Challenges in Platform-Independent UWB Ranging and Localization Systems

Laura Flueratoru
laura.flueratoru@tuni.fi
University Politehnica of Bucharest, Romania
Tampere University, Finland

Elena Simona Lohan
elena-simona.lohan@tuni.fi
Tampere University
Tampere, Finland

Dragoș Niculescu
dragos.niculescu@upb.ro
University Politehnica of Bucharest
Bucharest, Romania

ABSTRACT

The Ultra-Wideband (UWB) technology has grown in popularity to the point in which there are numerous UWB transceivers on the market that use different center frequencies, bandwidths, or hardware architectures. At the same time, efforts are made to reduce the ranging and localization errors of UWB systems. Until now, not much attention has been dedicated to the cross-platform compatibility of these methods. In this paper, we discuss for the first time the challenges in obtaining platform-independent UWB ranging and localization systems. We derive our observations from a measurement campaign conducted with UWB devices from three different developers. We evaluate the differences in the ranging errors and channel impulse responses of the devices and show how they can affect ranging mitigation methods customized for one device only. Finally, we discuss possible solutions towards platform-independent UWB localization systems.

CCS CONCEPTS

• **Hardware → Wireless integrated network sensors**; • **Networks → Location based services**.

KEYWORDS

ultra-wideband, ranging, indoor localization

ACM Reference Format:
Laura Flueratoru, Elena Simona Lohan, and Dragoș Niculescu. 2022. Challenges in Platform-Independent UWB Ranging and Localization Systems. In *16th ACM Workshop on Wireless Network Testbeds, Experimental evaluation & CHaracterizatiion (WiNTECH '22), October 17, 2022, Sydney, NSW, Australia.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3556564.3558238

1 INTRODUCTION

Ultra-wideband (UWB) devices are increasing in popularity for their centimeter-level ranging and localization capabilities. At this date, there are at least four major smartphone brands equipped with UWB chipsets and at least six developers of UWB platforms. Therefore, localization systems must ensure interoperability among different device brands [22]. For this purpose, the FiRa™ Consortium [4] provides specifications and certifications that ensure interoperability between different UWB solutions.

UWB systems can provide centimeter-level ranging and localization accuracy in line-of-sight (LOS) conditions but their performance is affected by obstacles between the transmitter (TX) and the receiver (RX), which is known as non-line-of-sight (NLOS) propagation. Multiple techniques have been proposed for dealing with errors caused by NLOS propagation [8, 21, 24]. Most of them rely on the channel impulse response (CIR) of the signal, since it offers rich information about the propagation path of the signal. However, so far, little to no attention has been dedicated to ensuring the proposed methods perform well on different UWB platforms.

To the best of our knowledge, we are the first to look at the challenges faced by ranging and localization systems when dealing with devices from different vendors. We derive our observations from a measurement campaign using UWB devices developed by three companies: Qorvo, TDSR, and 3db Access. We acquired measurements with pairs of devices from each brand at the same locations under different LOS and NLOS conditions. The measurements are made open-source [6] to facilitate future research in platform-independent UWB localization systems. The goal of the paper is to evaluate the differences in terms of distance errors and CIRs between different device models under the same propagation conditions (i.e., the same location, furniture arrangement, crowdedness, and, if applicable, obstacles).

First, we illustrate the danger in ignoring the cross-platform compatibility of a centralized localization system. We evaluate the performance of a neural network (NN) trained to achieve a good distance error prediction for one device when it is tested on measurements from different devices (acquired at the same locations), simulating the scenario in which the system disregards the users' device models. Our results show that an error mitigation technique that disregards the device model of incoming measurements may actually *degrade* the final performance of the system.

Second, we look at the root causes of cross-platform compatibility issues in UWB systems and we identify four main challenges:

(1) The same environmental conditions lead to different distance errors for different devices, especially in NLOS conditions. Although great efforts are made to reduce ranging and localization errors through NLOS detection and mitigation, not all types of obstacles introduce significant ranging errors and, perhaps more interestingly, they do not introduce the same error in all devices even under the same conditions.
(2) The same (NLOS) environmental conditions can lead to a different time of arrival (TOA) estimation in consecutive

WiNTECH '22, October 17, 2022, Sydney, NSW, Australia

ACM ISBN 978-1-4503-9527-4/22/10.
https://doi.org/10.1145/3556564.3558238

measurements even for the same device. Therefore, NLOS errors depend not only on the obstacle and environment, but also on the hardware and its TOA estimation algorithm.

(3) CIRs acquired by different platforms at the same locations have different lengths, shapes, and statistics. This can be due to different center frequencies, pulse shapes and bandwidths, nonlinearities imprinted by the different front-end architectures on the CIR, antenna propagation characteristics, etc.

(4) Different vendors provide different types of diagnostics (e.g., signal power, noise level), sometimes using incompatible or undisclosed units of measurement, which can hinder the cross-platform compatibility of localization methods based on such diagnostics [21].

Finally, we use these observations to make recommendations towards platform-independent UWB localization systems.

