations are used to assess the performance of the DL-based method, showing significant gains over conventional techniques.

Notation

We denote with $j = \sqrt{-1}$ the imaginary unit. Column vectors and matrices are denoted by lower- and upper-case characters, respectively. Matrix conjugation, transposition and conjugate transposition are indicated as \mathbf{A}^H , \mathbf{A}^T and \mathbf{A}^* ,

respectively. We indicate the Hadamard and Kronecker product between two matrices with \odot and \otimes , respectively. The symbol $\mathbb{E}[\cdot]$ is used for expectation of random variable, whereas \mathbb{C} and \mathbb{R} are the set of complex and real numbers, respectively. We denote with $\mathcal{N}(x; \mu, \sigma^2)$ the distribution of a Gaussian random variable x with mean μ and standard deviation σ . $\delta(\cdot)$ and $\delta[\cdot]$ indicate the Dirac delta and Kronecker functions, respectively, while $\lfloor x \rfloor$ represents the largest integer not greater than x .

III. SYSTEM MODEL

In this section, we define the channel model and the location-related channel fingerprints that will be used by the proposed DL positioning algorithms in Sec. IV. For simplicity of notation, we here report the multi-user single-BS case, leaving the extension to a set of cooperative BSs in Sec. V.

A. Channel Model

Let us consider a wide-bandwidth multi-user MIMO-OFDM system operating at carrier wavelength λ_c where K UEs communicate with a BS in uplink direction. The UE is equipped with an omni-directional antennas, whereas a uniform planar array (UPA) with $N \times M$ isotropic antenna elements (i.e., N and M elements in the vertical and horizontal directions, respectively) is installed in the cell panel of the BS. The antenna elements are separated by a distance $d^{(h)}$ and $d^{(v)}$ in the horizontal and vertical direction, respectively. A multipath channel with N_k paths is present between the BS and the UE k . The overall scenario is represented in Fig. 1, where the generic p -th path of the k -th UE channel is represented, jointly with the direction of arrivals (DoAs) of the signal impinging

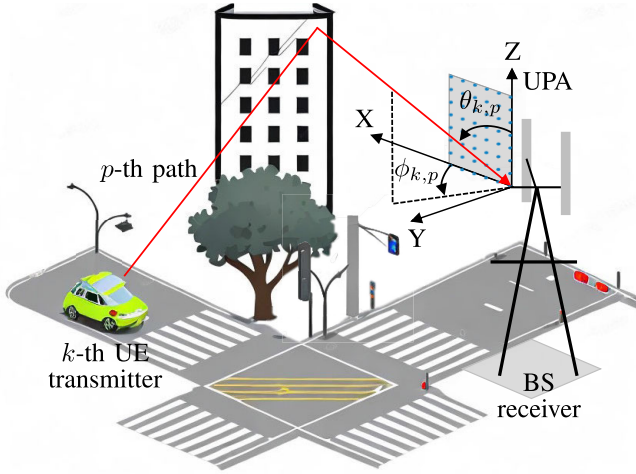


Fig. 1. BS receiving the uplink signal from the k -th UE through an UPA. The DoA of the p -th path is composed by the zenith angle $\theta_{k,p}$ and the azimuth angle $\phi_{k,p}$.

the antenna panel, composed by a zenith angle $0 \leq \theta_{k,p} \leq \pi$ and an azimuth angle $0 \leq \phi_{k,p} \leq \pi$.

We consider an OFDM modulation with sampling interval T_s , N_c sub-carriers and symbol duration $T_c = N_c T_s$. The ℓ -th sub-carrier has frequency $f_\ell = \frac{\ell}{T_c}$, $\ell = 0, \dots, N_c - 1$ and we assume that the cyclic-prefix duration $T_g = N_g T_s$ is greater than the maximum channel delay among all UEs, τ_{MAX} , while N_g is the number of sampling interval that composes a guard interval. The temporally resolvable propagation delay related to the p -th path and k -th UE is indicated with $r_{k,p} = \lfloor \frac{\tau_{k,p}}{T_s} \rfloor$. According to this notation, we model the baseband CIR of user k as [59]:

$$\mathbf{q}_k(\tau) = \sum_{p=1}^{N_k} a_{k,p} e^{-j2\pi(\frac{d_{k,p}}{\lambda_c} - \nu_{k,p}\tau_{k,p})} \mathbf{e}(\theta_{k,p}, \phi_{k,p}) \delta(\tau - \tau_{k,p}), \quad (1)$$

where $\alpha_{k,p} = a_{k,p} e^{-j2\pi(\frac{d_{k,p}}{\lambda_c} - \nu_{k,p}\tau_{k,p})}$ is the complex gain of p -th path which also includes the frequency shift due to Doppler $\nu_{k,p}$ and has average power $\sigma_{k,p}^2 = \mathbb{E}[|a_{k,p}|^2]$, $d_{k,p} = c\tau_{k,p}$ is the traveled distance (with c being the speed of light in air), $\delta(\tau - \tau_{k,p})$ is the delta Dirac function and $\mathbf{e}(\theta_{k,p}, \phi_{k,p}) \in \mathbb{C}^{MN}$ the array response vector [60]. We assume that over the time interval τ_{MAX} , the rotation due to the Doppler is almost constant.

Considering sampling time with rate $1/T_s$ and assuming each different path independent and wide sense stationary [59], we can write the CFR at the ℓ -th sub-carrier as [61]:

$$\mathbf{h}_{k,\ell} = \sum_{p=1}^{N_k} \bar{\alpha}_{k,p} \mathbf{e}(\theta_{k,p}, \phi_{k,p}), \quad (2)$$

where $\bar{\alpha}_{k,p} = \alpha_{k,p} e^{-j2\pi\tau_{k,p}f_\ell}$ are the equivalent channel gains in the frequency domain. Concatenating the different CFRs at each sub-carrier, we obtain the space-frequency channel response matrix (SFCRM):

$$\mathbf{H}_k = [\mathbf{h}_{k,0} \ \mathbf{h}_{k,1} \ \dots \ \mathbf{h}_{k,N_c-1}] \in \mathbb{C}^{MN \times N_c}, \quad (3)$$

which will be used in the next section for ADCPM extraction.

B. ADCPM Location Fingerprints

For location estimation, it is convenient to convert the channel response to the angle-delay domain, where the identification of the LOS component (if present) and secondary NLOS macro-paths is facilitated. In fact, in LOS condition, AoA and ToF can be effectively used to localize a UE thanks to their geometric relationship with the location. At the same time, in NLOS circumstances, different surrounding environments hold different channel parameters, acting as location-dependent features (or fingerprints). Therefore, we transform the SFCRM in (3) into an angle-delay channel response matrix (ADCRM) by introducing the phase-shifted discrete Fourier transform (DFT) matrices $\mathbf{V}_M \in \mathbb{C}^{M \times M}$ and $\mathbf{V}_N \in \mathbb{C}^{N \times N}$, where $[\mathbf{V}_M]_{i,j} = \frac{1}{\sqrt{M}} e^{-j2\pi \frac{i(j-\frac{M}{2})}{M}}$ and $[\mathbf{V}_N]_{i,j} = \frac{1}{\sqrt{N}} e^{-j2\pi \frac{i(j-\frac{N}{2})}{N}}$. We denote with $\mathbf{F} \in \mathbb{C}^{N_c \times N_g}$ the matrix formed by the first N_g columns of N_c dimensional unitary DFT matrix where $[\mathbf{F}]_{i,j} = \frac{1}{\sqrt{N_c}} e^{-j2\pi \frac{ij}{N_c}}$. Finally, we compute the ADCRM as [53]:

$$\mathbf{G}_k = \frac{1}{\sqrt{MNN_c}} (\mathbf{V}_M^H \otimes \mathbf{V}_N^H) \mathbf{H}_k \mathbf{F}^* \in \mathbb{C}^{MN \times N_g}, \quad (4)$$

where \mathbf{F}^* and $(\mathbf{V}_M^H \otimes \mathbf{V}_N^H)$ project the SFCRM into the delay and angle domain, respectively.

For positioning purposes, we propose to use power-angle-delay profile, represented by the ADCPM:

$$\mathbf{P}_k = \mathbb{E}[\mathbf{G}_k \odot \mathbf{G}_k^*] \in \mathbb{C}^{MN \times N_g}, \quad (5)$$

where $[\mathbf{P}_k]_{i,j} = \mathbb{E}[|[\mathbf{G}_k]_{i,j}|^2]$. It can be shown indeed that for N , M and $N_g \rightarrow \infty$, the ADCPM approximates a sparse matrix with elements $[\mathbf{P}_k]_{i,j}$ matching the i -th AoA and the j -th ToF [53]:

$$\lim_{M,N,N_g \rightarrow \infty} [\mathbf{P}_k]_{i,j} = \sum_{p=1}^{N_k} \sigma_{k,p}^2 \delta[i - m_{k,p}N - n_{k,p}] \delta[j - r_{k,p}], \quad (6)$$

where $r_{k,p}$ indicates the index of the j -th ToF and $m_{k,p}N + n_{k,p}$ refers to the index of the i -th AoA. Therefore, the statistical information of the ADCPM enables the learning by the DL model of the location-dependent characteristics, delivering steady and trustworthy fingerprints for positioning.

C. DL Model Input

We propose to employ the ADCPM as a set of measurements for location estimation. This sparse matrix provides indeed a visual image of the multipath configuration in the power-angle-delay domain from which a DL model such as CNN can extract the most representative location-related features. Additionally, since the first layers of CNN are often sparse and collect the highly discriminating features, the CNN eases features extraction from the ADCPM sparse channel matrix [62]. Moreover, the ADCPM embodies all the relevant information (i.e., ToF, AoA and RSS for every path) with small storage and low complexity characteristics thanks to the channel sparsity. To highlight this aspect, in Fig. 2 we show an example of ADCPM \mathbf{P}_k composed by $N_g = 352$ delay

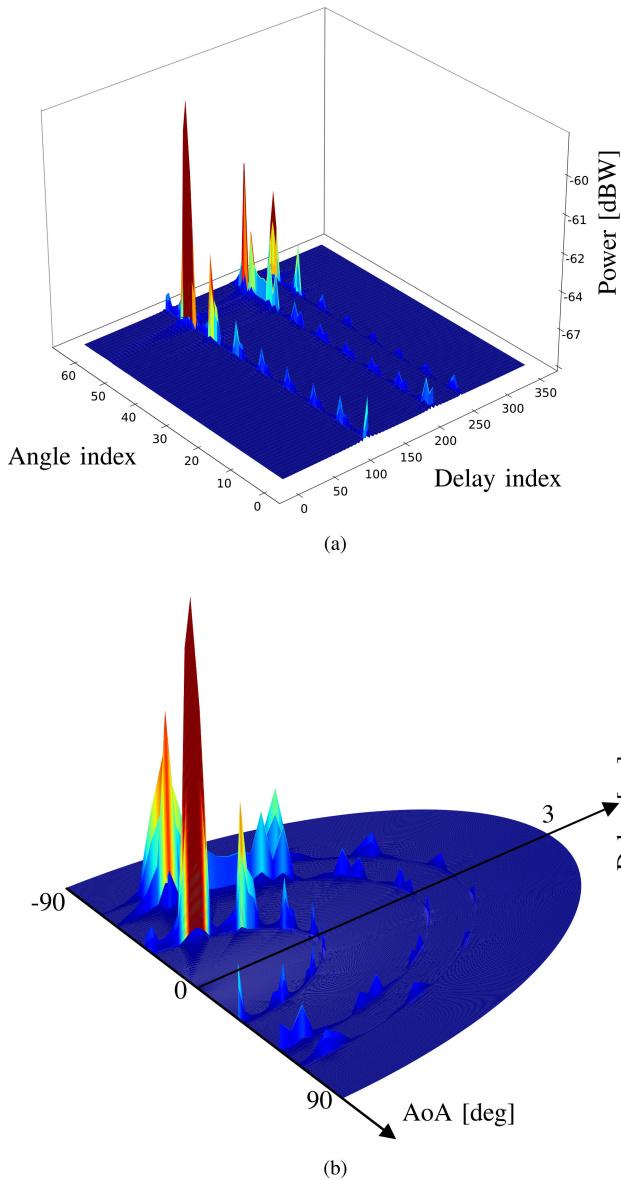


Fig. 2. Example of sparse channel power-delay-angle profile encoding the UE location, represented by an ADCPM with $N_g = 352$ temporal samples and $NM = 64$ spatial samples (resulting from $N = M = 8$ antennas) (a) and in the transformed angle-delay domain (b).

samples and $MN = 64$ angle directions. The actual model input can be seen in Fig. 2a, while the polar representation of the ADCPM with physical AoAs and ToFs is in Fig. 2b. Even in the absence of a large number of antennas or a high sample resolution, the matrix sparsity is clearly visible. Therefore, we propose to use the ADCPM as input to the model and we denote the i -th sample with $\mathbf{x}_i = \mathbf{P}_i$, dropping the index k for ease of notation.

