Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Institut für
Technische Informatik und
Kommunikationsnetze

TotTernary: A wearable platform for social interaction tracking

Master Thesis

Andreas Biri

abiri@ethz.ch

Computer Engineering and Networks Laboratory
Department of Information Technology and Electrical Engineering
ETH Zürich

Supervisors:
Pat Pannuto
Prof. Dr. Prabal Dutta
Prof. Dr. Lothar Thiele

December 21, 2018

Acknowledgements

I would like to express my gratitude to Pat Pannuto for his supervision and guidance throughout this thesis. His wealth of practical experience and in-depth technical knowledge enriched our conversations and helped tremendously in refining this work. I have strongly benefited from his dedication to his work and his personal interest into the research questions at hand. The constant encouragement to discuss, challenge and test proposed ideas is extremely stimulating and inspired many fruitful exchanges and constructive trains of thought.
Furthermore, the entire research group of Lab11 was of great help during the thesis and enriched my stay at the University of California in Berkeley immensely. In particular, I would like to thank William Huang and Branden Ghena for their advice and feedback while designing and implementing this thesis as well as the technical knowledge and experience they shared with me.

I am extremely grateful to Prof. Dr. Prabal Dutta for the opportunity to join his group for the duration of this thesis and the wealth of experiences I could gather along the way. I will always remember this academically and culturally enriching time and take away countless invaluable memories.
I would further like to show my appreciation to Prof. Dr. Lothar Thiele for enabling me to pursue the thesis abroad and the support he provided throughout this time. Without his help, this amazing adventure would not have been possible.

This work has been partially supported by a grant of the Swiss Study Foundation, an institution which supports distinguished students at Swiss universities.

Abstract

Commercial ultra-wideband radios leverage the availability of accurate timestamps to estimate distances based on two-way time of flight measurements with centimeter accuracy and hence enable high-precision localization. Their superior performance compared to previous technology encourages the use of such devices for social interaction tracking and the study of human behaviour in various environments with high fidelity. To enable the observation of cohorts under real-life circumstances, domain scientists require an infrastructure-free and configurable solution which satisfies both the severe space and energy constraints common for such applications. Currently, no system can satisfy lifetime requirements for deployments of more than a week and achieve the decimeter accuracy sought by sociologists and epidemiologists while respecting mobility and latency demands of diverse populations.

TotTernary provides a mobile, accurate, responsive, and reliable platform for ranging measurements which allows users to gather both distance and position information with decimeter accuracy. Measuring only 61 mm x 35 mm and weighing 7.7 g, it integrates two radios to achieve both low-power neighbour discovery and direct user interactions using Bluetooth Low Energy as well as precise ultra-wideband ranging measurements. The system leverages a novel energy-efficient ranging protocol with linear message complexity to achieve lifetimes of up to 39 days. We show how the platform can be tailored to varying scenarios by adjusting the ranging fidelity and the update rate and demonstrate the use of dynamic adaptions to changing environments. Furthermore, leveraging both antenna as well as frequency diversity, we demonstrate that the median ranging error can be reduced by up to 86% and that our system is capable of highly reliable and consistent measurements with as little as 14.8 cm of ranging error in the 99th percentile and a 90% confidence interval of 11.3 cm.

This work presents the first infrastructure-free, long-term research platform which enables domain scientists to measure human interactions with previously unprecedented accuracy and flexibility.

Contents

Chapter 1

Introduction

"The contrast between genetic and environmental, between nature and nurture, is not a contrast between fixed and changeable. It is a fallacy of biological determinism to say that if differences are in the genes, no change can occur." — Richard C. Lewontin [1]

The human species is considered to be one of the most successful developments in evolutionary history and yielded thousands of different races and cultures. While there are multiple physical factors which facilitated the development progress, a primary factor which enabled the advance was the development of tight social networks and complex interactions between members. By building highly structured networks involving high-frequency contacts, humans became capable of extended collaboration, were able to aggregate cultural knowledge, and exchanged information beyond their direct kin [2]. However, despite this evident success and an extensive research effort over the past thousand years, the underlying reasons for how and why humans develop to such a vast variety of heterogeneous individuals despite near-equivalent backgrounds is still open for debate.
The early Greeks already debated about the origin of this difference: While Plato proposed that the organism's *nature* (i.e. genetics) is the source of its structure and function, his student Aristotle advocated that the human being is rather shaped by its *nurture* [3]. Since then, generations of researchers have investigated this topic and attempted to unravel the true origin of human behaviour. As the human brain is an enormously complex system and consists of countless factors which influence its development, determining and validating the importance of individual effects is challenging.

Current research splits from the dichotomous separation between *nature* and *nurture* and rather embraces a developmental systems approach in which dynamic, co-actional processes form our behaviour [3]. Such reasoning focuses on a rather probabilistic model based on the constellation of organismic and environmental variables [4]. This shift advocates for long-term, diversified, and data-driven studies to provide the required measurements for behaviour analysis and factor separation to enable a more complete, empirical model.

1.1 Motivation

Up until recently, such studies were hard to conduct, difficult to quantify objectively, severely restricted in scale and extent, and often lacking long-term data for quantitative analysis. Before the technological advances in wearable technology in recent years, human behaviour research historically relied almost exclusively on either self reporting and public surveys [5, 6] or external observations in heavily controlled settings such as laboratories [7]. Inherently, those experiments are highly influenced by the chosen cohort and the cooperation of participants [7], are difficult to verify and scale, and often lack fine-grained quantitative metrics [6]. They further suffer from biases due to the supervised environment [8] as well as the Hawthorne effect, in which individuals modify their behaviour upon being aware of observation [7].
Social scientists hypothesise that wearable sensors could help to gather real-world data and avoid the variance induced by laboratory experiments in previous studies [8]. It has been shown that such automated methods using wearable sensors achieve higher reliability and accuracy compared to self reporting [9]. Therefore, researchers started to leverage modern smartphones and advanced RF technology for more scalable and diversified measurement methods.

Wearable sensors have been actively used in research deployments for person-to-person interaction with many studies in recent years [2, 10]. New advances in mobile sensor technology and the increasing spread of mobile devices enable data gathering at unprecedented scales [11, 12]. Infrastructure-based systems have been proposed which provide sufficient resolution in time and space [13, 14] but severely restrict wide-spread deployment due to large deployment and maintenance costs. However, while there are multiple available mobile systems for infrastructure-free environments using Infrared [10], RFID [12, 15], Bluetooth [6, 16] or custom RF signals [2, 17], these current systems are almost exclusively limited to binary proximity detection. Apart from one notable exception [5], even the most recent study [18] which targets close interactions of household members was restricted to proximity instead of fine-grained distance measurements. In addition to spacial imprecision, traditional systems furthermore suffer from fixed and low time resolutions in the order of dozens of seconds to minutes.

Research in various fields such as epidemiology, sociology, and robotics [5] has shown that high-fidelity interaction distance measurements could result in new insights by creating models based on real data. Fine-scaled temporal and spatial resolution is significant for tracking human interaction [19]. While proximity detection inherently limits the range to a person's immediate surroundings, information about other people at room- or building-level extends the contact network and shows how the entire periphery changes over time.

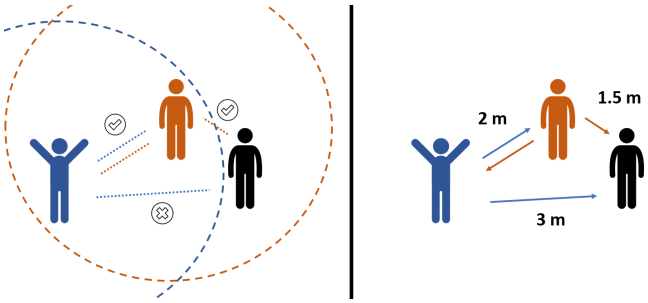

Figure 1.1: Proximity-based detection (left) does not provide any additional information about either close nor distant interactions, whereas distance measurements (right) show detailed connection networks.

The additional dimension which such distance measurements make accessible can allow research applications to improve in multiple fields:

- **Accuracy:** Distance measurements enable unmatched precision in profiling human interactions and allow fine-grained separation between contacts based on the actual distance. This permits the distinction of bins [20] such as touching distance, conversation, peripheral, same room, and same house solely based on a single measurement.
- **Differentiation:** High spatial and temporal resolution enables the detection of differences between interactions due to factors such as culture, race, age, and sex [21]. This opens up a new dimension for human contact studies and might provide an improved understanding of local and global variations for fields such as disease spreading, racial tensions, stress, and verbal communication [5].
- **Reliability:** Distinctions based on signal strength are known to be highly unreliable and strongly vary both over time and due to changing environments [22]. This makes such techniques ill-suited for mobile tracking and might result in neighbours constantly drifting in and out of range over short distances. Distance estimation based on time-of-flight measurements however benefit from longer ranges and high spacial precision.
- **Flexibility:** A fixed range limitation, as applied by previous work [2, 17], is not required. This enables a complete comparison of all interactions and adjustable thresholding into range bins later on in post-processing, delivering unparalleled flexibility to adjust to different scenarios.

A mobile, infrastructure-free and ranging-based platform for social interaction studies could significantly help researchers to gather new insights into how humans behave in real-world environments, interact inside networks, and develop behaviours over time.

1.2 Key challenges

As with most sensor platforms, the primary objective is to gather high-fidelity data in an accurate and reliable manner. To maximize the usefulness of the system during deployments, we also have to include additional considerations such as adaptability to different scenarios, lifetime requirements, and form factors. While an analysis of high-frequency networks such as sports team requires sampling rates in the order of Hertz and generates a large volume of data, long-term surveillance of elderly people might include energy-efficient, automated notifications to caretakers upon the detection of unusual behaviour. This leads to the following key challenges which shape our design considerations:

- **Customization towards scenario:** To enhance usability, the system should be capable of addressing a wide span of potential applications. The design space therefore includes both infrastructure-based and infrastructure-free deployments and should be able to provide researchers with options for both deployment types. Furthermore, it must support measurements over an extended period of time in relation to the duration of interest by optimizing its energy efficiency. In addition, it should exhibit a range of tuning capabilities to configure features such as update rate, accuracy, discovery latency, and robustness.
- **Reliable performance:** To provide data with a sufficient validity for researchers, the distance estimations must be reliably available over time and assure a high level of both accuracy and precision. This strongly influences the range of suitable technologies as described later on.
- **High accessibility:** As applications might require real-time measurement updates and the availability of constant monitoring, such an interface should be provided in a convenient manner using common communication equipment. Additionally, the platform must also be capable of storing data locally to make it accessible at a later time without relying on external support; this enables extensive analysis and simplifies deployment scenarios.
- **Unobtrusiveness:** To achieve a high acceptance and reduce the burden on the wearer, the form factor is strongly restricted both in terms of size and weight and should be suitable for all scenarios, including children and infants. The platform should be unobtrusive and minimize a potential influence on the experiment itself by minimizing the extent that individuals are actively aware of an ongoing evaluation.

1.3 Contributions

This thesis presents a platform for human interaction tracking. The hardware and ranging scheme is based on previous work by Kempke et al. [23]. This thesis improves and extends their concepts in the following points:

1. We generalize the binary ranging model of tag and anchor and decouple logical functionality from the physical device. Furthermore, we demonstrate that this distinction opens up infrastructure-free communication by eliminating the assumption of permanently available nodes and show how this can be leveraged during scheduling to improve energy efficiency.

2. We introduce a heterogeneous radio architecture and leverage the usage of Bluetooth Low Energy (BLE) as a complementary wake-up radio to reduce idle listening costs by more than 95%. We show how this secondary radio can be employed to automatically merge clusters and how efficient leader re-election can be implemented using a supplementary leader election algorithm on the primary radio. Additionally, we use BLE as a user interface to configure and control devices throughout the deployment.

3. We add flexible design parameters to the state-of-the-art ultra-wideband ranging protocol and leverage exclusive broadcasts to achieve linear message complexity. In combination with other optimizations, this boosts lifetime by up to 77%, improves scalability to reliably support large networks, and optimizes the dynamic behaviour to reduce join latency.

4. We extend the system with permanent storage capability and the availability of global timestamps to enable offline analysis and a contextualization of the measurements.

5. We show that the concept of employing modules as building blocks is still valid for hardware design, but no longer requires physical modules, and leverage state-of-the art design capabilities to permit more agile solutions.

6. To evaluate our concepts, we implement a system which provides a mobile, accurate, reliable, and agile platform for ranging measurements. Testing both under controlled settings and in real-world scenarios underlines the importance of diversity ranging by reducing the median ranging error up to 86% and demonstrates the automatic extraction of interaction statistics.

Both our hardware and software applications are publicly available in an open-source repository [24]. TotTernary is actively undergoing preliminary testing as the primary tool for social interaction studies regarding the interaction of babies and caretakers [25] and is planned to be deployed in dozens of U.S. households.

1.4 Outline

The rest of this document is structured as follows: Chapter 2 *Background & Related Work* offers an overview of the available technologies and introduces previous work which influenced the current platform. In Chapter 3 *System Design*, we present the system design of our platform and detail the considerations behind our hardware and protocol decisions. Chapter 4 *Implementation* sketches the final product and gives an insight into its technical aspects. Chapter 5 *Evaluation* analyses the performance of the system under various scenarios and demonstrates how such a platform might be employed. In Chapter 6 *Summary & Conclusions*, we present our findings and identify possible extensions of the concept.

Chapter 2

Background & Related Work

As localization systems and wireless sensor networks have been intensively studied in recent years, this chapter first gives an introduction into the various technologies employed to measure physical distances and motivates our selection of time-of-flight ranging using ultra-wideband (UWB). Thereafter, we summarize the historical evolution and draw an overview of the state-of-the art in human interaction research and MAC protocols for wireless sensor networks. The last section delivers a more detailed insight into selected previous contributions which strongly influenced this thesis by demonstrating that range-based interaction tracking can give valuable insights and identifying that diversity ranging improves both accuracy and stability significantly for UWB.

2.1 Physical distance estimation

Most modern ranging solutions rely on the propagation of waves through space to deduce a distance estimation between devices. Most frequently, purely RF-based systems are used due to their ability for non-line-of-sight communication and high-frequency data transmissions. Some designs also leverage ultrasonic [5, 26, 27] and infrared [11] signals to gather information such as direct face-to-face interactions. We discuss the different schemes and underlying physical-layer technology and identify how our platform complements the existing design space.

2.1.1 Infrastructure-based solutions

An omnipresent infrastructure-based system which is used in thousands of applications is the Global Navigation Satellite System (GNSS) consisting of GPS and similar platforms. The Global Positioning System (GPS) relies on 31 satellites orbiting the globe to calculate a global position based on time difference of

arrival. While it can provide valuable contextual information for outdoor sensor data, GPS is strongly limited in urban areas due to insufficient signal strength, especially for indoor localization [28]. Furthermore, its energy footprint is orders of magnitude higher compared to solutions which integrate local infrastructure [29]. Attempts to provide equivalent indoor accuracy [13, 27] are still restricted by fixed infrastructure costs and limited coverage.

Early wireless sensor solutions for studying contact networks often relied on infrastructure-based solutions [11, 26]. This allowed a simplification of the mobile hardware and resulted in less expensive and more energy efficient solutions. However, due to their large deployment overhead, restricted mobility and high fixed costs, research started focusing on system designs which did not rely on stationary hardware and therefore enabled more scalable and mobile deployments [17, 30]. Previous research suggests that infrastructure-based deployments were short and heavily restricted in space to single locations, whereas infrastructure-free deployments achieved large scales with a life span of weeks to months [5].

2.1.2 Proximity detection

The detection of other devices within a given, fixed vicinity accounts for the majority of previous interaction research. This results in purely binary measurements of the co-location of two devices within a given radius (usually between 2 to 5 meters [5]) and therefore only delivers an abstracted view of the environment. Most Bluetooth-based systems rely on simple detection which can be implemented with chips already available in many commercial products such as smartphones [16] and smartwatches [6]. Similarly, many wearable, infrastructure-free sensor systems only estimate proximity instead of actual interaction distance [15, 17, 30]. While this greatly simplifies measurements, the coarse-grained measurements only deliver rough insights into social interactions [10, 19]. Studies have found that a more detailed distinction between proximity zones could offer new insights for various scenarios such as disease spreading, racial attitudes, stress, and verbal communication [5].

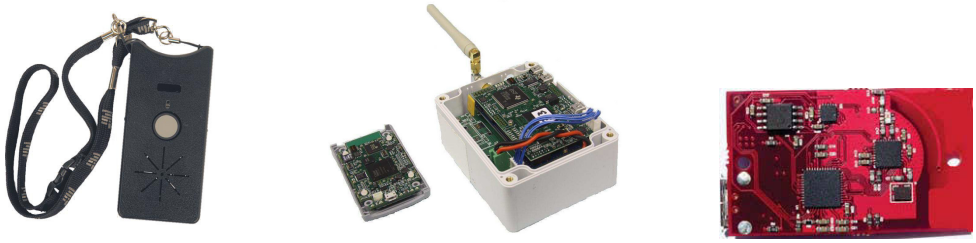

Figure 2.1: Examples of deployed tracking systems demonstrate various technologies: Sociometric badge [11, 31], WASP [13] and WREN [17] (left to right).

2.1.3 Signal strength measurements

To improve from all-or-none proximity sensing to gradual range estimations, systems started to exploit the received signal strength indicator (RSSI) values, which vary according to the path loss observed during transmission [30]. Based on the attenuation of the signal from sender to receiver, one can attempt to derive the physical distance between devices. However, in addition to distance, path loss is affected by strong multipath in indoor environments, the influence of antenna directionality, and the frequent occurrence of non-line-of-sight conditions. These effects reduce the usefulness of such estimations for scientific studies, as the measurements fluctuate widely over time in dynamic environments and are often unreliable and inaccurate [14]. Even in stable environments, such systems are only capable of 2 meters of ranging accuracy [5] and hence do not satisfy the desire for sub-meter accuracy requested by social researchers [20, 21].

2.1.4 Time of Flight measurements

To obtain a higher degree of accuracy and reliability, methods to estimate the distance based on the duration of the signal propagation instead of the attenuation became increasingly popular for research applications with the introduction of radios with highly accurate clocks and the ability to provide time stamps in the order of nanoseconds.

Time Difference of Arrival

Schemes based on the time difference of arrival (TDoA) focus on the timing between multiple received packets to determine the distance to one or more entities. Such systems offer an improvement in accuracy of more than an order of magnitude compared to RSSI-based ones and are often more temporally stable [14]. This comes at a trade-off of increased protocol complexity and a requirement for more specialized hardware.

TDoA can be achieved by exploiting the difference in propagation velocity of ultrasonic and electromagnetic waves [5]. The sender broadcasts two signals simultaneously on both channels and measures the delay of the sound arriving at the receiver compared to the RF packet which is travelling at the speed of light. As this method does not rely on synchronization with other nodes and can exclusively use broadcast messages, such networks are much more scalable and require less control overhead compared to other TDoA systems. However, the dependence on the successful reception of ultrasonic waves hampers deployments where packets cannot be detected in cases where line-of-sight is obstructed, such as when devices are in pockets, and can be influenced by other environmental

factors such as wind and noise. It is worth noting that in some fields such as disease spreading, the directionality of sound [5] or infrared [12] is explicitly leveraged to exclusively capture face-to-face interactions with direct line-of-sight.

It is possible to utilize pure RF transmissions for TDoA. This method only requires a single radio per device and therefore reduces complexity and hardware costs. By eliminating the dependence on the reception of ultrasonic signals, pure RF transmissions are also capable of transmitting at 360 degrees and despite non-line-of-sight (NLOS). For this, researchers exploit the reception of multiple, globally scheduled beacons at known times to calculate their position relative to the base stations [13]. This enables indoor localization accuracy within half a meter, suitable for most tracking and interaction applications. However, this gain in precision can usually only be achieved with infrastructure-based synchronization and requires the precise knowledge of base station positions.
In contrast to ultrasonic waves, differences of electromagnetic wave packets require precise measurements of the time of flight in the order of nanoseconds for meter accuracy. Therefore, it is crucial to correctly detect the exact point in time when a signal has been received. Additionally, these schemes require extremely tight time synchronization between base stations at the order of one nanosecond and can most often only be solved in the form of wired connections between base stations [13]. Therefore, they once again require fixed infrastructure and are hence not suitable for large-scale studies in shifting environments.

Time of Arrival

Instead of using the difference in reception of multiple packets, systems using *time of arrival* (ToA) use the absolute *time of flight* (ToF) to calculate the distance. We can distinguish two fundamentally different cases:

- **One-way ToF:** The simplest way to achieve time-of-flight measurements is by sending a single packet containing the timestamp at which it was sent. The receiver can then compare the transmit and receive timestamps and calculate the time the packet propagated to estimate the difference between sender and receiver. However, this requires tight time synchronization which is difficult to achieve for wearable, mobile systems and is therefore not directly applicable without further processing.
- **Two-way ToF:** To circumvent the synchronization requirement, many systems extend the scheme by sending an additional reply packet back. This allows the initial transmitter to calculate the distance based on the duration between sending a poll request and receiving a response. This round-trip time (RTT) must be corrected for the delay incurred by the device responding either by having a highly reliable response delay or by including the actual delay in the response packet.

Systems such as WASP [13] use coordinated pairwise two-way time-of-flight measurements. Requiring only a pair of packets per range, such localization systems can achieve highly reliable measurements in outdoor environments with an accuracy within single decimeters. Unfortunately, multipath, an effect prevalent in many indoor settings [23], often impedes the reception of a clean pulse, as multiple pulses overlap and the actual time of arrival is hard to estimate. There are methods to improve the result by filtering the raw channel impulse response (CIR), but those require extensive computational power and are difficult without detailed knowledge of the environment [13].

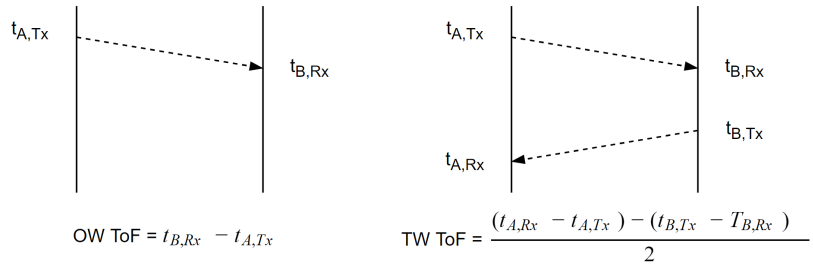

Figure 2.2: One-way ranging (left) requires tight time synchronization between devices, whereas two-way ranging (right) relies solely on local subtractions but needs two packet transmissions.

Instead of post-processing noisy signals on a resource-constrained device in unknown and time-varying settings, it is preferable to limit the overlap of multiple reflections from the beginning. Widening the signal in frequency domain corresponds to a reduction of the pulse length in time domain. Ultra-wideband (UWB) leverages this fact using its large bandwidth of more than 500 MHz to achieve exceptional timing resolution with very sharp pulses in the order of nanoseconds [32]. This largely eliminates multipath effects as reflections do not overlap and ensures a clean pulse detection [33]. Combined with its resilience to narrowband interference [32] and superior penetration of objects compared to traditional RF signals [34], UWB is a well-suited technology for reliable distance estimation in variable environments.
Unfortunately, the widespread deployment of UWB transceivers has long been hindered by high energy consumption in the order of multiple watts [13] and was restricted to niche applications due to its high system cost [35]. The approval of unlicensed usage in the 3.1 to 10.6 GHz band by the FCC [36] and the introduction of the 802.15.4a (UWB) standard [37] however increased the interest and research in this domain and recently resulted in the first viable solutions for embedded systems. The DecaWave DW1000 [38] used in this thesis is capable of timestamps with a resolution of 15 picoseconds, corresponding to a ranging accuracy of 4.5 millimeters. With a price point of less than $10 per chip and a maximal power draw of 560 milliwatts, it is suitable for large-scale, long-term deployments and can be integrated into classical circuit designs to create an robust and accurate wearable platform for interaction tracking.

2.2 Related work

Human interaction has a long history of wearable-assisted research. Personal electronic equipment is omnipresent nowadays and offers rich connection interfaces and multiple sensors. Both smartphones [16] and smartwatches [6] have been used to determine whether individuals are near each other. While such systems can provide additional contextual information and have been employed in small-scale trials for up to 15 months [16], large-scale deployments are limited due to the difficulty of supporting a broad spectrum of heterogeneous devices and the fixed range of Bluetooth [10].
To facilitate wide-spread and accurate studies, researchers started building dedicated wearable social sensing platforms with specialized hardware. The Sociometric badge [11, 12] enabled studies with a total of one thousand participants but required extensive infrastructure for distance estimation. Other systems such as OpenBeacon [15], WREN [17], and UCMote [2] only detect the binary presence of a device within a 5 m radius to establish contact networks inside a cohort and thus only allow limited insights into social interactions [10]. Infrastructure-based systems such as WASP [13] and GuoGuo [14] can achieve decimeter accuracy and high time-resolution on the order of up to 25 Hz, but are impractical due to extensive setup costs and a highly restricted environments. As the first infrastructure-free system with fine-grained distance measurement capabilities, Opo [5] leveraged a combination of ultrasound and RF channels to measure face-to-face interactions inside a cohort with centimeter accuracy to study the spread of diseases. However, its use case was limited due to its restricted range and the strong directionality of the ultrasonic waves.
Accurate indoor localization using UWB has been of great interest for many years [33] and demonstrates great potential due to its high range and accurate distance measurements despite multipath. To compensate for attenuation due to antenna polarization mismatches and local destructive interference, PolyPoint [35] introduced frequency and antenna diversity. Its successor SurePoint [23] supports multiple simultaneous nodes, adds robustness, and reduces the form factor to $3\,\mathrm{cm}^3$, but was still restricted by a requirement for infrastructure, fixed settings and a significant power draw of 280 mW.

To maximize the benefits of a system, suitable hardware must be paired with a fitting protocol to exploits its strengths. **Wireless sensor networks** have been extensively studied; Langendoen et al. [39] give a good overview over MAC protocols and their design considerations, whereas Huang et al. [40] include more modern advances in the field. In general, research is divided into asynchronous protocols, which aim to efficiently communicate despite different active/sleep cycles, and synchronous protocols, which rely on scheduled transmissions and focus on latency and throughput improvements.
B-MAC [41] first demonstrated a robust way of asynchronous packet distribu-

tion by alerting the channel using preambles. Successors such as SpeckMAC [42] tried to reduce long listening operations after a detected preamble using continuous preamble sampling, in which they include information about the actual sent operation. To get away from constant channel occupation, WiseMAC [43] introduced schedule learning to only initiate transmission shortly before nodes start to listen but required infrastructure nodes for bootstrapping. As an attempt to avoid long preamble transmission and reduce the discrepancy in energy costs between sender and receivers, protocols such as A-MAC [44] initiated transmissions on the receiver side by signalling their ability to receive using a probe. This does however lead to frequent collision at the receiver and requires long listening periods for transmitters.
To reduce idle listening and sending times, protocols such as S-MAC [45] use explicit time synchronization to align active periods for neighbouring nodes by forming clusters. The known communication time makes such protocols excel for periodic traffic when transmissions can be well predicted. While S-MAC [45] used explicit Request-to-Send (RTS) and Clear-to-Send (CTS) frames to efficiently contend for the channel, its adaptability to run-time changes was restricted due to the fixed period. T-MAC [46] introduced flexibility for the length of the active period, but suffers from overhearing and the overhead of explicit channel reservation.
As transmissions are often periodic, the application of TDMA and global schedules eliminates the contention overhead and improves reliability, but can suffer from pre-determined and fixed allocations. Z-MAC [47] additionally incorporates Carrier Sense Multiple Access (CDMA) and uses slot stealing to increase channel efficiency under low contention. TRAMA [48] adds a contention access period where nodes can try to claim slots for the following scheduled period to increase the flexibility of the schedule. LWB [49] reduces the overhead of such scheduling by exploiting the regularity of purely periodic traffic and only using a brief contention slot.

While an efficient networking protocol is essential, it is nevertheless wasteful to run a radio when no neighbouring devices are around. Wake-up radios such as the scheme used in Opo [5] theoretically permit large power savings by only turning on the primary radio when actually required. However, such systems mandate additional circuitry and often suffer from frequent erroneous activation. Early **neighbour discovery protocols** such as the birthday protocols [50] offered good average case performance but suffered from nondeterministic maximal discovery delays. Disco [51] used the Chinese Remainder Theorem to tightly upperbound the worst-case discovery. With the introduction of the Nihao family [52] and the "talk more and listen less" principle, protocols started to increasingly leverage cheap, short transmissions compared to extended idle listening. BLEnd [53] builds on this concept and extends it for the pervasive case of Bluetooth Low Energy (BLE) to provide a practical and predictable discovery mechanism for wearables.

2.3 Primary research influences

2.3.1 Opo

Opo [5] is the first wearable that achieves high-precision distance measurements without relying any supporting infrastructure. It captures ranges up to 3 m with a 5% average error as well as a high temporal fidelity of 2 seconds.

Figure 2.3: Opo only weights 11.4 g with a size of 14 cm^2 (left) and achieves highly accurate ranging results over a distance of up to 3 m (right) [5].

Opo makes use of a combination of ultrasonic and RF radios, where ultrasound is used both for ranging and as a wakeup source. Thanks to an efficient wakeup circuit with a low power draw of 19 µA, the system avoids a more expensive neighbour discovery protocol.
Opo shows that the additional spatial and temporal granularity compared to proximity-based systems offers more detailed insights into collaboration intensity, group dynamics, and behaviour variation throughout conversations. It also demonstrates the value of infrastructure-free deployments and investigates prolonged interactions outside of controllable environments such as while going out or travelling.

This thesis distinguishes itself in three crucial points from Opo. Due to Opo's focus on face-to-face interaction using ultrasonic pulses, the detection angle is by design strongly restricted to an angle of 60°, whereas we can achieve full coverage in all directions even in non-line-of-sight conditions and hence provide more constant measurements. Secondly, the propagation of ultrasonic waves is influenced by external factors such as wind and noise and can therefore only be reliably used for up to 3 m, whereas TotTernary achieves ranges of over 60 meters and allows room-level observations. Lastly, our platform provides multiple opportunities for extensions and direct interactions such as real-time tracking over BLE, extensive logging over external SD cards with a storage capacity which is three orders of magnitude larger and an accelerometer which allows us to enrich the ranging estimates with mobility data.

2.3.2 SurePoint

The introduction of a legal framework [36], the release of the official IEEE 802.15.4a standard [37], and the availability of commercial UWB transceiver [38] have enabled high fidelity distance measurements on embedded devices. However, those measurements often suffer from significant sporadic outliers due to signal attenuation [35] and strong accuracy variations in real-world applications depending on the environment [33].
SurePoint [23] is a UWB localization platform that leverages antenna and frequency diversity to achieve an order of magnitude improvement in median ranging accuracy. A novel, efficient ranging protocol limits the overhead of the additional messages used to diversify ranging and enables update frequencies up to 12 Hz. To ensure exclusive channel access despite multiple devices, the system makes use of LWB [49] to guarantee reliable and energy-efficient schedule distribution and tight time synchronization.

Diversity exploitation

While UWB offers a tighter distribution of ranges and hence a better precision than narrowband ToF estimates, the accuracy of the results can still be off as the arrival time is incorrectly detected. Because of signal attenuation due to destructive multipath interference and antenna nulls, range estimates often contain a positive error bias as detected pulses appear delayed or possess a shallow pulse slope. This results in an artificially prolonged time-of-flight estimate and a corresponding overestimation of the true distance. As seen in Figure 2.4, the distributions of range estimate errors observed from both a 802.15.4a channel model by Molisch et al. [54] as well as from real-world experiments reflect this fact and demonstrate that simple averaging of single channels results in flawed estimates. To reliably establish the true range, it is advantageous to collect multiple diverse range estimates. Based on empirical and simulated data, SurePoint found that the 12th percentile yields the best range estimate in real-world conditions.

