Robust ToA-Estimation using Convolutional Neural Networks on Randomized Channel Models

Tobias Feigl
Friedrich-Alexander-University of Erlangen-Nürnberg

Sebastian Kram
Fraunhofer Institute for Integrated Circuits IIS

Christopher Mutschler
Fraunhofer Institute for Integrated Circuits IIS

Robust ToA-Estimation using Convolutional Neural Networks on Randomized Channel Models

Tobias Feigl*, Ernst Eberlein*, Sebastian Kram†*, and Christopher Mutschler*
{ tobias.feigl | sebastian.kram | christopher.mutschler }@iis.fraunhofer.de
{ ernst.eberlein }@iis-extern.fraunhofer.de

*Fraunhofer IIS,
Fraunhofer Institute for Integrated Circuits IIS,
Division Positioning and Networks

‡Friedrich-Alexander University Erlangen-Nürnberg (FAU)
Institute of Information Technology
Erlangen, Germany

***Abstract*—Many radio-based positioning systems use time-of-arrival (ToA). We obtain it from the first and direct path of arrival (FDPoA) in a corresponding set of multipath components (MPC) of the underlying channel state information (CSI). While detection of the FDPoA under Line-of-Sight (LoS) is simple, it is prone to errors in environments with specular and diffuse reflections, as well as nonlinear diffraction, absorption, and transmission of a signal. Such Obstructed- or Non-Line-of-Sight (OLoS, NLoS) situations lead to incorrect FDPoA and consequently to incorrect ToA estimates and inaccurate positions. State-of-the-art estimators are computationally expensive and usually fail with O/NLoS at low signal-to-noise ratios (SNRs).**

We propose a deep learning (DL) approach to identify optimal FDPoAs as ToA directly from the raw CSI. Our 1D Convolutional Neural Network (CNN) learns the spatial distribution of MPCs of the CSI to predict correct estimates of the ToA. To train our DL model, we use QuaDRiGa to generate datasets with CIRs and ground truth ToAs for realistic 5G channel models. We found that Delay Spread (DS), k-Factor (kF), and SNR are appropriate metrics to cover most LoS-NLoS constellations in realistic datasets. We compare our DL model with state-of-the-art estimators such as threshold (PEAK), inflection point (IFP), and MUSIC and show that we consistently outperform them by about 17% for SNRs below -10 dB.

***Index Terms*—ToA Channel Parameter Estimation, Inflection Point, MUSIC, Machine Learning, Deep Learning.**

I. Introduction

Radio-frequency (RF) positioning relies on geometric relationship between the positions of transmitters and receivers. There are 3 main approaches [1]: Angle-of-Arrival (AoA), Received Signal Strength (RSS), and ToA. While AOA requires expensive directional antennas or antenna arrays [2], [3] RSS struggles from significant fluctuations over short distances and over time [2], [4], ToA relies on the efficient estimation of the time-of-flight and typically provides position accuracy in the centimeter range with good propagation conditions [3]. In a typical RF positioning system, several synchronized transceiver pipelines provide sets of channel impulse response (CIR) data, of which ToAs are estimated, which are then used for positioning tasks. However, a reliable ToA estimate is a challenge as it is a compromise between the legal (feasible) transmission power, the available bandwidth, and the size of the propagation environment. Fig. 1 shows 3 typical scenarios: A transmitter TX (right) emits a radio burst that travels through the environment to reach different receivers RX (left): LoS, NLoS, and OLoS. For the (green) LoS case, as the direct path and so the first and strongest peak in a CIR are not affected by a MPC, naive methods easily identify the correct FDPoA, i.e., ToA [5]. However, in the OLoS and NLoS cases, a path is affected by absorption, diffraction, reflection or transmission. For OLoS the direct path is difficult to distinguish from MPC components. And for NLoS the FDPoA is no longer available as it is significantly delayed and attenuated. In such practical cases it is difficult to estimate the true ToA from the CIR [6].

978-1-6654-0402-0/21/$31.00 © 2021 IEEE

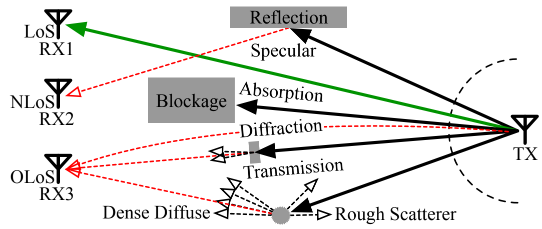


Fig. 1. Multipath propagation scenarios with specular and dense multipaths and respective LoS, OLoS, and NLoS (red: affected; green: unaffected).

Previous work estimates the ToA from multiple peaks in the CIR by selecting the maximum peak [7], applying thresholds along with the SNR and power w.r.t. the strongest peak [8] (PEAK), or by estimating the peak based on the maximum gradient along the first rising edge w.r.t. the inflection point (IFP) [9]. However, they do not work well, since multipath violates their core assumptions and their performance decreases with the SNR. In contrast, the high-resolution MUSIC [10], ESPRIT [11], RARE [12], SAGE [13], and RiMAX [14] methods estimate correct ToAs in the frequency and time domain, especially in multipath scenarios [7]. However, they require channel-specific a-priori information and their impractical computational complexity increases with the number of MPCs [10]. Recent DL [15] methods estimate the position directly from CIRs or derive parameters such as ToA [16], [17]. However, they only work for the training environment.

We propose a DL-based ToA estimator, as well as a training concept with synthetic data including metrics that simplify its deployment. The novel concept generates comprehensive, realistic synthetic data to implicitly identify and mitigate errors in ToA estimation in LoS, OLoS, and NLoS radio propagation environments by learning a mapping of synthetic CIRs to ToAs. The key ideas are to simulate the distributions of all realistic CIRs and their corresponding FDPoAs as ToAs for a variety of typical 5G channel models and to use the interpolation capability of an adequate CNN to reduce the cost of the data acquisition process. We propose a metric of 4 parameters for QuaDRiGa that represent every possible realistic MPC distribution and enable the generalizability to unknown propagation environments, represented by different bandwidth, DS, Ricean kF, and SNR. Our experiments show that our DL approach outperforms state-of-the-art methods such as PEAK, IFP, and MUSIC at SNRs below -10 dB by about 17$\%$ on average. Unlike PEAK and IFP, our method is not limited to individual parameters as it learns all possible parameters from data. And unlike MUSIC, its performance does not deteriorate with the number of MPCS. In contrast to existing DL-based methods, our approach generalizes to unknown propagation environments and reduces data acquisition costs.

The rest of this paper is structured as follows. Sec. II reviews related work. Sec. III describes the problem. Sec. IV introduces our novel architecture. Sec. V describes our experiments and discusses the results. Sec. VI concludes.