2 RELATED WORK

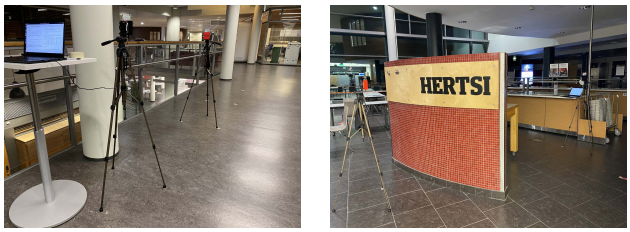
Measurements with different UWB devices at the same locations have been previously performed in [7, 11, 16, 19]. However, these studies were focused on comparing the performance of the devices in terms of ranging and/or localization accuracy or energy consumption. In this paper, we look not only at the *average* ranging error of each device, but also at the error distribution at selected locations, which gives insight into their different operating modes. To the best of our knowledge, we are the first to also compare the CIRs of different devices and to highlight how these differences can prevent the cross-platform compatibility of ranging and localization systems. We also compare other platforms than in the previous works and provide one of the few open-source datasets of this kind.

The localization accuracy of UWB-based systems can be improved by filtering or correcting individual distance measurements used in multilateration algorithms and/or by filtering or estimating locations directly. Distance errors can be reduced through NLOS detection and mitigation [24], data-driven TOA estimation algorithms [5], or models trained for distance error prediction [8, 21]. Other approaches estimate directly the location based on CIR features [12]. In recent years, ML-driven approaches for ranging and localization have taken precedence over statistics-based techniques [17]. However, all the cited solutions have been trained on data from only one device model. Previous work has also revealed that models trained in one environment have issues in adapting to different environments [3]. Given the diversity of UWB devices on the market, we deem necessary to evaluate the differences between various UWB devices that might affect the performance of error mitigation methods when applied on unknown devices.

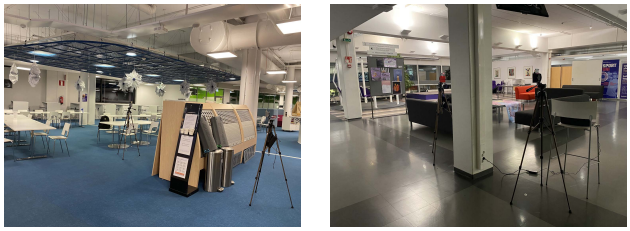
3 DATA COLLECTION

In this section, we briefly present the experimental setup used to collect the data. The dataset is open-source and a detailed description of all the locations in which we acquired measurements can be found in the repository [6]. The goal of the measurement campaign is to verify the consistency of CIRs and ranging errors from different devices at the same locations.

We performed ranging measurements between pairs of devices belonging to each brand in several locations at Tampere University.



(a) LOS (b) NLOS with half wall



(c) NLOS with bar refrigerator (d) NLOS with pillar

Fig. 1: Examples of environments from the measurement campaign.

Table 1: Device settings used in the experiments: the center frequency (f_c), the pulse bandwidth (B_p), and the CIR sampling period (T_s^{CIR}).

Device	f_c [GHz]	B_p [MHz]	T_s^{CIR} [ns]
DW3000	6.5	600	1
TDSR	4.3	620	0.061
3db	6.5	380	1

We targeted LOS and NLOS scenarios, for the latter using as obstructions walls, pillars, furniture, a TV screen, and room divider panels. We performed measurements in nine different spaces which included meeting rooms, offices, corridors, or cafeterias. Fig. 1 shows some of the locations in which we acquired measurements.

We used three models of UWB devices from different developers: the Qorvo (formerly Decawave) DW3000, the 3db Access 3DB6830C, and the TDSR P452A. For brevity, we will further refer to the devices, respectively, as DW3000 (or DW), 3db, and TDSR. Table 1 summarizes the characteristics and settings of the devices. DW3000 and 3db devices have the same center frequency of 6.5 GHz but different pulse bandwidths of 600 MHz and 380 MHz, respectively. TDSR devices have a different center frequency of 4.3 GHz but a bandwidth close to the DW3000 (620 MHz). DW3000 and 3db devices are compliant with the IEEE 802.15.4z standard [1], while TDSR devices implement a proprietary physical interface.

We took several measures to ensure that the devices from different manufacturers are placed at the same locations for each measurement. First, we mounted the devices on tripods such that the center of their antennas are aligned with respect to the tripod. Second, we marked the locations of the tripod legs on the floor for each test point. We estimate that, during the measurement process, there can be a maximum error of 1 cm between the antenna centers of two different devices.

At each test point, we kept a minimum recording time of 30 s in order to capture the CIR variations over multiple measurements. Because of some software limitations, in a part of the measurements

the ranging update rate of DW3000 was lower than the one of 3db and TDSR devices (2 s vs. approx. 300 ms, respectively). In total, our database contains approx. 30,000 measurements with 3db and TDSR devices each and 2,800 measurements with Decawave. The proportion of LOS and NLOS measurements is approx. 50/50.