Traditional radio systems only make use of a single antenna for transmitting and receiving packets. SurePoint employs three antennas with an offset of 120° to diversify its effective antenna polarization and circumvent constant misalignments. By using 3 transmit antennas, 3 frequency channels as well as 3 different antennas at the receiver, SurePoint collects $3^3 = 27$ separate samples of the physical UWB channel. This diversity results in a 54% reduction of the median ranging error and decreases the 99th percentile error to within 53 cm [23].

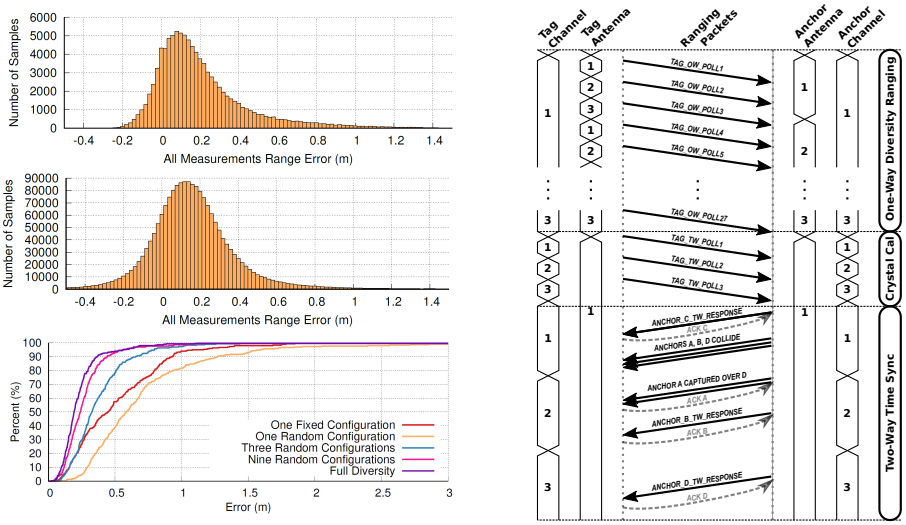

Figure 2.4: Simulations [54] (top left) as well as empirical measurements (middle left) suggest that the 12th percentile minimizes the estimation error. Diversity has a large influence on the overall accuracy (bottom left) and can reduce the 90th percentile error by over 60%. To this end, the ranging protocol uses 30 poll messages from the tag which are answered by the anchors (right) [23].

Network coordination

To incorporate 27 individual rangings within a network of unsynchronized devices, Kempke et al. developed an efficient ranging protocol which utilizes a combination of one-way and two-way rangings. A single two-way ranging is sufficient to solve the problem of unsynchronized clocks. Using one-way rangings for the remaining measurements reduces the number of packets by half and cuts down on air time and energy costs. To accomplish this, each round of ranging is split up into three different phases as seen in Figure 2.4:

1. **One-way diversity ranging:** By cycling through all of the 27 configurations and sending a single packet on each one of them, the tag can estimate the time of flight on multiple frequency bands and antenna pairings.

2. **Clock drift estimation:** To compensate for the varying clock rate on different devices, the tag repeats the transmission of the first 3 packets. This allows the tag to estimate the clock drifts between the tag and individual anchors by comparing the difference in timestamps on both devices and correct all other timestamps accordingly.

3. **Two-way time sync:** The last returning packet from the anchors allows the computation of a two-way ToF range estimation only based on the difference of timestamps on each device. This enables distance estimations despite unsynchronized clocks and can be leveraged to successfully compute the estimates for each of the 27 one-way rangings in hindsight.

The ranging protocol can then be used as a primitive inside a larger round structure. Inspired by LWB [49], a master first distributes a schedule in which it details the assigned slots for the rest of the round. With a Glossy flood [55], the schedule is then reliably distributed throughout the network. Each tag can use the incorporated contention slot to request an opportunity to perform ranging itself and will receive one of the twelve available indexes.
The schedule is also used for network-wide clock synchronization and to calibrate individual clock speeds to minimize clock drift. This is both essential for the successful constructive interference used by LWB and ensures that the entire network starts individual ranging slots simultaneously.

Hardware: Module and carrier board

To facilitate the use of their system, the authors built a hardware module which incorporates the crucial part: the radio and delicate RF circuitry such as the RF switch as well as the dedicated controller. The module abstracts the complexity of the system by having a proven hardware and software solution which is easily accessible over an I^2C bus. This abstraction layer ensures that time sensitive radio operations are not influenced by application-level decisions and simplify software design and debugging. A separate interrupt line allows for energy-efficient communication to the carrier board and integrates easily into common embedded software design principles.

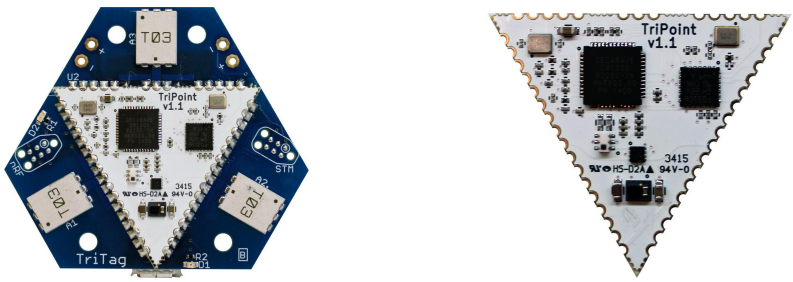

Figure 2.5: The carrier (left) provides a user interface via BLE, power supply circuitry as well as the 3 UWB antennas, while the module (right) contains the radio including its controller as well as the RF switch [23].

A carrier board integrates the hardware module and invites for a flexible design according to the specific application requirements. It provides the three UWB antennas at appropriate angles to leverage antenna diversity and makes them accessible to the module. A communication chip (integrated radio and MCU) allows direct interaction with other devices such as smartphones and gateways over BLE. Furthermore, the carrier includes the necessary power supply circuitry and battery connectors as well as connectors for programming both the communication MCU and the controller on the module.

SurePoint demonstrated that wearable UWB system is feasible; with its 4.5 g and 14 cm^3, it demonstrates that accurate and reliable ToF ranging is possible despite a restricted form factor.

Going further

While SurePoint achieves many requirements for a tracking system in terms of reliabiliy and form factor, it still lacks key features which prevent its direct application to our use case:

- **Infrastructure-free deployment:** The SurePoint architecture assumes the two distinct, fixed device classes tag and anchor. We separate the logical function from the physical device and add the ability to switch roles flexibly and assume them simultaneously during run-time.
- **Neighbour discovery:** Leader election is static and occurs during compile-time. During operation, a node assumes a constant connection to the network and does not conserve energy by turning off its radio when there are no other devices within range. This thesis adds energy-efficient discovery and dynamic election by using its secondary BLE radio.
- **Substantiated flexibility:** SurePoint does not incorporate any way for system designers to estimate the influence of protocol adaptations in advance when adapting the system to their desired use case. We propose various ways to boost the system lifetime by adjusting parameters and evaluate how such changes affect the overall performance using a detailed energy model based on extensive hardware measurements.
- **Data aggregation:** While real-time data could be accessed over BLE, the TriTag hardware does not allow local storage without additional gateways or similar external infrastructure. We directly incorporate local permanent storage capacity which allows the system to operate independently and enables detailed offline analysis in combination with global timestamps.

In addition, flexibility is further hampered by the limited Flash and RAM available, which is at its absolute limit in terms of application size. Furthermore, the implementation overclocks its controller to achieve a sufficiently high switching speed of less than 1 ms for the radio communication, which can lead to severe and unpredictable power issues and potentially results in failures during long-term deployments.

2.3.3 Low-Powered Wireless Bus (LWB)

Most traditional networking protocols try to establish an accurate representation of their neighbours or even the global network to optimize routing over links chosen depending on different link metrics. This can incur high protocol overhead and results in unstable connections for highly mobile networks where neighbourhoods constantly change. The Low-Power Wireless Bus (LWB) [49] leverages synchronized flooding to achieve highly reliable, energy-efficient multi-hop communication despite being completely agnostic to the physical network structure. It achieves this by globally scheduling slots in which designated nodes are allowed to initiate a Glossy flood [55], which then propagates throughout the network as shown in Figure 2.6. As a result of precise time synchronization due to the time-triggered transmissions, the simultaneous broadcasts of an identical packet constructively interfere at the receiver and result in an improved SNR, increasing the likelihood of successful reception up to 99.99%.

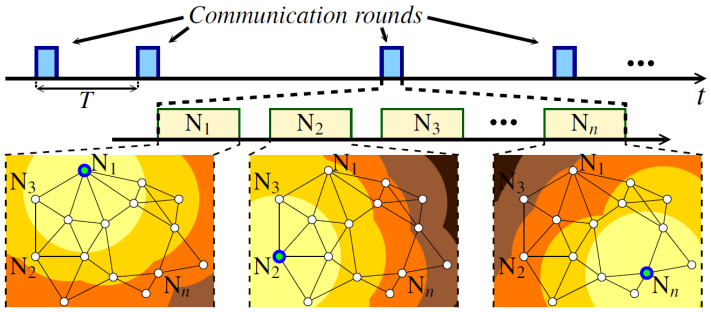

Figure 2.6: LWB uses Glossy [55] to distribute packets over multiple hops while being oblivious to the actual physical network structure [49]. We leverage Glossy in our protocol for topology-agnostic and reliable schedule distribution as well as its tight time synchronization.

At the beginning of each round during the active phase, illustrated in Figure 2.7, the logically centralized host node distributes a schedule detailing the length of the round as well as a precise allocation of slots to participating nodes. Similar to a time-division multiple access (TDMA) scheme, the assigned nodes can then use their slots exclusively. During the subsequent contention slot, interested nodes can attempt to access the channel and send a signalling packet to the host. Due to the capture effect [56], one of the packets will likely be recovered by the host and can affect the schedule for the next round.

LWB is an excellent choice for a highly reliable, energy-efficient distribution of periodic schedules. In addition, it inherently comes along with tight global time synchronization and is not affected by high node mobility due to its agnostic view on networking state.

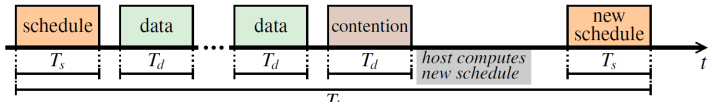

Figure 2.7: The host node determines the schedule and distributes it throughout the network to uniquely assign slots to other nodes. At the end of the round, contention allows nodes to request additional slots [49].

2.3.4 BLEnd

BLEnd [53] is a neighbour discovery protocol which is specifically tailored towards the constraints of Bluetooth Low Energy (BLE). It takes protocol peculiarities such as minimal advertisement intervals and random additive slack into account and optimizes protocol parameters towards minimal energy consumption while still respecting given design requirements. This allows application designers to define service level agreements, flexibly adjust the protocol according to their needs, and get an a priori, accurate estimate of the expected performance.

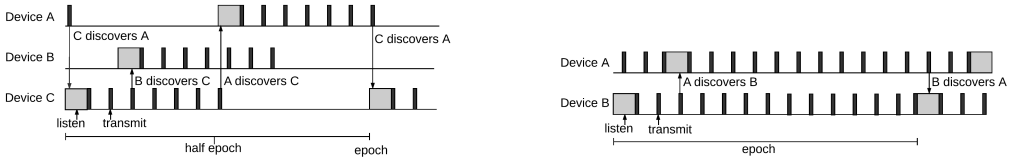

Figure 2.8: While U-BLEnd (left) only works for uni-directional discovery, F-BLEnd (right) achieves bi-directional discovery within a single epoch [53]. TotTernary combines U-BLEnd with connectable advertisements to still achieve bi-directional discovery at a lower cost.

To achieve this, BLEnd utilizes non-connectable advertisements and follows the "talk more and listen less" approach of the Nihao family [52]. For bi-directional discovery (F-BLEnd), nodes listen for a single scan interval and then periodically advertise for the rest of the epoch (see Figure 2.8). To reduce energy costs, uni-directional discovery (U-BLEnd) restricts advertisements to slightly more than half the epoch and sleeps for the remainder. An optimizer is employed at compile time to choose the epoch duration as well as the advertisement and scan intervals which minimize energy consumption.

As BLE is a low-power technology which permits connections to a majority of mobile devices, we leverage it in combination with BLEnd for neighbour discovery and user interaction. To integrate it into our energy model, we build our own version of the optimizer, as the authors are still working on making the original open-source. We improve upon their design by exploiting connectable advertisements to use the more efficient U-BLEnd for bi-directional discovery, combining both the energy savings and discovery guarantees of the proposed variants.

Chapter 3

System Design

The preceding chapters showed that while previous research has gone a large way towards realising a reliable human interaction tracking platform, it still failed to address key requirements that would both lead to an applicable system and open up the available technologies to a wider array of applications. In the following sections, we first analyse the problem space and detail both requirements and current shortcomings. We then present the concepts behind TotTernary, a system which realizes those goals and can be considered as a first step towards a real-world sensor platform which enables a wide range of human interaction studies.

3.1 Problem setting

Researchers studying social interactions lack a reliable, scalable, and mobile platform for measuring distances and indoor positions of a cohort. An effective system needs to be deployable in various scenarios such as educational research, sport team interaction analysis, elderly care, relationship dynamics, and workplace settings. For such systems, it is of special importance that individuals can be observed in their natural environment, that is during their day-to-day working schedules as well as on weekends when they break with their routines and conduct activities at various locations. This requires infrastructure-free deployments with potentially long lifetimes, as people are mobile and devices might not be charged regularly.
Furthermore, subjects should be oblivious to the recordings, as this can strongly influence their behaviour [8] and consequentially cause scientists to miss cases such as neglect of a child by their caretakers [7]. Interaction studies are currently mostly conducted in lab settings where people leave their normal environment and consciously switch to another location and possible mental state, a phenomenon known as the Hawthorne effect [7]. To avoid this and obtain unbiased data, we need to part with the lab setting and develop a system that is capable

of flexible, autonomous data collection for various settings in a small form factor. These considerations restrict potential form factors and guide the system design.

3.1.1 Application space

As this thesis is targeted specifically towards human interaction studies, we must consider different use cases and aggregate both their commonalities as well as their different peculiarities. The final protocol should support a union of all requirements specified below:

- **Team dynamics:** Sport teams can benefit from detailed analysis of the players behaviour. The availability of automated, player-specific data helps both each individual as well as the trainer and allows for quantifiable comparisons of game dynamics and performance developments. Such environments require high time resolution and precise localization to compare player statistics, whereas energy efficiency is secondary.
- **Elderly care:** The ability to unobtrusively observe people in their own, private environments opens up possibilities for automated behaviour analysis and alerts in case of abnormalities such as falls. Furthermore, it enables research into the development of mental illnesses such as Alzheimer's and their affect on people's mobility and daily life patterns. In such scenarios, simple deployability including infrastructure as well as maintainability by laymen without specialized equipment are key for long-term deployments.
- **Workplace evaluation:** Social interactions in companies (also called organizational behaviours) have always been one of the focus points of behavioural scientists [11, 12]. Automated data collection enables deployments on large scales and offers researchers a wealth of opportunities to study co-worker interaction and movement dynamics. Short network join latencies when clusters meet and separate again as well as diverse settings make such a deployment challenging.
- **Private homes:** Family dynamics can currently only be studied in limited lab settings. Having the opportunity to follow people in their daily lives and see how they interact throughout the entire day in their normal environment provides previously unobtainable data on private interactions. Often locations change and the requirement for mobile devices as well as the absence of trained personnel provide a challenge in such a scenario.

Based on these use cases, we distinguish the key requirements for an effective interaction tracking system as follows:

- *Distance estimation with decimeter accuracy* in a range of up to 30 meters despite non-line-of-sight conditions and multipath
- A *flexible, energy-efficient protocol* permitting hours to days of continued sensing, sporting low latency and update frequencies in the order of Hz
- *Infrastructure-based and -free deployment modes*, allowing for a wide range of distance measurements to positioning scenarios
- *Dynamic neighbour discovery* and leader election which ensure robustness to nomadic nodes and allow for quick reactions to environmental changes
- Ability to *log all data permanently* in local storage and off-load measurements using wired and wireless connections
- *Small physical size* and limited weight to allow for unobtrusive, long-term measurements

3.1.2 Ranging

In SurePoint [23], nodes behave either as tags, trying to initiate rangings, or as anchors which are stationary and respond to ranging requests. As the paper targets localization exclusively, which implies fixed nodes at known positions, separating those two functions physically does not restrict the operational space. However, more generally one can logically define the functions as **initiator**, which starts communication, and **responder**, which replies. While a tag only functions as an initiator and an anchor is exclusively a responder, many real-world scenarios may require mobile nodes to act as both initiator and responder in parallel, a configuration we will call **hybrid** henceforth. Furthermore, one can conceptually imagine a device which exhibits neither behaviour and therefore does not take part in ranging activities; nevertheless, it can still play a **support** role in the infrastructure. Therefore, we can distinguish four different classes of nodes and will use this distinction in the protocol we propose in this thesis.

3.1.3 Energy consumption

The power budget is one of the defining parameters for embedded systems and a major constraint for deployment scenarios due to its far-reaching influence on weight, size, and lifetime of the system. Hence, a detailed analysis of the energy usage in an application scenario is crucial. For this reason, we start with a detailed power analysis of a state-of-the-art localization system and draw conclusions on how a protocol can put its resources to optimal use.

3.2 Characterizing the state-of-the-art

From a comparison of both state-of-the-art human interaction systems [5] and localization systems [23], we conclude that SurePoint [23] offers both excellent SWaP (Size, Weight, and Power) and ranging performance compared to competing systems. In addition, it uses UWB ToF ranging which we showed in Section 2.1 to be best suited for our use case compared to other technologies and can easily be coherently extended to infrastructure-free operation. Therefore, we use SurePoint as a starting point, analyse its performance in terms of localization accuracy and energy consumption, and then adapt and extend its design principles to our use case.

3.2.1 Localization accuracy

SurePoint is a highly accurate system with a median error of only 29 cm in stationary environments and 19 cm for moving devices. While median accuracy is an important metric for describing measurement quality, this single value can be misleading, especially for techniques using RF transmissions which can display unpredictable patterns with high spatial and temporal variability. An analysis of the precision is hence crucial to estimate the reliability of the system.

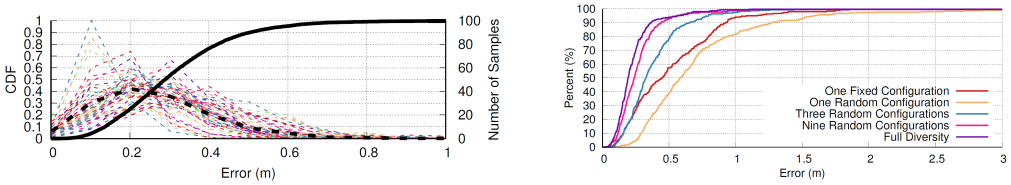

Figure 3.1: The results from the original SurePoint experiments show that it achieves an accuracy of 0.50 m for the 90th percentile as well as a precision of 0.27 m for stationary objects (left). It leverages antenna and frequency diversity to reduce its 99th percentile error from 2.86 m with a single configuration to 0.86 m using full diversity (right) [23].

We can see in Figure 3.1 that SurePoint achieves a high degree of precision and only deviates by 27 cm from its median value in 90% of the cases. This is largely a direct benefit of the extensive diversity afforded by the three antennas and three frequency bands. Leveraging full diversity reduces the 99th percentile error by 70%, while limited diversity with three samples still achieves a 59% reduction. This adjustable amount of diversity will offers us a promising trade-off between accuracy and energy consumption as we will demonstrate in Section 5.2.

3.2.2 Energy consumption

To create a model for the energy consumption of SurePoint, we must isolate the different hardware components and analyse them separately. In the current *TriTag* hardware, we can distinguish two major subsystems as shown in Figure 2.5:

- **Ranging module**: This component is located on a separate PCB and soldered on top of the carrier. It contains the DecaWave DW1000 [38] radio for UWB ranging, the STMicroelectronics STM32F031G6 MCU [57] running the controlling software as well as a circuit for switching between antennas. Except for the antennas themselves, all the hardware and software dedicated to ranging operations is contained in the module.
- **Carrier board**: The carrier board provides a sample platform which integrates the ranging module. Apart from three antennas and battery connectors, it also facilitates interaction with other devices over its BLE antenna and the corresponding Nordic Semiconductor nRF51822 radio [58] and offers multiple connectors for micro USB and debugging tools.

The three electrical components consuming the vast majority of power are the three microchips. We can see in Table 3.1 that the nRF51822 only requires a minor part of the energy budget, while the STM and the DW1000 demand an order of magnitude more power for their operations. Furthermore, one can deduce that the listening duration on the UWB channel will have the greatest impact on the lifetime of the system due to its high current draw.

Component	**Steady-state current Module** [mA]	**Steady-state current Carrier** [mA]
nRF51822	3.2	
STM32F031G6		30.4
DW1000 idle		12.9
DW1000 in Tx		34.1
DW1000 in Rx		151.8

Table 3.1: The characteristic currents of the individual components on full load show that reception on the DW1000 draws almost five times as much current as transmission and that computations on the STM are energy-expensive.

Table 3.2 demonstrates the different power levels of the system. While one can achieve a low power draw of 3 mA by turning off peripherals and putting the MCUs to sleep, keeping the two MCUs awake but idle only increases the current draw to 17 mA, compared to the 270 mA required for some equivalent commercial products [59].
The component which clearly dominates the power budget is the UWB transceiver, as turning it on nearly doubles the energy consumption to 30 mA even when simply residing in idle state. The transmission of a packet, the primary task of tags, once again temporarily doubles this figure to 54 mA. Rather surprisingly, putting the device into receiving mode results in a current draw of 171 mA, a tenfold increase over the original idle power draw and more than three times the power required for sending. Furthermore, as listening is a steady stage that a node can reside in for long periods of time (as opposed to the sporadic transmissions with short bursts in the order of tens of microseconds), continuous reception quickly drains the battery and must be minimized.

Power level	**Steady-state current** [mA]
Peripherals off	3.1
MCUs idle, DW1000 off	17.3
DW1000 idle	30.2
DW1000 in Tx	54.4
DW1000 in Rx	171.0

Table 3.2: The holistic view of the platform on different power levels of the complete system provides intuition for the energy consumption in practice and shows that, as expected, the DW1000 dominates the power budget.

Comparison between Tag and Anchor

Table 3.3 displays the current draw of both anchors and tags during the ranging protocol. While the anchor has a significantly higher idle current, it is interesting to see that at high utilization rates (all slots used), anchors overall require 15% less energy compared to tags in the current configuration. Although both consume a comparable amount during one-way diversity ranging, they differ considerably during the two following periods; whereas anchors idle in the Rx state and tags in the Tx state, the situation reverses during the two-way time synchronisation as only tags are required to receive. Due to the difference in length of the two stages, anchors overall spend 10 ms less time listening during a round and consume 17% less energy.

Stage	**Duration** [ms]	**Tag avg** [mA]	**Anchor avg** [mA]
1-Way Diversity Ranging	30	54.6	67.2
2-Way Time Sync	30	172.0	34.9
Idle	20	35.9	168.5
Long-term average		94.5	80.3
Schedule distribution	10	54.0	37.5
Contention	10	35.4	169.0

Table 3.3: During different stages of both the ranging protocol and the LWB round, the current draw varies heavily. Remaining in Rx mode for prolonged period of time while waiting for the next round or during contention significantly influences the average consumption.

The results in Table 3.3 have been gathered with a single tag running on the maximal duty cycle, i.e. one which uses every available slot in a round. In practical applications with a more lifetime-oriented update frequency, the energy requirements will decrease almost linearly with the update rate. Periodic schedule distribution is independent of the ranging events and adds a fixed overhead. While the energy consumption drops for tags with fewer ranging events, energy demand increases for anchors. As anchors remain in their default idle listening mode if no node is scheduled, their power draw massively increases. At a protocol level, this cost could be trivially evaded by enabling sleeping in unused slots. Unfortunately, the DW1000 experiences practical problems with the implementation of the sleep mode, incurring even higher current draw than during idle (39 mA in sleep vs. 36 mA in idle). Hence, the current implementation leads to the counter-intuitive situation that constant ranging outperforms occasional sleeping in absolute energy consumption and results in a more balanced depletion of energy reserves over multiple nodes.
By implementing a better-suited sleeping behaviour, the anchor's consumption ought to be reduced drastically. Nevertheless, as anchors must take part in all ranging events whereas tags are only active during their designated slot, the energy consumption of anchors will be higher overall, leading to a reduced lifetime. We address this inequality later on in Section 3.5.2.

Similar to the consumption during ranging events, tags also exhibit an increased power draw during the schedule distribution compared to some anchors. In contrast to the original Glossy paper [55], the current implementation does not include the repetition of the schedule by the Glossy host itself, but only by neighbouring nodes which first received the packet. This results in an unequal energy consumption for all tags and non-master anchors.

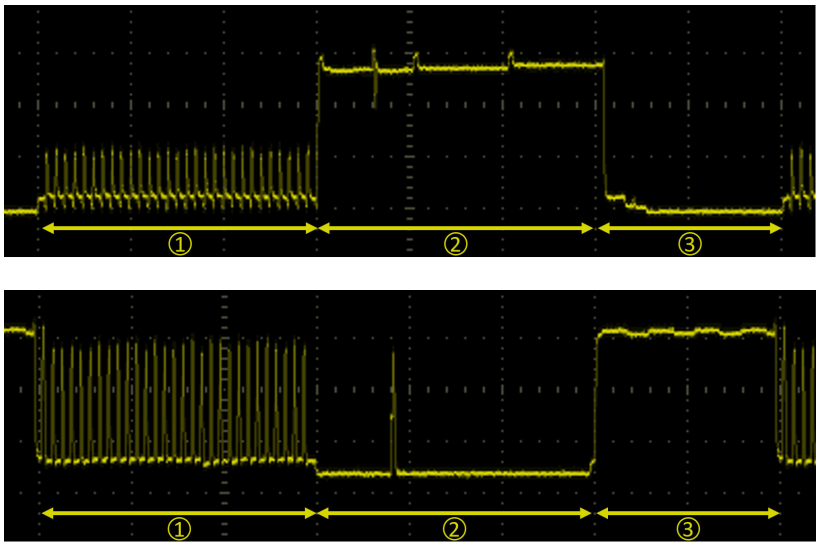

Figure 3.2: This figure illustrates the three phases of each slot for tags (top) and anchors (bottom): 1) One-Way Diversity Ranging, 2) Two-Way Time Sync and 3) Idle period. The active reception of ranging packets (bottom, Phase 1) requires significantly less power than the listening for responses (top, Phase 2) despite an equivalent duration.

Furthermore, nodes currently perform a maximum number of transmissions by repeating the packet until the end of the slot while disregarding the actual transmission number. By reducing the number of global retransmissions N to a fixed integer and trading reliability for energy efficiency, one can achieve the same performance over a given number of hops with only a fraction of the distribution costs and reduce the wasted transmissions at the border of the network.

The contention slot incurs high energy costs for all involved parties, as they must constantly listen to be prepared for flooding a schedule request throughout the network. By scheduling a contention period more infrequently and dynamically adapting to the needs of the network, the unused overhead of synchronized scheduling can be drastically reduced.

Bluetooth

A Bluetooth advertisement can be divided into a startup phase and three advertisement periods, each on another channel. While the startup takes 1.6 ms, the duration of the three periods can vary depending on whether the device receives a scan request from another device. In such a case, after 1.0 ms of sending the advertisement and listening for responses, the device answers with a scan response, adding another 0.5 ms.

Action	**Percentage**	**Avg current** [mA]	**Duration** [ms]
Peak Tx		17.0	
Peak Rx		13.5	
0 scan responses	28%	7.9	4.7
1 scan responses	52%	8.5	5.2
2 scan responses	20%	9.0	5.7

Table 3.4: In stark contrast to UWB, the current draw for BLE is almost equivalent for transmitting and receiving. Notice that while theoretically possible, we never observed three scan responses for an advertisement in practice.

The probability of receiving responses is strongly influenced by the environment. Each device which listens and actively responses to advertisements can maximally send a request on a single channel; therefore, the number of scan requests observed by the tag heavily depends on the amount of devices in its surroundings. The results in Table 3.4 represent an ordinary office environment with multiple active Bluetooth devices responding to advertisements. In busier environments, an increased volume of responses could significantly raise the energy consumption of advertisements as the additional 7.1 µC per response adds 20% to the base cost of an advertisement. The worst-case is a new request on each of the three channels by three separate devices, which would average an estimated 9.5 mA over 6.2 ms.

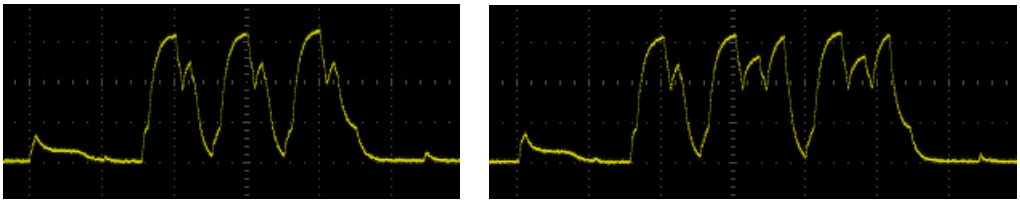

Figure 3.3: On the left, we can see the three advertisements without any scan requests and responses. The right figure shows two scan requests on separate channels and the corresponding Tx peak, prolonging the advertisement duration.

As SurePoint does not use BLE for neighbour discovery, the device is never scanning itself but only advertising. Using values provided by the datasheet [60], we can estimate a Rx current draw of 12.6 -13.4 mA. This also agrees with our measurements of the peak Rx current while listening for a scan request as shown in Table 3.4. Just as for the DW1000, Rx operations are continuous and draw power over longer periods of time; one must therefore find the optimal trade-off between “talking” and “listening” [52].

Analysis and implications

Based on the previous power analysis, we can conclude the following primary implications for the design of an improved protocol:

- As listening and receiving packets is more expensive than transmitting by a factor of 3, reducing the time where nodes are waiting for the transmission of other packets on the UWB channel is essential for substantial lifetimes. While it is well-known that idle listening periods often consume a dominant part of the power budget [41] due to their length, the strong restrictions on transmit power in UWB [36] further increase listening costs as reception is more than 4 times as expensive as transmission.
- Because of the increased energy consumption of anchors following the imbalance between sending and listening, there is a significant danger of lifetime inequality in infrastructure-free, battery-powered networks. This might lead to network failures despite the fact that the majority of nodes would still possess sufficient energy for continued operation. It is therefore essential to ensure that the energy demand is evenly distributed. Furthermore, if nodes have the possibility to reduce their consumption upon detecting a depletion of their batteries, the network can better adapt to unexpected scenarios.
- To enable energy-aware behaviour, nodes must be able to sense their current energy store. Voltage sensing at the battery can indicate an advanced level of depletion and provide valuable information for application-level interventions.
- While synchronized usage of the network leads to clear channel access routines and synchronized active radio periods, it also induces management overhead. For an energy efficient protocol, this trade-off space must be carefully evaluated to find the optimal protocol design principle for the given use case.
- Even under worst-case circumstances, the Bluetooth channel offers communication which is more energy efficient by a factor of 4 - 12 depending on the communication direction. Additionally, BLE also displays a more balanced power expenditure for sending and receiving packets with a ratio of 1.26 compared to 4.45 for UWB. Therefore, this channel should be leveraged in all cases where unidirectional communication is sufficient and does not require tight timing, especially for long-term listening such as neighbour discovery where idle listening can dominate the power budget. UWB should exclusively be used where it has a unique advantage, such as for accurate ranging or time-triggered flooding.