II. Related Work

Related work estimated positions directly from CIRs [1], [15] or from channel parameters (CP) such as AoA [2], [3], RSS [4], ToA [2], [18] or combinations thereof [3]. Since these end-to-end methods require an expensive data acquisition campaign and estimate environment-specific positions and so do not generalize to other environments, we will only discuss model-(Sec. II-A) and data-driven (Sec. II-A) CP estimators.

A. Model-driven CP Estimation

Guvenc et al. [7] use the maximum peak of the correlation function as ToA. Similar threshold-based methods use fixed SNR and power values [8]. However, the handcrafted parameters only work reliably in LoS situations. Instead, the IFP method [9] estimates the peak based on the maximum gradient along the first rising edge. This helps to account for the peak detection errors caused by MPCs with similar delays. Although IFP is an improvement over PEAK, it still suffers from multipath and low SNRs. Thus, to estimate correct ToAs even in multipath scenarios, high-resolution algorithms were proposed [6], [19]: subspace-based algorithms such as ESPRIT [20], MUSIC [21], and RARE [12] or SAGE-based methods [13]. However, only the maximum likelihood-based RiMAX [14], [22] methods integrate diffuse multipath scattering effects into their model and so surpass others [23]. Although these super-resolution techniques increase the time-domain resolution and thus estimate ToAs more accurately in multipath environments, their computational complexity increases significantly, making them impractical for many real-world applications [24]. In contrast, our data-driven method only estimates a single relative ToA, thus keeping the computational complexity low, is not limited to a predefined number of paths and therefore generalizes to other environments.

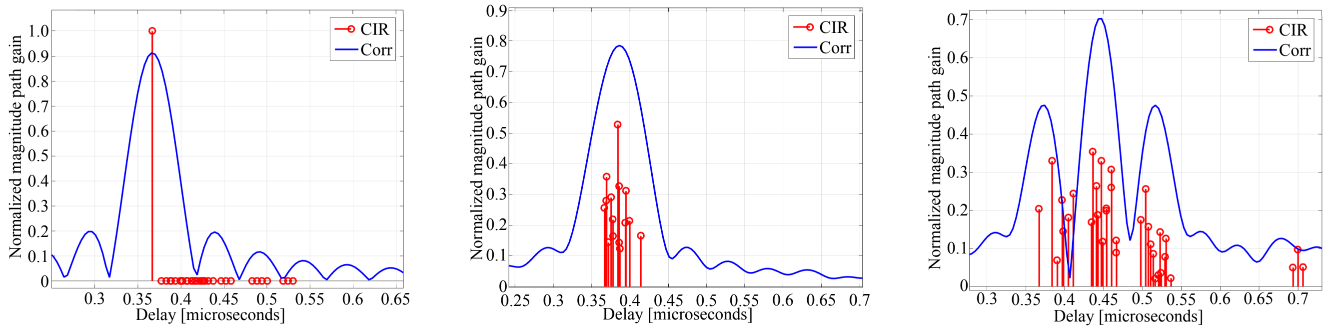
B. Data-driven CP Estimation

Others used supervised DL methods to extract CPs from CSI [16]. Wang et al. [25] trained a CNN to estimate AoAs from phase fingerprints. Comiter et al. [26] derive AoAs from two NNs that estimate the antenna beam. In contrast, we focus on ToA estimation as it typically yields higher accuracies in a downstream positioning task [3]. To the best of our knowledge, only Sun et al. [17] examined DL to directly estimate ToAs from CIRs. In (LoS) experiments with wired and radio transmission, they showed that their expensive data acquisition process offers higher accuracy and less computational effort at low SNRs than state-of-the-art methods at the cost of practicability. However, since they did not report important details about their experimental setup, the reproduction and quantitative comparison are impossible. Since their CNN architecture uses max pooling, that negatively impacts the time-critical feature extraction process [27], we do not evaluate their method. We derive our computationally efficient DL model on synthetic data to reduce data acquisition costs and to generalize it to combinations of synthetic and real random LoS and NLoS MPCs. We also report on reproducible parameters that we derive from real-world propagation scenarios and that best cover realistic channel models [28].

III. Problem Description

Typically, we estimate a ToA from a CIR that is extracted by decorrelating a received signal with a known pseudo-random sequence. Each MPC of a CIR describes the influence of path-loss and material interactions for a particular signal path. The ToA estimator identifies the delta time, i.e., the relative ToA, that corresponds to the FDPoA from a CIR. The distances between multiple transceiver lines are then determined using the speed of light to multilaterate a position.

However, CIRs differ considerably in a bandwidth limited channel, since the decorrelation leads to a significant overlap of the spatial information of the CIR and smears different impulses with one another. Thus, under multipath propagation the CIR contains many different (nondeterministic) MPCs. The limited temporal resolution limits the accuracy of ToA. Hence, extracting the information that represents the correct ToA is challenging and the expected ToA estimation performance depends on the overall channel statistics, which in turn depends on the environment and the deployment parameters. The root-mean-square DS and the *Ricean* kF characterize a CIR. These statistical parameters are derived from channel measurements and best describe the propagation conditions within any environment. DS and KF are both environment-dependent and are modeled by statistical distributions. The KF indicates how many MPCs negatively affect the FDPoA.



(a) No multipath. (b) Low delay spread. (c) High delay spread.

Fig. 2. CIR and related correlator output (Corr) of different delay spreads (red: unlimited bandwidth; blue: limited bandwidth).

The DS indicates the delay between the MPCs that influence the FDPoA most positively and negatively.

Figs. 2(a-c) show the effects of (bandwidth limited) DS on the magnitudes of three CIRs: (a) the CIR contains only a single LoS MPC (red) and its correlation signal Corr (blue); (b) shows a single cluster of CIRs that form a single peak in Corr. Instead, Fig. 2(c) shows multiple clusters that form three peaks in the magnitude of Corr w.r.t. the CIR.: the first cluster, an OLoS case, in which the FDPoA is less delayed but also weaker due to diffraction and transmission, see also the paths between $TX - RX_3$ in Fig. 1; and the second and third clusters, several NLoS cases, are higher delayed but are more powerful due to reflection and scattering, see also the paths between $TX - RX_2$ and $TX - RX_3$ in Fig. 1. We see that higher DS yields more MPCs clusters, and worsens the identification of the correct FDPoA. Of course, high DS and low KF values represent a worst-case situation to estimate a correct FDPoA. This becomes worse with lower SNRs.

IV. Method

A. Processing Pipeline of our Framework

Urban, industrial, and other environments with many absorbing, specular, and diffuse scattering and reflecting objects increase the variety of propagation paths and MPCs and thus increase the ToA estimation error. Bandwidth and transmission power limitation also worsen the situation. Our DL-based approach identifies correct FDPoAs even in these complicated scenarios, as it learns the complex spatial correlation of MPCs from snapshots of CIRs. Fig. 3 shows our pipeline: We train our models to map a CIR to a corresponding reference ToA. To train and evaluate our models, we use QuADRiGa [29] to generate training, validation, and test data with real-world channel statistics that represent predefined [28] propagation conditions for every possible environment. We sample different (μ and σ) combinations of KF and DS together with a varying SNR to generate a variety of realistic CIRs and reference ToAs.