We need to preprocess the CIRs in order to compare them. TDSR and 3db devices provide the real part of the CIR, while the DW3000 stores the complex CIR. To unify the three representations, we use the absolute value of the CIR. The TDSR CIRs have a sampling period of 61 ps so we downsampled them to 1 ns using linear interpolation. We also rescale the CIRs of each each device to have amplitudes between $[0, 1]$, taking into account the global minimum and maximum of the entire dataset for each device. The rescaling enables us to compare the CIR statistics from different devices and does not change the statistics of CIRs from the same dataset (i.e., from one manufacturer).

4 CROSS-PLATFORM PERFORMANCE OF AN ERROR PREDICTION MODEL

In this section, we want to highlight the danger in disregarding the cross-platform compatibility of a centralized localization system (LS). In many LSs, the anchors collect ranging measurements from a user, then transmit them to a central server, which further processes them and estimates the user's location [8]. Let us consider that the processing step consists of a NN which predicts the distance error[1] based on the CIR and uses the prediction to refine the estimated distance. Since the data collection process is difficult for one device, let alone for multiple brands of devices, we can assume that the NN was trained on a dataset acquired with only one type of device, as it is frequently the case in the literature [5, 21, 24]. However, in a public space, the system might have to deal with measurements from users with different UWB devices. So we evaluate the possible outcome if the NN naively tried to predict errors based on measurements from devices that it was not trained on. We will compare the performance of the model when applied on measurements from the device used for training vs. on measurements from different devices, both *acquired at the same locations*.

Let $D = \{\langle \boldsymbol{x}_i, y_i \rangle\}_{i=1}^{N}$ be our training set, where $\boldsymbol{x}_i$ is the feature vector (a portion of the CIR) from the feature space $\mathcal{X}$ and y_i is an instance of the target variable Y (the distance error) defined on the domain $\mathcal{Y}$. Our goal is to learn the function $f : \mathcal{X} \rightarrow \mathcal{Y}$, that maps a CIR to its distance error.

We use a fully connected network with three dense layers, each using a rectified linear unit (ReLU) activation function plus an additional fully connected layer. Each layer uses 256 filters. We used the Focal-R loss function proposed in [23] for imbalanced regression to deal with the long-tailed distribution of the target variable. We used an input size of 40 CIR samples aligned to the TOA, keeping 30 samples before and 10 samples after the TOA. Although using more samples after the TOA might show benefits, we preferred to use the minimum common CIR length among all devices in order to avoid zero-padding shorter CIRs.

[1] Among the possible error mitigation strategies, we chose predicting the distance *error* to better control the target variable distribution compared to directly predicting the (correct) distance. NLOS detection is another popular method but, as per Section 5.1, it often discards many useful measurements.

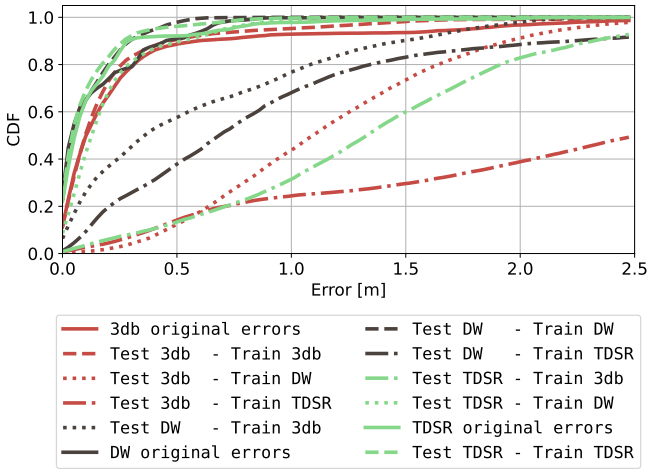


Fig. 2: CDF of the original distance errors ($|e_i|$) vs. the corrected errors ($|\hat{e}_i - e_i|$) using models trained on the same or different device(s).

We remind that our dataset includes multiple measurements performed with the three brands of devices from the same physical location, from multiple locations inside the same room, and from different rooms. To avoid overfitting, we include measurements from different locations in the training, test, and validation sets. We select at random one location for the test set, one for the validation set, and assign the rest to the training set. We repeat this procedure 10 times, generating 10 different splits each time. The same split is used for all devices, meaning that split number $k \in \{1, ..., 10\}$ from each device contains exactly the same combinations of locations and rooms inside the train, validation, and test sets. Therefore, if a model M trained on the set $D_k^{\text{train}}(d_1)$ from device d_1 achieves a good performance on the test set $D_k^{\text{test}}(d_1)$ but a bad performance on $D_k^{\text{test}}(d_2)$ from device d_2, the difference in the performance will be due to the hardware since $D_k^{\text{test}}(d_1)$ and $D_k^{\text{test}}(d_2)$ contain measurements acquired at exactly the same locations.

We evaluate the initial distance errors and the errors after correction using the model's predictions. Let $e_i = \hat{d}_i - d_i$ be the distance error of measurement $i \in \{1, ..., N\}$, i.e., the difference between the measured distance $\hat{d}_i$ and the true distance d_i. The average initial (absolute) error is $\mu_{\text{init}} = \sum_{i=1}^{N} |e_i|$, where we aggregate the errors over all test set splits. The average corrected (absolute) error is $\mu_{\text{corr}} = \sum_{i=1}^{N} |\hat{e}_i - e_i|$, where $\hat{e}_i$ is the error predicted by the model for measurement i. If the prediction perfectly matches the true error, then the distance error is completely mitigated.