3.3 Hardware platform

To support the desired functionality as stated in Section 3.1.1, we required a new hardware platform which provided us with both robust range measurements as well as an energy-efficient wake-up capability. Furthermore, it must be able to log data over longer periods of time and should be capable of wireless configuration over Bluetooth connections. The addition of an accelerometer enables supplementing features such as energy savings by triggering event-based ranging (by only enabling ranging when the device is actually in motion) and interaction initiation detection (by distinguishing which person approached others, valuable information for the study of interactions with children [7]). A secondary benefit of such a sensor is the option to leverage it for sensor fusion during data processing to achieve more stable movement trajectories [61].

More precisely, we can state our hardware requirements as follows:

- **High accuracy:** We require decimeter-accurate ranging capability to distinguish movement patterns among humans, allow for precise localization and identification of standard behaviours as well as measure conversation distances and collaboration.
- **Wide-area ranging:** The platform must allow for node-to-node ranges up to 30 m to sufficiently span the median US single-family house [62].
- **Local storage:** The device must be able to provide long-term storage throughout the deployment without relying on other infrastructure. Data should be gathered without any connectivity requirements to gateways or external memory.
- **Weeklong lifetime:** To enable long-term studies and reduce the deployment overhead, the platform should be capable of constant ranging with 1 Hz for up to a week at a time.
- **Portable package:** The device should have the dimension and weight of comparable research devices such as LENA [63] (8.6 cm x 5.6 cm x 1.5 cm, 59 g), a field-tested sensor platform for infants, to remain unobtrusive and be portable by adults as well as children.

3.3.1 Component specification

Dedicated ranging and communication subsystems enable designers to match their requirements with specialized hardware and ease development, as components can be build and tested individually and do not cause unintended, complex interference. Furthermore, such a design segregates components into different

schedule, error and power domains and simplifies the reuse of existing firmware such as the controller logic for other projects. To support such modularity, the functionality of our board design is separated into two different components, the ranging module and the carrier board. This distinction follows an increasing trend in embedded systems architecture [23, 64, 65] and allows the separation of software domains in hardware.
An abstraction layer using a custom API for interaction with other subcomponents allows for an integration of the module without requiring insight knowledge of the underlying implementation and therefore simplifies the main application. Due to the timing sensitive nature of both UWB and other communication channels, a dedicated MCU for each of them greatly facilitates reliable connections as real-time scheduling in software is hard to accomplish. As we will see later on, this is especially of a concern when working with the DW1000 radio [38], as it is a first-generation architecture and suffers from silicon bugs and challenging software implementations [23].

	Ranging module	**Carrier board**
Feature set	UWB ranging over an API for both raw and pre-processed distance estimates	Communication to external systems, power supply as well as logging and debug capabilities
Hardware requirements	Interrupt interface, UWB ranging chip, controller MCU and RF switch	Antenna for wireless communication, connector for wired communication, communication MCU, long-term storage capability, battery supply
Additional features		Accelerometer, serial debugging

Table 3.5: The specification of the two subcomponents used to build a complete tracking system. API commands are controlling the interaction in-between and allow for interrupt-driven events.

Interaction with the system should be simple for the end user as well as the system operators and provide extensive debugging capabilities to extract additional information during active deployments and help in improving stability. This demands compatibility with common protocols such as Bluetooth Low Energy (BLE) or the 802.15.4 standard for wireless interfaces with other devices and universal wired connectors such as microUSB for large data offload and charging.

3.3.2 Ranging module

Previous research advocated the use of hardware as individual building blocks to enable even novice designers to compose advanced hardware systems [64]. This approach of separating functionality onto multiple boards has also been followed with SurePoint [23] and resulted in a hardware module called TriPoint [66], named after its triangular shape.

To ease integration into other designs, we leverage new capabilities of state-of-the-art PCB design software and use an EAGLE design block [67] to create a module in software instead of a separate hardware module. In comparison to the integration of the module in hardware, the software module integration does not require additional design efforts. Design blocks support simple drop-in functionality and therefore drastically reduce implementation burdens by allowing a designer to use an existing, proven layout and integrate it effortlessly into a new board. An implementation in software decreases the module's space requirements by eliminating the connectors. Furthermore, it offers more freedom for the designer to perform small adaptations on the board such as spacing out components for easier hand assembly if space constraints are less rigid or migrate the power regulator to the power supply block of the carrier board to centralize power management. In addition to this freedom in design, keeping hardware modules in stock and up-to-date is expensive and requires considerable administrative overhead. Software modules on the other hand allow just-in-time acquisition and are as easy to keep up-to-date as updating the software library.

The freedom to change layouts can however also be counter-productive and invite inexperienced hardware designers to alter and potentially impair the performance of a design. This is especially precarious with complicated RF designs whose characteristics can be highly dependent on the exact layout; hence, it is advised to clearly mark critical sections and make use of supporting tools such as keepout zones to enforce implicit requirements.

3.3.3 Carrier board

The carrier board provides two primary functions to the ranging module. In a first step, it serves as a wake-up mechanism to complement the ranging protocol with a secondary, more-energy efficient radio and detect other networks. Secondly, it can be used as an interface to enable access to data originating from the ranging module. To collect distance estimation for offline analysis, it must support a permanent storage option capable of logging all information throughout the deployment period. Furthermore, a wireless user interface must enable external connections during the deployment to permit real-time tracking as well as direct control and configuration options.

3.4 Protocol design space

To complement the hardware design, we require a protocol which allows enough flexibility to be used in various situations and meets their individual requirements while still remaining adequate for an embedded system, which is highly constraint in size and power. This section explores the design space in detail to enable well-founded decisions for a suitable protocol which fits our application space.

Our preliminary testing in Section 3.2 indicates that a solution which purely relies on a single radio cannot satisfy both lifetime and measurement quality requirements. Hence, we need both a protocol for energy-efficient neighbour discovery to bootstrap a network as well as one for high accuracy, time-efficient ranging. While network join latency is of little concern for statically deployed wireless sensor networks, the introduced delay is critical in ranging applications as passers-by need to be contacted quickly and therefore require latencies in the order of single seconds. The separation of the network discovery and the ranging protocol enables an additional degree of freedom in the total join time and therefore permits separate energy trade-offs to leverage the individual strengths of the components in the dual-radio architecture. We will later on provide means to investigate this balance quantitatively and allow applications to find pareto-optimal design points.

It is important to notice that in contrast to many existing sensor network protocols [43, 45, 46], data routing is not of primary concern as we leverage local data storage. The main focus lies in the interaction with other devices in radio range and the exchange of collision avoidance and scheduling information regarding their ranging behaviour. Furthermore, the protocol should either comply with well-known industry standards such as IEEE 802.15.4, which limits the maximal length of preambles and hence prohibits certain low-power techniques [68], or Bluetooth to make use of commercially available, compatible radios such as the Nordic Semiconductor nRF52840 [69].
While many research efforts have targeted customized systems which only interact with devices of similar capabilities and can therefore leverage highly specialized protocols, this thesis aspires to enable researchers to easily set-up and test devices in the field. Therefore, we strive for an application interface which is easily accessible over pervasive, pre-existing solutions such as the omnipresent BLE chips in smartphones and laptops. Such an interface does not only simplify deployment scenarios for system operators, but also enables real-time analysis and adaptations on-the-fly by the cohort itself; this includes synchronizing the embedded devices with up-to-date timing information or sending automated debugging information back to researchers while the study is still on-going. Furthermore, it opens the platform up to integration into other systems and enables applications such as real-time monitoring using only commercial off-the-shelf (COTS) products such as laptops as additional tools.

3.4.1 Application parameter space

This thesis aims to provide application designers with a wide variety of options to choose from and adapt the protocol to their specific needs. Therefore, we can define the following metrics which span the design space of such a scheme and incorporate both hard and soft requirements for the end product:

- **Update frequency:** variable sample rate in a range from Hz to mHz, as well as the possibility of heterogeneous settings inside a single cluster
- **Accuracy and precision:** introduction of a variable for the amount of diversity and post-processing
- **Effective range:** usage of different transmission modes which also affects energy consumption and data rate
- **Discovery & join latency:** probabilistic and deterministic discovery as well as the ability to scale in both time and numbers
- **Network size & topology:** scalability of the ranging protocol in the number of active devices, as well as the incorporation of a network structure
- **Infrastructure:** potential (partial) dependency on external systems and the degree of mobility of the devices through physical space
- **Lifetime:** offers a trade-off in battery size and sample granularity, determined by the average power consumption
- **Data availability:** whether data is aggregated for real-time access or stored in a distributed fashion
- **Reliability:** influenced by the channel access method and the decisions of other devices on their availability for ranging

In all cases, the network should try to evenly distribute energy consumption over devices, as all nodes share the same hardware. In case where differences are unavoidable due to factors such as the role inside the ranging protocol (initiator or responder), the system should achieve graceful degradation; we address this issue in Section 3.5.2. Furthermore, as we assume potential mobility for all nodes, it is important to have the possibility to merge multiple clusters and synchronize them on a common time base. We further assume that nodes inside a deployment share the same parameter setting (while not relying on any shared history and similar state) to limit merge complexity.

While a protocol should provide the given properties, it must also aim to reduce unnecessary energy and time consumption. In general, the usual sources of overhead for such a protocol can be aggregated as follows [39, 70]:

- *Idle listening:* the main source of energy wastage is usually due to receivers being active over long periods of time without actually receiving packets
- *Overhearing:* waking-up to and listening to all messages inside the network instead of only the ones which a device should actually receive
- *Overemitting:* sending packets while the intended receiver is not listening
- *Collision:* occurs when nodes try to simultaneously access the channel
- *Traffic fluctuation:* while usually providing for the worst-case, one can try to dynamically adapt and optimize to varying requirements on-the-fly
- *Protocol overhead:* control packet overhead which does not directly contribute to the useful data

In conclusion, our protocol must both provide flexibility in the defined metrics as well as limit its exposure to the sources of inefficiency as detailed above. In the rest of this chapter, we will further explore how previous protocols covered this design space. The next chapter will show how our system tackles each of the metric and offers a wide span of configurations.

We can separate the problem statement into two phases: First efficiently discovering that other nodes are in communication distance and then conducting range measurements with them. For those two objectives, we leverage our dual-radio hardware design as well as specialized protocols and therefore discuss both applications separately in the following subsections.

3.4.2 Neighbour discovery

Neighbour discovery focuses predominantly on two factors: high energy efficiency and low discovery latency [52]. This is primarily determined by the duty cycle of the radio, the fraction of time that it is turned on for either transmitting or receiving packets.

Neighbour discovery has been treated repeatedly from the original birthday protocols [50] to the recent Nihao family [52] and therefore offers us tested solutions as a source of inspiration. As time synchronization is difficult and energy-expensive, most practical algorithms rely on asynchronous packet exchanges and can be classified into probabilistic and deterministic protocols. Dutta et al. [51] note that probabilistic discovery increases the unpredictability of discovery latency and can lead to arbitrarily long tails in discovery probability. They propose a deterministic schedule based on pairs of primes and the Chinese remainder theorem to limit the maximal discovering latency and allow for different duty-cycle

requirements. This is especially relevant in our case, as reliable and bounded discovering latency is a crucial factor for tracking systems. Inspired by this work, Qiu et al. [52] define the concept of *talk more listen less*, founded on the realisation that short beacons require less energy than long listening periods. They part from the paradigm that listening and beaconing must be balanced and start sending an increasing number of beacons to improve discovery probability and energy efficiency. While the Nihao family outperforms previous solutions and leads to shorter discovery latencies compared to previous work, it is also highly variable and can be adjusted to the desired scenario. This concept is continued in BLEnd [53] as described in Section 2.3.4.

BLE neighbour discovery

As our platform contains a Bluetooth chip for communication with mobile devices in any case, our use case encourages the usage of such a dual-radio scheme of an energy-expensive (UWB) radio and a light-weight (BLE) wake-up source. However, we only want to wake up when we can also perform ranging to prevent wasting energy; as Bluetooth range [71] is generally drastically lower than UWB [38], especially in in-doors settings where high frequency signals have superior penetration [23], this is implicitly guaranteed.

BLEnd [53] is designed with the peculiarities of the BLE specifications in mind and uses them to optimize parameters for lifetime. It further achieves a deterministic latency bound in the absence of collisions and accurately predicts detection probabilities in environments with a high node density. In addition, the authors propose an optimizer which permits a wide parameter range and computes the advertisement characteristics with the highest energy efficiency. BLEnd provides superior energy-efficiency over other state-of-the-art protocols and aligns well with our goal of providing a wide design space of efficient solutions.

Wake-up radios

First introduced by Guo et al. [72], an always-on wake-up radio requires only a bare minimum of energy to detect the presence of another node and subsequently trigger the powering-on of the main radio, which helps reducing the power consumption of high bandwidth links. However, due to additional hardware costs and complexity, implementations are rare in practice and are highly susceptible to noise [39]. Some implementations reduce the additional hardware overhead by directly leveraging the already present ranging transceivers and only complemented them with low-power detection circuitry [5].

Leader election

One of the complications when initializing a scheduled network is the unanimous decision on a leader. As all devices run the same code and therefore do not have inherent diversity, the deciding factor to break a tie during the election is usually either randomness or a unique feature of the device such as its Bluetooth address. As random backoff introduces non-determinism in the runtime of the algorithm and true randomness is non-trivial for embedded systems, we will utilize unique identifiers to decide between leaders.

3.4.3 Ranging protocol

Ranging applications for mobile nodes in low-power networks introduce significant challenges. In addition to physical constraints on devices such as weight and size to develop an applicable platform, addressed in Section 4.2, the system must exhibit the following properties:

- **Mobility:** All nodes should be unrestricted in their movement and can experience dislocation speeds up to 5 m/s, the speed of a running human. This also explicitly includes responders which, in contrast to infrastructure-based systems, are not assumed to remain stationary.
- **Scalability:** The protocol should gracefully adjust to larger networks as well as increased traffic loads and enable applications to effortlessly adapt to changing environments within a reasonable time scale.
- **Deterministic update rate:** In order to provide reliable and conclusive data, clear upper-bounds on data update latency must be provided and a uniform sampling distribution maintained.
- **Synchronized rangings:** For localization to work, the lateration algorithm requires quasi-simultaneous rangings with all participating responders. Furthermore, initiators should update their position in a close time interval to enable conclusive meta-data such as node density and event response analysis throughout the network.
- **Global time scale:** Researchers must be able to temporally compare ranges between devices, add known positions of stationary responders to calculate locations and match timestamps with external events.

As communication on the UWB layer provides high data rates but also incurs substantial energy costs, an efficient ranging protocol is crucial for the design of a low-power localization system. As soon as a device discovers neighbouring nodes, it will wake-up its main radio and initialize the protocol layer for accessing this secondary channel.

Media access control (MAC)

In general, MAC protocols can be separated into three common styles: random, slotted, and frame-based [39]. In random protocols, nodes neither coordinate time nor contend for channel access and therefore enjoy simplicity and flexibility at the cost of frequent collisions and long idle listening or frequent packet retransmissions. For slotted access, nodes maintain a synchronized clock domain and regularly wake up at the beginning of planned rounds to exchange messages in an contention-based manner. Finally, frame-based (or schedule-based) protocols such as LWB [49] schedule channel access down to individually assigned slots inside a larger frame, eliminating collisions and idle listening at the cost of organisational complexity and the involved messaging overhead [73].

This distinction between contention- and schedule-based approaches is primarily a trade-off between management overhead and collision occurrence. Scheduling requires precise coordination, which adds protocol overhead for the distribution of control information and might lead to overprovisioning as resources are not only acquired on-demand. By deterministically assigning slots, this reservation scheme avoids spending energy and time resources for collisions and reduces idle listening to an absolute minimum by defining the exact points in time when transmissions occur.
Contention-based protocols leverage randomization to distribute channel access over time and usually employ a back-off procedure in case of collisions. For low contention scenarios, this method achieves a high probability of successful transmission with minimal messaging overhead; however, it also leads to unknown send times and hence can force receivers to listen idly for long periods of time.

It has been shown that schedule-based protocols outperform contention-based ones in terms of throughput, especially when being used in high-load scenarios, at the cost of network-wide synchronization [73]. While strict scheduling solves the hidden node problem (not realizing that a transmission collided with another one), it introduces the exposed node problem (inefficient waiting, even though nodes might not interfere at all as they communicate outside of each others range). Scheduling is specifically superior for interactions which can be well predicted, such as periodic traffic, and if the interactions frequency lies above a sufficient level to warrant maintaining time synchronization [40]. Otherwise, contention-based channel access might be more energy-efficient because of inefficient assignments and protocol overhead.

Collision avoidance

While the previous discussion on MAC protocols focused on the usage of an individual channel, the concept of structured channel usage to avoid or resolve

collisions can be broken down more fundamentally into different ways of accessing the medium itself. Depending on the spacial direction, the location in time and frequency domain, as well as the exploitation of mathematical properties, we can distinguish different channels and must deliberately choose how to utilize the entire design space. To simplify this, the space is often separated into four division multiple access schemes [74]:

- **Space (SDMA)**: This dimension leverages directional antennas to transmit to and receive from physically distinctive regions and therefore avoid the interference of electromagnetic waves. While the concept is simple and effective, antenna directions are usually fixed and hinder its exploitation. Recent advances in antenna arrays opened up the possibility of directional gain by purely varying electrical instead of mechanical properties and might enable tight UWB arrays in the near future [75]. Currently however, there are no practical solutions to utilize SDMA for mobile sensor platforms.
- **Time (TDMA)**: TDMA schemes segment the time domain into disjoint slots. While this trivially solves the issue of contention, the separation requires sufficient guard space to prevent partial overlaps due to clock drift [70] and tight synchronization between devices.
- **Frequency (FDMA)**: This approach splits the frequency spectrum into orthogonal sub-bands which do not interfere with each other. Due to protocol and implementation complexity and restricted flexibility of such solutions [70], partly due to the scarcity of available frequencies [74], adoption for sensor networks is limited. We will however show that the diversity of multiple frequency bands can strongly increase the accuracy of time-of-flight ranging and can be integrated into TDMA schemes.
- **Code (CDMA)**: The usage of a spread spectrum based on orthogonal codes can establish additional channels on top of previous solutions. However, due to the high computational costs of the involved algorithms and a wasteful usage of bandwidth [70], they are ill-suited for most sensor networks and not supported in common standard such as IEEE 802.15.4 [37].

As discussed, SDMA and CDMA do not match our criteria and are therefore disregarded. However, even with TDMA and FDMA, care has to be taken when applying them for ranging protocols. As these mechanisms aim to prevent simultaneous channel usage and offer orthogonal channels, they inherently impede broadcasting operations where entire clusters (in our case responders) intend to receive the same packet. As we will see later on, many of the traditional MAC schemes are not applicable to our use case for the same reason.

Topology and communication patterns

As mentioned above, MAC protocols must also take the communication patterns at hand into consideration. For this, we distinguish between unicast (1:1), broadcast (1:N), aggregation (N:1), and mesh (N:N) communication. While aggregation to gateways used to dominate for wireless sensor networks [70], mesh communication is increasingly interesting with emerging specialized protocols for many-to-many communication [49, 76] where traffic flows from multiple source to multiple sink nodes. This diversification can better tolerate node failures and high mobility as it does not rely on fixed routing paths.

Modern ranging protocols [5, 23] depart from the classical pair-wise ranging primitive which relied on unicast messages and leverage broadcasts to drastically reduce the message complexity. As we will see later on, the protocol proposed in this thesis takes advantage of broadcasts to further reduce the previous message complexity of $O(n^2)$ to a linear $O(n)$ for any realistic number of nodes inside a network. Furthermore, we show that such patterns can have a strong influence on the choice of a suitable MAC protocol and the resulting energy consumption.

Topology independence

Most protocols rely on a certain structure of the network, usually by constructing spanning trees or mesh networks, to route packets. Especially for mobile nodes, this increases complexity and overhead, as link quality might change frequently and links can break down at any moment, causing the forwarding path to change and requiring a re-routing of packets. A novel idea based on constructive interference by Ferrari et al. [55] achieves highly reliable routing without requiring any network state information. Removing the bookkeeping of neighbour information and the reliance on single sink nodes for global scheduling furthermore reduces message complexity and increases robustness to errors.

Time synchronization

A common time base is a fundamental requirement for many use cases: Multi-node coordination (e.g. for active and passive phases), the scheduling behaviour of TDMA and matchable sensor data timestamping call for a known time throughout the deployment. Reliable timing demands compensation and constant recalibration for clock drift due to differences in clock frequencies as well as varying environmental variables such as temperature and the stability of the power supply. Furthermore, the network must be initialized upon startup to a common value to prevent clock offset. This can be achieved either for external (i.e. global UTC time) or internal (local network time) synchronization depending on the

requirements and available information inside the network. A common way of solving clock drift is through periodic beaconing messages which include timestamps [55] or by leveraging round-trip times [23]. For multiple hops, the added jitter grows with the square-root of the distance [77].

3.4.4 MAC protocol comparison

Choosing between different MAC protocols or the parameter values for a specific MAC protocol always includes trade-offs. Some protocols, such as *MiX-MAC* [68], try to choose the optimal variant on the fly out of a pool of compatible protocols and adapt to the current network state and data requirements so that transitions are transparent to the end user. Similarly, we will introduce fixed parameters for system operators to decide on such trade-offs and propose dynamic adaptions of the performance based on the distributed energy consumption.

In the following paragraphs, we explore previous design considerations and analyse their underlying principles. We will show how protocol design evolved and draw conclusions for our particular requirements which we will then apply to the protocol proposed in this thesis.

Sender-initiated: Low Power Listening

Low Power Listening (LPL) (sometimes also called preamble sampling) protocols [41] were introduced as a reaction to the realisation that idle listening is not only non-negligible, but may easily dominate the power budget if not taken care of [68]. Therefore, a node which possesses data will access the channel and transmit a continuous signal until it can guarantee that all receivers, which periodically sample the channel, have noticed its desire (see Figure 3.4) and will receive its data packet. It was shown that this principle can shift the energy balance for the benefit of the overall network: the single transmitting node (with long preambles) will spend much more power compared to the receiving nodes (which only briefly sample the channel). This asymmetry can be exploited for low duty cycles when the majority of nodes are not actively trying to access the channel but only sample it regularly. However, the preamble introduces long latencies and therefore requires latency tolerance from the application, as the average delay between sampling and data reception directly correlates with the preamble duration. SpeckMAC [42] builds on the concept of LPL, but leverages time synchronization to go to sleep between preamble and packet transmission and only wakes up when the data packet is actually sent. MiX-MAC [68] internalizes this idea and suggests adapting the MAC layer protocol on the sender side to the current network characteristics for more flexibility.

Long preambles have become impractical with the increasing use of packetizing radios, which take the input as the payload of the packet and add their own preamble, header information and CRC [40]. This development lessens the direct control overhead of the microcontroller and therefore enables it to abstract low-level radio control; however, it also reduces the flexibility of the developer and prevents implementations to use continuous data streams. LPL protocols such as X-MAC started to adjust to this [44] but still suffered from channel occupation, low energy efficiency, and a high chance of collisions with higher duty cycles.

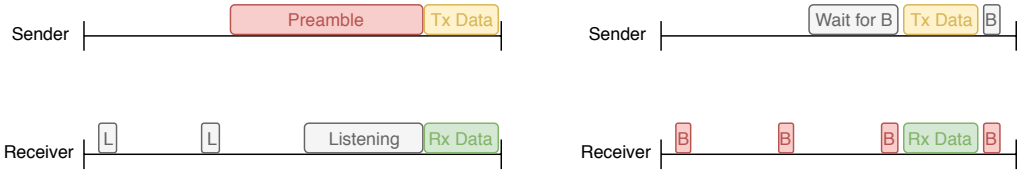

Figure 3.4: Sender-initiated transmissions (left) broadcast a long preamble to make receivers which periodically sample the channel aware of an impending transmission. Receiver-initiated packets (right) use beacons to signal that a specific node is currently listening for data.

Receiver-initiated: Low Power Probing

After LPL, research migrated towards low power probing (LPP) [44, 78]. This shift occurred due to a demand for improved throughput; the long preamble blocked the channel for long periods of time and proved to be inefficient for higher traffic demands [40]. The preamble transmissions blocked the channel for neighbouring nodes and required nodes to stay awake for long times while they might not even be the intended receiver. Receiver-initiated protocols reduce the cost on the sender side by forcing receivers to signal when they are ready for reception and keep the channel free otherwise [44]. A node with pending data will remain awake and listen for a beacon from the target node, signalling that it is ready to receive the packet (see Figure 3.4).

With this method, LPP solves the hidden terminal problem and supports asynchronous communication while avoiding long channel occupation [77]. Even though this suits wireless sensor network applications where data only originates sporadically and has a clear intended receiver (unicast), ranging demands a broadcast-like transmission pattern where all receivers are listening simultaneously. Furthermore, as all initiators demonstrate the equivalent traffic pattern, this will result in a temporal clustering of requests [40]. Nodes will implicitly synchronize upon waiting for receiver indications and start answering such beacons simultaneously. Therefore, LPP would result in heavy contention of initiators while not achieving the necessary synchronization of responders. Hence, we

can conclude that receiver-initiated communication does not suit our application style despite its advantages.

We verified using simulations that contention is already preventing deployments using asynchronous protocols at a sufficient scale. Additional factors such as poor energy efficiency and a limited duty cycle of responders would impede such schemes even further. LPL would only increase contention due to long preambles and therefore worsen the problem; even further improvements to decrease the preamble length such as schedule learning [43] cannot solve this particular problem, as all initiators target the same group and would therefore trigger simultaneously. LPP on the other hand would worsen contention even more, as all initiators would respond to receiver probes and therefore repeatedly collide.

Schedule-based protocols

In both previous solutions based on asynchronous communication, we either synchronized senders (LPP) or receivers (LPL). However, our use case requires that the entire cluster is synchronized: Responders must be available simultaneously to gather rangings at a single point in time (as otherwise, localization will include errors due to time inconsistency), while clustered requests from initiators will benefit the responders as they can aggregate responses into a single packet (as demonstrated later on). As such high activity in a short time frame would lead to extreme contention without explicit coordination, and the previous solutions only synchronize either one or the other side, a schedule is required which clearly defines the entire course of action. This requirement for synchronized measurements is the primary reason why schedule-based protocols are advantageous for our target application space. In addition to this, scheduling explicitly allocates resources to devices and therefore prevents high contention costs. As FDMA can be used to separate clusters in frequency domain but would prevent the reception of broadcast packets, most often TDMA is applied to divide time into slots and reserve them for specific nodes in a collision-free manner [73]. This allows all nodes to act as sender and receiver and communicate with any given other node inside the network, eliminating the need for any unicasted packets which would otherwise drastically increase message complexity.

Such synchronous MAC protocols are usually not fully deterministic. As the environment changes over time, energy-efficient protocols must be able to react to variations and adjust their scheduling. Signalling messages such as requests to join and leave the schedule as well as special protocol messages (e.g. blaming devices for not responding, not using their slot, or interfering with other nodes) still require flexibility. Therefore, in addition to dedicated slots, most protocols implement a so-called common active period which is open to all devices and often relies on back-off approaches to resolve simultaneous channel access [40].

Early work in synchronous MAC protocols mainly focused on high throughput and explicit signalling before transmitting data through requests-to-send [45]. While additional work accelerated distribution over multiple hops and introduced adaptive period lengths [46], protocols still focused on strongly varying traffic streams and explicit channel reservation preceding any transmission and therefore did not optimize towards more deterministic schedules with less protocol overhead. Furthermore, those protocols concentrated on multi-hop tree structures common in wireless sensor networks. As our target application is strongly spatially concentrated, and the radio range of UWB exceeds 300 m [38], such optimizations do not suit our purpose and would decrease performance due to superfluous messaging. Z-MAC [47] introduced the concept of slot stealing and a combination of CDMA for low contention and TDMA for high channel utilization. While this increases slot usage in case of an abandoned slot due to a failing node or no available data, the additional idle listening time can create a sizeable energy overhead. The protocol proposed in this thesis solves the same problem by leveraging the periodic sampling pattern using one-time signalling to permanently solve the issue by adjusting the schedule.

Some TDMA protocols assign time slots according to known neighbours and a fixed topology [48]. As such topology information is neither available nor stable in highly mobile networks, channel allocations due to known hop distances cannot be applied. We therefore require a protocol which can abstract the network links away and does not require explicit topology awareness. Systems that are based on broadcast traffic patterns can benefit from such simplifications, as they must reach all nodes irrespective of the available routing pattern. Multiple approaches in recent years have targeted the all-to-all traffic domain [49, 55, 79, 80]. While those protocols specify means for efficient packet distribution, they do not address low latency reply patterns such as the ones required for two way ranging. In this thesis, we leverage the existing research in this domain for synchronized scheduling and use individual slots to enable high frequency, high accuracy and low power ranging.

3.5 Software Platform

In the following subsections, we present our protocol, based on the metrics and considerations detailed in section Section 3.4. We combine a flexible neighbour discovery protocol using BLE and an efficient and dynamic ranging protocol using UWB and propose how additional, built-in hardware on our current platform can extend its functionality.

3.5.1 Neighbour discovery

For our neighbour discovery protocol, we opted for BLEnd [53] as already presented in Section 2.3.4. It is a state-of-the art protocol which both optimizes energy efficiency and meets high discovery probability. Furthermore, it allows seamless interaction with consumer hardware such as smartphones and laptops using the BLE standard and therefore simplifies interaction with users. However, the original authors disregard wake-up energy costs which distort the optimization towards too many advertisements and does not correspond to real scenarios. Additionally, we implement bi-directional discovery using U-BLEnd and connectable discoveries instead of the F-BLEnd alternative, as such advertisements are required for interconnection with mobile devices and reduce energy consumption by as much as half the energy of F-BLEnd.

The Bluetooth specifications allows a range of 20 ms - 10.24 s between single advertisements [81]; as this directly corresponds to the discovery latency in BLEnd, these numbers constitute the border of the design space. Utilizing this with 5 nodes and a discovery latency of maximally 5 seconds, we can achieve a discovery probability of over 90 % while only requiring an average current draw of 0.31 mA. Compared to hardware with dedicated ultra-low power wake-up circuits [5], this is only an increase by one order of magnitude and not significant enough in the overall power budget of the platform to justify dedicated circuitry.

Leader election

Leader election is accomplished by leveraging the existence of two separate communication channels. As the schedule distribution occurs over UWB, Bluetooth advertisements can still be maintained after joining a network and will allow new nodes to discover the network without affecting other communication.
As a general rule, the node with the highest assigned unique identifier is chosen to become the master of the network. As soon as a node joined a network, it will include the master's ID in its BLE advertisements. Whenever a node discovers an advertisement with a identifier higher than the previous maximum, it will leave its own network and join the newly discovered one. This merges clusters and therefore prevents network fragmentation. However, it might lead to an unintended and unnecessary break-down of an existing network if a node with a high unique identifier wakes up and starts advertising. To minimize such unwanted interference, nodes inside an already existing network will only switch if they discover a new network and not just a single node; complementary, a node with a high unique identifier will join an existing network if it itself is not yet affiliated with any.