B. Data Acquisition

3GPP TR38.90 [28] defines realistic reference channel models that describe small scale parameters (SSP), DS, KF, and Doppler and large scale parameters (LSP), path loss and shadowing. SSP and LSP define the statistical properties of corresponding CIRs with probability density functions (PDFs) for DS and KF to represent the properties of a random realistic environment w.r.t. LoS, OLoS, and NLoS conditions, the number of relevant reflectors, and the (distance to reflecting) objects around the transmitter. A detailed overview of the statistics for KF and DS are shown in Figs. 4a and 4b. Since KF and DS are typically correlated, we visualize their joint PDFs with pseudo 3D plots, wherein the probability is represented by the color, see Figs. 4c and 4d.

To generate realistic synthetic data we use the geometry-based stochastic channel model QuADRiGa [29]. QuADRiGa consists of a stochastic component, that creates a random propagation environment and random 3D positions at constant velocity of fixed scattering clusters within, and a deterministic part, that describes the interaction of transmitters and receivers within this environment over time. QuADRiGa's realism was validated based on real measurements in a coherent LTE Advanced Testbed [29], QuADRiGa Ind. in Fig. 4a. QuADRiGa determines the channel coefficients and yields CIR and ToA.

C. Datasets

To enable generalizability of DL models, we avoid environment and implementation-specific training datasets. Thus, we created a composite dataset that covers various scenarios as subsets. Each subset is generated using a specific SSP/LSP table that represents a realistic underlying channel statistic. We describe our synthetic *urban*, *indoor*, and *real-world* datasets with distributions of $\mu(KF)$, $\mu(DS)$, their σ, and SNR. The *urban* macro (`UMa`) scenario assumes high basestation (BS) towers, with high DS, and areas with narrow streets. The urban micro (`UMi`) assumes dense deployments of the BS, with medium DS, and rural areas. Fig. 4c visualizes the map that covers the channel characteristics of `UMi`. The *indoor* factory `InF` dataset with low DS represents typical indoor industrial (factory) applications [30]. Instead, indoor open office (`InO`) assumes open environments and includes more data [28]. Fig. 4d visualizes the map that covers the channel characteristics of InF. Fig. 5a shows the complete map of the channel characteristics `UMa`, `UMi`, `Inf`, and `InO`. We generated similar maps for our *real-world* data, see

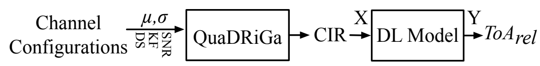
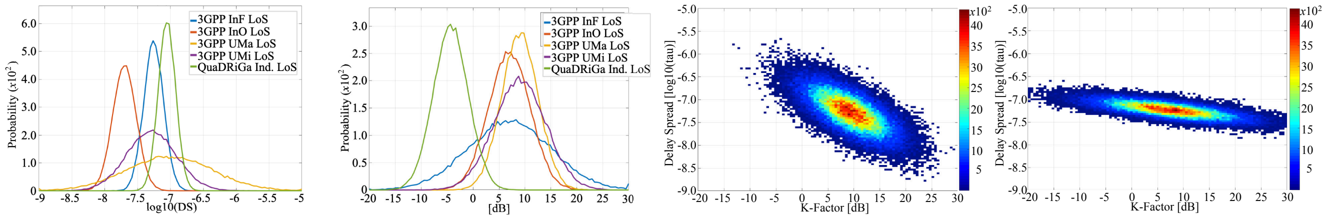


Fig. 3. Processing pipeline of our framework.



(a) PDF (DS) for 3GPP. (b) PDF (KF) for 3GPP. (c) Joint PDFs for UMi. (d) Joint PDFs for InF.

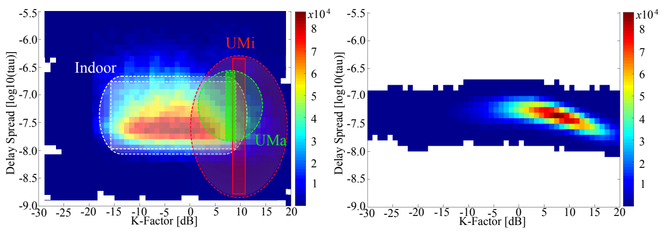
Fig. 4. PDFs and joint PDFs of DS and KF for the 3GPP (TR38.901) scenarios: `InF`, `InO`, `UMa`, `UMi`, and QuaDRiGa Ind.

Fig. 5b. We derived the statistics by analyzing data from a real measurement study of a typical indoor scenario. Our real-world system generates 200 CIRs per second on 16 RX for one TX in the ISM band of 2.4 GHz with a bandwidth of 80 MHz [15]. The real data are part of the subsets `InF` and `InO`, compare how the synthetic map in Fig. 5a covers the entire map of the real indoor scenario in Fig. 5b.

Since we vary the CIR properties within a certain environment, we enable the selection of the suitable training dataset for application-specific environmental conditions. Our synthetic and real maps visualize and ensure that the synthetic training data contain the channel properties of a real target environment. In this way, a DL model that is trained on the entire training data generalizes and we avoid complex scenario-specific training during a system deployment phase.

D. Data Pre-Processing

To evaluate our models on the trade-off between information gain, accuracy, and computational effort, we pre-process the input data in 4 variants: 2 work on 1D sequences of CIRs with dimensions of 60 samples over time (width, w=60 and height, $h \in [1, 2]$) and 2 work on 1D sequences of resampled CIRs with dimensions w=120 and $h \in [1, 2]$. We resample the input sequence X at p/q times the original sampling rate of 60 (with upsampling p=8 and downsampling q=1). So we insert zeros to increase the signal by p=8. We apply a FIR anti-aliasing lowpass filter to the upsampled signal (normalized cutoff frequency fc=$\pi/max(p, q)$ and gain p). We approximate the anti-aliasing filter with the Kaiser window method (filter order is $2 \times n \times max(p, q)$, w. n=50 and shape parameter β=5). We discard samples by q=1 to downsample the filtered signal. We shift the signal in time to compensate for the delay introduced by the FIR filter. In the case of h=1 we compute the magnitude vector of I, Q, whereas in the case of h=2 we use the raw I, Q vectors.



(a) Simulated data. (b) Real data.

Fig. 5. Distribution of channel characteristics for simulated and real data. The rectangles and circles classify `UMi`, `UMa`, and Indoor models spatially in this world map. Note that rectangles represent the area covered by the parameters μ(KF) and μ(DS), and the circles represent their corresponding σ.