IV. METHODOLOGY: SINGLE-BS LOCALIZATION

In this section, we first tackle the localization problem in a supervised setting in which a single BS has to locate the UE. We propose a DL model (Sec. IV-A) and a loss function for the joint task of position estimation and LOS identification (Sec. IV-B). The approach will then be extended to the multi-BS, i.e., cooperative, case in Sec. V.

A. Deep Learning Model

We assume a supervised ML setting in which both a regression and a classification problem have to be solved by a single BS. The regression problem refers to the estimation of the UE's position, while the classification problem concerns the LOS/NLOS identification. We define the training dataset as $\mathcal{S}^{\text{train}} = \{(\mathbf{x}_i, \mathbf{u}_i, s_i)\}_{i=1}^{N_{\text{train}}}$, where $\mathbf{x}_i = \mathbf{P}_i$ denotes the i -th input sample (i.e., ADCPM channel response), \mathbf{u}_i represents the 3D position and $s_i \in \{0, 1\}$ is the Boolean identifier of the sight condition. To validate the performances, we also hold a similar test dataset $\mathcal{S}^{\text{test}}$ composed of N_{test} samples.

We note that the regression and classification tasks are two interrelated problems that must be addressed accordingly. In fact, if the UE lies in a LOS condition, its position can be directly computed from the geometrical features of the direct path (i.e., ToF, AoA and mean power), while NLOS typically requires a finer training based on more complex multipath fingerprints.

To extract such key features from the ADCPM samples \mathbf{x}_i , we propose to employ an AE, with structure represented in Fig. 3. The encoder $\mathcal{E}(\cdot)$ is used to produce the hidden (or latent) features \mathbf{z}_i (including the location-related information embedded in the channel), while the decoder $\mathcal{D}(\cdot)$ tries to reconstruct the input samples obtaining $\hat{\mathbf{x}}_i$. The AE is designed so as to minimize a metric of the reconstruction error $\|\mathbf{x}_i - \hat{\mathbf{x}}_i\|_2^2$ [63], making the model able to reconstruct the input \mathbf{x}_i from the low-dimensional data \mathbf{z}_i . This guarantees that \mathbf{z}_i contains all the necessary and sufficient information to accomplish the specific task.

The tasks herein considered are position regression and LOS identification. Therefore, in principle, two NNs with \mathbf{z}_i as input, should be sufficient. However, due to the major difference between position estimation in LOS and NLOS conditions, we propose to use three separate NNs: one for LOS identification, one for position regression in LOS conditions and one for position regression in NLOS conditions. Given the overall model with parameters defined as \mathbf{W} , the output of the NNs is respectively: $\hat{p}_{s,i} \in [0, 1]$, $\hat{\mathbf{u}}_{\text{LOS},i} \in \mathbb{R}^3$ and $\hat{\mathbf{u}}_{\text{NLOS},i} \in \mathbb{R}^3$. Specifically, $\hat{p}_{s,i} = p(s_i)$ predicts the probability that the sample \mathbf{x}_i relates to a UE in LOS condition, while $\hat{\mathbf{u}}_{\text{LOS},i}$ and $\hat{\mathbf{u}}_{\text{NLOS},i}$ predict the 3D position of the UE in LOS and NLOS settings, respectively. The overall position estimate, indicated with $\hat{\mathbf{u}}_i$, is obtained by applying a threshold Γ to $\hat{p}_{s,i}$ and considering only $\hat{\mathbf{u}}_{\text{LOS},i}$ or $\hat{\mathbf{u}}_{\text{NLOS},i}$ according to the result:

$$\hat{\mathbf{u}}_i = \Gamma(\hat{p}_{s,i})\hat{\mathbf{u}}_{\text{LOS},i} + \Gamma(1 - \hat{p}_{s,i})\hat{\mathbf{u}}_{\text{NLOS},i}. \quad (7)$$

The position estimates in LOS and NLOS are then adopted in the loss function for the model training, described in the next section.

B. Loss Function for Joint Sight Detection and Localization

In order to train the model, we have to define a loss function whose objective is to enable learning the correct representation of the latent features and, at the same time, jointly estimating the sight condition and the location. To this aim, we first treat separately the classification and regression tasks and then we combine them with an overall objective

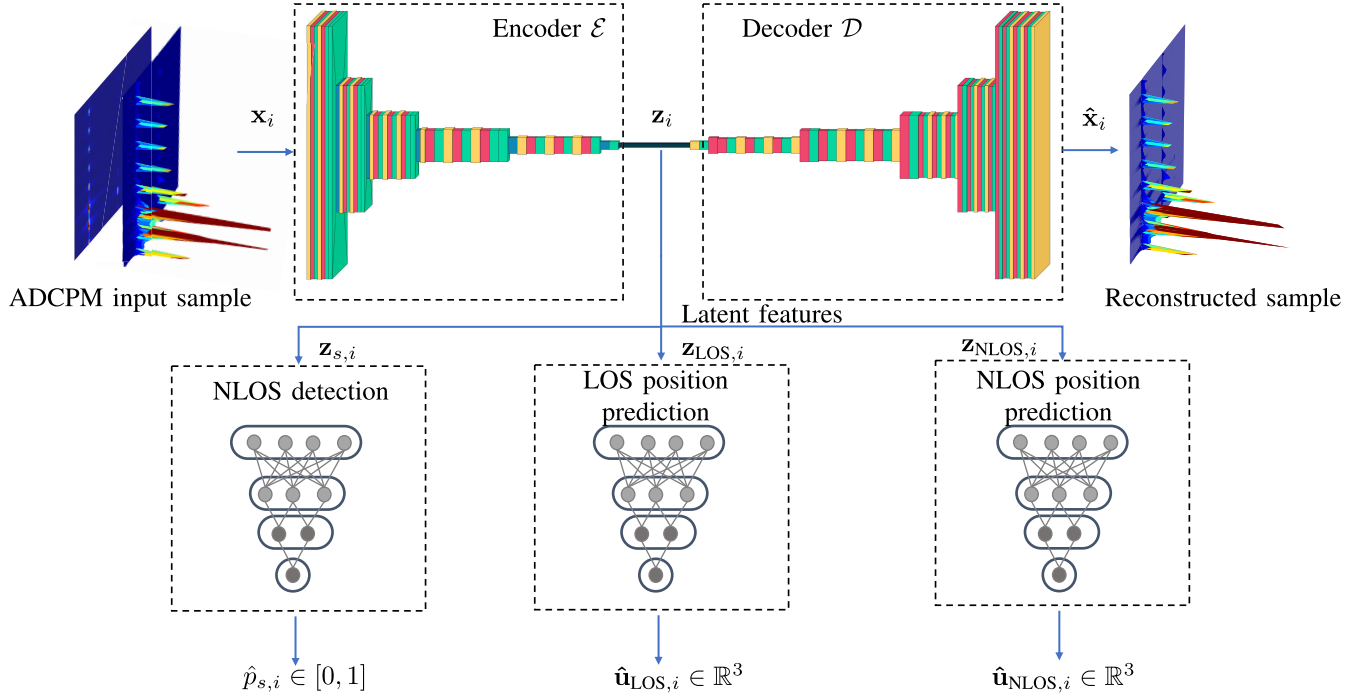


Fig. 3. Overview of the proposed model composed by an autoencoder (AE) for feature extraction, and 3 neural networks (NNs) for NLOS classification and position estimation in both LOS and NLOS conditions.

function. For the classification task, we propose a discriminative probabilistic approach, namely the maximum likelihood [64], where we directly define the posterior conditional probability $p(s_i|\mathbf{x}_i)$ using a parametric model \mathbf{W} , i.e., $p(s_i|\mathbf{x}_i) = p(s_i|\mathbf{x}_i, \mathbf{W})$. Subsequently, we maximize the likelihood of the model $p(s_i|\mathbf{x}_i, \mathbf{W})$, by optimizing the parameters \mathbf{W} over the training set. For the specific binary classification problem, the likelihood function is:

$$p(s_i|\mathbf{x}_i, \mathbf{W}) = \hat{p}_{s,i}\delta[s_i - 1] + (1 - \hat{p}_{s,i})\delta[s_i]. \quad (8)$$

For the regression task, we reiterate the discriminative approach with the constraint of belonging to a specific condition, either LOS or NLOS. Again, this is done considering that the two circumstances hold different statistics. We assume that in the LOS case, the target variable \mathbf{u}_i (i.e., the location) is Gaussian distributed with a deterministic mean $\hat{\mathbf{u}}_{\text{LOS},i}(\mathbf{x}_i, \mathbf{W})$:

$$\mathbf{u}_i = \hat{\mathbf{u}}_{\text{LOS},i}(\mathbf{x}_i, \mathbf{W}) + \boldsymbol{\epsilon}_{\text{LOS}}, \quad (9)$$

where $\boldsymbol{\epsilon}_{\text{LOS}}$ is a zero-mean random Gaussian variable with covariance $\mathbf{C}_{\text{LOS}} = \mathbf{I}_3\sigma_{\text{LOS}}^2$, and \mathbf{I}_3 denoting the 3×3 identity matrix. The likelihood function is thus given by:

$$p(\mathbf{u}_i|s_i = 1, \mathbf{x}_i, \mathbf{W}, \mathbf{C}_{\text{LOS}}) = \mathcal{N}(\mathbf{u}_i; \hat{\mathbf{u}}_{\text{LOS},i}(\mathbf{x}_i, \mathbf{W}), \mathbf{C}_{\text{LOS}}). \quad (10)$$

Similarly, the likelihood function of the regression problem in NLOS conditions can be written as:

$$p(\mathbf{u}_i|s_i = 0, \mathbf{x}_i, \mathbf{W}, \mathbf{C}_{\text{NLOS}}) = \mathcal{N}(\mathbf{u}_i; \hat{\mathbf{u}}_{\text{NLOS},i}(\mathbf{x}_i, \mathbf{W}), \mathbf{C}_{\text{NLOS}}). \quad (11)$$

Note that the requirement on the mono-modality of \mathbf{u}_i , both in LOS and NLOS, can be easily disregarded by applying for example a *mixture of experts* model [65].

Combining (8), (10) and (11), we define the joint likelihood for the variables (\mathbf{u}_i, s_i) as:

$$p(\mathbf{u}_i, s_i|\mathbf{x}_i, \mathbf{W}, \mathbf{C}_{\text{LOS}}, \mathbf{C}_{\text{NLOS}}) = \hat{p}_{s,i}\delta[s_i - 1]\mathcal{N}(\mathbf{u}_i; \hat{\mathbf{u}}_{\text{LOS},i}(\mathbf{x}_i, \mathbf{W}), \mathbf{C}_{\text{LOS}}) + (1 - \hat{p}_{s,i})\delta[s_i]\mathcal{N}(\mathbf{u}_i; \hat{\mathbf{u}}_{\text{NLOS},i}(\mathbf{x}_i, \mathbf{W}), \mathbf{C}_{\text{NLOS}}). \quad (12)$$

A representation of the distribution can be found in Fig. 4. In order to maximize the likelihood, we insert the negative log-likelihood in the loss function. It can be shown that the the negative log-likelihood of a batch of independent samples can be written as (see Appendix A):

$$\begin{aligned} L_{\text{task}} &\simeq -\log \prod_{i=1}^{N_b} p(\mathbf{u}_i, s_i|\mathbf{x}_i, \mathbf{W}, \mathbf{C}_{\text{LOS}}, \mathbf{C}_{\text{NLOS}}) \\ &= \sum_{i=1}^{N_b} \left\{ s_i \cdot \left[-\log(\hat{p}_{s,i}) + \frac{\|\mathbf{u}_i - \hat{\mathbf{u}}_{\text{LOS},i}\|_2^2}{2\sigma_{\text{LOS}}^2} \right] \right. \\ &\quad \left. + (1 - s_i) \cdot \left[-\log(1 - \hat{p}_{s,i}) + \frac{\|\mathbf{u}_i - \hat{\mathbf{u}}_{\text{NLOS},i}\|_2^2}{2\sigma_{\text{NLOS}}^2} \right] \right\}, \end{aligned} \quad (13)$$

where N_b is the batch size. Whenever the LOS condition occurs, the second right-hand side of (13) does not contribute to the likelihood, whereas if NLOS holds, only the first right hand side is considered.

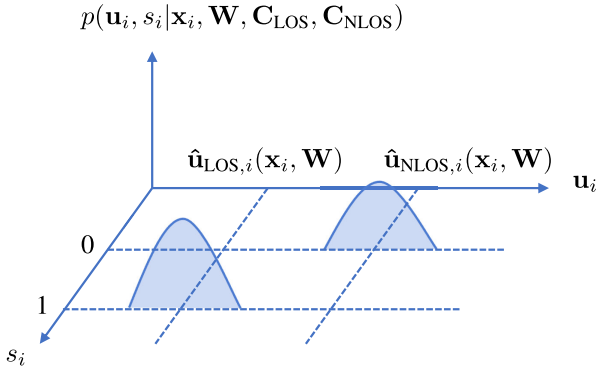


Fig. 4. Joint likelihood distribution of (\mathbf{u}_i, s_i) . Here, for sake of simplicity, the multi-dimensional Gaussian distributions are represented as uni-variate.