Connectivity loss to the scheduling master can occur due to two reasons: Either the master itself or nodes on the path experienced a power or similarly catastrophic failure and are (temporally) unable to maintain its service. While those cases differ in the cause of failure, the ultimate result is equivalent: A cluster of nodes is still synchronized but does not receive any further schedules.
We detect such a condition after a node fails to receive a schedule when it expected it. As this most probably demonstrates a temporal issue of the node itself (e.g. an object crossed line-of-sight and prevents communication), it does not immediately resort to new leader election but tries to regain synchronization by initiating a listening period. In case it receives poll requests or responses from other nodes, at least part of the network received a schedule and should have been flooding the schedule packet. Therefore, we can conclude that the master is still available and will assume that we are out of sync with the rest of the network. If the node however does not receive any packets for the duration of the last schedule (our best guess in steady-state condition), there is a high probability that the master node is unavailable. After a fixed period (which is configurable by the system operator) lacking any packet reception, the node will assume a loss of the master and attempt to re-establish the network.
As our UWB layer distributes all scheduling packets to the entire network, every single node possesses the complete network state. Therefore, the node assigned the highest unique identifier in the network takes over the master role and starts broadcasting a schedule. In case a node also does not receive any schedules from the newly assigned master, we must assume that the node itself migrated and resort to the most drastic measure of restarting neighbour discovery.

3.5.2 Ranging protocol

For the ranging protocol, we can conclude from Section 3.4 that scheduling-based approaches with an emphasis on many-to-many broadcasting patterns provide optimal results for our targeted scenario. Other solutions proved to be unsuitable due to their lacking support for coexisting synchronization of both senders and receivers as well as strong contention for periodic traffic rates. Our application domain demands for stable broadcasting capability to highly mobile nodes as well as low power consumption under various traffic loads; we find that those properties can only be reliably guaranteed by scheduling-based protocols. Especially the *Low-Power Wireless Bus* [49], as explained in Section 2.3.3, provides us with many primitives required for our application space: high energy-efficiency, network-wide synchronization, reliable schedule distribution and topology-unaware routing to support high node mobility. LWB uses constructive interference to flood broadcast messages throughout the network, thereby avoiding the overhead required for establishing and maintaining network topology and routing state, and enables previously unprecedented reliabil-

ity rates of up to 99.99 % despite low energy consumption. Another benefit of LWB is its round-based structure and tight time synchronization on a network scale, independent of initiator or responder roles, reducing idle listening times and minimizing overemitting costs by clearly defining radio transmission times.

Similar to SurePoint [23], we leverage LWB for the reliable distribution of a schedule and then substitute the data packets with a custom transmission scheme. However, we aim at reducing idle listening to a minimum by leveraging global time synchronization to avoid recurring contention. Our work distinguishes itself in the following points:

- Inside the scheduling packet, we do not only fix initiators, but also define a clear order for responders. This avoids the costly, contention-based replies by defining global, known transmit times. The scheduling of responses significantly boosts the energy efficiency of both initiators as it removes idle listening costs and responders through the elimination of retransmissions.
- Instead of response messages for each initiator, we aggregate timestamps into a single response packet per responder after all initiators finished their broadcasts. This reduces message complexity from $O(n^2)$ to $O(n)$ at practical scales (for up to 100 nodes).
- We leverage the known schedule to reduce inefficiencies according to what we coined the *ranging pyramid* principle. To address unequal energy expenditure, we alternate between increasing and decreasing ordering for the individual sets.
- We separate schedule distribution and the contention slot and place the latter at the end of the round to reduce join latencies. Furthermore, instead of a fixed, single slot, we adapt to contention levels and vary their amount accordingly.
- We propose the concept of energy-aware ranging to allow application designers to decrease energy cost and obtain more flexibility when designing their system. This enables devices to gracefully degrade their service with decreasing power instead of instantaneous drop-offs.

Our platform can be used for general purposes, as it does not restrict node mobility for any device in the network, permits a large heterogeneity in energy supply from different battery sizes to stationary, wall-powered infrastructure and does not rely on a specific network topology. Unlike some other protocols (such as for schedule learning [43]), all devices have homogeneous hardware and therefore do not require special maintenance or charging schedules.

Inspired by LWB [49], our protocol uses a round-based approach. However, we reduce overhead by only transmitting the schedule once as we do not adjust the round period. Each round consists of three phases as shown in Figure 3.5, whereby the individual phases differ in length depending on the network. The complete set of information to determine the currently valid structure is defined in the schedule and gathered by a centralized organizer to prevent inconsistent views. The round period itself is equal to the measurement update frequency and depends on current traffic demands.

Figure 3.5: The round is divided into three phases: schedule distribution, ranging and contention access. In strict contrast to SurePoint, poll requests (blue) and responses (green) are clearly separated in time.

Scheduling

The overarching guideline when designing the presented protocol is the aim to reduce idle listening to its minimum using synchronization, thereby following a similar concept as the recent design paradigm of "talk more listen less" [52] which we already followed for neighbour discovery. As UWB reception energy consumption dwarfs transmission expenditures both in prolonged duration as well as current draw, we focused on eliminating contention-based access under all circumstances. Tight time synchronization paired with deterministically scheduled behaviour enables us to reduce inefficiencies due to collisions to a minimum and provides temporally stable connections to other nodes.
Deterministic and strictly periodic data acquisition furthermore eliminates possible error sources such as instability and gaps in a measurement series. The regularity on the other hand simplifies extrapolation and data analysis: a periodic sample distribution over time allows for easier signal reconstruction compared to aperiodic sampling [82] and gives a clear upper-bound on the staleness of data for real-time measurements.

Our scheduling method guarantees a fair service for all devices in the network with equal update rates and balanced energy consumption for nodes in the same classes as described below. Measurement intervals are fixed and therefore provide a known, bounded latency for real-time applications. Leveraging the knowledge of dedicated communication slots, we experience neither overhearing nor overemitting and can limit contention to the explicit contention-access phase where we intentionally use its indeterministic behaviour.

The schedule is defined by a central master entity which initiates a Glossy flood [55] at the beginning of the round. It consists of the following parts:

- **Initiator schedule:** Number of scheduled initiators as well as byte array containing their UIDs. Nodes are ordered by classes, whereby pure initiators come first so that they can directly go to sleep afterwards, followed by devices which both initiate and respond (hybrids) and hence need to remain awake in any case.
- **Responder schedule:** Number of scheduled responders as well as byte array containing their UIDs. Initiators only have to follow it until they received sufficient replies and can decide when to go to sleep independently.
- **Contention slots:** Depending on the contention experienced in the previous round, the number of contention slots will vary and allows quick network expansions at low constant costs.
- **Global timestamp:** As nodes experience clock drift over time and individual resets might occur, causing nodes to lose track of time, the master regularly broadcasts its time base to sync timestamps on the entire network. This packet is optional and is only added infrequently to the schedule to reduce overhead.

We require as many retransmissions as the maximal hop distance to flood each packet. However, restrictions by the FCC [36] require a minimal message interval of 1 millisecond between transmissions from the same node. Hence, the schedule distribution overhead accumulates to 5 ms of channel occupation for our assumed use cases (distances between nodes of maximally 150 meters). By setting the number of LWB network retransmissions per node to $N = 1$ and sacrificing reliability, we could, however, in theory shorten the message interval to as little as 200 µs, reducing the overhead to 1 ms. Therefore, schedule distribution could potentially be reduced to as little as a 1/30th of a single ranging event.

To successfully receive the next schedule, nodes must switch on their receivers slightly in advance. As the clock drift of the oscillators we employ can accumulate to up to 6 ms over a round period of 5 minutes, earlier wake-up times due to this imprecision can outweigh the schedule distribution itself. This overhead is directly inversely proportional to the update rate and is negligible for update rates in the order of Hertz, as clocks drift by a maximal amount of 20 parts per million in relation to each other.

Ranging

Distance estimation based on time-of-flight relies on local timestamps to calculate the difference between sending a packet and receiving a response. As shown in Figure 2.2, this delay is composed of the travel time of the electromagnetic

waves between the devices and the processing time at the responding device. By precisely defining this turn-around time, one can extract the pure time of flight and deduce the corresponding distance.
While the principle is simple, such two-way ranging does not scale well due to the two packets required between each pair of devices. SurePoint [23] solves this issue by introducing an innovative ranging protocol which relies on broadcasting the poll request and then individually responds with one packet per anchor, thus nearly reducing the number of packets by half. Multiple rangings between the same pair of nodes can further be aggregated into a single response. This effectively results in a single two-way ranging which synchronizes the two devices and additional one-way rangings which can then make use of that synchronization. Unfortunately, the clocks on the two devices can be affected by external influences and slightly differ in frequency. Hence, clock cycles are not equivalent and the protocol must correct them by estimating the drift during the ranging event before the final calculations.

Inspired by SurePoint, we also make use of frequency and antenna diversity to improve the quality of the ranging estimates both in terms of accuracy and precision. As both sender and receiver offer three antenna to make use of the optimal antenna polarization and we can send on three separate frequency bands to evade interference with other devices as well as local multipath, we must send $3^3 = 27$ poll packets per initiator. In order to compensate for the specific clock drift of each device, we send 3 extra packets as a reference which allow the initiator to calculate the relation between clock domains using the equations below. Therefore, a ranging event is split into three types of ranging messages (also shown in Figure 3.6):

1. **Polling:** The initiator sends 27 broadcast packets on all physical channels to perform one-way diversity ranging using maximal diversity. It afterwards uses the single response to simulate a two-way ranging and synchronize the clock domains.

2. **Reference:** The initiator then adds an additional three messages, which are a repetition of messages 1, 10, and 19. These specific packets are chosen for robustness because they each make use of a different frequency band. The contained timestamps are then used to calibrate the crystals by knowing the exact offset between them in both clock domains.

3. **Response:** The responder answers the polls in a single packet, in which it includes the necessary receive timestamps of the 27 poll requests for the range calculations as well as the 3 reference packets. The packet contains responses for all initiators it heard during the previous phase; in case they exceed the maximum packet length of UWB, the packet is fragmented.

Based on those timestamps from both initiator and responder, we can compensate for the clock drift by multiplying with the crystal offset factor K and then compute the resulting time of flight:

$$ToF = [(Init_{RX\_Resp} - Init_{TX\_Poll}) \ - \ K * (Resp_{TX\_Resp} - Resp_{RX\_Poll})] \quad / \quad 2$$

$$K = \frac{Init_{TX\_Ref} \ - \ Init_{TX\_Poll}}{Resp_{RX\_Ref} \ - \ Resp_{RX\_Poll}}$$

While this scheme is similar to SurePoint, we differ in the way responders send their replies. First, we introduce scheduling for responders as well, which enables us to eliminate the long (and fixed) contention windows. As responders know exactly when to transmit their packets, we can prevent any idle listening costs and additionally exclude inefficiency due to collisions and subsequent retransmissions. Secondly, responders do not reply immediately to a single initiator, but aggregate all timestamps. After all initiators have completed their poll requests, the node then broadcasts a single response to all interested initiators. The broadcast allows us to send a single reply packet (albeit a longer one) instead of one to each initiator. This batching of responses has only been enabled by deterministically scheduling responders; in the old contention-based scheme, it would not have been possible to acknowledge the entire packet.
Using broadcasts exclusively results in a significant reduction of the message complexity for predominantly hybrid networks from $O(n * n)$ (quadratic) to $O(n + n) = O(n)$ (linear) and hence improves the scalability of the system. Previously, the number of responders was further limited by the contention slot length and could result in a degrading performance due to excessive contention. The elimination of contention between responders enables a seamless extension of such responses, leading to additional data and higher precision. The response messages themselves do become longer however, as multiple of them are bundled together. Nevertheless, thanks to the high 6.8 Mbps datarate of UWB, the packet duration increases less than 100 us for each additional initiator. This only has a minor influence with an overhead of less than 20% percent for 5 nodes and leads to even higher efficiency gains for larger networks as guard time between individual packets is eliminated.

By exploiting symmetry, we can reduce energy consumption by an additional 50%. Devices that act both as initiators *and* responders can sleep after their ranging slot until they are scheduled to reply as responders. For a network made up entirely of hybrids, such as those employed for infrastructure-free interaction tracking applications, this will halve the aggregated on-time of the network.
We term the underlying insight the *ranging pyramid* principle, shown in Figure 3.7: As the distance measurement from node A to node B is equivalent to the one from node B to node A, such a node A acting as initiator *and* responder

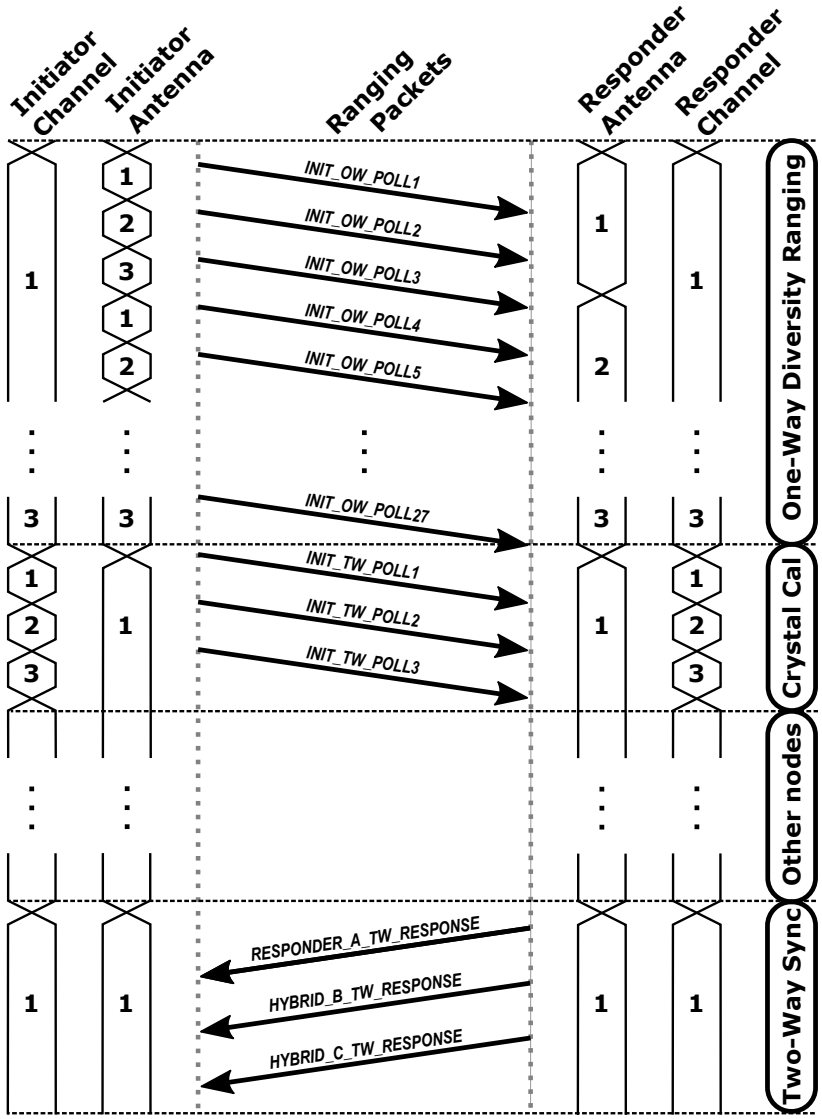

Figure 3.6: Each initiator sends 27 + 3 packets in the initial phase. After all other initiators finished the first half, responders will send their messages to complete the two-way measurement. Figure adapted from [23].

already knows the distance to all responders after its poll requests and therefore need not wait for nodes it has already heard from. It is, however, important to notice that this only works because scheduling allows all pure initiators to execute their rangings first. Discontinuing the listening period does not lead to any loss of data, as post-processing of data can compare timestamps for neighbours and complement each others ranging logs with the appropriate measurements. Throughout this first phase, this allows hybrids to progressively go to sleep until all initiators have ranged and it is time to send responses.

In the second part of the ranging phase, we can leverage a similar pattern and build the second half of the pyramid. Pure responders did not have the previous opportunity to go to sleep during ranging. However, as they performed ranging with all initiators, their packet is the most valuable for all other nodes and will therefore be scheduled first. For hybrids, we equivalently argue that nodes which could only go to sleep at the very end possess the most information and are therefore prioritized. In reverse order, those nodes are now scheduled to broadcast their responses and finish the two-way ranging. This ordering compensates hybrids which were scheduled late during the round and increases fairness.

In order to further distribute energy savings as evenly as possible, we can alternate the order in which nodes are scheduled. While nodes might be scheduled in increasing order of their UIDs in one round, the opposite will occur in the next. This simple and deterministic scheme guarantees that independent of the placement in its class order, a node will achieve the same lifetime as other nodes in the same class.

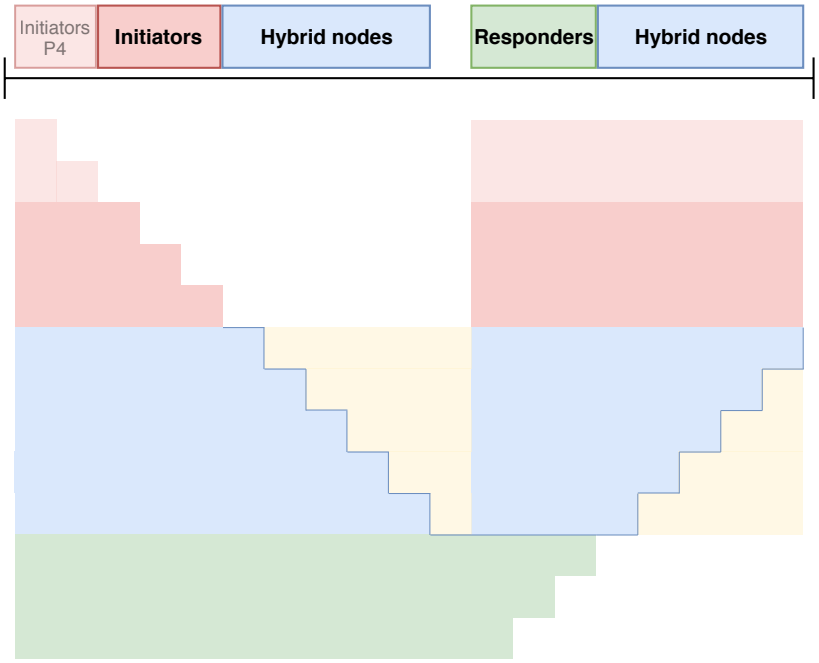

Figure 3.7: Initiators which sent out a request can go to sleep and will wake up for responses (red). Similarly, hybrids (blue) stop listening and can therefore save energy (yellow). Therefore, the number of hybrids which are awake is steadily declining. When hybrid nodes start their broadcasts, they are scheduled in the opposite order; this allows nodes which were forced to remain awake during the first half to quickly go to sleep and therefore evens out energy consumption across the network. As pure responders (green) constantly listen during the first half, they must return to sleep as quickly as possible during the second phase.

An initiator can dynamically adjust the number of rangings it uses for position estimation, as we use multilateration for location estimation. While it continues to receive more responses over time, a node obtains additional range measurements and will be able to generate more accurate results by including this data in the final calculations. This allows an application designer to directly trade-off energy consumption with localization accuracy.

Contention access phase

As contention is a potential cause for large protocol overhead, it is of crucial importance to minimize its effect as quickly as possible. Nodes which do not participate in Glossy floods (i.e. P3-4 as described below) can immediately go to sleep after their turn and are not affected by contention at all, which avoids idle listening completely.

While contention has been eliminated for the rest of the protocol, we deliberately use it in this phase in order to respond to varying traffic requirements and network sizes. A node will try to access the channel and send a signalling message if one of the following cases occurs:

- **Join the network:** A new node has synchronized to the Glossy floods, received the previous schedule and would now like to join the network.
- **Leave the network:** In case a node decides to leave the network (in order to join another one, due to an imminent lack of energy to support its functions or simply because it is switching off), it can request itself to be removed from the schedule.
- **Change parameters:** If a node decides to change its parameters (change its class, its power level, or its UID), it can notify the network using this packet type.
- **Complain:** This packet type can optionally be used to call the master's attention to a missing connection if an initiator did not use its slot or a responder did not follow its duties. If no other node in the next round asserts that it did in fact reach the device in question, one must assume power failure or other malfunction and deschedule the node. Such an algorithm could improve network stability by excluding unreliable nodes and prevents other nodes from wasting energy.

To dynamically adjust to current traffic conditions, the contention window size is adapted as follows: it doubles in size if half or more slots have been used, remains stable if between quarter and half of the slots were used and otherwise

decreases the width by half. This mechanism ensures that sufficient slots can be allocated while avoiding counter-productive size oscillations. If none of the slots were used, we can conclude that there is no more use for the contention window and its size is once again reduced to the default value.
We choose the approach of variation depending on usage instead of based on contention as collision is hard to detect in LWB networks where capture effect [56] still allows packets to be delivered successfully despite collisions. Therefore, we opt to not use explicit contention notifications (e.g. by seconding every slot with a another one in which devices might report the detection of collision) to increase efficiency and reduce protocol complexity.

Network discovery latency solely depends on the parameters set for neighbour discovery. Network join latency on the other hand is also influenced by the success probability of claiming a contention slot. By setting the standard amount of contention slots to a higher number, the expected latency drops at the expense of increased energy consumption. Therefore, this offers another opportunity for application designers to adjust the protocol to their specific use case.

Energy-aware ranging

To compensate the unequal effect of the ranging events, power levels allow devices to switch hybrid and initiator roles regularly to evenly distribute energy cost and strive to keep the entire network alive as long as possible. In general, it is hard to precisely predict the effective lifetime of systems and requires various considerations such as role-dependent variations, unknown recovery costs from error states, and unpredictable hardware. As a new concept for mobile ranging, we propose energy-aware ranging using the principle of power levels as an approach to mitigate this issue. As our protocol is targeted towards providing applications with the ability for guaranteed performance metrics, it does so far not incorporate the notion of probabilistic service. While determinism is an important option in many applications such as control loops and high-frequency tracking, some cases might prefer the trade-off of higher energy efficiency at the cost of reliability. To complement our protocol, we therefore argue that an additional layer based on strictly local decisions delivers a trade-off between lifetime and data acquisition stability and therefore broadens the platform's application space.

In some application scenarios such as multilateration, we require a minimum number of nodes in the perimeter to be active in order to ensure a sufficient number of rangings to achieve a given goal. As can be seen in Figure 3.7, such responders (green) are usually on for the entire first half and therefore suffer from quicker battery depletion compared to initiators and hybrid nodes.
Energy-aware ranging aims at giving application designers a tool to balance

this. If only a subset of the responders is actually required for localization or to create a connectivity graph in an overprovisioned deployment, we can switch to a lower power level for some rounds and only actively take part as a responder in a fraction of the rounds. This switch can occur in a distributed fashion and is therefore independent of the rest of the network. To ensure that a sufficient number of nodes still remain to respond to initiators, such a switch occurs probabilistically according to values set by the application designers. As an example, one could imagine that a node remains in power level P1 for 50% of its time and 25% each in P2 and P3. As our hardware can sense its battery level at run-time, one could also imagine that probabilities are adjusted to the current charge and the chance of lower power levels will increase when the battery shows indications of depletion, therefore leading to a gracefully degrading ranging service instead of sudden drop-offs. This also takes unbalanced battery depletion into account, as nodes with more charge left will be increasingly relied upon as responders, leading to a prolonged sustainability of the network as a whole.

Furthermore, we can use power levels for requesting special treatment to the network. As an example, P4 is intended as a level for nodes demanding ultra-efficient ranging and should therefore be prioritized in the schedule with an early slot so it can range and then go to sleep as quickly as possible (also see Figure 3.7). Such a node might either be powered by a smaller battery, employ less reliable energy-harvesting, or aim at longer deployments compared to other nodes.

The definition of power levels is further useful in deployments with mixed mobile and infrastructure nodes. It allows a simple inclusion of infrastructure nodes which can rely on wall power instead of batteries. Such devices can declare the highest power level in which they will always be used for both flooding and responding (P1) and can lessen the burden on other, more energy-restricted nodes.

	Polling	Flooding	Responding	Special remark
P1	X	X	X	
P2	X	X		
P3	X			
P4	X			Request special schedule

Table 3.6: While devices in P1 operate as usual, P2 nodes reduce extensive listening operations. P3 on the other hand additionally do not contribute to the schedule and signalling packet distribution. For maximal efficiency, P4 nodes can request to be scheduled at the beginning of the schedule.

We deliberately do not create a power level for nodes which act as responders, but do not contribute to Glossy floods. As nodes are required to listen and receive the scheduling packet in any case, the additional expenditure to continue the flood is small compared to acting as a responder which must remain awake for prolonged times. This trade-off would be unreasonable, as an increased probability of corrupted or even missed schedule reception could have far-reaching consequences for the entire network compared to minimal energy savings of single devices.

Devices which expose neither initiator nor responder functionality will not take part in ranging events. However, such nodes can support the network by helping with the flooding of the schedule using Glossy. This relieves other mobile nodes from additional energy consumption, improves delivery probabilities, and increases the potential range of the network. Especially in networks where low power levels (P3 and P4) are frequent, such devices can improve system lifetime and stability and are useful to guarantee coverage as infrastructure nodes.

Additional features

To maintain a global time scale, the device time must be regularly synchronized throughout the network due to clock drift of the oscillators. Such differences between devices can accumulate up to half a minute over the course of a week and are strongly influenced by external factors such as temperature and power supply stability. Our protocol mitigates prolonged clock drift by synchronizing the clock to the master node and therefore maintaining a single time source throughout the network. While being important for general timestamping, the synchronization advantage of schedule-based protocols is essential for multilateration as well as to complete ranging logs which miss entries because of the ranging pyramid principle.

To prevent wasting energy by performing ranging when nodes are stationary, we could leverage available accelerometer information on the hardware for event-based scheduling. The protocol allows us to deschedule a node from the network when no activity is registered for a given threshold time and quickly reschedule it as soon as renewed activity has been detected. This could significantly reduce the energy cost in stable situations such as offices where people remain seated most of the time and allows for focusing energy on high update rates when interesting actions actually occur.
This feature requires quick network join times to track people as soon as they start moving; our protocol inherently supports this even during highly contested times using its dynamic contention window. In the case of an entire cluster of nodes suddenly jumping into action, we can quickly and automatically allocate more slots and already schedule the discovered nodes in the subsequent round.

Chapter 4

Implementation

After having shown that the current state-of-the-art is lacking key requirements for a versatile, infrastructure-free tracking system, we argue in the last chapter how a new systems-level approach could mitigate those deficiencies.
We implement a prototype of our proposed system to demonstrate its viability and evaluate its performance in real-world settings. We begin with the high-level system architecture and outline the interaction schemes both between components on the same board and between multiple devices. We then introduce *TotTag*, a new hardware platform for social interaction studies. Next, we provide software tools for design space exploration. Finally, we treat embedded systems programming in more detail and discuss the modular state machine of the protocol implementation.

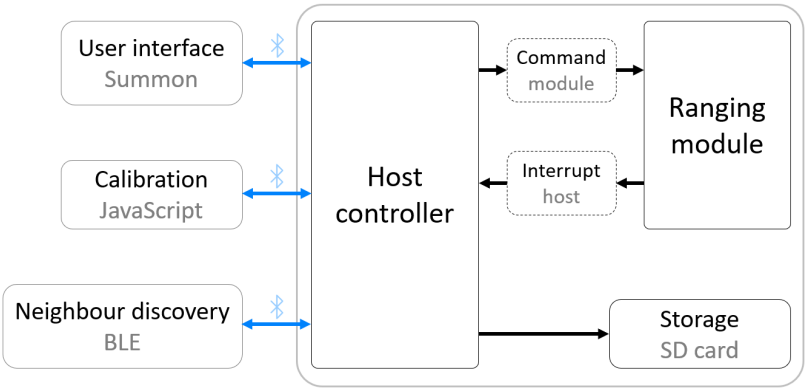

Figure 4.1: Block diagram of the major subsystems on an individual node. The host controller manages the interactions of the board with components outside of the network such as users and undiscovered nodes over BLE (blue links). It uses the command API as an abstraction layer to send instructions to the module and subsequently receives ranging data over interrupts, which it can log locally and relay further to the user interface.

4.1 System architecture

Figure 4.1 illustrates the data flow throughout the system and shows how other devices interact with a node. The host controller on the carrier board adopts the role of a central hub and acts as an indirection layer for external access to ranging data over the user interface and for calibration purposes. At the same time, it implements the neighbour discovery protocol over BLE, discussed previously in Section 3.5, and informs the module about changes in its surroundings such as newly discovered networks.

4.1.1 User interface

To facilitate system health monitoring and offer real-time tracking to the end user, our system leverages its BLE radio to support direct communication with smartphones and laptops. Users can use the available Summon app [83] to automatically discover devices in their proximity by leveraging the Google Eddystone interface. This provides users with the opportunity to interact effortlessly with our platform over their personal phones. As shown in Figure 4.2, the app discovers all TotTernary devices in its environment (five in the given example) and then loads a custom JavaScript application which visualizes the current ranges to other devices. If an infrastructure-based network of anchor nodes is available, we leverage the iterative solver presented in previous work [23] to directly update the node location.

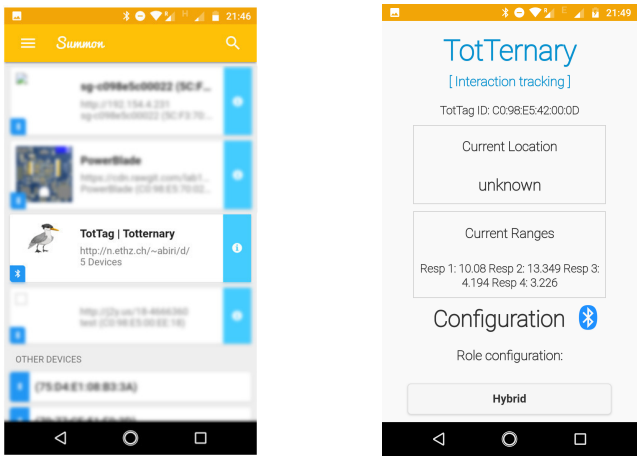

Figure 4.2: Together with the Summon app, our user interface enables simple interactions with devices: After having discovered them over BLE (left), users can observe ranges in real-time and configure the device on-the-fly (right).

In addition to receiving ranging information from the node, we can also employ the user interface to send commands and configurations directly to individual devices. This allows researchers to set up nodes in the field, configure their modes at run-time, and establish global time synchronization. An additional use case is wireless debugging, which permits testing nodes during dynamic deployments and retrieving information about errors in real-time to identify failure cases.
We leverage this two-way communication to use the application as a remote control to start and stop the device at will, thereby conserving power when it is not intended to be used actively such as during charging. It further allows users to back-up and restore the current system state from permanent storage, e.g. in case where batteries should be exchanged.

4.1.2 Calibration

As individual boards and radio ICs slightly differ in their electrical properties, calibration is required to correct for the antenna delays which could otherwise distort ranging measurements by up to 150 meters. A device-specific factory calibration entirely compensates for the particular variations of antenna trace lengths and solder joints. However, a one-time, universal characterization of the *design* drastically reduces deployment overhead and suffices for 30 cm accuracy [84] as it already takes the general board layout into account. To support simple deployments for future studies, we perform calibrations for 16 individual boards and provide the averaged results as default values.

To calibrate, we set up three nodes in an equilateral triangle with a known side length of 1 meter. A JavaScript application then configures the nodes over BLE and triggers a calibration event. During each event, one node will send a packet to synchronize its neighbours through simultaneous reception. Using this known point in time, the node to be calibrated sends two packets at fixed offsets. The resulting time stamps are then gathered over BLE and can be used to compute the antenna delays while compensating for clock drift between nodes.
This method only provides the combined measurement of transmission *and* reception delays, a fundamental limitation of timestamp-based calibration which can usually only be circumvented using expensive, high-frequency spectrum analysers [23]. However, we rely on separated TX and RX delays to correctly compensate for both frequency and antenna rotations during a two-way ranging. To solve this issue, we leverage the radio manufacturer's deep knowledge of the internal chip layout [84]: following their guidelines, we calculate $t_X = 0.44 * tr_X$ and $r_X = 0.56 * tr_X$ and store these individual TX and RX values in the node's flash memory.