E. Data-driven Main-Processing

The key idea is that a DL-based method learns to identify spatial correlations of MPCs of different propagation scenarios that cover any realistic channel configuration to provide correct FDPoA (i.e., ToA) predictions. We formulate the problem of mapping a CIR to a ToA as a supervised regression problem. During a training phase, the model learns to map the data sequence (CIR) to a corresponding ToA. At the inference time, the model then predicts a ToA from an unknown CIR.

Model Selection. In a preliminary study, we evaluated models such as Linear and Gaussian Process Regression, SmallNet [15], ResNet18 [31], and RNNs [31], CNNs [17], and their combination. In a large scale grid search[1], we optimize each architecture and parameters on each of the 4 input variants to find the architecture that does not require local pooling, minimizes under- and overfitting, and yields the lowest ToA error on stratified randomized training data [32]. Since our study revealed that our 1D-CNN, inspired by temporal CNNs [33], offers both the highest computing efficiency and the highest accuracy, we only discuss it in detail below.

Final Model Architecture. 100 1D-convolution filter kernels k detect 100 potential features and achieve the highest accuracy. A k=10×1 (for 60x1 input, respectively k=10×2 for 120x2) kernel slides alongside 99 others, directly on a 1D input sequence. They yield an output array, i.e., feature map fm, with dimensions: *input dimension*-ks+1. 4 conv. layers yield the highest accuracy. To transform an input sequence to the first conv. layer (*CL1*), n=100 filter kernels of size k=10x1 (respectively 20x2 for inputs with dim. 60x2 and 120x2) slide through the input sequence of length 60 for 51 steps (=60-10+1). This results in an output array of fm=51 × 100 of *CL1*. We apply the same number and size of k to process each *CL(1-4)* and obtain the following architecture: *CL1* (fm=100x51), *CL2* (fm=100x42), *CL3* (fm=100x31), and *CL4* (fm=100x24). Instead of a pooling we use a *dropout* layer (dropout rate dr=0.2) to prevent overfitting before we

[1] Grid Search for our 1D-CNN: Number of Conv. Layers: **4** $\in [1 : 1 : 20]$, residual layer in-between $\in$ [*yes*, **no**], k=**10** $\in [2 : 1 : 20]x1, x2$, number of kernels: **100** $\in [1 : 10 : 200]$, dr=**0.2** $\in [0 : 0.1 : 0.9]$, lr $\in$ [0.0001, **0.001**, ..., 1.0], activation $\in$ [**relu**, *sgd*], epochs $\in$ [early stopping w. patience=3, max. 1000], optimizer $\in$[**adam**, rmsprop]. Bold text highlights the configuration that yields the highest accuracy on **S5** that we use in Sec. V.

TABLE I
RESULTS OF OUR EXPERIMENTS S1: AWGN, S2: UMI, S3: UMA, AND S4: INDOOR.

	S1: AWGN Error [ns]					**S2: UMi Error** [ns]					**S3: UMa Error** [ns]					**S4: Indoor Error** [ns]					
	CEP_{50}	CEP_{75}	CEP_{95}	MAE	RMSE	CEP_{50}	CEP_{75}	CEP_{95}	MAE	RMSE	CEP_{50}	CEP_{75}	CEP_{95}	MAE	RMSE	CEP_{50}	CEP_{75}	CEP_{95}	MAE	RMSE	SNR
PEAK	0.90	1.10	1.39	1.24	1.36	1.74	2.73	2.74	2.66	2.92	2.91	3.31	3.76	3.45	4.36	3.35	3.99	4.84	4.70	5.84	+20
	1.82	2.11	3.62	3.45	3.93	2.47	3.55	4.96	4.45	4.64	3.43	4.29	5.52	4.86	5.42	4.84	4.94	6.35	5.64	6.76	+10
	5.12	7.11	11.23	8.54	12.67	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
IFP	**0.89**	**1.08**	1.32	**1.17**	**1.31**	1.53	**2.34**	2.83	2.75	2.89	2.46	3.45	4.56	4.16	5.34	3.33	4.12	5.84	5.42	6.66	+20
	1.45	2.08	2.45	2.23	2.67	2.91	3.63	3.97	3.51	3.84	4.36	5.32	7.31	6.50	7.35	5.45	6.86	8.45	7.55	8.74	+10
	3.23	4.34	5.23	4.65	6.56	4.76	5.66	6.87	5.34	7.97	9.25	10.58	12.81	11.63	13.14	9.53	11.33	12.85	12.74	14.66	0
MUSIC	0.91	1.12	**1.31**	1.27	1.34	1.65	2.43	**2.65**	**2.21**	2.87	3.85	4.27	5.47	4.79	6.58	4.86	5.46	6.38	5.96	7.86	+20
	1.51	2.11	2.45	2.23	2.67	2.67	3.34	4.79	3.65	**3.72**	4.36	6.14	8.47	7.74	8.25	5.55	7.87	9.46	8.66	9.54	+10
	3.56	4.12	5.89	5.78	7.34	4.43	5.76	6.23	6.21	8.34	9.84	11.52	14.36	13.41	14.41	10.92	12.76	15.99	14.79	15.45	0
DL	0.90	1.15	1.33	1.23	1.48	**1.62**	**2.34**	2.67	2.54	**2.73**	**1.92**	**2.46**	**3.61**	**3.25**	**3.61**	**2.27**	**2.64**	**3.88**	**3.76**	**3.85**	+20
	1.23	**1.34**	**2.23**	**2.12**	**2.43**	**2.23**	**2.87**	**3.54**	**3.43**	3.96	**2.36**	**3.01**	**3.85**	**3.58**	**3.90**	**2.73**	**3.16**	**4.28**	**3.98**	**4.13**	+10
	1.67	**2.78**	**2.96**	**2.94**	**3.75**	**2.47**	**3.32**	**4.51**	**3.78**	**4.65**	**2.52**	**3.36**	**4.53**	**4.26**	**4.47**	**2.91**	**3.46**	**4.77**	**4.25**	**4.45**	0
	2.32	**3.65**	**4.78**	**3.65**	**4.76**	**3.56**	**4.34**	**5.23**	**4.67**	**5.98**	**3.84**	**4.72**	**5.37**	**5.14**	**5.56**	**3.34**	**4.07**	**4.96**	**4.74**	**5.86**	-10
	3.45	**4.11**	**5.97**	**4.86**	**6.84**	**4.34**	**5.56**	**6.38**	**5.45**	**7.56**	**4.41**	**5.76**	**6.41**	**6.01**	**6.43**	**3.72**	**4.56**	**5.89**	**4.88**	**6.73**	-20

apply a fully connected layer to flatten its input. ADAM [34] and the root mean square error (RMSE) loss function optimize the regression layer to predict ToAs.