Finally, the complete adopted loss function is:

$$\begin{aligned} L_{\text{tot}} = & \frac{w_{\text{rec}}}{N_b} \sum_{i=1}^{N_b} \|\mathbf{x}_i - \hat{\mathbf{x}}_i\|_2^2 \\ & + \frac{1}{N_b} \sum_{i=1}^{N_b} \left\{ w_s s_i \cdot [-\log(\hat{p}_{s,i}) + w_u \|\mathbf{u}_i - \hat{\mathbf{u}}_{\text{LOS},i}\|_2^2] \right. \\ & \left. + (1 - s_i) \cdot [-\log(1 - \hat{p}_{s,i}) + w_u \|\mathbf{u}_i - \hat{\mathbf{u}}_{\text{NLOS},i}\|_2^2] \right\} \end{aligned} \quad (14)$$

where w_{rec} regulates the sample reconstruction, w_s compensates the unbalances between LOS and NLOS samples and w_u controls the uncertainty of the model on the position estimation and includes both σ_{LOS}^2 and σ_{NLOS}^2 .

V. METHODOLOGY: COOPERATIVE LOCALIZATION

In this section, we extend the approach for UE positioning to a multi-BS scenario. We present at first the proposed cooperative AE architecture and then the cooperative training procedure that can be used in practice to deploy the set of BSs composing the localization infrastructure.

A. Cooperative Architecture

We propose a cooperative architecture where each BS adopts a cell-specific DL model. The main assumptions behind the proposed architecture are the following. First, each BS is able to evaluate whether a UE is in LOS or NLOS condition based on the observed ADCPM and the procedure described in Section IV. Second, in case of NLOS, a position estimate is obtained by the BS by means of a previous fingerprint training procedure. Third, the latent geometrical features adopted by the LOS position estimation are combined in order to get a more accurate inference. This is somehow intuitive if we consider the latent features as a sort of non-linear combination of ToF and AoA measurements that each BS can share with the neighbors' cells.

Let us denote by $\mathcal{S}_{\text{BS},i} = \{1, \dots, S_{\text{BS},i}\}$ the set of BSs which detect a UE at timestep i , and gather a batch of samples $\mathbf{X}_i = \{\mathbf{x}_i^{(j)}\}_{j=1}^{S_{\text{BS},i}}$. Note that the number of detected BSs can vary at every timestep, without being constrained to detect a minimum number of BSs, as opposed to classical geometrical approaches. As illustrated in the pseudo-code in

Algorithm 1 Cooperative Position Inference

Input: sample $\mathbf{x}_i^{(j)}$, neighbors' BS $\mathcal{N}_i^{(j)}$ ▷ Run on BS j at timestep i

Output: estimated position $\hat{\mathbf{u}}_i^{(j)}$

- 1: Encode sample $\mathbf{x}_i^{(j)}$ through encoder \mathcal{E} : $\mathbf{z}_i^{(j)} \leftarrow \mathcal{E}(\mathbf{x}_i^{(j)})$
- 2: Initialize the latent features $\mathbf{z}_{s,i}^{(j)} \leftarrow \mathbf{z}_i^{(j)}$, $\mathbf{z}_{\text{LOS},i}^{(j)} \leftarrow \mathbf{z}_i^{(j)}$, $\mathbf{z}_{\text{NLOS},i}^{(j)} \leftarrow \mathbf{z}_i^{(j)}$
- 3: Predict probability of LOS condition $\hat{p}_{s,i}^{(j)}$ from $\mathbf{z}_{s,i}^{(j)}$
- 4: **for** $j' = 1, 2, \dots, \mathcal{N}_i^{(j)}$ **do**
- 5: Send $\{\hat{p}_{s,i}^{(j')}, \mathbf{z}_{\text{LOS},i}^{(j')}\}$ to j'
- 6: Receive $\{\hat{p}_{s,i}^{(j')}, \mathbf{z}_{\text{LOS},i}^{(j')}\}$ from j'
- 7: **end for**
- 8: Assign $\hat{\mathbf{z}}_{\text{PLOS},i}^{(j)} \leftarrow \frac{1}{\sum_{j'=1}^{S_{\text{BS},i}} \hat{p}_{s,i}^{(j')}} \sum_{j'=1}^{S_{\text{BS},i}} \hat{p}_{s,i}^{(j')} \mathbf{z}_{\text{LOS},i}^{(j')}$
- 9: Predict LOS position $\hat{\mathbf{u}}_{\text{LOS},i}^{(j)}$ from $\hat{\mathbf{z}}_{\text{PLOS},i}^{(j)}$
- 10: Predict NLOS position $\hat{\mathbf{u}}_{\text{NLOS},i}^{(j)}$ from $\mathbf{z}_{\text{NLOS},i}^{(j)}$
- 11: Estimate position $\hat{\mathbf{u}}_i^{(j)} \leftarrow \Gamma(\hat{p}_{s,i}^{(j)}) \hat{\mathbf{u}}_{\text{LOS},i}^{(j)} + \Gamma(1 - \hat{p}_{s,i}^{(j)}) \hat{\mathbf{u}}_{\text{NLOS},i}^{(j)}$

Algorithm 1, the main idea is that, for position inference, each BS j computes its latent features $\mathbf{z}_i^{(j)}$. At this stage, there is no difference between $\mathbf{z}_i^{(j)}$, $\mathbf{z}_{s,i}^{(j)}$, $\mathbf{z}_{\text{LOS},i}^{(j)}$, and $\mathbf{z}_{\text{NLOS},i}^{(j)}$ and the inference proceeds as in the single-BS method. Then, if the BS predicts a NLOS condition, the latent features extracted by that BS are not combined with other cells (e.g., as a multi-lateration) and position estimation continues according to the $\hat{\mathbf{u}}_{\text{NLOS},i}^{(j)}$ prediction. On the contrary, if we are in LOS condition, then the latent LOS features $\mathbf{z}_{s,i}^{(j)}$ are exchanged with the neighbors' cells, defined with $\mathcal{N}_i^{(j)} = \mathcal{S}_{\text{BS},i} \setminus \{j\}$. After averaging the latent LOS features $\mathbf{z}_{s,i}^{(j)}$, the position is estimated with $\hat{\mathbf{u}}_{\text{LOS},i}^{(j)}$.

As an example of cooperative inference, we refer to the scenario shown in Fig. 5 where three BSs detect a UE, here represented by a vehicle. NLOS and LOS links are indicated with red and green dashed arrows, respectively. Since BS j and j'' are in LOS condition with respect to the vehicle, the contribution of $\mathbf{z}_{s,i}^{(j)}$ and $\mathbf{z}_{s,i}^{(j')}$ to the position estimation will be higher than $\mathbf{z}_{s,i}^{(j')}$. The negligible contribution of NLOS BSs is highlighted with red solid arrows while significant latent features are colored in green. This is due to the fact that the probability of LOS condition of BS j' will be low, i.e., $\hat{p}_{s,i}^{(j')} < \hat{p}_{s,i}^{(j)}$ and $\hat{p}_{s,i}^{(j')} < \hat{p}_{s,i}^{(j')}$. The outcome is an increased accuracy on the position estimation which is not or little affected by NLOS BSs.

B. Cooperative Training

For the training of the cooperative set of BSs we propose the following procedure. During the data gathering phase, a UE (e.g., a vehicle) moves along specified trajectories (a-priori divided into LOS and NLOS segments) sending uplink signals, i.e., sounding reference signal (SRS), to every BS in its range. The time instant of the transmission, the coarse position obtained from global navigation satellite systems (GNSS) and

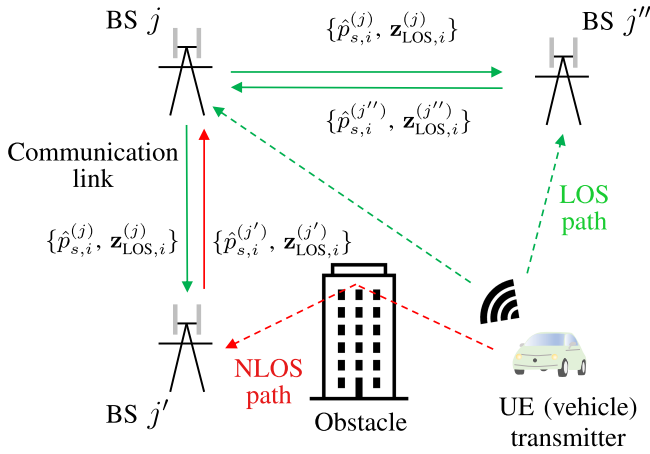


Fig. 5. Cooperative inference scenario composed of three BSs.

the LOS identification are either sent to the BSs as auxiliary data through the communication link, or stored inside the UE for the post-processing. As for the inference, the BSs evaluate the ADCPM samples and store the data. In order to perform back-propagation at BS j , the loss function is computed as:

$$\begin{aligned} L_{\text{tot}}^{(j)} &= \frac{w_{\text{rec}}}{N_b} \sum_{i=1}^{N_b} \left\| \mathbf{x}_i^{(j)} - \hat{\mathbf{x}}_i^{(j)} \right\|_2^2 \\ &+ \frac{1}{N_b} \sum_{i=1}^{N_b} \left\{ w_s s_i^{(j)} \left[w_u \left\| \mathbf{u}_i^{(j)} - \hat{\mathbf{u}}_{\text{LOS},i}^{(j)}(\bar{\mathbf{z}}_{\text{LOS},i}^{(j)}, s_i^{(j)}) \right\|_2^2 - \log(\hat{p}_{s,i}^{(j)}) \right] \right. \\ &\left. + (1 - s_i^{(j)}) \left[w_u \left\| \mathbf{u}_i^{(j)} - \hat{\mathbf{u}}_{\text{NLOS},i}^{(j)}(\mathbf{z}_{\text{NLOS},i}^{(j)}) \right\|_2^2 - \log(1 - \hat{p}_{s,i}^{(j)}) \right] \right\}, \end{aligned} \quad (15)$$

where $\bar{\mathbf{z}}_{\text{LOS},i}^{(j)} = \sum_{j'=1}^{S_{\text{BS},i}} s_i^{(j')} \mathbf{z}_{\text{LOS},i}^{(j')} / \sum_{j'=1}^{S_{\text{BS},i}} s_i^{(j')}$. Note that, to speed up the training phase, a centralized loss computation can be performed in a batch-manner, i.e., in parallel way, using $S_{\text{BS},i}$ as batch size. The rationale behind this approach is that while the models for NLOS position estimation are trained to be BS-specific, the LOS networks can be shared as they have consistent parameters among all the BSs.

VI. PERFORMANCE ASSESSMENT IN A 5G NETWORK

To assess the performance of the proposed cooperative DL positioning system, we employ a data generator based on the 5G new radio (NR) clustered delay line (CDL) channel model [66], which is defined over a bandwidth of 2 GHz in the frequency range from 0.5 GHz to 100 GHz. The radio wave propagation is simulated using a ray-tracing method [67], [68], [69] from the Matlab package, which plots the propagation paths from the UE to the BSs based on the surface geometry from a map file. The ray-tracing method employs the shooting and bouncing rays (SBR) method with up to 10 path reflections [70]. The channel model is then generated by combining all the paths and taking into account the small-scale fading caused by multipath and UE's movement. With this method, adjacent positions will have similar channel characteristics due to the similar scattering environment, ensuring spatial consistency.

TABLE III
EXPERIMENT PARAMETERS

Parameter	Value
Carrier frequency f_c	30 GHz
Bandwidth B	400 MHz
BS noise figure	7 dB
BS antenna configuration	$M = N = 8$, $d^{(h)} = d^{(v)} = 0.5\lambda_c$
BS antenna radiation pattern	Directional, 8 dBi – (Table 6.1.1-5 of [57])
UE noise figure	13 dB
UE antenna configuration	Omni
UE radiation pattern	0 dBi
UE max. TX power	23 dBm
Layout	Hexagonal grid, 19 macro sites, 3 sectors per site, ISD = 200 m
BS clock offset	0 ns (perfect synchronization)
Penetration loss	0 dB
UE mobility	Max 30 km/h
UE height	1.5 m
Min. BS-UE distance (2D)	10 m
BS antenna height	25 m

A. Simulated Scenario

For the experiments, we simulate a 3GPP urban micro (UMi) scenario in a 1000×1000 m area, near the Leonardo's campus of Politecnico di Milano, in Milano, Italy, using the parameters described in [57]. The specific values are summarized in Table III. As shown in Fig. 6, the scenario consists of 19 sites with an inter-site distance (ISD) of 200 m, placed in an hexagonal layout, each equipped with 3 cells and separated by 120 deg in azimuth. Each cell antenna uses an UPA configuration with $N = M = 8$ elements and a downtilt of 15 deg. A macro image of the antenna array pattern can be found in Fig. 7. Each antenna element was defined using specifications in [71], providing a front-to-back ratio of about 30 dB and a maximum gain of 8 dBi.