Fig. 2 shows the CDF of the original distance errors of all devices vs. the corrected errors using the trained NNs. First, when using the same device for training and testing, the error after correcting the distance using the model's prediction is 22–27 % smaller in the mean and 31–42 % smaller in the standard deviation compared to the initial error. This shows that the model is able to generalize to unknown locations from similar environments.

When applying the error prediction model on an unknown device, i.e., that the model was not trained on, the average *corrected* error is always higher than the initial error, in most cases 3–8× larger. This shows that models not trained on a particular device

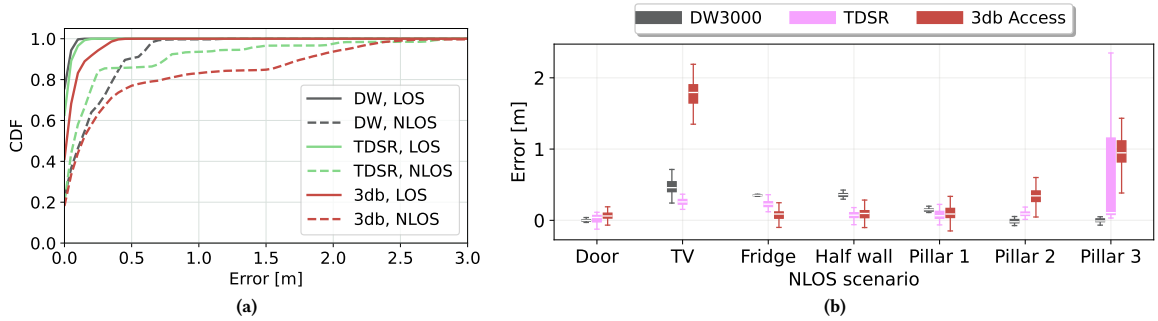


Fig. 3: (a) CDF of errors in LOS and NLOS for all devices. (b) Distribution of errors from selected locations with different obstacles.

might *degrade* the performance of a ranging system if we disregard information about the platforms used for training and testing. Therefore, the cross-platform compatibility of error mitigation techniques in LSs should be taken into account from the design stage.

5 CHALLENGES

In this section, we analyze the collected data and highlight four main challenges in obtaining cross-platform compatible UWB-based ranging and localization systems.

5.1 Same Conditions, Different Distance Errors for Different Devices

NLOS propagation is defined by the absence of a *visual* direct propagation path between the TX and the RX. Obstructions can introduce distance errors in UWB measurements in two ways. First, objects can attenuate the signal traveling through the direct path, making it indistinguishable from noise. When this happens, the device can incorrectly estimate the TOA as corresponding to a later reflection which has a higher amplitude than the direct path, causing a time (and distance) error. Second, some obstacles can decrease the propagation speed of signals traveling through them. In this case, the direct signal will arrive later than it would have arrived through air, introducing a distance error.

Fig. 3a shows the cumulative distribution function (CDF) of distance errors obtained with the three devices in LOS and NLOS scenarios. Indeed, the errors obtained in NLOS scenarios are always larger than those obtained in LOS. However, it is worth noting that at least 50 % of NLOS measurements are under 15 cm, which is usually deemed an acceptable ranging error [19]. Even if we developed a perfect NLOS detection algorithm, discarding or assigning lower weights to measurements with small errors might decrease the accuracy of the localization system, especially in cases with few anchors available. Therefore, a binary LOS/NLOS classification might not always be the best strategy for distance error mitigation.

Table 2 shows the median error of each device for all types of obstacles considered here. While it is true that certain types of obstacles (e.g., concrete walls, TV screen) generally introduce higher errors than others (e.g., door) [11, 20], we noted that not all devices yielded the same ranging errors even when placed under exactly the same conditions. Fig. 3b shows the distribution of distance errors (illustrated as box plots using Tukey's definition) obtained by the three devices at one location for several obstacles. There is a high variability in the distance errors obtained by different devices at the same location and, sometimes, a high spread in the errors obtained by a *single* device at one location.

Table 2: Median distance errors with various obstacles.

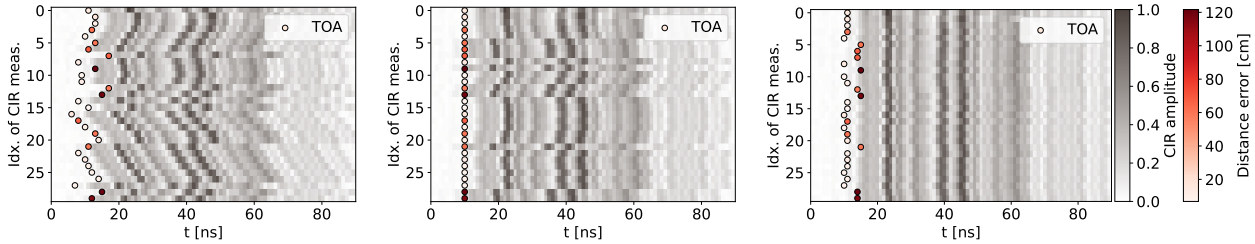
	Median distance error [m]					
Device	LOS	TV screen	Fridge	Door	Pillar	Half wall
DW	0.06	0.45	0.31	0.10	0.57	0.08
TDSR	0.00	0.22	0.33	0.08	0.28	0.01
3db	−0.02	1.89	0.00	−0.23	0.42	0.05

5.2 Same Conditions, Different TOAs

It is perhaps useful to explain the root cause of the differences in the measured distance of various devices. Since the distance between two devices is computed based on the (round-trip) time of flight between the TX and the RX, any error in the TOA estimation will also yield a distance error.