4.1.3 Inter-chip communication

Communication between the host controller on the carrier board and the radio controller on the module takes place using an I^2C interface and a single interrupt line. To signal the target IC the correct way of interpreting the data coming over the bus, we take the standard approach of prepending an operation code to the message which defines the structure of the subsequent bytes.

Module commands

As the I^2C master, the host manages the communication and is by definition the one to initiate a transaction. For many of the operations defined in Table 4.1, such as `CONFIG` or `SLEEP`, a simple, uni-directional packet transmission from the host to the module is sufficient. Some interactions such as `INFO` additionally require reading information from the module and are therefore split into a `write` operation, followed by a `read` transaction in which a known, fixed number of bytes is expected on the bus.
In the case of data with unknown size, such as for the `READ` command, an extra third step is required. After first informing the module, as the I^2C slave, that it intends to read further data, the host then receives the number of bytes available in a first `read` transaction. This information is then applied for a second `read`, allowing the bus to dynamically adjust transmission sizes according to the type and amount of data present at the module.

Command	**Usage**
`INFO`	Respond with protocol identifier and version number
`CONFIG`	Setup the node with the given mode and initiate the state machine
`READ`	Send the data which was signalled over the host interrupt
`SLEEP`	Stop protocol execution and halt radio transmissions
`RESUME`	Restart the state machine with the previous configuration
`TIME`	Update the internal epoch time and trigger network-wide distribution

Table 4.1: Using various commands of a module API, the host interacts with the module over an I^2C bus using serial communication. This high-level interface hides the protocol complexity on the radio controller and enables a simple inclusion of the module into new designs.

Host interrupts

As the module cannot initiate a transaction over the bus itself, it must have a separate means of communication at its disposal. An interrupt line allows the radio controller to inform the host that it possesses data that should be relayed further to either the BLE channel or permanent storage.
Depending on the type of data, such as range measurements or an updated master EUI, the module will store an interrupt reason as specified in Table 4.2 as well as the data bytes in an internal transmission buffer. When the interrupt is triggered on the host, it will dispatch a `READ` command (see Table 4.1) and copy the data over for further processing. Similarly to the operation code, the interrupt reason allows the host to interpret the data and forward it to the designated destinations.

Interrupt	**Usage**
`RANGES`	Signal the host that ranging measurements are available
`RAW`	Notify the host about updated raw diversity ranging results
`MASTER`	Update the master information on the host
`CALIBRATION`	Report calibration data for further processing

Table 4.2: To inform the host about pending data, the module triggers an interrupt and then expects the host to initiate a transaction over the `READ` command so it can empty its buffer.

4.1.4 Persistent storage

Long-term studies require logging large quantities of data as they possibly gather several weeks worth of measurements. Microcontrollers are not well suited for such permanent storage due to their limited, volatile memory which is erased upon power cycling. SD cards offer a cheap, high-density, and non-volatile storage medium which can easily be detached from the board for data post-processing on more capable machines.

Unfortunately, these cards are also highly inconsistent in power draw and can be difficult to reliably power off while not being used. To mitigate this issue, we power gate the entire SD card connector and locally buffer data on the host controller until sufficient data is gathered. This batch processing allows a highly efficient usage of the capacity in small bursts without running the risk of unnecessarily depleting the system's battery over a prolonged period of time.

4.2 Hardware

The TriTag board used by SurePoint [23] already provides three UWB antennas to achieve antenna diversity. For interaction tracking, it does however lack key elements such as a medium for permanent storage as well as a battery charge controller to guarantee safe usage in longer deployments. Additionally, we have identified during testing that the limited memory of the ICs does not permit additional functionality. Therefore, we redesign the board to support a broader range of applications.

4.2.1 Naming conventions

In the naming of the hardware designed for this thesis, we follow the conventions set by its predecessors. The localization module is branded SquarePoint to emphasise its symmetry in RF trace lengths and its separation of controller and ranging functionality into two squares. The carrier board is labelled TotTag to exhibit its ability to be "tagged" to people and to underscore its portable size, as a tot is a synonym for "a small amount." Furthermore, it reflects one primary use case: a collaboration with psychologists at Vanderbilt University [25] who wish to measure the interactions between care takers and toddlers, or *tots*.

For the entire platform, we depart from the *-Point* nomenclature of its ideological parents [23, 35] to underline the shift away from infrastructure-based solutions to a mobile platform. As we use three antennas and three frequency bands, the usage of ternary ("composed of three parts") was combined with the underlying tot theme to form the platform name, TotTernary.

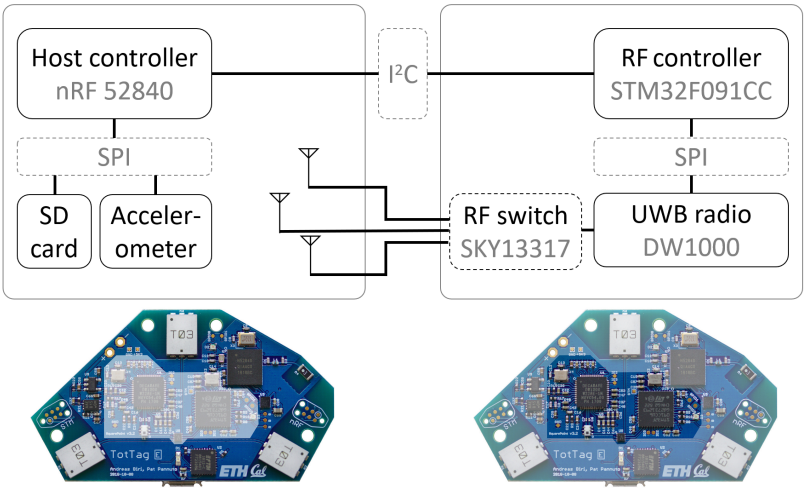

Figure 4.3: The architecture diagram shows the carrier on the left and the module on the right. The bottom emphasises the corresponding sections on the physical hardware and demonstrates how the module can be incorporated into a design.

4.2.2 Hardware design

To verify the usability of the system, we design and build a custom PCB according to the considerations detailed in Section 3.3. Figure 4.4 shows the physical board and the underlying technical drawing, on which the connections between components on the two outer layers are perceptible. Both the schematic and the final PCB layout are open-source and can be found publicly on Github [85]. We refer to the documentation available online [24] for detailed guidelines on how to adjust and deploy the software onto the platform.

As visualized in Figure 4.3, the carrier board and the module are merged on a single PCB. We use the SquarePoint module as a design block and then optimize its layout for the constraints of the carrier board by shifting connectors to the edge of the board for improved accessibility. Detailed suggestions for how to achieve good ranging performance can be found towards the end of this document in Section 7.2 where we also highlight insights into our development process and point out potential pitfalls.

Ranging module

The ranging module uses a DecaWave DW1000 [38] as a radio chip and the STMicroelectronic STM32F091CCU6 [57] as a controller MCU. We build upon the ideas of the original TriPoint [66] and improve its design in various regards:

- The MCU is upgraded from a STM32F031G6U6 to the STM32F091CCU6 [86], providing 256 kB Flash and 32 kB SRAM (an eight-fold increase). This overprovisioning simplifies development while still permitting a trivial downgrade for economical benefits as a 128 kB drop-in option is available.
- The SLEEP and DEEP SLEEP behaviour is correctly implemented according to the manufacturer's design guidelines [38]. Previously, low power settings were prohibited by noisy signalling.
- To enable extensive debugging, we add multiple, easily accessible UART test pads for the STM plus dedicated GPIOs for the DW1000, as well as a tri-color LED for visual feedback.
- The RF design has been improved to match the impedance requirements of the communication antenna and provide additional trace shielding using via stitching and keepout zones to other components. Additionally, the antenna traces are routed in a more symmetric and straight layout to reduce the variation of antenna delays and reflections.

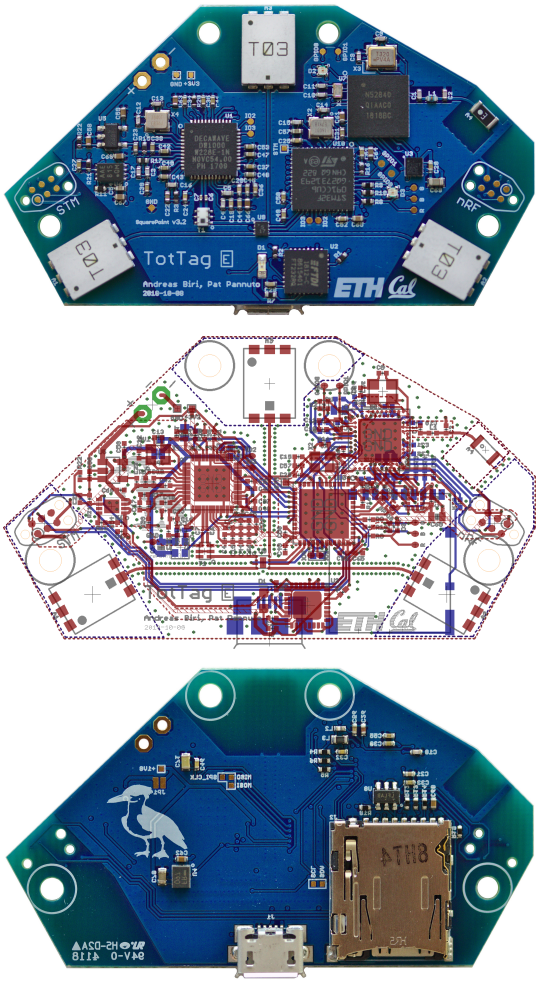

Figure 4.4: The PCB layout of the TotTag board: the red traces show the top layer of the board, while the blue ones reside on the bottom side.

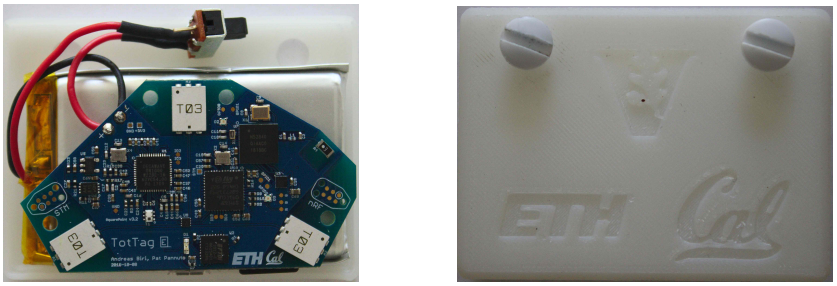

Figure 4.5: The board sits on top of the battery and is tightly fitted inside the casing (left). Plastic screws hold the top lid (right) and the bottom part together.

The previous design utilizes a triangular shape to simplify the arrangement of the three antennas with a phase offset of 120° as required for minimal cross-polarization. While keeping the phase offset, we diverge from the triangular shape as it incurs a central position of the RF switch required for the three antennas. Due to the potential interference of the RF traces with other signals and components, one usually tries to limit the intersection with such sensitive parts. This, however, implies that the circuitry should reside in the corners of the triangle, resulting in large areas of underutilized PCB space. To optimize the board for size as well as less interference with the RF traces, we therefore switch to a more rectangular shape.

Carrier board

For the carrier board, we again take inspiration from the TriTag carrier board [87]. We use an I^2C bus to communicate with the ranging module over the API described in Table 4.1. Apart from the MCU interconnect, SPI buses are used for high-speed communication inside the components, i.e. between the DW1000 and the STM as well as between the nRF, the SD card, and the accelerometer.

Despite the conceptual similarities, we upgrade the the carrier board to boost performance and add several new features to achieve the required specifications:

- We upgrade the MCU from a nRF51822 to the nRF52840, containing 1 MB Flash and 256 kB RAM (a 4- and 8-fold increase respectively). The new chip furthermore supports more powerful options such as a dedicated floating point unit (FPU) which allows the computation of the location with an iterative solver directly on the embedded device. It also adds the option for 802.15.4 communication in addition to Bluetooth, providing more flexibility for wireless connections which could be exploited in the future for customized neighbour discovery or meshing protocols.
- We add an SD card to locally log and permanently store data in large quantities. A separate transistor permits power gating the entire SD card and ensures that it only draws power when it is actually in use.
- We add an ultra-low-power accelerometer to obtain 3-axis acceleration information and enable features such as event-triggered ranging.
- We extend debugging capabilities by adding headers for 2 UART high-frequency outputs as well as a tri-color LED for direct visual status information.
- We upgrade the micro USB connector to support data by adding a USB-to-Serial UART using the FTDI FT232R [88] for data offload and serial debugging over USB.

- We improve the battery charging stability by adding the Microchip MCP73831 [89] as a battery charge controller to improve battery life and prevent battery overcharge. This extends battery lifetime and reduces the risk of harm for people wearing the device.

The nRF52840 offers IEEE 802.15.4 with the same 2.4 GHz antenna already installed for Bluetooth. While the 250 kbps data rate of 802.15.4 is lower than the 2 Mbps maximal bitrate of BLE, it offers the possibility to form multiple different topologies such as star, cluster tree or even full mesh [90] and can reach up to four times further [71]. Furthermore, the combination of 802.15.4 and BLE offers additional energy-latency trade-offs. The Nordic software stack already provides the popular Thread mesh network for 802.15.4 out-of-the-box and therefore allows for easy interchangeability between the two protocols.

4.2.3 Wearability

The board itself measures 61 mm x 35 mm and weights 7.7 g. With a 2000 mAh battery weighing 34.1 g and a custom built, 3D-printed casing, the dimensions of the total deployable system are 67 mm x 48 mm x 19 mm with a weight of 59.8 g. These properties are equivalent to LENA [63], a comparable wearable sensor, in weight and even achieve a decrease in volume by 18%. As the battery accounts for the majority of the weight, deployments which do not require multiple days to weeks of measurement can make use of smaller LiPo batteries such as a 350 mAh one at 7.9 g [91]. This can reduce the overall system weight by up to 43.8% while still providing more than a full day of measurements.

For testing, we design and build a 3D-printed casing which contains the board, the battery, and an internal power switch in a protected and small package. An opening on the side allows for direct charging by plugging in a micro USB cable as well removing the SD card without opening the case. Plastic screws keep the two parts together and ensure no interference with the electromagnetic properties of the system. Figure 4.5 shows the realized inside the case.
We utilize white plastic to make the status LEDs directly visible to the user wearing the device. While this has proven useful for debugging during initial deployments, preliminary tests with external participants report that the lights can be perceived as distracting and might especially disturb children. To mitigate this, we suggest turning off the LEDs after an initial verification stage to reduce the cohort awareness of ongoing measurements and to improve system lifetime.

4.3 Protocol configuration

We design our protocol to be adaptable to various deployment scenarios and offer system designers the opportunity of choosing a design point according to their own criteria. To support them in finding optimal parameters during design space exploration, we provide a detailed energy model which predicts the lifetime of the system and can be used to investigate the trade-off space in simulation.

4.3.1 Energy model

Directly based on extensive power measurements with our hardware platform, we build a tool to predict the energy consumption of the system during deployments. The simulation uses MathWorks' Matlab to evaluate various configurations and present the results in visualizations for direct comparison. We support the following parameters:

- **Network size:** As the number of initiators, responders, and hybrids strongly influences the estimated lifetime, investigating the effect of large clusters joining a network is critical and can strongly influence other considerations.
- **Update frequency:** Decreasing the sample rate can heavily influence node lifetime. A reduced percentage of active periods, however, also raises idle power consumption. In Chapter 5, we show that idle currents can have large, non-intuitive impacts on the system performance and must not be neglected.
- **Accuracy and precision:** Based on our large data set of over one million range measurements, we predict the influence of antenna and frequency diversity on the ranging accuracy and precision. Note that the increased use of diversity does imply more ranges and hence results in an additional energy trade-off that must be carefully evaluated.
- **Discovery latency:** Using our own implementation of the BLEnd optimizer [53], we can automatically determine the optimal parameters for the neighbour discovery protocol on our hardware platform. This also includes discovery probability estimations to give system designers a complete picture of the parameter effects on their future system.

Our implementation as a Matlab function enables users to trivially integrate the energy model into their own evaluation scripts. This tool allows them to sweep a large design space of their choosing and accurately predict deployment outcomes.

To facilitate data processing, we generate detailed graphs which split up energy consumption into individual parts such as the expense of the ranging protocol, the idle current consumption, the neighbour discovery overhead, and the cost of logging data locally. All simulations include predictions for the different operating modes (initiator, responder, and hybrid) so that the effect on different nodes can be seen in advance and appropriate measures, such as wall-powered anchors, can be considered during planning.

4.3.2 Module configuration

In addition to the parameters mentioned above, there are multiple supplementary configurations which can be exploited to customize the protocol behaviour in specific cases. While some of them are also available in the energy model, not all of them directly impact the energy consumption.

- *Real-time availability* of all pairwise distances requires disabling the ranging pyramid, thereby increasing energy consumption for hybrids.
- A *maximal number of responses* can permit initiators to discontinue the reception of responses and allows them to sleep earlier, saving energy at the cost of a reduced number of ranges.
- Enabling *network takeovers* by the second-highest node in the network prevents nodes from completely abandoning the network upon master migration, but might potentially lead to temporal fragmentation if the master should only be absent for parts of the network.
- Adjusting *timeouts* results in a trade-off between spending energy in the hope of still receiving a schedule from a known master after being out of sync and leaving the network early due to temporal instability.
- Configuring the *automatic timeout* after which a master periodically eliminates a node from the schedule is a balance between additional schedule overhead and efficient reaction to failures and migration.
- The amount of *standard contention slots* determines the immediate response latency of joining nodes but can add significant overhead in steady-state behaviour.
- The *UWB Tx power* can be reduced to save power if only short-distance interactions are of interest.
- Enabling *raw ranges* gives access to the complete diversity ranges but raises the activity rate on the inter-IC buses and further data paths (e.g. to SD cards).

Currently, all nodes possess a homogeneous set of configurations, such as update frequency, which is defined at compile-time. We discuss in Section 6.2 how our system could be extended with a more sophisticated scheduler to support heterogeneous node configurations. This would prove especially useful in combination with energy-aware ranging, as a node might reduce its energy consumption on the fly by requesting a reduced scheduling frequency. In contrast to our proposed concept of power levels, this method does not potentially waste the energy of other nodes which expect it to be available.

4.4 Embedded Systems programming

The software for both the host controller (the Nordic Semiconductor nRF52840) and the radio controller (the STMicroelectronics STM32F091CC) is implemented as bare-metal C code. This gives us full access to the available hardware and ensures that interactions with the UWB radio occur without interference from OS-triggered events or delays. For the Nordic chip, we include their newest SDK `15` libraries and make use of the Softdevice `s140` for the management of BLE communication. The radio controller uses the version `1.5` peripheral libraries from STM and integrates the newest DecaWave Device API `2.4` to support our silicon version `E` of the DecaWave DW1000.

4.4.1 Module initialization

In contrast to SurePoint [23] with its limited memory, we are able to always flash the complete code for any situation onto the radio controller and then choose the appropriate application on the fly. This flexibility is valuable from a deployment perspective, as neither the application (i.e. standard ranging, testing, or calibration) nor the mode (initiator, responder, or hybrid) has to be determined at compile-time and therefore does not require any existing knowledge about future use cases when flashing the devices. Nodes can be configured centrally over the wireless user interface, which enables batch adaptations of entire networks in seconds.
This run-time configuration is of particular importance when considering the role a node plays inside the network. As SurePoint hard-codes the role of a node either as master or slave, multiple clusters can permanently interfere as they cannot be combined. Furthermore, single master node failures or migrations result in a collapse of the entire network. Our solution on the other hand offers the possibility to adopt roles depending on the current state of the network, as detected by our neighbour discovery protocol, and can adapt to changing environments. This guarantees that in any case where nodes are in communication distance with one another, they will eventually be able to communicate and range.

Our modular mode design also allows us to enable and disable additional functionality during active operation. Using the concept of energy-aware ranging, a node can simply temporally switch its mode by changing a single variable and achieve power savings by operating on a lower power level. Likewise, a node which detects that it is being charged can enable responses on the fly and therefore support other nodes in its surroundings.

4.4.2 State diagram

On the host controller, the control loop remains simple: After the launch of individual services and a check that the module is available and initialized, the node remains idle and waits for the discovery of neighbours. BLE scanning and advertising triggers periodically as seen in Figure 4.6 and ensures that nodes detect each other. After successful discovery, the host configures the module and then awaits future interrupts. In the case where the connection over UWB is lost and the module times out, the host simply returns to its initial state.

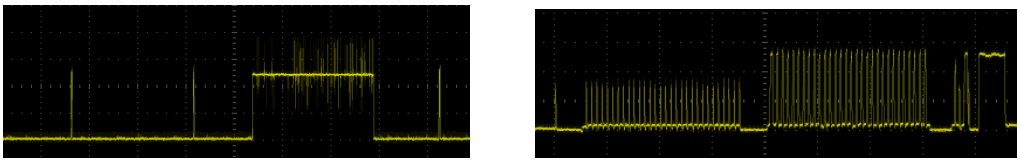

Figure 4.6: While discovering, we observe the individual current spikes of the advertisements as well as the scanning interval with a duration which is slightly longer than the advertisement period (left). During ranging between two hybrids, we see how the node first transmits 30 poll requests, then listens for another node's polls and finally transmits and receives the responses (right).

The module on the other hand uses a synchronous protocol for communication and therefore requires a sophisticated state machine. Figure 4.7 illustrates how, just as in a LWB round, the schedule is distributed and flooded throughout the network at the beginning of a round; similarly, contention access at the end enables nodes to send signalling packets to the master.
In stark contrast to LWB, the ranging events, detailed in Figure 4.8 and Figure 4.9, are *not* flooded and are only observed by neighbouring nodes. Following the previously received schedule, nodes interact with one another according to their modes and exchange time-of-flight ranging packets. An example of such a procedure can be observed in Figure 4.6 through a power trace. For each enabled mode (initiator and responder), the node advances a linear internal state machine to keep track of the gathered time stamps and imminent interactions. Depending on the currently active mode, Tx and Rx callbacks functions are switched on the fly to guarantee that events are correctly processed for the corresponding mode.

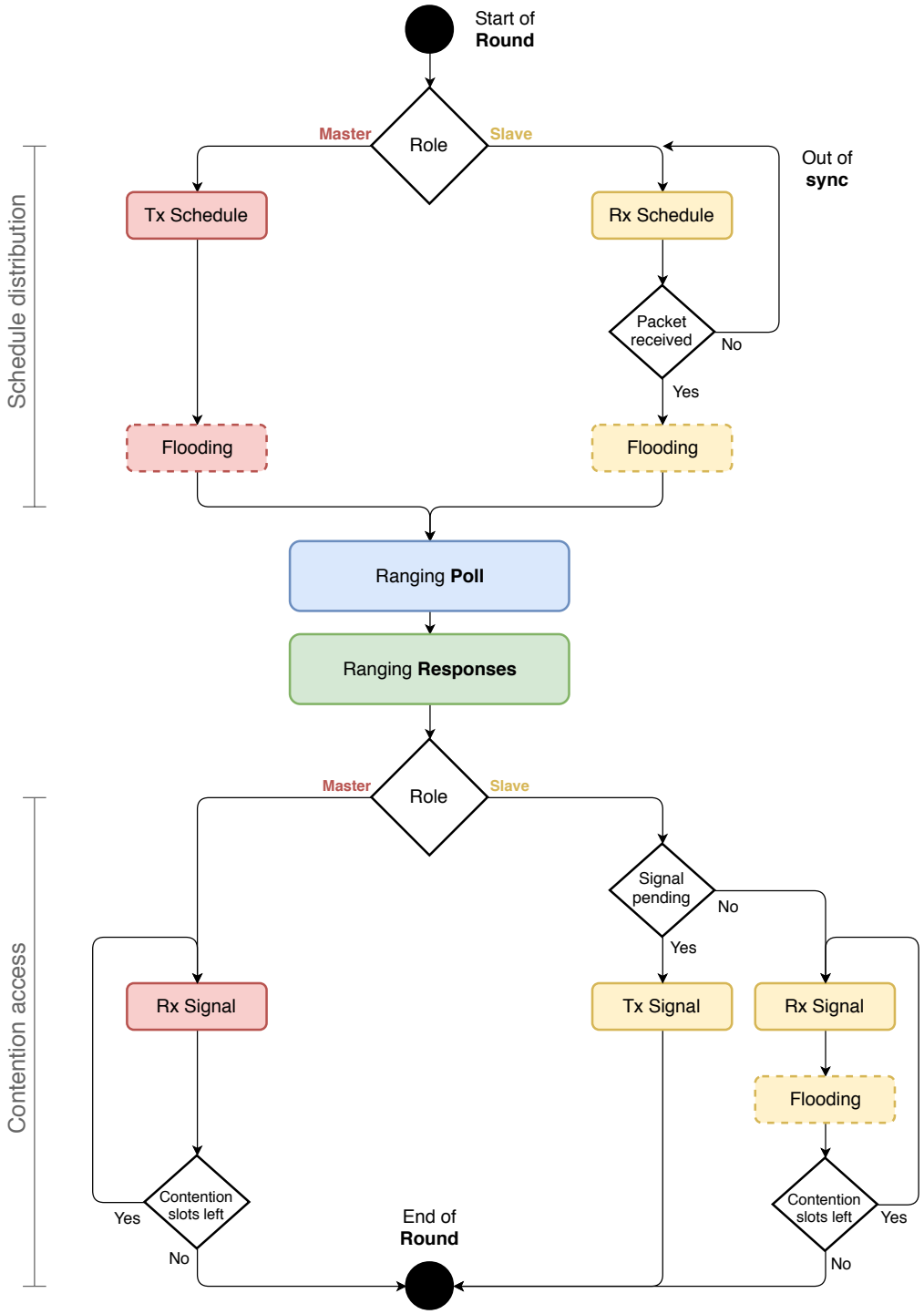

Figure 4.7: The round is separated into three phases: schedule distribution, ranging, and contention access. While the start and the end of the round is globally orchestrated and packets are flooded over Glossy, the ranging polls and responses are local and only affect one-hop neighbours.

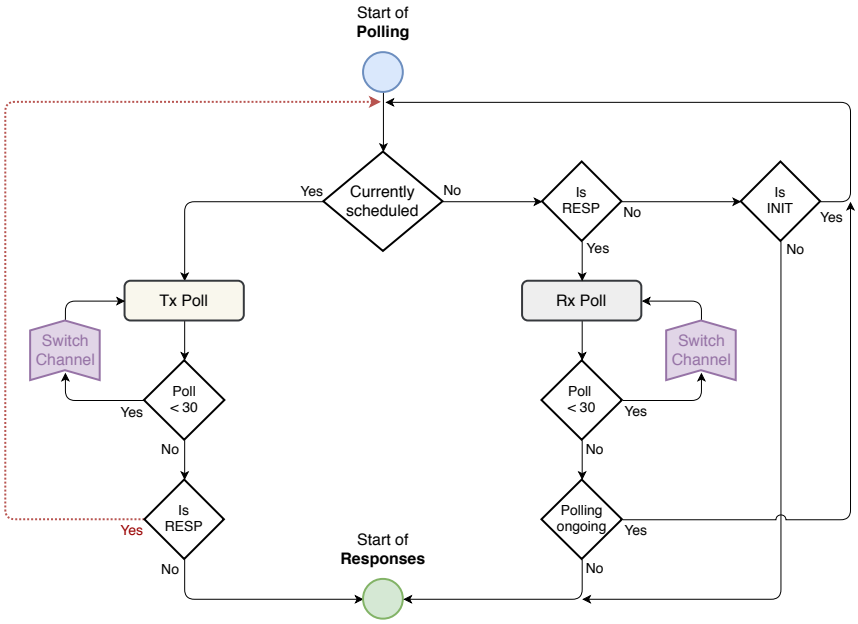

Figure 4.8: During the polling phase, initiators send 30 polling packets during their slot. Notice that the red return loop on the left is not required because of the ranging pyramid principle, therefore reducing energy consumption for all hybrids.

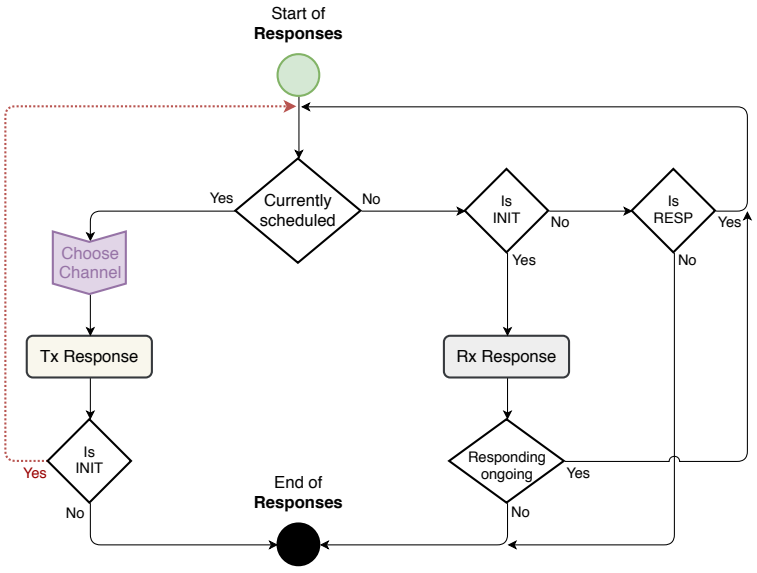

Figure 4.9: During the responding phase, responders send back the gathered timestamps so that initiators can calculate the ranges based on the time of flight. As in Figure 4.8, be aware that hybrids can save energy thanks to the ranging pyramid principle.

Chapter 5

Evaluation

To demonstrate the feasibility of our system, we test TotTernary both in controlled experiments as well as real-world deployments. We first evaluate its performance under the influence of varying range and physical channels and show that we can achieve decimeter ranging accuracy by leveraging frequency and antenna diversity.
Using the energy model introduced in Section 4.3 and additional simulations, we then compare our protocol to SurePoint and an asynchronous ranging protocol. We show that we improve the system lifetime by up to 77% and can achieve deployments of up to 39 days. With the analysis, we demonstrate that our energy model presents itself as an essential tool in identifying and quantifying the effects of protocol adjustments. We observe that idle currents must not be underestimated, as they can have a significant impact on lifetime.
As a proof-of-concept, we perform two long-term real-world deployments gathering 644,956 measurements with nodes worn over a period of up to 12 hours. Using the aggregated data, we demonstrate that our system is capable of detecting interactions with high temporal and spatial precision and achieves a stable sensing frequency with a schedule reception rate of 96.1% and a successful sample reception rate of 85.7%.

5.1 System characterization

Over the course of three hardware revisions, we iteratively enhance the original design and increase the initial transmission range from 5 meters to 64 meters, which is close to the maximum supported distance of the DW1000 for the chosen operational mode. The BLE radio can achieve up to 28 meters of reliable device-to-device communication and hence guarantees room-level discovery. As its range is shorter than UWB, BLE discovery inherently averts an accidental activation of the UWB channel while latter is out of range.