The UEs move around the region following various routes generated by the SUMO simulator that reproduces the vehicular traffic flow over the considered road network. The simulation runs for 170 seconds and generates up to 50 trajectories, which are sampled every second. Each UE uses a carrier frequency of $f_c = 30$ GHz and a transmission bandwidth of $B = 400$ MHz to transmit 5G standard compliant SRSs to all nearby BSs. Finally, each BS demodulates the signals and obtains a channel estimation, i.e., SFCRM, via the received pilot signals through a LS estimator. The ADCPM is then obtained according to (4) and (5).

B. Positioning Tests

We divided the experiments into offline and online phases. In the offline (training) phase, the UE moves along the trajectories and each BS gathers both LOS and NLOS channel realizations according to the specific UE's position. We considered the training positions as perfectly known (i.e., no error was introduced to the ground truth positions) as we aimed at assessing the lower bound performances of the method. In total, we gathered about $5.9 \cdot 10^4$ samples in 1659 positions only for the training phase. In the online (test) phase, the NLOS position capabilities were verified in the same trajectories but in random positions not adopted in the

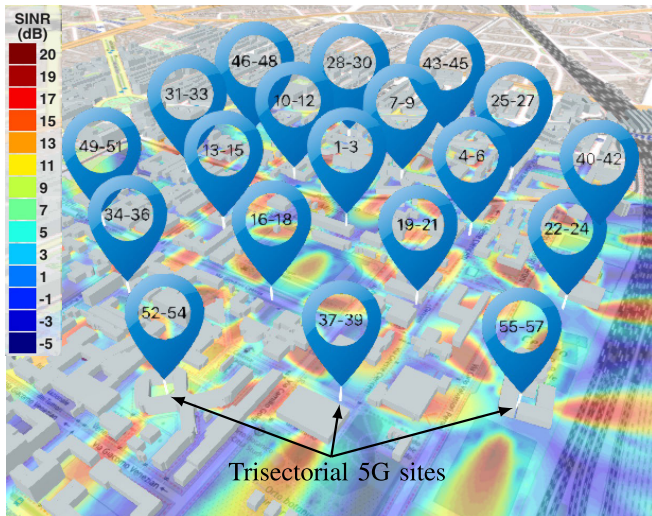


Fig. 6. UMi scenario simulated for performing analysis composed of 19 sites in the area of Politecnico di Milano, Leonardo campus, Milano, Italy.

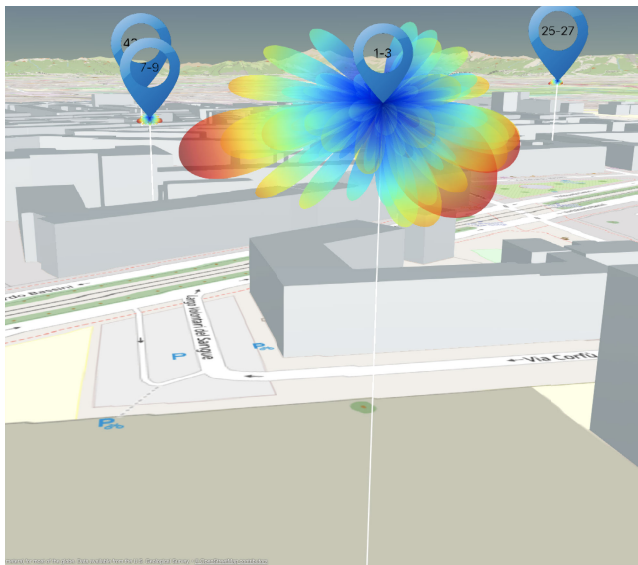


Fig. 7. Antenna array patterns of 3 cells (1-3) forming a site.

training. Specifically, we adopted about $2.5 \cdot 10^4$ samples in 711 positions for NLOS testing. In this way, we assess whether or not, each BS can learn the environment, i.e., the channel characteristics, around it. On the other hand, for LOS positioning, we validated completed different trajectories to analyze the capabilities of estimating the position from learned geometrical features. Here, the total number of LOS tested positions was 867. A representation of the adopted trajectories can be found in Fig. 8. Note that in the training trajectories, i.e., red markers, the number of detected BSs can vary significantly: we measured from 1 up to 13 BSs in the collected samples. The test samples for LOS evaluation, highlighted with purple placeholders, are located in the top-left corner of the map. To avoid biases and improve model convergence, before model training, we standardized with zero mean and unitary standard deviation all the samples, and we shuffled the dataset at each epoch.

The toolbox *Antenna Toolbox* of MATLAB 2022b is used to generate the channel fingerprints for the data points, while the model for training and testing is implemented using Pytorch [72] (v1.12 with Python 3.7.11). The simulations are run on a workstation with an Intel(R) Xeon(R) Silver 4210R CPU @ 2.40 GHz, 96 GB of RAM, and a Quadro RTX 6000 24 GB GPU. The training and testing times refer only to the runtime on Pytorch 1.12. Unless otherwise noted, the model is trained for a total $E = 60$ epochs with a batch size $N_b = 64$. The Adam optimization algorithm [73] is used with an initial learning rate $lr = 10^{-4}$ and momentum values $\beta_1 = 0.9$, $\beta_2 = 0.999$.

For what concerns the hyper-parameters choice of w_{rec} , w_s and w_u , we clarify that their values depend on the specific dataset at hand and thereby tuning has been performed as follows. Starting from w_s , since this parameters regulates how much weight is given to the LOS samples class in order to compensate the class unbalances, it is computed as in weighted cross-entropy loss: $w_s = N_{LOS}/N_{NLOS}$, where N_{LOS} and N_{NLOS} are the number of LOS and NLOS samples in the training batch, respectively. On the contrary, w_{rec} and w_u have been chosen empirically using a grid-approach in $[0.1, 1]$ with step size 0.1 and assigned as: $w_{rec} = 0.1$, $w_u = 0.9$. This can be intuitively explained by two reasons. First, the AE model is much more complex than the MLPs for positioning, leading to a fast drop of the reconstruction error. Second, the number of features in \mathbf{x}_i is larger than the dimension of \mathbf{u}_i , which automatically increases the weight of the reconstruction error with respect to the positioning error. In order to balance these two quantities, we suggest values that satisfy $w_{rec} < w_u$.

C. DL Model Design

The adopted DL model architecture is as follows. The AE is the most critical component as a good feature extraction is essential to enable precise positioning. Driven by the necessity of handling sparse data, i.e., ADCPM input, we select the Segnet architecture [74] where the upsampling layers employ the encoder pool indices to create ad-hoc sparse feature mapping. At testing time, we can completely discard the decoder part as the input reconstruction is only adopted in the training phase to learn the latent features representation.

The choice of the NNs for NLOS classification and position estimation is dictated by the specific task to accomplish. A BS must be able to localize a UE regardless of whether it is in ego mode, i.e., only a single BS detects the UE, or in cooperative mode, i.e., the UE is detected by multiple BSs. In the former case, the latent features should be a non-linear composition of ToF and AoA for each of the multipaths. On the other hand, in the latter case, multiple ToF must be exploited to localize the UE. To assess these concepts, we experiment different multi-layer perceptrons (MLPs) architectures trying to localize a UE with only a ToF and AoA or 3 or more ToF as inputs.

In Fig. 9, we show the test results of the MLP described in Table IV which was trained on a synthetic dataset using as input either ToF and AoA only (Fig. 9a), or 3 measurements of ToFs (Fig. 9b) or 10 ToFs (Fig. 9c). This was done in order to verify the positioning capabilities of the models, i.e.,

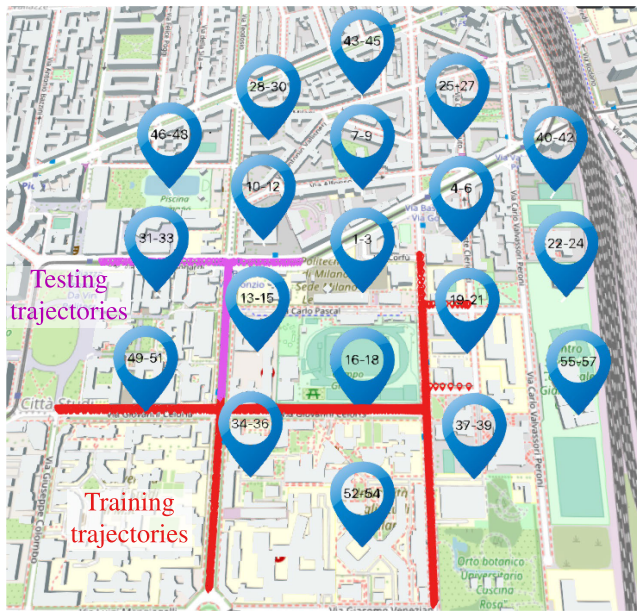


Fig. 8. Scenario comprising training trajectories, i.e., red placeholders, and LOS testing trajectories, i.e., purple placeholders.

to assess the bias of the MLPs for LOS and NLOS position estimation. Since the testing positions (red circles) are never seen in the training phase, we can conclude that the model is able to learn the geometrical meaning of the input features and perform multi-angulation and circular lateration. It is worth noticing that, in the case of cooperative architecture, the BSs' positions (black squares) are never used as input features but they are automatically learned by the training. This can be an advantage in case the coordinates of the BSs are not known or partially known.

Given these results, we employed the MLP in Table IV for the LOS and NLOS position prediction NNs, while the MLP in Table V is adopted for NLOS detection NN. The key difference between these two types of NNs is the size and layer composition. First, we assumed that single supervised NLOS classification is a simpler task if compared with 3D position regression, thus resulting in a different number of layers and neurons. Second, we employed Tanh (instead of ReLu or GeLu) activation functions in the NLOS detection NN since blockage detection should be performed in the fastest and most efficient way possible, as the whole cooperative prediction highly depends on it. On the contrary, GeLu is more computationally expensive but can capture more complex patterns (as needed for regression) due to its smooth and differentiable nature [75]. Finally, both adopt dropout to avoid overfitting.

D. Baselines

For performance assessment, we compare the proposed method with the following algorithms/models:

- Ego DL model: the proposed DL model (Fig. 3) trained with the loss (14) for single-BS positioning. The model does not communicate with any other BS and has to rely only on its prediction.

- Cooperative DL model: the proposed DL-based cooperative architecture described in Sec. V.
- Single-BS ToF-AoA: conventional positioning obtained by a single BS using the ToF estimate obtained from the cross-correlation with the SRS according to 5G NR Rel. 16 and the AoA estimated through multiple signal classification (MUSIC) algorithm [76].
- Multi-BS time difference of flight (TDoF): conventional hyperbolic-multilateration obtained using TDoFs estimated by all the BSs receiving SRS.

E. Simulation Results

1) *Training Convergence*: This first assessment has the aim of verifying the training convergence of the proposed ego DL model, i.e., the correct behavior of the loss function and of the root mean square error (RMSE) performance metric. In Fig. 10, we report the testing results separately for LOS and NLOS samples in the test set, for varying number of training epochs. We notice that the loss function decreases quite rapidly in just 60 epochs and does not show signs of overfitting. This may be due to the fact that the dataset is very large (about 50 GB) and thus it is difficult to memorize every sample. Furthermore, the model regularization, i.e., dropout, helps with this aspect.

Referring to the performance metric, we can see that the NLOS case holds much higher oscillations with respect to the LOS one. While learning geometrical patterns and associated positions is a more standard task, associating NLOS samples with the UE location is much more difficult. The DL model has not only to understand how the environment is configured but also how it could change between the training positions. Finally, we note that the LOS RMSE is slightly worse than the NLOS. This is due to the LOS performance bounds imposed by the physical layer parameters, i.e., the resolution of the ToF (limited by the signal bandwidth) and AoA (by the number of antennas). On the other hand, the NLOS RMSE highly depends on the training position resolution, i.e., the training spatial sampling, as the denser the training points, the better the performances.