Although the TOA estimation algorithms used by the devices are closed-source, the "Applications of the IEEE 802.15.4 standard" document [9] describes some approaches that are likely followed by the manufacturers. In LOS, the maximum peak of the CIR usually corresponds to the direct path, so the (correct) TOA is straight-forward to obtain. In NLOS, however, the direct path can be attenuated by an obstruction and later paths, which correspond to longer travel times, often have higher amplitudes. Therefore, the receiver needs to also consider other peaks in the vicinity of the strongest one as possible candidates for the first path. Popular approaches to implement the back-search include a sequential linear cancellation scheme [18] or threshold-based techniques [10]. Both techniques are model-based and require assumptions about the propagation conditions, such as the power ratio between two consecutive paths, the number of MPCs to be subtracted, the maximum peak to earlier peak ratio, or the peak to average power ratio. These methods perform well in LOS propagation but can lead to large errors in NLOS scenarios due to the inability of the fixed parameters to adapt to different propagation conditions.

We found that the most common reason for differences in ranging accuracy between the devices is the different TOA estimation, especially in the case of direct paths that have low amplitudes (close



(a) Unaligned CIRs (b) CIRs aligned to the TOA (c) CIRs aligned to the maximum cross-correlation index

Fig. 4: The plots show stacked CIRs (with their amplitude encoded through the color depth) acquired during a series of measurements in which there was a TV screen between two DW3000 devices. The circle markers represent the TOA estimated by the hardware and its color encodes the distance error of that measurement. Fig. 4a shows a portion of the unaligned CIRs (i.e., as dumped in the buffer by the device), Fig. 4b shows the CIRs aligned to the TOA, and Fig. 4c shows the CIRs aligned to the maximum value of the cross-correlation function computed between the first CIR and all of the following.

to the noise threshold). In our measurements, the "global" CIR shape in NLOS propagation is relatively stable for a single device and it is only the estimated TOA that fluctuates in some NLOS scenarios.

To illustrate this, let us consider a series of CIRs acquired with one pair of devices from the same brand in a NLOS scenario with a TV screen between the transmitter and the receiver. The devices are left unmoved during the experiment. We use the DW3000 for this example, but all devices present similar patterns. Fig. 4 shows series of 30 CIRs stacked vertically with different alignments. The MPCs from each CIR have color-encoded amplitudes and the TOA is denoted by a circle marker. The color of the circle marker encodes the distance error of the measurement corresponding to that CIR.

Fig. 4a shows the "raw" CIR alignment. In theory, the TOA can occur at any sample in the internal buffer which stores the CIR. In practice, we found that all DW3000 TOAs in our dataset occur between samples 710 and 750 in a quasi-random manner. In Fig. 4a, batches of consecutive CIRs appear to be shifted versions of each other, but the pattern changes every 5–10 measurements. In this representation, we cannot distinguish a clear pattern between CIRs that lead to large errors or between the MPCs of consecutive CIRs.

In most works, the CIRs are cropped around the TOA identified by the device, as in Fig. 4b. This provides a convenient representation for LOS/NLOS classification or for distance error prediction, since CIRs are aligned to a common denominator and deviations from it (e.g., the maximum peak occurring earlier or later in the buffer, shorter or longer tails) can indicate different error magnitudes. Using this representation, we could believe that different CIR shapes are correlated with different magnitudes of ranging errors.

Instead, if we align the CIRs according to the index which yields the maximum cross-correlation value, i.e., according to their similarity, we obtain Fig. 4c. In this figure, it is evident that, in fact, consecutive CIRs acquired in the same scenario are almost identical. What differs in this representation is the estimated TOA. The measurements with the lowest distance errors have the earliest TOA. However, because these early paths have amplitudes close to the noise threshold, they are ignored in some measurements and the TOA is identified as corresponding to a later reflection. Note that, since the devices perform two-way ranging, they must identify the correct TOA during both the poll *and* the response messages. In the plot, we see only the CIRs of the initiator, but large errors can also be caused by an incorrect TOA estimation at the responder. Therefore, the devices can experience different errors even at the same location because of the different SNR of the first path (which is influenced by the hardware architecture and signal processing steps) or because of the internal TOA estimation algorithm.

5.3 Same Conditions, Different CIRs for Different Devices

At first sight, CIRs seem convenient for platform-independent learning problems, since they provide a representation of the paths through which the signal travels from the TX to the RX. Therefore, we would perhaps expect CIRs acquired by different devices at the same location to be similar. However, we will see why this is not the case. We focus on two main differences: the effective duration and the shape of CIRs.