5.1.1 Distance estimation

Ranging accuracy

To estimate the accuracy and the precision of the system, we gather data inside a busy office space with common obstacles such as tables and monitors as well as multiple moving people which temporarily obstruct line-of-sight and influence multipath. Recall that SurePoint [23] collects 27 range estimates and then takes the 12th percentile of the range measurements as an estimate for the true distance. This bias is based on a channel model by Molisch [54] and has been verified in real-world experiments for individual physical channels [23]. The intuitive reason for this shift can be found in the nature of time-of-flight measurements: the earliest a signal might possibly arrive at the receiver is by crossing the line-of-sight (LOS) path at the speed of light. However, as the direct path might be obstructed and hence suffers from strong attenuation, a multipath element with a delayed arrival time could be mistaken as the LOS component and will result in a distance which overestimates the true range. Hence, the measurement distribution is usually shifted towards being too long when observing environments with heavy multipath. We will encounter this common effect later on in Section 5.3.

However, the model by Molisch assumes that the overall measurements resemble a single Gaussian distribution. As we are aggregating measurements over multiple different physical channels which vary in antenna and frequency and hence superimpose *multiple* Gaussian curves, the theoretical channel model might no

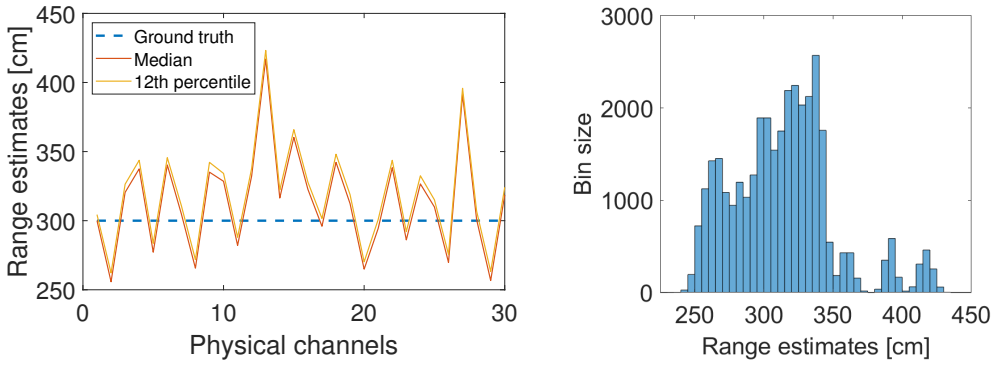

Figure 5.1: For each individual physical channel, the distributions are tight, as the median and the 12th percentile are closely grouped. However, when analysing the data across multiple physical channels, it becomes apparent that the range distribution cannot be extrapolated from a single channel to multiple ones, as the mean of the Gaussians are shifted based on frequency and antenna used.

longer apply and might require a re-evaluation of the optimal processing method. This is in fact indicated by our own experiments, as we do observe *shifted* Gaussian distributions for each physical channel. As shown by an example with 34,680 measurements over a distance of 3 meters in Figure 5.1, the median value, equivalent to the 50th percentile, and the 12th percentile are closely grouped for all individual channels and therefore do not suggest any strong multipath effect, but suggests that the channels have a tight distribution. As shown in the histogram of all ranges, the aggregation of *multiple* channels does however result in a superposition of Gaussian distributions with differing means and deviations which distorts the analysis. Therefore, the channel assumption by Molisch [54] is no longer valid for ranges aggregated over different physical channels.

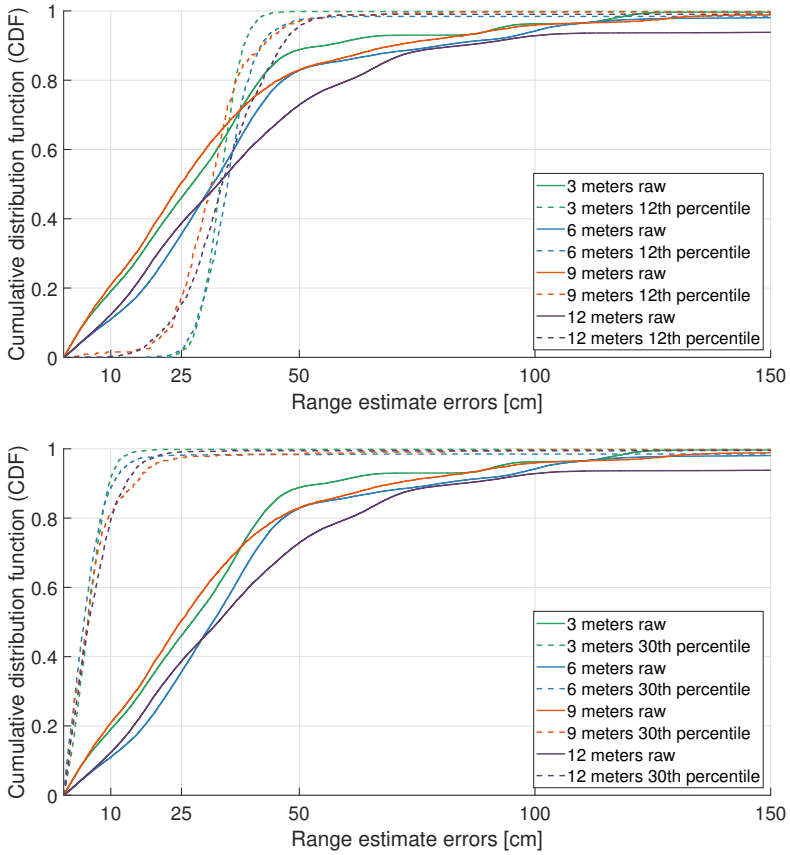

Figure 5.2: Over all distances, the average error remains below 35 cm with a 90th percentile error below 50 cm. By adjusting the processing from taking the 12th percentile (top) to taking the 30th percentile (bottom), we can improve the accuracy of the system and achieve a 90th percentile error of 15 cm.

This conclusion is also supported by our evaluation of the ranging accuracy over multiple distances. Recognizing that we are most often under-estimating the true distance, we conclude that the aggregated estimates are in fact not optimized by

choosing the 12th percentile. We could clearly improve the quality of the range estimates by adjusting the aggregation to use the 30th percentile, visualized in Figure 5.2. This improves the average 90th percentile error by 71.5% and achieves a 99th percentile error of only 14.8 cm over 3 m. It is important to note that this change is not simply adjusted for a single measurement series, but gives optimal results for all four tested ranges.

Diversity gain

We have seen in the previous section that the distribution of the measurements is non-homogeneous and heavily affected by the specific physical channel. As shown in the box plot in Figure 5.3, the individual physical channels are tightly grouped and highly precise; however, the *accuracy* varies strongly and can contain significant outliers such as the ones seen in channels 13 and 27. If one would have chosen such a physical channel by mischance, measurements would have shown centimeter precision while being off by more than one meter. Only dynamic system diversity ensures that the ranging estimate error is reduced for *any* potential environment. Our experiments show that the distributions vary heavily despite only slight alterations in the environment (see Figure 5.5). Hence, our application of diversity ranging is crucial to obtain applicable data for interaction tracking in changing surroundings.

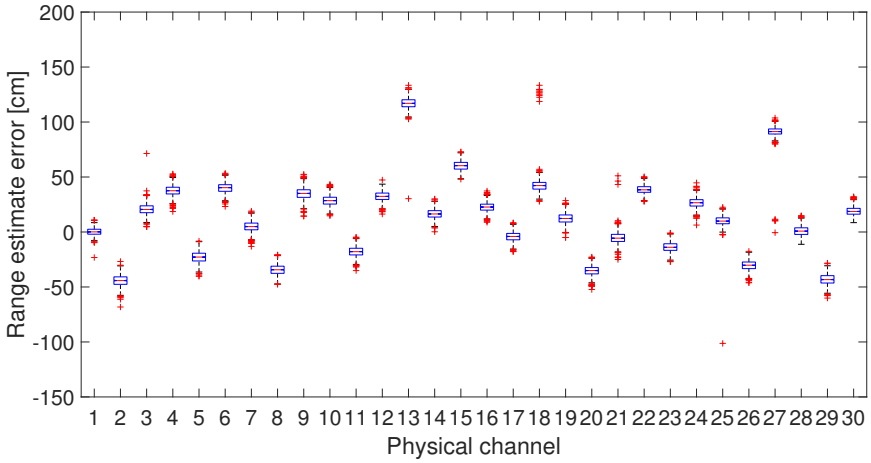

Figure 5.3: We can see the error distribution for 34,680 range estimations over a distance of 3 meters. Any given combination of antenna and frequency yields in precise but potentially inaccurate results.

To better understand the effect of diversity on accuracy and precision, we analyse frequency and antenna variation separately. Figure 5.4 shows how varying antennas while fixing the frequency only has a limited effect on accuracy; while

frequency bands 1 and 3 always have a positive error bias, frequency band 2 demonstrates a primarily negative deviation. On the other hand, antenna variations deviate both negatively and positively. Therefore, it is crucial for accuracy gains to include frequency diversity to cancel out biases. This insight can also be seen in Figure 5.11, where we show quantitatively that additional frequency bands can increase accuracy by up to 66% and further explore the trade-off between diversity and lifetime.

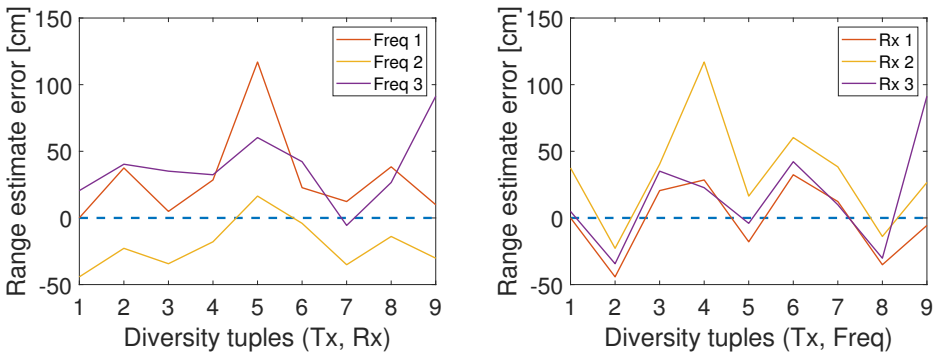

Figure 5.4: Using an inappropriate frequency channel for the given environment can result in significant accuracy variation which cannot be detected without switching the frequency bands (left). Changing an antenna does however not significantly affect the resulting accuracy (right).

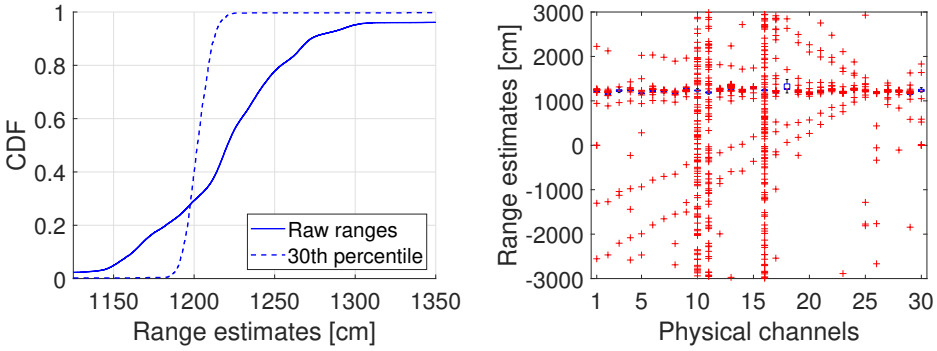

Figure 5.5: As seen on the left, measurements over 12 meters result in a significant percentage of estimates that are substantially off. We suspect that the errors are due to severe destructive interference, as a wall at a distance of only 30 cm reflects the signal. Inspecting the different physical channels, we can see that only channels **10**, **11** and **16** show significant errors. Those physical channels vary both in frequency and antenna selection. Therefore, the measurements show that one needs to make use of the full diversity in order to compensate for such isolated behaviour and cannot simply restrict oneself to either frequency or antenna diversity.

5.1.2 Physical channel stability

Reception quality

While accuracy and precision are important metrics, reliable reception is a prerequirement for successful ranging. Table 5.1 shows the packet reception rate over multiple distances. While most distances do not show significant losses, only half of the packets arrive at 9 meters. In an attempt to improve this reception, we also test out longer preamble settings as they increase the effective range at which packets can be received [38]. These raise the sent time and equivalently the energy consumption by a factor of 4 (260 µs instead of 65 µs) and result in virtually the same packet reception (49.6% compared to 49.3%). This experiment verifies that the decreased reception is not due to an insufficient preamble length, but because of unsuccessful reception of the packet's data portion [92]. Because we do not observe a decrease of similar scale at farther ranges, we presume that the node experiences local destructive interference on the fixed physical channel used for scheduling and that the distance itself is not directly related to the loss. To mitigate this issue, we discuss further suggestions regarding adaptations to the physical parameters in Section 6.2.

Distance	Expected packets	Received packets	Packet reception [%]
3 m	1230	1156	94.0%
6 m	1350	1258	93.2%
9 m	1170	576	49.3%
12 m	1350	1241	91.9%

Table 5.1: Packet reception is influenced both by distance and the given surroundings. The anomalies at 9 meters are replicable with equivalent results.

We suspect that the reason for the overall downward trend in connectivity over range is primarily due to a loss of the longer scheduling and response packets. Due to the way FCC regulations [36] are implemented, short packets with a transmit time below 1 ms can boost their transmit power beyond the nominal maximal values while still conforming to the maximally permitted energy per millisecond. Longer packets therefore linearly decrease in transmit power and hence also suffer from decreased range.
As a solution to fix such issues with the response packet, we propose to leverage our local storage to record the time stamps of the received polling packets directly at the responders and only respond with a short message. This ensures

that the response achieves equivalent range as normal polling messages while still retaining the rest of the protocol properties. Local logging further improves the protocol's scalability as it completely eliminates limitations due to maximal packet length of UWB and truly renders the system $O(n)$ for any number of members inside the network. This gain comes at the cost of additional complexity during analysis, as the range computation must occur retrospectively. It would additionally prevent real-time measurements, as local storage must first be merged before the timestamps can be converted into ranging estimates.

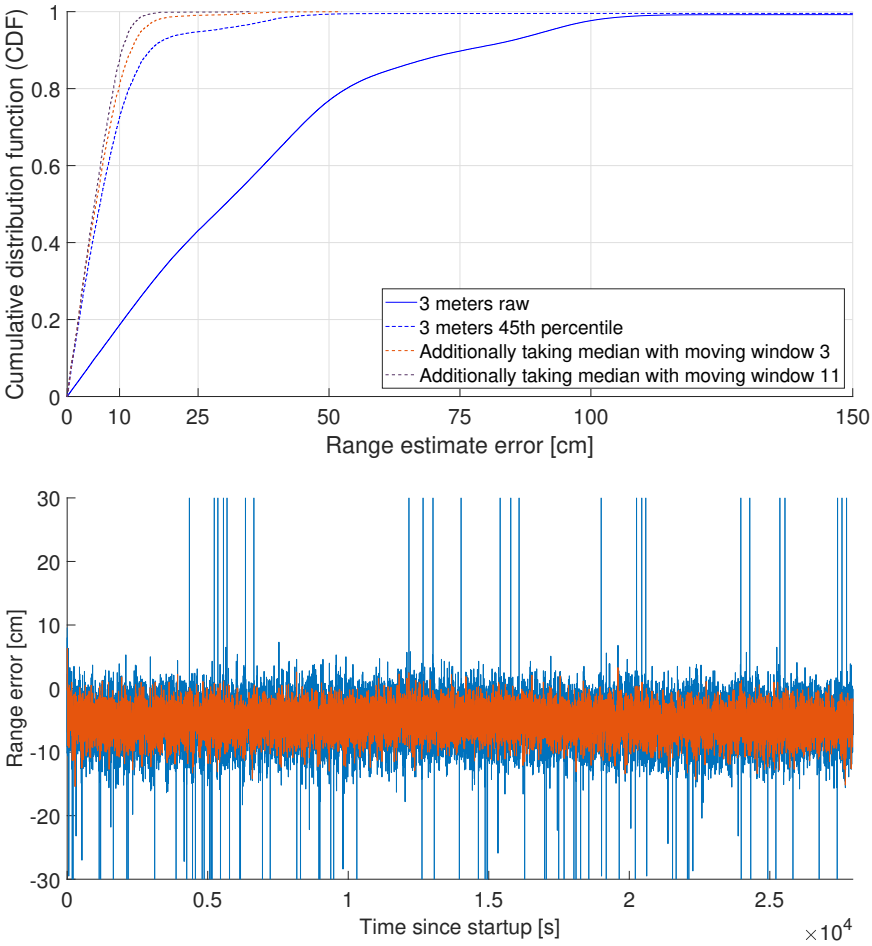

Figure 5.6: To demonstrate that additional temporal filtering can eliminate outliers effectively, we use a median filter with width 3 and 11. This achieves a reduction of the 99th percentile error by 78% to merely 15 cm with a confidence interval of less than 10 cm. The bottom plot shows that even for 840,600 measurements over more than 7 hours (blue), a simple median filter with a width of 3 (orange) can effectively filter out all outliers and achieve a maximum error of 12.1 cm.

Temporal variation

To verify that the 30 one-way rangings are stable across time and do not depend on the channel ordering, we also investigate clock drift stability. We observe in Figure 5.5 that single measurements (6 out of 37,230) have an error in clock drift estimation, visible by the linearly increasing / decreasing estimates. While such occurrences are rare, they could be trivially filtered out if the variance between samples is sufficiently high (more than 20 m for the given examples).

As a demonstration that further post-processing such as aggregation over time can improve the readability and simplify analysis, we show in Figure 5.6 that a simple median filter of width 3 can completely eradicate outliers and reduce the 99th percentile error by 78%. This illustrates that outliers are not temporally correlated and shows that accurate measurements can be gathered anytime throughout the experiment. For unknown reasons which might hint at issues with the DW1000, we observe an oscillating pattern in our measurements, shown in Figure 5.7. While its amplitude is limited to 10 cm, such a behaviour could be observed with varying periods between 200-400 seconds.

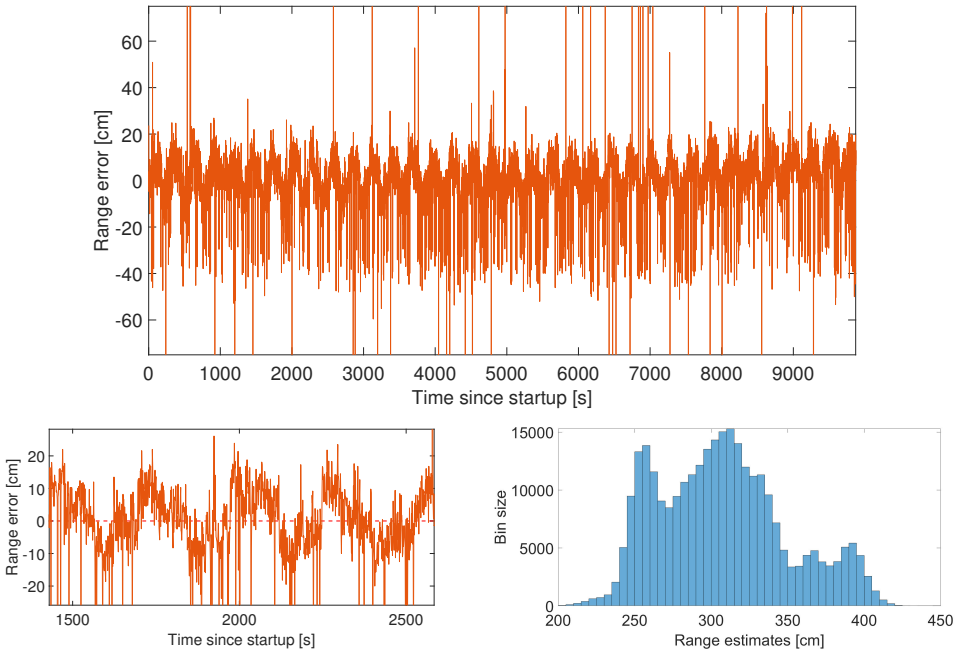

Figure 5.7: We take 300,000 measurements over a time span of three hours at 3 m to investigate how range measurements are distributed over time. As can be seen, large-scale errors do not correlate with one another and are randomly distributed (top). However, we can observe a periodic, sinusoidal pattern with amplitude of 10 cm and an approximate period of 300 s (left). This is reflected in the bimodal distribution of the ranges in the histogram (right).

5.2 Protocol

While the previous section focused on the ranging performance, a careful analysis of the protocol and its tunable parameter space is essential to identify the impact that a customizable scheme can have on the given application. To demonstrate its adaptability, we leverage our detailed energy model as well as additional simulations to investigate TotTernary's design space and evaluate it according to the MAC protocol properties defined by Demirkol et al. [70]:

- **Energy efficiency** and its effect on the network lifetime
- **Adaptability** to a change in parameters and communication patterns
- **Scalability** to a large number of nodes with graceful accommodation
- **Latency** of joining the network and transmitting packets
- **Reliability** in network communication

Where applicable, we compare our performance with SurePoint [23] as well as asynchronous variants of the ranging diversity protocol to highlight the effect that different approaches can have on the performance.

5.2.1 Energy efficiency and adaptability

As one of the central trade-offs in protocol design, a scheme usually aims at maximizing lifetime while satisfying the restrictions of parameters such as nodes per network and update frequency. Mathematically, this can be regarded as a globally dependant variable (the lifetime) which is influenced by several independent variables (batter size, network size, number of diversity ranges, discovery latency, sampling rate etc.).

As a first insight from our energy model, we observe in Figure 5.8 that while the active part of the round, during which nodes perform ranging and transmit scheduling messages, naturally consumes a significant part of the available energy, the long passive phase *between* them incurs the highest energy costs due to high sleep currents of 11.1 mA in our current implementation. The chips support theoretical sleep currents which are three orders of magnitude lower at 4.1 µA which could be achieved by carefully disabling peripherals and leveraging our host controller to wake up the module on demand. While limited engineering resources prevented us from fully leveraging those gains, we were able to verify a sleep current of 2.1 mA on the current hardware. To show how such an optimization could improve the lifetime of the system, we provide an option in our energy model to flexibly investigate both the current implementation as well as the potential, experimentally verified gains, visualized in Figure 5.8.

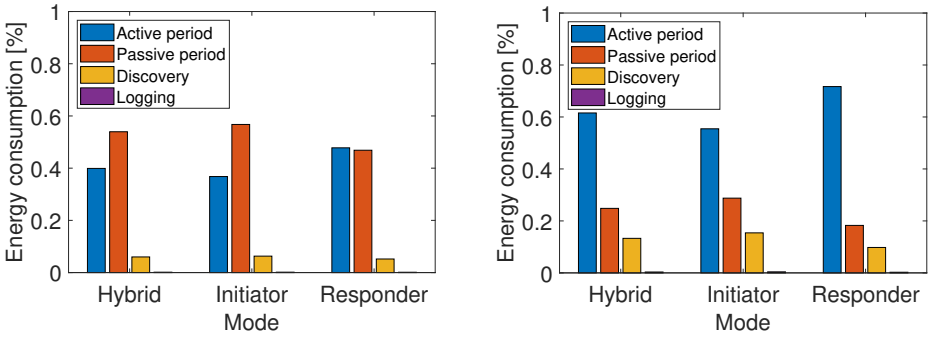

Figure 5.8: Idle currents can take up a large part of the energy budget. Despite ranging at 1 Hz between two hybrids, the majority of the power is consumed between rounds by sleep current. With an additional engineering effort, this problem could be mitigated.

To investigate the influence of the network size on the platform lifetime, we simulate various numbers of nodes while using full diversity and keeping the update frequency constant at 0.5 Hz; this guarantees that SurePoint is still capable of scheduling all nodes in a single round throughout the comparison. Figure 5.9 shows networks with an increasing number of hybrid nodes both for the implemented protocol as well as for the optimized variant. As expected, TotTernary strictly outperforms SurePoint for all scenarios and modes thanks to its elimination of the contention-based responses and the aggregated response packets. Two hybrids can achieve a lifetime of up to 6.12 days of constant ranging with full diversity and almost three weeks using the optimized parameters. Even for larger networks of 6 nodes, TotTernary provides a full work week (4.7 days) of data with a complete connection graph between all members; our optimized variant reaches a similar lifetime for an entire school class of 20 children.

We observe in Figure 5.9 that the influence of fragmented response packet per responder, due to the maximal packet length, plays a much stronger role than anticipated. Using the model, we see that an initiator mostly idles and is therefore strongly affected when its active time increases from 85 to 115 ms, an increase by 35.3% due to a single additional node in the network. At more than 15 nodes per network, initiators start to exceed the active time of SurePoint with its fixed response slot per initiator. While the additional reception does increase energy costs, the node also obtains significantly more valid ranges as TotTernary uses TDMA to guarantee successful reception, whereas SurePoint will observe strong collisions in its 10 slots when 20 nodes try to access the channel. Even when gathering additional ranges, our protocol still outperforms SurePoint for initiators because the overall round length is shorter as there is only a single such response slot.

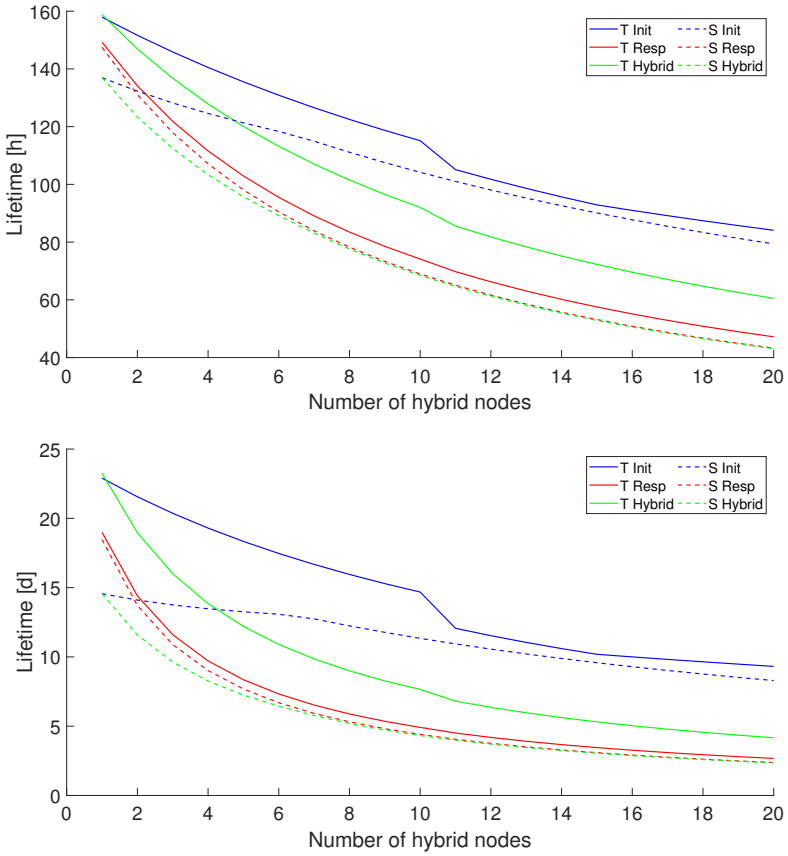

Figure 5.9: We can see that our protocol outperforms SurePoint for any number of nodes, especially for the case of hybrids where we can leverage the ranging pyramid to reduce consumption. With optimized idle currents (bottom), we achieve a 63.7% higher lifetime for two nodes and can sustain even 77.0% longer for a network of 10 hybrids. However, we can also see that the influence of long responses, which need to be fragmented into multiple packets (visible between 10 and 11 nodes), considerably influences the lifetime of initiators.

The influence of fragmented packets for TotTernary is especially pronounced in the optimized case. While initiators first perform more than 57% worse, they slowly close the gap with an increased number of nodes to create the impression that SurePoint can scale almost equally well up to a high number of responses. Initially, SurePoint's efficiency is poor as it pre-allocates a fixed, long contention window for each initiator which is then only rarely used. Counter-intuitively though, it is not only because of a better slot utilization that SurePoint is becoming more competitive, but because the effective energy invested for receiving the responses decreases with higher utilization. While idle listening requires 151 mA, active packet reception drops to mere 63 mA as seen in Figure 4.6. This

	SurePoint	**SurePoint+**	**TotTernary**	**TotTernary+**
Schedule	1768	1768	1768	1768
Rx Polls	67770	51138	67770	51138
Tx Response	4018	**4018**	189	**151**
Idle	10018	**4814**	1964	**754**
Active phase	83575	61738	71691	53811
Passive phase	1998	1890	6993	2873
Sum	85573	63628	78684	56684

Table 5.2: Energy consumption of a responder in a round of a network with 18 hybrids and an update rate of 0.5 Hz, given in µC. We compare both the implemented variant and the ones with optimized parameters and see that TotTernary improves the lifetime of a responder by 8.7% and 12.3% respectively. While the large savings during the responses reduce the power draw by 84.6% and 89.8%, an increased time spent waiting for the next round and the dominating costs of listening for polls limit the global effect of those savings.

anomaly reduces the overhead of the static contention window and improves the efficiency of initiators; as our protocol completely eliminates idle listening during ranging, we can guarantee a constant benefit of this effect.
Additionally, while SurePoint's energy consumption only scales slightly worse than our protocol, a high percentage of the replies will not be successfully received due to the contention-based delivery. This entails retransmissions for responders as well as a maximal theoretical reception of 30 responses if every single slot could be utilized successfully. In comparison, our protocol supports up to 396 responses every second deterministically without requiring retransmissions or acknowledgements.

We can also observe the benefit of the reduced response slots for responders, as they boost the lifetime gain from 1.2% for 1 node to 9.1% for 20 nodes and even 12.4% for the optimized variant. However, this is still significantly less than what we were intuitively expecting. Again, leveraging our energy model allows us to dissect the consumption in Table 5.2 into the individual parts of a round. Hence, we can see that while we do have high gains up to 89.8% for the responses, they are dwarfed by the high reception expense and additionally diminished due to added, non-negligible idle time.

For hybrids, we can primarily make use of the ranging pyramid to achieve a substantial performance improvement of 63.7% to 77.0% for all network sizes. Again, large idle currents of 11.1 mA have a significant influence and prevent the system from achieving the full potential gains of the concept.

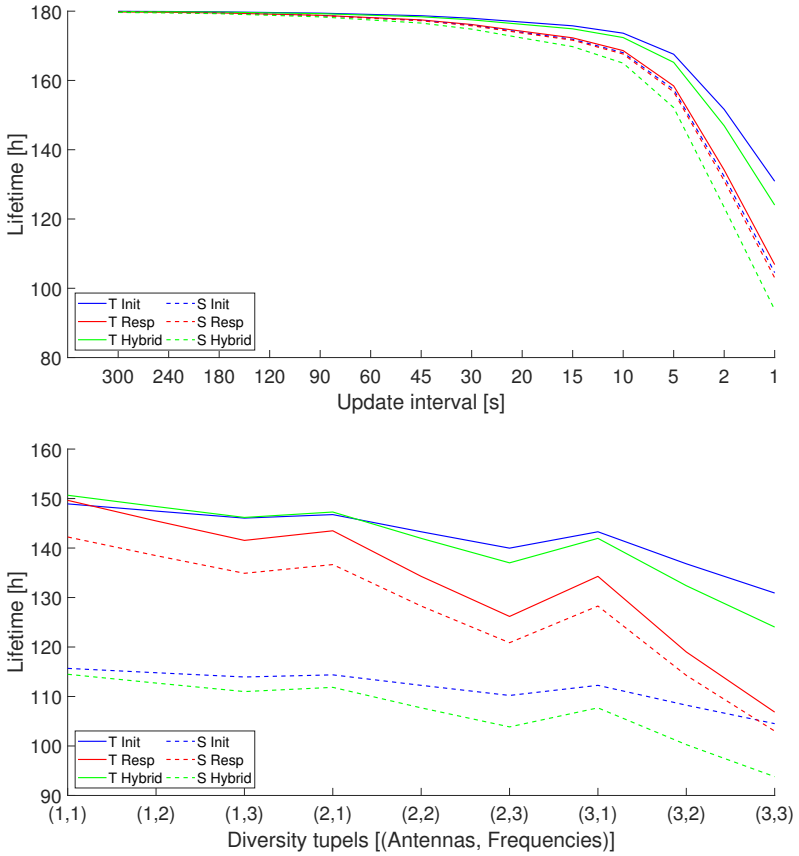

Figure 5.10: Increasing the update interval can result in large energy savings and increase lifetime by 45% for our realized system, demonstrated at the top for two hybrids, and up to 316% for optimized parameters (not shown). On the bottom, we show that the degree of diversity can be used as an effective trade-off tool. Especially for responders, the reduced listening time due to less polling packets can increase lifetime by up to 40.1%. Due to the large, contention-based response period overhead, SurePoint cannot leverage this gain as effectively.