2) *Blockage Detection*: This experiment has the objective of assessing the steering model capabilities in discerning LOS and NLOS conditions. The steering is taken into account in both the smart weighting of the geometrical latent features and in the final position estimate, i.e., lines 8 and 11 of Algorithm 1, respectively. To this aim, in Fig. 11 we show the testing accuracy, precision and recall for varying number of training epochs. The numerical results show that the model reaches an accuracy of approximately 85%, validating the capability of the proposed approach of learning the multi-task problem within such a realistic environment. In terms of the parameter Γ , we implemented a conservative yet effective approach, as detailed subsequently. We selected $\Gamma = 0.6$ not with the aim of matching the values of precision and recall, but rather to adjust it towards a lower number of false positives compared to false negatives. This approach is driven by the rationale that if the model is uncertain about the visibility condition, it slightly leans towards the NLOS case. As a result,

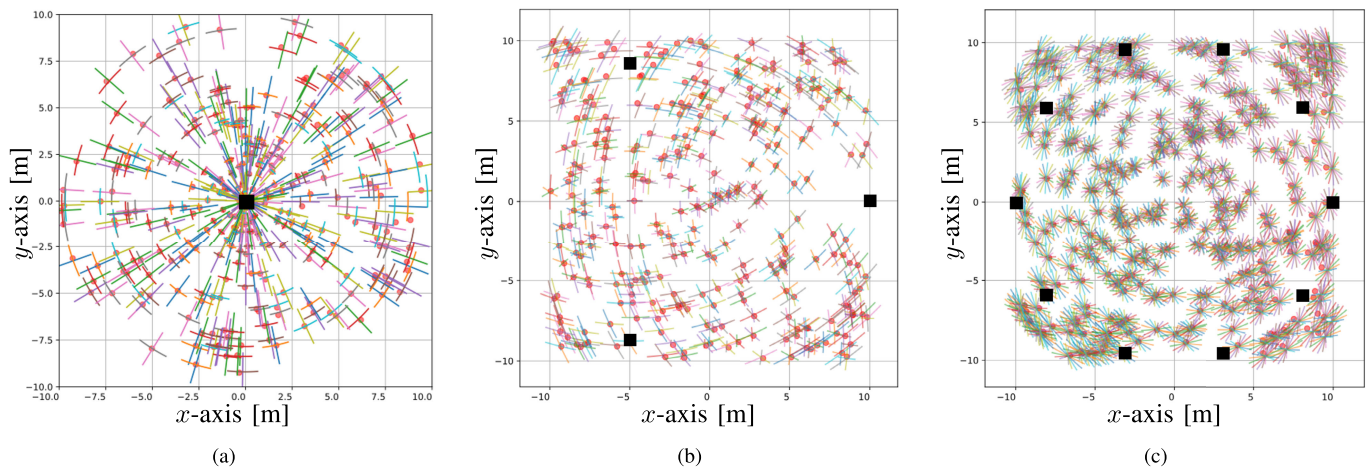


Fig. 9. Testing positioning results of an MLP with input (a) AoA and ToF, (b) 3 ToFs (c) 10 ToFs. The BSs are located in the black squares.

TABLE IV
LOS AND NLOS POSITIONING NETWORK

Layer Num.	Type	Output Size
0	Input	8×1
1	Linear + GeLu + Dropout	8×1
2	Linear + GeLu + Dropout	16×1
3	Linear + GeLu + Dropout	32×1
4	Linear + GeLu + Dropout	64×1
5	Linear + GeLu + Dropout	128×1
6	Linear + GeLu + Dropout	256×1
7	Linear + GeLu + Dropout	128×1
8	Linear + GeLu + Dropout	64×1
9	Linear + GeLu + Dropout	32×1
10	Linear + GeLu + Dropout	16×1
11	Linear	3×1

TABLE V
NLOS CLASSIFICATION NETWORK

Layer Num.	Type	Output Size
0	Input	8×1
1	Linear + Tanh + Dropout	8×1
2	Linear + Tanh + Dropout	5×1
3	Linear	1×1
4	BatchNorm1d	1×1
5	Sigmoid	1×1

substantial geometric errors are circumvented, and only the fingerprint technique is employed.

3) *Online Computational Complexity*: In this section, we analyze the computational complexity of the proposed localization method by assessing the number of floating point operations per second (FLOPS) and the inference time required for performing positioning. Starting from the computational complexity, we can divide the computational load into measurement extraction and position estimation. For measurement extraction, assuming to have N_g signal samples and MN antenna elements, the ADCPM can be efficiently computed adopting a 2D-inverse fast Fourier transform (IFFT), with a complexity of $O(MN N_g \cdot (\log(N_g MN)))$. On the other hand, time-based measurements obtained with cross-correlation, hold a complexity in the order of $O(N_g^2)$ (or $O(N_g \cdot (\log(N_g)))$ with efficient methods as FFT). Finally, angle-based measurements obtained by the MUSIC algorithm using N and M scanning directions in the azimuth and zenith domain, respectively, involve an overall complexity of $O((MN)^3)$ [77], considering the eigenvalue decomposition

(EVD) on the signal covariance matrix as predominant on the scanning search for source directions. In order to have a reference measure of time required by our system (described in Sec. VI-B) to compute the measurements, we clarify that on average, the ADCPM, ToF and AoA computations took about 0.4, 0.3 and 0.4 ms, respectively. As expected, the complexity of the ADCPM results higher than single time-based or angle-based measurements, as it embodies all the location-related information provided by the propagation channel, i.e., including the ToFs, AoAs and received power of all multipaths.

For what concerns the algorithms for position estimation, we compare non-linear least square (NLS) methods, adopted in Single-BS ToF-AoA and Multi-BS TDoF, and the proposed DL model. Denoting with N_u the unknown location coordinates to be estimated, N_{meas} the number of measurements and N_{iter} the number iterations for NLS convergence, for Gauss-Newton method we hold an overall complexity of $O(N_{\text{iter}}(N_{\text{meas}} \cdot N_u + N_{\text{meas}} \cdot N_u^2 + N_u^3))$ (assuming that the cost of computing the Jacobian is roughly proportional to $O(N_{\text{meas}} \cdot N_u)$, the cost of Jacobian matrix multiplication to $O(N_{\text{meas}} \cdot N_u^2)$, and the matrix inversion to $O(N_u^3)$). However, estimating the complexity of the DL models is not straightforward as it would involve a detailed breakdown of the operations performed. To this aim, in Fig. 12, we empirically assess the inference times using the system hardware described in Sec. VI-B and we compare it with NLS solutions, for varying number of measurements. For NLS method, we empirically choose $N_{\text{iter}} = 50$ and $N_u = 3$ (3D position). Results show that the proposed DL model is able to perform the position inference in about 1.5-2 ms which corresponds to a NLS with about 10 measurements. We point out that these results highly depend on the specific hardware and implementation of both the DL and NLS algorithms. Nevertheless, the overall conclusion is that, despite being slightly more complex, the proposed method has the main advantages of having greater accuracy with respect to classical geometrical algorithms and retaining the ability to localize in NLOS conditions (as described in Sec. VI-E.4), while being at the same time compliant with the strict requirements given

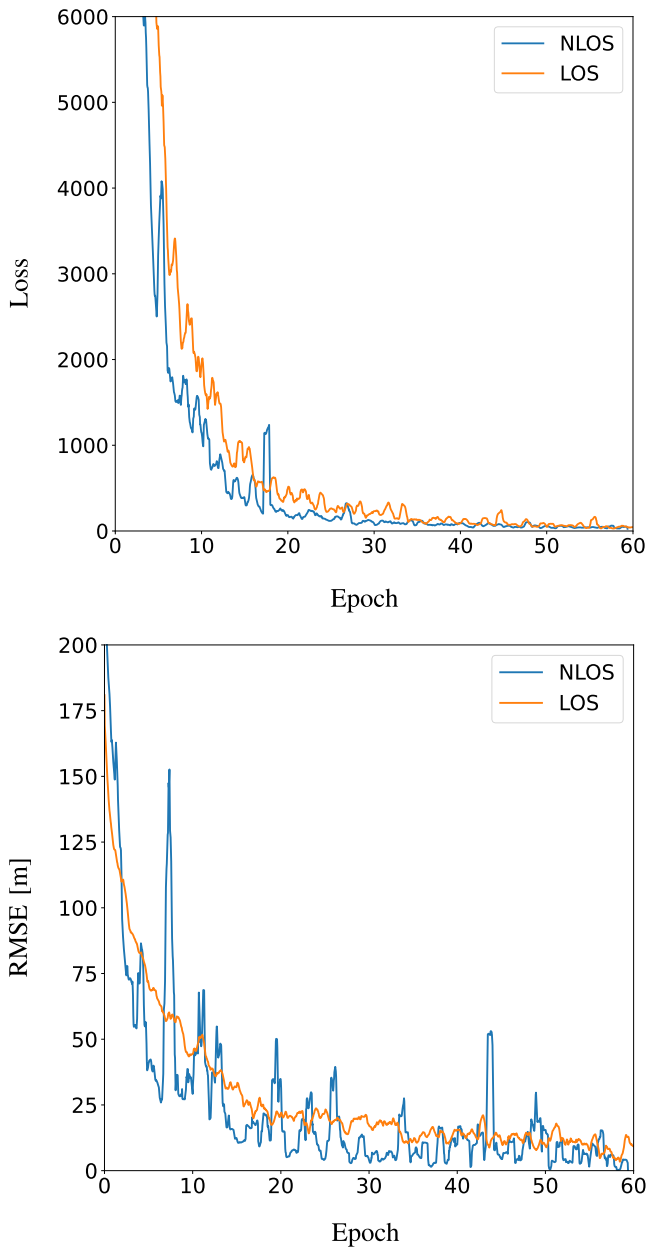


Fig. 10. Testing results, i.e., loss on top and RMSE on bottom, of LOS and NLOS samples varying the number of epochs in the training.

by cooperative, connected, and automated mobility (CCAM) of 5 ms [78].

4) *Positioning Accuracy*: In this assessment we test the positioning accuracy of the methods described in Sec. VI-D. To this aim, in Fig. 13 and 14, we show the cumulative density functions (CDFs) of the location error in the LOS and complete test trajectory, respectively. For the cooperative methods, we consider a position LOS if all the BSs detecting the UE are in LOS. To better compare the results, in both the figures, we report both the 2D and 3D error for each method.

Starting from the position methods with single BS in the LOS positions, we notice a huge improvement between the Single-BS ToF-AoA and the ego DL model, passing from a mean error of 26.38 m to 5.99 m in 3 dimensions. With the 2D error metric, the performances are slightly better,

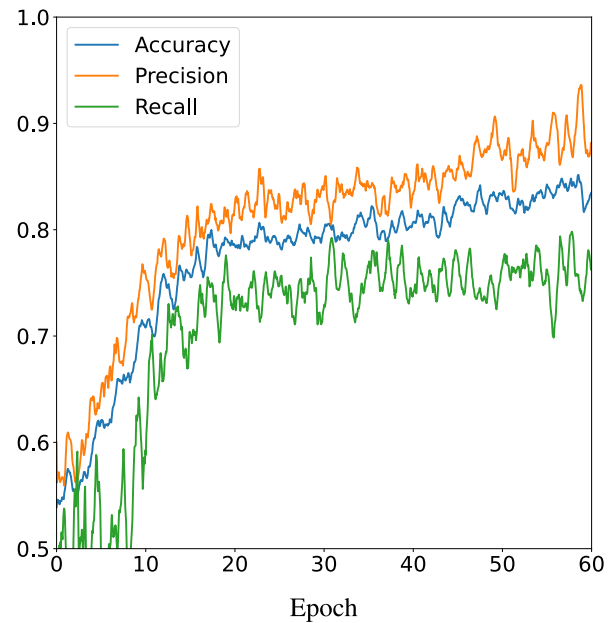


Fig. 11. Testing results of classification accuracy, precision and recall varying the number of epochs in the training.

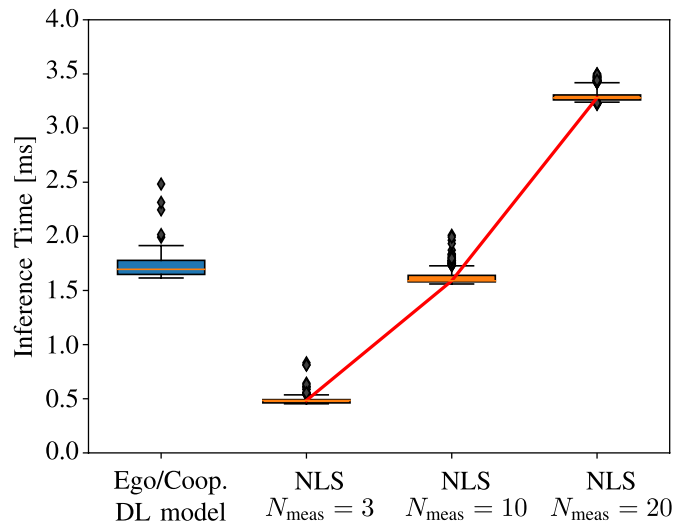


Fig. 12. Boxplot of the distribution of the inference time per sample [ms].

obtaining 24.76 m and 5.35 m, respectively. The ego DL model automatically extracts a compact non-linear representation of the overall multipath profile (i.e., ToF, AoA and power of all paths in LOS conditions), which is much more informative than the direct path information, thus outperforming the traditional single-BS based positioning. Moreover, in 3D positioning, it outperforms even the Multi-BS TDoF. This is due to the fact that, in classical TDoF approaches, the vertical geometrical dilution of precision (GDoP) is very limited due to the poor geometrical arrangement of the BSs over the vertical dimension, as BSs are usually located at similar heights. On the contrary, the ML approach is able to learn the usual altitude of the user and exploit this information a-priori in the position's computation.