V. EVALUATION

We discuss the ToA error, i.e., identification error of the FDPoA in ns=0.30 m, in terms of mean absolute error (MAE), the circular error probabilities (CEP) of 50%, 75%, and 95%, and the RMSE for each experiment w.r.t. varying SNRs in Tables I and II [35]. We report best results in bold. Our grid search provides optimized versions of PEAK [7], IFP [9], MUSIC [2] [6], and DL. For fair comparability and reproducibility, we evaluate these publicly available methods, as preliminary studies showed the lowest computational effort (PEAK, IFP) and highest accuracy (IFP, MUSIC). For SNR<0 dB they did not yield plausible results. We visualize the errors as cumulative distribution functions (CDFs) and *error world map* graphs, i.e., a heatmap of ToA errors (in ns), see Fig. 7.

For all experiments, both (S)ynthetic and (R)eal-world, we split each dataset into 60% for training, 10% for validation, and 30% for testing based on a fixed random seed from evenly distributed random samples. For a fair comparison, we only show the results on the test datasets. We evaluate the ToA estimators on individual and combined synthetic datasets (Sec. V-A) and the importance of KF on their accuracy (Sec. V-B). We also evaluate their ability to interpolate missing data (Sec. V-C) and to generalize to the real-world (Sec. V-D).

A. Synthetic Experiments

We evaluate the ToA error of all methods for individual synthetic datasets with channel configurations: S1 (`AWGN`), S2 (`UMi`), S3 (`UMa`), and S4 (`InF`). Each of these datasets contains 3,202,000 training samples (1601×2000 uncorrelated QuADRiGa drops that provide random motion sequences with a constant acceleration of 2 $\mathrm{m/s^2}$) and 1,601,000 test samples (801×2000 uncorrelated sequences). We also combine S1 to S4 in dataset S5 to assess the generalizability of the methods.

[2]For MUSIC, we use the MDL algorithm [36] to determine the number of signals and we derive full-rank matrices according to Pillai et al. [37].

For each of the 5 experiments we optimized, trained, and evaluated all methods for different SNRs.

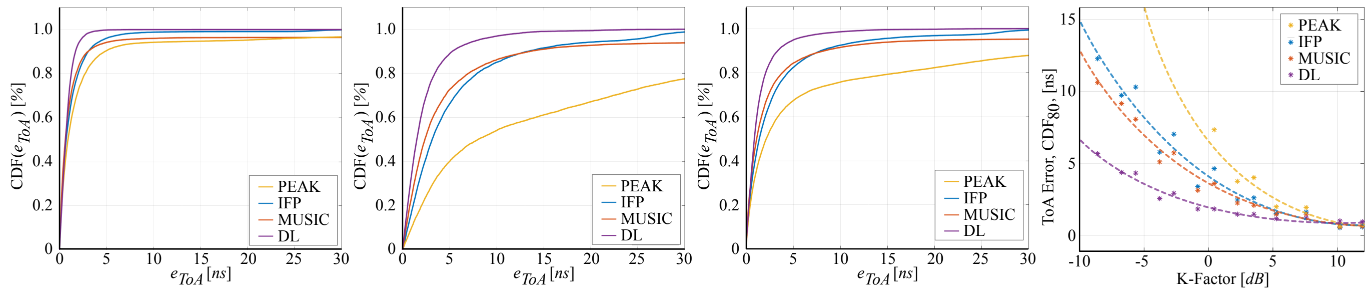
S1: `AWGN`. S1 represents a typical `AWGN` channel [28] with KF: μ=0.1, σ=1.1, and DS: μ=0.2, σ=2. At SNR$\geq$0, all methods perform similarly (SD=0.53 ns). At SNR=20, PEAK and IFP outperform DL. This is to be expected as both PEAK and IFP simply separate peaks (FDPoAs) from the noise floor, while MUSIC and DL also simply optimize their mapping. However, at lower SNRs, DL performs best, and with SNRs<0, only DL yields plausible results at all.

S2: `UMi`. Fig. 5a shows the channel configuration that S2 covers. The (red) rectangular area represents the μ of KF=$[8.4, -11]$ and DS=$[-8.82, -6.62]$, while the (red) circle represents the σ of KF=$[2.2, 7]$ and DS=$[0.03, 0.54]$.

On average, all methods provide lower accuracies on S2 than on S1, since the impact of KF and DS increases significantly. The results show that at SNR$\geq$0 all methods increase in error, but the error varies between the individual methods (SD=2.87). Interestingly, both PEAK and IFP keep up with MUSIC and DL at SNR=+20. However, PEAK yields implausible results at SNR$\leq$0. At SNR$\leq$+10, DL outperformed all other methods. This is caused by the NLoS density, the smaller KF, and larger DS values. We found similar effects in the synthetic experiments S3 to S5.

S3: `UMa`. Fig. 5a shows the channel configuration, that S3 covers, in green: μ of KF=$[7, 9]$ and DS=$[-7.76, -6.44]$ and the σ of KF=$[3, 5.7]$ and DS=$[-7.76, -6.44]$.

The error behavior of all methods is similar to that of S2. The ToA errors are slightly higher on average than for S1 and S2. This is caused by the increase in NLoS density, the decreasing KF, and much larger DS values. At SNR$\geq$0, all methods increase the error. However, their errors differ amongst them (SD=3.34 ns). Again, PEAK returns implausible results at SNR$\leq$0. In contrast to S1 and S2, DL always yields the highest accuracies. At SNR=0, IFP (7.35 to 13.14) and MUSIC (8.25 to 14.41) showed a significant increase in the RMSE. For DL there is no such effect (3.90 to 4.47). At SNR≤ 0 DL outperformed all others by about 160% (RMSE: IFP=7.35; DL=4.47). Again, at SNR<0, DL is the



(a) High KF, i.e., KF>10. (b) Low KF, i.e., KF≤0. (c) All KFs. (d) KFs versus ToA error.

Fig. 6. CDF of IFP, MUSIC, PEAK, and DL methods on S6.1 (a), S6.2 (b), and S6.3 (c), each with SNR= 0 dB for different KFs.

only method that yields plausible results.

S4: `InF`. Fig. 5a shows the channel configuration, in white: μ of KF=$[-15, 10]$, and DS=$[-8.25, -6.45]$ and σ of KF=$[0.2, 3.1]$ and DS=$[0.76, 0.44]$.

The ToA errors slightly increase on average over S1 to S3. Similar to S2 and S3, the errors of all methods increase almost linear from SNR=20 to SNR=0. This is caused by the very high NLoS density, the very small KF, and very large DS values. Again, at SNR≥0, IFP (8.74 to 14.66) and MUSIC (9.54 to 15.45) show a strong increase of the error, whereas PEAK returns implausible result, and DL shows no significant error (3.85 to 4.45). However, all results for each method w.r.t. SNR are less consistent than in S1 to S3 (SD=4.11 ns). Similar to S3, DL always yields the lowest errors. Although DL also slightly raises the error w.r.t. SNR, DL at SNR≥0 outperforms all others by more than 210% (RMSE: IFP=14.66; DL=4.45). And at SNR<0), only DL yields plausible results at all, as it separates MPCs from the noise floor.