Next, we analyse in Figure 5.10 how a network of two hybrids performs under varying sampling frequencies. While we see that a decrease can initially have a significant impact, ranging less frequently than every 15 seconds does not result in relevant lifetime gains as the idle current dominates the power budget and limits the platform to 180 hours of operation. With optimized parameters and an idle current of 2.1 mA, we could achieve system lifetimes of over a month with update rates up to 6 times per minute. Our results clearly suggest that additional engineering effort into a reduction of the idle currents could diversify this design axis and enable a more fine-grained parameter choice.

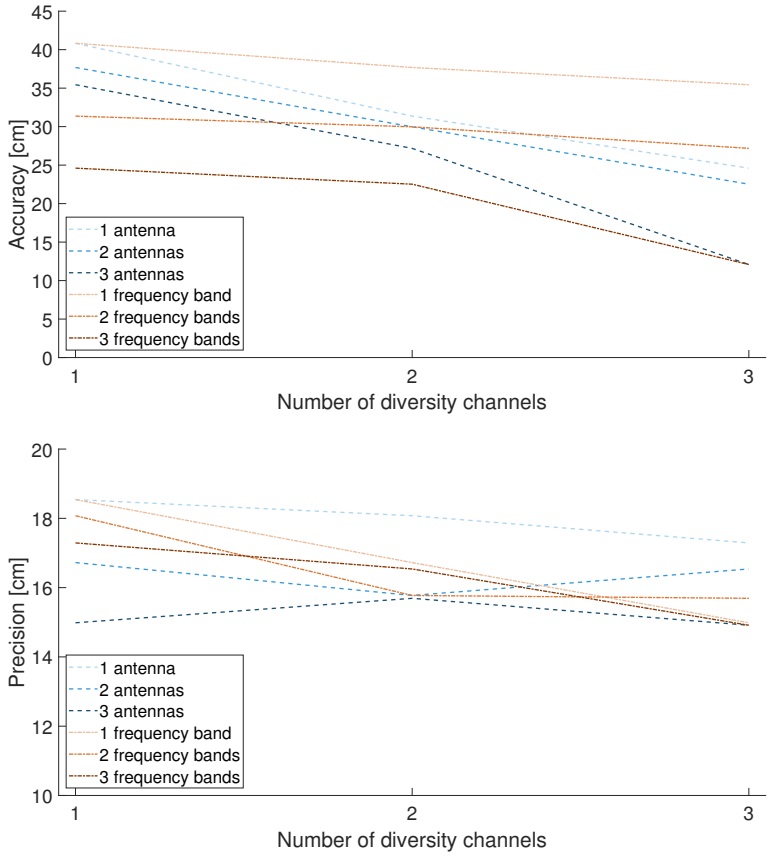

Figure 5.11: By raising diversity, we increase both the accuracy and the precision of the system drastically. While the accuracy is improved by up to 66% through an increased use of frequency bands, precision (defined as the difference between 90th and 10th ranging percentile) can be boosted by more than 19% with additional antenna diversity.

As the degree of diversity has a direct impact on the duration of the round, we further show in Figure 5.10 that reducing both frequency and antenna diversity can improve lifetime by up to 40% for a network of two hybrids ranging with 1 Hz. The effect on both accuracy and precision of the ranging results, as illustrated in Figure 5.11 based on extensive empirical measurements, shows a gracefully reducing quality of measurements with decreased diversity. It is crucial to notice that system designers must consider the correct trade-off for their application; while accuracy heavily benefits from additional frequency bands which correct biases (see Figure 5.4), precision is primarily influenced by additional antennas as the results from the diversity ranging are more strongly correlated.

5.2.2 Scalability and latency

We have shown in the previous subsection that owing to the contention-free responses, TotTernary can support more than 13 times as many responses per second compared to the theoretical maximum of SurePoint, in which every contention slot is successfully used. We achieve this scalability through the aggregation of responses and only send the packet once according to a known schedule. Switching all packets exclusively to broadcasts allows us to improve the message complexity effectively to $O(n+m)$ with n initiators and m responders. All previous diversity ranging solutions such as the naive implementation presented in PolyPoint [35] as well as SurePoint [23] required $O(n*m)$ messages due to the unicast responses. The impact of this improvement on the number of messages is visualized in Figure 5.12. Due to a maximal packet length of 1023 bytes, the practical implementation of our protocol still requires $O(n+m*\lceil n*96/1023\rceil)$; however, the law of small numbers and the high data rate of UWB renders the additional factor negligible for reasonable network sizes.

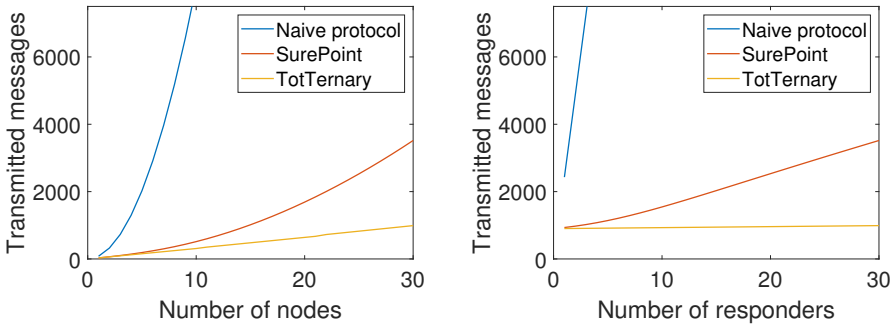

Figure 5.12: Our protocol drastically reduces the number of messages compared to both a naive two-way ranging approach using unicast packets, described in PolyPoint [35], as well as the protocol proposed in SurePoint [23]. For a network of 30 hybrids, we decrease complexity by factors of 73.6 x and 3.6 x respectively (left). This difference is even more pronounced when only adding responders (right); with a fixed number of 30 initiators, our protocol only requires 9.63% more packets with 30 responders compared to a single one, whereas SurePoint increases by 377.0%.

A network supporting mobile nodes must also be able to handle the sudden appearance of many nodes simultaneously [70]. Figure 5.13 demonstrates the effectiveness of our adaptive contention window. Despite the fact that we disregard the much heavier contention in a single slot for SurePoint, our protocol outperforms a static contention window in maximal join latency by a factor of 8.3 for 50 nodes and 14.3 for 100 nodes. Furthermore, the join latency differs much less between new nodes due to the aggressive contention window increase, seen by the similarity of average and maximal latency for TotTernary.

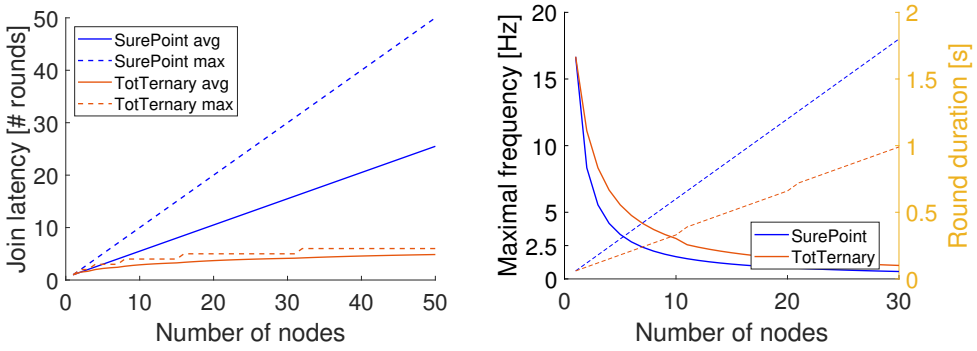

Figure 5.13: The left figure illustrates that our protocol handles simultaneous requests smoothly and allows up to 50 instantaneously appearing nodes to be scheduled within maximally 6 rounds, whereas SurePoint requires 8.3 x longer. On the right, we demonstrate that with a constant number of 15 responders (towards the upper limit SurePoint can handle reliably), our protocol strictly outperforms SurePoint and achieves 83% more measurements for 10 tags, providing update rates of over 1 Hz for up to 30 nodes.

While energy efficiency often is the dependent variable, there are scenarios where high-resolution data is the primary goal and rangings should occur as often as possible. Figure 5.13 shows that TotTernary supports up to 3 Hz update rates from 15 responders for 10 tags in parallel, enabling high-frequency applications such as the tracking of a basketball team while playing. For a single tag on e.g. a quadrocopter [35], we achieve up to 16 Hz and can support control loops for robots which can directly integrate our ranging data over BLE or micro USB.

5.2.3 Reliability and scheduling

After having shown that the protocol provides an energy-efficient and scalable ranging scheme, we investigate how the protocol performs in terms of reliability. The inclusion of constructive interference for flooding has been shown to enable highly reliably communication [49, 55]. We test our protocol with a real-world deployment and successfully receive 96.1% of the flooded packets over various distances up to 15 meters and in heavy multipath environments. While the flooding based on Glossy repeatedly transmits the same packet to propagate it throughout the network and therefore increases the likelihood of successful interception, our ranging packets are only transmitted once. Hence, we observe a slightly lower packet reception rate of 85.7% successful measurements, which implicitly also includes a successful schedule reception. We further observe that nodes in the center of the room experience higher reception probabilities as they have a shorter communication distance to neighbours (see Figure 5.14).

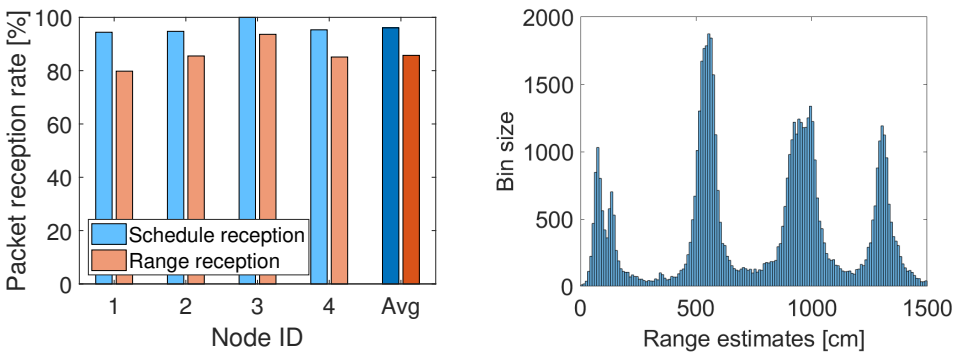

Figure 5.14: Despite high node mobility and many interfering objects in the deployed environment, packet reception is high with an average of 96.1% of schedules received and 85.7% of two-way rangings accomplished (left). As demonstrated on the right graph, the location of the cohort when sitting at their corresponding desks is clearly visible and differs considerably (right). Nodes which are further apart (e.g. nodes 1 and 4) experience decreased packet reception rates compared to nodes in the center of the room (nodes 2 and 3).

During our system design in Section 3.4, we opted for a schedule-based scheme as it increases reliability, improves energy efficiency for our target duty cycles, and guarantees performance despite high node densities. Figure 5.15 compares TotTernary both to SurePoint as well as an asynchronous variant of the diversity ranging scheme for a fixed update rate of 1 Hz. We give the asynchronous protocol the assumption that it knows the absolute number of nodes and can hence optimize its sampling frequency to compensate for collisions; however, even if we grant the asynchronous system wall-powered anchors which would constantly listen for requests and therefore put it at a significant advantage, its performance collapses for networks of more than 4 nodes due to collisions. We further show using our energy model that by reducing the scheduling packet length compared to SurePoint by up to 60%, the scheduling overhead is less than 3.2% for all nodes for networks larger than 20 and maximally reaches 6.2% for a minimal network. Additionally, nodes can opt out of the contention-access period at the end of the round and hence significantly reduce protocol overhead.

5.3 Real-world deployment

As a final verification step of our platform, we perform two real-world experiments where we deploy 5 nodes over 10 and 12 hours to a cohort of five students working in the same open-plan office. Over a time period of two working days, we gather 644,956 distance measurements out of over 19 million time-of-flight

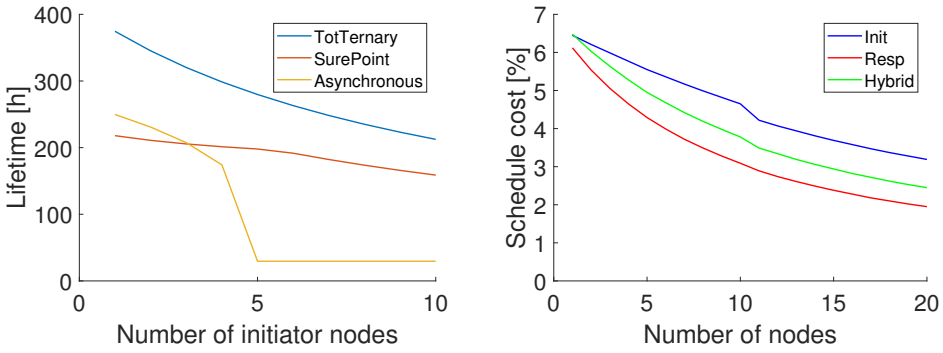

Figure 5.15: When comparing our scheduled approach with an asynchronous implementation of our diversity ranging protocol, we see that latter cannot support a network size larger than six due to collisions (left). The scheduling overhead of transmitting and receiving a schedule is small and only requires less than 2% of the energy budget for responders in networks of 20 nodes (right).

rangings. During this time, none of the nodes ran out of battery or experienced a hardware failure. Due to a known bug of the UWB radio occurring every few hours on the master node which prevents further reception (that could not be fixed with either an automatic software nor a hardware reset), we occasionally have to manually power cycle the master upon detecting the failure case over the status LED. Four subjects opt for a lanyard to wear the nodes, while one person chooses to wear it in a chest pocket (see Figure 5.16).

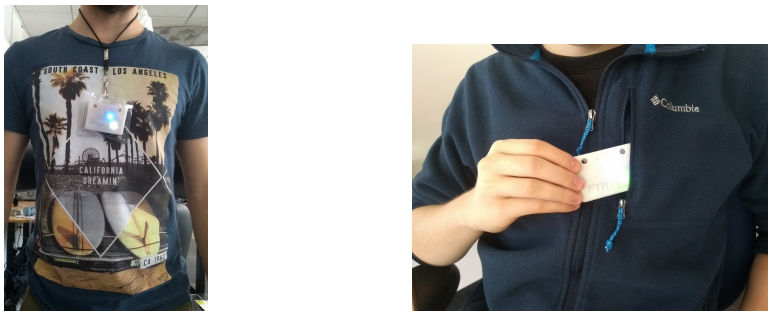

Figure 5.16: To find a suitable carrying mechanism, we test different approaches such as storing the node in the key pouch of trousers and using a pair of magnets to attach it to any kind of clothing. The two most popular methods are the use of a lanyard which is applicable to any kind of clothing (left) and the use of breast pockets found in shirts and jackets (right). Clearly visible in the picture on the left are the two status LEDs of the node: the upper one (blue) signals active neighbour discovery, while the lower one (green) indicates that the node is connected to an UWB network and performing rangings.

5.3.1 Interaction tracking

With a total of 81 interactions detected throughout a 10-hour deployment, we demonstrate that the system is capable of automated interaction tracking and analysis over a prolonged deployment scenario. Cohorts can be observed directly in their day-to-day routine without additional effort by wearing the device on a lanyard. Researchers can avoid the manual recording of observations and simply focus on the logistical aspects of the study. Thanks to the timestamping feature of the platform, which is configurable over the user interface, specific interactions and known external events can be used to contextualize measurements and mapped to the data, which allows for an in-depth analysis of individual interactions.

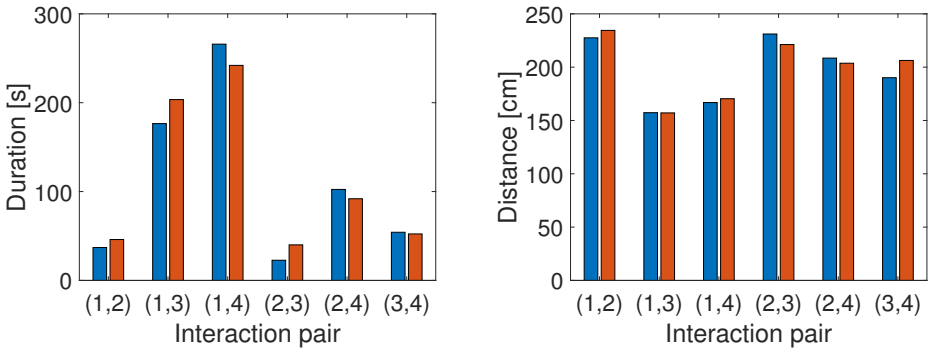

Figure 5.17: We can observe that the average interaction duration varies strongly between subjects (left). Furthermore, interactions can be clearly distinguished into sitting (1.5 m) and standing (2 m) (right). The fidelity of the data is cross-validated by comparing the detected interactions from both sides (blue and orange): we see that the average difference in duration is merely 14 seconds (left) and on average only varies by 69 mm in distance (right).

Figure 5.17 shows that the average interaction is rather short and lasts 111 seconds at a distance of 198 cm. Leveraging that our data allows us to distinguish interactions between different nodes, we observe that node 1 interacts much longer with node 3 and 4 than any other pair of nodes and that their interaction distance is significantly lower. A comparison with a manual daily log confirms that subject 1 and subject 4 were collaborating on a paper during the experiment and hence have been sitting next to each other over longer periods of time. Similarly, subject 1 and 3 were sitting next to each other several times during gatherings and hence also exhibit longer interactions at a shorter distance. All other meetings occurred casually throughout the day while at least one person was standing, reflected by the increased distance and decreased duration.

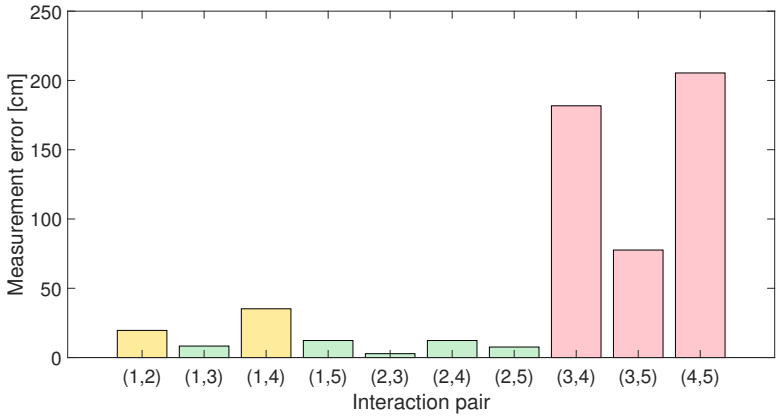

Figure 5.18: Comparing ranges to ground-truth measurements with a laser range finder, we find that while only lightly obstructed paths (green) result in highly accurate distance estimates, performance seems to decrease for node pairs which have multiple obstacles such as monitors and desks in their way (yellow). Although even under such conditions, the ranging error remains below 40 cm, we can see that results can significantly deviate from the ground truth for heavy NLOS paths where the direct path is blocked my multiple desks and the bodies of the members themselves (red).

To estimate the accuracy of our system in a real deployment with heavy non-line-of-sight influence, we compare a 20 minute steady-state period during the deployment where all members remained at their corresponding desks to known ground truth values in Figure 5.18. We observe a high general accuracy of 14 cm in most cases, but occasional measurements which are significantly influenced by the environmental conditions. This is primarily due to non-line-of-sight paths which can positively bias the estimates as pointed out in Section 5.1; with a single exception, we over-estimate the distances between nodes as being longer than they actually are. We have purposefully chosen one of the subjects because of its separated seating and see that it experiences heavily restricted line-of-sight to the majority of the cohort. Its corresponding node 4 is hence also represented in the two strongest deviations from the ground truth measurements due to the highly attenuated direct path and unobstructed reflections from the ceiling which experience a longer time-of-flight. Such NLOS influences are a known problem for UWB ranging; existing solutions [34, 93] have shown success in automatically detecting and correcting such errors and could be applied to this system in addition to the diversity ranging to mitigate this issue.

5.3.2 Activity observation

As a demonstration of a real-world application in different surroundings, we track 5 subjects over a period of 12 hours in both various indoors and outdoors settings. Figure 5.19 shows the interactions of a single node with the four other devices throughout the day. As the cohort is working in a shared office space, we can clearly distinguish the individual distances to the other subjects while each of them are working on their assigned desk, also illustrated in Figure 5.14 as peaks. The concurrent ranging with multiple nodes also allows us to infer the behaviour of others; a simultaneous distance reduction to everyone suggests a common gathering, while constant measurements with the majority and singular deviations such as the spikes of node 4 indicate that only a single node moved (and suggests that it frequently interacted with node 2 at the latter's desk, as is verified by the data on those devices). Of special interest is that such data even allows coarse-grained localization based on the known stationary location of nodes such as while subjects are at their desks. An example of this can be found in the orange activity in the middle of the day which shows that node 2 was between its own desk and the desk of node 4 for 10 minutes, which correlates with the position of the coffee machine.

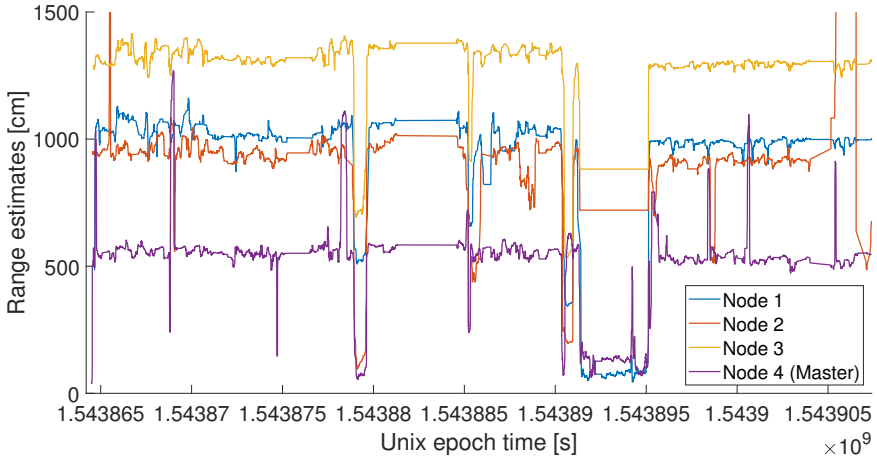

Figure 5.19: Temporal filtering with a median filter allows a clear assessment of the activity even over long periods of time. Single points in time can be attributed to known external events: as an example, at 1543868800 (12:26 local time) node 4 checks correct operation of all devices by walking to each one of them. The long interaction with nodes 2 and 4 around 1543879200 (15:20) shows a discussion about the previous weekend between co-workers. From 1543891350 (18:42) until 1543895150 (19:45), three members of the group went out for dinner together; one can even detect the short separation during that period where a single member went to get extra ice cream as a spike.

5.4 Long-term study

The initial input for this thesis came from researchers at Vanderbilt University [25], showing a demand for infrastructure-free mobile tracking applications with high accuracy. Their focus on observing the interaction between toddlers and caretakers considers a deployment horizon of multiple years to investigate the influence of interactions onto the children's development. While the study is still in its infancy and no data is yet available, we have been collaborating closely throughout this thesis and are actively helping the researchers with testing and deploying TotTernary. At the time of writing, the first boards are undergoing preliminary checks and we are confident that the project will be scaled up in the near future.

5.5 Design modularity

During our development process, we could show that the principle of modular hardware design to simplify the design of new hardware, introduced to sensor networks among others by EPIC [64], is still valid, especially in regards to highly delicate RF layouts. We have demonstrated that even in the absence of striking design flaws, a careful board design can have a drastic influence on performance. By strictly following manufacturer guidelines (and increasing safety margins when there are none), limiting noise coupling between oscillators, PLLs and power circuitry as well as a precise analysis of a component's functionality and potential effect on others, we could increase communication range by more than 1250%. An existing module can prevent future hardware designers from running into similar pitfalls and therefore increases the reliability of products while reducing their development time.

Due to technological advances, we argue that the concept of highly specialized breakout modules for an integration into production-quality hardware, as proposed in EPIC [64] and SurePoint [23], has been superseded by newly introduced features in hardware design software. Design blocks allow a detailed specification of modules purely in software and can be directly integrated into new designs. This facilitates sharing proven designs and enables simple verification of the compatibility with other components without running the risk of incorrectly setting up signals and power connections. These trusted modules permit designers to obtain a high degree of confidence that a new design will be functional before producing it in hardware.

Chapter 6

Summary & Conclusions

We motivate the development of a range-based tracking system in Chapter 1 and investigate the state-of-the-art in Chapter 2 to derive that time-of-arrival based systems show the most promise for this application. Thereafter, we analyse the design considerations in detail in Chapter 3 and propose both hardware and software for a platform which covers a large design space of tracking applications. We detail our implementation in Chapter 4 and demonstrate that the platform achieves its goal of an infrastructure-free, scalable and highly accurate system for interaction tracking in Chapter 5. The following sections summarize our achievements during the course of this thesis and give suggestions for prospective extensions.

6.1 Findings

This thesis presents TotTernary, a tracking platform which leverages an infrastructure-free, scalable design based on ultra-wideband diversity ranging to detect interactions reliably and with high accuracy. Inspired by SurePoint [23], we incorporate frequency and antenna diversity to achieve an average 95th percentile ranging error of 14.8 cm with an average precision of 7.3 cm.

We enable infrastructure-free deployments with a dual-radio design; energy-efficient BLE neighbour discovery in BLEnd [53] avoids power-consuming UWB listening to find and setup a network, while UWB diversity ranging gathers distance estimates with high reliability. The concept of hybrids permits nodes to operate independently and eliminates the requirement for anchor nodes. Despite high node mobility and obstructed lines-of-sight, we achieve a scheduling reliability of 96% using constructively-interfering Glossy floods [55].

We demonstrate that the exceptionally high idle listening costs of UWB severely impact the energy efficiency and must be evaded by all means. To achieve this, we use a schedule-based system and assign deterministic channel access to all parties.

By aggregating messages and exclusively using broadcasts, we improve diversity ranging from quadratic to linear message complexity. A dynamic adjustment of the contention slot reduces the network join latency and enables nomadic nodes to quickly interact with a new cluster. Together with additional optimizations such as the ranging pyramid principle, we boost lifetime by up to 77% compared to the current state-of-the-art.

The addition of local permanent storage permits detailed data logging for offline post-processing and automatic interaction detection. The data is enriched by global timestamps and hence allows researchers to contextualize measurements with known external events.

We build an energy model based on detailed measurements of the implemented system to predict the influence of adaptable design parameters on the system performance. The covered application space ranges from highly active networks of 30 nodes ranging with 1 Hz update rates up to long-term observations with more than a month of estimated lifetime. We further show that diversity allows a trade-off between ranging quality and energy consumption which can increase the maximum deployment time by up to 40% and observe that a reduction in the sampling rate boosts lifetime by up to 45% for the current implementation.

Our end product is a complete platform for interaction tracking. It provides an adaptable protocol which can be fine-tuned to the specific application based on a detailed energy model, a tested hardware and software implementation which permits direct connectivity with end users over a provided user interface as well as analysis tools which support researchers to evaluate the resulting data. The platform is currently undergoing testing with social scientists and shows promising preliminary results.

6.2 Future work

We propose the usage of energy-aware ranging and conceptualize how an integration into our design can boost overall network lifetime. To verify its usability, an implementation and extensive real-world testing could evaluate the impact of such a scheme and demonstrate how different energy consumption due to varying usage could be balanced.

The current platform uses two separate chips to provide both energy-efficient discovery over BLE as well as precise ranging estimates over UWB. While those two primitives used to require separate hardware, recent developments in the specifications of established protocols allow for both accurate time-of-flight rangings using Bluetooth 5 [32] as well as IEEE 802.11 (WiFi) [94]. Hence, a single IC might be able to support both services in the future. This combination could

simplify hardware design and will lead to new protocol designs which incorporate both communication and ranging on the same physical layer. However, such technology still requires significant effort to improve reliability until it can be deployed in real-world systems [95].

To increase the effective range of the platform, one could combine short-ranged, high speed 6.8 Mbps transmissions for energy efficiency and occasionally leverage long-range 110 kbps for improved range and material penetration. Recent research indicates that only a decreased packet rate might deliver an increase in packet reception rate [92], as our system already uses the lower carrier frequencies and higher PRF. The consequential drastic increase in transmission times would, however, require adjustments to the protocol and its timing behaviour.

In some cases, our protocol experienced severe packet loss due to local destructive interference on the response channel. As we are aspiring to minimize energy consumption, we did explicitly avoid congestion and acknowledgements and only used reliable flooding for system critical transmissions such as scheduling. However, if packet loss on the responses should dominate, one could re-introduce the repeated transmission of the response on multiple frequency bands. This could still occur in a deterministic way and would only have a limited impact on the energy consumption; nodes which successfully received the packet can sleep during the subsequent repetitions and therefore trivially avoid additional energy-expensive listening.

A highly useful extension of the protocol would be the support for dynamic and heterogeneous parameters such as update rates. This could be solved by leveraging existing schedulers such as BLINK [96], which schedule nodes with provably minimal energy costs. With additional information in the signalling packet, the system could support individual requests for sampling rates and respond to the demand on-the-fly.

6.3 Conclusions

In this thesis, we show that time-of-flight based distance estimation can provide social scientists with valuable insights at unprecedented accuracy for mobile systems. Using a dual-radio architecture, we demonstrate that BLE can be employed as a low-power wake-up circuit and is valuable as a user interface to configure networks during active deployments and achieve real-time tracking. In combination with a strict scheduling of all parties, we can eliminate high idle listening costs and conduct true infrastructure-free deployments. A detailed energy model enables us to identify the strong impact of sleep currents on the system lifetime and predict the influences of protocol parameter variations.

Chapter 7

Appendix

7.1 Technical information

7.1.1 Software

The software written for this thesis is open source and can be found on Github [24]. It is primarily written in C and uses additional helper scripts in JavaScript, Python and Matlab. The code relies on external libraries for the DecaWave DW1000, the STMicroelectronics STM32F091CC and the Nordic Semiconductors nRF52840; these code bases are included as submodules and are maintained by the *Lab11* research group at the University of California, Berkeley.

7.1.2 Hardware

The design files for the hardware developed for this thesis is open source and can be found on Github [85]. We used EAGLE [67] for designing and verifying the hardware as well as to create the Gerber files for production.

Power supply

The platform runs with four different voltage levels:

- V_{USB} **(5V):** For charging, we use a USB connector supplying 5 volts, which are then dropped by a battery charge controller to 4.2 V for the battery.
- V_{BAT} **(3.7V):** Voltage supplied by the Lithium-Ion battery; supplies an low-dropout regulator (LDO) which transforms the voltage to the main voltage level of the system.

- V_{DD} **(3.3V):** Main voltage level which supplies all integrated circuits as well as the accelerometer.
- V_{LDO} **(1.8V):** Lower voltage supplied by a switching regulator for internal digital and analog circuitry of the DW1000. Using an external regulator allows for increased power efficiency compared to the internal LDO of the DW1000.