Moving to the cooperative methods in the LOS testing trajectory, we observe that the proposed cooperative DL model

outperforms the Multi-BS TDoF even in the 2D positioning case, i.e., holding 66% of the points with an error of less than 0.81 m and 90% with an error of less than 1.3 m. The mean error decreases from 3.21 m with Multi-BS TDoF, to 1.68 m with the cooperative DL solution, with an improvement of above 47%. The cooperative DL model, in fact, holds a common-sense of where the UE could be, i.e., discarding or not considering possible unfeasible solutions that could be obtained by geometrical algorithms.

This is confirmed by the results on the complete testing trajectory in Fig. 14, composed by both LOS and NLOS positions. Whenever one or more BSs are in NLOS, we suffer a severe degradation of performances. In these conditions, the ego DL model reaches a median 2D error of 3.01 m with respect to an error of 3.74 m in case of Multi-BS TDoF. Comparing the cooperative and ego DL models, we notice an approaching of the two CDFs, mainly due to the higher NLOS performances in case of ego-positioning. This is because, even when LOS positions are inaccurately classified, the positioning is corrected through fingerprint training without being impacted by errors in geometrical features. Essentially, the position estimation solely relies on either the LOS MLP or the NLOS MLP, but never a combination of the two. On the other hand, the cooperative DL model suffers slightly higher errors since in NLOS positions the BSs cannot cooperate.

In case the achieved performances do not satisfy the target accuracy for the specific location-sensitive service, several strategies can be implemented in order to improve the proposed method without changing the structure of the model. First, starting from the physical layer, increasing the bandwidth and number of antennas at the BSs would permit to enhance the space-time system resolution, and thereby the ability of the model to resolve the multipath and extract location information from the ADCPM. Second, from a design point of view, we can increase the DL model dimension (especially the AE part), which reduces the model bias, and simultaneously increase the number of collected data, thus reducing the variance. However, we need to keep in mind that this comes at the price of higher training and inference times, as well as higher costs of dataset creation, resulting in a performance-complexity trade-off.

5) *Tracking Performance*: This experiment compares the performances of the positioning methods in the testing trajectory where a UE, i.e., a CAV, moves along a road at variable speed. In Fig. 15, we can see in pink the covered ground-truth trajectory both in 3D (a) and in 2D (b). The CAV moves faster in the north part of the trajectory and then slows down in the southern part of the road. Since the objective is to assess the point-position estimation of each method, we do not implement any tracking filter and we rely exclusively on the channel realization at specific samples of the trajectory. Moreover, passing from the 3D to the 2D representation, we discard unlikely estimated positions just for easy-visualization purposes of the 2D trajectory. The Single-BS ToF-AoA and the ego DL model results are obtained from the BS number 48, while the cooperative methods consider only the positions where the CAV detect the BS number 48.

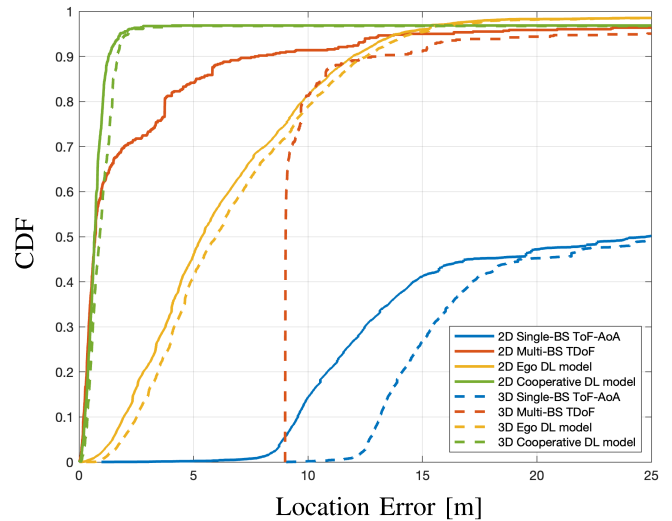


Fig. 13. Positioning performances in terms of CDF of the distance error in the LOS testing trajectory. The solid and dotted lines are the 2D and 3D errors, respectively.

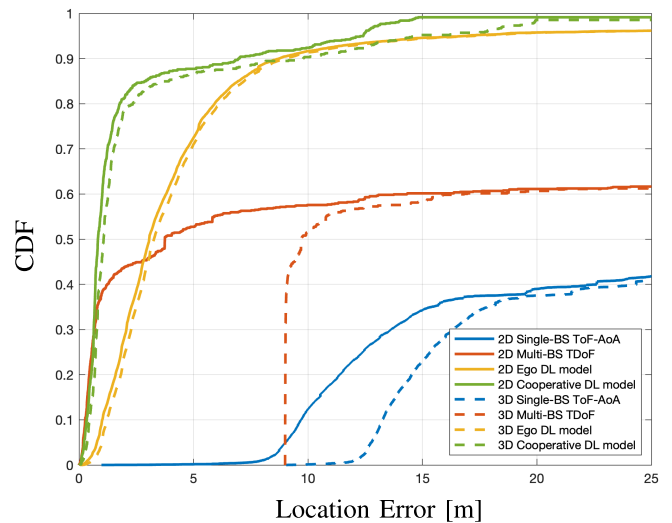


Fig. 14. Positioning performances in terms of CDF of the location error over the whole testing trajectory. The solid and dotted lines are the 2D and 3D errors, respectively.

Observing Fig. 15a, we notice that the Single-BS ToF-AoA method has the worst performances since it locates the CAV outside the road for most of the trajectory. This is mainly due to AoA estimations, which become worst the higher the distance from the BS, and ToF estimations. In LOS positions, i.e., north part of the trajectory, the ToF error is only due to the autocorrelation of SRSs and peak estimation. On the contrary, in NLOS positions, i.e., south part of the trajectory, the major source of error is represented by reflections. The ego DL model improves this aspect by halving the error (see Fig. 15b) especially in NLOS sections. The Multi-BS TDoF struggles in high-speed conditions, i.e., east part of LOS trajectory, and in NLOS positions where the number of cooperative BSs is limited or the range-biases are severe, i.e., top-right corner of Fig. 15a. Finally, the proposed cooperative DL model (green markers) achieves the higher positioning accuracy in almost all conditions, combining

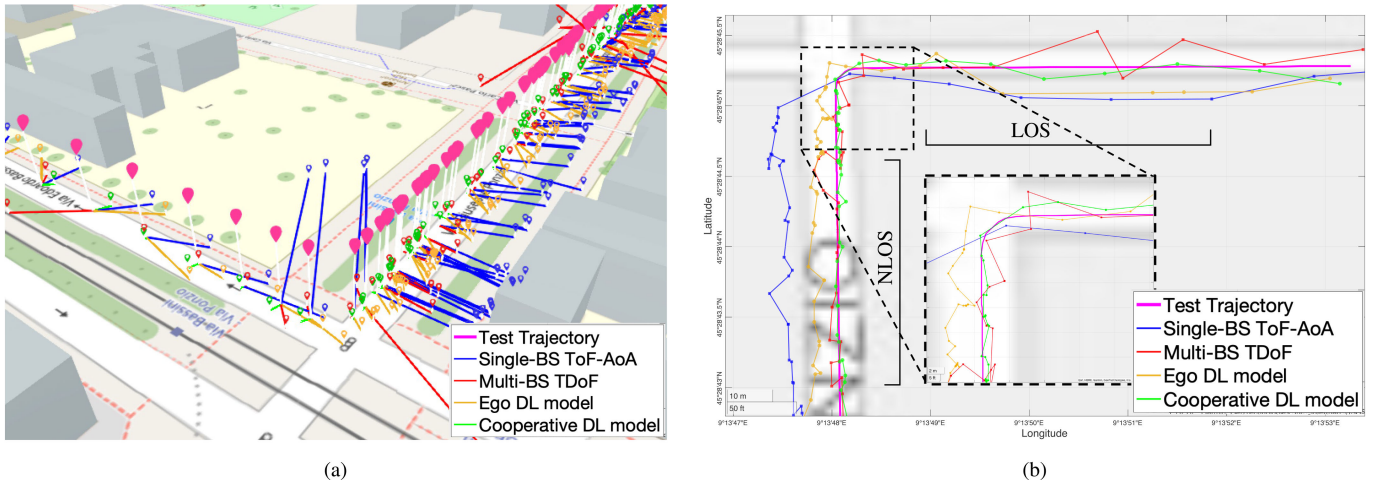


Fig. 15. Positioning performance in a 5G urban scenario: (a) 3D testing trajectory and related estimate obtained by the positioning methods (represented by different colors), with location errors represented as solid lines. (b) Bird-eye view of the testing and estimated trajectories.

cooperative LOS measurements with egocentric NLOS predictions.

VII. CONCLUSION AND FUTURE WORKS

Given the paramount importance of providing enhanced solutions for high-precision positioning in next 3GPP Releases, in this paper, we propose a cooperative positioning network architecture based on DL. Each BS composing the localization infrastructure has the same proposed DL model which solves the joint task of NLOS identification and position estimation. Depending on the condition, LOS or NLOS, the task is solved in cooperative or egocentric (ego) mode, respectively. In order to cooperate, the architecture internally exchanges only the compact latent feature representation of the channel obtained with an AE structure, permitting to combine the complete non-linear measurements and enhance positioning accuracy.

The proposed cooperative architecture is suitable and fully-compliant with 5G massive-MIMO OFDM systems, where sparse space-time channel responses, i.e., ADCPM, are adopted as input-images to the DL model. The ADCPM embodies position-dependent features, such as ToF, AoA and RSS of each propagation path, which can be automatically extracted by the proposed DL model. With the use of Matlab ray-tracing and SUMO software, we simulate a complex and realistic C-ITS scenario where some CAVs create multiple trajectories and communicate with a set of BSs, i.e., 3GPP UMi scenario. Results show that the proposed cooperative architecture is able to improve upon classical geometrical algorithms, e.g., TDoF multi-lateration, both in LOS and NLOS conditions, by increasing the accuracy of 47%. Moreover, the cooperation overcomes the limitations of single-BS prediction based on DL by automatically switching between egocentric and cooperative mode.

ML, and more especially DL techniques, are foreseen to have a huge impact on next-generation cellular networks. This work is thereby a first attempt to implement a cooperative high-precision positioning system towards that direction. Further developments could be the integration of different

DL models into the architecture or the tracking of many simultaneous targets with automatic data-association.

APPENDIX A

JOINT LOG-LIKELIHOOD

To prove (13), we start by rewriting the likelihood distribution (12) as:

$$p(\mathbf{u}_i, s_i | \mathbf{x}_i, \mathbf{W}, \mathbf{C}_{\text{LOS}}, \mathbf{C}_{\text{NLOS}}) = [\hat{p}_{s,i} \mathcal{N}(\mathbf{u}_i; \hat{\mathbf{u}}_{\text{LOS},i}(\mathbf{x}_i, \mathbf{W}), \mathbf{C}_{\text{LOS}})]^{s_i} \times [(1 - \hat{p}_{s,i}) \mathcal{N}(\mathbf{u}_i; \hat{\mathbf{u}}_{\text{NLOS},i}(\mathbf{x}_i, \mathbf{W}), \mathbf{C}_{\text{NLOS}})]^{(1-s_i)}. \quad (16)$$

For simplicity of notation, we drop the dependencies on \mathbf{x}_i and \mathbf{W} . The negative log-likelihood of the overall batch of samples, using (16) is:

$$\begin{aligned} \mathcal{L}_{\text{task}} &= - \sum_{i=1}^{N_b} \log \left\{ [\hat{p}_{s,i} \mathcal{N}(\mathbf{u}_i; \hat{\mathbf{u}}_{\text{LOS},i}, \mathbf{C}_{\text{LOS}})]^{s_i} \right. \\ &\quad \left. \times [(1 - \hat{p}_{s,i}) \mathcal{N}(\mathbf{u}_i; \hat{\mathbf{u}}_{\text{NLOS},i}, \mathbf{C}_{\text{NLOS}})]^{(1-s_i)} \right\} \quad (17) \\ &= \sum_{i=1}^{N_b} \left\{ s_i [-\log(\hat{p}_{s,i}) - \log(\mathcal{N}(\mathbf{u}_i; \hat{\mathbf{u}}_{\text{LOS},i}, \mathbf{C}_{\text{LOS}}))] \right. \\ &\quad \left. + (1-s_i) [-\log(1-\hat{p}_{s,i}) - \log(\mathcal{N}(\mathbf{u}_i; \hat{\mathbf{u}}_{\text{NLOS},i}, \mathbf{C}_{\text{NLOS}}))] \right\}. \quad (18) \end{aligned}$$

Now we can explicitly compute the logarithm of the multi-variate Gaussian distribution and obtain:

$$\begin{aligned} \mathcal{L}_{\text{task}} &= \sum_{i=1}^{N_b} \left\{ s_i \cdot \left[-\log(\hat{p}_{s,i}) + \frac{\log(2\pi|\mathbf{C}_{\text{LOS}}|)}{2} \right. \right. \\ &\quad \left. \left. + \frac{\|\mathbf{u}_i - \hat{\mathbf{u}}_{\text{LOS},i}\|_2^2}{2\sigma_{\text{LOS}}^2} \right] \right. \\ &\quad \left. + (1-s_i) \cdot \left[-\log(1-\hat{p}_{s,i}) + \frac{\log(2\pi|\mathbf{C}_{\text{NLOS}}|)}{2} \right. \right. \\ &\quad \left. \left. + \frac{\|\mathbf{u}_i - \hat{\mathbf{u}}_{\text{NLOS},i}\|_2^2}{2\sigma_{\text{NLOS}}^2} \right] \right\}. \quad (19) \end{aligned}$$

By discarding the terms that do not depend on \mathbf{u}_i or s_i , we obtain:

$$\begin{aligned} \mathcal{L}_{\text{task}} \simeq & \sum_{i=1}^{N_b} \left\{ s_i \cdot \left[-\log(\hat{p}_{s,i}) + \frac{\|\mathbf{u}_i - \hat{\mathbf{u}}_{\text{LOS},i}\|_2^2}{2\sigma_{\text{LOS}}^2} \right] \right. \\ & \left. + (1 - s_i) \cdot \left[-\log(1 - \hat{p}_{s,i}) + \frac{\|\mathbf{u}_i - \hat{\mathbf{u}}_{\text{NLOS},i}\|_2^2}{2\sigma_{\text{NLOS}}^2} \right] \right\}, \end{aligned} \quad (20)$$

concluding the derivation.