S5: `Combination of AWGN, UMa, UMi, InF`. We combine S1 to S4 to create the new dataset S5, to investigate whether a DL model trained on S5 yields more accurate ToAs as it may learn the properties of all channel configurations. This time we cannot report the performance of PEAK, IFP, and MUSIC as they require an individual channel configuration to return plausible results. S5 consists of 12,808,000 samples for training and 6,404,000 samples for testing (60%/10%/30% split). Fig. 5a shows the entire area (S1 to S4) of data.

Interestingly, the S5 experiment yields, on average, an accuracy similar to that of S4. One reason for this is that S4 already covers a large part of the entire area. Thus, similar to S4, the error of DL increases almost linearly. Similar to S1 to S4, the results show that the error increases with increasing SNR. All errors are slightly higher in S5 than in S4 (RMSE from 5.04 to 5.78). This implies that more knowledge (data) does not necessarily provide more accurate estimates, as the data in S5 are much more diverse and sparse than in S4.

Conclusion w.r.t. Execution Times: S1 to S5 show that PEAK (inference time[3]of 0.7 ms per CIR) and IFP at high SNRs=20 perform quite well, as they simply detect FDPoAs. IFP and MUSIC perform similar at high to medium SNR (+20 to +10). However, we recommend IFP (1.9 ms) as it runs much faster than MUSIC (52,000 ms). DL yields similar accuracies at high SNRs, but it outperforms others at medium to low SNR (0 to -20) with acceptable inference times (5.3 ms). Hence, we recommend DL with decreasing SNR, if the channel configuration (i.e., environment) are known, since its first layers extract shapes and patterns of different FDPoAs best and its deeper layers distinguish them from OLoS and NLoS MPCs and the (diffuse) noise floor with little error.

B. Importance of KF in Indoor Environments

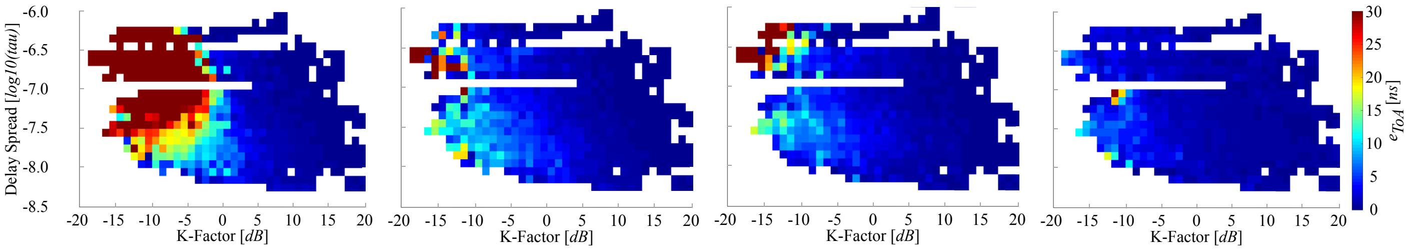
Results of S1 to S5 show that with a lower KF and higher DS, i.e., with stronger MPCs, the performance of all methods decreases. DS correlates with the bandwidth and does not affect our experiments with fixed bandwidth. To investigate ToA errors w.r.t. variations of only KF, we evaluate all methods on special subsets of the challenging S4 dataset: 3 intervals of KF split S4 to create 3 types of datasets: S6.1 (KF>10), S6.2 (KF≤0), and S6.3 (all KFs). For reasons of fairness and comparability, we have optimized and trained all methods for subsets of S6.x scenarios with KF≥0.

Figs. 6(a-c) show CDFs of the ToA errors (ns) of all methods on the 30% test datasets from S6.x. At KFs>10 (S6.1), all methods work equally well. In contrast, at KF≤0 (S6.2), DL performs better than the others. With all KFs (S6.3), the results are more accurate than with S6.2, as S6.3 contains many KF> 0, which has a positive effect on the overall result. However, S6.3 is still worse than S6.1. Fig. 6(d) shows the CDF$_{80}$ accuracy of the methods across different KFs. At KF>10 (right), all methods work similarly. At KF<7.5, DL always outperforms the others and Peak always yields the worst results. At KF<5, Music slightly outperforms IFP. At KF<0, PEAK already yields implausible results. We think that DL for S6 as with S1 to S5 uses its ability to precisely

[3]We measured the average inference time over all test samples of S1, processed with an Intel Core i7 1×3.6 GHz, excluding data load times.

TABLE II
RESULTS OF S5: COMBINED DATASET EXPERIMENT.

Method	Error [ns]					
	CEP$_{50}$	CEP$_{75}$	CEP$_{95}$	MAE	RMSE	SNR
DL	2.46	2.82	3.84	3.99	4.72	+20
	2.96	3.32	4.31	4.47	4.56	+10
	3.17	3.63	4.87	4.66	5.26	0
	3.43	3.95	4.92	5.06	6.91	-10
	3.81	4.47	5.99	5.27	7.45	-20



(a) PEAK. (b) IFP (c) MUSIC. (d) DL.

Fig. 7. ToA error map that shows the correlation of KF, DS, and SNR on S5 and an SNR = 0 dB.

extract, memorize, and interpolate complex inter- and intra-dependencies of MPCs in CIRs. However, it is unclear how well DL performs when the noise floor completely hides the MPCs. Figs. 7(a-d) support these results and show the ToA error distribution as a heatmap over KF and DS. We conclude that a lower KF and a higher DS result in a higher error for all methods, with DL having the lowest error variance. The diagrams from left (a) to right (d) show that the error decreases with the complexity of the method from PEAK to DL.

C. Generalization

PEAK, IFP, and MUSIC cannot generalize to unknown channel characteristics as they require individual adaptation to work correctly. To examine the generalizability of DL to unknown data, we consider two experiments: G1 and G2.

G1: General Interpolation Ability. We investigate whether DL interpolates between 2 general channel configurations (see Fig. 8). We use the test data (-7.75<DS<-7.5 and -10>KF>10) from the framed area and the remaining data for the training (-7.75>DS>-7.5 and -10>KF>10) from S6 at SNR≥0. Thus, we test whether DL interpolates between the left and right side (KF) and the upper and lower side (DS).

For SNR=20, DL returns lower errors (CEP_{50}=1.23, CEP_{75}=2.56, CEP_{95}=3.33, MAE=2.89, RMSE=4.84) than for SNR=0 (CEP_{50}=1.87, CEP_{75}=2.99, CEP_{95}=3.78, MAE=3.89, RMSE=5.66). The results suggest that DL interpolates quite well between known data distributions, as the error increases only slightly compared to the original results from DL on S6 (RMSE at SNR=20: +0.22; RMSE at SNR=0: +0.40). Thus, DL (re)constructs (new) channel models if it learns surrounding information in the training phase.