5.3.1 Effective duration. One crucial difference between CIRs acquired by different devices is their length. Let us call the *effective duration* of the CIR the portion starting from the estimated TOA until the end of buffer in which CIRs are stored, since this is usually the portion of most interest. The buffers have a length of 1016 samples for DW3000, 1632 samples for TDSR, and 256 samples for 3db. However, because we downsampled the TDSR from 61 ps to 1 ns, they will have a shorter (effective) duration than the other devices. In practice, we found that TDSR devices capture, on average, the paths which arrive within 17 ns of the TOA. In comparison, DW3000 and 3db devices have an average effective length of 278 ns and 125 ns, respectively. This difference is evident in Fig. 5, which shows an example of CIRs acquired with the three devices at the same location in one LOS scenario.

Different effective CIR lengths can cause issues in NNs which receive as input the raw CIR from different platforms. If the networks are designed to work on CIRs with a pre-defined length, shorter CIRs must be padded until the desired length, which can change the input sample distribution and the model's performance on those inputs. Alternatively, recursive NNs can be used on inputs of different lengths.

5.3.2 CIR shape. The received signal can be modeled as [13]:

$$r(t) = s(t) * h(t) + n(t), \tag{1}$$

where $s(t)$ is the transmitted pulse which is convolved with the CIR $h(t)$. The signal $n(t)$ represents sensor or environmental noise (so not related to the propagation path), usually modeled as zero-mean white Gaussian noise.

The CIR can be decomposed into K multipath components (MPCs) with delays τ_k, $k = 1, ..., K$ and amplitudes α_k and a stochastic process $\nu(t)$ which results in diffuse multipath caused by scattering and diffraction:

$$h(t) = \sum_{k=1}^{K} \alpha_k \delta(t - \tau_k) + \nu(t). \quad (2)$$

The estimated CIR, denoted by $\hat{h}(t)$, is obtained by decorrelating the received signal $r(t)$ with the known template pulse $s(t)$.

Because we acquired measurements with the three platforms at the same location, we would perhaps expect to see MPCs with the same delays τ_k in their CIRs. However, because the devices have different center frequencies and/or pulse bandwidths, there will be different constructive or destructive interference patterns reflected in $\nu(t)$ or even in the observed delays of the main MPCs τ_k. In addition, the amplitudes α_k are influenced by the different front-end circuits. The various signal processing components (e.g., low-noise amplifiers, mixers, automatic gain controls, analog filters, analog-to-digital converters) and digital processing can introduce different linear and/or non-linear distortions depending on the architecture. Therefore, even if the channel conditions and environment are the same, we will, in fact, see different patterns in the CIRs acquired by different platforms.

To characterize CIRs acquired at the same location with different devices, we look at the average number of main peaks (or MPCs) and the average delay between the first and the last significant peaks, which indicate whether we can identify the same main MPCs in all CIRs and how long it takes until their amplitudes decay to an insignificant level. We also compute the energy, mean excess delay (MED), and root-mean square (RMS) delay spread of the CIRs, which have been previously used to characterize CIRs in LOS/NLOS detection problems [17]. The MED and RMS delay spread are, respectively, the first and second moments of the power delay profile of the signal and characterize its delay statistics. The RMS delay spread captures the temporal dispersion of the signal's energy.

We search for the number of significant MPCs in each CIR, which we define as the peaks that exceed 25 % of the maximum amplitude of the CIR with a minimum time separation between peaks of 2 ns (to avoid detecting peaks belonging to the same path). We compute the average number of significant peaks (N_p) and the average time delay between the first and the last significant peak (μ_δ). The peak search is performed on the raw CIR, while the energy, MED, and RMS delay spread are computed on the CIR starting from the TOA until the end of the buffer, to mitigate the influence of the TOA index in the raw CIR buffer on the CIR statistics.

Fig. 5 shows the most significant peaks in a triplet of CIRs acquired at the same location with different devices and Table 3 shows the average CIR statistics over all locations. 3db CIRs have more significant MPCs than the other two devices and their amplitudes take longer to decay, reflected in a higher μ_δ and energy. TDSR CIRs have the shortest duration and therefore capture the fewest significant peaks, so they have the lowest energy, MED, and RMS delay spread.

In the literature, a low signal energy and high delay statistics have been associated with signal attenuation occurring in NLOS propagation [17]. However, if CIRs acquired in the *same* conditions with different brands of devices have, on average, different statistics, these can introduce issues in error mitigation methods customized for one device but applied on different brands of devices.