REFERENCES

- [1] J. A. del Peral-Rosado, R. Raulefs, J. A. López-Salcedo, and G. Seco-Granados, "Survey of cellular mobile radio localization methods: From 1G to 5G," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 2, pp. 1124–1148, 2nd Quart., 2018. [Online]. Available: <https://ieeexplore.ieee.org/document/8226757/>
- [2] O. Kanhere and T. S. Rappaport, "Position location for futuristic cellular communications: 5G and beyond," *IEEE Commun. Mag.*, vol. 59, no. 1, pp. 70–75, Jan. 2021. [Online]. Available: <https://ieeexplore.ieee.org/document/9356512/>
- [3] A. Nessa, B. Adhikari, F. Hussain, and X. N. Fernando, "A survey of machine learning for indoor positioning," *IEEE Access*, vol. 8, pp. 214945–214965, 2020. [Online]. Available: <https://ieeexplore.ieee.org/document/9264122/>
- [4] R. Di Taranto, S. Muppirisetty, R. Raulefs, D. Slock, T. Svensson, and H. Wymeersch, "Location-aware communications for 5G networks: How location information can improve scalability, latency, and robustness of 5G," *IEEE Signal Process. Mag.*, vol. 31, no. 6, pp. 102–112, Nov. 2014. [Online]. Available: <https://ieeexplore.ieee.org/document/6924849>
- [5] A. M. Nor, S. Halunga, and O. Fratu, "Survey on positioning information assisted mmWave beamforming training," *Ad Hoc Netw.*, vol. 135, Oct. 2022, Art. no. 102947. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S1570870522001287>
- [6] J. Zhang, E. Björnson, M. Matthaiou, D. W. K. Ng, H. Yang, and D. J. Love, "Prospective multiple antenna technologies for beyond 5G," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1637–1660, Aug. 2020. [Online]. Available: <https://ieeexplore.ieee.org/document/9113273/>
- [7] L. Yin, Q. Ni, and Z. Deng, "A GNSS/5G integrated positioning methodology in D2D communication networks," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 2, pp. 351–362, Feb. 2018.
- [8] F. Mogyorósi et al., "Positioning in 5G and 6G networks—A survey," *Sensors*, vol. 22, no. 13, p. 4757, Jun. 2022. [Online]. Available: <https://www.mdpi.com/1424-8220/22/13/4757>
- [9] A. M. Nor and E. M. Mohamed, "Li-Fi positioning for efficient millimeter wave beamforming training in indoor environment," *Mobile Netw. Appl.*, vol. 24, no. 2, pp. 517–531, Apr. 2019, doi: [10.1007/s11036-018-1154-4](https://doi.org/10.1007/s11036-018-1154-4).
- [10] C. Morelli, M. Nicoli, V. Rampa, and U. Spagnolini, "Hidden Markov models for radio localization in mixed LOS/NLOS conditions," *IEEE Trans. Signal Process.*, vol. 55, no. 4, pp. 1525–1542, Apr. 2007. [Online]. Available: <http://ieeexplore.ieee.org/document/4133045/>
- [11] L. Barbieri, M. Brambilla, A. Trabattoni, S. Mervic, and M. Nicoli, "UWB localization in a smart factory: Augmentation methods and experimental assessment," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–18, 2021. [Online]. Available: <https://ieeexplore.ieee.org/document/9409138/>
- [12] F. Liu et al., "Integrated sensing and communications: Toward dual-functional wireless networks for 6G and beyond," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 6, pp. 1728–1767, Jun. 2022. [Online]. Available: <https://ieeexplore.ieee.org/document/9737357/>
- [13] Q. Zhang, H. Sun, X. Gao, X. Wang, and Z. Feng, "Time-division ISAC enabled connected automated vehicles cooperation algorithm design and performance evaluation," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 7, pp. 2206–2218, Jul. 2022. [Online]. Available: <https://ieeexplore.ieee.org/document/9728752/>
- [14] S. Shi, Z. Cheng, L. Wu, Z. He, and B. Shankar, "Distributed 5G NR-based integrated sensing and communication systems: Frame structure and performance analysis," in *Proc. 30th Eur. Signal Process. Conf. (EUSIPCO)*, Belgrade, Serbia, Aug. 2022, pp. 1062–1066. [Online]. Available: <https://ieeexplore.ieee.org/document/9909950/>
- [15] Z. Yang, C. Wu, and Y. Liu, "Locating in fingerprint space: Wireless indoor localization with little human intervention," in *Proc. 18th Annu. Int. Conf. Mobile Comput. Netw.*, Istanbul Turkey, Aug. 2012, pp. 269–280, doi: [10.1145/2348543.2348578](https://doi.org/10.1145/2348543.2348578).
- [16] M. M. Butt, A. Rao, and D. Yoon, "RF fingerprinting and deep learning assisted UE positioning in 5G," in *Proc. IEEE 91st Veh. Technol. Conf. (VTC-Spring)*, May 2020, pp. 1–7.
- [17] J. Sangthong, J. Thongkam, and S. Promwong, "Indoor wireless sensor network localization using RSSI based weighting algorithm method," in *Proc. 6th Int. Conf. Eng., Appl. Sci. Technol.*, Chiang Mai, Thailand, Jul. 2020, pp. 1–4. [Online]. Available: <https://ieeexplore.ieee.org/document/9165300/>
- [18] *Moderator's Summary for Discussion [Ran93E-R18Prep-10] Expanded and Improved Positioning*, document TSG-RAN meeting (RP) 0.0.4, Version 0.0.4, 3GPP, Sep. 2021. [Online]. Available: https://www.3gpp.org/ftp/tsg_ran/TSG_RAN/TSGR_93e/Docs/
- [19] A. Bourdoux et al., "6G white paper on localization and sensing," 2020, *arXiv:2006.01779*.
- [20] T. Wild, V. Braun, and H. Viswanathan, "Joint design of communication and sensing for beyond 5G and 6G systems," *IEEE Access*, vol. 9, pp. 30845–30857, 2021. [Online]. Available: <https://ieeexplore.ieee.org/document/9354629/>
- [21] S. Rezaie, E. de Carvalho, and C. N. Manchón, "A deep learning approach to location- and orientation-aided 3D beam selection for mmWave communications," *IEEE Trans. Wireless Commun.*, vol. 21, no. 12, pp. 11110–11124, Dec. 2022. [Online]. Available: <https://ieeexplore.ieee.org/document/9832551/>
- [22] M. A. Arfaoui et al., "Invoking deep learning for joint estimation of indoor LiFi user position and orientation," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 9, pp. 2890–2905, Sep. 2021.
- [23] S. V. Balkus, H. Wang, B. D. Cornet, C. Mahabal, H. Ngo, and H. Fang, "A survey of collaborative machine learning using 5G vehicular communications," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 2, pp. 1280–1303, 2nd Quart., 2022. [Online]. Available: <https://ieeexplore.ieee.org/document/9706268/>
- [24] L. Barbieri, B. C. Tedeschini, M. Brambilla, and M. Nicoli, "Implicit vehicle positioning with cooperative LiDAR sensing," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Jun. 2023, pp. 1–5.
- [25] S. Bartoletti et al., "Positioning and sensing for vehicular safety applications in 5G and beyond," *IEEE Commun. Mag.*, vol. 59, no. 11, pp. 15–21, Nov. 2021. [Online]. Available: <https://ieeexplore.ieee.org/document/9665433/>
- [26] B. C. Tedeschini, M. Brambilla, L. Barbieri, and M. Nicoli, "Addressing data association by message passing over graph neural networks," in *Proc. 25th Int. Conf. Inf. Fusion (FUSION)*, Linköping, Sweden, Jul. 2022, pp. 1–7. [Online]. Available: <https://ieeexplore.ieee.org/document/9841233/>
- [27] P. Zhang, J. Lu, Y. Wang, and Q. Wang, "Cooperative localization in 5G networks: A survey," *ICT Exp.*, vol. 3, no. 1, pp. 27–32, Mar. 2017. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2405959517300346>
- [28] G. Soatti, M. Nicoli, N. Garcia, B. Denis, R. Raulefs, and H. Wymeersch, "Implicit cooperative positioning in vehicular networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 12, pp. 3964–3980, Dec. 2018. [Online]. Available: <https://ieeexplore.ieee.org/document/8307478/>
- [29] M. Brambilla, D. Pardo, and M. Nicoli, "Location-assisted subspace-based beam alignment in LOS/NLOS mm-wave V2X communications," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Dublin, Ireland, Jun. 2020, pp. 1–6. [Online]. Available: <https://ieeexplore.ieee.org/document/9148587/>
- [30] M. Brambilla, L. Combi, A. Matera, D. Tagliaferri, M. Nicoli, and U. Spagnolini, "Sensor-aided V2X beam tracking for connected automated driving: Distributed architecture and processing algorithms," *Sensors*, vol. 20, no. 12, p. 3573, Jun. 2020. [Online]. Available: <https://www.mdpi.com/1424-8220/20/12/3573>
- [31] B. C. Tedeschini, M. Nicoli, and M. Z. Win, "On the latent space of mmWave MIMO channels for NLOS identification in 5G-advanced systems," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 6, pp. 1655–1669, May 2023. [Online]. Available: <https://ieeexplore.ieee.org/document/10121016/>
- [32] S. Marano, W. M. Gifford, H. Wymeersch, and M. Z. Win, "NLOS identification and mitigation for localization based on UWB experimental data," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 7, pp. 1026–1035, Sep. 2010. [Online]. Available: <http://ieeexplore.ieee.org/document/5555901/>