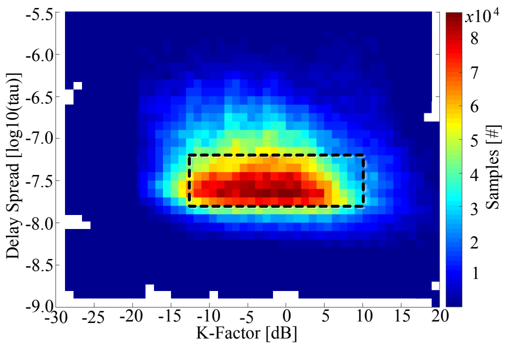


Fig. 8. Joint PDFs for the generalization benchmark (dataset split into data covered by the black rectangle and the remaining data).

G2: Specific Interpolation Ability. To also investigate whether DL processes unknown channel configurations, we trained with the S2 (UMi) dataset and tested with S3 (UMa), see Fig. 5a. Thus, as S2 (UMi) and S3 (UMa) overlap (similar to G1), we test if DL interpolates between them.

For SNR=20, DL returns lower errors (CEP_{50}=1.35, CEP_{75}=2.51, CEP_{95}=3.31, MAE=3.76, RMSE=3.87) than for SNR=0 (CEP_{50}=1.41, CEP_{75}=2.98, CEP_{95}=3.41, MAE=3.72, RMSE=4.52). Hence, DL may interpolate and reconstruct a gap in a data distribution. The results also show that DL predicts almost as accurate (SD<0.16 ns) as in S3.

D. Real-World Applicability

In OLoS and NLoS scenarios, it is expensive to obtain a reference FDPoA. Thus, this experiment evaluates whether DL (trained on realistic synthetic data) predicts accurate ToAs on real-world data to reduce the data acquisition cost. We derive a high KF (>0) and a low DS (<7.75) from the real-world LoS-NLoS *Rectangles* dataset [38]. From there we estimate channel properties, configure QuADRiGa, and generate data. Fig. 5b shows that the synthetic environment covers the real-world data completely. We also derive the reference ToA_{rel} of LoS measurements of the *Rectangles*.[4] We randomly select 10,000 LoS samples for training and 82,724 for tests. The training procedure is unchanged, only the maximum training epochs (= 100) are fixed for better comparability.

DL (pre-trained on synthetic data) on synthetic test data shows slightly lower errors (CEP_{50}=0.98, CEP_{75}=1.21, CEP_{95}=1.65, MAE=1.42, RMSE=2.13) than for S5. The same DL model results in higher accuracies (CEP_{50}=0.51, CEP_{75}=0.72, CEP_{95}=1.08, MAE=0.53, RMSE=0.64) on real test data than on synthetic test data. We think this is as the synthetic scenario is more complex than the real scenario and covers all types of the real FDPoAs. However, retraining the entire model provides the highest accuracies (CEP_{50}=0.001, CEP_{75}=0.001, CEP_{95}=0.007 MAE=0.004, RMSE=0.022). Instead, resetting or freezing parameters or layers results in worse accuracy and unstable training. And the training from scratch, with randomly mixed (batches of synthetic and real)

[4]We subtract the arrival (start) times of the correlation windows CIR^{1}_{abs} of RX1 from the arrival times CIR^{1-12}_{abs} of the (time) synchronized sets of 12 CIRs with optical reference transmitter-receiver distances d^{1-12}_{ref} to calculate the transmission duration CIR^{2-12}_{rel}. Then, we determine the reference FDPoAs: $ToA^{2-12}_{rel} = (d^{2-12}_{ref} - c \cdot CIR^{2-12}_{rel})/c$, w.r.t. the speed of light c.

data, also does not provide (significant) improvements on either test dataset. Hence, the results indicate that already a retraining of the entire network with few real LoS labels (10,000/200 Hz=50 s) yields the best results.

VI. Conclusion

We propose a novel data-driven ToA estimator that extracts optimal FDPoAs even on real-world CIRs. Our experiments show that even simple DL architectures estimate ToAs accurately and outperform the state-of-the-art. The difference becomes significant (26% higher accuracy) in scenarios with multipath. This even holds for SNRs<-10 dB: 17% on average. We also show that DL trained on synthetic data does not require (but does benefit) fine-tuning to work with real data.

Future work must examine the applicability of our method to various real-world channel configurations w.r.t. OLoS and NLoS propagation. In any case, our DL-based approach lowers all errors in the ToA estimation indoors and outdoors.

Acknowledgments

This work was supported by the Bavarian Ministry for Economic Affairs, Infrastructure, Transport and Technology through the Center for Analytics Data Applications (ADA-Center) within the framework of "BAYERN DIGITAL II".

References

[1] G. Mendoza Silva, J. Torres-Sospedra, and J. Huerta, "A Meta-Review of Indoor Positioning Systems," *Sensors J.*, vol. 19, no. 20, pp. 4507–4531, 2019.

[2] J. Rovňáková and D. Kocur, "Toa estimation and data association for through-wall tracking of moving targets," *European Association for Signal Processing*, vol. 2010, no. 1, pp. 1–11, 2010.

[3] F. Zafari, A. Gkelias, and K. K. Leung, "A survey of indoor localization systems and technologies," *IEEE Comm. Surveys & Tutorials*, vol. 21, no. 3, pp. 2568–2599, 2019.

[4] Radiocommunication Sector of ITU-R Spectrum Management, "Comparison of time-difference-of-arrival and angle-of-arrival methods of signal geolocation," tech. rep., SM.2211-2, 2018.

[5] Z. He, Y. Ma, and R. Tafazolli, "Improved high resolution TOA estimation for OFDM-WLAN based indoor ranging," *IEEE Wireless Comm. Letters*, vol. 2, no. 2, pp. 163–166, 2013.

[6] X. Li and K. Pahlavan, "Super-resolution TOA estimation with diversity for indoor geolocation," *IEEE Trans. on Wireless Comm.*, vol. 3, no. 1, pp. 224–234, 2004.

[7] I. Guvenc, S. Gezici, and Z. Sahinoglu, "Ultra-wideband range estimation: Theoretical limits and practical algorithms," in *Proc. Intl. Conf. on Ultra-Wideband*, vol. 3, pp. 93–96, Hannover, Germany, 2008.

[8] C. Falsi, D. Dardari, L. Mucchi, and M. Z. Win, "Time of arrival estimation for UWB localizers in realistic environments," *European Association for Signal Processing*, vol. 34, no. 1, pp. 32–82, 2006.

[9] G. Hofmann and M. Breiling, "Device and method for determining a time of arrival of a receive sequence," 2009. US Patent 7,627,063.