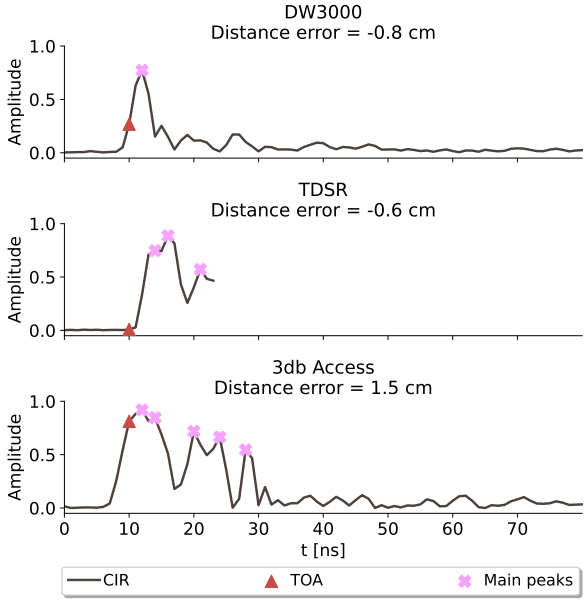


Fig. 5: CIRs and their main peaks acquired with each device at the same location in a LOS scenario.

Table 3: Statistics of the CIRs: number of significant paths (N_p), delay between the first and the last significant path (μ_δ), energy (E, quantized), mean excess delay (τ_{MED}), and RMS delay spread (τ_{RMS}).

Device	N_p	μ_δ [ns]	E [-]	τ_{MED} [ns]	τ_{RMS} [ns]
DW3000	3	11.2	2.4	13.5	21.7
TDSR	2	5.3	1.6	8.2	3.0
3db	5	24.4	8.5	13.3	13.8

5.4 Different Devices, Different Diagnostics

Since the CIRs are long (256–1632 samples for the devices we used), for some applications it can be too time consuming and computationally expensive to process them. Therefore, some works instead propose using other diagnostics provided by the devices (for instance, power or noise figures) for ranging/localization error mitigation [21]. However, this can pose issues for platform-independent localization systems because not all devices provide the same metrics and often not in the same units of measurement.

For instance, the DW3000 provides the power, maximum amplitude, and phase of arrival (POA) computed on the preamble for regular ranging and, additionally, the power and POA computed on the scrambled timestamp sequence for the secure ranging mode. The manufacturer provides formulas for converting the power to dBm [15]. TDSR devices provide the maximum value in the leading edge (LE) window of the received CIR and the noise level. However, the unit of measurement of these parameters is not specified in the documentation [14], so it is not clear if these parameters could be compared, for instance, with the ones of the DW3000 chipset.

TDSR devices also provide the noise amplitude and a coarse and a precise estimate of the range, of the distance error, and of the tag's velocity. The 3db chip provides the peak and the LE amplitudes (but not expressed in dBm) and other additional diagnostics [2].

Given the different diagnostics provided by the devices and the lack of a common measurement unit for the power and noise figures, it would be difficult to use this kind of information in a platform-independent UWB localization system. Therefore, the manufacturers should invest more effort to provide unified metrics for additional diagnostics.

6 DISCUSSION

As we showed in Section 4, error mitigation techniques that disregard the hardware with which incoming measurements were acquired can actually *degrade* the final accuracy of the system for devices that the system was not customized for. Therefore, error mitigation techniques developed for only one UWB hardware should implement a fallback plan to avoid degrading the accuracy of measurements acquired by other devices. An alternative that requires further research is to develop error mitigation solutions that can work across multiple platforms. Although different platforms have different CIRs and ranging errors under the same conditions, there is a common reason for all errors: the incorrect TOA estimation. Therefore, there might exist a set of device-agnostic CIR features related to the estimated vs. the correct TOA that could yield the correct distance (or TOA) errors across multiple devices. However, even if a device-agnostic set of features were found for a set of devices, it is unlikely to perform equally well on an unknown hardware.

An alternative to centralized solutions is to implement error correction only at the edge and use models customized for each target device. However, in this case, more work is needed to create environment-independent models [3], since it is unreasonable to collect data from every new location with every device on the market. Data-driven TOA estimation shows promising results [5] but it must be lightweight enough to be implemented on-chip and fast enough to minimize clock drift errors.

7 CONCLUSIONS

In this paper, we provided an overview of the challenges faced in obtaining platform-independent UWB ranging or localization systems. We derived our observations from a database of measurements acquired with three brands of UWB devices at exactly the same locations. We evaluated the differences in their ranging errors, CIRs, and diagnostics. We show that applying error mitigation models on devices not included in the training phase might result in a severe degradation of the ranging accuracy.

The cross-platform compatibility of error mitigation methods should not be taken for granted; instead, ranging and localization systems deployed in environments where users can have different UWB platforms must take into account cross-platform compatibility from the design phase. Future research should focus more on learning methods that can perform equally well on a wide range of devices or that degrade gracefully in the presence of an unknown device. Alternatively, error mitigation techniques could be moved to the edge and customized on the target device with the caveat that they must be environment-independent.

ACKNOWLEDGMENTS

The authors gratefully acknowledge funding from European Union's Horizon 2020 Research and Innovation programme under the Marie Sklodowska Curie grant agreement No. 813278 (A-WEAR: A network for dynamic wearable applications with privacy constraints, http://www.a-wear.eu/). This work does not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of its content.