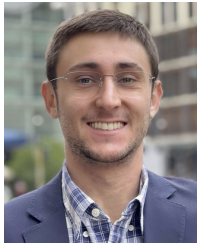
- [33] H. Wymeersch, S. Marano, W. M. Gifford, and M. Z. Win, "A machine learning approach to ranging error mitigation for UWB localization," *IEEE Trans. Commun.*, vol. 60, no. 6, pp. 1719–1728, Jun. 2012. [Online]. Available: <http://ieeexplore.ieee.org/document/6192275/>
- [34] T. Van Nguyen, Y. Jeong, H. Shin, and M. Z. Win, "Machine learning for wideband localization," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 7, pp. 1357–1380, Jul. 2015. [Online]. Available: <http://ieeexplore.ieee.org/document/7102989/>
- [35] E. Kurniawan, L. Zhiwei, and S. Sun, "Machine learning-based channel classification and its application to IEEE 802.11ad communications," in *Proc. IEEE Global Commun. Conf.*, Singapore, Dec. 2017, pp. 1–6. [Online]. Available: <http://ieeexplore.ieee.org/document/8254052/>
- [36] X. Yang, "NLOS mitigation for UWB localization based on sparse pseudo-input Gaussian process," *IEEE Sensors J.*, vol. 18, no. 10, pp. 4311–4316, May 2018. [Online]. Available: <https://ieeexplore.ieee.org/document/832223/>
- [37] P. Bahl and V. N. Padmanabhan, "RADAR: An in-building RF-based user location and tracking system," in *Proc. IEEE INFOCOM. Conf. Comput. Commun., 19th Annu. Joint Conf. IEEE Comput. Commun. Societies*, Tel Aviv, Israel, Mar. 2000, pp. 775–784. [Online]. Available: <http://ieeexplore.ieee.org/document/832252/>
- [38] M. Youssef and A. Agrawala, "The Horus WLAN location determination system," in *Proc. 3rd Int. Conf. Mobile Syst., Appl.*, Jun. 2005, pp. 205–218, doi: [10.1145/1067170.1067193](https://doi.org/10.1145/1067170.1067193).
- [39] X. Wang, L. Gao, S. Mao, and S. Pandey, "DeepFi: Deep learning for indoor fingerprinting using channel state information," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Mar. 2015, pp. 1666–1671.
- [40] Y. Chapre, A. Ignjatovic, A. Seneviratne, and S. Jha, "CSI-MIMO: An efficient Wi-Fi fingerprinting using channel state information with MIMO," *Pervas. Mobile Comput.*, vol. 23, pp. 89–103, Oct. 2015. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1574119215001406>
- [41] X. Wang, L. Gao, and S. Mao, "PhaseFi: Phase fingerprinting for indoor localization with a deep learning approach," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, San Diego, CA, USA, Dec. 2014, pp. 1–6. [Online]. Available: <http://ieeexplore.ieee.org/document/7417517/>
- [42] S. Savazzi, S. Sigg, M. Nicoli, V. Rampa, S. Kianoush, and U. Spagnolini, "Device-free radio vision for assisted living: Leveraging wireless channel quality information for human sensing," *IEEE Signal Process. Mag.*, vol. 33, no. 2, pp. 45–58, Mar. 2016. [Online]. Available: <http://ieeexplore.ieee.org/document/7426567/>
- [43] H. Chen, Y. Zhang, W. Li, X. Tao, and P. Zhang, "ConFi: Convolutional neural networks based indoor Wi-Fi localization using channel state information," *IEEE Access*, vol. 5, pp. 18066–18074, 2017. [Online]. Available: <http://ieeexplore.ieee.org/document/8027020/>
- [44] X. Wang, X. Wang, and S. Mao, "ResLoc: Deep residual learning for indoor localization with CSI tensors," in *Proc. IEEE 28th Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC)*, Montreal, QC, USA, Oct. 2017, pp. 1–6. [Online]. Available: <http://ieeexplore.ieee.org/document/8292236/>
- [45] X. Wang, X. Wang, and S. Mao, "CiFi: Deep convolutional neural networks for indoor localization with 5 GHz Wi-Fi," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Paris, France, May 2017, pp. 1–6. [Online]. Available: <http://ieeexplore.ieee.org/document/7997235/>
- [46] Y. Jing, J. Hao, and P. Li, "Learning spatiotemporal features of CSI for indoor localization with dual-stream 3D convolutional neural networks," *IEEE Access*, vol. 7, pp. 147571–147585, 2019. [Online]. Available: <https://ieeexplore.ieee.org/document/8864050/>
- [47] M. Malmstrom, I. Skog, S. M. Razavi, Y. Zhao, and F. Gunnarsson, "5G positioning—A machine learning approach," in *Proc. 16th Workshop Positioning, Navigat. Commun. (WPNC)*, Bremen, Germany, Oct. 2019, pp. 1–6. [Online]. Available: <https://ieeexplore.ieee.org/document/8970186/>
- [48] J. Gante, G. Falcão, and L. Sousa, "Deep learning architectures for accurate millimeter wave positioning in 5G," *Neural Process. Lett.*, vol. 51, no. 1, pp. 487–514, Feb. 2020, doi: [10.1007/s11063-019-10073-1](https://doi.org/10.1007/s11063-019-10073-1).
- [49] Y. Li, S. Mazuelas, and Y. Shen, "Deep generative model for simultaneous range error mitigation and environment identification," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Madrid, Spain, Dec. 2021, pp. 1–6. [Online]. Available: <https://ieeexplore.ieee.org/document/9685255/>
- [50] M. Stahlke, S. Kram, F. Ott, T. Feigl, and C. Mutschler, "Estimating TOA reliability with variational autoencoders," *IEEE Sensors J.*, vol. 22, no. 6, pp. 5133–5140, Mar. 2022. [Online]. Available: <https://ieeexplore.ieee.org/document/9503384/>
- [51] N. Lv, F. Wen, Y. Chen, and Z. Wang, "A deep learning-based end-to-end algorithm for 5G positioning," *IEEE Sensors Lett.*, vol. 6, no. 4, pp. 1–4, Apr. 2022. [Online]. Available: <https://ieeexplore.ieee.org/document/9706343/>
- [52] K. Gao, H. Wang, H. Lv, and W. Liu, "Toward 5G NR high-precision indoor positioning via channel frequency response: A new paradigm and dataset generation method," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 7, pp. 2233–2247, Jul. 2022. [Online]. Available: <https://ieeexplore.ieee.org/document/9729785/>
- [53] C. Wu et al., "Learning to localize: A 3D CNN approach to user positioning in massive MIMO-OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 20, no. 7, pp. 4556–4570, Jul. 2021. [Online]. Available: <https://ieeexplore.ieee.org/document/9364875/>
- [54] B. El Boudani et al., "Implementing deep learning techniques in 5G IoT networks for 3D indoor positioning: DELTA (deep learning-based co-operative architecture)," *Sensors*, vol. 20, no. 19, p. 5495, Sep. 2020. [Online]. Available: <https://www.mdpi.com/1424-8220/20/19/5495>
- [55] H. Kim, S. H. Lee, and S. Kim, "Cooperative localization with distributed ADMM over 5G-based VANETs," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Apr. 2018, pp. 1–5. [Online]. Available: <https://ieeexplore.ieee.org/document/8377454/>
- [56] L. Barbieri, S. Savazzi, M. Brambilla, and M. Nicoli, "Decentralized federated learning for extended sensing in 6G connected vehicles," *Veh. Commun.*, vol. 33, Jan. 2022, Art. no. 100396. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S2214209621000656>
- [57] *Study on NR Positioning Support*, document TR 38.855, Version 16.0.0, 3GPP, Sep. 2019. [Online]. Available: <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3501>
- [58] P. Alvarez Lopez et al., "Microscopic traffic simulation using SUMO," in *Proc. 21st Int. Conf. Intell. Transp. Syst.*, Nov. 2018, pp. 2575–2582. [Online]. Available: <https://elib.dlr.de/124092/>
- [59] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.
- [60] H. L. Van Trees, *Detection, Estimation, and Modulation Theory. 4: Optimum Array Processing*. New York, NY, USA: Wiley, 2002.
- [61] X. Sun, X. Gao, G. Y. Li, and W. Han, "Single-site localization based on a new type of fingerprint for massive MIMO-OFDM systems," *IEEE Trans. Veh. Technol.*, vol. 67, no. 7, pp. 6134–6145, Jul. 2018. [Online]. Available: <https://ieeexplore.ieee.org/document/8307353/>
- [62] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Computer Vision—ECCV (Lecture Notes in Computer Science)*, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds. Cham, Switzerland: Springer, 2014, pp. 818–833.
- [63] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Netw.*, vol. 61, pp. 85–117, Jan. 2015. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0893608014002135>
- [64] C. M. Bishop, *Pattern Recognition and Machine Learning*, vol. 4, no. 4. Berlin, Germany: Springer, Aug. 2006. [Online]. Available: <https://link.springer.com/book/9780387310732>
- [65] R. A. Jacobs, M. I. Jordan, S. J. Nowlan, and G. E. Hinton, "Adaptive mixtures of local experts," *Neural Comput.*, vol. 3, no. 1, pp. 79–87, Mar. 1991. [Online]. Available: <http://ieeexplore.ieee.org/document/6797059/>
- [66] *Study on Channel Model for Frequencies From 0.5 to 100 GHz (Rel-16)*, document TR 38.901, Version 16.1.0, 3GPP, Nov. 2020. [Online]. Available: https://www.etsi.org/deliver/etsi_tr/138900_138999/138901/16.01.00_60/tr_138901v1601000.pdf
- [67] C.-F. Yang and C.-J. Ko, "A ray tracing method for modeling indoor wave propagation and penetration," in *Proc. IEEE Antennas Propag. Soc. Int. Symp.*, Baltimore, MD, USA, Jul. 1996, pp. 441–444. [Online]. Available: <http://ieeexplore.ieee.org/document/549632/>
- [68] H.-J. Li, C.-C. Chen, T.-Y. Liu, and H.-C. Lin, "Applicability of ray-tracing technique for the prediction of outdoor channel characteristics," *IEEE Trans. Veh. Technol.*, vol. 49, no. 6, pp. 2336–2349, Nov. 2000. [Online]. Available: <http://ieeexplore.ieee.org/document/901902/>
- [69] A.-Y. Hsiao, C.-F. Yang, T.-S. Wang, I. Lin, and W.-J. Liao, "Ray tracing simulations for millimeter wave propagation in 5G wireless communications," in *Proc. IEEE Int. Symp. Antennas Propag. USNC/URSI Nat. Radio Sci. Meeting*, San Diego, CA, USA, Jul. 2017, pp. 1901–1902. [Online]. Available: <http://ieeexplore.ieee.org/document/8072993/>
- [70] Z. Yun and M. F. Iskander, "Ray tracing for radio propagation modeling: Principles and applications," *IEEE Access*, vol. 3, pp. 1089–1100, 2015. [Online]. Available: <http://ieeexplore.ieee.org/document/7152831/>

- [71] *Guidelines for Evaluation of Radio Interface Technologies for IMT-2020*, document SG05 ContribuQon 57, International Telecommunication Union (ITU)-R M, 2017. [Online]. Available: <https://www.itu.int/md/R15-SG05-C-0057>
- [72] A. Paszke et al., “Automatic differentiation in PyTorch,” in *Proc. NIPS*, Oct. 2017, pp. 1–4. [Online]. Available: <https://openreview.net/forum?id=BJJsrnfCZ>
- [73] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” 2014, *arXiv:1412.6980*.
- [74] V. Badrinarayanan, A. Kendall, and R. Cipolla, “SegNet: A deep convolutional encoder–decoder architecture for image segmentation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017. [Online]. Available: <http://ieeexplore.ieee.org/document/7803544/>
- [75] D. Hendrycks and K. Gimpel, “Gaussian error linear units (GELUs),” 2016, *arXiv:1606.08415*.
- [76] R. Schmidt, “Multiple emitter location and signal parameter estimation,” *IEEE Trans. Antennas Propag.*, vol. AP-34, no. 3, pp. 276–280, Mar. 1986. [Online]. Available: <http://ieeexplore.ieee.org/document/1143830/>
- [77] F.-G. Yan, J. Wang, S. Liu, Y. Shen, and M. Jin, “Reduced-complexity direction of arrival estimation using real-valued computation with arbitrary array configurations,” *Int. J. Antennas Propag.*, vol. 2018, pp. 1–10, Apr. 2018. [Online]. Available: <https://www.hindawi.com/journals/ijap/2018/3284619/>
- [78] *Study on Enhancement of 3GPP Support for 5G V2X Services*, document (TR) 22.886, Version 16.2.0, 3GPP, Dec. 2018. [Online]. Available: <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3108>



Monica Nicoli (Senior Member, IEEE) received the M.Sc. (Hons.) and Ph.D. degrees in communication engineering from Politecnico di Milano, Milan, Italy, in 1998 and 2002, respectively. She was a Visiting Researcher with ENI Agip, from 1998 to 1999, and Uppsala University, in 2001. In 2002, she joined Politecnico di Milano as a Faculty Member. She is currently an Associate Professor in telecommunications with the Department of Management, Economics and Industrial Engineering.

Her research interests include signal processing, machine learning, and wireless communications, with emphasis on smart mobility and Internet of Things (IoT). She was a recipient of the Marisa Bellisario Award, in 1999, and a co-recipient of the best paper awards of the EuMA Mediterranean Microwave Symposium, in 2022, the IEEE Symposium on Joint Communications and Sensing, in 2021, the IEEE Statistical Signal Processing Workshop, in 2018, and the *IET Intelligent Transport Systems journal*, in 2014. She is an Associate Editor of the IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS. She has also served as an Associate Editor for the *EURASIP Journal on Wireless Communications and Networking*, from 2010 to 2017, and a Lead Guest Editor for the Special Issue on Localization in Mobile Wireless and Sensor Networks, in 2011.



Bernardo Camajori Tedeschini (Graduate Student Member, IEEE) received the B.Sc. (Hons.) in Computer Science and M.Sc. (Hons.) degrees in Telecommunications Engineering from the Politecnico di Milano, Italy, in 2019 and 2021, respectively. From November 2021 he started as PhD fellow in Information Technology at Dipartimento di Eeeronica, Informazione e Bioingegneria (DEIB), Politecnico di Milano. He is currently a visiting researcher with the Laboratory for Information & Decision Systems at the Massachusetts Institute of

Technology (MIT), Cambridge, MA.

His research interests include federated learning, machine learning and localization methods. He was a recipient of the Ph.D. grant from the ministry of the Italian government Ministero dell’Istruzione, dell’Università e della Ricerca (MIUR) and the Roberto Rocca Doctoral Fellowship granted by MIT and Politecnico di Milano.