[10] M. Pourkhaatoun and S. A. Zekavat, *A review on TOA estimation techniques and comparison*. Hoboken, New Jersey: Wiley, 2011.

[11] A. J. Van der Veen, M. C. Vanderveen, and A. Paulraj, "Joint angle and delay estimation using shift-invariance techniques," *IEEE Trans. on Signal Processing*, vol. 46, no. 2, pp. 405–418, 1998.

[12] C. M. S. See and A. B. Gershman, "Direction-of-arrival estimation in partly calibrated subarray-based sensor arrays," *IEEE Trans. on Signal Processing*, vol. 52, no. 2, pp. 329–338, 2004.

[13] B. H. Fleury, M. Tschudin, R. Heddergott, D. Dahlhaus, and K. I. Pedersen, "Channel parameter estimation in mobile radio environments using the sage algorithm," *J. Selected Areas in Comm.*, vol. 17, no. 3, pp. 434–450, 1999.

[14] A. Richter, *Estimation of radio channel parameters: Models and algorithms*. Ilmenau, Germany: ISLE, 2005.

[15] A. Niitsoo, T. Edelhäußer, and C. Mutschler, "Convolutional neural networks for position estimation in tdoa-based locating systems," in *Intl. Conf. Indoor Positioning and Indoor Navigation*, (Nantes, France), 2018.

[16] T. O'shea and J. Hoydis, "An introduction to deep learning for the physical layer," *IEEE Trans. on Cognitive Comm. and Networking*, vol. 3, no. 4, pp. 563–575, 2017.

[17] H. Sun, A. O. Kaya, M. Macdonald, H. Viswanathan, and M. Hong, "Deep learning based preamble detection and TOA estimation," in *Global Comm. Conf.*, pp. 1–6, Waikoloa, HI, 2019.

[18] T. Feigl, T. Nowak, M. Philippsen, T. Edelhäußer, and C. Mutschler, "Recurrent neural networks on drifting time-of-flight measurements," in *Intl. Conf. Ind. Positioning and Ind. Navigation*, Nantes, France, 2018.

[19] R. Wang, C. U. Bas, Z. Cheng, T. Choi, H. Feng, Z. Li, X. Ye, P. Tang, S. Sangodoyin, J. Gómez-Ponce, *et al.*, "Enabling Super-Resolution Parameter Estimation for mm-Wave Channel Sounding," *IEEE Trans. on Wireless Comm.*, vol. 19, no. 5, pp. 3077–3090, 2020.

[20] M. Haardt, "Efficient one-, two-, and multidimensional high-resolution array signal processing," *Ph. D. Thesis, Shaker Verlag, Aachen*, 1996.

[21] A. Jakobsson, A. L. Swindlehurst, and P. Stoica, "Subspace-based estimation of time delays and Doppler shifts," *IEEE Trans. on Signal Processing*, vol. 46, no. 9, pp. 2472–2483, 1998.

[22] B. Hanssens, E. Tanghe, D. P. Gaillot, M. Liénard, C. Oestges, D. Plets, L. Martens, and W. Joseph, "An extension of the RiMAX multipath estimation algorithm for ultra-wideband channel modeling," *Wireless Comm. and Networking*, vol. 2018, no. 1, pp. 164–176, 2018.

[23] D. P. Gaillot, E. Tanghe, P. Stefanut, W. Joseph, M. Lienard, P. Degauque, and L. Martens, "Accuracy of specular path estimates with ESPRIT and RiMAX in the presence of measurement-based diffuse multipath components," in *Proc. European Conf. on Antennas and Propagation*, pp. 3619–3622, 2011.

[24] J. Salmi, A. Richter, and V. Koivunen, "Enhanced tracking of radio propagation path parameters using state-space modeling," in *European Signal Processing Conf.*, pp. 1–5, Florence, Italy, 2006.

[25] W. Xuyu, G. Lingjun, M. Shiwen, and P. Santosh, "CSI-Based Fingerprinting for Indoor Localization: A Deep Learning Approach," *IEEE Trans. Vehicular Technology*, vol. 66, no. 1, pp. 763–776, 2017.

[26] M. Comiter, M. Crouse, and H. T. Kung, "A Structured Deep Neural Network for Data Driven Localization in High Frequency Wireless Networks," *J. Comp. Netw. and Comm.*, vol. 9, no. 1, pp. 21–39, 2017.

[27] C. Pelletier, G. I. Webb, and F. Petitjean, "Temporal convolutional neural network for the classification of satellite image time series," *J. Remote Sensing*, vol. 11, no. 5, pp. 523–548, 2019.

[28] 3GPP Radio Access Network Working Group, *et al.*, "Study on channel model for frequencies from 0.5 to 100 GHz (Release 16)," tech. rep., 3GPP TR 38.901, 2020.

[29] S. Jaeckel, L. Raschkowski, K. Börner, and L. Thiele, "QuaDRiGa: A 3-D multi-cell channel model with time evolution for enabling virtual field trials," *IEEE Trans. on Antennas and Propagation*, vol. 62, no. 6, pp. 3242–3256, 2014.

[30] K. Haneda, J. Zhang, L. Tan, G. Liu, Y. Zheng, H. Asplund, J. Li, Y. Wang, D. Steer, C. Li, *et al.*, "5g 3gpp-like channel models for outdoor urban microcellular and macrocellular environments," in *Vehicular Technology Conf. (VTC)*, pp. 1–7, Montreal, Canada, 2016.

[31] T. Feigl, S. Kram, P. Woller, R. H. Siddiqui, M. Philippsen, and C. Mutschler, "RNN-aided human velocity estimation from a single IMU," *Sensors*, vol. 20, no. 13, pp. 3656–3690, 2020.

[32] V. N. Vapnik, "An overview of statistical learning theory," *Trans. Neural Networks*, vol. 10, no. 5, pp. 988–999, 1999.

[33] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," *arXiv:1803.01271*, 2018.

[34] K. Diederik P. and B. Jimmy, "Adam: A method for stochastic optimization," *arXiv:1412.6980*, 2014.

[35] F. Potortì, A. Crivello, P. Barsocchi, and F. Palumbo, "Evaluation of indoor localisation systems: Comments on the ISO/IEC 18305 standard," in *Intl. Conf. Ind. Positioning and Ind. Nav.*, Nantes, France, 2018.

[36] D. B. Williams and V. K. Madisetti, *Detection: Determining the number of sources*. Boca Raton, FL: CRC Press Inc., 1999.

[37] S. U. Pillai and B. H. Kwon, "Forward/backward spatial smoothing techniques for coherent signal identification," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 37, no. 1, pp. 8–15, 1989.

[38] A. Niitsoo, T. Edelhäußer, E. Eberlein, N. Hadaschik, and C. Mutschler, "A deep learning approach to position estimation from channel impulse responses," *Sensors J.*, vol. 19, no. 5, p. 1064, 2019.