REFERENCES

[1] 2020. IEEE Standard for Low-Rate Wireless Networks. *Amendment 1: Enhanced Ultra Wideband (UWB) Physical Layers (PHYs) and Associated Ranging Techniques (IEEE Std 802.15. 4z)* (2020).
[2] 3db Access AG. 2018. 3DB6830D User Guide. Version 4.0.
[3] Valentín Barral, Carlos J Escudero, José A García-Naya, and Pedro Suárez-Casal. 2019. Environmental cross-validation of NLOS machine learning classification/mitigation with low-cost UWB positioning systems. *Sensors* 19, 24 (2019), 5438.
[4] FiRa Consortium. 2022. Retrieved 19 August 2022 from firaconsortium.org
[5] Tobias Feigl, Ernst Eberlein, Sebastian Kram, and Christopher Mutschler. 2020. Robust ToA-Estimation using Convolutional Neural Networks on Randomized Channel Models. In *IPIN*. IEEE, 1–8.
[6] Laura Flueratoru. 2022. Dataset: Ultra-wideband ranging measurements acquired with three different platforms (Qorvo, TDSR, 3db Access). https://doi.org/10.5281/zenodo.6984698
[7] Laura Flueratoru, Silvan Wehrli, Michele Magno, Elena Simona Lohan, and Dragos Niculescu. 2021. High-accuracy ranging and localization with ultra-wideband communications for energy-constrained devices. *IEEE IoT-J* (2021).
[8] Jaron Fontaine, Matteo Ridolfi, Ben Van Herbruggen, Adnan Shahid, and Eli De Poorter. 2020. Edge inference for UWB ranging error correction using autoencoders. *IEEE Access* 8 (2020), 139143–139155.
[9] IEEE P802.15 Working Group for Wireless Personal Area Networks (WPANs). 2014. Application of IEEE Std 802.15.4.
[10] Ismail Guvenc, Sinan Gezici, and Zafer Sahinoglu. 2008. Ultra-wideband range estimation: Theoretical limits and practical algorithms. In *IEEE Int'l Conf. on UWB*, Vol. 3. IEEE, 93–96.
[11] Antonio Ramón Jiménez and Fernando Seco. 2016. Comparing Decawave and Bespoon UWB location systems: Indoor/outdoor performance analysis. In *IPIN*. IEEE, 1–8.
[12] Sebastian Kram, Maximilian Stahlke, Tobias Feigl, Jochen Seitz, and Jörn Thielecke. 2019. UWB channel impulse responses for positioning in complex environments: A detailed feature analysis. *Sensors* 19, 24 (2019), 5547.
[13] Erik Leitinger, Paul Meissner, Christoph Rüdisser, Gregor Dumphart, and Klaus Witrisal. 2015. Evaluation of position-related information in multipath components for indoor positioning. *IEEE JSAC* 33, 11 (2015), 2313–2328.
[14] TDSR LLC. 2020. TDSR RangeNet Application Programming Interface (API) Specification. Version 320-0313I.
[15] Decawave Ltd. 2021. DW3000 User Manual. Version 1.1.
[16] G MacGougan, K O'Keefe, and R Klukas. 2009. UWB ranging and ranging measurement accuracy. *Measurement Science and Technology* 20, 9 (2009), 0957–0233.
[17] Stefano Marano, Wesley M Gifford, Henk Wymeersch, and Moe Z Win. 2010. NLOS identification and mitigation for localization based on UWB experimental data. *IEEE JSAC* 28, 7 (2010), 1026–1035.
[18] Yihong Qi, Hisashi Kobayashi, and Hirohito Suda. 2006. On time-of-arrival positioning in a multipath environment. *IEEE Trans. on Vehicular Technology* 55, 5 (2006), 1516–1526.
[19] Antonio Ramón Jiménez Ruiz and Fernando Seco Granja. 2017. Comparing Ubisense, Bespoon, and Decawave UWB location systems: Indoor performance analysis. *IEEE TIM* 66, 8 (2017), 2106–2117.
[20] Anthony Schenck, Edwin Walsh, Jonas Reijniers, Ted Ooijevaar, Risang Yudanto, Erik Hostens, Walter Daems, and Jan Steckel. 2018. Information Theoretic Framework for the Optimization of UWB Localization Systems. In *IPIN*. 1–8.
[21] Lorenz Schmid, David Salido-Monzú, and Andreas Wieser. 2019. Accuracy assessment and learned error mitigation of UWB ToF ranging. In *IPIN*. IEEE, 1–8.
[22] Maximilian Schuh, Hannah Brunner, Michael Stocker, Markus Schuß, Carlo Alberto Boano, and Kay Römer. 2022. First Steps in Benchmarking the Performance of Heterogeneous Ultra-Wideband Platforms. In *CPS-IoTBench*. IEEE, 34–39.
[23] Yuzhe Yang, Kaiwen Zha, Yingcong Chen, Hao Wang, and Dina Katabi. 2021. Delving into deep imbalanced regression. In *Int'l Conf. on Machine Learning*. 11842–11851.
[24] Kegen Yu, Kai Wen, Yingbing Li, Shuai Zhang, and Kefei Zhang. 2018. A novel NLOS mitigation algorithm for UWB localization in harsh indoor environments. *IEEE Trans. on Vehicular Technology* 68, 1 (2018), 686–699.