IC specifications

	UWB radio	**Radio controller**
SoC Name	DecaWave Ltd DW1000 [38]	STMicroelectronics STM32F091CCU6 [86]
Processor Type	Custom host processor	ARM 32bit Cortex M0
Max frequency	38.4 MHz	16 MHz
SRAM	-	32 kB
Flash	-	256 kB
Active energy consumption	12 mA	10.2 mA
Sleep mode consumption	1 µA	6.4 mA
Deep sleep mode consumption	50 nA	4.1 µA
Communication	Ultra Wideband (3.5 - 6.5 GHz), SPI, GPIOs	I^2C, UART, SPI, GPIOs

Table 7.1: Technical specifications of the two processors used for the **ranging module**, including the UWB radio and its radio controller.

	Host controller	**Debugger**
SoC Name	Nordic Semiconductor nRF52840 [97]	FTDI Ltd FT232RQ [88]
Processor Type	ARM 32bit Cortex M4	Serial Interface Engine (SIE)
Max frequency	32 MHz	12 MHz
SRAM	256 kB	-
Flash	1 MB	-
Active energy consumption	3.3 mA	15 mA
Sleep mode consumption	1.5 µA	70 µA
Additional features	*Bluetooth* 5, 802.15.4 floating-point unit (FPU)	Serial-to-USB

Table 7.2: Technical specifications of the two processors used on the **carrier board**, including the host processor and an optional debugging chip.

Manufacturing costs

Allocation	**Current cost**	**Projected cost**
PCB	$47	$30
Components	$94	$40
Assembly	$69	$10
Fees & transport	$19	$5
Board cost	$229	$85
Battery	$12	$11
Case	$1	$1
Total	$242	$97

Table 7.3: With the current small batches of less than 20 boards, a complete node costs $242 dollars to manufacture. As component costs drop significantly with price breaks at 100 and 250 pieces and assembly is primarily a one-time, fixed charge, we estimate that the board could be manufactured for less than $100 at scales of a few hundred nodes.

7.2 Hardware evolution

7.2.1 Revision C

As a first trial run, we hand assembled five devices for verification purposes and testing before ordering a larger badge. Due to issues with old solder paste which bloated when heated up, only three devices could undergo testing, of which only a single one had both ICs as well as the UWB radio working. Regrettably, even this board had issues with the I^2C connection between the MCUs and could therefore not be used for testing the full functionality. Nevertheless, we could verify that all of the circuits, including the SD card and the accelerometer, were working and were able to identify an issue with the routing of the reset pin on the host controller.
During assembly, we quickly realised that the solder mask was ill-suited for the UWB radio, the DecaWave DW1000, as it involved a very dense pin layout and therefore easily bridged on multiple occasions. Furthermore, hand assembly proved to be demanding for the ball-grid array (BGA) of the Nordic nRF52840 and resulted in a yield of only a single device which could use the specific chip.

7.2.2 Revision D

Improvements for the second generation hardware included adaptations of the silk screen to fix switched up labels, an increased number of testpoints to expose all buses and voltage levels improve debugging as well as an adjustment of the solder mask to prevent bridging during the manufacturing process and to facilitate hand assembly. In addition, we shifted the balun further away from its filtering capacitors for easier hand assembly and fixed a hardware bug of the SD card causing automatic resets of the controller by adding a dedicated hardware reset line and switching the connected pins.
As an extra feature, we added a battery voltage probing pin to assess the charge level on the embedded device directly from the host controller; however, this last-minute addition turned out to haunt us later-on. During testing, we realised that due to the high battery voltage of 4.2 V which exceeded the maximal input tolerance of the nRF52840 of 3.6 V, we partially damaged the IC and reduced the resistance between the supply voltage and ground from 300-400 kΩ to mere 30-110 Ω. While the voltage protection circuit of the controller succeeded in containing the damage of the chip and allowed us to use it almost without any perceptible impact, the low resistance between the two voltage levels resulted in a constant idle currents in the order of 40-100 mA. While this resulted in a noticeable heating of the board during testing, it also prevented the system from achieving any reasonable life time expectations.

7.2.3 Revision E

While the previous generation suffered from a constant high power draw due to a faulty voltage probing pin, its main drawback was poor radio communication range. Both the BLE radio and the DW1000 radio were only capable of reliable transmission within 5 meters, even when using maximal transmission power. Hence, our main focus for the third hardware generation was on improving transmission range through eliminating noise sources and reducing the signal attenuation on the RF traces. Among the adjustments were:

- Fixed the erroneous voltage probing pin with a voltage divider to prevent 4.2 V at an input pin with a limit of 3.6 V.
- Improved the BLE transmission range:
  - Increased the keepout zone around the antenna by extending the PCB to improve the distance from other components such as the SD card and to avoid (debugging) traces below the antenna itself.
  - Increased the metal-free zone on all four layers around the antenna to conform to the evaluation board and achieve the performance already proven by PermaMote [98].
  - Closely replicated the matching circuitry layout of the evaluation board to reduce capacitive coupling between layers which can reduce performance. Further facilitated automatic filtering of harmonic components directly through the IC by connecting a dedicated RF filtering pin.
  - Repositioned the decoupling capacitors and power routing for the Nordic nRF52840 to instantly provide charge pools upon request at all pins.
  - Decreased the traces leading to the oscillator by leveraging the extended PCB to decrease the impact of noise on the clocks.
- Improved UWB transmission distances:
  - Corrected RF trace width between the DecaWave DW1000 and its balun to match the datasheet description. While the previous revision calculated the trace width as *single-ended* 100 Ω traces, they should have been calculated as *differential* 100 Ω traces. This lead to an incorrect trace width of 4 mil compared to the correct 10 mil traces.
  - Increased the keepout zones around the antennas on all planes to reduce the influence of the metal from shielding RF transmission.
  - Repositioned the balun to avoid bends in the RF traces and add the spacing for via shielding near the filtering capacitors at the cost of closer input and output traces of the RF switch.

- Repositioned the switching regulator, used for an efficient conversion of 3.3 V to 1.8 V ,to the back to shield the RF traces from its influence. We could show during testing that the 3 MHz switching frequency strongly influenced the received signal strength by decreasing it by a factor of 3.
- Improved the noise separation of the PLLs by further spacing them out and adding separate connections to ground for each of them. Similarly, the oscillator could be moved closer to the IC and achieved shorter connections to its decoupling capacitors to improve clock stability.
- Reducing noise on the power plane by routing signals first through the decoupling capacitors before they entered the vias, thus preventing any noise from propagating throughout the power plane .
- Small adjustments and addition of decoupling capacitors to better filter out noise.

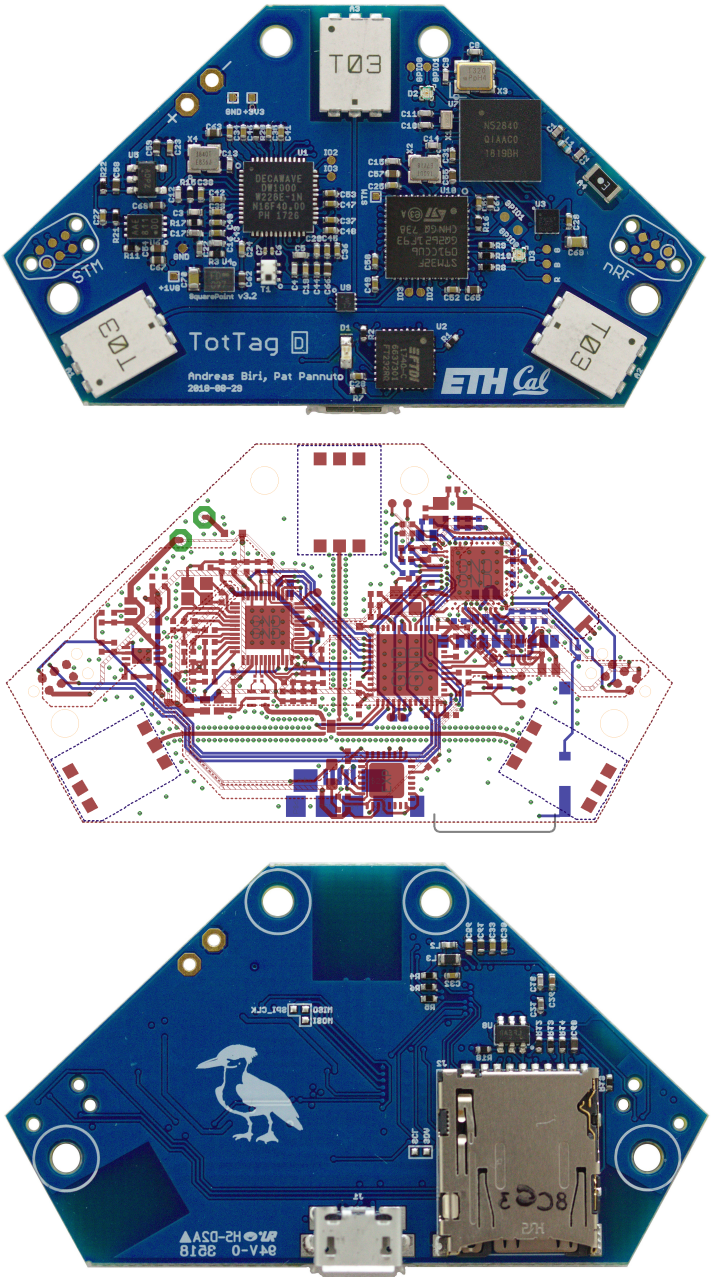

Figure 7.1: The PCB layout of the *TotTag revD* board: the red traces show the top layer of the board, while the blue ones reside on the bottom side. Compared to Figure 4.4, the BLE antenna on the upper right side of the top layer does not have sufficient spacing from other components on the opposite side of the board. Similarly, the UWB antennas only have tight keepout zones on the top layer and the switching regulator (directly above the board name) is still situated directly next to the balun, adding significant noise to the signal.

Bibliography

[1] R. Lewontin. *Biology As Ideology: The Doctrine of DNA*. CBC Massey Lectures. House of Anansi Press Incorporated, 1996.

[2] A. Migliano, A. Page, J. Gómez-Gardeñes, G. Salali, S. Viguier, M. Dyble, J. Thompson, N. Chaudhary, D. Smith, J. Strods, et al. Characterization of hunter-gatherer networks and implications for cumulative culture. *Nature Human Behaviour*, 1(2):0043, 2017.

[3] D. J. Lewkowicz. The biological implausibility of the nature–nurture dichotomy and what it means for the study of infancy. *Infancy*, 16(4):331–367, 2011.

[4] A. J. Sameroff. The science of infancy: Academic, social, and political agendas. *Infancy*, 7(3):219–242, 2005.

[5] W. Huang, Y.-S. Kuo, P. Pannuto, and P. Dutta. Opo: A wearable sensor for capturing high-fidelity face-to-face interactions. In *Proceedings of the 12th ACM Conference on Embedded Network Sensor Systems*, pages 61–75. ACM, 2014.

[6] A. Montanari, S. Nawaz, C. Mascolo, and K. Sailer. A Study of Bluetooth Low Energy performance for human proximity detection in the workplace. In *Pervasive Computing and Communications (PerCom), 2017 IEEE International Conference on*, pages 90–99. IEEE, 2017.

[7] C. Zeanah. *Handbook of Infant Mental Health, Fourth Edition*. Guilford Publications, 2018.

[8] D. A. Christakis, J. Gilkerson, J. A. Richards, F. J. Zimmerman, M. M. Garrison, D. Xu, S. Gray, and U. Yapanel. Audible television and decreased adult words, infant vocalizations, and conversational turns: A population-based study. *Archives of Pediatrics & Adolescent Medicine*, 163(6):554–558, 2009.

[9] R. Mastrandrea, J. Fournet, and A. Barrat. Contact Patterns in a High School: A Comparison between Data Collected Using Wearable Sensors, Contact Diaries and Friendship Surveys. *PLoS ONE*, 10, Sept. 2015.

[10] M. Starnini, B. Lepri, A. Baronchelli, A. Barrat, C. Cattuto, and R. Pastor-Satorras. Robust modeling of human contact networks across different scales and proximity-sensing techniques. In *International Conference on Social Informatics*, pages 536–551. Springer, 2017.

[11] D. Olguin and A. S. Pentland. Social sensors for automatic data collection. *AMCIS 2008 Proceedings*:171, 2008.

[12] B. Waber, S. Aral, D. Olguin, L. Wu, E. Brynjolfsson, and A. Pentland. Sociometric badges: A new tool for IS research, 2011.

[13] T. Sathyan, D. Humphrey, and M. Hedley. WASP: A system and algorithms for accurate radio localization using low-cost hardware. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 41(2):211–222, 2011.

[14] K. Liu, X. Liu, and X. Li. Guoguo: Enabling fine-grained indoor localization via smartphone. In *Proceeding of the 11th annual international conference on Mobile systems, applications, and services*, pages 235–248. ACM, 2013.

[15] C. Cattuto, W. Van den Broeck, A. Barrat, V. Colizza, J.-F. Pinton, and A. Vespignani. Dynamics of person-to-person interactions from distributed RFID sensor networks. *PloS one*, 5(7), 2010.

[16] N. Aharony, W. Pan, C. Ip, I. Khayal, and A. Pentland. Social fMRI: Investigating and shaping social mechanisms in the real world. *Pervasive and Mobile Computing*, 7(6):643–659, 2011.

[17] K. T. Min, A. Forys, A. Luong, E. Lee, J. Davies, and T. Schmid. WRENSys: Large-scale, rapid deployable mobile sensing system. In *2014 IEEE 39th Conference on Local Computer Networks Workshops (LCN Workshops)*, pages 557–565. IEEE, 2014.

[18] L. Ozella, F. Gesualdo, M. Tizzoni, C. Rizzo, E. Pandolfi, I. Campagna, A. E. Tozzi, and C. Cattuto. Close encounters between infants and household members measured through wearable proximity sensors. *PLOS ONE*, 13, June 2018.

[19] J. Stehlé, N. Voirin, A. Barrat, C. Cattuto, V. Colizza, L. Isella, C. Régis, J.-F. Pinton, N. Khanafer, W. Van den Broeck, et al. Simulation of an SEIR infectious disease model on the dynamic contact network of conference attendees. *BMC medicine*, 9(1):87, 2011.

[20] M. L. Walters, K. Dautenhahn, R. Te Boekhorst, K. L. Koay, C. Kaouri, S. Woods, C. Nehaniv, D. Lee, and I. Werry. The influence of subjects' personality traits on personal spatial zones in a human-robot interaction experiment. In *Robot and Human Interactive Communication, 2005. ROMAN 2005. IEEE International Workshop on*, pages 347–352. IEEE, 2005.

[21] N. Bruno and M. Muzzolini. Proxemics Revisited: Similar Effects of Arms Length on Men's and Women's Personal Distances. *Universal Journal of Psychology*, 1(2):46–52, 2013.

[22] D. Stojanovic and N. Stojanovic. Indoor localization and tracking: Methods, technologies and research challenges. *Facta Universitatis, Series: Automatic Control and Robotics*, 13(1):57–72, 2014.

[23] B. Kempke, P. Pannuto, B. Campbell, and P. Dutta. SurePoint: Exploiting ultra wideband flooding and diversity to provide robust, scalable, high-fidelity indoor localization. In *Proceedings of the 14th ACM Conference on Embedded Network Sensor Systems*, pages 137–149. ACM, 2016.

[24] A. Biri, B. Campbell, B. Kempke, and P. Pannuto. Software repository for TotTernary - Github. Oct. 20, 2018. URL: `https://github.com/lab11/totternary/`.

[25] Vanderbilt University. Disabilities and Human Development - Vanderbilt University. Nov. 9, 2018. URL: `https://vkc.mc.vanderbilt.edu/vkc/research/`.

[26] A. Chen, R. R. Muntz, S. Yuen, I. Locher, S. I. Sung, and M. B. Srivastava. A support infrastructure for the smart kindergarten. *IEEE Pervasive Computing*, 1(2):49–57, 2002.

[27] P. Lazik, N. Rajagopal, O. Shih, B. Sinopoli, and A. Rowe. ALPS: A bluetooth and ultrasound platform for mapping and localization. In *Proceedings of the 13th ACM conference on embedded networked sensor systems*, pages 73–84. ACM, 2015.

[28] J. Gonzalez, J. Blanco, C. Galindo, A. Ortiz-de-Galisteo, J. Fernández-Madrigal, F. Moreno, and J. Martinez. Combination of UWB and GPS for indoor-outdoor vehicle localization. In *Intelligent Signal Processing, 2007. WISP 2007. IEEE International Symposium on*, pages 1–6. IEEE, 2007.

[29] Z. Zhuang, K.-H. Kim, and J. P. Singh. Improving energy efficiency of location sensing on smartphones. In *Proceedings of the 8th international conference on Mobile systems, applications, and services*, pages 315–330. ACM, 2010.

[30] M. A. Kazandjieva, J. W. Lee, M. Salathé, M. W. Feldman, J. H. Jones, and P. Levis. Experiences in measuring a human contact network for epidemiology research. In *Proceedings of the 6th Workshop on Hot Topics in Embedded Networked Sensors*, page 7. ACM, 2010.

[31] D. Magazine. Designing a Happier Workplace - MIT Media Lab. 2014. URL: `http://discovermagazine.com/2014/march/11-designing-a-happy-workplace`.

[32] F. Zafari, A. Gkelias, and K. Leung. A survey of indoor localization systems and technologies. *arXiv preprint arXiv:1709.01015*, 2017.

[33] A. Alarifi, A. Al-Salman, M. Alsaleh, A. Alnafessah, S. Al-Hadhrami, M. A. Al-Ammar, and H. S. Al-Khalifa. Ultra wideband indoor positioning technologies: Analysis and recent advances. *Sensors*, 16(5):707, 2016.

[34] C. Di Franco, A. Prorok, N. Atanasov, B. Kempke, P. Dutta, V. Kumar, and G. J. Pappas. Calibration-free network localization using non-line-of-sight ultra-wideband measurements. In *Proceedings of the 16th ACM/IEEE International Conference on Information Processing in Sensor Networks*, pages 235–246. ACM, 2017.

[35] B. Kempke, P. Pannuto, and P. Dutta. PolyPoint: Guiding indoor quadrotors with ultra-wideband localization. In *Proceedings of the 2nd International Workshop on Hot Topics in Wireless*, pages 16–20. ACM, 2015.

[36] First Report and Order in the Matter of Revision of Part 15 of the Commission's Rules Regarding Ultrawideband Transmission Systems, Federal Communications Commission (FCC), 2002. URL: `https://transition.fcc.gov/Bureaus/Engineering_Technology/Orders/2002/fcc02048.pdf`.

[37] E. Karapistoli, F.-N. Pavlidou, I. Gragopoulos, and I. Tsetsinas. An overview of the IEEE 802.15.4a standard. *IEEE Communications Magazine*, 48(1), 2010.

[38] DecaWave Limited. DW1000 Datasheet — DecaWave. July 10, 2018. URL: `https://decawave.com/content/dw1000-datasheet`.

[39] K. Langendoen. Medium access control in wireless sensor networks. *Book chapter in Medium access control in wireless networks, Volume II: Practice and Standards*, 2:535–560, 2008.

[40] P. Huang, L. Xiao, S. Soltani, M. W. Mutka, and N. Xi. The evolution of MAC protocols in wireless sensor networks: A survey. *IEEE communications surveys & tutorials*, 15(1):101–120, 2013.

[41] J. Polastre, J. Hill, and D. Culler. Versatile low power media access for wireless sensor networks. In *Proceedings of the 2nd international conference on Embedded networked sensor systems*, pages 95–107. ACM, 2004.

[42] K.-J. Wong and D. Arvind. SpeckMAC: Low-power decentralised MAC protocols for low data rate transmissions in specknets. In *Proceedings of the 2nd international workshop on Multi-hop ad hoc networks: from theory to reality*, pages 71–78. ACM, 2006.

[43] A. El-Hoiydi and J.-D. Decotignie. WiseMAC: An ultra low power MAC protocol for multi-hop wireless sensor networks. In *International symposium on algorithms and experiments for sensor systems, wireless networks and distributed robotics*, pages 18–31. Springer, 2004.

[44] P. Dutta, S. Dawson-Haggerty, Y. Chen, C.-J. M. Liang, and A. Terzis. A-MAC: A versatile and efficient receiver-initiated link layer for low-power wireless. *ACM Transactions on Sensor Networks (TOSN)*, 8(4):30, 2012.

[45] W. Ye, J. Heidemann, and D. Estrin. An energy-efficient MAC protocol for wireless sensor networks. In *INFOCOM 2002. Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, volume 3, pages 1567–1576. IEEE, 2002.

[46] T. Van Dam and K. Langendoen. An adaptive energy-efficient MAC protocol for wireless sensor networks. In *Proceedings of the 1st international conference on Embedded networked sensor systems*, pages 171–180. ACM, 2003.

[47] I. Rhee, A. Warrier, M. Aia, J. Min, and M. L. Sichitiu. Z-MAC: A hybrid MAC for wireless sensor networks. *IEEE/ACM Transactions on Networking (TON)*, 16(3):511–524, 2008.

[48] V. Rajendran, K. Obraczka, and J. J. Garcia-Luna-Aceves. Energy-efficient, collision-free medium access control for wireless sensor networks. *Wireless networks*, 12(1):63–78, 2006.

[49] F. Ferrari, M. Zimmerling, L. Mottola, and L. Thiele. Low-power wireless bus. In *Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems*, pages 1–14. ACM, 2012.

[50] M. J. McGlynn and S. A. Borbash. Birthday protocols for low energy deployment and flexible neighbor discovery in ad hoc wireless networks. In *Proceedings of the 2nd ACM international symposium on Mobile ad hoc networking & computing*, pages 137–145. ACM, 2001.

[51] P. Dutta and D. Culler. Practical asynchronous neighbor discovery and rendezvous for mobile sensing applications. In *Proceedings of the 6th ACM conference on Embedded network sensor systems*, pages 71–84. ACM, 2008.

[52] Y. Qiu, S. Li, X. Xu, and Z. Li. Talk more listen less: Energy-efficient neighbor discovery in wireless sensor networks. In *INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications, IEEE*, pages 1–9. IEEE, 2016.

[53] C. Julien, C. Liu, A. L. Murphy, and G. P. Picco. BLEnd: Practical continuous neighbor discovery for Bluetooth Low Energy. In *Proceedings of the 16th ACM/IEEE International Conference on Information Processing in Sensor Networks*, pages 105–116. ACM, 2017.

[54] A. F. Molisch, K. Balakrishnan, D. Cassioli, C.-C. Chong, S. Emami, A. Fort, J. Karedal, J. Kunisch, H. Schantz, U. Schuster, et al. IEEE 802.15.4a channel model-final report. *IEEE P802*, 15(04):0662, 2004.

[55] F. Ferrari, M. Zimmerling, L. Thiele, and O. Saukh. Efficient network flooding and time synchronization with Glossy. In *Information Processing in Sensor Networks (IPSN), 2011 10th International Conference on*, pages 73–84. IEEE, 2011.

[56] K. Leentvaar and J. Flint. The capture effect in FM receivers. *IEEE Transactions on Communications*, 24(5):531–539, 1976.

[57] STMicroelectronics. "STM32F091CCU6 - STMicroelectronics". July 4, 2018. URL: https://www.st.com/content/st_com/en/products/microcontrollers/stm32-32-bit-arm-cortex-mcus/stm32-mainstream-mcus/stm32f0-series/stm32f0x1/stm32f091cc.html.

[58] Nordic Semiconductors. "nRF51822 - Ultra Low Power Wireless Solutions from Nordic Semiconductors". July 11, 2018. URL: https://www.nordicsemi.com/eng/Products/Bluetooth-low-energy/nRF51822.

[59] Time Domain. PulsON 440. July 30, 2018. URL: https://timedomain.com/products/pulson-440/.

[60] Nordic Semiconductors. nRF51822 Datasheet. July 11, 2018. URL: https://www.nordicsemi.com/eng/nordic/download_resource/62726/14/73592363/13358.

[61] J. Gonzalez, J.-L. Blanco, C. Galindo, A. Ortiz-de-Galisteo, J.-A. Fernandez-Madrigal, F. A. Moreno, and J. L. Martinez. Mobile robot localization based on ultra-wide-band ranging: A particle filter approach. *Robotics and autonomous systems*, 57(5):496–507, 2009.

[62] U.S. Department of Commerce. 2017 Characteristics of new housing. June 1, 2018. URL: https://www.census.gov/construction/chars/.

[63] M. Ford, C. T. Baer, D. Xu, U. Yapanel, and S. Gray. The LENATM language environment analysis system: Audio specifications of the DLP-0121. *Boulder, CO: Lena Foundation*, 2008.

[64] P. Dutta, J. Taneja, J. Jeong, X. Jiang, and D. Culler. A building block approach to sensornet systems. In *Proceedings of the 6th ACM conference on Embedded network sensor systems*, pages 267–280. ACM, 2008.

[65] F. Sutton, M. Zimmerling, R. Da Forno, R. Lim, T. Gsell, G. Giannopoulou, F. Ferrari, J. Beutel, and L. Thiele. Bolt: A stateful processor interconnect. In *Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems*, pages 267–280. ACM, 2015.

[66] B. Kempke. TriPoint reference layout - Github. July 20, 2018. URL: https://github.com/lab11/totternary/tree/master/hardware/tripoint.

[67] Autodesk Inc. "EAGLE - PCB Design & Schematic Software". July 20, 2018. URL: https://www.autodesk.com/products/eagle/overview.

[68] C. J. Merlin and W. B. Heinzelman. Schedule adaptation of low-power-listening protocols for wireless sensor networks. *IEEE Transactions on Mobile Computing*, 9(5):672–685, 2010.

[69] Nordic Semiconductors. "nRF52840 - Ultra Low Power Wireless Solutions from Nordic Semiconductors". July 11, 2018. URL: https://www.nordicsemi.com/eng/Products/nRF52840.

[70] I. Demirkol, C. Ersoy, F. Alagoz, et al. MAC protocols for wireless sensor networks: a survey. *IEEE Communications Magazine*, 44(4):115–121, 2006.

[71] Link Labs. ZigBee Vs. Bluetooth: A Use Case with Range Calculations. Aug. 1, 2018. URL: `https://www.link-labs.com/blog/zigbee-vs-bluetooth`.

[72] C. Guo, L. C. Zhong, and J. M. Rabaey. Low power distributed MAC for ad hoc sensor radio networks. In *Global Telecommunications Conference, 2001. GLOBECOM'01. IEEE*, volume 5, pages 2944–2948. IEEE, 2001.

[73] A. Bachir, M. Dohler, T. Watteyne, and K. K. Leung. MAC essentials for wireless sensor networks. *IEEE Communications Surveys & Tutorials*, 12(2):222–248, 2010.

[74] J. H. Schiller. *Mobile communications*. In Pearson education, 2003. Part 3, pages 69–92.

[75] G. Adamiuk, T. Zwick, and W. Wiesbeck. Compact, dual-polarized UWB-antenna, embedded in a dielectric. *IEEE Transactions on Antennas and Propagation*, 58(2):279–286, 2010.

[76] L. Mottola and G. P. Picco. MUSTER: Adaptive energy-aware multisink routing in wireless sensor networks. *IEEE Transactions on Mobile Computing*, 10(12):1694–1709, 2011.

[77] J. Beutel. Lecture notes for Low-power System Design, 2017.

[78] E.-Y. Lin, J. M. Rabaey, and A. Wolisz. Power-efficient rendez-vous schemes for dense wireless sensor networks. In *Communications, 2004 IEEE International Conference on*, volume 7, pages 3769–3776. IEEE, 2004.

[79] O. Landsiedel, F. Ferrari, and M. Zimmerling. Chaos: Versatile and efficient all-to-all data sharing and in-network processing at scale. In *Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems*, page 1. ACM, 2013.

[80] F. Mager, C. Herrmann, and M. Zimmerling. One for All, All for One: Toward Efficient Many-to-Many Broadcast in Dynamic Wireless Networks. In *Proceedings of the 4th ACM Workshop on Hot Topics in Wireless*, pages 19–23. ACM, 2017.

[81] Bluetooth Special Interest Group (SIG). Core Specifications - Bluetooth Technology. 2014. URL: `https://www.bluetooth.com/specifications/bluetooth-core-specification`.

[82] S. Maymon and A. V. Oppenheim. Sinc interpolation of nonuniform samples. *IEEE Transactions on Signal Processing*, 59(10):4745–4758, 2011.

[83] T. Zachariah. Summon - Browser for the Local Web of Things. Dec. 5, 2018. URL: `https://github.com/lab11/summon`.

[84] DecaWave Limited. Application note 14: Antenna delay calibration of DW1000-based products and systems. Oct. 30, 2018. URL: `https://www.decawave.com/application-notes/`.

[85] A. Biri and P. Pannuto. TotTag reference layout - Github. Oct. 20, 2018. URL: `https://github.com/lab11/totternary/tree/master/hardware/tottag/rev_e`.

[86] STMicroelectronics. STM32F091CCU6 Datasheet. July 4, 2018. URL: `https://www.st.com/resource/en/datasheet/stm32f091cc.pdf`.

[87] B. Campbell. TriTag reference layout - Github. July 20, 2018. URL: `https://github.com/lab11/totternary/tree/master/hardware/tritag`.

[88] Future Technology Devices International Ltd. FT232R. July 20, 2018. URL: `http://www.ftdichip.com/Support/Documents/DataSheets/ICs/DS_FT232R.pdf`.

[89] Microchip Technology. "MCP73831 - Battery Charger". July 6, 2018. URL: `https://www.microchip.com/wwwproducts/en/MCP73831`.

[90] P. Narendra, S. Duquennoy, and T. Voigt. BLE and IEEE 802.15.4 in the IoT: Evaluation and Interoperability Considerations. In *International Internet of Things Summit*, pages 427–438. Springer, 2015.

[91] Adafruit. Litioum Ion Polymer Battery - 3.7v 350mAh. Dec. 19, 2018. URL: `https://www.adafruit.com/product/2750`.

[92] B. Großwindhager, C. A. Boano, M. Rath, and K. Römer. Enabling Runtime Adaptation of Physical Layer Settings for Dependable UWB Communications. *2018 IEEE 19th International Symposium on "A World of Wireless, Mobile and Multimedia Networks" (WoWMoM)*:01–11, 2018.

[93] S. Marano, W. M. Gifford, H. Wymeersch, and M. Z. Win. NLOS identification and mitigation for localization based on UWB experimental data. *IEEE Journal on Selected Areas in Communications*, 28(7), 2010.

[94] R. Want, W. Wang, and S. Chesnutt. Accurate Indoor Location for the IoT. *Computer*, 51(8):66–70, Aug. 2018.

[95] M. Ibrahim, H. Liu, M. Jawahar, V. Nguyen, M. Gruteser, R. Howard, B. Yu, and F. Bai. Verification: Accuracy Evaluation of WiFi Fine Time Measurements on an Open Platform. In *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*, pages 417–427. ACM, 2018.

[96] M. Zimmerling, L. Mottola, P. Kumar, F. Ferrari, and L. Thiele. Adaptive real-time communication for wireless cyber-physical systems. *ACM Transactions on Cyber-Physical Systems*, 1(2):8, 2017.

[97] Nordic Semiconductors. nRF52840 Datasheet. July 11, 2018. URL: `http://infocenter.nordicsemi.com/pdf/nRF52840_PS_v1.0.pdf`.

[98] N. Jackson. PermaMote - New generation 10-year lifetime sensor node. Oct. 29, 2018. URL: `https://github.com/lab11/permamote`.